

Macroeconomic forecasting in the euro area using predictive combinations of DSGE models

Capek, Jan; Crespo Cuaresma, Jesus; Hauzenberger, Niko; Reichel, Vlastimil

DOI:
[10.57938/86ddaf39-548c-4bb2-87bd-52579a41e9bb](https://doi.org/10.57938/86ddaf39-548c-4bb2-87bd-52579a41e9bb)

Published: 01/11/2020

Document Version:
Publisher's PDF, also known as Version of record

Document License:
Unspecified

[Link to publication](#)

Citation for published version (APA):
Capek, J., Crespo Cuaresma, J., Hauzenberger, N., & Reichel, V. (2020). *Macroeconomic forecasting in the euro area using predictive combinations of DSGE models*. WU Vienna University of Economics and Business. Department of Economics Working Paper Series No. 305 <https://doi.org/10.57938/86ddaf39-548c-4bb2-87bd-52579a41e9bb>

Department of Economics
Working Paper No. 305

Macroeconomic forecasting in the euro area using predictive combinations of DSGE models

Jan Čapek
Jesús Crespo Cuaresma
Niko Hauzenberger
Vlastimil Reichel

November 2020



Macroeconomic forecasting in the euro area using predictive combinations of DSGE models*

Jan Čapek
Masaryk University

Jesús Crespo Cuaresma
*Vienna University of Economics and Business
Masaryk University, IIASA, WIFO*

Niko Hauzenberger
*University of Salzburg
Masaryk University*

Vlastimil Reichel
Masaryk University

Abstract

We provide a comprehensive assessment of the predictive ability of combinations of Dynamic Stochastic General Equilibrium (DSGE) models for GDP growth, inflation and the interest rate in the euro area. We employ a battery of static and dynamic pooling weights based on Bayesian model averaging principles, prediction pools and dynamic factor representations, and entertain eight different DSGE specifications and four prediction weighting schemes. Our results indicate that exploiting mixtures of DSGE models tends to achieve superior forecasting performance over individual specifications for both point and density forecasts. The largest improvements in the accuracy of GDP growth forecasts are achieved by the prediction pooling technique, while the results for the weighting method based on dynamic factors partly leads to improvements in the quality of inflation and interest rate predictions.

Keywords: Forecasting, model averaging, prediction pooling, DSGE models
JEL Codes: E37, E47, C53

*Corresponding author: Jesús Crespo Cuaresma, Vienna University of Economics and Business. E-mail: jrcrespo@wu.ac.at. Financial support from the Czech Science Foundation, Grant 17-14263S, is gratefully acknowledged. Computational resources were provided by the CESNET LM2015042 and the CERIT Scientific Cloud LM2015085, provided under the programme "Projects of Large Research, Development, and Innovations Infrastructures". Jesús Crespo Cuaresma gratefully acknowledges funding from IIASA and the National Member Organizations that support the institute.

1 Introduction

Due to their internal consistency and their ability to assess the effects of policy shocks in a rigorous manner, Dynamic Stochastic General Equilibrium (DSGE) models have become the workhorse of modern macroeconomic research.

In spite of their importance in modern economic analysis, the existing results concerning their out-of-sample forecasting ability are mixed. A series of studies have assessed the predictive ability of different types of DSGE models. Christoffel et al. (2011) examine the out-of-sample predictive ability of the European Central Bank’s New Area-Wide Model (NAWM), the DSGE model used to create projections of macroeconomic variables by the monetary authority of the euro area. The results in Christoffel et al. (2011) indicate that the DSGE model, as compared to other alternative reduced-form specifications, provides good predictions for twelve different macroeconomic variables. The predictive accuracy of DSGE models, however, does not necessarily remain stable over time. Del Negro et al. (2016) provide evidence that forecasts created using a Smets-Wouters type of DSGE model (Smets and Wouters, 2003, 2007) with financial frictions performs particularly well in periods of financial turmoil (in particular in the Great Recession), but that the predictive accuracy of the model tends to suffer in tranquil periods. The forecasting quality of DSGE structures which include financial frictions has also been assessed by Kolasa and Rubaszek (2015), and improvements in forecasting ability are reported in episodes of financial turmoil when housing market frictions are included in the model, although no systematic gains in predictive performance are found in more stable periods.

In parallel to the development of the literature on forecasting with DSGE models, there has been a revived interest among researchers in analysing the combination of predictions based on different modelling frameworks, an idea that dates back to the work by Bates and Granger (1969). Amisano and Geweke (2017), for instance, find improvements in out-of-sample prediction errors for macroeconomic variables in the euro area by pooling forecasts from different macroeconomic models using Bayesian predictive distributions.

In this study, we evaluate the forecasting ability of weighted combinations of DSGE models for GDP growth, inflation and the interest rate in the euro area, making use of several prediction combination techniques. Our analysis expands the work by Wolters (2015), which assesses the forecasting ability of four DSGE models for the US, as well as the potential predictive gains obtained by using combinations of these. We entertain eight different DSGE specifications for the euro area and four forecast combination methods, both static and dynamic, and evaluate point forecasts as well as density predictions. Our set of prediction combination techniques contains some of the forecast pooling techniques entertained in existing studies for DSGE models (Wolters, 2015, for example), as well as more novel methods based on optimization of weights and time-varying weighting. In particular, we use static weights based on principles of Bayesian model averaging and prediction pools, and dynamic weights that build upon dynamic factor representations of the variables of interest.

The combination techniques employed in our analysis result in significantly different weighting schemes across models. While Bayesian model averaging and combinations based on dynamic factors lead to pooled forecasts which assign positive weights to all of the DSGE specifications, the technique based on prediction pools acts as a dynamic model selection tool, assigning weights which are close to zero to most individual model predictions over the out-of-sample period. Our results indicate that substantial gains in predictive ability can be achieved using combination methods. The largest improvements in the accuracy of GDP growth forecasts (both in the one-quarter-ahead and in the one-year-ahead horizons) are achieved by the prediction pooling technique, while the results for the weighting method based on dynamic factors is particularly promising for inflation and interest rate predictions.

The paper is organized as follows. Section 2 presents the weighting schemes used to aggregate the DSGE models, which are introduced in section 3. Section 4 presents the results of the out-of-sample forecasting exercise and section 5 concludes.

2 Predictive combinations of DSGE models

In our analysis, we consider forecast combination methods for a set of L different variables of interest at an h -step-ahead horizon, $\{y_{1,t+h}, \dots, y_{L,t+h}\}$. Consider predictive densities for $y_{i,t+h}$, $i = 1, \dots, L$, which are available

from K different DSGE models. Each DSGE model M_j , for $j = 1, \dots, K$, incorporates information up to time t to generate a predictive density $p(y_{i,t+h}|\mathcal{F}_j(t), M_j)$ for period $t+h$. The information set $\mathcal{F}_j(t)$ usually consists of $\mathbf{y}_{1:t} = (\mathbf{y}_{1,1:t}, \dots, \mathbf{y}_{L,1:t})$, as well as of the information provided by additional variables specific to model M_j , $\mathbf{x}_{j1:t}$, that is, $\mathcal{F}_j(t) = (\mathbf{y}_{1:t}, \mathbf{x}_{j1:t})$.

For the i -th variable, in the following, we aim at combing the K predictive densities $p(y_{i,t+h}|\mathcal{F}_j(t), M_j)$ using a K -dimensional weights vector $\boldsymbol{\omega}_{i,t}$, potentially varying over time. Formally, the forecast combination for variable $y_{i,t+h}$ is given by

$$p(y_{i,t+h}|\boldsymbol{\omega}_{i,1:t}, \mathcal{F}_{1:K}(t), M_{1:K}) = \sum_{j=1}^K \omega_{ij,t+h} p(y_{i,t+h}|\mathcal{F}_j(t), M_j), \quad (1)$$

with $\boldsymbol{\omega}_{i,t} = (\omega_{i1,t}, \dots, \omega_{iK,t})'$ being a multivariate function relating K different predictive densities $p(\mathbf{y}_{i,t+h}|\mathcal{F}_j(t), M_j)$ to the target density $p(\mathbf{y}_{i,t+h}|\boldsymbol{\omega}_{i,1:t}, \mathcal{F}_{1:K}(t), M_{1:K})$. It proves useful to define the likelihood of observations $\mathbf{y}_{i,1:t}$ conditional on $\boldsymbol{\omega}_{i,1:t}$,

$$p(\mathbf{y}_{i,1:t}|\boldsymbol{\omega}_{i,1:t}, M_{1:K}) = \prod_{\tau=s}^t \sum_{j=1}^K \omega_{ij,\tau} p(y_{i,\tau}|\mathcal{F}_j(\tau-h), M_j). \quad (2)$$

Equation (2) directly relates to the Bayesian predictive synthesis of McAlinn and West (2019), where $\boldsymbol{\omega}_{i,t}$ is described as a dynamic synthesis function.¹ The synthesis function allows to incorporate different objectives based on policy targets and historical performance, and nests traditional approaches to forecast combination, such as prediction pools (Hall and Mitchell, 2007; Geweke and Amisano, 2011) and Bayesian model averaging (McAlinn et al., 2019). Equation (2) is used to discuss the different approaches applied in our analysis in order to combine predictive densities. First, we discuss different static weighting schemes (that is, $\boldsymbol{\omega}_{i,t} = \boldsymbol{\omega}_i$) and then turn to more general approaches based on using dynamic weights for the predictive densities.

Equal weights

An obvious starting point to combine predictions from different DSGE models, which provides a benchmark to evaluate different weighting schemes, is to use

$$\omega_{i1}^{(EQ)} = \dots = \omega_{iK} = 1/K. \quad (3)$$

Since $\omega_{ij}^{(EQ)} > 0$ and $\sum_{j=1}^K \omega_{ij}^{(EQ)} = 1$, the combination of predictive densities also constitutes a predictive density (Hall and Mitchell, 2007; Geweke and Amisano, 2011). This agnostic approach neglects the fact that different models might not be equally suitable for prediction at different time periods, and does not provide updates of the corresponding weights as information is gained about the different predictive ability of model specifications. An equal weighting scheme is commonly found to be a good competitor in terms of out-of-sample accuracy, as it tends to hedge against large forecast errors of single specifications (see Timmermann, 2006).

Dynamic Bayesian model averaging

A natural choice of model weights can be achieved by pooling according to particular model selection criteria (for example, based on their predictive marginal likelihood or past forecast performance). For a given set of priors over specifications, traditional Bayesian model averaging (BMA) approaches give models with a higher marginal likelihood more support, while downweighting models with deficient predictive characteristics.

Following Raftery et al. (2010) and Koop and Korobilis (2012), we consider posterior weights for individual specifications based on their (discounted) historical predictive power over the last $t-s$ observations, a procedure

¹Del Negro et al. (2016) and McAlinn and West (2019) provide a formal treatment of the decision problem concerning the choice of (time-varying) weights $\boldsymbol{\omega}_{i,t}$.

known as dynamic model averaging (DMA). Relying on information starting from time s up to time t , $\omega_{i,t+h}$ can be computed as

$$\omega_{ij,t+h}^{(DMA)} = \frac{\prod_{\tau=s}^t p\left(y_{i,\tau}^{(r)} | \mathcal{F}_j(\tau-h), M_j\right)^{\delta^{t-\tau}}}{\sum_{k=1}^K \prod_{\tau=s}^t p\left(y_{i,\tau}^{(r)} | \mathcal{F}_k(\tau-h), M_k\right)^{\delta^{t-\tau}}} \quad \text{for } j = 1, \dots, K, \quad (4)$$

where $y_{i,\tau}^{(r)}$ denotes the realized value in period τ and $0 < \delta < 1$ denotes the forgetting factor which downweights the importance of past predictive evidence.² By construction, $\omega_{ij,t+h}^{(DMA)} > 0$ and $\sum_{j=1}^K \omega_{ij,t+h}^{(DMA)} = 1$.

Prediction pools

More recent approaches tend to view the set of model-specific forecasts as a portfolio of predictions which must be chosen optimally with respect to a particular loss function (see, inter alia, Hall and Mitchell, 2007; Geweke and Amisano, 2011, 2012; Pettenuzzo and Ravazzolo, 2016). Following Geweke and Amisano (2011), the loss function is defined as a function of historical log predictive scores, which gives rise to optimal weights after minimization. Similarly to BMA and DMA methods, this approach ensures that forecasts from DSGE models with poor predictive abilities are downweighted and those computed from specifications that predict more successfully receive higher weights.

Information from s up to time t is available in order to choose $\omega_{i,t}$ optimally for period t . The negative weighted historical log predictive scores is minimized with respect to the weights vector $\omega_{i,t+h}$,

$$\omega_{i,t+h}^{(POOL)} = \min_{\omega_{i,t+h}} \left\{ - \sum_{\tau=s}^t \delta^{t-\tau} \log \left(\sum_{j=1}^K \omega_{i,j,t+h} p(y_{i,\tau}^{(r)} | \mathcal{F}(\tau-h), M_j) \right) \right\},$$

where δ denotes a discount factor. Moreover, we impose the restriction given by $\omega_{ij,t+h}^{(POOL)} > 0$ and $\sum_{j=1}^K \omega_{ij,t+h}^{(POOL)} = 1$.

Dynamic weights

As noted by Del Negro et al. (2016), the predictive ability of particular specifications may be affected by structural breaks in the prevailing macroeconomic environment. Such changes in predictive power should be addressed when combining t predictive densities for the target vector $\mathbf{y}_{i,1:t}$ and the mapping from each model to the combined predictive density should be adjusted accordingly. Equation (2) can be directly related to a dynamic factor model representation, as proposed by McAlinn et al. (2019) in the context of dynamic Bayesian Predictive synthesis (BPS) methods, by defining the synthesis function as

$$y_{i,t+h}^{(r)} = \mathbf{F}'_{i,t+h} \boldsymbol{\omega}_{i,t+h}^{(BPS)} + \epsilon_{i,t+h}, \quad \epsilon_{i,t+h} \sim \mathcal{N}(\mathbf{0}, \xi_i),$$

with the latent dynamic factor that corresponds to the weighting scheme evolving according to a random walk

$$\boldsymbol{\omega}_{i,t}^{(BPS)} = \boldsymbol{\omega}_{i,t-1}^{(BPS)} + \boldsymbol{\eta}_{i,t} \quad \boldsymbol{\eta}_{i,t} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi}_i).$$

Here, ξ_i denotes the variance of the shock in the observation equation and $\boldsymbol{\Psi}_i$ refers to the variance-covariance matrix of the error term in the state equation. As in McAlinn and West (2019), we define the factors $\mathbf{F}_{i,t} = (\hat{y}_{i,1t}, \dots, \hat{y}_{i,Kt})$ with $\hat{y}_{i,jt}$, for $j = 1, \dots, K$ being a draw from the predictive density of each model M_j at period t . In contrast to equal weighting, DMA, and predictive pooling, the weights $\omega_{i,t+h}^{(BPS)}$ are no longer necessarily non-negative and do not need to sum up to one. $\omega_{i,t+h}^{(BPS)}$ are thus to be interpreted as (time-varying) calibration parameters relating

²If the discount factor is set to one, static versions of these weights are obtained. These do not incorporate a “forgetting” factor for past predictive accuracy.

draws from the predictive densities to the actual realization $y_{i,t+h}^{(r)}$. A further difference to other weighting schemes is that we consider a measurement error $\epsilon_{i,t+h}$ in the observation equation, which explicitly accounts for model incompleteness (see Aastveit et al., 2018; McAlinn and West, 2019). Moreover, the latent weights $\omega_{i,1:t}^{(BPS)}$ are allowed to be correlated among models via a full variance-covariance matrix Ψ_i .

We use weakly informative priors, which are standard in literature for dynamic factor models. That implies the use of a multivariate normal prior for $\omega_{i,0}^{(BPS)}$, an inverse gamma prior for ξ_i , and an inverse Wishart prior for Ψ_i . We repeat this procedure for R draws from the predictive density and explicitly account for a potentially non-trivial form of the predictive densities of DSGE models. To estimate the model we rely on a standard Bayesian estimation techniques used for time-varying parameter models.³

3 The battery of DSGE models

3.1 Individual DSGE models

For our empirical analysis, we use a battery of DSGE models for the euro area of different size, complexity, and with particular features. Since the analysis is conducted on a set of three core macroeconomic variables (GDP growth, inflation, and the interest rate), we ensure that these three observable variables are common across all models. The most sparse model entertained, presented in Cogley et al. (2010), only requires these three observable variables. The model by Benchimol and Fourçans (2017) additionally uses (real) money as fourth observable variable and the specification by Christensen and Dib (2008) further adds investment as fifth observable variable. The group of more complex models share the set of observable variables of the Smets and Wouters (2007) model: GDP growth, inflation, the interest rate, consumption growth, investment growth, real wage growth, and hours worked. The dataset employed is used not only to estimate the Smets and Wouters (2007) model, but also its earlier version designed for the euro area and described in Smets and Wouters (2003), as well as a model such as that proposed by De Graeve (2008), and a Smets-Wouters-type model as in Fève et al. (2013). The specification by Justiniano et al. (2011) is estimated with the relative price of consumption to investment as the eighth observable variable. Table 1 lists the models entertained, together with their corresponding abbreviations, which are used in the description of the results of the analysis and in all subsequent figures and tables, and summarizes information about the number of observable variables, number of exogenous shocks, and the main features of each model. The particular observable variables included in each one of the DSGE models are presented in Table 2.

Reference	Name	Observables	Shocks	Features
Benchimol and Fourçans (2017)	BF 2017	4	4	Money in the utility function
Christensen and Dib (2008)	CD 2008	5	5	Financial frictions as in Bernanke et al. (1999)
Cogley et al. (2010)	CPS 2010	3	5	Inflation target can change over time
De Graeve (2008)	DG 2008	7	9	Smets and Wouters (2007) model enriched with financial frictions as in Bernanke et al. (1999)
Fève et al. (2013)	FMS 2013	7	7	Justiniano et al. (2011) model with endogenous public spending and Edgeworth complementarity
Justiniano et al. (2011)	JPT 2011	8	8	Two investment shocks
Smets and Wouters (2003)	SW 2003	7	10	Numerous real and nominal frictions
Smets and Wouters (2007)	SW 2007	7	7	Deterministic growth rate driven by labor-augmenting technological progress

Table 1: Euro area DSGE Models used: Summary

³In particular, we use a Gibbs sampler, which iterates through these R draws. Conditional on all other quantities we update the latent states $\omega_{i,t}^{(BPS)}$ with a standard forward filtering backward sampling (FFBS) algorithm (Carter and Kohn, 1994; Frühwirth-Schnatter, 1994). In a next step, conditional on the time-varying calibration parameters we independently draw the observation equation variance ξ_i and the state equation variance-covariance matrix Ψ_i . All steps involve standard conditional posteriors (for details, see McAlinn and West, 2019).

The selection of models is limited by the availability of euro area data spanning a sufficiently long period of time. In particular, the financial time series used to identify financial frictions in models that account for such a mechanism are either unavailable for the euro area or available only for a much shorter time span than the rest of our variables and are thus not included in our specifications.⁴

3.2 Data

The models in Table 1 are estimated using quarterly data for the euro area in its 19-country composition. The database spans information from 1970Q3 to 2019Q1 and thus contains 195 quarterly observations. The core of the database is sourced from the Area Wide Model (AWM), presented in Fagan et al. (2005) and updated and extended by Brand and Toulemonde (2015). The original AWM database is updated using ECB or Eurostat data since the 1990s. The database is also extended by population and hours worked from the Total Economy Database and Eurostat. Data on monetary aggregates are obtained directly from the OECD.⁵ Growth rates are calculated as *quarter-on-quarter* differences of logs, and the interest rate is calculated per quarter.

The data transformations performed to the model variables correspond to those in Smets and Wouters (2007). Real consumption, investment and GDP are divided by population and transformed to growth rates. Hours worked are multiplied by employment, divided by population, and logged. Inflation is defined as the growth rate of the GDP deflator. The nominal wage is deflated by the GDP deflator, divided by population, and transformed to growth rates. The interest rates are short-term market interest rates. The monetary aggregates M1 and M3 are deflated by the GDP deflator, divided by population, and transformed to growth rates.⁶ Finally, the relative price of investment is calculated as the investment deflator divided by the consumption deflator, and transformed to growth rates.

	BF 2017	CD 2008	CPS 2010	DG 2008 FMS 2013 SW 2003 SW 2007	JPT 2011
Output	✓	✓	✓	✓	✓
Inflation	✓	✓	✓	✓	✓
Interest rate	✓	✓	✓	✓	✓
Consumption				✓	✓
Investment		✓		✓	✓
Hours worked				✓	✓
Wage				✓	✓
M1		✓			
M3	✓				
Relative investment price					✓

Table 2: DSGE Models: Observable variables

3.3 Detrending macroeconomic variables

In general, the time series of observable variables used in the estimation of DSGE models are detrended prior to estimation. Gorodnichenko and Ng (2010) compiles the detrending methods employed in 21 different models and the list of filters used in various models shows a predominance of detrending by linear trend, Hodrick-Prescott (HP) filter, and first difference transformations. For the set of models we employ, the original contributions use

⁴Additional financial variables that are used as observables in addition to the standard set of variables in Smets and Wouters (2007) include ten-year ahead inflation expectations and spreads in Del Negro et al. (2016) or credit to nonfinancial firms, ten-year constant maturity government bond, entrepreneurial net worth, and the credit spread in Christiano et al. (2014).

⁵The database mnemonic for the M1 aggregate is 'MANMM101' and M3 is coded as 'MABMM301'.

⁶Although the time series of monetary aggregates are described as seasonally adjusted in the OECD database, some parts of the series still exhibit a clear seasonal pattern, which we removed making use of the TRAMO-SEATS method in JDemetra+.

different detrending methods: Benchimol and Fourçans (2017); De Graeve (2008); Smets and Wouters (2003) use linear trends, Christensen and Dib (2008) use Hodrick-Prescott filters, and Smets and Wouters (2007) (and similar models) introduce some of the observable variables in first differences. Delle Chiaie (2009) investigates the effects of detrending observable variables with the HP filter and a linear trend in the model by Smets and Wouters (2003) model and finds that structural parameter estimates are rather sensitive to the choice of a particular detrending method.

In order to ensure the comparability of forecasts across different DSGE specifications, we conduct our analysis using the same filter across all models. The simplest detrending approach (“Const/HP”) demeans the time series which are expressed as growth rates and uses the HP filter for the rest of the variables. In a different detrending approach, we use HP filtering to all observable variables. Following the criticism of using two-sided HP filters for the estimation of DSGE models originally voiced by Stock and Watson (1999), we employ the one-sided version of the HP filter, which is in line with the recursive state-space representation of our DSGE model structures. Since the main aim of this study is the recursive evaluation of forecasts, the two-sided HP filter appears less adequate. Additionally, we reflect on the criticism of (both the one and two-sided versions of) the HP filter in Hamilton (2018) and implement the regression-based detrending approach introduced in that contribution. All models are thus estimated alternatively on datasets with variables detrended with the “Const/HP” approach, the “Hodrick-Prescott” approach, and the “Hamilton” approach. In the presentation of our results, we concentrate on the estimates based on variables detrended using the Hamilton approach. Detailed results for the other detrending methods can be found in the Appendix.

3.4 Estimation and calculating predictive densities

Each model is recursively estimated using Bayesian methods in Adjemian et al. (2011), starting from 70 observations (corresponding to the time frame 1970Q3 – 1987Q4) and adding one quarter at a time to the maximum of 195 observations (corresponding to the full time frame 1970Q3 – 2019Q1). The models are estimated using a minimum of one million Metropolis-Hastings replications in two chains each. To ensure the convergence of the Markov chain to its ergodic distribution, the models are checked following Gelman and Rubin (1992) and in cases of non-convergence, an additional million replications are added, up to a maximum of 4 million. We use a Monte-Carlo based optimization routine to ensure that the optimal acceptance ratio of the Metropolis-Hasting algorithm is reached. For the analysis, we discard 90 percent of the replications as a burn-in.

Forecasts are based on 10,000 draws from the posterior distribution for every estimated model on each time frame. In each instance, we calculate one to four-step-ahead out-of-sample forecasts of GDP growth, inflation, and the interest rate. The analysis is conducted after imposing back the trend of the observable variables to ensure that all models and all detrending approaches provide predictions of the same time series for the variables of interest.⁷

4 Combination strategies for forecasts of DSGE models: The evidence

4.1 The dynamics of predictive weights

We start by assessing the dynamics in the relative predictive ability of DSGE models by studying the evolution of weights along the hold-out sample for our three different target variables: GDP growth, inflation and the interest rate. We calibrate the weights for each forecast combination scheme setting $s = 20$ (i.e. five years) and $\delta = 0.9$. The hold-out-sample, which is used to evaluate the out-of-sample predictive power of our models and combinations thereof, spans the period 1995Q1 – 2018Q4,

Figure 1 and Figure 2 show the weights obtained for each model and target variable in the hold-out sample period for one-step-ahead (Figure 1) and four-steps-ahead forecasts (Figure 2). The weighting schemes across forecast horizons are relatively similar, indicating that the predictive power of DSGE models is roughly stable across forecasting horizons. In spite of the fact that the loss functions in the DMA and prediction pool method are

⁷As an illustration, Figure A1 in the Appendix shows one-step-ahead to four-steps-ahead out-of-sample recursive forecasts for euro area inflation in selected combinations of DSGE models and detrending methods.

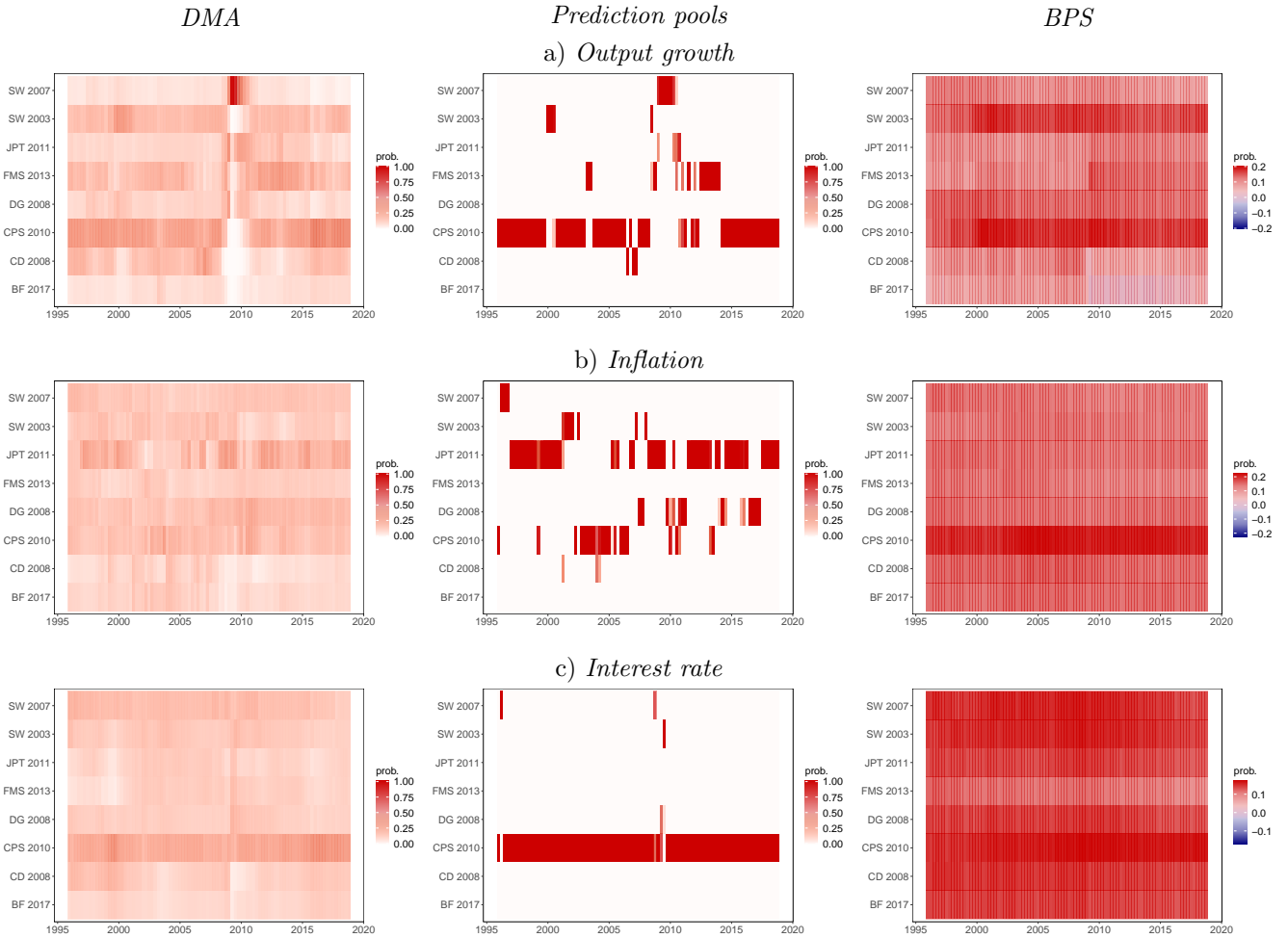


Figure 1: Evolution of the posterior mean of model weights over the hold-out sample for one-step-ahead predictions. The figure shows three different weighting schemes for the three target variables: output, inflation, and interest rate. Variables entering the DSGE models are detrended with the Hamilton filter.

both based on log-predictive scores, we observe substantial differences in the magnitude of the weights obtained for these two approaches. The prediction pool approach typically suggests a dynamic model selection scheme: in a given period of time, single models tend to receive a weight close to one, while DMA usually assigns positive weights to forecasts from all different DSGE models. In many periods, all models receive nearly equal weight when employing the DMA scheme. For the third combination approach, BPS, weights (corresponding to factor loadings), are positive and relatively similar across models for certain periods. However, during the financial crisis we observe individual negative factor loadings (see, for example, Figure 2 a)), implying a reversal of the sign of the prediction of the respective DSGE model in the combined forecast for these quarters.

Focusing on one-step-ahead weights, Figure 1 a) shows the results for the three different combination techniques for GDP growth. For DMA, we observe that CPS2010, CD2008, and SW2003 tend to dominate in terms of predictive ability prior to the financial crisis, with weights between 0.2 and 0.25. The weights of the other models are roughly equal, with a magnitude of approximately 0.075. During the financial crisis, the importance of these models in the combination decreases, and the SW2007 specification substantially gains in importance, having its maximum weight at around 0.8 in 2008Q3. In the subsequent years, the relevance of SW2007 within the combined predictions

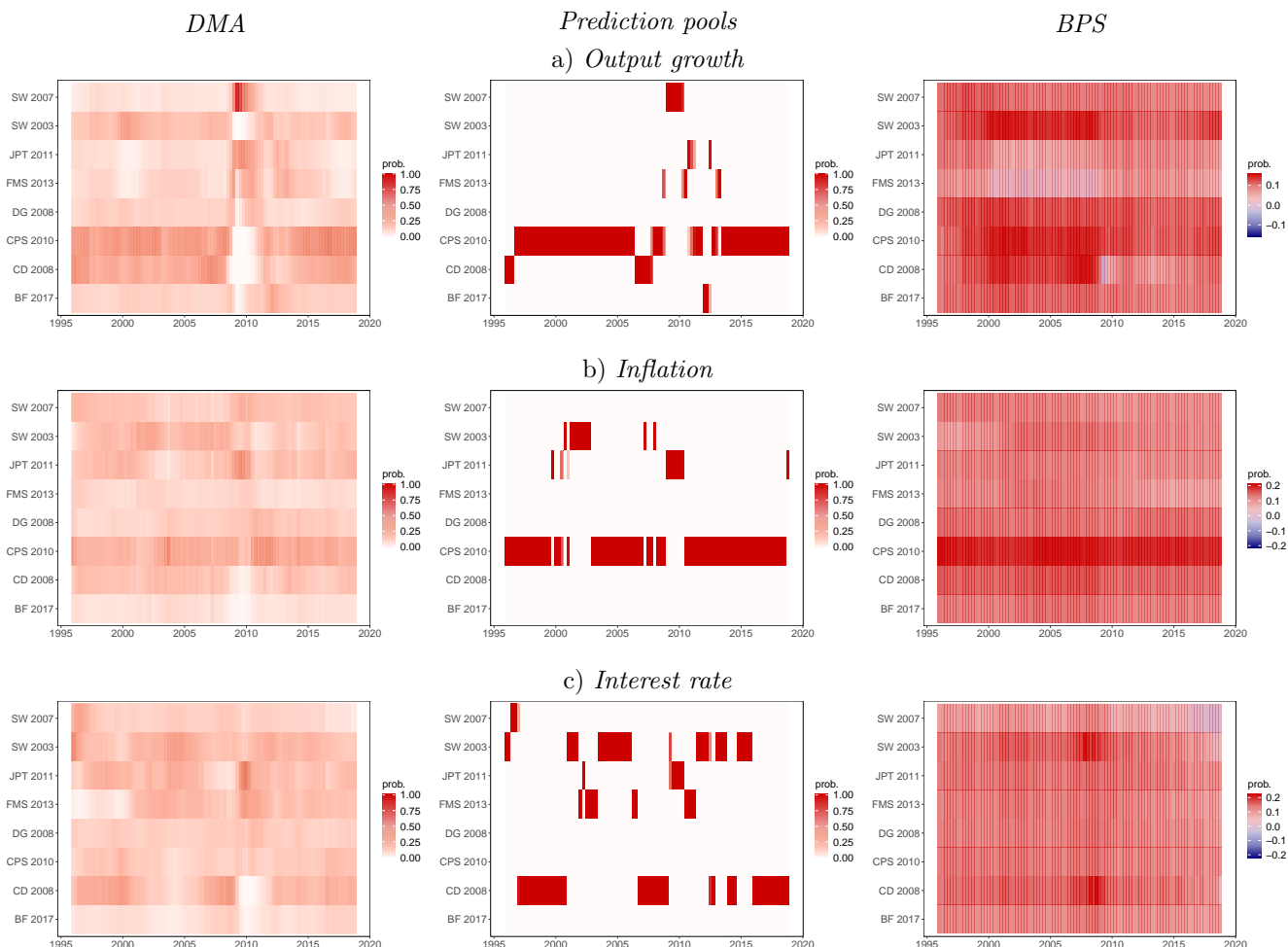


Figure 2: Posterior mean of model weights over the hold-out sample for four-steps-ahead predictions. The figure shows three different weighting schemes for the three target variables: output, inflation, and interest rate. Variables entering the DSGE models are detrended with the Hamilton filter.

decreases against other models, and in the aftermath of the financial crisis the weighting scheme becomes similar to that at the start of the hold-out sample. For prediction pooling, Figure 1 *a)* shows that before 2008 forecasts from the CPS2010 specification tend to dominate our weighted prediction scheme.

In contrast to the DMA approach, predictive pooling assigns a weight of unity to the forecasts of CPS2010, while DMA assigns a weight of approximately 0.25 to the best performing specification in this period. In line with the results for the DMA weighting scheme, the CPS2010 model loses importance during the financial crisis. In the crisis period, changes in weights occur more frequently, suggesting that no single model is systematically preferred over this time interval. Directly after the crisis and until the end of the hold-out sample, the FMS2013 specification dominates in terms of weights, followed by the CPS2010 model.

The right panel of Figure 1 *a)* shows the results for the weights corresponding to the BPS approach and reveals the relative importance of predictions of the CPS2010 model over the full period, closely followed by the SW2003 specification. Except for DG2008, which also receives a relatively high loading, all other models feature a weight between 0 and 0.15 at the beginning of the sample. After 2005, the CD2008 specification gains importance, but its

weight sharply drops at the outset of the financial crisis, in parallel with an increase of the weight of FMS2013. In the aftermath of the financial crisis, model loadings appear more stable.

Figure 1 *b)* depicts the dynamics of weighting schemes for inflation as a target variable. For DMA we observe weights ranging from 0.05 to 0.25 in all periods, although JPT2011 tends to attain the highest weight most of the time by a small margin. During the financial crisis period, weights appear more diverse. With prediction pools, in contrast to DMA, we see that most of the time the JPT2011 model tends to be selected, with a weight of unity. Until 2008, both SW2003 and SW2007 dominate in particular periods, while JPT2011 and DG2008 tend to be preferred after the financial crisis, in an alternating fashion. With BPS, CPS2010 receives the highest weight, ranging from 0.17 to 0.22, while all other specifications obtain a weight between 0.07 and 0.14.

Figure 1 *c)* depicts the 1-step-ahead model weights for interest rate prediction. In general, for the interest rate we observe a more persistent pattern in weighting schemes as compared to output growth and inflation. The DMA method leads to a near equal weighting scheme, with predictions obtained from CPS2010 receiving a relatively larger weight which substantially drops in the crisis. The results from prediction pools, in the middle panel of Figure 1, reflect the selection of the CPS2010 specification for practically the full period considered. The only exception is the very start of the hold-out sample and the financial crisis, in which SW2007, SW2003 and DG2008 receive a weight greater than zero. The BPS approach also shows a persistent evolution of the weights along the hold-out period. The weight assigned to predictions from CPS2010 tends to dominate that of the other specifications, while the FMS2013 model receives by far the lowest loading. Moreover, at the beginning of the sample, the SW2007 specification receives a high loading compared to other competing models, which falls substantially after the financial crisis.

For four-steps-ahead forecasts, Figure 2 *a)* shows a similar evolution of the weights for DMA combinations, with predictions from CD2008, CPS2010, and SW2003 dominating in the combination scheme at the beginning and the end of the hold-out sample, while the weight for SW2007 spikes during the financial crisis. For output, the combination chosen by prediction pooling leads to a more persistent weighting scheme as compared to one-step-ahead predictions. In tranquil periods, CPS2010 tends to be preferred, while in periods of turmoil the weights change more frequently. With the BPS combination approach, it is CPS2010 and SW2003 that dominate in most of the period, although the CD2008 specification gains relevance in the boom prior to the financial crisis. However, the predictions of this model experience a sharp decrease in its weight during the crisis. In the aftermath of the financial crisis, the weights are more stable. In Figure 2 *b)*, the same type of results in terms of the evolution of weighting schemes along the hold-out sample can be observed, with the pooling of three different models dominating throughout the entire hold-out sample and CPS2010 achieving the highest weight in most of the sample. The BPS combination method results for four-steps-ahead predictions of inflation show the highest loading for CPS2010 (ranging from 0.15 to 0.22), followed by CD2008. All other specifications receive a loading of similar magnitude.

Finally, the results for the interest rate indicate weights between 0.05 and 0.30 for the DMA method, without evidence for a clearly preferred model over the full period. The weight for CD2008 rises before the global financial crisis, but sharply drops afterwards. For the prediction pooling approach, there is a higher frequency of changing weights when compared to weights based on one-step-ahead predictions. BPS assigns the highest loadings to CD2008 and SW2003, with both weights reaching their maximum during the financial crisis. All other models receive loadings of around 0.1 over the full hold-out sample, except for the SW2007 specification, which always obtains the lowest loading.

The results of the analysis of the evolution of weight estimates for combinations of DSGE model predictions illustrate the stark differences existing across forecast pooling methods. These are not only reflected in the relative weight assigned to the predictions of different models, but also in the changes of these weights over time. The fact that the combination method based on prediction pools acts as a dynamic model selection device contrasts with the weighting schemes resulting from the other approaches entertained in the exercise, which tend to lead to composite predictions with positive weights for all specifications. The relative predictive performance of these approaches compared to each other, as well as to individual model forecasts, is explored in the next section.

4.2 Predictive ability improvements via DSGE combinations

In this section, we examine the predictive performance of the individual DSGE models and the forecast combinations by means of point and density forecasts. We calculate root mean squared error (RMSE) ratios as a point forecast

measure, and log predictive Bayes factors (BFs), as a density forecast measure for each observation in the hold-out-sample, using the forecast combination method based on equal weights as a benchmark.

Table 3 summarizes the predictive performance of forecasts based on the different weighting schemes across variables and forecast horizons. The table shows RMSE ratios and BFs, benchmarked against the equal weighting scheme. The results in Table 3 correspond to DSGE models whose variables have been detrended making use of the Hamilton filter.⁸

The results illustrate the predictive improvements that can be achieved making use of forecast combinations. On the one hand, the GDP growth forecasts obtained using prediction pooling methods present the best statistics in terms of point and density forecast accuracy for the two horizons considered. The use of weights based on dynamic factors leads to the lowest predictive error in interest rates for point forecasts and for density forecasts (at the four-steps-ahead horizon). For the case of inflation, the evidence for advantages of combining DSGE predictions is very limited, with clear improvements only visible for density forecasting using dynamic weights at the four-steps-ahead horizon. Inflation predictions from the CPS2010 lead to the most accurate point forecasts in our exercise for both horizons. This DSGE specification receives the highest weight in the dynamic weighting combination, which exhibits the best density forecasting statistic at the four-steps-ahead horizon.

Target Variable	Single								Combined		
	BF 2017	CD 2008	CPS 2010	DG 2008	FMS 2013	JPT 2011	SW 2003	SW 2007	DMA	POOL	BPS
<i>1-step-ahead</i>											
GDP growth	1.252 (-19.312)	1.284 (-13.685)	0.945 (11.696)	1.099 (-7.364)	0.996 (6.126)	1.073 (-5.446)	0.989 (3.973)	1.152 (-19.319)	0.945 (6.213)	0.918 (17.018)	1.021 (7.6)
Inflation	1.072 (-9.197)	1.108 (-13.42)	0.92 (4.517)	1.033 (4.984)	1.168 (-4.965)	1.122 (7.315)	1.15 (-0.746)	1.105 (0.903)	1.006 (1.573)	1.075 (5.841)	1.006 (5.042)
Interest rate	1.159 (-8.971)	1.168 (-3.001)	0.859 (16.035)	1.021 (-2.952)	1.35 (-6.031)	1.331 (-5.835)	1.104 (-0.622)	0.963 (5.292)	0.955 (2.827)	0.867 (15.577)	0.83 (10.242)
<i>4-step-ahead</i>											
GDP growth	0.989 (-5.331)	1.113 (-6.46)	0.972 (6.207)	1.047 (-6.634)	1.129 (-8.446)	1.139 (-13.356)	1.008 (-0.937)	1.098 (-17.979)	0.933 (6.675)	0.902 (18.67)	0.99 (7.763)
Inflation	1.424 (-20.706)	1.163 (0.864)	0.943 (13.122)	1.001 (-3.317)	1.286 (-11.716)	1.254 (4.279)	1.237 (4.713)	1.163 (1.932)	0.95 (3.764)	0.944 (13.037)	0.973 (17.031)
Interest rate	1.29 (-16.412)	1.103 (1.461)	1.01 (-1.211)	1.042 (-2.864)	1.135 (1.159)	0.986 (5.919)	0.993 (6.833)	1.022 (-6.629)	0.942 (3.666)	0.887 (11.908)	0.648 (29.397)

Table 3: Forecasting Performance of different DSGE models (with Hamilton filter detrending) and forecast combination approaches relative to the equal weight benchmark. The table shows RMSE ratios with BFs in parentheses. Bold numbers indicate the best performing models.

Figure 3 depicts the evolution of RMSE ratios benchmarked against the equal model weights alternative for Hamilton filter detrending, and Figure 4 shows their respective cumulative Bayes factors over the hold-out sample.⁹ Panel a) in Figure 3 depicts the evolution of one-step-ahead RMSE ratios. Over the full hold-out sample, all combination methods perform reasonably well and always produce lower RMSEs than the majority of single DSGE models. The mixed performance during the financial crisis, where weights tend to be particularly volatile, is also reflected in forecast comparisons based on predictive densities.

Focusing on the results for GDP growth predictions (the left panel of Figure 3 a)), CPS2010 presents the best predictive performance among individual DSGE specifications, with the SW2003 model also providing relatively accurate predictions. Interestingly, during the financial crisis, JPT2011, FMS2013, and SW2007 outperform the benchmark. For inflation predictions (middle panel of Figure 3 a)), the predictive accuracy of CD2008 and BF2017 is relatively poor in periods before and during the financial crisis. Over the full hold-out period, all four combination techniques lead to prediction accuracy improvements. For the interest rate (the right panel of Figure 3 a)), the BPS approach produces by far the lowest RMSEs, although other combination techniques perform reasonably well. For the interest rate, the flexibility of adjusting loadings using a dynamic factor model substantially pays off in terms of

⁸The results for models detrended with the HP filter and a “Const/HP” approach are depicted in the Appendix (Table A1 and Table A2).

⁹Results for other detrending methods can be found in the Appendix (see Figure A2, Figure A4, Figure A3 and Figure A5).

prediction quality. Focusing on the performance of each one of the DSGE models, the CD2008 specification performs particularly well at the beginning of our sample, but produces substantially higher RMSEs than the benchmark after the offset of the financial crisis.

The evolution of four-steps-ahead RMSEs is depicted in Figure 3 b), where it can be observed that up to the financial crisis SW2003 outperforms all other models for GDP growth prediction. However, the picture changes substantially during the financial crisis and in its aftermath. Here, predictions of the SW2003 model are outperformed by those of other DSGE models (for example, FMS2017 and SW2007). Forecast combination techniques tend to achieve low RMSE ratios after the financial crisis. For four-steps-ahead forecasts of inflation, the performance of single DSGE models is more mixed over the hold-out-sample. For instance, CD2008 leads to low RMSEs at the beginning of the sample, but its predictive performance is gradually reduced in the subsequent periods. On the other hand, DG2008 starts with a substantially higher RMSE ratio, but improves over the hold-out sample. Additionally, the middle panel of Figure 3 b) suggests that the use of prediction combination techniques leads to improvements in forecast accuracy. All four different techniques are capable to hedge against a poorer forecast performance of the BF2007 and other models after the crisis. The four-steps-ahead point forecast performance of the interest rate (right panel in Figure 3 b)) reveals predictive gains from the BPS approach, when compared to RMSE ratios based on one-step-ahead forecasts. The evolution over the hold-out sample features a quite persistent pattern in predictive ability.

As in the case of point forecasts, compared to individual DSGE forecasts, combinations based on DMA, prediction pooling, and BPS provide good results in terms of density forecasts. The left panel in Figure 4 a) indicates that, for GDP growth, CPS2010 produces the highest BFs compared to the rest of DSGE models. For many models, we observe a drop in predictive accuracy relative to the equal weighting benchmark during the financial crisis, a period where predictive pooling provides the best forecasting performance. BFs for one-step-ahead inflation forecasts (depicted in the middle panel of Figure 4 a)) indicate that predictions from JPT2011, the CPS2010, and SW2003 improve upon the benchmark over the hold-out sample. As a consequence, prediction pooling also performs well, since, as depicted in Figure 1, this combination scheme tends to favour at least one of these three models. At the other end of the spectrum, predictions from CD2008, FMS2011, and BF2017 perform poorly. The left panel of Figure 4 a) shows that that forecasts from the CPS2010 specification clearly dominate those of other DSGE models in terms of accuracy. Due to the dominance of a single model and the particular dynamic weighting scheme provided by prediction pooling, this technique achieves similar BFs to those of the best performing individual models.

The evolution of the BFs based on four-steps-ahead predictions is quite similar to that for point forecasts. In particular, for GDP growth, Figure 4 b) suggests that the same models dominate as in the case of shorter forecasting horizons. Prediction pooling substantially improves forecast accuracy after the crisis when compared to single DSGE models. For inflation, forecasts created using the BPS approach dominate all other predictions. Only prediction pooling, the CPS2010 and the SW2003 models are capable of competing with the BPS approach. The four-steps-ahead predictions of the interest rate (Figure 4 c)) indicate that BPS performs best, followed by the predictive pooling approach. Focusing on the performance of individual specifications, CPS2010 and JPT2011 achieve reasonable predictive accuracy, as reflected in their BFs.

The selection of a detrending method for the macroeconomic variables in DSGE models plays an important role in determining predictive ability for density forecasts of the interest rate (see Appendix). While the detrending approach introduced in Hamilton (2018) does not tend to lead to abrupt changes in forecasting performance, the use of HP filtering results in serious decreases of forecasting accuracy for many models, in particular during the financial crisis. The forecasting ability for inflation of the model by Smets and Wouters (2003), for instance, is seriously affected by the choice of a particular detrending approach. While the model does well in comparison to the benchmark for variables detrended using the method in Hamilton (2018), the performance is very deficient if the variables are detrended making use of the HP filter.

5 Conclusions

The results of our analysis show that combining forecasts from DSGE models tends to lead to improvements in predictive ability for macroeconomic variables. With only a few exceptions, predictive weighting schemes are able

to reach superior forecasting performance over individual DSGE specifications. The noteworthy exceptions are predictions of inflation making use of the model by Cogley et al. (2010), especially in shorter horizons.

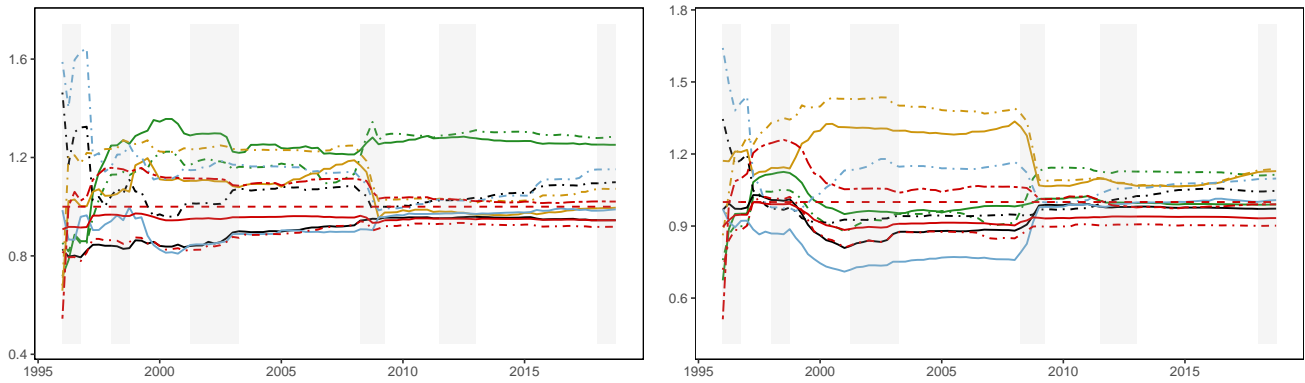
The weighting schemes implied by the combination methods employed are fundamentally different across techniques. Weighting based on prediction pools tends to lead to forecasts based on dynamic model selection, assigning zero weights to many individual model predictions over the out-of-sample period. DMA and weighting based on dynamic factors, on the other hand, results in combined forecasts with positive weights for all DSGE specifications.

The forecasting performance of individual DSGE models systematically worsens during the financial crisis. In contrast, the use of weighting across DSGE specifications tends to improve their forecasting ability even in crisis times. A special case is the model by De Graeve (2008), which exhibits an improvement in forecasting performance during this period, a result that can be explained by the fact that the model incorporates financial friction mechanisms in the spirit of Bernanke et al. (1999).

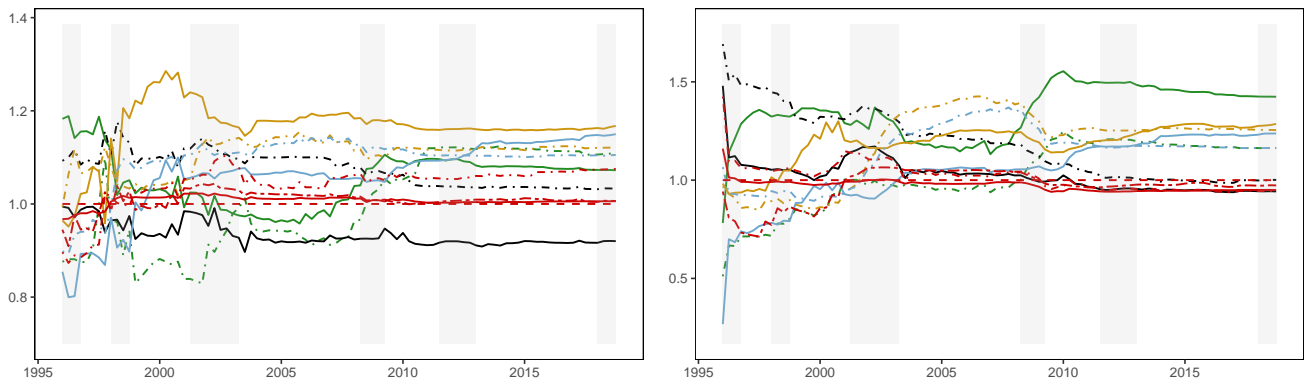
1-step-ahead

4-step-ahead

a) Output growth



b) Inflation



c) Interest rate

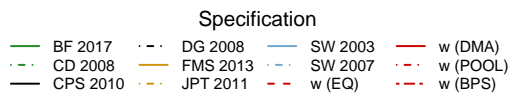
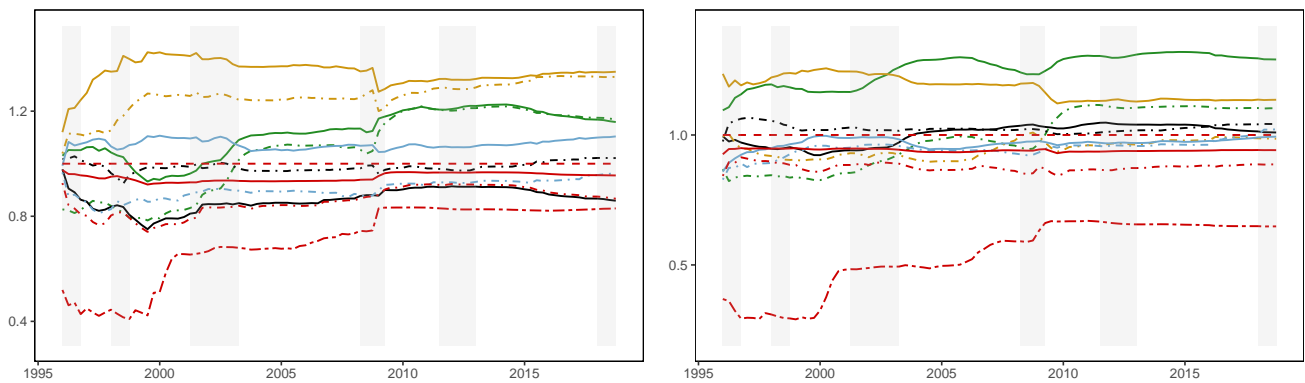
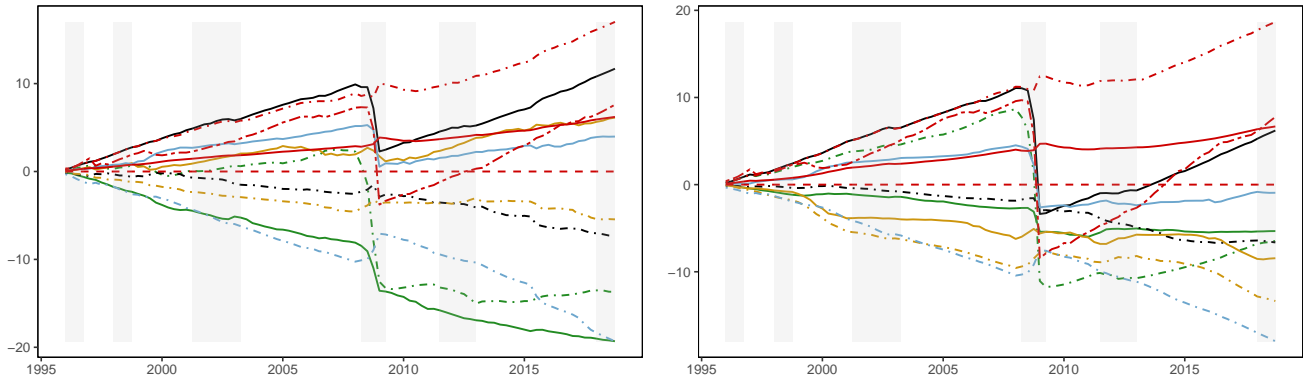


Figure 3: Evolution of root mean squared errors relative to the predictive ability of the equal weighting scheme. Hamilton filter detrending and four different weighting schemes. The gray shaded areas indicate OECD recessions for the euro area.

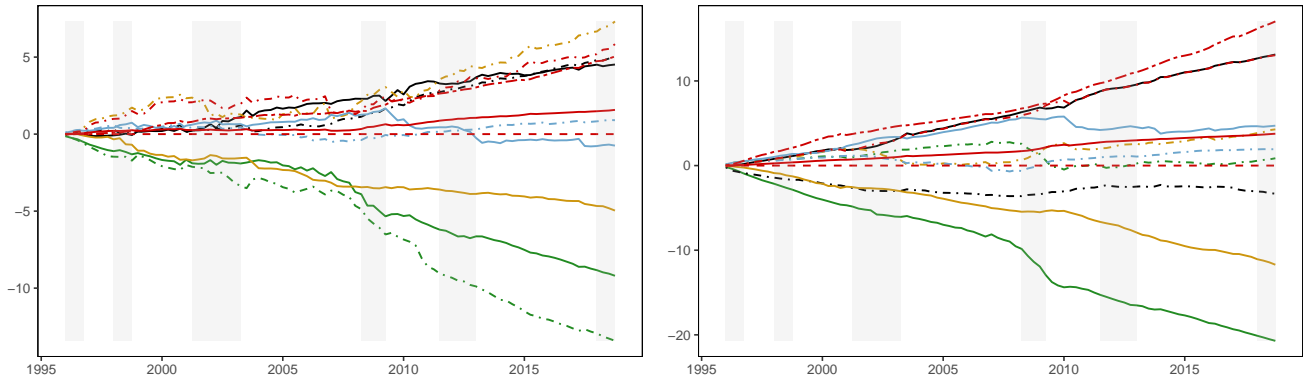
1-step-ahead

4-step-ahead

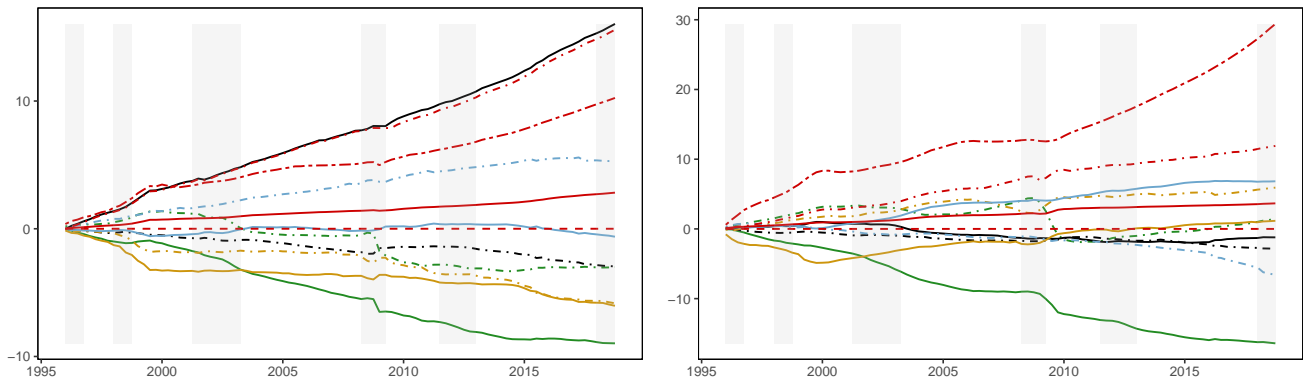
a) Output growth



b) Inflation



c) Interest rate



Specification
— BF 2017 - - - DG 2008 - - - SW 2003 - - - w (DMA)
- - - CD 2008 - - - FMS 2013 - - - SW 2007 - - - w (POOL)
— CPS 2010 - - - JPT 2011 - - - w (EQ) - - - w (BPS)

Figure 4: Evolution of Bayes factors relative to the predictive ability of the equal weighting scheme. Hamilton filter detrending and four different weighting schemes. The gray shaded areas indicate OECD recessions for the euro area.

References

- Aastveit KA, Ravazzolo F and Van Dijk HK (2018) Combined density nowcasting in an uncertain economic environment. *Journal of Business & Economic Statistics* 36(1), 131–145
- Adjemian S, Bastani H, Juillard M, Mihoubi F, Perendia G, Ratto M and Villemot S (2011) Dynare: Reference manual, version 4
- Amisano G and Geweke J (2017) Prediction using several macroeconomic models. *Review of Economics and Statistics* 99(5), 912–925
- Bates JM and Granger CW (1969) The combination of forecasts. *Journal of the Operational Research Society* 20(4), 451–468
- Benchimol J and Fourçans A (2017) Money and monetary policy in the eurozone: an empirical analysis during crises. *Macroeconomic Dynamics* 21(3), 677–707
- Bernanke BS, Gertler M and Gilchrist S (1999) Chapter 21 The financial accelerator in a quantitative business cycle framework. volume 1 of *Handbook of Macroeconomics*. Elsevier, 1341 – 1393
- Brand T and Toulemonde N (2015) Automating update of the Smets and Wouters (2003) database. Technical report
- Carter CK and Kohn R (1994) On Gibbs sampling for state space models. *Biometrika* 81(3), 541–553
- Christensen I and Dib A (2008) The financial accelerator in an estimated New Keynesian model. *Review of Economic Dynamics* 11(1), 155–178
- Christiano LJ, Motto R and Rostagno M (2014) Risk Shocks. *American Economic Review* 104(1), 27 – 65
- Christoffel K, Coenen G and Warne A (2011) Forecasting With DSGE Models. In *The Oxford Handbook of Economic Forecasting*. 89 – 126
- Cogley T, Primiceri GE and Sargent TJ (2010) Inflation-Gap Persistence in the US. *American Economic Journal: Macroeconomics* 2(1), 43–69
- De Graeve F (2008) The external finance premium and the macroeconomy: US post-WWII evidence. *Journal of Economic Dynamics and Control* 32(11), 3415–3440
- Del Negro M, Hasegawa RB and Schorfheide F (2016) Dynamic prediction pools: An investigation of financial frictions and forecasting performance. *Journal of Econometrics* 192(2), 391–405
- Delle Chiaie S (2009) The sensitivity of DSGE models’ results to data detrending. Technical Report 157, Oesterr. Nationalbank Working Paper Series
- Fagan G, Henry J and Mestre R (2005) An area-wide model for the euro area. *Economic Modelling* 22(1), 39 – 59
- Fève P, Matheron J and Sahuc JG (2013) A Pitfall with Estimated DSGE-Based Government Spending Multipliers. *American Economic Journal: Macroeconomics* 5(4), 141–178
- Frühwirth-Schnatter S (1994) Data augmentation and dynamic linear models. *Journal of time series analysis* 15(2), 183–202
- Gelman A and Rubin DB (1992) Inference from iterative simulation using multiple sequences. *Statistical science* 7(4), 457–472
- Geweke J and Amisano G (2011) Optimal prediction pools. *Journal of Econometrics* 164(1), 130–141

- Geweke J and Amisano G (2012) Prediction with misspecified models. *American Economic Review* 102(3), 482–86
- Gorodnichenko Y and Ng S (2010) Estimation of DSGE models when the data are persistent. *Journal of Monetary Economics* 57(3), 325 – 340
- Hall SG and Mitchell J (2007) Combining density forecasts. *International Journal of Forecasting* 23(1), 1–13
- Hamilton JD (2018) Why You Should Never Use the Hodrick-Prescott Filter. *The Review of Economics and Statistics* 100(5), 831–843
- Justiniano A, Primiceri GE and Tambalotti A (2011) Investment shocks and the relative price of investment. *Review of Economic Dynamics* 14(1), 102–121
- Kolasa M and Rubaszek M (2015) Forecasting using DSGE models with financial frictions. *International Journal of Forecasting* 31(1), 1–19
- Koop G and Korobilis D (2012) Forecasting inflation using dynamic model averaging. *International Economic Review* 53(3), 867–886
- McAlinn K, Aastveit KA, Nakajima J and West M (2019) Multivariate Bayesian Predictive Synthesis in Macroeconomic Forecasting. *Journal of the American Statistical Association* , 1–19
- McAlinn K and West M (2019) Dynamic Bayesian predictive synthesis in time series forecasting. *Journal of Econometrics* 210(1), 155–169
- Pettenuzzo D and Ravazzolo F (2016) Optimal Portfolio Choice Under Decision-Based Model Combinations. *Journal of Applied Econometrics* 31(7), 1312–1332
- Raftery AE, Kárný M and Ettler P (2010) Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics* 52(1), 52–66
- Smets F and Wouters R (2003) An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area. *Journal of the European Economic Association* 1(5), 1123–1175
- Smets F and Wouters R (2007) Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach. *American Economic Review* 97(3), 586–606
- Stock JH and Watson MW (1999) Forecasting inflation. *Journal of Monetary Economics* 44(2), 293 – 335
- Timmermann A (2006) Forecast combinations. *Handbook of Economic Forecasting* 1, 135–196
- Wolters MH (2015) Evaluating point and density forecasts of DSGE models. *Journal of Applied Econometrics* 30(1), 74–96

Appendix

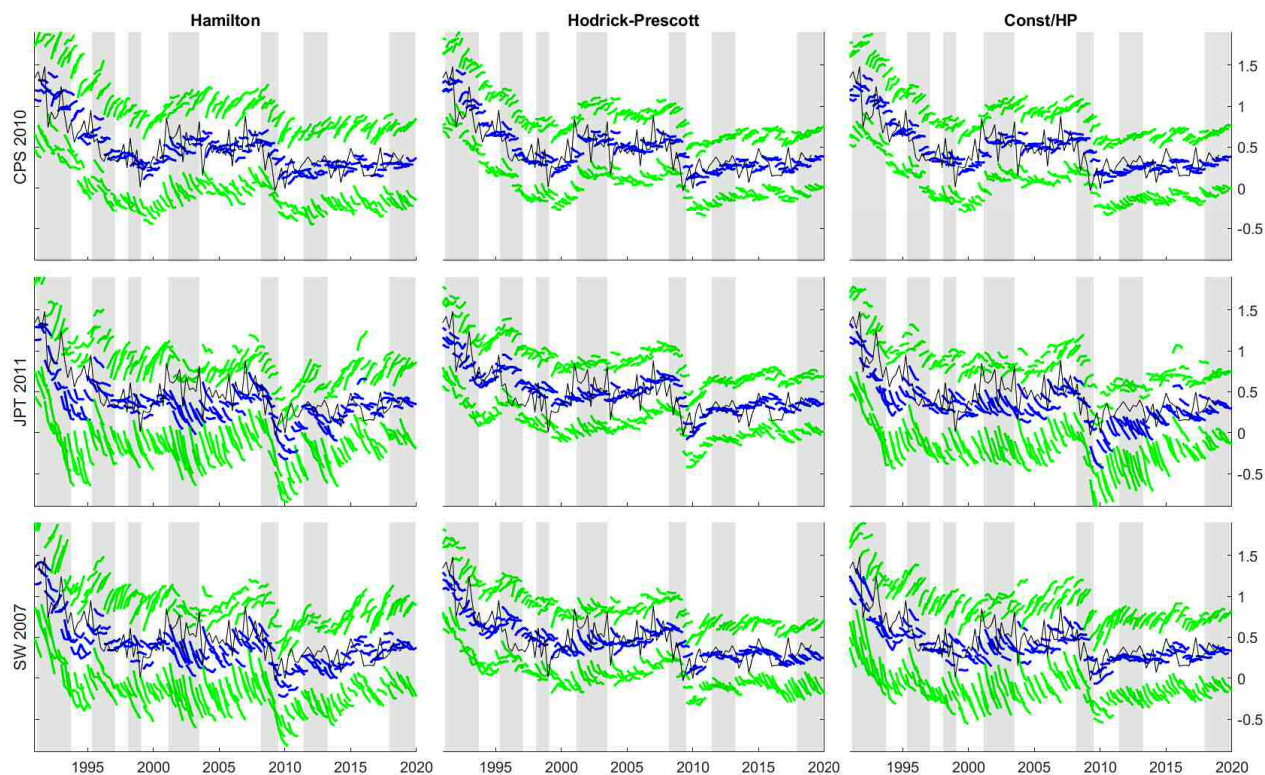


Figure A1: 1- to 4-step-ahead out-of-sample recursive forecasts of inflation for selected models and detrending approaches. Blue lines denote median forecasts, green lines denote 10th and 90th percentiles of the predictive density. The predictions are depicted for three models (CPS2010, JPT2011 and SW2007, in rows) and the three different detrending approaches (in columns). The forecasts of the CPS2010 model appear particularly robust to the use of different detrending methods, while those corresponding to the JPT2011 and SW2007 model exhibit particularly wide predictive densities when Hamilton or Const/HP detrending approaches are used. Detrending observable variables with the Hodrick-Prescott filter leads to predictive densities with the lowest variance across the three selected models.

Target Variable	Single								Combined		
	BF 2017	CD 2008	CPS 2010	DG 2008	FMS 2013	JPT 2011	SW 2003	SW 2007	DMA	POOL	BPS
<i>1-step-ahead</i>											
GDP growth	1.212 (-21.215)	1.126 (-16.98)	0.983 (-1.509)	1.003 (-3.993)	0.942 (2.512)	1.005 (-3.13)	0.973 (0.043)	0.996 (-6.924)	0.959 (5.963)	0.963 (11.83)	1.008 (-10.456)
Inflation	1.078 (-5.547)	0.958 (0.82)	0.936 (3.128)	1.083 (-2.306)	1.084 (-0.485)	1.064 (1.205)	1.189 (-6.516)	1.03 (2.065)	0.986 (0.712)	1.004 (2.313)	0.98 (8.083)
Interest rate	1.081 (-10.145)	1.12 (-4.51)	0.929 (9.287)	1.033 (-6.67)	1.08 (-3.149)	1.101 (-4.815)	1.052 (-0.201)	0.964 (1.667)	0.95 (2.643)	0.901 (9.599)	0.917 (6.82)
<i>4-step-ahead</i>											
GDP growth	1.048 (-12.534)	1.02 (-22.054)	1.047 (-24.706)	0.982 (-2.489)	0.999 (-7.438)	0.992 (-5.491)	1 (-5.656)	0.977 (-1.829)	0.965 (6.534)	0.945 (13.047)	0.881 (0.772)
Inflation	1.284 (-18.685)	0.986 (5.761)	0.953 (7.27)	1.096 (-17.986)	1.035 (7.677)	1.04 (6.294)	1.224 (-3.899)	1.033 (1.467)	0.959 (3.726)	0.924 (9.098)	0.918 (14.43)
Interest rate	1.12 (-12.335)	1.022 (-11.172)	0.968 (1.499)	1.011 (-2.422)	1.011 (-11.224)	1.022 (-10.329)	1.044 (-9.346)	1 (-3.887)	0.921 (11.094)	0.888 (18.24)	0.655 (29.778)

Table A1: Forecasting Performance of different DSGE models (with HP filter detrending) and forecast combination approaches relative to the equal weight benchmark. The table shows RMSE ratios with BFs in parentheses. Bold numbers indicate the best performing models. Bold numbers indicate the best performing models.

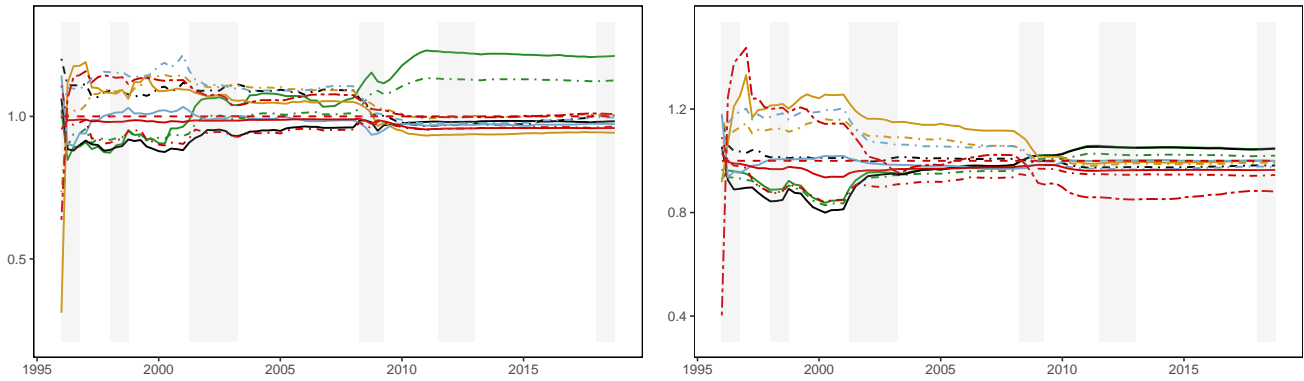
Target Variable	Single								Combined		
	BF 2017	CD 2008	CPS 2010	DG 2008	FMS 2013	JPT 2011	SW 2003	SW 2007	DMA	POOL	BPS
<i>1-step-ahead</i>											
GDP growth	1.181 (-16.357)	1.299 (-18.15)	0.922 (9.999)	1.041 (-5.544)	0.958 (4.244)	1.071 (-5.598)	0.955 (3.874)	1.158 (-16.417)	0.95 (4.903)	0.928 (15.874)	1.013 (4.719)
Inflation	1.049 (-2.938)	0.992 (1.148)	0.919 (10.07)	1.001 (7.401)	1.192 (-8.754)	1.145 (-3.086)	1.367 (-15.555)	1.06 (0.993)	0.98 (1.996)	0.945 (8.211)	0.995 (5.54)
Interest rate	1.124 (-7.433)	1.092 (-1.952)	0.973 (10.165)	1.104 (-5.335)	1.193 (-3.262)	1.195 (-5.611)	1.088 (-2.09)	0.947 (2.085)	0.97 (2.215)	0.927 (9.686)	0.928 (5.43)
<i>4-step-ahead</i>											
GDP growth	0.972 (-2.588)	1.107 (-9.291)	0.974 (3.594)	0.956 (-1.336)	1.079 (-10.811)	1.118 (-10.656)	1.1 (-7.656)	1.052 (-13.688)	0.943 (5.95)	0.903 (16.977)	0.977 (2.406)
Inflation	1.243 (-7.095)	1.083 (13.559)	0.947 (21.723)	0.906 (6.775)	1.416 (-23.708)	1.288 (-4.33)	2.264 (-39.408)	0.976 (0.375)	0.864 (10.07)	0.942 (20.415)	0.937 (17.535)
Interest rate	1.246 (-15.504)	1.097 (-1.901)	1.1 (-3.055)	1.087 (-6.751)	1.095 (-2.84)	0.986 (2.982)	1.109 (-7.199)	1.017 (-7.411)	0.89 (7.709)	0.822 (17.671)	0.755 (22.121)

Table A2: Forecasting Performance of different DSGE models (with “hybrid” detrending) and forecast combination approaches relative to the equal weight benchmark. The table shows RMSE ratios with BFs in parentheses. Bold numbers indicate the best performing models.

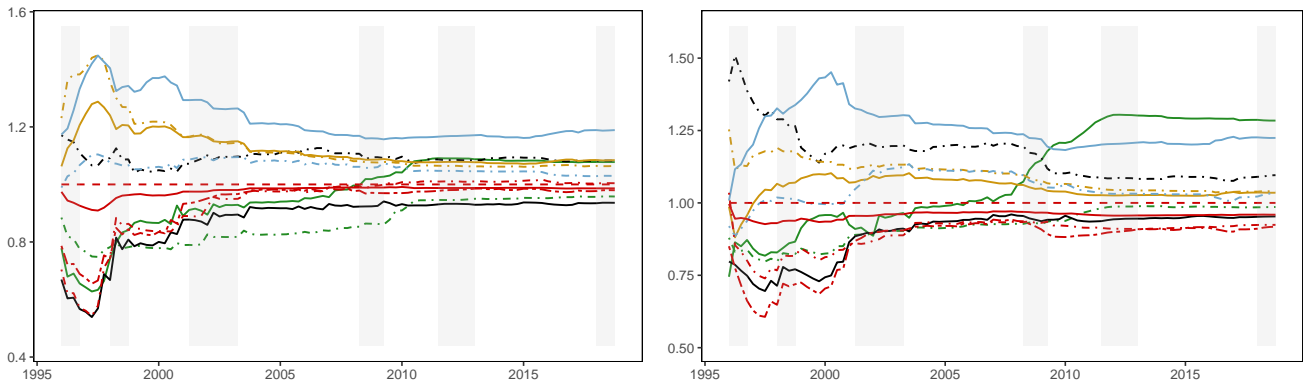
1-step-ahead

4-step-ahead

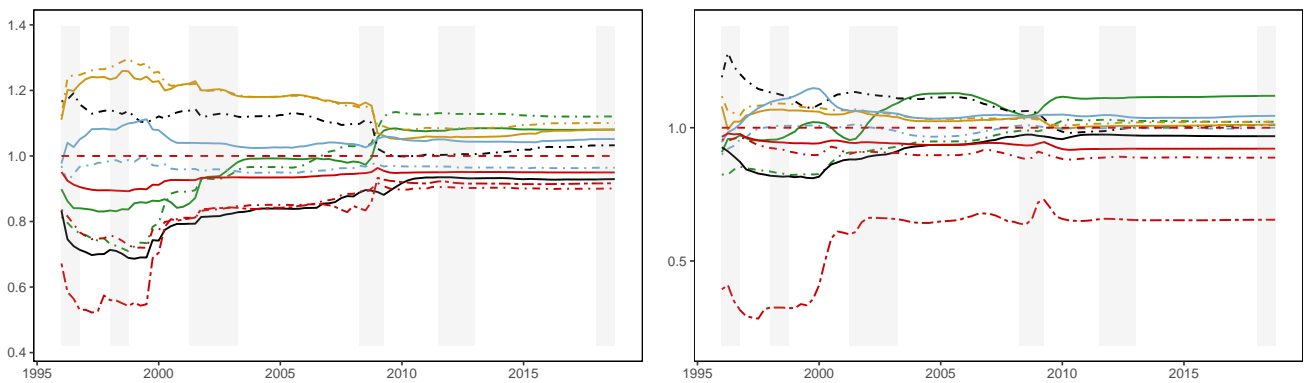
a) Output growth



b) Inflation



c) Interest rate



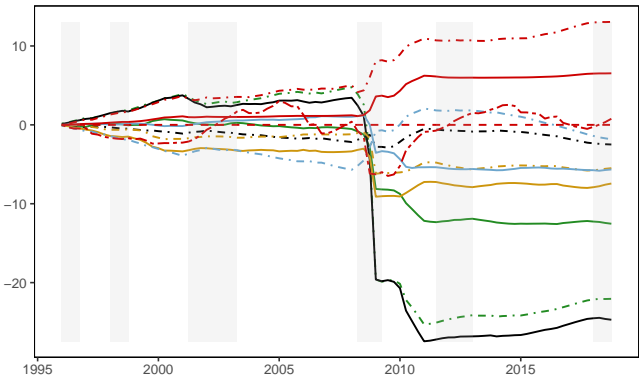
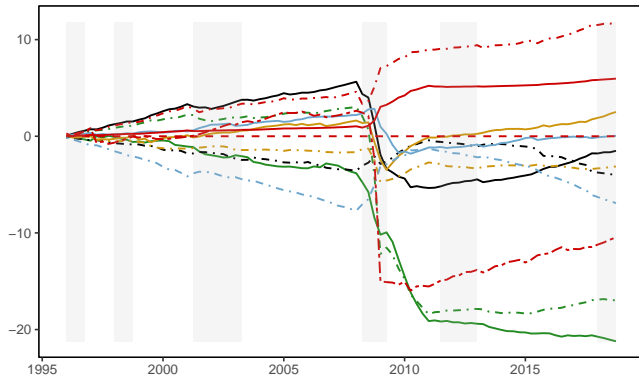
Specification
— BF 2017 - - - DG 2008 - - - SW 2003 - - - w (DMA)
- - - CD 2008 - - - FMS 2013 - - - SW 2007 - - - w (POOL)
— CPS 2010 - - - JPT 2011 - - - w (EQ) - - - w (BPS)

Figure A2: Evolution of root mean squared errors relative to the $w^{(EQ)}$ weighting scheme. First we average the squared errors over the hold-out and then we take the square root. The figure depicts different DSGE models with HP filter detrending and four different weighting schemes. The gray shaded areas indicate OECD recessions for the EA.

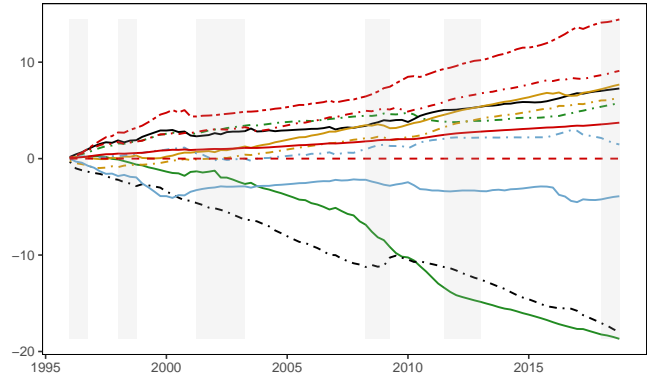
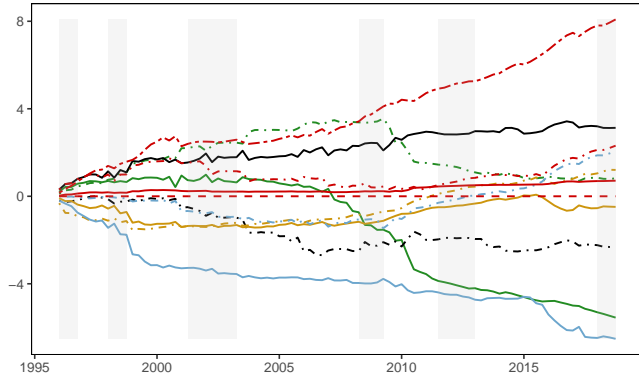
1-step-ahead

4-step-ahead

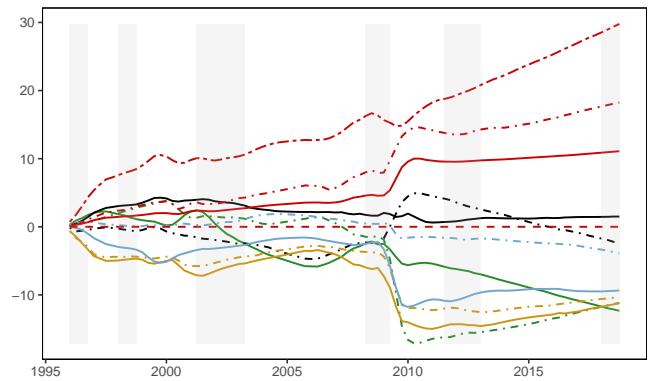
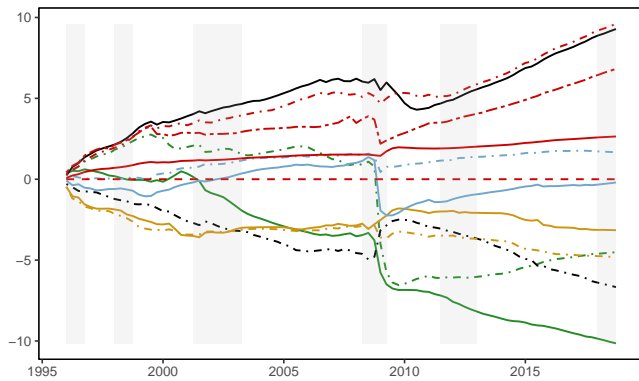
a) Output growth



b) Inflation



c) Interest rate



Specification

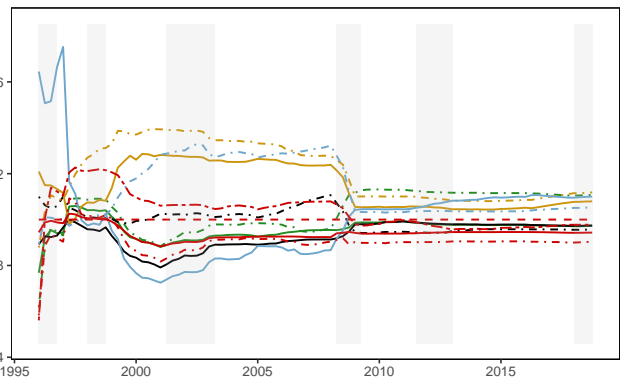
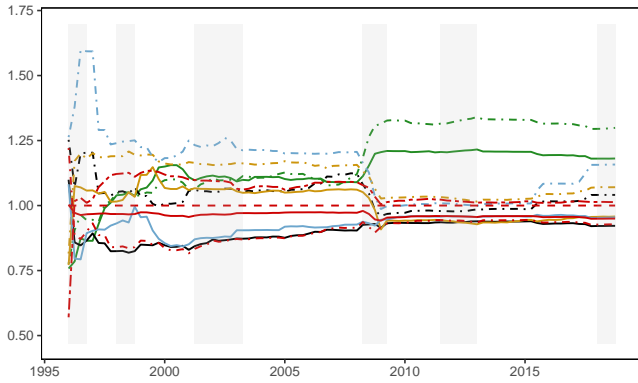
— BF 2017	- - - DG 2008	— SW 2003	— w (DMA)
- · - · CD 2008	— FMS 2013	- · - · SW 2007	- · - · w (POOL)
— CPS 2010	- · - · JPT 2011	- · - · w (EQ)	- · - · w (BPS)

Figure A3: Evolution of Bayes factors relative to the $w^{(EQ)}$ weighting scheme. The log predictive Bayes factors are cumulated over the hold-out. The figure depicts different DSGE models with HP filter detrending and four different weighting schemes. The gray shaded areas indicate OECD recessions for the EA.

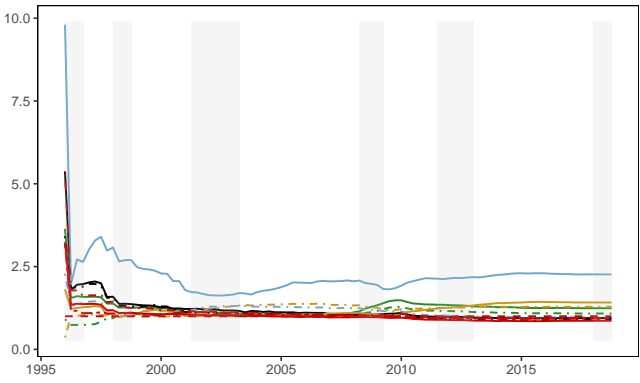
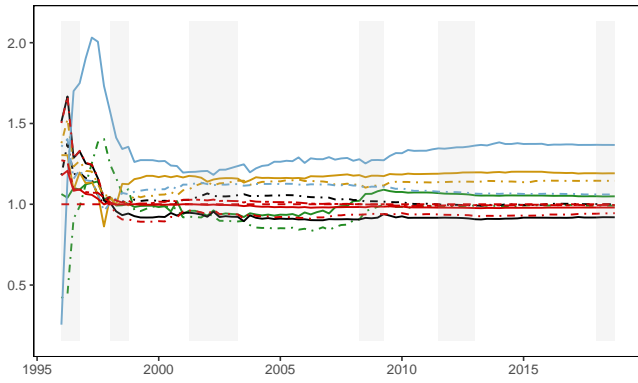
1-step-ahead

4-step-ahead

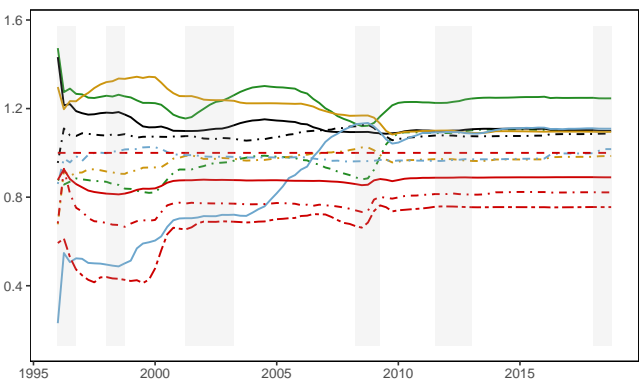
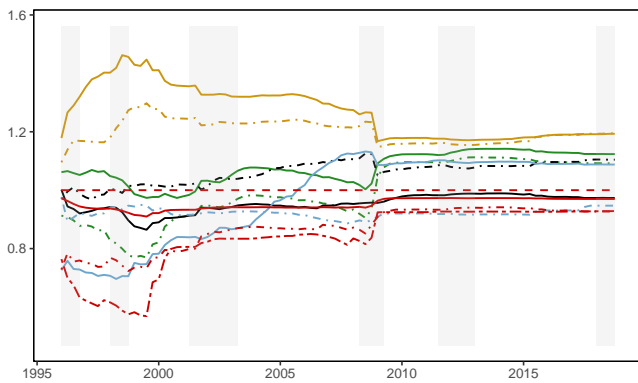
a) Output growth



b) Inflation



c) Interest rate



Specification

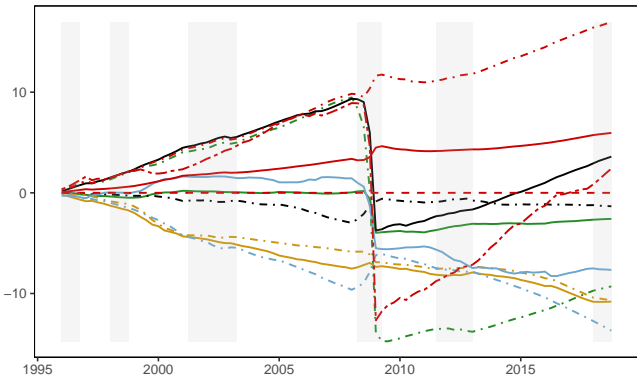
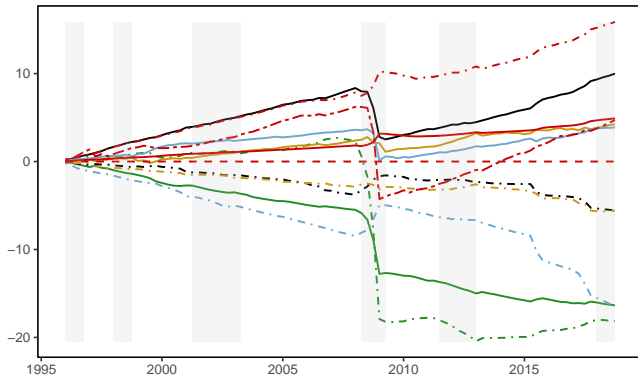
— BF 2017	- - - DG 2008	— SW 2003	— w (DMA)
- - - CD 2008	— FMS 2013	- - - SW 2007	- - - w (POOL)
— CPS 2010	- - - JPT 2011	- - - w (EQ)	- - - w (BPS)

Figure A4: Evolution of root mean squared errors relative to the $w^{(EQ)}$ weighting scheme. First we average the squared errors over the hold-out and then we take the square root. The figure depicts different DSGE models with “Const/HP” detrending and four different weighting schemes. The gray shaded areas indicate OECD recessions for the EA.

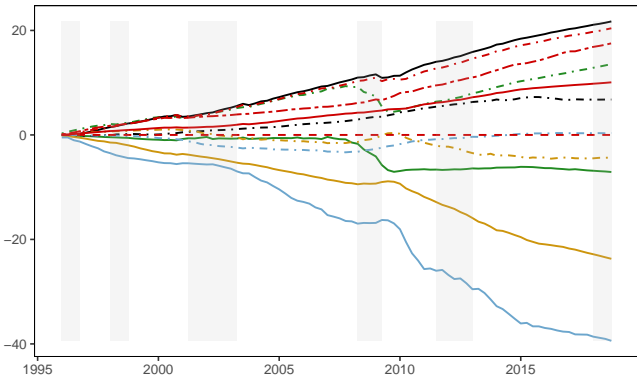
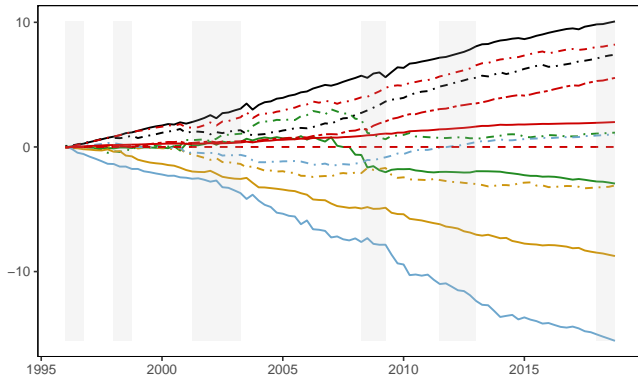
1-step-ahead

4-step-ahead

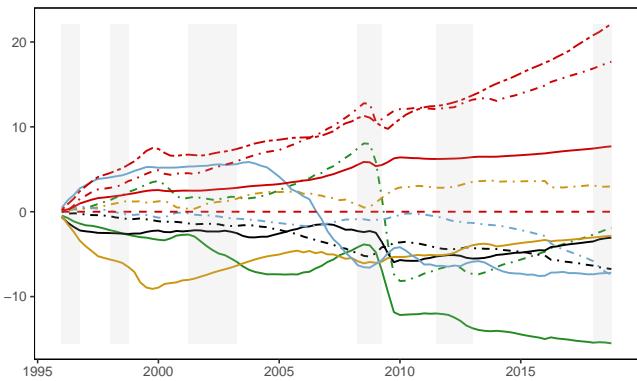
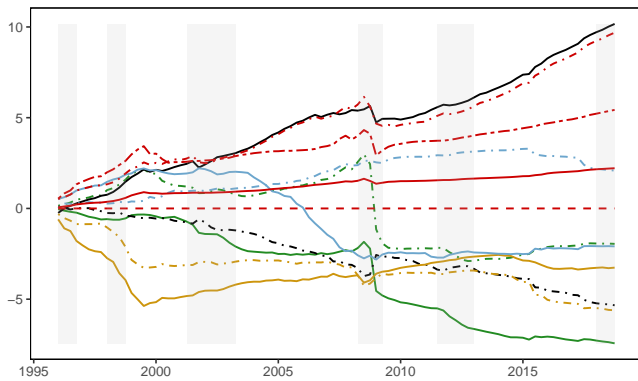
a) Output growth



b) Inflation



c) Interest rate



Specification

— BF 2017	- - - DG 2008	- · - SW 2003	— w (DMA)
- · - CD 2008	— FMS 2013	- · - SW 2007	- · - w (POOL)
— CPS 2010	- · - JPT 2011	- · - w (EQ)	- · - w (BPS)

Figure A5: Evolution of Bayes factors relative to the $w^{(EQ)}$ weighting scheme. The log predictive Bayes factors are cumulated over the hold-out. The figure depicts different DSGE models with “Const/HP” detrending and four different weighting schemes. The gray shaded areas indicate OECD recessions for the EA.