CQE Center for Quantitative Economics

Bayesian semiparametric multivariate stochastic volatility with an application to international stock-market co-movements

Martina Danielova Zaharieva [†], Mark Trede [†] und Bernd Wilfling [†]

62/2017

First Version: June 20, 2017
This Version: June 29, 2018

[†] Department of Economics, University of Münster, Germany

wissen.leben
WWU Münster

Bayesian semiparametric multivariate stochastic volatility with an application to international stock-market co-movements

Martina Danielova Zaharieva[a], Mark Trede[a], Bernd Wilfling[a,*]

[a] *Westfälische Wilhelms-Universität, Department of Economics (CQE), Am Stadtgraben 9, 48143 Münster, Germany*

(Date of this version: June 29, 2018)

Abstract

In this paper, we establish a Cholesky-type multivariate stochastic volatility estimation framework, in which we let the innovation vector follow a Dirichlet process mixture (DPM), thus enabling us to model highly flexible return distributions. The Cholesky decomposition allows parallel univariate process modeling and creates potential for estimating high-dimensional specifications. We use Markov Chain Monte Carlo methods for posterior simulation and predictive density computation. We apply our framework to a five-dimensional stock-return data set and analyze international stock-market co-movements among the largest stock markets. The empirical results show that our DPM modeling of the innovation vector yields substantial gains in out-of-sample forecst accuracy when compared with the prevalent benchmark models.

Keywords: Bayesian nonparametrics, Markov Chain Monte Carlo, Dirichlet process mixture, multivariate stochastic volatility, stock-market co-movements.

JEL classification: C11, C14, C53, C58, G10.

*Corresponding author.
E-mail addresses: martina.zaharieva@wiwi.uni-muenster.de (M. Danielova Zaharieva), mark.trede@uni-muenster.de (M. Trede), bernd.wilfling@wiwi.uni-muenster.de (B. Wilfling).

Bayesian semiparametric multivariate stochastic volatility with an application to international stock-market co-movements

Abstract

In this paper, we establish a Cholesky-type multivariate stochastic volatility estimation framework, in which we let the innovation vector follow a Dirichlet process mixture (DPM), thus enabling us to model highly flexible return distributions. The Cholesky decomposition allows parallel univariate process modeling and creates potential for estimating high-dimensional specifications. We use Markov Chain Monte Carlo methods for posterior simulation and predictive density computation. We apply our framework to a five-dimensional stock-return data set and analyze international stock-market co-movements among the largest stock markets. The empirical results show that our DPM modeling of the innovation vector yields substantial gains in out-of-sample forecst accuracy when compared with the prevalent benchmark models.

Keywords: Bayesian nonparametrics, Markov Chain Monte Carlo, Dirichlet process mixture, multivariate stochastic volatility, stock-market co-movements.

JEL classification: C11, C14, C53, C58, G10.

1 Introduction

Owing to increasingly integrated financial markets, both domestically and internationally, volatility modeling and the analysis of volatility co-movements and spillovers among multiple asset returns have become central topics for the last few decades (*inter alia* Ehrmann et al., 2011; Clements et al., 2015). The two by far the most popular volatility model classes discussed in the literature are the generalized autoregressive conditional heteroscedasticity (GARCH-type) models (Engle 1982; Bollerslev 1986) and the stochastic volatility (SV) models (Taylor, 1982; 1986), both in univariate and multivariate variants. Several in-depth overview articles on multivariate GARCH (Bauwens et al., 2006) and SV models (Chib et al., 2009) document the enormous professional interest in the field. While both model classes have distinct advantages on their own, a major characteristic of SV specifications is that they model the unobserved volatility directly as a separate stochastic process. This converts many SV specifications into discrete-time versions of continuous-time models that are well-established in finance theory, which constitutes the general attraction of SV models (Harvey et al., 1994; Kim et al., 1998, Asai et al., 2006).

Irrespective of model selection issues, various stylized empirical properties of asset returns have been discovered in real-world data, the most prominent being the fat-tail (kurtotic) nature of the return distribution. Cont (2001) reports that "... the (unconditional) distribution of returns seems to display a power-law or Pareto-like tail, with a tail index which is finite, higher than two and less than five for most data sets studied. In particular, this excludes stable laws with infinite variance and the normal distribution.". Interestingly, the fat-tail property even persists after correcting the financial returns for volatility clustering (e.g. via GARCH-type models), although to a less pronounced degree. Numerous attempts have been made to account for the fat-tail property by replacing the Gaussianity assumption with alternative parametric distributions for the return innovation in distinct volatility models. Recently, however, several authors have proposed the nonparametric modeling of return innovation as a Dirichlet process mixture (DPM) and emphasize the flexibility increase associated with this approach, compared to using parametric distributions. In particular, to date, the nonparametric DPM approach has been applied successfully (i) to univariate SV

modeling by Jensen and Maheu (2010, 2014) and Delatola and Griffin (2011, 2013), (ii) to univariate GARCH modeling by Ausín et al. (2014), and (iii) to multivariate GARCH modeling by Jensen and Maheu (2013) and Virbickaitė et al. (2016).

In this paper, we complete the above-described list by integrating the nonparametric DPM approach into a specific class of multivariate SV models with time-varying covariance components, based on the Cholesky decomposition of volatility matrices (see e.g. Nakajima, 2017). We establish a Bayesian estimation procedure for this semi-parametric overall framework and study its predictive abilities by means of predictive density evaluation. In the empirical part, we apply our econometric setup to a five-country data set, in order to analyze co-movements among the most important stock markets worldwide. In an out-of-sample forecasting comparison with two conventional models of the error-term distribution (multivariate normal and Student-t), we find that our DPM modeling yields substantial accuracy gains.

The paper is organized as follows. Section 2 reviews (i) the multivariate SV model based on Cholesky decomposition, and (ii) the Dirichlet process mixture. In Section 3, we present the Markov Chain Monte Carlo (MCMC) algorithm for Bayesian inference. Section 4 presents essential probabilistic features of our econometric framework and provides a simulation example. Section 5 contains the empirical application to daily returns from the five largest international stock markets. Section 6 concludes.

2 Model development

2.1 Cholesky Multivariate Stochastic Volatility (MSV)

In order to introduce Cholesky MSV modeling, we follow the approach of Primiceri (2005) and Nakajima (2017) and consider the $m \times 1$ vector $\mathbf{y}_t = (y_{1t}, \ldots, y_{mt})'$ of time series observations at date t, which we assume to follow an m-dimensional multivariate normal distribution with zero-expectation vector, $\mathbb{E}(\mathbf{y}_t) = \mathbf{0}$, and time-varying covariance matrix $\mathrm{Cov}(\mathbf{y}_t) = \mathbf{H}_t$, i.e. $\mathbf{y}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{H}_t)$. The Cholesky decomposition of $\mathbf{H}_t$ is given by the factorization

$$\mathbf{A}_t \mathbf{H}_t \mathbf{A}_t' = \mathbf{\Sigma}_t \mathbf{\Sigma}_t, \tag{1}$$

where $\mathbf{A}_t$ is the lower triangular matrix with 1s along the principal diagonal and $\boldsymbol{\Sigma_t}$ is a diagonal matrix with time-varying elements:

$$\mathbf{A}_t = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \alpha_{21,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \alpha_{m1,t} & \cdots & \alpha_{mm-1,t} & 1 \end{pmatrix}, \qquad \boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t' = \begin{pmatrix} \sigma_{1,t} & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \sigma_{m,t} \end{pmatrix}. \tag{2}$$

Via the Eqs. (1) and (2), the standard Cholesky MSV model is then defined as

$$\mathbf{y}_t = \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\epsilon}_t, \tag{3}$$

$$\mathbf{H}_t = \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})', \tag{4}$$

where the innovation vector $\boldsymbol{\epsilon}_t$ is assumed to follow the m-dimensional multivariate standard normal distribution $\mathbf{N}(\mathbf{0}, \mathbf{I})$. Based on Eqs. (2) and (3), several alternative Cholesky MSV models have been proposed in the literature, by letting the innovation vector $\boldsymbol{\epsilon}_t$ follow distributions other than the multivariate standard normal, for example, the multivariate t (originally Harvey et al., 1994, in a non-Cholesky MSV framework), and the multivariate generalized hyperbolic skew t distribution (Nakajima, 2017). The latter specification retains the essential Cholesky structure, but makes more realistic distributional assumptions, with the aim of more effectively capturing some stylized facts of financial return data (like leverage effects and skewness). In the next section, we define a new class of Cholesky MSV models by letting $\boldsymbol{\epsilon}_t$ follow a Dirichlet process mixture, in order to account for excess kurtosis in the data.

When it comes to Bayesian estimation of Cholesky MSV models with the time-varying parameters from Eq. (2), we adopt the common methodology of reducing the multivariate dynamics to univariate volatility processes that form a state-space representation (Lopes, 2012). Specifically, we collect the parameters from the matrix $\mathbf{A}_t$ in the $[m(m-1)/2] \times 1$ vector $\boldsymbol{\alpha}_t$, and for the elements from $\boldsymbol{\Sigma}_t$ we define the $m \times 1$ vector $\mathbf{h}_t$ as follows:

$$\boldsymbol{\alpha}_t = (\alpha_{21,t}, \alpha_{31,t}, \alpha_{32,t}, \ldots, \alpha_{m1,t}, \ldots, \alpha_{mm-1,t})', \tag{5}$$

$$\mathbf{h}_t = (\log(\sigma_{1,t}^2), \ldots, \log(\sigma_{m,t}^2))'. \tag{6}$$

(The parameters in $\boldsymbol{\alpha}_t$ are collected row by row from matrix $\mathbf{A}_t$.) We then specify the dynamics of the Cholesky parameters as the (stationary) AR(1) processes

$$\begin{aligned} \boldsymbol{\alpha}_t &= \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \boldsymbol{\Phi}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}) + \mathbf{e}_t, & (7) \\ \mathbf{h}_t &= \boldsymbol{\Phi}_{\mathbf{h}}\mathbf{h}_{t-1} + \boldsymbol{\eta}_t, & (8) \\ \begin{pmatrix} \mathbf{e}_t \\ \boldsymbol{\eta}_t \end{pmatrix} &\sim \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{\mathbf{e}} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\boldsymbol{\eta}} \end{bmatrix}\right), & (9) \end{aligned}$$

where we assume (i) that the matrices $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}, \boldsymbol{\Phi}_{\mathbf{h}}, \boldsymbol{\Sigma}_{\mathbf{e}}, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}$ are all diagonal, and (ii) that the $p = m(m-1)/2$ diagonal entries $\phi_{\boldsymbol{\alpha}1}, \ldots, \phi_{\boldsymbol{\alpha}p}$ of $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$ and the m diagonal entries $\phi_{\mathbf{h}1}, \ldots, \phi_{\mathbf{h}m}$ of $\boldsymbol{\Phi}_{\mathbf{h}}$ are all less than 1 in absolute value (stationarity conditions).[1]

2.2 Bayesian semiparametric Cholesky MSV

Finally, it remains to specify the distribution of the innovation vector $\boldsymbol{\epsilon}_t$ from Eq. (3), which we model as a nonparametric Dirichlet process mixture (DPM). The DPM represents an infinite mixture model and constitutes an extremely flexible extension of finite mixture models (Jensen and Maheu, 2010, 2013; Kalli et al., 2013; Maheu and Yang, 2016; Virbickaitė et al., 2016). In introducing the DPM, we need to consider the Dirichlet process $\text{DP}(c, G_0)$, defined in terms of the base distribution G_0 and the concentration parameter c (Ferguson,1973). In a Bayesian context, the base distribution G_0 represents the prior distribution of the component parameters in the infinite mixture, while the parameter c, roughly speaking, controls for the number of clusters in the mixture. A small value of c can be thought of as *a priori* indicating a small number of components with relatively large weights in the infinite mixture, whereas large values of c *a priori* assume many mixture components, all with relatively small weights.

Overall, our semiparametric Cholesky MSV specification, in which we model the $m \times m$ matrix $\mathbf{H}_t$ from Eq. (4) parametrically, while we let the distribution of the innovation vector $\boldsymbol{\epsilon}_t$ follow the nonparametric DPM as given in Eq. (17) below, has the

[1]Note that we specify the AR(1) process for $\mathbf{h}_t$ in Eq. (8) without an intercept term. This is due to an identification problem that would arise in the case of a non-zero intercept; see Jensen and Maheu (2010).

following hierarchical representation:

$$\begin{aligned}
\mathbf{y}_t|\,\boldsymbol{\Lambda}_t, \mathbf{A}_t, \boldsymbol{\Sigma}_t &\sim \mathbf{N}(\mathbf{0}, \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})'), && (10)\\
\mathbf{H}_t &= \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t(\mathbf{A}_t')^{-1}, && (11)\\
\boldsymbol{\Lambda}_t &= \mathrm{diag}(\lambda_{1,t}, \ldots, \lambda_{m,t}), && (12)\\
\lambda_{i,t} &\overset{\text{i.i.d.}}{\sim} G_i, \qquad (i = 1, \ldots, m) && (13)\\
G_i|\,G_0, c_i &\sim \mathrm{DP}(c_i, G_0), && (14)\\
G_0 &\overset{\text{d}}{=} \mathrm{Gamma}(\nu_0/2, s_0/2), && (15)\\
c_i &\sim \mathrm{Gamma}(a_0, b_0), && (16)
\end{aligned}$$

and where the elements of $\mathbf{A}_t$ and $\boldsymbol{\Sigma}_t$ collected in the vectors $\boldsymbol{\alpha}_t$ and $\mathbf{h}_t$ follow the AR(1) processes from Eqs. (7) and (8), respectively.[2] In Eqs. (10) and (12), the $m \times m$ matrix $\boldsymbol{\Lambda}_t$ is the precision matrix, which we assume to be diagonal, in order to ensure identification of the model.[3] We model the diagonal entries $\lambda_{1,t}, \ldots, \lambda_{m,t}$ as i.i.d. (with respect to t) and place a nonparametric Dirichlet process prior on the distribution of $\lambda_{i,t}$; see Eqs. (13) and (14). As in Ausín et al. (2014), we specify the base distribution G_0 for the diagonal elements of $\boldsymbol{\Lambda}_t$ as the gamma distribution in Eq. (15).

Following the line of argument in Jensen and Maheu (2013), we emphasize that our hierarchical model (10) to (16) can be expressed in the Sethuraman (1994) stick-breaking representation of the DPM mixture model. This allows us to write the density function of each component of the innovation vector $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \ldots, \epsilon_{mt})'$ as an infinite scale-mixture of Gaussian distributions. That is, for $i = 1, \ldots, m$ we have

$$f(\epsilon_{it}|\,\omega_{i1}, \omega_{i2}, \ldots, l_{i1}, l_{i2}, \ldots) = \sum_{j=1}^{\infty} \omega_{ij} f_N\left(\epsilon_{it}|\,0, l_{ij}^{-1}\right), \tag{17}$$

where $f_N\left(\cdot|\,0, l_{ij}^{-1}\right)$ denotes the density function of the univariate normal distribution with zero mean and variance l_{ij}^{-1}. The mixture parameters l_{ij} prior is given in Eq. (15). It follows from the stick-breaking representation that the mixture weights are dis-

[2] In the hierarchical representation, $\overset{\text{d}}{=}$ means 'has the distribution'. The operator $\mathrm{diag}(\lambda_1, \ldots, \lambda_m)$ creates the diagonal $m \times m$ matrix, say $\mathbf{M}$, with $\mathbf{M}_{ii} = \lambda_i$ and $\mathbf{M}_{ij} = 0$ for $i \neq j$ $(i, j = 1, \ldots, m)$.

[3] *Prima facie*, the diagonal structure of $\boldsymbol{\Lambda}_t$ might appear restrictive. However, as will become evident below, it does not impose any severe restriction on model flexibility.

tributed as $\omega_{i1} = \upsilon_{i1}$, $\omega_{ij} = (1-\upsilon_{i1})\cdots(1-\upsilon_{ij-1})\cdot\upsilon_{ij}$ for $j > 1$, where $\upsilon_{i1}, \upsilon_{i2}, \ldots$ are i.i.d. Beta$(1, c_i)$ (beta distribution with parameters 1 and c_i). As described above, the choice of the concentration parameter c_i is crucial. In line with Escobar and West (1995), we assume a gamma hyper-prior distribution $c_i \sim \text{Gamma}(a_0, b_0)$; see Eq. (16). We remark that each of the m mixtures in Eq. (17) is related to its own specific concentration parameter c_i.

For notational convenience, we collect the parameters from the parametric part of our Cholesky MSV model in the vector $\mathbf{\Phi}$ (i.e. $\mathbf{\Phi}$ contains all parameters from $\boldsymbol{\mu_\alpha}, \mathbf{\Phi_\alpha}, \mathbf{\Phi_h}, \mathbf{\Sigma_\eta}, \mathbf{\Sigma_e}$), and all parameters from the nonparametric specification in the infinite dimensional entity $\mathbf{\Omega} = \{\omega_{ij}, l_{ij}\}_{i=1,\ldots,m;j=1,2,\ldots,\infty}$. In cases where we need to address all model parameters, we merge the partial parameter entities $\mathbf{\Phi}$ and $\mathbf{\Omega}$ into the full-parameter vector $\mathbf{\Theta}$.

3 Bayesian inference

In this section, we present the samplers for the single parameter-components of the Cholesky-Dirichlet-Process-Mixture-Multivariate-Stochastic-Volatility (Cholesky-DPM-MSV) established in Section 2. In Section 3.1, we apply Forward-Filtering-Backward-Sampling (FFBS) to the elements of the matrix $\mathbf{A}_t$ (Carter and Kohn, 1994). In Section 3.2, we use the volatility sampler suggested by Jacquier et al. (2002) for the volatility processes in the matrix $\mathbf{\Sigma}_t$. For the nonparametric DPM part of the Cholesky-MSV model, we apply the efficient slice-sampler according to Walker (2007) and Kalli et al. (2011) in Section 3.3.

3.1 Sampling the $\mathbf{A}_t$-elements

In order to apply FFBS sampling to the elements of the $\mathbf{A}_t$-matrix, we need to embed the $\mathbf{A}_t$-parameters in an appropriate state-space model. To this end, we first rewrite Eq. (3) of our Cholesky-MSV-DPM model as

$$\mathbf{A}_t \mathbf{y}_t = \mathbf{\Sigma}_t \boldsymbol{\epsilon}_t, \tag{18}$$

where $\mathbf{y}_t$ is observable, and $\mathbf{A}_t$ has the lower triangular form given in Eq. (2). As in Primiceri (2005), we next define the $m \times m(m-1)/2$ matrix

$$\mathbf{Z}_t = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ -y_{1t} & 0 & \cdots & 0 \\ 0 & -\mathbf{y}_{[1:2]t} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\mathbf{y}_{[1:m-1]t} \end{pmatrix}, \tag{19}$$

in which $\mathbf{y}_{[1:i]t}$ denotes the row vector $(y_{1t}, y_{2t}, \ldots, y_{it})$, so that Eq. (18) can be written as

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\Sigma}_t \boldsymbol{\epsilon}_t, \tag{20}$$

where $\boldsymbol{\alpha}_t$, defined in Eq. (5), follows the AR(1) process specified in Eq. (7). Finally, we replace $\boldsymbol{\epsilon}_t$ in Eq. (20) with $\boldsymbol{\Lambda}_t^{-1/2}\mathbf{u}_t$, where $\mathbf{u}_t$ is assumed to follow the m dimensional multivariate standard normal distribution $\mathbf{N}(\mathbf{0}, \mathbf{I})$, and obtain the desired state-space model via Eqs. (20) and (7) with observation and transition equations given by

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\Sigma}_t \boldsymbol{\Lambda}_t^{-1/2} \mathbf{u}_t \equiv \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \tag{21}$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu}_{\boldsymbol{\alpha}} + \boldsymbol{\Phi}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_{\boldsymbol{\alpha}}) + \mathbf{e}_t, \tag{22}$$

with $\boldsymbol{\xi}_t \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t})$, $\boldsymbol{\Sigma}_{\boldsymbol{\xi}_t} = \boldsymbol{\Sigma}_t \boldsymbol{\Lambda}_t^{-1} \boldsymbol{\Sigma}_t$ and

$$\begin{pmatrix} \boldsymbol{\xi}_t \\ \mathbf{e}_t \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{e}} \end{bmatrix}\right). \tag{23}$$

In order to apply FFBS based on Kalman filter recursion, we denote the entire history of the vector $\mathbf{y}_t$ and the matrices $\mathbf{Z}_t, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t}$ to date s by $\mathbf{y}^{(s)} \equiv \{\mathbf{y}_0, \ldots, \mathbf{y}_{s-1}, \mathbf{y}_s\}$, $\mathbf{Z}^{(s)} \equiv \{\mathbf{Z}_0, \ldots, \mathbf{Z}_{s-1}, \mathbf{Z}_s\}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)} \equiv \{\boldsymbol{\Sigma}_{\boldsymbol{\xi}_0}, \ldots, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_{s-1}}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_s}\}$, , respectively, and let

$$\boldsymbol{\alpha}_{t|s} = \mathbb{E}(\boldsymbol{\alpha}_t | \mathbf{y}^{(s)}, \mathbf{Z}^{(s)}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)}, \boldsymbol{\Sigma}_{\mathbf{e}}) \tag{24}$$

$$\mathbf{V}_{t|s} = \mathrm{Cov}(\boldsymbol{\alpha}_t | \mathbf{y}^{(s)}, \mathbf{Z}^{(s)}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)}, \boldsymbol{\Sigma}_{\mathbf{e}}). \tag{25}$$

Furthermore, we define the $p \times 1$ vector

$$\mathbf{c}_{\boldsymbol{\alpha}} \equiv (\mu_{\boldsymbol{\alpha}1}(1 - \phi_{\boldsymbol{\alpha}1}), \ldots, \mu_{\boldsymbol{\alpha}p}(1 - \phi_{\boldsymbol{\alpha}p}))', \tag{26}$$

where $\mu_{\boldsymbol{\alpha}1}, \ldots, \mu_{\boldsymbol{\alpha}p}$ are the elements of the vector $\boldsymbol{\mu}_{\boldsymbol{\alpha}}$ and $\phi_{\boldsymbol{\alpha}1}, \ldots, \phi_{\boldsymbol{\alpha}p}$ the diagonal entries of the matrix $\boldsymbol{\Phi}_{\boldsymbol{\alpha}}$ as defined above. Then, given the starting values $\boldsymbol{\alpha}_{0|0}$ and $\mathbf{V}_{0|0}$, the standard Kalman filter can be summarized as follows:

$$\begin{aligned}
\boldsymbol{\alpha}_{t|t-1} &= \mathbf{c}_{\boldsymbol{\alpha}} + \boldsymbol{\Phi}_{\boldsymbol{\alpha}} \boldsymbol{\alpha}_{t-1|t-1}, && (27)\\
\mathbf{V}_{t|t-1} &= \boldsymbol{\Phi}_{\boldsymbol{\alpha}} \mathbf{V}_{t-1|t-1} \boldsymbol{\Phi}'_{\boldsymbol{\alpha}} + \boldsymbol{\Sigma}_{\mathbf{e}}, && (28)\\
\mathbf{K}_t &= \mathbf{V}_{t|t-1} \mathbf{Z}'_t (\mathbf{Z}_t \mathbf{V}_{t|t-1} \mathbf{Z}'_t + \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t})^{-1}, && (29)\\
\boldsymbol{\alpha}_{t|t} &= \boldsymbol{\alpha}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1}), && (30)\\
\mathbf{V}_{t|t} &= \mathbf{V}_{t|t-1} - \mathbf{K}_t \mathbf{Z}_t \mathbf{V}_{t|t-1}. && (31)
\end{aligned}$$

The final entities $\boldsymbol{\alpha}_{T|T}$ and $\mathbf{V}_{T|T}$ contain the mean and variances of the normal distribution, from which we draw $\boldsymbol{\alpha}_T$. We use this value in the first step of the backward recursion that yields $\boldsymbol{\alpha}_{T-1|T}$ and $\mathbf{V}_{T-1|T}$, which we then use to draw $\boldsymbol{\alpha}_{T-1}$. The backward recursion iterates from $T-1$ to 0, and at date t, the update step is given by

$$\begin{aligned}
\boldsymbol{\alpha}_{t|t+1} &= \boldsymbol{\alpha}_{t|t} + \mathbf{V}_{t|t} \boldsymbol{\Phi}'_{\boldsymbol{\alpha}} \mathbf{V}^{-1}_{t+1|t} (\boldsymbol{\alpha}_{t+1} - \mathbf{c}_{\boldsymbol{\alpha}} - \boldsymbol{\Phi}_{\boldsymbol{\alpha}} \boldsymbol{\alpha}_{t|t}), && (32)\\
\mathbf{V}_{t|t+1} &= \mathbf{V}_{t|t} - \mathbf{V}_{t|t} \boldsymbol{\Phi}'_{\boldsymbol{\alpha}} \mathbf{V}^{-1}_{t+1|t} \boldsymbol{\Phi}_{\boldsymbol{\alpha}} \mathbf{V}_{t|t}. && (33)
\end{aligned}$$

As the prior distribution of the initial state $\boldsymbol{\alpha}_{0|0}$ we use a multivariate normal distribution (see Section 5) and, as mentioned above, assume the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{e}}$ to be diagonal with entries $\sigma^2_{\mathbf{e}1}, \ldots, \sigma^2_{\mathbf{e}p}$. Note that for each $i = 1, \ldots, p$ the unconditional expectation of the α_{it}-process is $E(\alpha_{it}) = \mu_{\boldsymbol{\alpha}i} = \frac{c_{\boldsymbol{\alpha}i}}{1-\phi_{\boldsymbol{\alpha}i}}$, so that the $3p = 3m(m-1)/2$ parameters to be sampled are

$$c_{\boldsymbol{\alpha}1}, \ldots, c_{\boldsymbol{\alpha}p}, \phi_{\boldsymbol{\alpha}1}, \ldots, \phi_{\boldsymbol{\alpha}p}, \sigma^2_{\mathbf{e}1}, \ldots, \sigma^2_{\mathbf{e}p}.$$

The sampling strategy for these parameters is readily obtained from standard Bayesian estimation of the linear regression model. The prior distributions for the $c_{\boldsymbol{\alpha}i}$- (or $\mu_{\boldsymbol{\alpha}i}$-) and $\phi_{\boldsymbol{\alpha}i}$-parameters are normal distributions (where the prior for the $\phi_{\boldsymbol{\alpha}i}$-parameters have to be restricted to ensure the p stationarity conditions $|\phi_{\boldsymbol{\alpha}i}| < 1$), while the prior distribution for $\sigma^2_{\mathbf{e}i}$ is chosen as the inverse Gamma distribution. We sample the $c_{\boldsymbol{\alpha}i}$- and $\phi_{\boldsymbol{\alpha}i}$-parameters by the Metropolis-Hastings (MH) algorithm, while the

$\sigma^2_{\mathbf{e}i}$-parameters are sampled directly (given the conjugate prior).

3.2 Sampling the $\mathbf{\Sigma}_t$-elements

Defining $\widetilde{\mathbf{y}}_t = \mathbf{A}_t \mathbf{y}_t$, we note that $\widetilde{\mathbf{y}}_t$ has a diagonal covariance matrix, what enables us to independently estimate the m univariate stochastic volatility models.[4] The ith univariate stochastic volatility model is given by

$$\tilde{y}_{it} = \sigma_{i,t} \lambda_{i,t}^{-1/2} u_{it}, \qquad (i = 1, \ldots, m) \tag{34}$$

with $u_{it} \sim N(0,1)$. At this stage, we consider the matrix $\mathbf{A}_t$ as given and since $\mathbf{y}_t$ is observable, the values of $\tilde{y}_{it}$ can be computed. We note that the associated dynamic model (in state-space form) is nonlinear:

$$\begin{aligned} \tilde{y}_{it} &= \exp\{h_{it}/2\} \lambda_{i,t}^{-1/2} u_{it}, \qquad (i = 1, \ldots, m) && (35) \\ h_{it} &= \phi_{\mathbf{h}i} h_{it-1} + \eta_{it}, && (36) \end{aligned}$$

with $\eta_{it} \sim N(0, \sigma^2_{\boldsymbol{\eta}i})$ and $\sigma^2_{\boldsymbol{\eta}i}$ being the ith diagonal entry of the matrix $\mathbf{\Sigma}_{\boldsymbol{\eta}}$.

The m univariate SV models from Eqs. (35) and (36) can be estimated separately by consecutively sampling from the following conditionals, in the representation of which we use the m row vectors $\boldsymbol{\vartheta}_i = (\sigma^2_{\boldsymbol{\eta}i}, \phi_{\mathbf{h}i})$:

1. $\pi(\boldsymbol{\vartheta}_i | h_{i1}, \ldots, h_{iT})$, yielding the AR parameters.

2. $\pi(h_{i1}, \ldots, h_{iT} | \tilde{y}_{i1}, \ldots, \tilde{y}_{iT}, \boldsymbol{\vartheta}_i, \{l_{ij}\}_{j=1}^{\infty}, \{\omega_{ij}\}_{j=1}^{\infty})$, yielding the parametric volatility component.

3. $\pi(\{l_{ij}\}_{j=1}^{\infty}, \{\omega_{ij}\}_{j=1}^{\infty} | \tilde{y}_{i1}, \ldots, \tilde{y}_{iT}, h_{i1}, \ldots, h_{iT},$ yielding the nonparametric volatility component.

Sampling from the first conditional is straightforward and analogous to sampling the $\boldsymbol{\alpha}_t$-parameters in the previous section. Assuming a normal prior for the $\phi_{\mathbf{h}i}$- and an inverse Gamma prior for the $\sigma^2_{\boldsymbol{\eta}i}$-parameters, we sample the $\sigma^2_{\boldsymbol{\eta}i}$-parameters directly

[4] It is important to note that the stochastic volatility models in Eq. (34) do not correspond to the diagonal elements of the covariance matrix $\mathbf{H}_t$.

(given the conjugate prior), while we apply an MH step in order to sample the $\phi_{\mathbf{h}i}$-parameters from their posterior distribution, where we restrict these latter parameters to meet the stationarity conditions $|\phi_{\mathbf{h}i}| < 1$. The third conditional from above involves sampling the infinite mixture parameters, for which we introduce a complete sampling algorithm in Section 3.3.

In order to sample from $\pi(h_{i1}, \ldots, h_{iT}|\tilde{y}_{i1}, \ldots, \tilde{y}_{iT}, \boldsymbol{\vartheta}_i, \{l_{ij}\}_{j=1}^{\infty}, \{\omega_{ij}\}_{j=1}^{\infty})$, we follow Jensen and Maheu (2010) and apply our log volatility sampler to the transformation $y_{it}^* \equiv \tilde{y}_{it}\sqrt{\lambda_{i,t}}$ yielding the m simplified univariate models

$$y_{it}^* = \exp\{h_{it}/2\}\, u_{it}, \qquad (i = 1, \ldots, m) \tag{37}$$

$$h_{it} = \phi_{\mathbf{h}i} h_{it-1} + \eta_{it}, \tag{38}$$

so that our task reduces to sampling from $\pi(h_{i1}, \ldots, h_{iT}|\tilde{y}_{i1}, \ldots, \tilde{y}_{iT}, \boldsymbol{\vartheta}_i)$. We accomplish this by using a procedure from Jacquier et al. (2002), who propose a Bayesian approach, in which they construct a Markov chain for drawing directly from the joint posterior distribution of the latent volatility components. Specifically, let $\mathbf{h}_{-t}^{(i)} \equiv (h_{i0}, \ldots, h_{it-1}, h_{it+1}, \ldots, h_{iT})'$ and $\mathbf{y}_i^* \equiv (y_{i1}^*, \ldots, y_{iT}^*)'$, which are used to decompose $\pi(h_{i1}, \ldots, h_{iT}|\mathbf{y}_i^*, \boldsymbol{\vartheta}_i)$ into a set of conditionals of the form $\pi(h_{it}|\mathbf{h}_{-t}^{(i)}, \mathbf{y}_i^*, \boldsymbol{\vartheta}_i)$. The authors suggest a (hybrid) cyclic random walk Metropolis chain which uses a series of independent Metropolis acceptance/rejection chains, which do not directly sample from the univariate conditionals, but still ensure that the posterior is a stationary distribution.

Thus, in order to sample from the target distribution $\pi(h_{i1}, \ldots, h_{iT}|\mathbf{y}_i^*, \boldsymbol{\vartheta}_i)$, we follow the lines of argument in Jacquier et al. (2002) and sample from the auxiliary density $\pi(h_{it}|h_{it-1}, h_{it+1}, y_{it}^*, \boldsymbol{\vartheta}_i)$, which can be factorized for $t = 2, \ldots, T-1$ as follows:

$$\begin{aligned}
\pi(h_{it}|h_{it-1}, h_{it+1}, y_{it}^*, \boldsymbol{\vartheta}_i) &\propto \pi(y_{it}^*|h_{it})\pi(h_{it}|h_{it-1})\pi(h_{it+1}|h_{it}) \\
&\propto \frac{1}{\exp\{h_{it}/2\}} \exp\left\{-\frac{1}{2}\frac{(y_{it}^*)^2}{\exp\{h_{it}/2\}}\right\} \\
&\quad \times \exp\left\{\frac{-(h_{it} - \phi_{\mathbf{h}i} h_{it-1})^2 - (h_{it+1} - \phi_{\mathbf{h}i} h_{it})^2}{2\sigma_{\boldsymbol{\eta}i}^2}\right\}. \qquad (39)
\end{aligned}$$

The density (39) does not have a standard form and we apply a Metropolis Hastings algorithm for each of the latent volatility components $h_{i2}, \ldots, h_{iT-1}$. We sample the first and last latent volatility components from

$$\begin{aligned}\pi(h_{i1}|h_{i2}, y_{i1}^*, \boldsymbol{\vartheta}_i) &\propto \frac{1}{\exp\{h_{i1}/2\}} \exp\left\{-\frac{1}{2}\frac{(y_{i1}^*)^2}{\exp\{h_{i1}/2\}}\right\} \\ &\quad \times \exp\left\{\frac{-(h_{i2} - \phi_{\mathbf{h}i} h_{i1})^2}{2\sigma_{\boldsymbol{\eta}i}^2}\right\}, \end{aligned} \tag{40}$$

$$\begin{aligned}\pi(h_{iT}|h_{iT-1}, y_{iT}^*, \boldsymbol{\vartheta}_i) &\propto \frac{1}{\exp\{h_{iT}/2\}} \exp\left\{-\frac{1}{2}\frac{(y_{iT}^*)^2}{\exp\{h_{iT}/2\}}\right\} \\ &\quad \times \exp\left\{\frac{-(h_{iT} - \phi_{\mathbf{h}i} h_{iT-1})^2}{2\sigma_{\boldsymbol{\eta}i}^2}\right\}. \end{aligned} \tag{41}$$

As the proposal, used at each step of the random walk Metropolis Hastings algorithm, we use $N(0, \sigma_{\boldsymbol{\eta}i}^2)$.

3.3 Slice sampling the ϵ_t-DPM-elements

The slice sampler proposed by Walker (2007) and its more efficient version presented in Kalli et al. (2011) tackle the general issue of sampling the infinite number of DPM parameters. The idea behind the slice sampler is to introduce appropriate latent variables, with the objective of finding a finite set of DPM parameters, the sampling of which produces a valid Markov chain with a correct stationary distribution.

The first step of the slice-sampling procedure consists of introducing a latent variable ρ_{it} (with positive support), such that for $i = 1, \ldots, m$ the joint density of the innovation ϵ_{it} and the latent variable ρ_{it} is given by

$$\begin{aligned} f(\epsilon_{it}, \rho_{it}|\boldsymbol{\Theta}) &= \sum_{j=1}^{\infty} \mathbf{1}(\rho_{it} < \omega_{ij}) \cdot f_N(\epsilon_{it}|\, 0, l_{ij}^{-1}) \\ &= \sum_{j \in \mathcal{A}(\rho_{it})} f_N(\epsilon_{it}|\, 0, l_{ij}^{-1}), \end{aligned} \tag{42}$$

where (i) $\mathbf{1}(\cdot)$ is the indicator function, which is equal to 1 when its argument is true and 0 otherwise, and (ii) $\mathcal{A}(\rho_{it}) \equiv \{j : \omega_{ij} > \rho_{it}\}$, which becomes a finite set for any

given $\rho_{it} > 0$. We note that the conditional distribution of ϵ_{it} given ρ_{it} is a finite normal mixture with equal weights. Based on this result, the slice-sampling procedure then introduces a second latent variable ζ_{it} indicating the mixture component from which ϵ_{it} is observed to yield the joint density

$$f(\epsilon_{it}, \zeta_{it} = j, \rho_{it} | \, \boldsymbol{\Theta}) = f_N(\epsilon_{it} | \, 0, l_{ij}^{-1}) \, \mathbf{1}(j \in \mathcal{A}(\rho_{it})). \tag{43}$$

Specifically, after initializing the starting values $c_i^{(0)}, \zeta_{i1}^{(0)}, \ldots, \zeta_{iT}^{(0)}$, the slice sampler proposed by Kalli et al. (2011) and Walker (2007) proceeds as follows in iteration r of the MCMC algoritm ($r = 1, \ldots, R$):

1. Sampling c_i:
   We use the sampling strategy proposed in Escobar and West (1995) and start by sampling the auxiliary variable $\psi_i \sim \text{Beta}(c_i^{(r-1)} + 1, T)$ and then sample c_i from the Gamma mixture

   $$\pi_{\psi_i} \cdot f_\Gamma(c_i | a_0 + \zeta_i^*, b_0 - \log(\psi_i)) + (1 - \pi_{\psi_i}) \cdot f_\Gamma(c_i | a_0 + \zeta_i^* - 1, b_0 - \log(\psi_i)),$$

   where $f_\Gamma(\cdot | \alpha, \beta)$ denotes the density function of the Gamma(α, β) distribution, $\zeta_i^* = \max\left\{\zeta_{i1}^{(r-1)}, \ldots, \zeta_{iT}^{(r-1)}\right\}$ and $\pi_{\psi_i} = (a_0 + \zeta_i^* - 1)/(a_0 + \zeta_i^* - 1 + T(b_0 - \log(\psi_i)))$.

2. Sampling υ_{ij}:
   For $j = 1, 2, \ldots, \zeta_i^*$ we sample the υ_{ij} values from the conditional distribution

   $$\upsilon_{ij} | \, \zeta_{i1}^{(r-1)}, \ldots, \zeta_{iT}^{(r-1)} \sim \text{Beta}\left(n_{ij} + 1, T - n_{i\cdot} + c_i\right),$$

   where $n_{ij} = \sum_{t=1}^{T} \mathbf{1}(\zeta_{it}^{(r-1)} = j)$ is the number of observations belonging to the jth component of the ith variable, and $n_{i\cdot} = \sum_{k=1}^{j} n_{ik}$ is the cumulative sum of components in the groups. We compute the associated mixture weights according to the stick-breaking procedure, $\omega_{i1} = \upsilon_{i1}$, and $\omega_{ij} = (1 - \upsilon_{ij}) \ldots (1 - \upsilon_{ij-1}) \upsilon_j$ for $j = 2, \ldots, \zeta_i^*$.

3. Sampling ρ_{it}:
   We sample the latent variables ρ_{it} from the uniform distribution $U(0, \omega_{i\zeta_{it}^{(r-1)}})$

and set $\rho_i^* = \min\{\rho_{i1}, \ldots, \rho_{iT}\}$, which we use to truncate the sequence of mixture weights in the next step.

4. Updating the weights ω_{ij}:
   We determine the smallest integer j_i^* such that $\sum_{j=1}^{j_i^*} \omega_{ij} > (1 - \rho_i^*)$. For those ω_{ij} with $j > \zeta_i^*$, we draw υ_{ij} from the prior Beta$(c_i, 1)$ distribution and compute the associated weights ω_{ij} according to the stick-breaking procedure for $j = \zeta_i^* + 1, \ldots, j_i^*$. Thus, the latent variable ρ_{it} indicates how many weights need to be sampled.

5. Sampling the mixture parameters l_{ij}:
   The mixture parameters are sampled from the conditional posterior which, given the conjugate priors, has the following Gamma distribution:

$$
\begin{aligned}
l_{ij} &\sim \text{Gamma}(\bar{\nu}_{ij}/2, \bar{s}_{ij}/2), && (44)\\
\bar{\nu}_{ij} &= \nu_0 + n_{ij}, && (45)\\
\bar{s}_{ij} &= s_0 + \sum_{t=1}^{T} \epsilon_{it}^2 \cdot \mathbf{1}(\zeta_{it}^{(r-1)} = j). && (46)
\end{aligned}
$$

   We note that, according to Eq. (35), $\epsilon_{it} = \tilde{y}_{it} \exp\{-h_{it}/2\}$ is treated as observable at this stage of the algorithm. As in Step 4, if a new component has been formed, the mixture parameters are sampled from the prior.

6. Updating the indicator variables ζ_{it}:
   According to the weight truncation induced by the variable ρ_{it}, we update the indicator variables ζ_{it} by sampling from

$$
\Pr(\zeta_{it} = j | \{\epsilon_{it}\}_{t=1}^{T}, \{l_{ij}\}_{j=1}^{j_i^*}, \{\omega_{ij}\}_{j=1}^{j_i^*}, \{\rho_{it}\}_{t=1}^{T}) \propto f_N(\epsilon_{it}|0, l_{ij}^{-1}) \cdot \mathbf{1}(j \in \mathcal{A}(\rho_{it})).
$$

   The updated variables ζ_{it} indicate the component to which each observation belongs. Given ζ_{it}, we set $\lambda_{i,t} = l_{i\zeta_{it}}$.

4 Features of the Cholesky DPM-MSV model

4.1 Predictive density

A key issue in Bayesian nonparametric inference is the predictive density (Escobar and West, 1995). Denoting the sequence of all observations obtained through date T by $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \ldots, \mathbf{y}_T\}$, we write the one-step ahead predictive density as

$$f(\mathbf{y}_{T+1} | \mathbf{y}_{1:T}) = \int f(\mathbf{y}_{T+1} | \boldsymbol{\Theta}, \mathbf{y}_{1:T}) \pi(\boldsymbol{\Theta} | \mathbf{y}_{1:T}) \mathrm{d}\boldsymbol{\Theta}, \tag{47}$$

where (i) the density $f(\mathbf{y}_{T+1} | \boldsymbol{\Theta}, \mathbf{y}_{1:T})$ constitutes an infinite scale mixture, given the representation of the innovation term in Eq. (17), and (ii) the posterior $\pi(\boldsymbol{\Theta} | \mathbf{y}_{1:T})$ is defined on the infinitely dimensional parameter space $\boldsymbol{\Theta}$. Since the integral in Eq. (47) is analytically untractable, we approximate the predictive density via the MCMC output,

$$f(\mathbf{y}_{T+1} | \mathbf{y}_{1:T}) \approx \frac{1}{R} \sum_{r=1}^{R} f(\mathbf{y}_{T+1} | \boldsymbol{\Theta}^{(r)}, \mathbf{y}_{1:T}), \tag{48}$$

where R is the length of the Markov chain and $\boldsymbol{\Theta}^{(r)}$ denotes the parameter set in iteration r. We cope with the infinitely dimensional parameter space by introducing the latent variables according to Eq. (42) in each iteration r (which we denote by $\rho_{it}^{(r)}$) and thus for $i = 1, \ldots, m$ obtain the following (finite number of) DPM parameters in iteration r:

$$\left\{ \omega_{i1}^{(r)}, \omega_{i2}^{(r)}, \ldots, \omega_{ij_i^{*(r)}}^{(r)} \right\} \qquad \text{and} \qquad \left\{ l_{i1}^{(r)}, l_{i2}^{(r)}, \ldots, l_{ij_i^{*(r)}}^{(r)} \right\}.$$

Next, we implement the 3-step algorithm proposed by Jensen and Maheu (2013), in order to sample a single precision (mixture) parameter $l_i^{(r)}$ in iteration r for $i = 1, \ldots, m$:

1. We sample the random variable a_i from the uniform distribution $U(0, 1)$.
2. We compute the sum $\sum_{j=1}^{j_i^{*(r)}} \omega_{ij}^{(r)}$.
3. If $\sum_{j=1}^{j_i^{*(r)}} \omega_{ij}^{(r)} > a_i$, we find the index d_i such that

$$\sum_{j=1}^{d_i - 1} \omega_{ij}^{(r)} < a_i < \sum_{j=1}^{d_i} \omega_{ij}^{(r)}$$

and set the precision parameter $l_i^{(r)} = l_{id_i}^{(r)}$; else we draw $l_i^{(r)}$ from the prior distribution G_0 given in Eq. (15).

After having run the three steps for each $i = 1, \ldots, m$, we compose the predictive error term covariance matrix at iteration r as $(\mathbf{\Lambda}^{(r)})^{-1} \equiv \text{diag}(1/l_i^{(r)})$.

We now repeat the complete algorithm (i.e the 3 steps for each $i = 1, \ldots, m$) a number of times (say $B^{\max}$ times) and record at each iteration r, the $B^{\max}$ covariance matrices $(\mathbf{\Lambda}_1^{(r)})^{-1}, \ldots, (\mathbf{\Lambda}_{B^{\max}}^{(r)})^{-1}$. Denoting the density function of the m dimensional multivariate normal distribution by $f_{\mathbf{N}}(\cdot|\cdot,\cdot)$ and given sampled parameters, we approximate the one-step-ahead predictive density according to Eq. (48) as

$$f(\mathbf{y}_{T+1}|\mathbf{y}_{1:T}) \approx \frac{1}{R}\sum_{r=1}^{R} f^{(r)}(\mathbf{y}_{T+1}|\mathbf{y}_{1:T}) \tag{49}$$

with

$$f^{(r)}(\mathbf{y}_{T+1}|\mathbf{y}_{1:T}) = \frac{1}{B^{\max}}\sum_{k=1}^{B^{\max}} f_{\mathbf{N}}\left(\mathbf{y}_{T+1}|\,\mathbf{0}, (\mathbf{A}_{T+1}^{(r)})^{-1}\mathbf{\Sigma}_{T+1}^{(r)}(\mathbf{\Lambda}_k^{(r)})^{-1}\mathbf{\Sigma}_{T+1}^{(r)}[(\mathbf{A}_{T+1}^{(r)})^{-1}]'\right), \tag{50}$$

where, for the computation of $\mathbf{A}_{T+1}^{(r)}$ and $\mathbf{\Sigma}_{T+1}^{(r)}$, we draw each $\alpha_{iT+1}^{(r)}$ from $N(\mu_{\boldsymbol{\alpha}i}^{(r)} + \phi_{\boldsymbol{\alpha}i}^{(r)}\alpha_{iT}^{(r)}, \sigma_{\mathbf{e}}^{2(r)})$ for $i = 1, \ldots, p$, and each $h_{iT+1}^{(r)}$ from $N(\phi_{\mathbf{h}i}^{(r)}h_{iT}^{(r)}, \sigma_{\boldsymbol{\eta}}^{2(r)})$ for $i = 1, \ldots, m$. In our empirical application below, we choose $B^{max} = 3$.

4.2 Conditional moments

According to the hierarchical representation of our Cholesky DPM-MSV model from the Eqs. (10) to (17), the conditional mean of $\mathbf{y}_t$ is assumed to equal the zero vector, while the conditional covariance matrix is given by

$$\mathbf{H}_t^* = \text{Cov}(\mathbf{y}_t|\mathbf{\Theta}, \mathbf{y}_{1:t-1}) = \mathbf{A}_t^{-1}\mathbf{\Sigma}_t\text{Cov}(\boldsymbol{\epsilon}_t|\mathbf{\Omega})\mathbf{\Sigma}_t(\mathbf{A}_t^{-1})', \tag{51}$$

where

$$\text{Cov}(\boldsymbol{\epsilon}_t|\mathbf{\Omega}) = \text{diag}\left(\sum_{j=1}^{\infty}\omega_{ij}l_{ij}^{-1}\right).$$

Using our predictive density from the Eqs. (49) and (50), we may approximate conditional second-moment forecasts of the Cholesky DPM-MSV model by

$$\mathbb{E}(\mathbf{H}_{T+1}^{*}) \approx \frac{1}{R}\sum_{r=1}^{R}\mathbf{H}_{T+1}^{*(r)}, \tag{52}$$

where

$$\mathbf{H}_{T+1}^{*(r)} = (\mathbf{A}_{T+1}^{(r)})^{-1}\mathbf{\Sigma}_{T+1}^{(r)}\frac{1}{B^{\max}}\sum_{k=1}^{B^{\max}}(\mathbf{\Lambda}_{k}^{(r)})^{-1}\mathbf{\Sigma}_{T+1}^{(r)}[(\mathbf{A}_{T+1}^{(r)})^{-1}]'. \tag{53}$$

4.3 Ordering of variables

Owing to the lower triangular structure of the $\mathbf{A}_t$ matrix, the ordering of the variables in the vector $\mathbf{y}_t$ of the Cholesky DPM-MSV model is crucial (Primiceri, 2005). In the context of time-varying VAR models, Nakajima and Watanabe (2011) address the problem by analyzing the structure of the Japanese economy and monetary policy. When analyzing multiple financial time series data, it might sometimes appear problematic or arbitrary to use a specific ordering of variables *prima facie*. However, in our empirical application below, an obvious criterion for variable ordering is the chronological sequence, in which the various stock markets start their trading day.

Our chronological ordering is just one out of $5! = 120$ ordering permutations for the 5-dimensional stock-market data set. In order to assess the relevance of the ordering, we re-ran the estimation with the reverse ordering and compared the corresponding variance and covariance paths. Since these paths exhibit very similar patterns, we conclude that the variable ordering is not a critical factor in our empirical application.

4.4 Simulation

We illustrate the Cholesky DPM-MSV estimation framework by means of a simulation example. To this end, we simulated $T = 1000$ observations from a 5-dimensional model ($m = 5$) according to Eqs. (3) to (9) with parameter values $\phi_{\mathbf{h}i} = 0.95$, $\sigma_{\boldsymbol{\eta}i}^2 = 0.04$ for $i = 1, \ldots, 5$, and a (finite) mixture distribution for the error term given by

$$\boldsymbol{\epsilon}_t \sim \begin{cases} \mathbf{N}\left(\mathbf{0}, \mathrm{diag}(0.8, 0.9, 0.8, 0.9, 0.8)\right) & \text{with probability } 0.9 \\ \mathbf{N}\left(\mathbf{0}, \mathrm{diag}(2.8, 1.9, 2.8, 1.9, 2.8)\right) & \text{with probability } 0.1 \end{cases}.$$

We set the α_{ij}-processes ($i \neq j$) constantly equal to -0.5, implying positive correlations roughly between 0.5 and 0.9. We parametrized the prior distributions as $\phi_{\mathbf{h}i} \sim N(0.95, 25)\mathbf{1}(|\phi_{\mathbf{h}i}| < 1)$, $\sigma^2_{\boldsymbol{\eta}i} \sim \text{InverseGamma}(10/2, 0.5/2)$, $c_i \sim \text{Gamma}(4, 4)$, $G_0 \equiv \text{Gamma}(10/2, 10/2)$.[5] We estimated the model with our Cholesky DPM-MSV methodology from Section 3 with 50000 iterations (including a burn-in phase of 10000).

Table 1 about here

Figure 1 about here

Figure 2 about here

The upper block of Table 1 displays the posterior means of the AR parameters (along with 90% Bayesian intervals), which are close to the true values. The lower block of Table 1 compares the α_{ij}-processes (all set constantly equal to -0.5) with the posterior mean of the unconditional expectations given by $c_{\boldsymbol{\alpha}i}/(1 - \phi_{\boldsymbol{\alpha}i})$. Again, the estimation results are in close line with the true quantities. Figure 1 displays the posterior means of the 5 overall variance processes (red lines) in comparison with corresponding simulated paths (blue lines) and the 90% Bayesian intervals. Evidently, the estimated trajectories capture the simulated volatility dynamics satisfactorily. Figure 2 presents the analogous plots of the overall correlation processes, where we have thinned the number of draws from the posterior distribution by factor 50. In each panel of Figure 2, the simulated correlation path lies within the 90% Bayesian interval.

5 Empirical application

5.1 Data

In this section, we apply the Cholesky DPM-MSV model to stock-index data for the five most important international stock markets, with the objective of analyzing stock-market co-movements. In particular, our data set includes daily stock index values

[5] Ausín et al. (2014) provide a detailed discussion on the appropriate choice of prior distributions. We use the same prior distributions in our empirical application in Section 5.

between 17 February 2012 and 19 February 2016 (1046 observations for each time series) for (i) the US Dow Jones Industrial, (ii) the German DAX 30 Performance, (iii) the European EuroStoxx50 index, (iv) the Japanese Nikkei 225, and (v) the Chinese Shanghai Shenzen CSI 300. All data were collected from *Datastream* (daily closing prices).

Figure 3 about here

Figure 3 displays the five indices along with their daily returns (computed as the daily first differences in logs $\times 100$). The sampling period does not cover the global financial crisis, but includes two country-specific stock market turbulences, namely the European sovereign debt crisis in early 2012 and the Chinese stock market turmoil between June 2015 and February 2016. Both events are accompanied by phases of high return volatility, as is evident from the right panels in Figure 3.

Table 2 about here

Table 2 contains summary statistics and the sample correlation coefficients among the five return series. All return series exhibit negative skewness and excess kurtosis, indicating non-Gaussian behavior. Although all five sample means are close to zero, we use demeaned data in our estimation procedure. The sample correlation coefficients are all positive and lead us to expect particularly pronounced co-movements among the European and US markets. As argued in Section 4.3, the ordering of the 5 return series within the Cholesky DPM-MSV model could potentially matter. As the natural ordering, we choose the chronological sequence, in which the respective stock markets start their trading day, i.e. $y_{1t}, \ldots, y_{5t}$ are the return series for (1) the Nikkei, (2) the Shanghai Shenzen, (3) the EuroStoxx50, (4) the Dax, and (5) the Dow Jones.

Table 3 about here

5.2 Estimation results

According to Eqs. (5)–(9) and the exhibition in Section 3.3, the estimation of our five-dimensional Cholesky-DPM-MSV model involves the sampling of (i) 5 stochastic volatility processes ($\mathbf{h}_t$-processes), (ii) 10 $\boldsymbol{\alpha}_t$-processes, (iii) 40 AR-parameters (stemming from the $\mathbf{h}_t$- and $\boldsymbol{\alpha}_t$-processes), and (iv) 5 DPM sets $\{\omega_{ij}, l_{ij}\}_{j=1}^{\infty}$. We ran a total of 50000 + 50000 iterations, and deleted the first 50000 results as burn-in phase. As prior distributions, we chose

$$\begin{aligned}
c_{\boldsymbol{\alpha}i} &\sim N(0,1), \\
\phi_{\boldsymbol{\alpha}i} &\sim N(0.95,25)\mathbf{1}(|\phi_{\boldsymbol{\alpha}i}|<1), \\
\sigma^2_{\mathbf{e}i} &\sim \text{InverseGamma}(10/2, 0.5/2), \\
\phi_{\mathbf{h}i} &\sim N(0.95,25)\mathbf{1}(|\phi_{\mathbf{h}i}|<1), \\
\sigma^2_{\boldsymbol{\eta}i} &\sim \text{InverseGamma}(10/2, 0.5/2), \\
c_i &\sim \text{Gamma}(4,4),
\end{aligned}$$

and the base distribution G_0 as Gamma$(10/2, 10/2)$. Table 3 displays the posterior means and standard deviations of the 40 AR parameters.

Figure 4 about here

We assess the co-movements among the five markets via the pairwise in-sample time-varying correlation coefficients (denoted by $\text{Corr}_{\text{INDEX1, INDEX2};t}$), which we obtain from the overall time-varying covariance matrix $\mathbf{H}_t^*$ from Eq. (51) computed in each MCMC iteration and at every date t. Figure 4 displays the time-varying correlation coefficients for the 10 market pairs. In each panel, the solid line represents the correlation coefficients computed as an average of 333 posterior thinned draws (out of 50000), while the darkly and brightly shaded areas represent 50% and 90% Bayesian intervals, respectively.

Figure 5 about here

Figure 4 provides the following major findings: (i) The time-varying in-sample correlation coefficients appear surprisingly volatile. (ii) Except for $\text{Corr}_{\text{DJ, EU};t}$ (US/European markets), $\text{Corr}_{\text{DJ, DAX};t}$ (US/German markets) and $\text{Corr}_{\text{DAX, EU};t}$ (German/European markets), the time-varying correlation coefficients take on negative values strikingly often. (iii) The coefficients $\text{Corr}_{\text{EU, SHA};t}, \text{Corr}_{\text{DAX, SHA};t}$, $\text{Corr}_{\text{DJ, NIK};t}, \text{Corr}_{\text{DJ, SHA};t}$ appear to fluctuate around mean levels close to zero, indicating rather weak correlation among the corresponding markets. (iv) During the Chinese stock-market downturn between 2015 and 2016, the coefficients $\text{Corr}_{\text{SHA, NIK};t}$ take on substantially smaller values (close to zero) than during all other phases of the sampling period. (v) The most stable, positive correlation coefficients are found between the German and the European stock markets ($\text{Corr}_{\text{DAX, EU};t}$), the US and the European markets ($\text{Corr}_{\text{DJ, EU};t}$), and the US and the German markets ($\text{Corr}_{\text{DJ, DAX};t}$). As a robustness check, Figure 5 plots the sample correlations obtained from a rolling window of size 50 centered around t. Evidently, Figure 5 supports the findings from Figure 4.

Figure 6 about here

Figure 7 about here

Table 4 about here

Finally, we investigate the predictive ability of our Cholesky DPM-MSV model in terms of predictive density estimation. Figure 6 displays the nonparametric predictive densities of the elements of the covariance matrix $\mathbf{H}_t^*$, approximated according to Eqs. (52) and (53), while Figure 7 shows pairwise density contour plots. The covariances from the one-step-ahead prediction closely follow the patterns obtained from the in-sample estimation. For example, the contour plots for the European and the Chinsese markets (Panel EU, SHA), the German and the Chinese Markets (Panel DAX, SHA), the US and the Japanese markets (Panel DJ, NIK), and the US and the Chinese markets (Panel DJ, SHA) all reflect the lack of linear dependence, as mentioned in the above discussion on Figure 4. Table 4 summarizes the posterior information of the one-step-ahead predictive density. Our model predicts the highest variance for the

Japanese market (with the broadest 90% credibility interval), and the lowest variance for the US market.

Figure 8 about here

5.3 DP precision

The precision parameter c_i of the Dirichlet process controls the number of mixture components for each $i = 1, \ldots, 5$, where — under the specific model structure in Eqs. (10)–(16) — the limiting cases $c_i = 0$ and $c_i \to \infty$ correspond to the Gaussian and the Student-t distributions, respectively. Instead of c_i, we consider the one-to-one transformation $\tilde{c}_i \equiv \frac{c_i}{c_i+1}$ onto the interval $[0, 1)$. Along similar lines as in Jensen and Maheu (2013, 2014), we may use the Savage-Dickey density ratio to test for (i) normality ($\tilde{c}_i = 0$), and (ii) the Student-t distribution ($\tilde{c}_i \to 1$) each versus our general MSV-DPM model with $\tilde{c}_i \in (0, 1)$.[6] Figure 8 displays the posterior histograms of the $\tilde{c}_i$ after burn-in. We note that all five histograms exhibit zero-mass for both, $\tilde{c}_i = 0$ and $\tilde{c}_i \to 1$, thus yielding no in-sample indication in favor of the Gaussian or the Student-t distribution. All histograms disclose positive mass for $\tilde{c}_i$-values ranging between 0.1 and 0.9 and with modes around 0.4 and 0.5, suggesting distinct, stock-market specific numbers of mixture components.

5.4 Out-of-sample forecasting model comparison

5.4.1 Benchmark models

In order to analyze out-of-sample predictive power, we compare our Cholesky DPM-MSV model with two benchmark specifications. The first benchmark is the Cholesky Gaussian-MSV specification

$$\begin{aligned} \mathbf{y}_t | \, \mathbf{H}_t &\sim \mathbf{N}(\mathbf{0}, \mathbf{H}_t), \\ \mathbf{H}_t &= \mathbf{A}_t^{-1} \mathbf{\Sigma}_t \mathbf{\Sigma}_t (\mathbf{A}_t')^{-1}, \end{aligned}$$

where the latent processes are defined as in Eqs. (7)–(9). We estimate the model with the sampling algorithm proposed in Section 3 by setting the matrix $\mathbf{\Lambda}_t = \mathbf{I}$ (the

[6] Note that we imposed a uniform prior on $\tilde{c}_i$. For technical details, see Jensen and Maheu (2014).

identity matrix).

Our second benchmark model is the Cholesky Student-t-MSV specification

$$\begin{aligned} \mathbf{y}_t|\,\mathbf{H}_t &\sim \mathbf{St}(\mathbf{0},\mathbf{H}_t,\tilde{\boldsymbol{\nu}}), \\ \mathbf{H}_t &= \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Sigma}_t(\mathbf{A}_t')^{-1}, \end{aligned}$$

in which the conditional distribution of the return vector follows a multivariate Student-t distribution (denoted by $\mathbf{St}$) with mean vector $\mathbf{0}$, scale matrix $\mathbf{H}_t$ and $m \times 1$ degrees-of-freedom vector $\tilde{\boldsymbol{\nu}}$. In order to estimate this specifcation via our framework from Section 3 (without the slice sampler), we use the gamma-normal representation of the t-distribution (see, *inter alia*, Chib and Ramamurthy, 2014). To this end, we consider a gamma distributed latent variable q_{it} and the (independently distributed) standard normal variable u_{it} to write

$$\epsilon_{it} = q_{it}^{-1/2}u_{it},$$

and specify the following hierarchical prior

$$\begin{aligned} q_{it}|\tilde{\nu}_i &\sim \text{Gamma}\left(\tilde{\nu}_i/2,\tilde{\nu}_i/2\right), \\ \tilde{\nu}_i &\sim \pi, \end{aligned}$$

with π representing some prior distribution.

Conditional on q_{it}, the sampling steps can be performed in exactly the same way as described in Section 3.1. However, we rewrite the dynamic model from Eqs. (35) and (36) as

$$\begin{aligned} \tilde{y}_{it} &= \exp\left\{h_{it}/2\right\} q_{it}^{-1/2}u_{it}, \qquad (i=1,\ldots,m) \\ h_{it} &= \phi_{\mathbf{h}i}h_{it-1} + \eta_{it}, \end{aligned}$$

and the corresponding sampling steps as

1. $\pi(\boldsymbol{\vartheta}_i|h_{i1},\ldots,h_{iT})$,
2. $\pi(h_{i1},\ldots,h_{iT}|\tilde{y}_{i1},\ldots,\tilde{y}_{iT},\boldsymbol{\vartheta}_i,q_{i1},\ldots,q_{iT},\tilde{\nu}_i)$,

3. $\pi(\tilde{\nu}_i|\tilde{y}_{i1},\ldots,\tilde{y}_{iT},h_{i1},\ldots,h_{iT},q_{i1},\ldots,q_{iT})$,

4. $\pi(q_{i1},\ldots,q_{iT}|\tilde{y}_{i1},\ldots,\tilde{y}_{iT},h_{i1},\ldots,h_{iT},\tilde{\nu}_i)$.

Obviously, by defining $y^*_{it} \equiv \tilde{y}_{it}\sqrt{q_{it}}$, Steps 1 and 2 remain unchanged. Step 3 is a single Metropolis-Hastings step in order to sample from the posterior

$$\pi(\tilde{\nu}_i|\tilde{y}_{i1},\ldots,\tilde{y}_{iT},h_{i1},\ldots,h_{iT},q_{i1},\ldots,q_{iT}) \propto \pi(\tilde{\nu}_i) \times \prod_{t=1}^{T} \frac{(\tilde{\nu}_i/2)^{\tilde{\nu}_i/2}}{\Gamma(\tilde{\nu}_i/2)} q_{it}^{\tilde{\nu}_i/2-1} \exp\left\{-\frac{\tilde{\nu}_i q_{it}}{2}\right\}.$$

The posterior is defined for $\tilde{\nu}_i > 4$ and the proposal is a normal distribution truncated on $(4,\infty)$. Step 4 samples the latent variables q_{it} directly from the conditional

$$q_{it}|\tilde{y}_{it},h_{it},\tilde{\nu}_i \sim \text{Gamma}\left(\left[\tilde{\nu}_i+1\right]/2, \left[\tilde{\nu}_i + (\tilde{y}_{it}\exp\left\{-h_{it}/2\right\})^2\right]/2\right).$$

5.4.2 Predictive likelihoods

We use the cumulative log-predictive likelihoods (CPLs) to compare the out-of-sample 1-day-ahead predictive ability of our Cholesky DPM-MSV model with the two benchmark specifications (Cholesky Gaussian-MSV, Cholesky Student-t MSV). For our in-sample estimation, we use a daily expanding estimation window, which we initialize for 17 February 2012 and 19 February 2016 (1046 trading days, shortest in-sample estimation window). Our out-of-sample period ranges between 22 February 2016 and 08 July 2016, which amounts to 100 out-of-sample 1-day-ahead predictions.

Table 5 about here

Table 5 reports the out-of-sample CPL values for the three competing models. We note that the subtraction of two CPL values yields the Bayes factor, a widely used measure of relative predictive accuracy. Obviously, our Cholesky DPM-MSV specification substantially outperforms both benchmark models in terms of the log Bayes factors with values of 14.42 (Cholesky Gaussian-MSV) and 5.42 (Cholesky Student-t-MSV).

6 Conclusion

In this paper, we establish a Cholesky multivariate stochastic volatility model with a highly flexible nonparametric distribution for the innovation vector—based on the Dirichlet process mixture (DPM)—and implement a Bayesian semiparametric estimation procedure. A striking advantage of our modeling framework is that it allows us to estimate DPM-based volatility models of higher dimensions ($m > 3$), without imposing unnecessarily restrictive assumptions. More concretely, this is due to the Cholesky structure, under which the common assumption of uncorrelated DPM error terms does not entail a flexibility loss, insofar as our overall covariance matrix $\mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\boldsymbol{\Lambda}_t^{-1}\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})'$ contains DPM elements in its non-diagonal entries.

In the empirical section, we apply our estimation framework to five daily stock-index return series, with the aim of analyzing co-movements among international stock markets. As two major empirical results, we find (i) a reduction in the co-movement between the Chinese and the Japanese markets during the recent Chinese stock-market downturn, and (ii) distinctively stable, positive co-movements among the European (including the German) and the US stock markets. Our Cholesky DPM-MSV specification has appealing in-sample properties and, in an out-of-sample forecasting analysis, yields substantially improved forecast accuracy (in terms of Bayes factors) when compared with two benchmark models from the literature.

Two conceivable extensions of our modeling framework to be tackled in future research are worth mentioning. First, frequently observed volatility asymmetries could be modeled by integrating leverage effects into our Cholesky DPM-MSV framework. Second, our estimation framework could be applied to high-frequency data sets containing realized (co)variances along the lines of Shirota et al. (2016), who suggest estimating Cholesky realized stochastic volatility models.

References

Asai, M., McAleer, M., Yu, J., 2006. Multivariate stochastic volatility: A review. Econometric Reviews 25(2-3), 145-175.

Ausín, M.C., Galeano, P., Ghosh, P., 2014. A semiparametric Bayesian approach to

the analysis of financial time series with applications to value at risk estimation. European Journal of Operational Research 232(2), 350-358.

Bauwens, L., Laurent, S., Rombouts, J.V., 2006. Multivariate GARCH models: A survey. Journal of Applied Econometrics 21(1), 79-109.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307-327.

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81(3), 541-553.

Chib, S., Omori, Y., Asai, M., 2009. Multivariate stochastic volatility. In Handbook of Financial Time Series, pp. 365-400. Springer.

Chib, S., Ramamurthy, S., 2014. DSGE models with Student-t errors. Econometric Reviews 33, 152-171.

Clements, A.E., Hurn, A.S., Volkov, V.V., 2015. Volatility transmission in global financial markets. Journal of Empirical Finance 32, 3-18.

Cont, R., 2001. Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance 1, 223-236.

Delatola, E.-I., Griffin, J.E., 2011. Bayesian nonparametric modelling of the return distribution with stochastic volatility. Bayesian Analysis 6(4), 901-926.

Delatola, E.-I., Griffin, J.E., 2013. A Bayesian semiparametric model for volatility with a leverage effect. Computational Statistics & Data Analysis 60, 97-110.

Ehrmann, M., Fratzscher, M., Rigobon, R., 2011. Stocks, bonds, money markets and exchange rates: Measuring international financial transmission. Journal of Applied Econometrics 26, 948-974.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987-1007.

Escobar, M. D., West, M., 1995. Bayesian density estimation and inference using mixtures. Journal of the American Statistical Association 90(430), 577-588.

Ferguson, T. S., 1973. A Bayesian analysis of some nonparametric problems. The Annals of Statistics, 209-230.

Harvey, A., Ruiz, E., Shephard, N., 1994. Multivariate Stochastic Variance Models. Review of Economic Studies 61(2), 247-264.

Jacquier, E., Polson, N.G., Rossi, P.E., 2002. Bayesian analysis of stochastic volatility models. Journal of Business & Economic Statistics 20(1), 69-87.

Jensen, M.J., Maheu, J.M., 2010. Bayesian semiparametric stochastic volatility modeling. Journal of Econometrics 157(2), 306-316.

Jensen, M.J., Maheu, J.M., 2013. Bayesian semiparametric multivariate GARCH modeling. Journal of Econometrics 176(1), 3-17.

Jensen, M.J., Maheu, J.M., 2014. Estimating a semiparametric asymmetric stochastic volatility model with a Dirichlet process mixture. Journal of Econometrics 178, 523-538.

Kalli, M., Griffin, J.E., Walker, S.G., 2011. Slice sampling mixture models. Statistics and Computing 21(1), 93-105.

Kalli, M., Walker, S.G., Damien, P., 2013. Modeling the conditional distribution of daily stock index returns: An alternative Bayesian semiparametric model. Journal of Business & Economic Statistics 31(4), 371-383.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. Review of Economic Studies 65(3), 361-393.

Lopes, H., McCulloch, R., Tsay, R., 2012. Cholesky stochastic volatility models for high-dimensional time series. Discussion papers.

Maheu, J.M., Yang, Q., 2016. An infinite hidden Markov model for short-term interest rates. Journal of Empirical Finance 38, 202-220.

Nakajima, J., 2017. Bayesian analysis of multivariate stochastic volatility with skew return distribution. Econometric Reviews 36, 546-562.

Nakajima, J., Watanabe, T., 2011. Bayesian analysis of time-varying parameter vector autoregressive model with the ordering of variables for the Japanese economy and monetary policy. Global COE Hi-Stat Discussion Paper Series gd11-196, Institute of Economic Research, Hitotsubashi University.

Primiceri, G.E., 2005. Time varying structural vector autoregressions and monetary policy. The Review of Economic Studies 72(3), 821-852.

Sethuraman, J., 1994. A constructive definition of Dirichlet priors. Statistica Sinica 4, 639-650.

Shirota, S., Omori, Y., Lopes, H.F., Piao, H., 2016. Cholesky realized stochastic volatility model. Econometrics and Statistics (article in press). http://dx.doi.org/10.

1016/j.ecosta.2016.08.003.

Taylor, S.J., 1982. Financial returns modelled by the product of two stochastic processes - a study of the daily sugar prices 1961-75. In Anderson, O.D., ed. Time Series Analysis: Theory and Practice. Vol 1. North-Holland, pp. 203-226.

Taylor, S.J., 1986. Modelling Financial Time Series. Chichester: John Wiley.

Virbickaitė, A., Auín, M.C., Galeano, P., 2016. A Bayesian non-paprametric approach to asymmetric dynamic conditional correlation model with application to portfolio selection. Computational Statistics & Data Analysis 100, 814-829.

Walker, S.G., 2007. Sampling the Dirichlet mixture model with slices. Communications in Statistics - Simulation and Computation 36(1), 45-54.

Tables and Figures

Table 1: Parameter values, posterior means, 90% credibility intervals (CIs)

	True parameter	Posterior mean	90% CI
$\phi_{\mathbf{h}1}$	0.95	0.9439	(0.9102, 0.9691)
$\phi_{\mathbf{h}2}$	0.95	0.9231	(0.8700, 0.9595)
$\phi_{\mathbf{h}3}$	0.95	0.9040	(0.8277, 0.9569)
$\phi_{\mathbf{h}4}$	0.95	0.9251	(0.8894, 0.9571)
$\phi_{\mathbf{h}5}$	0.95	0.9698	(0.9447, 0.9873)
$\sigma^2_{\boldsymbol{\eta}1}$	0.04	0.0525	(0.0323, 0.0821)
$\sigma^2_{\boldsymbol{\eta}2}$	0.04	0.0810	(0.0433, 0.1367)
$\sigma^2_{\boldsymbol{\eta}3}$	0.04	0.0791	(0.0281, 0.1499)
$\sigma^2_{\boldsymbol{\eta}4}$	0.04	0.0680	(0.0436, 0.1002)
$\sigma^2_{\boldsymbol{\eta}5}$	0.04	0.0688	(0.0426, 0.1101)
	$\mathbb{E}(\alpha_{ij,t}) = \frac{c_{\boldsymbol{\alpha}i}}{(1-\phi_{\boldsymbol{\alpha}i})}$		
α_{12}	-0.5	-0.5014	$(-0.5595, -0.4447)$
α_{13}	-0.5	-0.5244	$(-0.5910, -0.4560)$
α_{23}	-0.5	-0.5275	$(-0.5840, -0.4653)$
α_{14}	-0.5	-0.4900	$(-0.5599, -0.4255)$
α_{24}	-0.5	-0.4424	$(-0.5108, -0.3826)$
α_{34}	-0.5	-0.4921	$(-0.5402, -0.4426)$
α_{15}	-0.5	-0.5425	$(-0.6092, -0.4748)$
α_{25}	-0.5	-0.4468	$(-0.5252, -0.3763)$
α_{35}	-0.5	-0.5453	$(-0.6056, -0.4878)$
α_{45}	-0.5	-0.4617	$(-0.5320, -0.4013)$

Notes: Simulated model according to Eqs. (3)–(9), with $m = 5, T = 1000$ and a mixture of two normal distributions for the error term, as specified in Section 4.4.

Table 2: Descriptive statistics

	NIK	SHA	EU	DAX	DJ
Mean	0.0509	0.0177	0.0125	0.0302	0.0226
Median	0.0086	0.0000	0.0111	0.0614	0.0066
Variance	1.9457	2.7957	1.5524	1.4216	0.6229
Skewness	−0.2386	−0.8491	−0.1151	−0.2329	−0.1961
Kurtosis	6.3634	8.1768	4.4545	4.2553	4.7188
Sample correlation:					
NIK	1.0000				
SHA	0.2160	1.0000			
EU	0.2158	0.1349	1.0000		
DAX	0.2194	0.1390	0.9526	1.0000	
DJ	0.1224	0.1418	0.5929	0.5743	1.0000

Notes: Daily returns between 17 February 2012 and 19 February 2016 (1046 observations) in percent. The indices are abbreviated as NIK (Nikkei 225), SHA (Shanghai Shenzen CSI 300), EU (EuroStoxx), DAX (DAX 30 Performance), DJ (Dow Jones Industrial).

Table 3: Posterior means and standard deviations (in parantheses)

i	$c_{\boldsymbol{\alpha}i}$	$\phi_{\boldsymbol{\alpha}i}$	$\sigma^2_{\mathbf{e}i}$	$\phi_{\mathbf{h}i}$	$\sigma^2_{\boldsymbol{\eta}i}$
1	−0.1455	0.0649	0.0474	0.9610	0.0510
	(0.0582)	(0.3532)	(0.0144)	(0.0158)	(0.0176)
2	−0.1614	0.0601	0.0657	0.9799	0.0364
	(0.0548)	(0.2965)	(0.0180)	(0.0087)	(0.0103)
3	−0.0662	−0.1444	0.0945	0.9323	0.0724
	(0.0338)	(0.1953)	(0.0234)	(0.0372)	(0.0499)
4	−0.0159	−0.0370	0.0261	0.9980	0.0330
	(0.0120)	(0.1403)	(0.0054)	(0.0013)	(0.0135)
5	0.0025	−0.0509	0.0310	0.9938	0.0674
	(0.0124)	(0.1679)	(0.0091)	(0.0039)	(0.0245)
6	−0.4111	0.5478	0.0254		
	(0.1070)	(0.1182)	(0.0069)		
7	0.0008	0.4161	0.0466		
	(0.0127)	(0.1831)	(0.0121)		
8	−0.0018	−0.4898	0.0277		
	(0.0218)	(0.1796)	(0.0107)		
9	−0.1832	0.2271	0.0498		
	(0.0986)	(0.3241)	(0.0159)		
10	−0.1745	−0.1649	0.0488		
	(0.0922)	(0.2732)	(0.0136)		

Notes: Estimates of parameters from Eqs. (7)–(8). $c_{\boldsymbol{\alpha}i}$ computed according to Eq. (26).

Table 4: Posterior summary of the elements of the one-step-ahead covariance matrix

$\mathbf{H}^{*}_{T+1}$	Mean	Median	90% CI
$\mathbf{H}^{*\text{NIK}}_{T+1}$	7.2337	6.1811	$(2.9575, 12.7184)$
$\mathbf{H}^{*\text{SHA}}_{T+1}$	3.2385	2.6356	$(1.0457, 6.1222)$
$\mathbf{H}^{*\text{EU}}_{T+1}$	3.7878	3.1497	$(1.4192, 6.8541)$
$\mathbf{H}^{*\text{DAX}}_{T+1}$	3.6246	2.9022	$(1.2207, 6.7965)$
$\mathbf{H}^{*\text{DJ}}_{T+1}$	1.6232	1.1733	$(0.3730, 3.3584)$
$\mathbf{H}^{*\text{SHA, NIK}}_{T+1}$	0.9987	0.7307	$(-0.8603, 3.2080)$
$\mathbf{H}^{*\text{EU, NIK}}_{T+1}$	1.5025	1.1124	$(-0.8145, 4.2972)$
$\mathbf{H}^{*\text{EU, SHA}}_{T+1}$	0.4096	0.2605	$(-0.9820, 1.9723)$
$\mathbf{H}^{*\text{DAX, NIK}}_{T+1}$	1.5457	1.1560	$(-0.8077, 4.3994)$
$\mathbf{H}^{*\text{DAX, SHA}}_{T+1}$	0.4052	0.2579	$(-1.0745, 2.0624)$
$\mathbf{H}^{*\text{DAX, EU}}_{T+1}$	3.5458	2.9134	$(1.2484, 6.5298)$
$\mathbf{H}^{*\text{DJ, NIK}}_{T+1}$	0.1496	0.0923	$(-1.8593, 2.2114)$
$\mathbf{H}^{*\text{DJ, SHA}}_{T+1}$	0.0052	-0.0103	$(-1.0902, 1.1261)$
$\mathbf{H}^{*\text{DJ, EU}}_{T+1}$	1.2784	0.9443	$(-0.1758, 3.1538)$
$\mathbf{H}^{*\text{DJ, DAX}}_{T+1}$	1.2232	0.8752	$(-0.1981, 3.1004)$

Notes: Forecast date $T+1$: 22 February 2016. The indices are abbreviated as NIK (Nikkei 225), SHA (Shanghai Shenzen CSI 300), EU (EuroStoxx), DAX (DAX 30 Performance), DJ (Dow Jones Industrial).

Table 5: Cumulative log predictive likelihoods (CPL)

Model	CPL
Cholesky Gaussian-MSV	−699.9647
Cholesky Student-t-MSV	−690.6329
Cholesky DPM-MSV	−685.5444

Notes: The two benchmark models (Cholesky Gaussian MSV, Cholesky Student-t MSV) estimated as described in Section 5.4.1. Out-of-sample period: 22 February 2016 – 08 July 2016 (100 observations).

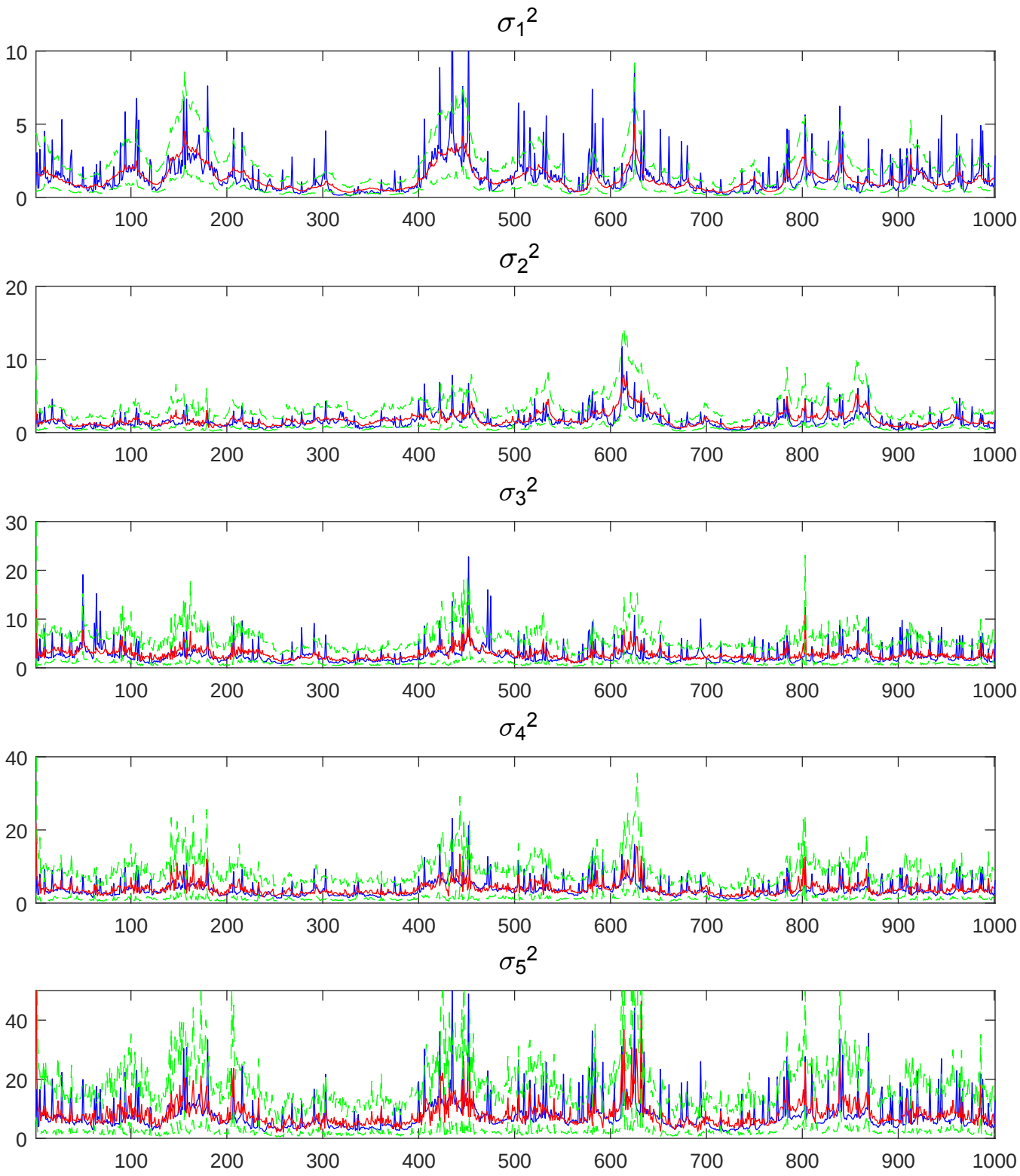

Figure 1. Simulated variance processes (blue lines), posterior means (red lines), and 90% Bayesian intervals (green lines).

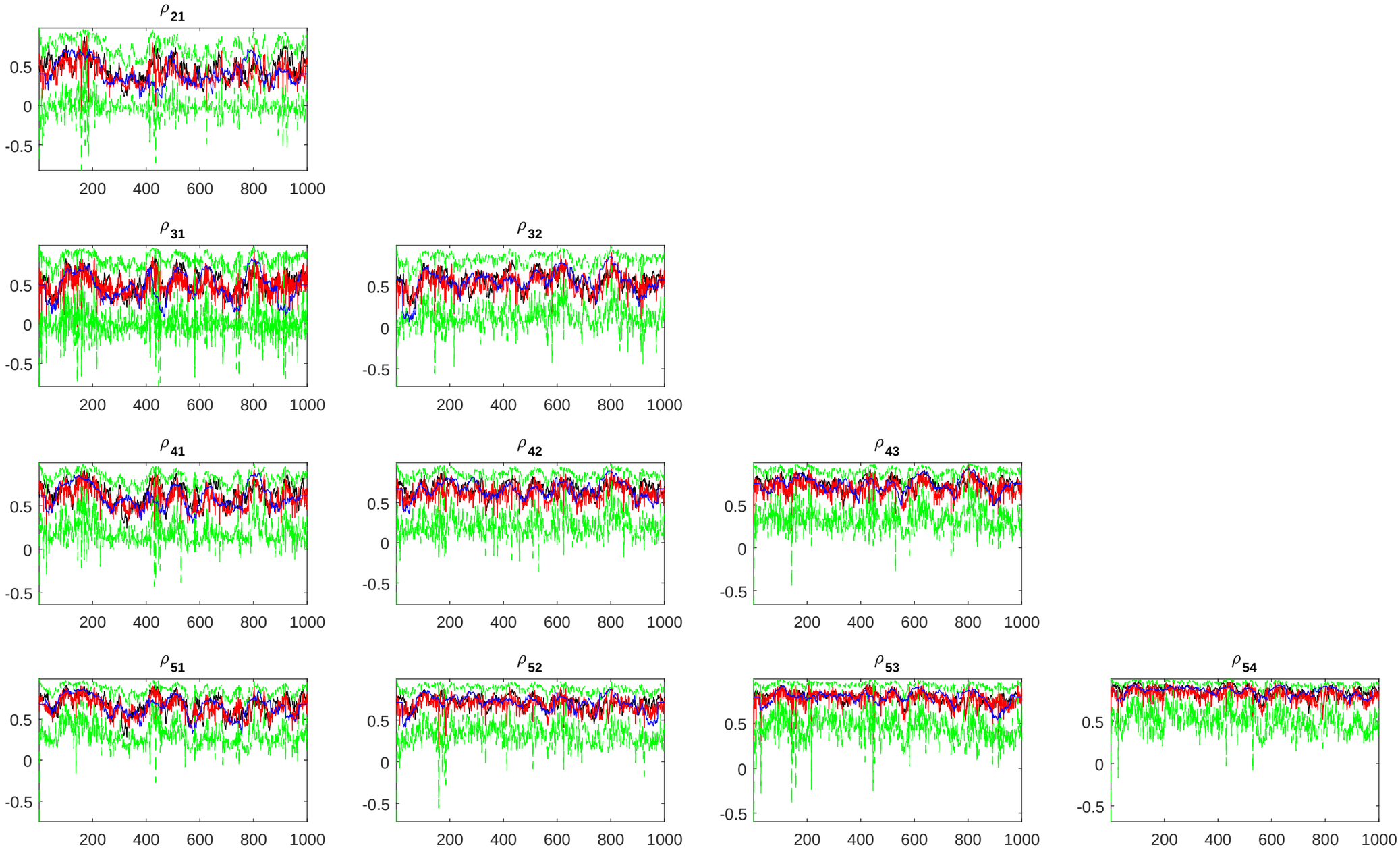

Figure 2. Simulated correlation processes (blue lines), posterior means (red lines), and 90% Bayesian intervals (green lines).

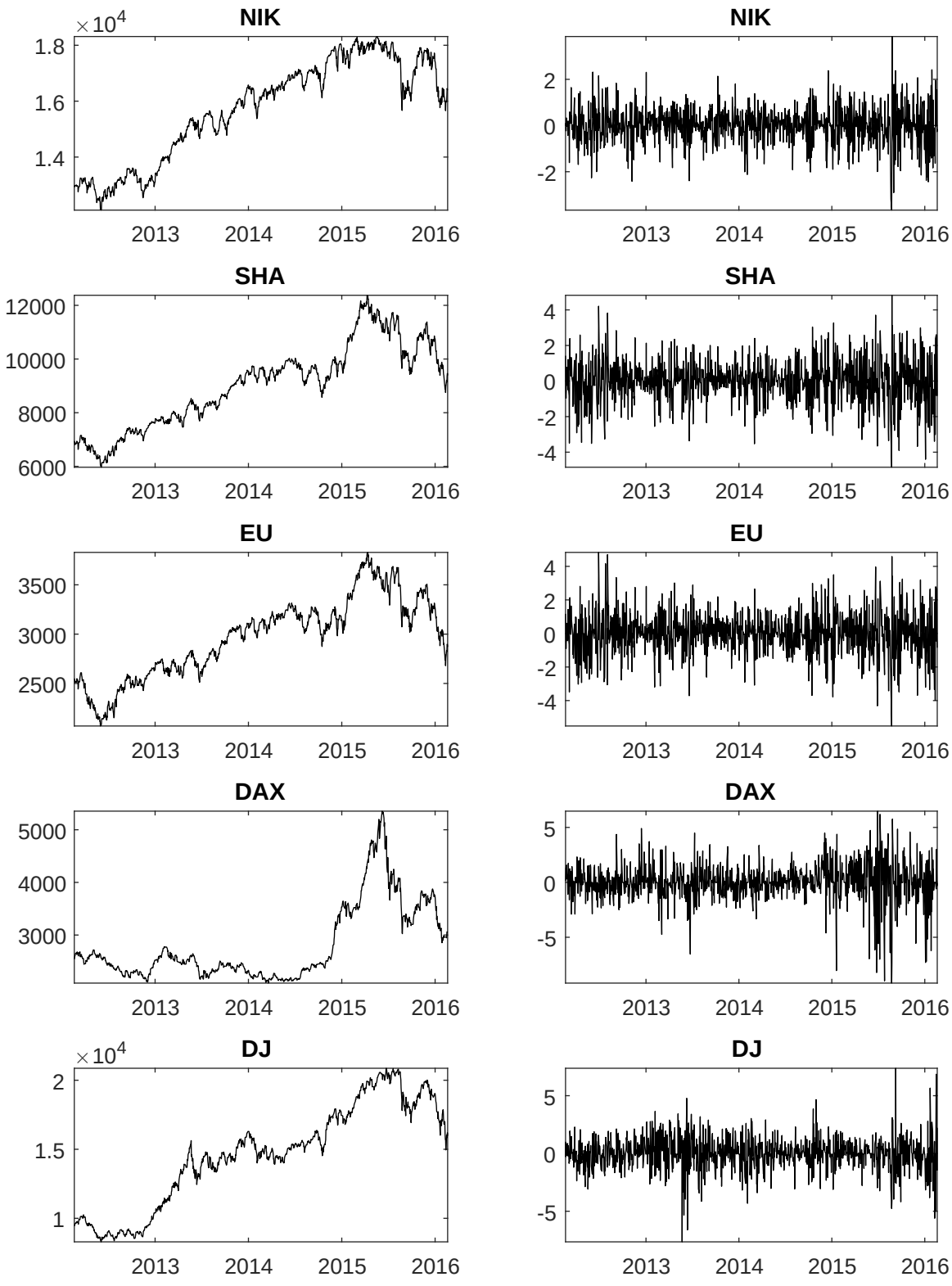

Figure 3. Index values and daily returns.

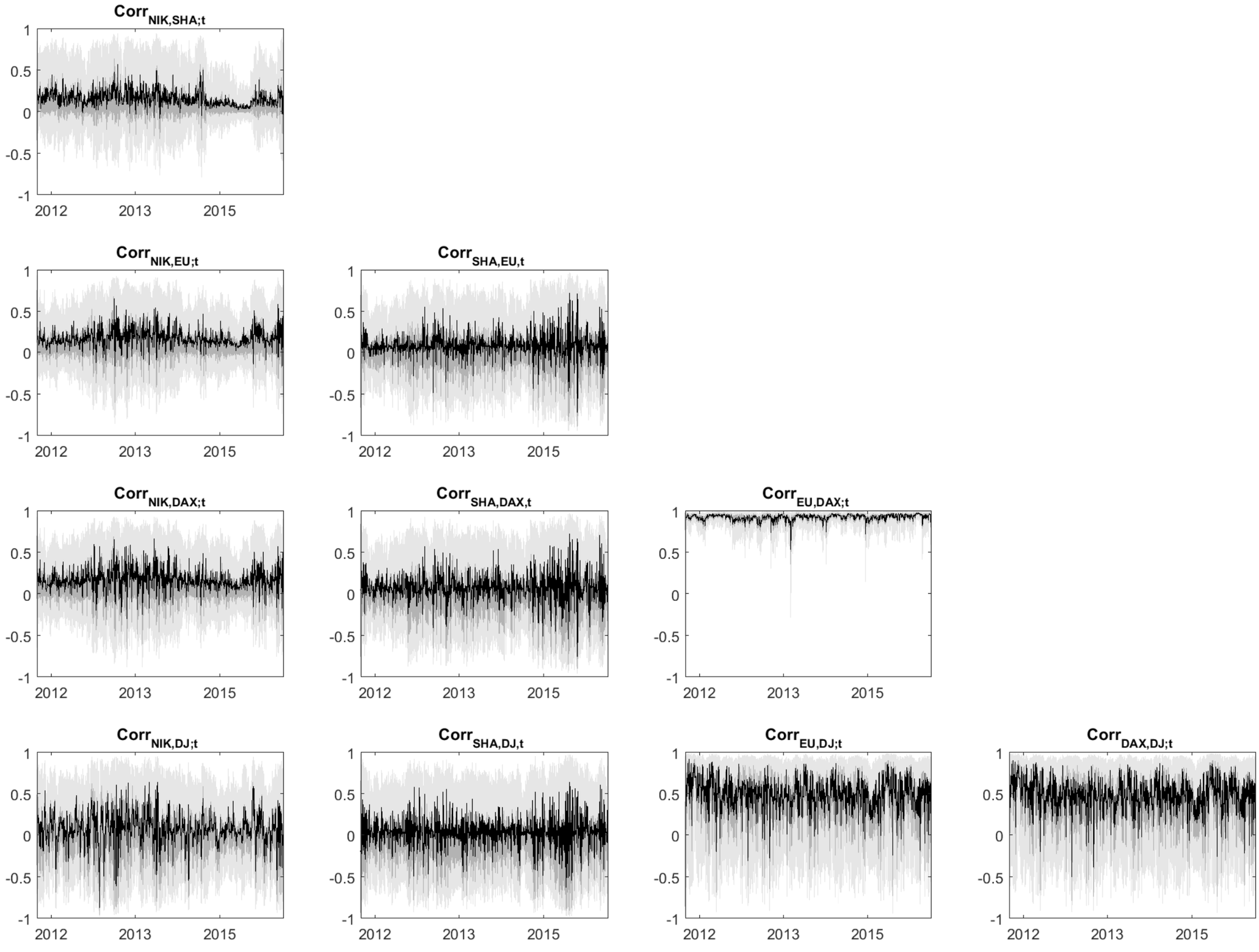

Figure 4. In-sample correlations: posterior means plus 50% and 90% Bayesian intervals.

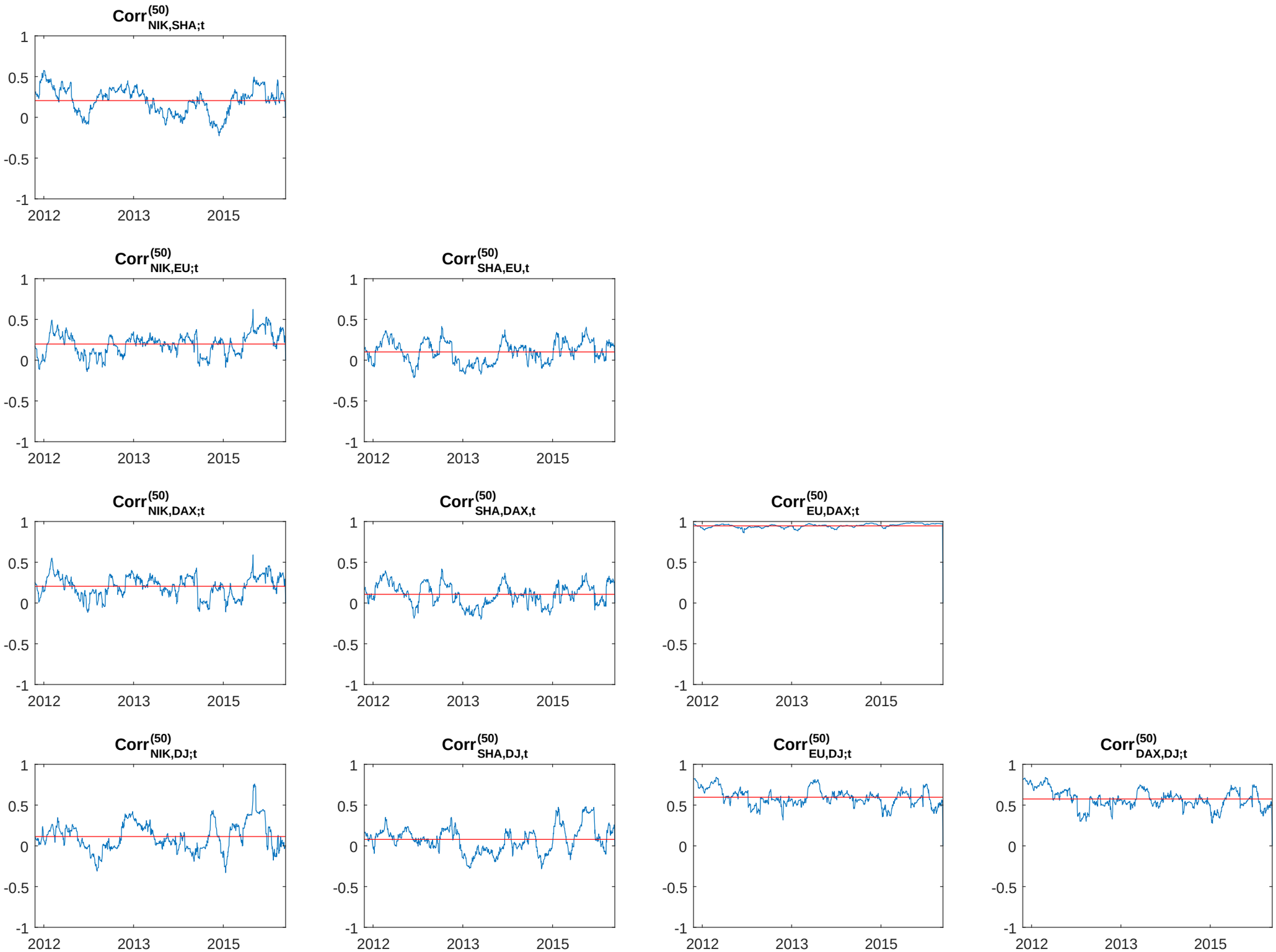

Figure 5. Sample correlations obtained from a rolling window of size 50 centered around the actual observation with the sample-correlation (horizontal line).

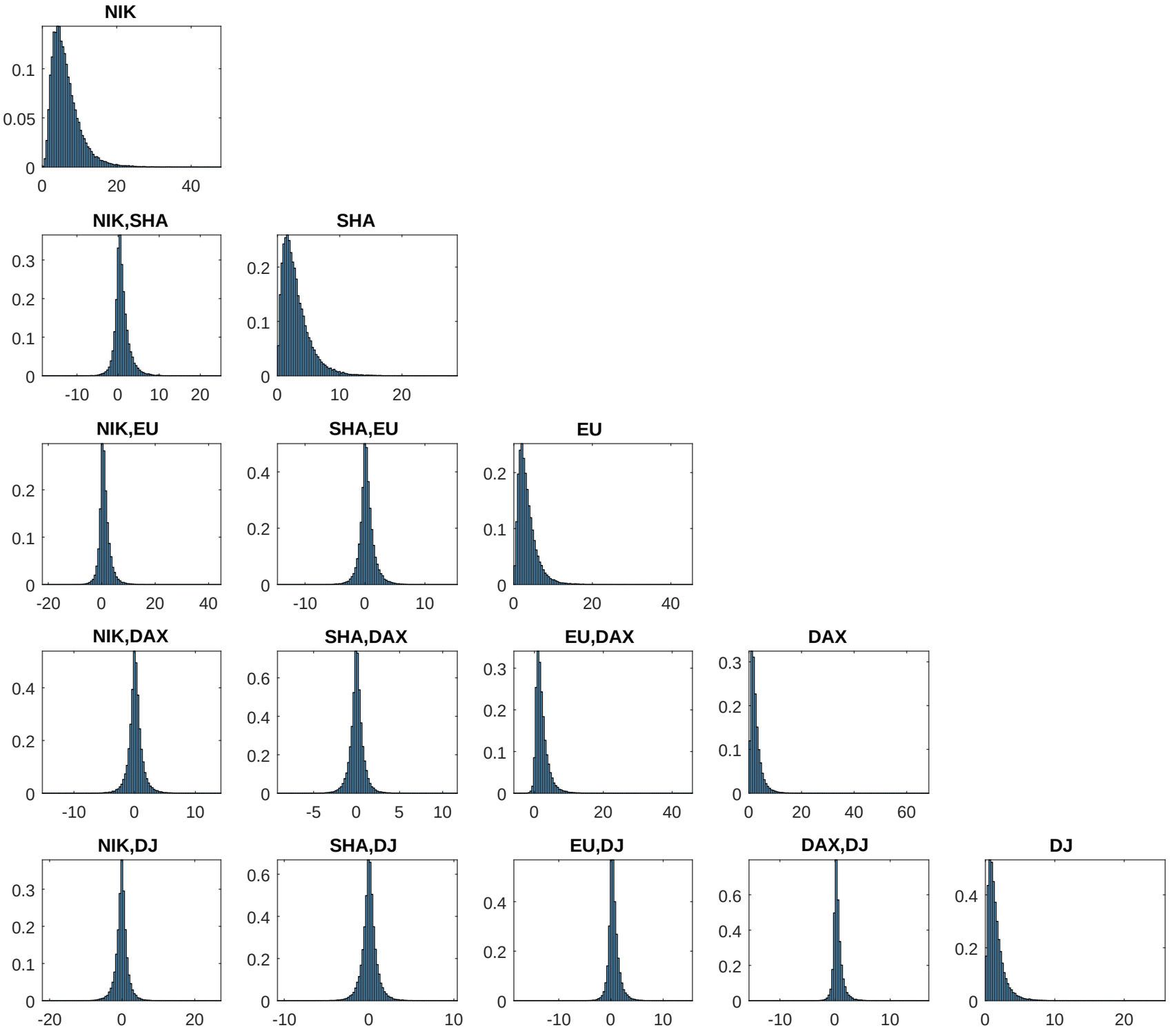

Figure 6. One-step-ahead density forecasts (of the elements of $\mathbf{H}^*_{T+1}$).

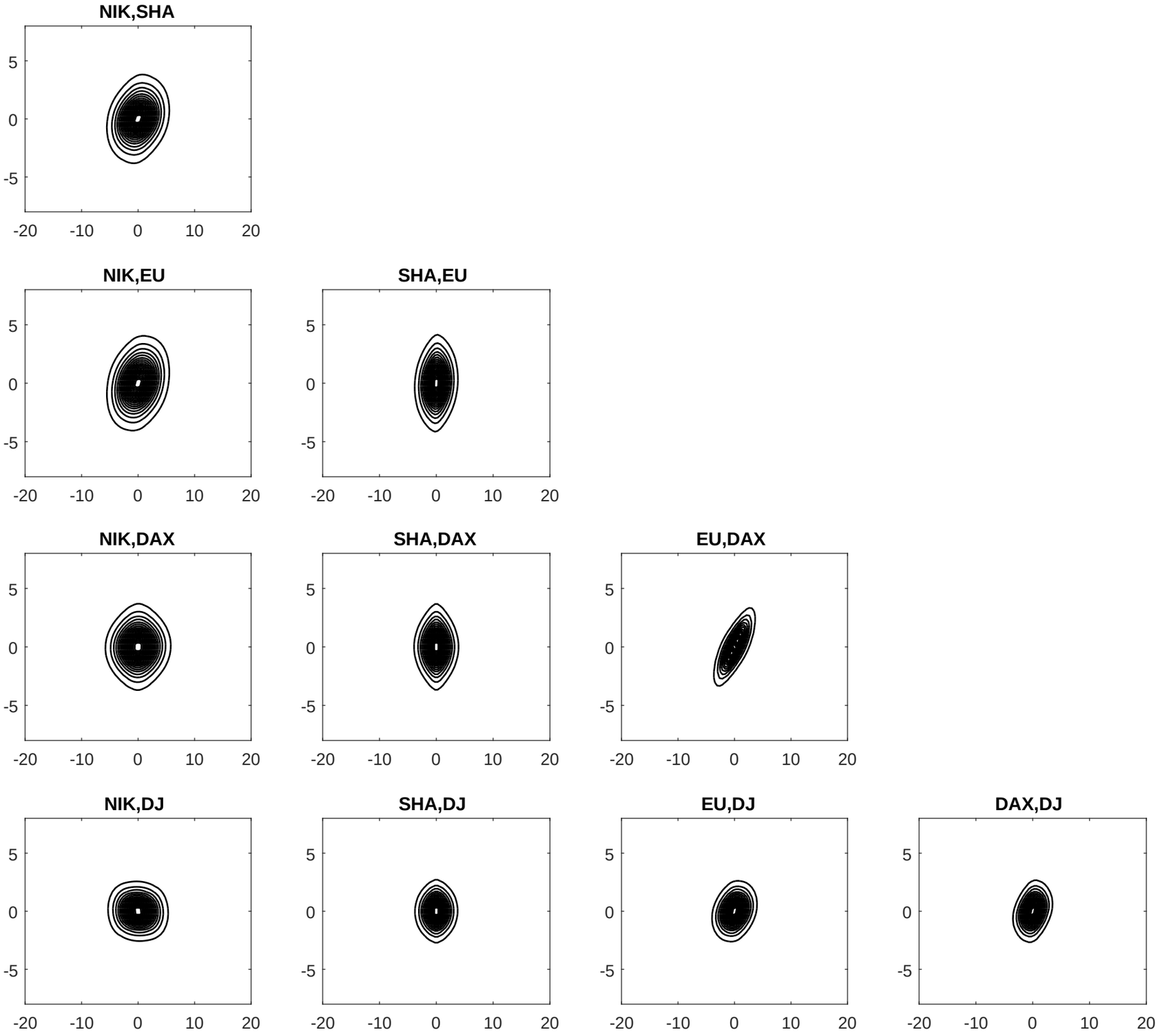

Figure 7. Contour plots of pairwise one-step-ahead density forecasts.

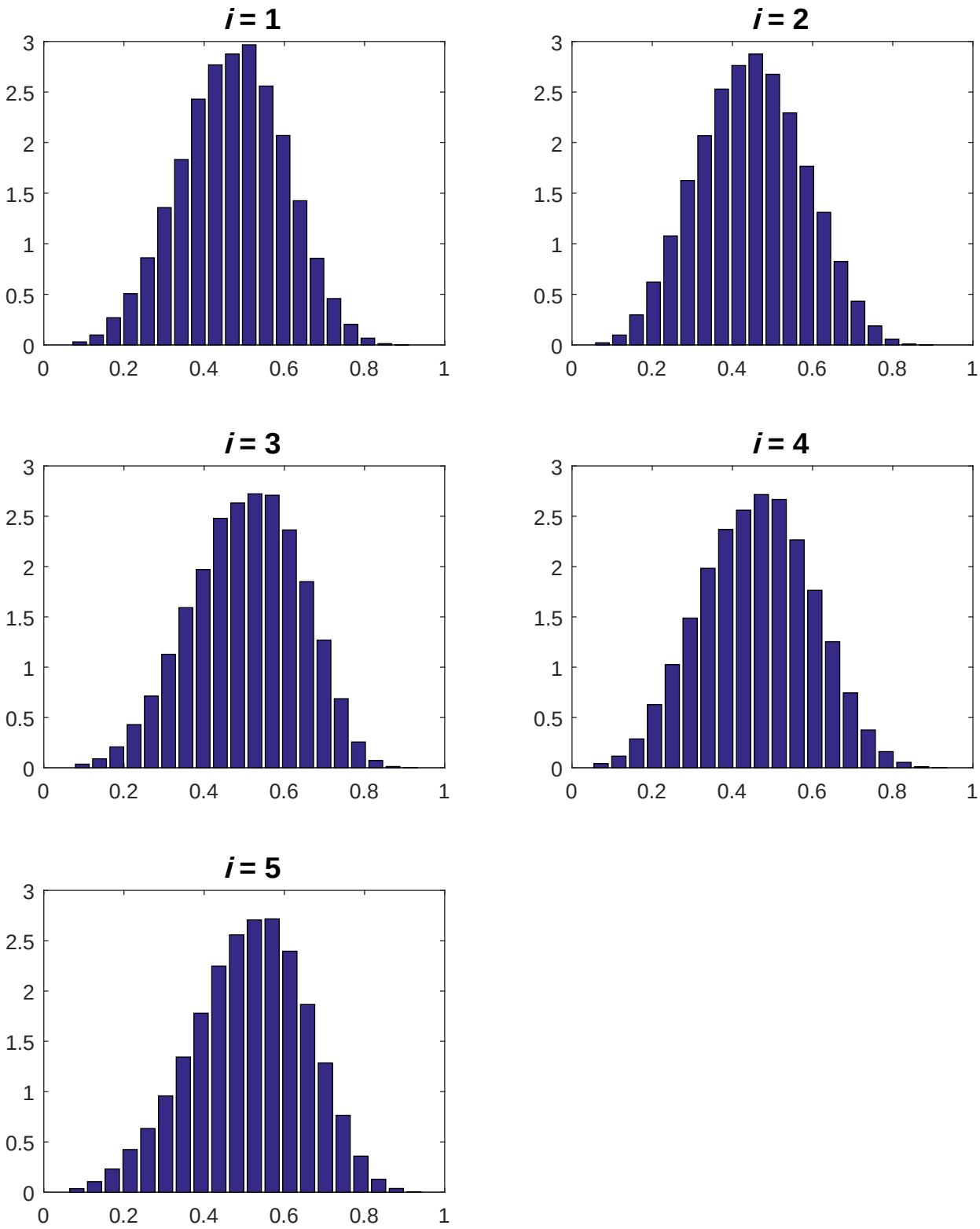

Figure 8. Transformed DP precision.