

DSGE MODEL FORECASTING: RATIONAL EXPECTATIONS VS. ADAPTIVE LEARNING

ANDERS WARNE*

JULY 29, 2024

ABSTRACT: This paper compares within-sample and out-of-sample fit of a DSGE model with rational expectations to a model with adaptive learning. The Galí, Smets and Wouters model is the chosen laboratory using quarterly real-time euro area data vintages, covering 2001Q1–2019Q4. The adaptive learning model obtains better within-sample fit for all vintages used for estimation in the forecast exercise and for the full sample. However, the rational expectations model typically predicts real GDP growth better as well as jointly with inflation. For the marginal inflation forecasts, the same holds for the inner quarters of the forecast horizon, while the adaptive learning model predicts better for the outer quarters.

KEYWORDS: Bayesian inference, CRPS, euro area, forecast comparison/evaluation, log score, real-time data.

JEL CLASSIFICATION NUMBERS: C11, C32, C52, C53, E37.

1. INTRODUCTION

Estimated dynamic stochastic general equilibrium (DSGE) models are regularly used by many central banks and other policy-oriented organizations to analyse macroeconomic conditions and to assess future developments; see, e.g., Lindé, Smets, and Wouters (2016) and Christiano, Eichenbaum, and Trabandt (2018) for recent surveys. Given the micro-foundations and optimization-based behavior of economic agents in such models, an important reason for their popularity is that they facilitate structural interpretations of the macroeconomic environment. Expectations are typically assumed to be rational such that the joint probability distribution of the model variables is fully consistent with the model and policy experiments are therefore not subject to the Lucas critique. Furthermore, the rational expectations (RE) assumption provides strong cross-equation restrictions on the resulting stochastic processes which describe the model variables and which help to better identify many of the parameters. Bayesian inference is frequently used for estimation, while calibration of some parameters that may be hard to identify from the macro data concerns relatively few parameters in the benchmark models, such as the well-known Smets and Wouters (2007) model, and where instead micro data may be consulted.

The main contribution of the current paper is the real-time density forecast comparison and evaluation of a benchmark DSGE model under RE and adaptive learning (AL) using the Slobodyan

ACKNOWLEDGEMENTS: I am grateful for discussions and correspondence with Günter Coenen, Matthieu Darracq Paries, Marco Del Negro, Mátyás Farkas, Sergio Santoro and Raf Wouters, as well as from seminar participants at the European Central Bank. This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors. The opinions expressed in this paper are those of the author and do not necessarily reflect the views of the European Central Bank or the Eurosystem.

*Forecasting and Policy Modelling Division of DG-Economics, European Central Bank, 60640 Frankfurt am Main, Germany; e-mail: anders.warne@ecb.europa.eu; Phone: +49-69 1344 8737; ORCID: 0009-0004-5594-4161.

and Wouters (2012a) approach. To my knowledge, such an empirical study is not available in the literature and, as a complement to the out-of-sample investigation, the models are also compared recursively within-sample to learn if the fit of a DSGE model subject to AL is better (or worse) than under RE for the euro area data. The selected DSGE model is the euro area version of the Galí, Smets, and Wouters (2012, SWU) model, which extends the Smets and Wouters (SW) model to incorporate unemployment; see Smets, Warne, and Wouters (2014) and McAdam and Warne (2019) for details on the euro area version of the SWU model. The real-time database for the euro area, introduced by Giannone, Henry, Lalik, and Modugno (2012), is considered for the forecast exercise, covering backcasts, nowcasts and up to eight-quarter-ahead forecasts over the forecast sample 2001Q1–2019Q4. The density forecasts are compared using the log score as well as the continuous ranked probability score (CRPS) and its multivariate version, the energy score, while the predictive distributions are evaluated using formal tests based on the standard probability integral transform and, in the case of multivariate distributions, Box’s (1980) density ordinate transform.

The forecasting methodology for DSGE models is discussed in Del Negro and Schorfheide (2013), which also provides illustrations on the empirical performance of such models relative to survey data, professional forecasts and reduced form models for the US; see Adolfson, Lindé, and Villani (2007) and Christoffel, Coenen, and Warne (2011) for discussions on methodology and empirical evidence from the euro area. The point and density forecasting performance of DSGE models compares well with reduced form benchmark models, such as BVARs, especially over the medium term. This general finding is also supported by Warne, Coenen, and Christoffel (2017) for a larger dimension DSGE model, which is compared with a large BVAR using the Bańbura, Giannone, and Reichlin (2010) approach, a DSGE-VAR model and a multivariate random-walk model for euro area data covering the sample 1999Q1–2011Q4.

The above euro area studies do not make use of real-time or survey data. Smets et al. (2014) utilizes such euro area data with the SWU model and focuses on point forecasts for the sample 2001Q1–2010Q4. One of the findings in that study is that adding one to two-year-ahead professional forecasts on real GDP growth, inflation and unemployment to the conditioning information, without otherwise changing the model, overall improves the forecasts of the SWU model, with the main deterioration appearing in the short-term nominal interest rate forecasts. McAdam and Warne (2019) extends the euro area real-time sample to 2014Q4 and focuses on comparing the density forecasts of the SW model to the SWU model, and an extension which includes financial frictions of the Bernanke, Gertler, and Gilchrist (1999, BGG) type. Using the log score to measure forecast performance, their study reports that adding financial frictions of the BGG-type to the SW model leads to a deterioration of the forecasts, not only of the density forecasts but also for the point forecasts. Modelling and measuring unemployment, on the other hand, tends to moderately improve the density forecasts of the SW model, especially since the onset of the Great Recession in 2008. This finding is the main motivation for choosing the SWU model in the current paper.

It is arguably unrealistic to assume that agents in a model know all its details and collect all necessary information such that their expectations about the future of the variables are fully model-consistent. There is a growing empirical literature on replacing the RE hypothesis with alternative

models of expectations. Evans and Honkapohja (2009) and Woodford (2013) provide surveys of such alternatives, while Evans and Honkapohja (2001) gives a textbook treatment of the foundations of learning theory in macroeconomics and Evans and McGough (2020) provides a recent survey of studies that use AL in macroeconomic and financial applications.¹

From an empirical perspective, it has been pointed out that a shortcoming of estimated DSGE models under RE is that they require highly persistent exogenous shock processes to explain observed persistence in the data. It is claimed by, for instance, Milani (2007, 2009) and Orphanides and Williams (2005) that learning can influence the dynamic responses to shocks and thereby increase the persistence in the shock responses. Milani (2007) estimates a small DSGE model on US data and finds that learning reduces structural frictions (habit persistence and price indexation) while also improving within-sample fit as measured by the marginal likelihood. Slobodyan and Wouters (2012b) studies the influence of constant gain learning relative to RE and finds that learning has little influence on the dynamics when the information set used for the learning process is the same as under RE, while a restricted information set can improve the fit without having a sizeable effect on the structural parameters related to real and nominal frictions.

Slobodyan and Wouters (2012a) estimate the SW model on US data under both RE and AL with a small forecasting model and where the belief coefficients are time-varying. The AL specification with the greatest marginal likelihood outperforms the RE version and displays much lower persistence of the wage and price markup shocks. In fact, a variant which restricts the ARMA parameters to zero for both these shock processes has a marginal likelihood very close to the best performer and, in addition, has lower wage and price indexation than the RE version. Their study also finds that the good within-sample fit is reflected in the out-of-sample point forecasting performance, especially for the shorter-term horizons.²

The remainder of the paper is organized as follows. Section 2 discusses expectation formation in log-linearized DSGE models with a focus on the AL approach considered in Slobodyan and Wouters (2012a) and the importance of the selection of forward looking variables. The DSGE model is briefly outlined in Section 3 and specification differences between the RE and AL cases are discussed. Estimation of the DSGE model parameters is thereafter presented in Section 4 for the full sample, including specification analysis, and within-sample fit for recursive annual vintages over the forecast sample. The comparisons of the point and density forecasts for the baseline RE and AL models are covered in Section 5, including density forecast evaluation and some sensitivity analysis. The

¹ By utilizing survey data and adopting a DSGE-VAR approach to assess the extent and sources of model misspecification Cole and Milani (2019) finds that the RE assumption is the main reason their new Keynesian model fails to match the data well.

² Forecast comparisons for RE and AL models are otherwise very rare in the literature. In a recent paper, Carvalho, Eusepi, Moench, and Preston (2023) develops a Rotemberg-pricing model with imperfect information to describe long-term trends in inflation expectations and shows that it can explain developments in long-term forecast data from US professional forecasters better than under RE. Elias (2022) estimates a small-scale new Keynesian model with heterogeneous Euler equation AL on US data and finds that a model with heterogeneous AL fits the data better than a model with homogeneous AL. Point forecasts of the output gap and inflation made by the three types of agents in the model are compared over the pre-Great Moderation, Great Moderation and pre-Great Recession periods and the RMSEs suggests heterogeneity matters. Still, these exercises are not carried out with an RE version of the model.

main findings are summarized in Section 6, while technical details on DSGE models with AL-based expectations as well as some additional empirical results are available in the Online Appendix.

2. EXPECTATIONS IN DSGE MODELS

DSGE models are usually estimated and analysed under the assumption that expectations are formed rationally and therefore model consistent. The structural form of a log-linearized DSGE model is represented by:

$$H_{-1}z_{t-1} + H_0z_t + H_1E_t[z_{t+1}] = D\eta_t, \quad t = 1, 2, \dots, \quad (1)$$

where z_t is a p -dimensional vector of model or state variables, E_t represents expectations at t and η_t is a q -dimensional vector of i.i.d. Gaussian structural shocks with zero mean and identity covariance matrix. Under the assumption of rational expectations (RE), E_t is given by the mathematical expectations operator using the history of z_t and all parameters of the model as input. The unique and convergent solution to this stochastic difference equation, when it exists, is given by

$$z_t = Fz_{t-1} + B\eta_t, \quad (2)$$

where F and B satisfy the cross equation restrictions $B = (H_0 + H_1F)^{-1}D$ and $H_{-1} + H_0F + H_1F^2 = 0$. When estimating such a DSGE model, the solution corresponds to the state equation, while the measurement equation is given by

$$y_t = \mu + H'z_t + w_t, \quad (3)$$

where y_t is an n -dimensional vector of observable variables and w_t an i.i.d. Gaussian measurement error, independent of η_t and with mean zero and covariance matrix R .

Over the last decades, alternative approaches to modelling expectations have been suggested in the literature. These include but are not limited to the bounded rationality model of Sargent (1993), rational inattention as in Sims (2003), the sticky information model of Mankiw and Reis (2002), partial information as in Svensson and Woodford (2003) and the learning approach of Evans and Honkapohja (2001). For a recent survey on adaptive learning (AL) in macroeconomics as well as references to alternatives to AL, see Evans and McGough (2020). Below, the AL approach suggested by Slobodyan and Wouters (2012a) is employed.

To relax the strict implications of the RE assumption, Slobodyan and Wouters (2012a) assume that agents forecast the forward looking variables of the model as a reduced form of the lagged state variables. A special case of this is given by the expression in equation (2), but it is also possible that the reduced form model differs from the RE solution. First, the parameters of the reduced form need not satisfy the cross-equation restrictions of the RE solution. Second, the reduced form may involve additional lags of the state variables and/or include the deterministic variables, which influence the law of motion of the forward looking variables in z_t .

Direct use of the representation in (1) to solve the model with an assumed learning process is called the Euler-equation approach to AL since the representation resembles a set of first-order conditions, while agents behave like econometricians who estimate the parameters of their model of the economy and use it to make forecasts; see, e.g., Evans and Honkapohja (2001) and Eusepi

and Preston (2018b).³ The forward looking variables of the DSGE model in (1) may be equal to the variables that appear in the expectation term, but need not be. The minimum number of forward looking variables is often equal to the rank of H_1 . Potentially, all variables appearing in expectations can be forward looking, but this need not be the case. For example, variables that are assumed to be shock processes are exogenous and are therefore not forward looking although they may appear in the expectations part of the model. In addition, such variables may affect the rank of H_1 . Ruling out such variables, the rank of H_1 is equal to the minimum number of forward looking variables supported by the model, while the maximum is equal to the number of non-zero columns of H_1 . Hence, the precise specification of the model matters when trying to identify possible forward looking variables. The Online Appendix discusses how the model in (1) can be transformed based on a given selection of forward looking variables.

The observation that the DSGE model can support several selections of forward looking variables points at a deeper issue concerning the uniqueness of solutions under AL. Different but equivalent formulations of a DSGE model with RE have the same unique solution, but this is typically not the case when RE is replaced with AL. For example, the choice of forward looking variables may have a direct implication for the solution of the model. How different the solutions are for the various choices of forward looking variables depends on the assumed expectation process, but it needs to be recognized. For example, the matrix H_1 has rank five in the Slobodyan and Wouters (2012a) model while the number of nonzero columns of this matrix is seven. This means that we may choose between five and seven forward looking variables. In that paper the authors chose seven, but the choices five and six are also supported by suitable model transformations as two of the variables appearing in the expectation term can be substituted for.

The AL approach in Slobodyan and Wouters (2012a) involves a number of decisions that the researcher needs to make. Apart from choosing the forward looking variables, the next step is to specify the perceived law of motion (PLM) for these variables. In Slobodyan and Wouters (2012a), the authors consider the simple approach of assuming that each forward looking variable is explained by a constant term, two of its own lags and a residual. The belief coefficients in these equations are time-varying and follow a VAR(1) process. This provides a state-space setup for the evolution of expectations where an additional Kalman filter is needed. Some details on this are found in their paper and additional details are available in the Online Appendix. The simple PLM from Slobodyan and Wouters will also be used in the empirical part as the benchmark specification, but alternative PLMs will also be considered in Section 5.4.

The state-space equations for the PLM and belief coefficients involve some additional parameters, covering the innovations to the PLM, the VAR coefficients and innovations of the belief coefficient equations. Slobodyan and Wouters (2012a) restricts the potential number of parameters to three, two scale parameters for the initial belief coefficient covariance matrix (σ_r) and the innovation covariance matrix to the belief coefficients (σ_ε) and one common autoregressive parameter (ρ) for the assumed

³ See, for instance, Eusepi and Preston (2018b, Section 4.2) and references therein regarding the impact of the so-called anticipated utility approach relative to the Euler-equation approach for the DSGE model solution(s) under learning or imperfect information.

diagonal VAR(1) matrix. Among these the scale parameters are calibrated while the autoregressive parameter is estimated.⁴ Furthermore, the innovation matrices of the PLM and belief coefficients are not diagonal and instead these moments are computed using the RE solution as well as the SURE-type formulation of the PLM of the DSGE model.

The DSGE model can now be solved under AL and the solution is called the actual law of motion (ALM). Since expectations under AL are typically not model consistent, the ALM and PLM differ for the forward looking variables. The time-variation of the belief coefficients imply that also the ALM coefficients are time-varying. The constant term in the PLM also feeds into the ALM such that it has a time-varying drift term. Provided that the maximum number of lags in the PLM does not exceed two, the ALM has the following form:

$$z_t = \mu_t + F_t z_{t-1} + B_t \eta_t, \quad (4)$$

where μ_t , F_t and B_t depend on the update estimates at t of the belief coefficients in the PLM and the structural form matrices of the DSGE model. Should the PLM have more than two lags, then additional lags of the state variables would be included in (4).

Once the solution is available, the parameters of the DSGE model can be estimated via a double Kalman filter algorithm, where first the state variables and likelihood function are estimated using on the ALM from the previous time period. The belief coefficients are thereafter computed using the estimated state variables from step one while the ALM is obtained in the last and third step of the algorithm, taking the new belief coefficients into account. This ALM is used as input for the next time period. The mathematical details of this filter, including its initialization, and the corresponding Kalman smoother are described in the Online Appendix.

3. THE DSGE MODEL

The well-known Smets and Wouters (2007) (SW) model is founded on a continuum of utility-maximizing households and profit-maximizing intermediate-good-producing firms who, respectively, supply labor and intermediate goods in monopolistic competition and set wages and prices. Final-good-producers use these intermediate goods and operate under perfect competition. The model incorporates several real and nominal rigidities, such as habit formation, investment adjustment costs, variable capital utilization and Calvo staggering in prices and wages. The monetary authority follows a Taylor-type rule when setting the nominal interest rate. There are seven stochastic processes: a TFP shock; a price and a wage markup shock; a risk premium (preference) shock; an exogenous spending shock; an investment-specific technology shock; and a monetary policy shock. The observed variables of the euro area version in McAdam and Warne (2019) are: real GDP, real private consumption, real total investment, total employment, real wages, the GDP deflator and the short-term nominal interest rate.

The euro area version of the Galí et al. (2012) extension of the SW model is presented in Smets et al. (2014). It explicitly provides a mechanism for explaining unemployment. This is accomplished

⁴ The σ_r parameter is denoted by σ_0 in Slobodyan and Wouters (2012a) while σ_ε is denoted by σ_v .

by modelling the labor supply decisions on the extensive margin (whether to work or not) rather than on the intensive margin (how many hours to work). As a consequence, the unemployment rate is added as an observable variable, while labor supply shocks are allowed for. The extension is motivated by the critique that the SW model cannot properly identify wage markup shocks; see Chari, Kehoe, and McGrattan (2009). The inclusion of observable unemployment into the wage equation can help to overcome that problem. Henceforth and following McAdam and Warne (2019), this extension is called the SWU model.

Smets et al. (2014) conducts a point forecast study with this model and a few benchmarks utilizing the real-time database (RTD) of the euro area for the sample 2001Q1–2011Q4; see Giannone et al. (2012).⁵ McAdam and Warne (2019) also make use of the SWU model to compare its real-time density forecasting properties with the SW model as well as with an extension of the latter subject to financial frictions based on the financial accelerator mechanism of Bernanke et al. (1999). Estimating these models on the vintages 2001Q1–2014Q4 from the euro area RTD, they find that the inclusion of unemployment gives some, albeit small, improvements over the SW model when comparing density forecasts of inflation and real GDP growth, while the financial frictions based extension overall has much worse forecasting properties.

Slobodyan and Wouters (2012a) estimate the parameters of a variant of the SW model on US data until 2008Q4 with RE and compare its behavior to an AL version. They deviate from the standard SW model by assuming that the output gap is not the flexible-price-gap, but with potential output determined by setting labor and capital to zero in the log-linearized production function such that potential output is proportional to the TFP shock. For the AL case this has the advantage of decreasing the number of forward looking variables as all those connected with the flexible price part are excluded. Their baseline variant under AL assumes that the price and wage markup shocks are i.i.d. rather than ARMA(1,1), while the two scale parameters for the initial belief coefficient covariance matrix and the innovation covariance matrix of the belief coefficients are calibrated to $\sigma_r = 0.03$ and $\sigma_\varepsilon = 0.003$, respectively, and the prior of the autoregressive parameter ρ is assumed to be standard uniform. The AL version of the SWU model makes use of some of these assumptions, such as the output gap assumption, while the RE version is specified exactly like in McAdam and Warne (2019). The estimation findings are discussed in the next section, including some sensitivity analysis with respect to the baseline specifications.

4. ESTIMATION OF THE SWU MODEL

The construction of the RTD combined with the AWM database is discussed in some detail by Smets et al. (2014) and the extension until 2015Q4 is covered in McAdam and Warne (2019). For the current study, the data vintages have been extended until 2020Q4 and the variables involved are real GDP, real private consumption, real total investment, total employment, the GDP deflator, real wages, the nominal 3-month Euribor interest rate and the unemployment rate. These variables are

⁵ A somewhat more detailed presentation of the log-linearized SWU model is available in the Online Appendix to McAdam and Warne (2019), while the Appendix to McAdam and Warne (2018) contains even further details, including the flexible price part of the model.

transformed exactly as in the above two studies. For some further details on the data, see Table B.1 in the Online Appendix.

One important difference for the additional vintages concerns the total investment data. Due to accounting issues regarding mainly multinational corporations in Ireland, but also for the Netherlands, some periods display extreme spikes. This is illustrated in Figure 1 for real total investment growth over the full sample discussed in Section 4.1. It is noteworthy that the numbers in the time periods 2015Q2–Q3, 2017Q2–Q3, and 2019Q2–Q4 are very big in an absolute sense compared with the overall volatility of the variable. Moreover, the growth rate is unusually high (low) in Q2 (Q3), consistent with accounting related data problems. Finally, these outliers appear since the 2018Q3 vintage and, hence, the number of affected vintages is fairly small.

Although the forecasting exercise does not attempt to predict investment growth, such outliers can influence the estimated parameters in a recursive exercise. In this paper, the outlier values have been replaced with the sample mean, but also the alternatives of keeping the data for those periods as well as treating the outlier periods as missing data have been examined. The choice between the mean-value replacement approach and the missing data approach is of lesser importance as the parameter estimates for the full sample discussed below are robust to these outlier corrections.⁶

4.1. FULL SAMPLE ESTIMATES AND MARGINAL LIKELIHOOD

The forecasting exercise in Section 5 uses annual revisions for the actual (true) values of the variables with the consequence that the last vintage used for prediction is 2019Q4. In this subsection the parameter estimates of the SWU model under RE and AL are discussed. The focus is on the full sample 1979Q4–2019Q4 using the latest update (2018) of the Area-Wide Model (AWM) database combined with the RTD vintage 2020Q3.⁷ The third quarter corresponds to the time period when the AWM database has typically been frozen with data until the end of the previous year and, hence, mimics an update from 2020.

The estimated SWU model in Smets et al. (2014) is based on the assumption of log utility in consumption and this variant is also used by McAdam and Warne (2019). This means that the inverse elasticity of intertemporal substitution for constant labor, denoted by σ_c , is equal to 1. When estimating the SWU model under AL this parameter is instead estimated and the prior distribution is given by $\sigma_c \sim N(1, 0.25^2)$ so that it is centered around unity. Furthermore, and as in McAdam and Warne (2019), the MA parameters of the price and wage markup shocks are set to 0 under RE as well as under AL. Finally, and as mentioned above, the baseline RE model uses the flexible-price

⁶ Estimating the AL baseline model with the outliers has an impact on investment related parameters. Both the elasticity of the capital adjustment and of the capital utilization cost functions drop as does the estimated capital share. Furthermore, the estimated persistence parameter of the investment-specific technology shock falls, while the estimated shock standard deviation increases. For the RE baseline version the use of the outliers have a somewhat smaller effect on the two elasticities, while the impacts on the capital share and investment-technology shock parameters are very present. In addition, the steady-state growth rate drops for the RE baseline model.

⁷ The observed variables are plotted in the Online Appendix, Figure C.1, where total investment growth is plotted with the sample mean correction for the selected outlier periods.

output gap and is therefore the same specification as in McAdam and Warne (2019), while under AL the output gap is defined as in Slobodyan and Wouters (2012a).⁸

Concerning the AL baseline model, the scale parameters are calibrated to $\sigma_r = 0.03$ and $\sigma_\varepsilon = 0.003$, as in Slobodyan and Wouters (2012a), while the prior distribution of the autoregressive parameter ρ is a standard beta with mean 0.25 and standard deviation 0.1 in the baseline version.⁹ This informative prior implies that the belief coefficients return fairly quickly to their steady-state values. Moreover, the forward looking variables are eight and given by consumption, investment, inflation, hours worked, the value of the capital stock, the rental rate of capital, real wages and employment. The PLM for each of these variables is equal to a constant and two own lags with time-varying parameters. Finally, the autoregressive parameter of the price markup shock process is calibrated to zero, such that this process is i.i.d. in the baseline model.

A selection of parameter estimates of the baseline models is shown in Table 1, which also provides some details on the estimation procedure.¹⁰ The focus is on the wage and price markup shock related parameters, but also the inverse elasticity of intertemporal substitution (σ_c) and the autoregressive parameter of the belief coefficients (ρ) are included. The baseline model under RE is shown in the first column and the baseline model under AL in the second column. Concerning the inverse elasticity of intertemporal substitution, the posterior mean under AL is somewhat greater than 1 and a 90 percent equal-tails credible region lies quite tightly around the posterior mean and above unity.

Regarding the wage related parameters, the indexation and stickiness parameters are lower under AL than under RE, while the wage markup shock persistence is higher. The price indexation and stickiness parameters are almost equal under AL and RE, while the price markup persistence is quite low under RE and nearly equal to the belief coefficient persistence under AL. These parameters are very different from an economic viewpoint, but seem to serve a similar purpose from a statistical perspective. In Slobodyan and Wouters (2012a), the baseline AL version has zero persistence for both the price and wage markup shocks and their estimate of ρ is around 0.96. From Table 1, the estimate of the baseline model is around 0.17, while the 90 percent credible region lies far from the estimate in Slobodyan and Wouters (2012a).

In terms of log marginal likelihood, the AL model obtains a much greater value with a difference relative to the RE model of about 22 log-units. The question is then if some of the specification differences between the RE and AL models can account for the large posterior odds ratio? In column three of Table 1, the RE model with the output gap specification from the AL model is shown. For this model the wage markup shock persistence is close to the AL model, while wage stickiness is now somewhat lower than for the AL model. However, the log marginal likelihood is substantially lower than for the baseline RE model. Hence, the within-sample improvement in fit of the AL model relative to the RE model is not explained by having a “better” definition of the output gap.

⁸ The prior distributions of all parameters are shown in the Online Appendix, Tables B.3–B.4.

⁹ For comparisons, keep in mind that a standard uniform distribution is equivalent to a standard beta distribution with mean 1/2 and standard deviation $1/\sqrt{12} \approx 0.28868$.

¹⁰ All parameter estimates are available in the Online Appendix, Tables B.5–B.6.

Turn next to the RE case when the inverse elasticity of intertemporal substitution, σ_c , is estimated using the same prior as for the AL model. In column four of Table 1, the posterior mean of this parameter is somewhat greater than unity, while the other parameters are essentially unchanged relative to the baseline RE model. There is some improvement in the log marginal likelihood but not on the scale of the baseline AL. Hence, it is probably the expectations mechanism that accounts for the better within-sample fit under AL.

Given the informative prior on the common persistence parameter of the belief coefficients in the baseline AL model, the fifth column of Table 1 covers the case of a standard uniform prior of ρ , as in Slobodyan and Wouters (2012a). It is noteworthy that the posterior mean at 0.10 is lower than in the baseline model, but the width of the marginal posterior distribution is such that the posterior mean in the baseline AL model is located at around the 80th percentile. The other parameters in the Table for the two AL models are very close and their marginal likelihoods are roughly of the same magnitude. Hence, the selected prior for ρ in the baseline AL model does not affect the finding that the persistence of the belief coefficients is low in the euro area, especially when compared with Slobodyan and Wouters (2012a).¹¹

4.2. RECURSIVE ESTIMATES OF THE MARGINAL LIKELIHOOD

Based on the evidence in the previous section, there is empirical support for the view that learning improves the within-sample fit of DSGE models. This is in line with empirical studies on US data, such as Milani (2007) and Slobodyan and Wouters (2012a,b). In this subsection, the within-sample fit of the baseline SWU models under RE and AL are estimated on vintages from the RTD that are used in the forecasting exercises in Section 5. Following McAdam and Warne (2019), the models are re-estimated on an annual frequency using the Q1 vintage with a sample covering the vintages from 2001 to 2019. The selected frequency mimics, under the best of circumstances, how often structural models are re-estimated in practise at policy institutions.

The recursive estimates are displayed in Figure 2. In the upper panel posterior-mode-based goodness-of-fit estimates are shown, while in the lower panel two estimators of the log marginal likelihood, the modified harmonic mean and the Laplace approximation, are plotted. Turning first to the upper panel, the estimated log likelihood (solid lines), the log posterior kernel (dash-dotted lines) and the Laplace approximation of the log marginal likelihood (dashed lines) are plotted with red lines for the RE model and blue for the AL model.¹² It is noteworthy that the difference between the blue and the red lines for each statistica over the sample is visually very similar. This suggests that mainly the data are driving these results, while the influence of the prior is negligible. Moreover, the AL model always has a greater value and the gap to the RE model is slowly increasing over time.

¹¹ The log marginal likelihood for the case when $\rho = 0$ (the belief coefficients are white noise around the steady-state) is estimated at -423.79. The posterior parameter estimates are generally similar to the models where ρ is estimated with the exception of the wage markup shock persistence parameter, which is substantially lower for the $\rho = 0$ case with a posterior mean of 0.22.

¹² The inverse Hessian used for computing the Laplace approximation is based on a finite difference estimator available in the well-known software Dynare.

In the lower panel we find both the Laplace approximation of the log marginal likelihood from the upper panel with dashed lines, and the modified harmonic mean (MHM) estimator using a selection of the post-burnin sample of draws from the posterior distributions. It is striking how small the differences between the two estimators are, with mainly visible differences for the AL model. The finding from the full sample that the AL model fits the euro area data better than the RE model is therefore also supported by the recursive estimates from the real-time sample 2001–2019.

5. FORECASTING UNDER RE AND AL

In this Section four topics are covered. First, we examine the point forecasts of the RE and AL models using the recursive posterior mean paths. Second, we turn to the density forecast where two scoring rules are considered: (i) the log score, and (ii) the continuous rank probability score for univariate forecasts and the energy score for multivariate forecasts; see, for instance, Gneiting and Raftery (2007) for a survey on scoring rules and Gneiting and Katzfuss (2014) for a review on probabilistic forecasting. Third, we consider some sensitivity analysis with respect to the PLM by including some additional variables. While these topics cover forecast comparisons, the last topic is concerned with forecast evaluation using the probability integral transform (PIT) in the univariate and Box’s density ordinate transform (BOT) in the multivariate case.

The forecast sample involves a total of 76 vintages from 2001Q1 until 2019Q4, while annual revisions are used as actual (true) values for the variables, i.e., actual values in 2001Q1 are taken from vintage 2002Q1, and so on. The focus is on quarterly real GDP growth and quarterly GDP deflator inflation, both individually and for the density forecasts also jointly. Given the ragged edge of the real-time data, the analysis involves mainly nowcasts and one- to eight-quarter-ahead forecasts, but some vintages also allow for backcasts. Since the number of vintages with backcasts is quite low (3 for real GDP growth and 22 for inflation), formal tests are not applied for them.¹³

5.1. POINT FORECASTS

The point forecast is given by the mean of the predictive distribution and is estimated by averaging the point forecast conditional on the parameters over the posterior draws; see, e.g., McAdam and Warne (2019, Section 5.2). These calculations have been performed using 10,000 equally spaced posterior draws from the 500,000 post burn-in draws of the random-walk Metropolis sampler, where the size of the burn-in sample is 250,000 draws. The resulting “spaghetti” plots are displayed in Figure 3 for real GDP growth in the upper panel and GDP deflator inflation in the lower panel. The forecasts from the baseline RE model are plotted to the left with red solid lines and the forecasts from the baseline AL model to the right with blue solid lines.¹⁴ The actual values are given by the black solid line, while the black dashed line provides the recursive posterior mean estimates of

¹³ The ragged edge of the real-time data is displayed in Table B.2 of the Online Appendix.

¹⁴ It should be noted that the forecasts for the AL model are computed under the assumption that the belief coefficients are held fixed over the forecast horizon. This is common in the learning literature; see, e.g., Eusepi and Preston (2018b) for discussions on this and related matters. The dynamic responses to an impulse/shock in any given time period are also computed under the assumption of fixed beliefs while updating of the beliefs occurs once the time period for the impulse moves forward one period.

steady-state real GDP growth and GDP deflator inflation, respectively. Similarly, the mean forecast errors (actual values minus prediction) and the Diebold-Mariano tests for equal mean square forecast errors (MSFE) are located in Table 2.¹⁵

Turning first to the real GDP growth forecasts in the upper panel of Figure 3, the forecasts from the AL model are overall flatter and are typically greater than the ones from the RE model. This is confirmed in Table 2 where the mean errors are negative and bigger in absolute terms for the AL model. In addition, the difference in the mean errors is overall increasing with the horizon. This is further supported by the Diebold-Mariano tests, where the small cdf-values for the longer horizons indicate that the RE model forecasts real GDP growth better than the AL model.

The point forecasts of inflation in the lower panel of Figure 3 visualizes the path contrasts between the RE and AL models. While the former are typically strongly upward-sloping, concave and tend to over-predict inflation, the latter paths are often u-shaped and fairly close to the actual values for the outer quarters of the horizon. This visual impression is confirmed by the results in Table 2 where the RE model has negative mean errors which increase in absolute terms with the forecast horizon. By contrast, the AL model under-predicts inflation but the mean errors are approaching zero as the horizon increases. The Diebold-Mariano tests support the view that the point forecasts from the AL model are better (worse) than those of the RE model for the outer (inner) quarters of the forecast horizon.¹⁶

On balance, however, the point forecasts give more support for the RE model than the AL model. The evidence is not clear-cut, but especially the real GDP growth mean forecast errors are big for the AL model and frequently more than 50 percent greater than those from the RE model. Given the typically flatter forecast paths from the AL model, it is tempting to consider higher persistence of the latter model as a possible explanation for the outcome.

In fact, the dynamic responses to the underlying shocks can be highly dissimilar under RE and AL. In Figure 4 the impulse responses of quarterly real GDP growth and inflation to a monetary policy shock are shown for the full sample posterior mode estimates of the baseline AL model. As a comparison, the responses under RE are also depicted using the posterior mode estimates from the AL model to avoid any influence from using different parameter estimates.¹⁷ The responses in real GDP growth are plotted in the upper panel and those of inflation in the lower. Since the ALM is time-varying with the belief coefficients, the responses to a one standard deviation shock change over time, but for real GDP growth the changes are not substantial and overall the response curve for each quarter is quite flat. Under RE, the response is initially much stronger, but also returns more quickly to the steady-state than for the AL baseline model.

¹⁵ The prediction errors for the individual horizons of real GDP growth and inflation are plotted in the Online Appendix, Figures C.2–C.3.

¹⁶ See Diebold (2015) for discussions on the use and abuse of the Diebold-Mariano test and more generally on the usefulness of out-of-sample forecast comparisons.

¹⁷ The overall dynamic pattern of the impulse responses under RE does not change much from the choice of estimates. For instance, the real GDP impulse responses in the baseline RE model at its posterior mode estimates are approximately the same, while the inflation responses have the same shape as for RE in Figure 4, but the initial response is much weaker at around -0.015 .

Turning to inflation, the pattern under AL is a gradual drop from a one standard deviation monetary policy shock before it begins to return to steady-state after about 20 quarters. The RE case is again a stronger early response of inflation and a faster convergence to steady-state. Overall, these impulse responses portray that the AL model has greater persistence than the RE model, a finding already supported by the point forecasts paths.

5.2. THE LOG SCORE

Density forecasts are typically compared across models using a scoring rule and below two such rules are utilized, focusing on either a single point of the predictive distribution or the full distribution. Technically, a scoring rule is said to be *proper* if a forecaster who maximizes the expected score provides the true subjective predictive distribution, and it is said to be *local* if the rule only depends on the predictive density at the realized value of the predicted variables. In this section the log score is used and, being the only proper local scoring rule, is equal to the sum (over the vintages in the comparison sample) of the log predictive likelihood evaluated at the actual value. This scoring rule is estimated using the approach described by, for example, Warne et al. (2017) and McAdam and Warne (2019), i.e., by averaging the predictive likelihood conditional on the parameter vector over the posterior draws. Section 5.3 discusses the empirical evidence using a proper scoring rule which also gives weight to other values of the predictive distribution than at the actual value. This aspect may be especially relevant in a real-time setting where the measured value of a variable is typically subject to revision.

The log scores for all 76 vintages in the forecast sample 2001Q1–2019Q4 are shown in Table 3 along with cdf-values of the Amisano and Giacomini (2007) weighted LR test and the Diebold-Mariano test of the hypothesis that the log scores of the RE and AL models are equal. The evidence from the density forecasts of the individual variables is broadly in line with the point forecast results, albeit with less extreme cdf-values for the real GDP growth tests. The joint real GDP and inflation density forecasts strongly support the RE baseline model up to the one-year-ahead forecasts, and weakly for the longer term forecasts. The first finding appears to be related to the inflation forecasts under RE which obtain a much larger log score than under AL. The real GDP growth up to the one-year horizon are similar under the two expectation formations. For the longer-term forecasts, the real GDP growth log scores under RE are larger than those under AL, while the inflation log scores are smaller under RE than AL. The joint log scores are consistent with these marginal log scores.

It is well known for a single database (vintage) that the log marginal likelihood over a sample is equal to the sum of the one-step-ahead log predictive likelihoods over the same sample. It may therefore seem surprising that the AL model with consistently higher marginal likelihoods has lower log scores than the RE model over the forecast sample. In the case of multiple databases, this result does not hold as the measured values of the observed variables are revised. Second, even without revisions, the ragged edge of the real-time database also breaks the relationship between these two likelihood concepts for one-step-ahead forecasts. Third, the likelihood relationship is only valid when the same variables are compared. In the SWU model case, the predictive likelihoods concern at most two of the observed variables, while the marginal likelihoods cover all eight observed

variables. Hence, even if we ignore the real-time dimension and only treat the issue in terms of a single database, the likelihood relationship implies that we would need to include the predictive likelihood of the remaining six (or seven) variables conditional on the actual value of real GDP growth and/or inflation for the equality to hold for each one of the two models at the one-step-ahead horizon.¹⁸ It is therefore not contradictory to find a reversal of ranking between the models when studying both within-sample and out-of-sample likelihood-based evidence.

The recursive estimates of the average log scores of the joint density forecasts are displayed in Figure 5; the case “Alt. PLM” with green dashed lines is discussed in Section 5.4. For the nowcast and up to four-quarter-ahead forecasts, the results from the full sample are confirmed. The ranking of the models is overall unchanged, especially since the onset of the Great Recession in 2008Q4. Turning to the longer-term joint density forecasts, the AL model appears to be better until 2008Q4, when the RE model takes the lead and then in 2016Q1 the average log score of the RE model decreases temporarily while that of the AL model increases.

The fall in log score of the RE model in 2016Q1 is also matched by the inflation density forecasts, but not by the real GDP forecasts; see Figures C.4–C.5 in the Online Appendix. Moreover, the point forecasts are the main reason for the performance fall of the RE model with big negative mean errors. This can also be seen in Figure 3, where actual inflation falls at least temporarily in early 2016 while the longer-term point forecasts for the RE model appear to increase, a pattern which is in sharp contrast to the point forecasts for the AL model. Moreover, the predictive inflation variance of the RE model is lower than the variance of the AL model, which suggests that the RE model is punished also by its lower forecast uncertainty.

As a possible explanation, it may first of all be kept in mind that the short-term nominal interest rate fell below zero in 2015Q2 and has remained negative throughout the remainder of the sample. Second, the models do not take the effective lower bound (ELB) on the short-term nominal interest rate into account when forecasting. It is conceivable that the less persistent RE model is more sensitive to the very low interest rates than the AL model and, hence, more prone to forecast higher inflation when the short-term rate is far from its steady-state value.

The forecasts of inflation and the short-term nominal interest rate from the 2014Q4 vintage are displayed in Figure 6, along with equal-tails credible intervals at 70 and 90 percent. In the upper left chart, the inflation forecast of the RE model are plotted as a red solid line along with its grey credible intervals. The forecast from the AL model are given by the blue dash-dotted line, 70 and 90 percent equal-tails credible intervals are portrayed by the region between the dashed and solid blue lines, respectively, whereas the black solid line with x-markers for each quarter gives the actual values of inflation. Notice that for the five-quarter to seven-quarter-ahead forecasts, the actual values lie at the 5:th percentile of the predictive distribution from the RE model and close to the predictive

¹⁸ It should also be kept in mind that the samples over which the log marginal likelihood and the log score are computed differ. However, the evidence in Figure 2 shows that the difference in log marginal likelihoods between the AL and RE models is increasing since 2001. This is consistent with a log score for one-step-ahead forecasts of all eight observed variables being higher for the AL model than for the RE model. Density forecast evidence on the remaining six observed variables is provided in Table B.8 of the Online Appendix, focusing on one of the scoring rules discussed in Section 5.3.

mean forecasts of the AL model. Overall, the RE model over-predicts inflation substantially and especially in the quarters of 2016, while the AL model slightly over-predicts in 2016. In terms of the log predictive likelihood, the RE model suffers considerably in 2016 from both its poor point forecasts and its comparatively narrow predictive distribution, while the AL model performs much better.¹⁹

Turning to the forecasts of the short-term nominal interest rate in the upper right corner, the RE model's point forecasts are firmly upward sloping and the predicted value in 2016Q4 is equal to 1.9%, while the AL model forecasts are moderately upward sloping with a predicted value of 0.6% in 2016Q4. Hence, the AL model predicts stronger monetary accommodation than the RE model and, *ceteris paribus*, this seems contradictory since lower interest rates are expected to boost inflation. However, the dynamic responses in the AL model are more persistent than in the RE model with the effect that the predictions are expected to be smoother and more drawn-out in the former model.

A simple and approximate approach to investigate the effects on the predictive distribution from imposing the ELB is to condition the forecasts of the model on the actual values of the short-term nominal interest rate over the forecast horizon 2015Q1–2016Q4, i.e., on the black solid line with x-marks in the upper right chart. Turning to the lower left chart, the forecasts of inflation under RE are compared with the conditional point and interval forecasts from the same model, where the latter point forecasts are plotted as the green dash-dotted line. The conditional forecasts are computed with the Waggoner and Zha (1999) approach and the 70 and 90 percent equal-tails credible intervals from the predictive distribution are represented by the region between the dashed and solid green lines, respectively. For this case, the actual values of inflation lie around the 15:th percentile rather than the 5:th percentile of the relevant predictive distribution. In addition, the conditional point forecasts of inflation are well below the unconditional point forecasts (red solid line) and closer to the actual values.

The relevance of the ELB for the predictive distributions may also be studied by dividing the forecast sample into two parts, where the first covers the period until 2014Q4 and the second sample the remaining 20 quarters with nominal interest rates close to or at the ELB. The log scores and test statistics from these sub-samples are shown in Table 4, where the evidence on the first sub-sample is widely consistent with the results from the full sample.

Turning to the ELB sub-sample since 2015, it is interesting to note that the RE model seems to predict real GDP growth better than the AL model when compared with the first sub-sample, also at the shorter horizons. Concerning inflation, the evidence in favor of the AL model is now also supported for the shorter horizons, while the joint scores are consistent with the marginal scores and do not favor any model.

¹⁹ The estimated log predictive likelihood values of the RE (AL) model in 2016 are for each quarter respectively -0.726 (0.012), -1.108 (-0.104), -0.942 (-0.131) and -0.243 (-0.140). The numerical standard errors of these estimates are very small and around 0.01 for the RE model and less than half of that for the AL model.

5.3. CRPS AND THE ENERGY SCORE

The log score uses only information about the predictive density at the actual value while, for instance, values near it and that have a high likelihood are ignored. Scoring rules based on the predictive cdf allows for a wide range of possible values and therefore provide a more comprehensive measure of density forecast performance than the log score. Furthermore, it is shown by Krüger, Lerch, Thorarinsdottir, and Gneiting (2021) that the theoretical conditions for the log score to be consistent when using MCMC draws of the parameters are stronger than for some alternative scoring rules. For univariate forecasts, they show that the *continuous ranked probability score* (CRPS) is consistent under weaker conditions. In addition, the CRPS is proper and, under certain conditions, even strictly proper; see Gneiting and Raftery (2007) for details. This scoring rule can be expressed as:

$$\text{CRPS}(h, i) = \frac{1}{2} E_{PD} \left[\left| y_{i,T+h} - y_{i,T+h}^{(c)} \right| \right] - E_{PD} \left[\left| y_{i,T+h} - y_{i,T+h}^{(o)} \right| \right], \quad i = 1, \dots, n_p, \quad (5)$$

where n_p is the number of predicted variables, $y_{i,T+h}$ and $y_{i,T+h}^{(c)}$ are independent copies of the predicted variable i in time period $T+h$ given the information at T with (posterior) predictive distribution PD , and with actual values $y_{i,T+h}^{(o)}$ for the h -quarter-ahead forecasts. For a given vintage T , this scoring rule can be consistently estimated using P simulated forecasts paths based on the posterior draws of the parameters by

$$\text{CRPS}(h, i) = \frac{1}{2P^2} \sum_{j_1=1}^P \sum_{j_2=1}^P \left| y_{i,T+h}^{(j_1)} - y_{i,T+h}^{(j_2)} \right| - \frac{1}{P} \sum_{j=1}^P \left| y_{i,T+h}^{(j)} - y_{i,T+h}^{(o)} \right|, \quad i = 1, \dots, n_p, \quad (6)$$

where $y_{i,T+h}^{(j)}$ is the value of variable i in simulated path j at $T+h$. For the distributional forecast comparisons, the CRPS is given by the sum of (6) over the 76 vintages at hand.²⁰

A multivariate extension of the CRPS, called the *energy score*, was introduced by Gneiting and Raftery (2007) and applied by, e.g., Gneiting, Stanberry, Gritmit, Held, and Johnson (2008). It can be expressed as

$$\text{ES}(h) = \frac{1}{2} E_{PD} \left[\left\| y_{1:n_p,T+h} - y_{1:n_p,T+h}^{(c)} \right\| \right] - E_{PD} \left[\left\| y_{1:n_p,T+h} - y_{1:n_p,T+h}^{(o)} \right\| \right], \quad (7)$$

where $\|x\| = \sqrt{x'x}$ is the Euclidean norm, and $y_{1:n_p,T+h}$ is a vector with the n_p predicted variables. This expression can be consistently estimated using P paths from the predictive distribution as

$$\text{ES}(h) = \frac{1}{2P^2} \sum_{j_1=1}^P \sum_{j_2=1}^P \left\| y_{1:n_p,T+h}^{(j_1)} - y_{1:n_p,T+h}^{(j_2)} \right\| - \frac{1}{P} \sum_{j=1}^P \left\| y_{1:n_p,T+h}^{(j)} - y_{1:n_p,T+h}^{(o)} \right\|. \quad (8)$$

²⁰ If the predictive distribution is normal, the CRPS has, as noted by Gneiting, Raftery, Westveld III, and Goldman (2005), a closed-form expression and can be computed directly from the first two moments of the distribution; see also Jordan, Krüger, and Lerch (2019) for closed-form expressions of the CRPS for a wide range of distributions. With $\varphi(x)$ being the standard normal pdf and $\Phi(x)$ the cdf, the CRPS is now given by:

$$\begin{aligned} \text{CRPS}(h, i) = & -\sigma_{y_{i,T+h}} \left(\frac{y_{i,T+h}^{(o)} - \mu_{y_{i,T+h}}}{\sigma_{y_{i,T+h}}} \left[2\Phi \left(\frac{y_{i,T+h}^{(o)} - \mu_{y_{i,T+h}}}{\sigma_{y_{i,T+h}}} \right) - 1 \right] \right. \\ & \left. + 2\varphi \left(\frac{y_{i,T+h}^{(o)} - \mu_{y_{i,T+h}}}{\sigma_{y_{i,T+h}}} \right) - \frac{1}{\sqrt{\pi}} \right), \end{aligned}$$

where $\mu_{y_{i,T+h}}$ and $\sigma_{y_{i,T+h}}$ are the predictive mean and standard deviation, respectively, of $y_{i,T+h}$ for $i = 1, \dots, n$.

It can be noted that the CRPS for a deterministic forecast system reduces to minus the absolute point forecast error. Hence, the CRPS (and ES) encompasses both probabilistic and deterministic scoring rules.

Distributional forecasts can be simulated with DSGE models using the algorithm proposed by Adolfson et al. (2007) and which is an extension of the Thompson and Miller (1986) approach. The nowcasts and backcasts are not included in the exercise below but may be simulated through an algorithm based on, for instance, the Waggoner and Zha (1999) conditioning approach for those time periods and where the conditioning variables are given by the vintage data for the observed variables. The precise choice of conditioning methodology influences the CRPS and ES as it has a direct impact on the distribution of the shocks over the conditioning period. Below we instead simulate shocks from period $T + 1$ and initial state variables for period T using the Kalman filter estimates which take the ragged edge into account.

For the RE and AL baseline models a total of 100,000 prediction paths per vintage and model have been computed using 1,000 parameter draws of the available 500,000 post-burnin draws and 100 path draws per parameter value based on the distributions of the structural shocks of the model and the state variables at T . The estimated CRPS and ES for the full sample are shown in Table 5 for the individual variables as well as the joint case.

The RE baseline model again obtains the highest score for the real GDP forecasts and its inflation forecast scores are greater than those from the AL model up to the four-quarter-ahead horizon. For the longer horizon inflation forecasts the AL model again has the largest scores. Turning to the joint forecasts, the RE model has a higher score for all horizons, but the gap to the AL model decreases for the longer horizons, also taking the number of observations into account.

The recursively estimated average ES of the models are shown in Figure 7. For the shorter-term forecasts it can be seen that the difference in score is similar over the sample, while the Great Recession appears to strengthen the first rank of the RE model over the longer-term forecasts. The influence of inflation in 2016 can be seen also for the recursive ES, albeit that this score appears to be less affected than the log score. It may also be noted that the impact of the inflation forecasts since 2016 are strong on the CRPS for the RE model.²¹

Overall, the forecast comparisons using point and density forecasts agree that the RE model forecasts real GDP growth better than the AL model and especially for the outer quarters of the forecast horizon, while the performance is mixed for inflation with RE winning for the shorter-term and AL for the longer term. In addition, the RE model obtains larger scores than the AL model for the joint forecasts.

5.4. ALTERNATIVE PLMS

The PLM for the adaptive learning case has so far been assumed to be an AR(2) process with a constant term for each forward looking variable. This simple PLM was suggested by Slobodyan and

²¹ The recursive average CRPS values for real GDP growth and inflation, respectively, are shown in Figures C.6–C.7 of the Online Appendix. In addition, Table B.8 provides the CRPS scores and Diebold-Mariano test p -values for the remaining six observed variables.

Wouters (2012a), but given the relatively poor forecasting performance of the AL model compared with the RE version an important reason for this failure may be the assumed PLM.

The theoretical learning literature provides some guidance on the selection of a PLM, such as using a forecasting model based on the RE solution form in (2) where the agents know the functional form for the forward looking variables but not the belief coefficients; see, e.g., Hommes and Zhu (2014) and Evans, Evans, and McGough (2022). This RE-based form of AL includes the time-varying specification of the belief coefficients and having a constant term which underlies the ALM in (4). When the functional form of the PLM is correctly specified, convergence of AL to a RE equilibrium can occur.

From a time series perspective we may instead consider, for example, VARs of the forward looking variables. Such PLMs are, like the benchmark PLM, misspecified since some variables which are typically present in the RE solution equations of the forward looking variables are themselves not forward looking. In addition, it can be very cumbersome from a computational perspective to estimate DSGE models with many belief coefficients and, as noted by Slobodyan and Wouters (2012a), VARs are therefore likely to suffer from overfitting. The RE-based functional form of the PLM is also likely to involve overfitting as each equation can have many belief coefficients. In the case of the baseline SWU model for AL with 8 forward looking variables, each equation of the PLM has 17 unknown belief coefficients when using the functional form of the RE solution. This amounts to 136 belief coefficients in total compared with 24 for the benchmark PLM. It may therefore be useful to investigate the economic model for simpler suggestions than the full-blown functional form of the RE solution.

The alternative PLM considered in this section uses the AR(2) process with a constant term as a basis and adds at most one variable to each one of the eight processes and limited to the first lag. Specifically, for real private consumption the lag of the nominal interest rate is included, for real investment the lag of the real value of the capital stock is used, inflation is linked to the lag of unemployment, hours worked makes use of the lag of employment, the rental rate of capital considers the lag of capital services used in production as potentially informative, real wages to the lag of inflation and, finally, employment to the lag of real wages. The only process for a forward looking variable which remains unchanged is the real value of the capital stock.²²

For the results on the average log scores of the joint real GDP growth and inflation case we return to Figure 5. Generally, the alternative PLM appears to provide some improvement over the benchmark PLM, especially for the nowcasts and the one-quarter-ahead forecasts. While the average log scores of the two- and three-quarter-ahead forecasts are similar, the benchmark PLM tends to forecast better until the Great Recession for the four-quarter-ahead to eight-quarter-ahead forecasts, while the alternative PLM overtakes the benchmark PLM thereafter and up to the last periods of the forecast sample. Still and with the exception of the nowcasts, the alternative PLM does not

²² Note that this alternative PLM is similar to one of the alternatives discussed in Slobodyan and Wouters (2012a, Section V.A). The structural equations for the forward looking variables provide the inspiration for this PLM; see, for instance, Smets et al. (2014, Section 2).

alter the main findings regarding log score comparisons with the RE and AL models for the selected sample.

Turning to the average energy scores in Figure 7, the alternative PLM is competitive with the RE model for the one-quarter-ahead forecasts and better than the benchmark PLM. For the other forecast horizons the alternative PLM remains competitive with the RE model until around 2012–13, where it gradually loses ground and approaches the score of the benchmark PLM. Hence, the alternative PLM can in part improve the density forecasts relative to the benchmark PLM, but not to the extent that it implies an AL model which forecasts better than the RE model.

5.5. PITs AND BOTs

The *probability integral transform* (PIT) has long been used to assess if a forecasting model is well calibrated. An early paper which considered this idea for density forecasting purposes in econometrics is Diebold, Gunther, and Tay (1998), but it has earlier been emphasized by Dawid (1984). Rosenblatt (1952) shows that for a correctly specified model

$$\pi_{i,T+1|T} = F(y_{i,T+1}|\mathcal{Y}_T), \quad i = 1, \dots, n_p,$$

is independent and uniformly distributed on the unit interval, where $F(\cdot)$ is the cdf and \mathcal{Y}_T is the information set available at T . Smith (1985) further noted that $z_{i,T+1|T} = \Phi^{-1}(\pi_{i,T+1|T})$, where $\Phi(\cdot)$ is the cdf of the normal distribution, is i.i.d. $N(0, 1)$.

Amisano and Geweke (2017) construct a test statistic based on the normality property of the inverse cdf; details are available in their Online Appendix. Specifically, let

$$\pi_{i,T+1|T}^{(j)} = \Phi \left(y_{i,T+1}^{(o)} \left| \mu_{i,T+1|T}^{(j)}, \sigma_{ii,T+1|T}^{(j)} \right. \right), \quad j = 1, \dots, N$$

where $\mu_{i,T+1|T}^{(j)}$ is the one-quarter-ahead point forecast of predicted variable $i = 1, \dots, n_p$ using the j :th posterior draw of θ , the vector of estimated parameters, while $\sigma_{ii,T+1|T}^{(j)}$ is the one-quarter-ahead forecast standard deviation of variable i using $\theta^{(j)}$.

Next, the Monte Carlo average of the N values of the uniform variable is taken such that

$$\pi_{i,T+1|T} = \frac{1}{N} \sum_{j=1}^N \pi_{i,T+1|T}^{(j)}, \quad (9)$$

while

$$z_{i,T+1|T} = \Phi^{-1}(\pi_{i,T+1|T}). \quad (10)$$

Under the assumption that the model forecasts are well calibrated, the variable $z_{i,T+1|T}$ is normally distributed with zero mean and unit variance. This assumption is tested in Amisano and Geweke (2017) using the first q moments and p lags of the $z_{i,T+1|T}$ process with a test statistic which is asymptotically χ_{q+p}^2 .

The uniformity of the PIT does not hold for multivariate forecasts; see Genest and Rivest (2001). As pointed out by Gneiting et al. (2008), an option is then to utilize the *Box density ordinate*

transform (BOT). The BOT was, e.g., proposed by Box (1980), and is defined as

$$\pi_{T+1|T} = 1 - \Pr \left[p(y_{1:n_p, T+1} | \mathcal{Y}_T) \leq p(y_{1:n_p, T+1}^{(o)} | \mathcal{Y}_T) \right].$$

If $y_{1:n_p, T+1}$ is distributed as $p(\cdot)$ and this density is continuous, then $\pi_{T+1|T}$ is standard uniform. For example, if the density $p(\cdot)$ is Gaussian with mean vector μ and covariance matrix Σ , then

$$\pi_{T+1|T} = 1 - \chi_{n_p}^2 \left((y_{1:n_p, T+1}^{(o)} - \mu)' \Sigma^{-1} (y_{1:n_p, T+1}^{(o)} - \mu) \right). \quad (11)$$

Using the Kolmogorov-Smirnov test with the SWU model under RE for the euro area RTD McAdam and Warne (2019) finds that the empirical distribution of the predictive likelihood is not statistically different from a normal distribution with predictive mean and covariance matrix estimated with Monte Carlo integration and evaluated at the actual value. These tests have been recomputed with the extended real-time sample under the RE and the AL baseline models and the conclusion is confirmed also for these cases; see Table B.7 in the Online Appendix. In the current study, the BOT expression in (11) is therefore used to estimate $\pi_{T+1|T}$ for the multivariate case. Furthermore, the expression in (10) is applied to transform this variable to a standard normal and the test statistic in Amisano and Geweke (2017) is thereafter used.

The evidence from the tests is provided in Table 6 for the nowcasts and the one-quarter-ahead forecasts with $p = q = 2$. The 95:th percentile value of the χ_4^2 distribution is around 9.49 and for the RE model the test value for the nowcasts of inflation is far below this value, while the value for the one-quarter-ahead forecasts is closer and above the 90:th percentile value. All other test values are larger than the 95:th percentile value and the overall assessment suggest is that the predictions of the two models are not well calibrated.

Histograms of the underlying π estimates for the BOTs are shown in Figure 8 using ten bins. In the case of the RE model the frequency of occurrence for the higher value bins is considerably greater than for the lower value bins, while the AL model displays greater occurrences for some of the bins above 0.5. Hence both models tend to have too many realizations in the tails of the joint predictive distributions. One important source for this is likely to be that they overpredict real GDP growth.²³

6. SUMMARY AND CONCLUSIONS

This paper compares and evaluates real-time density forecasts of real GDP growth and inflation for the euro area using an unemployment-based extension of the well-known Smets and Wouters (2007) model, which supports either rational expectations (RE) or adaptive learning (AL), as modelled in Slobodyan and Wouters (2012a). The forecast comparison sample begins with the first real-time database vintage in 2001Q1 and ends in 2019Q4, thereby covering the period of wage moderation, the global financial crisis, the Great Recession that followed, the European sovereign debt crisis, and the period in which policy rates are close to the effective lower bound (ELB). The selection of actual values of the observed variables follows “good practise” in the real-time literature and therefore takes the annual revisions such that vintages until 2020Q4 are also required. This means that noisy

²³ Histograms of the π estimates for the PITs are provided in the Online Appendix, Figures C.8–C.9.

early estimates are avoided, such as first releases, but also that major changes to the computational methodology are less likely to influence the outcome of the forecast analysis, as may be expected for the euro area if the latest vintage would instead be used for the actual values; see, e.g., Croushore and Stark (2001) and Croushore (2006, 2011) for discussions on using real-time data.

Concerning within-sample fit, the AL model has a greater log marginal likelihood than the RE model by about 22 log-units for the full estimation sample, i.e., when both models are estimated until 2019Q4.²⁴ This confirms the findings from US data by, e.g., Milani (2007) and Slobodyan and Wouters (2012a) that DSGE models subject to AL fit the estimation data better than under RE. Furthermore, the two models are also compared within-sample for each Q1 vintage over the real-time forecast sample, 2001–2019. Again, the AL model obtains larger log marginal likelihood values for each such vintage.

Turning to the forecast comparison exercise, the RE model overall predicts real GDP growth more accurately than the AL model from a point forecast perspective. Both models over-predict real GDP growth on average and the AL model substantially for the four-quarter-ahead to eight-quarter-ahead forecasts. The inflation forecast paths reflect a sharp difference between the dynamic behavior of the two expectation mechanisms. The paths from the RE model are typically upward sloping with a concave curvature during the comparison sample and result in an over-prediction of inflation on average, while the AL model paths tend to be u-shaped and lead to an under-prediction of inflation, albeit with average forecast errors close to zero for the outer quarters. As a consequence, the Diebold-Mariano tests indicate that the RE model forecasts inflation better for the short-term forecasts, while the AL model predicts better for the outer quarters of the horizon.

The density forecasts are compared using the log score as well as the continuous ranked probability score (CRPS) and the corresponding multivariate energy score (ES). The log score measures the height of the predictive density at the actual values, while the CRPS/ES covers the full predictive distribution. Overall, these scoring rules agree with the results from the point forecast exercise and therefore lend support to the RE model having a better out-of-sample fit than the AL model. Since the within-sample and out-of-sample fit measures do not cover the same variables, the model ranking findings are not contradictory.

Finally, it should be stressed that the AL model used in this study represents only one approach to taking learning or incomplete information into account and that other approaches, such as constant gain learning or the learning about long-term drifts approach in Eusepi and Preston (2018a), may fare better when faced with the baseline RE model in a forecast comparison exercise. An important feature of the AL model is that its dynamics shows greater persistence than the RE model's and this may be one reason why the AL model overall has greater difficulties forecasting real GDP growth, especially during a period as challenging as the one examined in this paper. At the same time, the backward expectations formation under AL and the model's greater persistence may explain why it provides better within-sample fit than the RE model. We leave these interesting questions open for future research.

²⁴ In terms of the base-10 log-scale, 22.07 natural log-units corresponds to about 9.58 log-10 units.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the author upon reasonable request.

CONFLICT OF INTEREST STATEMENT

The author declares no conflict of interest.

FUNDING INFORMATION

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

REFERENCES

- Adolfson, M., Lindé, J., and Villani, M. (2007), “Forecasting Performance of an Open Economy DSGE Model,” *Econometric Reviews*, 26(2), 289–328.
- Amisano, G. and Geweke, J. (2017), “Prediction Using Several Macroeconomic Models,” *The Review of Economics and Statistics*, 99(5), 912–925.
- Amisano, G. and Giacomini, R. (2007), “Comparing Density Forecasts via Weighted Likelihood Ratio Tests,” *Journal of Business & Economic Statistics*, 25(2), 177–190.
- Bañbura, M., Giannone, D., and Reichlin, L. (2010), “Large Bayesian Vector Auto Regressions,” *Journal of Applied Econometrics*, 25(1), 71–92.
- Bernanke, B. S., Gertler, M., and Gilchrist, S. (1999), “The Financial Accelerator in a Quantitative Business Cycle Framework,” in J. B. Taylor and M. Woodford (Editors), *Handbook of Macroeconomics*, volume 1C, 1341–1393, North Holland, Amsterdam.
- Box, G. E. P. (1980), “Sampling and Bayes’ Inference in Scientific Modelling and Robustness,” *Journal of the Royal Statistical Society Series A*, 143(4), 383–430.
- Carvalo, C., Eusepi, S., Moench, E., and Preston, B. (2023), “Anchored Inflation Expectations,” *American Economic Journal: Macroeconomics*, 15(1), 1–47.
- Chari, V. V., Kehoe, P. J., and McGrattan, E. R. (2009), “New Keynesian Models: Not Yet Useful for policy Analysis,” *American Economic Journal: Macroeconomics*, 1(1), 242–266.
- Christiano, L. J., Eichenbaum, M. S., and Trabandt, M. (2018), “On DSGE Models,” *Journal of Economic Perspectives*, 32(3), 113–140.
- Christoffel, K., Coenen, G., and Warne, A. (2011), “Forecasting with DSGE Models,” in M. P. Clements and D. F. Hendry (Editors), *The Oxford Handbook of Economic Forecasting*, 89–127, Oxford University Press, New York.
- Cole, S. J. and Milani, F. (2019), “The Misspecification of Expectations in New Keynesian Models: A DSGE-VAR Approach,” *Macroeconomic Dynamics*, 23(3), 974–1007.
- Croushore, D. (2006), “Forecasting with Real-Time Macroeconomic Data,” in G. Elliot, C. W. J. Granger, and A. Timmermann (Editors), *Handbook of Economic Forecasting*, 961–982, North Holland, Amsterdam.
- Croushore, D. (2011), “Frontiers of Real-Time Data Analysis,” *Journal of Economic Literature*, 49(1), 72–110.
- Croushore, D. and Stark, T. (2001), “A Real-Time Data Set for Macroeconomists,” *Journal of Econometrics*, 105(1), 111–130.
- Dawid, A. P. (1984), “Statistical Theory: The Prequential Approach,” *Journal of the Royal Statistical Society, Series A*, 147(2), 278–292.
- Del Negro, M. and Schorfheide, F. (2013), “DSGE Model-Based Forecasting,” in G. Elliott and A. Timmermann (Editors), *Handbook of Economic Forecasting*, volume 2, 57–140, North Holland, Amsterdam.

- Diebold, F. X. (2015), “Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests,” *Journal of Business & Economic Statistics*, 33(1), 1–9, with discussion, pages 9–24.
- Diebold, F. X., Gunther, T. A., and Tay, A. S. (1998), “Evaluating Density Forecasts with Applications to Financial Risk Management,” *International Economic Review*, 39(4), 863–883.
- Elias, C. J. (2022), “Bayesian Estimation of a Small-Scale New Keynesian Model with Heterogeneous Expectations,” *Macroeconomic Dynamics*, 26(4), 920–944.
- Eusepi, S. and Preston, B. (2018a), “Fiscal Foundations of Inflation: Imperfect Knowledge,” *American Economic Review*, 108(9), 2551–2589.
- Eusepi, S. and Preston, B. (2018b), “The Science of Monetary Policy: An Imperfect Knowledge Perspective,” *Journal Of Economic Literature*, 56(1), 3–59.
- Evans, D., Evans, G. W., and McGough, B. (2022), “The RPEs of RBCs and other DSGEs,” *Journal of Economic Dynamic & Control*, 143, Article 104492.
- Evans, G. W. and Honkapohja, S. (2001), *Learning and Expectations in Macroeconomics*, Princeton University Press, Princeton.
- Evans, G. W. and Honkapohja, S. (2009), “Learning and Macroeconomics,” *Annual Review of Economics*, 1, 421–451.
- Evans, G. W. and McGough, B. (2020), “Adaptive Learning in Macroeconomics,” in *Oxford Research Encyclopedia of Economics and Finance*, Oxford University Press.
- Galí, J., Smets, F., and Wouters, R. (2012), “Unemployment in an Estimated New Keynesian Model,” in D. Acemoglu and M. Woodford (Editors), *NBER Macroeconomics Annual 2011*, 329–360, University of Chicago Press.
- Genest, C. and Rivest, L. P. (2001), “On The Multivariate Probability Integral Transformation,” *Statistics & Probability Letters*, 53(4), 391–399.
- Giannone, D., Henry, J., Lalik, M., and Modugno, M. (2012), “An Area-Wide Real-Time Database for the Euro Area,” *The Review of Economics and Statistics*, 94(4), 1000–1013.
- Gneiting, T. and Katzfuss, M. (2014), “Probabilistic Forecasting,” *The Annual Review of Statistics and Its Applications*, 1, 125–151.
- Gneiting, T. and Raftery, A. E. (2007), “Strictly Proper Scoring Rules, Prediction, and Estimation,” *Journal of the American Statistical Association*, 102(477), 359–378.
- Gneiting, T., Raftery, A. E., Westveld III, A. H., and Goldman, T. (2005), “Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation,” *Monthly Weather Review*, 133(5), 1098–1118.
- Gneiting, T., Stanberry, L. I., Gritmit, E. P., Held, L., and Johnson, N. A. (2008), “Assessing Probabilistic Forecasts of Multivariate Quantities, with an Application to Ensemble Predictions of Surface Winds,” *TEST*, 17(2), 211–235.

- Harvey, D., Leybourne, S., and Newbold, P. (1997), “Testing the Equality of Prediction Mean Squared Errors,” *International Journal of Forecasting*, 13, 281–291.
- Hommes, C. and Zhu, M. (2014), “Behavioral Learning Equilibria,” *Journal of Economic Theory*, 150, 778–814.
- Jordan, A., Krüger, F., and Lerch, S. (2019), “Evaluating Probabilistic Forecasts with scoringRules,” *Journal of Statistical Software*, 90(12), 1–37.
- Krüger, F., Lerch, S., Thorarinsdottir, T., and Gneiting, T. (2021), “Predictive Inference Based on Markov Chain Monte Carlo Output,” *International Statistical Review*, 89(2), 274–301.
- Lindé, J., Smets, F., and Wouters, R. (2016), “Challenges for Central Banks’ Macro Models,” in J. B. Taylor and H. Uhlig (Editors), *Handbook of Macroeconomics*, volume 2B, 2185–2262, North Holland, Amsterdam.
- Mankiw, N. G. and Reis, R. (2002), “Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve,” *Quarterly Journal of Economics*, 117(4), 1295–1328.
- McAdam, P. and Warne, A. (2018), “Euro Area Real-Time Density Forecasting with Financial or Labor Market Frictions,” ECB Working Paper Series No. 2140.
- McAdam, P. and Warne, A. (2019), “Euro Area Real-Time Density Forecasting with Financial or Labor Market Frictions,” *International Journal of Forecasting*, 35(2), 580–600.
- Milani, F. (2007), “Expectations, Learning and Macroeconomic Persistence,” *Journal Of Monetary Economics*, 54(7), 2065–2082.
- Milani, F. (2009), “Adaptive Learning and Macroeconomic Inertia in the Euro Area,” *Journal of Common Market Studies*, 47(3), 579–599.
- Orphanides, A. and Williams, J. C. (2005), “Inflation Scares and Forecast-Based Monetary Policy,” *Review of Economic Dynamics*, 8(2), 498–527.
- Rosenblatt, M. (1952), “Remarks on a Multivariate Transformation,” *The Annals Of Mathematical Statistics*, 23(3), 470–472.
- Sargent, T. J. (1993), *Bounded Rationality in Macroeconomics*, Clarendon Press, Oxford.
- Sims, C. S. (2003), “Implications of Rational Inattention,” *Journal of Monetary Economics*, 50(3), 665–690.
- Slobodyan, S. and Wouters, R. (2012a), “Learning in a Medium-Scale DSGE Model with Expectations Based on Small Forecasting Models,” *American Economic Journal: Macroeconomics*, 4(2), 65–101.
- Slobodyan, S. and Wouters, R. (2012b), “Learning in an Estimated Medium-Scale DSGE Model,” *Journal of Economic Dynamics and Control*, 36(1), 26–46.
- Smets, F., Warne, A., and Wouters, R. (2014), “Professional Forecasters and Real-Time Forecasting with a DSGE Model,” *International Journal of Forecasting*, 30(4), 981–995.
- Smets, F. and Wouters, R. (2007), “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach,” *American Economic Review*, 97(3), 586–606.

- Smith, J. Q. (1985), “Diagnostic Checks of Non-Standard Time Series Models,” *Journal Of Forecasting*, 4(3), 283–291.
- Svensson, L. E. O. and Woodford, M. (2003), “Indicator Variables for Optimal Policy,” *Journal of Monetary Economics*, 50(3), 691–720.
- Thompson, P. A. and Miller, R. B. (1986), “Sampling the Future: A Bayesian Approach to Forecasting from Univariate Time Series Models,” *Journal of Business & Economic Statistics*, 4(4), 427–436.
- Waggoner, D. F. and Zha, T. (1999), “Conditional Forecasts in Dynamic Multivariate Models,” *Review of Economics and Statistics*, 81(4), 639–651.
- Warne, A., Coenen, G., and Christoffel, K. (2017), “Marginalized Predictive Likelihood Comparisons of Linear Gaussian State-Space Models with Applications to DSGE, DSGE-VAR and VAR Models,” *Journal of Applied Econometrics*, 32(1), 103–119.
- Woodford, M. (2013), “Macroeconomic Analysis without the Rational Expectations Hypothesis,” *Annual Review of Economics*, 5, 303–246.

TABLE 1: Selected parameter estimates of the SWU model of the euro area under rational expectations (RE) and adaptive learning (AL) for the sample 1985Q1–2019Q4.

	RE baseline	AL baseline	RE alt: gap	RE alt: σ_c	AL alt: ρ
subst. elasticity (σ_c)	1 [calibrated]	1.07 [1.04, 1.09]	1 [calibrated]	1.14 [1.05, 1.27]	1.07 [1.04, 1.09]
belief persistence (ρ)	–	0.17 [0.07, 0.28]	–	–	0.10 [0.01, 0.25]
Wage markup (AR)	0.73 [0.56, 0.86]	0.85 [0.80, 0.88]	0.88 [0.74, 0.96]	0.77 [0.63, 0.88]	0.84 [0.80, 0.87]
Wage indexation	0.25 [0.12, 0.40]	0.19 [0.10, 0.29]	0.25 [0.12, 0.40]	0.24 [0.12, 0.39]	0.19 [0.10, 0.30]
Wage stickiness	0.62 [0.54, 0.70]	0.54 [0.48, 0.61]	0.48 [0.40, 0.59]	0.59 [0.52, 0.67]	0.55 [0.49, 0.62]
Price markup (AR)	0.19 [0.05, 0.35]	0 [calibrated]	0.42 [0.18, 0.59]	0.20 [0.07, 0.36]	0 [calibrated]
Price indexation	0.22 [0.09, 0.35]	0.23 [0.14, 0.33]	0.15 [0.06, 0.32]	0.21 [0.09, 0.34]	0.23 [0.15, 0.33]
Price stickiness	0.80 [0.76, 0.84]	0.80 [0.76, 0.84]	0.72 [0.64, 0.80]	0.79 [0.74, 0.84]	0.80 [0.76, 0.84]
Log marg. like.	–443.08	–421.01	–462.81	–441.53	–421.46

NOTES: Posterior mean estimates are reported along with the 5th and 95th percentiles from the posterior distribution in brackets. The models are estimated with the random-walk Metropolis algorithm using 750,000 posterior draws, where the first 250,000 are treated as burn-in sample. The data from 1980Q1–1984Q4 are used as a training sample for the Kalman filter state variables. The log marginal likelihood is estimated with the modified harmonic mean estimator. The gap alternative model under RE uses the output gap specification from AL, while the σ_c alternative model under RE estimates the inverse elasticity of intertemporal substitution for constant labor σ_c using the prior $\sigma_c \sim N(1, 0.25^2)$. The ρ alternative model under AL uses a standard uniform prior distribution for the belief coefficients persistence parameter.

TABLE 2: Mean forecast errors based on the posterior predictive mean as the point forecast and p -values for the Diebold-Mariano test of equal mean squared forecast errors for the sample 2001Q1–2019Q4.

h	Real GDP growth			Inflation		
	Mean errors		cdf-values	Mean errors		cdf-values
	RE	AL	DM	RE	AL	DM
-1	-0.218	-0.219	–	0.087	0.140	–
0	-0.250	-0.348	0.38	0.019	0.123	0.00
1	-0.357	-0.450	0.01	-0.028	0.146	0.00
2	-0.388	-0.478	0.15	-0.072	0.152	0.02
3	-0.380	-0.483	0.16	-0.112	0.144	0.11
4	-0.346	-0.474	0.03	-0.153	0.124	0.29
5	-0.311	-0.470	0.00	-0.187	0.100	0.67
6	-0.276	-0.465	0.00	-0.214	0.076	0.90
7	-0.241	-0.458	0.00	-0.240	0.046	0.99
8	-0.206	-0.446	0.00	-0.260	0.016	1.00

NOTES: The cdf-values from the Diebold-Mariano (DM) test for the null hypothesis of equal mean squared forecast errors (MSFE) are calculated as in Harvey, Leybourne, and Newbold (1997), equation (9). The cdf-values shown above are taken from the Student's t -distribution with $N_h - 1$ degrees of freedom, with N_h being the number of h -quarter-ahead forecasts, $N_h = 76 - h$. A cdf-value close to zero suggests that the predictions of the RE model are better in an MSFE sense than the those of the AL model, and a value close to one that the reverse case is supported.

TABLE 3: Log scores as well as cdf-values of the Amisano-Giacomini weighted LR test and of the modified Diebold-Mariano test of the null hypothesis that the log scores of the RE and AL baseline models are equal for the full sample 2001Q1–2019Q4.

h	Real GDP growth				Inflation				Joint			
	Log scores		cdf-values		Log scores		cdf-values		Log scores		cdf-values	
	RE	AL	LR	DM	RE	AL	LR	DM	RE	AL	LR	DM
–1	–2.38	–2.14	–	–	–1.01	–3.17	–	–	–8.99	–10.12	–	–
0	–63.12	–63.30	0.47	0.47	21.07	11.74	0.00	0.00	–42.25	–51.02	0.01	0.01
1	–73.45	–75.60	0.19	0.15	16.57	6.79	0.00	0.00	–56.36	–71.86	0.00	0.00
2	–75.57	–76.79	0.34	0.36	9.88	0.62	0.01	0.01	–65.21	–79.68	0.00	0.00
3	–75.47	–76.52	0.34	0.38	3.78	–3.97	0.04	0.06	–71.47	–83.85	0.01	0.01
4	–73.37	–75.92	0.20	0.25	–1.34	–6.43	0.13	0.15	–74.92	–84.88	0.03	0.03
5	–70.59	–75.30	0.11	0.14	–5.15	–6.38	0.37	0.39	–76.38	–83.28	0.08	0.08
6	–68.79	–74.46	0.08	0.09	–9.24	–8.33	0.60	0.60	–78.90	–83.81	0.16	0.13
7	–67.36	–73.47	0.06	0.07	–13.02	–9.89	0.80	0.82	–81.41	–84.10	0.30	0.30
8	–66.16	–72.74	0.06	0.07	–16.50	–12.69	0.84	0.87	–83.72	–85.75	0.35	0.35

NOTES: The full sample covers the vintages 2001Q1–2019Q4. The cdf-values for the Amisano-Giacomini weighted LR test are taken from its asymptotic normal distribution using 1 lag for the HAC estimator and equal weights; see Amisano and Giacomini (2007) for details. The cdf-values for the Diebold-Mariano test are taken from the Student’s t -distribution with $N_h - 1$ degrees of freedom, with N_h being the number of h -quarter-ahead forecasts, $N_h = 76 - h$. The DM test statistic is based on the difference of minus the log scores rather than the difference in mean-squared forecast errors as in Harvey et al. (1997), equation (9). A cdf-value close to zero suggests that the predictions of the RE model are better in a log score sense than those of the AL model, and a value close to one that the reverse case is supported.

TABLE 4: Log scores as well as cdf-values of the Amisano-Giacomini weighted LR test and of the modified Diebold-Mariano test of the null hypothesis that the log scores of the RE and AL baseline models are equal for the sub-samples 2001Q1–2014Q4 and 2015Q1–2019Q4.

2001Q1-2014Q4													
h	Real GDP growth				Inflation				Joint				
	Log scores		cdf-values		Log scores		cdf-values		Log scores		cdf-values		
	RE	AL	LR	DM	RE	AL	LR	DM	RE	AL	LR	DM	
-1	-2.38	-2.14	-	-	-3.95	-5.71	-	-	-8.99	-10.12	-	-	
0	-52.24	-52.24	0.50	0.50	13.30	5.87	0.01	0.00	-39.17	-46.33	0.02	0.03	
1	-60.69	-61.41	0.38	0.36	11.57	1.54	0.00	0.00	-48.71	-62.49	0.00	0.00	
2	-62.88	-62.69	0.53	0.52	7.59	-3.01	0.00	0.00	-54.92	-68.91	0.00	0.00	
3	-63.44	-63.08	0.56	0.54	3.42	-6.36	0.01	0.01	-59.88	-72.66	0.00	0.00	
4	-61.96	-63.20	0.34	0.37	-0.36	-7.70	0.03	0.04	-62.60	-73.44	0.01	0.01	
5	-59.84	-63.37	0.18	0.21	-4.05	-7.20	0.18	0.20	-64.60	-72.23	0.05	0.02	
6	-58.71	-63.49	0.11	0.13	-8.27	-8.02	0.53	0.53	-67.95	-72.59	0.16	0.09	
7	-57.90	-63.28	0.08	0.10	-12.32	-9.09	0.81	0.85	-71.36	-73.17	0.36	0.35	
8	-57.31	-63.27	0.08	0.09	-15.45	-11.23	0.87	0.91	-73.95	-74.89	0.43	0.42	

2015Q1-2019Q4													
h	Real GDP growth				Inflation				Joint				
	Log scores		cdf-values		Log scores		cdf-values		Log scores		cdf-values		
	RE	AL	LR	DM	RE	AL	LR	DM	RE	AL	LR	DM	
-1	-	-	-	-	2.94	2.54	-	-	-	-	-	-	
0	-10.89	-11.06	0.32	0.29	7.77	5.86	0.15	0.15	-3.08	-4.69	0.18	0.18	
1	-12.76	-14.18	0.01	0.00	5.00	5.25	0.56	0.56	-7.64	-9.38	0.18	0.17	
2	-12.69	-14.09	0.01	0.01	2.29	3.62	0.78	0.78	-10.30	-10.77	0.40	0.40	
3	-12.03	-13.44	0.02	0.04	0.36	2.38	0.86	0.86	-11.58	-11.19	0.58	0.57	
4	-11.40	-12.72	0.02	0.06	-0.98	1.27	0.88	0.78	-12.32	-11.44	0.68	0.60	
5	-10.75	-11.93	0.02	0.06	-1.09	0.82	0.88	0.81	-11.79	-11.04	0.68	0.60	
6	-10.08	-10.97	0.04	0.16	-0.97	-0.30	0.73	0.65	-10.95	-11.22	0.42	0.46	
7	-9.47	-10.19	0.08	0.25	-0.70	-0.80	0.45	0.48	-10.04	-10.93	0.20	0.37	
8	-8.84	-9.47	0.09	0.30	-1.05	-1.46	0.31	0.42	-9.78	-10.86	0.16	0.36	

NOTES: The cdf-values for the Diebold-Mariano test are taken from the Student's t -distribution with $N_h - 1$ degrees of freedom, with N_h being the number of h -quarter-ahead forecasts, $N_h = 56 - h$ for the first sub-sample and $N_h = 20 - h$ for the second. The cdf-values should be interpreted with a large grain of salt for the short samples. See also Table 3 for details.

TABLE 5: Continuous ranked probability and energy scores for the full sample 2001Q1–2019Q4.

h	Real GDP growth			Inflation			Joint		
	CRPS		cdf-values	CRPS		cdf-values	ES		cdf-values
	RE	AL		RE	AL		RE	AL	
1	-24.99	-26.09	0.03	-7.33	-9.12	0.00	-27.33	-29.47	0.00
2	-26.03	-26.75	0.23	-7.90	-9.78	0.02	-28.59	-30.43	0.01
3	-25.80	-26.76	0.20	-8.70	-10.13	0.09	-28.85	-30.66	0.01
4	-24.70	-26.27	0.08	-9.50	-10.18	0.28	-28.30	-30.34	0.01
5	-23.48	-25.95	0.00	-10.10	-9.72	0.64	-27.49	-29.72	0.00
6	-22.50	-25.64	0.00	-10.79	-9.61	0.91	-27.00	-29.33	0.02
7	-21.67	-25.25	0.00	-11.45	-9.40	0.99	-26.67	-28.80	0.06
8	-21.08	-24.76	0.00	-12.07	-9.71	0.99	-26.51	-28.51	0.09

NOTES: The cdf-values from the Diebold-Mariano (DM) test for the null hypothesis of equal CRPS/ES are calculated as in Harvey et al. (1997), equation (9), with minus the CRPS/ES values replacing the mean-squared forecast errors. The cdf-values shown above are taken from the Student's t -distribution with $N_h - 1$ degrees of freedom, with N_h being the number of h -quarter-ahead forecasts, $N_h = 76 - h$. A cdf-value close to zero suggests that the predictions of the RE model are better in a CRPS/ES sense than those of the AL model, and a value close to one that the reverse case is supported.

TABLE 6: PIT and BOT tests for the marginal and joint nowcasts and one-quarter-ahead density forecasts of real GDP growth and GDP deflator inflation over the vintages 2001Q1–2019Q4.

h	Model	Real GDP growth		Inflation		Joint	
		AG	p -value	AG	p -value	AG	p -value
0	RE	32.37	0.00	3.30	0.50	49.51	0.00
	AL	61.66	0.00	16.80	0.00	27.21	0.00
1	RE	36.49	0.00	7.29	0.12	152.85	0.00
	AL	57.88	0.00	26.11	0.00	26.98	0.00

NOTES: The Amisano-Geweke tests are described in the Online Appendix of Amisano and Geweke (2017). The statistics above are based on $p = 2$ lags and $q = 2$ moments such that the asymptotic distribution is χ_4^2 . The PIT (BOT) tests concern the marginal (joint) nowcasts and forecasts.

FIGURE 1: The euro area data on total investment growth, 1980Q1–2019Q4.

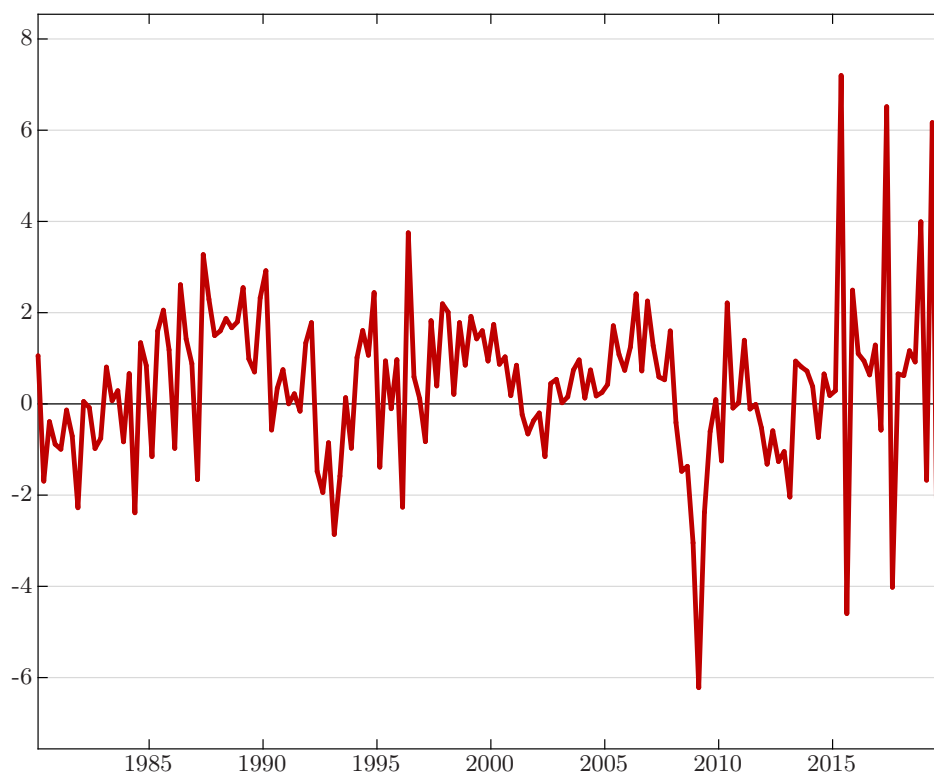
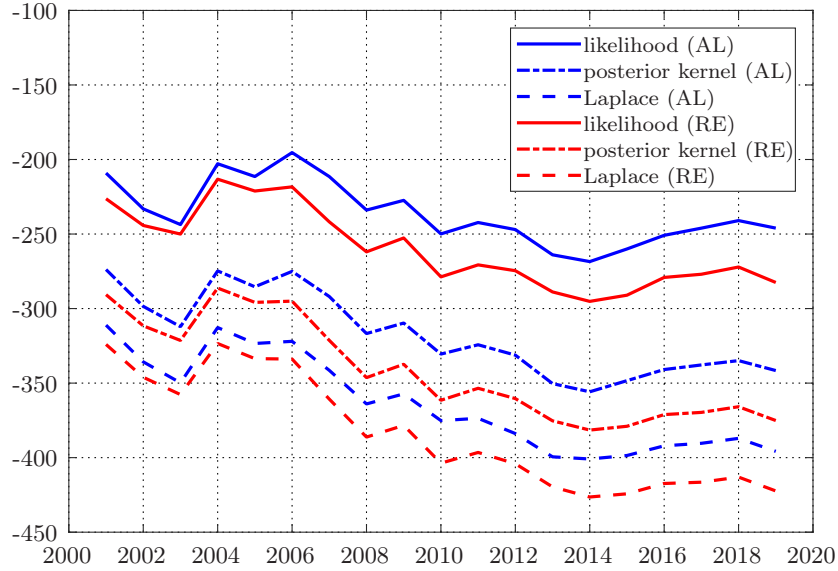
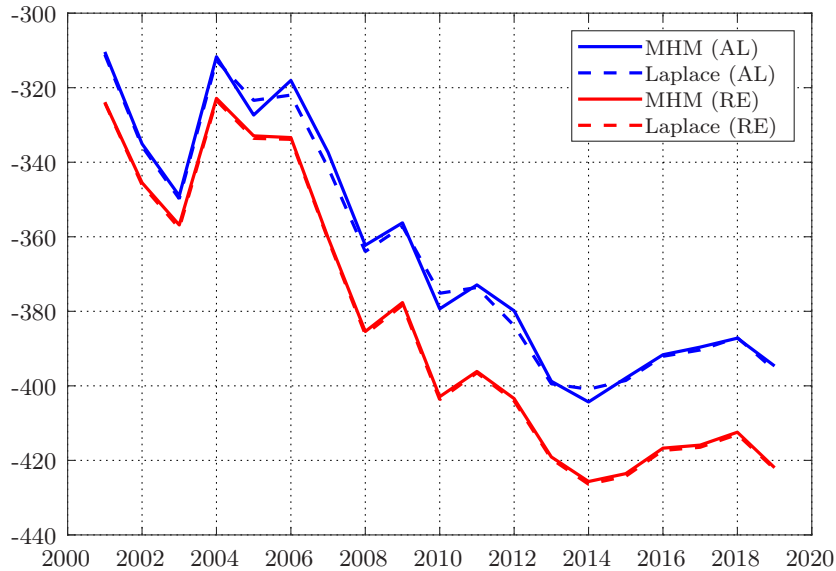


FIGURE 2: Recursive estimates of the log marginal likelihood, the log posterior kernel and the log likelihood of the rational expectations (RE) and adaptive learning (AL) models using the RTD annual vintages 2001–2019.

A. Posterior mode based estimates.



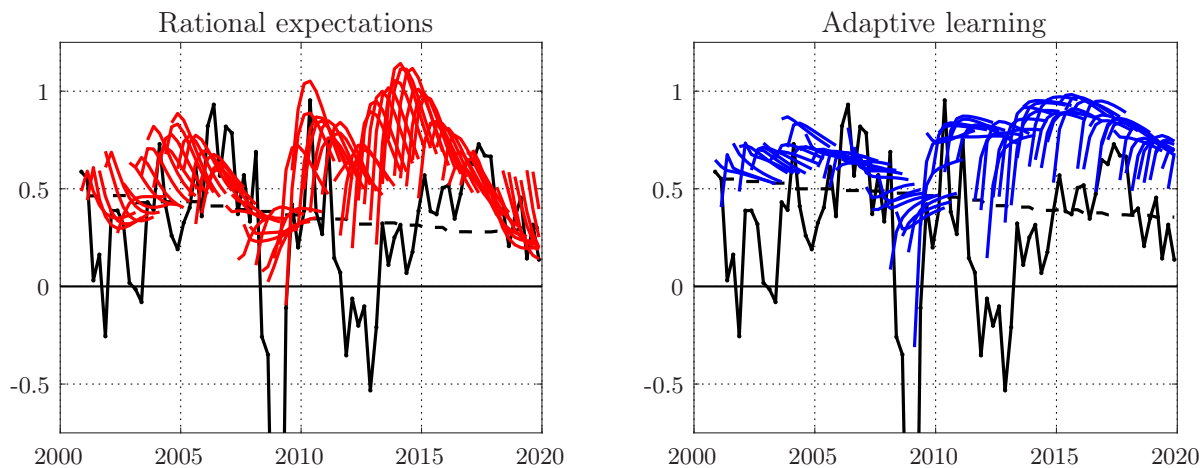
B. Posterior draws/mode based estimates



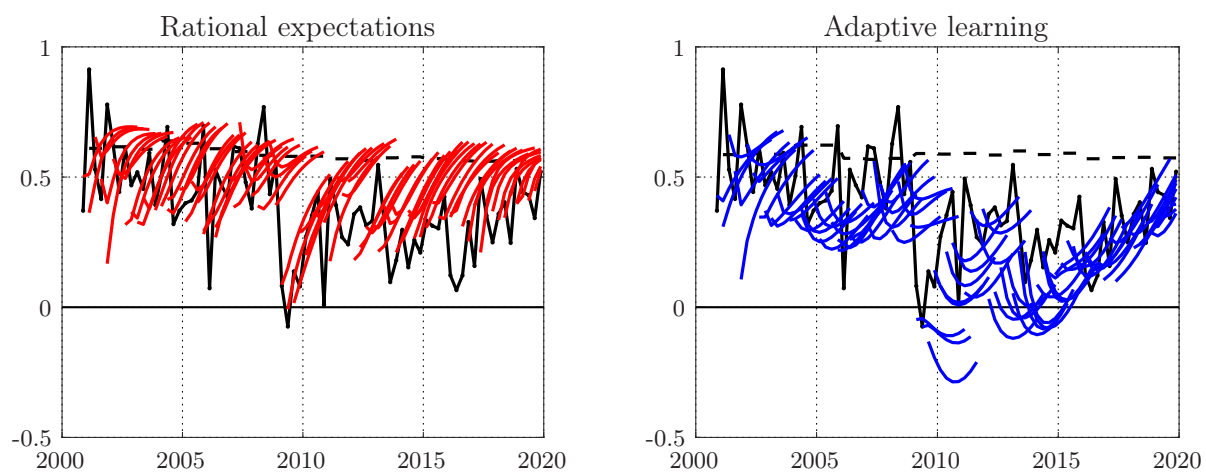
NOTES: The log marginal likelihood is estimated with (i) the modified harmonic mean (MHM) estimator using 10,000 equally spaced posterior draws from the 500,000 post burn-in draws of the random-walk Metropolis sampler with a burn-in sample of 250,000 draws, and (ii) the Laplace approximation for the joint posterior mode. The log likelihood and the log posterior kernel are also estimated at the joint posterior mode of the respective annual vintage, each taken from Q1.

FIGURE 3: Recursive point backcasts, nowcasts and up to eight-quarters-ahead forecasts of quarterly real GDP growth and quarterly GDP deflator inflation using the RTD vintages 2001Q1–2019Q4.

A. Real GDP Growth



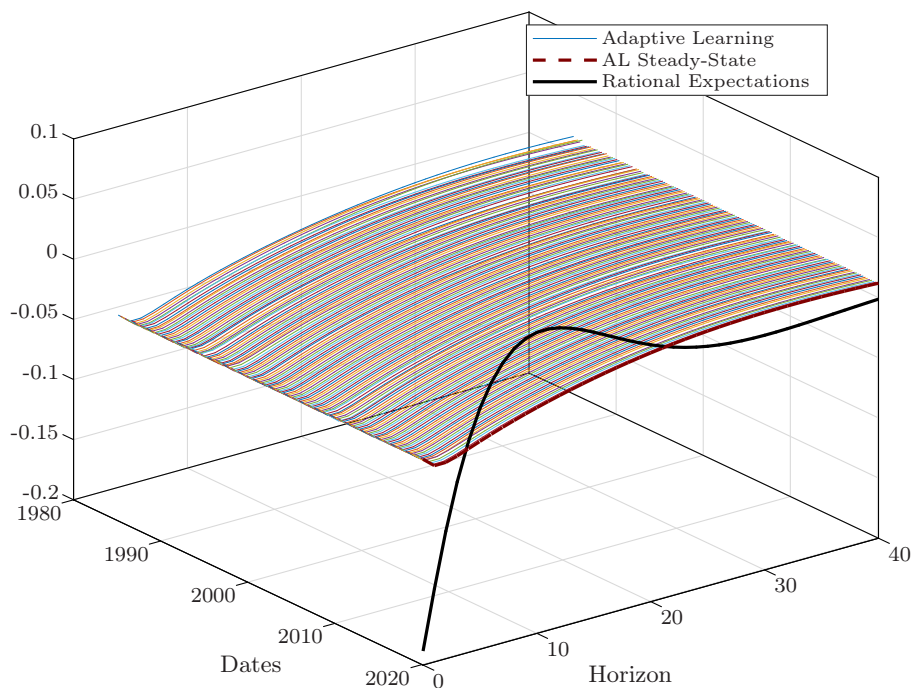
B. GDP Deflator Inflation



NOTES: The actual values are plotted as solid black lines. Recursively estimated posterior predictive mean values of real GDP growth and inflation are plotted as solid red and blue lines under RE and AL, respectively, while the dashed black lines are the recursive posterior mean estimates of steady-state real GDP growth and inflation.

FIGURE 4: Responses to real GDP growth and inflation from a one standard deviation shock to monetary policy using the full sample posterior mode estimates from the baseline AL model.

A. Real GDP Growth



B. GDP Deflator Inflation

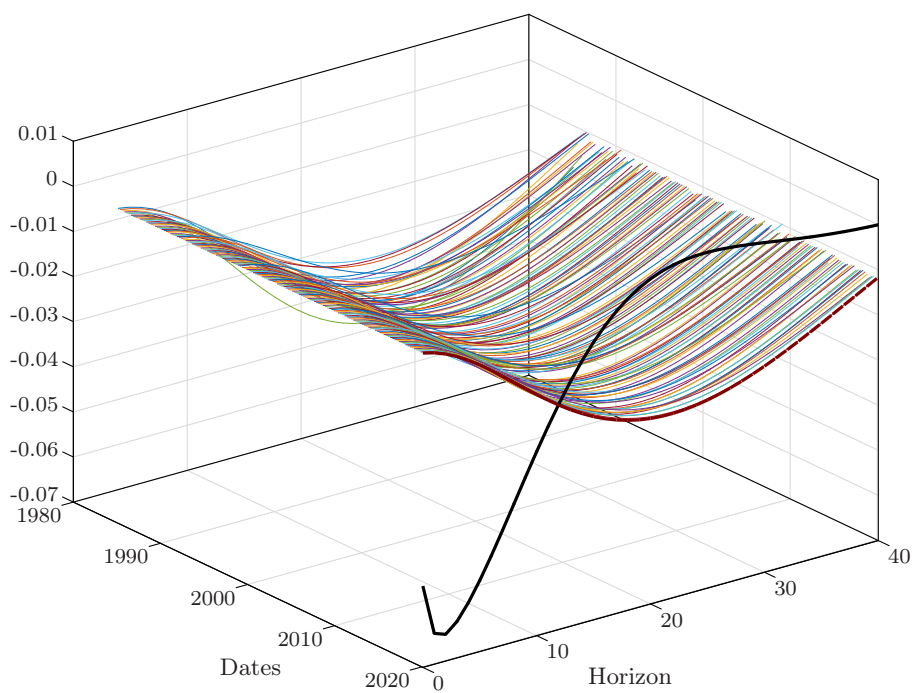


FIGURE 5: Recursive estimates of the average log scores of the joint real GDP growth and inflation density forecasts covering the vintages 2001Q1–2019Q4.

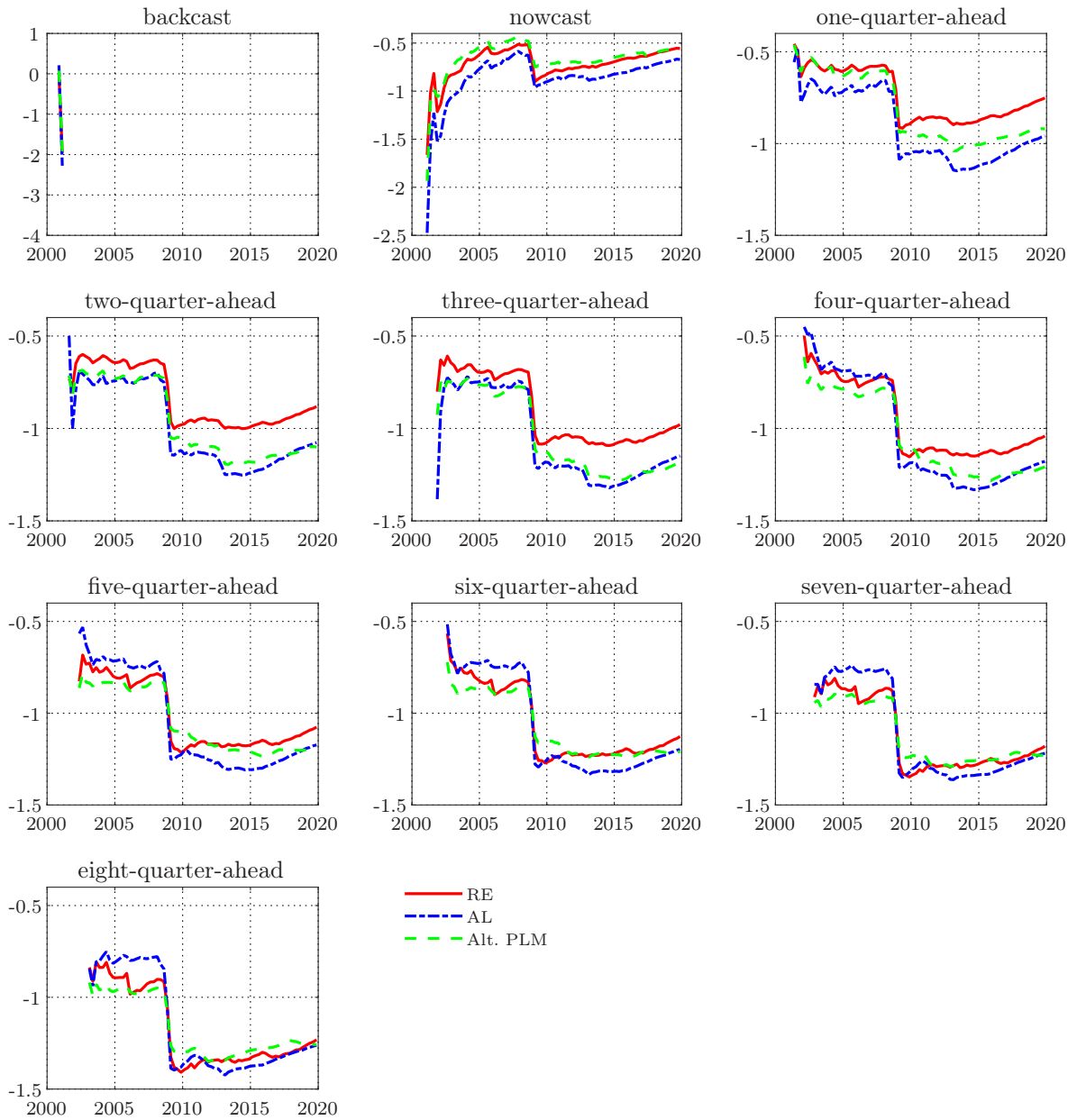
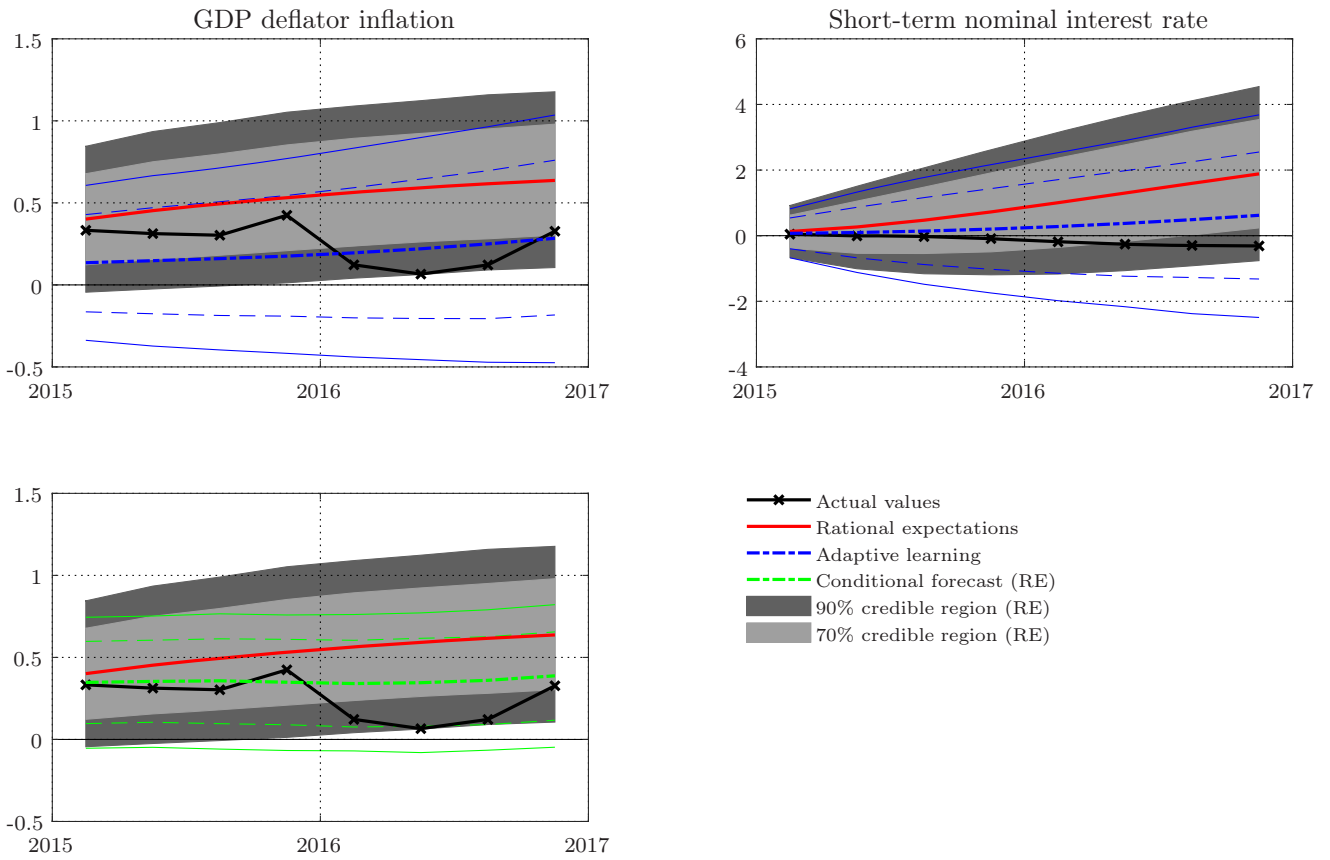


FIGURE 6: Predictive distributions of one-quarter until eight-quarter-ahead GDP deflator inflation and the short-term nominal interest rate for the 2014Q4 vintage.



NOTES: The unconditional forecasts of the baseline RE model are plotted as a solid red line, while the 90 % (70 %) equal-tails credible interval of these forecasts are shown as a dark-grey (light-grey) area. The unconditional forecasts of the baseline AL model are given by a dash-dotted blue line, while the borders of the 90 % (70 %) equal-tails credible interval are plotted as solid (dashed) blue lines. Finally, the conditional forecasts of the baseline RE model are given by the dash-dotted green line, while the borders of its 90 % (70 %) equal-tails credible interval are plotted as solid (dashed) green lines. The conditional forecast of inflation are based on the actual values of the short-term nominal interest rate in 2015Q1–2016Q4 using the Waggoner and Zha (1999) conditioning approach.

FIGURE 7: Recursive estimates of the average energy scores of the joint real GDP growth and inflation distributional forecasts covering the vintages 2001Q1–2019Q4.

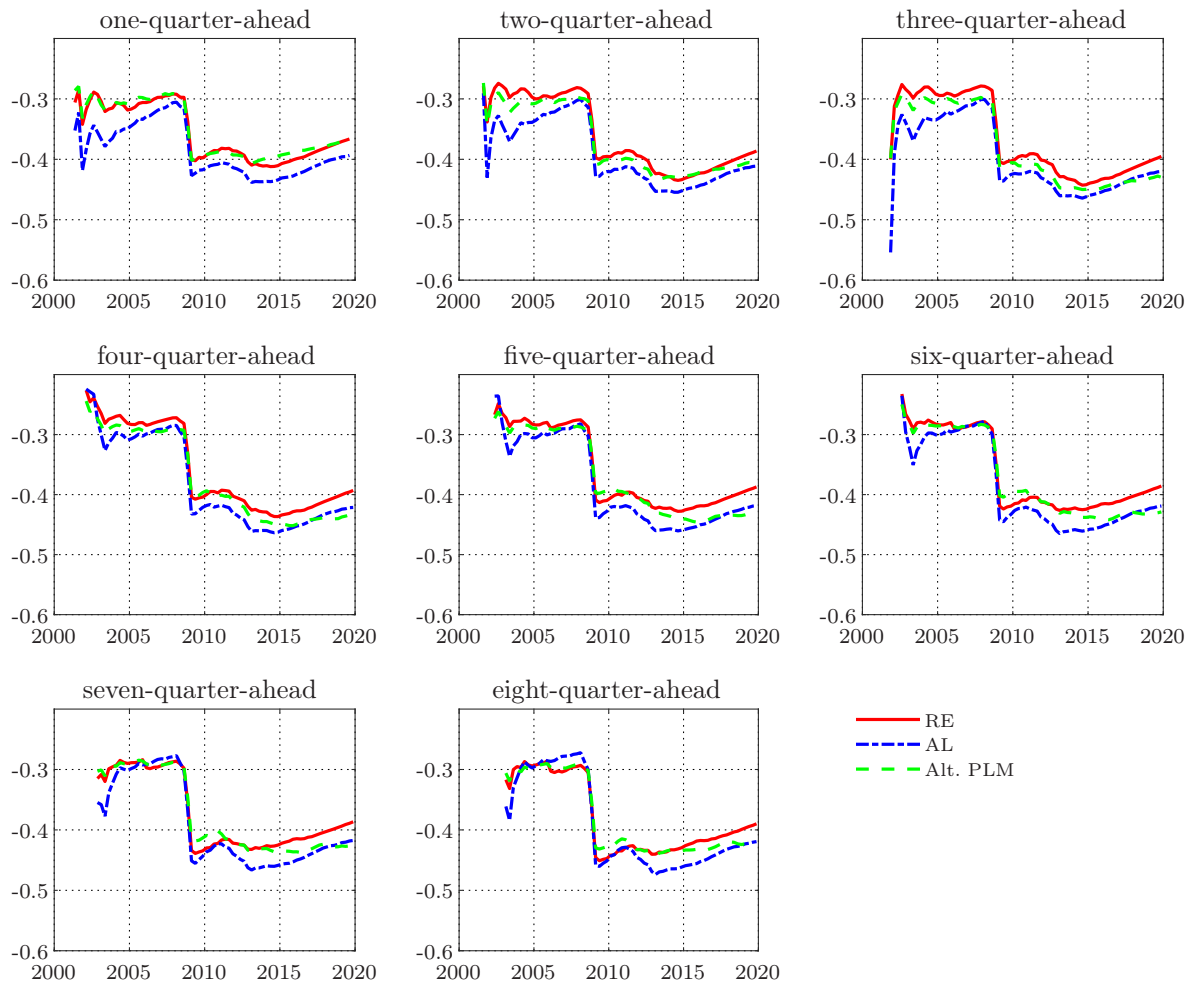
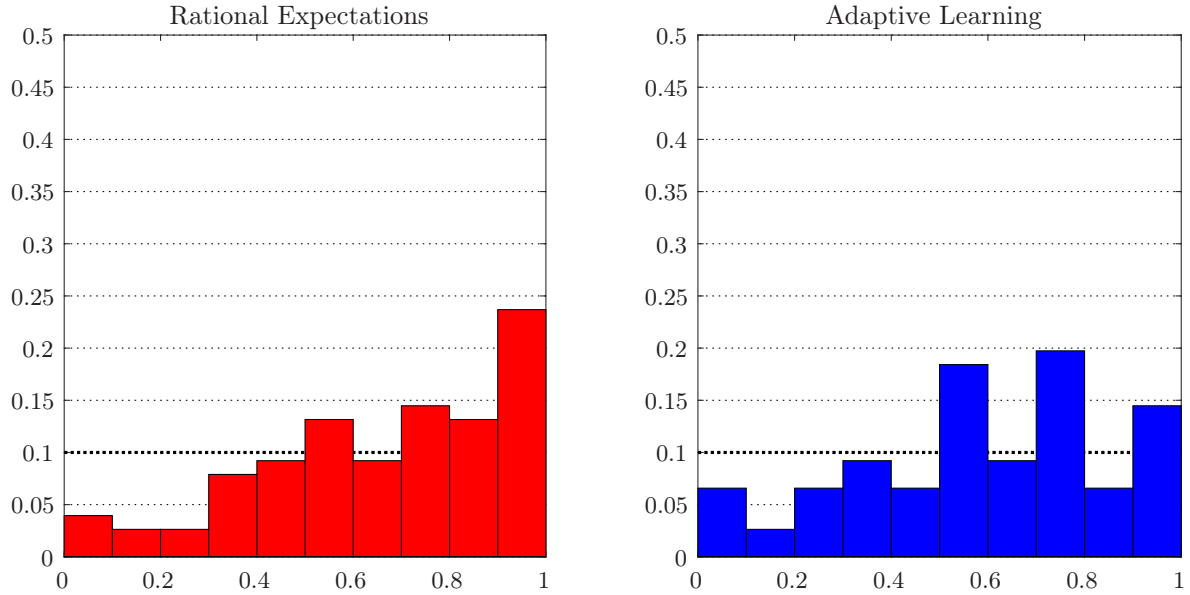
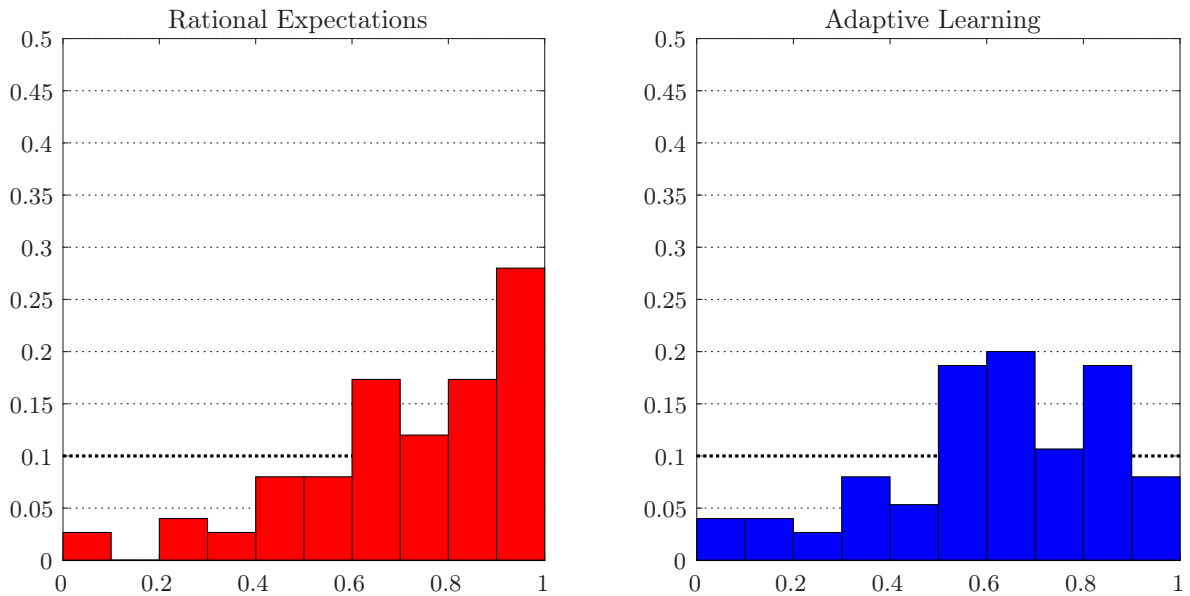


FIGURE 8: Histograms of the estimated $\pi_{T+h|T}$ values for the BOTs of the joint real GDP growth and inflation density forecasts at the nowcast ($h = 0$) and one-step-ahead ($h = 1$) horizons for 2001Q1–2019Q4.

A. Nowcasts



B. One-quarter-ahead forecasts



NOTES: The horizontal axis shows the 10 bins while the vertical axis shows the occurrence frequency for the estimated π 's. If these variables are uniformly distributed for a model, then the occurrence in large samples is 0.10 for all bins.