Journal of Econometrics 210 (2019) 98–115

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Sparse Bayesian time-varying covariance estimation in many dimensions

Gregor Kastner *

WU Vienna University of Economics and Business, Austria

ARTICLE INFO

Article history:
Available online 12 November 2018

JEL classification:
C32
C38
C53
C58
G11

Keywords:
Dynamic correlation
Factor stochastic volatility
Curse of dimensionality
Shrinkage
Minimum variance portfolio

ABSTRACT

We address the curse of dimensionality in dynamic covariance estimation by modeling the underlying co-volatility dynamics of a time series vector through latent time-varying stochastic factors. The use of a global–local shrinkage prior for the elements of the factor loadings matrix pulls loadings on superfluous factors towards zero. To demonstrate the merits of the proposed framework, the model is applied to simulated data as well as to daily log-returns of 300 S&P 500 members. Our approach yields precise correlation estimates, strong implied minimum variance portfolio performance and superior forecasting accuracy in terms of log predictive scores when compared to typical benchmarks.

© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

The joint analysis of hundreds or even thousands of time series exhibiting a potentially time-varying variance–covariance structure has been on numerous research agendas for well over a decade. In the present paper we aim to strike the indispensable balance between the necessary flexibility and parameter parsimony by using a factor stochastic volatility (SV) model in combination with a global–local shrinkage prior. Our contribution is threefold. First, the proposed approach offers a hybrid cure to the curse of dimensionality by combining *parsimony* (through imposing a factor structure) with *sparsity* (through employing computationally efficient absolutely continuous shrinkage priors on the factor loadings). Second, the efficient construction of posterior simulators allows for conducting Bayesian inference and prediction in very high dimensions via carefully crafted Markov chain Monte Carlo (MCMC) methods made available to end-users through the R (R Core Team, 2017) package **factorstochvol** (Kastner, 2017). Third, we show that the proposed method is capable of accurately predicting covariance and precision matrices which we assess via statistical and economic forecast evaluation in several simulation studies and an extensive real-world example.

Concerning factor SV modeling, early key references include Harvey et al. (1994), Pitt and Shephard (1999) and Aguilar and West (2000) which were later picked up and extended by e.g. Philipov and Glickman (2006), Chib et al. (2006), Han (2006), Lopes and Carvalho (2007), Nakajima and West (2013), Zhou et al. (2014) and Ishihara and Omori (2017). While reducing the dimensionality of the problem at hand, models with many factors are still rather rich in parameters. Thus, we further shrink unimportant elements of the factor loadings matrix to zero in an automatic way within a Bayesian framework. This approach

* Correspondence to: Department of Finance, Accounting and Statistics, WU Vienna University of Economics and Business, Welthandelsplatz 1/D4/4, 1020 Vienna, Austria.
E-mail address: gregor.kastner@wu.ac.at.

https://doi.org/10.1016/j.jeconom.2018.11.007
0304-4076/© 2018 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

is inspired by high-dimensional regression problems where the number of parameters frequently exceeds the size of the data. In particular, we adopt the approach brought forward by Caron and Doucet (2008) and Griffin and Brown (2010) who suggest to use a special continuous prior structure – the Normal-Gamma prior – on the regression parameters (in our case the factor loadings matrix). This shrinkage prior is a generalization of the Bayesian Lasso (Park and Casella, 2008) and has recently received attention in the econometrics literature (Bitto and Frühwirth-Schnatter, 2018; Huber and Feldkircher, 2017).

Another major issue for such high-dimensional problems is the computational burden that goes along with statistical inference, in particular when joint modeling is attempted instead of multi-step approaches or rolling-window-like estimates. Suggested solutions include Engle and Kelly (2012) who propose an estimator assuming that pairwise correlations are equal at every point in time, Pakel et al. (2017) who consider composite likelihood estimation, Gruber and West (2016) who use a decoupling–recoupling strategy to parallelize estimation (executed on graphical processors), Lopes et al. (2018) who treat the Cholesky-decomposed covariance matrix within the framework of Bayesian time-varying parameter models, and Oh and Patton (2017) who choose a copula-based approach to link separately estimated univariate models. We propose to use a Gibbs-type sampler which allows to jointly take into account both parameter as well as sampling uncertainty in a finite-sample setup through fully Bayesian inference, thereby enabling inherent uncertainty quantification. Additionally, this approach allows for fully probabilistic in- and out-of-sample density predictions.

For related work on sparse Bayesian prior distributions in high dimensions, see e.g. Kaufmann and Schumacher (2018) who use a point mass prior specification for factor loadings in dynamic factor models or Ahelegbey et al. (2016) who use a graphical representation of vector autoregressive models to select sparse graphs. From a mathematical point of view, Pati et al. (2014) investigate posterior contraction rates for a related class of continuous shrinkage priors for static factor models and show excellent performance in terms of posterior rates of convergence with respect to the minimax rate. All of these works, however, assume homoskedasticity and are thus potentially misspecified when applied to financial or economic data. For related methods that take into account heteroskedasticity, see e.g. Nakajima and West (2013, 2017) who employ a latent thresholding process to enforce time-varying sparsity. Moreover, Zhao et al. (2016) approach this issue via dependence networks, Loddo et al. (2011) use stochastic search for model selection, and Baştürk et al. (2018) use time-varying combinations of dynamic models and equity momentum strategies. These methods are typically very flexible in terms of the dynamics they can capture but are applied to moderate dimensional data only.

We illustrate the merits of our approach through extensive simulation studies and an in-depth financial application using 300 S&P 500 members. In simulations, we find considerable evidence that the Normal-Gamma shrinkage prior leads to substantially sparser factor loadings matrices which in turn translate into more precise correlation estimates when compared to the usual Gaussian prior on the loadings.[1] In the real-world application, we evaluate our model against a wide range of alternative specifications via log predictive scores and minimum variance portfolio returns. Factor SV models with sufficiently many factors turn out to imply extremely competitive portfolios in relation to well-established methods which typically have been specifically tailored for such applications. Concerning density forecasts, we find that our approach outperforms all included competitors by a large margin.

The remainder of this paper is structured as follows. In Section 2, the factor SV model is specified and the choice of prior distributions is discussed. Section 3 treats statistical inference via MCMC methods and sheds light on computational aspects concerning out-of-sample density predictions for this model class. Extensive simulation studies are presented in Section 4, where the effect of the Normal-Gamma prior on correlation estimates is investigated in detail. In Section 5, the model is applied to 300 S&P 500 members. Section 6 wraps up and points out possible directions for future research.

2. Model specification

Consider an m-variate zero-mean return vector $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{mt})'$ for time $t = 1, \ldots, T$ whose conditional distribution is Gaussian, i.e.

$$\boldsymbol{y}_t | \boldsymbol{\Sigma}_t \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{\Sigma}_t)\,.$$

2.1. Factor SV model

To reduce dimensionality, factor SV models utilize a decomposition of the $m \times m$ covariance matrix $\boldsymbol{\Sigma}_t$ with $m(m+1)/2$ free elements into a factor loadings matrix $\boldsymbol{\Lambda}$ of size $m \times r$, an r-dimensional diagonal matrix $\boldsymbol{V}_t$ and an m-dimensional diagonal matrix $\boldsymbol{U}_t$ in the following fashion:

$$\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}\boldsymbol{V}_t\boldsymbol{\Lambda}' + \boldsymbol{U}_t. \tag{1}$$

This reduces the number of free elements to $mr + m + r$. Because r is typically chosen to be much smaller than m, this specification constrains the parameter space substantially, thereby inducing parameter parsimony. For the paper at hand, $\boldsymbol{\Lambda}$ is considered to be time invariant whereas the elements of both $\boldsymbol{V}_t$ and $\boldsymbol{U}_t$ are allowed to evolve over time through

[1] Note that in contrast to e.g. Frühwirth-Schnatter and Tüchler (2008), we do not attempt to identify exact zeros in a covariance matrix. Rather, we aim to find a parsimonious factor representation of the underlying heteroskedastic data which may (or may not) imply covariances that are close to zero at times.

parametric stochastic volatility models, i.e. $\boldsymbol{U}_t = \text{diag}(\exp(h_{1t}), \ldots, \exp(h_{mt}))$ and $\boldsymbol{V}_t = \text{diag}\big(\exp(h_{m+1,t}), \ldots, \exp(h_{m+r,t})\big)$ with

$$h_{it} \sim \mathcal{N}\big(\mu_i + \phi_i(h_{i,t-1} - \mu_i), \sigma_i^2\big), \quad i = 1, \ldots, m, \tag{2}$$

$$h_{m+j,t} \sim \mathcal{N}\big(\phi_{m+j} h_{m+j,t-1}, \sigma_{m+j}^2\big), \quad j = 1, \ldots, r. \tag{3}$$

More specifically, $\boldsymbol{U}_t$ describes the idiosyncratic (series-specific) variances while $\boldsymbol{V}_t$ contains the variances of underlying orthogonal factors $\boldsymbol{f}_t \sim \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{V}_t)$ that govern the contemporaneous dependence. The autoregressive process in (3) is assumed to have mean zero to identify the unconditional scaling of the factors.

This setup is commonly written in the following hierarchical form (e.g. Chib et al., 2006):

$$\boldsymbol{y}_t | \boldsymbol{\Lambda}, \boldsymbol{f}_t, \boldsymbol{U}_t \sim \mathcal{N}_m(\boldsymbol{\Lambda}\boldsymbol{f}_t, \boldsymbol{U}_t), \quad \boldsymbol{f}_t | \boldsymbol{V}_t \sim \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{V}_t),$$

where the distributions are assumed to be conditionally independent for all points in time. To make further exposition clearer, let $\boldsymbol{y} = (\boldsymbol{y}_1 \cdots \boldsymbol{y}_T)$ denote the $m \times T$ matrix of all observations, $\boldsymbol{f} = (\boldsymbol{f}_1 \cdots \boldsymbol{f}_T)$ the $r \times T$ matrix of all latent factors and $\boldsymbol{h} = (\boldsymbol{h}_1' \cdots \boldsymbol{h}_{m+r}')'$ the $(T+1) \times (m+r)$ matrix of all $m+r$ log-variance processes $\boldsymbol{h}_i = (h_{i0}, h_{i1}, \ldots, h_{iT})$, $i = 1, \ldots, m+r$. The vector $\boldsymbol{\theta}_i = (\mu_i, \phi_i, \sigma_i)'$ is referred to as the vector of parameters where μ_i is the level, ϕ_i the persistence, and σ_i^2 the innovation variance of $\boldsymbol{h}_i$. To denote specific rows and columns of matrices, we use the "dot" notation, i.e. $\boldsymbol{X}_{i\cdot}$ refers to the ith row and $\boldsymbol{X}_{\cdot j}$ to the jth column of $\boldsymbol{X}$. The proportions of variances explained through the common factors for each component series, $C_{it} = 1 - U_{ii,t}/\Sigma_{ii,t}$ for $i = 1, \ldots, m$, are referred to as the *communalities*. Here, $U_{ii,t}$ and $\Sigma_{ii,t}$ denote the ith diagonal element of $\boldsymbol{U}_t$ and $\boldsymbol{\Sigma}_t$, respectively. As by construction $0 \leq U_{ii,t} \leq \Sigma_{ii,t}$, the communality for each component series and for all points in time lies between zero and one. The joint (overall) communality $C_t = m^{-1} \sum_{i=1}^m C_{it}$ is simply defined as the arithmetic mean over all series.

Three comments are in order. First, the variance–covariance decomposition in (1) can be rewritten as $\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}_t \boldsymbol{\Lambda}_t' + \boldsymbol{U}_t$ with $\boldsymbol{\Lambda}_t := \boldsymbol{\Lambda}\boldsymbol{V}_t^{1/2}$. An essential assumption within the factor framework is that both $\boldsymbol{V}_t$ as well as $\boldsymbol{U}_t$ are diagonal matrices. This implies that the factor loadings $\boldsymbol{\Lambda}_t$ are dynamic but can only vary column-wise over time. Consequently, the time-variability of $\boldsymbol{\Sigma}_t$'s off-diagonal elements are cross-sectionally restricted while its diagonal elements are allowed to move independently across series. Hence, the "strength" of a factor, i.e. its cross-sectional explanatory power, varies jointly for all series loading on it. Consequently, it is likely that more factors are needed to properly explain the co-volatility dynamics of a multivariate time series than in models which allow for completely unrestricted time-varying factor loadings (Lopes and Carvalho, 2007), correlated factors ($\boldsymbol{V}_t$ not diagonal, see Zhou et al., 2014), or approximate factor models ($\boldsymbol{U}_t$ not diagonal, see Bai and Ng, 2002). Our specification, however, is less prone to overfitting and has the significant advantage of vastly simplified computations.

Second, identifying loadings for latent factor models is a long-standing issue that goes back to at least Anderson and Rubin (1956). Even though this problem is alleviated somewhat when factors are allowed to exhibit conditional heteroskedasticity (Sentana and Fiorentini, 2001; Rigobon, 2003), most authors have chosen an upper triangular constraint of the loadings matrix with unit diagonal elements, thereby introducing dependence on the ordering of the data (see Frühwirth-Schnatter and Lopes, 2018). However, when estimation of the actual factor loadings is not the primary concern (but rather a means to estimate and predict the covariance structure), this issue is less striking because a unique identification of the loadings matrix is not necessary.[2] This allows leaving the factor loadings matrix completely unrestricted, thus rendering the method invariant with respect to the ordering of the series.

Third, note that even though the joint distribution of the data is conditionally Gaussian, its stationary distribution has thicker tails. Nevertheless, generalizations of the univariate SV model to cater for even more leptokurtic distributions (e.g. Liesenfeld and Jung, 2000) or asymmetry (e.g. Yu, 2005) can straightforwardly be incorporated in the current framework. All of these extensions, however, tend to increase both sampling inefficiency as well as running time considerably and could thus preclude inference in very high dimensions.

2.2. Prior distributions

The usual prior for each (unrestricted) element of the factor loadings matrix is a zero-mean Gaussian distribution, i.e. $\Lambda_{ij} \sim \mathcal{N}\big(0, \tau_{ij}^2\big)$ independently for each i and j, where $\tau_{ij}^2 \equiv \tau^2$ is a constant specified a priori (e.g. Pitt and Shephard, 1999; Aguilar and West, 2000; Chib et al., 2006; Ishihara and Omori, 2017; Kastner et al., 2017). To achieve more shrinkage, we model this variance hierarchically by placing a hyperprior on τ_{ij}^2. This approach is related to Bhattacharya and Dunson (2011) and Pati et al. (2014) who investigate a similar class of priors for homoskedastic factor models. More specifically, let

$$\Lambda_{ij} | \tau_{ij}^2 \sim \mathcal{N}\big(0, \tau_{ij}^2\big), \quad \tau_{ij}^2 | \lambda_i^2 \sim \mathcal{G}\big(a_i, a_i\lambda_i^2/2\big), \quad \lambda_i^2 \sim \mathcal{G}(c_i, d_i). \tag{4}$$

Intuitively, each prior variance τ_{ij}^2 provides element-wise shrinkage governed independently for each row by λ_i^2. By integrating out τ_{ij}^2 it can be seen that the conditional variance of $\Lambda_{ij} | \lambda_i^2$ is $2/\lambda_i^2$ and the excess kurtosis of Λ_{ij} is $3/a_i$. The hyperparameters a_i, c_i, and d_i are fixed a priori, whereas a_i in particular plays a crucial role for the amount of shrinkage this

[2] The conditional covariance matrix $\boldsymbol{\Sigma}_t = \boldsymbol{\Lambda}\boldsymbol{V}_t\boldsymbol{\Lambda}' + \boldsymbol{U}_t$ involves a rotation-invariant transformation of $\boldsymbol{\Lambda}$.

prior implies. Choosing a_i small enforces strong shrinkage towards zero, while choosing a_i large imposes little shrinkage. For more elaborate discussions on Bayesian shrinkage in general and the effect of a_i specifically, see Griffin and Brown (2010) and Polson and Scott (2011). Note that the Bayesian Lasso prior (Park and Casella, 2008) arises as a special case when $a_i = 1$.

One can see prior (4) as *row-wise shrinkage* with *element-wise adaption* in the sense that all variances in row i can be thought of as "random effects" from the same underlying distribution. In other words, each series has high and a priori independent mass not to load on any factors and thus can be thought of as *series-specific* shrinkage. For further aspects on introducing hierarchical prior structure via the Normal-Gamma distribution, see Griffin and Brown (2017) and Huber and Feldkircher (2017). Analogously, it turns out to be fruitful to also consider *column-wise shrinkage* with element-wise adaption, i.e.

$$\Lambda_{ij}|\tau_{ij}^2 \sim \mathcal{N}\big(0, \tau_{ij}^2\big), \quad \tau_{ij}^2|\lambda_j^2 \sim \mathcal{G}\big(a_j, a_j\lambda_j^2/2\big), \quad \lambda_j^2 \sim \mathcal{G}\big(c_j, d_j\big).$$

This means that each factor has high and a priori independent mass not to be loaded on by any series and thus can be thought of as *factor-specific* shrinkage.

Concerning the univariate SV priors, we follow Kastner and Frühwirth-Schnatter (2014). For the m idiosyncratic and r factor volatilities, the initial states h_{i0} are distributed according to the stationary distributions of (2) and (3), respectively. Furthermore, $p(\mu_i, \phi_i, \sigma_i) = p(\mu_i)p(\phi_i)p(\sigma_i)$, where the level $\mu_i \in \mathbb{R}$ is equipped with the usual Gaussian prior $\mu_i \sim \mathcal{N}\big(b_\mu, B_\mu\big)$, the persistence parameter $\phi_i \in (-1, 1)$ is implied by $(\phi_i + 1)/2 \sim \mathcal{B}(a_0, b_0)$ and the volatility of volatility parameter $\sigma_i \in \mathbb{R}^+$ is chosen according to $\sigma_i^2 \sim B_\sigma \chi_1^2 = \mathcal{G}(1/2, 1/(2B_\sigma))$.

3. Statistical inference

There are a number of methods to estimate factor SV models such as quasi-maximum likelihood (e.g. Harvey et al., 1994), simulated maximum likelihood (e.g. Liesenfeld and Richard, 2006; Jungbacker and Koopman, 2006), and Bayesian MCMC simulation (e.g. Pitt and Shephard, 1999; Aguilar and West, 2000; Chib et al., 2006; Han, 2006). For high dimensional problems of this kind, Bayesian MCMC estimation proves to be a very efficient estimation method because it allows simulating from the high dimensional joint posterior by drawing from lower dimensional conditional posteriors.

3.1. MCMC estimation

One substantial advantage of MCMC methods over other ways of learning about the posterior distribution is that it constitutes a modular approach due to the conditional nature of the sampling steps. Consequently, conditionally on the matrix of variances $\boldsymbol{\tau} = (\tau_{ij})_{1\le i\le m;\ 1\le j\le r}$, we can adapt the sampling steps of Kastner et al. (2017). For obtaining draws for $\boldsymbol{\tau}$, we follow Griffin and Brown (2010). The MCMC sampling steps for the factor SV model are:

1. For factors and idiosyncratic variances, obtain m conditionally independent draws of the idiosyncratic log-volatilities from $\boldsymbol{h}_i|\boldsymbol{y}_{i\cdot}, \boldsymbol{\Lambda}_{i\cdot}, \boldsymbol{f}, \mu_i, \phi_i, \sigma_i$ and their parameters from $\mu_i, \phi_i, \sigma_i|\boldsymbol{y}_{i\cdot}, \boldsymbol{\Lambda}_{i\cdot}, \boldsymbol{f}, \boldsymbol{h}_i$ for $i = 1, \dots, m$. Similarly, perform r updates for the factor log-volatilities from $\boldsymbol{h}_{m+j}|\boldsymbol{f}_{m+j,\cdot}, \phi_{m+j}, \sigma_{m+j}$ and their parameters from $\phi_{m+j}, \sigma_{m+j}|\boldsymbol{f}_{m+j,\cdot}, \boldsymbol{h}_{m+j}$ for $j = 1, \dots, r$. This amounts to $m + r$ univariate SV updates.[3]

2a. Row-wise shrinkage only: For $i = 1, \dots, m$, sample from

$$\lambda_i^2|\boldsymbol{\tau}_{i\cdot} \sim \mathcal{G}\left(c_i + a_i\tilde{r}, d_i + \frac{a_i}{2}\sum_{j=1}^{\tilde{r}}\tau_{ij}^2\right),$$

where $\tilde{r} = \min(i, r)$ if the loadings matrix is restricted to have zeros above the diagonal and $\tilde{r} = r$ in the case of an unrestricted loadings matrix. For $i = 1, \dots, m$ and $j = 1, \dots, \tilde{r}$, draw from $\tau_{ij}^2|\lambda_i, \Lambda_{ij} \sim \text{GIG}(a_i - \frac{1}{2}, a_i\lambda_i^2, \Lambda_{ij}^2)$.[4]

2b. Column-wise shrinkage only: For $j = 1, \dots, r$, sample from

$$\lambda_j^2|\boldsymbol{\tau}_{\cdot j} \sim \mathcal{G}\left(c_j + a_j(m - \tilde{j} + 1), d_j + \frac{a_j}{2}\sum_{i=\tilde{j}}^{m}\tau_{ij}^2\right),$$

where $\tilde{j} = j$ if the loadings matrix is restricted to have zeros above the diagonal and $\tilde{j} = 1$ otherwise. For $j = 1, \dots, r$ and $i = \tilde{j}, \dots, r$, draw from $\tau_{ij}^2|\lambda_j, \Lambda_{ij} \sim \text{GIG}(a_j - \frac{1}{2}, a_j\lambda_j^2, \Lambda_{ij}^2)$.[4]

[3] There is a vast body of literature on efficiently sampling univariate SV models. For the paper at hand, we use R package **stochvol** (Kastner, 2016).

[4] The Generalized Inverse Gaussian distribution GIG(m, k, l) has a density proportional to $x^{m-1}\exp\left\{-\frac{1}{2}(kx + l/x)\right\}$. To draw from this distribution, we use the algorithm described in Hörmann and Leydold (2013) which is implemented in the R package **GIGrvg** (Leydold and Hörmann, 2017).

Table 1
Empirically obtained runtime per MCMC iteration on a single i5-5300U CPU (2.30 GHz) core running Xubuntu Linux 16.04 using R 3.3.3 linked against the default linear algebra packages as well as Intel MKL (single thread). Measured in milliseconds for $m \in \{10, 100, 500\}$ dimensional time series of length $T = 1000$.

	$r = 0$	$r = 1$	$r = 5$	$r = 10$	$r = 20$	$r = 50$
Plain R, $m = 10$	4	5	8	14		
MKL, $m = 10$	5	6	10	15		
Plain R, $m = 100$	43	46	56	74	131	451
MKL, $m = 100$	45	49	57	69	90	185
Plain R, $m = 500$	222	240	279	361	600	1993
MKL, $m = 500$	225	240	269	310	389	693

3. Letting $\Psi_i = \text{diag}\left(\tau_{i1}^{-2}, \tau_{i2}^{-2}, \ldots, \tau_{i\tilde{r}}^{-2}\right)$, draw $\boldsymbol{\Lambda}_{i\cdot}' | \boldsymbol{f}, \boldsymbol{y}_{i\cdot}, \boldsymbol{h}_i, \Psi_i, \sim \mathcal{N}_{\tilde{r}}(\boldsymbol{b}_{iT}, \boldsymbol{B}_{iT})$ with $\boldsymbol{B}_{iT} = (\boldsymbol{X}_i'\boldsymbol{X}_i + \Psi_i)^{-1}$ and $\boldsymbol{b}_{iT} = \boldsymbol{B}_{iT}\boldsymbol{X}_i'\tilde{\boldsymbol{y}}_{i\cdot}$. Hereby, $\tilde{\boldsymbol{y}}_{i\cdot} = (y_{i1}e^{-h_{i1}/2}, \ldots, y_{iT}e^{-h_{iT}/2})'$ denotes the ith normalized observation vector and
$$\boldsymbol{X}_i = \begin{bmatrix} f_{11}e^{-h_{i1}/2} & \cdots & f_{\tilde{r}1}e^{-h_{i1}/2} \\ \vdots & & \vdots \\ f_{1T}e^{-h_{iT}/2} & \cdots & f_{\tilde{r}T}e^{-h_{iT}/2} \end{bmatrix}$$
is the $T \times \tilde{r}$ design matrix. This constitutes a standard Bayesian regression update.

3*. When inference on the factor loadings matrix is sought, optionally redraw $\boldsymbol{\Lambda}$ using *deep interweaving* (Kastner et al., 2017) to speed up mixing. This step is of less importance if one is interested in the (predictive) covariance matrix only.

4. Draw the factors from $\boldsymbol{f}_t | \boldsymbol{\Lambda}, \boldsymbol{y}_t, \boldsymbol{h}_t \sim \mathcal{N}_r(\boldsymbol{b}_{mt}, \boldsymbol{B}_{mt})$ with $\boldsymbol{B}_{mt}^{-1} = \boldsymbol{X}_t'\boldsymbol{X}_t + \boldsymbol{V}_t^{-1}$ and $\boldsymbol{b}_{mt} = \boldsymbol{B}_{mt}\boldsymbol{X}_t'\tilde{\boldsymbol{y}}_t$. Hereby, $\tilde{\boldsymbol{y}}_t = (y_{1t}e^{-h_{1t}/2}, \ldots, y_{mt}e^{-h_{mt}/2})'$ denotes the normalized observation vector at time t and
$$\boldsymbol{X}_t = \begin{bmatrix} \Lambda_{11}e^{-h_{1t}/2} & \cdots & \Lambda_{1r}e^{-h_{1t}/2} \\ \vdots & & \vdots \\ \Lambda_{m1}e^{-h_{mt}/2} & \cdots & \Lambda_{mr}e^{-h_{mt}/2} \end{bmatrix}$$
is the $m \times r$ design matrix. This constitutes a standard Bayesian regression update.

The above sampling steps are implemented in an efficient way within the R package **factorstochvol**. Table 1 displays the empirical run time in milliseconds per MCMC iteration. Note that using more efficient linear algebra routines such as Intel MKL leads to substantial speed gains only for models with many factors. To a certain extent, computation can further be sped up by computing the individual steps of the posterior sampler in parallel. In practice, however, doing so is only useful in shared memory environments (e.g. through multithreading/multiprocessing) as the increased communication overhead in distributed memory environments easily outweighs the speed gains.

3.2. Prediction

Given draws of the joint posterior distribution of parameters and latent variables, it is in principle straightforward to predict future covariances and consequently also future observations. This gives rise to the predictive density (Geweke and Amisano, 2010), defined as
$$p(\boldsymbol{y}_{t+1} | \boldsymbol{y}_{[1:t]}^o) = \int_{\boldsymbol{K}} p(\boldsymbol{y}_{t+1} | \boldsymbol{y}_{[1:t]}^o, \boldsymbol{\kappa}) \times p(\boldsymbol{\kappa} | \boldsymbol{y}_{[1:t]}^o)\, \mathrm{d}\boldsymbol{\kappa}, \tag{5}$$
where $\boldsymbol{\kappa}$ denotes the vector of all unobservables, i.e. parameters and latent variables. The superscript o in $\boldsymbol{y}_{[1:t]}^o$ denotes *ex post* realizations (observations) for the set of points in time $\{1, \ldots, t\}$ of the *ex ante* random values $\boldsymbol{y}_{[1:t]} = (\boldsymbol{y}_1 \cdots \boldsymbol{y}_t)$. The integration space $\boldsymbol{K}$ simply stands for the space of the possible values for $\boldsymbol{\kappa}$. Because (5) is the integral of the likelihood function where the values of $\boldsymbol{\kappa}$ are weighted according to their posterior distribution, it can be seen as the forecast density for an unknown value $\boldsymbol{y}_{t+1}$ after accounting for the uncertainty about $\boldsymbol{\kappa}$, given the history $\boldsymbol{y}_{[1:t]}^o$.

As with most quantities of interest in Bayesian analysis, computing the predictive density can be challenging because it constitutes an extremely high-dimensional integral which cannot be solved analytically. However, it may be approximated at a given "future" point $\boldsymbol{y}^f$ through Monte Carlo integration,
$$p(\boldsymbol{y}^f | \boldsymbol{y}_{[1:t]}^o) \approx \frac{1}{K} \sum_{k=1}^{K} p(\boldsymbol{y}^f | \boldsymbol{y}_{[1:t]}^o, \boldsymbol{\kappa}_{[1:t]}^{(k)}), \tag{6}$$
where $\boldsymbol{\kappa}_{[1:t]}^{(k)}$ denotes the kth draw from the posterior distribution up to time t. If (6) is evaluated at $\boldsymbol{y}^f = \boldsymbol{y}_{t+1}^o$, it is commonly referred to as the (one-step-ahead) *predictive likelihood* at time $t + 1$, denoted PL_{t+1}. Also, draws from (5) can

straightforwardly be obtained by generating values $\boldsymbol{y}_{t+1}^{(k)}$ from the distribution given through the (in our case multivariate Gaussian) density $p(\boldsymbol{y}_{t+1}|\boldsymbol{y}_{[1:t]}^{o}, \boldsymbol{\kappa}_{[1:t]}^{(k)})$.

For the model at hand, two ways of evaluating the predictive likelihood particularly stand out. First, one could average over $k = 1, \ldots, K$ densities of

$$\mathcal{N}_m\left(\boldsymbol{\Lambda}_{[1:t]}^{(k)}\boldsymbol{f}_{t+1,[1:t]}^{(k)}, \boldsymbol{U}_{t+1,[1:t]}^{(k)}\right),$$

evaluated at $\boldsymbol{y}_{t+1}^{o}$, where the subscript $t+1$ denotes the corresponding one-step ahead predictive draws and $\boldsymbol{U}_{t+1,[1:t]}^{(k)} = \operatorname{diag}\left(\exp h_{1,t+1,[1:t]}^{(k)}, \ldots, \exp h_{m,t+1,[1:t]}^{(k)}\right)$. Note that because $\boldsymbol{U}_{t+1,[1:t]}^{(k)}$ is by construction diagonal, this method only requires univariate Gaussian evaluations and is thus computationally efficient. Nevertheless, because evaluation is done conditionally on realized values of $\boldsymbol{f}_{t+1,[1:t]}$, it is extremely unstable in many dimensions. Moreover, since the numerical inaccuracy increases with an increasing number of factors r, this approach can lead to systematic undervaluation of PL_{t+1} for larger r. Thus, in what follows, we recommended an alternative approach.

To obtain PL_{t+1}, we suggest to average over $k = 1, \ldots, K$ densities of

$$\mathcal{N}_m\left(\boldsymbol{0}, \boldsymbol{\Lambda}_{[1:t]}^{(k)}\boldsymbol{V}_{t+1,[1:t]}^{(k)}(\boldsymbol{\Lambda}_{[1:t]}^{(k)})' + \boldsymbol{U}_{t+1,[1:t]}^{(k)}\right),$$

evaluated at $\boldsymbol{y}_{t+1}^{o}$, where $\boldsymbol{V}_{t+1,[1:t]}^{(k)} = \operatorname{diag}\left(\exp h_{m+1,t+1,[1:t]}^{(k)}, \ldots, \exp h_{m+r,t+1,[1:t]}^{(k)}\right)$. This form of the predictive likelihood is obtained by analytically performing integration in (5) with respect to $\boldsymbol{f}_{t+1,[1:t]}$. Consequently, it is numerically more stable, irrespectively of the number of factors r. However, it requires a full m-variate Gaussian density evaluation for each k and is thus computationally much more expensive. To a certain extent, the computational burden can be mitigated by using the Woodbury matrix identity, $\boldsymbol{\Sigma}_t^{-1} = \boldsymbol{U}_t^{-1} - \boldsymbol{U}_t^{-1}\boldsymbol{\Lambda}\left(\boldsymbol{V}_t^{-1} + \boldsymbol{\Lambda}'\boldsymbol{U}_t^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}'\boldsymbol{U}_t^{-1}$, along with the matrix determinant lemma, $\det(\boldsymbol{\Sigma}_t) = \det(\boldsymbol{V}_t^{-1} + \boldsymbol{\Lambda}'\boldsymbol{U}_t^{-1}\boldsymbol{\Lambda})\det(\boldsymbol{V}_t)\det(\boldsymbol{U}_t)$. This substantially speeds up the repetitive evaluation of the multivariate Gaussian distribution if $r \ll m$.

We apply these results for comparing competing models A and B between time points t_1 and t_2 and consider cumulative log predictive Bayes factors defined through $\log BF_{t_1,t_2}(A, B) = \sum_{t=t_1+1}^{t_2} \log PL_t(A) - \log PL_t(B)$, where $PL_t(A)$ and $PL_t(B)$ denote the predictive likelihood of model A and B at time t, respectively. When the cumulative log predictive Bayes factor is greater than 0 at a given point in time, there is evidence in favor of model A, and vice versa. Thereby, data up to time t_1 is regarded as prior information and out-of-sample evaluation starts at time $t_1 + 1$.

4. Simulation studies

The aim of this section is to apply the model to a simulated data set in order to illustrate the shrinkage properties of the Normal-Gamma prior for the factor loadings matrix elements. For this purpose, we first illustrate several scenarios on a single ten dimensional data set. Second, we investigate the performance of our model in a full Monte Carlo simulation based on 100 simulated data sets. Third, and finally, we investigate to what extent these results carry over to higher dimensions.

In what follows, we compare five specific prior settings. Setting 1 refers to the usual standard Gaussian prior with variance $\tau_{ij}^2 \equiv \tau^2 = 1$ and constitutes the benchmark. Setting 2 is the row-wise Bayesian Lasso where $a_i = 1$ for all i. Setting 3 is the column-wise Bayesian Lasso where $a_j = 1$ for all j. Setting 4 is the Normal-Gamma prior with row-wise shrinkage where $a_i = 0.1$ for all i. Setting 5 is the Normal-Gamma prior with column-wise shrinkage where $a_j = 0.1$ for all j. Throughout this section, prior hyperparameters are chosen as follows: $b_\mu = 0$, $B_\mu = 1000$, $B_\sigma = 1$. The prior hyperparameters for the persistence of the latent log variances are fixed at $a_0 = 10$, $b_0 = 2.5$ for the idiosyncratic volatilities and $a_0 = 2.5$, $b_0 = 2.5$ for the factor volatilities; note that the parameters of the superfluous factor are only identified through the prior. The shrinkage hyperparameters are set as in Belmonte et al. (2014), i.e. $c_i = c_j = d_i = d_j = 0.001$ for all applicable i and j. For each setting, the algorithm is run for 110 000 iterations of which the first 10 000 draws are discarded as burn-in.

4.1. The shrinkage prior effect: An illustration

To investigate the effects of different priors on the posteriors of interest, we simulate a single data set from a two factor model for $m = 10$ time series of length $T = 1000$. For estimation, an overfitting model with three latent factors is employed. The nonzero parameter values used for simulation are picked randomly and are indicated as black circles in Fig. 1; some loadings are set to zero, indicated by black dots. We set Λ_{ij} to zero if $j > i$ for simulation and estimation.

Fig. 1 shows smoothed kernel density estimates of posterior loadings under the different prior assumptions. The signs of the loadings have not been identified so that a multimodal posterior distribution hints at a "significant" loading whereas a unimodal posterior hints at a zero loading, see also Frühwirth-Schnatter and Wagner (2010). It stands out that only very little shrinkage is induced by the standard Gaussian prior. The other priors, however, impose considerably tighter posteriors. For the nonzero loadings on factor one, e.g., the row-wise Bayesian Lasso exhibits the strongest degree of shrinkage. Little difference between the various shrinkage priors can be spotted for the nonzero loadings on factor two.

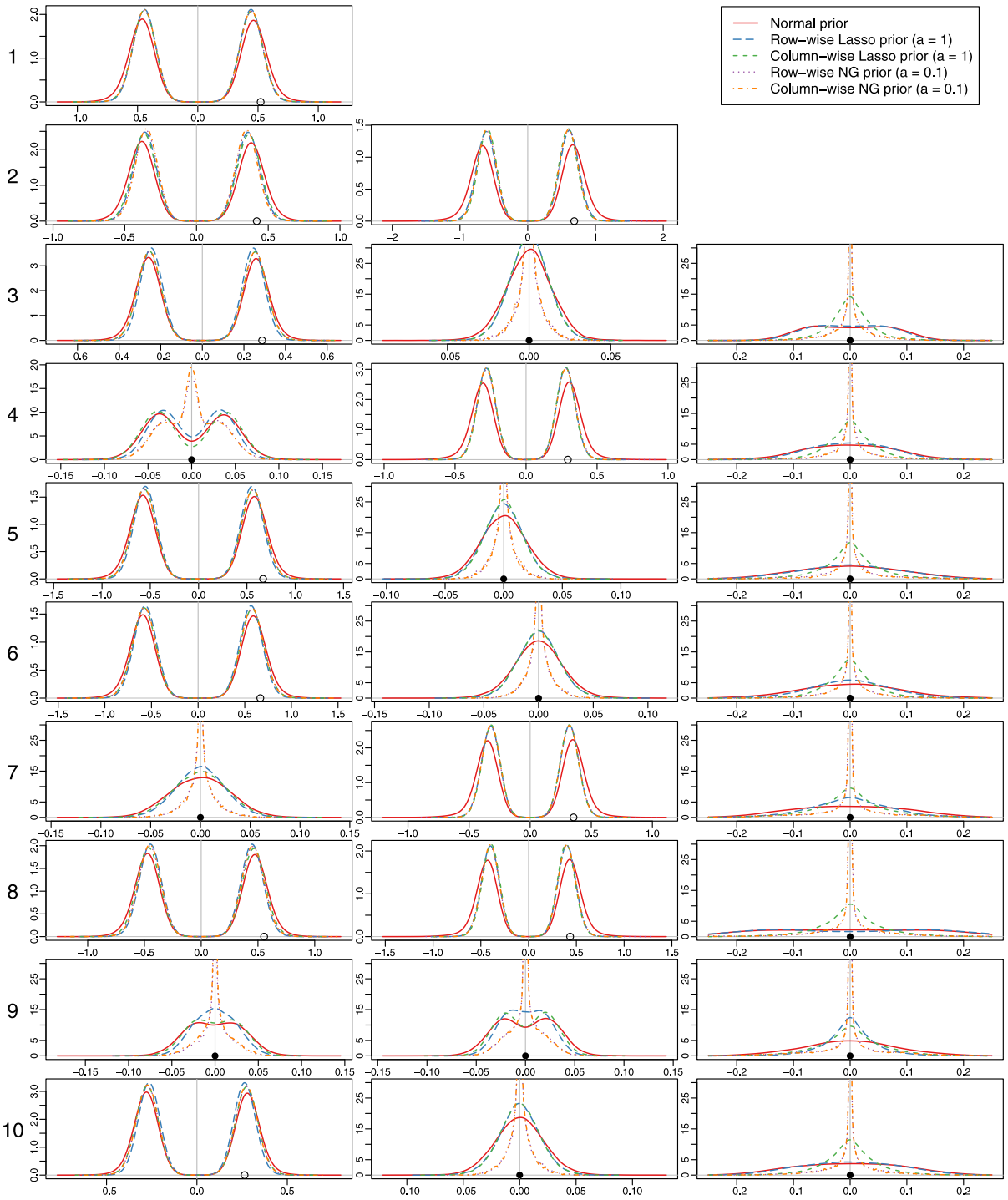

Fig. 1. Kernel density estimates of posterior factor loadings under different priors. The standard Gaussian prior (setting 1) in red solid strokes, the row-wise Lasso prior (setting 2) in blue long-dashed strokes, the column-wise Lasso prior (setting 3) in green short-dashed strokes, the row-wise Normal-Gamma prior (setting 4) in purple dotted strokes, the column-wise Normal-Gamma prior (setting 5) in orange dashed–dotted strokes. The vertical axis is capped at 30.

Turning towards the zero loadings, the strongest shrinkage is introduced by both variants of the Normal-Gamma prior, followed by the different variants of the Bayesian Lasso and the standard Gaussian prior. This is particularly striking for the loadings on the superfluous third factor. The difference between row- and column-wise shrinkage for the Lasso variants can

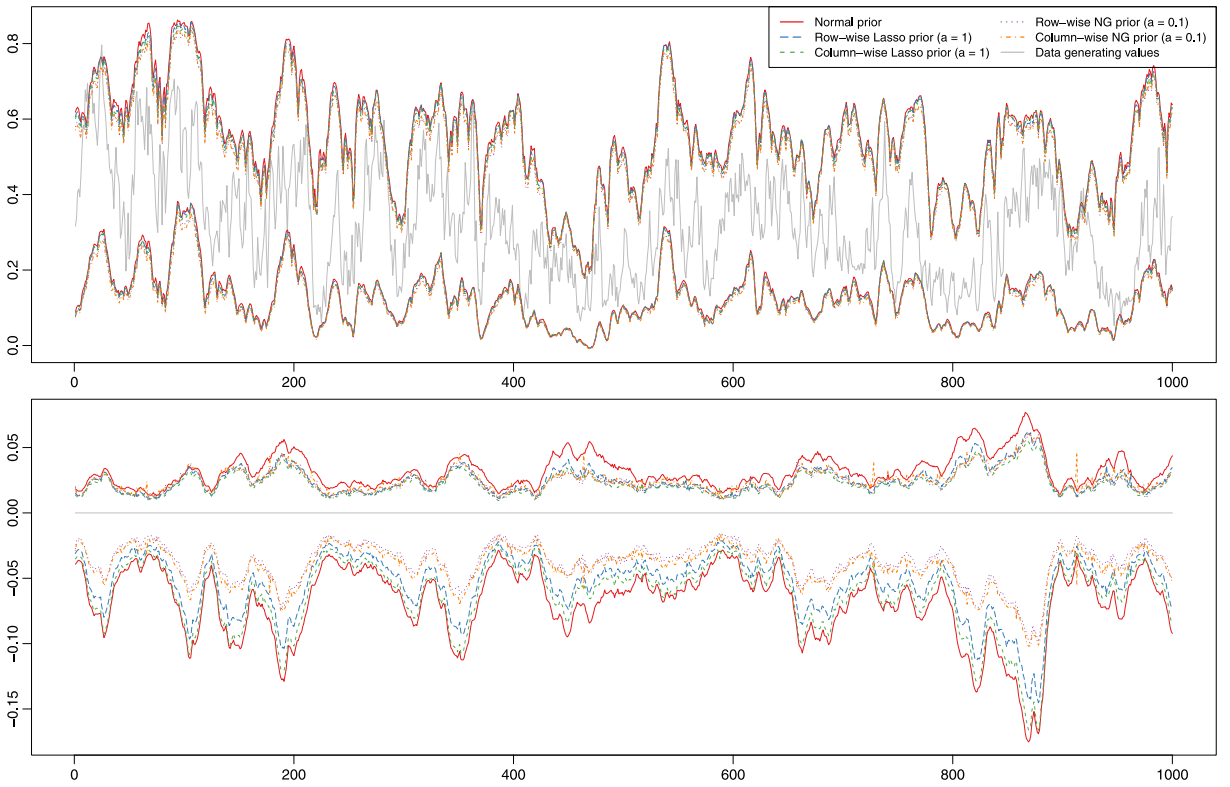

Fig. 2. "True" (gray, solid) and estimated posterior correlations between series 1 and series 2 (top) as well as series 9 and series 10 (bottom). To illustrate estimation uncertainty, posterior means plus/minus 2 times posterior standard deviations are displayed.

Table 2
Estimated log Bayes factors at $t = 1500$ against the 2-factor model using a standard Gaussian prior, where data up to $t = 1000$ is treated as the training sample. Lines 1 to 5 correspond to cumulative 1-day-ahead Bayes factors, lines 6 to 10 correspond to 10-days-ahead predictive Bayes factors.

	no factor	1 factor	2 factors	3 factors	4 factors	5 factors
Standard Gaussian	−845.80	−316.41		−2.81	−5.83	−8.38
Row-wise NG ($c_i = d_i = 1$)	−845.80	−317.75	0.27	−0.44	−1.74	−2.34
Col-wise NG ($c_j = d_j = 1$)	−845.80	−317.45	0.38	−0.77	−1.61	−2.06
Row-wise NG ($c_i = d_i = 0.01$)	−845.80	−317.71	0.40	−0.29	−1.36	−1.97
Col-wise NG ($c_j = d_j = 0.01$)	−845.80	−317.52	1.04	−0.12	−0.88	−2.07
Standard Gaussian	−787.41	−255.92		−4.98	−7.75	−12.67
Row-wise NG ($c_i = d_i = 1$)	−787.41	−256.85	2.90	1.73	0.87	−0.05
Col-wise NG ($c_j = d_j = 1$)	−787.41	−256.44	3.14	1.17	0.90	0.09
Row-wise NG ($c_i = d_i = 0.01$)	−787.41	−256.81	3.18	1.98	1.28	0.53
Col-wise NG ($c_j = d_j = 0.01$)	−787.41	−255.90	2.85	1.61	0.84	0.16

most clearly be seen in row 9 and column 3, respectively. The row-wise Lasso captures the "zero-row" 9 better, while the column-wise Lasso captures the "zero-column" 3 better. Because of the increased element-wise shrinkage of the Normal-Gamma prior, the difference between the row-wise and the column-wise variant are minimal.

In the context of covariance modeling, however, factor loadings can be viewed upon as a mere means to parsimony, not the actual quantity of interest. Thus, Fig. 2 displays selected time-varying correlations. The top panel shows a posterior interval estimate (mean plus/minus two standard deviations) for the correlation of series 1 and series 2 (which is nonzero) under all five prior settings; the bottom panel depicts the interval estimate for the correlation of series 9 and 10 (which is zero). While the relative differences between the settings in the nonzero correlation case are relatively small, the zero correlation case is picked up substantially better when shrinkage priors are used. Posterior means are closer to zero and the posterior credible intervals are tighter.

To conclude, we briefly examine predictive performance by investigating cumulative log predictive Bayes factors. Thereby, the first 1000 points in time are treated as prior information, then 1-day- and 10-days-ahead predictive likelihoods are recursively evaluated until $t = 1500$. Table 2 displays the sum of these values for the respective models in relation to the 2-factor model with the standard Gaussian prior. This way, numbers greater than zero can be interpreted as evidence in favor

Table 3
Relative RMSEs of pairwise correlations. Above the diagonal: Row-wise Lasso ($a_i = 1$) vs. benchmark standard Gaussian prior with geometric means (first row). Below the diagonal: Row-wise Normal-Gamma ($a_i = 0.1$) vs. row-wise Lasso prior ($a_i = 1$) with geometric means (last row). Numbers greater than one mean that the former prior performs better than the latter. All values reported are medians of 100 repetitions where $c_i = d_i = 0.001$.

	1	2	3	4	5	6	7	8	9	10
Average 1	1.05	1.03	1.05	1.07	1.04	1.05	1.08	1.03	1.32	1.05
1		1.00	1.00	1.08	1.00	1.00	1.10	1.00	1.29	1.00
2	1.00		1.00	1.00	1.00	1.00	1.00	1.00	1.26	1.00
3	1.00	1.01		1.07	1.00	1.00	1.09	1.00	1.31	0.99
4	3.04	1.00	2.85		1.06	1.07	1.00	1.00	1.30	1.10
5	1.00	1.01	1.00	3.01		1.00	1.06	1.00	1.31	1.00
6	1.00	1.01	1.00	3.00	1.00		1.07	1.00	1.32	0.99
7	3.02	1.00	2.81	1.00	2.87	2.77		1.00	1.28	1.09
8	1.00	1.00	1.00	1.01	1.01	1.01	1.01		1.27	1.00
9	2.85	2.61	2.70	2.69	2.73	2.71	2.77	2.54		1.31
10	1.00	1.01	1.00	2.84	1.00	1.00	2.89	1.01	2.72	
Average 2	1.45	1.12	1.40	2.01	1.40	1.41	2.04	1.12	2.79	1.40

Table 4
Different error measures ($\times 10^{-2}$) of posterior mean correlation estimates under various priors. For all scenarios, 3-factor SV models are fit to 10-dimensional data of length 1000 simulated from a 2-factor SV model. The column titled "GSH" contains the values of $c_i = c_j = d_i = d_j$. All values reported are medians of 100 repetitions.

	GSH	Abs. RMSE	Rel. RMSE	Abs. MAE	Rel. MAE
Standard Gaussian		7.581		5.397	
Row-wise Lasso	0.001	7.587	100.449	5.324	101.850
Col-wise Lasso	0.001	7.593	100.176	5.372	100.622
Row-wise Lasso	1.000	7.602	99.995	5.416	100.146
Col-wise Lasso	1.000	7.584	99.987	5.415	100.119
Row-wise NG	0.001	7.351	102.793	4.891	110.252
Col-wise NG	0.001	7.395	102.525	4.957	109.912
Row-wise NG	1.000	7.392	102.641	4.941	109.735
Col-wise NG	1.000	7.384	102.622	4.925	109.722

of the respective model. Not very surprisingly, log Bayes factors are highest for the 2-factor model; within this class, models imposing stronger shrinkage perform slightly better, in particular when considering the longer 10-day horizon. Underfitting models predict very poorly both on the short and the longer run, while overfitting models appear almost en par with the baseline model when shrinkage priors are used. This suggests that shrinkage safeguards against overfitting, at least to a certain extent.

4.2. Medium dimensional Monte Carlo study

For a more comprehensive understanding of the shrinkage effect, the above study is repeated for 100 different data sets where all latent variables are generated randomly for each realization. In Table 3, the medians of the respective relative RMSEs (root mean squared errors, averaged over time) between the true and the estimated pairwise correlations are depicted. The part above the diagonal represents the relative performance of the row-wise Lasso prior (setting 2) with respect to the baseline prior (setting 1), the part below the diagonal represents the relative performance of the row-wise Normal-Gamma prior (setting 4) with respect to the row-wise Lasso prior (setting 2). Clearly, gains are highest for series 9 which is by construction completely uncorrelated to the other series. Additionally, geometric averages of these performance indicators are displayed in the first row (setting 2 vs. baseline) and in the last row (setting 4 vs. baseline). They can be seen as the average relative performance of one specific series' correlation estimates with all other series. We remark that choosing c_i and d_i small is crucial for the shrinkage effect of the Bayesian Lasso while the performance of the Normal-Gamma prior is relatively robust with regard to these choices. Simulation studies with less extreme hyperparameter choices ($c_i = d_i = 1$) show that performance gains of the Bayesian Lasso over the standard Gaussian prior nearly vanish while performance gains of the Normal-Gamma prior remain stable. This indicates that the shrinkage effect of the Bayesian Lasso is strongly dependent on the particular choice of these hyperparameters (governing row-wise shrinkage) while the Normal-Gamma can adapt better through increased element-wise shrinkage.

An overall comparison of the errors under different priors is provided in Table 4 which lists RMSEs and MAEs for all prior settings, averaged over the non-trivial correlation matrix entries as well as time. Note again that results under the Lasso prior are sensitive to the particular choices of the global shrinkage hyperparameters as well as the choice of row- or column-wise shrinkage, which is hardly the case for the Norma-Gamma prior. Interestingly, the performance gains achieved through shrinkage prior usage are higher when absolute errors are considered. This is coherent with the extremely high kurtosis of Normal-Gamma-type priors which, while placing most mass around zero, allow for large values.

Table 5
Different error measures ($\times 10^{-2}$) of posterior mean correlation estimates under various priors. For all scenarios, unrestricted 11-factor SV models are fit to 100-variate data of length 1000 simulated from 10-factor SV models. The column titled "GSH" contains the values of $c_i = c_j = d_i = d_j$. All values reported are medians of 100 repetitions.

	GSH	Abs. RMSE	Rel. RMSE	Abs. MAE	Rel. MAE
Standard Gaussian		8.906		7.074	
Row-wise Lasso	0.001	8.293	107.047	6.553	107.844
Col-wise Lasso	0.001	8.425	105.642	6.659	106.125
Row-wise Lasso	1.000	8.215	107.811	6.496	108.392
Col-wise Lasso	1.000	8.345	106.508	6.609	107.058
Row-wise NG	0.001	7.773	114.466	6.074	116.411
Col-wise NG	0.001	7.802	114.525	6.085	116.768
Row-wise NG	1.000	7.799	114.386	6.099	116.391
Col-wise NG	1.000	7.778	114.121	6.070	116.410

Table 6
GICS sectors and the amount of members within the S&P 500 data set.

GICS sector	Members
Consumer discretionary	45
Consumer staples	28
Energy	23
Financials	54
Health care	30
Industrials	42
Information technology	27
Materials	23
Telecommunications services	3
Utilities	25

4.3. High dimensional Monte Carlo study

The findings are similar if dimensionality is increased; in analogy to above, we report overall RMSEs and MAEs for 495 000 pairwise correlations, resulting from $m = 100$ component series at $T = 1000$ points in time. The factor loadings for the $r = 10$ factors are again randomly sampled with 43.8% of the loadings being equal to zero, resulting in about 2.6% pairwise correlations that vanish. Consequently, 100 data sets are generated; for each of these, a separate (overfitting) factor SV model using $r = 11$ factors without any prior restrictions on the factor loadings matrix is fit. Table 5 reports the medians of the aggregated error measures. In this setting, the shrinkage priors outperform the standard Gaussian prior by a relatively large margin; the effect of the specific choice of the global shrinkage hyperparameters is less pronounced.

5. Application to S&P 500 data

In this section we apply the SV factor model to stock prices listed in the Standard & Poor's 500 index. We only consider firms which have been continuously included in the index from November 1994 until December 2013, resulting in $m = 300$ stock prices on 5001 days, ranging from 11/1/1994 to 12/31/2013. The data was obtained from Bloomberg Terminal in January 2014. Instead of considering raw prices we investigate percentage log-returns which we demean a priori.

The presentation consists of two parts. First, we exemplify inference using a multivariate stochastic volatility model and discuss the outcome. Second, we perform out-of-sample predictive evaluation and compare different models. To facilitate interpretation of the results discussed in this section, we consider the GICS[5] classification into 10 sectors listed in Table 6.

5.1. A four-factor model for 300 S&P 500 members

To keep graphical representation feasible, we only focus on the latest 2000 returns of our data set, i.e. 5/3/2006 to 12/31/2013. This time frame is chosen to include both the 2008 financial crisis as well as the period before and thereafter. Furthermore, we restrict our discussion to a four-factor model. This choice is somewhat arbitrary but allows for a direct comparison to a popular model based on four observed (Fama–French plus Momentum) factors. A comparison of predictive performance for varying number of factors is discussed in Section 5.2; the Fama–French plus Momentum model is introduced in Section 5.3.

[5] Global Industry Classification Standard, retrieved from https://en.wikipedia.org/w/index.php?title=List_of_S%26P_500_companies&oldid=589980759 on April 11, 2016.

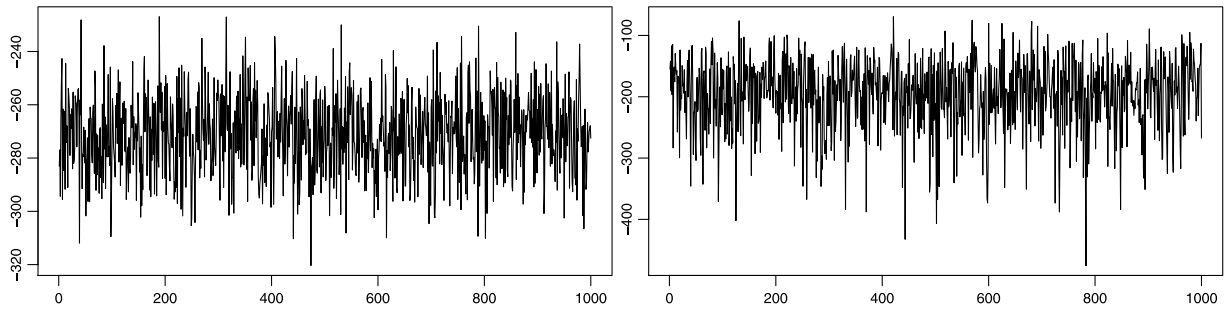

Fig. 3. Trace plots for 1000 draws after burn-in and thinning of the log determinant of the covariance (left panel) and the correlation matrix (right panel) for $t = T$.

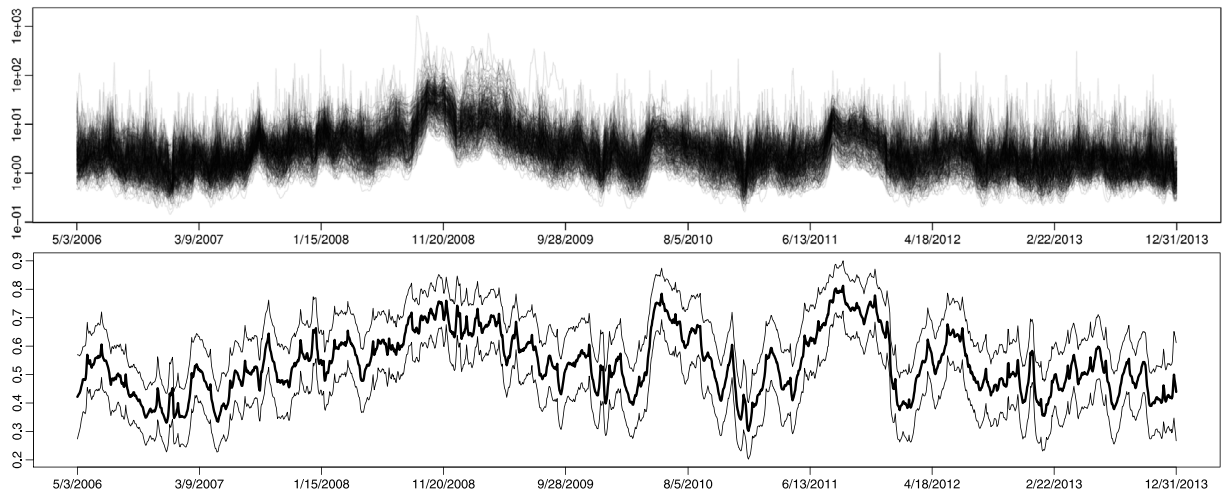

Fig. 4. Top panel: 300 mean posterior variances, i.e. $\mathbb{E}(\text{diag}(\boldsymbol{\Sigma}_t)\,|\boldsymbol{y})$ for $t = 1, \ldots, T$ (logarithmic scale). Bottom panel: Posterior mean of the joint communality C_t (bold line) along with mean plus/minus two posterior standard deviations (light lines).

We run our sampler employing the Normal-Gamma prior with row-wise shrinkage for 110 000 draws and discard the first 10 000 draws as burn-in.[6] Of the remaining 100 000 draws every 10th draw is kept, resulting in 10 000 draws used for posterior inference. Hyperparameters are set as follows: $a_i \equiv a = 0.1$, $c_i \equiv c = 1$, $d_i \equiv d = 1$, $b_\mu = 0$, $B_\mu = 100$, $a_0 = 20$, $b_0 = 1.5$, $B_\sigma = 1$, $B_{m+j} = 1$ for $j = 1, \ldots, r$. To prevent factor switching, we set all elements above the diagonal to zero. The leading series are chosen manually after a preliminary unidentified run such that series with high loadings on that particular factor (but low loadings on the other factors) become leaders. Note that this intervention is only necessary for interpreting the factor loadings matrix but not for covariance estimation or prediction. We observe excellent mixing for both the covariance as well as the correlation matrix MCMC draws. To exemplify, trace plots of the first 1000 draws after burn-in and thinning for posterior draws of the log determinant distribution of the covariance and correlation matrices at $t = T$ are displayed in Fig. 3.

To illustrate the substantial degree of volatility co-movement, mean posterior variances are depicted in the top panel of Fig. 4. Clear spikes can be spotted during the financial crisis in late 2008 but also in early 2010 and late 2011. This picture is mirrored (to a certain extent) in the bottom panel which displays the posterior distribution of the joint communality C_t. At times, the joint communality reaches high values of 0.7 and more.

Median posterior factor loadings are visualized in Fig. 5. In the top panel it can be seen that all series significantly load on the first factor which consequently could be interpreted to represent the joint dynamics of the wider US equity market. Investigating the second factor, it stands out that due to the use of the Normal-Gamma prior a considerable amount of loadings are shrunk towards zero. Main drivers are all in the sector *Utilities*. Also, companies in sectors *Consumer Staples*, *Health Care* and (to a certain extent) *Financials* load positively here. Both the loadings on factor 3 as well as the loadings on factor 4 are substantially shrunk towards zero. Exceptions are *Energy* and *Materials* companies for factor 3 and *Financials* for factor 4.

[6] To keep presentation at a reasonable length and because qualitative as well as quantitative results are very similar, we omit details about the Normal-Gamma prior with column-wise shrinkage.

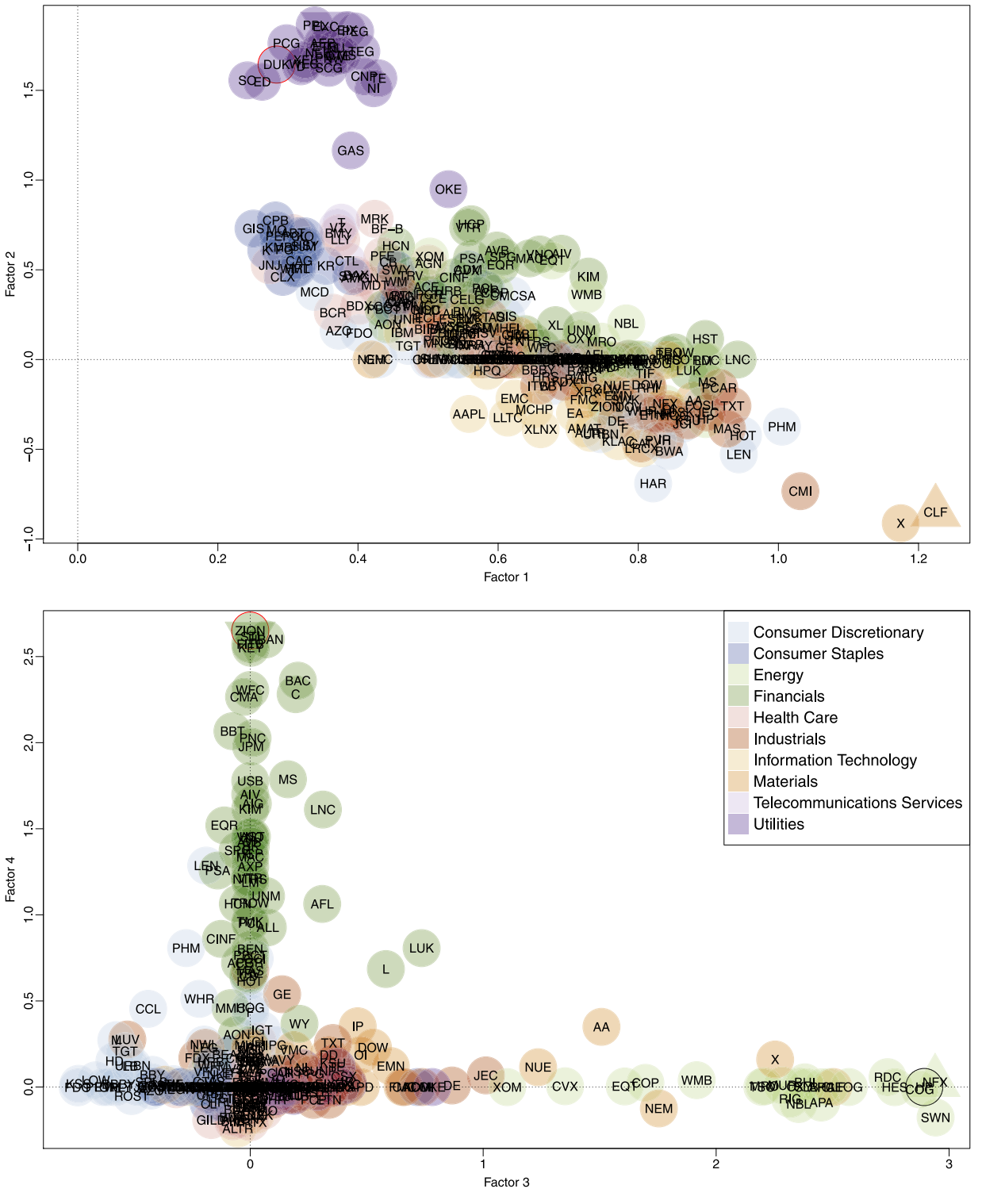

Fig. 5. Median loadings on the first two factors (top) and the last two factors (bottom) of a 4-factor model applied to $m = 300$ demeaned stock price log-returns listed in the S&P 500 index. Shading: Sectors according to the Global Industry Classification Standard.

The corresponding factor log variances are displayed in Fig. 6. Apart from featuring similar low- to medium frequency properties, each process exhibits specific characteristics. First, notice the sharp increase of volatility in early 2010 which is

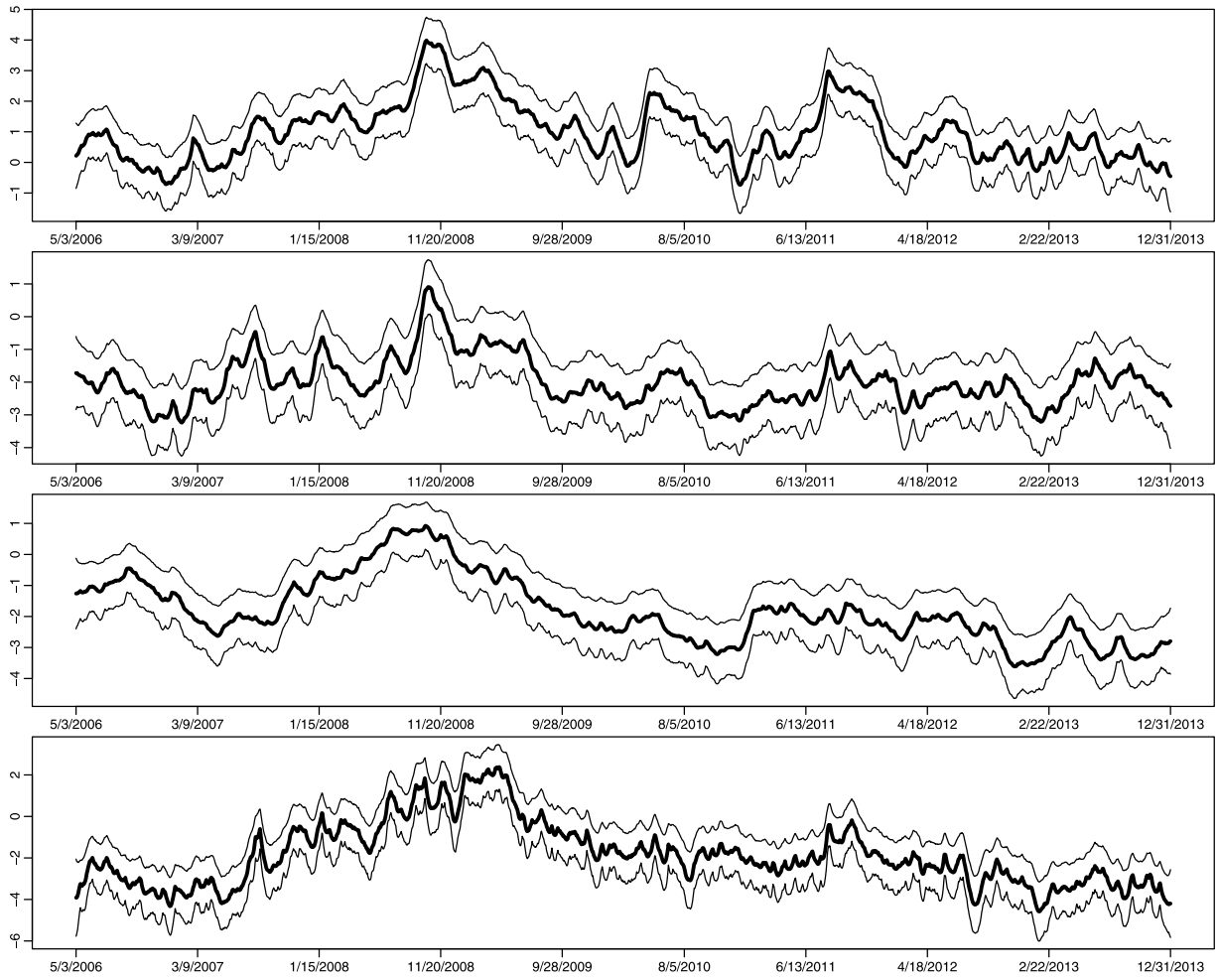

Fig. 6. Latent factor log variances $\boldsymbol{h}_{m+j,\cdot}$, $j = 1, \ldots, 4$ (top to bottom). Bold line indicates the posterior mean; light lines indicate mean $\pm$ 2 standard deviations.

mainly visible for the "overall" factor 1. The second factor (*Utilities*) displays a pre-crisis volatility peak during early 2008. The third factor, driven by *Energy* and *Materials*, shows relatively smooth volatility behavior while the fourth factor, governed by the *Financials*, exhibits a comparably turbulent volatility evolution.

Finally, we show three examples of the posterior mean of the correlation matrix $\boldsymbol{\Sigma}_t$ in Fig. 7. The series are grouped according to the alphabetically ordered industry sectors (and simply sorted according to their ticker symbol therein). An animation displaying the mean correlation matrix for all points in time is available at https://vimeo.com/217021226. Considering the last trading day in 2006, highly correlated clusters appear within *Energy* and *Utilities*, to a certain extent also within *Financials*, *Industrials* and *Materials*. Not very surprisingly, there exists only low correlation between companies in the sectors *Consumer Discretionary/Staples* and *Energy* but higher correlation between *Energy*, *Industrials* and *Materials*. Looking at the last trading day of 2008, the overall picture changes radically. Higher correlation can be spotted throughout, both within sectors but also between sectors. There are only few companies that show little and virtually no companies that show no correlation with others. Another two years later, we again see a different overall picture. Lower correlations throughout become apparent with moderate correlations remaining within the sectors *Energy*, *Utilities*, and *Financials*.

5.2. Predictive likelihoods for model selection

Even for univariate volatility models evaluating in- or out-of-sample fit is not straightforward because the quantity of interest (the conditional standard deviation) is not directly observable. While in lower dimensions this issue can be circumvented to a certain extent using intraday data and computing realized measures of volatility, the difficulty becomes more striking when the dimension increases. Thus, we focus on iteratively predicting the observation density out-of-sample

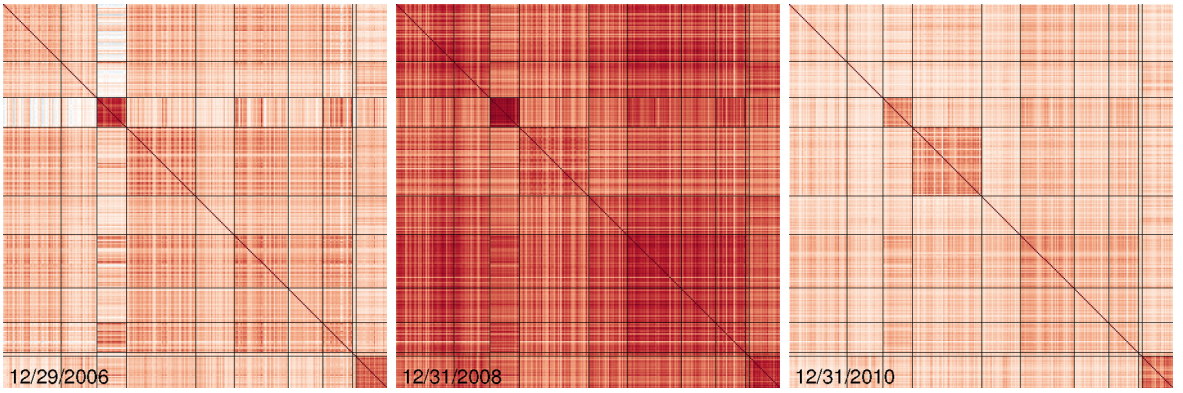

Fig. 7. Posterior mean of the time-varying correlation matrix $\mathbb{E}(\boldsymbol{\Sigma}_t|\boldsymbol{y})$, exemplified for $t \in \{173, 696, 1218\}$ which corresponds to the last trading day in 2006, 2008, 2010, respectively. The matrix has been rearranged to reflect the different GICS sectors in alphabetical order, cf. Table 6. An animation depicting all time points is available at https://vimeo.com/217021226.

which is then evaluated at the actually observed values. Because this approach involves re-estimating the model for each point in time, it is computationally costly but can be parallelized in a trivial fashion on multi-core computers.

For the S&P 500 data set, we begin by using the first 3000 data points (until 5/2/2006) to estimate the one-day-ahead predictive likelihood for day 3001 as well as the ten-day-ahead predictive likelihood for day 3010. In a separate estimation procedure, the first 1001 data points (until 5/3/2006) are used to estimate the one-day-ahead predictive likelihood for day 3002 and the corresponding ten-day-ahead predictive likelihood for day 3011, etc. This procedure is repeated for 1990 days until the end of the sample is reached.

We use a no-factor model as the baseline which corresponds to 300 individual stochastic volatility models fitted to each component series separately. For each date, values greater than zero mean that the model outperforms the baseline model up to that point in time. Competitors of the no-factor SV model are r-factor SV models with $r = 1, \ldots, 20$ under the usual standard Gaussian prior and under the Normal-Gamma prior with $a_i \equiv 0.1$ and $c_i \equiv d_i \equiv 1$ employing row-wise-shrinkage. All other parameters are kept identical, i.e. $b_\mu = 0$, $B_\mu = 100$, $a_0 = 20$, $b_0 = 2.5$ (idiosyncratic persistences), $a_0 = 2.5$, $b_0 = 2.5$ (factor persistences), $B_\sigma = 0.1$, $B_{m+j} = 0.1$ for $j = 1, \ldots, r$. Note that because the object of interest in this exercise does not require the factor loadings matrix to be identified, no a priori restrictions are placed on $\boldsymbol{\Lambda}$. This alleviates the problem of arranging the data in any particular order before running the sampler. Other competing models are discussed in the following sections.

Accumulated log predictive likelihoods for the entire period are displayed in Fig. 8. Gains in predictive power are substantial up to around 8 factors with little difference for the two priors. After this point, the benefit of adding even more factors turns out to be less pronounced. On the contrary, the effect of the priors becomes more pronounced. Again, while differences tend to be muted for models with fewer factors, the benefit of shrinkage grows when r gets larger.

While joint models with $r > 0$ outperform the marginal (no-factor) model for all points in time, days of particular turbulence particularly stand out. To illustrate this, we display average and top three log predictive gains over the no-factor model in Table 7. The biggest gains can be seen on "Black Monday 2011" (August 8), when US stock markets tumbled after a credit rating downgrade of US sovereign debt by Standard and Poor's. The trading day before this, August 4, displays the third highest gain. The 27th of February in 2007 also proves to be an interesting date to consider. This day corresponds to the burst of the Chinese stock bubble that led to a major crash in Chinese stock markets, causing a severe decline in equity markets worldwide. It appears that joint modeling of stock prices is particularly important on days of extreme events when conditional correlations are often higher.

5.3. Using observable instead of latent factors

An alternative to estimating latent factors from data is to use observed factors instead (cf. Wang et al., 2011). To explore this route, we investigate an alternative model with four observed factors, the three Fama–French plus the Momentum factor.[7] The Fama–French factors consist of the excess return on the market, a size factor SMB, and a book-to-market factor HML (Fama and French, 1993); the momentum factor MOM (Carhart, 1997) captures the empirically observed tendency for falling asset prices to fall further, and rising prices to keep rising. For estimation of this model we proceed exactly as before, except that we omit the last step of our posterior sampler and keep $\boldsymbol{f}$ fixed at the observed values.

[7] The Fama–French+Momentum factors are available at a daily frequency at Kenneth French's web page at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html. The data was downloaded on February 21, 2017; missing values were replaced with zeros and the data was standardized to have unconditional mean zero and variance one.

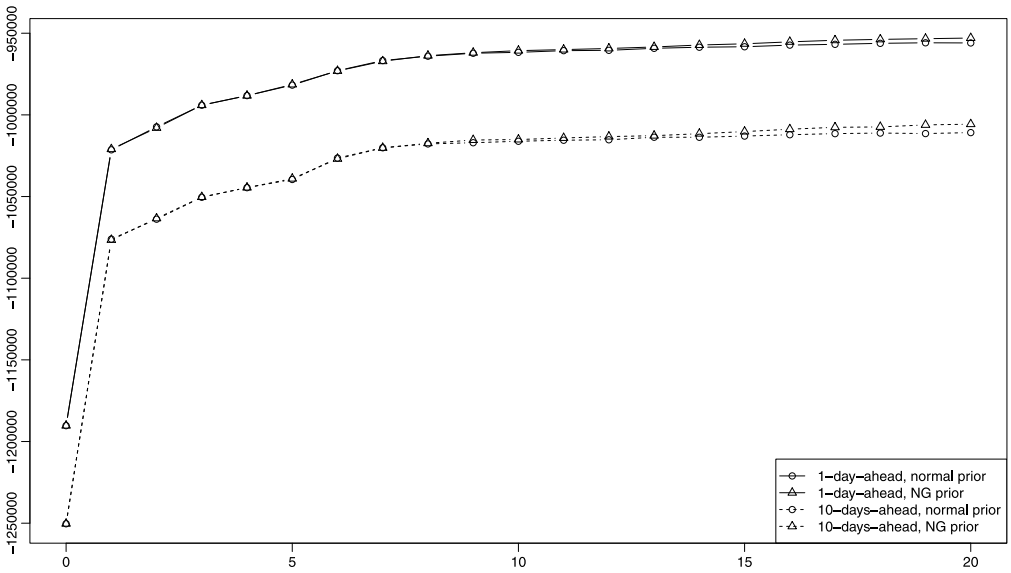

Fig. 8. Accumulated 1-day-ahead and 10-days-ahead log predictive likelihoods for models with $0, 1, \ldots, 20$ factors. Data until $t = 3000$ is treated as training data.

Table 7
Average and top 3 daily 1-day-ahead log predictive gains over the no-factor SV model.

	Mean	2/27/2007	8/4/2011	8/8/2011
1 factors, Gaussian	84.83	1415.13	1189.78	1446.03
2 factors, Gaussian	91.88	1455.39	1261.38	1462.23
3 factors, Gaussian	98.55	1449.31	1237.97	1479.84
4 factors, Gaussian	101.50	1449.21	1254.36	1502.17
10 factors, Gaussian	114.81	1482.36	1271.59	1557.63
20 factors, Gaussian	117.69	1481.15	1293.98	1568.60
1 factors, NG prior	85.03	1421.45	1194.08	1445.51
2 factors, NG prior	91.56	1450.88	1239.42	1461.73
3 factors, NG prior	98.57	1448.90	1231.94	1471.58
4 factors, NG prior	101.47	1443.86	1240.68	1536.77
10 factors, NG prior	115.32	1482.32	1283.90	1546.43
20 factors, NG prior	119.21	1499.08	1293.74	1578.54

Without presenting qualitative results in detail due to space constraints, we note that both the loadings on and the volatilities of the market excess returns show a remarkably close resemblance to those corresponding to the first latent factor displayed in Figs. 5 and 6. To a certain extent (although much less pronounced), this is also true for SMB and the second latent factor as well as HML and the fourth latent factor. However, most of the loadings on MOM are shrunk towards zero and there is no recognizable similarity to the remaining latent factor from the original model. Log predictive scores for this model are very close to those for the SV model with two latent factors. In what follows, we term this approach FF+MOM.

5.4. Comparison to other models

We now turn to investigating the statistical performance of the factor SV model via out-of-sample predictive measures as well as its suitability for optimal asset allocation. The competitors are: Moving averages (MAs) of sample covariance matrices over a window of 500 trading days, exponentially weighted moving averages (EWMAs) of sample covariances defined by $\boldsymbol{\Sigma}_{t+1} = (1-\alpha)\boldsymbol{y}_t'\boldsymbol{y}_t + \alpha\boldsymbol{\Sigma}_t$, the Ledoit–Wolf shrinkage estimator (Ledoit and Wolf, 2004) and FF+MOM described above. While this choice is certainly not exhaustive, it includes many of the approaches most widely used in practice.

For comparison, we use two benchmarking methods. First, we consider the minimum variance portfolio implied by $\hat{\boldsymbol{\Sigma}}_{t+1}$ (the point estimate or posterior mean estimate, respectively) which uniquely defines the optimal portfolio weights

$$\boldsymbol{\omega}_{t+1} = \frac{\hat{\boldsymbol{\Sigma}}_{t+1}^{-1}\boldsymbol{\iota}}{\boldsymbol{\iota}'\hat{\boldsymbol{\Sigma}}_{t+1}^{-1}\boldsymbol{\iota}},$$

where $\boldsymbol{\iota}$ denotes an m-variate vector of ones. Using these weights, we compute the corresponding realized portfolio returns r_{t+1} for $t = 3000, 3001, \ldots, 3999$, effectively covering an evaluation period from 5/3/2006 to 3/1/2010. In the first three

Table 8
Predictive performance measures, averaged over 1000 trading days after 5/3/2006. SD: Annualized empirical standard deviations of portfolio returns. Avg: Annualized average excess returns over the equal weight portfolio. Sharpe: Quotients of Avg and SD. PLPS: Average one-day-ahead pseudo log predictive scores over the no-factor SV model. N-FSV stands for the factor SV model with the standard normal prior. NG-FSV stands for the factor SV model with row-wise Normal-Gamma prior ($a_i \equiv 0.1$).

	SD	Avg	Sharpe	PLPS
Equal weight portfolio	29.02	0.00	0.00	
MA (500 days)	16.56	−2.39	−0.14	−320.17
EWMA ($\alpha = 0.94$)	35.90	−4.43	−0.12	-3×10^9
EWMA ($\alpha = 0.99$)	18.42	−4.18	−0.23	−1409.97
Ledoit–Wolf	12.53	6.14	0.49	25.68
FF+MOM	16.82	12.68	0.75	74.56
no-factor SV	22.53	1.02	0.05	0.00
N-FSV 1	19.03	9.45	0.50	66.29
N-FSV 2	18.92	9.32	0.49	75.25
N-FSV 3	15.50	7.36	0.48	81.47
N-FSV 4	14.58	8.26	0.57	87.07
N-FSV 10	12.94	12.45	0.96	97.74
N-FSV 20	12.57	7.81	0.62	101.82
N-FSV 50	12.41	5.03	0.40	106.57
NG-FSV 1	18.98	9.44	0.50	66.29
NG-FSV 2	18.84	9.37	0.50	75.21
NG-FSV 3	15.52	7.20	0.46	81.39
NG-FSV 4	14.85	8.36	0.56	86.80
NG-FSV 10	12.94	12.00	0.93	97.74
NG-FSV 20	12.02	13.33	1.11	102.30
NG-FSV 50	12.42	10.78	0.87	107.23

columns of Table 8, we report annualized empirical standard deviations, annualized average excess returns over those obtained from the equal weight portfolio, and the quotient of these two measures, the Sharpe ratio (Sharpe, 1966).

Considering the portfolio standard deviation presented in the first column of Table 8, it turns out that the Ledoit–Wolf shrinkage estimator implies an annualized standard deviation of about 12.5 which is only matched by factor SV models with many factors. Lower-dimensional factor SV models, including FF+MOM, as well as simple MAs and highly persistent EWMAs do not perform quite as well but are typically well below 20. Less persistent EWMAs, the no-factor SV model and the naïve equal weight portfolio exhibit standard deviations higher than 20. Considering average returns, FF+MOM and factor SV models with around 10 to 20 factors tend to do well for the given time span. In the third column, we list Sharpe ratios, where factor SV models with 10 to 20 factors show superior performance, in particular when the Normal-Gamma prior is employed. Note however that column two and three have to be interpreted with some care, as average asset and portfolio returns generally have a high standard error.

Second, we use what we coin *pseudo log predictive scores* (PLPSs), i.e. Gaussian approximations to the actual log predictive scores. This simplification is necessary because most of the above-mentioned methods only deliver point estimates of the forecast covariance matrix and it is not clear how to properly account for estimation uncertainty. Moreover, the PLPS is simpler to evaluate as there is no need to numerically solve a high-dimensional integral. Consequently, it is frequently used instead of the actual LPS in high dimensions while still allowing for evaluation of the covariance accuracy (Adolfson et al., 2007; Carriero et al., 2016; Huber, 2016). More specifically, we use data up to time t to determine a point estimate $\hat{\boldsymbol{\Sigma}}_{t+1}$ for $\boldsymbol{\Sigma}_{t+1}$ and compute the logarithm of the multivariate Gaussian density $\mathcal{N}_m\big(\mathbf{0}, \hat{\boldsymbol{\Sigma}}_{t+1}\big)$ evaluated at the actually observed value $\boldsymbol{y}^o_{t+1}$ to obtain the one-day-ahead PLPS for time $t+1$.

In terms of average PLPSs (the last column in Table 8), factor SV models clearly outperform all other models under consideration. In particular, even if r is chosen as small as $r = 4$ to match the number of factors, the model with latent factors outperforms FF+MOM. Using latent factors is generally preferable; note however that the 4-factor FF+MOM does better than the single- and no-factor SV models. Generally speaking, many factors appear to be needed for accurately representing the underlying data structure, irrespectively of the prior choice. Considering the computational simplicity of the Ledoit–Wolf estimator, its prediction accuracy is quite remarkable. It clearly outperforms the no-factor SV model which, in turn, beats simple MAs and EWMAs.

6. Conclusion and outlook

The aim of this paper was to present an efficient and parsimonious method of estimating high-dimensional time-varying covariance matrices through factor stochastic volatility models. We did so by proposing an efficient Bayesian MCMC algorithm that incorporates parsimony by modeling the covariance structure through common latent factors which themselves follow univariate SV processes. Moreover, we added additional sparsity by utilizing a hierarchical shrinkage prior, the Normal-Gamma prior, on the factor loadings. We showed the effectiveness of our approach through simulation studies and illustrated the effect of different shrinkage specifications. We applied the algorithm to a high-dimensional data set consisting of stock returns of 300 S&P 500 members and conducted an out-of-sample predictive study to compare different

prior settings and investigate the choice of the number of factors. Moreover, we discussed the out-of-sample performance of a minimum variance portfolio constructed from the model-implied weights and related it to a number of competitors often used in practice.

Because the algorithm scales linearly in both the series length T as well as the number of component series m, applying it to even higher dimensions is straightforward. We have experimented with simulated data in thousands of dimensions for thousands of points in time and successfully recaptured the time-varying covariance matrix.

Further research could be directed towards incorporating prior knowledge into building the hierarchical structure of the Normal-Gamma prior, e.g. by choosing the global shrinkage parameters according to industry sectors. Alternatively, Villani et al. (2009) propose a mixture of experts model to cater for smoothly changing regression densities. It might be fruitful to adopt this idea in the context of covariance matrix estimation by including either observed (Fama–French) or latent factors as predictors and allowing for other mixture types than the ones discussed there. While not being the focus of this work, it is easy to extend the proposed method by exploiting the modular nature of Markov chain Monte Carlo methods. In particular, it is straightforward to combine it with mean models such as (sparse) vector autoregressions (e.g., Bańbura et al., 2010; Kastner and Huber, 2018), dynamic regressions (e.g., Korobilis, 2013), or time-varying parameter models (e.g., Koop and Korobilis, 2013; Huber et al., 2019).

Acknowledgments

Variants of this paper were presented at the 6th European Seminar on Bayesian Econometrics (ESOBE 2015), the 2015 NBER-NSF Time Series Conference, the 2nd Vienna Workshop on High-Dimensional Time Series in Macroeconomics and Finance 2015, the 2015 International Work-Conference on Time Series Analysis, the 2016 ISBA World Meeting, the 10th International Conference on Computational and Financial Econometrics 2016, the CORE Econometrics & Finance Seminar 2017 and the 61st World Statistics Congress 2017. The author thanks all participants, in particular Luc Bauwens, Manfred Deistler, John Geweke, Florian Huber, Sylvia Kaufmann, Hedibert Lopes, Ruey Tsay, Stefan Voigt, Mike West, as well as the handling editor Herman van Dijk and two anonymous referees for crucially valuable comments and suggestions. Special thanks go to Mark Jensen who discussed this paper at the ESOBE 2015 and Sylvia Frühwirth-Schnatter who supported the author throughout the development of this work.

References

Adolfson, M., Lindé, J., Villani, M., 2007. Forecasting performance of an open economy DSGE model. Econometric Rev. 26 (2–4), 289–328.
Aguilar, O., West, M., 2000. Bayesian dynamic factor models and portfolio allocation. J. Bus. Econ. Stat. 18 (3), 338–357.
Ahelegbey, D.F., Billio, M., Casarin, R., 2016. Bayesian graphical models for structural vector autoregressive processes. J. Appl. Econometrics 31 (2), 357–386.
Anderson, T.W., Rubin, H., 1956. Statistical inference in factor analysis. In: Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 5: Contributions to Econometrics, Industrial Research, and Psychometry. University of California Press, Berkeley, Calif., pp. 111–150.
Bai, J., Ng, S., 2002. Determining the number of factors in approximate factor models. Econometrica 70 (1), 191–221.
Bańbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector auto regressions. J. Appl. Econometrics 25 (1), 71–92.
Baştürk, N., Borowska, A., Grassi, S., Hoogerheide, L., van Dijk, H.K., 2018. Forecast density combinations of dynamic models and data driven portfolio strategies. J. Econometrics.
Belmonte, M.A., Koop, G., Korobilis, D., 2014. Hierarchical shrinkage in time-varying parameter models. J. Forecast. 33 (1), 80–94.
Bhattacharya, A., Dunson, D.B., 2011. Sparse Bayesian infinite factor models. Biometrika 98 (2), 291–309.
Bitto, A., Frühwirth-Schnatter, S., 2018. Achieving shrinkage in a time-varying parameter model framework. J. Econometrics.
Carhart, M.M., 1997. On persistence in mutual fund performance. J. Finance 52 (1), 57–82.
Caron, F., Doucet, A., 2008. Sparse Bayesian nonparametric regression. In: Proceedings of the 25th International Conference on Machine Learning (ICML'2008). Helsinki, Finland, pp. 88–95.
Carriero, A., Clark, T.E., Marcellino, M., 2016. Common drifting volatility in large Bayesian VARs. J. Bus. Econom. Statist. 34 (3), 375–390.
Chib, S., Nardari, F., Shephard, N., 2006. Analysis of high dimensional multivariate stochastic volatility models. J. Econometrics 134 (2), 341–371.
Engle, R.F., Kelly, B., 2012. Dynamic equicorrelation. J. Bus. Econom. Statist. 30 (2), 212–228.
Fama, E.F., French, K.R., 1993. Common risk factors in the returns on stocks and bonds. J. Financ. Econ. 33 (1), 3–56.
Frühwirth-Schnatter, S., Lopes, H.F., 2018. Sparse Bayesian factor analysis when the number of factors is unknown. arXiv pre-print 1804.04231. URL: https://arxiv.org/abs/1804.04231.
Frühwirth-Schnatter, S., Tüchler, R., 2008. Bayesian parsimonious covariance estimation for hierarchical linear mixed models. Stat. Comput. 18 (1), 1–13.
Frühwirth-Schnatter, S., Wagner, H., 2010. Stochastic model specification search for Gaussian and partial non-Gaussian state space models. J. Econometrics 154 (1), 85–100.
Geweke, J., Amisano, G., 2010. Comparing and evaluating Bayesian predictive distributions of asset returns. Int. J. Forecast. 26 (2), 216–230.
Griffin, J.E., Brown, P.J., 2010. Inference with normal-gamma prior distributions in regression problems. Bayesian Anal. 5 (1), 171–188.
Griffin, J.E., Brown, P.J., 2017. Hierarchical shrinkage priors for regression models. Bayesian Anal. 12 (1), 135–159.
Gruber, L., West, M., 2016. GPU-accelerated Bayesian learning and forecasting in simultaneous graphical dynamic linear models. Bayesian Anal. 11 (1), 125–149.
Han, Y., 2006. Asset allocation with a high dimensional latent factor stochastic volatility model. Rev. Financ. Stud. 19 (1), 237–271.
Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Rev. Econom. Stud. 61 (2), 247–264.
Hörmann, W., Leydold, J., 2013. Generating generalized inverse Gaussian random variates. Stat. Comput. 24 (4), 1–11.
Huber, F., 2016. Density forecasting using Bayesian global vector autoregressions with stochastic volatility. Int. J. Forecast. 32 (3), 818–837.
Huber, F., Feldkircher, M., 2017. Adaptive shrinkage in Bayesian vector autoregressive models. J. Bus. Econ. Stat.
Huber, F., Kastner, G., Feldkircher, M., 2019. Should I stay or should I go? A latent threshold approach to large-scale mixture innovation models. J. Appl. Econometrics.
Ishihara, T., Omori, Y., 2017. Portfolio optimization using dynamic factor and stochastic volatility: Evidence on fat-tailed error and leverage. Jpn. Econ. Rev. 68 (1), 63–94.

Jungbacker, B., Koopman, S.J., 2006. Monte Carlo likelihood estimation for three multivariate stochastic volatility models. Econometric Rev. 25 (2–3), 385–408.
Kastner, G., 2016. Dealing with stochastic volatility in time series using the R package **stochvol**. J. Stat. Softw. 69 (5), 1–30.
Kastner, G., 2017. **factorstochvol**: Bayesian estimation of (sparse) latent factor stochastic volatility models. R package version 0.8.4. URL: https://CRAN.R-project.org/package=factorstochvol.
Kastner, G., Frühwirth-Schnatter, S., 2014. Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. Comput. Statist. Data Anal. 76, 408–423.
Kastner, G., Frühwirth-Schnatter, S., Lopes, H.F., 2017. Efficient Bayesian inference for multivariate factor stochastic volatility models. J. Comput. Graph. Statist. 26 (4), 905–917.
Kastner, G., Huber, F., 2018. Sparse Bayesian vector autoregressions in huge dimensions. arXiv pre-print 1704.03239. URL: https://arxiv.org/abs/1704.03239.
Kaufmann, S., Schumacher, C., 2018. Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification. J. Econometrics.
Koop, G., Korobilis, D., 2013. Large time-varying parameter VARs. J. Econometrics 177 (2), 185–198.
Korobilis, D., 2013. Hierarchical shrinkage priors for dynamic regressions with many predictors. Int. J. Forecast. 29 (1), 43–59.
Ledoit, O., Wolf, M., 2004. Honey, I shrunk the sample covariance matrix. J. Portfolio Manage. 30 (4), 110–119.
Leydold, J., Hörmann, W., 2017. **GIGrvg**: Random variate generator for the GIG distribution. R package version 0.5. URL: http://CRAN.R-project.org/package=GIGrvg.
Liesenfeld, R., Jung, R.C., 2000. Stochastic volatility models: Conditional normality versus heavy-tailed distributions. J. Appl. Econometrics 15 (2), 137–160.
Liesenfeld, R., Richard, J.-F., 2006. Classical and Bayesian analysis of univariate and multivariate stochastic volatility models. Econometric Rev. 25 (2–3), 335–360.
Loddo, A., Ni, S., Sun, D., 2011. Selection of multivariate stochastic volatility models via Bayesian stochastic search. J. Bus. Econom. Statist. 29 (3), 342–355.
Lopes, H.F., Carvalho, C.M., 2007. Factor stochastic volatility with time varying loadings and markov switching regimes. J. Statist. Plann. Inference 137 (10), 3082–3091.
Lopes, H.F., McCulloch, R.E., Tsay, R.S., 2018. Parsimony inducing priors for large scale state-space models. Working paper.
Nakajima, J., West, M., 2013. Dynamic factor volatility modeling: A Bayesian latent threshold approach. J. Financ. Econ. 11 (1), 116–153.
Nakajima, J., West, M., 2017. Dynamics and sparsity in latent threshold factor models: A study in multivariate EEG signal processing. Braz. J. Probab. Statist. 31 (4), 701–731.
Oh, D.H., Patton, A.J., 2017. Modeling dependence in high dimensions with factor copulas. J. Bus. Econom. Statist. 35 (1).
Pakel, C., Shephard, N., Sheppard, K., Engle, R.F., 2017. Fitting vast dimensional time-varying covariance models. Working paper.
Park, T., Casella, G., 2008. The Bayesian lasso. J. Amer. Statist. Assoc. 103 (452), 681–686.
Pati, D., Bhattacharya, A., Pillai, N.S., Dunson, D., 2014. Posterior contraction in sparse Bayesian factor models for massive covariance matrices. Ann. Statist. 42 (3), 1102–1130.
Philipov, A., Glickman, M.E., 2006. Factor multivariate stochastic volatility via Wishart processes. Econometric Rev. 25 (2–3), 311–334.
Pitt, M.K., Shephard, N., 1999. Time-varying covariances: A factor stochastic volatility approach. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics 6 – Proceedings of the Sixth Valencia International Meeting. Oxford University Press, pp. 547–570.
Polson, N.G., Scott, J.G., 2011. Shrink globally, act locally: Sparse Bayesian regularization and prediction. In: Bernardo, J.M., Bayarri, M.J., Berger, J.O., P., D.A., Heckerman, D., Smith, A.F.M., West, M. (Eds.), Bayesian Statistics 9 – Proceedings of the Ninth Valencia International Meeting. Oxford University Press, Oxford, pp. 501–538.
R Core Team, 2017. R: A Language and environment for statistical computing. R Foundation for Statistical Computing. URL: https://www.R-project.org/.
Rigobon, R., 2003. Identification through heteroskedasticity. Rev. Econ. Stat. 85 (4), 777–792.
Sentana, E., Fiorentini, G., 2001. Identification, estimation and testing of conditionally heteroskedastic factor models. J. Econometrics 102 (2), 143–164.
Sharpe, W.F., 1966. Mutual fund performance. J. Bus. 39 (1), 119–138.
Villani, M., Kohn, R., Giordani, P., 2009. Regression density estimation using smooth adaptive Gaussian mixtures. J. Econometrics 153 (2), 155–173.
Wang, H., Reeson, C., Carvalho, C.M., 2011. Dynamic financial index models: Modeling conditional dependencies via graphs. Bayesian Anal. 6 (4), 639–664.
Yu, J., 2005. On leverage in a stochastic volatility model. J. Econometrics 127 (2), 165–178.
Zhao, Z.Y., Xie, M., West, M., 2016. Dynamic dependence networks: Financial time series forecasting and portfolio decisions. Appl. Stoch. Models Bus. Ind. 32 (3), 311–332.
Zhou, X., Nakajima, J., West, M., 2014. Bayesian forecasting and portfolio decisions using dynamic dependent sparse factor models. Int. J. Forecast. 30 (4), 963–980.