Check for updates

Received: 28 June 2023 | Revised: 13 February 2024 | Accepted: 16 February 2024

DOI: 10.1002/for.3121

RESEARCH ARTICLE

WILEY

Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian vector autoregressions?

Martin Feldkircher[1] | **Luis Gruber**[2] | **Florian Huber**[3] | **Gregor Kastner**[2]

[1]Vienna School of International Studies (DA), Vienna, Austria

[2]Department of Statistics, University of Klagenfurt, Klagenfurt am Wörthersee, Austria

[3]University of Salzburg, Salzburg, Austria

Correspondence
Gregor Kastner, Department of Statistics, University of Klagenfurt, Universitätsstraße 65-67, 9020 Klagenfurt am Wörthersee, Austria.
Email: gregor.kastner@aau.at

Funding information
Austrian Science Fund

Abstract

We assess the relationship between model size and complexity in the time-varying parameter vector autoregression (VAR) framework via thorough predictive exercises for the euro area, the United Kingdom, and the United States. It turns out that sophisticated dynamics through drifting coefficients are important in small data sets, while simpler models tend to perform better in sizeable data sets. To combine the best of both worlds, novel shrinkage priors help to mitigate the curse of dimensionality, resulting in competitive forecasts for all scenarios considered. Furthermore, we discuss dynamic model selection to improve upon the best performing individual model for each point in time.

KEYWORDS

density predictions, dynamic model selection, global–local shrinkage priors, hierarchical modeling, stochastic volatility

1 | INTRODUCTION

In predictive inference, two main bearings can be found. First, simple models are increasingly replaced by more sophisticated versions in order to avoid functional misspecification. Second, due to increased data availability, information sets become more sizeable, and models thus higher dimensional, which in turn decreases the likelihood of omitted variable bias. The goal of this paper is a systematic assessment of the relationship between model size and complexity using different macroeconomic data sets by comparing the predictive performance of multivariate time series models that range from being relatively simple (i.e., feature constant parameters and heteroskedastic shocks) to very flexible (time-varying parameters and heteroskedastic shocks). Our conjecture is that the introduction of drifting coefficients in multivariate time series regressions can control for an omitted variable bias in small-scale models or, conversely, larger information sets can substitute for nonlinear model dynamics.

The choice between model size and model complexity is not innocuous. Large models, such as vector autoregressions (VARs) with many endogenous variables, naturally avoid an omitted variable bias. This often translates into superior predictive performance (Bańbura et al., 2010) and avoids puzzles commonly observed in empirical macroeconomics (such as the price puzzle; see Sims, 1992). Adding stochastic volatility (SV) often further improves forecast performance (Clark, 2011; Clark & Ravazzolo, 2015) but only slightly increases the computational burden if state-of-the-art sampling algorithms are used (for recent examples, see Chan et al., 2023; Kastner & Huber, 2020).

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2024 The Authors. *Journal of Forecasting* published by John Wiley & Sons Ltd.

Both VARs and VARs with SV, however, are over-parameterized if the number of endogenous variables becomes moderately large and can be miss-specified if the data generating process (DGP) features nonlinearities. This calls for appropriate regularization techniques, and the Bayesian literature has proposed appropriate shrinkage priors to overcome overfitting issues. Particular examples are the popular class of Minnesota-type priors (Doan et al., 1984; Litterman, 1986; Sims & Zha, 1998), stochastic search variable selection priors (George et al., 2008; George & McCulloch, 1993), and global–local shrinkage priors (see, e.g., Bhattacharya et al., 2015; Brown & Griffin, 2010; Follett & Yu, 2019; Huber & Feldkircher, 2019; Kastner & Huber, 2020; Polson & Scott, 2010; Zhang et al., 2022). More recently, approaches combining the merits of Minnesota-type shrinkage priors and global–local priors have been developed and used successfully in constant-parameter VARs (see, e.g., Chan, 2021; Cross et al., 2020; Gruber & Kastner, 2022).

Complicated models, by contrast, are often difficult to estimate and interpret and do not scale well to large data sets. In particular, time-varying parameter VARs (TVP-VARs) are flexible models that, as shown in Granger (2008), allow for capturing nonlinearities of unknown form. In macroeconomic data, these nonlinearities often relate to parameter change and structural breaks over longer time periods. However, TVP-VARs are often applied to small data sets with only a handful of endogenous variables. The main reason is that the computational burden increases appreciable with increasing model dimension. However, there exists a large body of literature (see, e.g., D'Agostino et al., 2013; Huber et al., 2021) that shows that small to moderately sized TVP-VARs can compete with large-dimensional VARs in predictive performance. In recent papers, researchers increasingly adopt shrinkage priors in TVP regressions to force the model toward a time-invariant specification (see, e.g., Belmonte et al., 2014; Cadonna et al., 2020; Huber et al., 2021; Kalli & Griffin, 2014). These papers leverage the non-centered parameterization of the state space model (see Frühwirth-Schnatter & Wagner, 2010) to disentangle the time-invariant component of the model from the dynamic part. In another recent paper, Chan (2023) provides methods that allow for estimating TVP-VARs where some equations feature TVPs, while other equations are characterized by constant coefficients.

To sum up this discussion, applied researchers face a trade-off. On the one hand, using large data sets alleviates the risk of missing important information that might be relevant for macroeconomic forecasting. On the other hand, estimating sophisticated models such as TVP-VARs in large dimensions is extremely challenging, and the risk of overfitting is further intensified. Table 1 provides a summary of the pros and cons of these different general modeling strategies.

In this paper, we explore this relationship empirically by considering three well-known data sets for the euro area (EA), the United Kingdom, and the United States. We consider VARs that differ in terms of their model dimensions. Specifically, we consider for each data set a small model that features three variables, a moderately sized one with seven variables, and a large model with 15 variables. To focus on the relationship between model size and complexity, we differentiate between models that feature constant and drifting parameters. In both specifications, we assume that the shocks are heteroskedastic and follow an SV model. To control for overfitting, two recent shrinkage priors, the Normal-Gamma (NG) prior (Brown & Griffin, 2010) and the Dirichlet–Laplace (DL) prior (Bhattacharya et al., 2015), are used to induce shrinkage in our different model specifications.

Our results are threefold: First, we show that the proposed TVP-VAR-SV shrinkage models improve one-step ahead forecasts. Allowing for time variation and using shrinkage priors leads to smaller drops in forecast performance during the global financial crisis—a finding that is also corroborated by looking at model weights in a dynamic model selection exercise. Second, comparing the proposed priors, we find that the DL prior shows a strong performance in small-scale applications, while the NG prior outperforms using larger information sets. This is driven by the higher degree of shrinkage the NG prior provides, which is especially important for large scale applications. Last, we demonstrate that the larger the information set, the stronger the forecast performance of a simple, constant parameter VAR with SV. However, also here, the NG-VAR-SV model turns out to be a valuable alternative, providing forecasts that are not far off those of the constant parameter competitor. To allow for different models at different points in time, we also discuss the possibility of dynamic model selection (Koop et al., 2009).

TABLE 1 Trade-offs in VAR modeling.

		Model	
		Simple	**Sophisticated**
Information set	Small	Omitted variables + misspecification	Omitted variables
	Sizeable	Misspecification	Overfitting

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The remainder of the paper is structured as follows. The second section sets the stage, introduces a standard TVP-VAR-SV model, and highlights typical estimation issues involved. Section 3 describes in detail the prior setup adopted. Section 4 presents the necessary details to estimate the model, including an overview of the Markov chain Monte Carlo (MCMC) algorithm and the relevant conditional posterior distributions. Section 5 provides empirical results alongside the main findings of our forecasting comparison. Furthermore, it contains a discussion of dynamic model selection. Finally, the last section summarizes and concludes the paper.

2 | ECONOMETRIC FRAMEWORK

In this paper, the model of interest is a TVP-VAR with SV in the spirit of Primiceri (2005). The model summarizes the joint dynamics of an M-dimensional zero-mean vector of macroeconomic time series $\{\boldsymbol{y}_t\}_{t=1}^{T}$ as follows[1]:

$$\boldsymbol{y}_t = \boldsymbol{A}_{1t}\boldsymbol{y}_{t-1} + \ldots + \boldsymbol{A}_{pt}\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}_M, \boldsymbol{\Sigma}_t). \quad (1)$$

The $M \times M$ matrix $\boldsymbol{A}_{jt}$ $(j = 1, \ldots, p)$ contains time-varying autoregressive coefficients. $\boldsymbol{\varepsilon}_t$ is a vector white noise error with zero mean and a time-varying variance–covariance matrix $\boldsymbol{\Sigma}_t = \boldsymbol{H}_t \boldsymbol{V}_t \boldsymbol{H}_t'$. $\boldsymbol{H}_t$ is a lower unitriangular matrix, and $\boldsymbol{V}_t = \text{diag}(e^{v_{1t}}, \ldots, e^{v_{Mt}})$ denotes a diagonal matrix with time-varying shock variances. The model in Equation (1) can be cast in a standard regression form as follows:

$$\boldsymbol{y}_t = \boldsymbol{A}_t \boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \quad (2)$$

with $\boldsymbol{A}_t = (\boldsymbol{A}_{1t}, \ldots, \boldsymbol{A}_{pt})$ being an $M \times (pM)$ matrix and $\boldsymbol{x}_t = \left(\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}'\right)$. Following Cogley and Sargent (2005), we can rewrite Equation (2) as

$$\boldsymbol{y}_t - \boldsymbol{A}_t \boldsymbol{x}_t = \boldsymbol{H}_t \boldsymbol{\eta}_t, \text{ with } \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_t), \quad (3)$$

and multiplying from the left with $\widetilde{\boldsymbol{H}}_t := \boldsymbol{H}_t^{-1}$ yields

$$\widetilde{\boldsymbol{H}}_t \boldsymbol{\varepsilon}_t = \boldsymbol{\eta}_t. \quad (4)$$

For further illustration, note that the first two equations of the system are given by

$$\varepsilon_{1t} = \eta_{1t}, \quad (5)$$

$$\widetilde{h}_{21,t}\varepsilon_{1t} + \varepsilon_{2t} = \eta_{2t}, \quad (6)$$

with $\widetilde{h}_{21,t}$ denoting the second element of the first column of $\widetilde{\boldsymbol{H}}_t$. Equation (6) can be rewritten as

$$y_{2t} = \boldsymbol{A}_{2\bullet,t}\boldsymbol{x}_t - \widetilde{h}_{21,t}\varepsilon_{1t} + \eta_{2t}, \quad (7)$$

where $\boldsymbol{A}_{i\bullet,t}$ denotes the ith row of $\boldsymbol{A}_t$. More generally, the ith equation of the system is a standard regression model augmented with the residuals of the preceding $i-1$ equations:

$$y_{it} = \boldsymbol{A}_{i\bullet,t}\boldsymbol{x}_t - \sum_{s=1}^{i-1} \widetilde{h}_{is,t}\varepsilon_{st} + \eta_{it}. \quad (8)$$

Thus, the ith equation is a standard regression model with $K_i = pM + i - 1$ explanatory variables given by $\boldsymbol{z}_{it} = \left(\boldsymbol{x}_t', -\varepsilon_{1t}, \ldots, -\varepsilon_{i-1,t}\right)'$ and a K_i-dimensional time-varying coefficient vector $\boldsymbol{B}_{it} = \left(\boldsymbol{A}_{i\bullet,t}, \widetilde{h}_{i1,t}, \ldots, \widetilde{h}_{ii-1,t}\right)'$. For each equation $i > 1$, the corresponding dynamic regression model is then given by

$$y_{it} = \boldsymbol{B}_{it}'\boldsymbol{z}_{it} + \eta_{it}. \quad (9)$$

The states in $\boldsymbol{B}_{it}$ evolve according to a random walk process:

$$\boldsymbol{B}_{it} = \boldsymbol{B}_{it-1} + \boldsymbol{v}_t, \text{ with } \boldsymbol{v}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_i), \quad (10)$$

where $\boldsymbol{\Omega}_i = \text{diag}(\omega_1, \ldots, \omega_{K_i})$ is a diagonal variance–covariance matrix. Note that if a given diagonal element of $\boldsymbol{\Omega}_i$ is zero, the corresponding regression coefficient is assumed to be constant over time.

Typically, conjugate inverted Gamma priors are specified on ω_j $(j = 1, \ldots, K_i)$. However, as Frühwirth-Schnatter and Wagner (2010) demonstrate, this choice is suboptimal if ω_j equals zero, since the inverted Gamma distribution artificially places prior mass away from zero and thus introduces time variation even if the likelihood points toward a constant parameter specification. To alleviate such concerns, Frühwirth-Schnatter and Wagner (2010) exploit the non-centered parameterization of Equations (9) and (10):

$$y_{it} = \boldsymbol{B}_{i0}'\boldsymbol{z}_{it} + \widetilde{\boldsymbol{B}}_{it}'\sqrt{\boldsymbol{\Omega}_i}\boldsymbol{z}_{it} + \eta_{it}. \quad (11)$$

We let $\sqrt{\boldsymbol{\Omega}_i}$ denote the matrix square root such that $\boldsymbol{\Omega}_i = \sqrt{\boldsymbol{\Omega}_i}\sqrt{\boldsymbol{\Omega}_i}$ and $\widetilde{\boldsymbol{B}}_{it}$ has typical element j given by $\widetilde{b}_{ij,t} = \frac{b_{ij,t} - b_{ij,0}}{\sqrt{\omega_{ij}}}$. The corresponding state equation is given by

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\widetilde{\boldsymbol{B}}_{it} = \widetilde{\boldsymbol{B}}_{it-1} + \boldsymbol{u}_{it}, \text{ with } \boldsymbol{u}_{it} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{K_i}). \quad (12)$$

Moving from the centered to the non-centered parameterization allows us to treat the (signed) square root of the state innovation variances as additional regression parameters to be estimated. Moreover, this parameterization also enables us to control for model uncertainty associated with whether a given element of $\boldsymbol{z}_{it}$, that is, both autoregressive coefficients and covariance parameters, should be included or excluded from the model. This can be achieved by noting that if $b_{ij,0} \neq 0$, the jth regressor is included. The second dimension of model uncertainty stems from the empirically relevant question whether a given regression coefficient should be constant or time-varying. Thus, if $\omega_{jj} \neq 0$, the jth regressor drifts smoothly over time. Especially for forecasting applications, appropriately selecting which subset of regression coefficients should be constant or time-varying proves to be one of the key determinants in achieving superior forecasting properties (Belmonte et al., 2014; Bitto & Frühwirth-Schnatter, 2019; D'Agostino et al., 2013; Korobilis, 2013).

Finally, we also have to introduce a suitable law of motion for the diagonal elements of $\boldsymbol{V}_{it}$. Here, we assume that the v_{it}'s evolve according to independent AR(1) processes:

$$v_{it} = \mu_i + \rho_i(v_{it-1} - \mu_i) + w_{it}, \quad w_{it} \sim \mathcal{N}(0, \sigma_i^2), \quad (13)$$

for $i = 1, \ldots, M$. The parameter μ_i denotes the mean of the ith log variance, ρ_i is the corresponding persistence parameter, and σ_i^2 stands for the error variance of the relevant shocks.

3 | PRIOR SPECIFICATION

We opt for a fully Bayesian approach to estimation, inference, and prediction. This calls for the specification of suitable priors on the parameters of the model. Typically, inverse Gamma or inverted Wishart priors are used for the state innovation variances in Equation (10). However, as discussed above, such priors bound the diagonal elements of $\boldsymbol{\Omega}_i$ artificially away from zero, always inducing at least some movement in the parameters of the model.

We proceed by utilizing two flexible global–local (GL) shrinkage priors (see Polson & Scott, 2010) on $\boldsymbol{B}_{i0}$ and $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{iK_i})'$. A GL shrinkage prior comprises a global scaling parameter that pushes all elements of the coefficient vector toward zero and a set of local scaling parameters that enable coefficient-specific deviations from this general pattern.

3.1 | The NG shrinkage prior

The first prior we consider is a modified variant of the NG shrinkage prior proposed in Brown and Griffin (2010) and adopted within the general class of state space models in Bitto and Frühwirth-Schnatter (2019). In what follows, we let $\boldsymbol{a}_0 = \text{vec}(\boldsymbol{A}_0)$ denote the time-invariant part of the VAR coefficients with typical element a_{0j} for $j = 1, \ldots, K = pM^2$. The corresponding signed squared root of the state innovation variance is consequently denoted by $\pm\sqrt{\omega_j}$ or simply $\sqrt{\omega_j}$. Thus, $\sqrt{\omega_j}$ crucially determines the amount of time variation in the jth element of $\boldsymbol{a}_t$.

With this in mind, our prior specification is a scale mixture of Gaussians:

$$a_{0j} \mid \tau_{aj}^2, \lambda_a \sim \mathcal{N}\left(0, 2/\lambda_a \tau_{aj}^2\right), \quad \tau_{aj}^2 \sim \mathcal{G}(\vartheta_a, \vartheta_a) \quad (14)$$

$$\sqrt{\omega_j} \mid \tau_{\omega j}^2, \lambda_\omega \sim \mathcal{N}\left(0, 2/\lambda_\omega \tau_{\omega j}^2\right), \quad \tau_{\omega j}^2 \sim \mathcal{G}(\vartheta_\omega, \vartheta_\omega) \quad (15)$$

$$\lambda_a \sim \mathcal{G}(c_{\lambda_a}, d_{\lambda_a}), \quad (16)$$

$$\lambda_\omega \sim \mathcal{G}(c_{\lambda_\omega}, d_{\lambda_\omega}), \quad (17)$$

where τ_{sj}^2 for $s \in \{a, \omega\}$ is a set of local scaling parameters that follow a Gamma distribution and λ_s is a set of global shrinkage parameters. The hyperparameter ϑ_a controls the excess kurtosis of the marginal prior,

$$p(a_{0j} \mid \lambda_a) = \int p\left(a_{0j} \mid \tau_{aj}^2, \lambda_a\right) d\tau_{aj}^2, \quad (18)$$

obtained after integrating out the local scaling parameters. For the marginal prior, λ_a controls the overall degree of shrinkage. Lower values of ϑ_a place increasing prior mass at zero while at the same time lead to heavy tails of $p(a_{0j} \mid \lambda_a)$.

On the covariance parameters $\widetilde{h}_{is,0}$ $(i = 2, \ldots, M; s = pM + 1, \ldots, K_i)$ and the associated innovation standard deviations $\gamma_{is} = \sqrt{\omega_{is}}$, we impose the standard implementation of the NG prior. To simplify prior implementation we collect the $v = M(M-1)/2$ free covariance parameters in a vector $\widetilde{\boldsymbol{h}}_0$ and the corresponding elements of $\boldsymbol{\Omega} = \text{diag}(\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_M)$ in a v-dimensional vector $\boldsymbol{\gamma}$ with typical elements $\widetilde{h}_{i0}$ and γ_i:

$$\widetilde{h}_{i0} \mid \tau_{hi}^2, \varpi_h \sim \mathcal{N}\left(0, 2/\varpi_h \tau_{hi}^2\right), \quad \tau_{hi}^2 \sim \mathcal{G}(\vartheta_h, \vartheta_h), \quad (19)$$

$$\gamma_i \mid \tau_{\gamma i}^2, \varpi_\gamma \sim \mathcal{N}\left(0, 2/\varpi_\gamma \tau_{\gamma i}^2\right), \quad \tau_{\gamma i}^2 \sim \mathcal{G}\left(\vartheta_\gamma, \vartheta_\gamma\right), \quad (20)$$

$$\varpi_h \sim \mathcal{G}(c_{\varpi_h}, d_{\varpi_h}), \quad (21)$$

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\varpi_\gamma \sim \mathcal{G}(c_{\varpi_\gamma}, d_{\varpi_\gamma}). \tag{22}$$

Here, for $s \in \{h,\gamma\}$, τ_{si}^2 is the set of local scaling parameters, and ϖ_h and ϖ_γ are global shrinkage parameters that push all covariance parameters and the corresponding state innovation standard deviations across equations to zero, respectively. The set of hyperparameters ϑ_s for $s \in \{h,\gamma\}$ again controls the excess kurtosis of the marginal priors.

3.2 | The DL shrinkage prior

The NG prior possesses good empirical properties. However, from a theoretical point of view, its properties are still not well understood. In principle, GL shrinkage priors aim to approximate a standard spike and slab prior (George et al., 2008; George & McCulloch, 1993) by introducing suitable mixing distributions on the local and global scaling parameters of the model. Bhattacharya et al. (2015) introduce a prior specification and analyze its properties within the stylized normal means problem. Their prior, the DL shrinkage prior, excels in both theory and empirical applications, especially in very high dimensions. Thus, for the TVP-VAR-SV, it seems to be well suited given the large-dimensional parameter and state space.

Similarly to the NG prior, the DL prior also depends on a set of global and local shrinkage parameters:

$$a_{0j} \mid \psi_{aj}, \xi_{aj}^2, \widetilde{\lambda}_a^2 \sim \mathcal{N}\left(0, \psi_{aj}\xi_{aj}^2\widetilde{\lambda}_a^2\right), \quad \psi_{aj} \sim Exp(1/2), \quad \xi_j \sim Dir(n_a, \ldots, n_a), \tag{23}$$

$$\sqrt{\omega_j} \mid \psi_{\omega j}, \xi_{\omega j}^2, \widetilde{\lambda}_\omega^2 \sim \mathcal{N}\left(0, \psi_{\omega j}\xi_{\omega j}^2\widetilde{\lambda}_\omega^2\right), \quad \psi_{\omega j} \sim Exp(1/2), \quad \xi_j \sim Dir(n_\omega, \ldots, n_\omega), \tag{24}$$

$$\widetilde{\lambda}_a \sim \mathcal{G}(Kn_a, 1/2), \tag{25}$$

$$\widetilde{\lambda}_\omega \sim \mathcal{G}(Kn_\omega, 1/2). \tag{26}$$

Hereby, for $s \in \{a,\omega\}$, ξ_{sj} is the set of local scaling parameters defined on the $(K-1)$-dimensional unit simplex $\mathcal{S}^{K-1} = \left\{\boldsymbol{x} = (x_1, \ldots, x_K)' : x_j \geq 0, \sum_{j=1}^K x_j = 1\right\}$ with $\boldsymbol{\xi}_s = (\xi_{s1}, \ldots, \xi_{sK})'$. The parameter n_s controls the overall tightness of the prior. ψ_{sj} is the set of auxiliary scaling parameters to achieve conditional normality.

For the variance–covariance matrix, we also impose the DL prior,

$$\widetilde{h}_{i0} \mid \psi_{hi}, \xi_{hi}^2, \widetilde{\varpi}_h^2 \sim \mathcal{N}\left(0, \psi_{hi}\xi_{hi}^2\widetilde{\varpi}_h^2\right), \quad \psi_{hi}^2 \sim Exp(1/2), \quad \xi_{hi} \sim Dir(n_h, \ldots, n_h), \tag{27}$$

$$\gamma_i \mid \psi_{\gamma i}, \xi_{\gamma i}^2, \widetilde{\varpi}_\gamma^2 \sim \mathcal{N}\left(0, \psi_{\gamma i}\xi_{\gamma i}^2\widetilde{\varpi}_\gamma^2\right), \quad \psi_{\gamma i}^2 \sim Exp(1/2), \quad \xi_{\gamma i} \sim Dir(n_\gamma, \ldots, n_\gamma), \tag{28}$$

$$\widetilde{\varpi}_h \sim \mathcal{G}(\nu n_h, 1/2), \tag{29}$$

$$\widetilde{\varpi}_\gamma \sim \mathcal{G}(\nu n_\gamma, 1/2). \tag{30}$$

The local shrinkage parameters ψ_{si} and ξ_{si}^2 for $s \in \{h,\gamma\}$ are defined analogously to the case of the regression coefficients described above. We let $\widetilde{\varpi}_s$ denote the set of global shrinkage parameters with large values, implying heavy shrinkage on the covariance parameters of the model.

The main differences of the NG and the DL prior are the presence of the Dirichlet components that introduce even more flexibility. Bhattacharya et al. (2015) show that in the framework of the stylized normal means problem, this specification yields excellent posterior contraction rates in light of a sparse DGP. Within an extensive simulation exercise, they moreover provide some evidence that this prior also works well in practice.

Finally, the prior setup on the coefficients in the state equation of the log volatilities closely follows Kastner (2016). Specifically, we place a weakly informative Gaussian prior on μ_i, $\mu_i \sim \mathcal{N}(0, 10^2)$ and a Beta prior on $\frac{\rho_i+1}{2} \sim \mathcal{B}(25, 1.5)$. Additionally, $\sigma_i^2 \sim \mathcal{G}(1/2, 1/2)$ introduces some shrinkage on the process innovation variances of the log-volatilities. This setup is used for all equations.

4 | BAYESIAN INFERENCE

The joint posterior distribution of our model is analytically intractable. Fortunately, however, the full conditional posterior distributions mostly belong to some well-known family of distributions, implying that we can set up a conceptually straightforward Gibbs sampling algorithm to estimate the model.

4.1 | A brief sketch of the MCMC algorithm

Our algorithm is related to the MCMC scheme put forward in Carriero et al. (2019) and estimates the latent states on an equation-by-equation basis. Specifically,

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

conditional on a suitable set of initial conditions, the algorithm cycles through the following steps:

1. Draw $\left(\boldsymbol{B}_{i0}', \omega_{i1}, ..., \omega_{iK_i}\right)'$ for $i = 1, ..., M$ from $\mathcal{N}(\boldsymbol{\mu}_{Bi}, \boldsymbol{V}_i)$ with $\boldsymbol{V}_i = \left(\boldsymbol{Z}_i'\boldsymbol{Z}_i + \underline{\boldsymbol{V}}_i^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_{Bi} = \boldsymbol{V}_i(\boldsymbol{Z}_i\boldsymbol{Y}_i)$. We let $\boldsymbol{Z}_i$ be a $T \times (2K_i)$ matrix with typical tth row $\left[\boldsymbol{z}_{it}', (\boldsymbol{B}_{it} \bigodot \boldsymbol{z}_{it})'\right] e^{-(v_{it}/2)}$, $\boldsymbol{Y}_i$ is a T-dimensional vector with element $y_{it} e^{-(v_{it}/2)}$, and $\underline{\boldsymbol{V}}_i$ is a prior covariance matrix that depends on the prior specification adopted. Note that in contrast to Carriero et al. (2019) who sample the VAR parameters in $\boldsymbol{A}_0$ and the elements of $\widetilde{\boldsymbol{H}}_0$ conditionally on each other, we follow Eisenstat et al. (2016) and sample these jointly, which speeds up the mixing of the sampler.
2. Simulate the full history of $\left\{\widetilde{\boldsymbol{B}}_{it}\right\}_{t=1}^{T}$ by means of a forward filtering backward sampling algorithm (see Carter & Kohn, 1994; Frühwirth-Schnatter, 1994) per equation.
3. The log volatilities and the corresponding parameters of the state equation in Equation (13) are simulated using the algorithm put forward in Kastner and Frühwirth-Schnatter (2014) via the R package stochvol (Hosszejni & Kastner, 2021; Kastner, 2016).
4. Depending on the prior specification adopted, draw the parameters used to construct $\underline{\boldsymbol{V}}_i$ using the conditional posterior distributions detailed in Section 4.2 (NG prior) or Section 4.3 (DL prior).

This algorithm produces draws from the joint posterior distribution of the states and the model parameters. In the empirical application that follows we use 30,000 iterations where we discard the first 15,000 as burn-in.

4.2 | Conditional posterior distributions associated with the NG prior

Conditional on the full history of all latent states in our model as well as the global shrinkage parameters, it is straightforward to show that the conditional posterior distributions of τ_{sj}^2 for $s \in \{a, \omega\}$ and $j = 1, ..., K$ are given by

$$\tau_{aj}^2 \mid \bullet \sim \mathcal{GIG}\left(\vartheta_a - 1/2, a_{0j}^2, \vartheta_a\lambda_a\right), \quad \tau_{\omega j}^2 \mid \bullet \sim \mathcal{GIG}\left(\vartheta_\omega - 1/2, \sqrt{\omega}_j^2, \vartheta_\omega\lambda_\omega\right), \tag{31}$$

where $\bullet$ indicates conditioning on the remaining parameters and states of the model. Moreover, $\mathcal{GIG}(\zeta, \chi, \varrho)$ denotes the generalized inverse Gaussian distribution with density proportional to $x^{\zeta-1}\exp\{-(\chi/x + \varrho x)/2\}$. To draw from this distribution, we use the algorithm of Hörmann and Leydold (2013) implemented in the R package GIGrvg (Leydold & Hörmann, 2017).

The conditional posteriors of the local scalings for the covariance parameters and their corresponding innovation standard deviations also follow GIG distributions:

$$\tau_{hi}^2 \mid \bullet \sim \mathcal{GIG}\left(\vartheta_h - 1/2, \widetilde{h}_{i0}^2, \vartheta_h\varpi_h\right), \quad \tau_{\gamma i}^2 \mid \bullet \sim \mathcal{GIG}\left(\vartheta_\gamma - 1/2, \gamma_i^2, \vartheta_\gamma\varpi_\gamma\right). \tag{32}$$

The conditional posterior of λ_s for $s \in \{a, \omega\}$ is of well-known form, namely, a Gamma distribution:

$$\lambda_s \mid \bullet \sim \mathcal{G}\left\{c_{\lambda_s} + \vartheta_s K, d_{\lambda_s} + \frac{\vartheta_s}{2}\sum_{j=1}^{K}\tau_{sj}^2\right\}. \tag{33}$$

Likewise, for $s \in \{h, \gamma\}$, the conditional posterior of ϖ_s is given by

$$\varpi_s \mid \bullet \sim \mathcal{G}\left\{c_{\varpi_s} + \vartheta_s v, d_{\varpi_s} + \frac{\vartheta_s}{2}\sum_{i=1}^{v}\tau_{sj}^2\right\}. \tag{34}$$

4.3 | Conditional posterior distributions associated with the DL prior

Following Bhattacharya et al. (2015), we update jointly the full conditional posterior distribution $p\left(\xi_{sj}, \lambda_s, \psi_{sj} \mid \bullet\right)$ for $s \in \{a, \omega\}$. For the Dirichlet components, the conditional posterior distribution is obtained by sampling a set of K auxiliary variables $N_{aj}, N_{\omega j}$ $(j = 1, ..., K)$:

$$N_{aj} \mid a_{0j} \sim \mathcal{GIG}\left(n_a - 1, 2 \mid a_{0j} \mid, 1\right), \quad N_{\omega j} \mid \sqrt{\omega}_j \sim \mathcal{GIG}\left(n_\omega - 1, 2 \mid \sqrt{\omega}_j \mid, 1\right). \tag{35}$$

After obtaining the K scaling parameters, we set $\xi_{aj} = N_{aj}/N_a$ and $\xi_{\omega j} = N_{\omega j}/N_\omega$ with $N_a = \sum_{j=1}^{K} N_{aj}$ and $N_\omega = \sum_{j=1}^{K} N_{\omega j}$.

The conditional posterior of the global shrinkage parameter $\widetilde{\lambda}_s$ for $s \in \{a, \omega\}$ follows a GIG distribution:

$$\widetilde{\lambda}_a \mid \xi_{aj}, a_{0j} \sim \mathcal{GIG}\left\{K(n_a - 1), 2\sum_{j=1}^{K} \mid a_{0j} \mid / \xi_{aj}, 1\right\},$$
$$\widetilde{\lambda}_\omega \mid \xi_{\omega j}, \sqrt{\omega}_j \sim \mathcal{GIG}\left\{K(n_\omega - 1), 2\sum_{j=1}^{K} \mid \sqrt{\omega}_j \mid / \xi_{\omega j}, 1\right\}. \tag{36}$$

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The full conditional posterior distributions of ψ_{aj}^{-1} and $\psi_{\omega j}^{-1}$ are inverse Gaussian:

$$\psi_{aj}^{-1} \mid \bullet \sim \mathcal{IG}\left(\xi_{aj}\widetilde{\lambda}_a/|a_{0j}|,1\right), \quad \psi_{\omega j}^{-1} \mid \bullet \sim \mathcal{IG}\left(\xi_{\omega j}\widetilde{\lambda}_\omega/|\sqrt{\omega}_j|,1\right). \tag{37}$$

To sample from the conditional posterior distribution of ξ_{si} for $s \in \{h,\gamma\}$, again, we introduce a set of auxiliary variables $N_{hi}, N_{\gamma i}$,

$$N_{hi}|\tilde{h}_{i0} \sim \mathcal{GIG}(n_h - 1, 2|\tilde{h}_{i0}|, 1), \quad N_{\gamma i}|\gamma_i \sim \mathcal{GIG}(n_\gamma - 1, 2|\gamma_i|, 1), \tag{38}$$

and obtain draws from ξ_{hi} and $\xi_{\gamma i}$ by using $\xi_{hi} = N_{hi}/\sum_{i=1}^{v} N_{hi}$ and $\xi_{\gamma i} = N_{\gamma i}/\sum_{i=1}^{v} N_{\gamma i}$. The global shrinkage parameter on the covariance parameters and the process innovation variances follows again a GIG distribution:

$$\begin{aligned} \widetilde{\varpi}_h \mid \widetilde{h}_{i0}, \xi_{hi} &\sim \mathcal{GIG}\left\{v(n_h - 1), 2\sum_{j=1}^{v}\left(\frac{|\widetilde{h}_{i0}|}{\xi_{hi}}\right), 1\right\}, \\ \widetilde{\varpi}_\gamma \mid \gamma_i, \xi_{\gamma i} &\sim \mathcal{GIG}\left\{v(n_\gamma - 1), 2\sum_{j=1}^{v}\left(\frac{|\gamma_i|}{\xi_{\gamma i}}\right), 1\right\}. \end{aligned} \tag{39}$$

The full conditional posterior distributions of ψ_{hi}^{-1} and $\psi_{\gamma i}^{-1}$ for $i = 1,\ldots,v$ are given by

$$\psi_{hi}^{-1} \mid \bullet \sim \mathcal{IG}\left(\widetilde{\varpi}_h\zeta_{hi}/|\widetilde{h}_{i0}|,1\right), \quad \psi_{\gamma i}^{-1} \mid \bullet \sim \mathcal{IG}\left(\widetilde{\varpi}_\gamma\zeta_{\gamma i}/|\gamma_i|,1\right). \tag{40}$$

5 | FORECASTING MACROECONOMIC QUANTITIES FOR THREE MAJOR ECONOMIES

In what follows we systematically assess the relationship between model size and model complexity by forecasting several macroeconomic indicators for three large economies, namely, the EA, the United Kingdom, and the United States. In Section 5.1, we briefly describe the different data sets and discuss model specification issues. Section 5.2 deals with simple visual summaries of posterior sparsity in terms of the VAR coefficients and their time variation for the two shrinkage priors proposed. The main forecasting results are discussed in Section 5.3. Finally, Section 5.4 discusses the possibility to dynamically select among different specifications in an automatic fashion.

5.1 | Data and model specification

We use prominent macroeconomic data sets for the EA, the United Kingdom, and the United States. All three data sets are on a quarterly frequency but span different periods of time. For the EA, we take data from the area wide model (Fagan et al., 2005) and additionally include equity prices available from 1987Q1 to 2015Q4. UK data stem from the Bank of England's "A millennium of macroeconomic data" (Thomas et al., 2010) and covers the period from 1982Q2 to 2016Q4. For the United States, we use a subset from the FRED QD database (McCracken & Ng, 2016), which covers the period from 1959Q1 to 2015Q1.

For each of the three cases we use three subsets, a small (three variables), a medium (seven variables), and a large (15 variables) subset. The small subset covers only real activity, prices, and short-term interest rates. The medium models cover in addition investment and consumption, the unemployment rate, and either nominal or effective exchange rates. For the large models, we add wages, money (measured as M2 or M3), government consumption, exports, equity prices, and 10-year government bond yields.

To complete the data set for the large models, we include additional variables depending on data availability for each country set. For example, the UK data set offers a wide range of financial data, so we complement the large model by including also data on mortgage rates and bond spreads. For the EA data set, we include also a commodity price indicator and labor market productivity, while for the United States, we add consumer sentiment and hours worked. In what follows, we are interested not only in the relative performance of the different priors but also in the forecasting performance using different information sets. Thus, we have opted firstly to strike a good balance between different types of data (e.g., real, labor market, and financial market data) and secondly to alter variables for the large data sets slightly. This is done to rule out that performance between information sets depends crucially on the type of information that is added (e.g., labor market data vs. financial market data).

For data that are nonstationary, we take first differences; see Table A1 in the appendix for more details. Consistent with the literature (Cogley & Sargent, 2005; D'Agostino et al., 2013; Primiceri, 2005), we include $p = 2$ lags of the endogenous variables in all models. For both the NG prior and the DL prior, we use lag-wise specifications, similarly to those in Huber and Feldkircher (2019), in order to increase shrinkage for higher lags.[2]

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

5.2 | Inspecting posterior sparsity

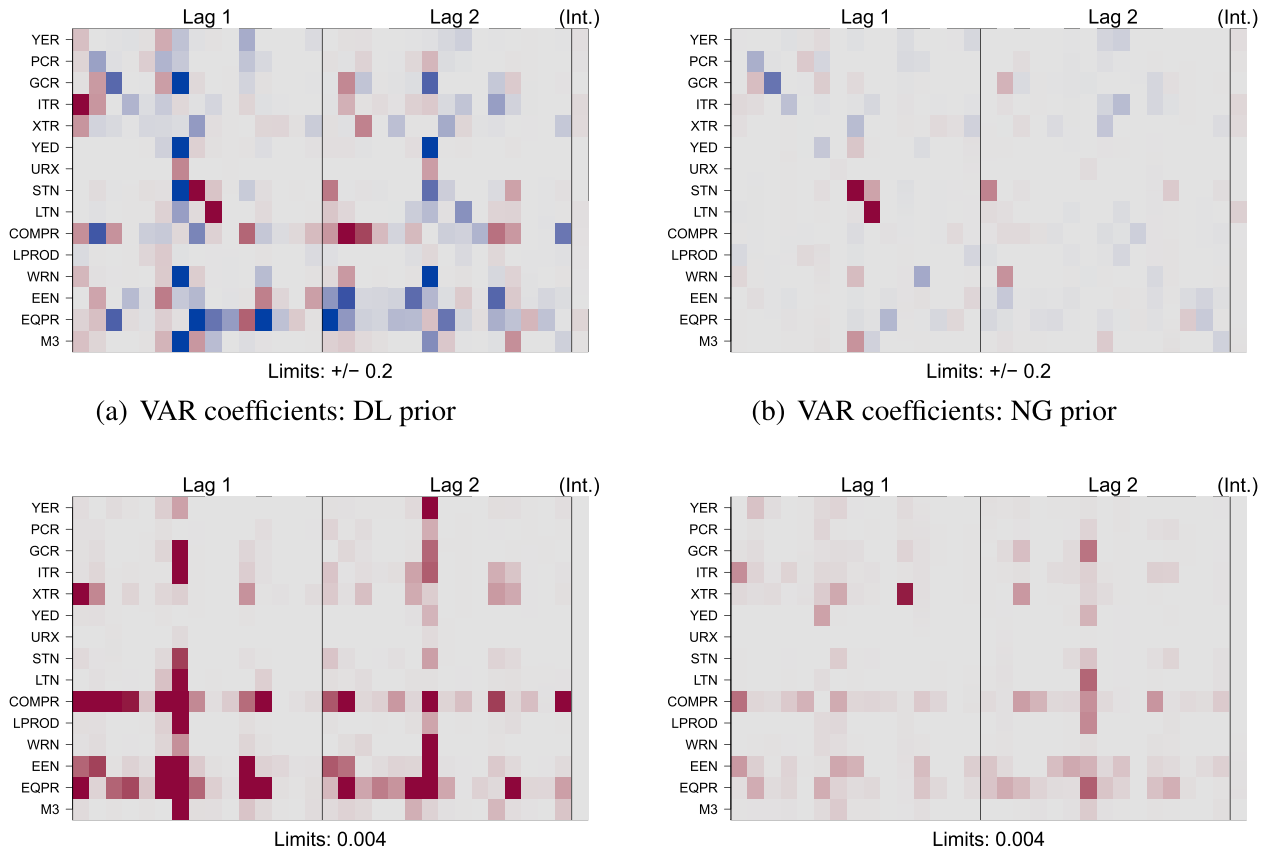
Before we turn to the forecasting exercise, we assess the amount of sparsity induced by our two proposed global–local shrinkage specifications, labeled TVP-SV NG and TVP-SV DL. This analysis is based on inspecting heatmaps that show the posterior mean of the coefficients as well as the posterior mean of the standard deviations that determine the amount of time variation in the dynamic regression coefficients. Figures 1–3 show the corresponding heatmaps. Red and blue squares indicate positive and negative values, respectively. To permit comparability, we use the same scaling across priors within a given country.

We start by inspecting posterior sparsity attached to the time-invariant part of the models, provided in the upper panels of Figures 1–3. We generally find that the first own lag of a given variable appears to be important, while the second lag is slightly less important in most equations. This can be seen by dense (i.e., colored) main diagonals elements. Turning to variables along the off-diagonal elements, that is, the coefficients associated with variables $j \neq i$ in equation i, we find considerable evidence that the (un)employment rate as well as long-term interest rates appears to load heavily on the other quantities in most country models, as indicated by relatively dense columns associated with the first lag of unemployment and interest rates.

Equations that are characterized by a large amount of nonzero coefficients (i.e., dense rows) are mostly related to financial variables, namely, exchange rates, equity, and commodity prices. These observations are general in nature and relate to all three countries considered.

In the next step, we investigate sparsity in terms of the degree of time variation of the VAR coefficients (see the lower panels of Figures 1–3). Here, we observe that, consistent with the dense pattern in $\boldsymbol{a}_0$, equations associated with financial variables display the largest amount of time-variation. Interestingly, the results suggest that coefficients in the EA tend to display a greater propensity to drift as compared to the coefficients of the UK country model.

Comparing the degree of shrinkage between both the DL and the NG prior reveals that the latter specification induces much more sparsity in large-dimensional systems. While both priors yield rather sparse models, the



(a) VAR coefficients: DL prior

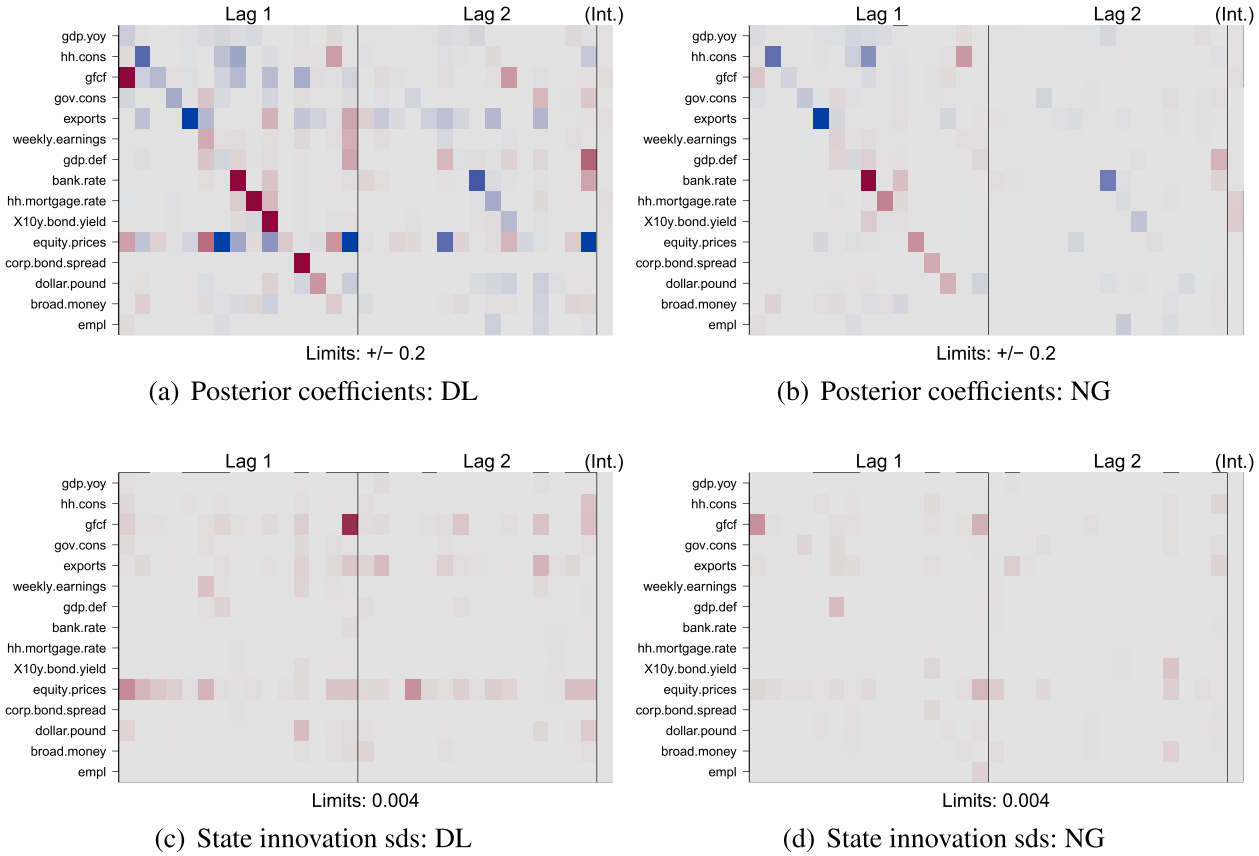
(b) VAR coefficients: NG prior

(c) State innovation sds: DL prior

(d) State innovation sds: NG prior

FIGURE 1 Posterior means in the large model: euro area.

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Posterior coefficients: DL (b) Posterior coefficients: NG

(c) State innovation sds: DL (d) State innovation sds: NG

FIGURE 2 Posterior means in the large model: United Kingdom.

findings point toward a much stronger degree of shrinkage of the NG prior. Notice that the NG prior also favors constant parameter specifications. This suggests that in large-scale applications, the NG prior might be particular useful when issues of overparametrization are more of a concern, while in smaller models, the flexibility of the DL prior might be beneficial.

5.3 | Forecasting results

In this section, we examine the forecasting performance of the proposed prior specifications. The forecasting setup largely follows Huber and Feldkircher (2019) and focuses on the one-quarter- and one-year-ahead forecast horizons and three different information sets: small (three variables), medium (seven variables), and large (15 variables). We use an expanding window and a hold-out sample of 80 quarters, which results into the following hold out samples: 1995Q4-2015Q3 for the EA, 1997Q1-2016Q4 for the United Kingdom, and 1995Q4-2015Q3 for the United States.

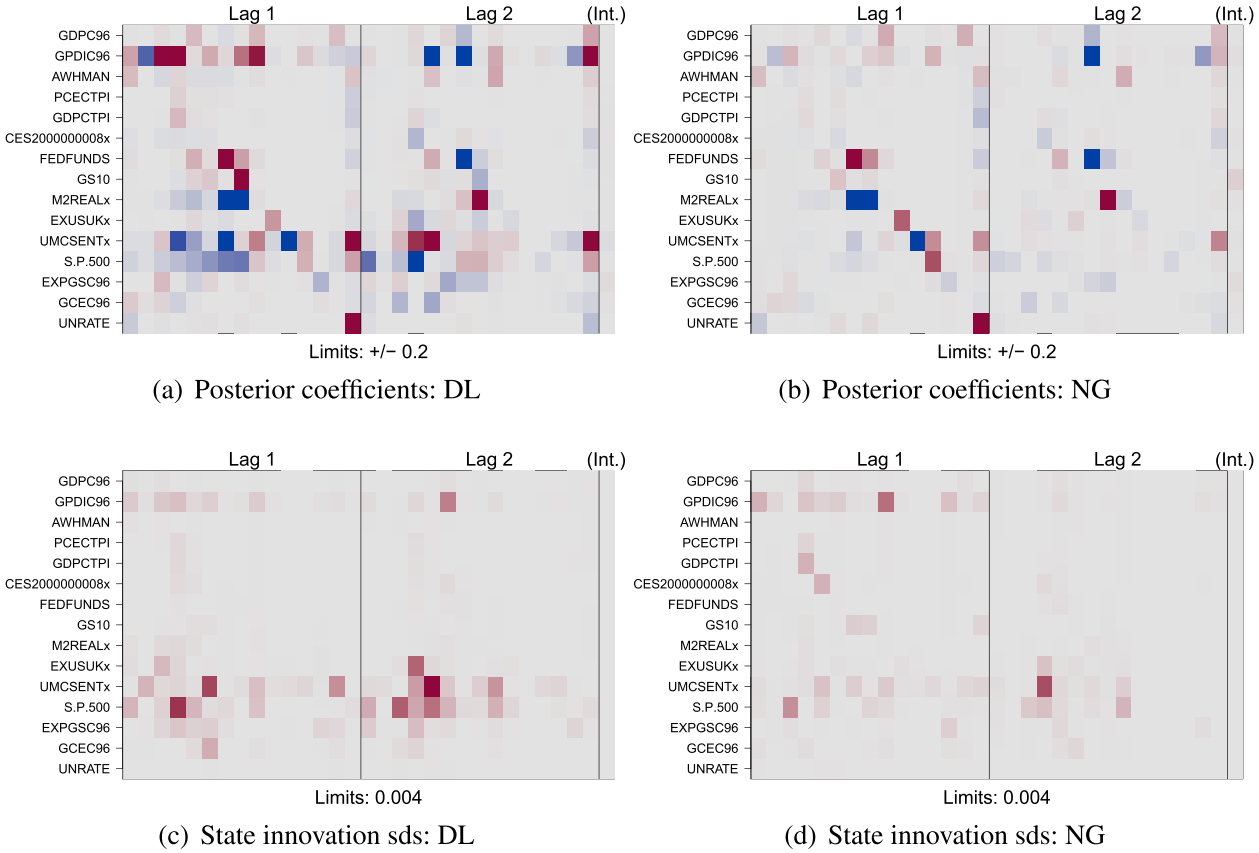
Forecasts are evaluated using log predictive scores (LPSs), a widely used metric to measure density forecast accuracy (see, e.g., Geweke & Amisano, 2010). The LPSs are closely related to the marginal likelihood, a standard Bayesian measure for discriminating between competing models. It is also worth stressing that, as opposed to focusing on point forecasts, considering LPSs allows us to factor in how well a given model captures higher-order features of the predictive distribution. These higher-order features are important if the researcher is interested in reporting functions of the predictive distribution, such as growth at risk (Adrian et al., 2019) or other measures relevant for capturing tail risks to macroeconomic outcomes (Clark et al., 2023). Last, relying on point forecasts requires the researcher to choose a loss function (in addition to model and prior choice) in order to evaluate their quality, whereas using LPSs amounts to choosing the model-implied loss function.

We compare the NG and DL specifications with a simpler constant parameter Bayesian VAR (BVAR-SV) and a time-varying parameter VAR with a loose prior setting (TVP-SV) as a general benchmark. Specifically, this benchmark model assumes that the prior on $\sqrt{\omega_j}$ is given by

$$\omega_j \sim \mathcal{G}(1/2, 1/2) \Leftrightarrow \pm\sqrt{\omega_j} \sim \mathcal{N}(0,1). \tag{41}$$

On $\boldsymbol{a}_0$ and for the BVAR-SV, we use the NG shrinkage prior described in Section 3. For the evaluation, we

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Posterior coefficients: DL

(b) Posterior coefficients: NG

(c) State innovation sds: DL

(d) State innovation sds: NG

FIGURE 3 Posterior means in the large model: United States.

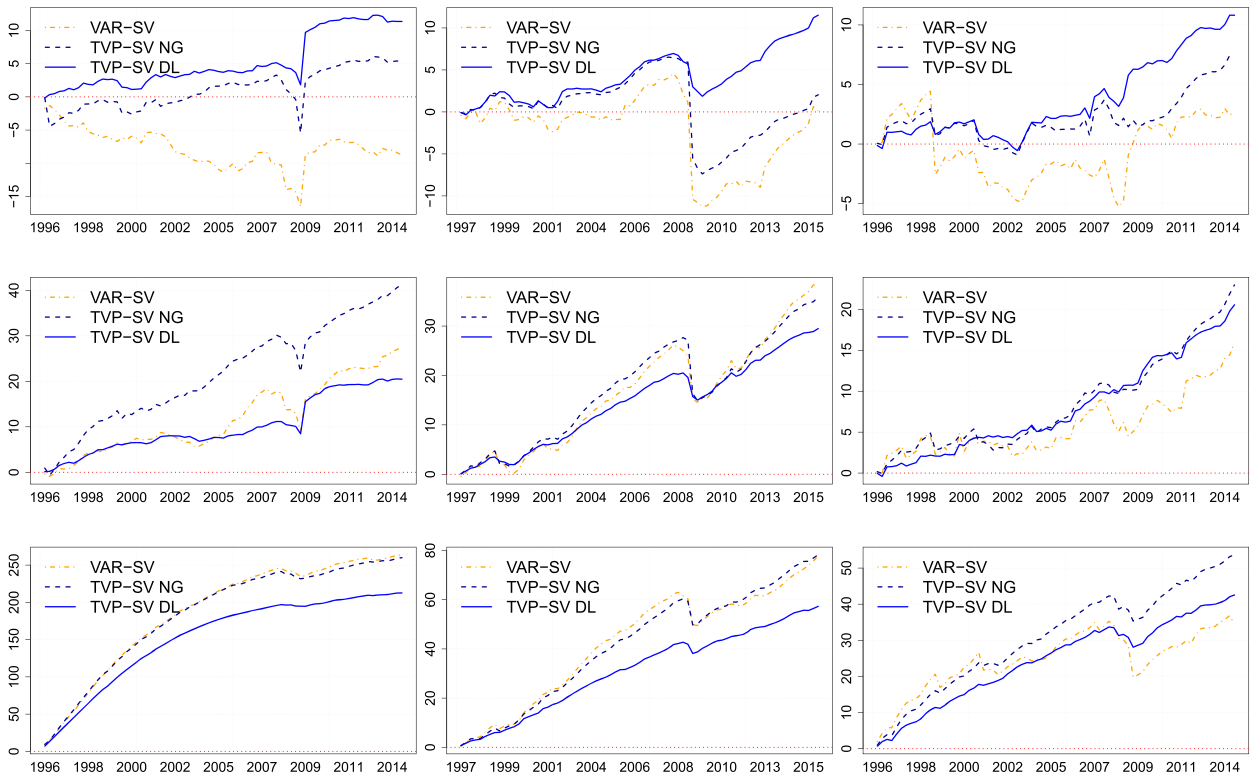


FIGURE 4 One-quarter-ahead cumulative log predictive Bayes' factors for the three focus variables GDP growth, inflation, and short-term interest rates over time, relative to the TVP-SV-VAR without shrinkage. Top row: Small model (three variables). Middle row: Medium model (seven variables). Bottom row: Large model (15 variables). Left column: Euro area (EA). Middle column: United Kingdom (UK). Right column: United States (US).

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

focus on the joint predictive distribution of three focal variables, namely, GDP growth, inflation, and short-term interest rates. This allows us to assess the predictive differences obtained by switching from small to large information sets. Figure 4 summarizes the results for the one-step-ahead forecast horizon. All panels display LPSs for the three focus variables relative to the TVP-SV specification. To assess the overall forecast performance over the hold-out sample, particularly consider the rightmost point in the respective figures.

Doing so reveals that the time-varying parameter specifications, TVP-SV NG and TVP-SV DL outperform the benchmark for all three countries and information sets as indicated by positive log predictive Bayes' factors. Except for the EA and the small information set, this finding holds also true for the constant parameter VAR-SV specification. Zooming in and looking at performance differences among the priors reveals that the TVP-SV DL specification dominates in the case of small models. The TVP-SV NG prior ranks second, and the constant parameter VAR-SV model performs worst. The dominance of the DL prior stems from the performance during the period of the global financial crisis 2008/2009. While predictions from all model specifications worsen, they deteriorate the least for the DL specification. In particular for the EA and the United Kingdom, the dominance of the DL prior stems mainly from improved forecast for short-term interest rates, see Figures A1 and A2 in Appendix B.

It is worth noting that in small-dimensional models, the TVP-SV specification also performs quite well and proves to be a competitive alternative relative to the BVAR-SV model. This is due to the fact that parameters are allowed to move significantly with only little punishment introduced through the prior, effectively controlling for structural breaks and sharp movements in the underlying structural parameters. This result corroborates findings in D'Agostino et al. (2013) and appears to support our conjecture that for small information sets, allowing for time variation proves to dominate the detrimental effect of the large number of additional parameters to be estimated.

In the next step, we enlarge the information set and turn our focus to the seven variable VAR specifications. Here, the picture changes slightly, and the NG prior outperforms forecasts of its competitors. Depending on the country, either forecasts of the DL specification or

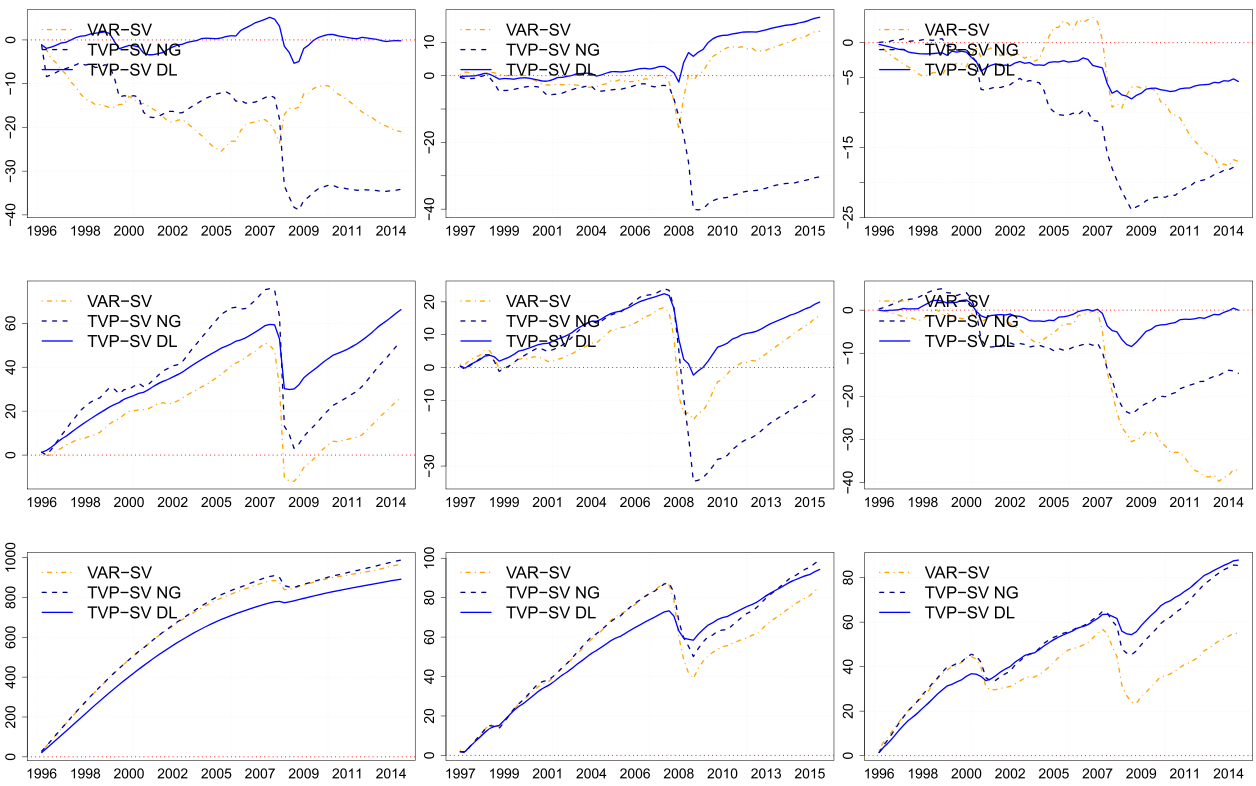


FIGURE 5 Four-quarter-ahead cumulative log predictive Bayes' factors for the three focus variables GDP growth, inflation, and short-term interest rates over time, relative to the TVP-SV-VAR with loose shrinkage. Top row: Small model (three variables). Middle row: Medium model (seven variables). Bottom row: Large model (15 variables). Left column: Euro area (EA). Middle column: United Kingdom (UK). Right column: United States (US).

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the constant parameter VAR-SV model rank second. For US data, it pays off to use a time-varying parameter specification since—as with the small information set—the BVAR-SV model performs worst. Finally, we turn to the large VAR specifications, featuring 15 variables. Here, we see a very similar picture as with the seven-variable specification. The TVP-SV NG prior yields the best forecasts, with the constant parameter model turning out to be a strong competitor. Only for US data, both time-varying parameter specifications clearly outperform the constant parameter competitor.

We now briefly examine forecasts for the four quarter horizon displayed in Figure 5. For the small and medium-sized models, all competitors yield forecasts that are close or worse compared to the loose shrinkage benchmark prior model. The high degree of shrinkage induced by the NG prior yields particularly poor forecasts, especially for observations that fall in the period of the global financial crisis. The picture slightly reverses when considering the large-scale models. Here, all competitors easily outperform forecasts of the loose benchmark model, implying that shrinkage pays off. Viewed over all settings, the DL prior does a fine job in balancing the degree of shrinkage across model sizes.

5.4 | Improving predictions through dynamic model selection

The discussion in the previous subsection highlighted the marked heterogeneity of model performance over time. In terms of achieving superior forecasting results, one could ask whether there are gains from dynamically selecting models.

Following Raftery et al. (2010); Koop and Korobilis (2012); Onorante and Raftery (2016) we perform dynamic model selection by computing a set of weights for each model within a given model size. These weights are based on the predictive likelihood for the three focus variables at $t-1$. Intuitively speaking, this combination scheme implies that if a given model performed well in predicting last quarter's output, inflation and interest rates, it receives a higher weight in the next period. By contrast, models that performed badly receive less weight in the model pool. We further employ a so-called forgetting factor that induces persistence in the model weights over time. This implies that the weights are not only shaped by the most recent forecast performance of the underlying models, but also by their historical forecasting performance. Finally, to select a given model, we simply pick the one with the highest weight.

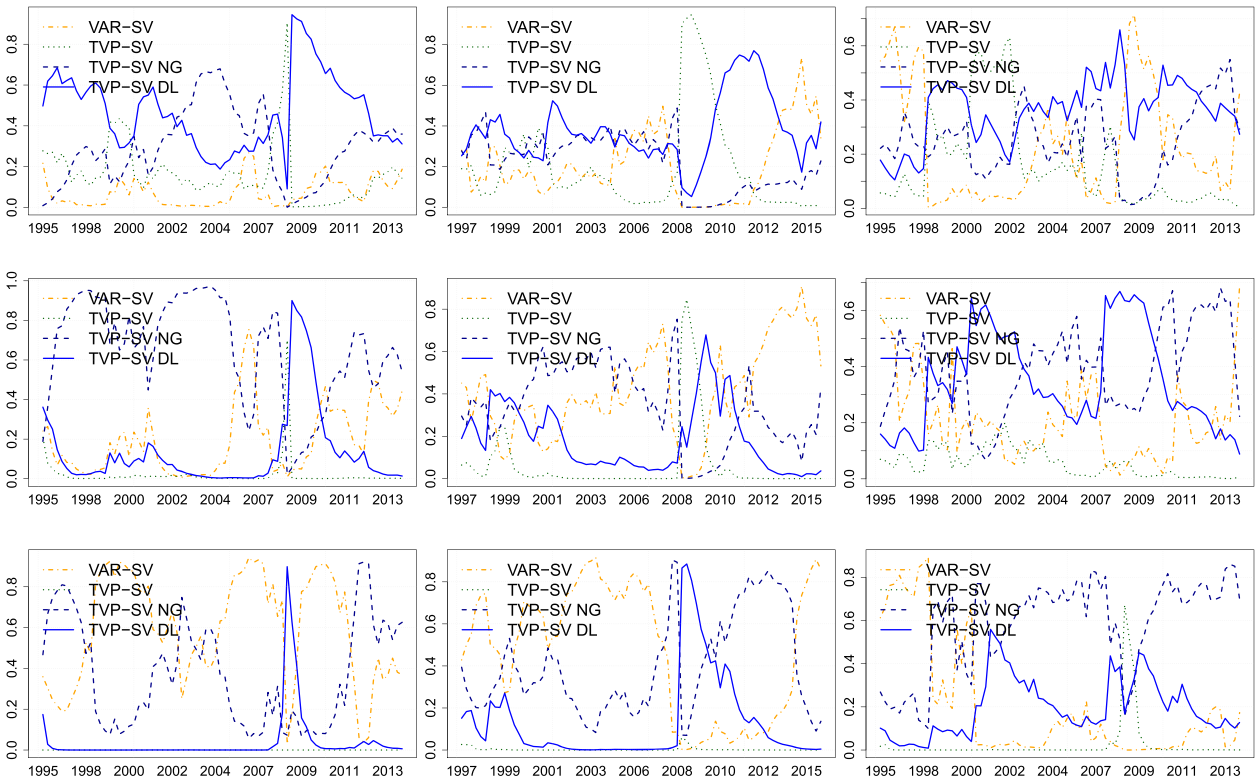


FIGURE 6 Model weights over time as implied by the predictive performance for the three focus variables GDP growth, inflation, and short-term interest rates. Top row: Small model (three variables). Middle row: Medium model (seven variables). Bottom row: Large model (15 variables). Left column: Euro area (EA). Middle column: United Kingdom (UK). Right column: United States (US).

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The predicted weight associated with model i is computed as follows

$$\mathfrak{w}_{t|t-1,i} := \frac{\mathfrak{w}^{\alpha}_{t-1|t-1,i}}{\sum_{i \in \mathscr{M}} \mathfrak{w}^{\alpha}_{t-1|t-1,i}}, \tag{42}$$

with $\alpha = 0.99$ denoting a forgetting factor close to unity and $\mathfrak{w}_{t-1|t-1,i}$ is given by

$$\mathfrak{w}_{t-1|t-1,i} = \frac{\mathfrak{w}_{t-1|t-2,i} p_{t-1|t-2,i}}{\sum_{i \in \mathscr{M}} \mathfrak{w}_{t-1|t-2,i} p_{t-1|t-2,i}}.$$

Here, $p_{t-1|t-2,i}$ denotes the one-step-ahead predictive likelihood for the three focus variables in $t-1$ for model i within the model space $\mathscr{M}$. Letting t_0 stand for the final quarter of the training sample, the initial weights $\mathfrak{w}_{t_0+1|t_0,i}$ are assumed to be equal for each model.

Before proceeding to the forecasting results, Figure 6 shows the model weights over time. One interesting regularity for small-scale models is that especially during the crisis period, the algorithm selects the benchmark, weak shrinkage TVP-SV model. This choice, however, proves to be of transient nature and the algorithm quickly adapts and switches back to either the TVP-SV NG or the TVP-SV DL model. We interpret this finding to be related to the necessity to quickly adjust to changes in the underlying macroeconomic conditions in light of the small information set adopted. The TVP-SV model allows for large shifts in the underlying regression coefficients, whereas the specifications based on hierarchical shrinkage priors introduce shrinkage, which excels over the full hold-out period but proves to be detrimental during crisis episodes.

For the medium and large information set, the model weights corroborate the results reported in the previous subsection. Specifically, we see that TVP-SV NG and TVP-SV DL receive high weights during the global financial crisis, while the BVAR-SV receives large shares of posterior probability during the remaining periods. This implies that during periods with overall heightened uncertainty, gains from using a time-varying parameter framework are sizable.

We now turn to the forecasting results using DMS, provided in Figure 7. The figure shows the log predictive Bayes' factors relative to the best performing models over the whole sample period. These correspond to those

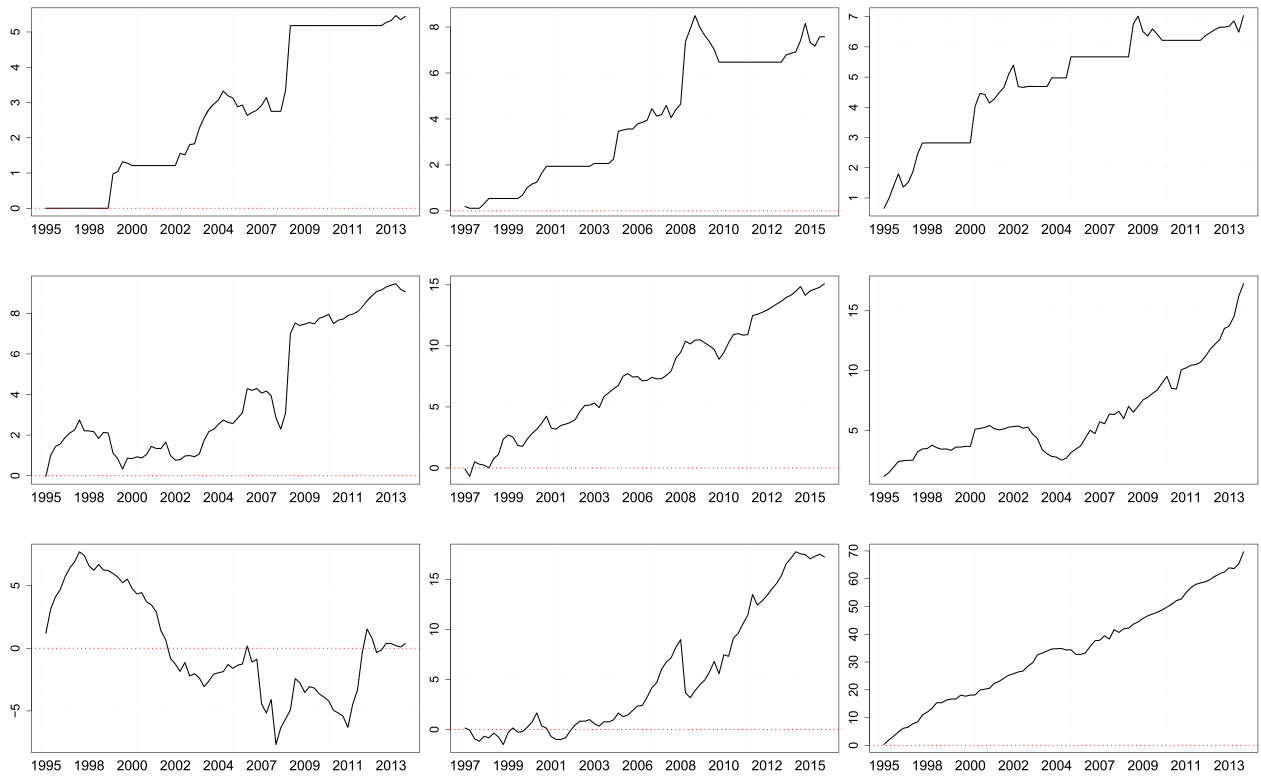


FIGURE 7 Log predictive Bayes' factors of dynamic model selection relative to the best performing model over time, as implied by the predictive performance for the three focus variables GDP growth, inflation, and short-term interest rates. Top row: Small model (three variables). Middle row: Medium model (seven variables). Bottom row: Large model (15 variables). Left column: Euro area (EA). Middle column: United Kingdom (UK). Right column: United States (US).

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

achieving the highest cumulative log predictive Bayes' factors in Figure 4.

The results indicate that dynamic model selection tends to improve forecasts throughout all model sizes and for all three country data sets. In particular, during the period of the global financial crisis, selecting from a pool of model pays off. Forecast gains during the crisis are more pronounced for the EA and the United Kingdom, whereas with US data, forecasts are more gradually improving over the sample period. Forecasts for the EA that are based on the large information set are less precise during the period from 2000 to 2012 compared to the benchmark models. This might be related to the creation of the euro, which in turn has triggered a fundamental shift in the joint dynamics of the EA's macro model. Due to the persistence in the models' weights, the model selection algorithm takes some time to adapt to the new regime. This can be seen by investigating the latest period in the sample, in which dynamic model selection again outperforms forecasts of the benchmark model. In other words, for EA data, either restricting the sample period to post 2000 or reducing the persistence via the forgetting factors might improve forecasting results.

6 | CONCLUSIVE REMARKS

In this paper, our goal is to investigate how model complexity (loosely defined as the dimension of the state space) interacts with model size. The initial conjecture is that small but flexible (TVP) models are capable of achieving predictive performances that are comparable to large-scale models that assume constant parameters (but time-varying error variances). In three different data sets, we find precisely this pattern. The larger the model gets, the less we require time-varying parameters and the closer the forecasting performance of the large TVP-VAR is to the constant parameter large VAR with SV.

This is mainly driven by the fact that modern global–local shrinkage priors introduce substantial shrinkage on the process innovation variances and the degree of shrinkage increases rapidly in the number of coefficients. In case that the model is small, the TVPs control for omitted variable biases, and this proves to be a good substitute for using larger data sets in combination with simpler models.

ACKNOWLEDGMENT

Luis Gruber, Florian Huber, and Gregor Kastner acknowledge funding from the Austrian Science Fund (FWF) for the project "High-dimensional statistical learning: New methods to advance economic and sustainability policy" (ZK 35), jointly carried out by the University of Klagenfurt, Paris Lodron University Salzburg, TU Wien, and the Austrian Institute of Economic Research (WIFO).

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available. A detailed description is given in Appendix A.

ORCID

Martin Feldkircher https://orcid.org/0000-0002-5511-9215
Luis Gruber https://orcid.org/0000-0002-2399-738X
Florian Huber https://orcid.org/0000-0002-2896-7921
Gregor Kastner https://orcid.org/0000-0002-8237-8271

ENDNOTES

[1] To simplify the model exposition, we omit an intercept term in this section. Irrespectively of this, we allow for nonzero intercepts in the empirical applications that follow.

[2] Given that the lag length is set equal to two, this choice does not matter much empirically.

REFERENCES

Adrian, T., Boyarchenko, N., & Giannone, D. (2019). Vulnerable growth. *American Economic Review*, *109*, 1263–1289. https://doi.org/10.1257/aer.20161923

Bańbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto-regressions. *Journal of Applied Econometrics*, *25*, 71–92. https://doi.org/10.1002/jae.1137

Belmonte, M. A., Koop, G., & Korobilis, D. (2014). Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting*, *33*, 80–94. https://doi.org/10.1002/for.2276

Bhattacharya, A., Pati, D., Pillai, N. S., & Dunson, D. B. (2015). Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association*, *110*, 1479–1490. https://doi.org/10.1080/01621459.2014.960967

Bitto, A., & Frühwirth-Schnatter, S. (2019). Achieving shrinkage in a time-varying parameter model framework. *Journal of Econometrics*, *210*, 75–97. https://doi.org/10.1016/j.jeconom.2018.11.006

Brown, P. J., & Griffin, J. E. (2010). Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis*, *5*, 171–188. https://doi.org/10.1214/10-BA507

Cadonna, A., Frühwirth-Schnatter, S., & Knaus, P. (2020). Triple the gamma—A unifying shrinkage prior for variance and variable selection in sparse state space and TVP models. *Econometrics*, *8*, 20. https://doi.org/10.3390/econometrics8020020

Carriero, A., Clark, T. E., & Marcellino, M. G. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, *212*, 137–154. https://doi.org/10.1016/j.jeconom.2019.04.024

Carter, C. K., & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, *81*, 541–553. https://doi.org/10.1093/biomet/81.3.541

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Chan, J. C. (2021). Minnesota-type adaptive hierarchical priors for large Bayesian VARs. *International Journal of Forecasting*, *37*, 1212–1226. https://doi.org/10.1016/j.ijforecast.2021.01.002

Chan, J. C. (2023). Large hybrid time-varying parameter VARs. *Journal of Business & Economic Statistics*, *41*, 890–905. https://doi.org/10.1080/07350015.2022.2080683

Chan, J. C., Koop, G., & Yu, X. (2023). Large order-invariant Bayesian VARs with stochastic volatility. *Journal of Business & Economic Statistics*, 1–13. https://doi.org/10.1080/07350015.2023.2252039

Clark, T. E. (2011). Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, *29*, 327–341. https://doi.org/10.1198/jbes.2010.09248

Clark, T. E., Huber, F., Koop, G., Marcellino, M., & Pfarrhofer, M. (2023). Tail forecasting with multivariate Bayesian additive regression trees. *International Economic Review*, *64*, 979–1022. https://doi.org/10.1111/iere.12619

Clark, T. E., & Ravazzolo, F. (2015). Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, *30*, 551–575. https://doi.org/10.1002/jae.2379

Cogley, T., & Sargent, T. J. (2005). Drifts and volatilities: Monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, *8*, 262–302. https://doi.org/10.1016/j.red.2004.10.009

Cross, J. L., Hou, C., & Poon, A. (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, *36*, 899–915. https://doi.org/10.1016/j.ijforecast.2019.10.002

D'Agostino, A., Gambetti, L., & Giannone, D. (2013). Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, *28*, 82–101. https://doi.org/10.1002/jae.1257

Doan, T. R., Litterman, B. R., & Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, *3*, 1–100. https://doi.org/10.1080/07474938408800053

Eisenstat, E., Chan, J. C. C., & Strachan, R. W. (2016). Stochastic model specification search for time-varying parameter VARs. *Econometric Reviews*, *35*, 1638–1665. https://doi.org/10.1080/07474938.2015.1092808

Fagan, G., Henry, J., & Mestre, R. (2005). An area-wide model for the euro area. *Economic Modelling*, *22*, 39–59. https://doi.org/10.1016/j.econmod.2004.04.002

Follett, L., & Yu, C. (2019). Achieving parsimony in Bayesian vector autoregressions with the horseshoe prior. *Econometrics and Statistics*, *11*, 130–144. https://doi.org/10.1016/j.ecosta.2018.12.004

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, *15*, 183–202. https://doi.org/10.1111/j.1467-9892.1994.tb00184.x

Frühwirth-Schnatter, S., & Wagner, H. (2010). Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics*, *154*, 85–100. https://doi.org/10.1016/j.jeconom.2009.07.003

George, E. I., & McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, *88*, 881–889. https://doi.org/10.1080/01621459.1993.10476353

George, E. I., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, *142*, 553–580. https://doi.org/10.1016/j.jeconom.2007.08.017

Geweke, J., & Amisano, G. (2010). Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting*, *26*, 216–230. https://doi.org/10.1016/j.ijforecast.2009.10.007

Granger, C. W. (2008). Non-linear models: Where do we go next – Time varying parameter models? *Studies in Nonlinear Dynamics and Econometrics*, *12*. https://doi.org/10.2202/1558-3708.1639

Gruber, L., & Kastner, G. (2022). Forecasting macroeconomic data with Bayesian VARs: Sparse or dense? It depends! arXiv preprint arXiv:2206.04902. https://arxiv.org/abs/2206.04902

Hörmann, W., & Leydold, J. (2013). Generating generalized inverse Gaussian random variates. *Statistics and Computing*, *24*, 1–11. https://doi.org/10.1007/s11222-013-9387-3

Hosszejni, D., & Kastner, G. (2021). Modeling univariate and multivariate stochastic volatility in R with stochvol and factorstochvol. *Journal of Statistical Software*, *100*, 1–34. https://doi.org/10.18637/jss.v100.i12

Huber, F., & Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics*, *37*, 27–39. https://doi.org/10.1080/07350015.2016.1256217

Huber, F., Koop, G., & Onorante, L. (2021). Inducing sparsity and shrinkage in time-varying parameter models. *Journal of Business & Economic Statistics*, *39*, 669–683. https://doi.org/10.1080/07350015.2020.1713796

Kalli, M., & Griffin, J. E. (2014). Time-varying sparsity in dynamic regression models. *Journal of Econometrics*, *178*, 779–793. https://doi.org/10.1016/j.jeconom.2013.10.012

Kastner, G. (2016). Dealing with stochastic volatility in time series using the R package stochvol. *Journal of Statistical Software*, *69*, 1–30. https://doi.org/10.18637/jss.v069.i05

Kastner, G., & Frühwirth-Schnatter, S. (2014). Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis*, *76*, 408–423. https://doi.org/10.1016/j.csda.2013.01.002

Kastner, G., & Huber, F. (2020). Sparse Bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*, *39*, 1142–1165. https://doi.org/10.1002/for.2680

Koop, G., & Korobilis, D. (2012). Forecasting inflation using dynamic model averaging. *International Economic Review*, *53*, 867–886. https://doi.org/10.1111/j.1468-2354.2012.00704.x

Koop, G., Leon-Gonzalez, R., & Strachan, R. W. (2009). On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control*, *33*, 997–1017. https://doi.org/10.1016/j.jedc.2008.11.003

Korobilis, D. (2013). Hierarchical shrinkage priors for dynamic regressions with many predictors. *International Journal of Forecasting*, *29*, 43–59. https://doi.org/10.1016/j.ijforecast.2012.05.006

Leydold, J., & Hörmann, W. (2017). GIGrvg: Random variate generator for the GIG distribution. The package version (note) is: R package version 0.5.

Litterman, R. B. (1986). Forecasting with Bayesian vector autoregressions – Five years of experience. *Journal of Business*

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and Economic Statistics, *4*, 25–38. https://doi.org/10.1080/07350015.1986.10509491

McCracken, M. W., & Ng, S. (2016). FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics*, *34*, 574–589. https://doi.org/10.1080/07350015.2015.1086655

Onorante, L., & Raftery, A. E. (2016). Dynamic model averaging in large model spaces using dynamic Occam's window. *European Economic Review*, *81*, 2–14. https://doi.org/10.1016/j.euroecorev.2015.07.013

Polson, N. G., & Scott, J. G. (2010). Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics*, *9*, 501–538. https://doi.org/10.1093/acprof:oso/9780199694587.003.0017

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *Review of Economic Studies*, *72*, 821–852. https://doi.org/10.1111/j.1467-937X.2005.00353.x

Raftery, A. E., Karny, M., & Ettler, P. (2010). Online prediction under model uncertainty via dynamic model averaging: Application to a cold rolling mill. *Technometrics*, *52*, 52–66. https://doi.org/10.1198/TECH.2009.08104

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, *36*, 975–1000. https://doi.org/10.1016/0014-2921(92)90041-T

Sims, C. A., & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, *39*, 949–968. https://doi.org/10.2307/2527347

Thomas, R., Hills, S., & Dimsdale, N. (2010). The UK recession in context – What do three centuries of data tell us? *Bank of England Quarterly Bulletin*, *50*, 277–291.

Zhang, Y. D., Naughton, B. P., Bondell, H. D., & Reich, B. J. (2022). Bayesian regression using a prior on the model fit: The R2-D2 shrinkage prior. *Journal of the American Statistical Association*, *117*, 862–874. https://doi.org/10.1080/01621459.2020.1825449

AUTHOR BIOGRAPHIES

Martin Feldkircher holds the Chair in International Economics at the Vienna School of International Studies (Diplomatische Akademie Wien – DA). He graduated from the Institute for Advanced Studies in Vienna (Economics Program), completed his PhD at the University of Innsbruck, and holds a venia docendi from the Department of Economics at Vienna University of Business and Economics. Before joining the DA in October 2020, he spent 12 years in the Austrian central bank's Foreign Research Division. His research interests are in the fields of international macroeconomics, monetary policy, and Bayesian econometrics.

Luis Gruber is a research assistant at the Department of Statistics at the University of Klagenfurt with a focus on Bayesian modeling of time series.

Florian Huber is a professor of economics at the University of Salzburg, specializing in Bayesian time series econometrics and its applications in macroeconomics and finance. Before joining the University of Salzburg, he served as an assistant professor at WU Vienna and as an economist at the Oesterreichische Nationalbank. Florian is a scientific consultant to international organizations like the European Commission and the European Central Bank. He holds a PhD and a venia docendi (habilitation) in economics from WU Vienna.

Gregor Kastner is a professor of statistics at the University of Klagenfurt and works on the Bayesian modeling of time series and spatiotemporal data with a focus on computational aspects. Before joining the University of Klagenfurt, he served as an assistant professor and senior lecturer at WU Vienna. He holds a PhD in Applied Mathematics from JKU Linz and several graduate degrees from TU Wien and the University of Vienna.

How to cite this article: Feldkircher, M., Gruber, L., Huber, F., & Kastner, G. (2024). Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients in Bayesian vector autoregressions? *Journal of Forecasting*, *43*(6), 2126–2145. https://doi.org/10.1002/for.3121

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

APPENDIX A: DATA OVERVIEW

TABLE A1 Data overview.

Variable	EA	UK	US	m = 3	m = 7	m = 15	T-code
Real gross domestic product	YER, GDP at market prices, chain linked volumes, sa	GDP at market prices.	GDPC96, bn of chained 2009 USD	x	x	x	1
Prices	YED, GDP deflator index	GDP deflator at market prices	GDPCTPI, chain-type price index	x	x	x	1
Short-term interest rates	STN, 3-month euribor in % p.a.	Bank rate, avg. of monthly series	FEDFUNDS, eff. federal funds rate in %	x	x	x	2
Investment	ITR, gross fixed capital formation, chain linked volume, sa	Gross fixed capital formation, chained volume measure	GPDIC96, bn of chained 2009 USD		x	x	1
Consumption	PCR, individual consumption expenditure, chain linked volume, sa	Household consumption, chained volume measure	PCECTPI, personal consumption expenditures: chain-type price index		x	x	1
Exchange rate	EEN, nominal eff. Exchange rate vis-a-vis 38 non-euro area trading partners	Nominal dollar per pound exchange rate	EXUSUKx, dollar pound foreign exchange rate		x	x	1
Employment/ unemployment	URX, percentage of civilian workforce, sa	Employment in heads	UNRATE, civilian unempl. rate in %		x	x	2
Wages	WRN, wages per head	Spliced average weekly earnings series	CES2000000008x, real avg. hourly earnings			x	1
						x	
Money	M2, sa, OECD data	Stock of "broad money", break-adjusted	M2REALx, real M2 money stock (bn of 1982–1984 USD), deflated by CPI			x	1
Equity prices	Equity share price index, OECD data	Spliced monthly share price index weighted by market capitalization	S&P 500 common composite stock price index			x	1
Long-term interest rate	LTN, 10-year gov. bond yields in % p.a.	10-year gov. bond yields in % p.a.	GS10, 10-year gov. bond yields in %			x	2
Government consumption	GCR, general gov. expenditure, chain linked volumes, sa	Gov. consumption, chained volume measure	GCEC96, real gov. spending, in bn of 2009 chained USD			x	1
Exports	XTR, exports of goods and services, chain linked volumes, sa	Export volumes, chained volume measures.	EXPGSC96, bn of chained 2009 USD			x	1
Commodity price index	COMPR, weighted sum of oil prices and non-oil commodity prices in US dollars					x	1
Labour productivity	LPROD, ratio of real GDP and total employment					x	1

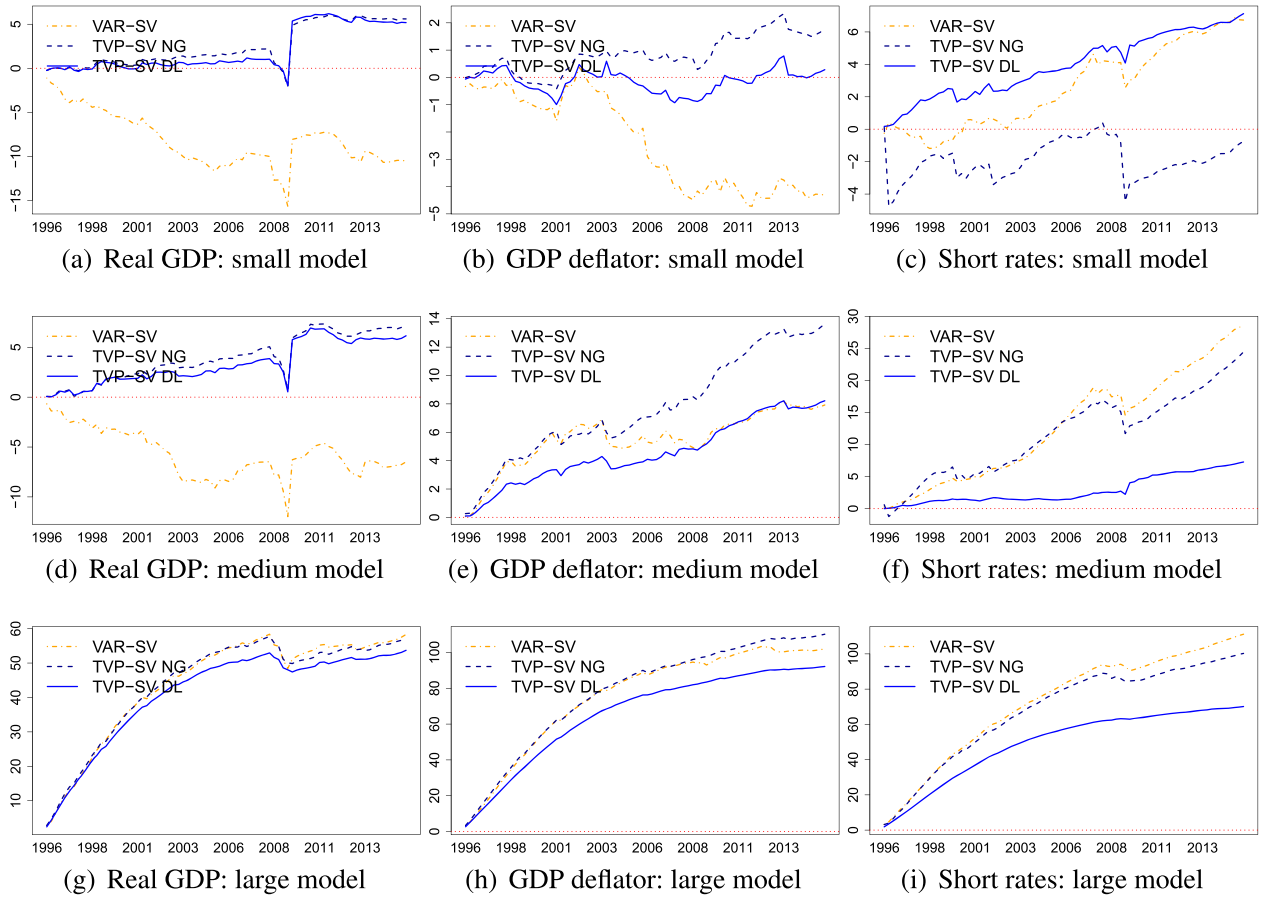
1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A1 (Continued)

Variable	EA	UK	US	m = 3	m = 7	m = 15	T-code
Mortgage rate		Household var. mortg. rate in % p.a.				x	2
Corporate bond spread		Spliced interpolated corporate bond spreads				x	2
Hours worked			AWHMAN, avg. weekly hours of production and nonsup. employees: manufacturing			x	1
Consumer sentiment			UMCSENTx, consumer sentiment index			x	1

Notes: Data are obtained from https://eabcn.org/page/area-wide-model for the euro area, from https://www.bankofengland.co.uk/statistics/research-datasets for the United Kingdom, and from https://research.stlouisfed.org/econ/mccracken/fred-databases/ for the United States. For the euro area, equity prices and M2 obtained from https://data.oecd.org/; employment data for UK transformed using log differences. Transformation codes (T-codes) are as follows: 1, log differences; 2, raw data.

APPENDIX B: ADDITIONAL EMPIRICAL RESULTS



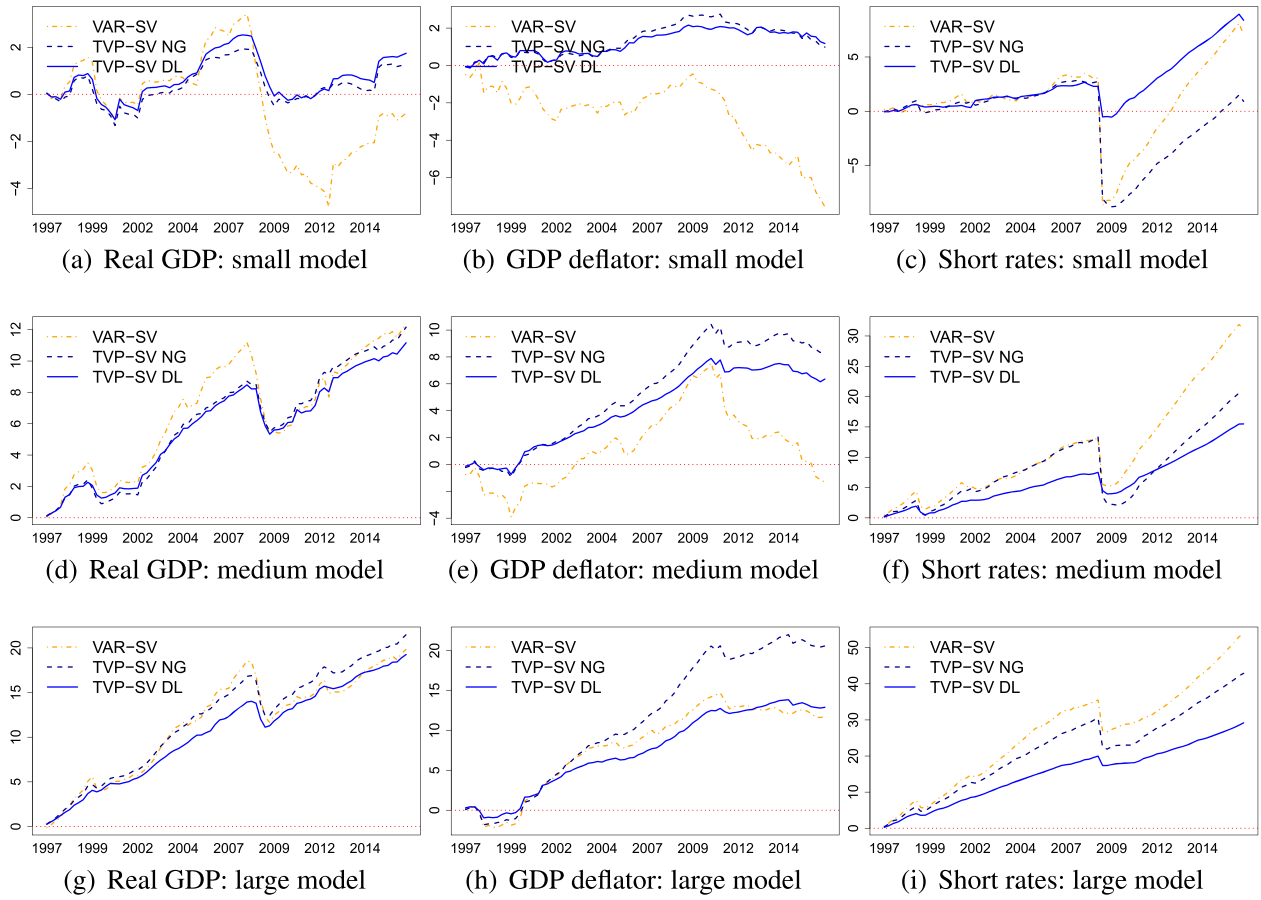
(a) Real GDP: small model (b) GDP deflator: small model (c) Short rates: small model

(d) Real GDP: medium model (e) GDP deflator: medium model (f) Short rates: medium model

(g) Real GDP: large model (h) GDP deflator: large model (i) Short rates: large model

FIGURE A1 Euro area (EA): Univariate cumulative log predictive one-quarter-ahead Bayes factors over time relative to the TVP-SV-VAR with loose shrinkage.

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



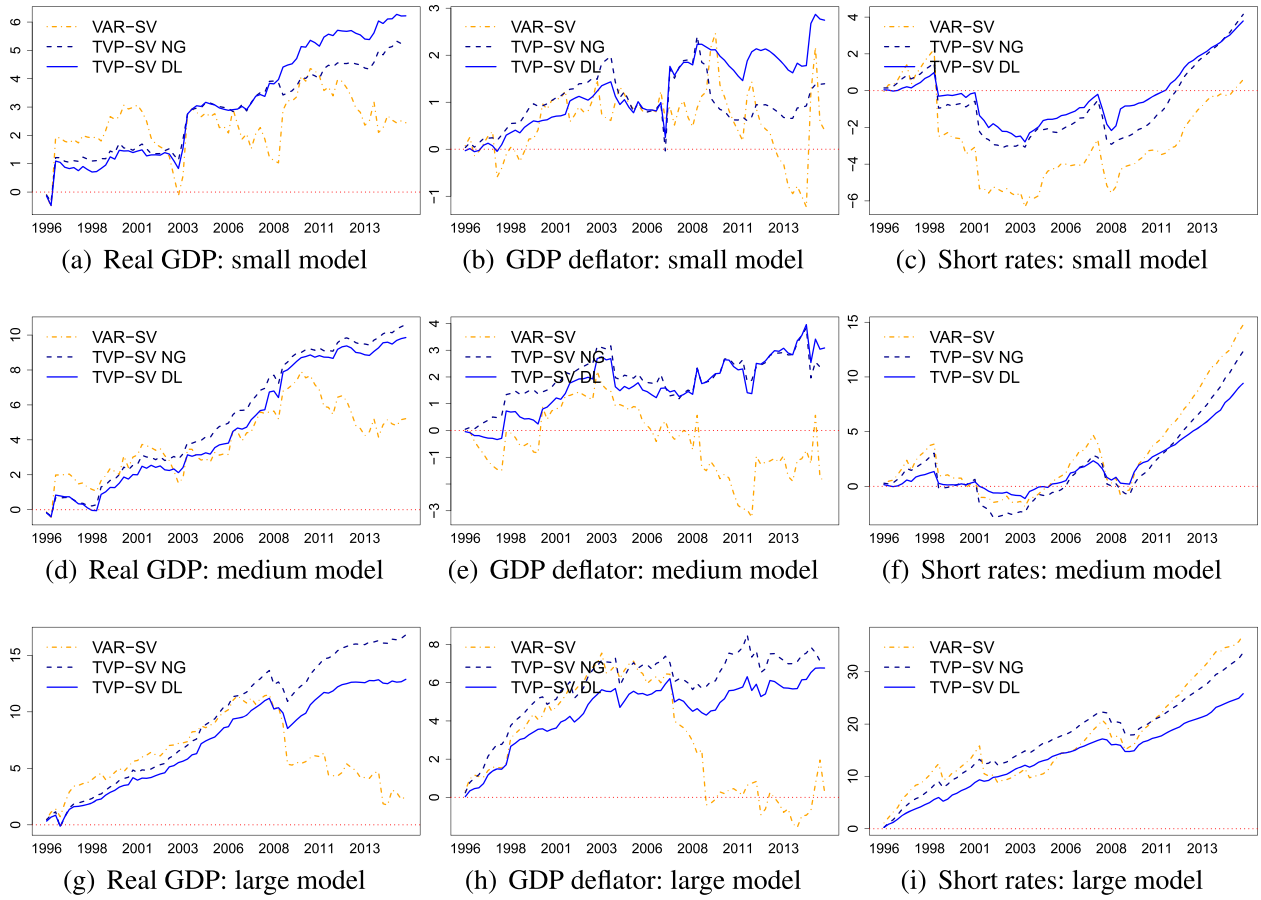
(a) Real GDP: small model (b) GDP deflator: small model (c) Short rates: small model

(d) Real GDP: medium model (e) GDP deflator: medium model (f) Short rates: medium model

(g) Real GDP: large model (h) GDP deflator: large model (i) Short rates: large model

FIGURE A2 United Kingdom (UK): Univariate cumulative log predictive one-quarter-ahead Bayes factors over time relative to the TVP-SV-VAR with loose shrinkage.

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License



(a) Real GDP: small model (b) GDP deflator: small model (c) Short rates: small model

(d) Real GDP: medium model (e) GDP deflator: medium model (f) Short rates: medium model

(g) Real GDP: large model (h) GDP deflator: large model (i) Short rates: large model

FIGURE A3 United States (US): Univariate cumulative log predictive one-quarter-ahead Bayes factors over time relative to the TVP-SV-VAR with loose shrinkage.

We display forecasting performances for three economies (EA, UK, US) and different model specifications for three key variables (real GDP, GDP deflator, short rates) in Figures A1–A3.

1099131x, 2024, 6, Downloaded from https://onlinelibrary.wiley.com/doi/10.1002/for.3121 by CochraneAustria, Wiley Online Library on [30/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License