Universität St.Gallen

Multivariate Stochastic Volatility with Dynamic Cross Leverage

Sebastian Trojan

August 2014 Discussion Paper no. 2014-24

School of Economics and Political Science,
Department of Economics

University of St Gallen

Editor:	Martina Flockerzi University of St.Gallen School of Economics and Political Science Department of Economics Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25 Fax +41 71 224 31 35 Email seps@unisg.ch
Publisher:	School of Economics and Political Science Department of Economics University of St.Gallen Bodanstrasse 8 CH-9000 St. Gallen Phone +41 71 224 23 25
Electronic Publication:	Fax +41 71 224 31 35 http://www.seps.unisg.ch

Multivariate Stochastic Volatility with Dynamic Cross Leverage[1]

Sebastian Trojan

Author's address: Sebastian Trojan
Institute of Mathematics and Statistics
University of St.Gallen
Bodanstrasse 6
CH-9000 St. Gallen
Email sebastian.trojan@student.unisg.ch
sebastian.trojan@web.de

[1] The author thanks Francesco Audrino, Michael Lechner, Nikolaus Hautsch, and seminar participants at the University of St. Gallen for their helpful comments and suggestions.

Abstract

A multivariate stochastic volatility (MSV) model based on a Cholesky-type decomposition of the covariance matrix to model dynamic correlation in the observation and transition error as well as in cross leverage terms is proposed. The empirically relevant asymmetric concept of cross leverage is defined as a nonzero correlation between the i^{th} asset return at time t and the j^{th} log-volatility at time $t+1$. Volatilities and covariances are modeled separately, which makes an interpretation of leverage parameters straightforward. The model is applied on a three-dimensional portfolio consisting of the S&P 500 sector indices Financials, Industrials and Healthcare, spanning the recent financial crisis 2008/09. During and in the aftermath of market turmoil, increased cross leverage effects, higher unconditional kurtosis and stronger correlated information flow are observed. However, there is risk of overfitting and restricting time variation to elements governing dynamics of the observation error may be advisable.

Keywords

Multivariate stochastic volatility, dynamic correlation, cross leverage, Cholesky decomposition, nonlinear state space model, Markov chain Monte Carlo, block sampler, particle filter.

JEL Classification

C11, C15, C32, C58.

1 Introduction

Multivariate Stochastic Volatility (MSV) has more recently became an active and exiting research field. Reviews are given by e.g. Chib, Omori, and Asai, 2009, or Asai, McAleer, and Yu, 2006. Thereby, the asymmetric concept of cross leverage is empirically relevant but relatively less explored. It is defined as a nonzero correlation between the i^{th} asset return at time t and the j^{th} log-volatility at time $t+1$. The current paper proposes a MSV model based on a Cholesky-type decomposition of the covariance matrix to model dynamic correlation in the observation and transition error as well as in cross leverage terms. Volatilities and covariances are modeled separately, which makes an interpretation of the leverage parameters straightforward. A latent and nonlinear nature of stochastic volatility prevents direct computation of the likelihood in general. Accordingly, simulation based techniques, among those Markov chain Monte Carlo (McMC) certainly one of the more efficient and flexible, are commonly employed for estimation.

Three main strands of MSV models may be identified in the literature. The first, generally referred to as the basic MSV model, is a direct extension of the univariate SV model by introducing correlation in the observation and transition errors and dates back to Harvey, Ruiz, and Shephard (1994), who employ (inefficient) quasi maximum likelihood (QML) for estimation. Asai and McAleer (2006) extend this class of models to incorporate leverage effects, which is empirically important at least for stock indices (see e.g. Yu, 2005, or Omori, Shephard, and Nakajima, 2007). They additionally consider size effects and use Monte Carlo likelihood (MCL) for estimation. Chan, Kohn, and Kirby (2006), and Ishihara and Omori (2012) estimate a complete specification of the covariance matrix using McMC, including cross leverage effects, but the former specify leverage in a less common, contemporaneous way. The latter further introduce a heavy tailed observation error into the model and find evidence of both in a five-dimensional time series of S&P 500 sector index returns. However, a major drawback of the basic MSV model is that, although the covariance of returns is time varying due to the volatility dynamics, the implied conditional correlation matrix is constant.

Volatility mean factor models (e.g. Pitt and Shephard, 1999c, and Aguilar and West, 2000) offer a possible remedy to the afore mentioned drawback. In this class of models, volatility comovements are determined by a combination of underlying latent factors. Thereby, factor and idiosyncratic observation errors evolve according to univariate stochastic volatility dynamics. Chib, Nardari, and Shephard (2006) generalize this approach to allow for fat tails and jumps in the observation equation. Han (2006) models factor evolution as an $\mathrm{AR}(1)$ process. Lopes and Carvalho (2007) introduce time varying factor loadings and Markov switching in the common factor volatilities. Philipov and Glickman (2006b) let the factor covariance matrix evolve according to an inverse Wishart autoregressive process (WAR). All preceding models are estimated by McMC. A great advantage of factor models, if specified appropriately, is their scalability in terms of factors and assets. However, as Chan et al. (2006) note, interpretation of a possible leverage effect would be difficult due to now two sources of volatility.

A further approach to generate time varying correlation is by modeling correlation (or functions of correlation) directly. Yu and Meyer (2006) model correlation in a bivariate SV model using the fisher transformation. Asai and McAleer (2009) propose two dynamic correlation MSV models involving the Wishart distribution, building on ideas of Engle (2002). An advantage of both works is that volatility and correlation dynamics are modeled separately. However, extending the former to higher dimensions is difficult due to positive definiteness constraints on the correlation matrix. A feature of the latter, instead, is its parsimonious parameterization, making it potentially capable to handle a larger number of assets. Philipov and Glickman (2006a) let the conditional covariance matrix follow an inverse Wishart distribution with autoregressive evolution. Ishihara, Omori, and Asai (2013) model dynamics in the covariance matrix of returns using the matrix exponential transformation. A feature of the latter work is the inclusion of (time invariant) cross leverage effects into the model. Tsay (2005) introduces a Cholesky decomposition to model correlation dynamics. Lopes, McCulloch, and Tsay (2013) present a remarkably large scale application of this method, fitting a multivariate series of 94 members of the S&P 100 index. In a different development, Gouriéroux, Jasak, and Sufana (2009), and Gouriéroux (2006) propose the Wishart autoregressive process. Except the latter two, all works presented in the current paragraph apply again McMC for model estimation.

The approach taken in this work is an application of the Cholesky decomposition method, whereby the complete covariance matrix of returns and log-volatilities is decomposed. In this way one is not only able to model the covariance dynamics of returns, but also dynamics in the (cross) leverage effects and transition error. Estimation is by McMC, where the block sampler proposed by Ishihara and Omori (2012) is extended to handle time varying covariances.

The dynamic cross leverage model is introduced in Section 2. Section 3 presents a McMC algorithm for posterior computation and Section 4 an associated particle filter. Empirical applications on S&P 500 sector indices data follow in Section 5, including estimation of a model variant with constant correlation. Model fit is investigated in Section 6. Section 7 concludes. The Appendix contains technical details.

2 The Model

Let y_t denote an asset return at time t. The basic univariate stochastic volatility (SV) model with normal error and leverage can then be specified as follows,

$$y_t = \exp(\alpha_t/2)\epsilon_t, \qquad t = 1, \dots, n, \tag{1}$$

$$\alpha_{t+1} = \phi\alpha_t + \eta_{t+1}, \qquad t = 1, \dots, n-1, \tag{2}$$

$$\alpha_1 \backsim \mathcal{N}(0, \sigma^2_{\text{ini}}), \tag{3}$$

$$\begin{pmatrix} \epsilon_t \\ \eta_{t+1} \end{pmatrix} \backsim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} \sigma^2_\epsilon & \rho\sigma_\epsilon\sigma_\eta \\ \rho\sigma_\eta\sigma_\epsilon & \sigma^2_\eta \end{pmatrix}, \tag{4}$$

where α_t is the unobserved volatility.[1] The expected value of the volatility evolution process, Eq. (2), must be set to 0 for identification.[2] Further, Eq. (1)-(4) constitute

[1]Strictly speaking, it is the unobserved log-variance, but as in the literature I speak about volatility throughout. This simplifies terminology and should not cause conceptual confusion as the respective measures are linked by monotonic transformation.

[2]Pitt and Shephard (1999b) show that estimating the volatility mode of SV models in the transition rather than observation equation is superior in terms of efficiency. However, a reparameterization is needed that would result in a constraint covariance matrix. Chan and Jeliazkov (2009) propose McMC techniques, but an extension to time varying covariance matrices appears rather infeasible.

a state space model with measurement equation (1) and transition equation (2) (see e.g Durbin and Koopman, 2008, or Harvey, 1989, for a comprehensive treatment). Stationarity of the volatility process is imposed ($|\phi| < 1$), and the initial value α_1 in Eq. (3) is assumed to follow the respective unconditional distribution, with $\sigma^2_{\text{ini}} = \sigma^2/(1-\phi^2)$. Parameter ρ measures the correlation between ϵ_t and η_{t+1}. We have volatility asymmetry if $\rho \neq 0$ and specifically, if $\rho < 0$, we speak of the leverage effect.

The basic SV model of Eq. (1)-(4) is now extended to the multivariate case with time varying covariance matrix $\boldsymbol{\Sigma}_{t:t+1}$. Let $\boldsymbol{y}_t = (y_{1,t}, \ldots, y_{p,t})'$ denote a p dimensional vector of asset returns and $\boldsymbol{\alpha}_t = (\alpha_{1,t}, \ldots, \alpha_{p,t})'$ the corresponding stochastic volatility. The proposed MSVdc model is then given by

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{V}_t^{1/2}\boldsymbol{\epsilon}_t, && t = 1, \ldots, n, && (5)\\ \boldsymbol{\alpha}_{t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\eta}_{t+1}, && t = 1, \ldots, n-1, && (6)\\ \boldsymbol{\alpha}_1 &\backsim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\text{ini}}), \end{aligned}$$

with

$$\begin{aligned} \boldsymbol{V}_t &= \operatorname{diag}(\exp(\alpha_{1,t}), \ldots, \exp(\alpha_{p,t})),\\ \boldsymbol{\Phi} &= \operatorname{diag}(\phi_1, ..., \phi_p),\\ \begin{pmatrix} \boldsymbol{\epsilon}_t \\ \boldsymbol{\eta}_{t+1} \end{pmatrix} &\backsim \mathcal{N}_{2p}(\boldsymbol{0}, \boldsymbol{\Sigma}_{t:t+1}), \quad \boldsymbol{\Sigma}_{t:t+1} = \begin{pmatrix} \boldsymbol{\Sigma}_{\epsilon\epsilon,t} & \boldsymbol{\Sigma}_{\epsilon\eta,t:t+1} \\ \boldsymbol{\Sigma}_{\eta\epsilon,t:t+1} & \boldsymbol{\Sigma}_{\eta\eta,t+1} \end{pmatrix}, \end{aligned} \tag{7}$$

where operator $\operatorname{diag}(\cdot)$ creates a diagonal matrix. Again, the underlying latent volatility process is assumed stationary, $|\phi_i| < 1, i = 1, \ldots, p$, with $\boldsymbol{\Sigma}_{\text{ini}}$ the unconditional distribution, $\operatorname{vec}(\boldsymbol{\Sigma}_{\text{ini}}) = (\boldsymbol{I}_{p^2} - \boldsymbol{\Phi} \otimes \boldsymbol{\Phi})^{-1}\operatorname{vec}(\boldsymbol{\Sigma}_{\eta\eta,1})$, where $\operatorname{vec}(\cdot)$ stacks a matrix column-wise, $\otimes$ denotes the Kronecker product, and $\boldsymbol{I}_p$ is the identity matrix of dimension p. To obtain $\boldsymbol{\Sigma}_{\eta\eta,1}$, it is assumed that $\boldsymbol{\Sigma}_{0:1} = \boldsymbol{\Sigma}_{1:2}$, a reasonable strategy for initialization. Observe that volatilities and covariances are modeled separately, which makes an interpretation of leverage straightforward. Note further that leaving $\boldsymbol{\Sigma}_{t:t+1}$ time invariant recovers the multivariate stochastic volatility model with cross leverage proposed by Ishihara and Omori (2012).

Time varying covariance matrix $\boldsymbol{\Sigma}_{t:t+1}$ is modeled by Cholesky-type decomposition following Chan and Jeliazkov (2009), which ensures positive semi-definiteness and is amenable to Bayesian estimation,

$$\boldsymbol{\Sigma}^{-1}_{t:t+1} = \boldsymbol{A}'_{t:t+1}\boldsymbol{H}^{-1}\boldsymbol{A}_{t:t+1},$$

where $\boldsymbol{A}_{t:t+1}$ is a lower triangular matrix with ones on the diagonal and $\boldsymbol{H}$ a diagonal matrix with positive elements,

$$\boldsymbol{A}_{t:t+1} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ a_{21,t} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ a_{q1,t:t+1} & \cdots & a_{q,q-1,t+1} & 1 \end{pmatrix}, \quad \boldsymbol{H} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_q \end{pmatrix}, \tag{8}$$

with $q = 2p$ the dimension of $\boldsymbol{\Sigma}_{t:t+1}$. This reparameterization is in fact an orthogonal transformation, and the dynamics specified in Eq. (5)-(6) can now be formulated as

$$\boldsymbol{A}_{t:t+1}\boldsymbol{z}_{t:t+1} \backsim \mathcal{N}(\boldsymbol{0}, \boldsymbol{H}), \quad \boldsymbol{z}_{t:t+1} = \begin{pmatrix} \boldsymbol{V}_t^{-1/2}\boldsymbol{y}_t \\ \boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t \end{pmatrix}. \tag{9}$$

Rewriting Eq. (9) leads to a linear Gaussian state space model as follows,

$$
\begin{aligned}
\boldsymbol{z}_{t:t+1} &= \boldsymbol{X}_{t:t+1}\boldsymbol{a}_{t:t+1} + \boldsymbol{H}^{1/2}\mathbf{e}_{t:t+1}, \quad \mathbf{e}_{t:t+1} \backsim \mathcal{N}(\mathbf{0}, \boldsymbol{I}_q), && (10)\\
\boldsymbol{X}_{t:t+1} &= -\begin{pmatrix}
0 & \cdots & & & & & & 0\\
z_1 & 0 & 0 & \cdots & & & & \vdots\\
0 & z_1 & z_2 & 0 & \cdots & & & \\
0 & 0 & 0 & z_1 & \cdots & & & \\
\vdots & & & & \ddots & 0 & \cdots & 0\\
0 & \cdots & & & 0 & z_1 & \cdots & z_{q-1}
\end{pmatrix},\\
\boldsymbol{a}_{t:t+1} &= \begin{pmatrix} a_{21} & a_{31} & \cdots & a_{q,q-1}\end{pmatrix}', && (11)
\end{aligned}
$$

with $\boldsymbol{X}_{t:t+1}$ a $q \times q(q-1)/2$ matrix containing elements z_i of $\boldsymbol{z}_{t:t+1}$, and $\boldsymbol{a}_{t:t+1}$ a $q(q-1)/2 \times 1$ vector having the a_{ij}'s of $\boldsymbol{A}_{t:t+1}$ stacked row-wise (time subscripts dropped for clarity). The recursive conditional (or triangular) regressions in Eq. (10) constitute the observation equations and $\boldsymbol{a}_{t:t+1}$ the state variables. Model specification is then completed by specifying simple independent random walk dynamics for the state transitions,

$$
a_{ij,t+1} = a_{ij,t} + \mathrm{u}_{ij,t+1}, \quad \mathrm{u}_{ij,t+1} \backsim \mathcal{N}(0, \sigma^2_{ij}), \tag{12}
$$

with $a_{ij,\mathrm{ini}} \backsim \mathcal{N}(0, 100^2)$, $i \le p$, and $a_{ij,\mathrm{ini}} \backsim \mathcal{N}(0,1)$, $i > p$, set.[3] The initial variances can be regarded as diffuse considering empirical relevant values for $\boldsymbol{a}_{t:t+1}$.

In the current application, the elements of $\boldsymbol{A}_{t:t+1}$ corresponding to dynamics of transition error $\boldsymbol{\eta}_{t+1}$ are kept constant over time, introducing some parsimony. More formally, a_{ij} fix for $i, j > p$. Such a stepwise approach to complexity is a sensible strategy, regarding that these dynamics are completely second order latent and inference can be expected difficult. Note however that $\boldsymbol{\Sigma}_{\eta\eta,t+1}$ continues to be time varying due to the specific structure of the Cholesky decomposition. Similarly, we obtain time varying volatility modes for $i = 2, \ldots, p$.

3 McMC Algorithm

Define parameters $\boldsymbol{\theta} \equiv \{\boldsymbol{\phi}, \boldsymbol{\lambda}, \boldsymbol{\sigma}\}$, with $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_p)'$, $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_q)'$, and $q(q-1)/2 \times 1$ vector $\boldsymbol{\sigma} = (\sigma_{21}, \ldots, \sigma_{q,q-1})'$. Further let $\boldsymbol{y}_{1:n} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_n)'$, $\boldsymbol{\alpha}_{1:n} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n)'$, and $\boldsymbol{a}_{1:n} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n)'$, the latter a $n \times q(q-1)/2$ matrix containing the elements of $\boldsymbol{a}_{t:t+1}$ in Eq. (11), $t = 1, \ldots, n$, stacked appropriately. Denote by $\boldsymbol{\theta}_{-r}$ all parameters in $\boldsymbol{\theta}$ without r. Prior distributions are assumed to be independent and generically denoted by $\pi(\cdot)$, with

$$
\begin{aligned}
\lambda_i &\backsim \mathcal{IG}\left(\frac{\nu_{0,i}}{2}, \frac{\delta_{0,i}}{2}\right), \quad i = 1, \ldots, q, && (13)\\
\sigma^2_{ij} &\backsim \mathcal{IG}\left(\frac{b_{0,ij}}{2}, \frac{c_{0,ij}}{2}\right), \quad i = 2, \ldots, q, \quad j < i,
\end{aligned}
$$

where $\mathcal{IG}(\cdot)$ is the inverse gamma distribution. In the case of a_{ij} fix,

$$
a_{ij} \backsim \mathcal{N}(a_{0,ij}, \sigma^2_{a_0,ij}).
$$

[3] One may consider such an approach to initialization ad hoc and decide to implement exact state filtering and smoothing algorithms as e.g. proposed by de Jong (1991a), and outlined in Durbin and Koopman (2008).

The McMC algorithm then consists of the following steps:[4]

1. Initialize $\boldsymbol{\theta}$, $\boldsymbol{\alpha}_{1:n}$, $\boldsymbol{a}_{1:n}$.
2. Generate $\boldsymbol{\alpha}_{1:n} \,|\, \boldsymbol{\theta}, \boldsymbol{a}_{1:n}, \boldsymbol{y}_{1:n}$.
3. Generate $\boldsymbol{a}_{1:n} \,|\, \boldsymbol{\theta}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{y}_{1:n}$.
4. Generate $\boldsymbol{\lambda} \,|\, \boldsymbol{\theta}_{-\lambda}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{a}_{1:n}, \boldsymbol{y}_{1:n}$.
5. Generate $\boldsymbol{\sigma} \,|\, \boldsymbol{\theta}_{-\sigma}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{a}_{1:n}, \boldsymbol{y}_{1:n}$.
6. Generate $\boldsymbol{\phi} \,|\, \boldsymbol{\theta}_{-\phi}, \boldsymbol{\alpha}_{1:n}, \boldsymbol{a}_{1:n}, \boldsymbol{y}_{1:n}$.
7. Go to 2.

In the sequel each updating step of the McMC algorithm is presented in detail.

Generation of latent volatilities $\boldsymbol{\alpha}_{1:n}$

Step 2. The multi-move or block sampler proposed by Shepard and Pitt (1997) and modified by Watanabe and Omori (2004), which samples from the true conditional distribution, is applied to extract latent volatility $\boldsymbol{\alpha}_{1:n}$. Ishihara and Omori (2012) propose a multi-move sampler for the efficient estimation of a multivariate stochastic volatility model with cross leverage. Their sampler is extended to handle time varying covariances. It is detailed in App. A.

Generation of Cholesky decomposition matrices $\boldsymbol{A}_{t:t+1}$ and $\boldsymbol{H}$

Due to the initial condition, a final Metropolis-Hastings (MH) step (see e.g. Chib and Greenberg, 1995) must be added to sample $\boldsymbol{A}_{1:2}$ and $\boldsymbol{H}$, except for $a_{ij,1}$, $i \leq p$. The latter do not appear in the calculation of $\boldsymbol{\Sigma}_{\text{ini}}$ and can be sampled directly.

Step 3.1 Sampling of $\boldsymbol{a}_{ij,1:n}$, $i \leq p$, and $\boldsymbol{a}_{ij,3:n}$, $i > p$, is a standard application of forward filtering, backward sampling (FFBS) as Eq. (10) and Eq. (12) constitute an ordinary multivariate linear Gaussian state space system.[5] However, sampling the elements of $\boldsymbol{a}_{t:t+1}$ row-wise soon starts to slow down the sampler with an increasing number of assets p. An one-by-one strategy avoids computationally costly multivariate matrix operations and is applied here. Moreover, no significant deterioration of sampler performance appeared using the latter approach.

Step 3.2 Proposals for $a_{ij,1:2}$ time varying, $i > p$, form $a^*_{ij,1:2} \backsim \mathcal{N}(\mu_{a_{ij}}, \sigma^2_{a_{ij}})$, with

$$\begin{aligned}
\mu_{a_{ij}} &= \sigma^2_{a_{ij}}[\lambda_i^{-1} z_{j,1:2}(z_{i,1:2} - \boldsymbol{X}_{i\backslash j,1:2}\boldsymbol{a}_{\backslash ij,1:2}) + \sigma_{ij}^{-2} a_{ij,2:3}],\\
\sigma^{-2}_{a_{ij}} &= \lambda_i^{-1} z^2_{j,1:2} + \sigma^{-2}_{ij,\text{ini}} + \sigma^{-2}_{ij},
\end{aligned}$$

[4] Implementation (also the particle filter in Sec. 4) is in stand alone C++ code developed by the author using the freely available Scythe statistical library (Pemstein, Quinn, and Martin, 2011).

[5] Frühwirth-Schnatter (1994), and Carter and Kohn (1994) independently developed methods for simulation smoothing of the state vector. de Jong and Shephard (1995) proposed refinements by increasing efficiency through sampling the uncorrelated disturbances instead. A yet simpler but computationally highly efficient approach is presented in Durbin and Koopman (2002).

where $z_{j,1:2}$ denotes the j^{th} element of $\boldsymbol{z}_{1:2}$, $\boldsymbol{X}_{i\backslash j,1:2}$ the i^{th} row of $\boldsymbol{X}_{1:2}$ without $z_{j,1:2}$, and $\boldsymbol{a}_{\backslash ij,1:2}$ denotes $\boldsymbol{a}_{1:2}$ without $a_{ij,1:2}$. Further, current proposals are used in the sequential calculations.

Step 3.3 Proposals for a_{ij} time invariant, $i > p$, is a routine application of linear regression theory. The conditional posterior of a_{ij} is

$$\pi(a_{ij}|\cdot) \propto \pi(a_{ij}) \times \exp\left\{-\frac{1}{2\lambda_i}(\boldsymbol{Y}_{a_{ij}} - \boldsymbol{X}_{a_{ij}}a_{ij})'(\boldsymbol{Y}_{a_{ij}} - \boldsymbol{X}_{a_{ij}}a_{ij})\right\},$$

with

$$\begin{aligned}
\boldsymbol{X}_{a_{ij}} &= \begin{pmatrix}\boldsymbol{z}_{j,1:2} & \cdots & \boldsymbol{z}_{j,n-1:n}\end{pmatrix}', \\
\boldsymbol{Y}_{a_{ij}} &= \begin{pmatrix}\boldsymbol{z}_{i,1:2} - \boldsymbol{X}_{i\backslash j,1:2}\boldsymbol{a}_{\backslash ij,1:2} & \cdots & \boldsymbol{z}_{i,n-1:n} - \boldsymbol{X}_{i\backslash j,n-1:n}\boldsymbol{a}_{\backslash ij,n-1:n}\end{pmatrix}'.
\end{aligned}$$

Hence $a^*_{ij}|\cdot \backsim \mathcal{N}(\mu_{a_{ij}}, \sigma^2_{a_{ij}})$, with

$$\mu_{a_{ij}} = \sigma^2_{a_{ij}}(\sigma^{-2}_{a_0,ij}a_{0,ij} + \lambda_i^{-1}\boldsymbol{X}'_{a_{ij}}\boldsymbol{Y}_{a_{ij}}), \quad \sigma^{-2}_{a_{ij}} = \sigma^{-2}_{a_0,ij} + \lambda_i^{-1}\boldsymbol{X}'_{a_{ij}}\boldsymbol{X}_{a_{ij}}.$$

Step 4. To sample $\boldsymbol{\lambda}^*$, let $\boldsymbol{w}_{t:t+1} = \boldsymbol{A}_{t:t+1}\boldsymbol{z}_{t:t+1}$ and $\boldsymbol{w} = \{\boldsymbol{w}_{1:2}, \ldots, \boldsymbol{w}_{n-1:n}, \boldsymbol{w}_n\}$. Further observe that $|\boldsymbol{\Sigma}^{-1}_{t:t+1}| = |\boldsymbol{A}'_{t:t+1}||\boldsymbol{H}^{-1}||\boldsymbol{A}_{t:t+1}| = |\boldsymbol{H}|^{-1} = \prod_{i=1}^q \lambda_i^{-1}$. The conditional likelihood without initial condition can then be expressed as

$$\begin{aligned}
f(\boldsymbol{w}|\cdot) &\propto \left(\prod_{t=1}^{n-1}|\boldsymbol{\Sigma}_{t:t+1}|^{-1/2}\right)\exp\left\{-\frac{1}{2}\sum_{t=1}^{n-1}\boldsymbol{z}'_{t:t+1}\boldsymbol{\Sigma}^{-1}_{t:t+1}\boldsymbol{z}_{t:t+1}\right\}\times \\
&\qquad |\boldsymbol{\Sigma}_n|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{z}'_n\boldsymbol{\Sigma}_n^{-1}\boldsymbol{z}_n\right\} \\
&\propto \left(\prod_{i=1}^q \lambda_i^{-n_i/2}\right)\exp\left\{-\frac{1}{2}\left[\sum_{t=1}^{n-1}\boldsymbol{w}'_{t:t+1}\boldsymbol{H}^{-1}\boldsymbol{w}_{t:t+1} + \boldsymbol{w}'_n\boldsymbol{H}_n^{-1}\boldsymbol{w}_n\right]\right\},
\end{aligned}$$

where $n_i = n$ if $1 \le i \le p$ else $n_i = n-1$, and $\boldsymbol{H}_n = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$. Rewriting the above expression in terms of λ_i, $i = 1, \ldots, q$ yields

$$\begin{aligned}
f(\boldsymbol{w}|\cdot) \propto \left(\prod_{i=1}^q \lambda_i^{-n_i/2}\right)&\exp\left\{-\frac{1}{2}\left[\operatorname{tr}\left(\boldsymbol{H}^{-1}\sum_{t=1}^{n-1}\boldsymbol{w}_{t:t+1}\boldsymbol{w}'_{t:t+1}\right)+\right.\right. \\
&\left.\left.\operatorname{tr}\left(\boldsymbol{H}_n^{-1}\boldsymbol{w}_n\boldsymbol{w}'_n\right)\right]\right\} = \prod_{i=1}^q \lambda_i^{-n_i/2}\exp\left\{-\frac{r_i + s_i}{2\lambda_i}\right\},
\end{aligned}$$

where r_i denotes the (i,i)-element of $\sum_{t=1}^{n-1}\boldsymbol{w}_{t:t+1}\boldsymbol{w}'_{t:t+1}$, and s_i the (i,i)-element of $\boldsymbol{w}_n\boldsymbol{w}'_n$ if $1 \le i \le p$ else $s_i = 0$. Hence, the conditional posterior of $\boldsymbol{\lambda}$ is

$$\pi(\boldsymbol{\lambda}|\cdot) \propto \pi(\boldsymbol{\lambda}) \times |\boldsymbol{\Sigma}_{\text{ini}}|^{-1/2}\exp\left\{-\frac{1}{2}\boldsymbol{\alpha}'_1\boldsymbol{\Sigma}^{-1}_{\text{ini}}\boldsymbol{\alpha}_1\right\} \times \prod_{i=1}^q \lambda_i^{-n_i/2}\exp\left\{-\frac{r_i + s_i}{2\lambda_i}\right\}.$$

Under the inverse gamma prior of Eq. (13), proposals λ^*_i are sampled independently,

$$\lambda^*_i|\cdot \backsim \mathcal{IG}\left(\frac{\nu_{0,i} + n_i}{2}, \frac{\delta_{0,i} + r_i + s_i}{2}\right), \quad i = 1, \ldots, q.$$

Step 5. A final MH step must be applied. Proposals are $\{(a^*_{ij,1:2} \vee a^*_{ij})|i > p, \boldsymbol{\lambda}^*\}$. These are accepted with probability

$$\min\left\{\frac{|\boldsymbol{\Sigma}_{\text{ini}}|^{-1/2^*}\exp\{-\frac{1}{2}\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\text{ini}}^{-1^*}\boldsymbol{\alpha}_1\}}{|\boldsymbol{\Sigma}_{\text{ini}}|^{-1/2}\exp\{-\frac{1}{2}\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\text{ini}}^{-1}\boldsymbol{\alpha}_1\}}, 1\right\},$$

where $\boldsymbol{\Sigma}_{\text{ini}}$ is the current unconditional variance of $\boldsymbol{\alpha}_1$.

Generation of $\boldsymbol{\sigma}^2$ governing the random walk dynamics of $\boldsymbol{a}_{1:n}$

Step 6. Random walk variances $\sigma^2_{ij} \in \boldsymbol{\sigma}^2$ are independent and sampled one-by-one using standard results of linear regression theory. The conditional posterior probability of σ^2_{ij} is

$$\begin{aligned}
\pi(\sigma^2_{ij}|\cdot) &\propto \pi(\sigma^2_{ij}) \times (\sigma^2_{ij})^{-n_i/2}\exp\left\{-\sum_{t=k_i}^{n-1}\frac{(a_{ij,t+1}-a_{ij,t})^2}{2\sigma^2_{ij}}\right\} \\
&= (\sigma^2_{ij})^{-b_{ij}/2-1}\exp\left\{-\frac{c}{2\sigma^2_{ij}}\right\},
\end{aligned}$$

with

$$b_{ij} = b_{0,ij} + n_i, \quad c_{ij} = c_{0,ij} + \sum\nolimits_{t=k_i}^{n-1}(a_{ij,t+1}-a_{ij,t})^2,$$

where $n_i = n-1 \wedge k_i = 1$ if $1 \leq i \leq p$ else $n_i = n-2 \wedge k_i = 2$, and $\sigma_{ij,1} = \sigma_{ij,\text{ini}}$ known. Hence, $\sigma^2_{ij}|\cdot \backsim \mathcal{IG}(b_{ij}/2, c_{ij}/2)$.

Generation of volatility persistence $\boldsymbol{\Phi}$

Step 7. The conditional posterior probability of $\boldsymbol{\Phi}$ does not have closed form, and sampling is by the MH algorithm. Let $\boldsymbol{\Sigma}^{ij}_{t:t+1}$ be a $p \times p$ matrix that denotes the (i,j)-block of $\boldsymbol{\Sigma}^{-1}_{t:t+1}$. The posterior of $\boldsymbol{\Phi}$ can then be written as

$$\pi(\boldsymbol{\Phi}|\cdot) \propto g(\boldsymbol{\Phi}) \times \exp\left\{-\frac{1}{2}(\boldsymbol{\Phi}-\boldsymbol{\mu}_{\boldsymbol{\Phi}})'\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}^{-1}(\boldsymbol{\Phi}-\boldsymbol{\mu}_{\boldsymbol{\Phi}})\right\},$$

with

$$\begin{aligned}
\boldsymbol{\mu}_{\boldsymbol{\Phi}} &= \boldsymbol{\Sigma}_{\boldsymbol{\Phi}} \times \underline{\text{diag}}\left(\sum\nolimits_{t=1}^{n-1}\left\{\boldsymbol{\alpha}_t\boldsymbol{y}_t'\boldsymbol{V}_t^{-1/2}\boldsymbol{\Sigma}^{12}_{t:t+1} + \boldsymbol{\alpha}_t\boldsymbol{\alpha}_{t+1}'\boldsymbol{\Sigma}^{22}_{t:t+1}\right\}\right), \\
\boldsymbol{\Sigma}_{\boldsymbol{\Phi}}^{-1} &= \sum\nolimits_{t=1}^{n-1}\boldsymbol{\Sigma}^{22}_{t:t+1} \odot \boldsymbol{\alpha}_t\boldsymbol{\alpha}_t', \\
g(\boldsymbol{\Phi}) &= \pi(\boldsymbol{\Phi}) \times |\boldsymbol{\Sigma}_{\text{ini}}|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}\boldsymbol{\alpha}_{\mathbf{1}}'\boldsymbol{\Sigma}_{\text{ini}}^{-1}\boldsymbol{\alpha}_{\mathbf{1}}\right\},
\end{aligned}$$

where $\underline{\text{diag}}(\cdot)$ returns the diagonal as column vector, and $\odot$ denotes the Hadamard product. Then propose a candidate from the multivariate normal distribution truncated over the region $R = \{\boldsymbol{\Phi} : |\phi_i| < 1, i = 1,\dots,p\}$, $\boldsymbol{\Phi}^* \backsim \mathcal{TN}_R(\boldsymbol{\mu}_{\boldsymbol{\Phi}}, \boldsymbol{\Sigma}_{\boldsymbol{\Phi}})$, and accept it with probability $\min\{\text{g}(\boldsymbol{\Phi}^*)/\text{g}(\boldsymbol{\Phi}), 1\}$, where $\boldsymbol{\Phi}$ is the current sample.

4 Particle Filter

An auxiliary particle filter (Pitt and Shephard, 1999a) is developed that recursively delivers draws of the unobservable volatilities $\boldsymbol{\alpha}_t$, given parameter values $\boldsymbol{\theta}$ and observables up to time t, $\boldsymbol{y}_{1:t}$. These can be utilized for likelihood evaluation, goodness-of-fit statistics, and forecasting. The auxiliary particle filter is a sequential Monte Carlo (SMC) technique that increases efficiency in the propagation process by weighting particles according to an importance function dependent on the subsequent observation. A key characteristic of the proposed filter is that, given draws of latent volatilities $\boldsymbol{\alpha}_t$, latent covariance dynamics $\boldsymbol{a}_t$ can be extracted in analytical form using the standard Kalman Filter prediction and updating equations. For more information on particle filtering, consider e.g. Doucet, de Freitas, and Gordon (2001), or Whiteley and Johansen (2011).

Observe that the model as defined in Eq. (5)-(7) can be reformulated having measurement density

$$f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{V}_t^{1/2}\boldsymbol{\Sigma}_{\epsilon\epsilon,t}\boldsymbol{V}_t^{1/2})$$

and transition density

$$f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\Sigma}_{t:t+1}, \boldsymbol{\alpha}_t, \boldsymbol{y}_t) = \mathcal{N}(\boldsymbol{\mu}_{\alpha,t+1}, \boldsymbol{\Sigma}_{\alpha,t+1}), \tag{14}$$

where

$$\begin{aligned}\boldsymbol{\mu}_{\alpha,t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\boldsymbol{\Sigma}_{\epsilon\epsilon,t}^{-1}\boldsymbol{V}_t^{-1/2}\boldsymbol{y}_t,\\ \boldsymbol{\Sigma}_{\alpha,t+1} &= \boldsymbol{\Sigma}_{\eta\eta,t+1} - \boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\boldsymbol{\Sigma}_{\epsilon\epsilon,t}^{-1}\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\end{aligned}$$

(for clarity, time invariant parameter vector $\boldsymbol{\theta}$ may be dropped from conditioning arguments in this section). Applying Bayes Theorem we obtain the target posterior density,

$$f(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_t|\boldsymbol{y}_{1:t+1}) \propto f(\boldsymbol{y}_{t+1}|\boldsymbol{\alpha}_{t+1})f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t, \boldsymbol{y}_t)f(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:t}), \tag{15}$$

where we assume that we have samples (particles) from $f(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:t})$ and a discrete uniform approximation $\hat{f}(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:t})$ to $f(\boldsymbol{\alpha}_t|\boldsymbol{y}_{1:t})$. Latent covariance dynamics $\boldsymbol{a}_{t:t+1}$ appear in linear Gaussian state space form and have been integrated out of Eq. (15) analytically.

More precisely, measurement equation (10) is treated row-wise, having density

$$f(z_{i,t:t+1}|\boldsymbol{a}_{i,t:t+1}) = \mathcal{N}(\boldsymbol{X}'_{i,t:t+1}\boldsymbol{a}_{i,t:t+1}, \lambda_i), \quad i = 2, \ldots, q,$$

with $\boldsymbol{a}_{i,t:t+1}$ a $i-1\times 1$ column vector containing the i^{th} row of the lower triangular of matrix $\boldsymbol{A}_{t:t+1}$, and $\boldsymbol{X}_{i,t:t+1}$ a column vector of same dimension stacking all nonzero elements of the i^{th} row of $\boldsymbol{X}_{t:t+1}$. The corresponding sequential learning process has priors (prediction)

$$\breve{\boldsymbol{a}}_{i,t:t+1} = \boldsymbol{a}_{i,t:t-1}, \tag{16}$$

$$\breve{\boldsymbol{P}}_{i,t:t+1} = \boldsymbol{P}_{i,t:t-1} + \boldsymbol{Q}_i, \tag{17}$$

with $\boldsymbol{P}_{i,t:t-1}$ the error covariance of $\mathrm{E}(\boldsymbol{a}_{i,t:t-1}|z_{i,t:t-1})$, and diagonal process error covariance matrix $\boldsymbol{Q} = \mathrm{diag}(\sigma^2_{i1}, \ldots, \sigma^2_{i,i-1})$. Posteriors (correction) are

$$\boldsymbol{a}_{i,t:t+1} = \breve{\boldsymbol{a}}_{i,t:t+1} + \boldsymbol{K}_{i,t:t+1}(z_{i,t:t+1} - \boldsymbol{X}'_{i,t:t+1}\breve{\boldsymbol{a}}_{i,t:t+1}), \tag{18}$$

$$\boldsymbol{P}_{i,t:t+1} = (\boldsymbol{I}_{i-1} - \boldsymbol{K}_{i,t:t+1}\boldsymbol{X}'_{i,t:t+1})\breve{\boldsymbol{P}}_{i,t:t+1}, \tag{19}$$

where $\boldsymbol{K}_{i,t:t+1}$ is the Kalman gain,

$$\boldsymbol{K}_{i,t:t+1} = \breve{\boldsymbol{P}}_{i,t:t+1}\boldsymbol{X}_{i,t:t+1}(\boldsymbol{X}'_{i,t:t+1}\breve{\boldsymbol{P}}_{i,t:t+1}\boldsymbol{X}_{i,t:t+1} + \lambda_i)^{-1}.$$

Note that Eq. (16), (17) and (18), (19) are optimal in the mean square error sense. Further, if elements of $\boldsymbol{a}_{i,t:t+1}$ are kept constant over time, the Kalman filter operates on the associated lower dimensional system only.

By including $\boldsymbol{y}_{t+1}$ in the importance probability density function more weight is given to particles with stronger predictive power. The importance function that is used to sample from Eq. (15) then has the following form (superscripts $k = 1, \ldots, K$ denote particles),

$$\begin{aligned} g(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_t^{(k)}|\boldsymbol{y}_{1:t+1}) \propto f(\boldsymbol{y}_{t+1}|\breve{\boldsymbol{\mu}}^{(k)}_{\alpha,t+1}, \breve{\boldsymbol{a}}^{(k)}_{t+1}, \breve{\boldsymbol{a}}^{(k)}_{t:t+1}) \times \\ f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t^{(k)}, \boldsymbol{y}_t)\hat{f}(\boldsymbol{\alpha}_t^{(k)}|\boldsymbol{y}_{1:t}), \end{aligned} \tag{20}$$

with

$$\begin{aligned} \breve{\boldsymbol{V}}^{(k)}_{t+1} &= \boldsymbol{V}_{t+1}\big|_{\boldsymbol{\alpha}_{t+1}=\breve{\boldsymbol{\mu}}^{(k)}_{\alpha,t+1}}, \\ \breve{\boldsymbol{\Sigma}}^{(k)}_{\epsilon\epsilon,t+1} &= \boldsymbol{\Sigma}_{\epsilon\epsilon,t+1}\big|_{\boldsymbol{a}_{i,t+1}=\breve{\boldsymbol{a}}^{(k)}_{i,t+1},\ i=2,\ldots,p}, \end{aligned}$$

and

$$\begin{aligned} \breve{\boldsymbol{\mu}}^{(k)}_{\alpha,t+1} &= \boldsymbol{\Phi}\boldsymbol{\alpha}_t^{(k)} + \breve{\boldsymbol{\Sigma}}^{(k)}_{\eta\epsilon,t:t+1}\boldsymbol{\Sigma}^{(k)-1}_{\epsilon\epsilon,t}\boldsymbol{V}_t^{(k)-1/2}\boldsymbol{y}_t, \\ \breve{\boldsymbol{\Sigma}}^{(k)}_{\alpha,t+1} &= \breve{\boldsymbol{\Sigma}}^{(k)}_{\eta\eta,t+1} - \breve{\boldsymbol{\Sigma}}^{(k)}_{\eta\epsilon,t:t+1}\boldsymbol{\Sigma}^{(k)-1}_{\epsilon\epsilon,t}\breve{\boldsymbol{\Sigma}}^{(k)}_{\epsilon\eta,t:t+1}, \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{V}_t^{(k)} &= \boldsymbol{V}_t\big|_{\boldsymbol{\alpha}_t=\boldsymbol{\alpha}_t^{(k)}}, \\ \boldsymbol{\Sigma}^{(k)}_{\epsilon\epsilon,t} &= \boldsymbol{\Sigma}_{\epsilon\epsilon,t}\big|_{\boldsymbol{a}_{i,t}=\boldsymbol{a}^{(k)}_{i,t},\ i=2,\ldots,p}, \\ \breve{\boldsymbol{\Sigma}}^{(k)}_{\eta\epsilon,t:t+1} &= \boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\big|_{\boldsymbol{a}_{i,t}=\boldsymbol{a}^{(k)}_{i,t},\ i=2,\ldots,p,\ \boldsymbol{a}_{i,t:t+1}=\breve{\boldsymbol{a}}^{(k)}_{i,t:t+1},\ i=p+1,\ldots,q}, \end{aligned}$$

$\breve{\boldsymbol{\Sigma}}^{(k)}_{\epsilon\eta,t:t+1}$, $\breve{\boldsymbol{\Sigma}}^{(k)}_{\eta\eta,t+1}$ in analogy to $\breve{\boldsymbol{\Sigma}}^{(k)}_{\eta\epsilon,t:t+1}$, and

$$\begin{aligned} \breve{\boldsymbol{a}}^{(k)}_{i,t+1} &= \boldsymbol{a}^{(k)}_{i,t}, & \boldsymbol{a}^{(k)}_{i,t} &= \boldsymbol{a}_{i,t}\big|_{\boldsymbol{\alpha}_t=\boldsymbol{\alpha}_t^{(k)},\ \breve{\boldsymbol{a}}_{i,t}=\breve{\boldsymbol{a}}^{(k)}_{i,t}}, & i &= 2, \ldots, p, \\ \breve{\boldsymbol{a}}^{(k)}_{i,t:t+1} &= \boldsymbol{a}^{(k)}_{i,t-1:t}, & \boldsymbol{a}^{(k)}_{i,t-1:t} &= \boldsymbol{a}_{i,t-1:t}\big|_{\boldsymbol{\alpha}_t=\boldsymbol{\alpha}_t^{(k)},\ \breve{\boldsymbol{a}}_{i,t-1:t}=\breve{\boldsymbol{a}}^{(k)}_{i,t-1:t}}, & i &= p+1, \ldots, q. \end{aligned}$$

Note that in Eq. (20) and in the sequel, as there is a one-to-one relationship between $\boldsymbol{\alpha}_t^{(k)}$ and $\boldsymbol{a}^{(k)}_{i,t}$, $i = 2, \ldots, q$, the latter may be omitted in the arguments to not clutter notation. Moreover, note that as the initial states of the random walk dynamics $\boldsymbol{a}_t$ are centered at zero, an appropriate convergence period for the particle filter must be chosen.

The proposed implementation follows:

1. Initialization, $t = 1$ $(k = 1, \ldots, K)$:

   (a) Calculate $\boldsymbol{\Sigma}_{0:1}$ by setting random walk elements $a_{ij,0:1} = 0$, $j \leq p$, and compute corresponding $\boldsymbol{\Sigma}_{\text{ini}}$. Then generate $\boldsymbol{\alpha}_1^{(k)} \backsim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\text{ini}})$.

   (b) Initialize $\breve{\boldsymbol{a}}_{i,1}^{(k)} = \mathbf{0}$, $\breve{\boldsymbol{P}}_{i,1} = \text{diag}(\boldsymbol{I}_{i-1})$, and update $\boldsymbol{a}_{i,1}^{(k)} = \boldsymbol{a}_{i,1}|_{\boldsymbol{\alpha}_1 = \boldsymbol{\alpha}_1^{(k)},\ \breve{\boldsymbol{a}}_{i,1} = \breve{\boldsymbol{a}}_{i,1}^{(k)}}$, $i = 2, \ldots, p$.

   (c) Compute $w_1^{(k)} = f(\boldsymbol{y}_1|\boldsymbol{\alpha}_1^{(k)})$, and let $\hat{f}(\boldsymbol{\alpha}_1^{(k)}|\boldsymbol{y}_1) = w_1^{(k)} / \sum_{l=1}^K w_1^{(l)}$.

   (d) Initialize $\breve{\boldsymbol{a}}_{i,1:2}^{(k)} = \mathbf{0}$, $\breve{\boldsymbol{P}}_{i,1:2} = \text{diag}(0.1^2 \boldsymbol{I}_{i-1})$, $i = p+1, \ldots, q$.

2. Iterate, $t = 1, \ldots, n-1$ $(k, l = 1, \ldots, K)$:

   (a) Generate $\{\boldsymbol{\alpha}_{t+1}^{(k)}, \boldsymbol{\alpha}_t^{(k)}\}$ from $g(\boldsymbol{\alpha}_{t+1}, \boldsymbol{\alpha}_t^{(l)}|\boldsymbol{y}_{1:t+1})$, with $\boldsymbol{\alpha}_t^{(l)}$ the current particle set at time t, as follows. First, evaluate importance weights (pre-weighting)

   $$w_t^{(l)} \propto f(\boldsymbol{y}_{t+1}|\breve{\boldsymbol{\mu}}_{\alpha,t+1}^{(l)}, \breve{\boldsymbol{a}}_{t+1}^{(l)}, \breve{\boldsymbol{a}}_{t:t+1}^{(l)}) \hat{f}(\boldsymbol{\alpha}_t^{(l)}|\boldsymbol{y}_{1:t}),$$

   and resample $\{\boldsymbol{\alpha}_t^{(l)}, w_t^{(l)}\}$ to get $\{\boldsymbol{\alpha}_t^{(k)}, K^{-1}\}$. Then generate $\boldsymbol{\alpha}_{t+1}^{(k)}$ from $f(\boldsymbol{\alpha}_{t+1}|\boldsymbol{\alpha}_t^{(k)}, \boldsymbol{y}_t)$ given in Eq. (14), and update
   $\boldsymbol{a}_{i,t+1}^{(k)} = \boldsymbol{a}_{i,t+1}|_{\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_{t+1}^{(k)},\ \breve{\boldsymbol{a}}_{i,t+1} = \breve{\boldsymbol{a}}_{i,t+1}^{(k)}}$, $i = 2, \ldots, p$,
   $\boldsymbol{a}_{i,t:t+1}^{(k)} = \boldsymbol{a}_{i,t:t+1}|_{\boldsymbol{\alpha}_{t+1} = \boldsymbol{\alpha}_{t+1}^{(k)},\ \breve{\boldsymbol{a}}_{i,t:t+1} = \breve{\boldsymbol{a}}_{i,t:t+1}^{(k)}}$, $i = p+1, \ldots, q$.

   (b) Compute weights (correcting for the pre-weighting)

   $$w_{t+1}^{(k)} = \frac{f(\boldsymbol{y}_{t+1}|\boldsymbol{\alpha}_{t+1}^{(k)}, \boldsymbol{a}_{t+1}^{(k)}, \boldsymbol{a}_{t:t+1}^{(k)})}{f(\boldsymbol{y}_{t+1}|\breve{\boldsymbol{\mu}}_{\alpha,t+1}^{(k)}, \breve{\boldsymbol{a}}_{t+1}^{(k)}, \breve{\boldsymbol{a}}_{t:t+1}^{(k)})},$$

   and let $\hat{f}(\boldsymbol{\alpha}_{t+1}^{(k)}|\boldsymbol{y}_{t+1}) = w_{t+1}^{(k)} / \sum_{l=1}^K w_{t+1}^{(l)}$.

Draws $\boldsymbol{\alpha}_{t|t}^{(k)}$, which are utilized to calculate posterior log-likelihood ordinate $\log f_{\text{post}}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}) = \sum_{t=1}^n \log \hat{f}(\boldsymbol{y}_t|\boldsymbol{y}_{1:t}, \boldsymbol{\theta})$, can be obtained by resampling from $\hat{f}(\boldsymbol{\alpha}_t^{(k)}|\boldsymbol{y}_t)$ after step 1.(c) and 2.(b). These particles are propagated forward in time to obtain $\boldsymbol{\alpha}_{t+1|t}^{(k)}$, by first calculating priors as given by Eq. (16) and (17), and then sampling from Eq. (14), to obtain an estimate of the one-step-ahead prediction density $f(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta})$, which in turn is needed for forecasting and calculation of predictive log-likelihood ordinate $\log f_{\text{prior}}(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}) = \sum_{t=1}^n \log \hat{f}(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \boldsymbol{\theta})$.

5 Empirical Application

5.1 Data

Descriptive statistics of the dataset are presented. In Sec. 5.2 and 5.3 estimation results of the MSV model with constant correlation and dynamic correlation are discussed, respectively. The dataset consists of three S&P 500 sector indices,

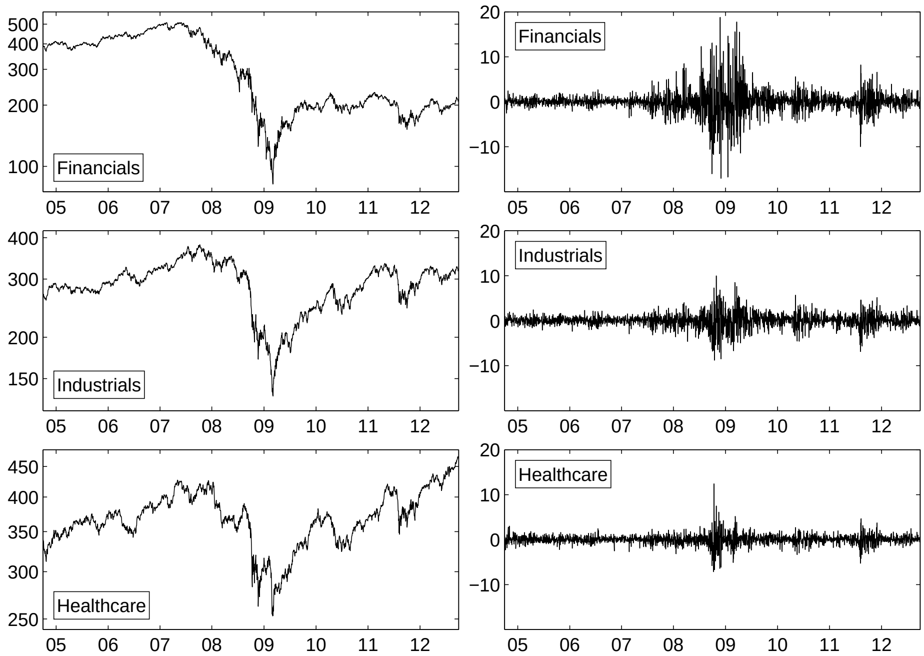

Figure 1: Price (log-scale) and Return (%) of S&P 500 Sector Indices

Table 1: Descriptive Statistics (%-returns) (10/2004-09/2012, 2,015 obs.)

	Financials	Industrials	Healthcare	S&P 500
Mean	0.0017	0.0209	0.0227	0.0224
Stdev	2.5172	1.5180	1.1195	1.3922
Skew	0.4215	-0.2130	0.2341	-0.0548
Kurtosis	14.731	8.4833	16.261	12.964
Min	-17.005	-8.8032	-7.1469	-9.0350
Max	18.769	9.9840	12.427	11.580

namely Financials (series 1), Industrials (series 2) and Healthcare (series 3), obtained from Thomson Reuters Datastream. The choice was motivated by Industrials and Financials being main pillars of the economy, supposed to exhibit strong interdependence, and Healthcare having distinct characteristics due to being countercyclical. The period analyzed covers 8 years from 2004/10/01 to 2012/09/28, a total of $2{,}015$ observations. Market holidays were excluded. Returns are calculated as $(P_{t+1}/P_t - 1) \times 100$, where P_t denotes the asset price at time t. Further, the data is demeaned and standardized by its sample standard deviation. Working on a common scale facilitates prior choice, especially for the dynamic correlation model using Cholesky decomposition, however it is not necessary.

Descriptive statistics are given in Tab. 1, including the S&P 500 index as reference. Financials clearly is the most volatile sector, while Healthcare is the least volatile. Notably, the largest return in absolute value is positive for all sector indices considered and the S&P 500, despite the period under analysis containing the financial crisis 2008/09. A quite strong observed positive skew for Financials and less so for Healthcare complements the picture. As usual, all time series show kurtosis larger than 3, with Financials and Healthcare having the heaviest tails, the latter influenced largely by one extraordinary positive outlier 2008/10/13. Excluding this outlier would actually turn the sample skewness of Healthcare the most negative. Fig. 1 visualizes distinct volatility characteristics of the series. Comovement of volatility among the sector indices is clearly visible and most pronounced during the 2008/09 financial crisis with significantly increased volatility in all three sectors. Further, volatility surges following the rating downgrade of U.S. Treasury bonds from their AAA status by Standard & Poor's in early August 2011. The latter extended period of financial turmoil may be explained with deepening global macro dislocations like a continuing European debt crisis and the prospect of a slowing global economy. Note also that Financials did not recover to previous pricing levels, in contrast to Industrials and Healthcare.

5.2 The MSVcc Model (Constant Correlation)

This section reports estimation results of the MSV model with constant correlation and cross leverage proposed by Ishihara and Omori (2012) (henceforth MSVcc), which serves as a baseline for the more complex MSVdc model with dynamic correlation and cross leverage. The afore mentioned authors apply the model to a five-dimensional portfolio of S&P 500 sector indices covering a slightly different time span, but estimation results may be compared with those obtained here. The MSVcc model is recovered from the MSVdc model as defined in Eq. (5)-(12) by replacing time varying covariance matrix $\boldsymbol{\Sigma}_{t:t+1}$ in Eq. (7) with time invariant

$$\boldsymbol{\Sigma}_{\mathrm{MSVcc}} = \begin{pmatrix} \boldsymbol{\Sigma}_{\epsilon\epsilon} & \boldsymbol{\Sigma}_{\epsilon\eta} \\ \boldsymbol{\Sigma}_{\eta\epsilon} & \boldsymbol{\Sigma}_{\eta\eta} \end{pmatrix}.$$

Accordingly, the proposed McMC algorithm of Sec. 3 simplifies. Specifically, Steps 3 to 5 are replaced by an appropriate sampling step as outlined in Sec. C of the Appendix. Then, the following priors are assumed,

$$\frac{\phi_i + 1}{2} \backsim \mathcal{B}(20, 1.5), \quad i = 1, \ldots, p,$$
$$\boldsymbol{\Sigma}^{-1} \backsim \mathcal{W}(10, (10\boldsymbol{\Sigma}_0)^{-1}), \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} 0.4^2 \boldsymbol{I}_p & \boldsymbol{0} \\ \boldsymbol{0} & 0.15^2 \boldsymbol{I}_p \end{pmatrix},$$

where $\mathcal{B}(\cdot)$ denotes the beta and $\mathcal{W}(\cdot)$ the Wishart distribution. The prior on ϕ_i implies a belief in moderate persistence with mean 0.86 and standard deviation 0.11. The prior on $\boldsymbol{\Sigma}^{-1}$ conservatively assumes zero correlation and is chosen such that $\mathrm{E}(\boldsymbol{\Sigma}^{-1}) = \boldsymbol{\Sigma}_0^{-1}$. Beliefs in volatility of volatility of the latent processes, expressed by the lower block diagonal of $\boldsymbol{\Sigma}_0$, are rather standard. However, the specific prior choice for the volatility modes in the upper diagonal block matrix of $\boldsymbol{\Sigma}_0$ demands some explanation. A unit prior for the standardized data would be an obvious choice, but did not yield satisfactory estimation results for the specific dataset at hand. Specifically, a unit prior led to overestimation of the volatility modes, offset by latent volatility processes with unconditional posterior sample averages considerably below their model theoretic value of zero. Hence the choice of a smaller prior to pull the sample averages of $\boldsymbol{\alpha}_{1:n}$ closer to zero, where 'close' is of course subject to the researcher's opinion, say a threshold of approximately 10^{-1}. One may choose a larger sample size to alleviate such model fitting problems and associated prior tuning, but this would increase the computational burden. Moreover, regarding forecasting, one

Table 2: Estimation Results of the MSVcc Model (constant correlation)

	Series i/ij	Mean	Stdev	95% c.b.	IF	CD
	1	0.9818	0.0036	[0.9743, 0.9883]	18.7	0.99
ϕ_i	2	0.9745	0.0048	[0.9643, 0.9831]	22.0	0.97
	3	0.9631	0.0066	[0.9493, 0.9751]	21.2	0.97
	1	0.5814	0.0503	[0.4876, 0.6858]	170.4	0.94
$\sigma_{\epsilon,i}$	2	0.7537	0.0468	[0.6672, 0.8519]	100.7	0.90
	3	0.7574	0.0368	[0.6861, 0.8309]	56.1	0.95
	1	0.2045	0.0188	[0.1707, 0.2435]	113.0	0.97
$\sigma_{\eta,i}$	2	0.1865	0.0183	[0.1542, 0.2243]	121.0	0.94
	3	0.2080	0.0187	[0.1727, 0.2467]	104.4	0.96
	12	0.8320	0.0073	[0.8175, 0.8459]	1.6	0.98
$\rho_{\epsilon\epsilon,ij}$	13	0.7236	0.0113	[0.7009, 0.7452]	1.7	0.97
	23	0.7656	0.0100	[0.7456, 0.7847]	1.8	0.95
	12	0.8282	0.0350	[0.7518, 0.8878]	76.2	0.95
$\rho_{\eta\eta,ij}$	13	0.7507	0.0503	[0.6400, 0.8357]	76.6	0.83
	23	0.8050	0.0409	[0.7142, 0.8734]	82.2	0.92
	11	-0.4901	0.0663	[-0.6117, -0.3550]	54.0	0.82
	12	-0.4772	0.0677	[-0.6037, -0.3410]	48.9	0.80
	13	-0.5142	0.0659	[-0.6353, -0.3769]	55.0	0.78
	21	-0.4956	0.0667	[-0.6207, -0.3610]	56.6	0.84
$\rho_{\epsilon\eta,ij}$	22	-0.5437	0.0626	[-0.6605, -0.4176]	51.3	0.87
	23	-0.5539	0.0641	[-0.6716, -0.4219]	64.9	0.86
	31	-0.4528	0.0725	[-0.5877, -0.3074]	61.9	0.77
	32	-0.4401	0.0702	[-0.5733, -0.2995]	50.0	0.91
	33	-0.5082	0.0668	[-0.6314, -0.3718]	60.3	0.89

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value).

may feel more comfortable working with a dataset closer to the actual state. The proposed approach is coherent and systematic using pilot runs.

The first $10{,}000$ draws are discarded as burn-in, collecting the next $40{,}000$ draws for parameter inference. In the multi-move sampler, average block size as a tuning parameter is set to 20 and number of iterations to achieve convergence to 5. Posterior means, standard deviations, 95% credible bounds, p-values of Geweke's (1992) convergence diagnostic (CD), and inefficiency factors (IF) (see App. D for the deployed formulae) are reported. For all models estimated in this work, the null of the CD statistic ('chain has converged') can not be rejected at conventional significance levels for all parameters.

Turning to results in Tab. 2, persistence ϕ_i is high for all processes and highest for Financials. Higher volatility persistence for Financials is in accordance with this sector being the most volatile.[6] Interestingly, volatility mode $\sigma_{\epsilon,1}$ of Financials features the lowest value, but the data is standardized. Moreover, the high persistence of Financials' latent volatility makes it more difficult to extract the volatility mode, as mirrored by a higher inefficiency factor relative to Industrials and Healthcare.

[6] Jacquier and Miller (2012) estimate latent volatility of the S&P 500 index around the financial crisis using a sequential Monte Carlo (SMC) approach. They make volatility persistence parameter ϕ time varying and find increased persistence during the turmoil, high volatility period.

Ephemeral observation error $\boldsymbol{\epsilon}_t$ and transition error $\boldsymbol{\eta}_{t+1}$ both feature high positive correlations $\boldsymbol{\rho}_{\epsilon\epsilon}$ and $\boldsymbol{\rho}_{\eta\eta}$, respectively. Leverage and cross leverage effects are significant and moderate in size. Note that cross leverage effects to Healthcare, the least volatile asset, are highest, albeit not significantly. Ishihara and Omori (2012) further estimate univariate versions of the model and report stronger leverage $\rho_{\epsilon\eta,i}$, arguing that in case of the multivariate model, part of the variation of one series may be explained by those of other series through high correlations among $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_{t+1}$.

5.3 The MSVdc Model (Dynamic Correlation)

This section fits the MSVdc model, featuring dynamic correlation and cross leverage. Database remains the same, namely the S&P 500 sector indices Financials (series 1), Industrials (series 2) and Healthcare (series 3).[7] With 3 assets there are 3 latent volatility processes $\boldsymbol{\alpha}_t$ and 12 latent processes governing the covariance dynamics of $\boldsymbol{\Sigma}_{t:t+1}$ (the elements of $\boldsymbol{a}_{t:t+1}$ modeling the covariance dynamics of $\boldsymbol{\Sigma}_{\eta\eta,t+1}$ are kept constant for parsimony). With such many latent parameters, prior selection inevitably becomes important in the model fitting procedure. Through a careful choice of latent process variances σ_{ij}^2, Eq. (12), we are able to control the smoothness of $\boldsymbol{a}_{t:t+1}$ and consequently, that of $\boldsymbol{\Sigma}_{t:t+1}$. If we choose $\boldsymbol{\Sigma}_{t:t+1}$ too smooth, we risk to ignore potentially important dynamics. On the other hand, a $\boldsymbol{\Sigma}_{t:t+1}$ that is too adaptive may chase spurious noise in the data. Moreover, too diffuse dynamics for $\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}$ have the undesirable side effect of decreasing persistence $\boldsymbol{\Phi}$ of latent volatility, overfitting the model further. Such a sensibility to prior choice may be perceived as a disadvantage. However, with a careful and systematic tuning approach, these interdependencies allow the researcher to improve predictive performance.

The proposed approach to prior selection is explained in more detail. First, priors on the diagonal elements of $\boldsymbol{\Phi}$ are as in Sec. 5.2. For remaining parameters, the following priors are assumed,

$$\begin{aligned}\lambda_i &\backsim \mathcal{IG}(5/2, 3/2), \quad i = 1,\ldots,p, \quad \lambda_i \backsim \mathcal{IG}(5/2, 0.05/2), \quad i = p+1,\ldots,q,\\ \sigma_{ij}^2 &\backsim \mathcal{IG}(5/2, 0.005/2), \quad i \le p, \quad \sigma_{ij}^2 \backsim \mathcal{IG}(5/2, 0.000001/2), \quad i > p,\end{aligned} \tag{21}$$

and for a_{ij} fix ($i > p$, $j > p$)

$$a_{ij} \backsim \mathcal{N}(0, 10^{12}).$$

The priors for λ_i, $i = 1,\ldots,p$, have mean 1.00 and standard deviation 1.41, making them a direct choice for the standardized data as they imply unit variance for the observation error in the case of no correlation. Those for transition error related λ_i, $i = p+1,\ldots,q$, imply volatilities with mean 0.12 and standard deviation 0.05 in the case of zero correlation, and are rather standard in the literature (moments obtained by MC simulation). Time invariant elements a_{ij} of the lower triangular matrix $\boldsymbol{A}_{t:t+1}$ receive noninformative priors, centered around zero.

The priors for σ_{ij}^2, $i \le p$, imply volatilities with mean 0.038 and standard deviation 0.016. Those for σ_{ij}^2, $i > p$, a mean of 0.00053 and standard deviation of 0.00022 (all moments obtained by MC simulation). They control the smoothness of $\boldsymbol{\Sigma}_{t:t+1}$ and are regarded as tuning parameters. Accordingly, we may want

[7] Note that ordering of the assets matters when deploying Cholesky-type decomposition and is according to unconditional sample volatility in decreasing order. Reverse ordering has been tested, but yielded a worse in-sample fit according to the deviance information criterion (DIC).

Table 3: Estimation Results of the MSVdc Model (dynamic correlation)

	Series i/ij	Mean	Stdev	95% c.b.	IF	CD
	1	0.9810	0.0036	[0.9735, 0.9877]	22.6	0.89
ϕ_i	2	0.9728	0.0050	[0.9622, 0.9818]	32.2	0.94
	3	0.9625	0.0074	[0.9466, 0.9755]	36.2	0.90
	1	0.2625	0.0521	[0.1763, 0.3780]	244.6	0.91
$\lambda_{\epsilon,i}$	2	0.1558	0.0182	[0.1234, 0.1940]	111.6	0.93
	3	0.2097	0.0215	[0.1711, 0.2560]	74.2	1.00
	1	0.0412	0.0079	[0.0278, 0.0586]	114.3	0.85
$\lambda_{\eta,i}$	2	0.0062	0.0015	[0.0038, 0.0098]	167.1	0.96
	3	0.0090	0.0024	[0.0052, 0.0146]	195.5	0.76
	21	-0.7546	0.0671	[-0.8920, -0.6275]	325.8	0.90
$a_{\eta\eta,ij}$	31	0.3446	0.1244	[0.1097, 0.5912]	755.3	0.70
	32	-1.2499	0.1694	[-1.5621, -0.9524]	1563.8	0.66
	21	0.0265	0.0053	[0.0178, 0.0383]	321.9	0.94
$\sigma_{\epsilon\epsilon,ij}$	31	0.0241	0.0050	[0.0163, 0.0356]	340.1	0.62
	32	0.0258	0.0048	[0.0180, 0.0365]	263.7	0.88
	11	0.000123	0.000063	[0.000030, 0.000266]	57.7	1.00
	12	0.000570	0.000283	[0.000287, 0.001237]	996.3	0.97
	13	0.000538	0.000231	[0.000270, 0.001179]	1,245.3	0.90
	21	0.000521	0.000200	[0.000275, 0.001034]	1,217.0	0.86
$\sigma_{\eta\epsilon,ij}$	22	0.000503	0.000176	[0.000277, 0.000947]	831.6	0.58
	23	0.000545	0.000211	[0.000292, 0.001129]	1,109.6	0.91
	31	0.000494	0.000170	[0.000273, 0.000925]	689.0	0.82
	32	0.000506	0.000200	[0.000268, 0.001026]	951.7	0.97
	33	0.000516	0.000229	[0.000275, 0.001108]	1,297.8	0.81

Posterior mean, standard deviation, 95% credible bounds, inefficiency factor, and Geweke's convergence diagnostic (p-value). $a_{\eta\eta,ij}$, $\sigma_{\epsilon\epsilon,ij}$, $\sigma_{\eta\epsilon,ij}$ relate to lower triangular of $\boldsymbol{A}_{t:t+1}$.

σ^2_{ij} small in general. Thereby, choosing larger values for $\sigma^2_{i\cdot}$, $i = 2, \ldots, p$, and smaller values for $\sigma^2_{i\cdot}$, $i = p+1, \ldots, q$, is rather intuitive considering the specific structure of the triangular regressions in Eq. (10). More concretely, scale parameters for the inverse gamma priors in Eq. (21) have been found by conducting pilot runs, selecting combinations over the set $\{0.05, 0.01, 0.005, \ldots, 0.0001, 0.00005\}$ for $\sigma^2_{i\cdot}$, $i = 2, \ldots, p$, and $\{0.0005, 0.0001, 0.00005, \ldots, 0.0000005, 0.0000001\}$ for $\sigma^2_{i\cdot}$, $i = p+1, \ldots, q$.[8] Multiple selection criteria have been applied. Specifically, a sample average of latent volatility process $\boldsymbol{\alpha}_{1:n}$ close to zero, and in-sample fit as measured by the deviance information criterion (DIC), accounting for model complexity. Further, the predictive log-likelihood (in-sample), which may be seen as an indicator for out-of-sample performance. Comparability of resulting parameter estimates with the baseline MSVcc model has also been taken into account, inspecting volatility persistence $\boldsymbol{\Phi}$ and unconditional correlation and volatility sample averages as derived from the full covariance matrix including $\boldsymbol{\alpha}_t$. The proposed approach is

[8]Inaccurate priors may provoke numerical instabilities when block sampling the volatilities, especially for very extreme choices. Such problems can be alleviated using square root filtering and related techniques (de Jong, 1991b; for a more recent review, consider e.g. Grewal and Andrews, 2001). However, it is clearly preferable to avoid such problems in the first place through an adequate prior choice.

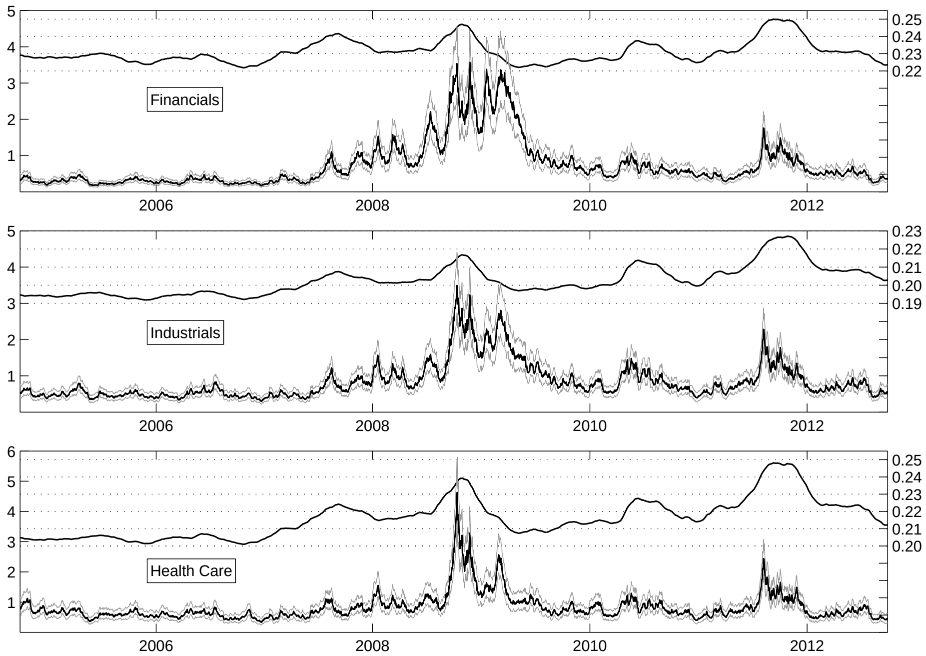

Figure 2: MSVdc - Volatility with 90% c.b. (left scale) and Volatility of Volatility (right scale). Data standardized to unit variance.

certainly not optimal, but systematic and only moderately time consuming. Moreover, as the data is standardized, it should be generally applicable.

The first $10{,}000$ draws are discarded as burn-in, collecting the next $200{,}000$ draws for parameter inference. Average block size and number of iterations to achieve convergence in the multi-move sampler are similar to Sec. 5.2. As the parameter estimates are mostly unintuitive due to the applied Cholesky decomposition, more emphasis is put on a graphical discussion.

Tab. 3 reports estimation results. Volatility persistence $\mathbf{\Phi}$ closely mirrors that of the MSVcc model. Moreover, time invariant elements $a_{\eta\eta,ij}$ associated with $\mathbf{\Sigma}_{\eta\eta,t+1}$ are all significant. Note also that except for $\sigma_{\eta\epsilon,11}$ (leverage Financials) variances governing the random walks related to cross leverage dynamics are mainly determined by their priors. The applied Cholesky-type decomposition reparameterizes the covariance matrix as a series of regressions, modeling the joint distribution as a marginal and then a series of conditionals. Accordingly, the latter series do not appear to carry significant new information.

Importantly, the MSVdc model tends to produce lower cross leverage effects compared to the MSVcc model, which are compensated by more pronounced transition error dynamics, featuring higher standard deviations and correlations. One may choose to equate parameters with the MSVcc model rather closely through a more and more tedious prior choice. However, at some point this is considered impractical. Instead, this shift of dynamics is attributed to the general structure of the applied Cholesky-type decomposition. Above all, referring to both models, dynamics are latent and in part to a significant degree prior dependent.

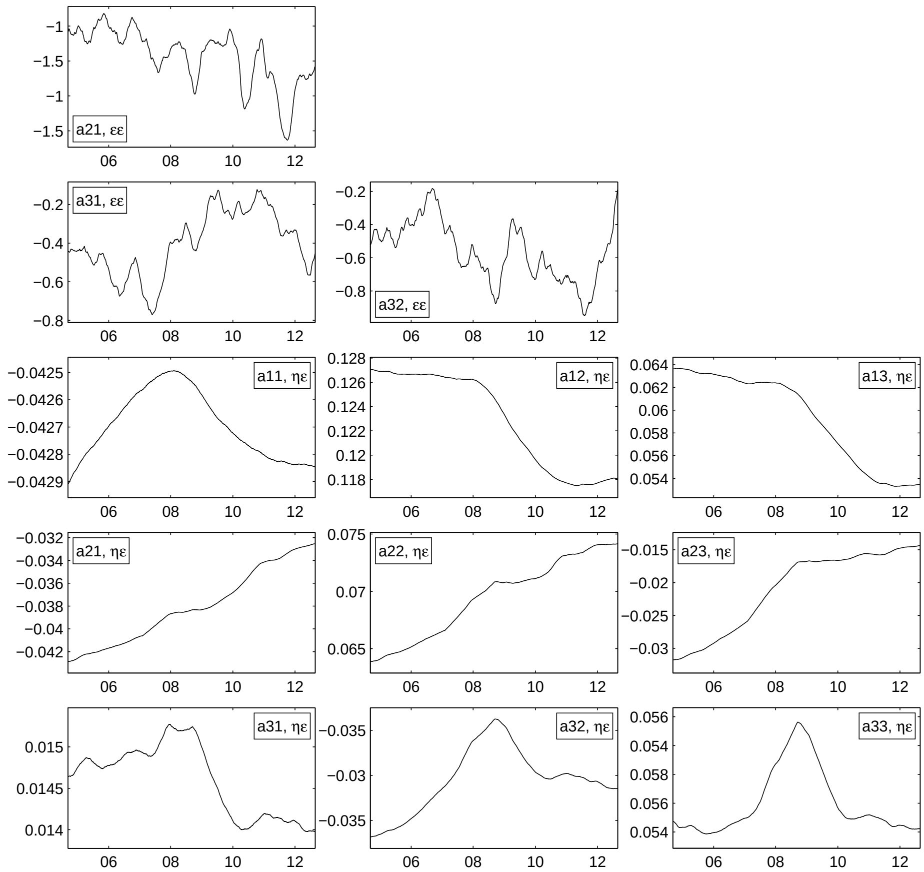

Figure 3: MSVdc - Cholesky Reparameterization Matrix $\boldsymbol{A}_{t:t+1}$

Fig. 2 plots latent volatility as calculated from the full covariance matrix including $\boldsymbol{\alpha}_t$, with associated 90% credible bounds.[9] For all three assets there is a pronounced increase in volatility during the financial crisis 2008/09. A more extended heavily volatile period for Financials during the crash period is visible. Interestingly, Healthcare shows the most distinct volatility spike, but volatility also tapers off fastest for this countercyclical sector. The high volatility period starting in early August 2011 and following the rating downgrade of U.S. Treasury bonds can also be clearly inferred.

As already mentioned, the current restricted model still produces dynamic volatility of volatility, and thus time varying unconditional kurtosis (for theoretical properties of SV models, see e.g. Harvey, 1994, or Taylor, 1994). The model generates increased unconditional tail risk during periods of market turmoil, which is intuitive. Interestingly, this is most pronounced for the period of increased macroeconomic uncertainty following the rating downgrade of U.S. Treasury bonds. Again, outlier risk is most distinctive for the Healthcare sector. Moreover, tail risk is overall higher after the crisis 2008/09, which is most pronounced for Industrials and Healthcare.

[9] Credible bounds have been calculated from a thinned sample of the posterior output, collecting every 100^{th} draw.

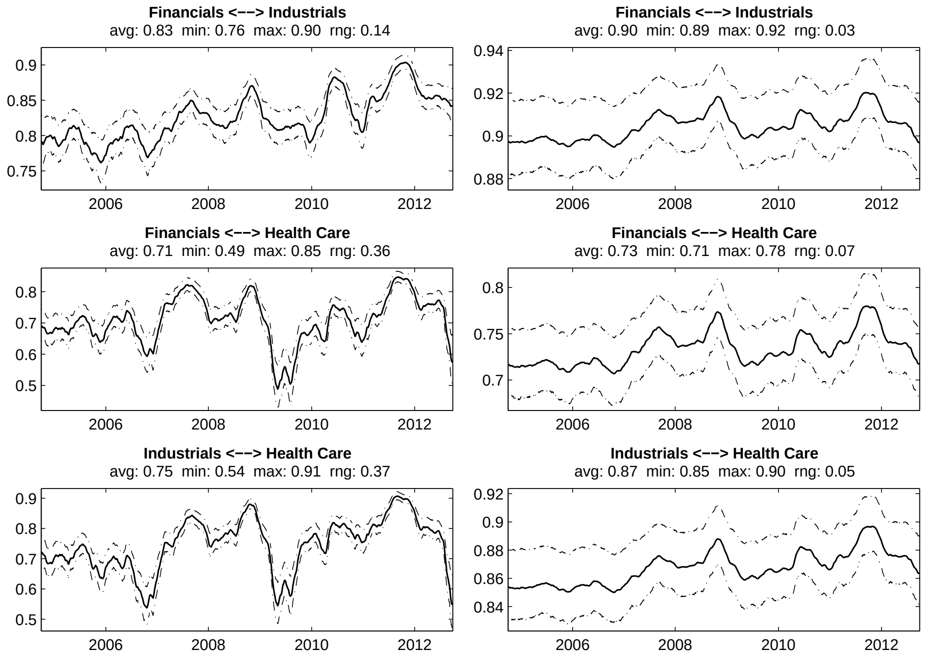

Figure 4: MSVdc - Correlation Observation Error $\boldsymbol{\rho}_{\epsilon\epsilon,t}$ (left col.) and Transition Error $\boldsymbol{\rho}_{\eta\eta,t}$ (right col.), with 50% c.b. Average, min., max., and range reported.

Fig. 3 plots dynamics of latent processes $\boldsymbol{a}_{t:t+1}$ governing the covariance evolution of $\boldsymbol{\Sigma}_{t:t+1}$ (the lower triangular of $\boldsymbol{A}_{t:t+1}$ excluding a_{ij}, $i > p$, $j > p$, is shown). The upper two rows responsible for observation error dynamics show strong time varying patterns. However, the elements of $\boldsymbol{A}_{t:t+1}$ in the lower three rows responsible for cross leverage dynamics feature some pronounced patterns as well. Observe e.g. a distinct spike of $a_{11,\eta\epsilon}$, corresponding to the leverage effect of Financials, or of $a_{32,\eta\epsilon}$ and $a_{33,\eta\epsilon}$, associated with cross leverage from Industrials to Healthcare and leverage of Healthcare, respectively, around the crisis.

Fig. 4 and Fig. 5 then visualize resulting correlations in $\boldsymbol{\Sigma}_{t:t+1}$. Inspecting observation error correlations $\boldsymbol{\rho}_{\epsilon\epsilon,t}$ in Fig. 4 first, we infer overall stronger interdependencies between Financials and Industrials than between Healthcare and the former two. Moreover, correlation between Financials and Industrials rises steadily during the examined period and shows a more stable pattern compared to the dependency structure between the latter two and Healthcare, which in turn look very similar. Correlation of Healthcare with both Financials and Industrials drops in the aftermath of the financial crisis 2008/09 and the increased volatility period following the debt downgrade. Prior to that it has reached local maxima, probably evidence of the well know contagion effect during turmoil periods. Generally note substantial uncertainty in the estimates, as indicated by 50% credible bounds. Correlation $\boldsymbol{\rho}_{\eta\eta,t}$ in the transition error is not modeled directly, but the generated dynamics are intuitive, with periods of increased uncertainty showing local correlation peaks in the information flow.

Turning to cross leverage dynamics visualized in Fig. 5, the conjecture that these are dominated by the series connected to the observation error and to some

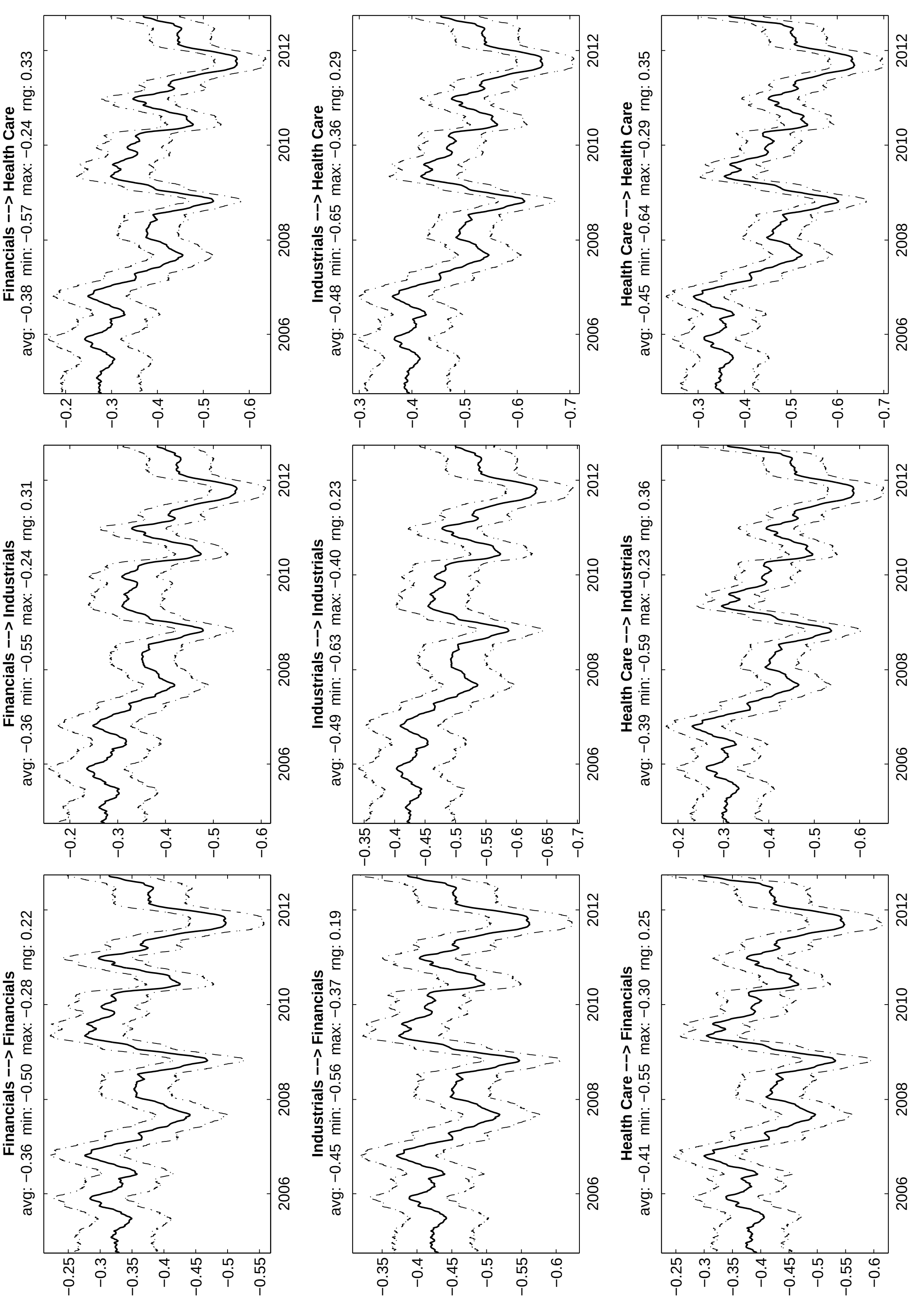

Figure 5: MSVdc - Cross Leverage $\boldsymbol{\rho}_{\epsilon\eta}$. Average, minimum, maximum, and range reported. 50% c.b.

extent by the first cross leverage element is confirmed. Accordingly, dynamics show large commonalities, but differences exist. Overall, there is a substantial increase in cross leverage around periods of market turmoil as the financial crisis 2008/09 and the high volatility period following the debt downgrade, which is intuitive. For example, it can be argued that investors are more alert during and shortly after market turmoil, rebalancing their market-wide portfolios significantly as a large price drop occurs in one sector. This would lead to the observed volatility increase in other sectors. Further and as in the MSVcc model, cross leverage effects to Healthcare appear stronger, albeit not significantly. Values also exhibit the largest fluctuations. Moreover, patterns in cross leverage dynamics may be distinguished between spillovers to the respective sectors each.

The ways in which dynamics feed through in the Cholesky-type approach are nontrivial, and the intuitive content of the generated cross leverage and transition error patterns reconfirms the usefulness of this technique. That point remains valid, even though, or indeed because the bulk of the dynamics in cross leverage effects is a result of the observation error dynamics being the dominant force in this factor-like approach.

6 Model Selection

Models are compared using the deviance information criterion (DIC) (Spiegelhalter, Best, Carlin, and van der Linde, 2002). It is defined by

$$DIC = \bar{D} + p_d,$$

where

$$p_d = \bar{D} - D(\bar{\boldsymbol{\theta}}), \quad \bar{D} = \mathrm{E}_{\theta|y}[D(\boldsymbol{\theta})], \quad D(\boldsymbol{\theta}) = -2\log f(\boldsymbol{y}_{1:n}|\boldsymbol{\theta}),$$

with $D(\boldsymbol{\theta})$ the deviance and $\bar{\boldsymbol{\theta}}$ the posterior mean. It is a Bayesian measure of model fit, attaining smaller values for better models, with p_d a penalty term for model complexity. Intuitively, higher parameter uncertainty lets $D(\boldsymbol{\theta})$ fluctuate more widely around $D(\bar{\boldsymbol{\theta}})$, yielding larger p_D values. Numerical standard errors of the estimates are obtained the following way. Posterior expectation of the deviance $\bar{D}$, readily available as a byproduct of the McMC run, is calculated by dividing the sample in 10 equally sized batches. Regarding deviance of the posterior mean $D(\bar{\boldsymbol{\theta}})$, the particle filter outlined in Sec. 4 is repeatedly applied 10 times to obtain estimates of log-likelihood ordinate $\log f_{\mathrm{post}}(\boldsymbol{y}_{1:n}|\bar{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \log \hat{f}(\boldsymbol{y}_t|\boldsymbol{y}_{1:t}, \bar{\boldsymbol{\theta}})$. The deviance of the predictive log-likelihood is calculated in a similar way from ordinate $\log f_{\mathrm{prior}}(\boldsymbol{y}_{1:n}|\bar{\boldsymbol{\theta}}) = \sum_{t=1}^{n} \log \hat{f}(\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1}, \bar{\boldsymbol{\theta}})$. The above approach yields 100 different DIC values, from which statistics are calculated.

In addition to the dynamic correlation model of Sec. 5.3, a MSVdc $\#2$ model is estimated, deploying larger priors for the latent random walks variances governing cross leverage dynamics,

$$\sigma_{ij}^2 \backsim \mathcal{IG}(5/2, 0.00001/2), \quad i > p, \quad j \leq p,$$

allowing for more diffuse dynamics. Another candidate is the MSVdc $\epsilon\epsilon$ model, with only the $a_{ij,t}$ associated with covariance dynamics $\boldsymbol{\Sigma}_{\epsilon\epsilon,t}$ time variant. Its performance is also investigated. For the initial convergence period of the particle

Table 4: Model Selection

	DIC	(s.e.)	Rank	p_D	(s.e.)	$\bar{D}$	(s.e.)	Rank	D_{pred}	(s.e.)	Rank
MSVcc	0	(6.4)	4	0	(3.4)	0	(3.1)	4	0	(1.4)	4
MSVdc $\epsilon\epsilon$	93.7	(5.2)	3	93.2	(2.9)	186.9	(2.5)	3	55.0	(1.2)	1
MSVdc	95.9	(7.0)	1	101.1	(4.2)	196.9	(3.2)	2	18.5	(6.0)	2
MSVdc $\#2$	94.1	(7.7)	2	110.3	(4.2)	204.3	(3.7)	1	17.1	(3.6)	3

Deviance information criterion, complexity penalty, posterior expectation of the deviance, and deviance of the predictive log-likelihood. Numerical standard error in parentheses.

filter a time frame of 500 observations prior to the dataset is chosen for all models, which is standardized by means and standard deviations calculated from a backward rolling window having size equal to the dataset.

Tab. 4 reports results, with the MSVcc as baseline model, and larger values indicating better fit. The DIC clearly favors the models featuring dynamic correlation and cross leverage over the MSVcc with time invariant correlation, with the presented MSVdc model the best. However, the dynamic correlation variants do not show significant differences among each other according to this criterion. Complexity penalty terms p_D are intuitive, with the MSVcc model having the lowest and the MSVdc $\#2$ with most diffuse dynamics the highest value. Inspecting posterior expectation of the deviance $\bar{D}$, we can infer a significant but relatively small improvement in in-sample fit for the MSVdc variants modeling dynamics in $\mathbf{\Sigma}_{\epsilon\eta,t:t+1}$ explicitly over the MSVdc $\epsilon\epsilon$ model. Not surprisingly then, the MSVdc $\#2$ variant fits the data best. Regarding forecasting ability, however, deviance of the predictive log-likelihood D_{pred} favors the MSVdc $\epsilon\epsilon$ variant, indicating improved predictive performance by making $\mathbf{\Sigma}$ time varying but also risk of overfitting when time dependency in cross leverage dynamics $a_{ij,t}$, $i > p$ is modeled explicitly.

7 Conclusion

A multivariate stochastic volatility model with dynamic correlation and cross leverage (MSVdc) is proposed, using a Cholesky-type decomposition. Modeling volatilities and covariances separately makes a direct interpretation of leverage possible. A McMC algorithm to efficiently estimate the model and an associated particle filter are proposed and outlined in detail. Parsimony is introduced by not modeling dynamics in the transition error explicitly. An empirical application based on three main S&P 500 sector indices illustrates features of the model. Cross leverage effects peak during and in the aftermath of financial turmoil. Moreover, the indirectly generated dynamics of volatility of volatility are intuitive. Specifically, increased unconditional tail risk and stronger correlated transition shocks are observed during distress periods.

Model selection indicates a superior fit of the proposed MSVdc model, benchmarked against the MSV model with constant correlation and cross leverage of Ishihara and Omori (2012). However, there is risk of overfitting when modeling the drivers of cross leverage explicitly, as the observation error plays a dominant role in generating the cross leverage and transition error dynamics. Consequently, a less complex model variant making only the Cholesky elements responsible for the observation error correlation dynamics time varying has shown to be competitive. This model variant also features dynamic cross leverage, due to the specific structure of the applied Cholesky decomposition. Moreover, the mechanism through which these dynamics are generated is nontrivial.

Dynamic sparsity could be introduced by deploying a latent threshold technique on elements of the Cholesky-type decomposed covariance matrix, as proposed by Nakajima (2012a), and Nakajima and West (2013), especially on those elements associated with cross leverage dynamics. Another way to improve on predictive performance could be to re-estimate daily, conducting one-step-ahead forecasts during the McMC run. An obvious advantage of such a strategy is the incorporation of parameter uncertainty into the forecasts. The smoothing approach of McMC estimation, which uses all available information of the sample, may also be beneficial in this regard. In contrast, the sequential particle filter conditions only on the most recent observation. Although computationally costly, the latter strategy is feasible, but not investigated in this work. Further, especially the drivers underlying cross leverage can be modeled as stationary processes, stabilizing estimates around the mean.

It would be interesting to compare performance of the proposed model with other related models in literature, especially the matrix exponential stochastic volatility model with cross leverage proposed by Ishihara et al. (2013), or the MSV model of Nakajima (2012b) with leverage and dynamic observation error. The former model features dynamics in the observation error with constant cross leverage effects. As such, a comparison with the MSVdc $\epsilon\epsilon$ model would deliver insights about the benefits of dynamics for $\mathbf{\Sigma}_{\epsilon\eta}$ generated indirectly via Cholesky-type decomposition. However, this is left for future research.

A Multi-Move Sampler

The multi-move sampler described below is an extension of Ishihara and Omori (2012) to handle time varying covariances and cross leverage effects. First divide $\boldsymbol{\alpha}_{1:n}$ into $K+1$ blocks at random, say $(\boldsymbol{\alpha}_{k_{i-1}+1}, \ldots, \boldsymbol{\alpha}_{k_i})'$, $i = 1, \ldots, K+1$, with $k_0 = 0$ and $k_{K+1} = n$. Knots are stochastic as proposed by Shephard and Pitt (1997) and are given by

$$k_i = \mathrm{int}[n(i + \mathcal{U}_i)/(K+2)], \quad i = 1, \ldots, K, \quad k_i - k_{i-1} \geq 2,$$

where $\mathcal{U}_i$ is a random sample from the uniform distribution $\mathcal{U}(0,1)$. This makes sampling more efficient as the points of conditioning (knots) change over the McMC iterations, with K as a tuning parameter.

Let $\boldsymbol{x}_t = \boldsymbol{R}_t^{-1}\boldsymbol{\eta}_t$, where matrix $\boldsymbol{R}_t$ denotes a Cholesky decomposition of $\boldsymbol{\Sigma}_{\eta\eta,t}$ such that $\boldsymbol{\Sigma}_{\eta\eta,t} = \boldsymbol{R}_t\boldsymbol{R}_t'$, $t > 1$, and $\boldsymbol{\Sigma}_{\eta\eta,\mathrm{ini}} = \boldsymbol{R}_1\boldsymbol{R}_1'$. Further suppose that $k_{i-1} = r$ and $k_i = r + d$ for the i^{th} block. The sampler then exploits the independence of disturbances $\boldsymbol{\eta}_t$, block sampling $\underline{\boldsymbol{x}} \equiv (\boldsymbol{x}_{r+1}', \ldots, \boldsymbol{x}_{r+d}')'$ instead of $\underline{\boldsymbol{\alpha}} \equiv (\boldsymbol{\alpha}_{r+1}', \ldots, \boldsymbol{\alpha}_{r+d}')'$ from its full joint conditional posterior density ($r \geq 0$, $d \geq 2$, $r + d \leq n$),

$$\pi(\underline{\boldsymbol{x}}|\boldsymbol{\alpha}_{r+d+1}, \boldsymbol{y}_{r+1}, \ldots, \boldsymbol{y}_{r+d}, \boldsymbol{\theta}) \propto$$
$$\prod_{t=r+1}^{r+d} f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1}) \times \prod_{t=r+1}^{r+d} f(\boldsymbol{x}_t) \times f(\boldsymbol{\alpha}_{r+d}), \qquad r = 0, \tag{22}$$
$$\pi(\underline{\boldsymbol{x}}|\boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+d+1}, \boldsymbol{y}_r, \ldots, \boldsymbol{y}_{r+d}, \boldsymbol{\theta}) \propto$$
$$\prod_{t=r}^{r+d} f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1}) \times \prod_{t=r+1}^{r+d} f(\boldsymbol{x}_t) \times f(\boldsymbol{\alpha}_{r+d}), \qquad r + d < n,$$
$$\pi(\underline{\boldsymbol{x}}|\boldsymbol{\alpha}_r, \boldsymbol{y}_r, \ldots, \boldsymbol{y}_{r+d}, \boldsymbol{\theta}) \propto$$
$$\prod_{t=r}^{r+d-1} f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1}) \times f(\boldsymbol{y}_n|\boldsymbol{\alpha}_n) \times \prod_{t=r+1}^{r+d} f(\boldsymbol{x}_t), \qquad r + d = n. \tag{23}$$

The logarithm of $f(\boldsymbol{y}_t|\boldsymbol{\alpha}_t, \boldsymbol{\alpha}_{t+1})$, $f(\boldsymbol{y}_n|\boldsymbol{\alpha}_n)$ in Eq. (22)-(23) can be written as (excluding constant terms)

$$l_t = -\frac{1}{2}\mathbf{1}_p'\boldsymbol{\alpha}_t - \frac{1}{2}(\boldsymbol{z}_t - \boldsymbol{m}_t)'\boldsymbol{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t),$$

where $\mathbf{1}_p$ is a $p \times 1$ vector of ones, and

$$\boldsymbol{z}_t = \boldsymbol{V}_t^{-1/2}\boldsymbol{y}_t, \quad \boldsymbol{m}_t = \boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)\mathrm{I}_{t<n} + \mathbf{0},$$
$$\boldsymbol{S}_t = \begin{cases} \boldsymbol{\Sigma}_{\epsilon\epsilon,t} - \boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1} & t < n \\ \boldsymbol{\Sigma}_{\epsilon\epsilon,t} & t = n. \end{cases}$$

Then the logarithmized posterior density given by Eq. (22)-(23) can be expressed as $-1/2\sum_{r+1}^{r+d}\boldsymbol{x}_t'\boldsymbol{x}_t + L$, where (excluding constant terms)

$$L = \log f(\boldsymbol{\alpha}_{r+d})\mathrm{I}_{r+d<n} + \begin{cases} \sum_{r+1}^{r+d} l_t & r = 0 \\ \sum_{r}^{r+d} l_t & r > 0, \end{cases} \tag{24}$$
$$\log f(\boldsymbol{\alpha}_{r+d}) = -\frac{1}{2}(\boldsymbol{\alpha}_{r+d+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{r+d})'\boldsymbol{\Sigma}_{\eta\eta,r+d+1}^{-1}(\boldsymbol{\alpha}_{r+d+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_{r+d}).$$

To construct a proposal density based on a normal approximation of the posterior density of $\underline{\boldsymbol{x}}$, define further

$$\boldsymbol{\delta} = (\boldsymbol{\delta}'_{r+1}, \ldots, \boldsymbol{\delta}'_{r+d})', \quad \boldsymbol{\delta}_t = \frac{\partial L}{\partial \boldsymbol{\alpha}_t},$$

$$\boldsymbol{Q} = -\mathrm{E}\left(\frac{\partial^2 L}{\partial \underline{\boldsymbol{\alpha}} \partial \underline{\boldsymbol{\alpha}}'}\right) = \begin{pmatrix} \boldsymbol{A}_{r+1} & \boldsymbol{B}_{r+2} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{B}_{r+2} & \boldsymbol{A}_{r+2} & \boldsymbol{B}_{r+3} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{B}_{r+3} & \boldsymbol{A}_{r+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \boldsymbol{B}_{r+d} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{B}_{r+d} & \boldsymbol{A}_{r+d} \end{pmatrix},$$

$$\boldsymbol{A}_t = -\mathrm{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_t}\right), \quad t = r+1, \ldots, r+d,$$

$$\boldsymbol{B}_t = -\mathrm{E}\left(\frac{\partial^2 L}{\partial \boldsymbol{\alpha}_t \partial \boldsymbol{\alpha}'_{t-1}}\right), \quad t = r+2, \ldots, r+d, \quad \boldsymbol{B}_{r+1} = \boldsymbol{0}.$$

First and second derivatives $\boldsymbol{\delta}_t$, $\boldsymbol{A_t}$, $\boldsymbol{B_t}$ are given in App. B. Applying a second order Taylor expansion to the log of the posterior density in Eq. (24) around mode $\hat{\underline{\boldsymbol{x}}}$ yields an approximating normal density as follows (excluding constant terms):

$$\begin{aligned} &\log \pi(\underline{\boldsymbol{x}}|\cdot) \\ &\approx \hat{L} + \left.\frac{\partial L}{\partial \underline{\boldsymbol{x}}'}\right|_{\underline{\boldsymbol{x}}=\hat{\underline{\boldsymbol{x}}}} (\underline{\boldsymbol{x}} - \hat{\underline{\boldsymbol{x}}}) + \frac{1}{2}(\underline{\boldsymbol{x}} - \hat{\underline{\boldsymbol{x}}})' \, \mathrm{E}\left(\frac{\partial^2 L}{\partial \underline{\boldsymbol{x}} \partial \underline{\boldsymbol{x}}'}\right)\bigg|_{\underline{\boldsymbol{x}}=\hat{\underline{\boldsymbol{x}}}} (\underline{\boldsymbol{x}} - \hat{\underline{\boldsymbol{x}}}) - \frac{1}{2}\sum_{t=r+1}^{r+d} \boldsymbol{x}'_t \boldsymbol{x}_t \\ &= \hat{L} + \hat{\boldsymbol{\delta}}'(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}}) - \frac{1}{2}(\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}})' \hat{\boldsymbol{Q}} (\underline{\boldsymbol{\alpha}} - \hat{\underline{\boldsymbol{\alpha}}}) - \frac{1}{2}\sum_{t=r+1}^{r+d} \boldsymbol{x}'_t \boldsymbol{x}_t \\ &\equiv \log q(\underline{\boldsymbol{x}}|\cdot), \end{aligned}$$

where $\hat{L}$, $\hat{\boldsymbol{\delta}}$ and $\hat{\boldsymbol{Q}}$ are the values of L, $\boldsymbol{\delta}$ and $\boldsymbol{Q}$ at $\underline{\boldsymbol{\alpha}} = \hat{\underline{\boldsymbol{\alpha}}}$ (or at $\underline{\boldsymbol{x}} = \hat{\underline{\boldsymbol{x}}}$, equivalently). It can be shown that proposal density $q(\underline{\boldsymbol{x}}|\cdot)$ is the posterior of $\underline{\boldsymbol{x}}$ obtained from an ordinary linear state space model given by Eq. (25)-(26) below. Mode $\hat{\underline{\boldsymbol{x}}}$ is found by repeating the following algorithm until convergence (usually 2 to 5 iterations).

Algorithm 1 (Disturbance smoother):

1. Initialize $\hat{\underline{\boldsymbol{x}}}$, and compute $\hat{\underline{\boldsymbol{\alpha}}}$ at $\underline{\boldsymbol{x}} = \hat{\underline{\boldsymbol{x}}}$ using state equation Eq. (6) recursively.

2. Evaluate $\hat{\boldsymbol{\delta}}_t$, $\hat{\boldsymbol{A}}_t$, and $\hat{\boldsymbol{B}}_t$ at $\underline{\boldsymbol{\alpha}} = \hat{\underline{\boldsymbol{\alpha}}}$, $t = r+1, \ldots, r+d$.

3. Set $\hat{\boldsymbol{D}}_{r+1} = \hat{\boldsymbol{A}}_{r+1}$ and $\hat{\boldsymbol{b}}_{r+1} = \hat{\boldsymbol{\delta}}_{r+1}$. Compute the following variables recursively for $t = r+2, \ldots, r+d$:

$$\hat{\boldsymbol{D}}_t = \hat{\boldsymbol{A}}_t - \hat{\boldsymbol{B}}_t \hat{\boldsymbol{D}}_{t-1}^{-1} \hat{\boldsymbol{B}}'_t, \quad \hat{\boldsymbol{K}}_t = \mathrm{chol}(\hat{\boldsymbol{D}}_t),$$
$$\hat{\boldsymbol{b}}_t = \hat{\boldsymbol{\delta}}_t - \hat{\boldsymbol{B}}_t \hat{\boldsymbol{D}}_{t-1}^{-1} \hat{\boldsymbol{b}}_{t-1},$$

   with $\mathrm{chol}(\hat{\boldsymbol{D}}_t)$ a Cholesky decomposition of $\hat{\boldsymbol{D}}_t$ such that $\hat{\boldsymbol{D}}_t = \hat{\boldsymbol{K}}_t \hat{\boldsymbol{K}}'_t$.

4. Define auxiliary vector $\hat{\boldsymbol{y}}_t = \hat{\boldsymbol{\varphi}}_t + \hat{\boldsymbol{D}}_t^{-1} \hat{\boldsymbol{b}}_t$, where $\hat{\boldsymbol{\varphi}}_t = \hat{\boldsymbol{\alpha}}_t + \hat{\boldsymbol{D}}_t^{-1} \hat{\boldsymbol{B}}_{t+1} \hat{\boldsymbol{\alpha}}_{t+1}$, $t = r+1, \ldots, r+d$, and set $\hat{\boldsymbol{B}}_{d+r+1} = 0$, $\hat{\boldsymbol{\alpha}}_{r+d+1} = \boldsymbol{\alpha}_{r+d+1}$.

5. Consider the linear Gaussian state space model formulated by

$$\hat{\boldsymbol{y}}_t = \boldsymbol{Z}_t\boldsymbol{\alpha}_t + \boldsymbol{G}_t\boldsymbol{\kappa}_t, \quad t = r+1, \ldots, r+d, \tag{25}$$
$$\boldsymbol{\alpha}_{t+1} = \boldsymbol{\Phi}\boldsymbol{\alpha}_t + \boldsymbol{H}_{t+1}\boldsymbol{\omega}_{t+1}, \quad t = r, \ldots, r+d-1,$$
$$\begin{pmatrix} \boldsymbol{\kappa}_t \\ \boldsymbol{\omega}_{t+1} \end{pmatrix} \backsim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_{2p}), \tag{26}$$

with

$$\boldsymbol{Z}_t = \boldsymbol{I}_p + \hat{\boldsymbol{D}}_t^{-1}\hat{\boldsymbol{B}}'_{t+1}\boldsymbol{\Phi}, \quad \boldsymbol{G}_t = (\hat{\boldsymbol{K}}_t^{-1}, \hat{\boldsymbol{D}}_t^{-1}\hat{\boldsymbol{B}}'_{t+1}\boldsymbol{R}_{t+1}), \quad \boldsymbol{H}_{t+1} = (\boldsymbol{0}, \boldsymbol{R}_{t+1}),$$

and $\boldsymbol{\alpha}_0 = 0$. Apply the Kalman filter (e.g. Durbin and Koopman, 2008) and disturbance smoother (Koopman, 1993) to this state space model to obtain posterior mode $\hat{\underline{\boldsymbol{x}}}$ or $\hat{\underline{\boldsymbol{\alpha}}}$, equivalently.

6. Go to 2.

In the McMC sampling procedure, the current sample of $\underline{\boldsymbol{x}}$ may be taken as initial value of $\hat{\underline{\boldsymbol{x}}}$ in Step 1. Further note that the above steps are equivalent to the method of scoring used to maximize the conditional posterior density. After convergence of Algorithm 1, $\underline{\boldsymbol{x}}$ is sampled from the conditional posterior density by conducting an AR (Accept - Reject) - MH algorithm (Tierney, 1994) using the simulation smoother (de Jong and Shephard, 1995, or Durbin and Koopman, 2002).

Algorithm 2 (AR-MH step and simulation smoother):

1. Let $\underline{\boldsymbol{x}}_0$ denote the current value. Find mode $\hat{\underline{\boldsymbol{x}}}$ using Algorithm 1.
2. Proceed with Steps 2-4 in Algorithm 1 to obtain the approximated linear Gaussian state space model of Eq. (25)-(26).
3. Propose a candidate $\underline{\boldsymbol{x}}^*$ by sampling from $\tilde{q}(\underline{\boldsymbol{x}}^*) \propto \min\{\pi(\underline{\boldsymbol{x}}^*|\cdot), cq(\underline{\boldsymbol{x}}^*|\cdot)\}$ using the AR algorithm as follows:
   (a) Generate $\underline{\boldsymbol{x}}^*$ using the simulation smoother for the approximated state space of model Eq. (25)-(26).
   (b) Accept $\underline{\boldsymbol{x}}^*$ with probability
   $$\frac{\min\{\pi(\underline{\boldsymbol{x}}^*|\cdot), cq(\underline{\boldsymbol{x}}^*|\cdot)\}}{cq(\underline{\boldsymbol{x}}^*|\cdot)},$$
   where c is a scaling constant. If it is rejected, go to (a).
4. Conduct the MH algorithm using candidate $\underline{\boldsymbol{x}}^*$, with acceptance probability
$$\min\left\{\frac{\pi(\underline{\boldsymbol{x}}^*|\cdot) \times \min\{\pi(\underline{\boldsymbol{x}}_0|\cdot), cq(\underline{\boldsymbol{x}}_0|\cdot)\}}{\pi(\underline{\boldsymbol{x}}_0|\cdot) \times \min\{\pi(\underline{\boldsymbol{x}}^*|\cdot), cq(\underline{\boldsymbol{x}}^*|\cdot)\}}, 1\right\}.$$

B Derivatives $\boldsymbol{\delta}_t$, $\boldsymbol{A}_t$, $\boldsymbol{B}_t$

Before calculating derivatives $\boldsymbol{\delta}_t$, $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$, definitions of the first and second derivative of a matrix and the product and chain rule as in Magnus and Abadir (2007) are provided. Let F be a twice differentiable $m \times p$ matrix function of an $n \times q$ matrix $\boldsymbol{X}$. Then the first derivative (Jacobian) of F at $\boldsymbol{X}$ is defined by the $mp \times nq$ matrix

$$\mathbf{D}F(\boldsymbol{X}) = \frac{\partial F(\boldsymbol{X})}{\partial \boldsymbol{X}} = \frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial \mathrm{vec}(\boldsymbol{X})'},$$

and the second derivative (Hessian) of F at $\boldsymbol{X}$ is defined by the $mnpq \times nq$ matrix

$$\mathbf{H}F(\boldsymbol{X}) = \mathbf{D}((\mathbf{D}F(\boldsymbol{X}))') = \frac{\partial}{\partial \mathrm{vec}(\boldsymbol{X})'} \mathrm{vec}\bigg(\bigg(\frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial \mathrm{vec}(\boldsymbol{X})'}\bigg)'\bigg).$$

Product Rule: Let S be a subset of $\mathbb{R}^{n\times q}$, and assume that $F : S \to \mathbb{R}^{m\times p}$ and $G : S \to \mathbb{R}^{p\times r}$ are differentiable at an interior point C of S. Then

$$\frac{\partial \mathrm{vec}(FG)}{\partial \mathrm{vec}(\boldsymbol{X})'} = (G' \otimes \boldsymbol{I}_m)\frac{\partial \mathrm{vec}(F)}{\partial \mathrm{vec}(\boldsymbol{X})'} + (\boldsymbol{I}_r \otimes F)\frac{\partial \mathrm{vec}(G)}{\partial \mathrm{vec}(\boldsymbol{X})'}. \tag{27}$$

Chain Rule: Let S be a subset of $\mathbb{R}^{n\times q}$, and assume that $F : S \to \mathbb{R}^{m\times p}$ is differentiable at an interior point C of S. Let T be a subset of $\mathbb{R}^{m\times p}$ such that $F(\boldsymbol{X}) \in T$ for all $\boldsymbol{X} \in S$, and assume that $G : T \to \mathbb{R}^{r\times s}$ is differentiable at an interior point $B = F(C)$ of T. Then the composite function $H : S \to \mathbb{R}^{r\times s}$ defined by $H(\boldsymbol{X}) = G(F(\boldsymbol{X}))$ is differentiable at C, and

$$\boldsymbol{D}H(\boldsymbol{X}) = (\boldsymbol{D}G(F(\boldsymbol{X})))(\boldsymbol{D}F(\boldsymbol{X})) = \frac{\partial \mathrm{vec}(G(F(\boldsymbol{X})))}{\partial \mathrm{vec}(F(\boldsymbol{X}))'}\frac{\partial \mathrm{vec}(F(\boldsymbol{X}))}{\partial \mathrm{vec}(\boldsymbol{X})'}.$$

If $q = 1$ then $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$, $f : \mathbb{R}^{n\times 1} \to \mathbb{R}^{m\times p}$, $g : \mathbb{R}^{m\times p} \to \mathbb{R}^{r\times s}$, and

$$\frac{\partial g(f(\boldsymbol{x}))}{\partial \boldsymbol{x}'} = \frac{\partial \mathrm{vec}(g(f(\boldsymbol{x})))}{\partial \mathrm{vec}(f(\boldsymbol{x}))'}\frac{\partial \mathrm{vec}(f(\boldsymbol{x}))}{\partial \mathrm{vec}(\boldsymbol{x})'}. \tag{28}$$

Derivatives $\boldsymbol{\delta}_t$, $\boldsymbol{A}_t$, $\boldsymbol{B}_t$
With L as given in Eq. (24) first derivatives are

$$\boldsymbol{\delta}_t = \frac{\partial L}{\partial \boldsymbol{\alpha}_t} = \frac{\partial l_t}{\partial \boldsymbol{\alpha}_t} + \frac{\partial l_{t-1}}{\partial \boldsymbol{\alpha}_t} + \boldsymbol{\Phi}\boldsymbol{\Sigma}^{-1}_{\eta\eta,t+1}(\boldsymbol{\alpha_{t+1}} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)\mathrm{I}_{t=r+d<n},$$

where, applying the chain rule of Eq. (28),

$$\begin{aligned}\frac{\partial l_t}{\partial \boldsymbol{\alpha}'_t} &= -\frac{1}{2}\mathbf{1}'_p - (\boldsymbol{z}_t - \boldsymbol{m}_t)'\boldsymbol{S}_t^{-1}\frac{\partial(\boldsymbol{z_t} - \boldsymbol{m_t})}{\partial \boldsymbol{\alpha}'_t} \\ &= -\frac{1}{2}\mathbf{1}'_p + \frac{1}{2}(\boldsymbol{z}_t - \boldsymbol{m}_t)'\boldsymbol{S}_t^{-1}\left[\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}^{-1}_{\eta\eta,t+1}\boldsymbol{\Phi}\mathrm{I}_{t<n}\right], \\ \frac{\partial l_{t-1}}{\partial \boldsymbol{\alpha}'_t} &= (\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})'\boldsymbol{S}^{-1}_{t-1}\frac{\partial \boldsymbol{m}_{t-1}}{\partial \boldsymbol{\alpha}'_t} \\ &= (\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})'\boldsymbol{S}^{-1}_{t-1}\boldsymbol{\Sigma}_{\epsilon\eta,t-1:t}\boldsymbol{\Sigma}^{-1}_{\eta\eta,t}\mathrm{I}_{t>1}.\end{aligned}$$

Thus we have

$$\begin{aligned}\boldsymbol{\delta}_t = &-\frac{1}{2}\mathbf{1}_p + \frac{1}{2}\left[\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\mathrm{I}_{t<n}\right]\boldsymbol{S}_t^{-1}(\boldsymbol{z}_t - \boldsymbol{m}_t)\\ &+ \boldsymbol{\Sigma}_{\eta\eta,t}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t-1:t}\boldsymbol{S}_{t-1}^{-1}(\boldsymbol{z}_{t-1} - \boldsymbol{m}_{t-1})\mathrm{I}_{t>1} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}(\boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t)\mathrm{I}_{t=r+d<n}.\end{aligned}$$

To compute second derivative $\boldsymbol{A}_t$ start by calculating the Hessian matrix

$$\frac{\partial^2 L}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}_t'} = \frac{\partial^2 l_t}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}_t'} + \frac{\partial^2 l_{t-1}}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}_t'} - \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\mathrm{I}_{t=r+d<n},$$

where, applying the product rule of Eq. (27),

$$\begin{aligned}\frac{\partial^2 l_t}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}_t'} &= \frac{1}{2}\left[(\boldsymbol{z}_t - \boldsymbol{m}_t)'\boldsymbol{S}_t^{-1}\otimes\boldsymbol{I}_p\right]\frac{\partial\mathrm{vec}(\mathrm{diag}(\boldsymbol{z}_t))}{\partial\boldsymbol{\alpha}_t'}\\ &\quad-\frac{1}{4}\left[\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\mathrm{I}_{t<n}\right]\\ &\quad\times\boldsymbol{S}_t^{-1}\left[\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\mathrm{I}_{t<n}\right],\\ \frac{\partial^2 l_{t-1}}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}_t'} &= -\boldsymbol{\Sigma}_{\eta\eta,t}\boldsymbol{\Sigma}_{\eta\epsilon,t-1:t}\boldsymbol{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\epsilon\eta,t-1:t}\boldsymbol{\Sigma}_{\eta\eta,t}\mathrm{I}_{t>1}.\end{aligned}\tag{29}$$

Now observe that

$$\frac{\partial\mathrm{vec}(\mathrm{diag}(\boldsymbol{z}_t))}{\partial\boldsymbol{\alpha}_t'} = -\frac{1}{2}\begin{pmatrix}z_{1,t}\boldsymbol{e}_1\boldsymbol{e}_1'\\ \vdots\\ z_{p,t}\boldsymbol{e}_p\boldsymbol{e}_p'\end{pmatrix},\quad \mathrm{E}\left(z_{i,t}(\boldsymbol{z}_t - \boldsymbol{m}_t)\right) = \boldsymbol{S}_t\boldsymbol{e}_i,$$

where $\boldsymbol{e}_i$ is a $p\times 1$ unit vector with the i^{th} component equal to 1. Then the expected value of the first term in Eq. (29) is

$$\mathrm{E}\left([(\boldsymbol{z}_t - \boldsymbol{m}_t)'\boldsymbol{S}_t^{-1}\otimes\boldsymbol{I}_p]\frac{\partial\mathrm{vec}(\mathrm{diag}(\boldsymbol{z}_t))}{\partial\boldsymbol{\alpha}_t'}\right) = -\frac{1}{2}\sum_{i=1}^{p}(\boldsymbol{e}_i'\boldsymbol{S}_t^{-1})(\boldsymbol{S}_t\boldsymbol{e}_i)\boldsymbol{e}_i\boldsymbol{e}_i' = -\frac{1}{2}\boldsymbol{I}_p.$$

Noting that $\mathrm{diag}(\boldsymbol{z}_t)\boldsymbol{S}_t^{-1}\mathrm{diag}(\boldsymbol{z}_t) = \boldsymbol{S}_t^{-1}\odot(\boldsymbol{z}_t\boldsymbol{z}_t')$, the expected value of the second term in Eq. (29) is

$$\begin{aligned}&\mathrm{E}\left(\left[\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\mathrm{I}_{t<n}\right]\boldsymbol{S}_t^{-1}\left[\mathrm{diag}(\boldsymbol{z}_t) - 2\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\mathrm{I}_{t<n}\right]\right)\\ &\quad= \boldsymbol{S}_t^{-1}\odot(\boldsymbol{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t') + 4\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\boldsymbol{S}_t^{-1}\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\mathrm{I}_{t<n}\\ &\qquad- 2\left[\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\boldsymbol{S}_t^{-1}\mathrm{diag}(\boldsymbol{m}_t) + \mathrm{diag}(\boldsymbol{m}_t)\boldsymbol{S}_t^{-1}\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\right]\mathrm{I}_{t<n},\end{aligned}$$

and finally we have

$$\begin{aligned}\boldsymbol{A}_t = &\frac{1}{4}\left[\boldsymbol{I}_p + \boldsymbol{S}_t^{-1}\odot(\boldsymbol{S}_t + \boldsymbol{m}_t\boldsymbol{m}_t')\right] + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\boldsymbol{S}_t^{-1}\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\mathrm{I}_{t<n}\\ &-\frac{1}{2}\left[\boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t:t+1}\boldsymbol{S}_t^{-1}\mathrm{diag}(\boldsymbol{m}_t) + \mathrm{diag}(\boldsymbol{m}_t)\boldsymbol{S}_t^{-1}\boldsymbol{\Sigma}_{\epsilon\eta,t:t+1}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\right]\mathrm{I}_{t<n}\\ &+\boldsymbol{\Sigma}_{\eta\eta,t}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t-1:t}\boldsymbol{S}_{t-1}^{-1}\boldsymbol{\Sigma}_{\epsilon\eta,t-1:t}\boldsymbol{\Sigma}_{\eta\eta,t}^{-1}\mathrm{I}_{t>1} + \boldsymbol{\Phi}\boldsymbol{\Sigma}_{\eta\eta,t+1}^{-1}\boldsymbol{\Phi}\mathrm{I}_{t=r+d<n}.\end{aligned}$$

In a similar fashion it is straightforward to show that

$$\boldsymbol{B}_t = -\mathrm{E}\left(\frac{\partial^2 l_{t-1}}{\partial\boldsymbol{\alpha}_t\partial\boldsymbol{\alpha}_{t-1}'}\right) = \frac{1}{2}\boldsymbol{\Sigma}_{\eta\eta,t}^{-1}\boldsymbol{\Sigma}_{\eta\epsilon,t-1:t}\boldsymbol{S}_{t-1}^{-1}\left[\mathrm{diag}(\boldsymbol{m}_{t-1}) - 2\boldsymbol{\Sigma}_{\epsilon\eta,t-1:t}\boldsymbol{\Sigma}_{\eta\eta,t}^{-1}\boldsymbol{\Phi}\right].$$

C Sampling from the Wishart Distribution

Sampling of $\boldsymbol{\Sigma}$ time invariant is done by the MH algorithm. Assuming a conjugate Wishart prior of the form $\boldsymbol{\Sigma}^{-1} \backsim \mathcal{W}(\nu_0, \boldsymbol{V}_0)$, the conditional posterior of $\boldsymbol{\Sigma}$ reads

$$\pi(\boldsymbol{\Sigma}|\cdot) \propto |\boldsymbol{\Sigma}|^{-\frac{\nu+2p+1}{2}} \exp\Big\{-\frac{1}{2}\mathrm{tr}(\boldsymbol{V}^{-1}\boldsymbol{\Sigma}^{-1})\Big\} \times \mathrm{c}(\boldsymbol{\Sigma}),$$

where

$$\mathrm{c}(\boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}_{\mathrm{ini}}|^{-\frac{1}{2}}|\boldsymbol{\Sigma}_{\epsilon\epsilon}|^{-\frac{1}{2}} \exp\Big\{-\frac{1}{2}\Big(\boldsymbol{\alpha}_1'\boldsymbol{\Sigma}_{\mathrm{ini}}^{-1}\boldsymbol{\alpha}_1 + \boldsymbol{y}_n'\boldsymbol{V}_n^{-1/2}\boldsymbol{\Sigma}_{\epsilon\epsilon}^{-1}\boldsymbol{V}_n^{-1/2}\boldsymbol{y}_n\Big)\Big\},$$
$$\boldsymbol{V}^{-1} = \boldsymbol{V}_0^{-1} + \sum_{t=1}^{n-1}\boldsymbol{v}_t\boldsymbol{v}_t', \quad \boldsymbol{v}_t = \begin{pmatrix} \boldsymbol{V}_t^{-1/2}\boldsymbol{y}_t \\ \boldsymbol{\alpha}_{t+1} - \boldsymbol{\Phi}\boldsymbol{\alpha}_t \end{pmatrix}, \quad \nu = \nu_0 + n - 1.$$

A proposal $\boldsymbol{\Sigma}^* \backsim [\mathcal{W}(\nu, \boldsymbol{V})]^{-1}$ is accepted with probability $\min\{c(\boldsymbol{\Sigma}^*)/c(\boldsymbol{\Sigma}), 1)\}$, where $\boldsymbol{\Sigma}$ is the current sample.

D Inefficiency Factor / Convergence Diagnostic

The inefficiency factor is the estimated ratio of the numerical variance of the posterior sample mean to the variance of the hypothetical sample mean from uncorrelated draws. With an inefficiency factor of m, we need to draw McMC samples m times as many as uncorrelated samples. The following formula is used to calculate the inefficiency factor,

$$IF = 1 + 2\frac{b_w}{b_w - 1}\sum_{i=1}^{b_w} K_{\mathrm{QS}}\left(\frac{i}{b_w}\right)\hat{\rho}_i, \tag{30}$$

where $\hat{\rho}_i$ is the sample autocorrelation at lag i, K_{QS} the Quadratic Spectral (QS) kernel and b_w the bandwidth parameter (the correction term $b_w/(b_w - 1)$ dropped if $b_w = 1$). Andrews (1991) finds the QS kernel to be superior in terms of an asymptotic truncated mean squared error criterion, relative to other kernels. The QS kernel is defined as

$$K_{\mathrm{QS}}(x) = \frac{25}{12\pi^2x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right).$$

The bandwidth b_w is determined automatically as a function of the data, following Andrews (1991). Specifically,

$$\hat{b}_w = 1.3221(\hat{\alpha}(2)N)^{1/5},$$

where N is the number of collected draws and

$$\hat{\alpha}(2) = \frac{4\hat{\rho}_a^2\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^8}\Big/\left[\frac{\hat{\sigma}_a^4}{(1-\hat{\rho}_a)^4}\right],$$

with $\{\hat{\rho}_a,\, \hat{\sigma}_a\}$ the autoregressive and variance innovation parameter, respectively, obtained by running a first-order autoregressive linear regression on the demeaned draws of the Markov chain. As Andrews (1991) notes, there can be sensitivity of the estimator to the choice of the bandwidth parameter. In the current context, this would suggest obtaining multiple bandwidth values by varying $\hat{\rho}_a$ by one or two standard deviations, $\hat{\rho}_a \pm 1/\sqrt{N}$ or $\hat{\rho}_a \pm 2/\sqrt{N}$, respectively.

Geweke (1992) proposes the widely used convergence diagnostic (CD) to test if a Markov chain actually has converged. The basic idea is to compare values sampled early in the sequence (subsample A) with those late in the sequence (subsample B). Denote by $\theta^{(i)}$ the i^{th} draw of parameter θ in the recorded N draws, and calculate $\bar{\theta}_A = \frac{1}{n_A} \sum_{i=1}^{n_A} \theta^{(i)}$, $\bar{\theta}_B = \frac{1}{n_B} \sum_{i=N-n_B+1}^{N} \theta^{(i)}$. Further, let σ_A and σ_B denote the standard errors of $\bar{\theta}_A$ and $\bar{\theta}_B$, respectively. The convergence diagnostic is then defined as

$$CD = \frac{\bar{\theta}_A - \bar{\theta}_B}{\sqrt{\sigma_A^2 + \sigma_B^2}} \xrightarrow{d} \mathcal{N}(0, 1),$$

under the assumption that sequence $\theta^{(i)}$ is stationary. For window lengths n_A and n_B Geweke (1992) proposes $0.1N$ and $0.5N$, respectively, which have become standard in the literature and are also employed in this work. With a corresponding estimate of the inefficiency factor in Eq. (30) at hand, variances σ_A^2, σ_B^2 are calculated as $\sigma_A^2 = s_A^2 (b_{w,A} + 1)^{-1} IF_A$, $\sigma_B^2 = s_B^2 (b_{w,B} + 1)^{-1} IF_B$, where s_A^2 and s_B^2 are the variances of subsamples A and B, respectively.

References

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

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

Asai, M. and M. McAleer (2006). Asymmetric multivariate stochastic volatility. *Econometric Reviews 25*(2-3), 453–473.

Asai, M. and M. McAleer (2009). The structure of dynamic correlations in multivariate stochastic volatility models. *Journal of Econometrics 150*(2), 182–192. Recent Development in Financial Econometrics.

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

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

Chan, D., R. Kohn, and C. Kirby (2006). Multivariate stochastic volatility models with correlated errors. *Econometric Reviews 2/3*(25), 245–274.

Chan, J. C.-C. and I. Jeliazkov (2009). McMC estimation of restricted covariance matrices. *Journal of Computational and Graphical Statistics 18*(2), 457–480.

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *Journal of the American Statistical Association 49*(4), 327–335.

Chib, S., F. Nardari, and N. Shephard (2006). Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics 134*(2), 341–371.

Chib, S., Y. Omori, and M. Asai (2009). Multivariate stochastic volatility. In T. G. Anderson, R. A. Davis, and J.-P. Kreis (Eds.), *Handbook of Fiancial Time Series*. New York, USA: Springer.

de Jong, P. (1991a). The diffuse Kalman filter. *The Annals of Statistics 19*(2), 1073–1083.

de Jong, P. (1991b). Stable algorithms for the state space model. *Journal of Time Series Analysis 12*(2), 143–157.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*(2), 339–350.

Doucet, A., N. de Freitas, and N. Gordon (2001). *Sequential Monte Carlo Methods in Practice*. New York, USA: Springer.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*(3), 603–615.

Durbin, J. and S. J. Koopman (2008). *Time Series Analysis by State Space Methods*. Oxford, UK: Oxford University Press.

Engle, R. F. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics 20*(3), 339–350.

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

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, Volume 4, pp. 169–188. New York, USA: Oxford University Press.

Gouriéroux, C. (2006). Continuous time Wishart process for stochastic risk. *Econometric Reviews 25*(2-3), 177–217.

Gouriéroux, C., J. Jasiak, and R. Sufana (2009). The Wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics 150*(2), 167–181. Recent Development in Financial Econometrics.

Grewal, M. S. and A. P. Andrews (2001). *Kalman Filtering: Theory and Practice*. New York, USA: John Wiley & Sons.

Han, Y. (2006). Asset allocation with a high dimensional latent factor stochastic volatility model. *Review of Financial Studies 19*(1), 237–271.

Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge, UK: Cambridge University Press.

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *The Review of Economic Studies 61*(2), 247–264.

Ishihara, T. and Y. Omori (2012). Efficient Bayesian estimation of a multivariate stochastic volatility model with cross leverage and heavy-tailed errors. *Computational Statistics & Data Analysis 56*(11), 3674–3689. 1st issue of the Annals of Computational and Financial Econometrics. Sixth Special Issue on Computational Econometrics.

Ishihara, T., Y. Omori, and M. Asai (2013). Matrix exponential stochastic volatility with cross leverage. Discussion paper series, CIRJE-F-904, Faculty of Economics, University of Tokyo.

Jacquier, E. and S. Miller (2012). The information content of realized volatility. Working paper, HEC Montreal.

Koopman, S. J. (1993). Disturbance smoother for state space models. *Biometrika 80*(1), 117–126.

Lopes, H. F. and C. M. Carvalho (2007). Factor stochastic volatility with time varying loadings and Markov switching regimes. *Journal of Statistical Planning and Inference 137*(10), 3082–3091. Special Issue: Bayesian Inference for Stochastic Processes.

Lopes, H. F., R. E. McCulloch, and R. S. Tsay (2013). Cholesky stochastic volatility models for high-dimensional time series. Discussion paper, Booth School of Business, The University of Chicago.

Magnus, J. R. and K. M. Abadir (2007). On some definitions in matrix algebra. Discussion paper, CIRJE-F-476, Faculty of Economics, University of Tokyo.

Nakajima, J. (2012a). Bayesian analysis of latent threshold models. PhD thesis, Duke University.

Nakajima, J. (2012b). Bayesian analysis of multivariate stochastic volatility with skew distribution. Discussion Paper Series 2013-01, Department of Statistical Science, Duke University.

Nakajima, J. and M. West (2013). Bayesian analysis of latent threshold dynamic models. *Journal of Business & Economic Statistics 31*(2), 151–164.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: Fast and efficient likelihood inference. *Journal of Econometrics 140*(2), 425–449.

Pemstein, D., K. M. Quinn, and A. D. Martin (2011). The Scythe statistical library: An open source C++ library for statistical computation. *Journal of Statistical Software 42*(12), 1–26.

Philipov, A. and M. E. Glickman (2006a). Multivariate stochastic volatility via Wishart processes. *Journal of Business and Economic Statistics 24*(3), 313–328.

Philipov, A. and M. E. Glickman (2006b). Factor multivariate stochastic volatility via Wishart processes. *Econometric Reviews 25*(2-3), 311–334.

Pitt, M. K. and N. Shephard (1999a). Filtering via simulation: Auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599.

Pitt, M. K. and N. Shephard (1999b). Analytic convergence rates and parameterization issues for the Gibbs sampler applied to state space models. *Journal of Time Series Analysis 20*(1), 63–85.

Pitt, M. K. and N. Shephard (1999c). Time varying covariances: A factor stochastic volatility approach (with discussion). In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Bayesian Statistics*, Volume 6, pp. 547–570. London, UK: Oxford University Press. Proceedings of the Sixth Valencia International Meeting.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*(3), 653–667.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64*(4), 583–639.

Taylor, S. J. (1994). Modeling stochastic volatility: A review and comparative study. *Mathematical Finance 4*(2), 183–204.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics 22*(4), 1701–1728.

Tsay, R. S. (2005). *Analysis of Financial Time Series: Financial Econometrics* (2nd ed.). New York, USA: John Wiley & Sons.

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard & Pitt (1997). *Biometrika 91*(1), 246–248.

Whiteley, N. and A. M. Johansen (2011). *Bayesian Time Series Models*. Cambridge, UK: Cambridge University Press.

Yu, J. (2005). On leverage in a stochastic volatility model. *Journal of Econometrics 127*(2), 165–178.