International Journal of Forecasting 39 (2023) 364–390

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Does the Phillips curve help to forecast euro area inflation?☆

Marta Bańbura, Elena Bobeica *

European Central Bank, Sonnemannstrasse 20, Frankfurt am Main, Germany

ARTICLE INFO

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

ABSTRACT

We find that it does, but choosing the right specification is not trivial. Based on an extensive forecast evaluation we document notable forecast instabilities for most simple Phillips curves. Euro area inflation was particularly hard to forecast in the run-up to the Economic and Monetary Union and after the sovereign debt crisis, when the trends—and, for the latter period, also the amount of slack—were harder to pin down. Yet, some specifications outperform a univariate benchmark and point to the following lessons: (i) the key type of time variation to consider is an inflation trend; (ii) a simple filter-based output gap works well, but after the Great Recession it is outperformed by endogenously estimated slack or by "institutional" estimates; (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; and (v) averaging over a wide range of modelling choices helps.

1. Introduction

It would be extraordinarily useful to discover a specification of the Phillips curve that fits the data reliably.... [A]s 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 & Mazumder, 2019b]

The Phillips curve has long been 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 whether 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 FED, where the validity of the Phillips curve has taken centre 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. This so-called "missing disinflation" puzzle was first discussed in the US case (see e.g. Ball & Mazumder, 2011; Blanchard, 2016; Hooper et al., 2019, and the references therein for a more detailed discussion) and soon found to characterise other advanced economies as well (Friedrich, 2014; IMF, 2013). Afterwards, in spite of the economic recovery, inflation

☆ 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.

* Corresponding author.
E-mail addresses: Marta.Banbura@ecb.int (M. Bańbura), Elena.Bobeica@ecb.int (E. Bobeica).

[1] Constancio (2015), Lane (2019), Powell (2020), and Yellen (2015) are examples of policymakers extensively referencing the Phillips curve framework.

https://doi.org/10.1016/j.ijforecast.2021.12.001
0169-2070/

appeared to be surprisingly low, and more so in the euro area compared to the US (Bobeica & Jarociński, 2019; Constancio, 2015).

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

In this paper we assess whether the Phillips curve is still alive through its usefulness in 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 in whether and when the Phillips curve can produce more accurate forecasts than those from the 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. Our contribution in this paper is to run a systematic comparison of a large number of specifications and investigate how various elements affect the forecast performance in the euro area. We start by taking a simple workhorse model and evaluate whether it can be improved by: (i) accounting for a time-varying inflation 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 important, as previous studies for the US economy have shown that the performance of Phillips curves tends to be episodic (see e.g. Fisher et al., 2002; Granziera & Sekhposyan, 2019; Stock & 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 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.

Admittedly, the empirical literature dealing with forecasting with Phillips curves is large (see Stock & Watson, 2009, for a comprehensive survey), but most of the lessons are derived for the US case. Our paper is linked to the seminal contributions of Stock and Watson (1999) and Stock and Watson (2009). In contrast to these papers, we also assess the forecasting performance of a set of newer-generation Phillips curve models that have not been systematically evaluated. In addition, we cover the post-Great Recession period, considering lessons derived from studying the puzzling behaviour of inflation during this time. Finally, we focus on the euro area, for which evidence is much scarcer, albeit growing. 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 that 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.

The available forecasting studies for the euro area often have a different focus, are based on a shorter evaluation sample, or consider a narrower set of model specifications. Many papers rely on in-sample analysis, trying to fit a Phillips curve model to euro area data, without the specific forecasting angle (see e.g. Ball & Mazumder, 2020; Doepke et al., 2008; Eser et al., 2020; Galí et al., 2001; Musso et al., 2009; Paloviita, 2008). Similar to our study, Moretti et al. (2019) explore a large space of euro area Phillips curves by employing numerous proxies for the inflation drivers to account for model uncertainty; yet, they do not perform a forecast evaluation for euro area inflation. Kulikov and Reigl (2019) and Álvarez and Correa-López (2020) study the usefulness of various measures of inflation expectations in the Phillips curve framework in a *conditional* forecast setup. There are, in general, fewer papers studying the out-of-sample forecasting performance. Hubrich (2005) employs Phillips curve models to study whether forecast gains are achieved by aggregating forecasts of subindices of the Harmonised Index of Consumer Prices (HICP), as opposed to a top-down approach. Buelens (2012), Jarociński and Lenza (2018), Marcellino and Musso (2010) focus on estimating the output gap for the euro area and delve into the issue of its predictive power for inflation. Similar to our paper, Bereau et al. (2018) perform a forecast evaluation for euro area inflation using Phillips curve models, but they focus on the role of global variables. Compared to these studies, the range of Phillips curve models that we cover is much wider. We also look more in depth at the issue of time variation than all the papers mentioned. 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 does not when

2 Some authors also propose alternative measures of inflation, such as Ball and Mazumder (2019a). We do not pursue this idea here.

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. Also, we draw attention to the fact that the forecasting performance tends to vary over time. Yet, some Phillips curve specifications can offer systematic improvements over a univariate model; in other words, measures of economic activity do help to forecast inflation. Unlike Stock and Watson (2010), for the euro area we do not find a link between the forecasting performance of the Phillips curve and certain states of the economy, namely expansion versus recession. Instead, euro area inflation appeared to be 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 the former period, inflation trended downwards and most models, being estimated on historical "high inflation" samples, resulted in overpredictions.[3] The "low inflation period" after the sovereign debt crisis was also particularly challenging for forecasters in terms of capturing both the inflation trend and the right measure of economic slack.

Our conclusions and lessons for practitioners can be summarised as follows.

First, accounting for a trend or time-varying mean of inflation leads to sizable improvements in forecast accuracy in certain periods. While this is in line with findings for the US (see e.g. Clark & Doh, 2014; Clark & McCracken, 2010a; Faust & Wright, 2013), we show that these gains are also time-varying. For the euro area, inflation detrending helps particularly in the periods marked by more pronounced changes in the underlying trend—namely, in the early part of the sample corresponding to the inflation convergence process, and, more recently, during the low inflation period. 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). 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.[4] The accelerationist version works well over an early part of the sample, reminding of the lessons drawn for the US (see the discussion in Ball & Mazumder, 2019b), but its performance deteriorates after the Great Recession; interestingly, it improves somewhat again during the latest low inflation period, suggesting that the persistence of inflation might have increased.

Second, the choice of slack matters. We employ a wide set of product and labour market slack indicators, some of which have been constructed and backcasted 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 a simple filter-based output gap performs well over the entire period. On the labour market it appears harder to single out one indicator that is superior over the whole sample. In particular, 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 & 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.

Third, 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 considerable explanatory power for domestic inflation in-sample (see e.g. Bobeica & Jarociński, 2019, for evidence for the euro area).

Fourth, new-generation Phillips curve models (such as the one proposed by Chan et al., 2018, 2016) 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.[5]

Finally, 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 and brings some forecast gains.

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, a 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. We also conduct a real-time evaluation of the new-generation Phillips curve models, as they appear to be a promising avenue for conducting forecasts. We find that the real-time forecast evaluation does not change the picture drawn based on pseudo-real-time data, with some models still offering consistent forecast gains with respect to univariate benchmarks.

3 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.

4 Bańbura and van Vlodrop (2018) reach a similar conclusion using different models.

5 Canova (2007) shows that despite the evident structural changes in the process of inflation, the contribution of time variations in the coefficients to the forecasting performance of multivariate models is limited.

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:

$$\pi_{t+h}^h = \frac{1}{h}\sum_{i=1}^{h}\pi_{t+i} = \frac{400}{h}\ln\left(\frac{P_{t+h}}{P_t}\right). \quad (2.1)$$

P_t is the seasonally and working day-adjusted euro area Harmonised Index of Consumer Prices (HICP) excluding energy. We disregard the energy component, as central banks are more concerned with persistent sources of inflationary pressure and less so with volatile components in the inflation rate, linked for instance to oil price movements.[6]

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 the unemployment rate, where we keep the cycles shorter than 15 years.[7] Other slack measures, including composite indicators, are considered in Section 3.2. In particular, 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 more detail, 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 the online appendix to this paper.

6 Ehrmann et al. (2018) note that the ECB monitors a wide range of underlying inflation measures, including several exclusion-based indicators. 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.

7 The Christiano–Fitzgerald filter is a nearly optimal one-sided band-pass filter; see Christiano and Fitzgerald (2003). Robustness checks were performed with the Hodrick–Prescott filter and by extending the underlying series based on an AR process prior to filtering.

All the data are quarterly and seasonally adjusted, and they cover the sample from 1980 to 2018. Details regarding the data are provided in Appendix A.

We focus on a one-year-ahead forecast horizon ($h = 4$).[8] Hence, the target variable is the annual inflation rate. 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 score (LPS) for the new-generation models. The evaluation period is from 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 naive random walk (RW) of Atkeson and Ohanian (2001) and the unobserved components stochastic volatility model (UCSV) of Stock and Watson (2007).

Let $\pi_{t+h|t}$ ($\pi_{t+h|t}^h$) denote the h-step-ahead forecast given the information at time t. For the RW benchmark, the forecast is set to the latest observed annual inflation rate, as below:

$$\pi_{t+h|t}^h = \pi_t^4. \quad (2.2)$$

The UCSV model is a "non-centred" version of the one in Stock and Watson (2007); see Chan (2018). The model decomposes each variable into a trend and a transitory component, where each component features an independent stochastic volatility:

$$\pi_t = \tau_t + e^{\frac{1}{2}(h_0+\omega_h\tilde{h}_t)}\varepsilon_t^{\pi}, \qquad \varepsilon_t^{y} \sim N(0,1), \quad (2.3)$$

$$\tau_t = \tau_{t-1} + e^{\frac{1}{2}(g_0+\omega_g\tilde{g}_t)}\varepsilon_t^{\tau}, \qquad \varepsilon_t^{\tau} \sim N(0,1), \quad (2.4)$$

$$\tilde{h}_t = \tilde{h}_{t-1} + \varepsilon_t^{h}, \qquad \varepsilon_t^{h} \sim N(0,1), \quad (2.5)$$

$$\tilde{g}_t = \tilde{g}_{t-1} + \varepsilon_t^{g}, \qquad \varepsilon_t^{g} \sim N(0,1). \quad (2.6)$$

To produce the forecast, we simulate π_t forward to obtain $\pi_{t+1|t}, \ldots, \pi_{t+h|t}$ and take the median of the draws of $\pi_{t+h|t}^h = \frac{1}{h}\sum_{i=1}^{h}\pi_{t+i|t}$ as the point forecast.

2.2. Forecasting performance

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

$$\tilde{\pi}_{t+1} = \alpha\tilde{\pi}_t + \beta y_{t+1} + \nu_{t+1} \quad (2.7)$$

where α and β are the coefficients, y_t is a proxy for the real marginal cost (economic slack), and $\tilde{\pi}_t = \pi_t - \mu_\pi$ denotes the de-meaned inflation rate.[9] μ_π can be interpreted as an inflation trend that is constant over the estimation sample. This assumption is relaxed in Section 3.1, where various forms for the trend process are evaluated.

8 Similar messages also apply for longer horizons. The results for $h = 8$ are available upon request.

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

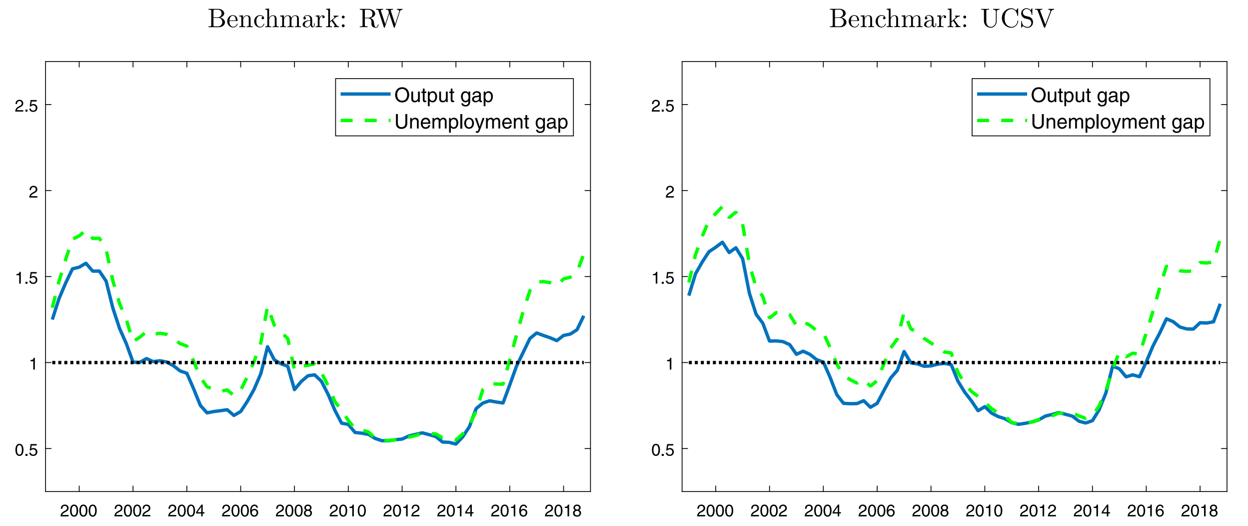

Fig. 2.1. Forecasting performance of the workhorse Phillips curve models. Note: RMSFEs over a rolling window of 20 quarters relative to the corresponding RMSFEs of the benchmark models. Forecast horizon: 4 quarters ahead.

The estimation is performed with a rolling window of 60 observations using the OLS.[10]

The forecasts h steps ahead are obtained in an iterative fashion. The explanatory variables are first forecasted with an AR(4) process.[11] Then we iteratively obtain the forecasts $\tilde{\pi}_{t+i|t}$, $i = 1, \ldots, h$, and set the forecast for the target variable to $\pi^h_{t+h|t} = \frac{1}{h}\sum_{i=1}^{h} \tilde{\pi}_{t+i|t} + \mu_\pi$ (the mean (or trend) is added back to the forecast).

Fig. 2.1 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 labour market, the unemployment gap, relative to the two benchmarks, the RW and the UCSV.

The first result that stands out in Fig. 2.1 is that the relative forecasting performance of the Phillips 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), and 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.[12] Over the entire sample, the workhorse model does not outperform a univariate benchmark in a statistically significant way (see Table B.1), and 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 when the model performs worse than the benchmark (see Table B.1 and Fig. B.2 in Appendix B). 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 that are very similar, but just marginally better than that of the random walk (see the reported RMSFEs in Table B.1). Thus, in what follows, we use only the UCSV as the benchmark.

In order to get further insights into the forecast performance, Fig. 2.2 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 introduction of the euro. At that time, the inflation rates in many euro area countries converged to lower levels, and inflation expectations were anchored by 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 strong improvements in the euro area labour market in the recent years (at least when looking solely at the headline unemployment

10 Bayesian estimation techniques, as applied in Section 3.4, yield similar results. These are available upon request.

11 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.

12 The online appendix includes results based on a narrower rolling window and shows, e.g., a marked worsening in relative forecast performance during the sovereign debt crisis.

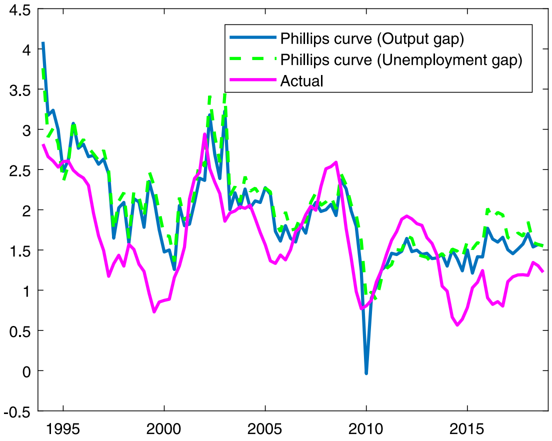

Fig. 2.2. Actual inflation versus Phillips curve forecasts. Note: Forecast horizon: 4 quarters ahead.

rate indicator), which makes measures based on statistical filters (as those applied here) point to rather buoyant labour market conditions at a time when inflation was stubbornly low.

There has been intense discussion surrounding the two puzzling inflation episodes in the aftermath of the Great Recession, namely the missing disinflation and subsequent missing inflation (see Bobeica & 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) raises 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 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 the 60 used above) and recursive estimation, with the estimation sample starting in 1980.[13]
2. Number of lags included and lag selection method: we allow up to four lags of inflation and of slack in Eq. (2.7), and at each step of the evaluation we choose the number by either the Akaike or Bayesian information criterion.
3. "Functional" form of the model: in addition to the one given by Eq. (2.7), we consider the following specifications:
   - *ADL model with "direct forecast" formulation*:
   $$\tilde{\pi}^h_{t+h} = \alpha(L)\tilde{\pi}_t + \beta(L)y_t + \nu^h_{t+h} \qquad (2.8)$$
   where $\alpha(L)$ and $\beta(L)$ are lag polynomials, and h is the forecast horizon, i.e. four quarters. Such "direct forecast" specifications have often been employed in forecasting applications, following Stock and Watson (1999) and Stock and Watson (2009).[14]
   - *VAR model*:
   $$X_{t+1} = \Phi(L)X_t + \nu_{t+1}\,, \qquad X_t = [\tilde{\pi}_t\ y_t]' \quad (2.9)$$
   "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 Eq. (2.7). 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. Fig. 2.3 presents the relative performance of the average point forecasts (across the specifications for each slack measure). The averaging is done using equal weights for simplicity and based on a proven track record for such an approach (see e.g. Clark & McCracken, 2010b). The messages remain qualitatively the same compared to those from the workhorse model in Fig. 2.1. It appears, nevertheless, that forecast averaging can provide some accuracy gains for the entire analysed period (see Table B.1 in Appendix B), but more so 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 Fig. B.3, which compares accuracy along one "dimension" of forecasts averaged along the remaining specification choices.[15]

3. Can one improve on the workhorse Phillips curve model?

The previous section documented the weak forecast performance of the simple Phillips curve model in some

13 Rossi and Inoue (2012) discuss 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.

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

15 Another issue is how robust the results are to choosing another univariate benchmark. The online appendix reports the RMSFEs of the Phillips curves relative to their univariate counterparts, i.e. models specified by a version of Eq. (2.7) 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 from including an unemployment gap are not visible all the time, with a deterioration especially towards the end of the sample, as discussed above.

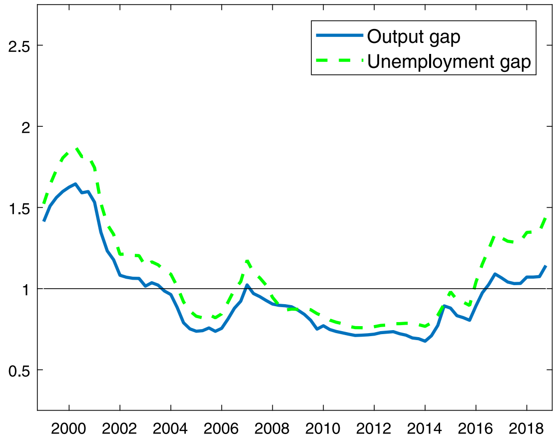

Fig. 2.3. Forecasting performance of the average over different specifications. Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows, and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. Forecast horizon: 4 quarters ahead.

periods and hinted at possible underlying reasons for this, such as changing inflation trends 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, and further features such as those 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 the average forecast across all the specifications considered in that section (different estimation windows, lag selection approaches, and functional forms), unless indicated otherwise.[16]

3.1. Inflation trend

The Phillips curve models presented so far assume a constant (unconditional) mean in inflation over the estimation sample. Yet many papers find forecast gains when assuming a time-varying inflation trend (see e.g. Bańbura & van Vlodrop, 2018; Clark & Doh, 2014; Clark & McCracken, 2008; Faust & Wright, 2013; Zaman, 2013). 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 rely on long-term survey expectations of inflation as proxies. Specifically, 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).[17] This is the *accelerationist* Phillips curve (Friedman, 1968), where gaps are related to the *changes* as opposed to the *level* of inflation.
- *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 a weighted average of past inflation rates with a smoothing parameter ϕ. Clark and McCracken (2008, 2010a) 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.[18]
- *Long-term survey inflation expectations*: $\pi_t^{TR} = \pi_t^{Cons} - (\mu_{Cons} - \mu_{\pi})$
  π_t^{Cons} is the forecast of average inflation six 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 that inflation excluding energy is systematically lower over the sample considered (bias correction).[19]
  Long-term survey-based inflation expectations have often been used to proxy trend inflation (see e.g. Faust & 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

16 The results remain qualitatively the same when we consider the specifications as in the workhorse model.

17 In order to have a specification as in Stock and Watson (1999, 2009) we also include an intercept in Eqs. (2.7)–(2.9). We also consider versions without an intercept, following, the "triangle" model of Gordon (1982). The forecast performance of both versions is compared in the online appendix. The latter version performs somewhat better than the former. However, relative to other measures of trend, the results are qualitatively similar.

18 Clark and McCracken (2008, 2010a) use a smoothing parameter of 0.05 or 0.07. We 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 & Watson, 2007). These choices did not result in consistently better forecast performance. In particular, the performance of the IMA(1,1) model was 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.

19 The bias-corrected version of the specification yields better results over the recent sample.

found improvements in the forecast performance by modelling the inflation gap (as opposed to inflation itself) as a deviation from a survey-based measure of inflation expectations (see e.g. Clark & Doh, 2014; Faust & Wright, 2013; Zaman, 2013). Our formulation is similar to that recently used by Hooper et al. (2019).[20] 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 it is 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 Fig. B.1 in Appendix B.[21] Notably, survey expectations declined to somewhat lower levels 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 & 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, the previous-period inflation rate is a very noisy measure of 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.

Fig. 3.1 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 1994).

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 (and in a statistically significant manner; see Table B.1). 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 towards the end of the sample. This is related to the differences in the behaviour of the two trend estimates in the period of low inflation discussed above.

[20] These authors include the long-term inflation expectations as a regressor, but 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 that they 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.

[21] 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.

An interesting result refers to the performance of the accelerationist Phillips curve models. They perform the best in the 1990s and early 2000s. A similar story was found for the US, where the accelerationist version appears to fit the data starting in the 1960s, but fails to characterise inflation in the last couple of decades. Ball and Mazumder (2019b) 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 the case of the euro area, the accelerationist models witnessed some improvement in their forecasting power towards the end of the sample. This might suggest that inflation persistence increased during the low inflation period in the euro area, which is in line with the findings of Ciccarelli and Osbat (2017).

Apart from the benchmarks discussed so far, we compared the performance of the non-detrended and detrended versions of the Phillips curve models to their univariate counterparts (without a slack measure). In the case of detrended versions, adding the output gap again helps in improving the forecast performance. In addition, we compared the performance of the Phllips curve models to a "pure trend forecast" and found 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 (see the online appendix for these results).

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 & 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 & Lenza, 2018). In addition, the challenge in reconciling nominal developments with traditional indicators of economic or labour market slack made researchers believe that perhaps alternative measures were needed. In this section, we evaluate a wide range of measures in terms of their 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, and (vi) economic sentiment.

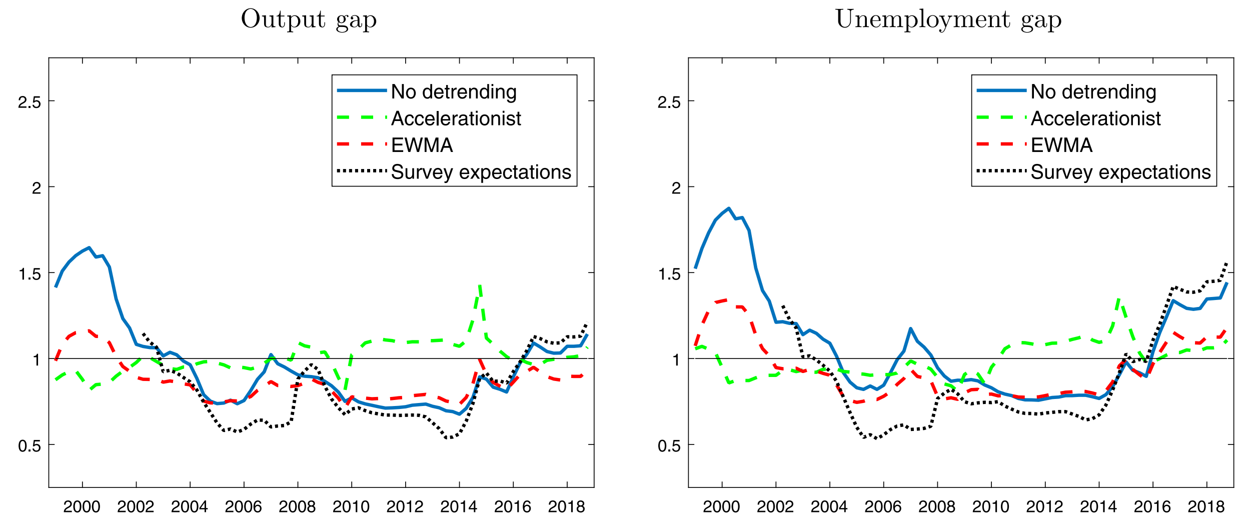

Fig. 3.1. Impact of inflation detrending on the forecasting performance. Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows, and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. Forecast horizon: 4 quarters ahead.

- *Labour market slack measures*
  We consider conventional indicators, as well as measures that were proposed more recently in the literature to better grasp the degree of labour underutilisation: (i) unemployment gap, (ii) unemployment rate, (iii) the unemployment recession gap of Stock and Watson (2010),[22] (iv) employment growth, (v) employment gap, (vi) short-term unemployment rate, (vii) broad unemployment rate (U6),[23] and (viii) U6 unemployment gap.
  The *short-term unemployment rate* was brought forward in the US case as an explanation for the "missing disinflation" episode, when inflation did not fall as much as the headline unemployment rate would have predicted (see Gordon, 2013 or Ball & Mazumder, 2019b). 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 in Bell & Blanchflower, 2018 for Europe and the US). These two indicators have not previously been considered in longer-term evaluations for the euro area due to their short history. For the purpose of the analysis here, they have been backcasted (see details in Appendix A and in the online appendix).
- *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 was 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 accounts for: the growth rate of real credit to households, growth rate of real credit to non-financial corporations (NFCs), growth rate of a real broad monetary aggregate (M3), growth rate of real residential property prices, and the spread between the short- and long-term risk-free interest rate.

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

Fig. 3.2 shows the forecasting performance of Phillips curve models with different slack measures (the average across the specifications listed in Section 2.3) for model versions without and with EWMA detrending. Detailed information is provided in Fig. B.4 in Appendix B. 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,

22 The unemployment recession gap is the difference between the current unemployment rate and the minimum unemployment rate over the current and previous 11 quarters.

23 The U6 measure adds three further categories to the unemployed: those who are available but not seeking work, those who seek work but are not available, and underemployed part-time workers.

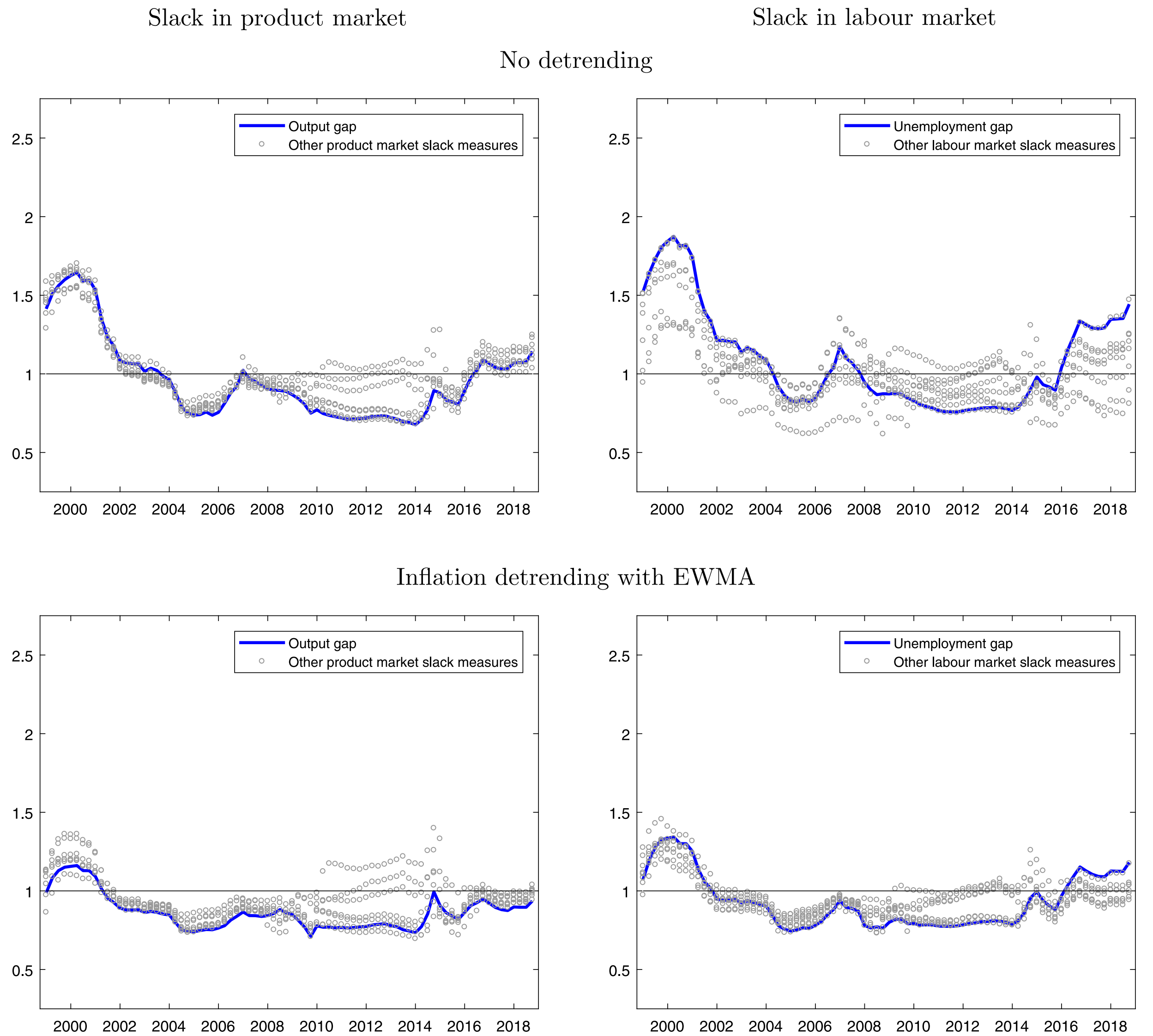

Fig. 3.2. Impact of slack measure choice on the forecasting performance. Note: For each measure of slack, the RMSFEs correspond to forecasts averaged over all considered functional forms, estimation windows, and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. Circles represent the RMSFEs of models with other measures of slack from a given group. The principal component is included in both panels; the output gap with financial variables is included in the panel with slack in the product market. Forecast horizon: 4 quarters ahead.

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 others. The output gap fares well compared to alternatives. On the labour market, it is harder to pin down an indicator that 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 the labour market in the aftermath of the Great Recession not captured by this conventional indicator. These results talk in favour 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 of estimating 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.

We exploit a novel dataset 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 with the fall 2001 OECD Economic Outlook.
- *IMF:*
  Output gap and unemployment gap: vintages starting with the October 2004 World Economic Outlook.
- *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 twice a year. We interpolate the series to quarterly frequency using the Stram–Wei method (Stram & Wei, 1986), and in order to produce forecasts at quarterly intervals, we assume that in between the updates, the estimates of the gaps are unchanged (the online appendix shows what 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 the IMF, and in March by the Eurosystem/ECB incorporate actual data up to the fourth quarter of the previous year.[26]

Fig. 3.3 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 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 Fig. B.5 in Appendix B.[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 Fig. 3.3 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 to that reported in Fig. 3.3 but based on real-time data for all the variables. For this purpose, we construct real-time vintages for inflation, real GDP, and the unemployment rate starting in mid-2005, using the information stored in the ECB's Statistical Data Warehouse (SDW); see Table A.1 in the online appendix for more details on the real-

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

[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 1991, and in the case of the EC and the Eurosystem/ECB, the series tend to start in 1995.

[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, the April 2010 forecast for the IMF, and the 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 a 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.

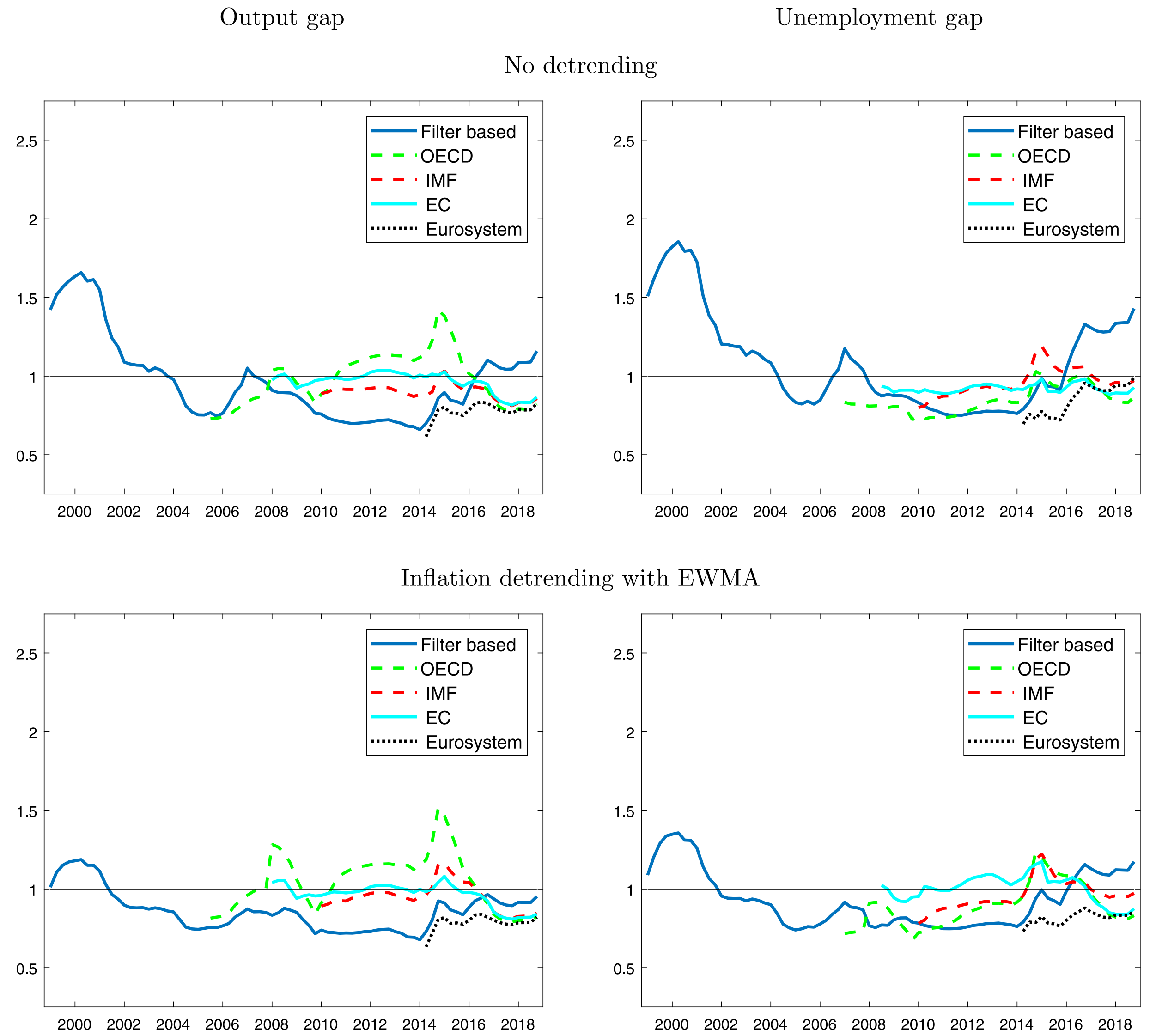

Fig. 3.3. Forecast performance with slack estimates from international institutions. Note: RMSFEs of forecasts averaged over all considered functional forms, estimation windows, and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. Forecast horizon: 4 quarters ahead.

time data.[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 forecasts by the international institutions.[29] Fig. B.6 in Appendix B 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".

3.3. Cost-push shocks and external developments

Ever since the development of the "triangle model" of inflation (Gordon, 1982) cost-push shocks, particularly coming from non-domestic drivers, have been considered 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

[28] Earlier vintages were not 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 are not available from the SDW.

[29] More precisely, we apply the following stylised "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 the 25th of January of that year, whereas in the case of the March projections by the Eurosystem/ECB, the cut-off is the 25th of February.

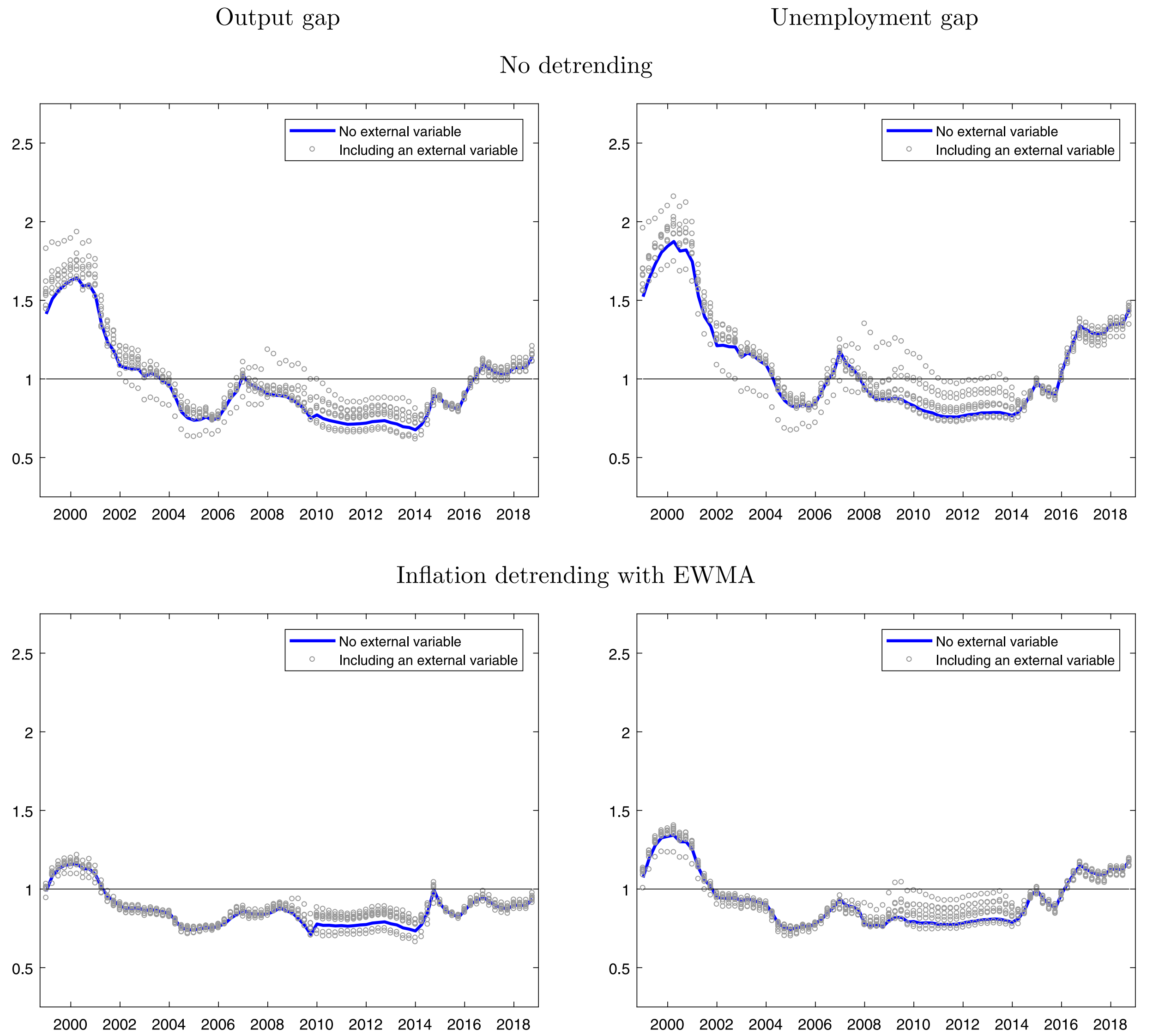

Fig. 3.4. Impact of considering external developments on the forecasting performance. Note: Average RMSFEs over all considered functional forms, estimation windows, and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. Circles represent the RMSFEs of models that include an external variable. Forecast horizon: 4 quarters ahead.

so for total inflation than for a core index stripped out of volatile items (see e.g. Bobeica & Jarociński, 2019; Forbes, 2019; Forbes et al., 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 external and externally driven variables:

- *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 (iv) the US CPI.
- *Commodity prices:*
  (i) the oil price in euros, and (ii) a 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 annualised quarter-on-quarter growth rates. We augment Eq. (2.7) and the versions presented in Section 2.3 with one external variable at a time.

30 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.

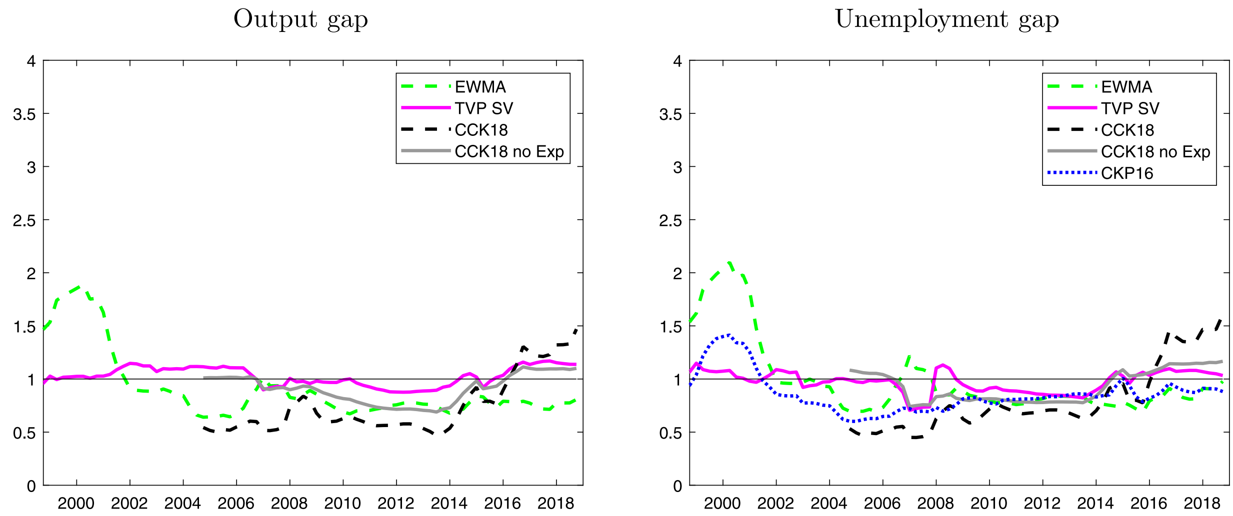

Fig. 3.5. Forecasting performance of new-generation Phillips curve models. Note: The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. All models are estimated via Bayesian methods, and RMSFEs are based on the median inflation forecast. Forecast horizon: 4 quarters ahead.

Fig. 3.4 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 (the 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 the non-detrended version by including US inflation (as shown in the online 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 the average forecast across the slack measures considered in Section 3.2. We thus confirm 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.

3.4. New-generation Phillips curve models

In the specifications considered so far, the only type of time variation allowed 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 those regarding 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. Ciccarelli & Osbat, 2017; Constancio, 2015).[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)*:

$$\pi_{t+1} = c_{t+1} + \alpha_{t+1}\pi_t + \beta_{t+1}y_{t+1} + \nu_{t+1} \tag{3.1}$$

  The Phillips curve coefficients and log volatility of the residuals follow random walks. The 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), and 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 this type of time variation helps when forecasting inflation.
- *A Phillips curve model à la (Chan et al., 2016)* (henceforth, *CKP16*):

$$(\pi_{t+1} - \pi^{TR}_{t+1}) = \alpha_{t+1}(\pi_t - \pi^{TR}_t) + \beta_{t+1}(u_{t+1} - u^{\star}_{t+1}) + \nu_{t+1} \tag{3.2}$$

[31] Earlier evidence on time variation in the Phillips curve 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 employing a smooth transition model. We do not consider this approach here.

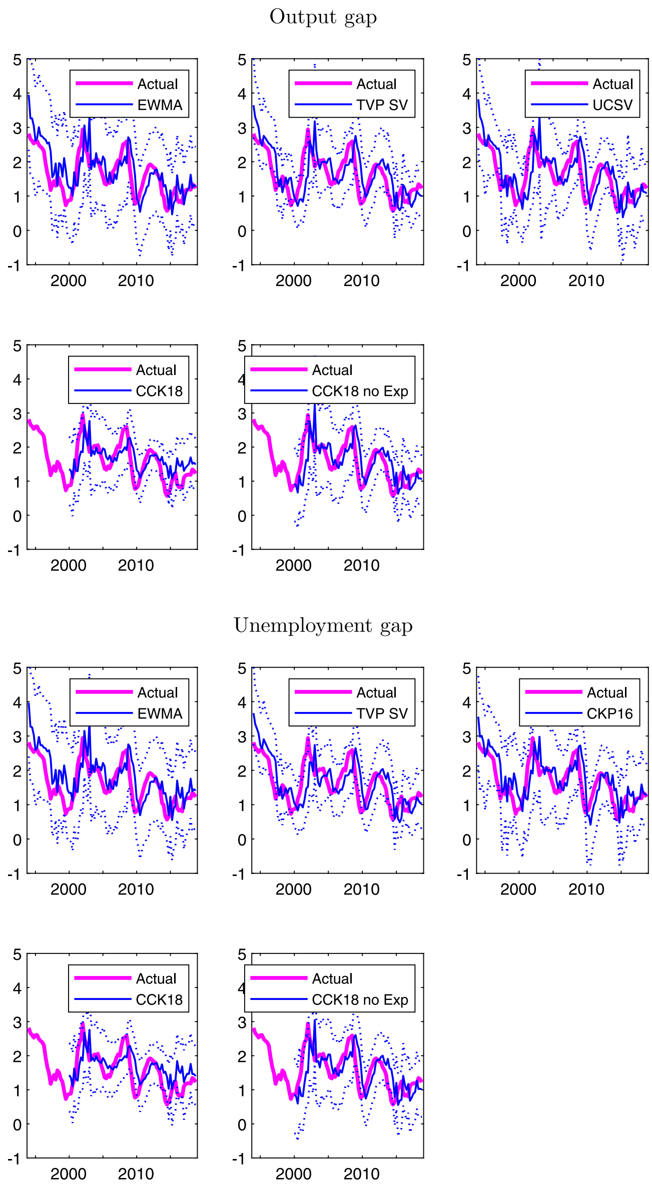

Fig. 3.6. Actual versus forecasts based on new-generation Phillips curve models. Note: The dotted lines show the 90% credible interval.

where $u_t^\star$ is an endogenous measure of the natural rate of unemployment (in other words, the measure of slack, $u_t - u_t^\star$, is estimated within the model). The 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 euro area inflation by Hindrayanto et al. (2019)[33] or by Stevens and Wauters (2018).[34]

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

$$(\pi_{t+1} - \pi_{t+1}^{TR}) = \alpha_{t+1}(\pi_t - \pi_t^{TR}) + \beta_{t+1}y_{t+1} + \nu_{t+1} \qquad (3.3)$$

$$\pi_{t+1}^{Cons} = a_{t+1} + b_{t+1}\pi_{t+1}^{TR} + u_{t+1} \qquad (3.4)$$

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.

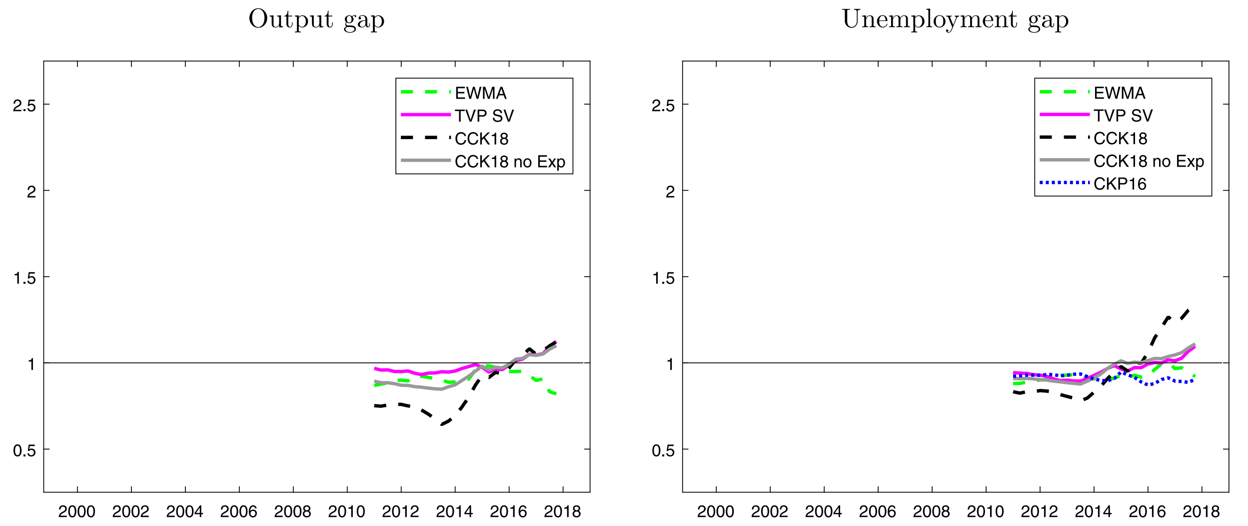

Fig. 3.7. Forecasting performance of new-generation Phillips curve models in real time. Note: The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. All models are estimated via Bayesian methods, and RMSFEs are based on the median inflation forecast. We use the vintages with the cut-off dates of the Eurosystem/ECB projections. Forecast horizon: 4 quarters ahead.

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 does not simply equate long-term forecasts from surveys (from Consensus Economics in our case) 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 à 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 1980s.[36] As *CCK18* relies on long-term inflation expectations from Consensus Economics, the estimation sample for this model starts in the 1990s and it is evaluated over a shorter period, similarly to the models used in Section 3.1. Fig. B.1 in the Appendix shows the pseudo-real-time estimates of trend inflation from both versions of the (Chan et al., 2018) model with output gap, as well as from the (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 are sizable differences in the results from the two versions of the (Chan et al., 2018) model: the trend estimates from *CCK18 no Exp* are much lower.[37]

Fig. 3.5 presents the forecasting performance of the models for the two main slack indicators. This performance is also compared to that 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).[38] 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 Fig. 3.1). 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 time-varying coefficients and stochastic volatility, at least

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 25,000 draws from the posterior distribution.

37 Fig. B.7 in the online appendix reports the parameter estimates from both versions of the (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).

38 We assume a normal inverse gamma prior for the coefficients α and β and for the variance of the residual ν_t. The mean and variance of the normal distribution are 0 and 0.2^2, respectively. The inverse gamma has three degrees of freedom and the scale is set so as to yield a unit prior variance for the residual.

Table A.1
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' calculations	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, backcasted
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 exchange rate vis–à–vis 19 trading partners
USD/EUR exchange rate	ECB	exchange rate against 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 backcasted 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 backcasted 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).

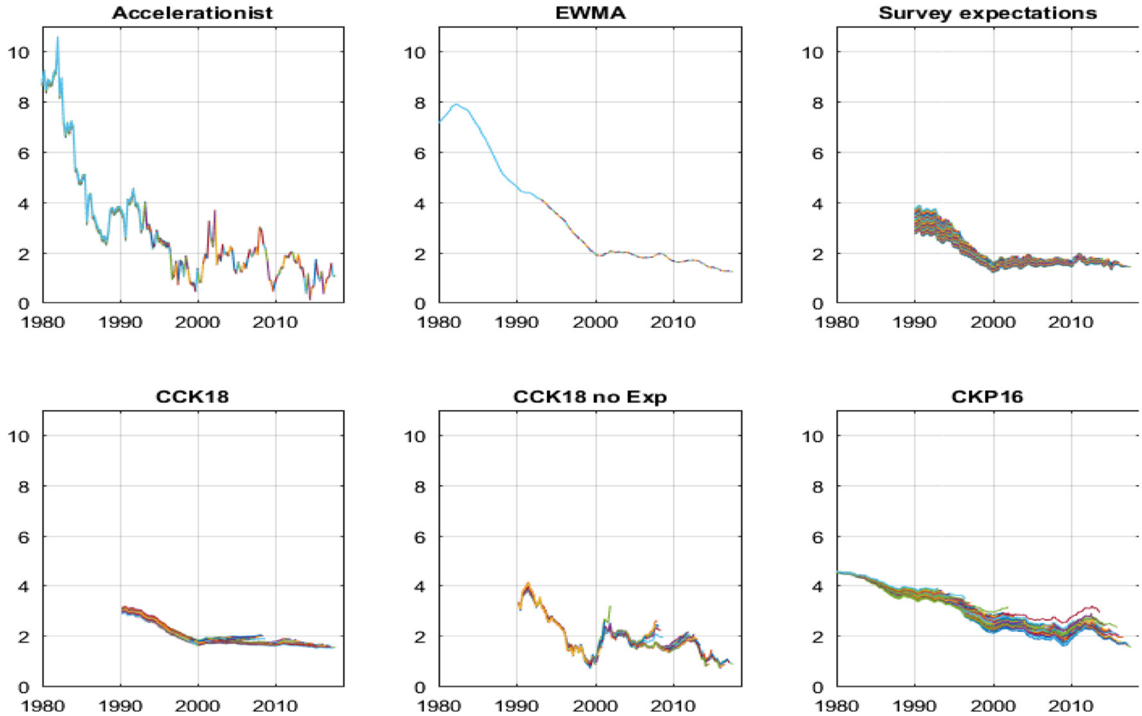

Fig. B.1. Pseudo-real-time estimates of inflation trends. Note: Under the accelerationist assumption, trend inflation is the past inflation rate. EWMA stands for the exponentially weighted moving average inflation trend (with a smoothing parameter equal to 0.05). The revisions in the long-term survey 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 here). *CCK18*, *CCK18 no Exp*, and *CKP16* correspond to the recursive estimates of inflation trends obtained from the models described in Section 3.4. The former two correspond to the version with output gap as the measure of slack.

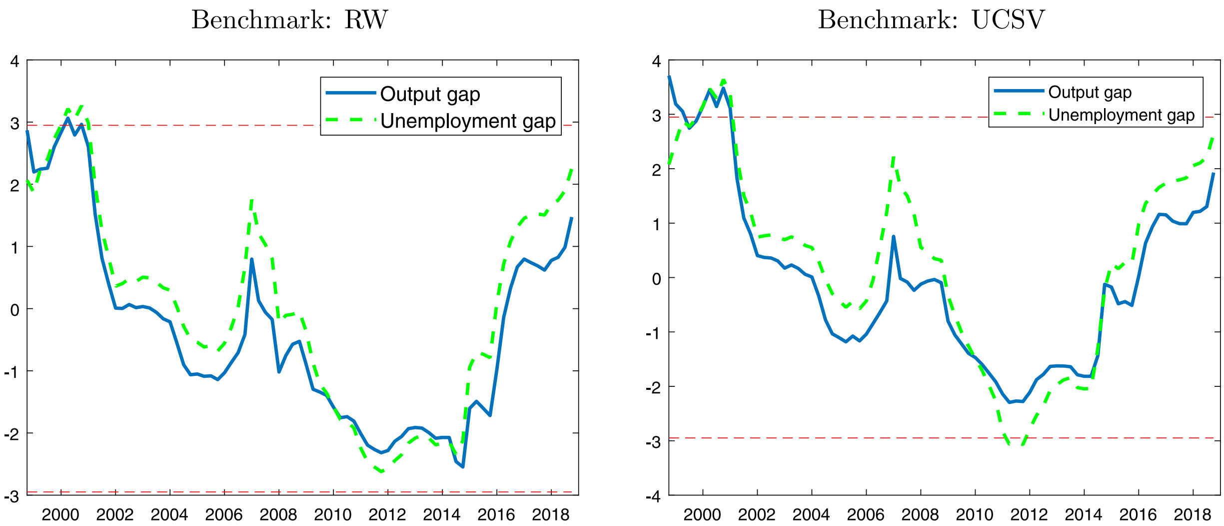

Fig. B.2. Giacomini–Rossi fluctuation test for the workhorse model. Note: Giacomini and Rossi (2010) fluctuation test for a rolling window of 20 quarters. Dashed lines show the critical values for the 90% confidence interval interval. The null of equal forecasting performance is rejected when the test 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. Forecast horizon: 4 quarters ahead.

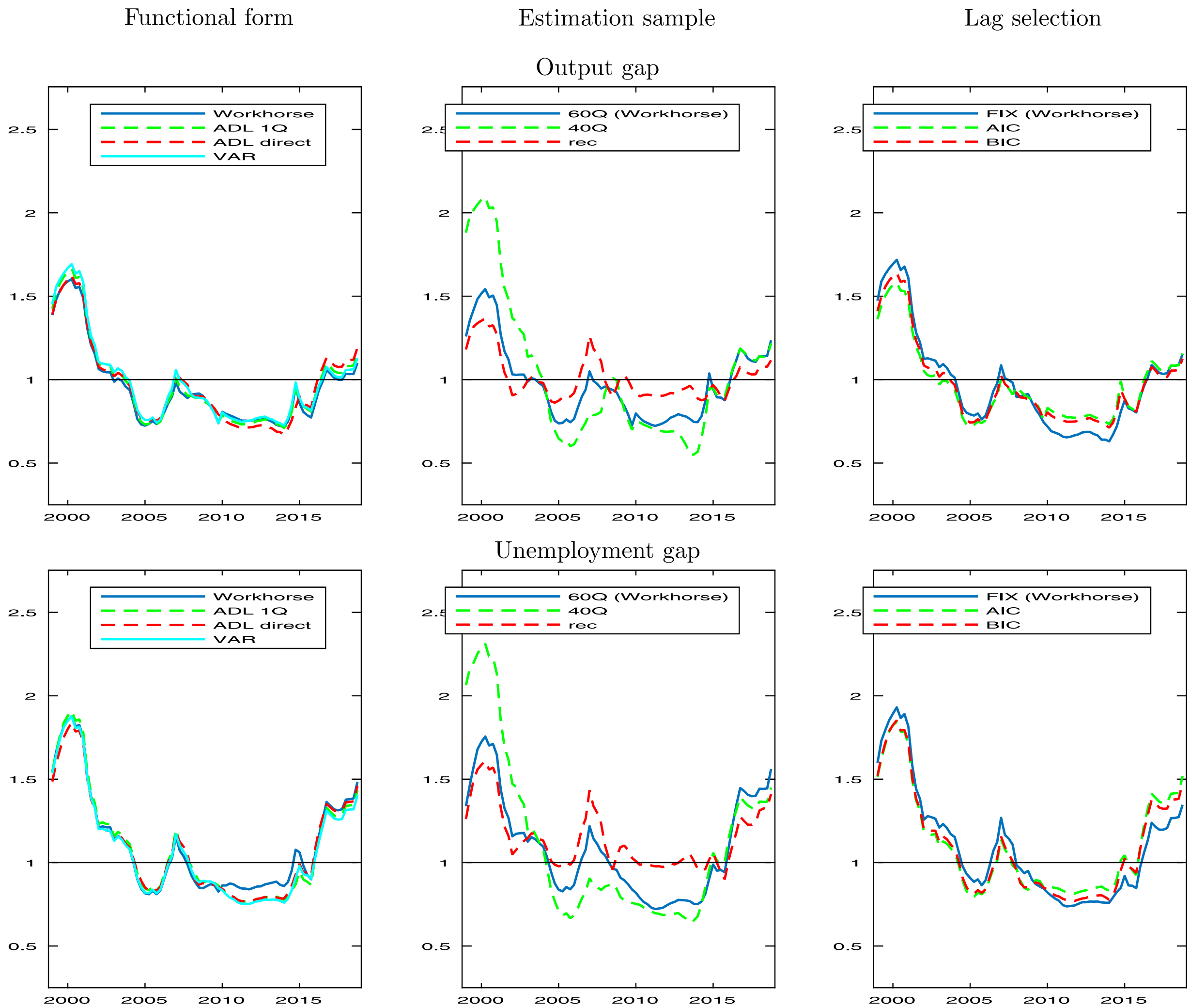

Fig. B.3. Impact of certain modelling choices on the forecasting performance. Note: The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. The chart compares accuracy across functional forms (first column), estimation samples (second column), and lag selection approaches (third column). The first column presents the 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 *ADL 1Q* it starts with the previous quarter, and for *ADL direct* it starts four 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 estimations (starting in 1980). Forecast horizon: 4 quarters ahead.

what regards the point forecasts.[39] *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 (see also the results presented in Table B.1). 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 & 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 Fig. B.7 in Appendix B, 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 Fig. B.8, indicate somewhat worse calibration of this model. This suggests advantages of the new-generation models when predictive distributions are of interest.

[39] 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.

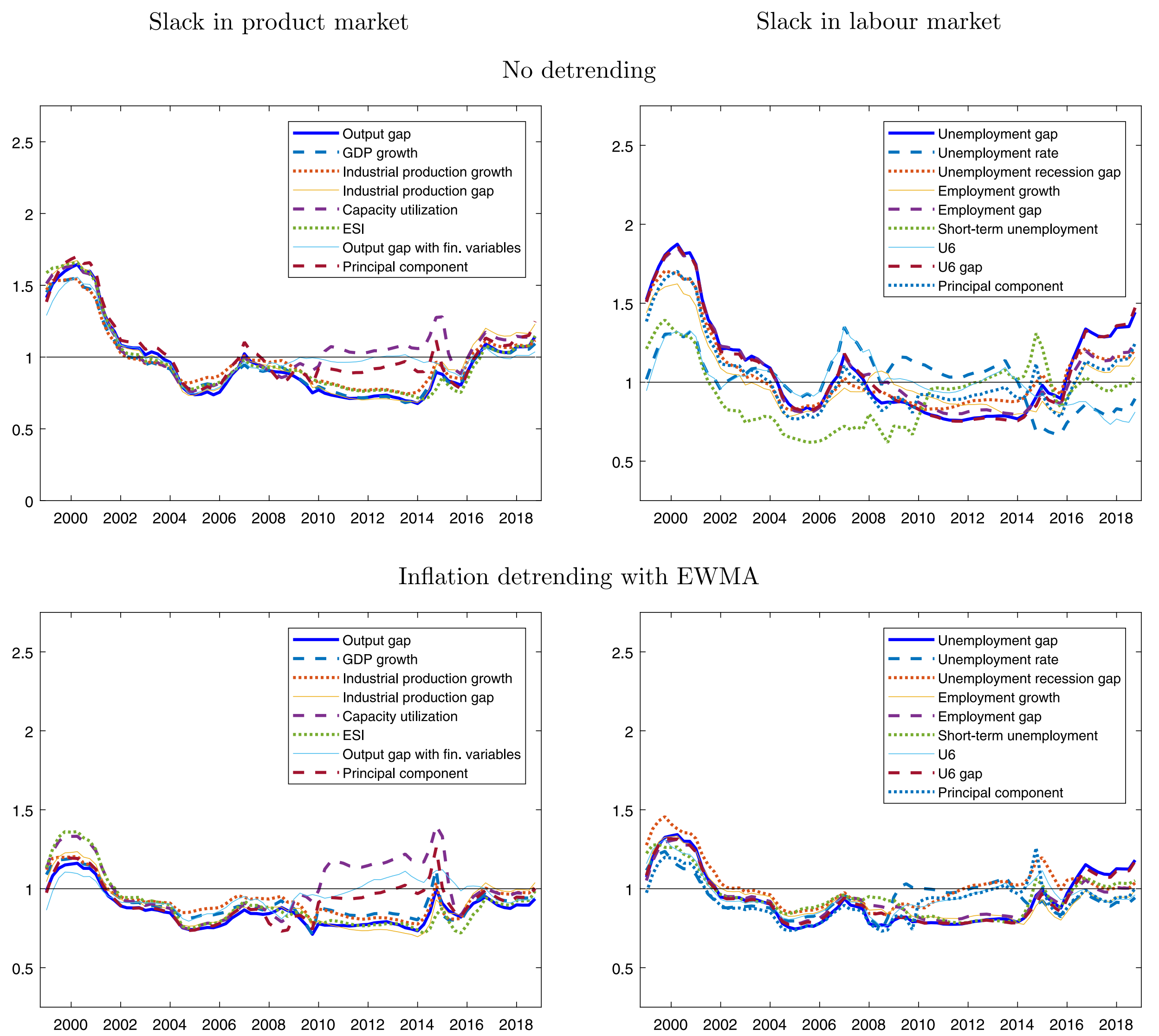

Fig. B.4. Impact of the choice of slack measure on the forecasting performance. Note: RMSFEs of the average forecast over all considered functional forms, estimation samples, and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. Forecast horizon: 4 quarters ahead.

The forecasts produced by these models are plausible, and, by and large, inflation outcomes lie within the credible intervals, as shown in Fig. 3.6. 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 better performance of these models over the recent period.

The new-generation Phillips curve models appear to be a promising avenue for conducting forecasts. 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.[40] Fig. 3.7 shows the results. The lessons drawn using pseudo-real-time data are also valid in a real-time setup. While most models worsen during the euro area post-sovereign debt crisis period, *Workhorse+EWMA* and *CKP16* still outperform the univariate benchmark.

4. Conclusion

In this paper we evaluated the forecasting performance of a wide range of Phillips curve specifications for euro

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

Table B.1
RMSFE for selected Phillips curve models.

		Benchmark		Workhorse model	Average across modelling choices				Workhorse & EWMA	New generation models			
		UCSV	RW		No detrending	Accelerationist	EWMA	Survey		TVP SV	CCK18	CCK18 no Exp	CKP16
Output gap	1994:2018	0.53	0.59	0.56	0.54	0.53	**0.46****	0.48	0.51	0.54	**0.45****	0.51	–
	2002:2018	0.51	0.57	0.47	0.45	0.53	**0.42****	**0.41***	**0.46****	0.51	0.43	0.49	–
Unemployment gap	1994:2018	0.53	0.59	0.63	0.60	0.53	0.49	0.56	0.53	0.52	0.47	0.43	0.47*
	2002:2018	0.51	0.57	0.56	0.51	0.53	0.46	0.46	0.48	0.49	0.46	0.49	**0.46*****

Note: Stars show that the Phillips curve models outperform the UCSV in a statistically significant manner with a 10, 5, 1% (*, **, ***) significance threshold according to the Diebold Mariano test. The Workhorse model corresponds to the one in Fig. 2.1 in the paper, the average across modelling choices with no detrending corresponds to Fig. 2.3, models with detrended inflation correspond to Fig. 3.1, the Workhorse & EWMA and the New generation models correspond to Fig. 3.5. For specifications with inflation detrended with survey measures the first evaluation sample starts in 1997. For the CCK18 models the first evaluation sample starts in 2000. Forecast horizon: 4 quarters ahead.

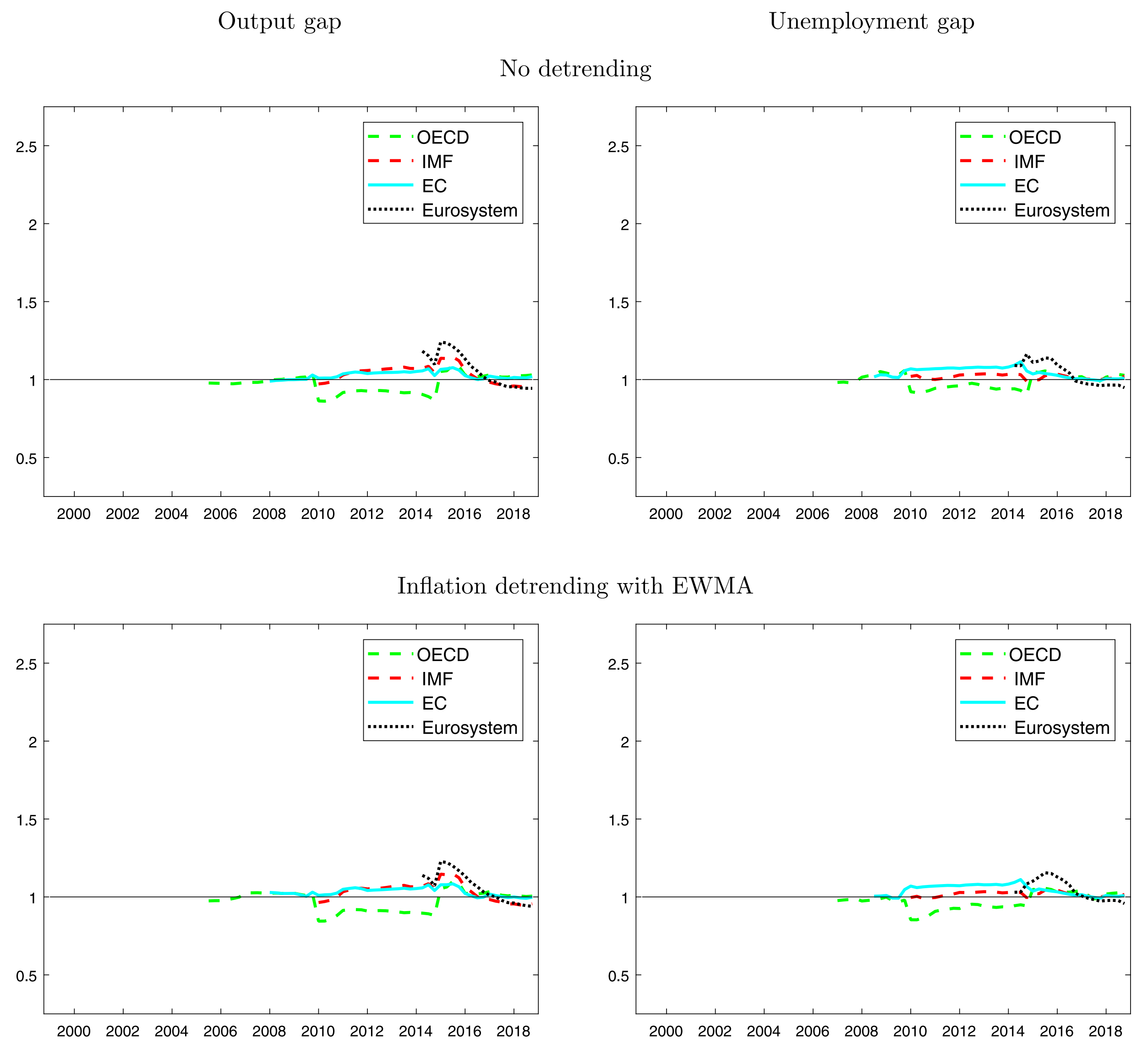

Fig. B.5. Accuracy of forecasts *conditional* on projections for slack from international economic institutions relative to their unconditional counterpart. Note: RMSFEs of the average forecast over ADL 1Q, VAR, and Workhorse model specifications, but also across estimation samples and lag choice approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE from the unconditional forecasts of the same models. Forecast horizon: 4 quarters ahead.

area HICP inflation excluding energy, over the period from 1994 to 2018. Our findings are relevant in the debate regarding the validity of the Phillips curve and are particularly useful for those building model toolboxes for forecasting inflation and crosschecking 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 trends—and, more recently, also 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

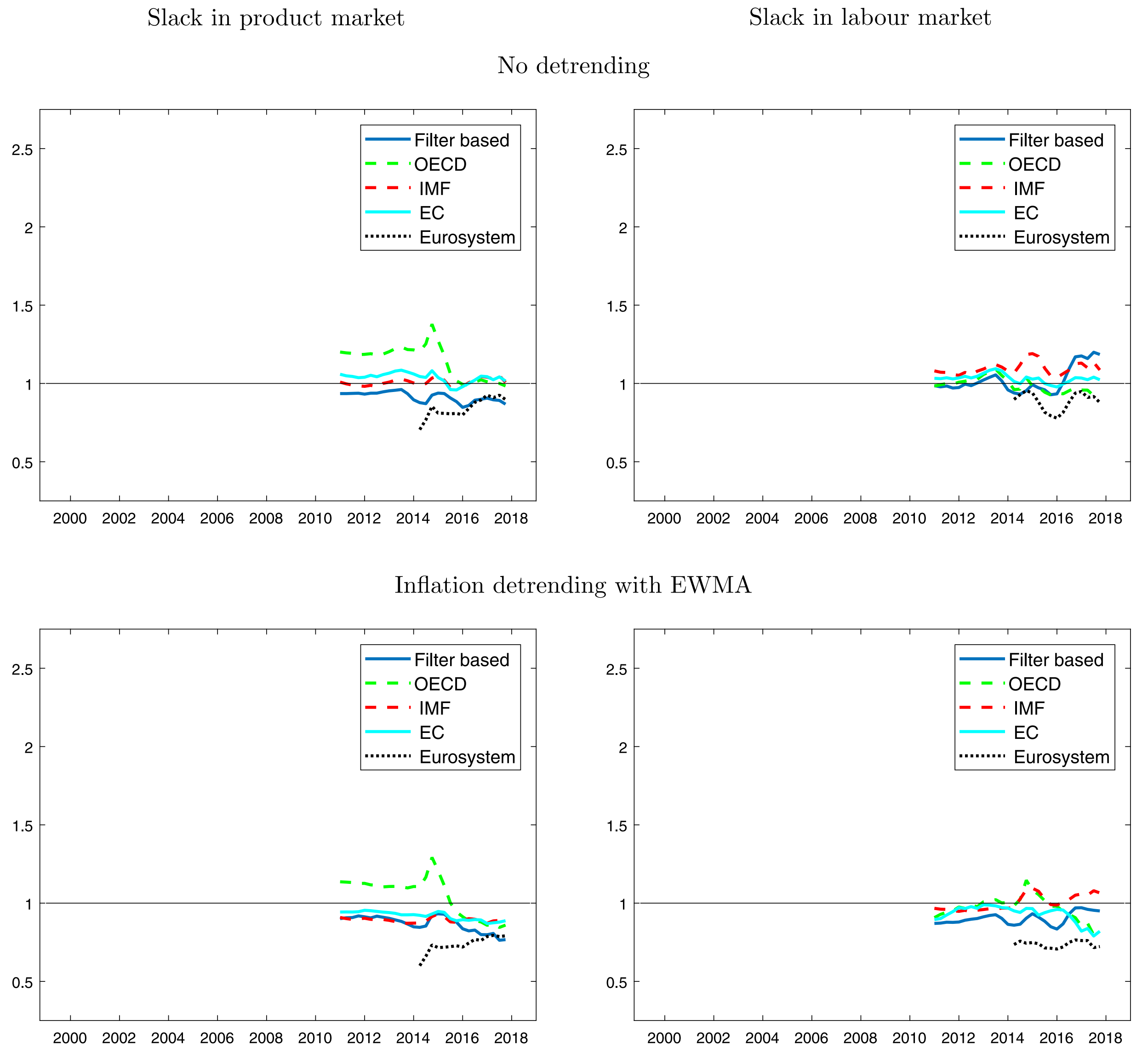

Fig. B.6. Real-time evaluation of "institutional" slack measures. Note: RMSFEs of the average forecast over all considered functional forms, estimation samples, and lag selection approaches. The RMSFEs were computed over a rolling window of 20 quarters relative to the corresponding RMSFE of the UCSV model. For filter-based measures of slack, we use the vintages with the cut-off dates of the Eurosystem/ECB projections. Forecast horizon: four quarters ahead.

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 trends 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 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

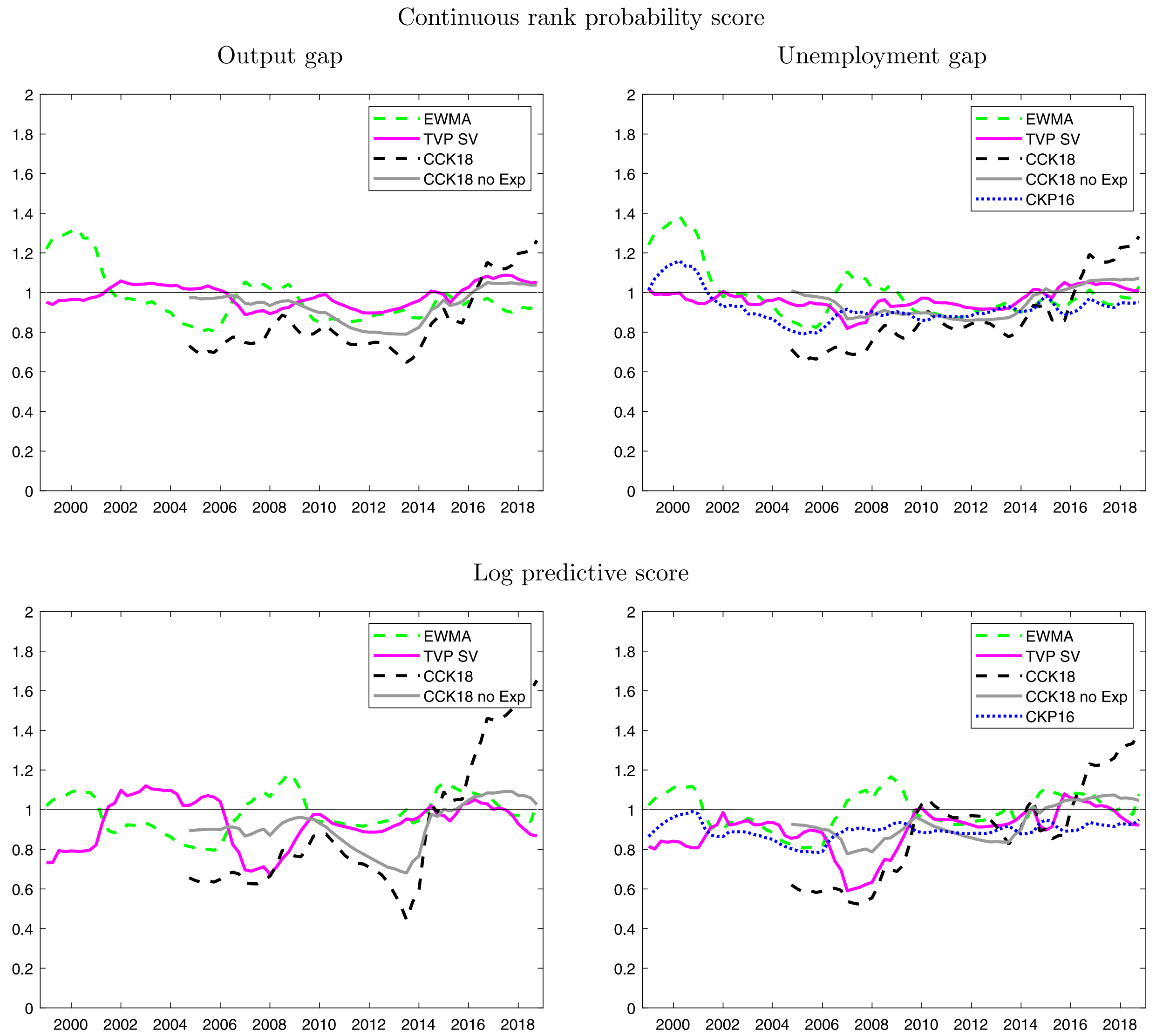

Fig. B.7. Forecasting performance of the new-generation Phillips curve models. Note: The continuous rank probability score and log predictive score were computed over a rolling window of 20 quarters relative to the ones corresponding to the UCSV model. Forecast horizon: 4 quarters ahead.

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 offer some hedge against instability in forecast performance. More sophisticated techniques, such as the approximate Bayesian model averaging employed by Garratt and Petrella (2019), could be explored going forward.

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 & Osbat, 2017; Hindrayanto et al., 2019). The analysis does not allow us to draw conclusions on the potential drivers of such declines. They could be related to the credibility of the monetary authority, demographics, financial factors, or the rise of globalisation or of e-commerce (see e.g. Cavallo, 2018; Ciccarelli & Osbat, 2017). Incorporating such structural drivers into forecasting models has not yet been thoroughly explored.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

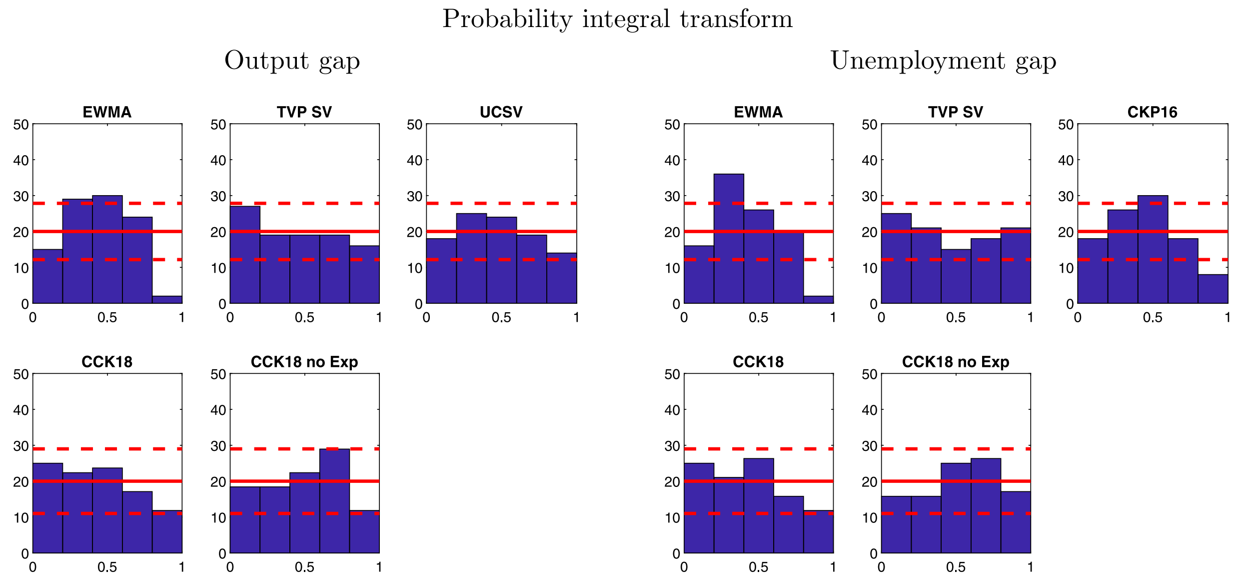

Fig. B.8. Forecasting performance of the new-generation Phillips curve models. Note: The intervals are computed following Diebold et al. (1998).

Appendix A. Data sources

See Table A.1.

Appendix B. Additional results

See Figs. B.1–B.8 and Table B.1.

Appendix C. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2021.12.001.

References

Álvarez, L. J., & Correa-López, M. (2020). Inflation expectations in euro area phillips curves. *Economics Letters*, *195*(C), http://dx.doi.org/10.1016/j.econlet.2020.10, https://ideas.repec.org/a/eee/ecolet/v195y2020ics0165176520302780.html.

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

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

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

Ball, L. M., & Mazumder, S. (2019). A Phillips curve with anchored expectations and short-term unemployment. *Journal of Money, Credit and Banking*, *51*(1), 111–137, https://ideas.repec.org/a/wly/jmoncb/v51y2019i1p111-137.html.

Ball, L. M., & Mazumder, S. (2020). A Phillips curve for the euro area. *ECB Working Papers 2354*, European Central Bank, https://ideas.repec.org/p/ecb/ecbwps/20202354.html.

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-casting and the real-time data flow. In A. T. Graham Elliott (Ed.), *Handbook of economic forecasting* (pp. 195–237). North Holland.

Bańbura, M., & 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., & Blanchflower, D. G. (2018). Underemployment in the US and Europe. *NBER Working Papers 24927*, National Bureau of Economic Research, Inc, https://ideas.repec.org/p/nbr/nberwo/24927.html.

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

Bereau, S., Faubert, V., & 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? *PB16-1*, Peterson Institute for International Economics, Working Paper WP15-19. https://ideas.repec.org/p/iie/pbrief/pb16-1.html.

Blanchard, O., Cerutti, E., & Summers, L. H. (2015). Inflation and activity: Two explorations and their monetary policy implications. *WP15-19*, Peterson Institute for International Economics, Working Paper WP15-19 https://ideas.repec.org/p/iie/wpaper/wp15-19.html.

Bobeica, E., & Jarociński, M. (2019). Missing disinflation and missing inflation: A VAR perspective. *International Journal of Central Banking*, *15*(1), 199–232, https://ideas.repec.org/a/ijc/ijcjou/y2019q1a5.html.

Bobeica, E., & 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.

Bracha, A., & 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, https://ideas.repec.org/p/fip/fedbwp/16-29.html.

Buelens, C. (2012). Inflation forecasting and the crisis: Assessing the performance of different forecasting models and methods. European Commission 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., & 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. In *Jackson hole economic symposium conference proceedings*. Federal Reserve Bank of Kansas City, Jackson Hole, Wyoming, Technical report.

Chan, J. C. C. (2018). Specification tests for time-varying parameter models with stochastic volatility. *Econometric Reviews*, *37*(8), 807–823. http://dx.doi.org/10.1080/07474938.2016.116, https://ideas.repec.org/a/taf/emetrv/v37y2018i8p807-823.html.

Chan, J. C. C., Clark, T. E., & 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, https://ideas.repec.org/a/wly/jmoncb/v50y2018i1p5-53.html.

Chan, J. C. C., Koop, G., & 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, https://ideas.repec.org/a/wly/japmet/v31y2016i3p551-565.html.

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

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

Clark, T. E., & Doh, T. (2014). Evaluating alternative models of trend inflation. *International Journal of Forecasting*, *30*(3), 426–448. http://dx.doi.org/10.1016/j.ijforecast.2013, https://ideas.repec.org/a/eee/intfor/v30y2014i3p426-448.html.

Clark, T. E., & McCracken, M. W. (2008). Forecasting with small macroeconomic VARs in the presence of instabilities. In D. Rapach, & M. Wohar (Eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty (Frontiers of Economics and Globalization)*: *vol. 3*, (pp. 93–147). Emerald Group Publishing Limited, Bingley.

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

Clark, T. E., & McCracken, M. W. (2010). Averaging forecasts from VARs with uncertain instabilities. *Journal of Applied Econometrics*, *25*(1), 5–29. http://dx.doi.org/10.1002/jae.1127, https://ideas.repec.org/a/jae/japmet/v25y2010i1p5-29.html.

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. European Central Bank.

Coibion, O., & 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, https://ideas.repec.org/a/aea/aejmac/v7y2015i1p197-232.html.

Coibion, O., Gorodnichenko, Y., & 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, https://ideas.repec.org/p/nbr/nberwo/25511.html.

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

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

Diebold, F. X., Gunther, T. A., & Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, *39*(4), 863–883.

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

Dotsey, M., Fujita, S., & Stark, T. (2018). Do Phillips curves conditionally help to forecast inflation? *International Journal of Central Banking*, *14*(4), 43–92, https://ideas.repec.org/a/ijc/ijcjou/y2018q3a2.html.

Ehrmann, M., Ferrucci, G., Lenza, M., & O'Brien, D. (2018). Measures of underlying inflation for the euro area. *Economic Bulletin Articles*, *4*, https://ideas.repec.org/a/ecb/ecbart/201800043.html.

Eser, F., Karadi, P., Lane, P. R., Moretti, L., & Osbat, C. (2020). The Phillips curve at the ECB. *Working Paper Series 2400*, European Central Bank, https://ideas.repec.org/p/ecb/ecbwps/20202400.html.

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

Faust, J., & Wright, J. H. (2013). Forecasting inflation. In G. Elliott, & A. Timmermann (Eds.), *Handbook of economic forecasting, vol. 2* (pp. 2–56). North Holland.

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

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

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

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

Friedrich, C. (2014). Global inflation dynamics in the post-crisis period: What explains the twin puzzle? *Staff Working Papers 14–36*, Bank of Canada, https://ideas.repec.org/p/bca/bocawp/14-36.html.

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

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

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

Garratt, A., & Petrella, I. (2019). Commodity prices and inflation risk. *EMF Research Papers 23*, Economic Modelling and Forecasting Group, https://ideas.repec.org/p/wrk/wrkemf/23.html.

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

Giannone, D., Henry, J., Lalik, M., & Modugno, M. (2012). An area-wide real-time database for the euro area. *The Review of Economics and Statistics*, *94*(4), 1000–1013, https://ideas.repec.org/a/tpr/restat/v94y2012i4p1000-1013.html.

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

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.

Granziera, E., & Sekhposyan, T. (2019). Predicting relative forecasting performance: An empirical investigation. *International Journal of Forecasting*, *35*(4), 1636–1657. http://dx.doi.org/10.1016/j.ijforecast.2019, https://ideas.repec.org/a/eee/intfor/v35y2019i4p1636-1657.html.

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

Hasenzagl, T., Pellegrino, F., Reichlin, L., & Ricco, G. (2018). A model of the Fed's view on inflation. *CEPR Discussion Papers 12564*, C.E.P.R. Discussion Papers, https://ideas.repec.org/p/cpr/ceprdp/12564.html.

Hindrayanto, I., Samarina, A., & 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., & 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* (pp. 1–17). Washington: International Monetary Fund.

Jarociński, M., & 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, https://ideas.repec.org/a/wly/jmoncb/v50y2018i6p1189-1224.html.

Kozicki, S., & Tinsley, P. A. (2001). Shifting endpoints in the term structure of interest rates. *Journal of Monetary Economics*, *47*(3), 613–652.

Kulikov, D., & Reigl, N. (2019). Inflation expectations in Phillips curves models for the euro area. *Working paper No. 8/2019*, Bank of Estonia.

Lane, P. R. (2019). The Phillips curve at the ECB. Speech at the 50th anniversary conference of the money, macro & finance research group, London school of economics.

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

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

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

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

Orphanides, A., & 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.

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

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. Federal Reserve System.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*(3), 821–852, https://ideas.repec.org/a/oup/restud/v72y2005i3p821-852.html.

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

Stella, A., & Stock, J. H. (2015). A state-dependent model for inflation forecasting. In S. J. Koopman, & N. Shephard (Eds.), *Unobserved components and time series econometrics*. Oxford University Press.

Stevens, A., & 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, https://ideas.repec.org/p/nbb/reswpp/201810-355.html.

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

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

Stock, J. H., & Watson, M. W. (2009). Phillips curve inflation forecasts. In J. Fuhrer, Y. Kodrzycki, J. S. Little, & G. Olivei (Eds.), *Understanding inflation and the implications for monetary policy, a Phillips curve retrospective* (pp. 101–186). The MIT Press.

Stock, J. H., & Watson, M. W. (2010). Modeling inflation after the crisis. *NBER Working Papers 16488*, National Bureau of Economic Research, Inc, https://ideas.repec.org/p/nbr/nberwo/16488.html.

Stram, D. O., & Wei, W. W. S. (1986). A methodological note on the disaggregation of time series totals. *Journal of Time Series Analysis*, *7*(4), 293–302. http://dx.doi.org/10.1111/j.1467-9892.1986, https://ideas.repec.org/a/bla/jtsera/v7y1986i4p293-302.html.

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

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

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