Online Appendix to
Density Forecast Combinations: The Real-Time Dimension

Peter McAdam[#] and Anders Warne*

October 31, 2023

Section A: Bayesian VAR Models with Homoskedastic Innovations

The General Normal-Inverted Wishart BVAR Model

Let y_t be an n-dimensional vector of observable variables such that its VAR representation is given by

$$y_t = \Phi_0 + \sum_{j=1}^{p} \Phi_j y_{t-j} + \epsilon_t, \quad t = 1, \ldots, T, \tag{A.1}$$

where $\epsilon_t \sim N_n(0, \Omega)$ and Φ_j are $n \times n$ matrices for $j \geq 1$ and an $n \times 1$ vector if $j = 0$. Let $X_t = [1 \; y_t' \; \cdots \; y_{t-p+1}']'$ be an $(np+1)$-dimensional vector, while the $n \times (np+1)$ matrix $\Phi = [\Phi_0 \; \Phi_1 \; \cdots \; \Phi_p]$ such that the VAR can be expressed as:

$$y_t = \Phi X_{t-1} + \epsilon_t. \tag{A.2}$$

Stacking the VAR system as $y = [y_1 \; \cdots \; y_T]$, $X = [X_0 \; \cdots \; X_{T-1}]$ and $\epsilon = [\epsilon_1 \; \cdots \; \epsilon_T]$, we can express this as

$$y = \Phi X + \epsilon, \tag{A.3}$$

with log-likelihood

$$\log p\big(y\big|X_0; \Phi, \Omega\big) = -\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log\big|\Omega\big| - \frac{1}{2}\mathrm{tr}\big[\Omega^{-1}\epsilon\epsilon'\big], \tag{A.4}$$

where, for convenience, we use the same notation for the random variables as their realizations.

The normal-inverted Wishart prior for (Φ, Ω) is given by

$$\mathrm{vec}\big(\Phi\big)\big|\Omega, \alpha \sim N_{n(np+1)}\big(\mathrm{vec}\big(\mu_\Phi\big), \big[\Omega_\Phi \otimes \Omega\big]\big), \tag{A.5}$$

$$\Omega\big|\alpha \sim IW_n\big(A, v\big), \tag{A.6}$$

Acknowledgements: We thank two anonymous referees and editor Massimiliano Marcellino. We have also benefitted from discussions with and suggestions by Gianni Amisano, Romain Aumond, Dean Croushore, Szabolcs Deak, Mátyás Farkas, Domenico Giannone, Gary Koop, Michele Lenza, Giorgio Primiceri, Bernd Schwaab and Allan Timmermann, as well as from participants at seminars at the European Central Bank and at the University of Kent, Canterbury. The opinions expressed in this paper are those of the authors and do not necessarily reflect the views of the European Central Bank or the Eurosystem, or the views of the Federal Reserve Bank of Kansas City or the Federal Reserve System.

Economic Research Department, Federal Reserve Bank of Kansas City, 1 Memorial Drive, Kansas City, Missouri 64198, USA; peter.mcadam@kc.frb.org; Phone: +1-816 585 0118.

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

where the prior parameters $(\mu_\Phi, \Omega_\Phi, A, v)$ are determined through a vector of hyperparameters, denoted by α.

The conjugate normal-inverted Wishart prior gives us a normal posterior for $\Phi|\Omega, \alpha$ and an inverted Wishart posterior for $\Omega|\alpha$. Specifically,

$$\operatorname{vec}\big(\Phi\big)\big|\Omega, y, X_0, \alpha \sim N_{n(np+1)}\big(\operatorname{vec}\big(\bar{\Phi}\big), \big[(XX' + \Omega_\Phi^{-1})^{-1} \otimes \Omega\big]\big), \tag{A.7}$$

$$\Omega\big|y, X_0, \alpha \sim IW_n\big(S, T + v\big). \tag{A.8}$$

Furthermore, it follows that the log marginal likelihood conditional on α is given by

$$\begin{aligned}\log p\big(y\big|X_0, \alpha\big) = &-\frac{nT}{2}\log(\pi) + \log\Gamma_n(T+v) - \log\Gamma_n(v) - \frac{n}{2}\log\big|\Omega_\Phi\big| \\ &+ \frac{v}{2}\log\big|A\big| - \frac{n}{2}\log\big|XX' + \Omega_\Phi^{-1}\big| - \frac{T+v}{2}\log\big|S\big|,\end{aligned} \tag{A.9}$$

where $\Gamma_b(a) = \prod_{i=1}^{b}\Gamma([a-i+1]/2)$ for positive integers a and b with $a \geq b$, while $\Gamma(\cdot)$ is the gamma function.

Prior A: Minnesota with Sum-Of-Coefficients and Dummy-Initial-Observation

In this paper we consider two ways of parameterizing the prior parameters $(\mu_\Phi, \Omega_\Phi, A, v)$. The first approach, called Prior A, is based on Giannone et al. (2015) with a Minnesota prior combined with the standard sum-of-coefficients prior by Doan et al. (1984), and the dummy-initial-observation prior by Sims (1993). As pointed out by Sims and Zha (1998), the latter part of the prior was designed to neutralize the bias against cointegration due to the sum-of-coefficients prior, while still treating the issue of overfitting of the deterministic component; see also Sims (2000).

Specifically, Prior A can be implemented through $T_d = n(p+2)+1$ dummy observations by prepending the y $(n \times T)$ and X $(np+1 \times T)$ matrices with the following:

$$\begin{aligned} y_{(d)} &= \begin{bmatrix} \lambda_o^{-1}\operatorname{diag}\big(\psi \odot \omega\big) & 0_{n\times n(p-1)} & \operatorname{diag}(\omega) & \delta^{-1}\bar{y}_0 & \mu^{-1}\operatorname{diag}\big(\psi \odot \bar{y}_0\big) \end{bmatrix}, \\ X_{(d)} &= \begin{bmatrix} 0_{1\times np} & 0_{1\times n} & \delta^{-1} & 0_{1\times n} \\ \lambda_o^{-1}\big(j_p \otimes \operatorname{diag}(\omega)\big) & 0_{np\times n} & \delta^{-1}\big(\imath_p \otimes \bar{y}_0\big) & \mu^{-1}\big(\imath_p \otimes \operatorname{diag}(\bar{y}_0)\big) \end{bmatrix}, \end{aligned} \tag{A.10}$$

where $\odot$ is the Hadamard product, i.e., element-by-element multiplication. The vector $\imath_p$ is a p-dimensional unit vector, while the $p \times p$ matrix $j_p = \operatorname{diag}[1 \cdots p]$. Notice that the first $n(p+1)$ columns of the matrices in (A.10) cover the Minnesota prior, the following column is the dummy-initial-observation prior, while the remaining n columns determine the sum-of-coefficients prior.

The hyperparameter $\lambda_o > 0$ gives the overall tightness in the Minnesota prior, the cross-equation tightness is set to unity, while the harmonic lag decay hyperparameter is equal to 2. The hyperparameter δ captures shrinkage for the dummy-initial-observation prior, where $\delta \to \infty$ gives the standard diffuse prior for Φ_0. The hyperparameter μ similarly determines shrinkage for the sum-of-coefficients prior, while the vector ω handles scaling issues. In this paper we focus on forecasting and let each element of ω be given by the estimated innovation standard deviation from AR processes of order p for the corresponding observed variable. The vector ψ is the prior mean of the diagonal

of Φ_1, and $\bar{y}_0$ is given by the pre-sample mean of y_t, i.e., $\bar{y}_0 = (1/p)\sum_{j=1}^{p} y_{j-p}$. This is consistent with the treatment in Bańbura et al. (2010) and Giannone et al. (2019).[1]

The vector ψ is given by ι_n under the orthodox Minnesota prior (random walk prior mean), but can also be given by, for instance, a 0-1 vector as in Bańbura et al. (2010), where ψ_i is set to unity if y_{it} is a levels variable and to zero if it is a first differenced variable. For the SoC prior, we let $\psi_i = 1$ for all variables that appear in first differences in the measurement equations of the DSGE models and as levels variables in the VAR models, while the remaining elements have $\psi_i = 0.9$. It now follows that the three-dimensional vector of hyperparameters to be estimated is given by $\alpha = [\lambda_o \; \delta \; \mu]'$ under the SoC prior.

From, e.g., Bańbura et al. (2010) we find that the relationship between the dummy observations and the prior parameters $(\mu_\Phi, \Omega_\Phi, A, v)$ are:

$$\mu_\Phi = y_{(d)}X'_{(d)}\left(X_{(d)}X'_{(d)}\right)^{-1}, \qquad \Omega_\Phi = \left(X_{(d)}X'_{(d)}\right)^{-1},$$
$$A = \left(y_{(d)} - \mu_\Phi X_{(d)}\right)\left(y_{(d)} - \mu_\Phi X_{(d)}\right)', \qquad v = T_d - (np+1).$$

If we make use of the expression for the number of degrees of freedom above, it follows that $v = 2n$ and the prior mean of Ω exists as $v > n+1$ when $n \geq 2$. However, the number of degrees of freedom can instead be selected as desired rather than taken literally from the dimensions of the dummy observation matrices. For example, the choice $v = n+2$ is sufficient to ensure that the expectation of $\Omega|\alpha$ under the prior density exists and this is the choice we make in this paper for the SoC prior as well as for the prior for the long run discussed below.

Given the dummy observations in equation (A.10), simple analytical expressions for the prior location matrices μ_Φ and A can be shown to be

$$\mu_\Phi = \Big[\big((\iota_n - \psi) \odot \bar{y}_0\big) \;\; \mathrm{diag}\big(\psi\big) \;\; 0_{n\times n(p-1)}\Big],$$
$$A = \mathrm{diag}\big(\omega\big)^2.$$

Furthermore, since Ω_Φ only depends on the parameters affecting $X_{(d)}$, the prior covariance matrix of Φ does *not* depend on the vector ψ.

Letting $y_\star = [y_{(d)} \; y]$ and $X_\star = [X_{(d)} \; X]$, it can be verified that the posterior parameters can be conveniently expressed as

$$\bar{\Phi} = y_\star X'_\star\big(X_\star X'_\star\big)^{-1},$$
$$XX' + \Omega_\Phi^{-1} = X_\star X'_\star,$$
$$S = \big(y_\star - \bar{\Phi}X_\star\big)\big(y_\star - \bar{\Phi}X_\star\big)'.$$

Prior B: Minnesota with a Prior for the Long Run

The second parameterization of the normal-inverted Wishart prior is based on the prior for the long run (PLR) suggested by Giannone et al. (2019). The PLR provides an alternative to the SoC prior for formulating the disbelief in an excessive explanatory power of the deterministic component of

[1] An alternative approach is considered by, e.g., Giannone et al. (2015) who treat ω as a hyperparameter to be estimated.

the model; see Sims (2000). Specifically, the PLR focuses on long-run relations, stationary as well as non-stationary, where economic theory can play an important role for eliciting the priors. The PLR does not impose the long-run relations but instead allows for shrinkage of the VAR parameters towards them.

Let B be an $n \times n$ nonsingular matrix with two blocks of rows

$$B = \begin{bmatrix} \beta_\perp' \\ \beta' \end{bmatrix}, \tag{A.11}$$

where β are $r \leq n$ potential cointegration relations (Johansen, 1996) and $\beta_\perp$ reflects coefficients on the $n - r$ possible stochastic trends, with $\beta'\beta_\perp = 0$ whenever $1 \leq r \leq n-1$. For the PLR with a diffuse prior for the constant term (Φ_0) we replace the last $n+1$ columns of $y_{(d)}$ and $X_{(d)}$ in equation (A.10) such that

$$\begin{aligned} y_{(d)} &= \begin{bmatrix} \lambda_o^{-1}\mathrm{diag}(\psi \odot \omega) & 0_{n\times n(p-1)} & \mathrm{diag}(\omega) & 0_{n\times 1} & B^{-1}\mathrm{diag}(\psi \odot B\bar{y}_0 \ominus \phi) \end{bmatrix}, \\ X_{(d)} &= \begin{bmatrix} 0_{1\times np} & 0_{1\times n} & \gamma^{-1} & 0_{1\times n} \\ \lambda_o^{-1}\big(j_p \otimes \mathrm{diag}(\omega)\big) & 0_{np\times n} & 0_{np\times 1} & \big(\imath_p \otimes B^{-1}\mathrm{diag}(B\bar{y}_0 \ominus \phi)\big) \end{bmatrix}, \end{aligned} \tag{A.12}$$

where element-by-element division is denoted by $\ominus$. The hyperparameter γ reflects overall tightness of Φ_0 such that a diffuse and improper prior is obtained when γ^{-1} is (arbitrarily close to) zero. The hyperparameter ϕ is an $n \times 1$ vector which captures shrinkage of the prior on the possibly non-stationary and stationary linear combinations of y in the rows of B. Since the PLR addresses the issue of the overfitting of the deterministic component, while also allowing for cointegration relations, there is no strong a priori reason for also including the dummy-initial-observation prior in this setup, other than it being an elegant approach for including a proper prior for the constant term of the VAR model.

Note that the original PLR is based on $\psi = \imath_n$, but we have introduced it here as a convenient way of allowing for non-unit means of the diagonal elements of Φ_1 also for this prior. As a consequence, it complements the treatment of possible cointegration relations, where the prior mean may otherwise imply a unit root. For instance, if a possible cointegration relation is a single variable, then having the corresponding ψ element set to some value less than one in absolute terms ensures that the prior mean of the VAR parameters is consistent with this variable being stationary. In this paper, we let $\psi_i = 0.8$ for such variables under the PLR; see also the specification of β below.

The vector of unknown hyperparameters is given by $\alpha = [\lambda_o \; \phi']'$, with $n + 1$ elements, while the matrix B is suggested by economic theory, such as from the three DSGE models considered in this paper. As pointed out by Giannone et al. (2019), the PLR simplifies to the sum-of-coefficients prior when $B = I_n$ and $\phi_i = \mu$ for $i = 1, \ldots, n$; see the last n rows of $y_{(d)}$ and $X_{(d)}$ in (A.10) and (A.12).

With the nine variables of y_t being ordered as real GDP, real private consumption, real total investment, GDP deflator inflation, total employment, real wages, the nominal short-term interest rate, the spread between the total lending rate and the policy rate, and unemployment, the DSGE

models may be used directly to suggest the following non-stationary long-run relations:

$$\beta_{\perp}' = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

This means that the two potential stochastic trends are given by a technology trend shared by GDP, consumption, investment and wages, and a labor supply (population) trend shared by GDP, consumption, investment and employment. The possibly stationary long-run relations are similarly given by

$$\beta' = \begin{bmatrix} -1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

These seven linear combinations yield (the log of) the consumption-output ratio, the investment-output ratio, the labor share, inflation, the short-term nominal interest rate, the spread, and unemployment.[2]

It should be noted that having an improper prior on Φ_0 through $\gamma^{-1} = 0$ means that $X_{(d)}X_{(d)}'$ is singular and this should be taken into account when computing, e.g., the log marginal likelihood in (A.9). To deal with this, let

$$X = \begin{bmatrix} \imath_T' \\ Y \end{bmatrix}, \qquad X_{(d)} = \begin{bmatrix} c_{(d)} \\ Y_{(d)} \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \Phi_1 & \cdots & \Phi_p \end{bmatrix}, \qquad \Omega_{\Phi} = \begin{bmatrix} \gamma^2 & 0_{1\times np} \\ 0_{np\times 1} & \Omega_{\Gamma} \end{bmatrix},$$

where $\imath_T$ is a $T \times 1$ unit vector, and where $c_{(d)}$ is the first row of $X_{(d)}$ which is zero except for position $n(p+1)+1$ being equal to γ^{-1}. The prior for the VAR parameters is now expressed as

$$\mathrm{vec}\big(\Gamma\big)\big|\Omega \sim N_{n^2p}\big(\mathrm{vec}\big(\mu_{\Gamma}\big), \big[\Omega_{\Gamma} \otimes \Omega\big]\big), \tag{A.13}$$

while $p(\Phi_0) \propto 1$ and the prior of Ω is given by (A.6). Let $Z = y - \Gamma Y$, $\bar{\Phi}_0 = T^{-1}Z\imath_T$, and let

$$D = I_T - T^{-1}\imath_T\imath_T',$$

[2] A possible contender to this setup is to follow Giannone et al. (2019) and instead consider also a third possible stochastic trend, given by a vector with unit coefficients on inflation and the short-term nominal interest rate and zeros elsewhere, i.e., a nominal stochastic trend. This means that the two vectors in β' above that pick these two variables (rows four and five) should be replaced with one vector taken as row five minus row four, providing a possibly stationary real interest rate.

a $T \times T$ symmetric and idempotent matrix. Furthermore, with D being symmetric and idempotent we may define $\tilde{Z} = ZD$, such that $\tilde{y} = yD$, $\tilde{Y} = YD$ and $ZDZ' = \tilde{Z}\tilde{Z}'$, while

$$\bar{\Gamma} = \left(\tilde{y}\tilde{Y}' + \mu_\Gamma \Omega_\Gamma^{-1}\right)\left(\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1}\right)^{-1}$$
$$S = \tilde{y}\tilde{y}' + A + \mu_\Gamma \Omega_\Gamma^{-1}\mu_\Gamma' - \bar{\Gamma}\left(\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1}\right)\bar{\Gamma}'.$$

It can be shown that the normal-inverted Wishart posterior for the VAR parameters is given by

$$\Phi_0 \big| \Gamma, \Omega, y, X_0, \alpha \sim N_n\big(\bar{\Phi}_0, T^{-1}\Omega\big), \tag{A.14}$$
$$\mathrm{vec}\big(\Gamma\big) \big| \Omega, y, X_0, \alpha \sim N_{n^2p}\big(\mathrm{vec}\big(\bar{\Gamma}\big), \big[(\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1})^{-1} \otimes \Omega\big]\big) \tag{A.15}$$
$$\Omega \big| y, X_0, \alpha \sim IW_n\big(S, T + v - 1\big). \tag{A.16}$$

Hence, the improper prior on Φ_0 results in a loss of degrees of freedom for the posterior of Ω.[3]

Furthermore, the log marginal likelihood is

$$\begin{aligned}\log p\big(y \big| X_0, \alpha\big) = &-\frac{n(T-1)}{2}\log(\pi) + \log\Gamma_n(T + v - 1) - \log\Gamma_n(v) - \frac{n}{2}\log\big|\Omega_\Gamma\big| \\ &+ \frac{v}{2}\log\big|A\big| - \frac{n}{2}\log(T) - \frac{n}{2}\log\big|\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1}\big| - \frac{T + v - 1}{2}\log\big|S\big|,\end{aligned} \tag{A.17}$$

where the term $\log(T)$ stems from $T = \imath_T'\imath_T$ and is obtained when integrating out Φ_0 from the joint posterior. The relationship between the dummy observations and the prior parameters is

$$\mu_\Gamma = y_{(d)}Y_{(d)}'\left(Y_{(d)}Y_{(d)}'\right)^{-1}, \qquad \Omega_\Gamma = \left(Y_{(d)}Y_{(d)}'\right)^{-1},$$
$$A = \left(y_{(d)} - \mu_\Gamma Y_{(d)}\right)\left(y_{(d)} - \mu_\Gamma Y_{(d)}\right)', \qquad v = T_d - (np + 1).$$

Letting $\tilde{y}_\star = [y_{(d)}\ \tilde{y}]$ and $\tilde{Y}_\star = [Y_{(d)}\ \tilde{Y}]$, it follows that the posterior parameters

$$\bar{\Gamma} = \tilde{y}_\star\tilde{Y}_\star'\big(\tilde{Y}_\star\tilde{Y}_\star'\big)^{-1},$$
$$\tilde{Y}\tilde{Y}' + \Omega_\Gamma^{-1} = \tilde{Y}_\star\tilde{Y}_\star',$$
$$S = \big(\tilde{y}_\star - \bar{\Gamma}\tilde{Y}_\star\big)\big(\tilde{y}_\star - \bar{\Gamma}\tilde{Y}_\star\big)'.$$

In the event that the improper prior on Φ_0 is replaced with a proper prior through the dummy-initial-observation prior, then the unknown $2(n+1) \times 1$ vector of hyperparameters is given by $\alpha = [\lambda_o\ \delta\ \omega'\ \phi']'$. For this case, $X_{(d)}X_{(d)}'$ is non-singular and we may use the same equations as in the case of Prior A when determining the marginal likelihood and the posterior distributions of the VAR parameters.

Hyperpriors and Posterior Inference

The use of hyperpriors is by no means new and has recently been employed in the BVAR models studied in the papers by Giannone et al. (2015, 2019) and Bańbura et al. (2015). Following Giannone et al. (2015), as hyperpriors for λ_o, δ and μ we use a Gamma distribution with mode 0.2, 1 and 1

[3] This loss of one degree of freedom is due to $(y_{(d)}, X_{(d)})$ having one observation less as $\gamma \to 0$, i.e., as the prior on Φ_0 becomes improper.

(also as in Sims and Zha, 1998) and standard deviations 0.4, 1 and 1, respectively.[4] Furthermore, following Giannone et al. (2019), the hyperprior for each element of ϕ is Gamma with mode and standard deviation equal to 1.

By combining the marginal likelihood in (A.9) with the SoC prior for the hyperparameters, or the expression in (A.17) for the marginal likelihood with the PLR for the hyperparameters, the hyperparameters in the vector α can be estimated from the corresponding log posterior kernel. A numerical optimizer, such as `csminwel` by Chris Sims, may now be used to compute the posterior mode of α as well as a suitable covariance matrix, such as the inverse Hessian at the mode. To obtain posterior draws of α one may, for instance, apply the standard random-walk Metropolis algorithm using the mode estimates, a normal proposal density and a suitable scaling parameter for the covariance matrix such that the acceptance rate lies within a suitable interval. Once these draws are available, posterior draws of Φ and Ω may be obtained from their posterior distributions conditional on α.

Dealing with the Ragged Edge

To formally deal with the ragged edge property of real-time data vintages, the methodology discussed above is not feasible and needs to be replaced with a computationally heavier approach. Specifically, the expression of the likelihood function[5] is no longer valid as it assumes that all variables have observations for the full sample. The likelihood function can instead be computed recursively with a Kalman filter that supports missing observations; see, e.g., Durbin and Koopman (2012, Chap. 4.10). This is technically uncomplicated, but the need to take missing data into account in a stepwise manner when computing the likelihood function means that an analytical expression for the marginal likelihood conditional on α is not available. The joint prior density of the parameters (Φ, Ω, α) can, of course, be computed and the product between it and the likelihood function yields the usual posterior kernel. Numerical optimization of all the parameters can now be applied to this kernel, yielding the posterior mode estimate of all parameters. Posterior sampling of the parameters can be conducted using, e.g., Markov Chain Monte Carlo (MCMC) or Sequential Monte Carlo (SMC) methods.

While this procedure is formally valid, the dimension of the parameter space is typically large, making posterior mode estimation and posterior sampling cumbersome and time consuming. Moreover, this is expected to be numerically very challenging since restrictions on Ω (positive definite) and α (positive) must hold. Since the real-time data vintages used by, for instance, our study

[4] Recall that the gamma distribution has the following density

$$p_G(z|a,b) = \frac{1}{\Gamma(a)b^a} z^{a-1} \exp\left(\frac{-z}{b}\right),$$

with shape parameter $a > 0$ and scale parameter $b > 0$. The mean is here ab while the variance is ab^2. The mode is unique when $a > 1$ and is then given by $b(a-1)$. With mode denoted by $\tilde{\mu}$ and standard deviation by σ, it holds that $b = (\sqrt{\tilde{\mu} + 4\sigma^2} - \tilde{\mu})/2$, while $a = (\sigma/b)^2$. For the case with $\sigma = 1$ and $\tilde{\mu} = 1$, it follows that $b = (\sqrt{5} - 1)/2 \approx 0.6180$, $a = 1/b^2 \approx 2.6180$, and the mean is close to 1.6180. The other case with $\sigma = 0.4$ and $\tilde{\mu} = 0.2$ means that $b = (\sqrt{17} - 1)/10 \approx 0.3123$ and $a = (2/5b)^2 \approx 1.6404$, such that the mean is approximately equal to 0.5123.

[5] See, e.g., equation (A.4).

only have missing data for the last two time periods, one option is to discard these periods when estimating the parameters. Doing so would allow for the approach advocated by Giannone et al. (2015, 2019), which reduces the posterior mode estimation and MCMC or SMC posterior sampling dimension problems from having $n^2p+n+\dim(\alpha)$ parameters to simply having $\dim(\alpha)$ parameters, while the VAR parameters can be obtained from direct sampling once the α parameters have been computed. The trade-off between using a formally valid approach with a high computational burden and a procedure based on the disposal of some information during the estimation stage for lower computational costs is expected to favor the latter case when n and p are large enough compared with the number of data points being disposed of. For the current study with $n = 9$ and $p = 4$, this amounts to having 333 fewer parameters in the latter case. From McAdam and Warne (2019, Table 4) the number of available observations on the different variables for the last time period is less than or equal to 2 of 9, while the corresponding number for the second last time period is 7 of 9; see also Warne (2022a, Table B.2) for an updated Table which includes also the vintages 2015Q1–2019Q4. Hence, the number of data points being discarded is at most 9. It should be kept in mind, however, that the discarded data may be highly important when forecasting and may also influence the parameter estimates, especially when the discarded data is sufficiently different from the utilized data. The direct effect is avoided by including all the available vintage data during the forecast stage, while the indirect effect through the parameter estimates is the cost of discarding data during the estimation stage.

Once posterior draws of the VAR parameters are available, the predictive likelihood can be estimated using the approach advocated by Warne et al. (2017) and McAdam and Warne (2019), where the last two time periods of each data vintage are now included in the information set. In this paper we do not evaluate the costs of using the two approaches with respect to computational time and difference between the predictive likelihood estimates. To save valuable computational time, we opt for the second approach and to further save time, we do not use posterior draws of the hyperparameters when sampling the VAR parameters from their posteriors, but fix them at the posterior mode estimates for each data vintage.

Additional Results on the Estimation of the BVAR Models

The recursive posterior mode estimates of the α hyperparameters are depicted in Figure I.2 for the BVAR with the SoC prior (top) and the PLR (bottom). The former model has three hyperparameters and from the plots we find that the estimates of the overall Minnesota tightness hyperparameter, λ_o, vary between roughly 0.2 and 0.3 with an average of 0.26. The shrinkage hyperparameter for the dummy-initial-observation part of the prior, δ, typically takes values between 1.5 and 2.0 (the mean is 1.61) with most of the values below 1.5 up to 2005, and values above thereafter. The μ hyperparameter related to the sum-of-coefficients part is estimated at about 2.30 on average.

Turning next to the hyperparameters of the PLR case, we find that the recursive estimates of the λ_o parameter are similar to those for the SoC prior with a mean of 0.27. Concerning the shrinkage hyperparameters on the long-run relations in the B matrix, we find that the ϕ_1, ϕ_2 and ϕ_4 are

all very close to unity. Recall that these parameters reflect shrinkage for the two possibly non-stationary relations reflecting a technology and a labor trend, as well as the potentially stationary investment-output ratio. The high stability of the posterior mode estimates around the prior mode for these hyperparameters reflects a lack of information from the (marginal) likelihood about these parameters.[6] Concerning the hyperparameter for the consumption-output ratio, ϕ_3, the recursive estimates are upward sloping until 2014 with an average value of 0.89. Next, for the labor-share hyperparameter, ϕ_5, the estimates are below unity with an average of 0.60, suggesting more shrinkage than at the prior mode, with a fairly large drop in 2003 from 0.9 to around 0.4 and a jump up to roughly 0.7 in 2010.

The last four hyperparameters concern shrinkage for specific variables: inflation, the short-term nominal interest rate, the spread and the unemployment rate, respectively. The average posterior mode estimate of ϕ_6 is 0.08, with a mild upward trend for the recursive estimates until 2014. The remaining three hyperparameters all have larger average values of 2.40, 0.87 and 5.00, respectively. Overall, there is some variation for these hyperparameters and in the cases of ϕ_7, for the short-term nominal interest rate, an upward trending path consistent with less shrinkage over the sample. For the shrinkage hyperparameter related to the spread, ϕ_8, there are instead two jumps in the path around 2003 and 2005, respectively, until a new plateau of approximately 0.9 is reached.

[6] Specifically, if one plots the log marginal likelihood function for this BVAR model, fixing all hyperparameters at their posterior mode values except one (for anyone of the vintages) and making use of a suitable grid around the mode value for, say, ϕ_1 gives a very flat profile, while the corresponding log posterior kernel has the same shape as the underlying gamma prior.

Section B: Real-Time Backcasts, Nowcasts and Forecasts from VAR Models

This Appendix first describes how the VAR model in equation (A.1) can be applied to backcast, nowcast and forecast when we have real-time data. To be specific, we shall assume that data on some variables are missing in periods $T-1$ and T, corresponding to the situation for the RTD of the euro area. The second part is concerned with the estimation of the marginal likelihood of the VAR model when taking the ragged edge of the real-time data into account.

In addition to being interested in the y_t variables, we are also interested in forecasting the first differences of some of these variables. To this end, let S be an $n \times s$ matrix with full column rank $s \leq n$ which selects unique elements of y_t such that

$$z_t = S'\big(y_t - y_{t-1}\big).$$

In other words, z_t is an s-dimensional vector whose elements are first differences of some of the elements in y_t. This means that

$$z_t = \Phi_0^* + \sum_{j=1}^{p} \Phi_j^* y_{t-j} + S'\epsilon_t,$$

where

$$\Phi_j^* = \begin{cases} S'\big(\Phi_1 - I_n\big), & \text{if } j = 1, \\ S'\Phi_j, & \text{otherwise.} \end{cases}$$

In order to derive a state-space system for the VAR model, including the first difference variables z_t, we stack these equations as follows:

$$\begin{bmatrix} z_t \\ 1 \\ y_t \\ y_{t-1} \\ \vdots \\ y_{t-p+2} \\ y_{t-p+1} \end{bmatrix} = \begin{bmatrix} 0 & \Phi_0^* & \Phi_1^* & \Phi_2^* & \cdots & \Phi_{p-2}^* & \Phi_{p-1}^* & \Phi_p^* \\ 0 & 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & \Phi_0 & \Phi_1 & \Phi_2 & \cdots & \Phi_{p-2} & \Phi_{p-1} & \Phi_p \\ 0 & 0 & I_n & 0 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & & \ddots & & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & & I_n & 0 & 0 \\ 0 & 0 & 0 & 0 & \cdots & 0 & I_n & 0 \end{bmatrix} \begin{bmatrix} z_{t-1} \\ 1 \\ y_{t-1} \\ y_{t-2} \\ \vdots \\ y_{t-p+1} \\ y_{t-p} \end{bmatrix} + \begin{bmatrix} S'\epsilon_t \\ 0 \\ \epsilon_t \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}.$$

This gives us the state equation of the system. More compactly, we express it as

$$\xi_t = F\xi_{t-1} + C\epsilon_t. \tag{B.1}$$

The measurement equation of the system is now given by

$$y_t = H'\xi_t, \quad t = 1, \ldots, T-2. \tag{B.2}$$

where

$$H' = \Big[0_{n\times(s+1)}\ I_n\ 0_{n\times n(p-1)}\Big].$$

For $t = T-1, T$, when some of the variables in y_t are unobserved, we introduce the matrices S_t, where $\tilde{y}_t = S_t' y_t$ includes all the observed values of y_t and none (NaN) of the unobserved. Accordingly, we have that for $\tilde{H}_t = HS_t$ the measurement equations are given by

$$\tilde{y}_t = \tilde{H}_t' \xi_t, \quad t = T-1, T. \tag{B.3}$$

We are now equipped with the state-space system and can proceed to setup a suitable Kalman filter, updater and smoother; see, e.g., Durbin and Koopman (2012) for details.

To this end, note that for $t = 1, \ldots, T-2$ the vector ξ_t is observed and determined as

$$\xi_t = KX_t,$$

where

$$K = \begin{bmatrix} 0 & S' & -S' & 0 \\ 1 & 0 & 0 & 0 \\ 0 & I_n & 0 & 0 \\ 0 & 0 & I_n & 0 \\ 0 & 0 & 0 & I_{n(p-2)} \end{bmatrix},$$

is an $(np + s + 1) \times (np + 1)$ matrix. Notice that the case $p = 1$ is treated as $p = 2$ with $\Phi_2 = 0$. Letting $\xi_{t|t-1}$ denote the standard Kalman filter projection of ξ_t based on the data up to period $t-1$ and taking the parameters as fixed, it follows that

$$\xi_{t|t-1} = F\xi_{t-1}, \quad t = 1, \ldots, T-2.$$

Similarly, let $P_{t|t-1}$ denote the covariance matrix of $\xi_{t|t-1}$ with the consequence that

$$P_{t|t-1} = C\Omega C', \quad t = 1, \ldots, T-2.$$

Furthermore, it holds that $\xi_{t|t} = \xi_t$ and $P_{t|t} = 0$ for the same time periods. It can also be seen that the covariance matrix of y_t given the data up to period $t-1$ is given by

$$\Sigma_{y,t|t-1} = H'C\Omega C'H = \Omega, \quad t = 1, \ldots, T-2.$$

Unless one is interested in computing the likelihood function, there is no need to run this Kalman filter recursively. Rather, the above filter equations are merely used as input for the interesting time periods $t = T-1, T, T+1, \ldots, T+h$, where we shall perform backcasting, nowcasting and forecasting taking the ragged edge into account.

For $t = T-1, T$, it no longer holds that all elements of y_t are observed. As mentioned above, the observations are instead given by $\tilde{y}_t$. For $t = T-1$, the filtering equations are:

$$\xi_{T-1|T-2} = F\xi_{T-2|T-2} = F\xi_{T-2}, \tag{B.4}$$

$$P_{T-1|T-2} = C\Omega C', \tag{B.5}$$

$$\tilde{y}_{T-1|T-2} = \tilde{H}'_{T-1}\xi_{T-1|T-2}, \tag{B.6}$$

$$\Sigma_{\tilde{y},T-1|T-2} = \tilde{H}'_{T-1}P_{T-1|T-2}\tilde{H}_{T-1} = S'_{T-1}\Omega S_{T-1}, \tag{B.7}$$

where $\Sigma_{\tilde{y},t|t-1}$ denotes the one-step-ahead forecast error covariance matrix of $\tilde{y}_t$. Concerning the update equations, these are given by

$$\xi_{T-1|T-1} = \xi_{T-1|T-2} + P_{T-1|T-2}\tilde{H}_{T-1}\Sigma^{-1}_{\tilde{y},T-1|T-2}\big(\tilde{y}_{T-1} - \tilde{y}_{T-1|T-2}\big), \tag{B.8}$$

$$P_{T-1|T-1} = P_{T-1|T-2} - P_{T-1|T-2}\tilde{H}_{T-1}\Sigma^{-1}_{\tilde{y},T-1|T-2}\tilde{H}'_{T-1}P_{T-1|T-2}. \tag{B.9}$$

Notice that $P_{T-1|T-1}$ is not a zero matrix and that its rank is expected to be $n - \text{rank}(S_{T-1})$.

Turning to $t = T$, the filtering equations are:

$$\xi_{T|T-1} = F\xi_{T-1|T-1}, \tag{B.10}$$

$$P_{T|T-1} = FP_{T-1|T-1}F' + C\Omega C', \tag{B.11}$$

$$\tilde{y}_{T|T-1} = \tilde{H}'_T\xi_{T|T-1}, \tag{B.12}$$

$$\Sigma_{\tilde{y},T|T-1} = \tilde{H}'_T P_{T|T-1}\tilde{H}_T, \tag{B.13}$$

while the update equations are given by:

$$\xi_{T|T} = \xi_{T|T-1} + P_{T|T-1}\tilde{H}_T\Sigma^{-1}_{\tilde{y},T|T-1}\big(\tilde{y}_T - \tilde{y}_{T|T-1}\big) = \xi_{T|T-1} + P_{T|T-1}r_{T|T}, \tag{B.14}$$

$$P_{T|T} = P_{T|T-1} - P_{T|T-1}\tilde{H}_T\Sigma^{-1}_{\tilde{y},T|T-1}\tilde{H}'_T P_{T|T-1} = P_{T|T-1} - P_{T|T-1}N_{T|T}P_{T|T-1}. \tag{B.15}$$

These equations define the vector $r_{T|T}$ and the matrix $N_{T|T}$ which are used as input for Kalman smoothing.

While the smooth estimates for period T are equal to the update estimates for T, we are also interested in the smooth estimates of the state variables and the corresponding covariance matrix for period $T-1$. These are determined from

$$\xi_{T-1|T} = \xi_{T-1|T-2} + P_{T-1|T-2}r_{T-1|T}, \tag{B.16}$$

$$P_{T-1|T} = P_{T-1|T-2} - P_{T-1|T-2}N_{T-1|T}P_{T-1|T-2}, \tag{B.17}$$

where

$$\begin{aligned}
r_{T-1|T} &= \tilde{H}_{T-1}\Sigma^{-1}_{\tilde{y},T-1|T-2}\big(\tilde{y}_{T-1} - \tilde{y}_{T-1|T-2}\big) + \big(F - K_{T-1}\tilde{H}'_{T-1}\big)'r_{T|T}, \\
K_{T-1} &= FP_{T-1|T-2}\tilde{H}_{T-1}\Sigma^{-1}_{\tilde{y},T-1|T-2}, \\
N_{T-1|T} &= \tilde{H}_{T-1}\Sigma^{-1}_{\tilde{y},T-1|T-2}\tilde{H}'_{T-1} + \big(F - K_{T-1}\tilde{H}'_{T-1}\big)'N_{T|T}\big(F - K_{T-1}\tilde{H}'_{T-1}\big).
\end{aligned}$$

The smooth estimates for period T are used for the nowcasts, while the smooth estimates for period $T-1$ are similarly employed for the backcasts.

Define the $(np+s+1)\times s$ matrix G such that

$$G' = \Big[I_s \; 0_{s,np+1} \Big],$$

with the consequence that $z_t = G'\xi_t$. The backcast $(T-1)$ and nowcast (T) of z_t are therefore given by

$$z_{t|T} = G'\xi_{t|T}, \quad t = T-1, T, \tag{B.18}$$

while the covariance matrices are

$$\Sigma_{z,t|T} = G' P_{t|T} G. \tag{B.19}$$

We can similarly compute the backcast and nowcast of y_t as

$$y_{t|T} = H'\xi_{t|T}, \tag{B.20}$$

and covariance matrices

$$\Sigma_{y,t|T} = H' P_{t|T} H. \tag{B.21}$$

Forecasting is also straightforward in this setup. Specifically, the forecasts of z_{T+h} and y_{T+h} for $h \geq 1$ are:

$$z_{T+h|T} = G'\xi_{T+h|T}, \tag{B.22}$$

$$y_{T+h|T} = H'\xi_{T+h|T}, \tag{B.23}$$

with covariance matrices

$$\Sigma_{z,T+h|T} = G' P_{T+h|T} G, \tag{B.24}$$

$$\Sigma_{y,T+h|T} = H' P_{T+h|T} H. \tag{B.25}$$

The required forecasts of the state variables and corresponding covariance matrices are determined from

$$\xi_{T+h|T} = F^h \xi_{T|T} = F\xi_{T+h-1|T}, \tag{B.26}$$

$$P_{T+h|T} = F^h P_{T|T} \big(F'\big)^h + \sum_{j=0}^{h-1} F^j C\Omega C' \big(F'\big)^j = F P_{T+h-1|T} F' + C\Omega C'. \tag{B.27}$$

To compute the projected value and the covariance matrix for a combination of variables in z_t and y_t, we simply construct a matrix D from the corresponding columns of G and H and use this matrix instead of G or H in the expressions above. We can thereafter proceed to compute the predictive likelihood of the combination of variables as in Warne et al. (2017) and McAdam and Warne (2019), using the actual values of the variables for the normal density with mean and covariance given by the values of these objects for a fixed posterior value of (Φ, Ω, α), and average these likelihoods conditional on the parameters over all the posterior draws.

Section C: Bayesian VAR Models with Stochastic Volatility

The VAR model with stochastic volatility we consider is given by equation (A.1), but we now assume that the innovations are independent and given by $\epsilon_t \sim N_n(0, \Omega_t)$. The matrix Ω_t is assumed to be positive definite and parameterized as

$$\Omega_t = R\Lambda_t R', \tag{C.1}$$

where R is an $n \times n$ lower triangular matrix with unit diagonal elements and with elements r_{ij} below the diagonal, $i > j$. The $n \times n$ matrix Λ_t is diagonal and specified as

$$\Lambda_t = \text{diag} \begin{bmatrix} \bar{s}_1 \exp\big(\lambda_{1,t}\big) \\ \vdots \\ \bar{s}_n \exp\big(\lambda_{n,t}\big) \end{bmatrix}, \quad t = 1, \ldots, T. \tag{C.2}$$

The $\bar{s}_i$ scaling parameters are assumed to be known, while $\lambda_{i,t}$ is a univariate process which generates heteroskedasticity with

$$\lambda_{i,t} = \gamma \lambda_{i,t-1} + \eta_{i,t}, \quad i = 1, \ldots, n, \quad t = 1, \ldots, T. \tag{C.3}$$

The innovation $\eta_{i,t}$ is assumed to be i.i.d. Gaussian with zero mean and variance ϕ_i, while the single autoregressive parameter γ is assumed to be known and shared across the n stochastic volatility components.

The modelling approach we follow relies greatly on Cogley and Sargent (2005) as it has been implemented in the BEAR Toolbox; see Dieppe et al. (2016). The technical details on estimation with stochastic volatility using the above setup are given in Dieppe et al. (2018), concerning their *standard model* in Section 5.2.

The parameters to be estimated are the VAR coefficients, Φ, the elements of $r^{-1} = (r_2^{-1}, \ldots, r_n^{-1})$ related to R^{-1} and to be discussed below, as well as the dynamic coefficients $\lambda = \{\lambda_{i,t}; i = 1, \ldots, n; t = 1, \ldots, T\}$ and the heteroskedasticity parameters $\phi = (\phi_1, \ldots, \phi_n)$. Regarding the prior distribution, it is assumed that Φ is independent of the other parameters. The same is true for r^{-1} and for ϕ, while the prior for the dynamic coefficients is conditional on the heteroskedasticity parameters. Furthermore, the individual r_i^{-1} are also assumed to be independent of each other and the same assumption is made regarding the heteroskedasticity parameters. Concerning the dynamic coefficients, λ_i is independent of λ_j and only dependent on ϕ_i, but not on ϕ_j for $i \neq j$.

A Minnesota-Type Prior for the VAR Parameters

The prior for the VAR coefficients is given by a standard Minnesota prior; see Doan et al. (1984). This means that Φ has a multivariate Gaussian distribution with mean μ_Φ and covariance Σ_Φ. A number of hyperparameters determine these matrices. With $\Sigma_{\Phi_{ij,k}}$ being the prior variance of $\Phi_{ij,k}$ for $i, j = 1, \ldots, n$ and $k = 1, \ldots, p$, it is first assumed that

$$\Sigma_{\Phi_{ii,k}} = \left(\frac{\lambda_1^*}{k^{\lambda_3^*}} \right)^2. \tag{C.4}$$

The hyperparameter λ_1^* gives the *overall tightness*, while λ_3^* is the *harmonic lag decay* hyperparameter. Furthermore,

$$\Sigma_{\Phi_{ij,k}} = \left(\frac{\sigma_i^2}{\sigma_j^2}\right)\left(\frac{\lambda_1^*\lambda_2^*}{k^{\lambda_3^*}}\right)^2, \quad i \neq j, \tag{C.5}$$

where σ_i^2 is the OLS residual variance, estimated from an AR(p) process with a constant for variable i, where $i = 1, \ldots, n$. Hence, the ratio σ_i^2/σ_j^2 deals with scaling of the variables. The hyperparameter λ_2^* gives the *cross-variable tightness*.

Concerning the parameters on the constant, the prior variance for element i is given by

$$\Sigma_{\Phi_{i,0}} = \sigma_i^2\left(\lambda_1^*\lambda_4^*\right)^2, \tag{C.6}$$

where λ_4^* is the tightness hyperparameter specific to Φ_0.

Concerning the prior mean of Φ it may be noted that the $n \times (np+1)$ matrix μ_Φ is given by

$$\mu_\Phi = \Big[\mu_{\Phi_0} \;\; \mu_{\Phi_1} \;\; 0 \;\; \cdots \;\; 0\Big]. \tag{C.7}$$

In the standard Minnesota prior case, the $n \times n$ matrix μ_{Φ_1} is diagonal with unit diagonal elements. When estimating the BVAR with the standard stochastic volatility model through the BEAR Toolbox, $\mu_{\Phi_1} = aI_n$, where a is a known scalar parameter. The BVAR with stochastic volatility that we estimate for the euro area real-time data is based on the same transformation of the observed variables as the three DSGE models, but using all nine variables, like for the two BVAR models with homoskedastic innovation. Since some of the observed variables are measured in first differences, while some appear in levels, one may wish to consider the diagonal elements of μ_{Φ_1} to differ across at least the type of variable, i.e., first difference or level. Although the BEAR Toolbox does not support this directly, small changes to the code make such a simple enhancement possible.

Concerning the prior mean for the parameters on the constant, μ_{Φ_0}, we have tested with the standard assumption of zero prior mean, as well as the assumption that $\mu_{\Phi_0} = (1-a)\bar{y}_0$. The vector $\bar{y}_0$ is calibrated and given by the sample mean of the p presample values of the observed variables, like in the case of the two BVARs with homoskedastic innovations. Overall, the former assumption gives better density forecasts, especially for inflation marginally and jointly with real GDP growth, and we have therefore decided to let $\mu_{\Phi_0} = 0$.

In the empirical study, we let $\lambda_1^* = 0.1$, $\lambda_2^* = 0.5$, $\lambda_3^* = 1$, $\lambda_4^* = 2$ and $a = 0.5$. We have studied alternative values for these hyperparameters, such as a taking on various values between 0 and 1, as well as testing with μ_{Φ_1} being diagonal and with the elements for the first differenced variables being equal to a_d, while the elements for the levels variables equal to a_l, but there does not appear to be any direct gains from letting $a_d \neq a_l$. The alternative calibrations of the prior did not improve the forecasting ability of the model overall. In addition, we have tested with a diffuse prior for the constant term with a value of $\lambda_4^* = 100$, as well as with λ_4^* being 100 for the first differenced variables and 2 for the levels variables. Although the selected value for λ_4^* is not the standard one for, e.g., US data, it allows for more sensible point and density forecasts for the euro area real-time data in connection with the chosen stochastic volatility setup.

The prior on R is implemented through a prior on each row of R^{-1}. To clarify notation, let

$$R^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ r_{21}^{-1} & 1 & 0 & \cdots & 0 & 0 \\ r_{3,1}^{-1} & r_{3,2}^{-1} & 1 & & 0 & 0 \\ \vdots & \vdots & & \ddots & & \vdots \\ r_{n-1,1}^{-1} & r_{n-1,2}^{-1} & r_{n-1,3} & & 1 & 0 \\ r_{n,1}^{-1} & r_{n,2}^{-1} & r_{n,3}^{-1} & \cdots & r_{n,n-1}^{-1} & 1 \end{bmatrix}.$$

Denote by r_i^{-1} a column vector with the non-zero and non-unit elements of row i in R^{-1} for $i = 2, \ldots, n$. Hence, this vector has $i-1$ elements and is given by

$$r_i^{-1} = \begin{bmatrix} r_{i,1}^{-1} \\ \vdots \\ r_{i,i-1}^{-1} \end{bmatrix}, \quad i = 2, \ldots n.$$

The parameters in r^{-1} above are given by the $n-1$ vectors r_i^{-1}.

The prior distribution for r_i^{-1} is assumed to be independent of r_j^{-1}, $i \neq j$, and to be multivariate normal with mean $\mu_{r_i^{-1}}$ and covariance $\Sigma_{r_i^{-1}}$. The BEAR Toolbox is calibrated to use a zero mean and covariance equal to 10000 times the identity matrix, as there are no obvious prior values. We have tested using a less diffuse covariance matrix for the euro area real-time data (10 times the identity), but in our experience this does not have a great effect on the density forecasts.

Concerning the dynamic coefficients, the prior for $\lambda_i|\phi_i$ can either be expressed jointly or period-by-period. Below we give the joint formulation and refer readers to the BEAR Toolbox technical manual for details on the period-by-period case; see Dieppe et al. (2018).

The joint formulation is based on the sparse matrix approach of Chan and Jeliazkov (2009), where the value of $\lambda_{i,t}$ depends on the value of $\lambda_{i,0}$ and the innovation values $\eta_{i,1}, \ldots, \eta_{i,t}$. The stochastic volatility process in (C.3) may be stacked as

$$\Gamma\lambda_i = \eta_i, \quad i = 1, \ldots, n, \tag{C.8}$$

with

$$\Gamma = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -\gamma & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\gamma & 1 & & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ 0 & & & & 1 & 0 \\ 0 & 0 & 0 & & -\gamma & 1 \end{bmatrix}, \quad \lambda_i = \begin{bmatrix} \lambda_{i,1} \\ \lambda_{i,2} \\ \lambda_{i,3} \\ \vdots \\ \lambda_{t,T-1} \\ \lambda_{i,T} \end{bmatrix}, \quad \eta_i = \begin{bmatrix} \gamma\lambda_{i,0} + \eta_{i,1} \\ \eta_{i,2} \\ \eta_{i,3} \\ \vdots \\ \eta_{i,T-1} \\ \eta_{i,T} \end{bmatrix}.$$

Notice that Γ is a $T \times T$ matrix and we let $\lambda_{i,0}$ be the initial value of the process. It is assumed that this scalar variable is normal with zero mean and variance

$$\sigma^2_{\lambda_{i,0}} = \frac{\phi_i(\omega - 1)}{\gamma^2},$$

where ω is a known variance parameter. It can furthermore be shown that

$$\eta_i \sim N_T\big(0, \phi_i I_\omega\big),$$

where I_ω is a $T \times T$ diagonal matrix with ω in element (1,1) and unity elsewhere on the diagonal. Consequently,

$$\lambda_i \sim N_T\big(0, \phi_i \Gamma^{-1} I_\omega (\Gamma')^{-1}\big).$$

The BEAR Toolbox has calibrated $\omega = 10000$ to allow for a very diffuse initial condition, but this value is easy to change. To examine the dependence on this value, we have also tried with $\omega = 10$, but this did not have any appreciable effect on the density forecasts for the vintages we tried. It may also be noted that with ω being very large, the value of ϕ_i will not matter much for the prior on the dynamic coefficients, unless it is small enough to counteract ω. Furthermore, our empirical study uses $\gamma = 0.85$, but we have also tried alternative values including unity for the random walk version of stochastic volatility.

Finally, the prior for ϕ_i is independent of ϕ_j, $i \neq j$, and assumed to follow an inverted Gamma distribution. The BEAR Toolbox uses the following parameterization of its density

$$p\big(\phi_i|\alpha_0, \delta_0\big) = \frac{\big(\delta_0/2\big)^{\alpha_0/2}}{\Gamma\big(\alpha_0/2\big)} \phi_i^{-(\alpha_0/2)-1} \exp\left(\frac{-\delta_0}{2\phi_i}\right),$$

where $\Gamma(\cdot)$ is the gamma function. The shape and scale parameters α_0 and δ_0, respectively, need to be determined by the user. If a weakly informative prior is preferred, then values such as $\alpha_0 = \delta_0 = 0.001$, or lower, may be considered. In our empirical study we use these values.

Posterior Sampling

The posterior distributions are provided in the technical manual of the BEAR Toolbox. The posterior sampler is mainly a Gibbs algorithm, but extended with a Metropolis-Hastings step for the dynamic coefficients, $\lambda_{i,t}$, as its conditional posterior distribution is not fully known; see also Cogley and Sargent (2005). The algorithm is initialized by, e.g., estimating the VAR model with OLS and computing the scaling coefficients from the OLS estimate of Ω through the Choleski decomposition $\Omega = R\Lambda R'$, where $\bar{s}_i$ is equal to the i^{th} diagonal element of the estimate of Λ. Details on the initialization and the posterior sampler are presented by Dieppe et al. (2018) in Algorithm 5.2.1.

Real-Time Backcasts, Nowcasts and Forecasts Under Stochastic Volatility

The procedure we employ for the forecasting purposes of the BVAR with stochastic volatility (SV) is nearly identical to the one in Appendix B. We therefore focus the discussion below on the features that are different.

First, since the stochastic volatility (SV) BVAR already has the z_t variables in the y_t vector, the s equations for z_t can be dispensed with, making the state-space formulation from Appendix B simpler here. For example, the matrix $K = I_{np+1}$, while the G matrix can be also skipped.

Second, as in the case of the SoC and PLR BVARs, the parameters of the SV BVAR are estimated up to period $T-2$ for each vintage T. At this stage, the innovation covariance matrix is given by Ω_{T-2}. If we assume that $\gamma = 1$, such that the n stochastic volatility processes in (C.3) are random walks, then projecting Ω_{T-2} forwards we note that $\lambda_{i,T+h|T-2} = \lambda_{i,T-2}$, $h = -1, 0, \ldots, h^*$ and $i = 1, \ldots, n$, with the consequence that $\Omega_{T+h|T-2} = \Omega_{T-2}$.

Suppose instead that $\gamma < 1$, as in our empirical application, and let λ_t and $\bar{s}$ be n-dimensional vectors with typical element $\lambda_{i,t}$ and $\bar{s}_i$, respectively. The projections of λ_{T-2} forwards are then

$$\lambda_{T+h|T-2} = \gamma\lambda_{T+h-1|T-2} = \gamma^{h+2}\lambda_{T-2}, \quad h = -1, 0, 1, \ldots, h^*.$$

It then follows that

$$\Lambda_{T+h|T-2} = \operatorname{diag}\left(\bar{s} \odot \exp\left(\lambda_{T+h|T-2}\right)\right),$$

where

$$\exp\left(\lambda_{T+h|T-2}\right) = \left(\exp\left(\lambda_{T-2}\right)\right)^{\gamma^{h+2}} = \left(\exp\left(\lambda_{T+h-1|T-2}\right)\right)^{\gamma}.$$

Finally, we have that

$$\Omega_{T+h|T-2} = R\Lambda_{T+h|T-2}R'.$$

Notice that the limit expression of the covariance matrix as $h \to \infty$ is simply $R\text{diag}(\bar{s})R'$, such that the scaling factors give the long run forecast error variances.

Section D: Probability Integral Transform

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

$$\pi_{j,T+1|T} = F_j\big(y_{j,T+1}\big|\mathcal{Y}_T\big), \quad j = 1, \ldots, n,$$

is independent and uniformly distributed on the unit interval, where $F_j(\cdot)$ is the cumulative distribution function (cdf). With $\Phi(\cdot)$ being the cdf of the normal distribution, Smith (1985) further noted that $z_{j,T+1|T} = \Phi^{-1}(\pi_{j,T+1|T})$ is i.i.d. $N(0,1)$; see also Berkowitz (2001).

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

$$\pi^{(i)}_{j,T+1|T} = \Phi\left(y^{(o)}_{j,T+1}\middle|\mu^{(i)}_{j,T+1|T}, \sigma^{(i)}_{jj,T+1|T}\right), \quad i = 1, \ldots, N$$

where $\mu^{(i)}_{j,T+1|T}$ is the one-step-ahead point forecast of observed variable $j = 1, \ldots, n$ using the i^{th} posterior draw of θ, while $\sigma^{(i)}_{jj,T+1|T}$ is the one-step-ahead forecast error standard deviation of variable j using $\theta^{(i)}$.

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

$$\pi_{j,T+1|T} = \frac{1}{N}\sum_{i=1}^{N} \pi^{(i)}_{j,T+1|T}, \tag{D.1}$$

while

$$z_{j,T+1|T} = \Phi^{-1}\big(\pi_{j,T+1|T}\big). \tag{D.2}$$

Under the assumption that the model is correctly specified, or similarly that the density forecasts are well calibrated, the variable $z_{j,T+1|T}$ 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 $z_{j,T+1|T}$ process. They now consider the test statistic

$$\text{AG} = T_1\big(\bar{m}_{T_1} - m_{q+p}\big)'\Omega^{-1}\big(\bar{m}_{T_1} - m_{q+p}\big) \xrightarrow{\text{d}} \chi^2_{q+p}, \tag{D.3}$$

where T_1 is the number of one-step-ahead forecasts,

$$\bar{m}_{T_1} = \begin{bmatrix} T_1^{-1}\sum_{T=1}^{T_1} z_{j,T+1|T} \\ \vdots \\ T_1^{-1}\sum_{T=1}^{T_1} z^q_{j,T+1|T} \\ \big(T_1-1\big)^{-1}\sum_{T=2}^{T_1} z_{j,T+1|T} z_{j,T|T-1} \\ \vdots \\ \big(T_1-p\big)^{-1}\sum_{T=p+1}^{T_1} z_{j,T+1|T} z_{j,T+1-p|T-p} \end{bmatrix}, \quad m_{q+p} = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_q \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

The scalar μ_r is the r^{th} population moment of a normal distribution. The matrix Ω is given by

$$\Omega = \begin{bmatrix} \mu_2 & \mu_3 & \cdots & \mu_{q+1} & 0_{1\times p} \\ \mu_3 & \mu_4 & \cdots & \mu_{q+2} & 0_{1\times p} \\ \vdots & \vdots & & \vdots & \vdots \\ \mu_{q+1} & \mu_{q+2} & \cdots & \mu_{2q} & 0_{1\times p} \\ 0_{p\times 1} & 0_{p\times 1} & \cdots & 0_{p\times 1} & I_p \end{bmatrix},$$

while

$$\mu_r = \begin{cases} \prod_{i=1}^{r/2}(2i-1) & \text{if } r \text{ is odd,} \\ 0 & \text{otherwise.} \end{cases}$$

In practice, Amisano and Geweke (2017) let $q = p = 4$, such that $\mu_1 = \mu_3 = \mu_5 = \mu_7 = 0$, while $\mu_2 = 1$, $\mu_4 = 3$, $\mu_6 = 15$ and $\mu_8 = 105$.

The vector $\bar{m}_{T_1}$ takes into account that if the model is well calibrated $z_{j,T+1|T}$ is not serially correlated. However, for $z_{j,T+h|T}$ is it generally not true as the forecast errors will follow a moving average process of order $h-1$ in that situation. Consequently, for h-step-ahead density forecasts the last p elements of $\bar{m}_{T_h}$ need to take this into account. Specifically, we have that for h-step-ahead forecasts:

$$\bar{m}_{T_h} = \begin{bmatrix} T_h^{-1}\sum_{T=1}^{T_h} z_{j,T+h|T} \\ \vdots \\ T_h^{-1}\sum_{T=1}^{T_h} z_{j,T+h|T}^{q} \\ \left(T_h - h\right)^{-1}\sum_{T=h+1}^{T_h} z_{j,T+h|T} z_{j,T|T-h} \\ \vdots \\ \left(T_h - (h-1+p)\right)^{-1}\sum_{T=h+p}^{T_h} z_{j,T+h|T} z_{j,T+h-p|T-p} \end{bmatrix},$$

In the empirical study, the nowcasts of real GDP growth and GDP deflator inflation are typically one-step-ahead forecasts once the theoretical autocorrelation pattern of the forecast errors is taken into account. Similarly, the h-step-ahead forecasts are therefore typically $(h+1)$-step-ahead. The reason for the word *typically* is that for some vintages, the data for the backcast period is also missing and needs to be predicted prior to predicting the nowcast and the forecasts; see McAdam and Warne (2019, Table 4). Hence, the one-step-ahead forecast is actually the backcast for the relevant vintages. For simplicity, these vintages are dropped for the PIT analysis, i.e., 3 vintages for real GDP growth and 22 vintages for GDP deflator inflation. Moreover, since the number of vintages is quite low, we limit the PIT analysis to nowcasts and one-step-ahead forecasts, and consider the cases $(q,p) = (2,2)$ and $(q,p) = (4,2)$.

The test results are displayed in Table I.9 for the marginal density forecasts of real GDP growth and GDP deflator inflation. The uniformity of the PIT does not hold for multivariate forecasts; see Genest and Rivest (2001). These forecasts are therefore not tested here, but as pointed out by

Gneiting et al. (2008), an option is then to utilize the *Box density ordinate transform* (BOT); see Box (1980) for details. Based on the asymptotic distribution of the tests, the marginal real GDP growth density forecasts appear *not* to be well calibrated for any model, with the BVAR models displaying particularly large test values, especially when four moments are included. Concerning the marginal inflation forecasts, only the SWFF model is found not to be well calibrated, as we may indeed "suspect" from the spaghetti plots in Figure 2 of the paper.

It is also noteworthy that some of the combination methods do better in the comparison exercise than the individual models for real GDP growth and the joint forecasts, while they cannot improve the density forecasts of inflation. This is consistent with the PIT results as most models that forecast inflation relatively well also seem to have fairly well calibrated inflation density forecasts.

As a complement to the formal tests, histograms are shown in Figures I.3–I.4 for the estimated $\pi_{j,T+h+1|T}$ values for $h = 0, 1$ that underlie the tests. The graphs are based on a bin width of 0.1 and if a model is well calibrated, i.e., the π's are uniform, then the frequency of occurrence in large samples should match the height of the uniform at 0.1. Concerning real GDP growth in Figure I.3, it can be seen that the bulk of estimates fall into the bins below 0.5. This is consistent with the mean errors being negative in Table I.3; see Appendix G for details. Concerning inflation in Figure I.4, the histograms display a more uniform-like shape for all models except the SWFF model. However and consistent with the mean errors, there is a tendency for a larger share of the estimated π's to lie in the bins above 0.5.

Section E: Density Forecast Combination Methods

Static Optimal Prediction Pools

Static optimal predictions pools were introduced by Hall and Mitchell (2007) as a means to improve the density forecasts of individual forecasters or models by computing the optimal linear combination of these forecasts. The basic idea is to maximize the log predictive score of a linear combination of the predictive likelihoods of the models in the pool. This linear combination is constrained such that the model weights are constant, non-negative and sum to unity. As a result, the combination of the predictive likelihoods is also a predictive likelihood and the log predictive score is formed by accumulating the log of the pooled predictive likelihoods.

Let $w_{i,h}$ be the weight on model i for h-step-ahead forecasts, satisfying $w_{i,h} \geq 0$ and $\sum_{i=1}^{M} w_{i,h} = 1$. The log predictive score of the static optimal pool is therefore given by

$$S_{T:T_h,h}^{(\text{SOP})} = \sum_{t=T}^{T_h} \log \left(\sum_{i=1}^{M} w_{i,h} p_{t+h|t}^{(i)} \right), \tag{E.1}$$

where the predictive likelihood of the pool is given by the term being logged on the right hand side of this equation. Estimates of the weights are obtained by maximizing the log score in (E.1) with respect to the weights and subject to their restrictions.

This forecast combination is referred to as static since the weights are treated as constant over time. A recursive approach to the estimation of these weights is more realistic when viewing the problem of comparing models in real time. In that case, the weights of the models can change due to re-optimization with more recent information. Hall and Mitchell (2007) motivate the use of static optimal pools on the grounds that the weights are chosen to minimize the "distance" between the forecasted and the unknown true predictive density in the sense of the Kullback-Leibler information criterion (KLIC); see Kullback and Leibler (1951). In contrast with combination approaches such as Bayesian Model Averaging (BMA), discussed below, Geweke and Amisano (2011, 2012) point out that static optimal prediction pools do not rely on the assumption that one of the models in the pool is true, i.e., the approach allows for incomplete models with the effect that all of the models in the pool may be false; see Geweke (2010).[7]

Dynamic Prediction Pools

Dynamic prediction pools, suggested by Del Negro et al. (2016), directly allow the weights to vary as well as to be correlated over time. While their setup is based on two models, Amisano and Geweke (2017) extends the dynamic prediction pool from two to three models. In fact, the approach in Amisano and Geweke allows for any finite number of models and it is for such a general case that we present dynamic prediction pools below. Accordingly, the number of models is, as before, equal

[7] See also Pauwels and Vasnev (2016) for further analysis of optimal prediction pool weights and the underlying optimization problem, and Opschoor et al. (2017) for extensions to alternative scoring rules.

to M and the log predictive score for the dynamic prediction pool is:

$$S^{(\text{DP})}_{T:T_h,h} = \sum_{t=T}^{T_h} \log\left(\sum_{i=1}^{M} w_{i,t+h|t} p^{(i)}_{t+h|t}\right), \tag{E.2}$$

where $w_{i,t+h|t}$ is the weight on model i for h-step-ahead density forecasts at t, where all weights are non-negative and sum to unity over i.

To model the time variation of the weights, we follow Amisano and Geweke (2017) and consider a parsimoniously parameterized M-dimensional process for the state variable

$$\xi_t = \rho\xi_{t-1} + \sqrt{(1-\rho^2)}\eta_t, \qquad t = T, \ldots, T_h, \tag{E.3}$$

where $0 \leq \rho \leq 1$ is a scalar and $\eta_t \sim$ i.i.d. $N(0, I_M)$.[8] The ρ parameter is interpreted as a "forgetting factor" for dynamic pools by Del Negro et al. (2016) since with $\rho < 1$ there is discounting of past information. The parameterization of ξ_t in (E.3) ensures that its covariance is the identity matrix when $\rho < 1$, while ξ_t is constant (static) otherwise.

The weights are determined from a logistic transformation of the individual elements of the state vector, $\xi_{i,t}$ which ensures that each element is non-negative and that the sum of the elements is unity:

$$w_{i,t} = \frac{\exp(\xi_{i,t})}{\sum_{j=1}^{M}\exp(\xi_{j,t})}, \quad i = 1, \ldots, M. \tag{E.4}$$

The vector w_t is consequently M-dimensional with individual entries given by the model weights. As pointed out by Amisano and Geweke (2017, Appendix E.4) the specification in (E.4) implies a symmetric prior across weights.

To estimate the log predictive score in (E.2) conditional on ρ, Amisano and Geweke (2017) follow the approach in Del Negro et al. (2016) and employ the *Bayesian bootstrap particle filter*; see, e.g., Gordon et al. (1993) and Herbst and Schorfheide (2016) for details and further references. This filter is *initialized* as follows: At $t = T-1$, draw N particles from the unconditional distribution of ξ_{T-1} and map these into $w_{i,T-1}$ using (E.4), while each particle is assigned equal weight; i.e., for $i = 1, \ldots, M$ and $n = 1, \ldots, N$

$$\begin{aligned}
\xi^{(n)}_{T-1} &\sim N(0, I_M), \\
w^{(n)}_{i,T-1} &= \frac{\exp(\xi^{(n)}_{i,T-1})}{\sum_{j=1}^{M}\exp(\xi^{(n)}_{j,T-1})}, \\
W^{(n)}_{T-1} &= 1.
\end{aligned}$$

During the *recursions* of the bootstrap particle filter for $t = T, \ldots, T_h$, each iteration involves three steps: forecasting, updating, and selection. The *forecasting* step concerns propagating the N

[8] For the implementation of the dynamic prediction pool in Del Negro et al. (2016), they have a univariate process similar to (E.3), but also allow for a drift parameter μ and a standard deviation σ. The process for ξ in (E.3) can be enriched in various ways, but for the sake of parsimony we only allow for one free parameter.

particles forward by drawing innovations $\eta_t^{(n)} \sim N(0, I_M)$ such that

$$\tilde{\xi}_t^{(n)} = \rho \xi_{t-1}^{(n)} + \sqrt{(1-\rho^2)}\eta_t^{(n)},$$

$$\tilde{w}_{i,t}^{(n)} = \frac{\exp\big(\tilde{\xi}_{i,t}^{(n)}\big)}{\sum_{j=1}^{M} \exp\big(\tilde{\xi}_{j,t}^{(n)}\big)},$$

for $i = 1, \ldots, M$ and $n = 1, \ldots, N$. Next, the incremental weights are calculated as

$$\tilde{\omega}_t^{(n)} = p\big(x_t^{(m)} \big| \tilde{w}_t^{(n)}, \mathcal{I}_{t-h}^{(\mathcal{P})}, \mathcal{P}\big) = \sum_{i=1}^{M} \tilde{w}_{i,t}^{(n)} p\big(x_t^{(m)} \big| \mathcal{I}_{t-h}^{(i)}, A_i\big), \quad n = 1, \ldots, N,$$

where these weights depend on ρ via $\tilde{w}_t^{(n)}$, a vector with $\tilde{w}_{i,t}^{(n)}$ in element i, and with $\mathcal{I}_t^{(\mathcal{P})}$ being the joint information set of the pooled models. Notice that the incremental weights are formed from the h-steps-ahead predictive likelihoods of the individual models based on measured values of the predicted variables in period t, $x_t^{(m)}$. Furthermore, we here let $x_t^{(m)}$ denote the observations used when computing the weights and these may be the actual values, $x_t^{(a)}$, and/or taken from the current vintage.

The *updating* step consists in recomputing the weights according to

$$\tilde{W}_t^{(n)} = \frac{\tilde{\omega}_t^{(n)} W_{t-1}^{(n)}}{(1/N)\sum_{j=1}^{N} \tilde{\omega}_t^{(j)} W_{t-1}^{(j)}}.$$

Notice that the sum of the updated weights $\tilde{W}_t^{(n)}$ over all particles is equal to N. If all the weights from recursion $t-1$ are equal, then the updated weights are proportional to the incremental weights and therefore the particle likelihood values.

For the *selection* step we first compute the effective sample size (ESS) according to

$$\mathrm{ESS}_t = \frac{N}{(1/N)\sum_{n=1}^{N} \big(\tilde{W}_t^{(n)}\big)^2}.$$

On the one hand, if $\mathrm{ESS}_t < \delta^* N$ for a suitable value of the hyperparameter δ^*, where $0 < \delta^* < 1$, the particles are resampled with multinomial resampling, characterized by support points and weights $\{\tilde{\xi}_t^{(n)}, \tilde{w}_t^{(n)}, \tilde{W}_t^{(n)}\}_{n=1}^N$. Let $\{\xi_t^{(n)}, w_t^{(n)}, W_t^{(n)}\}_{n=1}^N$ denote a swarm of N i.i.d. draws where the weights are given by $W_t^{(n)} = 1$. On the other hand, if $\mathrm{ESS}_t \geq \delta^* N$, the weights $W_t^{(n)} = \tilde{W}_t^{(n)}$ while the state variables $\xi_t^{(n)} = \tilde{\xi}_t^{(n)}$ and the corresponding model weights $w_t^{(n)} = \tilde{w}_t^{(n)}$.

The conditional predictive likelihood of the dynamic pool in recursion t is approximated with

$$p_{t+h|t}^{(\mathcal{P})}(\rho) = p\big(x_{t+h}^{(a)} \big| \mathcal{I}_t^{(\mathcal{P})}, \mathcal{P}; \rho\big) = \sum_{i=1}^{M} w_{i,t+h|t}(\rho) p\big(x_{t+h}^{(a)} \big| \mathcal{I}_t^{(i)}, A_i\big), \tag{E.5}$$

where the particle weights depend on the parameter ρ, and the weights $w_{i,t+h|t}^{(n)}$ are computed by iterating forward using the law of motion in (E.3). That is

$$\begin{aligned} w_{i,t+h|t}(\rho) &= E\big[w_{i,t+h} \big| \mathcal{I}_t^{(\mathcal{P})}, \mathcal{P}; \rho\big] \\ &\approx \frac{1}{N} \sum_{n=1}^{N} w_{i,t+h}^{(n)}(\rho) W_t^{(n)}, \end{aligned}$$

where for each particle n

$$\xi_{i,t+s}^{(n)}(\rho) = \rho\xi_{i,t+s-1}^{(n)} + \sqrt{\big(1-\rho^2\big)}\eta_{t+s}^{(n)},$$

$$w_{i,t+h}^{(n)}(\rho) = \frac{\exp\big(\xi_{i,t+h}^{(n)}(\rho)\big)}{\sum_{j=1}^{M}\exp\big(\xi_{j,t+h}^{(n)}(\rho)\big)},$$

for $s = 1, \dots, h$, and where $\eta_{t+s}^{(n)}$ is a draw from an M-variate standard normal distribution.

In a real-time setting, the information lag needs to be taken into account. Consider the convenient, albeit mechanical, assumption that the information lag is equal to the observation lag. With annual revisions data, this means that the predictive likelihoods for the individual models are lagged four periods when computing the incremental weights. Since measured values of x may be available for periods $\tau-3$, $\tau-2$ and $\tau-1$ in vintage τ, a shorter information lag is feasible. Unless the measured values are equal to the actual values for these time periods, however, the bootstrap particle filter requires two loops, where for each vintage τ the iterations from T until τ are (at least partly) revisited. The assumption that the information lag is equal to the observation lags means that the second loop can be avoided. Alternatively, it may also be skipped if one assumes that measured values are well approximated by the actual values, i.e., that the revisions are sufficiently small that they can be neglected.[9]

Amisano and Geweke (2017) apply the bootstrap particle filter over a fine grid of values for ρ and compute $\hat{\rho}_\tau$ by maximizing the log score

$$\hat{\rho}_{\tau,h} = \arg\max_{\rho} \sum_{t=T}^{\tau-h} \log\Big(p\big(x_{t+h}^{(m)}|\mathcal{I}_t^{(\mathcal{P})}, \mathcal{P}; \rho\big)\Big), \quad \tau = T, \dots, T_h \tag{E.6}$$

where the predictive likelihood on the right hand side are computed as in (E.5), but with the measured value instead of the actual value. This means that the first period τ when the h-step-ahead predictive likelihood of the individual models can be observed occurs at $\tau = T+h$. Taking the real-time aspect fully into account means that the information lag, $l \leq k$, is added to this number. It follows that for all τ less than $T+h+l$, a unique value of ρ cannot be determined as there are no data on the predictive likelihoods available. This initialization problem may be dealt with by replacing the unobserved predictive likelihood values with a positive constant, such as $1/M$, with the consequence that all values of ρ from T up to $T+h+l-1$ obtain the same log score when computing the weights. As the number of particles becomes very large, this amounts to giving all models the same weight. Furthermore, the upper limit of the sum on the right hand side of (E.6) should also take the information lag into account. Hence, we set it equal to $\tau-h-l$.

[9] If only data on x from vintage τ are used as measured values they are all subject to revision and the double loop has to be executed from period T for each vintage τ. On the other hand, if actuals are used up to period $\tau-4$ and measured values from vintage τ for periods $\tau-3$, $\tau-2$ and $\tau-1$, then the double loop would start in $\tau-4$ since the earlier computations were run for vintage $\tau-1$. Finally, if the measured values for $\tau-3$, $\tau-2$ and $\tau-1$ are approximated by the actual values, then the double loop can be dispensed with. For each of these cases, an information lag of one is applied.

With the sequence $\{\hat{\rho}_t\}_{t=T}^{\tau}$, the real-time log predictive score of the dynamic prediction pool is then estimated by

$$S_{T:\tau,h}^{(\mathrm{DP})} = \sum_{t=T}^{\tau} \log\Big(p_{t+h|t}^{(\mathcal{P})}\big(\hat{\rho}_{t,h}\big)\Big).$$

According to Amisano and Geweke (2017), this procedure may be interpreted as a Bayesian analysis based on a flat (uniform) prior on ρ.

Resampling was originally employed by Gordon et al. (1993) to reduce the effects from sample *degeneracy*—a highly uneven distribution of particle weights—as this part of the selection step adds noise; see Chopin (2004). It also allows for the removal of low weight particles with a high probability and this is very practical as it is preferable that the filter is focused on regions with a high probability mass. However, this also means that resampling may produce sample *impoverishment* as the diversity of the particle values is reduced. The hyperparameter δ^* provides a 'crude' tool for balancing the algorithm against the pitfalls of degeneracy or impoverishment by ensuring that resampling takes place but does not occur 'too often'.[10]

Del Negro et al. (2016) use a multinomial distribution for the selection step with $\delta^* = 2/3$, but as pointed out by, for example, Douc et al. (2005), this resampling algorithm produces an unnecessarily large variance of the particles. Moreover, and as emphasized by Hol et al. (2006), ordering of the underlying uniform draws improves the computational speed considerably. Commonly used alternative resampling algorithms are faster and have a smaller variance of the particles. Systematic resampling, introduced by Kitagawa (1996) and also emphasized by Carpenter et al. (1999), is one such approach and it is very easy to implement, comparatively fast, and, according to Doucet and Johansen (2011), as it often outperforms other sequential resampling schemes. It is a faster version of stratified resampling (Kitagawa, 1996), where instead of drawing N uniforms only one is required, while stratification and, simultaneously, sorting is dealt with via the same simple affine function. A drawback with systematic resampling is that it generates cross-sectional dependencies among the particles, which also makes it difficult to establish its theoretical properties; see Chopin (2004). For an overview of sequential resampling schemes see, e.g., Hol et al. (2006), who also discuss theoretical criteria for choosing between multinomial, stratified, systematic and residual resampling (suggested by Liu and Chen, 1998); Douc et al. (2005), who also study large sample behavior; and more recently Li et al. (2015), who also discusses distributed or parallel algorithms.[11]

Bayesian Model Averaging

BMA provides a coherent framework for accounting for model uncertainty; see Hoeting et al. (1999). The standard BMA weights rely on posterior model probabilities and therefore require the calculation of marginal likelihoods over a set of models which predict the same observables. The DSGE and

[10] A more direct approach to combatting sample impoverishment is based on taking the particle values into account when resampling; see, e.g., Doucet and Johansen (2011) for discussions on the *resample-move* algorithm of Gilks and Berzuini (2001) and the *block sampling* algorithm of Doucet et al. (2006).

[11] Since the bootstrap particle filter will spend a considerable share of the computational time in the resampling stage, especially when δ^* is high, the precise implementation of the selected resampling scheme is also important; see also Warne (2022b, Section 8.4) for some details on how to implement the standard sequential algorithms.

VAR models we examine, however, do not satisfy this condition: the VAR models predict all nine observable variables comprising the full information set, while the DSGE models do not predict all variables.

However, an alternative BMA weighting scheme may be obtained by only using the predictive likelihood values of the variables of interest (x_t) and a prior for the weights; see Eklund and Karlsson (2007) for discussions on the general idea. Specifically, BMA weights may be obtained as in Del Negro et al. (2016, Section 3) such that

$$\hat{w}_{i,t,h} = \frac{\hat{w}_{i,t-1,h} p\big(x_t^{(m)} | \mathcal{I}_{t-h}^{(i)}, A_i\big)}{\sum_{j=1}^{M} \hat{w}_{j,t-1,h} p\big(x_t^{(m)} | \mathcal{I}_{t-h}^{(j)}, A_j\big)}, \tag{E.7}$$

for $t = T+1, \ldots, T_h$, with $\hat{w}_{i,T,h}$ denoting the prior predictive probability that $w_i = 1$. In this case, the model weights at t are built using only the predictive likelihood values up to period t. As long as $\hat{w}_{i,T,h} = 1$ for some model i does *not* hold, these BMA weights will be positive for several models until the recursive log score of one model dominates the others sufficiently. In the empirical sections of the paper, we primarily let $\hat{w}_{i,T,h} = 1/M$.

It should be kept in mind that the predictive likelihood generated BMA weights are based on the following assumption:

$$p\big(x_t | \mathcal{I}_{t-h}^{(i)}, A_i\big) = p\big(x_t | \mathcal{I}_{t-h}^{(\mathcal{P})}, A_i\big).$$

That is, the additional information available in $\mathcal{I}_{t-h}^{(\mathcal{P})}$ relative to $\mathcal{I}_{t-h}^{(i)}$ does not change the density forecast of x_t for any model $i = 1, \ldots, M$. This holds trivially for the VAR models, but also for the three DSGE models since the "missing" variables in $\mathcal{I}_{t-h}^{(i)}$ are not predicted by these models.

The BMA weights in (E.7) are based on the assumption that $x_t^{(m)}$ is observed at t. As discussed in Section 2 of the paper, the real-time dimension means the BMA weights need to be adjusted by lagging the predictive likelihoods on the right hand side by the information lag. In other words, we replace equation (E.7) with

$$\hat{w}_{i,t,h} = \frac{\hat{w}_{i,t-1,h} p\big(x_{t-l}^{(m)} | \mathcal{I}_{t-h-l}^{(i)}, A_i\big)}{\sum_{j=1}^{M} \hat{w}_{j,t-1,h} p\big(x_{t-l}^{(m)} | \mathcal{I}_{t-h-l}^{(j)}, A_j\big)}. \tag{E.8}$$

Like in the cases of the prediction pools, the first period when an h-step-ahead predictive likelihood value is observed occurs at $T+h+l$. For $t = T+1, \ldots, T+h+l-1$ we can therefore replace the predictive likelihood values in (E.8) with a positive constant such that $\hat{w}_{i,t,h} = \hat{w}_{i,T,h}$, the prior predictive probability of model i.

The predictive likelihood for the BMA combination is given by

$$p\big(x_{t+h}^{(a)} | \mathcal{I}_t^{(\mathcal{P})}, \mathcal{P}\big) = \sum_{i=1}^{M} \hat{w}_{i,t,h} p\big(x_{t+h}^{(a)} | \mathcal{I}_t^{(i)}, A_i\big),$$

from which it is straightforward to compute the log predictive score, denoted by $S_{T:T_h,h}^{(\text{BMA})}$, of this density forecast combination method.

Dynamic Model Averaging

Raftery et al. (2010) proposed a model combination method, called dynamic model averaging (DMA), where the weights of the log predictive score may be regarded as depending on a hidden Markov process, s_t, as in Waggoner and Zha (2012), but where the estimates of the Markov transition probabilities are approximated. Specifically, Raftery et al. (2010) suggest to estimate the weights as follows

$$w_{i,t+h|t} = \frac{\hat{w}_{i,t,h}^{\varphi^h}}{\sum_{j=1}^{M} \hat{w}_{j,t,h}^{\varphi^h}}, \tag{E.9}$$

where φ is a parameter such that $0 \leq \varphi \leq 1$. Notice that the weights $\hat{w}_{i,t,h}$ on the right hand side of equation (E.9) correspond to the weights from the recursive BMA calculation. Accordingly, if $\varphi = 1$ then DMA is identical to BMA, and if $\varphi = 0$ then DMA implies that the weights are equal $(1/M)$ for all time periods. The parameter φ is interpreted as a *forgetting factor*, where lower values means that past forecast performance is given a lower weight; see also Koop and Korobilis (2012). Provided that $\varphi < 1$, it follows that the DMA weights approach the equal weights as h increases.

The predictive likelihood for the DMA forecast combination is given by

$$p\big(x_{t+h}^{(a)}|\mathcal{I}_t^{(\mathcal{P})}, \mathcal{P}\big) = \sum_{i=1}^{M} w_{i,t+h|t} p\big(x_{t+h}^{(a)}|\mathcal{I}_t^{(i)}, A_i\big),$$

from which the log predictive score, denoted by $S_{T:T_h,h}^{(\text{DMA})}$, can be directly computed.

Amisano and Geweke (2017) estimate the forgetting factor by recursively maximizing the log predictive score over a grid of φ values, similar to the case for ρ in in the subsection on dynamic prediction pools above. They suggest that the procedure can be given a Bayesian interpretation where the researcher assigns a flat (uniform) prior on φ. DMA is also considered by Del Negro et al. (2016), who consider three values of φ below but close to unity.

Log Score Weights and KLIC Weights

Jore et al. (2010) suggests to make use of weights based on the scoring function applied to compare models and density combination methods. For the log predictive score this amounts to setting the recursive weights equal to

$$w_{i,h,t} = \frac{\exp\left(S_{T:t-l-h,h}^{(i)}\right)}{\sum_{j=1}^{M} \exp\left(S_{T:t-l-h,h}^{(j)}\right)}, \quad i = 1, \ldots, M, \quad t = T + l + h, \ldots, T_h, \tag{E.10}$$

while the weights are initialized at fixed values for $t = T, \ldots, T + l + h - 1$. Jore et al. (2010) assume that the information lag is equal to $l = 1$ and they refer to this weighting scheme as a recursive one. Since all combination methods but the fixed weights are recursive, we refer to this density forecast combination method as the log score (LS) approach. It is interesting to note that the LS weights in (E.10) are identical to the BMA weights in (E.8), provided that equal weights are used for the initial conditions.

It may be noted that the LS weighting scheme is similar to the KLIC weighting scheme suggested by Mitchell and Hall (2005). The KLIC weights for model i and forecast horizon h are given by

$$w_{i,h,t} = \frac{\exp\big(-\Delta_{i,h,t}\big)}{\sum_{j=1}^{M} \exp\big(-\Delta_{j,h,t}\big)}, \tag{E.11}$$

where

$$\Delta_{i,h,t} = \mathrm{KLIC}_{i,h,t} - \min_{j=1,\ldots,M} \mathrm{KLIC}_{j,h,t}.$$

The KLIC estimate for model i at t, taking the information lag and the forecast horizon into account, is

$$\mathrm{KLIC}_{i,h,t} = \frac{1}{t-l-h-T} \sum_{\tau=T}^{t-l-h} \log f_\tau\big(x_{\tau+h}^{(m)} \big| \mathcal{I}_\tau\big) - \frac{1}{t-l-h-T} S_{T:t-l-h,h}^{(i)},$$

where $f_\tau(x_{\tau+h}^{(m)}|\mathcal{I}_\tau)$ is the "true" but unknown density for predicting $x_{\tau+h}$. For the KLIC based weights this true density cancels out when determining $\Delta_{i,h,t}$ and the resulting weight for model i at t for the h-step-ahead forecasts is therefore the average log score of model i minus the average log score of the worst performing model up to that period. The latter term cancels out in equation (E.11) with the effect that the KLIC weights are based on the *average* log score rather than the log score, as in equation (E.10).[12] This also means that the KLIC based approach has a DMA interpretation, but where the "forgetting factor" recursively becomes smaller with t in $1/(t-l-h-T)$ and larger with h, where the latter is a property shared with DMA.

[12] Keeping in mind that $\exp(ab) = (\exp(b))^a$ it follows that the LS and KLIC weights are not equal, except when $t = T + l + h + 1$.

Section F: Data and Transformations

Table I.1 lists the observable variables and expresses their different treatment across the DSGE and BVAR models. The SW model uses seven observables: real GDP, private consumption, total investment, real wages, total employment, the GDP deflator, and the short-term nominal interest rate (given by the three-month EURIBOR). For the SWU model there is the addition of the unemployment rate, and for SWFF there is the addition of the financial lending spread. The BVARs, by contrast, use all nine observables (albeit mostly with a different transformation, see below).

For the DSGE models and the BVAR with heteroskedastic innovations (stochastic volatility), the first five observables are transformed into quarterly growth rates by the first difference of their logarithm multiplied by 100, whilst the BVARs with homoskedastic innovations instead use the log level of these variables multiplied by 100. The inflation time series is obtained as the first difference of the log of the GDP deflator times 100 and the same transformation is used in all models. The final three observables (r, u and s) are also defined in the same manner across the DSGEs and BVARs with the interest rate and spread being expressed in annualized percentage terms. As in Smets et al. (2014), we only consider data from 1979Q4, such that the growth rates are available from 1980Q1. All variables are available at a quarterly frequency, except for the unemployment and the interest rate series which exist at a monthly frequency.

The euro area real-time database (RTD), on which these models are estimated and assessed, is described in Giannone et al. (2012). To extend the data back in time, we follow Smets et al. (2014) and McAdam and Warne (2019) and link the real-time data to various updates from the area-wide model (AWM) database; see Fagan et al. (2005). The exception is the spread which is not included in the RTD vintages nor in the AWM updates.

The RTD covers vintages starting in January 2001 and has been available on a monthly basis until early 2015 when the vintage frequency changed from three to two per quarter. We consider the vintages from 2001Q1–2019Q4 for estimation and forecasting. For more detailed information on the real-time vintages until 2014Q4, see McAdam and Warne (2019), while the Online Appendix in Warne (2022a, Table B.2) displays the ragged edge until 2019Q4.

Section G: Estimation of the Models & Point Forecasts

I. Estimation of the Models

The real-time euro area dataset is extensively discussed in McAdam and Warne (2019), including how these vintages can be extended back in time until 1970 using vintages from the area-wide model database; see also Smets et al. (2014). Below, we use an extended sample of vintages from the euro area real-time database, i.e., from 2001Q1 until 2019Q4 instead of until 2014Q4. The three DSGE models in McAdam and Warne (2019) are estimated using observations from 1985Q1, while 1980Q1–1984Q4 is used as a training sample. We also estimate the BVAR models using observations from 1985Q1, while the initialization vector X_0 is built from observations prior to this date. Specifically, the three BVAR models we study below have $p = 4$ lags such that X_0 is formed by observations from 1984. Details on the data transformations and the variables included in the various models are provided in the Online Appendix, Section F.

The predictive likelihoods of the DSGE models have previously been estimated until the vintage from 2014Q4 and the Bayesian estimation approach is discussed in McAdam and Warne (2019), while the extension until 2019Q4 used below follows the same procedure. In their study, 750,000 posterior draws of the parameters of each DSGE model—using the random-walk Metropolis sampler—have been computed on an annual basis for the Q1 vintages, reflecting how often such models are typically re-estimated by policy institutions in practice. The Q1 parameter draws are then also used for the Q2 until Q4 vintages within the same year. Treating the first 250,000 draws as a burn-in period of the sampler, the predictive likelihoods for backcasts, nowcasts and up to eight-quarter-ahead forecasts have thereafter been estimated using 10,000 of the remaining 500,000 draws, where each used draw is separated by 50 draws.

The parameters of the BVAR models are estimated by trimming the data such that the last two time periods of each vintage are discarded. Furthermore, the posterior draws of the parameters for the SoC and PLR models are obtained from their normal-inverted Wishart distributions by setting the vector of estimated hyperparameters to its posterior mode value. In contrast to the DSGE models, these two BVAR models are re-estimated for each vintage with 100,000 posterior draws. The density forecasts are based on the full vintage dataset; see the Online Appendix, Section B. Since the posterior draws are independent, all draws are used when estimating the predictive likelihood.

The SV model is estimated through a Gibbs algorithm, extended with a Metropolis-Hastings step. For this model, we have used 10,000 burn-in draws for the posterior sampler and 50,000 remaining draws, where every 10^{th} draw is used for the forecasting exercise and the model is re-estimated for each real-time vintage.

II. Point Forecasts

The point forecast, given by the mean of the predictive density, is estimated by averaging the point forecast conditional on the parameters over the posterior draws; see, for example, McAdam and Warne (2019, Section 5.2) and Appendix B and Appendix C of this Online Appendix. The recursively estimated paths are displayed in so-called spaghetti plots in Figure I.1 for the DSGE and

BVAR models, with the real GDP growth forecasts in Panel A (top) and the inflation forecasts in Panel B (bottom). The actual values are plotted with solid black lines, while the dashed lines are the recursive posterior estimates of the population mean of the two variables for the DSGE models. For the BVAR models, where some variables appear in levels rather than in first differences, the dashed lines instead trace out the vintage sample mean values of real GDP growth and inflation since 1995Q1.

Turning first to the real GDP growth forecasts in Panel A, the BVAR model forecasts tend to be lower than those obtained from the DSGE models. This is not only true in the aftermath of the Great Recession, but also prior to this episode. The mean forecast errors for the whole sample are shown in Table I.3 and this visual observation is indeed confirmed as the mean errors concern the difference between the actual value and the point forecast. The smallest mean errors are generated by the BVAR model with the PLR, while those of the BVAR with the SoC prior are roughly 0.05–0.15 percentage points larger in absolute terms and those of the SV model even larger. Since the mean errors are all negative it follows that all six models over-predict real GDP growth on average with overall larger errors in absolute terms for the DSGE models than for the BVAR models.[13]

Concerning the inflation point forecasts in Panel B of Figure I.1, the SW and SWU model forecast paths are, as pointed out by McAdam and Warne (2019), quite similar up to 2014, with a strong tendency for mean reversion over the forecast horizon. However, from around 2016 the SWU paths still follow the same pattern, while those of the SW model tend to be flatter and even downward sloping. The point forecasts of the SWFF model are quite different with v-shaped paths with a strong tendency to drift down from one vintage to the next. Turning to the BVAR models, the forecasts paths are flatter and more varied than those of the DSGE models. From the mean errors in Table I.3, it can be seen that the DSGE models tend to under-predict inflation for the shorter horizons and, in the cases of the SW and SWU models, over-predict inflation in the medium and long term. As can also be inferred from the spaghetti-plots for the SWFF model, it under-predicts inflation, albeit with a slightly decreasing error as the forecast horizon grows. By contrast, the mean errors of the BVAR models are all smaller in absolute terms than those for the DSGE models and suggest that the models weakly under-predict inflation. Recalling that both real GDP growth and inflation are measured in quarterly terms such that their values are comparable in terms of scale, the evidence in Table I.3 suggests that the six models tend to forecast inflation better than real GDP growth in terms of point forecasts.

[13] See McAdam and Warne (2019) for discussions on the finding that the DSGE models' point forecast paths of real GDP growth tend to jump up somewhat after the Great Recession, while the path of posterior population mean estimates is downward-sloping.

Section H: Impact of the Initial Weights on the Combination Method Results

Changing the initialization from equal weights to some other weighting scheme is straightforward when applying the ALS, the SOP as well as the BMA and the DMA combination methods. For the DP, however, the weight initialization depends on the parameters of the underlying process, ξ_t, as well as on the information available through the model predictive likelihood values. The default parameterization of ξ_t favors the propagation of equal weights through the initialization of ξ_{T-1} and the innovation η_t.

The predictive likelihood values of the models are initially set to a constant, such as unity or $1/M$, for the time periods $t = T, \ldots, T+h+l-1$ due to the information lag, l, and the forecast horizon, h, delaying when actual realizations of the model predictive likelihoods can be observed. When the number of particles is large enough, this initialization means that the model weights estimator

$$\frac{1}{N}\sum_{n=1}^{N} w_{i,t+h|t}^{(n)}(\rho) W_t^{(n)} \approx \frac{1}{M}, \quad t = T, \ldots, T+h+l-1.$$

One way to influence the weights during this initialization phase is to feed the bootstrap particle filter with alternative predictive likelihood values. For instance, the case when the SWFF model should be given zero initial weight while the other five models each have the weight 1/5 can, partially, be implemented by setting the predictive likelihood values of the models equal to these weights during the first $h+l$ periods. The reason why this only partially changes the weights to the desired values is related to the initialization of ξ_{T-1} and to the assumption about η_t. Since both vectors are assumed to be standard normal, it follows that for large N, the sample average of $w_{i,T-1} = 1/6$ ($M = 6$) and similarly for the candidate model weights $\tilde{w}_{i,t}$ at the beginning of the forecasting step in the Bayesian bootstrap filter. The expected value of the incremental weights in period T, $\tilde{\omega}_T$, is therefore given by the weighted average of the predictive likelihood values (1/6).

During the updating step, the particles are given a candidate re-weighting scheme, $\tilde{W}_T^{(n)}$, based on the old particle weights (unity) and the incremental weights. This candidate scheme will not favor equal weights since the predictive likelihood value of those particles based on a low weight on the SWFF model will achieve a larger predictive likelihood than average. For the selection step, two possibilities exist, but with similar outcomes. If the effective sample size remains above the selected threshold size, $\delta^* N$, the re-weighted candidate scheme of particle weights will be applied to the candidate model weights when estimating the model weights that are used to compute the predictive likelihood for the nowcast and the h-quarter-ahead forecasts. The corresponding model weights will have values less than 1/6 for the SWFF model but greater than zero since the candidate model weights have mean 1/6. The other models obtain approximately equal values somewhat larger than 1/6. On the other hand, if the effective sample size is below the threshold size, resampling occurs based on the swarm of candidate model and particle weights where model weights for particles with larger particle weights have a greater chance of survival, i.e., those with a low weight on the SWFF model. Again, the result is that average model weight has a value less than 1/6 for the SWFF model but greater than zero, while the average model weights on the other five models are approximately equal and somewhat larger than 1/6.

The inherent equal-weights force of ξ_t through its initialization and η_t can be guided towards other steady-state weights by tuning its distribution. One option is to change the mean of the normal distribution from zero to, say, -1.96 for the models where we wish to consider an initial weight close to zero. Since the purpose is to aim at a particular vector of near fixed weights during the first $h + l$ time periods of the forecast sample, the mean of the normal distribution need only be swapped during this period. Thereafter, the zero mean value is applied again, bearing in mind that the η_t random draws again push the dynamic pool weights towards equal weights. In this section, we consider the case when the mean of ξ_{T-1} and η_t are equal to -1.96 for the SWFF model and zero for the other models. This not only ensures that the weight on the SWFF model is relatively close to 0 while the other weights are close to 1/5 up to time period $T + h + l - 1$, but it also allows the $\xi_{i,t}$ process for the SWFF model to recover thereafter, provided that its predictive likelihood values give sufficient support for the model. Furthermore, by limiting the number of periods for the change of η_t to the initial $h + l$ periods of the forecast sample, it follows that this also applies when iterating the ξ_t process forward when computing the h-step-ahead weight.[14]

The full sample log predictive scores for the alternative initialization scheme with "zero weight" on the SWFF model and equal weights on the other models are shown in Table I.6 for the joint real GDP growth and inflation case.[15] The fixed weights cases in the table are denoted FW since this covers both the EW combination and the combination with zero weight on the SWFF model. It is striking that the log scores of all combination methods are positively affected by having a zero initial weight on the SWFF model, with the exception of the SOP forecasts at some horizons ($h = 4, 5, 7$); it should be kept in mind that by construction the selection of weights for the SOP does not depend on lagged weights. The fixed weight combination scheme obtains the highest log score for all horizons when the SWFF model is excluded, except when $h = 5$, and the improvements are particularly notable for the shorter horizons. Although the DP also records substantial gains, they are not sufficiently large for it to hold on to the first rank.

The posterior model weights for the joint real GDP growth and inflation density forecasts using the DP and the SWFF zero initialization are displayed in Figure I.26. The grey lines in the panels are the posterior weights based on the equal weights initialization; see also Figure 6 of our article. The SWFF zero initialization method we have proposed for the DP indeed approximately provides the intended properties, i.e., close to zero weights during the first $h+l$ periods of the forecast sample and a chance to recover thereafter. In general, the weights on the SWFF model gradually become larger over the sample, especially for the longer term forecasts. The weights on the other models shift proportionally relative to the equal weights initialization to offset the lower weight on the SWFF model, and with the shift becoming smaller as the weight on the SWFF model depends less on the selected initialization method.

[14] This implementation is equivalent to introducing a time-varying drift-term into the equation for $\xi_{i,t}$, when i is the SWFF model, and having a zero mean for the distributions of ξ_{T-1} and η_t.

[15] The results for the marginal cases for real GDP growth and inflation are provided in Tables I.7-I.8. Recursive estimates of the changes to the log predictive scores for selected horizons are shown in Figures I.24-I.25.

Section I: Tables and Figures

TABLE I.1: Observables used in the DSGE and BVAR models.

		Observed Variables					
Variable	Symbol	DSGEs:	SW	SWU	SWFF	SoC & PLR	SV
real GDP (log)	y	$100 \times \Delta y_t$	$\surd$	$\surd$	$\surd$	$100 \times y_t$	$100 \times \Delta y_t$
real private consumption (log)	c	$100 \times \Delta c_t$	$\surd$	$\surd$	$\surd$	$100 \times c_t$	$100 \times \Delta c_t$
real total investment (log)	i	$100 \times \Delta i_t$	$\surd$	$\surd$	$\surd$	$100 \times i_t$	$100 \times \Delta i_t$
real wages (log)	w	$100 \times \Delta w_t$	$\surd$	$\surd$	$\surd$	$100 \times w_t$	$100 \times \Delta w_t$
total employment (log)	e	$100 \times \Delta e_t$	$\surd$	$\surd$	$\surd$	$100 \times e_t$	$100 \times \Delta e_t$
GDP deflator inflation (log, quarterly)	π	$100 \times \pi_t$	$\surd$	$\surd$	$\surd$	$100 \times \pi_t$	$100 \times \pi_t$
short-term nominal interest rate (%)	r	r_t	$\surd$	$\surd$	$\surd$	r_t	r_t
unemployment rate (%)	u	$100 \times u_t$		$\surd$		$100 \times u_t$	$100 \times u_t$
spread (%)	s	s_t			$\surd$	s_t	s_t

TABLE I.2: The ragged edge of the euro area RTD: Vintages with missing data for the variables.

Date	y	c	i	π	e	w	r	u
Backcast	2001Q1–2001Q2	2001Q1–2001Q2	2001Q1–2001Q2	2001Q1–2003Q3	2001Q1–2017Q3	2001Q1–2017Q7	–	–
	2002Q1	2002Q1	2002Q1					
		2003Q3	2003Q3					
		2004Q3	2004Q3	2004Q3				
		2006Q1	2006Q1	2006Q1				
		2006Q3	2006Q3	2006Q3				
		2014Q3–2015Q4	2014Q3–2015Q4	2014Q3–2015Q4				
		2016Q2	2016Q2	2016Q2				
					2018Q1	2018Q1		
		2019Q1	2019Q1	2019Q1		2019Q1		
Total	3 of 76	15 of 76	15 of 76	22 of 76	68 of 76	69 of 76	0 of 76	0 of 76
Nowcast	2001Q1–2019Q4	2001Q1–2019Q4	2001Q1–2019Q4	2001Q1–2019Q4	2001Q1–2019Q4	2001Q1–2019Q4	–	2005Q1
								2005Q3–2005Q4
								2006Q3
								2008Q1
Total	76 of 76	76 of 76	76 of 76	76 of 76	76 of 76	76 of 76	0 of 76	5 of 76

TABLE I.3: Mean errors based on the predictive mean as the point forecast for the sample 2001Q1–2019Q4.

	Real GDP growth						Inflation					
	DSGE			BVAR			DSGE			BVAR		
h	SW	SWFF	SWU	SoC	PLR	SV	SW	SWFF	SWU	SoC	PLR	SV
−1	0.045	−0.337	−0.218	−0.243	−0.139	−0.159	0.133	0.237	0.087	0.104	0.071	0.420
0	−0.160	−0.372	−0.250	−0.110	−0.057	−0.158	0.076	0.256	0.019	0.055	0.024	0.046
1	−0.333	−0.580	−0.357	−0.148	−0.074	−0.228	0.042	0.320	−0.028	0.064	0.024	0.046
2	−0.375	−0.638	−0.388	−0.168	−0.071	−0.247	0.006	0.336	−0.072	0.069	0.024	0.046
3	−0.378	−0.646	−0.380	−0.186	−0.068	−0.250	−0.028	0.329	−0.112	0.076	0.027	0.046
4	−0.356	−0.628	−0.346	−0.190	−0.060	−0.241	−0.067	0.304	−0.153	0.074	0.024	0.040
5	−0.337	−0.612	−0.311	−0.199	−0.061	−0.239	−0.099	0.278	−0.187	0.073	0.024	0.036
6	−0.316	−0.595	−0.276	−0.206	−0.064	−0.239	−0.124	0.252	−0.214	0.073	0.026	0.035
7	−0.295	−0.577	−0.241	−0.209	−0.065	−0.238	−0.151	0.223	−0.240	0.070	0.024	0.030
8	−0.272	−0.555	−0.206	−0.206	−0.062	−0.233	−0.175	0.197	−0.260	0.068	0.026	0.028

TABLE I.4: Log predictive scores for the marginal density forecasts of real GDP growth and GDP deflator inflation over the vintages 2001Q1–2019Q4, with $h = -1$ being the backcast horizon and $h = 0$ the nowcast horizon.

	Real GDP growth								Inflation							
	DSGE			BVAR			Bounds		DSGE			BVAR			Bounds	
h	SW	SWFF	SWU	SoC	PLR	SV	Upper	Lower	SW	SWFF	SWU	SoC	PLR	SV	Upper	Lower
−1	−2.12	−2.32	−2.18	−0.85	−0.59	0.00			−3.91	−12.38	−2.65	−0.17	0.00	−1.37		
						−0.20	0.05	−2.79					1.64		5.82	−13.94
0	−21.98	−28.22	−26.18	−17.03	−7.94	0.00			−9.76	−45.38	0.00	−0.67	−2.14	−12.52		
						−36.95	−15.78	−92.45			21.07				34.66	−37.74
1	−13.51	−27.08	−16.11	−7.94	−6.41	0.00			−9.49	−49.88	−0.80	0.00	−1.58	−8.74		
						−57.34	−30.97	−107.22				17.37			29.88	−41.83
2	−8.13	−22.20	−9.48	−2.24	0.00	−0.37			−9.49	−46.41	−1.11	0.00	−4.90	−6.28		
					−66.08		−33.21	−112.59				10.99			25.43	−45.67
3	−11.09	−23.56	−10.52	−3.14	0.00	−2.73			−11.98	−41.92	−3.42	0.00	−6.12	−4.29		
					−64.96		−35.05	−112.76				7.19			22.00	−46.55
4	−11.90	−23.58	−10.18	−4.16	0.00	−4.13			−12.87	−34.33	−5.13	0.00	−6.04	−0.47		
					−63.19		−35.76	−110.58				3.79			19.14	−43.70
5	−11.16	−22.50	−8.15	−4.15	0.00	−6.19			−14.26	−28.86	−7.16	−1.58	−8.19	0.00		
					−62.44		−36.40	−109.16						2.01	16.49	−41.48
6	−9.89	−20.99	−6.52	−3.88	0.00	−11.92			−17.31	−25.49	−11.03	−3.26	−11.27	0.00		
					−52.31		−37.24	−107.61						1.78	14.11	−40.14
7	−9.63	−20.82	−6.52	−3.68	0.00	−7.68			−19.78	−21.16	−14.65	−6.26	−14.26	0.00		
					−60.85		−37.21	−105.74						1.63	12.75	−38.36
8	−7.34	−19.09	−5.26	−1.68	0.00	−5.74			−21.46	−18.04	−17.05	−8.20	−15.98	0.00		
					−60.90		−36.50	−103.52						0.55	11.15	−39.25

NOTES: The log scores are displayed in deviation from the largest value for a given forecasted vector of variables and horizon. The largest log score is shown in the row below for the model which achieves this value. The backcast horizon is denoted by $h = -1$ while the nowcast horizon is given by $h = 0$. The ex-post-based upper and lower bounds for combination methods are shown in the last two columns.

TABLE I.5: Log predictive scores for marginal density forecasts of real GDP growth and GDP deflator inflation based on different information lags over the vintages 2001Q1–2019Q4.

		Real GDP growth							Inflation						
h	l	SOP	DP	BMA	DMA	ALS	Upper	Lower	SOP	DP	BMA	DMA	ALS	Upper	Lower
0	1	−36.95	−40.34	−38.74	−35.64	−44.09	−15.78	−92.45	20.44	16.79	19.77	19.73	15.37	34.66	−37.74
	1*	−38.96	−39.82	−38.75	−37.99	−44.14			20.16	16.58	19.51	19.16	15.33		
	4	−41.98	−42.57	−39.74	−43.75	−44.58			20.78	16.62	19.71	19.61	15.35		
1	1	−61.60	−55.01	−65.66	−60.49	−58.24	−30.97	−107.22	16.31	12.61	15.58	15.68	11.51	29.88	−41.83
	1*	−61.63	−54.64	−62.35	−61.49	−58.24			16.29	12.41	14.81	14.52	11.49		
	4	−64.62	−56.12	−68.45	−64.68	−58.57			16.46	12.45	15.80	15.85	11.40		
2	1	−70.91	−58.48	−79.12	−68.28	−61.58	−33.21	−112.59	10.78	6.38	9.03	9.00	5.30	25.43	−45.67
	1*	−71.07	−58.35	−76.50	−68.33	−61.58			10.06	6.13	7.70	8.34	5.20		
	4	−72.98	−59.37	−80.86	−70.71	−61.85			10.29	6.24	8.59	9.76	5.17		
4	1	−71.42	−59.89	−76.49	−67.54	−61.65	−35.76	−110.58	0.58	−1.14	0.88	0.75	−1.63	19.14	−43.70
	1*	−71.46	−59.94	−72.92	−66.91	−61.62			1.03	−1.28	−1.40	1.05	−1.61		
	4	−72.57	−60.28	−75.96	−68.30	−61.78			−0.12	−1.45	0.20	0.34	−1.72		
8	1	−71.62	−57.84	−66.04	−59.96	−58.78	−36.50	−103.52	−1.79	−9.38	−2.22	−5.70	−9.67	11.15	−39.25
	1*	−71.72	−58.04	−63.92	−59.13	−58.74			−1.88	−9.58	−3.78	−7.24	−9.71		
	4	−70.93	−57.88	−63.99	−59.86	−58.85			−1.56	−9.44	−2.24	−5.71	−9.67		

TABLE I.6: Log predictive scores of joint real GDP growth and GDP deflator inflation for nowcasts and one-to eight-quarter-ahead forecasts of six combination methods based on different initializations over the vintages 2001Q1–2019Q4.

h	Initial weights	FW	SOP	DP	BMA	DMA	ALS
0	SWFF zero	−24.14	−31.56	−25.48	−36.80	−29.18	−28.18
	Equal	−29.88	−31.82	−27.11	−37.18	−30.13	−28.44
1	SWFF zero	−39.27	−50.66	−39.46	−58.10	−47.91	−43.97
	Equal	−47.12	−51.06	−41.97	−58.64	−48.93	−44.38
2	SWFF zero	−48.47	−68.87	−49.06	−72.49	−57.03	−53.78
	Equal	−56.16	−69.33	−51.70	−73.10	−58.22	−54.24
3	SWFF zero	−54.35	−73.72	−55.25	−77.18	−63.44	−59.34
	Equal	−61.36	−74.26	−57.79	−77.85	−64.63	−59.86
4	SWFF zero	−56.87	−70.68	−57.12	−78.47	−65.71	−60.91
	Equal	−63.22	−69.14	−59.54	−79.17	−66.54	−61.51
5	SWFF zero	−59.60	−71.43	−59.60	−81.34	−66.59	−63.54
	Equal	−65.33	−70.44	−61.86	−82.07	−67.34	−64.16
6	SWFF zero	−62.06	−71.22	−61.88	−79.73	−67.01	−65.52
	Equal	−67.14	−72.85	−63.80	−80.47	−68.20	−66.15
7	SWFF zero	−63.83	−72.92	−63.21	−79.79	−66.80	−66.87
	Equal	−68.12	−70.90	−64.95	−80.54	−67.93	−67.53
8	SWFF zero	−65.46	−69.97	−64.39	−78.23	−65.88	−67.87
	Equal	−69.17	−72.02	−65.78	−78.87	−66.57	−68.57

NOTES: The initial weight scheme denoted "SWFF zero" has equal weights 1/5 on the SW, SWU, BVAR SoC, BVAR PLR and BVAR SV models and weight 0 on the SWFF model, except for the dynamic pool, where the initialization scheme is random but where the weights approximate such fixed weights on average. The initial weight scheme "Equal" gives weight 1/6 to each model during the initialization sample. For the fixed weight (FW) combination method the weights are constant at the "initial weights".

TABLE I.7: Log predictive scores of real GDP growth for nowcasts and one-to eight-quarter-ahead forecasts of six combination methods based on different initializations over the vintages 2001Q1–2019Q4.

h	Initial weights	FW	SOP	DP	BMA	DMA	ALS
0	SWFF zero	−42.48	−41.92	−41.64	−39.50	−43.19	−44.52
	Equal	−45.29	−41.98	−42.57	−39.74	−43.75	−44.58
1	SWFF zero	−55.03	−64.43	−54.51	−68.05	−63.72	−58.38
	Equal	−58.98	−64.62	−56.12	−68.45	−64.68	−58.57
2	SWFF zero	−57.91	−73.38	−57.61	−80.38	−68.66	−61.59
	Equal	−61.96	−72.98	−59.37	−80.86	−70.71	−61.85
3	SWFF zero	−59.24	−75.54	−59.35	−79.16	−69.40	−62.75
	Equal	−63.17	−75.24	−61.10	−79.71	−71.12	−63.08
4	SWFF zero	−58.55	−73.08	−58.49	−75.39	−66.61	−61.35
	Equal	−62.38	−72.57	−60.28	−75.96	−68.30	−61.78
5	SWFF zero	−57.81	−73.18	−57.95	−71.96	−63.54	−60.62
	Equal	−61.50	−72.20	−59.71	−72.56	−65.26	−61.06
6	SWFF zero	−57.58	−70.87	−57.93	−68.62	−61.65	−60.26
	Equal	−61.07	−69.81	−59.60	−69.26	−63.14	−60.74
7	SWFF zero	−56.93	−73.58	−57.44	−64.59	−59.33	−59.60
	Equal	−60.32	−74.12	−59.01	−65.21	−60.86	−60.12
8	SWFF zero	−55.70	−71.50	−56.24	−63.38	−57.66	−58.28
	Equal	−59.01	−70.93	−57.88	−63.99	−59.86	−58.85

NOTES: See Table I.6 for details.

TABLE I.8: Log predictive scores of GDP deflator inflation for nowcasts and one- to eight-quarter-ahead forecasts of six combination methods based on different initializations over the vintages 2001Q1–2019Q4.

h	Initial weights	FW	SOP	DP	BMA	DMA	ALS
0	SWFF zero	17.92	20.95	18.19	20.06	19.88	15.52
	Equal	14.05	20.78	16.62	19.71	19.61	15.35
1	SWFF zero	14.68	16.68	14.36	16.17	16.14	11.62
	Equal	9.87	16.46	12.45	15.80	15.85	11.40
2	SWFF zero	8.31	10.48	7.74	8.96	10.14	5.36
	Equal	4.00	10.29	6.24	8.59	9.76	5.17
3	SWFF zero	3.78	5.29	2.48	3.33	4.66	1.20
	Equal	0.16	5.08	1.42	2.96	3.25	1.00
4	SWFF zero	0.72	0.10	−0.73	0.58	1.97	−1.49
	Equal	−2.21	−0.12	−1.45	0.20	0.34	−1.72
5	SWFF zero	−2.34	−0.37	−3.68	−1.00	−0.86	−4.08
	Equal	−4.64	−0.61	−3.99	−1.37	−2.24	−4.31
6	SWFF zero	−4.73	0.20	−5.99	−0.57	−1.92	−5.97
	Equal	−6.42	0.01	−5.88	−0.87	−3.94	−6.16
7	SWFF zero	−7.14	−0.29	−8.00	−0.71	−2.83	−7.56
	Equal	−7.99	−0.48	−7.51	−1.02	−4.93	−7.74
8	SWFF zero	−9.75	−1.39	−10.15	−2.07	−4.53	−9.51
	Equal	−10.01	−1.56	−9.44	−2.24	−5.71	−9.67

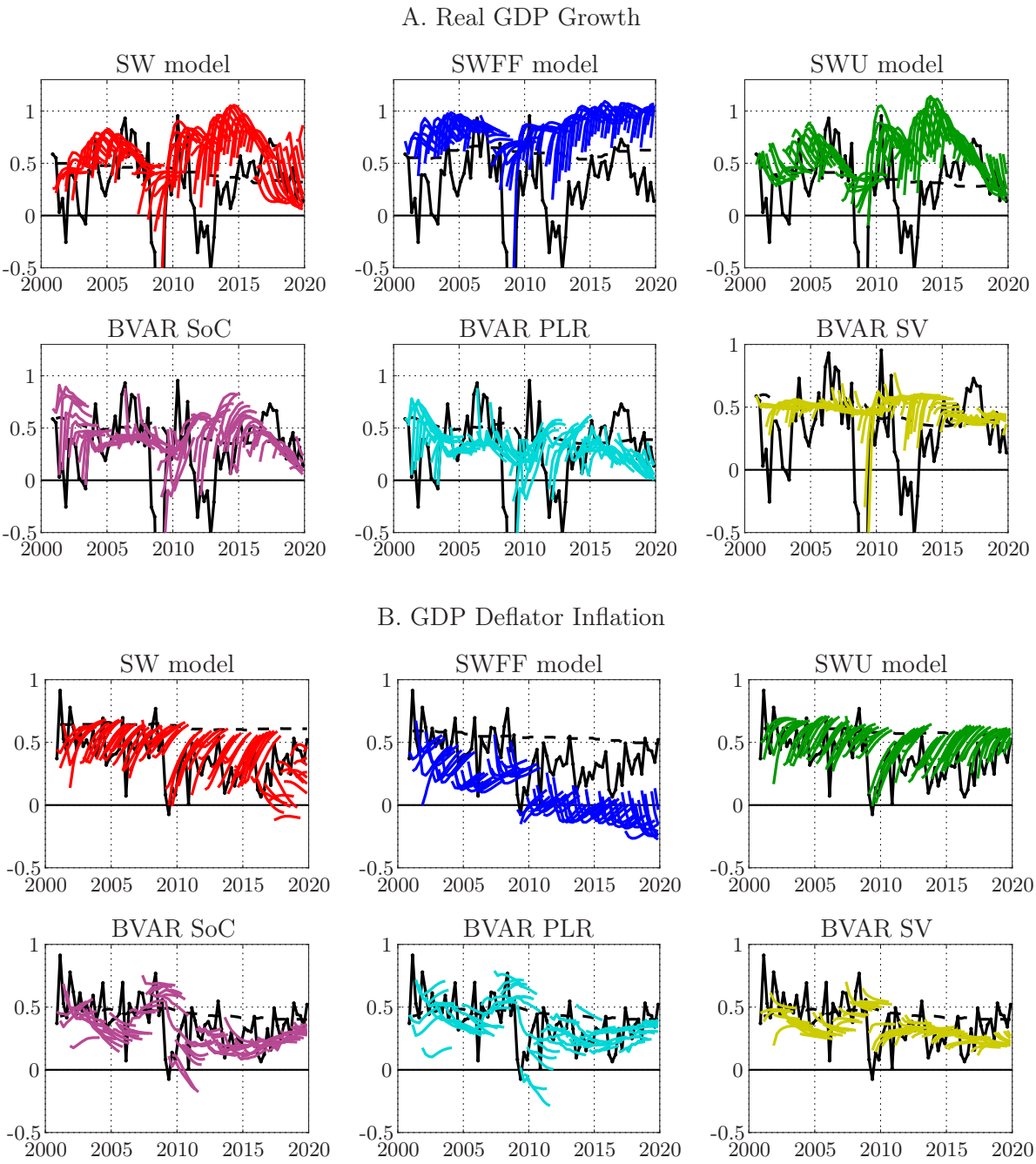
NOTES: See Table I.6 for details.

TABLE I.9: PIT tests for the marginal nowcasts and one-step-ahead density forecasts of real GDP growth and GDP deflator inflation over the vintages 2001Q1–2019Q4.

			Real GDP growth		Inflation	
(p,q)	h	Model	AG	p-value	AG	p-value
$(2,2)$	0	SW	19.22	0.00	6.37	0.17
		SWFF	46.87	0.00	143.70	0.00
		SWU	32.37	0.00	3.30	0.51
		SoC	69.88	0.00	3.83	0.43
		PLR	74.03	0.00	3.04	0.55
		SV	48.50	0.00	6.07	0.19
	1	SW	31.45	0.00	8.41	0.08
		SWFF	91.67	0.00	214.29	0.00
		SWU	36.49	0.00	7.29	0.12
		SoC	21.23	0.00	4.24	0.37
		PLR	17.50	0.00	5.41	0.25
		SV	64.56	0.00	7.90	0.10
$(4,2)$	0	SW	27.44	0.00	7.98	0.24
		SWFF	69.88	0.00	223.23	0.00
		SWU	52.56	0.00	4.42	0.62
		SoC	132.96	0.00	3.92	0.69
		PLR	144.41	0.00	3.33	0.77
		SV	90.43	0.00	6.09	0.41
	1	SW	46.07	0.00	8.30	0.22
		SWFF	144.41	0.00	353.53	0.00
		SWU	55.64	0.00	8.02	0.24
		SoC	36.04	0.00	4.24	0.64
		PLR	32.29	0.00	6.50	0.37
		SV	114.91	0.00	8.38	0.21

NOTES: The Amisano-Geweke PIT tests are described in the Online Appendix of Amisano and Geweke (2017); see also Appendix D above.

FIGURE I.1: Recursive point backcasts, nowcasts and forecasts of real GDP growth and GDP deflator inflation using the RTD vintages 2001Q1–2019Q4.



NOTES: The actual values are plotted as solid black lines. Recursively estimated posterior mean values of mean real GDP growth and mean inflation are plotted as dashed lines using the DSGE models. By contrast, the dashed lines are vintage sample mean values since 1995Q1 of real GDP growth and inflation for the BVAR plots.

FIGURE I.2: Recursive posterior mode estimates of the hyperparameters of the BVAR models for 2001Q1–2014Q4.

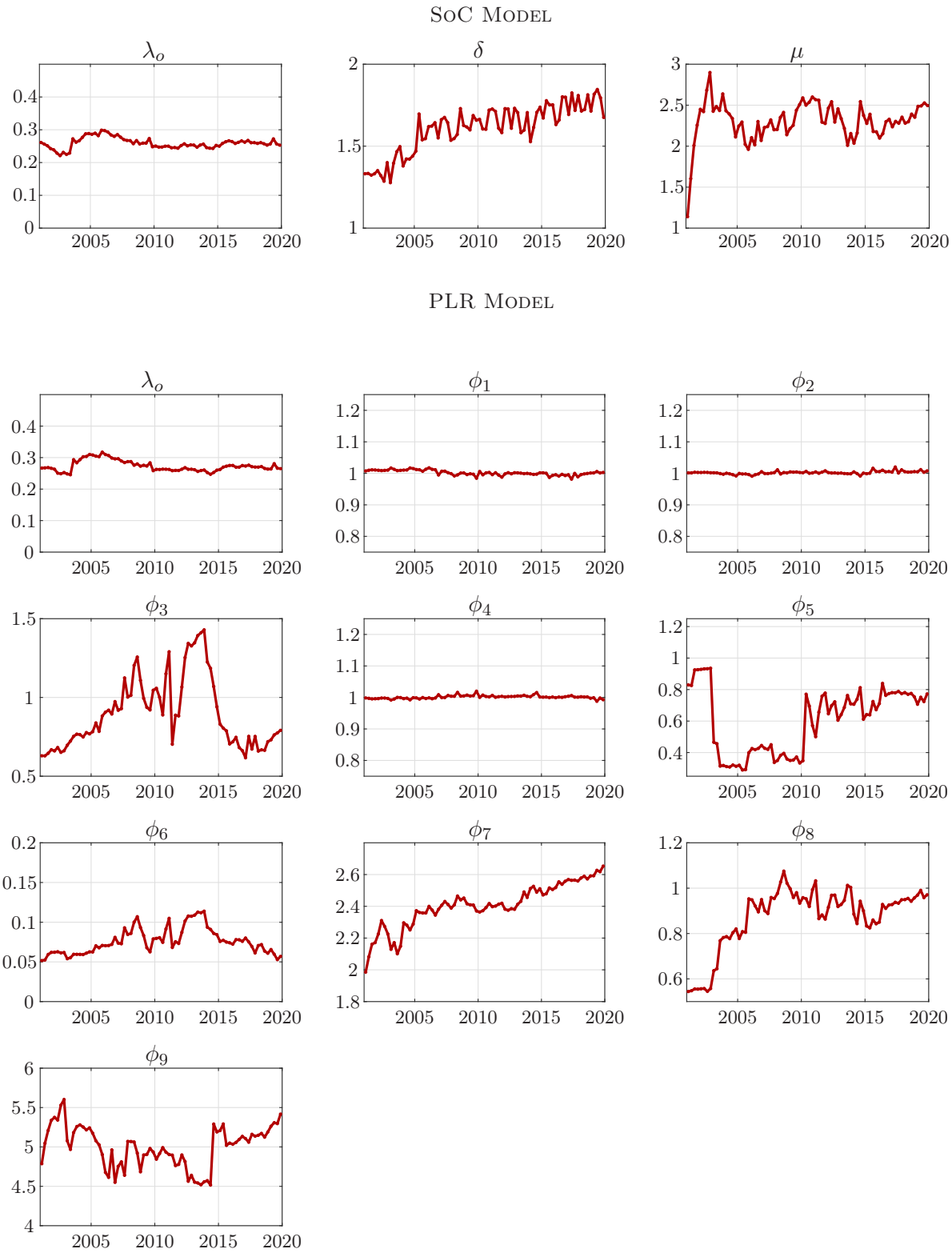
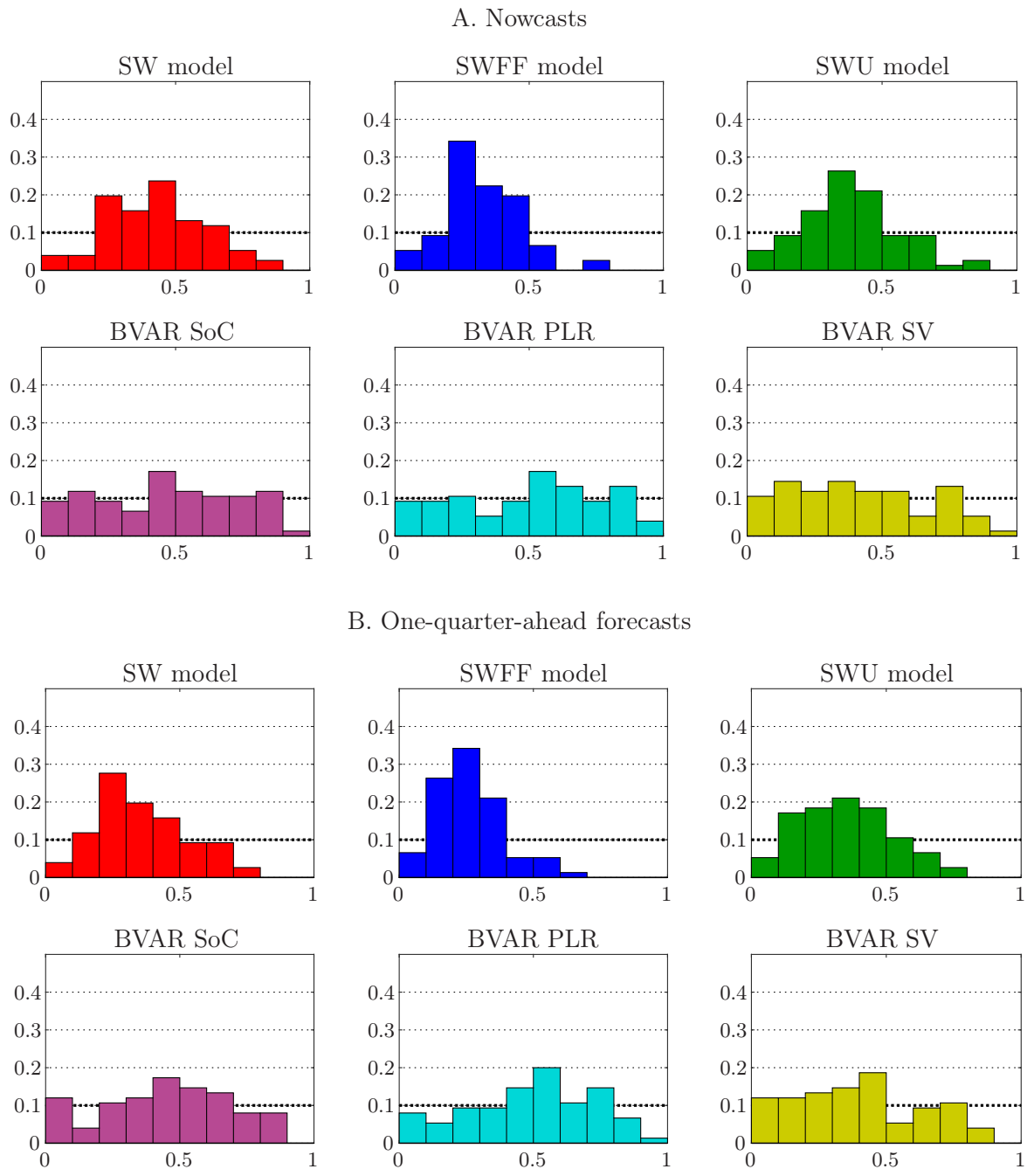
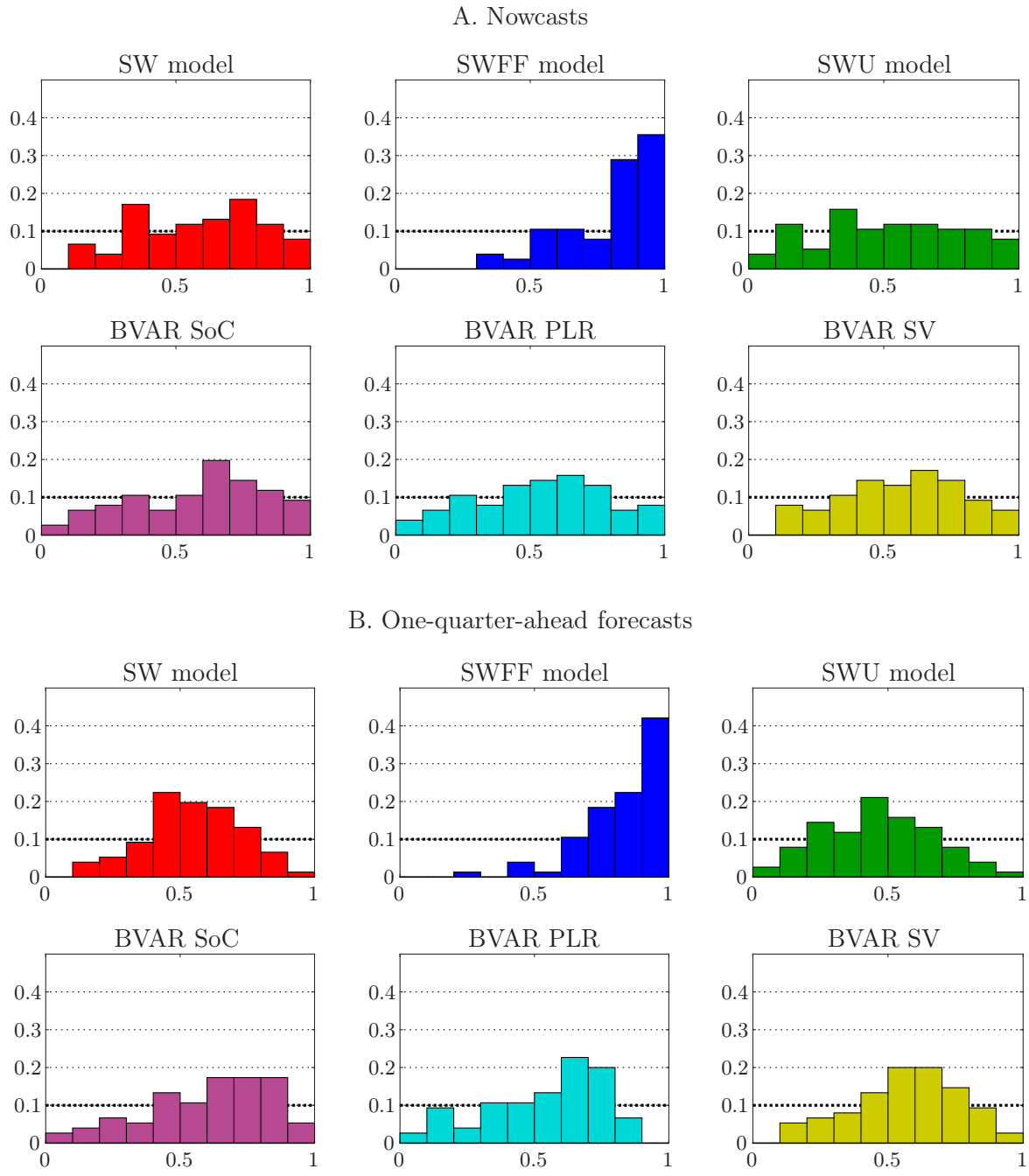


FIGURE I.3: Histograms of the estimated $\pi_{T+h+1|T}$ values for the marginal real GDP growth density forecasts at the nowcast ($h = 0$) and one-quarter-ahead ($h = 1$) horizons for 2001Q1–2019Q4.



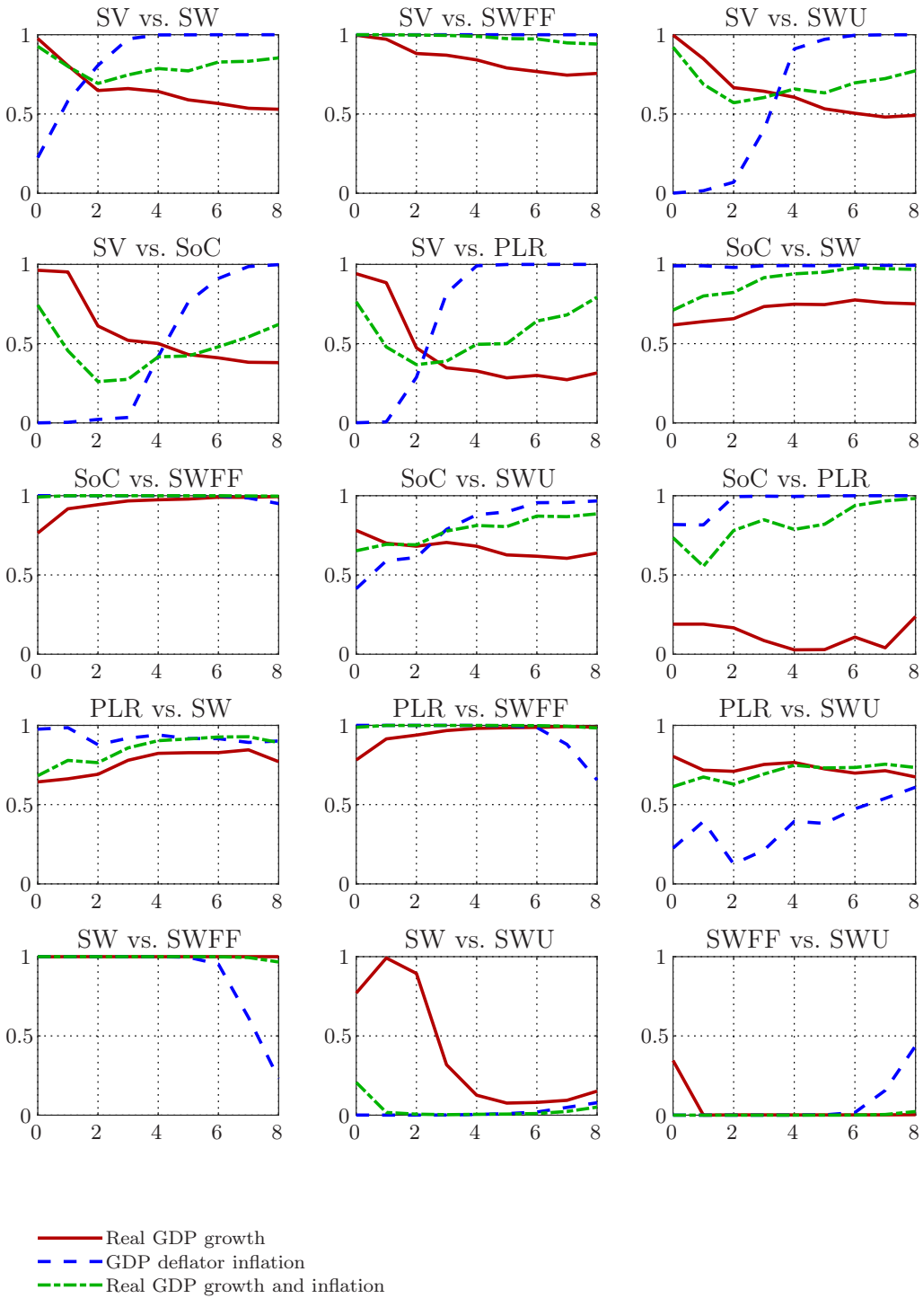
NOTES: The horizontal axis shows the 10 bins while the vertical axis shows the occurence frequency for the estimated π's. If these variables are uniformly distributed for a model, then the occurence in large samples is 0.10 for all bins.

FIGURE I.4: Histograms of the estimated $\pi_{T+h+1|T}$ values for the marginal GDP deflator inflation density forecasts at the nowcast ($h = 0$) and one-quarter-ahead ($h = 1$) horizons for 2001Q1–2019Q4.



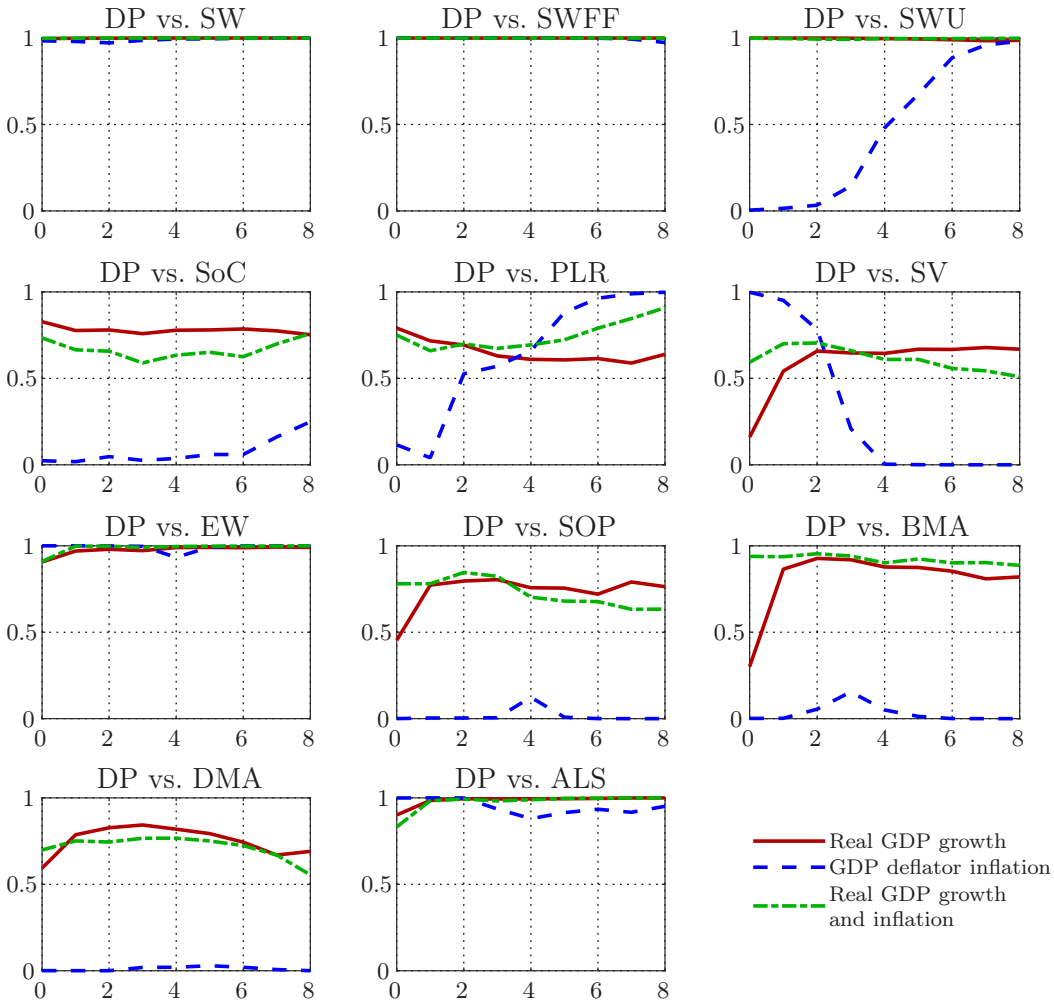
NOTES: See Figure I.3.

FIGURE I.5: Percentile values of the Amisano-Giacomini weighted likelihood ratio tests for the equality of the average log predictive scores of ten BVAR and DSGE model pairs for the sample 2001Q1–2019Q4.



NOTES: The test statistics have been computed with equal weights and using the Bartlett kernel for the HAC estimator (Newey and West, 1987). Following Amisano and Giacomini (2007) we use a short truncation lag, but rather than using their selection of zero lags we use one lag. The percentile value of the test statistic is taken from a standard normal distribution, where large percentile values favor the first model of the test and small percentile values the second model.

FIGURE I.6: Percentile values of the Amisano-Giacomini weighted likelihood ratio tests for the equality of the average log predictive scores of the dynamic prediction pool versus nine alternative forecast schemes for the sample 2001Q1–2019Q4.



NOTES: The density forecast combination methods are given by dynamic prediction pool (DP), equal weights (EW), static optimal prediction pool (SOP), Bayesian model averaging (BMA), dynamic model averaging (DMA) and average log score (ALS). The DP, SOP, BMA, DMA and ALS combination methods are based on an information lag of four quarters. See Figures I.5 for further details.

FIGURE I.7: Recursive estimates of the average log predictive scores of the marginal density forecasts of real GDP growth and GDP deflator inflation and in deviation from the recursive estimates of the average log scores from the dynamic prediction pool covering the vintages 2001Q1–2019Q4.

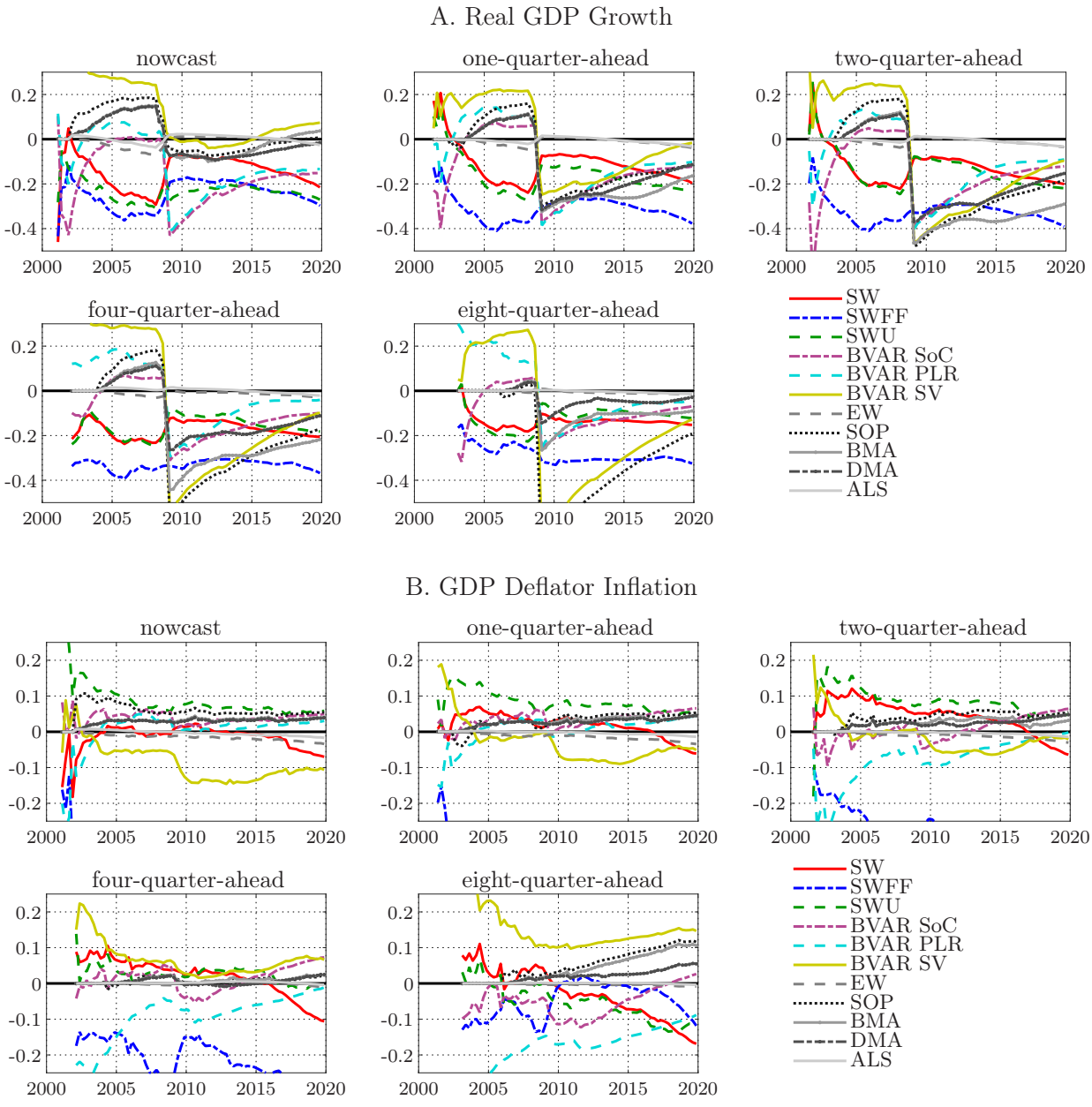


FIGURE I.8: Posterior estimates of the model weights for ALS of the joint density forecasts of real GDP growth and GDP deflator inflation covering the vintages 2001Q1–2019Q4.

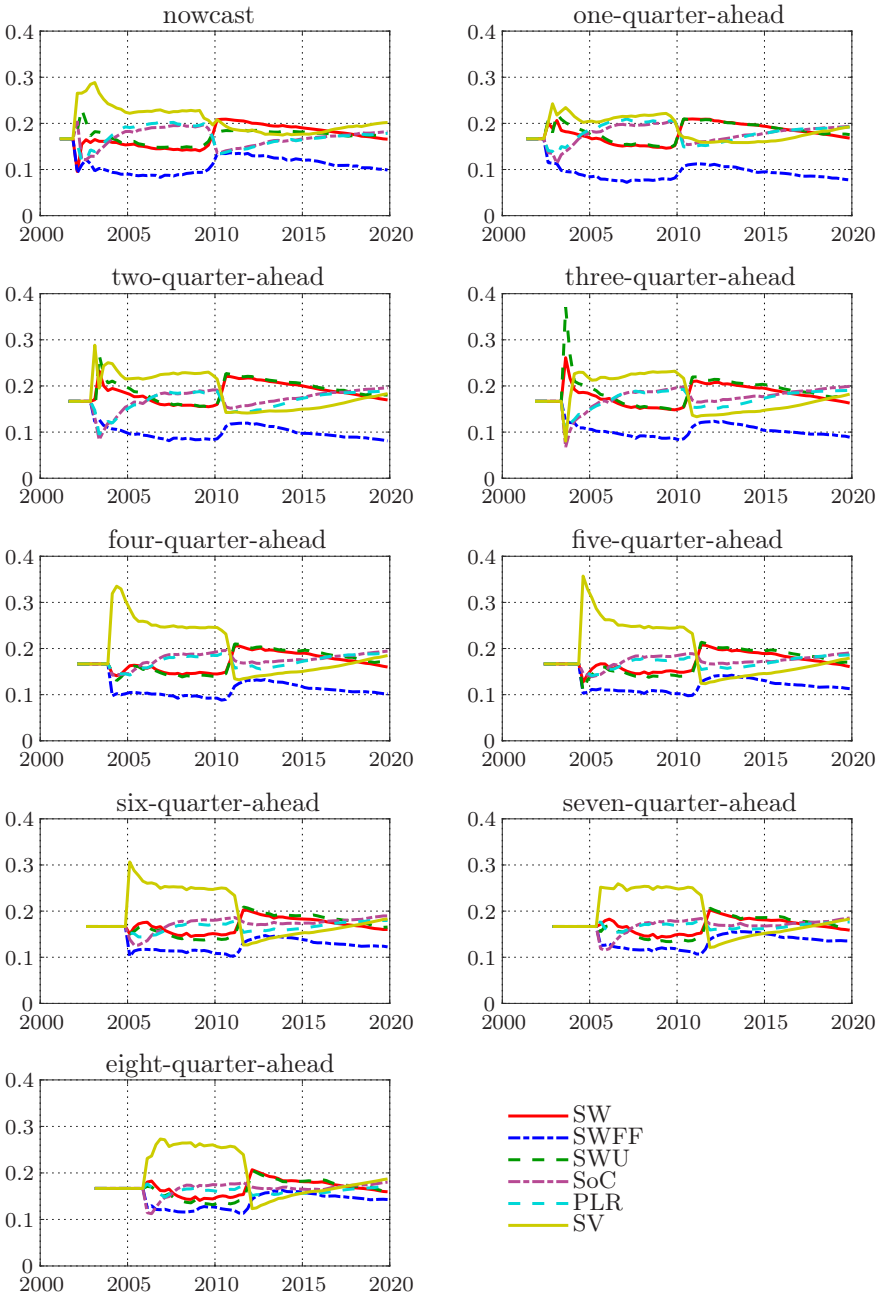


FIGURE I.9: Posterior estimates of the model weights for ALS of the marginal density forecasts of real GDP growth covering the vintages 2001Q1–2019Q4.

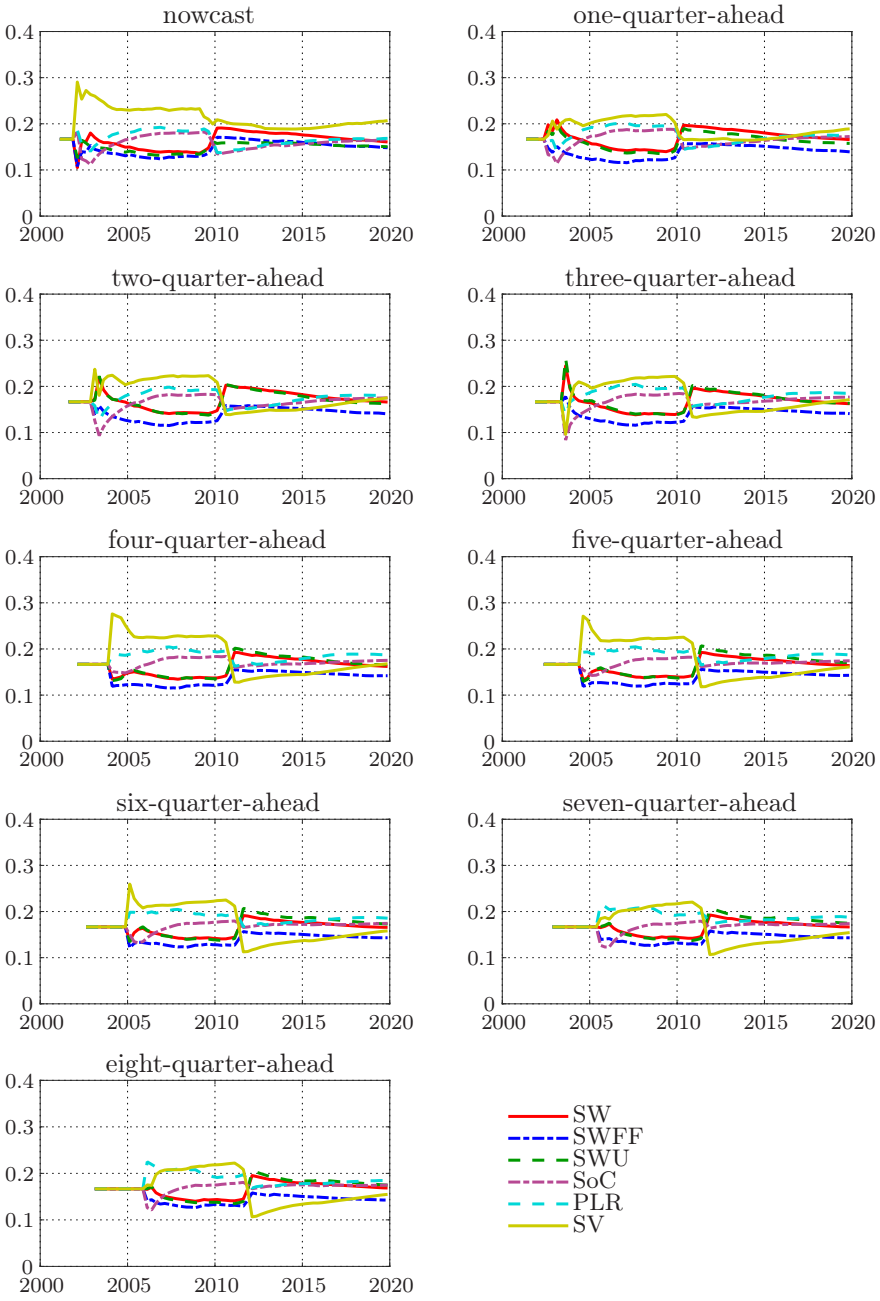


FIGURE I.10: Posterior estimates of the model weights for ALS of the marginal density forecasts of GDP deflator inflation covering the vintages 2001Q1–2019Q4.

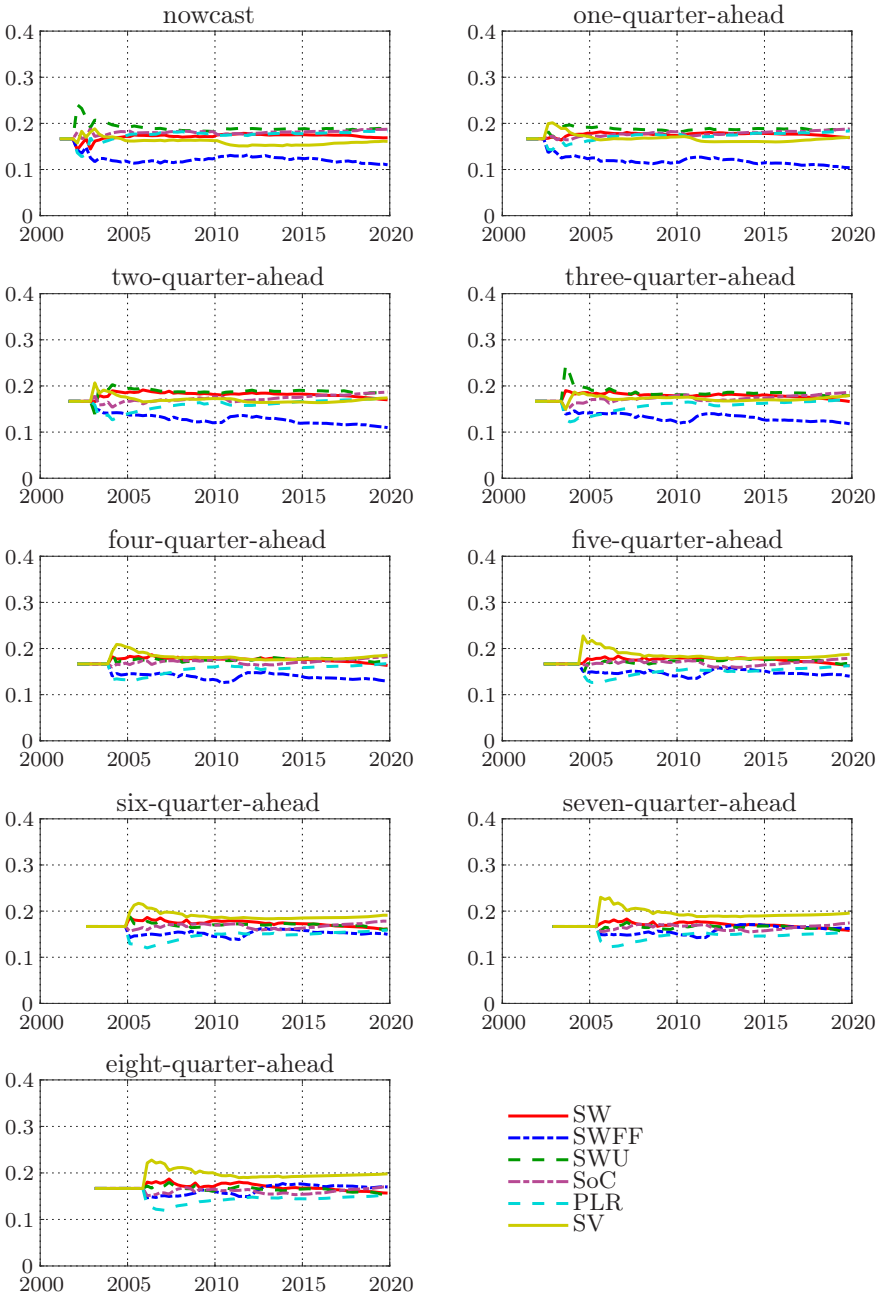


FIGURE I.11: Posterior estimates of the model weights for BMA of the joint density forecasts of real GDP growth and GDP deflator inflation covering the vintages 2001Q1–2019Q4.

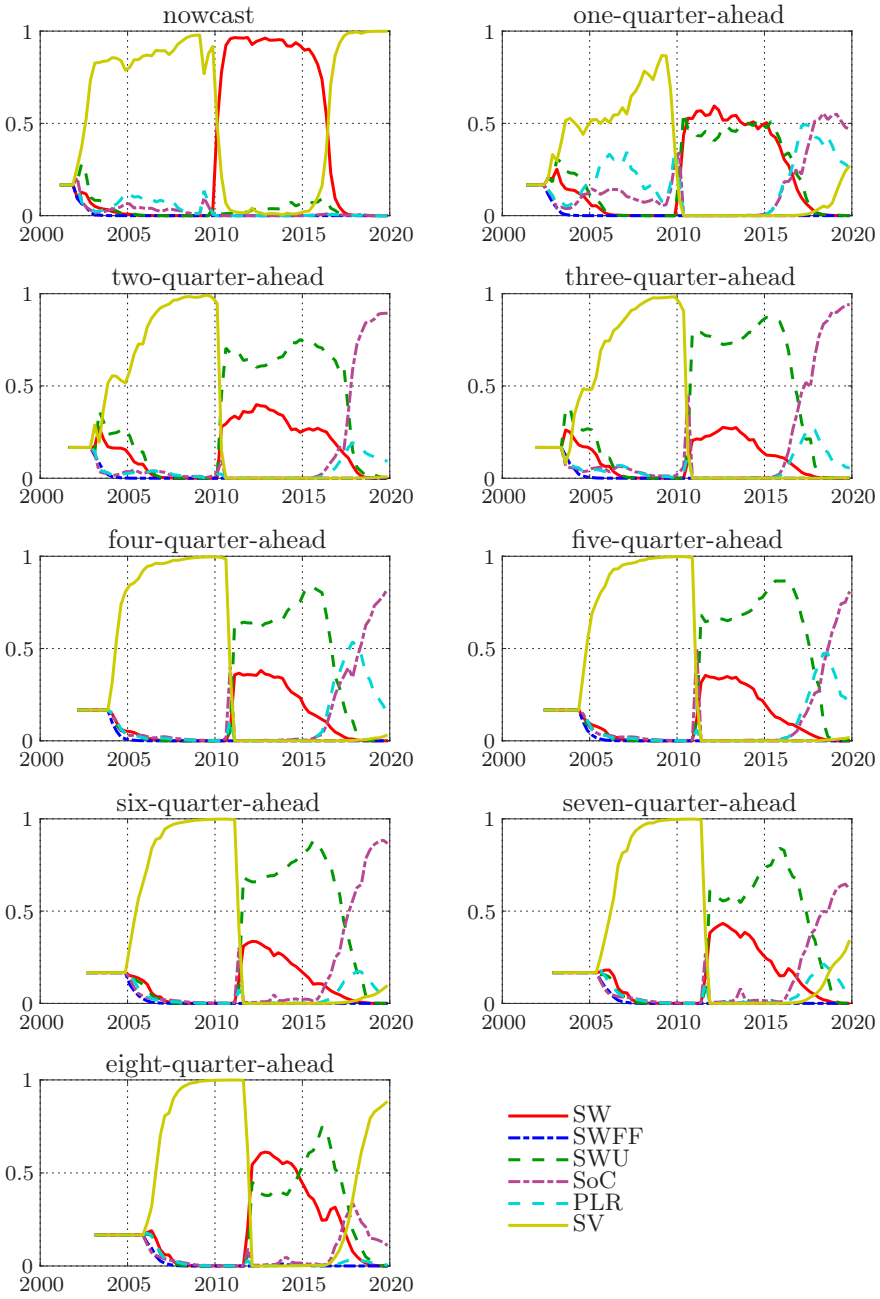


FIGURE I.12: Posterior estimates of the model weights for BMA of the marginal density forecasts of real GDP growth covering the vintages 2001Q1–2019Q4.

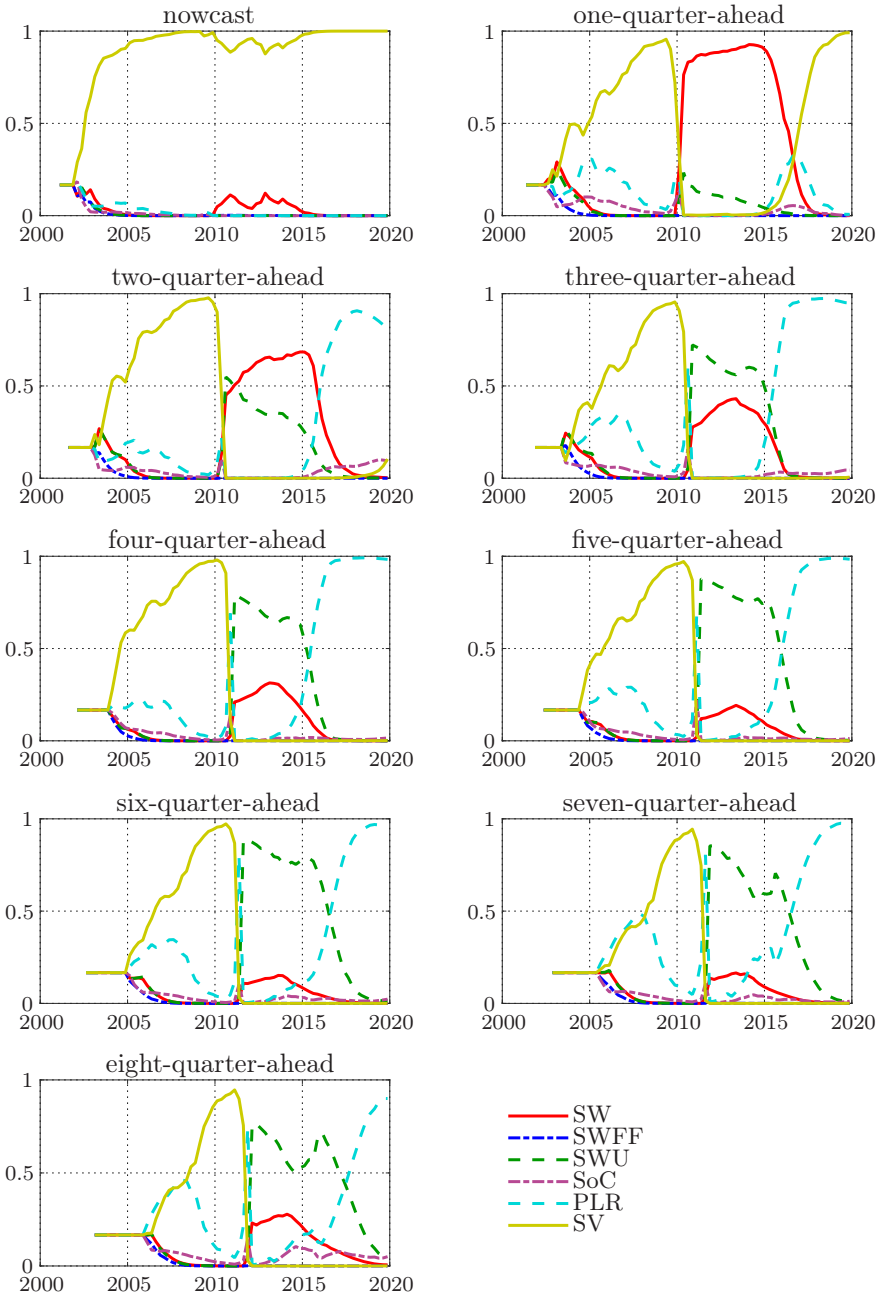


FIGURE I.13: Posterior estimates of the model weights for BMA of the marginal density forecasts of GDP deflator inflation covering the vintages 2001Q1–2019Q4.

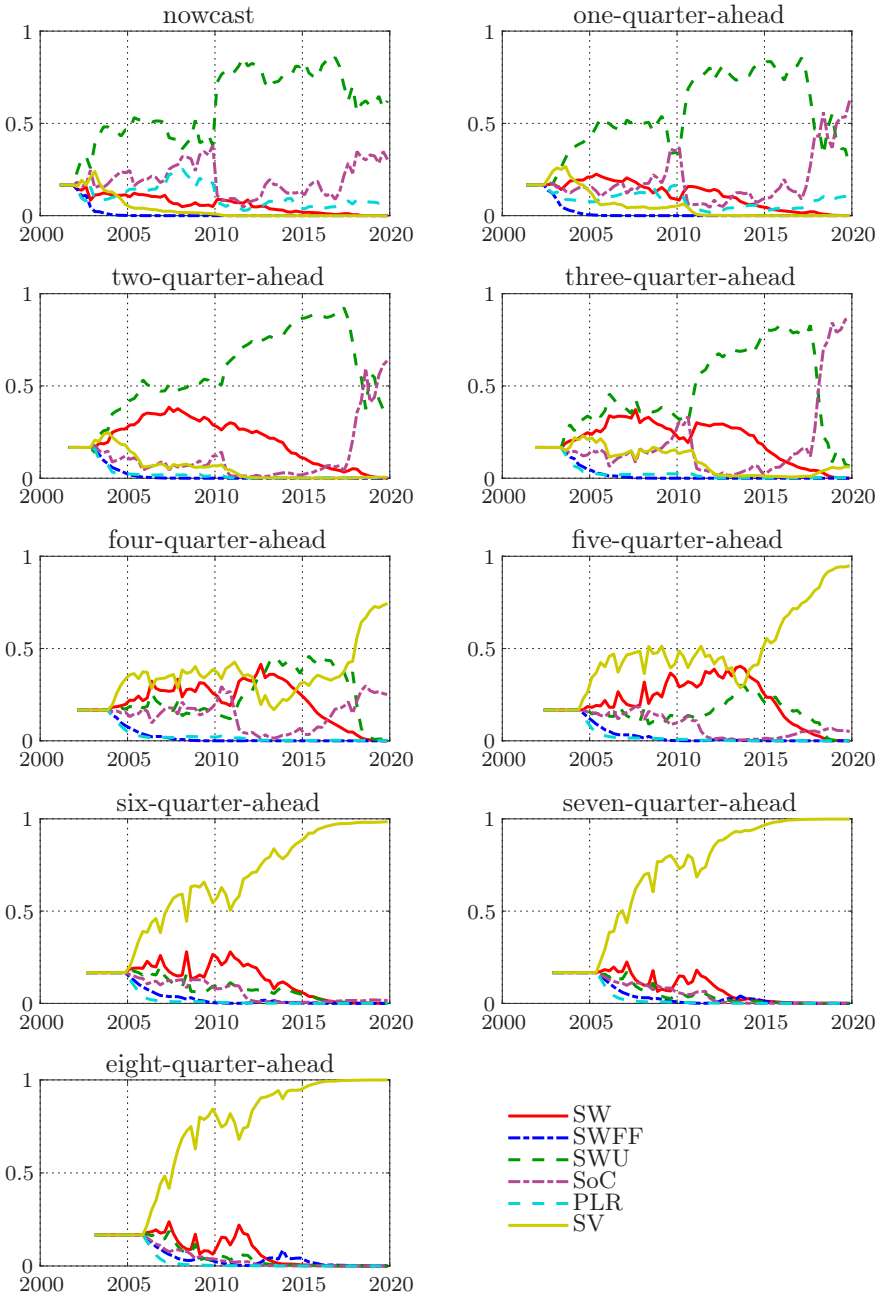


FIGURE I.14: Posterior estimates of the model weights for DMA of the joint density forecasts of real GDP growth and GDP deflator inflation covering the vintages 2001Q1–2019Q4.

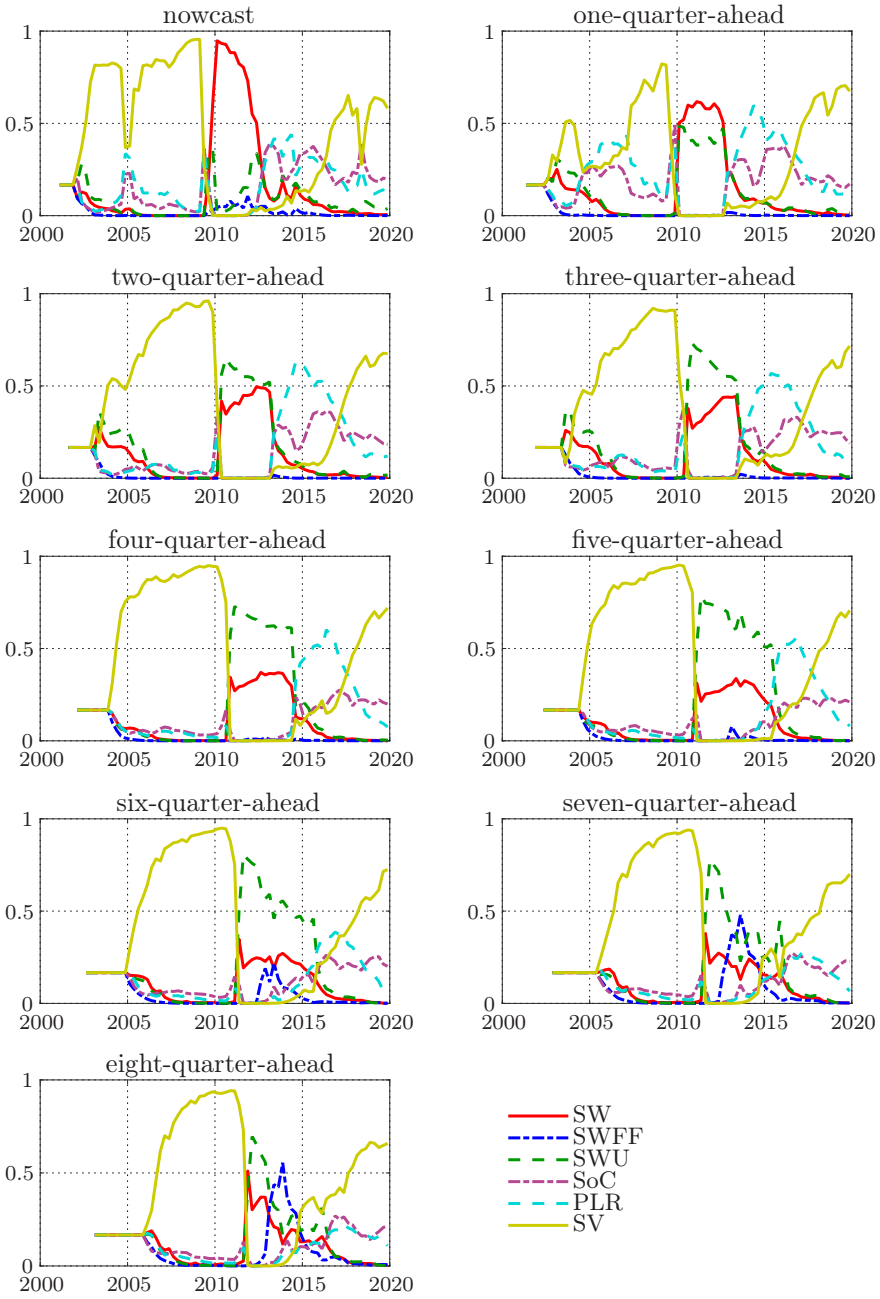


FIGURE I.15: Posterior estimates of the model weights for DMA of the marginal density forecasts of real GDP growth covering the vintages 2001Q1–2019Q4.

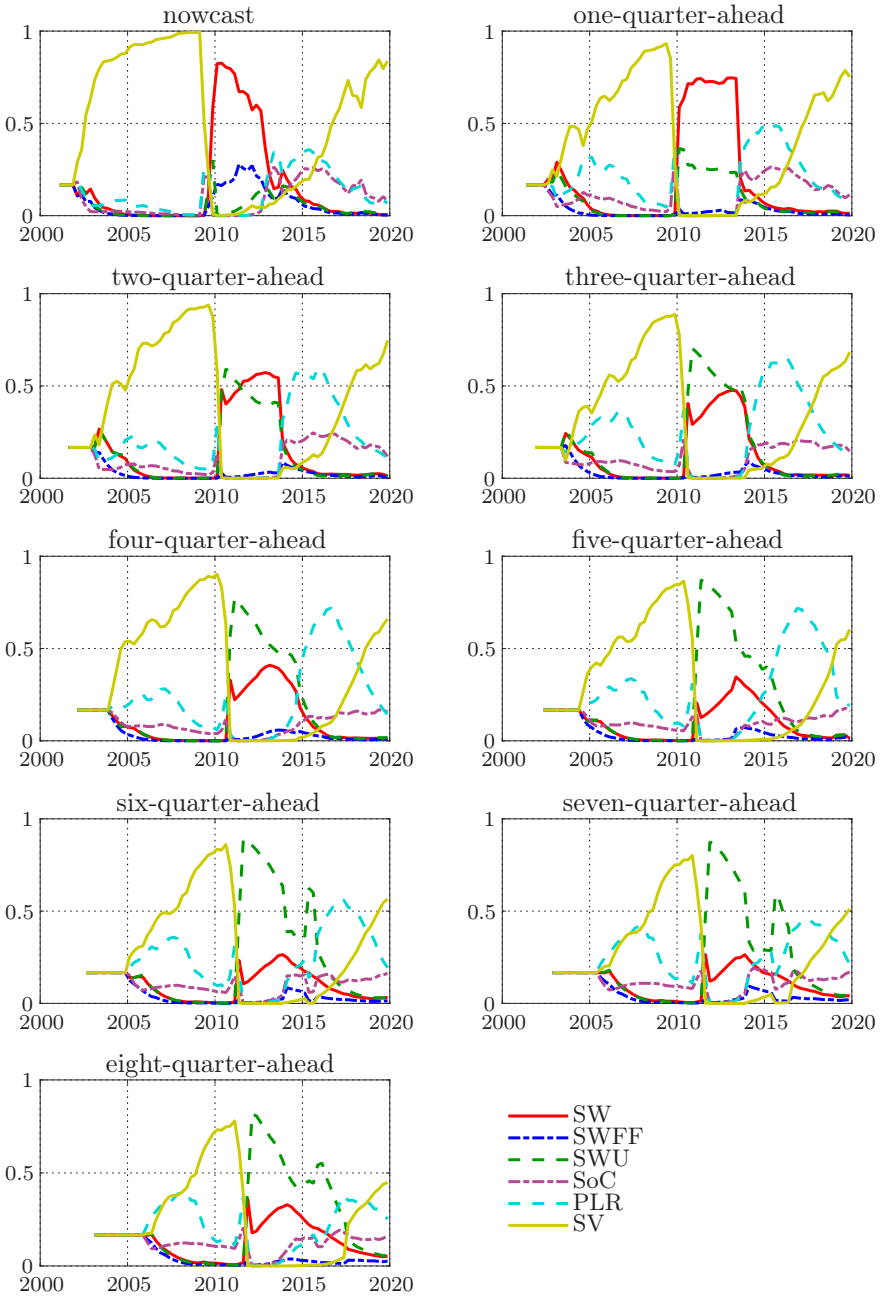


FIGURE I.16: Posterior estimates of the model weights for DMA of the marginal density forecasts of GDP deflator inflation covering the vintages 2001Q1–2019Q4.

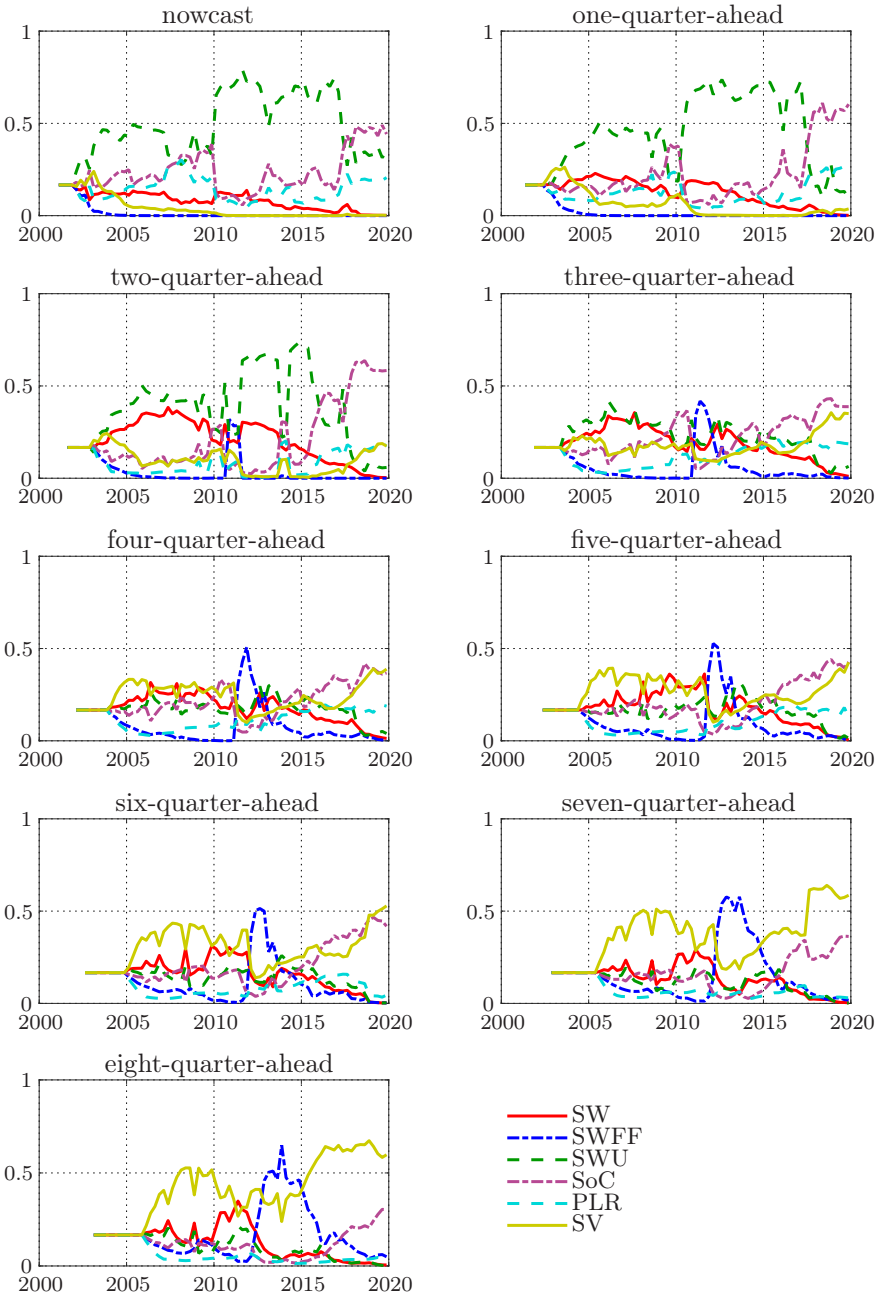


FIGURE I.17: Posterior estimates of the model weights for the DP of the joint density forecasts of real GDP growth and GDP deflator inflation covering the vintages 2001Q1–2019Q4.

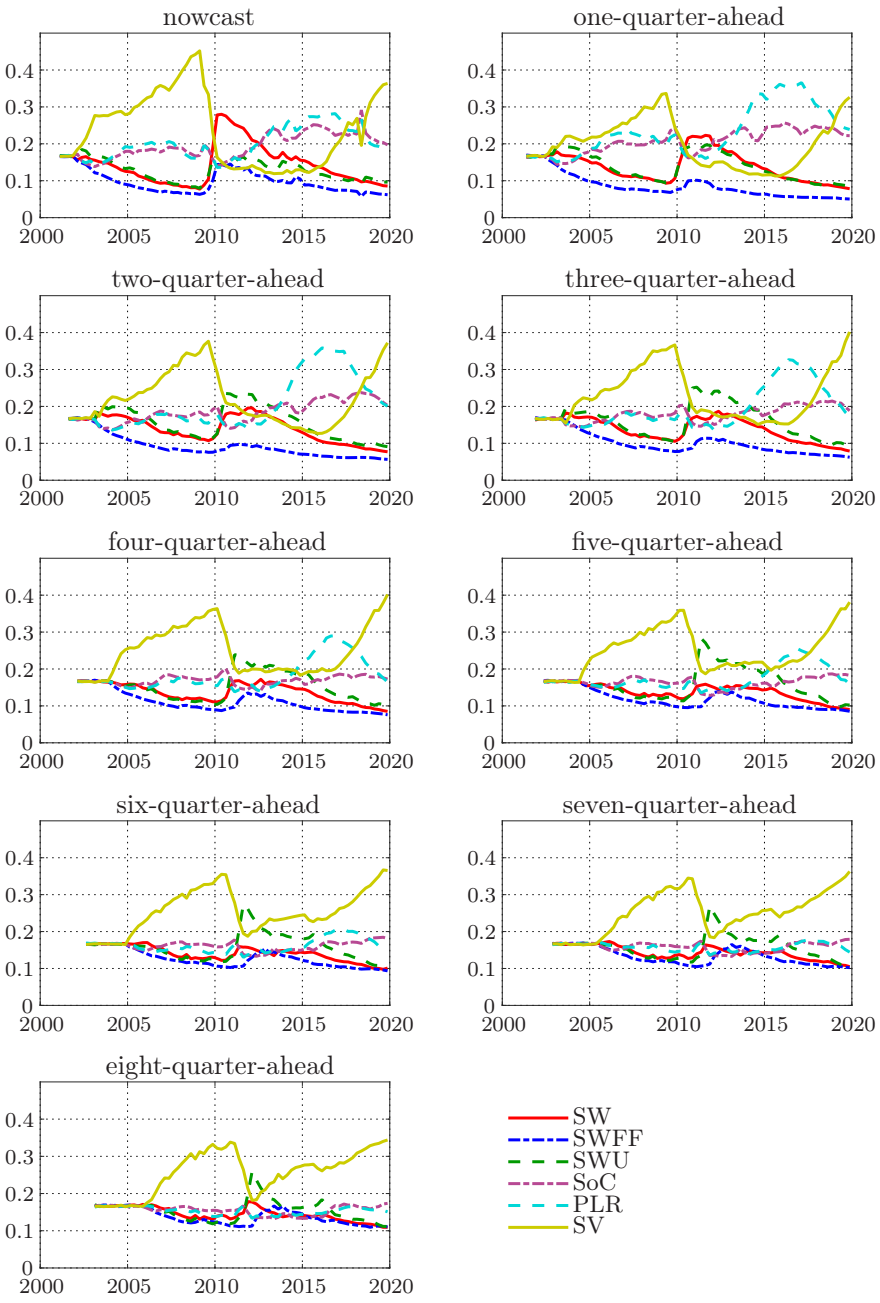


FIGURE I.18: Posterior estimates of the model weights for the DP of the marginal density forecasts of real GDP growth covering the vintages 2001Q1–2019Q4.

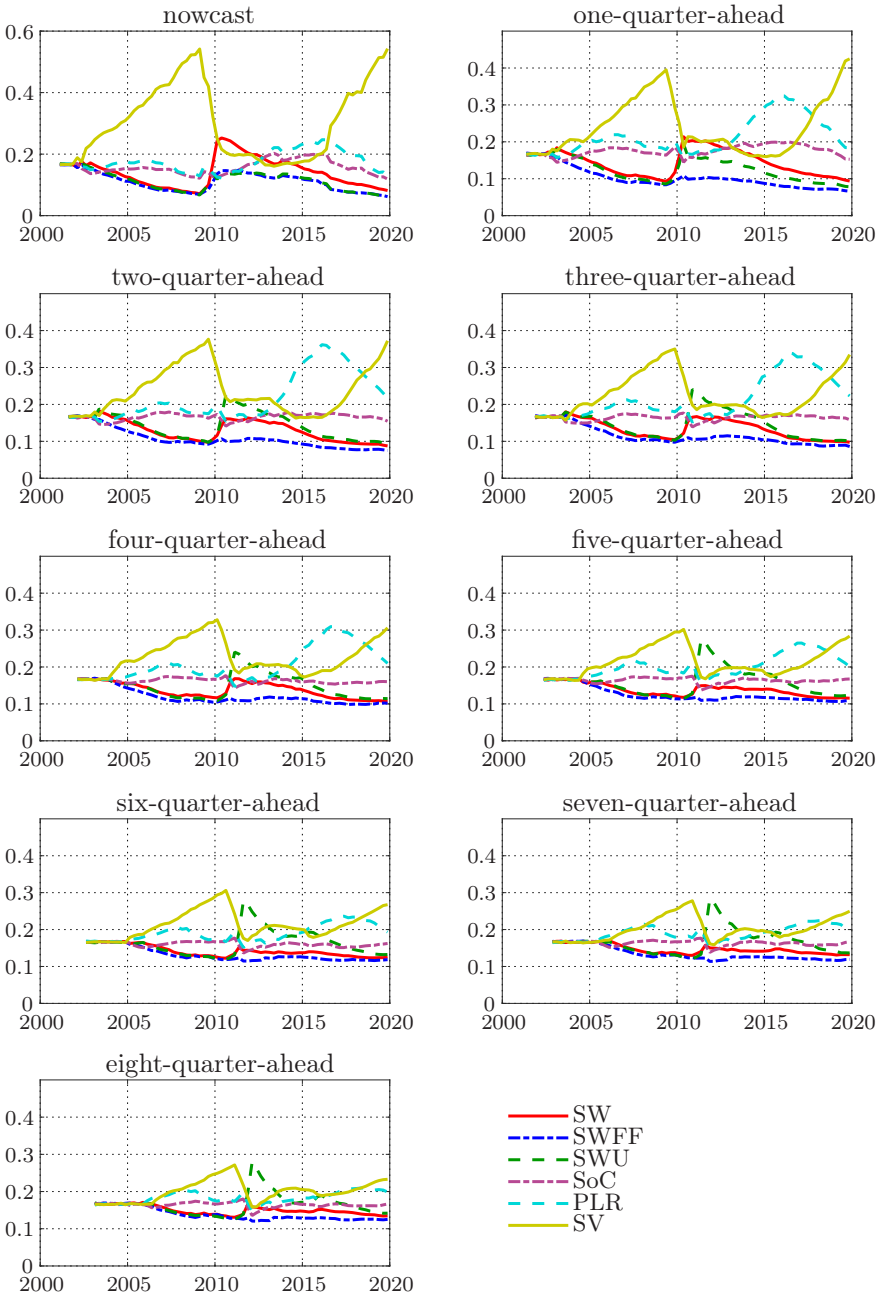


FIGURE I.19: Posterior estimates of the model weights for the DP of the marginal density forecasts of GDP deflator inflation covering the vintages 2001Q1–2019Q4.

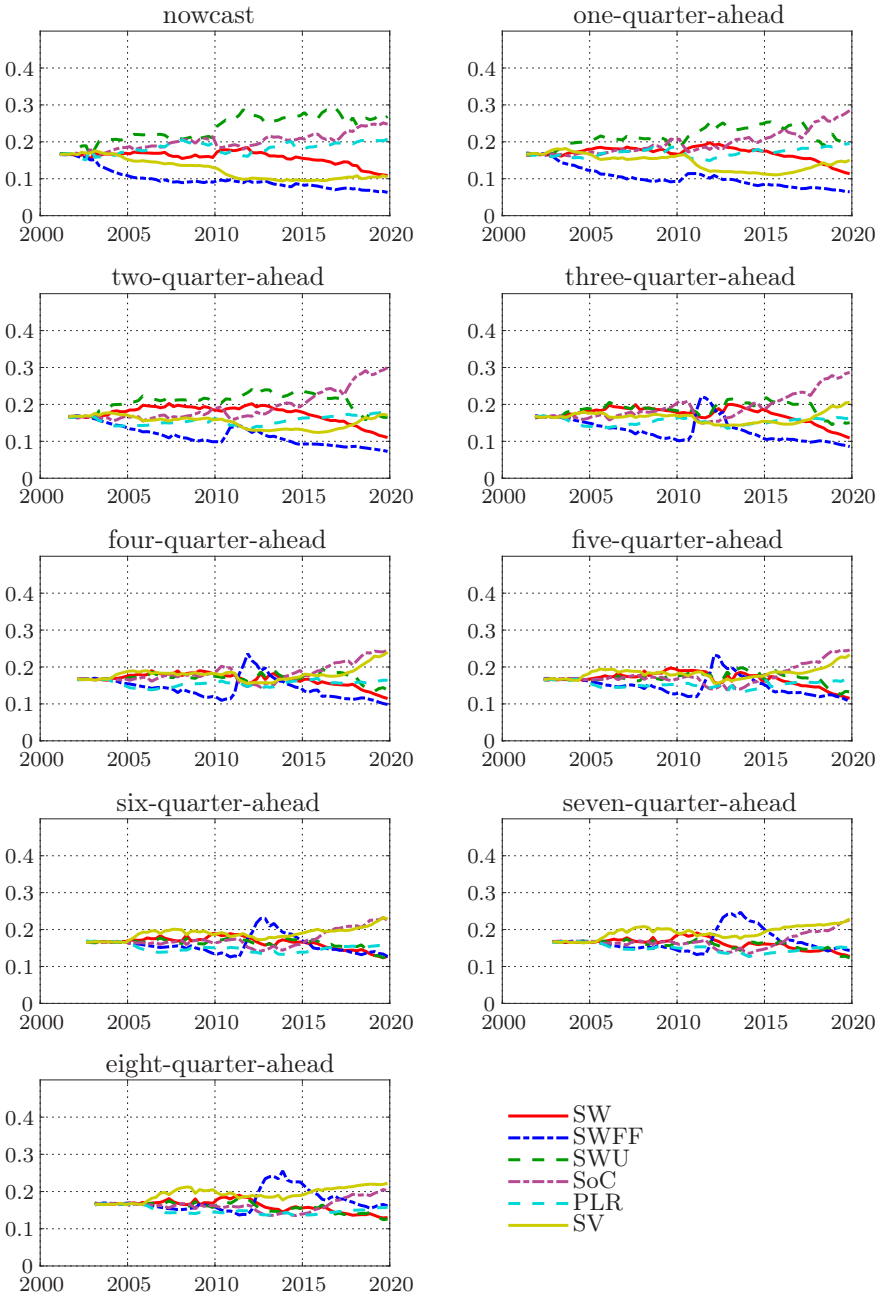


FIGURE I.20: Posterior estimates of the model weights for the SOP of the joint density forecasts of real GDP growth and GDP deflator inflation covering the vintages 2001Q1–2019Q4.

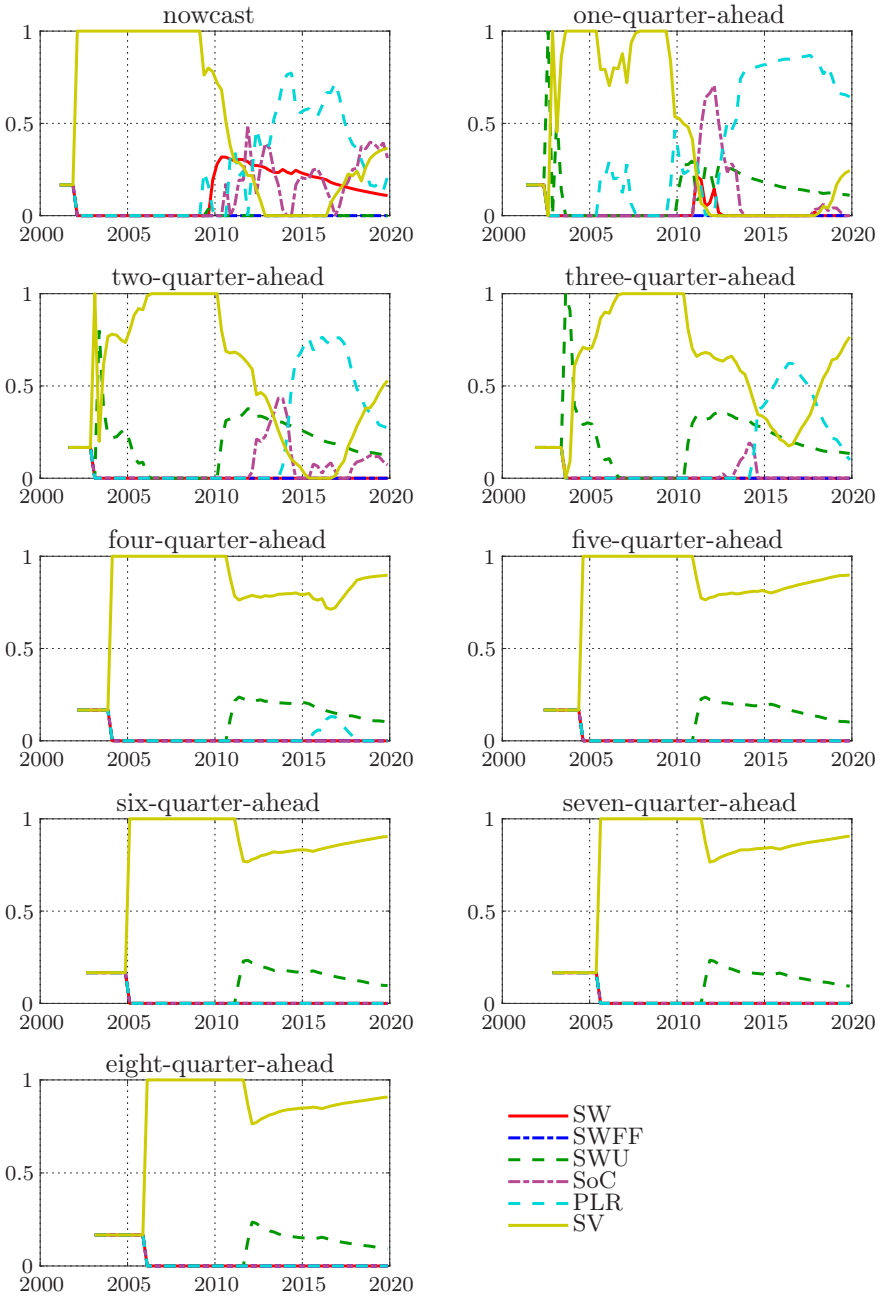


FIGURE I.21: Posterior estimates of the model weights for the SOP of the marginal density forecasts of real GDP growth covering the vintages 2001Q1–2019Q4.

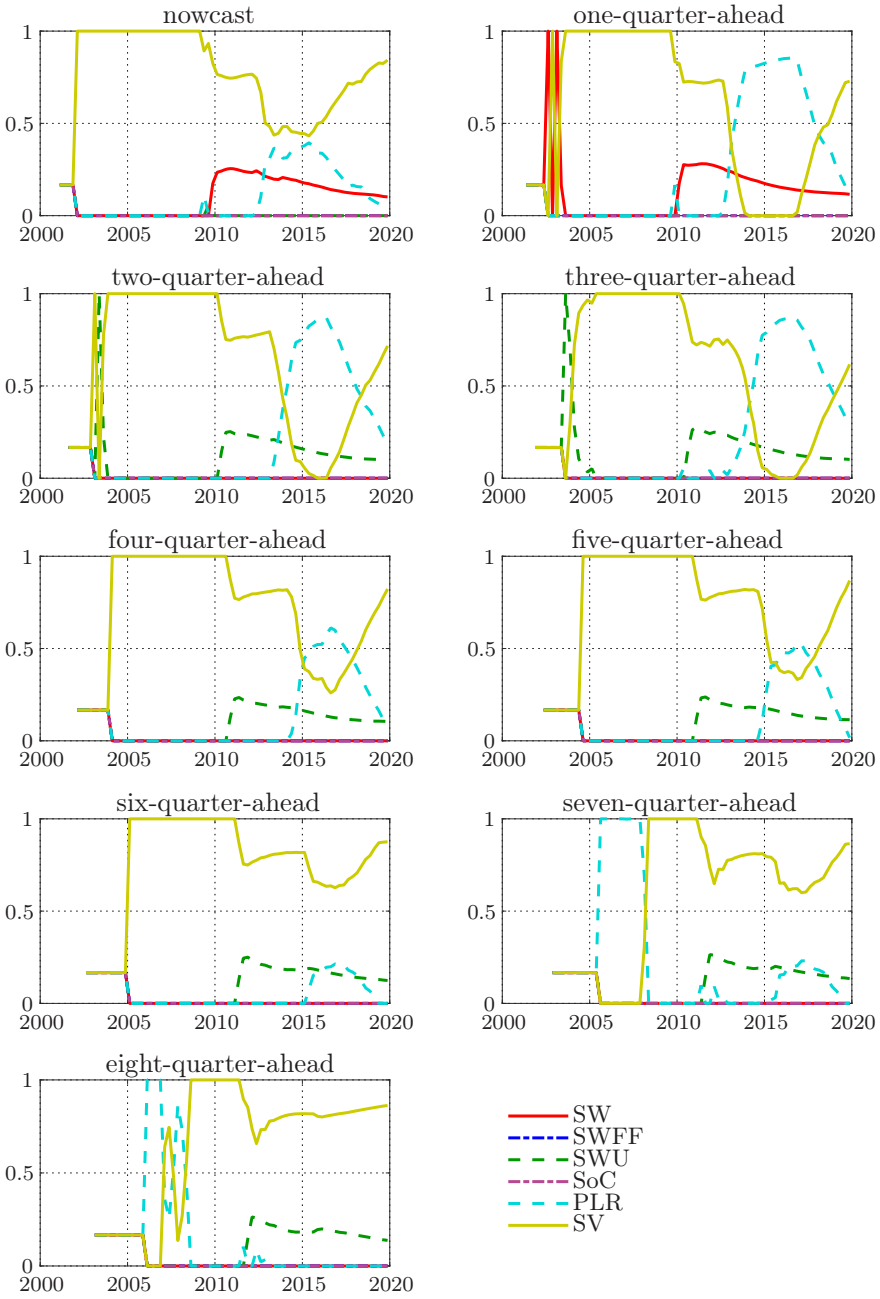


FIGURE I.22: Posterior estimates of the model weights for the SOP of the marginal density forecasts of GDP deflator inflation covering the vintages 2001Q1–2019Q4.

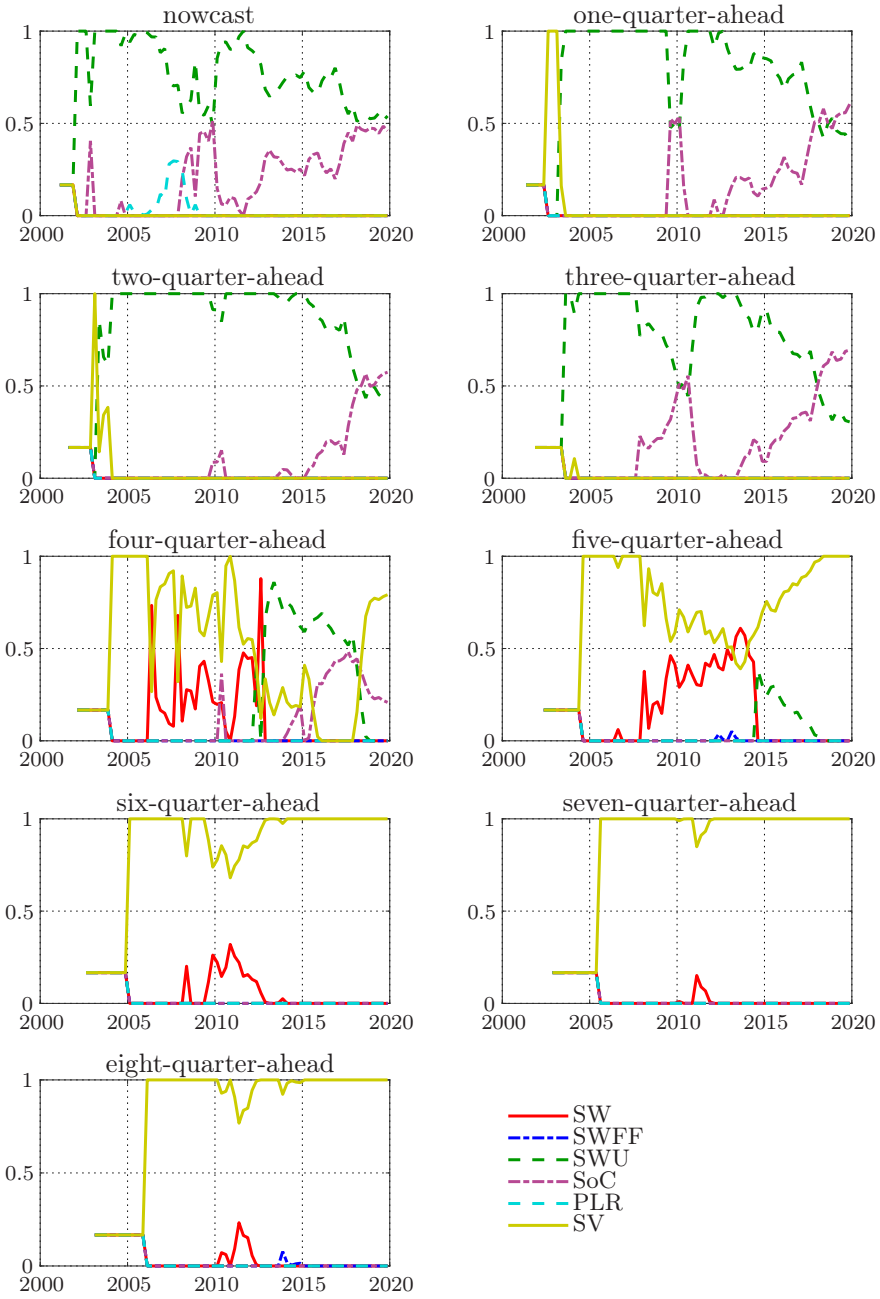


FIGURE I.23: Recursive estimates of the differences of log predictive scores of marginal real GDP growth and GDP deflator inflation density forecasts with information lag 1 and 4 covering the vintages 2001Q1–2019Q4.

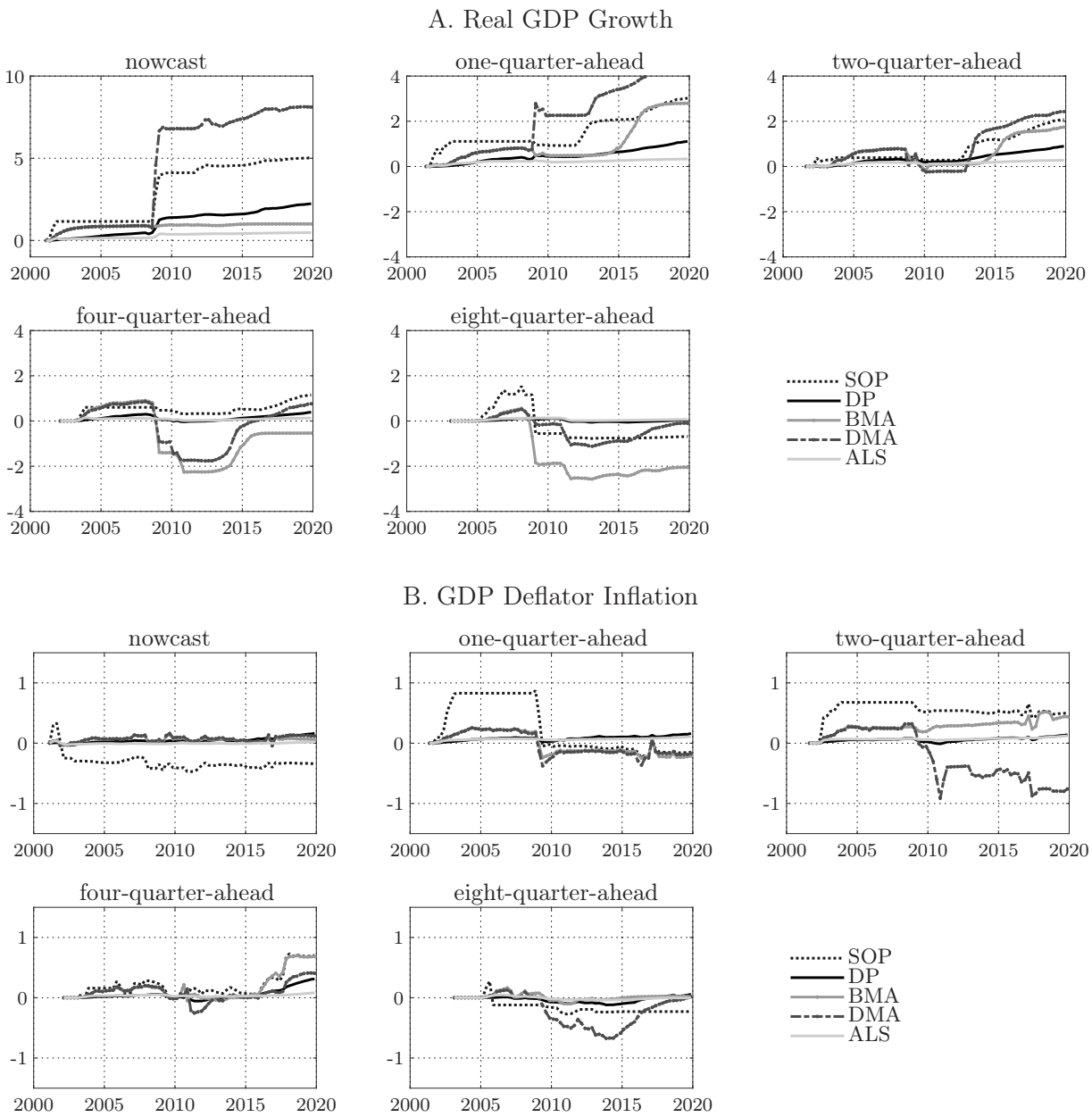


FIGURE I.24: Recursive estimates of the differences of log predictive scores of joint real GDP growth and GDP deflator inflation density forecasts with the SWFF zero initialization and the equal weights initialization covering the vintages 2001Q1–2019Q4.

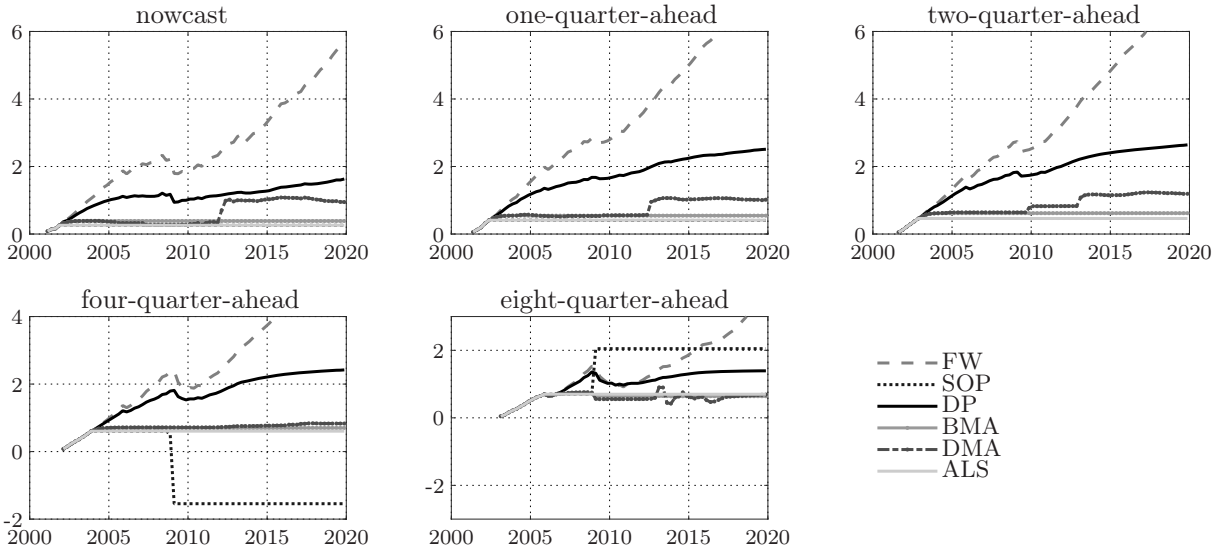


FIGURE I.25: Recursive estimates of the differences of log predictive scores of marginal real GDP growth and GDP deflator inflation density forecasts with the SWFF zero initialization and the equal weights initialization covering the vintages 2001Q1–2019Q4.

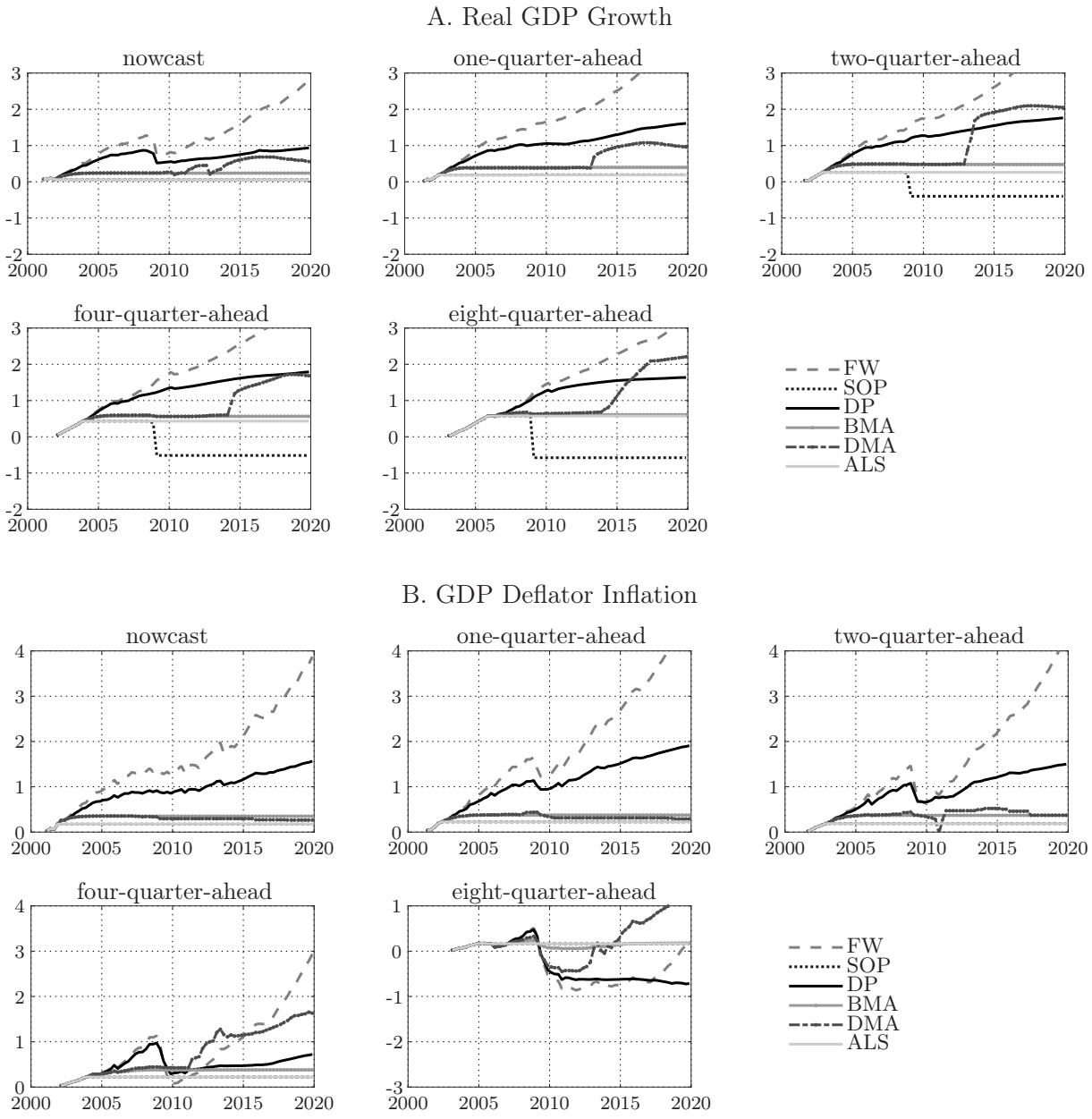


FIGURE I.26: Posterior estimates of the model weights for the dynamic prediction pool of joint real GDP growth and GDP deflator inflation based on the SWFF zero initialization and covering the vintages 2001Q1–2019Q4.

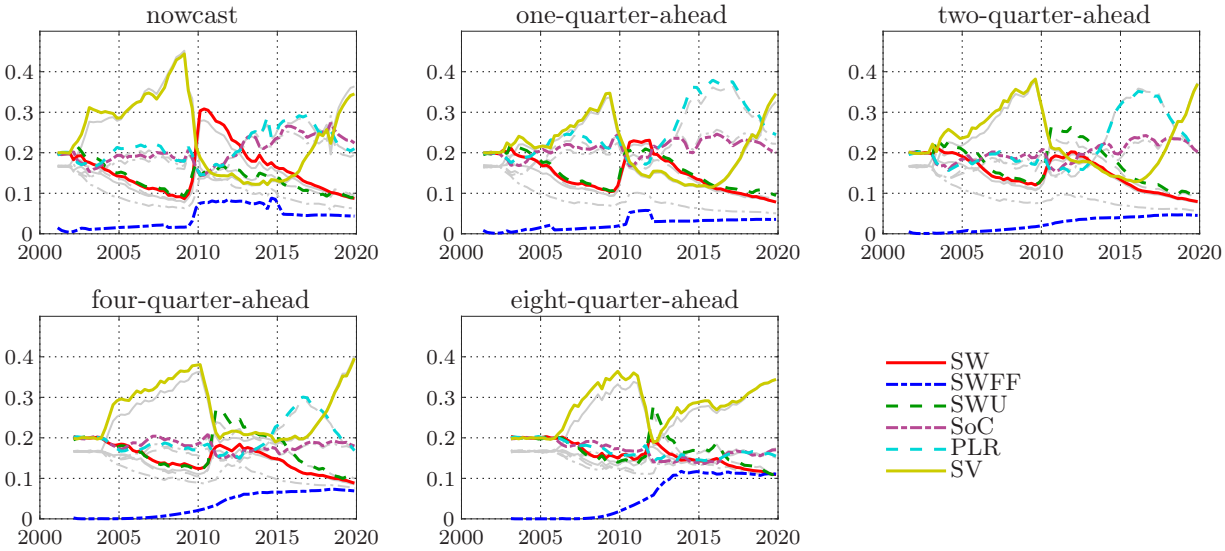
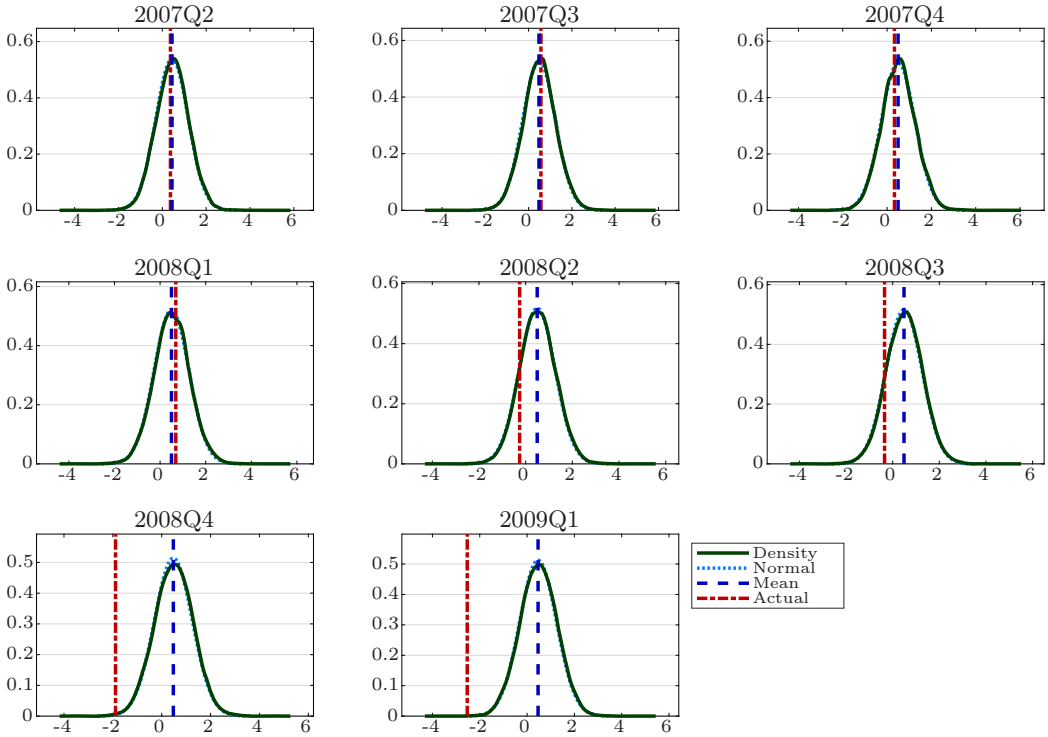


FIGURE I.27: Posterior predictive densities of real GDP growth in 2007Q2–2009Q1 from the SW, SWFF and SWU models based on the real-time database vintage from 2007Q1 along with the normal approximation using the posterior predictive mean and variance, the posterior mean and the actual value.

A. SW Model



B. SWFF Model

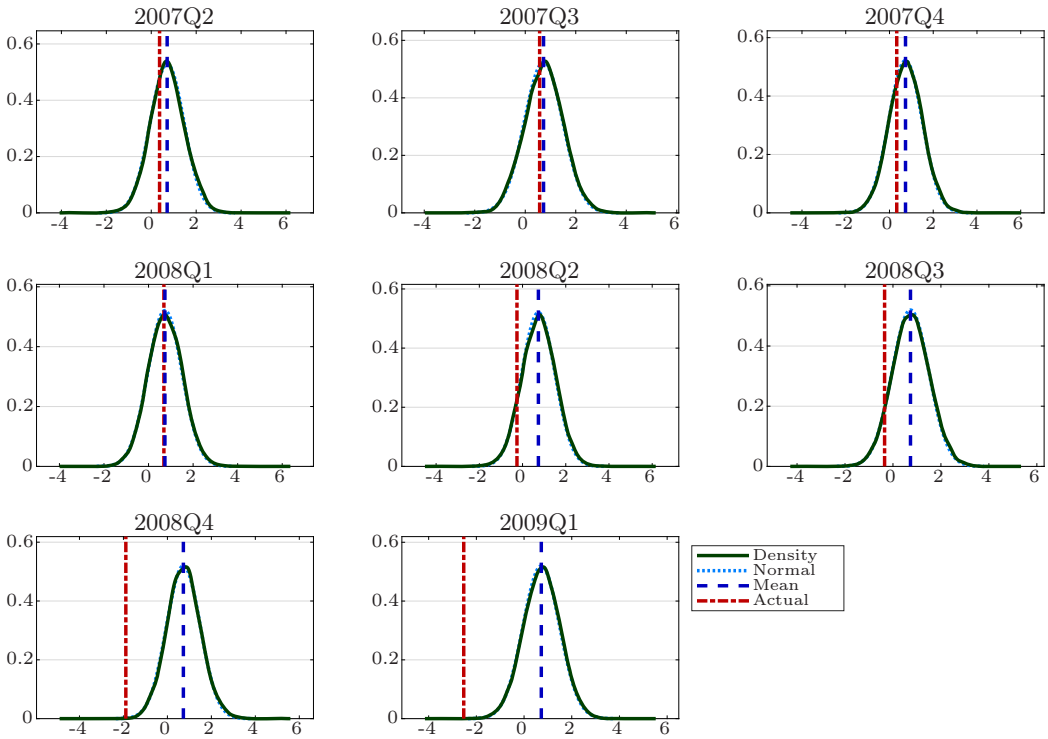


FIGURE I.27: Posterior predictive densities of real GDP growth in 2007Q2–2009Q1 from the SW, SWFF and SWU models based on the real-time database vintage from 2007Q1 along with the normal approximation using the posterior predictive mean and variance, the posterior mean and the actual value (Continued).

C. SWU Model

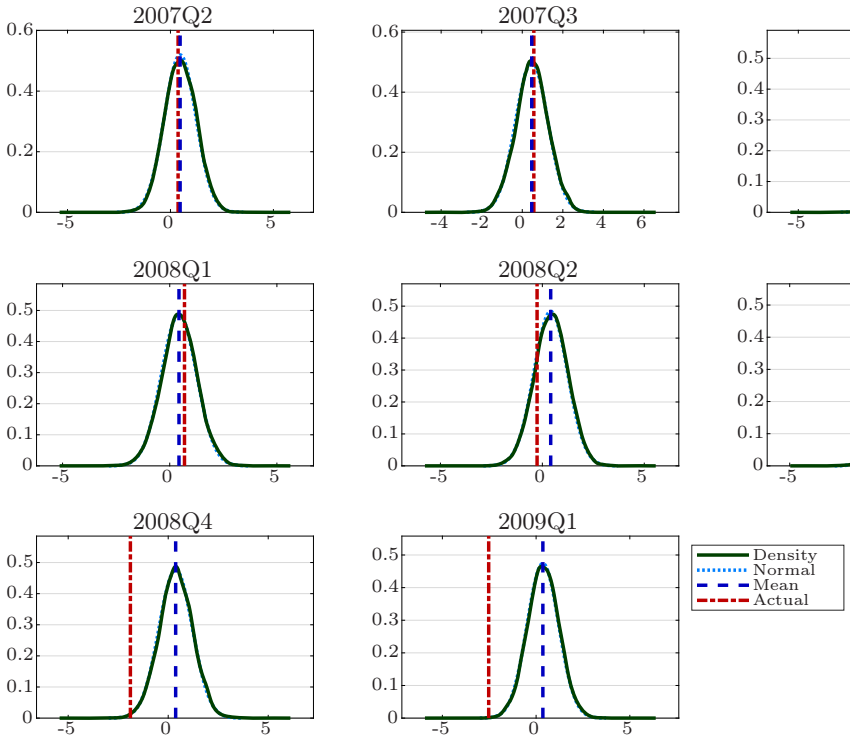


FIGURE I.28: Recursive estimates of the average log predictive scores of the joint density forecasts of real GDP growth and GDP deflator inflation for the vintages 2001Q1–2019Q4 and using the first release data for the actual values.

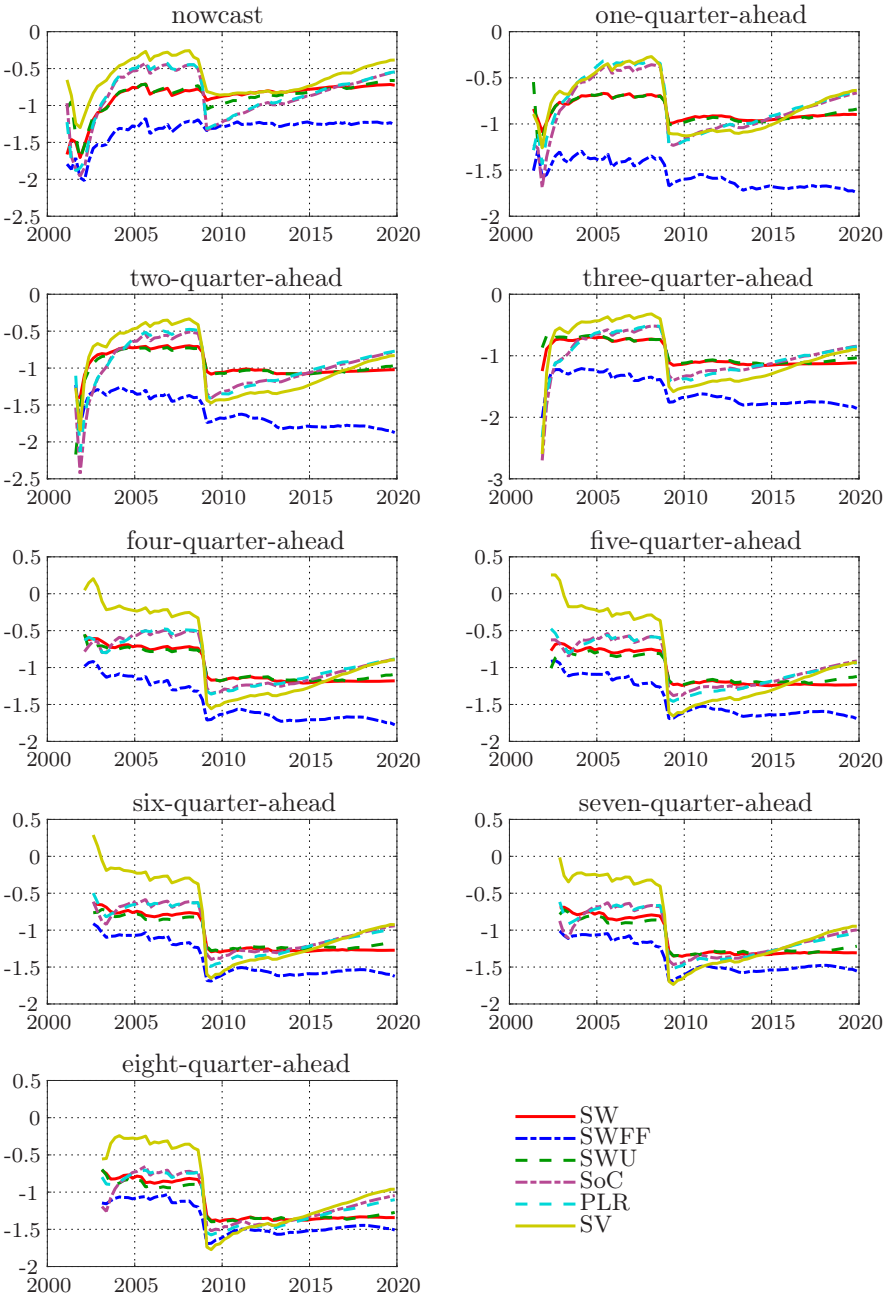


FIGURE I.29: Recursive estimates of the average log predictive scores of the joint density forecasts of real GDP growth and GDP deflator inflation for the vintages 2001Q1–2019Q4 and using the second quarter release data for the actual values.

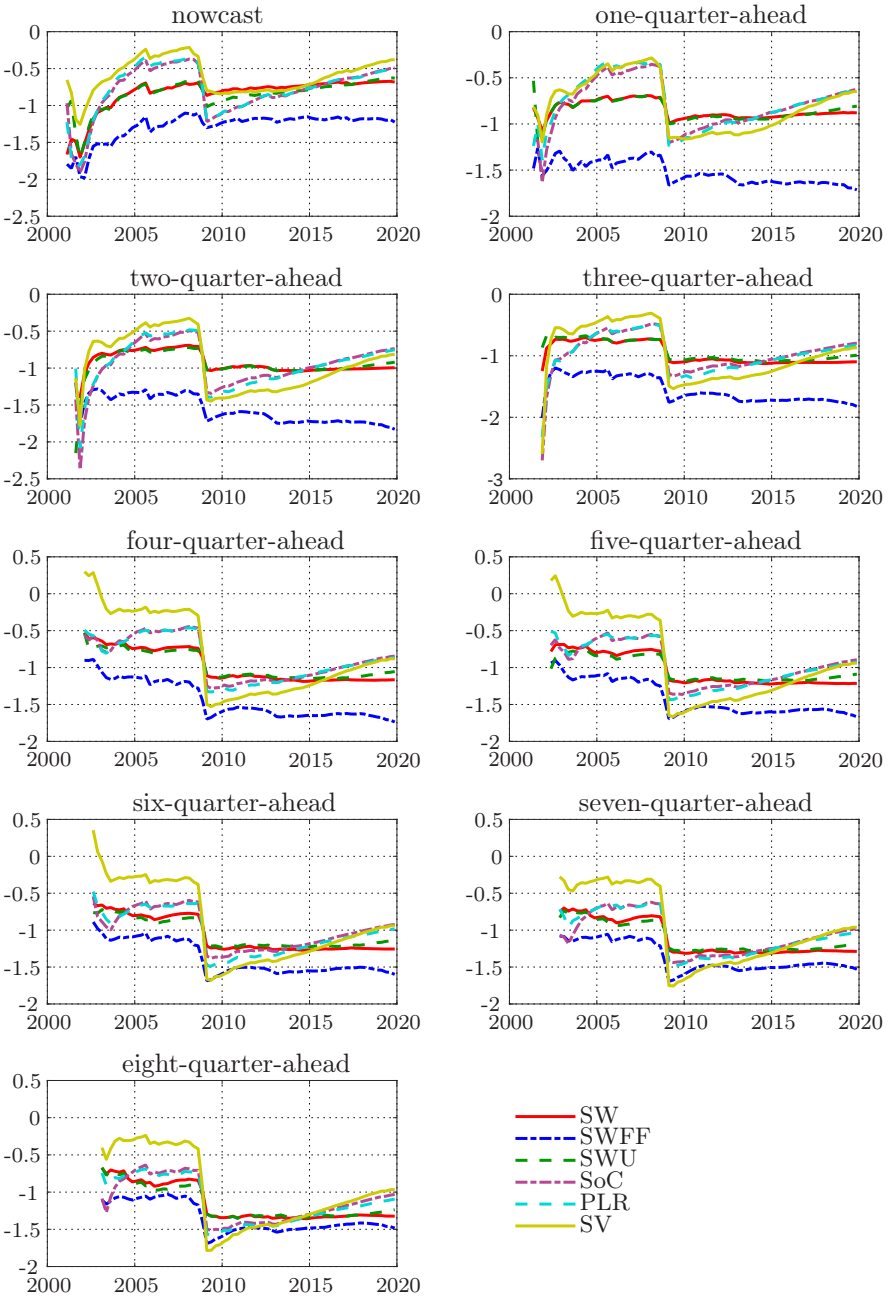


FIGURE I.30: Recursive estimates of the average log predictive scores of the joint density forecasts of real GDP growth and GDP deflator inflation for the vintages 2001Q1–2019Q4 and using the third quarter release data for the actual values.

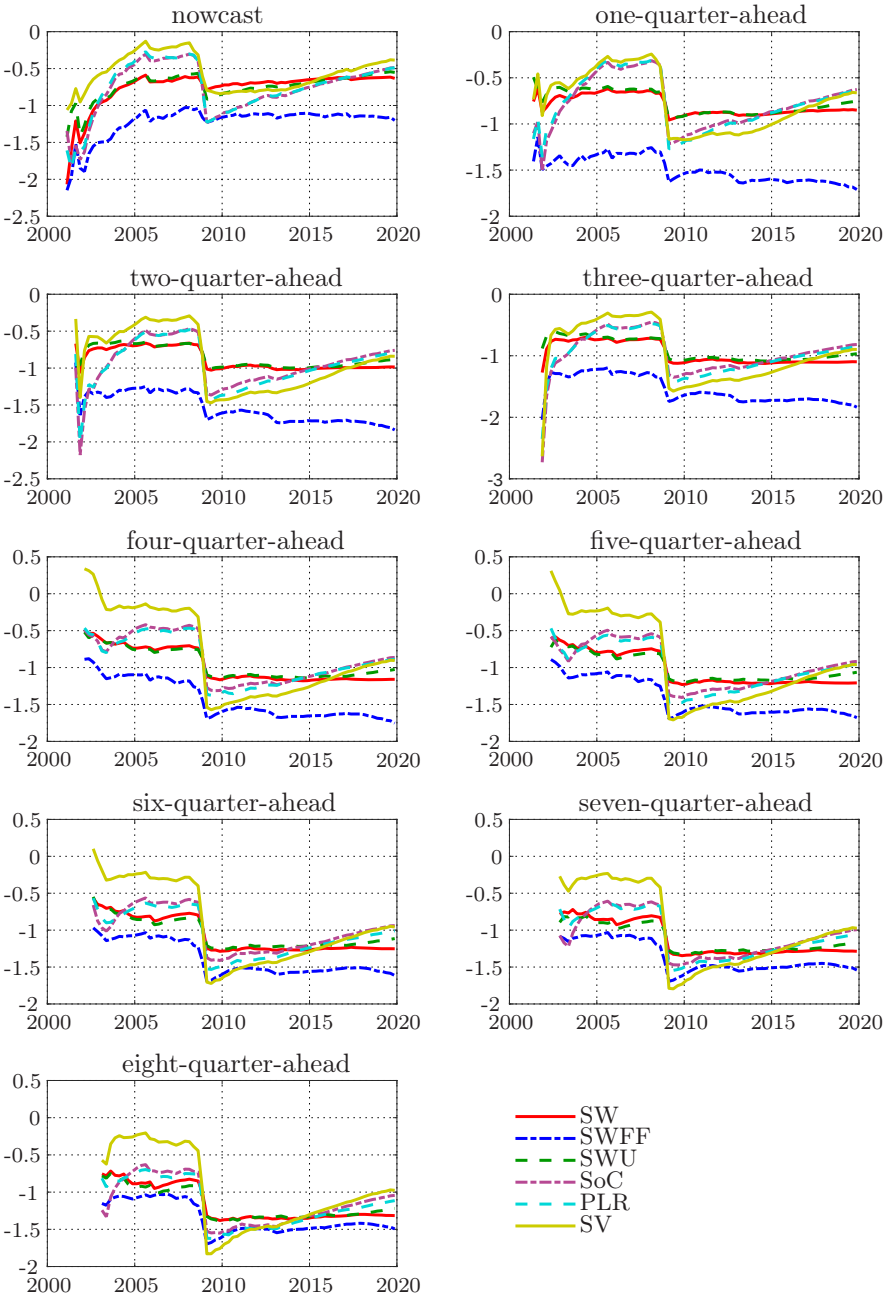
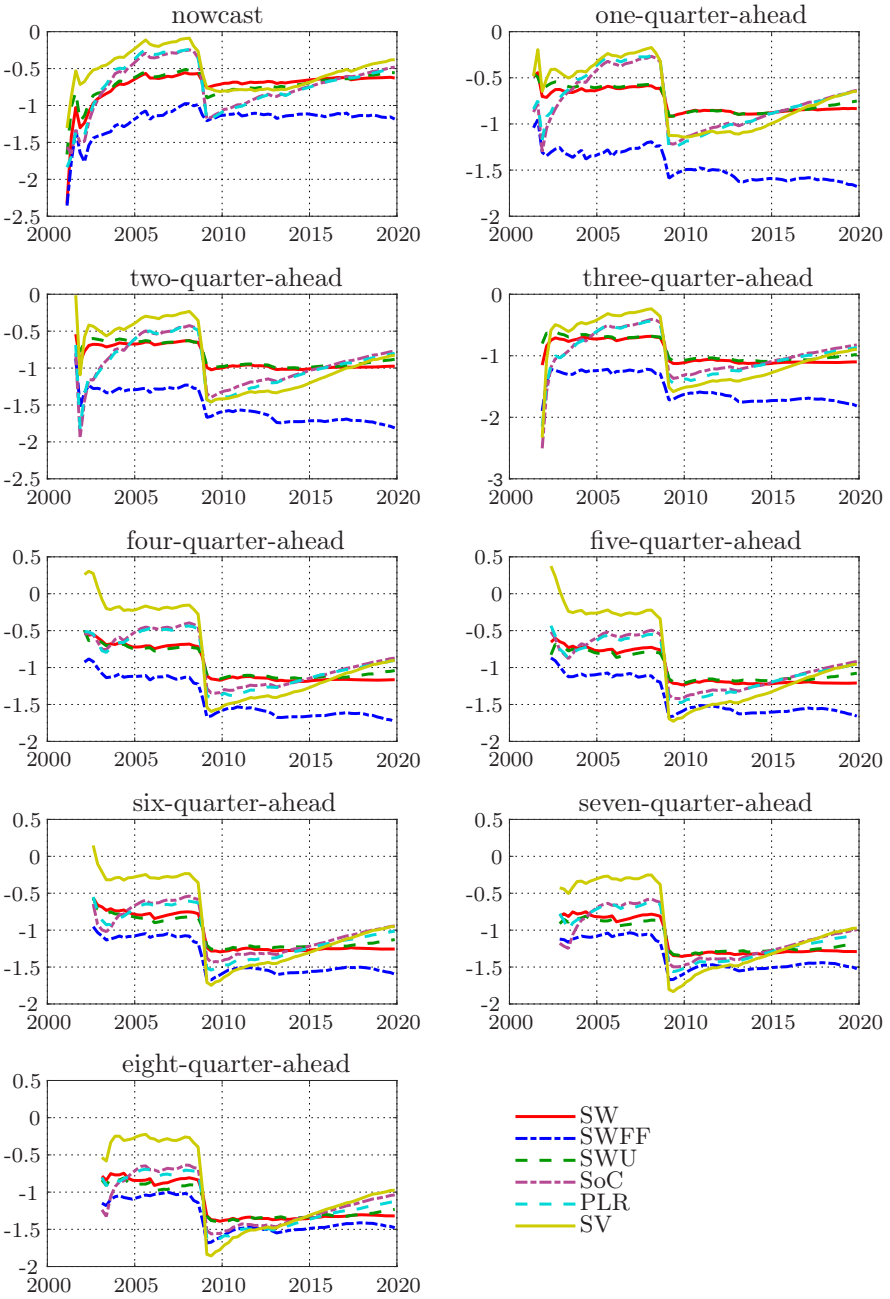


FIGURE I.31: Recursive estimates of the average log predictive scores of the joint density forecasts of real GDP growth and GDP deflator inflation for the vintages 2001Q1–2019Q4 and using the annual revision data for the actual values.



References

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.

Bańbura, M., Giannone, D., and Lenza, M. (2015), "Conditional Forecasts and Scenario Analysis with Vector Autoregressions for Large Cross-Sections," *International Journal of Forecasting*, 31(3), 739–756.

Bańbura, M., Giannone, D., and Reichlin, L. (2010), "Large Bayesian Vector Auto Regressions," *Journal of Applied Econometrics*, 25, 71–92.

Berkowitz, J. (2001), "Testing Density Forecasts, with Applications to Risk Management," *Journal of Business & Economic Statistics*, 19(4), 465–474.

Box, G. E. P. (1980), "Sampling and Bayes' Inference in Scientific Modelling and Robustness," *Journal of the Royal Statistical Society Series A*, 143, 383–430.

Carpenter, J., Clifford, P., and Fearnhead, P. (1999), "An Improved Particle Filter for Non-linear Prolems," *IEE Proceedings—Radar, Sonar and Navigation*, 146(1), 2–7.

Chan, J. C. C. and Jeliazkov, I. (2009), "Efficient Simulation and Integrated Likelihood Estimation in State Space Models," *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1–2), 101–120.

Chopin, N. (2004), "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *The Annals of Statistics*, 32(6), 2385–2411.

Cogley, T. and Sargent, T. J. (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US," *Review of Economic Dynamics*, 8(2), 262–302.

Dawid, A. P. (1984), "Statistical Theory: The Prequential Approach," *Journal of the Royal Statistical Society, Series A*, 147(2), 278–292.

Del Negro, M., Hasegawa, R. B., and Schorfheide, F. (2016), "Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance," *Journal of Econometrics*, 192(2), 391–403.

Diebold, F., Gunther, T. A., and Tay, A. S. (1998), "Evaluating Density Forecasts with Applications to Financial Risk Management," *International Economic Review*, 39(4), 863–883.

Dieppe, A., Legrand, R., and van Roye, B. (2016), "The BEAR Toolbox," ECB Working Paper Series No. 1934.

Dieppe, A., Legrand, R., and van Roye, B. (2018), "The Bayesian Estimation, Analysis and Regression (BEAR) Toolbox: Technical Guide," Technical Document, BEAR Toolbox, European Central Bank.

Doan, T., Litterman, R., and Sims, C. A. (1984), "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3(1), 1–100.

Douc, R., Cappé, O., and Moulines, E. (2005), "Comparison of Resampling Schemes for Particle Filtering," in *Proceedings of the 4th International Symposium on Image and Signal Processing and*

Analysis, ISPA 2005, IEEE, conference Location: Zagreb, Crotia.

Doucet, A., Briers, M., and Sénécal, S. (2006), "Efficient Block Sampling Strategies for Sequential Monte Carlo Methods," *Journal of Computational and Graphical Statistics*, 15(3), 693–711.

Doucet, A. and Johansen, A. M. (2011), "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," in D. Crisan and B. Rozovskiĭ (Editors), *The Oxford Handbook of Nonlinear Filtering*, volume 12, 656–704, Oxford University Press, Oxford.

Durbin, J. and Koopman, S. J. (2012), *Time Series Analysis by State Space Methods*, Oxford University Press, Oxford, 2nd edition.

Eklund, J. and Karlsson, S. (2007), "Forecast Combinations and Model Averaging using Predictive Measures," *Econometric Reviews*, 26(2–4), 329–363.

Fagan, G., Henry, J., and Mestre, R. (2005), "An Area-Wide Model for the Euro Area," *Economic Modelling*, 22(1), 39–59.

Genest, C. and Rivest, L. P. (2001), "On The Multivariate Probability Integral Transformation," *Statistics & Probability Letters*, 53(4), 391–399.

Geweke, J. (2010), *Complete and Incomplete Econometrics Models*, Princeton University Press, Princeton.

Geweke, J. and Amisano, G. (2011), "Optimal Prediction Pools," *Journal of Econometrics*, 164(1), 130–141.

Geweke, J. and Amisano, G. (2012), "Prediction and Misspecified Models," *American Economic Review*, 102(3), 482–486.

Giannone, D., Henry, J., Lalik, M., and Modugno, M. (2012), "An Area-Wide Real-Time Database for the Euro Area," *The Review of Economics and Statistics*, 94(4), 1000–1013.

Giannone, D., Lenza, M., and Primiceri, G. E. (2015), "Prior Selection for Vector Autoregressions," *The Review of Economics and Statistics*, 97(2), 436–451.

Giannone, D., Lenza, M., and Primiceri, G. E. (2019), "Priors for the Long Run," *Journal of the American Statistical Association*, 114(526), 565–580.

Gilks, W. R. and Berzuini, C. (2001), "Following a Moving Target—Monte Carlo Inference for Dynamic Bayesian Models," *Journal of the Royal Statistical Society Series B*, 63(1), 127–146.

Gneiting, T., Stanberry, L. I., Grimit, 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.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993), "Novel Approach to Nonlinear/non-Gaussian Bayesian State Estimation," *Radar and Signal Processing, IEE Proceedings-F*, 140(2), 107–113.

Hall, S. G. and Mitchell, J. (2007), "Combining Density Forecasts," *International Journal of Forecasting*, 23(1), 1–13.

Herbst, E. and Schorfheide, F. (2016), *Bayesian Estimation of DSGE Models*, Princeton University Press, Princeton.

Hoeting, J. A., Madigan, D., Raftery, A. E., and Volinsky, C. T. (1999), "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14(4), 382–417.

Hol, J. D., Schön, T. B., and Gustafsson, F. (2006), "On Resampling Algorithms for Particle Filters," *Proceedings of the IEEE Nonlinear Statistical Signal Processing Workshop*, Cambridge, U. K., Sept. 2006, 79–82.

Johansen, S. (1996), *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford, 2nd edition.

Jore, A. S., Mitchell, J., and Vahey, S. P. (2010), "Combining Forecast Densities from VARs with Uncertain Instabilities," *Journal of Applied Econometrics*, 25(4), 621–634.

Kitagawa, G. (1996), "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, 5(1), 1–25.

Koop, G. and Korobilis, D. (2012), "Forecasting Inflation using Dynamic Model Averaging," *International Economic Review*, 53(3), 867–886.

Kullback, S. and Leibler, R. A. (1951), "On Information and Sufficiency," *The Annals of Mathematical Statistics*, 22(1), 79–86.

Li, T., Bolić, M., and Djurić, P. M. (2015), "Resampling Methods for Particle Filtering: Classification, Implementation, and Strategies," *IEEE Signal Processing Magazine*, 32(3), 70–86.

Liu, J. S. and Chen, R. (1998), "Sequential Monte Carlo Methods for Dynamic Systems," *Journal of the American Statistical Association*, 93(443), 1032–1044.

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.

Mitchell, J. and Hall, S. G. (2005), "Evaluating, Comparing and Combining Density Forecasts Using the KLIC with an Application to the Bank of England and NIESR 'Fan' Charts of Inflation," *Oxford Bulletin of Economics and Statistics*, 67(S1), 995–1033.

Newey, W. K. and West, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Opschoor, A., van Dijk, D., and van der Wel, M. (2017), "Combining Density Forecasts using Focused Scoring Rules," *Journal of Applied Econometrics*, 32(7), 1298–1313.

Pauwels, L. L. and Vasnev, A. L. (2016), "A Note on the Estimation of Optimal Weights for Density Forecast Combinations," *International Journal of Forecasting*, 32(2), 391–397.

Raftery, A. E., 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.

Rosenblatt, M. (1952), "Remarks on a Multivariate Transformation," *The Annals Of Mathematical Statistics*, 23(3), 470–472.

Sims, C. A. (1993), "A Nine-Variable Probabilistic Macroeconomic Forecasting Model," in J. H. Stock and M. W. Watson (Editors), *Business Cycles, Indicators and Forecasting*, 179–212, University of Chicago Press, Chicago.

Sims, C. A. (2000), "Using a Likelihood Perspective to Sharpen Econometric Discourse: Three Examples," *Journal of Econometrics*, 95(2), 443–462.

Sims, C. A. and Zha, T. (1998), "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39(4), 949–968.

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.

Smith, J. Q. (1985), "Diagnostic Checks of Non-Standard Time Series Models," *Journal Of Forecasting*, 4(3), 283–291.

Waggoner, D. F. and Zha, T. (2012), "Confronting Model Misspecification in Macroeconomics," *Journal of Econometrics*, 171(2), 167–184.

Warne, A. (2022a), "DSGE Model Forecasting: Rational Expectations vs. Adaptive Learning," Manuscript, European Central Bank.

Warne, A. (2022b), "YADA Manual — Computational Details," Manuscript, European Central Bank. Available with the YADA distribution.

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.