Sparse Bayesian vector autoregressions in huge dimensions

Gregor Kastner* and Florian Huber

WU Vienna University of Economics and Business [gkastner@wu.ac.at|fhuber@wu.ac.at]

April 12, 2017

Abstract

We develop a Bayesian vector autoregressive (VAR) model that is capable of handling vast dimensional information sets. Three features are introduced to permit reliable estimation of the model. First, we assume that the reduced-form errors in the VAR feature a factor stochastic volatility structure, allowing for conditional equation-by-equation estimation. Second, we apply a Dirichlet-Laplace prior to the VAR coefficients to cure the curse of dimensionality. Finally, since simulation-based methods are needed to simulate from the joint posterior distribution, we utilize recent innovations to efficiently sample from high-dimensional multivariate Gaussian distributions that improve upon recent algorithms by large margins. In the empirical exercise we apply the model to US data and evaluate its forecasting capabilities.

Keywords: factor stochastic volatility, curse of dimensionality, shrinkage, Dirichlet-Laplace prior, MCMC.

JEL Codes: C11, C30, C53, E27.

1 Introduction

Previous research has identified two important features which macroeconometric models should possess: the ability to exploit high dimensional information sets (Bańbura et al., 2010; Stock and Watson, 2011) and the potentiality to capture non-linear features of the underlying macroeconomic time series (Cogley and Sargent, 2002; Primiceri, 2005; Clark, 2011; Clark and Ravazzolo, 2015). While the literature suggests

*Corresponding author. Address: Welthandelsplatz 1, Building D4, 4th Floor. 1020 Vienna, Austria. Phone: +43 1 31336-5593.

several paths to estimate large models, the majority of such approaches implies that once non-linearities are taken into account, analytical solutions are no longer available and the computational burden becomes prohibitive.[1] This implies that high dimensional non-linear models can practically be estimated only under strong (and often unrealistic) restrictions on the dynamics of the model. However, especially in forecasting applications or in structural analysis, recent literature suggests that successful models should be able to exploit lots of information and also control for breaks in the autoregressive parameters or, more importantly, changes in the volatility of economic shocks (Primiceri, 2005; Sims and Zha, 2006; Koop et al., 2009).

Two reasons limit the use of large (or even huge) non-linear models. The first reason is statistical. Since the number of parameters in a standard vector autoregression rises rapidly with the number of time series included and commonly used macroeconomic time series are rather short, in-sample overfitting proves to be a serious issue. As a solution, the Bayesian literature on VAR modeling (Doan et al., 1984; Litterman, 1986; Sims and Zha, 1998; George et al., 2008; Bańbura et al., 2010; Koop, 2013; Clark, 2011; Clark and Ravazzolo, 2015; Korobilis and Pettenuzzo, 2016; Huber and Feldkircher, 2017) suggests shrinkage priors that push the parameter space towards some stylized prior model like a multivariate random walk. This typically leads to much improved forecasting properties and more meaningful structural inference. Moreover, much of the literature on Bayesian VARs imposes conjugate priors on the autoregressive parameters, allowing for analytical posterior solutions and thus avoiding simulation based techniques like Markov chain Monte Carlo (MCMC).

The second reason is computational. Since non-linear models typically have to be estimated by means of MCMC, computational intensity increases vastly if the number of variables included becomes large. The increase in computational complexity stems from the fact that standard algorithms for multivariate regression models call for the inversion of large covariance matrices. Especially for large systems, this can quickly turn prohibitive since the inverse of the posterior variance-covariance matrix on the coefficients has to be computed for each sweep of the MCMC algorithm. For natural conjugate models, this step can be vastly simplified since the likelihood features a convenient Kronecker structure, implying that all equations in a VAR feature the same set of explanatory variables. This speeds up computation by large margins but restricts the flexibility of the model. Carriero et al. (2016), for instance, exploit this fact and introduce a simplified stochastic volatility specification. Another strand of the literature augment each equation of the VAR by including the residuals of the preceding equations (Carriero et al., 2015) which also provides significant improvements in terms of computational speed. Finally, in a recent contribution, Pettenuzzo et al. (2016) re-

[1]One exception is Koop and Korobilis (2013).

duce the dimensionality of the problem at hand by randomly compressing the lagged endogenous variables in the VAR.

All papers mentioned hitherto focus on capturing cross-variable correlation in the conditional mean through the VAR part and the co-movement in volatilities is captured by a rich specification of the error variance (Primiceri, 2005) or by a single factor (Carriero et al., 2016). Another strand of the literature, typically used in financial econometrics, utilizes factor models to provide a parsimonious representation of a covariance matrix, focusing exclusively on the second moment of the predictive density. For instance, Pitt and Shephard (1999) and Aguilar and West (2000) assume that the variance-covariance matrix of a broad panel of time series might be described by a lower dimensional matrix of latent static factors featuring stochastic volatility and a variable-specific idiosyncratic stochastic volatility process.

The present paper combines the virtues of exploiting large information sets and allowing for movements in the error variance. The overfitting issue mentioned above is solved as follows. First, we use a Dirichlet-Laplace (DL) prior specification (see Bhattacharya et al., 2015) on the VAR coefficients. This prior is a global-local shrinkage prior in the spirit of Polson and Scott (2011); it enables us to heavily shrink the parameter space but at the same time provides enough flexibility to allow for non-zero regression coefficients if necessary. Second, a factor stochastic volatility model on the VAR errors grants a parsimonious representation of the time-varying error variance-covariance matrix of the VAR. To deal with the computational complexity, we exploit the fact that, conditionally on the latent factors and their loadings, equation by equation estimation becomes possible within each MCMC iteration. Moreover, we apply recent advances for fast sampling from high dimensional multivariate Gaussian distributions (Bhattacharya et al., 2016) that enables us to estimate a model featuring almost $200\,000$ autoregressive parameters and an error covariance matrix with over $20\,000$ nontrivial time-varying elements on a quarterly US dataset in a reasonable amount of time. In a careful analysis we show to what extent our proposed method improves upon a set of standard algorithms typically used to simulate from the joint posterior distribution of large dimensional Bayesian VARs.

In an empirical application we use a modified version of the quarterly dataset proposed by Stock and Watson (2011) and McCracken and Ng (2016) and analyze the dynamic properties of the US economy within our Bayesian VAR with factor stochastic volatility. Our model allows new insights on the different driving forces of macroeconomic volatility within the US. To illustrate the out-of-sample performance of our model, we forecast important economic indicators such as output, consumer price inflation and short-term interest rates, amongst others. The proposed model is benchmarked against several nested alternatives. Our findings suggests that the proposed model performs well, outperforming all benchmarks on all time horizons considered. In addition, investigating the time profile of the cumulative log predictive likelihood reveals that the factor structure especially pays off in times of economic stress.

The remainder of this paper is structured as follows. Section 2 introduces the econometric framework. The Bayesian estimation approach, including an elaborated account of the prior setup adopted and the corresponding conditional posterior distributions, is detailed in Section 3. Section 4 provides an analysis of the computational gains of our proposed algorithm relative to a wide set of different algorithms. Section 5, after giving a brief overview of the dataset used along with the model specification, illustrates our modeling approach by fitting a single factor model to US data. Moreover, we perform a forecasting exercise to assess the predictive performance of our approach and discuss the choice of the number of latent factors. Finally, the last section concludes.

2 Econometric framework

Suppose interest centers on modeling an $m \times 1$ vector of time series denoted by $\boldsymbol{y}_t$ with $t = 1, \dots, T$. We assume that $\boldsymbol{y}_t$ follows a VAR(p) process,[2]

$$\boldsymbol{y}_t = \boldsymbol{A}_1 \boldsymbol{y}_{t-1} + \cdots + \boldsymbol{A}_p \boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t. \tag{2.1}$$

Each $\boldsymbol{A}_j$ $(j = 1, \dots, p)$ is an $m \times m$ matrix of autoregressive coefficients and $\boldsymbol{\varepsilon}_t$ follows a factor stochastic volatility model (Pitt and Shephard, 1999; Aguilar and West, 2000),

$$\boldsymbol{\varepsilon}_t = \boldsymbol{\Lambda} \boldsymbol{f}_t + \boldsymbol{\eta}_t. \tag{2.2}$$

Here, $\boldsymbol{\Lambda}$ denotes the $m \times q$ matrix of factor loadings with typical element λ_{ij} $(i = 1, \dots, m; j = 1, \dots, q)$ and $\boldsymbol{f}_t \sim \mathcal{N}_q(\boldsymbol{0}_q, \boldsymbol{V}_t)$ is a vector of latent factors with q-dimensional zero mean vector $\boldsymbol{0}_q$ and variance-covariance matrix $\boldsymbol{V}_t = \text{diag}(e^{h_{1t}}, \dots, e^{h_{qt}})$. We also allow for heteroscedasticity in the idiosyncratic error term $\boldsymbol{\eta}_t$, more concretely we assume that $\boldsymbol{\eta}_t \sim \mathcal{N}_m(\boldsymbol{0}_m, \boldsymbol{\Sigma}_t)$ with $\boldsymbol{\Sigma} = \text{diag}(\sigma^2_{1t}, \dots, \sigma^2_{mt})$. The logarithms of the diagonal elements of $\boldsymbol{V}_t$ and $\boldsymbol{\Sigma}_t$ follow AR(1) processes,

$$h_{jt} = \rho_{hj} h_{j,t-1} + e_{hj,t}, \ j = 1, \dots, q, \tag{2.3}$$

$$\log \sigma^2_{it} = \mu_{\sigma i} + \rho_{\sigma i}(\log \sigma^2_{i,t-1} - \mu_{\sigma i}) + e_{\sigma i,t}, \ i = 1, \dots, m. \tag{2.4}$$

To identify the scaling of the latent factors, the process specified in Eq. (2.3) is assumed to have mean zero while $\mu_{\sigma j}$ is the mean of the idiosyncratic log volatility to be estimated from the data. The parameters ρ_{hj} and $\rho_{\sigma i}$ are a priori restricted to the interval

[2] For simplicity we omit the intercept term in the following discussion.

$(-1, 1)$ and denote the persistences of the latent log variances. The error terms $e_{hj,t}$ and $e_{\sigma i,t}$ denote independent zero mean innovations with variances ς_{hj}^2 and $\varsigma_{\sigma i}^2$, respectively. This specification implies that the log volatilities are mean reverting and thus bounded in the limit.

The variance-covariance matrix of $\boldsymbol{\eta}_t$, denoted by $\boldsymbol{\Omega}_t$, is given by

$$\boldsymbol{\Omega}_t = \boldsymbol{\Lambda} \boldsymbol{V}_t \boldsymbol{\Lambda}' + \boldsymbol{\Sigma}_t. \tag{2.5}$$

This structure makes it clear that off-diagonal entries of $\boldsymbol{\Omega}_t$ exclusively stem from the volatilities of the q factors, while the diagonal entries are allowed to feature idiosyncratic deviations.

The model described by Eqs. (2.1) to (2.5) is related to several alternative specifications commonly used in the literature. For instance, assuming that $\boldsymbol{V}_t = \boldsymbol{I}$ and $\boldsymbol{\Sigma}_t \equiv \boldsymbol{\Sigma}$ for all t leads to the specification adopted in Stock and Watson (2005). Setting $q = 1$ and $\boldsymbol{\Sigma}_t \equiv \boldsymbol{\Sigma}$ yields a specification that is similar to the one stipulated in Carriero et al. (2016), with the difference that our model imposes restrictions on the covariances whereas Carriero et al. (2016) estimate a full (but constant) covariance matrix and our model implies that the stochastic volatility enters $\boldsymbol{\Omega}_t$ in an additive fashion.

Before proceeding to the next subsection it is worth summarizing the key features of the model given by Eqs. (2.1) to (2.5). First, we capture cross-variable movements in the conditional mean through the VAR block of the model and assume that co-movement in conditional variances is captured by the factor model in Eq. (2.2). Second, the model introduces stochastic volatility by assuming that a large panel of volatilities may be efficiently summarized through a set of latent factors. This choice is more flexible than a single factor model for the volatility, effectively providing a parsimonious representation of $\boldsymbol{\Omega}_t$ that is flexible enough to replicate the dynamic behavior of the variances of a broad set of macroeconomic quantities.

3 Achieving shrinkage in large dimensional VAR models

Our approach to estimation and inference is Bayesian. This implies that we have to specify suitable prior distributions on the parameters of the model described in the previous subsection and combine the prior distributions with the likelihood to obtain the corresponding posterior distributions.

3.1 A global-local shrinkage prior

For prior implementation, it proves to be convenient to define a $k \times 1$ vector $\boldsymbol{x}_t = (\boldsymbol{y}'_{t-1}, \dots, \boldsymbol{y}'_{t-p})'$ and a $m \times k$ coefficient matrix $\boldsymbol{B} = (\boldsymbol{A}_1, \dots, \boldsymbol{A}_p)$ with $k = mp$ to rewrite the model in Eq. (2.1) more compactly as

$$\boldsymbol{y}_t = \boldsymbol{B}\boldsymbol{x}_t + \varepsilon_t. \tag{3.1}$$

Stacking the rows of $\boldsymbol{y}_t$, $\boldsymbol{x}_t$, and $\boldsymbol{\eta_t}$ yields

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B}' + \boldsymbol{E}, \tag{3.2}$$

where $\boldsymbol{Y} = (\boldsymbol{y}_1, \dots, \boldsymbol{y}_T)', \boldsymbol{X} = (\boldsymbol{x}_1, \dots, \boldsymbol{x}_T)'$ and $\boldsymbol{E} = (\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_T)'$ denote the corresponding full data matrices.

Typically, the matrix $\boldsymbol{B}$ is a sparse matrix with non-zero elements mainly located on the main diagonal of $\boldsymbol{A}_1$. In fact, existing priors in the Minnesota tradition tend to strongly push the system towards the prior model in high dimensions. However, especially in large models an extremely tight prior on $\boldsymbol{B}$ might lead to severe overshrinking, effectively zeroing out coefficients that might be important to explain $\boldsymbol{y}_t$. If the matrix $\boldsymbol{B}$ is characterized by a relatively low number of non-zero regression coefficients, a possible solution is a global-local shrinkage prior (Polson and Scott, 2011).

A recent variant that falls within the class of global-local shrinkage priors is the Dirichlet-Laplace (DL) prior put forward in Bhattacharya et al. (2015). This prior possesses convenient shrinkage properties in the presence of a large degree of sparsity of the parameter vector $\boldsymbol{b} = \text{vec}(\boldsymbol{B})$. In what follows we impose the DL prior on each of the $K = mk$ elements of $\boldsymbol{b}$, denoted as b_j for $j = 1, \dots, K$,

$$b_j \sim \mathcal{N}(0, \psi_j \vartheta_j^2 \zeta), \quad \psi_j \sim \text{Exp}(1/2), \tag{3.3}$$

where ψ_j are local scaling parameters, $\text{Exp}(\lambda)$ denotes the exponential distribution with rate λ, and the elements of $\boldsymbol{\vartheta} = (\vartheta_1, \dots, \vartheta_K)'$ and are additional auxiliary scaling parameters that are bounded to the $(K-1)$-dimensional simplex $\mathcal{S}^{K-1} = \{\boldsymbol{\vartheta} : \vartheta_j \geq 0, \sum_{j=1}^n \vartheta_j = 1\}$. A natural prior choice for ϑ_j is the (symmetric) Dirichlet distribution with hyperparameter a,

$$\vartheta_j \sim \text{Dir}(a, \dots, a). \tag{3.4}$$

In addition, ζ is a global shrinkage parameter that pushes all elements in $\boldsymbol{B}$ towards zero and exhibits an important role in determining the tail behavior of the marginal prior distribution on b_j, obtained after integrating out the ϑ_js. Thus, we follow Bhattacharya et al. (2015) and adopt a fully Bayesian approach by specifying a Gamma distributed prior on ζ,

$$\zeta \sim \mathcal{G}(Ka, 1/2). \tag{3.5}$$

It is noteworthy that this prior setup has at least two convenient features that appear to be of prime importance for VAR modeling. First, it exerts a strong degree of shrinkage on all elements of $\boldsymbol{B}$ but still provides additional flexibility such that non-zero regression coefficients are permitted. This critical property is a feature which a large class of global-local shrinkage priors share (Griffin and Brown, 2010; Carvalho et al., 2010; Polson and Scott, 2011) and has been recently adopted in a VAR framework by Huber and Feldkircher (2017) and within the general context of state space models by e.g. Bitto and Frühwirth-Schnatter (2016) and Kastner (2016). Second, implementation is simple and requires relatively little additional input from the researcher. In fact, the prior heavily relies on a single structural hyperparameter that has to be specified with some care, namely a.

The hyperparameter a influences the empirical properties of the proposed shrinkage prior along several important dimensions. Smaller values of a lead to heavy shrinkage on all elements of $\boldsymbol{B}$. To see this, note that lower values of a imply that more prior mass is placed on small values of ζ a priori. Similarly, when a is small, the Dirichlet prior places more mass on values of ϑ_j close to zero. Since lower values of ζ translate into thicker tails of the marginal prior on b_j, the specific choice of a not only influences the overall degree of shrinkage but also the tail behavior of the prior. Bhattacharya et al. (2015) show that if a is specified as $K^{-(1+\Delta)}$ for any $\Delta > 0$ to be small the DL prior displays excellent posterior contraction rates.[3]

For the factor loadings we independently use a standard normally distributed prior on each element $\lambda_{ij} \sim \mathcal{N}(0, 1)$ for $i = 1, \dots, m$ and $j = 1, \dots, q$. Likewise, we impose a normally distributed prior on the mean of the log-volatility $\mu_{\sigma j} \sim \mathcal{N}(0, M_\mu)$ with M_μ denoting the prior variance. Furthermore, we place the commonly employed Beta distributed prior on the transformed persistence parameter of the log-volatility $\frac{\rho_{sj}+1}{2} \sim \mathcal{B}(a_0, b_0)$ for $s \in \{h, \sigma\}$ and $a_0, b_0 \in \mathbb{R}^+$ to ensure stationarity. Finally, we use a restricted Gamma prior on the innovation variances in Eqs. (2.3) and (2.4), $\varsigma_{sj}^2 \sim \mathcal{G}(\frac{1}{2}, \frac{1}{2\xi})$. Here, ξ is a hyperparameter used to control the tightness of the prior. This choice, motivated in Frühwirth-Schnatter and Wagner (2010) implies that if the data is not informative on the degree of time variation of the log volatilities then we do not bound ς_{sj}^2 artificially away from zero, effectively applying more shrinkage than the standard inverted Gamma prior.

3.2 Full conditional posterior distributions

Conditional on the latent factors and the corresponding loadings, the model in Eq. (2.1) can be cast as a system of m unrelated regression models for the elements in $\boldsymbol{z}_t =$

[3] For a discussion of the shrinkage properties of the proposed prior within the context of factor models, see Pati et al. (2014)

$\boldsymbol{y}_t - \boldsymbol{\Lambda}\boldsymbol{f}_t$, z_{it}, with heteroscedastic errors,

$$z_{it} = \boldsymbol{B}_{i\bullet}\boldsymbol{x}_t + \eta_{it}, \text{ for } i = 1, \dots, m. \tag{3.6}$$

Here we let $\boldsymbol{B}_{i\bullet}$ denote the ith row of $\boldsymbol{B}$ and η_{it} is the ith element of $\boldsymbol{\eta}_t$. The corresponding posterior distribution of $\boldsymbol{B}'_{i\bullet}$ is a k-dimensional multivariate Gaussian distribution,

$$\boldsymbol{B}'_{i\bullet}|\bullet \sim \mathcal{N}(\boldsymbol{b}_i, \boldsymbol{Q}_i), \tag{3.7}$$

with $\bullet$ indicating that we condition on the remaining parameters and latent quantities of the model. The posterior variance and mean are given by

$$\boldsymbol{Q}_i = (\tilde{\boldsymbol{X}}'_j\tilde{\boldsymbol{X}}_j + \boldsymbol{\Phi}_i^{-1})^{-1}, \tag{3.8}$$
$$\boldsymbol{b}_i = \boldsymbol{Q}_i(\tilde{\boldsymbol{X}}'_j\tilde{z}_i). \tag{3.9}$$

The diagonal prior covariance matrix of the coefficients related to the ith equation is given by $\boldsymbol{\Phi}_i$, the respective $k \times k$ diagonal submatrix of $\boldsymbol{\Phi} = \zeta \times \text{diag}(\psi_1\vartheta_1^2, \dots, \psi_K\vartheta_K^2)$. Moreover, we let $\tilde{\boldsymbol{X}}_j$ be a $T \times K$ matrix with typical row t given by $\boldsymbol{X}_t/\sigma_{jt}$ and $\tilde{\boldsymbol{z}_{\bullet i}}$ is a T-dimensional vector with the tth element given by z_{it}/σ_{jt}. This normalization renders Eq. (3.6) conditionally homoscedastic with standard normally distributed white noise errors.

The full conditional posterior distribution of the inverse of ψ_j is inverse Gaussian distributed,

$$\psi_j^{-1}|\bullet \sim iG(\vartheta_j\zeta/|b_j|, 1) \text{ for } j = 1, \dots, K. \tag{3.10}$$

For the global shrinkage parameter ζ the conditional posterior follows a generalized inverted Gaussian (GIG) distribution,

$$\zeta|\bullet \sim \mathcal{GIG}\left(K(a-1), 1, 2\sum_{j=1}^{K}|b_j|/\vartheta_j\right). \tag{3.11}$$

To draw from this distribution, we use the R-package `GIGrvg` (Leydold and Hörmann, 2015) implementing the efficient algorithm of Hörmann and Leydold (2013). Moreover, we sample the scaling parameters ϑ_j by first sampling L_j from

$$L_j|\bullet \sim \mathcal{GIG}(a-1, 1, 2|b_j|), \tag{3.12}$$

and then setting

$$\vartheta_j = \frac{L_j}{\sum_{i=1}^{K} L_i}. \tag{3.13}$$

The conditional posterior distributions of the factors are Gaussian and thus straightforward to draw from. The factor loadings are sampled using "deep interweaving" (see Kastner et al., 2018), and the parameters in Eq. (2.3) and Eq. (2.4) along the full histories of the latent log-volatilities are sampled as in Kastner and Frühwirth-Schnatter (2014) using the R-package `factorstochvol` (Kastner, 2017).

Our MCMC algorithm iteratively draws from the conditional posterior distributions outlined above and discards the first J draws as burn-in. In terms of computational requirements, the single most intensive step is the simulation from the joint posterior of the autoregressive coefficients in $\boldsymbol{B}$. Because this step is implemented on an equation-by-equation basis, speed improvements relative to the standard approach are already quite substantial. However, note that if k is large (i.e. of the order of several thousands), even the commonly employed equation-by-equation sampling fails to deliver a sufficient amount of draws within a reasonable time window. Consequently, we outline an alternative algorithm to draw from a high-dimensional multivariate Gaussian distribution under a Bayesian prior that features a diagonal prior variance-covariance matrix in the upcoming section.

4 Computational aspects

The typical approach to sampling from Eq. (3.7) is based on the full system and simultaneously samples from the full conditional posterior of $\boldsymbol{B}$, implying that the corresponding posterior distribution is a K-dimensional normal distribution with a $K \times K$ dimensional variance-covariance matrix. Under a non-conjugate prior the computational difficulties arise from the need to invert the $K \times K$ variance-covariance matrix which requires operations of order $O(m^6 p^3)$ under Gaussian elimination.

If a conjugate prior in combination with a constant (or vastly simplified heteroscedastic, see Carriero et al., 2015) specification of $\boldsymbol{\Sigma}_t$ is used, the corresponding variance covariance features a Kronecker structure which is computationally cheaper to invert and scales better in large dimensions. Specifically, the manipulations of the corresponding covariance matrix are of order $O(m^3 + k^3)$, a significant gain relatively to the standard approach. However, this comes at a cost since all equations have to feature the same set of variables, the prior on the VAR coefficients has to be symmetric and any stochastic volatility specification is necessary overly simplistic to preserve conditional conjugacy.

By contrast, recent studies emphasize the computational gains that arise from utilizing a framework that is based on equation-by-equation estimation. Carriero et al. (2015) and Pettenuzzo et al. (2016) augment each equation of the system by either contemporaneous values of the endogenous variables of the preceding equations or the residuals from the previous equations. Here, our approach renders the equations of the system conditionally independent by conditioning on the factors. From a computational perspective, the differences between using a factor model to disentangle the equations and an approach based on augmenting specific equations by quantities that aim to approximate covariance parameters are negligible. If we sample from Eq. (3.7) directly, the computations involved are of order $O(mk^3) = O(m^4p^3)$. This already poses significant improvements relative to full system estimation.

One contribution of the present paper is the application of the algorithm proposed by Bhattacharya et al. (2016) and developed for univariate regression models under a global-local shrinkage prior.[4] This algorithm is applied to each equation in the system and cycles through the following steps:

1. Sample independently $\boldsymbol{u}_i \sim \mathcal{N}(\mathbf{0}_k, \boldsymbol{\Psi}_i)$ and $\boldsymbol{\delta}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_T)$.
2. Use $\boldsymbol{u}_i$ and $\boldsymbol{\delta}_i$ to construct $\boldsymbol{v}_i = \tilde{\boldsymbol{X}}\boldsymbol{u}_i + \boldsymbol{\delta}_i$.
3. Solve $(\tilde{\boldsymbol{X}}\boldsymbol{\Psi}_i\tilde{\boldsymbol{X}}' + \boldsymbol{I}_T)\boldsymbol{w}_i = (\tilde{\boldsymbol{z}}_{i\bullet} - \boldsymbol{v}_i)$ for $\boldsymbol{w}_i$.
4. Set $\boldsymbol{B}'_{i\bullet} = \boldsymbol{u}_i + \boldsymbol{\Psi}_i\tilde{\boldsymbol{X}}'\boldsymbol{w}_i$.

Bhattacharya et al. (2016) show that this algorithm yields a valid draw from Eq. (3.7). This algorithm outperforms all competing variants discussed previously in situations where $k \gg T$, a situation commonly encountered when dealing with large VAR models. In such cases, steps (1) to (4) can be carried out using $O(pm^2T^2)$ floating point operations. In situations where $k \approx T$, the computational advantages relative to the standard equation-by-equation algorithm mentioned above are modest or even negative. However, note that the cost is quadratic in m and linear in p and thus scales much better when the number of endogenous variables and/or lags thereof is increased. More information on the empirical performance of our algorithm can be found in Section 5.4.

5 Empirical application: A huge model of the US economy

In Section 5.1 we first summarize the data set adopted and outline specification choices made. The section that follows (Section 5.2) estimates a simple one factor model to outline the virtues of our proposed framework. Section 5.3 presents the main findings of our forecasting exercise and discusses the choice of the number of factors used for modeling the error covariance structure.

[4] The algorithm is not specific to global-local shrinkage priors but requires that the prior variance-covariance matrix is cheap to invert, i.e. diagonal.

5.1 Data, model specification and selection issues

In the empirical application we forecast a set of key US macroeconomic quantities. To this end, we use the quarterly dataset provided by McCracken and Ng (2016), a variant of the well-known Stock and Watson (2011) dataset for the US. The data spans the period ranging from 1959:Q1 to 2015:Q4. We include $m = 215$ quarterly time series, capturing information on 14 important segments of the economy. We follow McCracken and Ng (2016) and transform the data to be approximately stationary due to stability reasons; furthermore, we standardize each component series to have zero mean and variance one. In the empirical examples we include $p = 4$ lags of the endogenous variables, a standard choice for quarterly macroeconomic data. The hyperparameters are chosen as follows: $M_\mu = 10$, $a_0 = 20$, $b_0 = 1.5$, $\xi = 1$. For further information on the data and transformations used see McCracken and Ng (2016).

5.2 A one factor model of the US economy

To provide some intuition on how our modeling approach works in practice, we start by estimating a simple one factor model (i.e. $q = 1$) and investigate several features of our empirical model. In the next section we will perform an extensive forecasting exercise and discuss the optimal number of factors in terms of forecasting accuracy. We illustrate our approach using a single factor model because the factor can then be interpreted as being closely related to macroeconomic uncertainty since it summarizes co-movement in one-step-ahead prediction errors and is thus closely related to the measure proposed in Jurado et al. (2015) and Mumtaz and Theodoridis (2017).

We start by inspecting the posterior distribution of $\boldsymbol{\Lambda}$ and assess what variables appear to load heavily on the latent factor. It is worth emphasizing that most quantities[5] associated with real activity (i.e. industrial production and its components, GDP growth, employment measures) load heavily on the factor. Moreover, expectation measures, housing markets, equity prices and spreads also load heavily on the joint factor.

Since several components associated with most important segments of the economy display a robust relationship with the factor our economic interpretation of a broad based economic uncertainty measure appears to be reasonable. To assess whether spikes in the volatility associated with the factor coincide with major economic events, the bottom panel of Fig. 1 depicts the evolution of the posterior distribution of factor volatility over time. A few findings are worth mentioning. First, volatility spikes sharply during the midst of the 1970s, a period characterized by the first oil price shock and the bankruptcy of Franklin National Bank in 1974. After declining markedly during the second half of the 1970s, the shift in US monetary policy towards aggressively fight-

[5]Hereby we refer to the one-step-ahead forecast error related to a given time series.

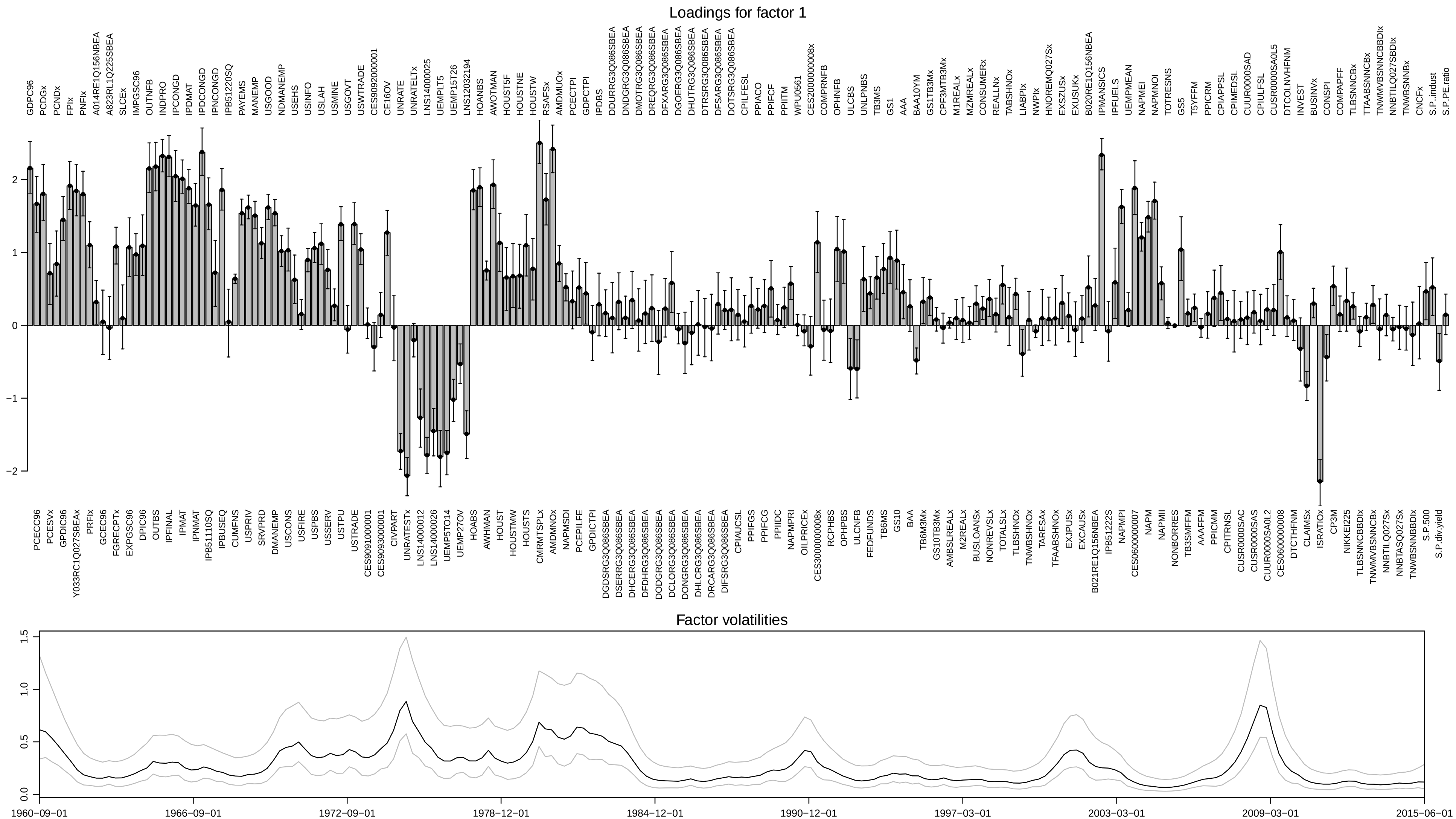

Fig. 1: 5th, 50th, and 95th posterior percentiles of factor loadings (upper panel) and factor volatility (lower panel).

ing inflation and the second oil price shock again translate into higher macroeconomic uncertainty. Note that from the mid 1980s onward we observe a general decline in macroeconomic volatility that lasts until the beginning of the 1990s. There we observe a slight increase in volatility possibly caused by the events surrounding the first gulf war. The remaining years up to the beginning of the 2000s has been relatively unspectacular, with volatility levels being muted most of the time. In 2000/2001, volatility again increases due to the burst of the dot-com bubble and the 9/11 terrorist attacks. Finally, we observe marked spikes in volatility during recessionary episodes like the recent financial crisis in 2008.

Finally, we assess how well the DL prior with $a = 1/K$ performs in shrinking the coefficients in $\boldsymbol{B}$ to zero. Fig. 2 depicts a heatmap that gives a rough feeling on the size of each regression coefficient based on the posterior median of $\boldsymbol{B}$ and Fig. 3 depicts the posterior interquartile range, providing some evidence on posterior uncertainty.[6] The DL prior apparently succeeds in shrinking the vast majority of the approximately $180\,000$ coefficients towards zero. Looking at the importance of each lag of $\boldsymbol{y}_t$ suggests that the first lag proves to be important for the vast majority of variables included whereas the second lag also plays a prominent role. For the third and fourth lag we occasionally observe non-zero signals but with no clear pattern along the main diagonal of $\boldsymbol{A}_3$ and $\boldsymbol{A}_4$. This indicates that our prior provides enough flexibility that the first "own" lag of a given variable is typically non-zero whereas this pattern vanishes if higher lag orders are considered. More specifically, for lag orders higher than two we see that higher lags of other variables tend to be included only for very few elements of $\boldsymbol{y}_t$.

Interestingly, for selected time series measuring inflation (both consumer and producer price inflation) we find that lags of monetary aggregates are allowed to load on the respective inflation series. This result points towards a big advantage of our proposed prior relative to standard VAR priors in the Minnesota tradition: while these priors also tend to work relatively well in huge dimensions they also show a tendency to overshrink when the overall tightness of the prior is integrated out in a Bayesian framework, effectively pushing the posterior distribution of $\boldsymbol{B}$ towards the prior mean and thus ruling out patterns observed under the DL prior.

Inspection of the interquartile range also indicates that the proposed shrinkage prior succeeds in reducing posterior uncertainty markedly. Note that the pattern found for the posterior median of $\boldsymbol{B}$ can also be found in terms of the posterior dispersion. We again see that the coefficients associated with the first, own lag of a given variable are allowed to be non-zero whereas in most other cases the associated posterior is strongly

[6]Since the corresponding posterior distribution is quite heavy-tailed, using posterior standard deviations, while providing a qualitatively similar picture, tend to be slightly exaggerated.

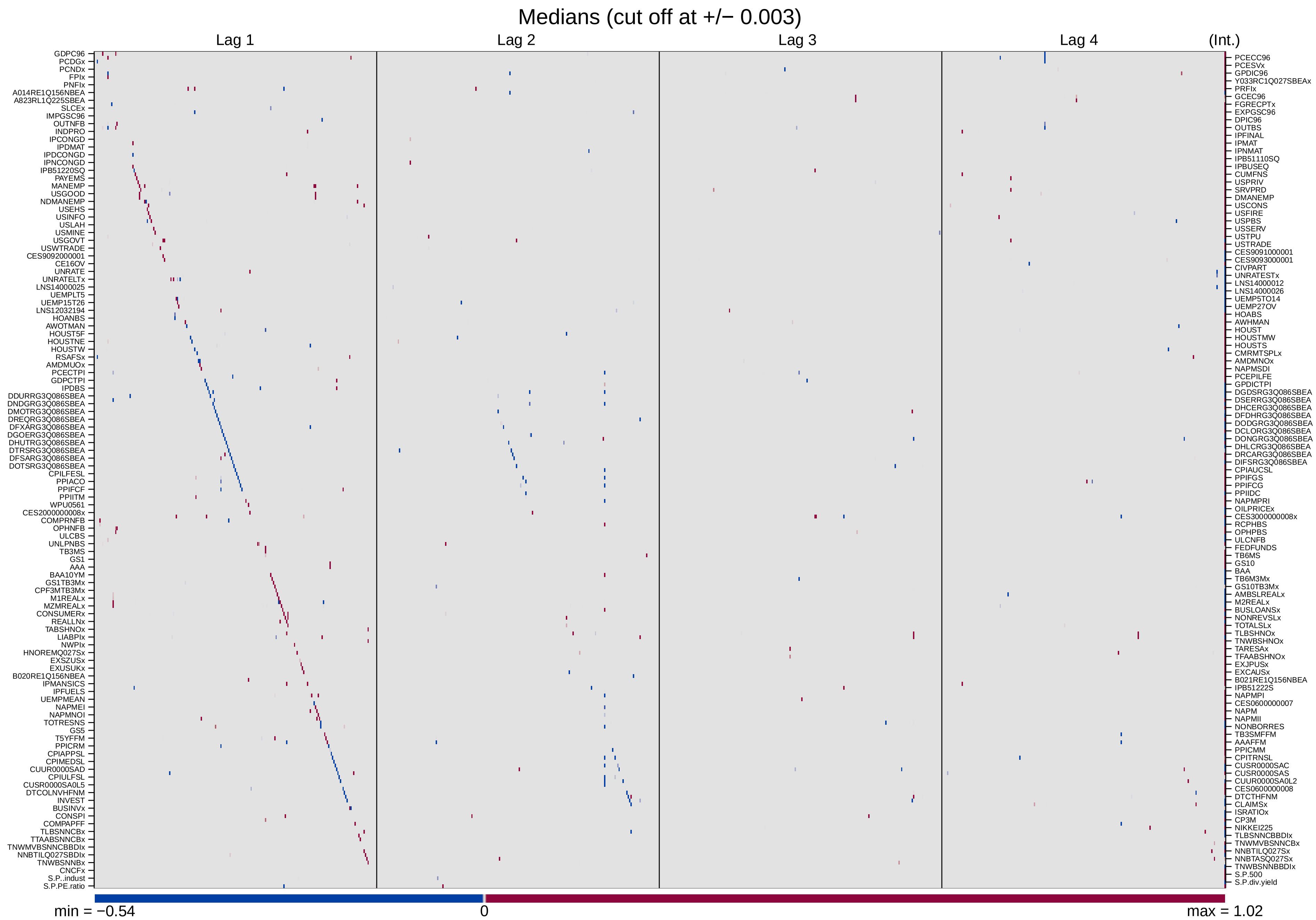

Fig. 2: Posterior medians of VAR coefficients, $a = 1/K = 1/185115$.

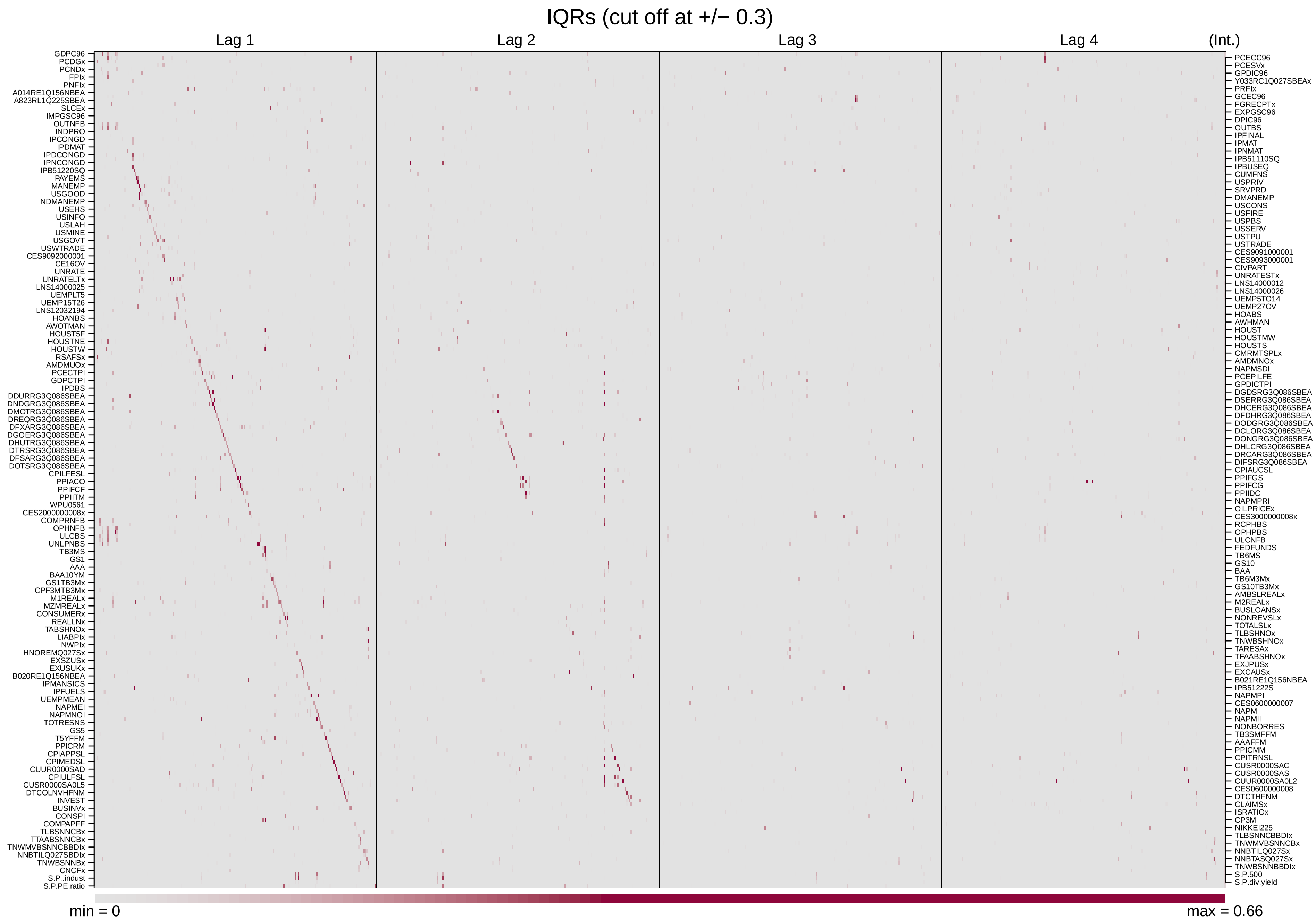

Fig. 3: Posterior interquartile ranges of VAR coefficients, $a = 1/K = 1/185115$.

concentrated around zero. For comparison, we provide heatmaps of posterior medians and interquartile ranges for $a = 1/2$ in the Appendix, see Figs. 6 and 7.

5.3 Some predictive evidence

We focus at forecasting gross domestic product (GDPC96), industrial production (INPRO), total nonfarm payroll (PAYEMS), civilian unemployment rate (UNRATE), new privately owned housing units started (HOUST), consumer price index inflation (CPIAUCSL), producer price index for finished goods inflation (PPIFGS), effective federal funds rate (FEDFUNDS), 10-year treasury constant maturity rate (GS10), U.S./U.K. exchange rate (EXUSUKx), and the S&P 500 (S.P.500). This choice includes the variables investigates by Pettenuzzo et al. (2016) and some additional important macroeconomic indicators that are commonly monitored by practitioners, resulting in a total of 11 series.

To assess the forecasting performance of our model, we conduct a pseudo out-of-sample forecasting exercise with initial estimation sample ranging from 1959:Q3 to 1990:Q2. Based on this estimation period we compute one-quarter-ahead predictive densities for the first period in the hold-out (i.e. 1990:Q3). After obtaining the corresponding predictive densities and evaluating the corresponding log predictive likelihoods, we expand the estimation period and re-estimate the model. This procedure is repeated 100 times until the final point of the full sample is reached. The quarterly scores obtained this way are then accumulated.

Our model with factors $q \in \{0, 1 \ldots, 5\}$ is benchmarked against the prior model, a pure factor stochastic volatility (FSV) model with conditional mean equal to zero (i.e. $\boldsymbol{B} = \mathbf{0}_{m \times k}$). In what follows we label this specification FSV 0. Moreover, we include two models that impose the restriction that $\boldsymbol{A}_1 = \boldsymbol{I}_m$ and $\boldsymbol{A}_1 = 0.8 \times \boldsymbol{I}_m$ while $\boldsymbol{A}_j$ for $j > 1$ are set equal to zero matrices in both cases. The first model, labeled FSV 1, assumes that the conditional mean of $\boldsymbol{y}_t$ follows a random walk process and the second specification, denoted as FSV 0.8, imposes the restriction that the variables in $\boldsymbol{y}_t$ feature a rather strong degree of persistence but are stationary. These two cases serve to evaluate whether it pays off to adopt our flexible shrinkage prior relative to a model that is closely related to a Minnesota prior in high dimensions. Since a hierarchical Minnesota prior with a single shrinkage parameter typically induces heavy amounts of shrinkage on $\boldsymbol{B}$, strongly pushing the posterior median of $\boldsymbol{B}$ towards the prior mean matrix, the different FSV models can also be understood as a Minnesota prior VAR with a shrinkage parameter very close to zero. The exercise serves to evaluate whether it pays off to impose a VAR structure on the first moment of the joint density of our

	0	1	2	3	4	5
VAR-FSV	-1225	-1108	-1033	-1026	-1014	-1026
FSV 0	-1208	-1108	-1102	-1066	-1061	-1071
FSV 0.8	-1178	-1132	-1106	-1087	-1096	-1125
FSV 1	-1171	-1128	-1095	-1095	-1107	-1118

Table 1: Overall log predictive scores for the number of factors $q \in \{0, 1, \ldots, 5\}$ for the VAR-FSV as well as the competing FSV models. Larger numbers indicate better joint predictive density performance for 11 variables of interest.

data and to assess how many factors are needed to obtain precise multivariate density predictions for our 11 variables of interest.

Overall log predictive scores are summarized in Table 1. An immediate finding is that ignoring the error covariance structure (using zero factors) produces rather inaccurate forecasts for all models considered. While a single factor model improves predictive accuracy by a large margin, even more factors (i.e. even more flexible modeling of the covariance structure) further increases the forecasting performance. For this exercise, we identify four factors to be a reasonable choice when the joint log predictive scores of the aforementioned variables are considered. We would like to stress that this choice critically depends on the number of variables we include in our prediction set. If we focus attention on the marginal predictive densities (i.e. the univariate predictive densities obtained after integrating out the remaining elements in $\boldsymbol{y}_t$) we find that fewer or even no factors receive more support whereas in the case of higher dimensional prediction sets we generally observe that more factors lead to more accurate density predictions.

Comparing the differences between the benchmark FSV 0 model and our proposed VAR-FSV we find that for the zero factor case, the FSV 0 model outperforms a specification that explicitly models the conditional mean of the process to follow an VAR(p) process while it is beaten by both specifications that introduce persistence on the first, own lag of a given variable. We conjecture that the inclusion of the factors has pronounced effects on the estimates of $\boldsymbol{B}$ and this might lead to an increase in predictive accuracy. Intuitively this might be explained by the fact that if the underlying data generating process suggests that $\boldsymbol{B}$ is time-varying, inclusion of the factors might alleviate issues associated with model misspecification by, at least to a certain extent, controlling for structural breaks in $\boldsymbol{B}$. Looking at the differences between both specifications that assume that the conditional mean of $\boldsymbol{y}_t$ is highly persistent and the VAR-FSV indicates that after including at least a single factor, the VAR-FSV and the FSV 0 outperform both, the FSV 0.8 and the FSV 1, competing specifications.

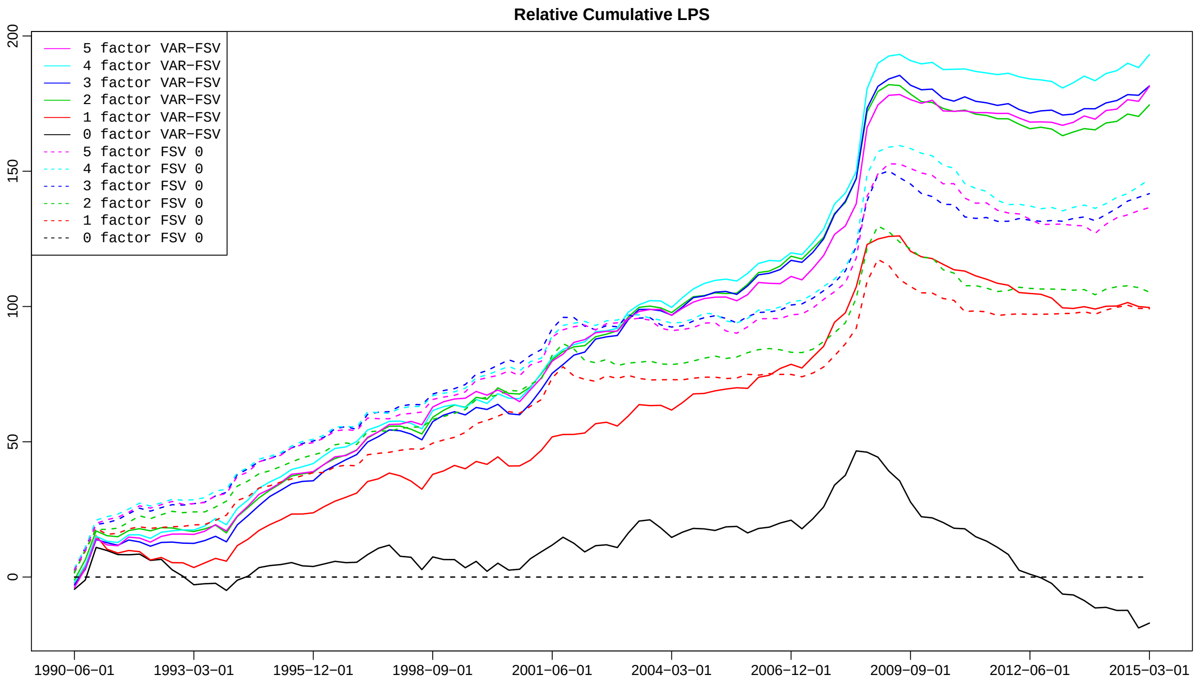

Fig. 4: Marginal cumulative log predictive scores over time, relative to a zero-mean model with independent stochastic volatility models for all component series. Higher values correspond to better one-quarter-ahead density predictions up to the corresponding point in time.

For a better understanding of how the values in Table 1 come to life, Fig. 4 visualizes the cumulative log predictive scores (LPS) relative to the zero-factor FSV model over time. Since the FSV 0 model proves to be the strongest competitor and the pattern of the LPS appears to be similar across specifications, we exclude the FSV 0.8 and the FSV 1 from this figure.

The benefit of the flexible SV structure in the VAR residuals is particularly pronounced during the 2008 financial crisis which can be seen by comparing the colored lines to black ones. During this period, time-varying covariance modeling appears to be of great importance and the performance of models that ignore contemporaneous dependence deteriorates. This finding is in line with, e.g., Kastner (2016), who reports analogous results for US asset returns. This increase in predictive accuracy can be traced back to the fact that within an economic downturn, the correlation structure of our dataset changes markedly, with most indicators that measure real activity sharply declining in lockstep. A model that takes contemporaneous cross-variable linkages seriously is thus able to fully exploit such behavior which in turn improves predictions.

A careful analysis of whether the VAR structure on the mean actually improves forecasting performance leads to an affirmative conclusion only after the turn of the millennium, where models featuring the vector autoregressive mean structure tend to do better than their zero-mean counterparts. Note that the period surrounding the burst of the dot-com bubble and the 9/11 terrorist attacks appears to be the first time frame where VAR models sharply improve upon the FSV specification. After 2002, most models that explicitly cater for the conditional mean continue to improve upon the different FSV specifications. The accuracy gains are particularly pronounced through the 2008 financial crisis, with the gap widening even further. This trend continues until the end of the sample if the VAR residuals are allowed to have two or more latent factors.

5.4 A note on the computational burden

Even though the efficient sampling schemes outlined in this paper help to overcome absolutely prohibitive computational burdens, the CPU time needed to perform fully Bayesian inference in a model of this size can still be considered substantial. In what follows we shed light on the estimation time required and how it is related to the length of the time series T, the lag length p and to the number of latent factors $q \in \{0, 50\}$. Fig. 5 shows the time needed to perform a single draw from the joint posterior distribution of the $215 + 215^2p$ coefficients and their corresponding $2(215 + 215^2p) + 1$ auxiliary shrinkage quantities, the qT factor realizations and the associated $215q$ loadings, alongside $(T + 1)(215 + q)$ latent volatilities with their corresponding $645 + 2q$ parameters. This amounts to $166\,841$ random draws for the smallest model considered (one lag, no factors, $T = 124$) and $776\,341$ random draws for the largest model (5 lags, 50 factors, $T = 224$) at each MCMC iteration.

As mentioned above, the computation time rises approximately linearly with the number of lags included. Dotted lines indicate the time in seconds needed to perform a single draw from a model with 50 factors included while solid lines refer to the time needed to estimate a model without factors and a diagonal time-varying variance-covariance matrix $\mathbf{\Sigma}_t$. Interestingly, the additional complexity when moving from a model without factors to a highly parameterized model with 50 factors appears to be negligible, increasing the time needed by a fraction of a second on average. The important role of the length of the sample can be seen by comparing the green, red and black lines. Here it can be seen that the time necessary to perform a simple MCMC draw quickly rises with the length of our sample, consistent with the statements made in Section 2.4. This feature of our algorithm, however, is convenient especially when researchers are interested in combining many short time series or performing recursive forecasting based on a tiny initial estimation sample.

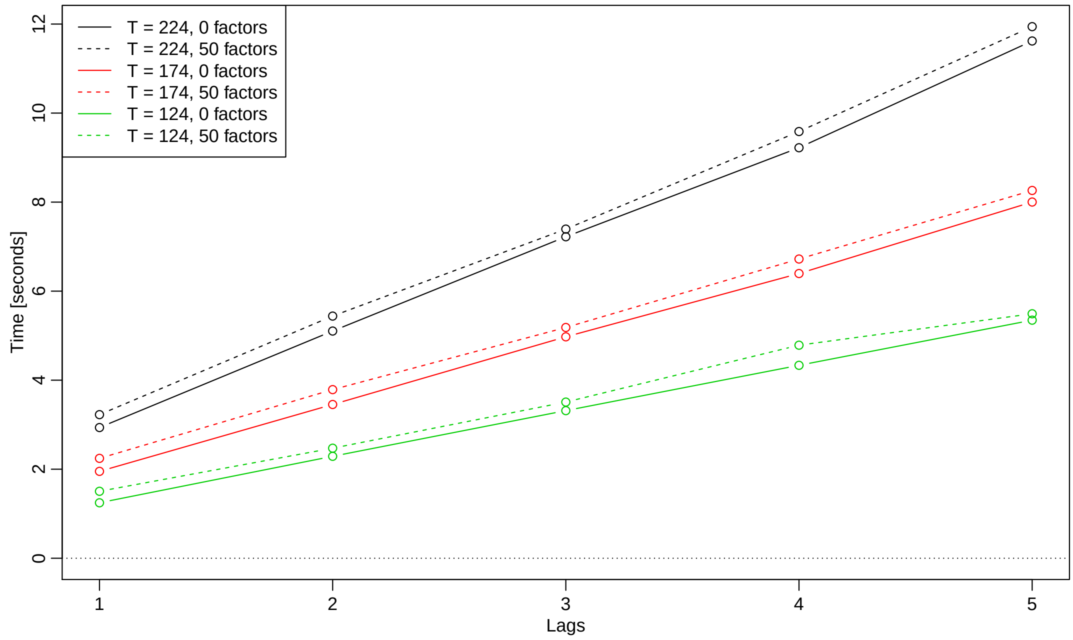

Fig. 5: Empirical CPU times for each MCMC iteration on a standard laptop computer using one core. Time series lengths are $T \in \{124, 174, 224\}$; the numbers of latent factors are $q \in \{0, 50\}$.

6 Closing remarks

In this paper we propose an alternative route to estimate huge dimensional VAR models that allow for time-variation in the error variances. The Dirichlet-Laplace prior, a recent variant of a global-local shrinkage prior, enables us to heavily shrink the parameter space towards the prior model while providing enough flexibility that individual regression coefficients are allowed to be unrestricted. This prior setup alleviates overfitting issues generally associated with large VAR models. To cope with computational issues we assume that the one-step-ahead forecast errors of the VAR feature a factor stochastic volatility structure that enables us to perform equation-by-equation estimation, conditional on the loadings and the factors. Since posterior simulation of each equation's autoregressive parameters involves manipulating large matrices, we implement an alternative recent algorithm that improves upon existing methods by large margins, rendering a fully fledged Bayesian estimation of truly huge systems possible.

In an empirical application we first present various key features of our approach based on a single factor model. This single factor which summarizes the joint dynamics of the VAR errors can be interpreted as an uncertainty measure that closely tracks observed factors such as the volatility index. The question whether such a simplistic structure proves to be an adequate representation of the time-varying covariance matrix naturally arises and we thus provide a detailed forecasting exercise to evaluate the merits of our approach relative to the prior model and a set of competing models with a different number of latent factors in the errors.

As a possible avenue of further research we would like to stress that our approach could also be used to estimate huge dimensional time-varying parameter VAR models with stochastic volatility. To cope with the computational difficulties associated with the vast state space, a possible solution could be to rely on an additional layer of hierarchy that imposes a factor structure on the time-varying autoregressive coefficients and thus reduce the computational burden considerably.

References

Aguilar O and West M (2000) Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics* 18(3), 338–357

Bańbura M, Giannone D and Reichlin L (2010) Large Bayesian vector auto regressions. *Journal of Applied Econometrics* 25(1), 71–92

Bhattacharya A, Chakraborty A and Mallick BK (2016) Fast sampling with Gaussian scale-mixture priors in high-dimensional regression. *Biometrika* 4(103), 985–991

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. ArXiv pre-print 1611.01310

Carriero A, Clark T and Marcellino M (2015) Large Vector Autoregressions with Asymmetric Priors. Working Papers 759, Queen Mary University of London, School of Economics and Finance

Carriero A, Clark TE and Marcellino M (2016) Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics* 34(3), 375–390

Carvalho CM, Polson NG and Scott JG (2010) The horseshoe estimator for sparse signals. *Biometrika* 97(2), 465–480

Clark TE (2011) Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics* 29(3), 327–341

Clark TE and Ravazzolo F (2015) Macroeconomic Forecasting Performance under Alternative Specifications of Time-Varying Volatility. *Journal of Applied Econometrics* 30(4), 551–575

Cogley T and Sargent TJ (2002) Evolving post-World War II U.S. inflation dynamics. In Bernanke BS and Rogoff K, eds., *NBER Macroeconomics Annual 2001*. National Bureau of Economic Research, Inc, 331–388

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

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

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

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* (forthcoming)

Jurado K, Ludvigson SC and Ng S (2015) Measuring uncertainty. *The American Economic Review* 105(3), 1177–1216

Kastner G (2016) Sparse Bayesian time-varying covariance estimation in many dimensions ArXiv preprint arXiv:1608.08468

Kastner G (2017) *factorstochvol: Bayesian estimation of (sparse) latent factor stochastic volatility models*. R package version 0.8.4

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

Kastner G, Frühwirth-Schnatter S and Lopes HF (2018) Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models. *Journal of Computational and Graphical Statistics* (forthcoming). ArXiv preprint arXiv:1602.08154

Koop G and Korobilis D (2013) Large time-varying parameter VARs. *Journal of Econometrics* 177(2), 185–198

Koop G, Leon-Gonzalez R and Strachan RW (2009) On the evolution of the monetary policy transmission mechanism. *Journal of Economic Dynamics and Control* 33(4), 997–1017

Koop GM (2013) Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics* 28(2), 177–203

Korobilis D and Pettenuzzo D (2016) Adaptive Minnesota Prior for High-Dimensional Vector Autoregressions. Essex finance centre working papers, University of Essex, Essex Business School

Leydold J and Hörmann W (2015) *GIGrvg: Random variate generator for the GIG distribution*. R package version 0.4

Litterman R (1986) Forecasting with Bayesian vector autoregressions – Five years of experience. *Journal of Business and Economic Statistics* 4(1), 25–38

McCracken MW and Ng S (2016) FRED-MD: A monthly database for macroeconomic research. *Journal of Business & Economic Statistics* 34(4), 574–589

Mumtaz H and Theodoridis K (2017) The changing transmission of uncertainty shocks in the US: An empirical analysis. *Journal of Business & Economic Statistics* (forthcoming), 1–39

Pati D, Bhattacharya A, Pillai NS and Dunson D (2014) Posterior contraction in sparse Bayesian factor models for massive covariance matrices. *Annals of Statistics* 42(3), 1102–1130

Pettenuzzo D, Koop G and Korobilis D (2016) Bayesian Compressed Vector Autoregressions. Working Papers 103, Brandeis University, Department of Economics and International Businesss School

Pitt MK and Shephard N (1999) Time-varying covariances: A factor stochastic volatility approach. In Bernardo JM, Berger JO, Dawid AP and Smith AFM, eds., *Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting*. Oxford University Press, 547–570

Polson NG and Scott JG (2011) Shrink globally, act locally: Sparse Bayesian regularization and prediction. In Bernardo JM, Bayarri MJ, Berger JO, Dawid AP, Heckerman D, Smith AFM and West M, eds., *Bayesian Statistics 9 – Proceedings of the Ninth Valencia International Meeting*. Oxford University Press, 501–538

Primiceri GE (2005) Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies* 72(3), 821–852

Sims CA and Zha T (1998) Bayesian methods for dynamic multivariate models. *International Economic Review* 39(4), 949–968

Sims CA and Zha T (2006) Were there regime switches in U.S. monetary policy? *American Economic Review* 96(1), 54–81

Stock JH and Watson MW (2005) Understanding changes in international business cycle dynamics. *Journal of the European Economic Association* 3(5), 968–1006

Stock JH and Watson MW (2011) Dynamic Factor Models. In Clements MP and Henry DF, eds., *Oxford Handbook of Economic Forecasting*. Oxford University Press, Oxford, 35–59

A Additional results

This appendix contains some additional results. Fig. 6 displays the posterior median estimates for the real data example discussed in Section 5 of the main paper when the shrinkage parameter a is chosen to be $1/2$ (cf. Bhattacharya et al., 2015, for a discussion of this choice). While $a = 1/2$ appears to provide a fair amount of shrinkage in other applications, for our huge dimensional example this prior exerts only relatively little shrinkage and appears to lead to overfitting. The diagonal pattern in the first two lags known from Fig. 2 in the main paper appears here as well, but there is a considerable amount of nonzero medians elsewhere. Correspondingly, the interquartile ranges visualized in Fig. 7 are also very large compared to those obtained under the much more tight prior used in the main paper (Fig. 3).

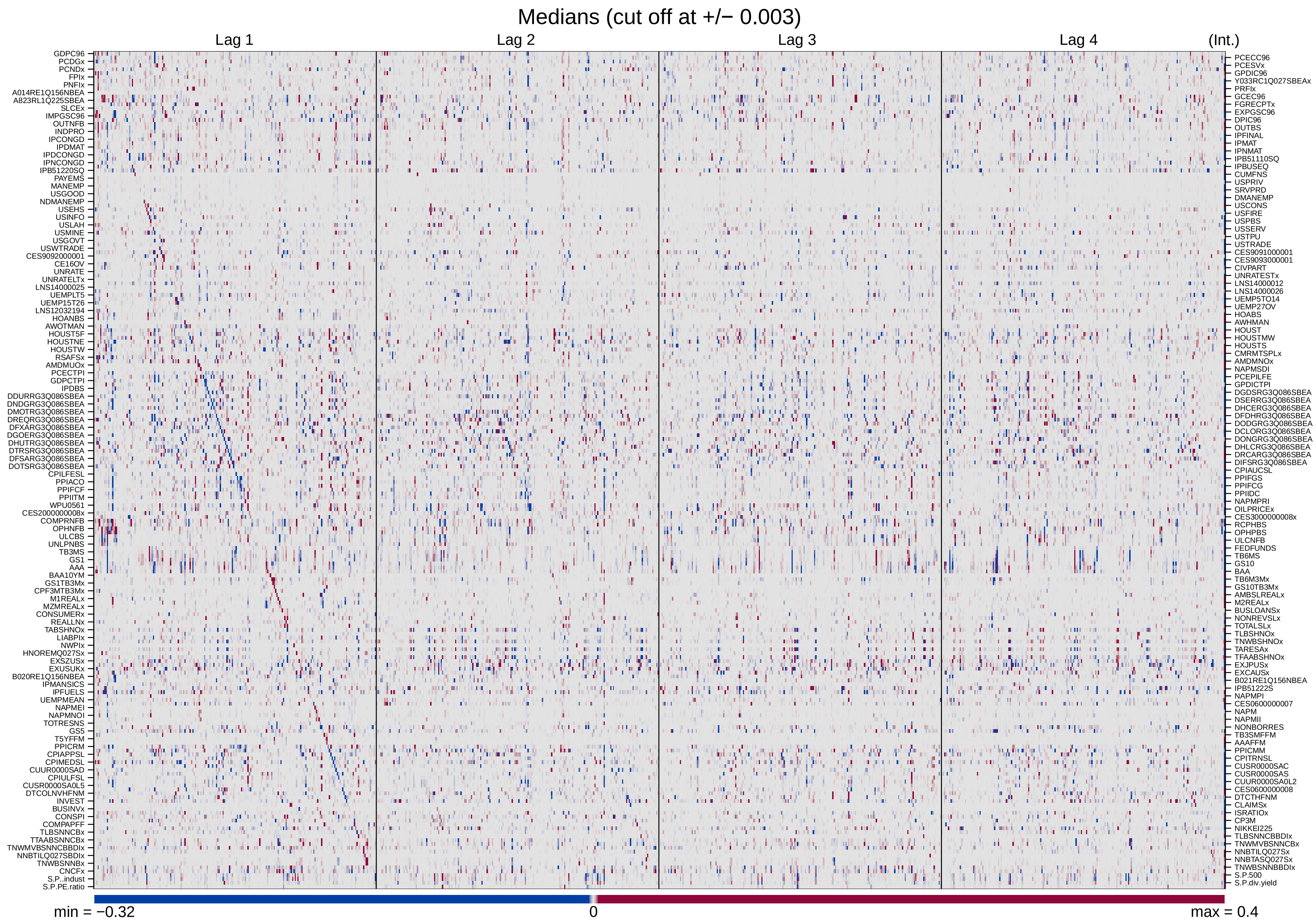

Fig. 6: Posterior medians of VAR coefficients, $a = 1/2$.

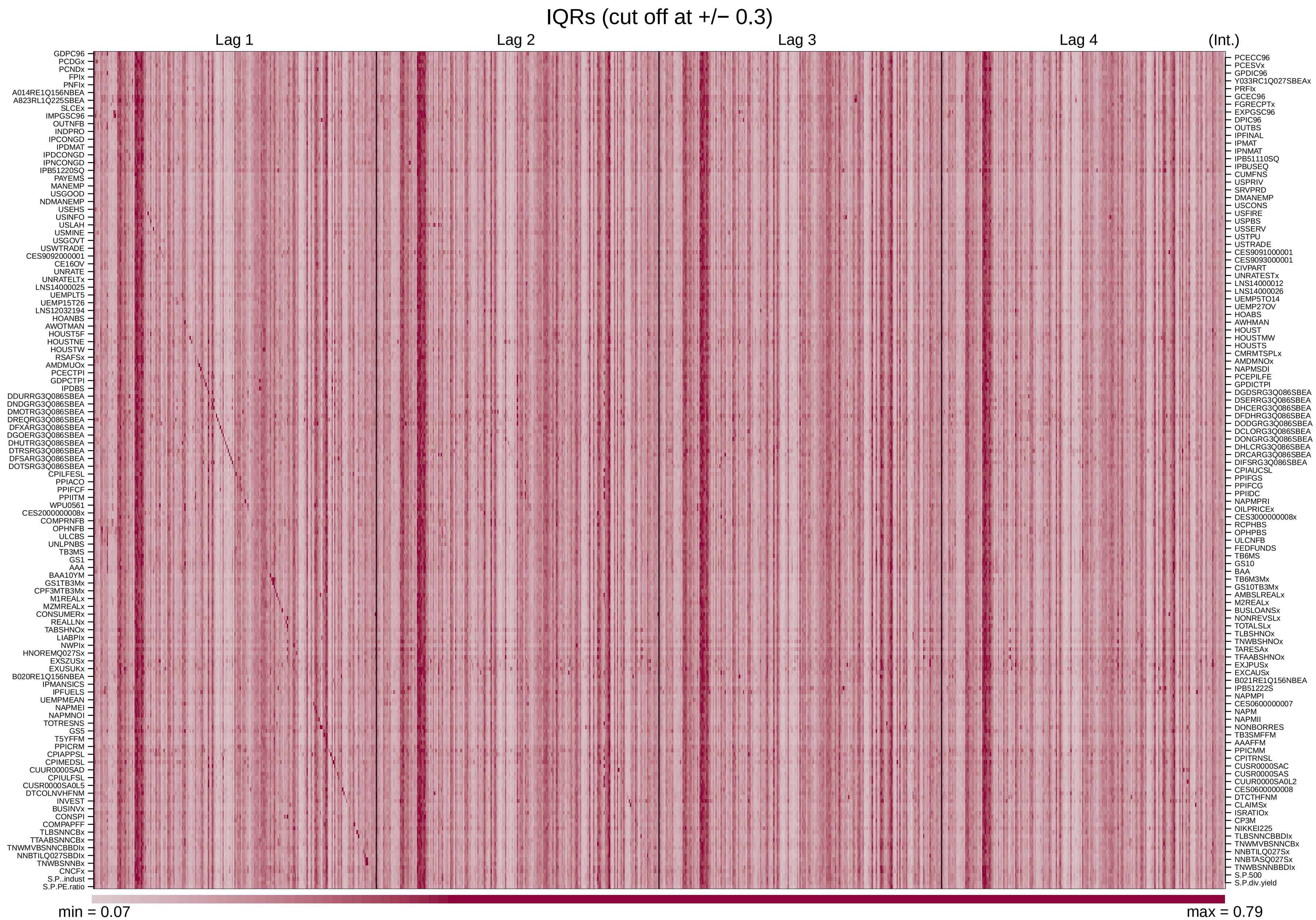

Fig. 7: Posterior interquartile ranges of VAR coefficients, $a = 1/2$.