Department of Economics
Working Paper No. 260

Sophisticated and small versus simple and sizeable: When does it pay off to introduce drifting coefficients

Martin Feldkircher
Florian Huber
Gregor Kastner

January 2018

WIRTSCHAFTS UNIVERSITÄT WIEN VIENNA UNIVERSITY OF ECONOMICS AND BUSINESS

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

Martin Feldkircher[1], Florian Huber*[2], and Gregor Kastner[2]

[1]Oesterreichische Nationalbank (OeNB)
[2]WU Vienna University of Economics and Business

Abstract

We assess the relationship between model size and complexity in the time-varying parameter 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 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: Global-local shrinkage priors, density predictions, hierarchical modeling, stochastic volatility, dynamic model selection

JEL Codes: C11, C30, C53, E52.

*Corresponding author: Florian Huber, WU Vienna University of Economics and Business. E-mail: fhuber@wu.ac.at.

1 Introduction

In contemporary econometrics, 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, small 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 in the popular time-varying parameter vector autoregressive framework with stochastic volatility (TVP-VAR-SV). Our conjecture is that the introduction of drifting coefficients can control for an omitted variable bias in small-scale models or conversely, larger information sets can substitute for non-linear model dynamics. Since recent research increasingly focuses on combining large models with non-linear model dynamics, appropriate solutions to combine the best of both worlds are needed to avoid overfitting and decreased predictive power.

Within a Bayesian framework, it is thus necessary to develop suitable shrinkage priors for the TVP-VAR-SV case that overcome issues related to overfitting. In this paper we exploit the non-centered parameterization of the state space model (see Frühwirth-Schnatter and Wagner, 2010) to disentangle the time-invariant component of the model from the dynamic part.[1] Shrinkage is achieved by modifying two global-local shrinkage priors to accommodate features of the Minnesota prior (Doan et al., 1984; Sims and Zha, 1998). The first specification proposed is a modified version of the Normal-Gamma (NG) shrinkage prior (Griffin and Brown, 2010; 2017; Bitto and Frühwirth-Schnatter, 2016) while the second version modifies the recent Dirichlet-Laplace (DL) shrinkage prior (Bhattacharya et al., 2015) to cater for lag-wise sh rinkage (Huber and Feldkircher, 2017). Both priors proposed combine recent advances on Bayesian VARs (Korobilis and Pettenuzzo, 2016) with the literature on infinite dimensional factor models (Bhattacharya and Dunson, 2011). Our prior controls for model uncertainty by pushing higher lag orders dynamically towards zero and in the same step applies shrinkage on the time-variation of the autoregressive coefficients and covariance parameters. Loosely speaking, we introduce a lag-specific shrinkage parameter that controls how much lags to include and to what extend the corresponding coefficients drift over time. This lag-specific shrinkage parameter is expected to grow at an undetermined rate, increasingly placing more mass around zero for coefficients associated with higher lags of endogenous variables. By contrast, the standard implementations

[1] For recent applications of this general modeling strategy within state space models, see Belmonte et al. (2014); Bitto and Frühwirth-Schnatter (2016); Eisenstat et al. (2016).

of the NG and the DL priors rely on a single global shrinkage parameter that pushes all coefficients to zero. To render computation feasible, we apply the algorithm put forward in Carriero et al. (2016) and estimate the TVP-VAR-SV on an equation-by-equation basis. This, in combination with the two proposed shrinkage priors, permits fast and reliable estimation of large-dimensional models.

In an empirical exercise, we examine the forecasting properties of the TVP-VAR-SV equipped with our proposed shrinkage priors using three well-known data sets for the Euro area (EA), the United Kingdom (UK) and the United States (US). We evaluate the merits of our model approach relative to a set of other forecasting models, most notably a constant parameter Bayesian VAR with SV and a TVP-VAR with a weakly informative shrinkage prior. Since the size of the information set could play a crucial role in assessing whether time-variation is necessary, we investigate for each data set a small model that features 3 variables, a moderately sized one with 7 variables and a large model with 15 variables. Our results are three-fold: 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 corrob orated 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.

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 stochastic volatility (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:[2]

$$\boldsymbol{y}_t = \boldsymbol{A}_{1t}\boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_{pt}\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t,\ \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}_M, \boldsymbol{\Sigma}_t). \tag{2.1}$$

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 Eq. (2.1) can be cast in a standard regression form as follows,

$$\boldsymbol{y}_t = \boldsymbol{A}_t\boldsymbol{x}_t + \boldsymbol{\varepsilon}_t, \tag{2.2}$$

with $\boldsymbol{A}_t = (\boldsymbol{A}_{1t}, \ldots, \boldsymbol{A}_{pt})$ being an $M \times (pM)$ matrix and $\boldsymbol{x}_t = (\boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_{t-p}')'$. Following Cogley and Sargent (2005) we can rewrite Eq. (2.2) as

$$\boldsymbol{y}_t - \boldsymbol{A}_t\boldsymbol{x}_t = \boldsymbol{H}_t\boldsymbol{\eta}_t,\ \text{with } \boldsymbol{\eta}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_t), \tag{2.3}$$

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

$$\tilde{\boldsymbol{H}}_t\boldsymbol{\varepsilon}_t = \boldsymbol{\eta}_t. \tag{2.4}$$

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

$$\varepsilon_{1t} = \eta_{1t}, \tag{2.5}$$

$$\tilde{h}_{21,t}\varepsilon_{1t} + \varepsilon_{2t} = \eta_{2t}, \tag{2.6}$$

with $\tilde{h}_{21,t}$ denoting the second element of the first column of $\tilde{\boldsymbol{H}}_t$. Eq. (2.6) can be rewritten as

$$y_{2t} = \boldsymbol{A}_{2\bullet,t}\boldsymbol{x}_t - \tilde{h}_{21,t}\varepsilon_{1t} + \eta_{2t}, \tag{2.7}$$

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

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,

$$y_{it} = \boldsymbol{A}_{i\bullet,t}\boldsymbol{x}_t - \sum_{s=1}^{i-1} \tilde{h}_{is,t}\varepsilon_{st} + \eta_{it}. \tag{2.8}$$

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

$$y_{it} = \boldsymbol{B}_{it}'\boldsymbol{z}_{it} + \eta_{it}. \tag{2.9}$$

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

$$\boldsymbol{B}_{it} = \boldsymbol{B}_{it-1} + \boldsymbol{v}_t, \quad \text{with } \boldsymbol{v}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega}_i), \tag{2.10}$$

where $\boldsymbol{\Omega}_i = \text{diag}(\omega_1, \dots, \omega_{K_i})$ 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 $(j = 1, \dots, K_i)$. 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 towards a constant parameter specification. To alleviate such concerns, Frühwirth-Schnatter and Wagner (2010) exploit the non-centered parameterization of Eqs. (2.9) and (2.10),

$$y_{it} = \boldsymbol{B}_{i0}'\boldsymbol{z}_{it} + \tilde{\boldsymbol{B}}_{it}'\sqrt{\boldsymbol{\Omega}_i}\boldsymbol{z}_{it} + \eta_{it}. \tag{2.11}$$

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 $\tilde{\boldsymbol{B}}_{it}$ has typical element j given by $\tilde{b}_{ij,t} = \frac{b_{ij,t} - b_{ij,0}}{\sqrt{\omega_{ij}}}$. The corresponding state equation is given by

$$\tilde{\boldsymbol{B}}_{it} = \tilde{\boldsymbol{B}}_{it-1} + \boldsymbol{u}_{it}, \quad \text{with } \boldsymbol{u}_{it} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{K_i}). \tag{2.12}$$

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 $\boldsymbol{z}_{it}$, i.e., 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 forecasti ng properties (D'Agostino et al., 2013; Korobilis, 2013; Belmonte et al., 2014; Bitto and Frühwirth-Schnatter, 2016)

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

$$v_{it} = \mu_i + \rho_i(v_{it-1} - \mu_i) + w_{it}, \quad w_{it} \sim \mathcal{N}(0, \sigma_i^2), \tag{2.13}$$

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 Eq. (2.10). 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 and Scott, 2010) on $\boldsymbol{B}_{i0}$ and $\boldsymbol{\omega}_i = (\omega_{i1}, \ldots, \omega_{iK_i})'$. A GL shrinkage prior comprises of a global scaling parameter that pushes all elements of the coefficient vector towards zero and a set of local scaling parameters that enable coefficient-specific deviations from this general pattern.

3.1 The Normal-Gamma shrinkage prior

The first prior we consider is a modified variant of the Normal-Gamma (NG) shrinkage prior proposed in Griffin and Brown (2010) and adopted within the general class of state space models in Bitto and Frühwirth-Schnatter (2016). 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 $j = 1, \dots, K = pM^2$. 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,

$$a_{0j}|\tau_{aj}^2, \lambda_l \sim \mathcal{N}(0, 2/\lambda_l\ \tau_{aj}^2), \quad \tau_{aj}^2 \sim \mathcal{G}(\vartheta_l, \vartheta_l) \tag{3.1}$$

$$\sqrt{\omega}_j|\tau_{\omega j}^2, \lambda_l \sim \mathcal{N}(0, 2/\lambda_l\ \tau_{\omega j}^2), \quad \tau_{\omega j}^2 \sim \mathcal{G}(\vartheta_l, \vartheta_l) \tag{3.2}$$

$$\lambda_l = \prod_{s=1}^{l} \nu_s,\ \ \nu_s \sim \mathcal{G}(c_\lambda, d_\lambda), \tag{3.3}$$

where τ_{aj}^2 and $\tau_{\omega j}^2$ denote a set of local scaling parameters that follow a Gamma distribution and λ_l is a lag-specific shrinkage parameter. Thus, if the jth element of $\boldsymbol{a}_0$ is related to the lth lag of the endogenous variables, λ_l applies a lag-specific degree of shrinkage to all coefficients associated to $\boldsymbol{y}_{t-l}$ as well as the corresponding standard deviations $\sqrt{\omega}_j$. The hyperparameter $\vartheta_l = \vartheta/l^2$ also depends on the lag length of the system and controls the excess kurtosis of the marginal prior,

$$p(a_{0j}|\lambda_l) = \int p(a_{0j}|\tau_{aj}^2, \lambda_l) d\tau_{aj}^2, \tag{3.4}$$

obtained after integrating out the local scaling parameters. For the marginal prior, λ_l controls the overall degree of shrinkage. Lower values of ϑ_l place increasing prior mass on zero while at the same time lead to heavy tails of $p(a_{0j}|\lambda_l)$. Thus, our specification implies that with increasing lag length we increasingly place more mass on zero while maintaining heavy tails.

In our case, we specify λ_l to be a lag-wise shrinkage parameter that follows a multiplicative Gamma process proposed in Bhattacharya and Dunson (2011),[3] with c_λ and d_λ denoting hyperparameters. As long as ν_s exceeds unity, this prior stochastically introduces more shrinkage for higher lag orders. Note that λ_l simultaneously pulls all elements in $\boldsymbol{a}_0$ associated with the lth lag and the corresponding $\sqrt{\omega}_j$s to zero. This implies that if a given lag of the endogenous variables is not included in the model, time-variation is also less likely. However, it could be the case that a given element in $\boldsymbol{a}_0$ associated with a higher lag

[3] See Korobilis (2014) for a recent application of a similar idea to the TVP-VAR-SV case.

order might be important to explain $\boldsymbol{y}_t$. In that case, the local scaling parameters introduce sufficient flexibility to pull sufficient posterior mass away from zero, enabling non-zero regression sign als if necessary.

On the covariance parameters $\tilde{h}_{is,0}$ $(i = 2, \ldots, M; s = pM+1, \ldots, K_i)$ 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 $\tilde{\boldsymbol{h}}_0$ 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 $\tilde{h}_{i0}$ and γ_i,

$$\tilde{h}_{i0}|\tau_{hi}^2, \varpi \sim \mathcal{N}(0, 2/\varpi\ \tau_{hi}^2), \quad \tau_{hi}^2 \sim \mathcal{G}(\vartheta_h, \vartheta_h), \tag{3.5}$$

$$\gamma_i|\tau_{\gamma i}^2, \varpi \sim \mathcal{N}(0, 2/\varpi\ \tau_{\gamma i}^2), \quad \tau_{\gamma i}^2 \sim \mathcal{G}(\vartheta_h, \vartheta_h), \tag{3.6}$$

$$\varpi \sim \mathcal{G}(c_\varpi, d_\varpi). \tag{3.7}$$

Here, τ_{hi}^2 and $\tau_{\gamma i}^2$ are local scaling parameters and ϖ is a global shrinkage parameter that pushes all covariance parameters and the corresponding state innovation standard deviations across equations to zero. The hyperparameter ϑ_h again controls the excess kurtosis of the marginal prior.

Note that this prior also captures several features of the Minnesota prior (Doan et al., 1984; Sims and Zha, 1998) since it captures the notion that more distant lags appear to be less relevant to predict the current value of $\boldsymbol{y}_t$. However, as opposed to the deterministic penalty function on higher lag orders introduced in a standard Minnesota prior our model specification entails an increasing degree of shrinkage in a stochastic manner, effectively allowing for deviations if the data suggests it.

3.2 The Dirichlet-Laplace 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 and McCulloch, 1993; George et al., 2008) 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 Dirichlet-Laplace (DL) shrinkage prior, excels both in 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,

$$a_{0j}|\psi_{aj}, \xi_{aj}^2, \tilde{\lambda}_l \sim \mathcal{N}(0, \psi_{aj}\xi_{aj}^2/\tilde{\lambda}_l^2), \quad \psi_{aj} \sim Exp(1/2), \quad \xi_j \sim Dir(n_a, \dots, n_a), \tag{3.8}$$

$$\sqrt{\omega}_j|\psi_{\omega j}, \xi_{\omega j}^2, \tilde{\lambda}_l \sim \mathcal{N}(0, \psi_{\omega j}\xi_{\omega j}^2/\tilde{\lambda}_l^2), \quad \psi_{\omega j} \sim Exp(1/2), \quad \xi_j \sim Dir(n_a, \dots, n_a), \tag{3.9}$$

$$\tilde{\lambda}_l = \prod_{s=1}^{l} \tilde{\nu}_s, \ \tilde{\nu}_s \sim \mathcal{G}(c_\lambda, d_\lambda). \tag{3.10}$$

Hereby, for $s \in \{a, \omega\}$, ψ_{sj} is again a set of local scaling parameters and ξ_{sj} constitutes an auxiliary scaling parameter defined on the $(K-1)$-dimensional unit simplex $\mathcal{S}^{K-1} = \{\boldsymbol{x} = (x_1, \dots, x_K)' : x_j \geq 0, \sum_{j=1}^{K} x_j = 1\}$ with $\boldsymbol{\xi}_s = (\xi_{s1}, \dots, \xi_{sK})'$. The lag-specific shrinkage parameter $\tilde{\lambda}_l$ is defined analogously to the NG prior. Our specification of the global-shrinkage parameter differs from the original implementation by assuming that $\tilde{\lambda}_l$ is applied to a subset of the regression coefficients only; the original variant of the prior features one single global shrinkage coefficient. The parameter n_a controls the overall tightness of the prior. Bhattacharya et al. (2015) show that if $n_a = K^{-(1+\epsilon)}$ for ϵ close to zero, the corresponding prior displays excellent theoretical shrinkage properties.

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

$$\tilde{h}_{i0}|\psi_{hi}, \xi_{hi}^2, \tilde{\varpi} \sim \mathcal{N}(0, \psi_{hi}\xi_{hi}^2/\tilde{\varpi}^2), \quad \psi_{hi}^2 \sim Exp(1/2), \quad \xi_{hi} \sim Dir(n_h, \dots, n_h), \tag{3.11}$$

$$\gamma_i|\psi_{\gamma i}, \xi_{\gamma i}^2, \tilde{\varpi} \sim \mathcal{N}(0, \psi_{\gamma i}\xi_{\gamma i}^2/\tilde{\varpi}^2), \quad \psi_{\gamma i}^2 \sim Exp(1/2), \quad \xi_{\gamma i} \sim Dir(n_h, \dots, n_h), \tag{3.12}$$

$$\tilde{\varpi} \sim \mathcal{G}^{-1}(2 v n_h, 1/2). \tag{3.13}$$

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 $\tilde{\varpi}$ denote a global shrinkage parameter 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 data generating process. 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 Markov chain Monte Carlo algorithm

Our algorithm is related to the MCMC scheme put forward in Carriero et al. (2016) and estimates the latent states on an equation-by-equation basis. Specifically, conditional on a suitable set of initial conditions, the algorithm cycles through the following steps:

1. Draw $(\boldsymbol{B}_{i0}', \omega_{i1}, \dots, \omega_{iK_i})'$ for $i = 1, \dots, M$ from $\mathcal{N}(\boldsymbol{\mu}_{Bi}, \boldsymbol{V}_i)$ with $\boldsymbol{V}_i = (\boldsymbol{Z}_i'\boldsymbol{Z}_i + \underline{\boldsymbol{V}}_i^{-1})^{-1}$ 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 $[\boldsymbol{z}_{it}', (\boldsymbol{B}_{it} \odot \boldsymbol{z}_{it})']\, e^{-(v_{it}/2)}$, $\boldsymbol{Y}_i$ 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. (2016) who sample the VAR parameters in $\boldsymbol{A}_0$ and the elements of $\tilde{\boldsymbol{H}}_0$ conditionally on each other, we propose to draw these jointly which speeds up the mixing of the sampler.

2. Simulate the full history of $\{\tilde{\boldsymbol{B}}_{it}\}_{t=1}^{T}$ by means of a forward filtering backward sampling algorithm (see Carter and Kohn, 1994; Frühwirth-Schnatter, 1994) per equation.

3. The log-volatilities and the corresponding parameters of the state equation in Eq. (2.13) are simulated using the algorithm put forward in Kastner and Frühwirth-Schnatter (2014) via the R package stochvol (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 Full conditional posterior distributions associated with the NG prior

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

$$\tau_{aj}^2 | \bullet \sim \mathcal{GIG}(\vartheta_l - 1/2, a_{0j}^2, \vartheta_l \lambda_l), \quad \tau_{\omega j}^2 | \bullet \sim \mathcal{GIG}(\vartheta_l - 1/2, \omega_j^2, \vartheta_l \lambda_l), \tag{4.1}$$

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 and Hörmann, 2017).

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

$$\tau_{hi}^2 | \bullet \sim \mathcal{GIG}(\vartheta_h - 1/2, \tilde{h}_{hi}^2, \vartheta_h \varpi), \quad \tau_{\gamma i}^2 | \bullet \sim \mathcal{GIG}(\vartheta_h - 1/2, \gamma_i^2, \vartheta_h \varpi). \tag{4.2}$$

Concerning the sampling of ν_l, note that combining each component of the Gamma likelihood given by $p(\tau_{aj}^2, \tau_{\omega j}^2 | \nu_l, \lambda_{l-1}) = p(\tau_{aj}^2 | \nu_l, \lambda_{l-1}) \times p(\tau_{\omega j}^2 | \nu_l, \lambda_{l-1})$ with the Gamma prior $p(\nu_l)$ yields a conditional posterior that itself follows a Gamma distribution,

$$\nu_1 | \bullet \sim \mathcal{G}\left\{c_\lambda + 2\vartheta_1 M^2, d_\lambda + \frac{\vartheta_1}{2} \sum_{j \in \mathcal{A}_1} (\tau_{aj}^2 + \tau_{\omega j}^2)\right\} \quad \text{for } l = 1, \tag{4.3}$$

where $\mathcal{A}_1$ denotes an index set that allows selecting all elements in $\boldsymbol{A}_0$ and $\sqrt{\boldsymbol{\Omega}}$ associated with the first lag of the endogenous variables. For lags $l > 1$, the conditional posterior is also Gamma distributed,

$$\nu_l | \lambda_{l-1}, \bullet \sim \mathcal{G}\left\{c_\lambda + 2\vartheta_l M^2, d_\lambda + \lambda_{l-1} \frac{\vartheta_l}{2} \sum_{j \in \mathcal{A}_l} (\tau_{aj}^2 + \tau_{\omega j}^2)\right\} \quad \text{for } l > 1. \tag{4.4}$$

Likewise, the conditional posterior of ϖ is given by

$$\varpi|\bullet \sim \mathcal{G}\left\{c_\varpi + 2\vartheta_h v, d_\varpi + \frac{\vartheta_h}{2}\sum_{i=1}^{v}(\tau_{hj}^2 + \tau_{\gamma j}^2)\right\}. \tag{4.5}$$

4.3 Full conditional posterior distributions associated with the DL prior

We start by outlining the conditional posterior distribution of ψ_{aj}. Similar to the NG case, Bhattacharya et al. (2015) show that ψ_{aj} and $\psi_{\omega j}$ follow a GIG distribution,

$$\psi_{aj}|\bullet \sim \mathcal{GIG}(1/2, |a_{j0}|\tilde{\lambda}_l/\xi_{aj}, 1), \quad \psi_{\omega j}|\bullet \sim \mathcal{GIG}(1/2, |\sqrt{\omega}_j|\tilde{\lambda}_l/\xi_{\omega j}, 1). \tag{4.6}$$

For the Dirichlet components, the conditional posterior distribution is obtained by sampling a set of K auxiliary variables $N_{aj}, N_{\omega j}$ $(j = 1, \dots, K)$,

$$N_{aj}|\bullet \sim \mathcal{GIG}(n_a - 1, 2|a_{j0}|, 1), \quad N_{\omega j}|\bullet \sim \mathcal{GIG}(n_a - 1, 2|\sqrt{\omega}_{j0}|, 1). \tag{4.7}$$

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 lag-specific shrinkage parameters under the DL prior are obtained by stating the DL prior in its hierarchical form,

$$a_{0j}|\tilde{\lambda}_l \sim DE(\xi_{aj}/\tilde{\lambda}_l), \quad \xi_{aj} \sim Dir(n_a, \dots, n_a), \tag{4.8}$$

$$\sqrt{\omega}_j|\tilde{\lambda}_l \sim DE(\xi_{\omega j}/\tilde{\lambda}_l), \quad \xi_{\omega j} \sim Dir(n_a, \dots, n_a), \tag{4.9}$$

with $DE(\lambda)$ denoting the double exponential distribution whose density is proportional to $\lambda^{-1}e^{-|x|/\lambda}$. Using the same prior representation for $\sqrt{\omega}_j$ and noting that $p(a_{0j}, \sqrt{\omega}_j|\tilde{\lambda}_l, \xi_{aj}, \xi_{\omega j}) = p(a_{0j}|\tilde{\lambda}_l, \xi_{aj}, \xi_{\omega j}) \times p(\sqrt{\omega}_j|\tilde{\lambda}_l, \xi_{aj}, \xi_{\omega j})$ yields

$$\prod_{j \in \mathcal{A}_l} p(a_{0j}, \sqrt{\omega_j}|\tilde{\lambda}_l, \xi_{\omega j}, \xi_{aj}) = \tilde{\lambda}_l^{2M^2} \exp\left\{-\tilde{\lambda}_l \sum_{j \in \mathcal{A}_l}\left(\frac{|a_{0j}|}{\xi_{aj}} + \frac{|\sqrt{\omega}_j|}{\xi_{\omega j}}\right)\right\}. \tag{4.10}$$

Combining Eq. (4.10) with Eq. (3.10) for $l = 1$ leads to

$$p(\tilde{\nu}_1|\bullet) \propto \tilde{\nu}_1^{(c_\lambda + 2M^2) - 1} \exp\left\{-\left[d_\lambda + \sum_{j \in \mathcal{A}_1}\left(\frac{|a_{0j}|}{\xi_{aj}} + \frac{|\sqrt{\omega}_j|}{\xi_{\omega j}}\right)\right]\tilde{\nu}_1\right\}, \tag{4.11}$$

which is the kernel of a Gamma density $\mathcal{G}\left\{c_\lambda + 2M^2, d_\lambda + \sum_{j\in\mathcal{A}_1}\left(\frac{|a_{0j}|}{\xi_{aj}} + \frac{|\sqrt{\omega}_j|}{\xi_{\omega j}}\right)\right\}$. For higher lag orders $l > 1$ we obtain

$$p(\tilde{\nu}_l|\tilde{\lambda}_{l-1}, \bullet) \propto \tilde{\nu}_l^{(c_\lambda+2M^2)-1} \exp\left\{-\left[d_\lambda + \tilde{\lambda}_{l-1}\sum_{j\in\mathcal{A}_l}\left(\frac{|a_{0j}|}{\xi_{aj}} + \frac{|\sqrt{\omega}_j|}{\xi_{\omega j}}\right)\right]\tilde{\nu}_l\right\}, \tag{4.12}$$

i.e. $\mathcal{G}\left\{c_\lambda + 2M^2, d_\lambda + \tilde{\lambda}_{l-1}\sum_{j\in\mathcal{A}_l}\left(\frac{|a_{0j}|}{\xi_{aj}} + \frac{|\sqrt{\omega}_j|}{\xi_{\omega j}}\right)\right\}$.

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

$$\psi_{hi}|\bullet \sim \mathcal{GIG}\left(1/2, |\tilde{h}_{i0}|/(\tilde{\varpi}\zeta_{hi}), 1\right), \quad \psi_{\gamma i}|\bullet \sim \mathcal{GIG}\left(1/2, |\gamma_i|/(\tilde{\varpi}\zeta_{\gamma i}), 1\right). \tag{4.13}$$

Again, we introduce a set of auxiliary variables $N_{hi}, N_{\gamma i}$,

$$N_{hi}|\bullet \sim \mathcal{GIG}(n_h - 1, 2|\tilde{h}_{i0}|, 1), \quad N_{\gamma i}|\bullet \sim \mathcal{GIG}(n_h - 1, 2|\gamma_i|, 1), \tag{4.14}$$

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 final component is the global shrinkage parameter on the covariance parameters and the process innovation variances which again follow a GIG distribution,

$$\tilde{\varpi}|\bullet \sim \mathcal{GIG}\left\{2v(n_h - 1), 2\sum_{j=1}^{v}\left(\frac{|h_{i0}|}{\xi_{hi}} + \frac{|\gamma_i|}{\xi_{\gamma i}}\right), 1\right\}. \tag{4.15}$$

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 UK and the US. 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

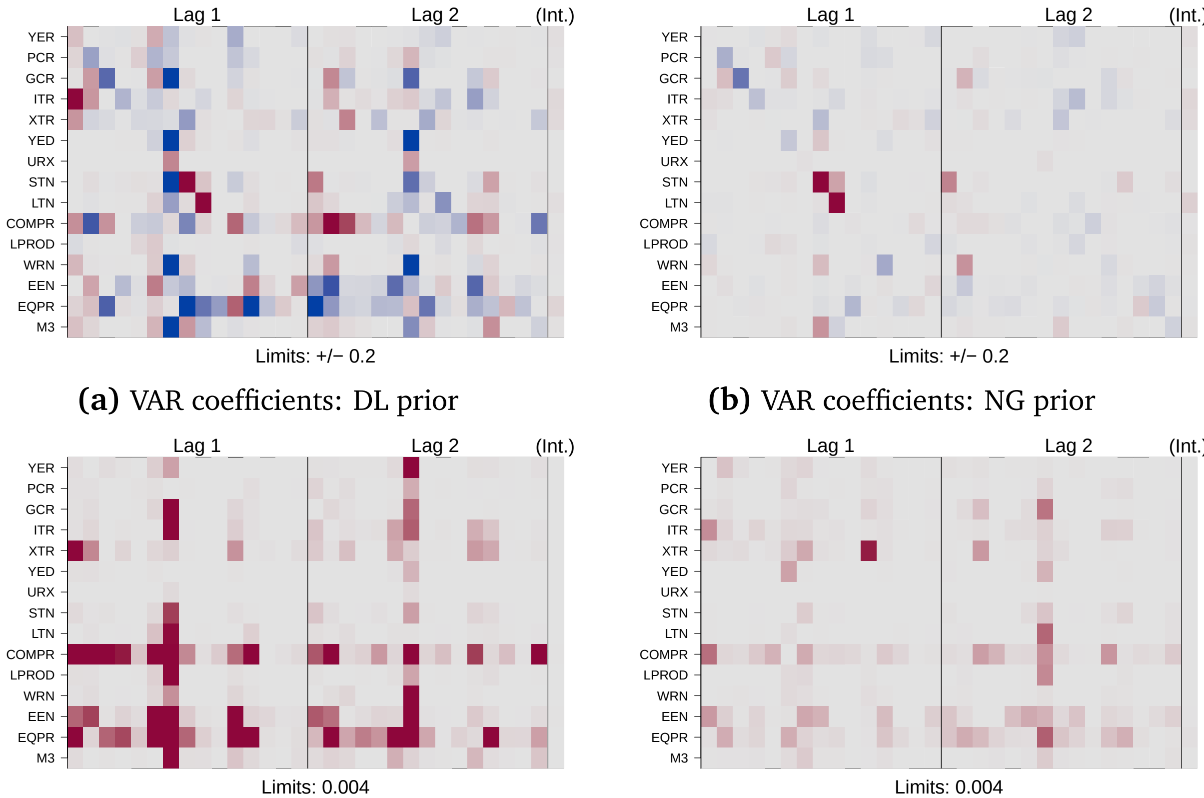
We use prominent macroeconomic data sets for the EA, the UK and the US. All three data sets are on a quarterly frequency but span different periods of time. For the euro area we take data from the area wide model (Fagan et al., 2001) and additionally include equity prices available from 1987Q1 to 2015Q4. UK data stem from the Bank of England's "A millenium of macroeconomic data" (Thomas et al., 2010) and covers the period from 1982Q2 to 2016Q4. For the US, we use a subset from the FRED QD data base (McCracken and Ng, 2016) which covers the period from 1959Q1 to 2015Q1.

For each of the three cases we use three subsets, a small (3 variables), a medium (7 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 US 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 to first 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 versus financial market dat a).

For data that are non-stationary we take first differences, see Table A.1 in the appendix for more details. Consistent with the literature (Cogley and Sargent, 2005; Primiceri, 2005; D'Agostino et al., 2013) we include $p = 2$ lags of the endogenous variables in all models.

Before proceeding to the empirical results, a brief word on the specific choice of the hyperparameters is in order. For the NG prior we set $\vartheta = \vartheta_h = 0.1$ and $c_\lambda = 1.5, d_\lambda = 1$. The first choice is motivated by recent empirical evidence provided in Huber and Feldkircher (2017) who integrate ϑ out of the joint posterior in a Bayesian fashion. The second choice is not critical empirically but serves to place sufficient prior mass on values of ν_s above unity.



(a) VAR coefficients: DL prior

(b) VAR coefficients: NG prior

(c) State innovation sds: DL prior

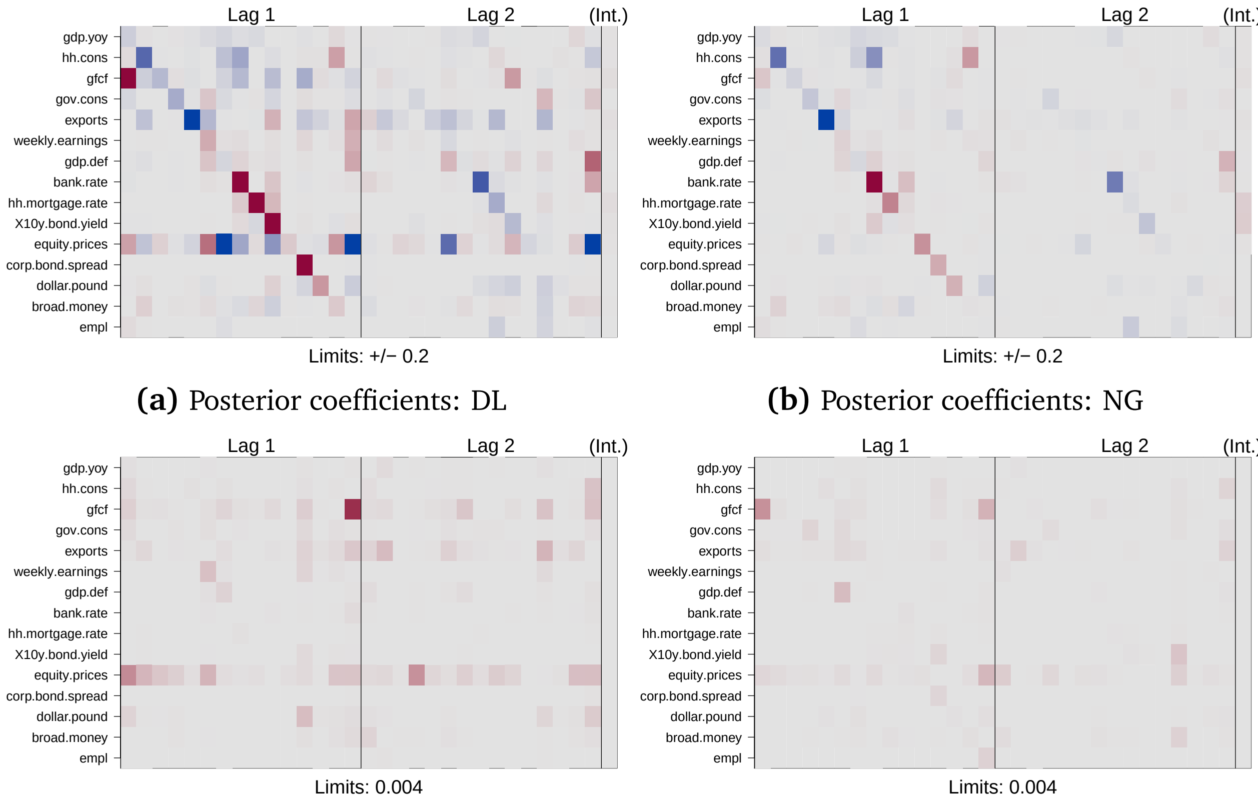
(d) State innovation sds: NG prior

Fig. 1: Posterior means in the large model – Euro area.

Moreover, we set $c_\varpi = d_\varpi = 0.01$ to induce heavy shrinkage on the covariance parameters. For the DL prior c_λ and d_λ are specified analogously to the NG case and $n_a = 1/K, n_h = 1/v$. Note that if n_a is set to larger values the degree of shrinkage is too small and the empirical performance of the DL prior becomes much worse.

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. Figs. 1 to 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.



(a) Posterior coefficients: DL

(b) Posterior coefficients: NG

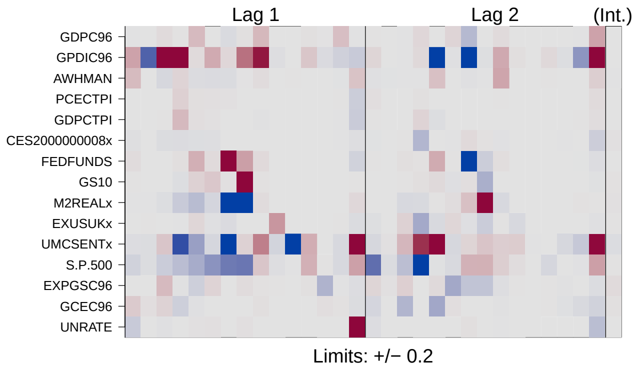
(c) State innovation sds: DL

(d) State innovation sds: NG

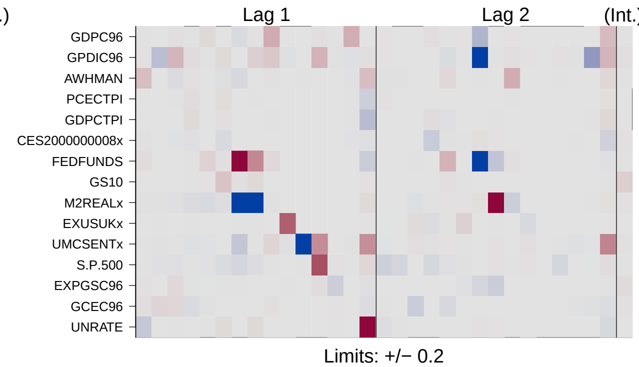
Fig. 2: Posterior means in the large model – UK.

We start by inspecting posterior sparsity attached to the time-invariant part of the models, provided in the upper panels of Figs. 1 to 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, i.e. 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 appear 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.

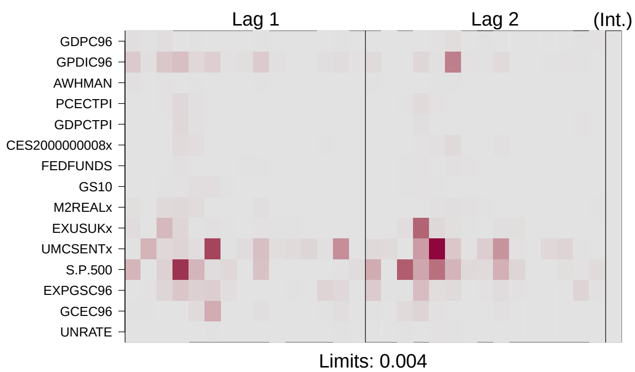
Equations that are characterized by a large amount of non-zero coefficients (i.e., dense rows) are mostly related to financial variables, namely exchange rates, equity and commodity prices. These observations are general in nature and relate to all three countries considered.



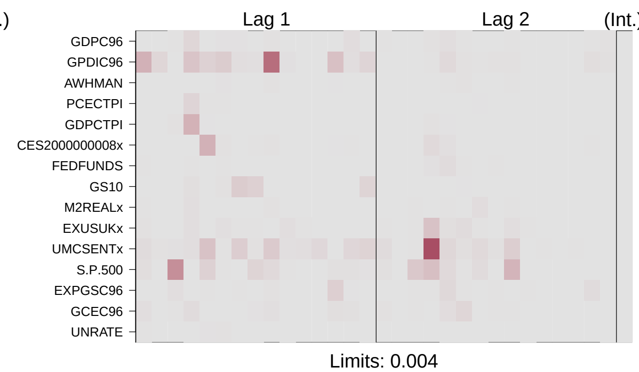
(a) Posterior coefficients: DL



(b) Posterior coefficients: NG



(c) State innovation sds: DL



(d) State innovation sds: NG

Fig. 3: Posterior means in the large model – US.

In the next step we investigate sparsity in terms of the degree of time variation of the VAR coefficients (see the lower panels of Figs. 1 to 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 euro area 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 findings point towards 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

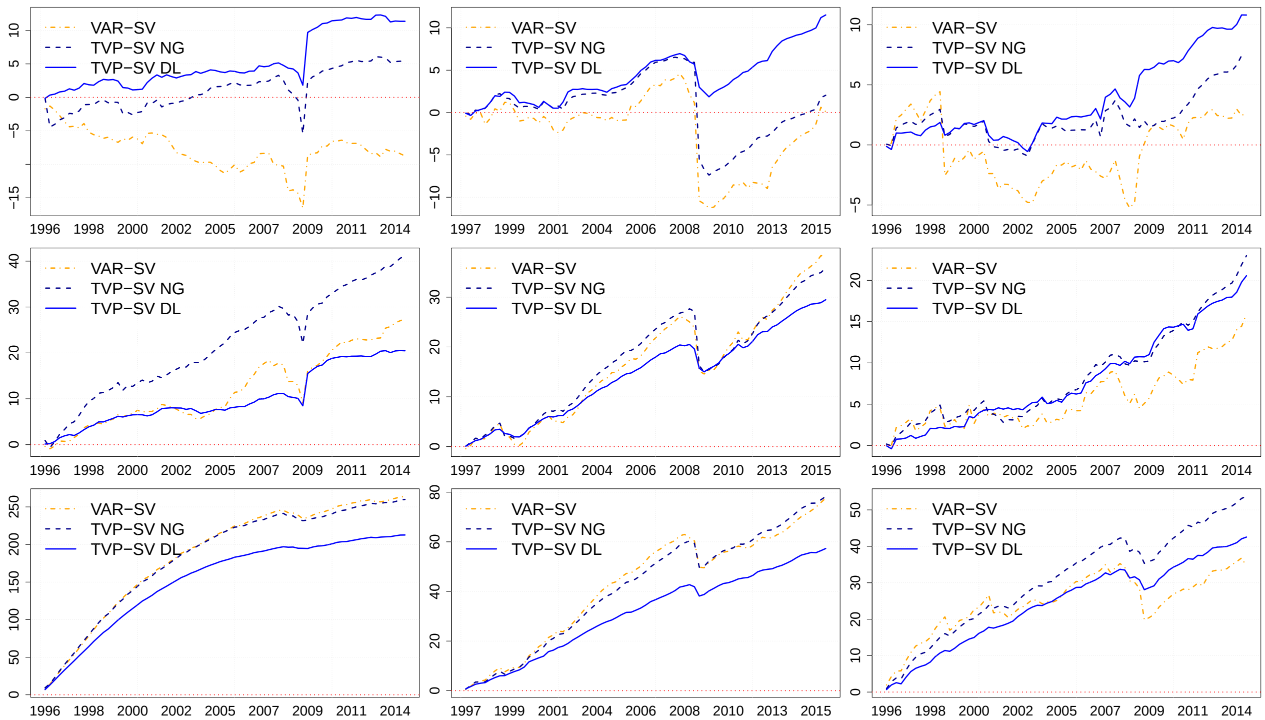
In this section we examine the forecasting performance of the proposed prior specifications. The forecasting set-up largely follows Huber and Feldkircher (2017) and focuses on the one-quarter and one-year ahead forecast horizons and three different information sets: small (3 variables), medium (7 variables) 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 UK and 1995Q4-2015Q3 for the US.

Forecasts are evaluated using log predictive scores (LPSs), a widely used metric to measure density forecast accuracy (see e.g., Geweke and Amisano, 2010). 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

$$\omega_j \sim \mathcal{G}(1/2, 1/2) \Leftrightarrow \pm\sqrt{\omega_j} \sim \mathcal{N}(0, 1). \tag{5.1}$$

On $\boldsymbol{a}_0$ and for the BVAR-SV we use the NG shrinkage prior described in Section 3. For the evaluation, we 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. Fig. 4 summarizes the results for the one-step-ahead forecast horizon. All panels display log predictive scores 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. With the exception of the euro area 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 2008/09. While predictions from all model specifications worsen, they deteriorate the least for the DL specification. In particular for the EA and the UK, the dominance of the DL prior



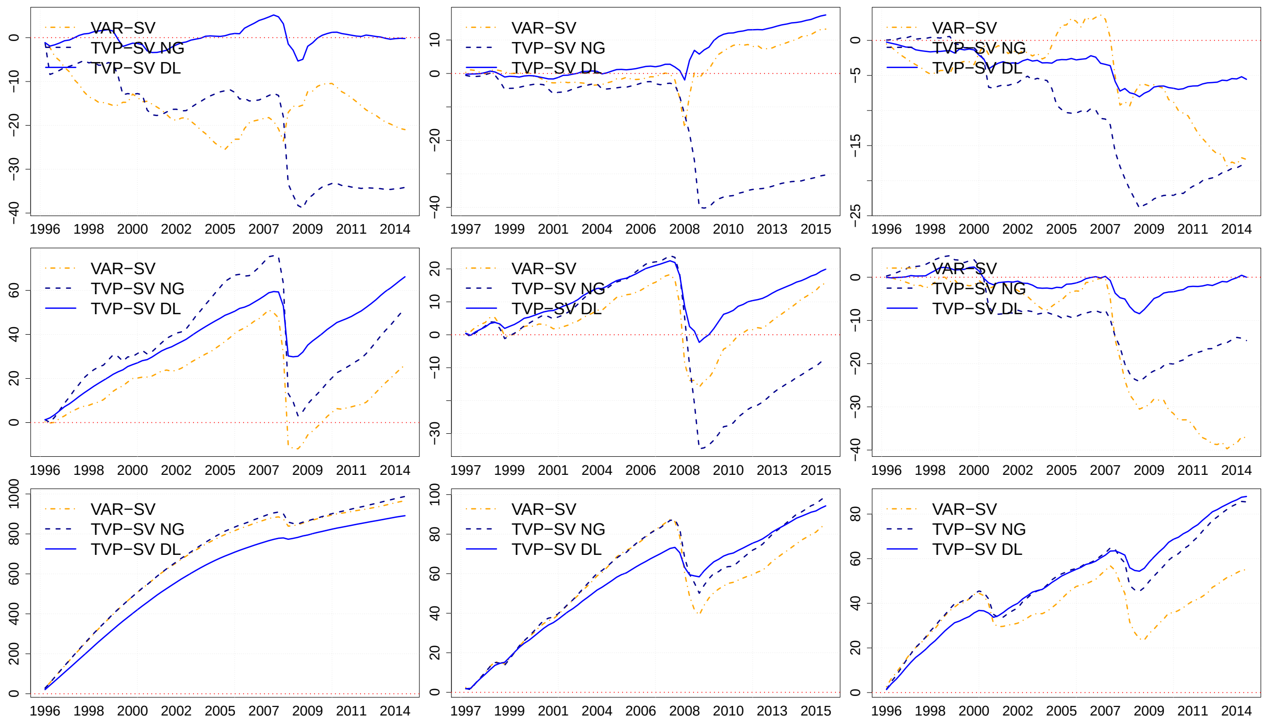
(a) EA **(b)** UK **(c)** US

Fig. 4: One-quarter-ahead cumulative log predictive Bayes factors over time relative to the TVP-SV-VAR without shrinkage. Top row: Small model (3 variables). Middle row: Medium model (7 variables). Bottom row: Large model (15 variables).

stems mainly from improved forecast for short-term interest rates, see Figs. B.1 and B.2 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 the constant parameter VAR-SV model rank second. For US data it pays off to



(a) EA **(b)** UK **(c)** US

Fig. 5: Four-quarter-ahead cumulative log predictive Bayes factors over time relative to the TVP-SV-VAR with loose shrinkage. Top row: Small model (3 variables). Middle row: Medium model (7 variables). Bottom row: Large model (15 variables).

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 Fig. 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 quarters 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.

The predicted weight associated with model i is computed as follows

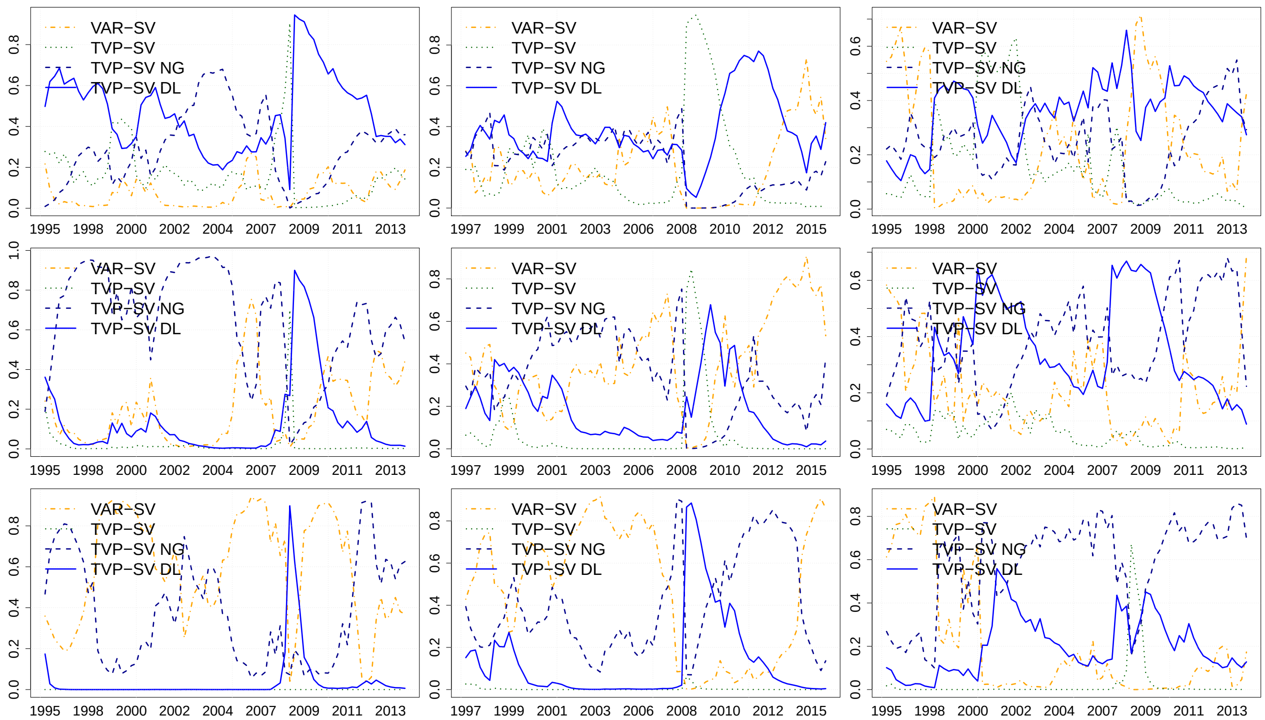
$$\mathfrak{w}_{t|t-1,i} := \frac{\mathfrak{w}^{\alpha}_{t-1|t-1,i}}{\sum_{i\in\mathcal{M}} \mathfrak{w}^{\alpha}_{t-1|t-1,i}}, \tag{5.2}$$

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

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

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 $\mathcal{M}$. 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, Fig. 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



(a) EA **(b)** UK **(c)** US

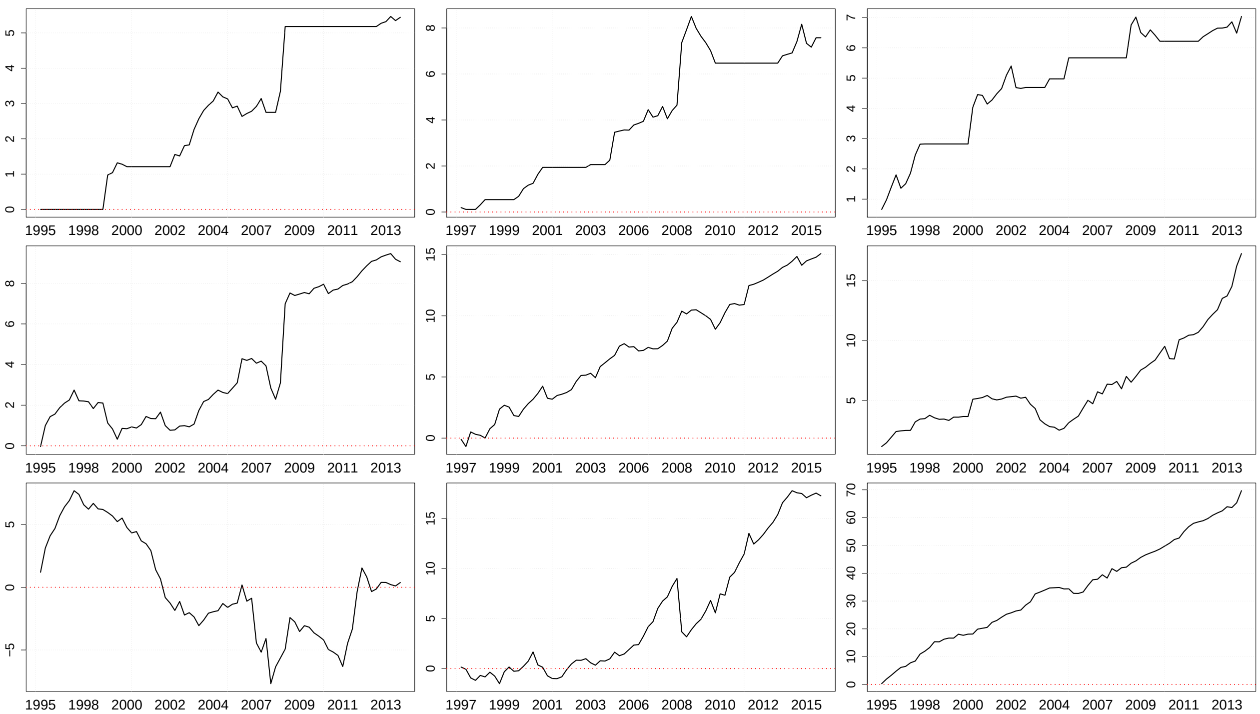
Fig. 6: Model weights over time. Top row: Small model (3 variables). Middle row: Medium model (7 variables). Bottom row: Large model (15 variables).

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 Fig. 7. The figure shows the log predictive Bayes factors relative to the best performing models over the whole sample period. These correspond to those achieving the highest cumulative log predictive Bayes factors in Fig. 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



(a) EA **(b)** UK **(c)** US

Fig. 7: Log predictive Bayes factor of dynamic model selection relative to the best performing model over time. Top row: Small model (3 variables). Middle row: Medium model (7 variables). Bottom row: Large model (15 variables).

crisis are more pronounced for the EA and the UK, 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 euro area'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 we have adapted two recent global-local shrinkage priors and used them to efficiently estimate time-varying parameter VARs of differing sizes and for three large economies. The priors capture convenient features of the traditional Minnesota prior, effectively pushing coefficients associated with higher lag orders as well as their propensity to drift towards zero.

Applying the proposed priors to three different data sets, we find improvements in one-step ahead forecasts from the time-varying parameter specifications against various competitors. Allowing for time variation and using shrinkage priors leads to smaller drops in forecast performance during the global financial crisis, while their forecasts remain competitive during the rest of the sample period. This finding is further corroborated by a dynamic model selection exercise which attaches sizable model weights to time-varying parameter models during the period of the global financial crisis. In that sense using flexible time-varying parameter models leads to large forecast gains during times of heightened uncertainty.

Finally, and comparing the two proposed priors, we find that the DL prior outperforms in small-scale VARs. By contrast, the TVP-VAR equipped with a NG prior shows the strongest performance in medium to large scale applications along with the constant parameter NG-VAR with SV. This is driven by the fact that the NG prior induces more shrinkage on the coefficients and pushes more strongly towards a constant parameter model and the payoffs of more shrinkage in larger scale models are well documented. The same holds true for the four steps ahead forecast horizon. The DL prior does a fine job in small to medium scale models, while the merits of the NG prior play out most strongly in large models.

That said, our results also point at a trade-off between complexity (i.e., allowing for time-varying parameters) and model size (i.e., data information). The larger the information set, the stronger the performance of the constant parameter model. In other words, within the VAR framework for macroeconomic time series, it is advisable to use sophisticated models for small data and simple models for sizeable data. For consistently good performance independently of the size of the data, we recommend to use sophisticated models with strong shrinkage priors such as the proposed NG shrinkage prior. This alleviates the problem of overfitting and provides a plethora of additional inferential opportunities.

References

Belmonte MA, Koop G and Korobilis D (2014) Hierarchical shrinkage in time-varying parameter models. *Journal of Forecasting* 33(1), 80–94

Bhattacharya A and Dunson DB (2011) Sparse Bayesian infinite factor models. *Biometrika* 98(2), 291–306

Bhattacharya A, Pati D, Pillai NS and Dunson DB (2015) Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association* 110(512), 1479–1490

Bitto A and Frühwirth-Schnatter S (2016) Achieving shrinkage in a time-varying parameter model framework. Technical report. ArXiv:1611.01310

Carriero A, Clark TE and Marcellino MG (2016) Large vector autoregressions with stochastic volatility and flexible priors. *FRB of Cleveland Working Paper* (16)

Carter CK and Kohn R (1994) On Gibbs sampling for state space models. *Biometrika* 81(3), 541–553

Cogley T and Sargent TJ (2005) Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic dynamics* 8(2), 262–302

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

Doan TR, Litterman BR and Sims CA (1984) Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3(1), 1–100

Eisenstat E, Chan JCC and Strachan RW (2016) Stochastic Model Specification Search for Time-Varying Parameter VARs. *Econometric Reviews* 35(8-10), 1638–1665

Fagan G, Henry J and Mestre R (2001) An Area Wide Model (AWM) for the Euro Area. ECB Working Paper 42, European Central Bank

Frühwirth-Schnatter S (1994) Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15(2), 183–202

Frühwirth-Schnatter S and Wagner H (2010) Stochastic model specification search for Gaussian and partial non-Gaussian state space models. *Journal of Econometrics* 154(1), 85–100

George EI and McCulloch RE (1993) Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88(423), 881–889

George EI, Sun D and Ni S (2008) Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* 142(1), 553–580

Geweke J and Amisano G (2010) Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26(2), 216–230

Griffin JE and Brown PJ (2010) Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* 5(1), 171–188

Griffin JE and Brown PJ (2017) Hierarchical shrinkage priors for regression models. *Bayesian Analysis* 12(1), 135–159

Hörmann W and Leydold J (2013) Generating Generalized Inverse Gaussian Random Variates. *Statistics and Computing* 24(4), 1–11

Huber F and Feldkircher M (2017) Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics* , 1–13

Kastner G (2016) Dealing with Stochastic Volatility in Time Series Using the R Package stochvol. *Journal of Statistical Software* 69(5), 1–30

Kastner G and Frühwirth-Schnatter S (2014) Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis* 76, 408–423

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

Korobilis D (2013) Hierarchical Shrinkage Priors for Dynamic Regressions with Many Predictors. *International Journal of Forecasting* 29, 43–59

Korobilis D (2014) Data-based priors for vector autoregressions with drifting coefficients. Technical

Report 14, Scottish Institute for Research in Economics (SIRE)
Korobilis D and Pettenuzzo D (2016) Adaptive Minnesota Prior for High-Dimensional Vector Autoregressions. *Essex Finance Centre Working Paper Series* (14)
Leydold J and Hörmann W (2017) *GIGrvg: Random Variate Generator for the GIG Distribution*. R package version 0.5
McCracken MW and Ng S (2016) FRED-MD: A Monthly Database for Macroeconomic Research. *Journal of Business & Economic Statistics* 34(4), 574–589
Onorante L and Raftery AE (2016) Dynamic model averaging in large model spaces using dynamic Occam's window. *European Economic Review* 81, 2–14
Polson NG and Scott JG (2010) Shrink globally, act locally: Sparse Bayesian regularization and prediction. *Bayesian Statistics* 9, 501–538
Primiceri GE (2005) Time Varying Structural Vector Autoregressions and Monetary Policy. *Review of Economic Studies* 72(3), 821–852
Raftery A, Karny M and Ettler P (2010) Online Prediction Under Model Uncertainty Via Dynamic Model Averaging: Application to a Cold Rolling Mill. *Technometrics* 52, 52–66
Sims CA and Zha T (1998) Bayesian Methods for Dynamic Multivariate Models. *International Economic Review* 39(4), 949–68
Thomas R, Hills S and Dimsdale N (2010) The UK Recession in Context – What do Three Centuries of Data Tell Us? *Bank of England Quarterly Bulletin* 50, 277–291

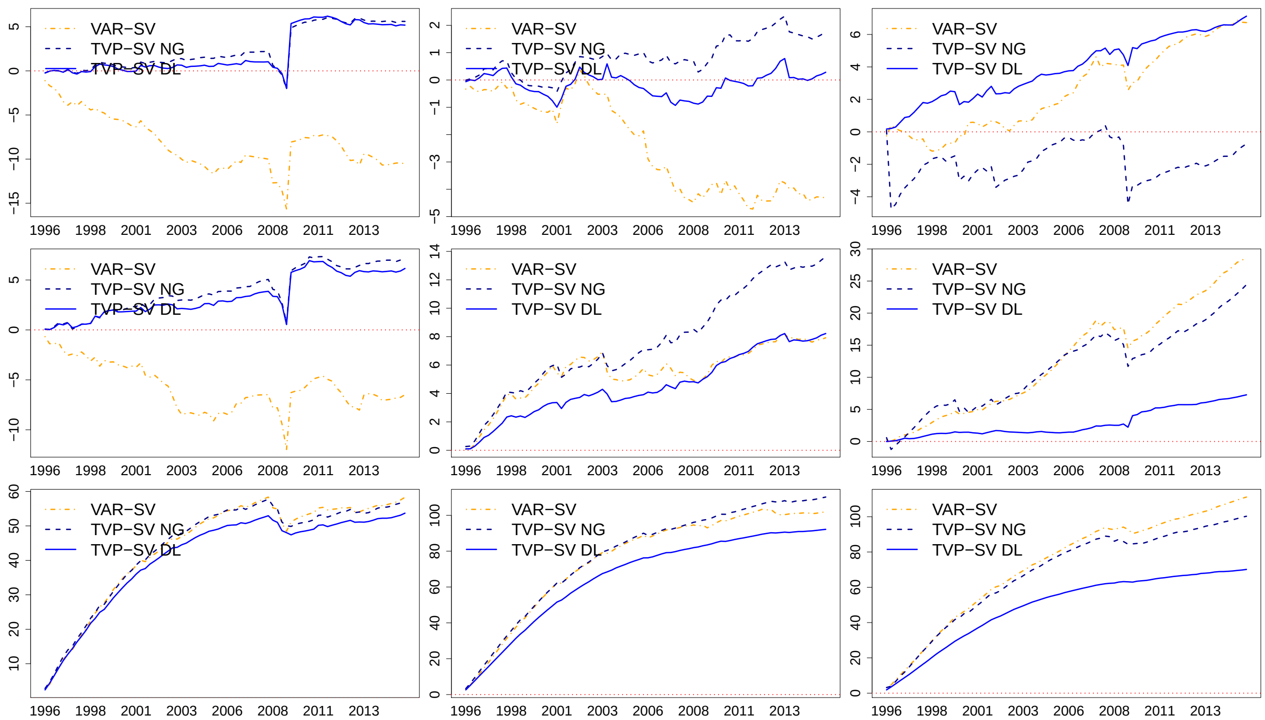
Appendix A Data overview

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	CES2000000008x, real avg.			x	1
		earnings series.	hourly earnings.			x	
Money	M2, sa, OECD data.	Stock of "broad money", break-adjusted.	M2REALx, real M2 money stock (bn of 1982-84 USD), deflated by CPI.			x	1
Equity prices	Equity share price index, OECD data.	Spliced monthly share price index weighted by market capitalisation.	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
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 http://eabcn.org/page/area-wide-model for the euro area, from http://www.bankofengland.co.uk/research/Pages/datasets/default.aspx for the UK and from https://research.stlouisfed.org/econ/mccracken/fred-databases/ for the US. For the euro area, equity prices and M2 obtained from the OECD; employment data for UK transformed using log differences. Transformation codes (T-codes) are as follows: 1 - log differences, 2 - raw data.

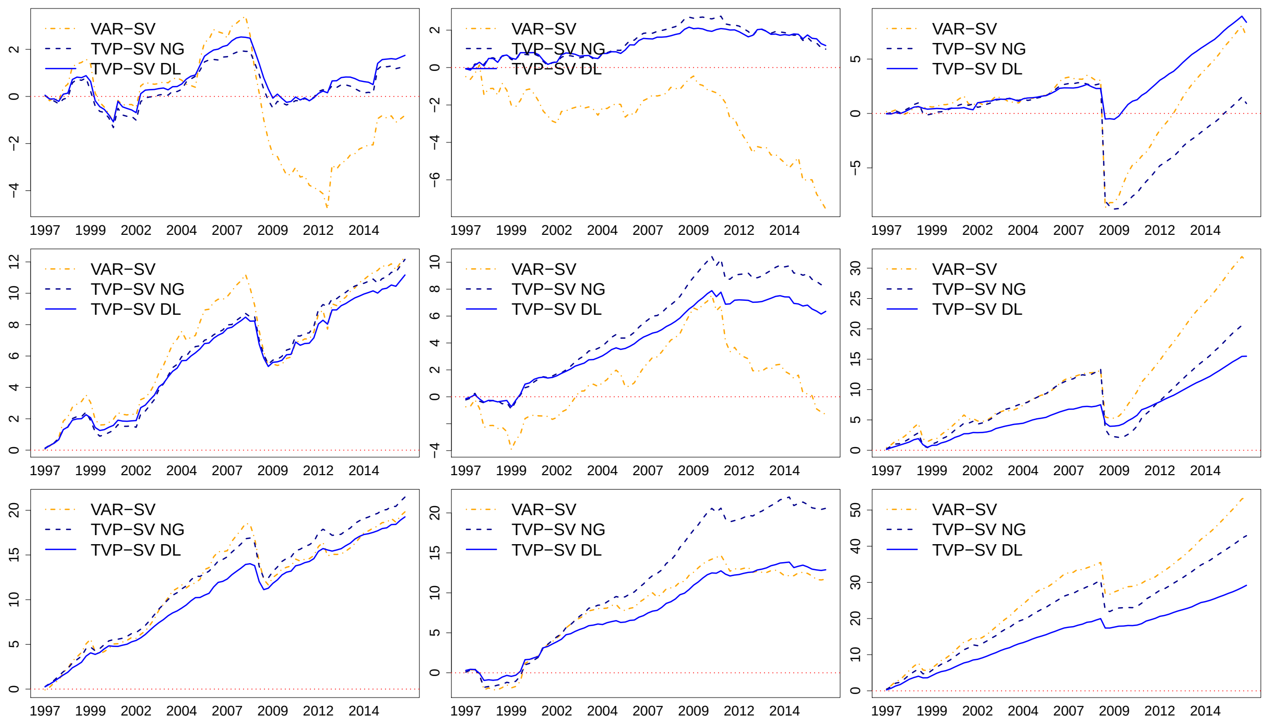
Table A.1: Data overview

Appendix B Additional empirical results



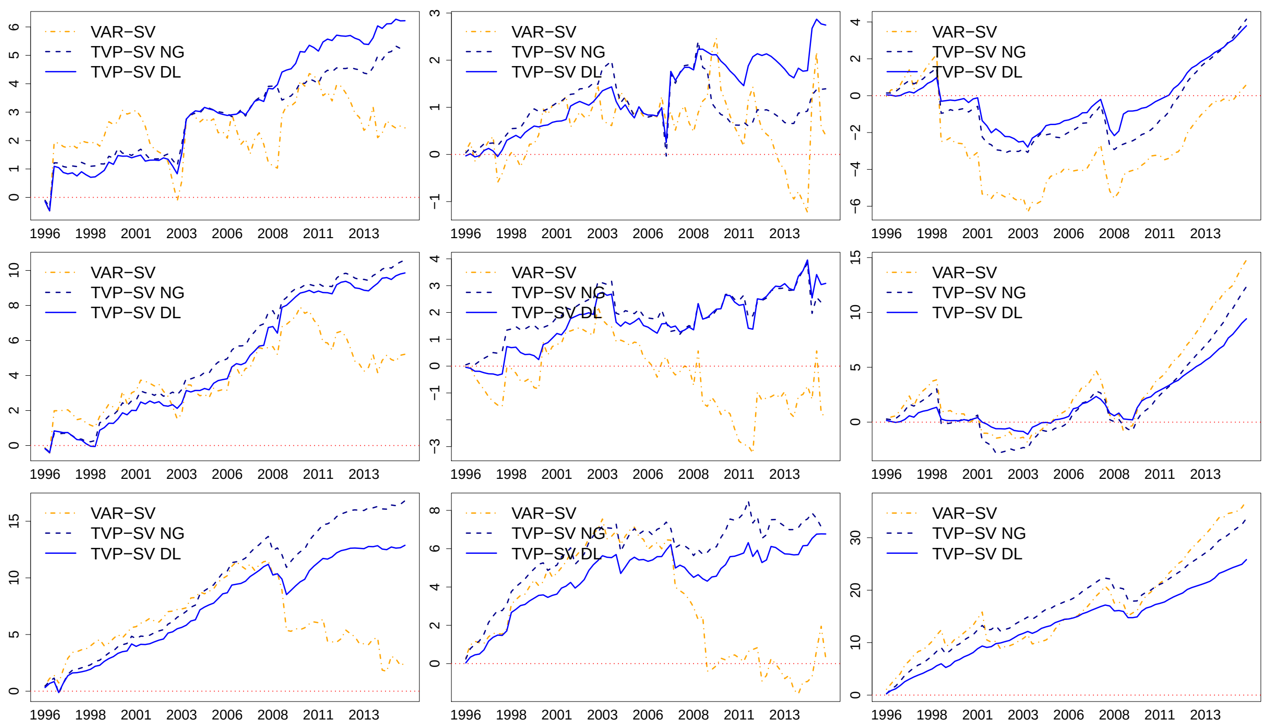
(a) Real GDP **(b)** GDP deflator **(c)** Short rates

Fig. B.1: Euro Area: Univariate cumulative log predictive one-quarter-ahead Bayes factors over time relative to the TVP-SV-VAR with loose shrinkage. Top row: Small model (3 variables). Middle row: Medium model (7 variables). Bottom row: Large model (15 variables).



(a) Real GDP **(b)** GDP deflator **(c)** Short rates

Fig. B.2: United Kingdom: Univariate cumulative log predictive one-quarter-ahead Bayes factors over time relative to the TVP-SV-VAR with loose shrinkage. Top row: Small model (3 variables). Middle row: Medium model (7 variables). Bottom row: Large model (15 variables).



(a) Real GDP **(b)** GDP deflator **(c)** Short rates

Fig. B.3: United States: Univariate cumulative log predictive one-quarter-ahead Bayes factors over time relative to the TVP-SV-VAR with loose shrinkage. Top row: Small model (3 variables). Middle row: Medium model (7 variables). Bottom row: Large model (15 variables).