



Bayesian semiparametric multivariate stochastic volatility with application

Martina Danielova Zaharieva, Mark Trede & Bernd Wilfling

To cite this article: Martina Danielova Zaharieva, Mark Trede & Bernd Wilfling (2020): Bayesian semiparametric multivariate stochastic volatility with application, *Econometric Reviews*, DOI: [10.1080/07474938.2020.1761152](https://doi.org/10.1080/07474938.2020.1761152)

To link to this article: <https://doi.org/10.1080/07474938.2020.1761152>



© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 19 May 2020.



Submit your article to this journal [↗](#)



View related articles [↗](#)



View Crossmark data [↗](#)

Bayesian semiparametric multivariate stochastic volatility with application

Martina Danielova Zaharieva^a, Mark Trede^b, and Bernd Wilfling^b

^aDepartment of Econometrics, Erasmus University Rotterdam, Rotterdam, The Netherlands; ^bDepartment of Economics (CQE), Westfälische Wilhelms-Universität, Münster, Germany

ABSTRACT

In this article, 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 density forecast accuracy when compared with the prevalent benchmark models.

KEYWORDS

Bayesian nonparametrics; Dirichlet process mixture; Markov chain Monte Carlo; 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* Clements et al., 2015; Ehrmann et al., 2011). The two by far most popular volatility model classes discussed in the literature are the generalized autoregressive conditional heteroscedasticity (GARCH-type) models (Bollerslev 1986; Engle 1982) 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 the SV framework is that it models 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 (Asai et al., 2006; Harvey et al., 1994; Kim et al., 1998).

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

CONTACT Bernd Wilfling  bernd.wilfling@wiwi.uni-muenster.de  Westfälische Wilhelms-Universität, Department of Economics (CQE), Am Stadtgraben 9, 48143 Münster, Germany.

© 2020 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

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, 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). All of these studies use infinite mixtures of normals, some authors analyze scale, and others location-scale mixtures. In their empirical applications to FOREX, stock-price and stock-index data, the articles unambiguously document the outperformance of the DPM approach over conventional parametric benchmark models in terms of multi-step ahead predictive power. Additionally, Delatola and Griffin (2013) and Jensen and Maheu (2014) model the leverage effect by means of a nonparametric prior, thus providing new insight into how the effect is linked to current market conditions.

In this article, 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 semiparametric 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 the wake of the European sovereign debt crisis and the Chinese stock-market bubble. In an out-of-sample forecasting comparison with two conventional models of the error-term distribution (multivariate normal and Student- t) and an asymmetric extension of our model, we find that our symmetric DPM model yields more accurate forecasts. While the accuracy gain is modest in comparison to the asymmetric case, there are substantial gains over the multivariate normal and the Student- t specifications.

The article is organized as follows. Section 2 reviews (i) the multivariate SV model based on Cholesky decomposition, and (ii) the Dirichlet process mixture. Section 3 presents essential probabilistic features of our econometric framework and provides a simulation example. Section 4 contains the empirical application to daily returns from the five largest international stock markets. Section 5 concludes. The appendix gives a concise description of the Bayesian estimation approach.

2. Model development

2.1. Cholesky (multivariate) SV

In order to introduce Cholesky SV modeling, we follow the approach of Primiceri (2005) and Nakajima (2017) and consider the $m \times 1$ vector $\mathbf{y}_t = (y_{1t}, \dots, y_{mt})'$ 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 $\text{Cov}(\mathbf{y}_t) = \mathbf{H}_t$, 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

$$\mathbf{A}_t \mathbf{H}_t \mathbf{A}_t' = \boldsymbol{\Sigma}_t \boldsymbol{\Sigma}_t', \quad (1)$$

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:

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

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

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

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

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 SV 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 SV 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 features of financial return data (like leverage effects and skewness). In the next section, we define a new class of Cholesky SV 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 SV 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 et al., 2012). Specifically, we collect the parameters from the matrix \mathbf{A}_t row-by-row 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:

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

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

We then specify the dynamics of the Cholesky parameters as the (stationary) AR(1) processes

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu}_\alpha + \boldsymbol{\Phi}_\alpha (\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_\alpha) + \mathbf{e}_t, \quad (7)$$

$$\mathbf{h}_t = \boldsymbol{\Phi}_h \mathbf{h}_{t-1} + \boldsymbol{\eta}_t, \quad (8)$$

$$\begin{pmatrix} \mathbf{e}_t \\ \boldsymbol{\eta}_t \end{pmatrix} \sim \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_e & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_\eta \end{bmatrix}\right), \quad (9)$$

where we assume (i) that the matrices $\boldsymbol{\Phi}_\alpha, \boldsymbol{\Phi}_h, \boldsymbol{\Sigma}_e, \boldsymbol{\Sigma}_\eta$ are all diagonal, and (ii) that the $p = m(m-1)/2$ diagonal entries $\phi_{\alpha_1}, \dots, \phi_{\alpha_p}$ of $\boldsymbol{\Phi}_\alpha$ and the m diagonal entries $\phi_{h_1}, \dots, \phi_{h_m}$ of $\boldsymbol{\Phi}_h$ are all less than 1 in absolute value (stationarity conditions).¹

By construction, our Cholesky-type covariance matrix imposes top-down dependency among the elements of \mathbf{y}_t , implying that the variable ordering affects inference. Chan et al. (2018a, 2018b) refer to this type of modeling as noninvariant specifications, provide an in-depth literature overview of the issue, and establish invariant inference in the context of static factor models and volatility shock decomposition. By contrast, we do not consider invariant inferential techniques, but discuss implications of the specifically chosen variable ordering in our empirical application in Sections 3 and 4.

¹We 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 nonzero intercept; see Jensen and Maheu (2010).

2.2. Bayesian semiparametric Cholesky SV

It remains to specify the distribution of the innovation vector ϵ_t from Eq. (3), which we model as a nonparametric Dirichlet process mixture. The DPM represents an infinite mixture model and constitutes an extremely flexible extension of finite mixture models in financial return modeling (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 $DP(c, G_0)$, 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 SV 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 ϵ_t follow the nonparametric DPM as given in Eq. (17) below, has the following hierarchical representation:

$$\mathbf{y}_t | \mathbf{\Lambda}_t, \mathbf{A}_t, \mathbf{\Sigma}_t \sim \mathbf{N}(\mathbf{0}, \mathbf{A}_t^{-1} \mathbf{\Sigma}_t \mathbf{\Lambda}_t^{-1} \mathbf{\Sigma}_t (\mathbf{A}_t^{-1})'), \quad (10)$$

$$\mathbf{H}_t = \mathbf{A}_t^{-1} \mathbf{\Sigma}_t \mathbf{\Lambda}_t^{-1} (\mathbf{A}_t')^{-1}, \quad (11)$$

$$\mathbf{\Lambda}_t = \text{diag}(\lambda_{1,t}, \dots, \lambda_{m,t}), \quad (12)$$

$$\lambda_{i,t} \stackrel{\text{i.i.d.}}{\sim} G_i, \quad (i = 1, \dots, m) \quad (13)$$

$$G_i | G_0, c_i \sim DP(c_i, G_0), \quad (14)$$

$$G_0 \stackrel{\text{d}}{=} \text{Gamma}(\nu_0/2, s_0/2), \quad (15)$$

$$c_i \sim \text{Gamma}(a_0, b_0), \quad (16)$$

and where the elements of \mathbf{A}_t and $\mathbf{\Sigma}_t$ 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 $\mathbf{\Lambda}_t$ is the precision matrix, which we assume to be diagonal, in order to ensure identification of the model.³ We model the diagonal entries $\lambda_{1,t}, \dots, \lambda_{m,t}$ 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 $\mathbf{\Lambda}_t$ 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's (1994) stick-breaking representation of the DPM mixture model. This allows us to write the density function of each component of the innovation vector $\epsilon_t = (\epsilon_{1t}, \dots, \epsilon_{mt})'$ as an infinite scale-mixture of Gaussian distributions. That is, for $i = 1, \dots, m$ we have

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

where $f_N(\cdot | 0, l_{ij}^{-1})$ denotes the density of the univariate normal distribution with expectation zero and variance l_{ij}^{-1} . The prior of these mixture parameters is given in Eq. (15). The weights ω_{ij} are

²In the hierarchical representation, $\stackrel{\text{d}}{=}$ means "has the distribution." The operator $\text{diag}(\lambda_{1,t}, \dots, \lambda_{m,t})$ creates the diagonal $m \times m$ matrix, say \mathbf{M} , with $\mathbf{M}_{ij} = \lambda_i$ and $\mathbf{M}_{ij} = 0$ for $i \neq j$ ($i, j = 1, \dots, m$).

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

distributed as $\omega_{i1} = v_{i1}, \omega_{ij} = (1 - v_{i1}) \cdots (1 - v_{ij-1}) \cdot v_{ij}$ for $j > 1$, where v_{i1}, v_{i2}, \dots are i.i.d. beta distributed with parameters 1 and c_i [in symbols: $\text{Beta}(1, c_i)$]. 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). Our notation distinguishes between the precision parameters l_{i1}, l_{i2}, \dots of the mixture components and the actually drawn precision parameter λ_{it} of the innovation ϵ_{it} (see Step 6 of the slice sampling algorithm described in Appendix A.3).

For notational convenience, we collect the parameters from the parametric part of our Cholesky SV model in the vector Φ (i.e. Φ contains all parameters from $\mu_x, \Phi_x, \Phi_h, \Sigma_\eta, \Sigma_\epsilon$), and all parameters from the nonparametric specification in the infinite dimensional entity $\Omega = \{\omega_{ij}, l_{ij}\}_{i=1, \dots, m; j=1, 2, \dots, \infty}$. In cases where we need to address all model parameters, we merge the partial parameter entities Φ and Ω into the full-parameter entity Θ .

The Cholesky Dirichlet-Process-Mixture-Stochastic-Volatility (Cholesky DPM-SV) model established in this section can be estimated by Bayesian methods. We describe each step of the MCMC approach in detail in Appendix A.

3. Features of the Cholesky DPM-SV model

3.1. 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 $\mathbf{y}_{1:T} = \{\mathbf{y}_1, \dots, \mathbf{y}_T\}$, we write the one-step ahead predictive density as

$$f(\mathbf{y}_{T+1} | \mathbf{y}_{1:T}) = \int f(\mathbf{y}_{T+1} | \Theta, \mathbf{y}_{1:T}) \pi(\Theta | \mathbf{y}_{1:T}) d\Theta, \quad (18)$$

where (i) the density $f(\mathbf{y}_{T+1} | \Theta, \mathbf{y}_{1:T})$ constitutes an infinite scale mixture, given the representation of the innovation term in Eq. (17), and (ii) the posterior $\pi(\Theta | \mathbf{y}_{1:T})$ is defined on the infinitely dimensional parameter space Θ . Since the integral in Eq. (18) is analytically untractable, we approximate the predictive density via the MCMC output,

$$f(\mathbf{y}_{T+1} | \mathbf{y}_{1:T}) \approx \frac{1}{R} \sum_{r=1}^R f(\mathbf{y}_{T+1} | \Theta^{(r)}, \mathbf{y}_{1:T}), \quad (19)$$

where R is the length of the Markov chain and $\Theta^{(r)}$ denotes the parameter set in iteration r . We cope with the infinitely dimensional parameter space by introducing the latent variables according to Appendix-Eq. (A.25) in each iteration r (which we denote by $\rho_{it}^{(r)}$) and thus for $i = 1, \dots, m$ obtain the following (finite number of) DPM parameters in iteration r :

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

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 $i = 1, \dots, m$:

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 $i = 1, \dots, m$, we compose the predictive error term covariance matrix at iteration r as $(\mathbf{A}^{(r)})^{-1} \equiv \text{diag}(1/l_i^{(r)})$.

We now repeat the complete algorithm (i.e. the three steps for each $i = 1, \dots, m$) a number of times (say B^{\max} times) and record at each iteration r , the B^{\max} covariance matrices $(\mathbf{A}_1^{(r)})^{-1}, \dots, (\mathbf{A}_{B^{\max}}^{(r)})^{-1}$. 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. (19) as

$$f(\mathbf{y}_{T+1} | \mathbf{y}_{1:T}) \approx \frac{1}{R} \sum_{r=1}^R f^{(r)}(\mathbf{y}_{T+1} | \mathbf{y}_{1:T}) \quad (20)$$

with

$$f^{(r)}(\mathbf{y}_{T+1} | \mathbf{y}_{1:T}) = \frac{1}{B^{\max}} \sum_{k=1}^{B^{\max}} f_{\mathbf{N}}\left(\mathbf{y}_{T+1} | \mathbf{0}, (\mathbf{A}_{T+1}^{(r)})^{-1} \boldsymbol{\Sigma}_{T+1}^{(r)} (\mathbf{A}_k^{(r)})^{-1} \boldsymbol{\Sigma}_{T+1}^{(r)} \left[(\mathbf{A}_{T+1}^{(r)})^{-1} \right]'\right), \quad (21)$$

where, for the computation of $\mathbf{A}_{T+1}^{(r)}$ and $\boldsymbol{\Sigma}_{T+1}^{(r)}$, we draw each $\alpha_{iT+1}^{(r)}$ from $N(\mu_{\alpha_i}^{(r)} + \phi_{\alpha_i}^{(r)} \alpha_{iT}^{(r)}, \sigma_{\alpha_i}^2)^{(r)}$ for $i = 1, \dots, p$, and each $h_{iT+1}^{(r)}$ from $N(\phi_{hi}^{(r)} h_{iT}^{(r)}, \sigma_{\eta}^2)^{(r)}$ for $i = 1, \dots, m$. In our empirical application below, we choose $B^{\max} = 3$.⁴

3.2. Conditional moments

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

$$\mathbf{H}_t^* = \text{Cov}(\mathbf{y}_t | \boldsymbol{\Theta}, \mathbf{y}_{1:t-1}) = \mathbf{A}_t^{-1} \boldsymbol{\Sigma}_t \text{Cov}(\boldsymbol{\epsilon}_t | \boldsymbol{\Omega}) \boldsymbol{\Sigma}_t' (\mathbf{A}_t^{-1})', \quad (22)$$

where

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

Using our predictive density from the Eqs. (20) and (21), we may approximate conditional second-moment forecasts of the Cholesky DPM-SV model by

$$\mathbb{E}(\mathbf{H}_{T+1}^*) \approx \frac{1}{R} \sum_{r=1}^R \mathbf{H}_{T+1}^{*(r)}, \quad (23)$$

where

$$\mathbf{H}_{T+1}^{*(r)} = (\mathbf{A}_{T+1}^{(r)})^{-1} \boldsymbol{\Sigma}_{T+1}^{(r)} \frac{1}{B^{\max}} \sum_{k=1}^{B^{\max}} (\mathbf{A}_k^{(r)})^{-1} \boldsymbol{\Sigma}_{T+1}^{(r)} \left[(\mathbf{A}_{T+1}^{(r)})^{-1} \right]'. \quad (24)$$

3.3. 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-SV model is crucial (Primiceri, 2005). In the context of time-varying

⁴In an ideal setting, we would set B^{\max} equal to the true number of components in the data-generating process. Since this number is unknown in our empirical setup, we experimented with several B^{\max} values and found that $B^{\max} = 3$ produces accurate results.

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*.

In our empirical application below, an obvious criterion for variable ordering appears to be the chronological sequence, in which the intercontinental stock markets start their trading days. Nevertheless, owing to the circular structure among the worldwide trading zones, even the concept of chronology does not tie down the-one-and-only reasonable variable ordering. We readdress this issue in Section 4.1, where we justify our explicit variable ordering by taking a concrete financial-market stance. Ultimately, however, our chosen chronological ordering is just one out of $5! = 120$ ordering permutations for our 5-dimensional stock-market data set.

3.4. Simulation

We illustrate the Cholesky DPM-SV estimation framework by means of a simulation example. To this end, we simulated $T = 1,000$ observations from a 5-dimensional model ($m = 5$) according to Eqs. (3)–(9) with parameter values $\phi_{hi} = 0.95$, $\sigma_{\eta_i}^2 = 0.04$ for $i = 1, \dots, 5$, and a (finite) location-scale mixture distribution for the error term given by

$$\epsilon_t \sim \begin{cases} \mathbf{N}(\boldsymbol{\mu}^{(1)}, \text{diag}(0.6, 0.7, 0.6, 0.7, 0.6)) & \text{with probability } 0.9 \\ \mathbf{N}(\boldsymbol{\mu}^{(2)}, \text{diag}(2.02, 1.2746, 2.02, 1.2746, 2.02)) & \text{with probability } 0.1 \end{cases},$$

$$\boldsymbol{\mu}^{(1)} = (0.156 \quad 0.1 \quad 0.156 \quad 0.1 \quad 0.156)',$$

$$\boldsymbol{\mu}^{(2)} = (-1.4 \quad -0.9 \quad -1.4 \quad -0.9 \quad -1.4)',$$

with the values chosen so that (roughly) $\mathbb{E}(\epsilon_t) = \mathbf{0}$. 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 $\phi_{hi} \sim N(0.95, 25)\mathbf{1}(|\phi_{hi}| < 1)$, $\sigma_{\eta_i}^2 \sim \text{InverseGamma}(10/2, 0.5/2)$, $c_i \sim \text{Gamma}(4, 4)$, $G_0 \equiv \text{Gamma}(10/2, 10/2)$.⁵ We estimated the model with 10,000 + 40,000 iterations using the method described in Appendix A.

The upper block of Table 1 displays the posterior means of the AR parameters (along with 90% highest posterior density intervals [HPDIs]), which appear close to the true values. The lower block of Table 1 compares the constant α_{ij} -processes (all set constantly equal to -0.5) with the theoretical expectations of the α_{ij} -processes that prevail upon replacing the theoretical parameters with our parameter estimates (the posterior means). Again, our estimation results appear to be in close line with the true quantities. The average number of mixture components is between 4 and 5 (not reported). Figure 1 displays the posterior means of the five overall variance processes (red lines) in comparison with corresponding simulated paths (blue lines) plus the 90% HPDIs. Evidently, the estimated trajectories capture the simulated volatility dynamics satisfactorily. Figure 2 presents the analogous plots of the overall correlation processes (denoted by $\rho_{21}, \dots, \rho_{54}$), which we obtain from the time-varying covariance matrix \mathbf{H}_t^* from Eq. (22), and where we have thinned the number of draws from the posterior distribution by factor 50. In each panel of Fig. 2, the simulated correlation path lies within the 90% HPDI.

⁵Ausín 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 4.

Table 1. Parameter values, posterior means, 90% highest posterior density intervals.

	True parameter	Posterior mean	90% HPDI
ϕ_{h1}	0.95	0.8694	(0.6001 0.9676)
ϕ_{h2}	0.95	0.9668	(0.9455 0.9874)
ϕ_{h3}	0.95	0.9786	(0.9623 0.9939)
ϕ_{h4}	0.95	0.9701	(0.9489 0.9908)
ϕ_{h5}	0.95	0.9886	(0.9779 0.9984)
$\sigma_{\eta 1}^2$	0.04	0.0822	(0.0158 0.3117)
$\sigma_{\eta 2}^2$	0.04	0.0438	(0.0209 0.0651)
$\sigma_{\eta 3}^2$	0.04	0.0514	(0.0219 0.1072)
$\sigma_{\eta 4}^2$	0.04	0.0340	(0.0180 0.0540)
$\sigma_{\eta 5}^2$	0.04	0.0238	(0.0117 0.0384)
$\mathbb{E}(\alpha_{ji,t}) = \frac{c_{\alpha_j}}{(1-\phi_{\alpha_j})}$			
α_{12}	-0.5	-0.4813	(-0.5411 -0.4220)
α_{13}	-0.5	-0.4759	(-0.5568 -0.4032)
α_{23}	-0.5	-0.4741	(-0.5388 -0.4134)
α_{14}	-0.5	-0.5171	(-0.6016 -0.4416)
α_{24}	-0.5	-0.4746	(-0.5430 -0.3975)
α_{34}	-0.5	-0.4784	(-0.5476 -0.4076)
α_{15}	-0.5	-0.4529	(-0.5411 -0.3538)
α_{25}	-0.5	-0.5284	(-0.6095 -0.4415)
α_{35}	-0.5	-0.4443	(-0.5199 -0.3646)
α_{45}	-0.5	-0.5020	(-0.5636 -0.4302)

Notes: Simulated model according to Eqs. (3)–(9), with $m = 5, T = 1000$, and a location-scale mixture of two normal distributions for the error term, as specified in Section 3.4.

4. Empirical application

4.1. Data

In this section, we apply the Cholesky DPM-SV 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 between February 17, 2012 and February 19, 2016 (1,046 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 Shenzhen CSI 300. All data were collected from *Datastream* (daily closing prices).

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 Fig. 3.

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 3.3, the ordering of the five return series in the vector $\mathbf{y}_t = (y_{1t}, \dots, y_{5t})'$ of the Cholesky DPM-SV model could matter considerably. We decided to choose that chronological sequence, according to which the intercontinental stock market consecutively start their trading on the same calendar day. This implies that y_{1t}, \dots, y_{5t} represent the return series for (1) the Nikkei, (2) the Shanghai Shenzhen, (3) the EuroStoxx50, (4) the DAX, and (5) the Dow Jones. A conceivable alternative ordering, which also meets the aspect of chronology, would be to start with (1) the Dow Jones at date t , followed by (2) the Nikkei, (3) the Shanghai Shenzhen, (4) the

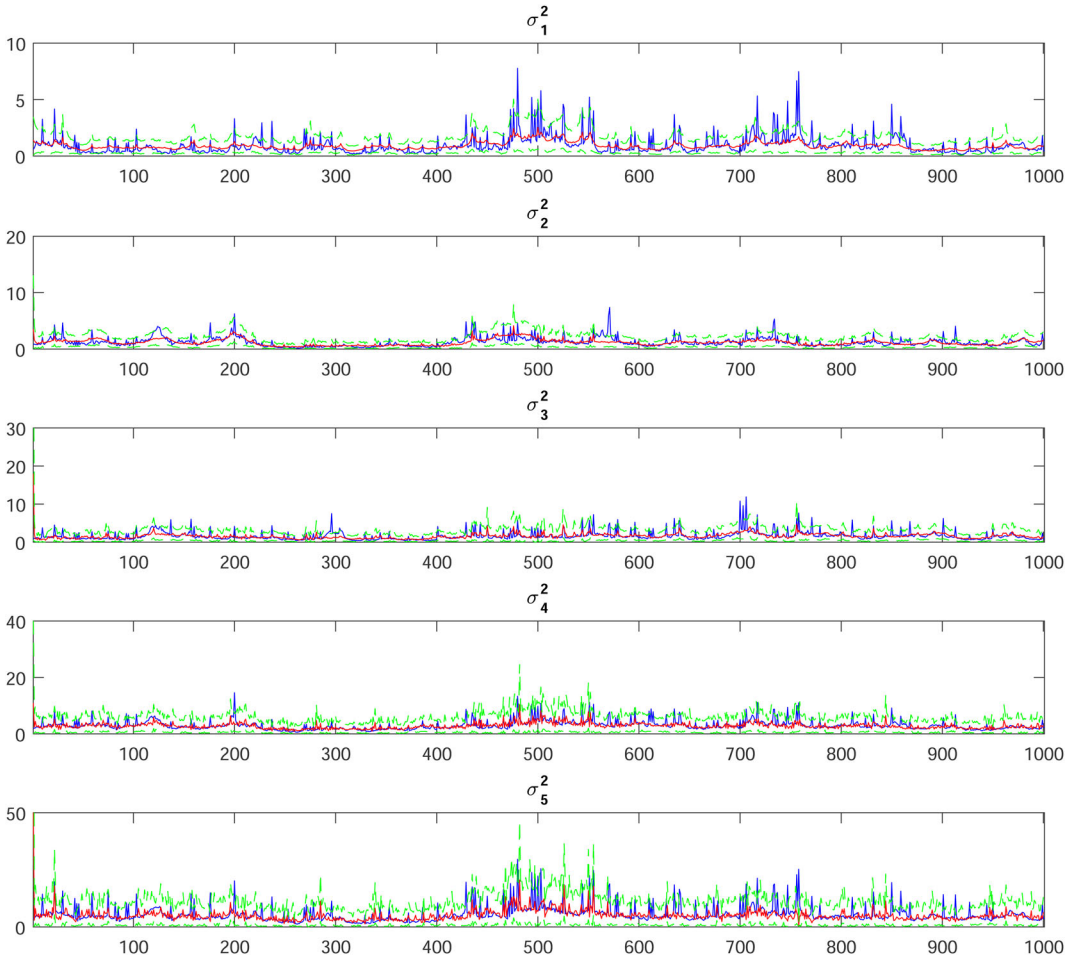


Figure 1. Simulated variance processes (blue lines), posterior means (red lines), and 90% highest posterior density intervals (green lines).

EuroStoxx50, and (5) the DAX, each of the latter four indices on the next trading day. However, in the subsequent analysis we take up a Eurasian financial-market perspective, by considering the European sovereign debt crisis and a bubbly period in the Chinese stock market—both events occurred during the sampling period—as important impulses to the chronologically consecutive markets (on the same calendar day) worldwide.⁶

4.2. Estimation results

According to Eqs. (5)–(9), the estimation of our five-dimensional Cholesky DPM-SV model involves the sampling of (i) five SV processes (\mathbf{h}_t -processes), (ii) ten α_t -processes, (iii) 40 AR-parameters (stemming from the \mathbf{h}_t - and α_t -processes), and (iv) five DPM sets $\{\omega_{ij}, l_{ij}\}_{j=1}^{\infty}$. We ran a total of 50,000 + 50,000 iterations and deleted the first 50,000 results as burn-in phase. As prior distributions, we chose

⁶We thank two reviewers for drawing our attention to this issue.

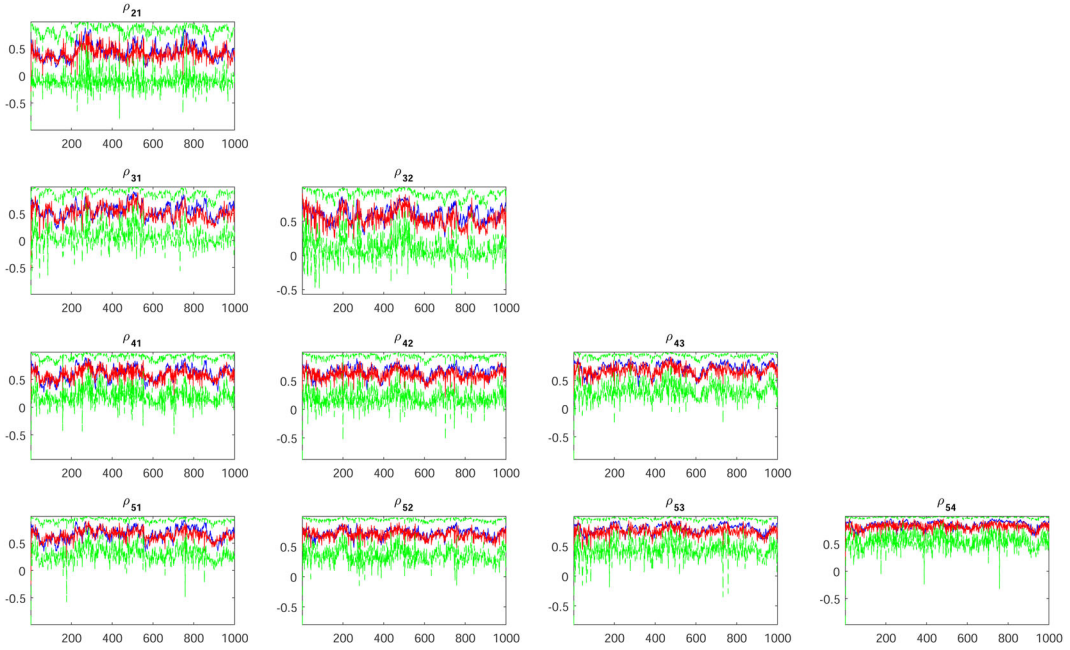


Figure 2. Simulated correlation processes (blue lines), posterior means (red lines), and 90% highest posterior density intervals (green lines).

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

and the base distribution G_0 as $\text{Gamma}(10/2, 10/2)$.⁷ Table 3 displays the posterior means and standard deviations of the 40 AR parameters. Clearly, the parameters do not provide direct interpretation with respect to the overall variance and covariance processes. We note, however, the higher standard errors of the persistence $\phi_{\alpha i}$ -parameter estimates as compared to the standard errors of the $\phi_{\mathbf{h}i}$ -parameter estimates. Since the $\phi_{\alpha i}$ parameters predominantly affect the return co-movements, we expect rather rough co-movement paths.

We assess the co-movements among the five markets via the pairwise in-sample time-varying correlation coefficients (denoted by $\text{Corr}_{\text{INDEX1}, \text{INDEX2}; t}$), which we obtain from the covariance matrix \mathbf{H}_t^* in Eq. (22) computed in each MCMC iteration and at every date t . Figure 4 displays the time-varying correlation coefficients for the ten market pairs. In each panel, the solid line represents the correlation coefficients computed as an average of 333 posterior thinned draws (out of 50,000), while the darkly and brightly shaded areas represent 50% and 90% HPDIs, respectively.

Figure 4 provides the following major findings: (i) The time-varying in-sample correlation coefficients appear surprisingly volatile. (ii) Except for $\text{Corr}_{\text{DJ}, \text{EU}; t}$ (US/European markets),

⁷Since the data turn out not to be very informative about the hyperparameters c_i , we also experimented with other priors for c_i . While the posterior distributions of the hyperparameters c_i are affected, the posterior distributions of the other model parameters do not change substantially.

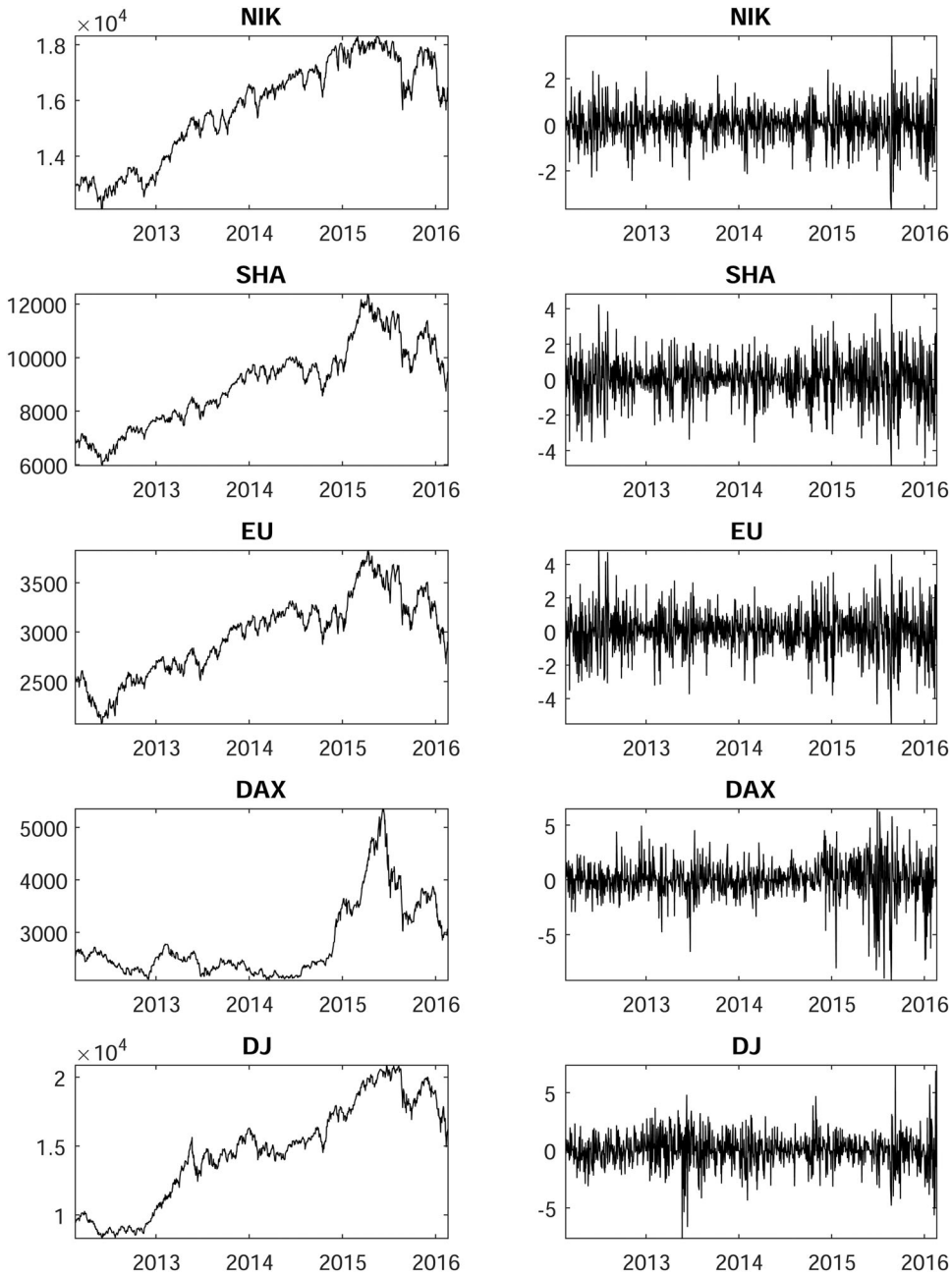


Figure 3. Index values and daily returns (in %).

$\text{Corr}_{\text{DJ},\text{DAX};t}$ (US/German markets) and $\text{Corr}_{\text{DAX},\text{EU};t}$ (German/European markets), the time-varying correlation coefficients take on negative values strikingly often. (iii) The coefficients $\text{Corr}_{\text{EU},\text{SHA};t}$, $\text{Corr}_{\text{DAX},\text{SHA};t}$, $\text{Corr}_{\text{DJ},\text{NIK};t}$, $\text{Corr}_{\text{DJ},\text{SHA};t}$ 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 $\text{Corr}_{\text{SHA},\text{NIK};t}$ 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

Table 2. Descriptive statistics of daily returns (in %).

	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

Note: The indices are abbreviated as NIK (Nikkei 225), SHA (Shanghai Shenzen CSI 300), EU (EuroStoxx), DAX (DAX 30 Performance), DJ (Dow Jones Industrial).

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, Fig. 5 plots the sample correlations obtained from a rolling window of size 50 centered around t . Evidently, Fig. 5 vastly confirms the findings from Fig. 4.

Finally, we investigate the predictive ability of our Cholesky DPM-SV 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. (23) and (24), while Fig. 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 Chinese markets (Panel SHA, EU), the German and the Chinese Markets (Panel SHA, DAX), the US and the Japanese markets (Panel NIK, DJ), and the US and the Chinese markets (Panel SHA, DJ) all reflect the lack of linear dependence, as mentioned in the above discussion on Fig. 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% HPDI), and the lowest variance for the US market.

4.3. DP precision

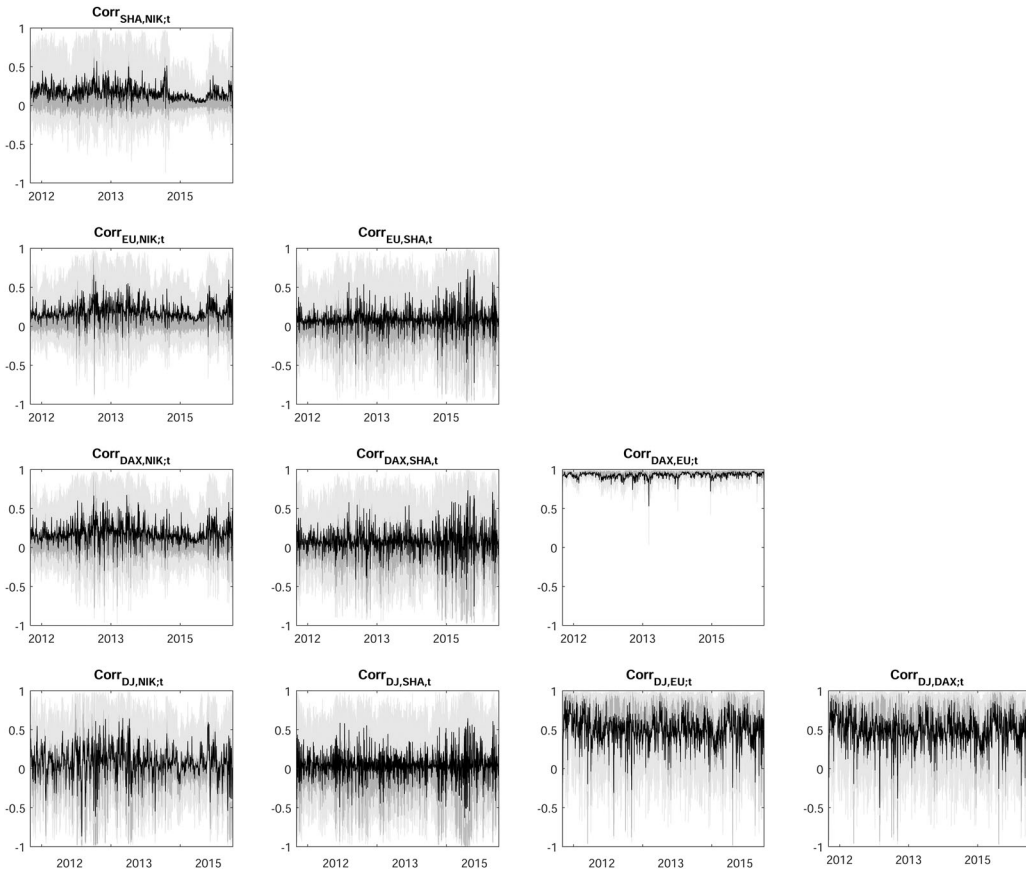
The precision parameter c_i of the Dirichlet process controls the number of mixture components for each $i = 1, \dots, 5$, where the two limiting cases $c_i = 0$ and $c_i \rightarrow \infty$ —under the specific model structure in Eqs. (10)–(16)—correspond to the Gaussian and the Student- t distributions, 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 ($\tilde{c}_i \rightarrow 1$) each versus our general Cholesky DPM-SV model with $\tilde{c}_i \in (0, 1)$. 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 $\tilde{c}_i \rightarrow 1$, 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.⁸

⁸A sensitivity analysis for the parameter c_i (or \tilde{c}_i) reveals that the shape of its posterior is strongly affected by the choice of the prior. This finding is, however, inconsequential, as different specifications of the prior for c_i have only minor impact on the posterior distributions of all (but one) remaining model parameters. The only parameter, affected by the prior of c_i , is the average number of nonempty clusters. This illustrates one of the most prominent features of the Bayesian nonparametric models, namely that the same density can be approximated using different numbers of clusters with different mixing parameters.

Table 3. Posterior means and standard deviations (in parentheses).

i	c_{xi}	ϕ_{xi}	σ_{ei}^2	ϕ_{hi}	σ_{hi}^2	\bar{n}_i
1	-0.1455 (0.0582)	0.0649 (0.3532)	0.0474 (0.0144)	0.9610 (0.0158)	0.0510 (0.0176)	9
2	-0.1614 (0.0548)	0.0601 (0.2965)	0.0657 (0.0180)	0.9799 (0.0087)	0.0364 (0.0103)	10
3	-0.0662 (0.0338)	-0.1444 (0.1953)	0.0945 (0.0234)	0.9323 (0.0372)	0.0724 (0.0499)	12
4	-0.0159 (0.0120)	-0.0370 (0.1403)	0.0261 (0.0054)	0.9980 (0.0013)	0.0330 (0.0135)	5
5	0.0025 (0.0124)	-0.0509 (0.1679)	0.0310 (0.0091)	0.9938 (0.0039)	0.0674 (0.0245)	10
6	-0.4111 (0.1070)	0.5478 (0.1182)	0.0254 (0.0069)			
7	0.0008 (0.0127)	0.4161 (0.1831)	0.0466 (0.0121)			
8	-0.0018 (0.0218)	-0.4898 (0.1796)	0.0277 (0.0107)			
9	-0.1832 (0.0986)	0.2271 (0.3241)	0.0498 (0.0159)			
10	-0.1745 (0.0922)	-0.1649 (0.2732)	0.0488 (0.0136)			

Note: \bar{n}_i denotes the average number of nonempty mixture components rounded to the nearest integer.

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

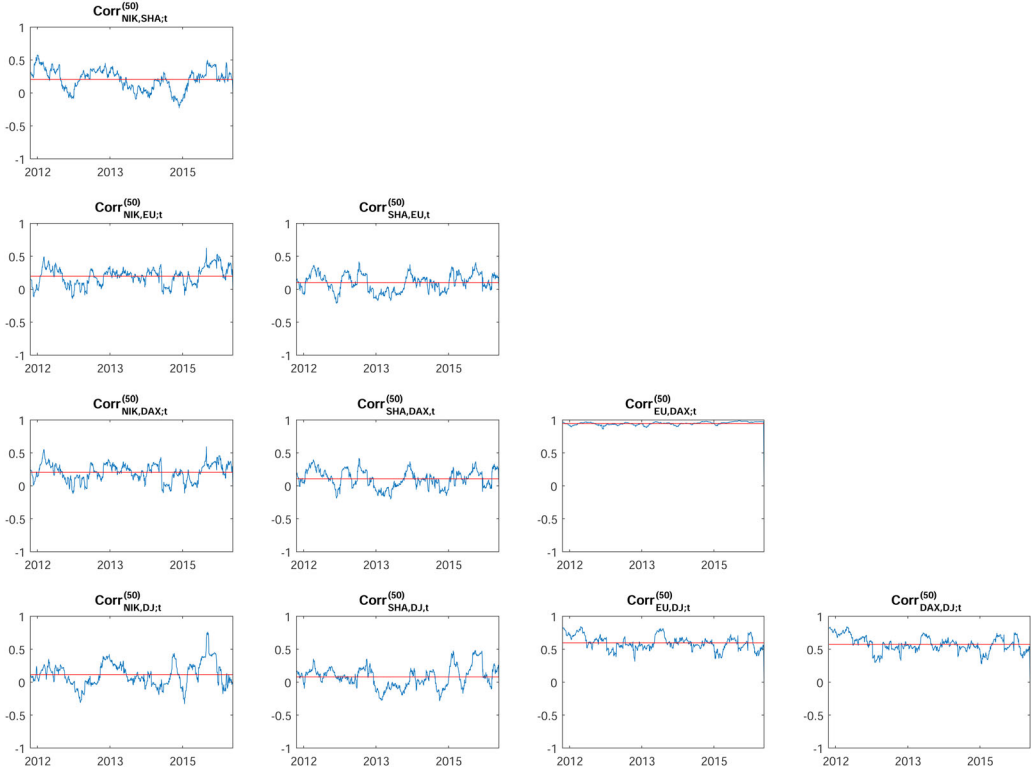


Figure 5. Sample correlations obtained from a rolling window of size 50 centered around the actual observation with the sample-correlation (horizontal line).

4.4. Out-of-sample forecasting model comparison

4.4.1. Benchmark models

In order to analyze out-of-sample predictive power, we compare our Cholesky DPM-SV model with three benchmark specifications.⁹ The first benchmark is the Gaussian Cholesky SV specification

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

where the latent processes are defined as in Eqs. (7)–(9). We estimate the model setting the matrix $\boldsymbol{\Lambda}_t = \mathbf{I}$.

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

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

in which the conditional distribution of the return vector follows a multivariate Student- t distribution (denoted by \mathbf{St}) with mean vector $\mathbf{0}$, covariance matrix \mathbf{H}_t and $m \times 1$ degrees-of-freedom vector $\tilde{\mathbf{v}}$. In order to estimate this specification (without the slice sampler), we use the gamma-normal representation of the t -distribution (see, *inter alia*, Chib and Ramamurthy, 2014). To this

⁹To economize on space, our choice of benchmark models is limited to these three specifications. Another model, not considered here, is the MGARCH-DPM model (Jensen and Maheu, 2013). Jensen and Maheu (2013) propose a nonstochastic (GARCH-type) approach to multivariate volatility modeling, which (i) is order invariant, and (ii) allows for nondiagonal mixing scale. A comparison of their GARCH-type model with our stochastic volatility approach will be subject to future research.

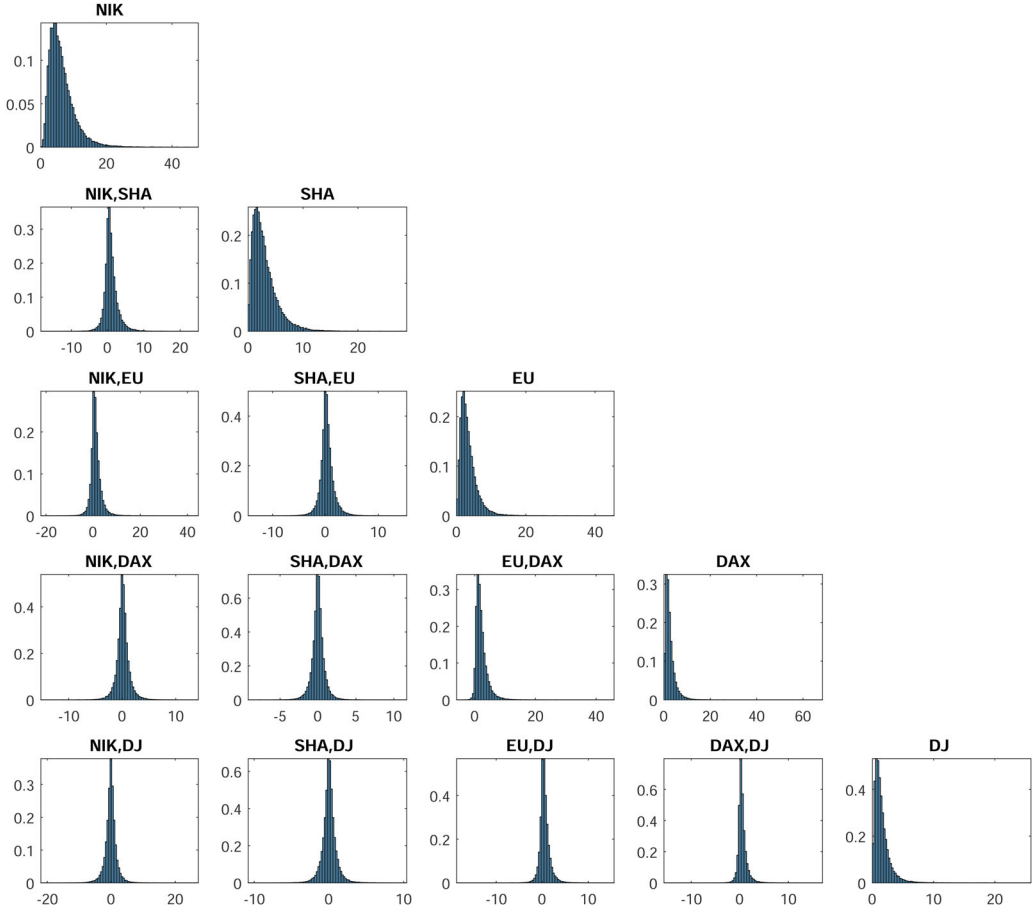


Figure 6. One-step-ahead density forecasts (of the elements of \mathbf{H}_{T+1}^*).

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

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

with π representing some prior distribution.

Conditional on q_{it} , the sampling steps can be performed in exactly the same way as described in [Appendix A.1](#). However, we rewrite the dynamic model from [Appendix-Eqs. \(A.18\) and \(A.19\)](#) as

$$\begin{aligned} \tilde{y}_{it} &= \exp\{h_{it}/2\} q_{it}^{-1/2} u_{it}, \quad (i = 1, \dots, m) \\ h_{it} &= \phi_{hi} h_{it-1} + \eta_{it}, \end{aligned}$$

and the corresponding sampling steps as

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

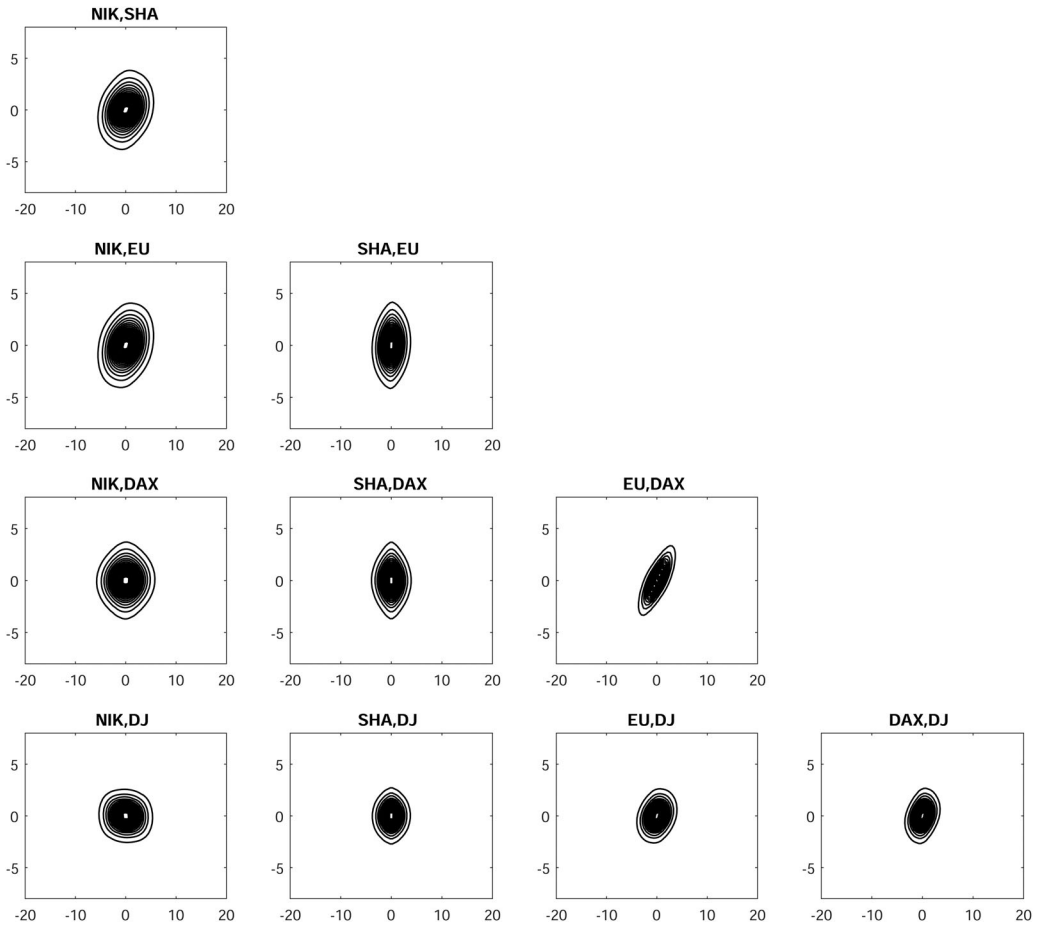


Figure 7. Contour plots of pairwise one-step-ahead density forecasts.

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

H_{T+1}^*	Mean	Median	90% HPDI
$H_{T+1}^* \text{ NIK}$	7.2337	6.1811	(1.5248, 13.2226)
$H_{T+1}^* \text{ SHA}$	3.2385	2.6356	(0.3246, 6.2110)
$H_{T+1}^* \text{ EU}$	3.7878	3.1497	(0.6449, 6.9720)
$H_{T+1}^* \text{ DAX}$	3.6246	2.9022	(0.4688, 6.8787)
$H_{T+1}^* \text{ DJ}$	1.6232	1.1733	(0.0774, 3.3822)
$H_{T+1}^* \text{ SHA, NIK}$	0.9987	0.7307	(-1.8207, 4.0100)
$H_{T+1}^* \text{ EU, NIK}$	1.5025	1.1124	(-1.9305, 5.3189)
$H_{T+1}^* \text{ EU, SHA}$	0.4096	0.2605	(-1.6861, 2.7071)
$H_{T+1}^* \text{ DAX, NIK}$	1.5457	1.1560	(-1.9358, 5.4155)
$H_{T+1}^* \text{ DAX, SHA}$	0.4052	0.2579	(-1.7940, 2.9082)
$H_{T+1}^* \text{ DAX, EU}$	3.5458	2.9134	(0.5579, 6.7130)
$H_{T+1}^* \text{ DJ, NIK}$	0.1496	0.0923	(-2.9625, 3.1029)
$H_{T+1}^* \text{ DJ, SHA}$	0.0052	-0.0103	(-1.5690, 1.7853)
$H_{T+1}^* \text{ DJ, EU}$	1.2784	0.9443	(-0.9013, 3.6663)
$H_{T+1}^* \text{ DJ, DAX}$	1.2232	0.8752	(-0.9412, 3.5763)

Note: The indices are abbreviated as NIK (Nikkei 225), SHA (Shanghai Shenzen CSI 300), EU (EuroStoxx), DAX (DAX 30 Performance), DJ (Dow Jones Industrial).

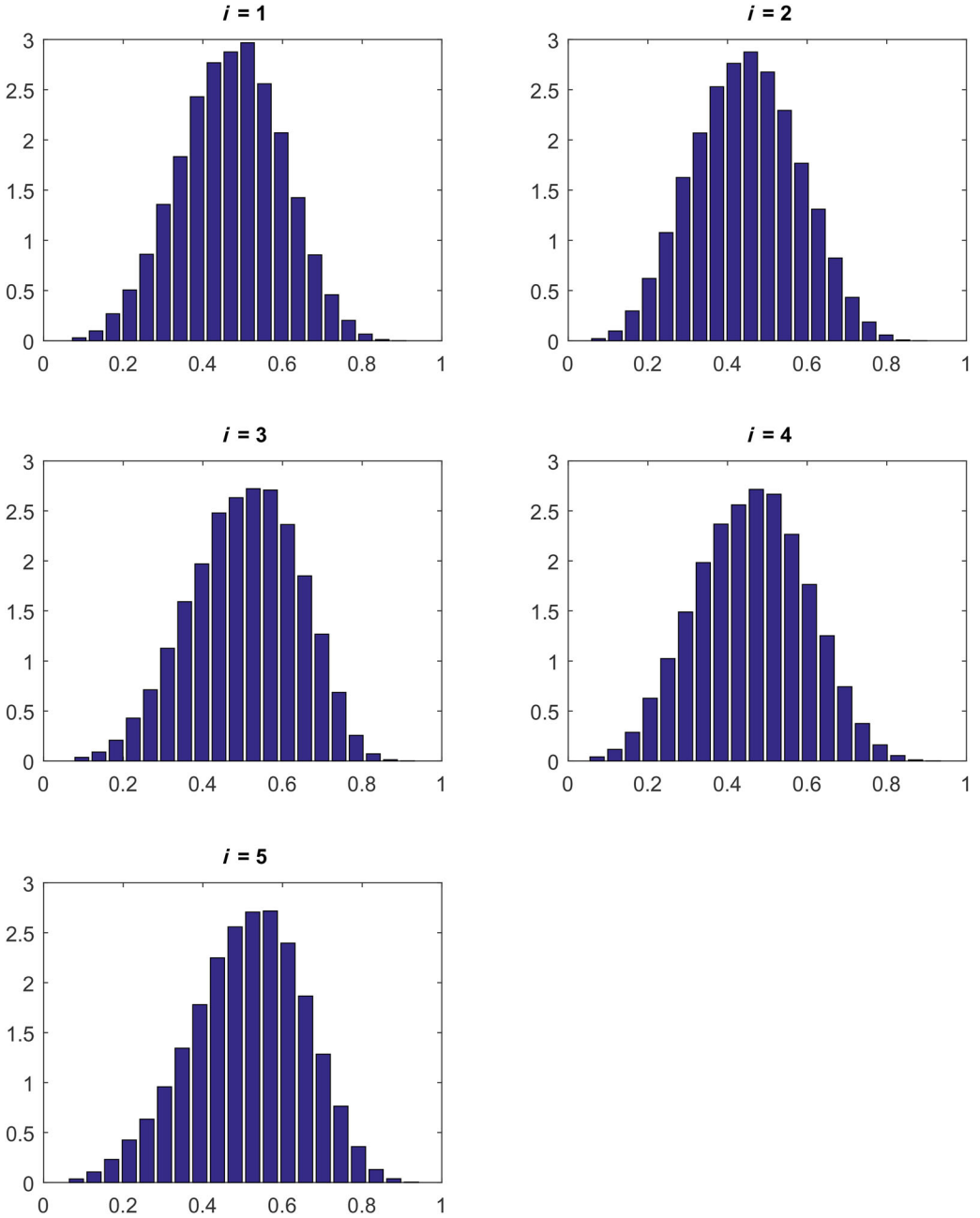


Figure 8. Transformed posterior DP precision.

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 (MH) step in order to sample from the posterior

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

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

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

As the third benchmark model, we consider an asymmetric extension of our Cholesky DPM-SV specification, by extending the infinite scale mixture for the error term to a location scale mixture:

$$\mathbf{y}_t|\boldsymbol{\mu}_t, \boldsymbol{\Lambda}_t, \mathbf{A}_t, \boldsymbol{\Sigma}_t \sim \mathbf{N}(\boldsymbol{\mu}_t, \mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})'), \quad (25)$$

$$\begin{pmatrix} \mu_{i,t} \\ \lambda_{i,t} \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} G_i, \quad (i = 1, \dots, m) \quad (26)$$

$$G_i|G_0, c_i \sim \text{DP}(c_i, G_0), \quad (27)$$

$$G_0 \stackrel{\text{d}}{=} N(b, (\tau\lambda_{i,t})^{-1}) \times \text{Gamma}(\nu_0/2, s_0/2), \quad (28)$$

where the right-hand side of Eq. (28) denotes the conjugate normal-gamma distribution. Besides kurtosis, this model extension is also able to capture skewness, a frequently observed feature in financial time series. The estimation algorithm remains the same as for our original Cholesky DPM-SV framework, except for the latent volatility sampler, which is now conditional on $\mu_{i,t}$, and the sampling of the mixture parameters, which is now

$$\bar{b} = \frac{\tau b + \sum_{t=1}^T \epsilon_{it} \cdot \mathbf{1}(\zeta_{it}^{(r-1)} = j)}{\tau + n_{ij}}, \quad (29)$$

$$\bar{\tau} = \tau + n_{ij}, \quad (30)$$

$$\bar{\nu}_{ij} = \nu_0 + n_{ij}, \quad (31)$$

$$\bar{s}_{ij} = s_0 + \tau^2 b - \bar{\tau}^2 \bar{b} + \sum_{t=1}^T \epsilon_{it}^2 \cdot \mathbf{1}(\zeta_{it}^{(r-1)} = j). \quad (32)$$

4.4.2. 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-SV model with the three benchmark specifications (Gaussian and Student- t Cholesky SV, asymmetric Cholesky DPM-SV). For our in-sample estimation, we use an estimation window from February 17, 2012 to February 19, 2016 (1046 trading days). Our out-of-sample period ranges between February 22, 2016 and July 8, 2016, which amounts to 100 out-of-sample 1-day-ahead predictions.

Table 5 reports the out-of-sample CPL values for the four competing models. We note that the subtraction of two CPL values yields the predictive log Bayes factor, a concept used (i) for measuring relative predictive accuracy, and (ii) for assessing a wide range of model comparison issues (Koop, 2003, Chapter 2.5). In terms of the predictive log Bayes factor, our Cholesky DPM-SV specification outperforms both symmetric benchmark models with the values 14.42 (Gaussian Cholesky SV) and 5.09 (Student- t Cholesky SV). As an appropriate statistical guideline for model comparison, Kass and Raftery (1995) suggest considering twice the (predictive) log Bayes factor. Here, these values are 28.84 and 10.18, both exceeding the threshold level of 10. According to Kass and Raftery's classification, our Cholesky DPM-SV model is therefore "very strongly" preferred to the Gaussian as well the Student- t Cholesky SV specification. Finally, we note that our symmetric Cholesky DPM-SV model also outperforms its asymmetric counterpart (third benchmark) with a (double) log Bayes factor of 2.68 (5.36), implying at least "positive" preference of our base specification according to Kass and Raftery's (1995) guidelines. Therefore, in this

Table 5. Cumulative log predictive likelihoods (CPL).

Model	CPL	Predictive log Bayes factor
Cholesky DPM-SV	-685.5444	
Gaussian Cholesky SV	-699.9647	14.4203
Student- t Cholesky SV	-690.6329	5.0885
Asymmetric Cholesky DPM-SV	-688.2255	2.6811

Note: The three benchmark models (Gaussian and Student- t Cholesky SV, asymmetric Cholesky DPM-SV) are estimated as described in Section 4.4.1. Out-of-sample period: February 22, 2016 – July 8, 2016 (100 observations).

particular case, we find that the additional modeling flexibility, provided by an asymmetric distribution, does not result in improved predictive power.

5. Conclusion

In this article, we establish a Cholesky SV model with a highly flexible nonparametric distribution for the innovation vector—based on the Dirichlet process mixture—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 $\mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t\mathbf{A}_t^{-1}\boldsymbol{\Sigma}_t(\mathbf{A}_t^{-1})'$ contains DPM elements in its nondiagonal 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-SV specification has appealing in-sample properties and, in an out-of-sample forecasting analysis, yields substantially improved density forecasts (in terms of predictive Bayes factors) when compared with two benchmark models from the literature. Our specification also has a higher predictive power than an asymmetric variant. However, the improvement is not as strong as in the case of the other benchmark models, indicating the potential importance of skewed errors. This issue needs to be addressed in future research.

Three conceivable extensions of our modeling framework to be tackled in future research are worth mentioning. (i) Frequently observed volatility asymmetries could be modeled by integrating leverage effects into our Cholesky DPM-SV framework. (ii) Our estimation framework could be applied to high-frequency data sets containing realized (co)variances along the lines of Shirota et al. (2017), who suggest estimating Cholesky realized SV models. (iii) The superior out-of-sample predictive ability of our Cholesky DPM-SV framework calls for investigating potential implications for international investors. Highly relevant research questions include, *inter alia*, the impact on (conditional) value at risk (VaR, CVaR) estimation.

Appendix A: Bayesian inference

This appendix presents the samplers for the Cholesky DPM-SV.

A.1. Sampling the \mathbf{A}_t -elements

In order to apply Forward-Filtering-Backward-Sampling to the elements of the \mathbf{A}_t -matrix, we need to set up an appropriate state-space model (Carter and Kohn, 1994). To this end, we first rewrite Eq. (3) as

$$\mathbf{A}_t\mathbf{y}_t = \boldsymbol{\Sigma}_t\boldsymbol{\epsilon}_t, \quad (\text{A.1})$$

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

$$\mathbf{Z}_t = \begin{pmatrix} 0 & \cdots & \cdots & 0 \\ -y_{1t} & 0 & \cdots & 0 \\ 0 & -\mathbf{y}_{[1:2]t} & \ddots & 0 \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -\mathbf{y}_{[1:m-1]t} \end{pmatrix}, \quad (\text{A.2})$$

in which $\mathbf{y}_{[1:i]t}$ denotes the row vector $(y_{1t}, y_{2t}, \dots, y_{it})$, so that

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\Sigma}_t \boldsymbol{\epsilon}_t, \quad (\text{A.3})$$

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. (A.3) with $\boldsymbol{\Lambda}_t^{-1/2} \mathbf{u}_t$, where \mathbf{u}_t is assumed to follow the m -dimensional $\mathbf{N}(\mathbf{0}, \mathbf{I})$ distribution, and obtain the observation and transition equations

$$\mathbf{y}_t = \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\Sigma}_t \boldsymbol{\Lambda}_t^{-1/2} \mathbf{u}_t \equiv \mathbf{Z}_t \boldsymbol{\alpha}_t + \boldsymbol{\xi}_t, \quad (\text{A.4})$$

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu}_x + \boldsymbol{\Phi}_x (\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}_x) + \mathbf{e}_t, \quad (\text{A.5})$$

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

$$\begin{pmatrix} \boldsymbol{\xi}_t \\ \mathbf{e}_t \end{pmatrix} \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}\left(\mathbf{0}, \begin{bmatrix} \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{\mathbf{e}} \end{bmatrix}\right). \quad (\text{A.6})$$

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 $\mathbf{y}^{(s)} \equiv \{\mathbf{y}_0, \dots, \mathbf{y}_{s-1}, \mathbf{y}_s\}$, $\mathbf{Z}^{(s)} \equiv \{\mathbf{Z}_0, \dots, \mathbf{Z}_{s-1}, \mathbf{Z}_s\}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)} \equiv \{\boldsymbol{\Sigma}_{\boldsymbol{\xi}_0}, \dots, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_{s-1}}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}_s}\}$, respectively, and let

$$\boldsymbol{\alpha}_{t|s} = \mathbb{E}(\boldsymbol{\alpha}_t | \mathbf{y}^{(s)}, \mathbf{Z}^{(s)}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)}, \boldsymbol{\Sigma}_{\mathbf{e}}) \quad (\text{A.7})$$

$$\mathbf{V}_{t|s} = \text{Cov}(\boldsymbol{\alpha}_t | \mathbf{y}^{(s)}, \mathbf{Z}^{(s)}, \boldsymbol{\Sigma}_{\boldsymbol{\xi}}^{(s)}, \boldsymbol{\Sigma}_{\mathbf{e}}). \quad (\text{A.8})$$

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

$$\mathbf{c}_x \equiv (\mu_{x1}(1 - \phi_{x1}), \dots, \mu_{xp}(1 - \phi_{xp}))', \quad (\text{A.9})$$

where $\mu_{x1}, \dots, \mu_{xp}$ are the elements of the vector $\boldsymbol{\mu}_x$ and $\phi_{x1}, \dots, \phi_{xp}$ the diagonal entries of the matrix $\boldsymbol{\Phi}_x$. Then, given the starting values $\boldsymbol{\alpha}_{0|0}$ and $\mathbf{V}_{0|0}$, the standard Kalman filter can be summarized as follows:

$$\boldsymbol{\alpha}_{t|t-1} = \mathbf{c}_x + \boldsymbol{\Phi}_x \boldsymbol{\alpha}_{t-1|t-1}, \quad (\text{A.10})$$

$$\mathbf{V}_{t|t-1} = \boldsymbol{\Phi}_x \mathbf{V}_{t-1|t-1} \boldsymbol{\Phi}_x' + \boldsymbol{\Sigma}_{\mathbf{e}}, \quad (\text{A.11})$$

$$\mathbf{K}_t = \mathbf{V}_{t|t-1} \mathbf{Z}_t' (\mathbf{Z}_t \mathbf{V}_{t|t-1} \mathbf{Z}_t' + \boldsymbol{\Sigma}_{\boldsymbol{\xi}_t})^{-1}, \quad (\text{A.12})$$

$$\boldsymbol{\alpha}_{t|t} = \boldsymbol{\alpha}_{t|t-1} + \mathbf{K}_t (\mathbf{y}_t - \mathbf{Z}_t \boldsymbol{\alpha}_{t|t-1}), \quad (\text{A.13})$$

$$\mathbf{V}_{t|t} = \mathbf{V}_{t|t-1} - \mathbf{K}_t \mathbf{Z}_t \mathbf{V}_{t|t-1}. \quad (\text{A.14})$$

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

$$\boldsymbol{\alpha}_{t|t+1} = \boldsymbol{\alpha}_{t|t} + \mathbf{V}_{t|t} \boldsymbol{\Phi}_x' \mathbf{V}_{t+1|t}^{-1} (\boldsymbol{\alpha}_{t+1} - \mathbf{c}_x - \boldsymbol{\Phi}_x \boldsymbol{\alpha}_{t|t}), \quad (\text{A.15})$$

$$\mathbf{V}_{t|t+1} = \mathbf{V}_{t|t} - \mathbf{V}_{t|t} \boldsymbol{\Phi}_x' \mathbf{V}_{t+1|t}^{-1} \boldsymbol{\Phi}_x \mathbf{V}_{t|t}. \quad (\text{A.16})$$

As the prior distribution of the initial state $\boldsymbol{\alpha}_{0|0}$, we use a multivariate normal distribution (see Section 4) and assume the covariance matrix $\boldsymbol{\Sigma}_{\mathbf{e}}$ to be diagonal with entries $\sigma_{e1}^2, \dots, \sigma_{ep}^2$. Note that for each $i = 1, \dots, p$ the unconditional expectation of the α_{it} -process is $\mathbb{E}(\alpha_{it}) = \mu_{xi} = \frac{c_{xi}}{1 - \phi_{xi}}$, so that the $3p = 3m(m-1)/2$ parameters to be sampled are $c_{x1}, \dots, c_{xp}, \phi_{x1}, \dots, \phi_{xp}, \sigma_{e1}^2, \dots, \sigma_{ep}^2$. The sampling strategy for these parameters is readily obtained from standard Bayesian estimation of the linear regression model. The prior distributions for the c_{xi} - (or μ_{xi} -) and ϕ_{xi} -parameters are normal (the priors for the ϕ_{xi} -parameters are restricted to ensure the p stationarity conditions $|\phi_{xi}| < 1$), while the prior distribution for σ_{ei}^2 is chosen as inverse Gamma. We sample the c_{xi} - and ϕ_{xi} -parameters by the MH algorithm, while the σ_{ei}^2 -parameters are sampled directly.

A.2. Sampling the Σ_t -elements

The vector $\tilde{\mathbf{y}}_t = \mathbf{A}_t \mathbf{y}_t$ has a diagonal covariance matrix. This enables us to independently estimate the m univariate SV models,

$$\tilde{y}_{it} = \sigma_{i,t} \lambda_{i,t}^{-1/2} u_{it}, \quad (i = 1, \dots, m) \quad (\text{A.17})$$

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

$$\tilde{y}_{it} = \exp\{h_{it}/2\} \lambda_{i,t}^{-1/2} u_{it}, \quad (i = 1, \dots, m) \quad (\text{A.18})$$

$$h_{it} = \phi_{\mathbf{h}_i} h_{it-1} + \eta_{it}, \quad (\text{A.19})$$

with $\eta_{it} \sim N(0, \sigma_{\eta_i}^2)$ and $\sigma_{\eta_i}^2$ being the i th diagonal entry of the matrix Σ_{η} .

The m univariate SV models from Eqs. (A.18) and (A.19) can be estimated separately by consecutively sampling from the following conditionals, in the representation of which we use the m row vectors $\boldsymbol{\vartheta}_i = (\sigma_{\eta_i}^2, \phi_{\mathbf{h}_i})$:

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

Sampling from the first conditional is straightforward and analogous to sampling the α_t -parameters in the previous section. The third conditional from above involves sampling the infinite mixture parameters, for which we present the sampling algorithm in Section A.3. As to the second conditional, 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

$$y_{it}^* = \exp\{h_{it}/2\} u_{it}, \quad (i = 1, \dots, m) \quad (\text{A.20})$$

$$h_{it} = \phi_{\mathbf{h}_i} h_{it-1} + \eta_{it}, \quad (\text{A.21})$$

so that our task reduces to sampling from $\pi(h_{i1}, \dots, h_{iT} | y_{i1}^*, \dots, y_{iT}^*, \boldsymbol{\vartheta}_i)$. We accomplish this by using the procedure of Jacquier et al. (2002) who construct a Markov chain for drawing directly from the joint posterior distribution of the latent volatility components.¹⁰ Specifically, let $\mathbf{h}_{-t}^{(i)} \equiv (h_{i0}, \dots, h_{it-1}, h_{it+1}, \dots, h_{iT})'$ and $\mathbf{y}_t^* \equiv (y_{i1}^*, \dots, y_{iT}^*)'$, which are used to decompose $\pi(h_{i1}, \dots, h_{iT} | \mathbf{y}_t^*, \boldsymbol{\vartheta}_i)$ into a set of conditionals of the form $\pi(h_{it} | \mathbf{h}_{-t}^{(i)}, \mathbf{y}_t^*, \boldsymbol{\vartheta}_i)$. 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 stationarity.

Thus, in order to sample from the target distribution $\pi(h_{i1}, \dots, h_{iT} | \mathbf{y}_t^*, \boldsymbol{\vartheta}_i)$, we follow 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 $t = 2, \dots, T-1$ as follows:

$$\begin{aligned} \pi(h_{it} | h_{it-1}, h_{it+1}, y_{it}^*, \boldsymbol{\vartheta}_i) &\propto \pi(y_{it}^* | h_{it}) \pi(h_{it} | h_{it-1}) \pi(h_{it+1} | h_{it}) \\ &\propto \frac{1}{\exp\{h_{it}/2\}} \exp\left\{-\frac{1}{2} \frac{(y_{it}^*)^2}{\exp\{h_{it}/2\}}\right\} \\ &\times \exp\left\{\frac{-(h_{it} - \phi_{\mathbf{h}_i} h_{it-1})^2 - (h_{it+1} - \phi_{\mathbf{h}_i} h_{it})^2}{2\sigma_{\eta_i}^2}\right\}. \end{aligned} \quad (\text{A.22})$$

The density (A.22) does not have a standard form and we apply a MH algorithm for each of the latent volatility components h_{i2}, \dots, h_{iT-1} . We sample the first and last latent volatility components from

$$\pi(h_{i1} | h_{i2}, y_{i1}^*, \boldsymbol{\vartheta}_i) \propto \frac{1}{e^{h_{i1}/2}} \exp\left\{-\frac{1}{2} \frac{(y_{i1}^*)^2}{e^{h_{i1}/2}}\right\} \exp\left\{\frac{-(h_{i2} - \phi_{\mathbf{h}_i} h_{i1})^2}{2\sigma_{\eta_i}^2}\right\}, \quad (\text{A.23})$$

$$\pi(h_{iT} | h_{iT-1}, y_{iT}^*, \boldsymbol{\vartheta}_i) \propto \frac{1}{e^{h_{iT}/2}} \exp\left\{-\frac{1}{2} \frac{(y_{iT}^*)^2}{e^{h_{iT}/2}}\right\} \exp\left\{\frac{-(h_{iT} - \phi_{\mathbf{h}_i} h_{iT-1})^2}{2\sigma_{\eta_i}^2}\right\}. \quad (\text{A.24})$$

As a proposal for the MH algorithm, we use $N(0, \sigma_{\eta_i}^2)$.

¹⁰It is well-known that the sampler of Jacquier et al. (2002) has some inefficiencies that slow down mixing. Chib et al. (2002) and Jensen and Maheu (2010) propose more efficient sampling algorithms. Chib et al. (2002) overcame the naturally built-up dependency between the parameters and the latent volatilities. Jensen and Maheu (2010) suggested a random-blocking approach so that the dependency on the beginning and ending volatilities are mixed over. However, due to the nonparametric part of our model, these algorithms cannot easily be adopted. Since our trace plots do not indicate poor mixing, we propose to use the sampler of Jacquier et al. (2002).

A.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 issue of sampling the infinite number of DPM parameters. The first step consists of introducing a latent variable ρ_{it} (with positive support), such that for $i = 1, \dots, m$ the joint density of the innovation ϵ_{it} and the latent variable ρ_{it} is given by

$$\begin{aligned} f(\epsilon_{it}, \rho_{it} | \Theta) &= \sum_{j=1}^{\infty} \mathbf{1}(\rho_{it} < \omega_{ij}) \cdot f_N(\epsilon_{it} | 0, l_{ij}^{-1}) \\ &= \sum_{j \in \mathcal{A}(\rho_{it})} f_N(\epsilon_{it} | 0, l_{ij}^{-1}), \end{aligned} \quad (\text{A.25})$$

where $\mathbf{1}(\cdot)$ is the indicator function, and $\mathcal{A}(\rho_{it}) \equiv \{j : \omega_{ij} > \rho_{it}\}$, which becomes a finite set for any given $\rho_{it} > 0$. 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

$$f(\epsilon_{it}, \zeta_{it} = j, \rho_{it} | \Theta) = f_N(\epsilon_{it} | 0, l_{ij}^{-1}) \mathbf{1}(j \in \mathcal{A}(\rho_{it})). \quad (\text{A.26})$$

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

1. Sampling c_i : As in Escobar and West (1995) we start by sampling the auxiliary variable $\psi_i \sim \text{Beta}(c_i^{(r-1)} + 1, T)$ and then sample c_i from the 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, $\zeta_i^* = \max\{\zeta_{i1}^{(r-1)}, \dots, \zeta_{iT}^{(r-1)}\}$ and $\pi_{\psi_i} = (a_0 + \zeta_i^* - 1) / (a_0 + \zeta_i^* - 1 + T(b_0 - \log(\psi_i)))$.

2. Sampling v_{ij} : For $j = 1, 2, \dots, \zeta_i^*$, we sample the v_{ij} values from

$$v_{ij} | \zeta_{i1}^{(r-1)}, \dots, \zeta_{iT}^{(r-1)} \sim \text{Beta}(n_{ij} + 1, T - n_i + c_i),$$

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

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 $\rho_i^* = \min\{\rho_{i1}, \dots, \rho_{iT}\}$, 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 v_{ij} from the prior Beta($c_i, 1$) distribution and compute the associated weights ω_{ij} according to the stick-breaking procedure for $j = \zeta_i^* + 1, \dots, j_i^*$. 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

$$l_{ij} \sim \text{Gamma}(\bar{v}_{ij}/2, \bar{s}_{ij}/2), \quad (\text{A.27})$$

$$\bar{v}_{ij} = \nu_0 + n_{ij}, \quad (\text{A.28})$$

$$\bar{s}_{ij} = s_0 + \sum_{t=1}^T \epsilon_{it}^2 \cdot \mathbf{1}(\zeta_{it}^{(r-1)} = j). \quad (\text{A.29})$$

We note that, according to Eq. (A.18), $\epsilon_{it} = \tilde{y}_{it} \exp\{-h_{it}/2\}$ 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 their 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 $\lambda_{i,t} = l_{\zeta_{it}}$.

Acknowledgments

We are grateful to Esfandiari Maasoumi, an anonymous associate editor, and three reviewers for their constructive and extensive comments, which greatly improved the paper. The usual disclaimer applies.

References

- Asai, M., McAleer, M., Yu, J. (2006). Multivariate stochastic volatility: A review. *Econometric Reviews* 25(2–3): 145–175. doi:10.1080/07474930600713564
- 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. doi:10.1016/j.ejor.2013.07.008
- Bauwens, L., Laurent, S., Rombouts, J. V. (2006). Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21(1):79–109. doi:10.1002/jae.842
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31(3): 307–327. doi:10.1016/0304-4076(86)90063-1
- Carter, C. K., Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81(3):541–553. doi:10.1093/biomet/81.3.541
- Chan, J., Doucet, A., Leon-Gonzalez, R., Strachan, R. W. (2018a). Multivariate stochastic volatility with co-heteroscedasticity. GRIPS Discussion Paper 18–12. Tokyo, Japan: National Graduate Institute for Policy Studies.
- Chan, J., Leon-Gonzalez, R., Strachan, R. W. (2018b). Invariant inference and efficient computation in the static factor model. *Journal of the American Statistical Association* 113(522):819–828. doi:10.1080/01621459.2017.1287080
- Chib, S., Nardari, F., Shephard, N. (2002). Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108(2):281–316. doi:10.1016/S0304-4076(01)00137-3
- Chib, S., Omori, Y., Asai, M. (2009). Multivariate stochastic volatility. In: Andersen, T.G., Davis, R. A., Kreiß, J.-P., Mikosch, T., eds., *Handbook of Financial Time Series*, New York: Springer, pp. 365–400.
- Chib, S., Ramamurthy, S. (2014). DSGE models with student-*t* errors. *Econometric Reviews* 33(1–4):152–171. doi:10.1080/07474938.2013.807152
- Clements, A. E., Hurn, A. S., Volkov, V. V. (2015). Volatility transmission in global financial markets. *Journal of Empirical Finance* 32:3–18. doi:10.1016/j.jempfin.2014.12.002
- Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1(2): 223–236. doi:10.1080/713665670
- Delatola, E.-I., Griffin, J. E. (2011). Bayesian nonparametric modelling of the return distribution with stochastic volatility. *Bayesian Analysis* 6(4):901–926. doi:10.1214/11-BA632
- Delatola, E.-I., Griffin, J. E. (2013). A Bayesian semiparametric model for volatility with a leverage effect. *Computational Statistics and Data Analysis* 60:97–110. doi:10.1016/j.csda.2012.10.023
- Ehrmann, M., Fratzscher, M., Rigobon, R. (2011). Stocks, bonds, money markets and exchange rates: Measuring international financial transmission. *Journal of Applied Econometrics* 26(6):948–974. doi:10.1002/jae.1173
- Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4):987–1007. doi:10.2307/1912773
- Escobar, M. D., West, M. (1995). Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association* 90(430):577–588. doi:10.1080/01621459.1995.10476550
- Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems. *The Annals of Statistics* 1(2):209–230. doi:10.1214/aos/1176342360
- Harvey, A., Ruiz, E., Shephard, N. (1994). Multivariate stochastic variance models. *The Review of Economic Studies* 61(2):247–264. doi:10.2307/2297980
- Jacquier, E., Polson, N. G., Rossi, P. E. (2002). Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 20(1):69–87. doi:10.1198/073500102753410408
- Jensen, M. J., Maheu, J. M. (2010). Bayesian semiparametric stochastic volatility modeling. *Journal of Econometrics* 157(2):306–316. doi:10.1016/j.jeconom.2010.01.014
- Jensen, M. J., Maheu, J. M. (2013). Bayesian semiparametric multivariate GARCH modeling. *Journal of Econometrics* 176(1):3–17. doi:10.1016/j.jeconom.2013.03.009
- 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. doi:10.1016/j.jeconom.2013.08.018
- Kalli, M., Griffin, J. E., Walker, S. G. (2011). Slice sampling mixture models. *Statistics and Computing* 21(1): 93–105. doi:10.1007/s11222-009-9150-y
- 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 and Economic Statistics* 31(4):371–383. doi:10.1080/07350015.2013.794142

- Kass, R. E., Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association* 90(430):773–795. doi:10.1080/01621459.1995.10476572
- Kim, S., Shepherd, N., Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies* 65(3):361–393. doi:10.1111/1467-937X.00050
- Koop, G. (2003). *Bayesian Econometrics*, Chichester: Wiley.
- Lopes, H., McCulloch, R., Tsay, R. (2012). Cholesky stochastic volatility models for high-dimensional time series. Discussion papers. Available at: www.researchgate.net. Last accessed 21 September 2016.
- Maheu, J. M., Yang, Q. (2016). An infinite hidden Markov model for short-term interest rates. *Journal of Empirical Finance* 38:202–220. doi:10.1016/j.jempfin.2016.06.006
- Nakajima, J. (2017). Bayesian analysis of multivariate stochastic volatility with skew return distribution. *Econometric Reviews* 36(5):546–562. doi:10.1080/07474938.2014.977093
- 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. Tokyo: 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. doi:10.1111/j.1467-937X.2005.00353.x
- Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica* 4:639–650.
- Shirota, S., Omori, Y., Lopes, H. F., Piao, H. (2017). Cholesky realized stochastic volatility model. *Econometrics and Statistics* 3:34–59. doi: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, Amsterdam: Elsevier, pp. 203–226.
- Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: Wiley.
- Virbickaitė, A., Ausín, M. C., Galeano, P. (2016). A Bayesian non-parametric approach to asymmetric dynamic conditional correlation model with application to portfolio selection. *Computational Statistics and Data Analysis* 100:814–829. doi:10.1016/j.csda.2014.12.005
- Walker, S. G. (2007). Sampling the Dirichlet mixture model with slices. *Communications in Statistics - Simulation and Computation* 36(1):45–54. doi:10.1080/03610910601096262