JOURNAL OF COMPUTATIONAL AND GRAPHICAL STATISTICS
2017, VOL. 26, NO. 4, 905–917
https://doi.org/10.1080/10618600.2017.1322091

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Efficient Bayesian Inference for Multivariate Factor Stochastic Volatility Models

Gregor Kastner [a], Sylvia Frühwirth-Schnatter[a], and Hedibert Freitas Lopes[b]

[a]Department of Finance, Accounting and Statistics, WU Vienna University of Economics and Business, Austria; [b]Insper, Brazil

ABSTRACT

We discuss efficient Bayesian estimation of dynamic covariance matrices in multivariate time series through a factor stochastic volatility model. In particular, we propose two interweaving strategies to substantially accelerate convergence and mixing of standard MCMC approaches. Similar to marginal data augmentation techniques, the proposed acceleration procedures exploit nonidentifiability issues which frequently arise in factor models. Our new interweaving strategies are easy to implement and come at almost no extra computational cost; nevertheless, they can boost estimation efficiency by several orders of magnitude as is shown in extensive simulation studies. To conclude, the application of our algorithm to a 26-dimensional exchange rate dataset illustrates the superior performance of the new approach for real-world data. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received February 2016
Revised December 2016

KEYWORDS
Ancillarity-sufficiency interweaving strategy (ASIS); Curse of dimensionality; Data augmentation; Dynamic correlation; Dynamic covariance; Exchange rate data; Markov chain Monte Carlo (MCMC)

1. Introduction

The analysis of multivariate time series has become a vivid research area over the last decades, where both methodological and computational advances have made it possible to estimate more and more complex models. In parallel, real-world applications with an ever-increasing amount of data call for the joint modeling of many simultaneous and often co-varying observations over time. However, already the number of pairwise co-movements increases quadratically with the number of time series, let alone higher-dimensional dependency structures. This property, often referred to as the *curse of dimensionality*, can often be mitigated in various ways by imposing a lower-dimensional latent factor structure, thereby effectively reducing the number of parameters to a feasible amount. In the article at hand, we particularly focus on the case where these factors are allowed to have time-varying variances which in turn drive the multivariate dynamics. To the best of our knowledge, models of this type have first been discussed by Jacquier, Polson, and Rossi (1994); Shephard (1996); Kim, Shephard, and Chib (1998). We particularly focus on the model formulation brought forward by Chib, Nardari, and Shephard (2006).

Applications of multivariate factor stochastic volatility models typically reside in the field of financial econometrics, most prominently in areas that involve accurate quantification of uncertainty and risk. Examples thereof are asset allocation (e.g., Aguilar and West 2000; Han 2006; Zhou, Nakajima, and West 2014) and asset pricing (e.g., Nardari and Scruggs 2007). These models extend standard factor pricing models such as the arbitrage pricing theory (Ross 1976) and the capital asset pricing model (Sharpe 1964; Lintner 1965) by relaxing the assumption that the multivariate volatility dynamics are constant over time.

Statistical estimation of these models can be challenging, and a variety of solutions such as quasi-maximum likelihood (e.g., Harvey, Ruiz, and Shephard 1994) or simulated maximum likelihood (e.g., Jungbacker and Koopman 2006; Liesenfeld and Richard 2006) have been proposed. For medium- to high-dimensional problems, Bayesian MCMC estimation (Pitt and Shephard 1999; Aguilar and West 2000; Chib, Nardari, and Shephard 2006; Han 2006; Omori et al. 2007) is probably the most efficient estimation method, however, it is associated with a considerable computational burden when the number of assets is moderate to large.

The aim of this work is to outline a reliable method for Bayesian inference that performs well for a wide range of datasets while at the same time being easy to implement and convenient to extend. Therefore, we combine an efficient method for estimating univariate stochastic volatility models introduced by Kastner and Frühwirth-Schnatter (2014) with a standard Gibbs sampler for regression problems. To ensure fast convergence and proper mixing of the MCMC chains, we augment this simple procedure with interweaving strategies brought forward by Yu and Meng (2011). Through extensive simulation studies and a real-world example, we demonstrate the effectiveness of our procedure which can boost sampling efficiency by a factor of 100 and more.

The remainder of this article is structured as follows. Section 2 establishes notation for the factor stochastic volatility model framework and discusses questions about model specification and identification. Section 3 gives an in-depth exposure to the estimation algorithm and its implementation, whereby the focus is placed on the novel interweaving strategies employed. Section 4 presents measures of sampling efficiency for simulated

CONTACT Gregor Kastner *gregor.kastner@wu.ac.at* Department of Finance, Accounting and Statistics, WU Vienna University of Economics and Business, Austria.
Color versions of one or more of the figures in the article can be found online at *www.tandfonline.com/r/JCGS*.
Supplementary materials for this article are available online. Please go to *www.tandfonline.com/r/JCGS*.

© 2017 The Authors. Published with license by Taylor & Francis Group, LLC
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (*http://creativecommons.org/licenses/by/4.0/*), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

datasets and compares the algorithms presented. Section 5 discusses a case study with 26 daily EUR exchange rates. Section 6 concludes.

2. The Multivariate Factor Stochastic Volatility Model

In a multivariate framework, the quadratic growth of the number of covariances alongside their inherent time-variability calls for a model which is sufficiently parsimoniously specified. At the same time, the model needs to be flexible enough to have the potential to capture typical features of financial and economic time series such as volatility clustering and volatility comovement. On top of that, common irregularities in the data require the model to be robust with respect to idiosyncratic shocks.

The multivariate factor stochastic volatility (SV) model (Chib, Nardari, and Shephard 2006) aims at uniting simplicity with flexibility and robustness. It is *simple* in the sense that the potentially high-dimensional observation space is reduced to a lower-dimensional orthogonal latent factor space, just like in the case of the classic factor model. It is *flexible* in the sense that these factors are allowed to exhibit volatility clustering, and it is *robust* in the sense that idiosyncratic deviations are themselves stochastic volatility processes, thereby allowing for the degree of volatility co-movement to be time-varying.

2.1. Model Specification

For each point in time $t = 1, \ldots, T$, let $\boldsymbol{y}_t = (y_{1t}, \ldots, y_{mt})'$ be a zero-mean vector of m observed returns and let $\boldsymbol{f}_t = (f_{1t}, \ldots, f_{rt})'$ be a vector of r unobserved latent factors. In analogy to the static factor model, the observations are assumed to be driven by the latent factors and the idiosyncratic innovations. In the case of the factor stochastic volatility model, however, both the idiosyncratic innovations as well as the latent factors are allowed to have time-varying variances, depending on $m + r$ latent volatilities $\boldsymbol{h}_t = (\boldsymbol{h}_t^U, \boldsymbol{h}_t^V)$, where $\boldsymbol{h}_t^U = (h_{1t}, \ldots, h_{mt})'$ and $\boldsymbol{h}_t^V = (h_{m+1,t}, \ldots, h_{m+r,t})'$. In short, we have

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{U}_t\left(\boldsymbol{h}_t^U\right)^{1/2}\boldsymbol{\epsilon}_t, \qquad \boldsymbol{f}_t = \boldsymbol{V}_t\left(\boldsymbol{h}_t^V\right)^{1/2}\boldsymbol{\zeta}_t, \tag{1}$$

where $\boldsymbol{\Lambda}$ is an unknown $m \times r$ factor loadings matrix, $\boldsymbol{U}_t(\boldsymbol{h}_t^U) = \text{diag}\,(\exp(h_{1t}), \ldots, \exp(h_{mt}))$ is a diagonal $m \times m$ matrix containing the idiosyncratic (series-specific) variances, and $\boldsymbol{V}_t(\boldsymbol{h}_t^V) = \text{diag}\,(\exp(h_{m+1,t}), \ldots, \exp(h_{m+r,t}))$ is a diagonal $r \times r$ matrix containing the factor variances. These variances are themselves modeled as latent variables whose logarithms follow independent autoregressive processes of order one, that is, for $i = 1, \ldots, m + r$:

$$h_{it} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + \sigma_i\eta_{it}, \tag{2}$$

with unknown initial value h_{i0}.

All innovations are assumed to follow independent standard normal distributions, that is, $\boldsymbol{\epsilon}_t \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{I}_m)$, $\boldsymbol{\zeta}_t \sim \mathcal{N}_r(\boldsymbol{0}, \boldsymbol{I}_r)$, and $\boldsymbol{\eta}_t \sim \mathcal{N}_{m+r}(\boldsymbol{0}, \boldsymbol{I}_{m+r})$, where $\boldsymbol{\eta}_t = (\eta_{1t}, \ldots, \eta_{m+r,t})'$. This implies the following structure:

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}\boldsymbol{f}_t + \boldsymbol{\epsilon}_t, \qquad \boldsymbol{f}_t|\boldsymbol{h}_t \sim \mathcal{N}_r\left(\boldsymbol{0}, \boldsymbol{V}_t\left(\boldsymbol{h}_t^V\right)\right), \tag{3}$$

with $\boldsymbol{\epsilon}_t|\boldsymbol{h}_t \sim \mathcal{N}_m(\boldsymbol{0}, \boldsymbol{U}_t(\boldsymbol{h}_t))$. One of the main reasons for estimating a factor SV model is to reliably estimate the potentially time-varying conditional covariance matrix of $\boldsymbol{y}_t$ which, for the model at hand, is given by $\text{cov}(\boldsymbol{y}_t|\boldsymbol{h}_t) = \boldsymbol{\Sigma}_t(\boldsymbol{h}_t) = \boldsymbol{\Lambda}\boldsymbol{V}_t(\boldsymbol{h}_t^V)\boldsymbol{\Lambda}' + \boldsymbol{U}_t(\boldsymbol{h}_t^U)$. Note that because $\boldsymbol{U}_t(\boldsymbol{h}_t^U)$ is diagonal, all covariances between the component series are governed by the latent factors. Marginally with respect to $\boldsymbol{h}_t$, $\boldsymbol{y}_t$ is a process with a non-Gaussian stationary distribution.

2.2. Identification Issues

Whenever certain combinations of parameter values result in (almost) identical maxima in the likelihood function, estimation of the corresponding parameter values from data can become impossible. Consequently, observationally equivalent parameter constellations must be ruled out for reliable statistical inference and a large body of literature dealing with this issue has arisen. In particular, Frühwirth-Schnatter and Lopes (2017) gave an overview of recent advances in the context of static Bayesian factor models, and Sentana and Fiorentini (2001) specifically discussed identification for models where the factors exhibit conditional heteroscedasticity (but innovations are assumed to be homoscedastic). If not dealt with properly, usually through certain restrictions on the parameter space, sensible interpretation of the posterior distribution is not possible ("nonidentifiability"). In less severe cases ("near-nonidentifiability"), MCMC algorithms and other estimation procedures often lack convergence and thus provide unreliable results. For the model at hand, we face several issues related to this problem.

First, to prevent factor rotation and column switching, one option is to follow the usual convention and set the upper triangular part of $\boldsymbol{\Lambda}$ to zero and diag $(\boldsymbol{\Lambda})$ nonzero (e.g., Geweke and Zhou 1996). Doing so, however, imposes an—often unwanted—order dependence. We therefore also discuss the possibility to leave the factor loadings matrix unrestricted and deal with column switching through post-processing of the MCMC draws.

Second, without identifying the scaling of either the jth column of $\boldsymbol{\Lambda}$ or the variance of f_{jt}, the model is not identified. The usual remedy (e.g., Aguilar and West 2000; Chib, Nardari, and Shephard 2006; Han 2006; Lopes and Carvalho 2007; Nakajima and West 2013; Zhou, Nakajima, and West 2014) is that the diagonal loading elements in model (3) are fixed to one, that is, $\Lambda_{jj} = 1$, for $j = 1, \ldots, r$, while the level μ_{m+j} of the factor volatilities $h_{m+j,t}$ in model (2) (which corresponds to the scaling of f_{jt}) is modeled to be unknown. This approach implies that the first r variables are leading the factors and thus makes variable ordering an even more important modeling decision. To alleviate this issue, we leave the diagonal elements Λ_{jj} in model (3) unrestricted, an intuitive interpretation being that "leadership" of a factor can be shared by several series. Instead, we fix the level μ_{m+j} of the factor volatilities $h_{m+j,t}$ at zero:

$$\begin{aligned} h_{it} &= (1 - \phi_i)\mu_i + \phi_i h_{i,t-1} + \sigma_i\eta_{it}, \quad i = 1, \ldots, m, \\ h_{m+j,t} &= \phi_{m+j}h_{m+j,t-1} + \sigma_{m+j}\eta_{m+j,t}, \quad j = 1, \ldots, r. \end{aligned} \tag{4}$$

This assumption, alongside the prior distribution on the loadings introduced in Section 3.1, identifies the factor variance.

Finally, each column of $\boldsymbol{\Lambda}$ is only identified up to a possible sign switch. We deal with this (lightweight) identification issue

a posteriori, meaning that we run our MCMC sampler in the unrestricted model and identify signs afterward, see Section A.4 in Appendix A for details.

Factor model (3) together with the $m + r$ SV models (4) defines our baseline parameterization, however alternative parameterizations will be exploited in Section 3.3 in the context of efficient MCMC estimation of the factor SV model.

3. Bayesian Inference

We perform Bayesian inference based on a set of carefully selected proper priors which are introduced in Section 3.1 and develop efficient schemes for full conditional MCMC sampling in the remaining subsections.

3.1. Prior Distributions

Independently for each $i \in \{1, \ldots, m + r\}$, priors for the univariate SV processes are chosen as in Kastner and Frühwirth-Schnatter (2014): $p(\mu_i, \phi_i, \sigma_i) = p(\mu_i)p(\phi_i)p(\sigma_i)$, where the level $\mu_i \in \mathbb{R}$ is equipped with the usual normal prior $\mu_i \sim \mathcal{N}(b_\mu, B_\mu)$, the persistence parameter $\phi_i \in (-1, 1)$ is chosen according to $(\phi_i + 1)/2 \sim \mathcal{B}(a_0, b_0)$ as in Kim, Shephard, and Chib (1998), and the volatility of log variance $\sigma_i \in \mathbb{R}^+$ is implied by $\sigma_i^2 \sim B_\sigma \times \chi_1^2 = \mathcal{G}(\frac{1}{2}, \frac{1}{2B_\sigma})$. The initial state h_{i0} is distributed according to the stationary distribution of the AR(1) process (2), that is, $h_{i0}|\mu_i, \phi_i, \sigma_i \sim \mathcal{N}(\mu_i, \sigma_i^2/(1 - \phi_i^2))$. For every unrestricted element of the factor loadings matrix, we choose independent zero-mean Gaussian distributions, that is, $\Lambda_{ij} \sim \mathcal{N}(0, B_\Lambda)$.

3.2. Full Conditional MCMC Estimation

Bayesian inference operates directly in the latent variable model (3) and (4) and relies on data augmentation by introducing the latent volatilities $\boldsymbol{h} = \{\boldsymbol{h}_{i,\bullet}\}$, $i = 1, \ldots, m + r$, where $\boldsymbol{h}_{i,\bullet} = (h_{i0}, h_{i1}, \ldots, h_{iT})'$, and the latent factors $\boldsymbol{f} = \{\boldsymbol{f}_{j,\bullet}\}$, $j = 1, \ldots, r$, where $\boldsymbol{f}_{j,\bullet} = (f_{j1}, \ldots, f_{jT})'$, as latent data. This allows to set up a simple scheme for full conditional MCMC sampling which is outlined in Algorithm 1 and discussed in detail thereafter.

Algorithm 1. Choose appropriate starting values for μ_i, $i \in \{1, \ldots, m\}$, ϕ_i and σ_i, $i \in \{1, \ldots, m + r\}$, as well as $\boldsymbol{\Lambda}$, $\boldsymbol{h}$, and $\boldsymbol{f}$ and repeat the following steps:

(a) Perform in total $m + r$ univariate SV updates of the m idiosyncratic variances $\boldsymbol{h}_{i,\bullet}$ as well as the parameters $(\mu_i, \phi_i, \sigma_i)$, independently for each $i = 1, \ldots, m$, and of the r factor variances $\boldsymbol{h}_{m+j,\bullet}$ as well as the parameters $(\phi_{m+j}, \sigma_{m+j})$, independently for each $j = 1, \ldots, r$.

(b) For $i = 1, \ldots, m$, sample each row $\boldsymbol{\Lambda}_{i,\bullet}$ of the factor loading matrix from $\boldsymbol{\Lambda}_{i,\bullet}|\boldsymbol{f}, \boldsymbol{y}_{i,\bullet}, \boldsymbol{h}_{i,\bullet}$. This step constitutes m independent $\tilde{r}_i$-variate regression problems with T observations, where $\tilde{r}_i$ denotes the number of unrestricted elements in $\boldsymbol{\Lambda}_{i,\bullet}$.

(b*) Redraw the diagonal elements of $\boldsymbol{\Lambda}$ through interweaving into the state equation for the latent factors (*shallow interweaving*) or through interweaving into the state equation for the latent volatilities (*deep interweaving*).

(c) For $t = 1, \ldots, T$, sample $\boldsymbol{f}_t$ from $\boldsymbol{f}_t|\boldsymbol{\Lambda}, \boldsymbol{y}_t, \boldsymbol{h}_t$, constituting T independent r-variate regression problems with m observations.

For Step (a), observe that conditional on knowing the latent factors $\boldsymbol{f}$ and the loadings $\boldsymbol{\Lambda}$, we are dealing with $m + r$ independent, univariate SV models where the latent state equations in (4) are combined with following observation equations:

$$\log(y_{it} - \boldsymbol{\Lambda}_{i,\bullet}\boldsymbol{f}_t)^2 = h_{it} + \log \epsilon_{it}^2, \qquad i = 1, \ldots, m, \quad (5)$$

$$\log f_{jt}^2 = h_{m+j,t} + \log \zeta_{jt}^2, \qquad j = 1, \ldots, r. \quad (6)$$

Hence, sampling the latent volatilities $\boldsymbol{h}_{i,\bullet}$ as well as the parameters $(\mu_i, \phi_i, \sigma_i)$ for $i = 1, \ldots, m + r$ (with $\mu_i = 0$ for $i > m$) in Step (a) amounts to $m + r$ univariate SV updates. Consequently, the substantial amount of research on this matter which has emerged in the last two decades can directly be applied. In particular, we follow recent findings in Kastner and Frühwirth-Schnatter (2014), where an efficient sampling scheme is proposed and evaluated, and simply use the implementation in the R package `stochvol` (Kastner 2016a) as a "plug-in" for Step (a) of the factor SV sampler presented in Algorithm 1; see Appendix A.1 for more details and additional references on MCMC estimation for univariate SV models.

On the other hand, conditional on knowing the latent volatilities $\boldsymbol{h}$, we are dealing in (3) with a factor model with heteroscedastic errors. Nevertheless, given $\boldsymbol{h}$, $\boldsymbol{f}$ and $\boldsymbol{\Lambda}$ may be sampled conditionally on each other from the respective multivariate normal distributions in a similar manner as for a standard factor model (Lopes and West 2004). This approach is conceptually straightforward, see Appendix A.2 for details how to sample in Step (b) each row $\boldsymbol{\Lambda}_{i,\bullet}$ of the factor loading matrix from $\boldsymbol{\Lambda}_{i,\bullet}|\boldsymbol{f}, \boldsymbol{y}_{i,\bullet}, \boldsymbol{h}_{i,\bullet}$, where $\boldsymbol{y}_{i,\bullet} = (y_{i1}, \ldots, y_{iT})'$, and Appendix A.3 for details how to sample in Step (c) the factor $\boldsymbol{f}_t$ from $\boldsymbol{f}_t|\boldsymbol{\Lambda}, \boldsymbol{y}_t, \boldsymbol{h}_t$ for $t = 1, \ldots, T$.

After discarding a certain amount of initial draws (the *burn-in*), the *standard full conditional sampler* iterates steps (a)–(c) of Algorithm 1, but not (b*), and should, in principle, yield draws from the joint posterior distribution. However, when estimating factor SV models through such an MCMC scheme, slow convergence and poor mixing (i.e., high correlation of posterior draws) can become a potentially prohibitive issue. This phenomenon substantiates in enormous autocorrelation of posterior draws—even after thinning—and can render MCMC output practically useless. For certain datasets, the burn-in phase may take extremely long and a huge amount of samples has to be discarded before the draws can be considered to emerge from the posterior distribution. Additionally, even after burn-in, these draws often show extraordinarily high autocorrelation and thus only explore the target distribution painstakingly slowly. These so-called *badly mixing* samplers do not only prolong computation time, they also frequently lead to unreliable estimates and misleading results. The simulation study in Section 4 illustrates that this can happen for the standard full conditional sampler even with data simulated from the true model, see, for example, the top of the two panels in Figure 1. Consequently, a carefully crafted posterior simulator is of utmost importance.

To overcome this problem, Chib, Nardari, and Shephard (2006) proposed to sample the factor loading matrix $\boldsymbol{\Lambda}$ from

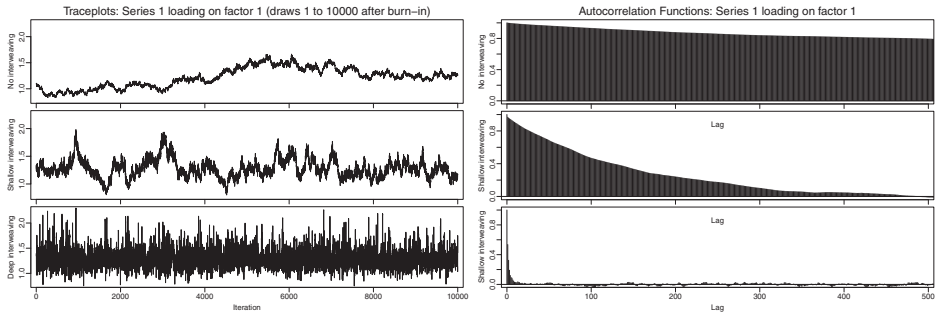

Figure 1. Trace plots of 10,000 draws from $p(\Lambda_{11}|\boldsymbol{y})$ (left-hand side) and empirical autocorrelation functions of all 5000000 draws (right-hand side) obtained via the standard sampler (top), shallow interweaving (middle), and deep interweaving (bottom).

the marginalized conditional posterior $p(\boldsymbol{\Lambda}|\boldsymbol{y}, \boldsymbol{h})$, without conditioning on the factors $\boldsymbol{f}$. This distribution, however, is not available in closed form, and to sample from it requires a rather involved Metropolis-Hastings update where the proposal distribution is based on numerically maximizing the often high-dimensional conditional likelihood function and approximating its Hessian matrix at every MCMC iteration. To avoid this potential bottleneck, we employ the simpler full conditional procedure outlined in Algorithm 1 but enhance it in Step (b*) by employing two variants of an ancillarity-sufficiency interweaving strategy (ASIS; Yu and Meng 2011), called *shallow interweaving* and *deep interweaving*, which are explained in detail in Section 3.3.

Applications to simulated data in Section 4 as well as to exchange rate data in Section 5 illustrate how adding Step (b*) boosts MCMC dramatically, in particular for deep interweaving; compare, for example, the top panel in Figure 1 to the remaining panels. Appendix A.5 provides comments on practical implementation of the boosted Algorithm 1 using the R package `factorstochvol` (Kastner 2016b).

3.3. Boosting Full Conditional MCMC Through Interweaving

As discussed in Section 3.2, the *standard full conditional sampler* outlined in Algorithm 1 is based on data augmentation in the parameterization (3) and (4) of the factor SV model and suffers from slow convergence like so many other MCMC schemes which alternate between sampling from the full conditionals of the latent states and the model parameters. A large literature has emerged discussing various techniques to improve such algorithms, in particular reparameterization (Papaspiliopoulos, Roberts, and Sköld 2007), marginal data augmentation (van Dyk and Meng 2001), and interweaving strategies (Yu and Meng 2011).

Reparameterization relies on data augmentation in a different parameterization of the model with alternative latent variables. In particular, so-called noncentered parameterizations, where unknown model parameters are moved from the latent state equation to the observation equation, proved to be useful, see, for example, Frühwirth-Schnatter and Wagner (2010) in the context of state-space modeling of time series. However, MCMC estimation based on different data augmentation schemes will often be efficient in separate regions of the parameter space, as demonstrated, for example, by Kastner and Frühwirth-Schnatter (2014) in the context of univariate SV models. This suggests to combine different data augmentation schemes to obtain an improved sampler.

Marginal data augmentation employs a randomly sampled "working parameter" to transform the baseline parameterization to an expanded, unidentified latent variable model in which the model parameters are updated conditional on the (randomly) transformed latent variables. This technique has been applied to the basic factor model, using the undefined scaling of the factors as a working parameter (Ghosh and Dunson 2009; Frühwirth-Schnatter and Lopes 2017), however, it is not easily extended to factor SV models, in particular if the latent volatilities should be part of the acceleration scheme.

The *ancillarity-sufficiency interweaving strategy (ASIS)*, introduced by Yu and Meng (2011), provides another principled way to interweave two different data augmentation schemes by resampling certain parameters conditional on the latent variables in an alternative parameterization of the model, thereby combining "best of different worlds." ASIS has been successfully employed in a variety of contexts such as univariate SV models (Kastner and Frühwirth-Schnatter 2014) and dynamic linear state-space models (Simpson 2015; Simpson, Niemi, and Roy 2017). To boost Algorithm 1, we apply ASIS to the factor SV model in the present article. Two interweaving strategies—called *shallow interweaving* and *deep interweaving*—are derived in Section 3.3.1, where the diagonal elements $\Lambda_{11}, \ldots, \Lambda_{rr}$ of the factor loadings matrix are resampled in Step (b*) in two alternative parameterizations of the model.

As will become clear in the following sections, deep interweaving typically yields the highest sampling efficiency gains and is thus the generally recommended strategy. However, also shallow interweaving has its merits. First, being conditionally conjugate, it is somewhat easier to implement. Second, it can be applied also to static factor models which are by construction not suited for deep interweaving (see also Bitto and Frühwirth-Schnatter 2016).

3.3.1. Shallow and Deep Interweaving

As discussed in Section 2.2, our baseline parameterization (3) and (4) is just one of several alternative ways to handle the scaling problem inherent in factor SV models and this identification issue is exploited by our schemes.

The parameterization underlying shallow interweaving constrains the diagonal elements of the factor loadings matrix to be equal to 1, whereas the variances of the factors depend on r unknown scaling parameters $\boldsymbol{D} = \text{diag}\,(\Lambda_{11}, \ldots, \Lambda_{rr})$.

The latent volatility processes are modeled as in the baseline parameterization (4), whereas the factor model takes a different form:

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}^\star \boldsymbol{f}_t^\star + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{f}_t^\star | \boldsymbol{h}_t, \Lambda_{11}, \ldots, \Lambda_{rr} \sim \mathcal{N}_r\big(\boldsymbol{0}, \boldsymbol{D}^2 \boldsymbol{V}_t\left(\boldsymbol{h}_t^V\right)\big), \tag{7}$$

with a lower triangular loading matrix $\boldsymbol{\Lambda}^\star$ where $\Lambda_{11}^\star = 1, \ldots, \Lambda_{rr}^\star = 1$. The idiosyncratic errors $\boldsymbol{\varepsilon}_t$ are distributed as in (3). Factor model (3) in the baseline parameterization can be transformed into factor model (7) through a simple linear transformation:

$$\boldsymbol{f}_t^\star = \boldsymbol{D}\boldsymbol{f}_t, \qquad t = 1, \ldots, T, \qquad \boldsymbol{\Lambda}^\star = \boldsymbol{\Lambda}\boldsymbol{D}^{-1}. \tag{8}$$

Boosting through shallow interweaving consists of three parts. First, transformation (8) is used to move the current posterior draws of the latent factors $\boldsymbol{f}_t$ and the factor loading matrix $\boldsymbol{\Lambda}$ from the baseline parameterization to parameterization (7). Second, the scale parameters $\Lambda_{11}, \ldots, \Lambda_{rr}$, contained in $\boldsymbol{D}$, are resampled in parameterization (7), conditionally on the transformed values $\boldsymbol{f}^\star$, from $p(\Lambda_{11}, \ldots, \Lambda_{rr} | \boldsymbol{f}^\star, \boldsymbol{\Lambda}^\star, \boldsymbol{h})$. Finally, the new values $\Lambda_{11}^{\text{new}}, \ldots, \Lambda_{rr}^{\text{new}}$ are used in transformation (8) to move $\boldsymbol{f}_t^\star$ and $\boldsymbol{\Lambda}^\star$ back to new draws $\boldsymbol{f}_t^{\text{new}}$ and $\boldsymbol{\Lambda}^{\text{new}}$ in the baseline parameterization.

It is evident from transformation (8) that shallow interweaving only affects the factors and the factor loading matrix, whereas the latent volatilities remain untouched. This is the feature that makes shallow interweaving also applicable to static factor models. However, to achieve boosting also for the r factor volatilities, deep interweaving is based on an alternative SV model for the factor volatilities, where the level is assumed to be unknown. The parameterization underlying deep interweaving relies on the factor model

$$\boldsymbol{y}_t = \boldsymbol{\Lambda}^\star \boldsymbol{f}_t^\star + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{f}_t^\star | \boldsymbol{h}_{m+j,\bullet}^\star \sim \mathcal{N}_r\big(\boldsymbol{0}, \text{diag}\,(e^{h_{m+1,t}^\star}, \ldots, e^{h_{m+r,t}^\star})\big), \tag{9}$$

where $\boldsymbol{\Lambda}^\star$ has the same structure as in the factor model (7) for shallow interweaving and the idiosyncratic errors $\boldsymbol{\varepsilon}_t$ are distributed as before, with the univariate SV models for the m underlying volatilities following (4). However, the r latent factor volatilities $h_{m+j,t}^\star$ follow alternative univariate SV models, where the level is $\mu_{m+j} = \log \Lambda_{jj}^2$ rather than zero:

$$h_{m+j,t}^\star = \mu_{m+j}(1 - \phi_{m+j}) + \phi_{m+j} h_{m+j,t-1}^\star + \sigma_{m+j}\eta_{m+j,t}. \tag{10}$$

This parameterization can be motivated by moving the parameters $\Lambda_{11}, \ldots, \Lambda_{rr}$ from factor model (7) into SV model (10), since:

$$f_{jt}^\star | \Lambda_{jj}, h_{m+j,t} \sim \mathcal{N}\left(0, \Lambda_{jj}^2 e^{h_{m+j,t}}\right) = \mathcal{N}\left(0, e^{\log \Lambda_{jj}^2 + h_{m+j,t}}\right) = \mathcal{N}\left(0, e^{h_{m+j,t}^\star}\right).$$

Hence, the baseline parameterization can be transformed into parameterization (9) and (10) by applying transformation (8) to the factors and the factor loadings, as well as the following transformation to the factor volatilities:

$$h_{m+j,t}^\star = h_{m+j,t} + \log \Lambda_{jj}^2, \quad t = 0, \ldots, T, \quad j = 1, \ldots, r. \tag{11}$$

Boosting through deep interweaving also consists of three parts. First, transformations (8) and (11) are used to move from the current draws of $\boldsymbol{f}_t$, $\boldsymbol{\Lambda}$ and the factor log-variances $h_{m+j,t}$ from the baseline parameterization to parameterization (9) and (10). Second, the scale parameters $\Lambda_{11}, \ldots, \Lambda_{rr}$ are resampled in parameterization (10) conditionally on the transformed values $\boldsymbol{h}_{m+j,\bullet}^\star = (h_{m+j,0}^\star, \ldots, h_{m+j,T}^\star)'$ from $p(\Lambda_{11}, \ldots, \Lambda_{rr} | \boldsymbol{h}_{m+1,\bullet}^\star, \cdots, \boldsymbol{h}_{m+r,\bullet}^\star, \boldsymbol{\Lambda}^\star)$. Third, based on the new values $\Lambda_{11}^{\text{new}}, \ldots, \Lambda_{rr}^{\text{new}}$, transformations (8) and (11) are inverted to move $\boldsymbol{f}_t^\star$, $\boldsymbol{\Lambda}^\star$, $h_{m+j,t}^\star$ back to new draws $\boldsymbol{f}_t^{\text{new}}$, $\boldsymbol{\Lambda}^{\text{new}}$, $h_{m+j,t}^{\text{new}}$ in the baseline parameterization.

Both interweaving strategies are summarized in Algorithm 2. Details on resampling $\Lambda_{jj}^{\text{new}}$ are provided in Section 3.3.2. It is evident that deep interweaving affects the factors, the factor loading matrix as well as the latent factor volatilities and for this reason is more effective in boosting MCMC for factor SV models than shallow interweaving.

Algorithm 2 (Shallow and Deep Interweaving). Denote the original posterior draws for $\boldsymbol{\Lambda}_{\bullet,j}$, $\boldsymbol{f}_{j,\bullet}$, and $\boldsymbol{h}_{m+j,\bullet}$ in Algorithm 1 by $\boldsymbol{\Lambda}_{\bullet,j}^{\text{old}}$, $\boldsymbol{f}_{j,\bullet}^{\text{old}}$, and $\boldsymbol{h}_{m+j,\bullet}^{\text{old}}$ and perform following steps independently for each $j = 1, \ldots, r$ in Step (b*):

(b*-1) Determine the vector $\boldsymbol{\Lambda}_{\bullet,j}^\star$, containing the k_j free parameters $\Lambda_{ij}^\star = \Lambda_{ij}^{\text{old}} / \Lambda_{jj}^{\text{old}}$ in the jth column of the transformed factor loading matrix $\boldsymbol{\Lambda}^\star$.

(b*-2) For shallow interweaving, define $\boldsymbol{f}_{j,\bullet}^\star = \Lambda_{jj}^{\text{old}} \boldsymbol{f}_{j,\bullet}^{\text{old}}$ and sample a new value $\Lambda_{jj}^{\text{new}}$ from $p(\Lambda_{jj} | \boldsymbol{\Lambda}_{\bullet,j}^\star, \boldsymbol{f}_{j,\bullet}^\star, \boldsymbol{h}_{m+j,\bullet})$. For deep interweaving, define $\boldsymbol{h}_{m+j,\bullet}^\star = \boldsymbol{h}_{m+j,\bullet}^{\text{old}} + 2\log|\Lambda_{jj}^{\text{old}}|$ and sample $\Lambda_{jj}^{\text{new}}$ from $p(\Lambda_{jj} | \boldsymbol{\Lambda}_{\bullet,j}^\star, \boldsymbol{h}_{m+j,\bullet}^\star, \phi_{m+j}, \sigma_{m+j})$; see Section 3.3.2 for details.

(b*-3) Update $\boldsymbol{\Lambda}_{\bullet,j}$, $\boldsymbol{f}_{j,\bullet}$, and, for deep interweaving, also $\boldsymbol{h}_{m+j,\bullet}$:

$$\boldsymbol{\Lambda}_{\bullet,j} = \frac{\Lambda_{jj}^{\text{new}}}{\Lambda_{jj}^{\text{old}}} \boldsymbol{\Lambda}_{\bullet,j}^{\text{old}}, \qquad \boldsymbol{f}_{j,\bullet} = \frac{\Lambda_{jj}^{\text{old}}}{\Lambda_{jj}^{\text{new}}} \boldsymbol{f}_{j,\bullet}^{\text{old}},$$

$$\boldsymbol{h}_{m+j,\bullet} = \boldsymbol{h}_{m+j,\bullet}^{\text{old}} + 2\log\left|\frac{\Lambda_{jj}^{\text{old}}}{\Lambda_{jj}^{\text{new}}}\right|.$$

3.3.2. Sampling the Scaling Parameters in the Alternative Representations

To derive the full conditional posterior distribution of Λ_{jj}, we combine the appropriate full conditional likelihood function with the Gaussian prior $\Lambda_{jj} \sim \mathcal{N}(0, B_\Lambda)$. In addition, the prior $\boldsymbol{\Lambda}_{\bullet,j}^\star | \Lambda_{jj}^2 \sim \mathcal{N}_{k_j}(0, B_\Lambda / \Lambda_{jj}^2 \boldsymbol{I}_{k_j})$ of the transformed factor loadings in column j contributes to the posterior distribution of Λ_{jj}^2 because its scale depends on Λ_{jj}^2.

For shallow interweaving, we sample Λ_{jj}^2 and define $\Lambda_{jj}^{\text{new}}$ as the square root of Λ_{jj}^2. Combining the likelihood obtained from factor model (7) with the implied prior $\Lambda_{jj}^2 \sim \mathcal{G}\,(1/2, 1/(2B_\Lambda))$

and $p(\boldsymbol{\Lambda}^{\star}_{\bullet,j}|\Lambda^2_{jj})$ yields

$$p\left(\Lambda^2_{jj}|\boldsymbol{\Lambda}^{\star}_{\bullet,j}, \boldsymbol{f}^{\star}_{j,\bullet}, \boldsymbol{h}_{m+j,\bullet}\right) \propto p\left(\boldsymbol{f}^{\star}_{j,\bullet}|\boldsymbol{h}_{m+j,\bullet}, \Lambda^2_{jj}\right) p\left(\boldsymbol{\Lambda}^{\star}_{\bullet,j}|\Lambda^2_{jj}\right) p\left(\Lambda^2_{jj}\right),$$

which is the product of T univariate Gaussian densities with Λ^2_{jj} appearing as part of the variance, k_j univariate Gaussian densities with Λ^2_{jj} appearing as part of the precision, and one Gamma density with Λ^2_{jj} appearing as argument. Thus, the resulting posterior distribution of Λ^2_{jj} is generalized inverse Gaussian, that is,

$$\Lambda^2_{jj}|\boldsymbol{\Lambda}^{\star}_{\bullet,j}, \boldsymbol{f}^{\star}_{j,\bullet}, \boldsymbol{h}_{m+j,\bullet} \sim \text{GIG}\left(\frac{1+k_j-T}{2}, \frac{1}{B_\Lambda}\left(1+(\boldsymbol{\Lambda}^{\star}_{\bullet,j})'\boldsymbol{\Lambda}^{\star}_{\bullet,j}\right), \sum_{t=1}^{T}\frac{f^{*2}_{jt}}{e^{h_{m+j,t}}}\right), \tag{12}$$

where $\text{GIG}(p, a, b)$ has a density proportional to $x^{p-1}\exp\{-\frac{1}{2}(ax+b/x)\}$. Given an efficient method to draw from the GIG such as the adaptive rejection sampling algorithm provided by Hörmann and Leydold (2013), sampling from (12) is straightforward. For practical implementation, we use the R package `GIGrvg` (Leydold and Hörmann 2015) which provides a C/C++ interface to avoid the cost of interpreting code at every MCMC iteration, thereby rendering the reupdating negligible in terms of overall computation time.

For deep interweaving, we sample Λ_{jj} indirectly through $\mu_{m+j} = \log\Lambda^2_{jj}$. Combining the implied prior $p(\mu_{m+j}) \propto \exp\{\mu_{m+j}/2 - e^{\mu_{m+j}}/(2B_\Lambda)\}$ with the likelihood obtained from SV model (10) and the priors $h^{\star}_{m+j,0}|\mu_{m+j}, \phi_{m+j}, \sigma^2_{m+j} \sim \mathcal{N}(\mu_{m+j}, \sigma^2_{m+j}/(1-\phi^2_{m+j}))$ and $\boldsymbol{\Lambda}^{\star}_{\bullet,j}|\mu_{m+j} \sim \mathcal{N}_{k_j}(0, B_\Lambda e^{-\mu_{m+j}}\boldsymbol{I}_{k_j})$ yields the posterior

$$p\left(\mu_{m+j}|\boldsymbol{\Lambda}^{\star}_{\bullet,j}, \boldsymbol{h}^{\star}_{m+j,\bullet}, \phi_{m+j}, \sigma^2_{m+j}\right) \propto p\left(\boldsymbol{h}^{\star}_{m+j,\bullet}|\mu_{m+j}, \phi_{m+j}, \sigma^2_{m+j}\right) p\left(\boldsymbol{\Lambda}^{\star}_{\bullet,j}|\mu_{m+j}\right) p(\mu_{m+j}),$$

which has a nonstandard form. To generate draws from this density, we consider an independence Metropolis-Hastings update in the spirit of Kastner and Frühwirth-Schnatter (2014). Since the likelihood $p(h^{\star}_{m+j,1}, \ldots, h^{\star}_{m+j,T}|h^{\star}_{m+j,0}, \mu_{m+j}, \phi_{m+j}, \sigma^2_{m+j})$ is the kernel of a Gaussian density in μ_{m+j}, it can be used to construct an auxiliary posterior under a conjugate auxiliary prior $p_{\text{aux}}(\mu_{m+j}|\sigma^2_{m+j}, \phi_{m+j}) \sim \mathcal{N}(0, B_0\sigma^2_{m+j}/(1-\phi_{m+j})^2)$ with B_0 large. Consequently, we draw a proposal μ^{prop}_{m+j} from the $\mathcal{N}(m^{\mu}_j, S^{\mu}_j)$ distribution with:

$$m^{\mu}_j = \frac{\sum_{t=1}^{T-1} h^{\star}_{m+j,t} + (h^{\star}_{m+j,T} - \phi_{m+j}h^{\star}_{m+j,0})/(1-\phi_{m+j})}{T + 1/B_0},$$

$$S^{\mu}_j = \frac{\sigma^2_{m+j}/(1-\phi_{m+j})^2}{T + 1/B_0}.$$

Denoting the old value of μ_{m+j} by μ^{old}_{m+j}, this proposal gets accepted with probability $\min(1, R)$, where

$$R = \frac{p(\boldsymbol{\Lambda}^{\star}_{\bullet,j}|\mu^{\text{prop}}_{m+j})p(h^{\star}_{m+j,0}|\mu^{\text{prop}}_{m+j}, \phi_{m+j}, \sigma^2_{m+j})p(\mu^{\text{prop}}_{m+j})}{p(\boldsymbol{\Lambda}^{\star}_{\bullet,j}|\mu^{\text{old}}_{m+j})p(h^{\star}_{m+j,0}|\mu^{\text{old}}_{m+j}, \phi_{m+j}, \sigma^2_{m+j})p(\mu^{\text{old}}_{m+j})} \times \frac{p_{\text{aux}}(\mu^{\text{old}}_{m+j}|\sigma^2_{m+j}, \phi_{m+j})}{p_{\text{aux}}(\mu^{\text{prop}}_{m+j}|\sigma^2_{m+j}, \phi_{m+j})}.$$

In case of acceptance, set $\Lambda^{\text{new}}_{jj} = e^{\mu^{\text{prop}}_{m+j}/2}$; otherwise, let $\Lambda^{\text{new}}_{jj} = \Lambda^{\text{old}}_{jj}$.

To conclude, we add two remarks. First, note that it is easy to combine both interweaving schemes within the MCMC sampler by daisy-chaining the corresponding steps. Second, note that any nonzero element of the jth factor column $\boldsymbol{\Lambda}_{\bullet,j}$ can be used to boost its mixing (not only the diagonal element Λ_{jj}). This is useful in particular when no loading matrix restrictions are enforced as it cannot be guaranteed that the diagonal elements are nonzero. Thus, in such situations, one could use a randomly selected (nonzero) element of each loadings column instead. Alternatively, one could also use the element whose absolute value is maximal.

4. Simulation Study

To compare the different algorithms in terms of sampling efficiency, a simple simulation experiment is conducted. We use $m = 10$ (simulated) series and $r = 2$ (simulated) factors to generate $T = 1000$ observations, thereby imposing the usual lower triangular constraint. The data-generating parameter values—listed in Table B.1 in Appendix B—are kept constant, whereas the data-generating process as well as the estimation procedure is repeated 100 times. Each time, the draws are initialized at the data-generating values; then, 5,100,000 draws are obtained of which 100,000 are discarded as burn-in. Prior hyperparameters are set as follows: $B_\Lambda = 1$, $b_\mu = 0$, $B_\mu = 100$, $a_0 = 20$, $b_0 = 1.5$, and $B_\sigma = 1$.

To gain insight about the mixing behavior of the different sampling strategies, trace plots (i.e., time series plots of the MCMC draws) for Λ_{11} are displayed in the left-hand panel of Figure 1. Even though the plots depict only the first 100,00 iterations after burn-in, it becomes very clear that the mixing of the noninterwoven sampler is extremely slow. The algorithm does not seem to explore the posterior distribution within a reasonable amount of draws which renders this output practically useless in terms of posterior inference. Moreover, the burn-in period for this sampler would need to be chosen extremely long to avoid strong dependence on the starting values. This situation is slightly mitigated when using shallow interweaving; nevertheless, mixing is still poor and for reliable posterior inference many draws are required. Turning toward the deeply interwoven sampler, one can observe quick mixing and hardly any visible autocorrelation.

Investigating autocorrelations of the draws via the empirical autocorrelation function confirms this picture; the right-hand panel of Figure 1 shows that the empirical autocorrelation function for draws from $p(\Lambda_{11}|\boldsymbol{y})$ decays very quickly for the sampler using deep interweaving which is not the case for the other two samplers, where visible autocorrelation remains even at large lags.

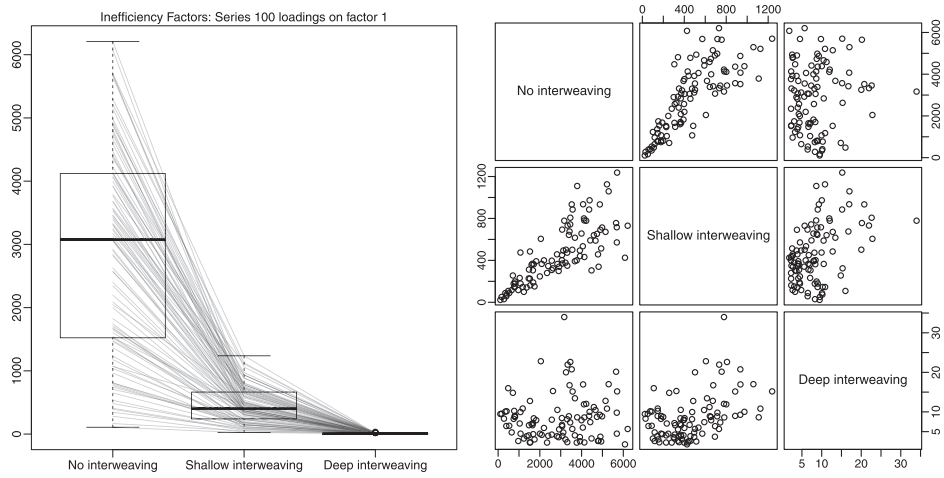

Figure 2. Left: boxplots of estimated inefficiency factors for posterior draws from $p(\Lambda_{11}|\boldsymbol{y}^{[i]})$ where $\boldsymbol{y}^{[i]}$, $i \in \{1, \ldots, 100\}$, denote artificially generated datasets whose underlying parameters are identical, see Table B.1 in Appendix B. Right: pairwise scatterplots thereof.

A convenient and common way of measuring sampling (in)efficiency is by means of the *inefficiency factor (IF)*, sometimes called the *(integrated) autocorrelation time*. It is defined as the ratio of the numerical variance of a statistic which is estimated from the Markov chain to the variance of that statistic when estimated from independent draws, thereby quantifying the relative loss of efficiency when inferring from correlated as opposed to independent samples. In other words, to achieve the same inferential accuracy about some posterior moment of some parameter as with k independent samples, IF $\times$ k MCMC draws are required. For the article at hand, we use the R package `coda` (Plummer et al. 2006) to estimate the inefficiency factors.

Moreover, when investigating performance of MCMC samplers through simulation studies, it is of great importance to take sample variation into account; even when identical parameter values are used for the generation of latent variables and data, sampling (in)efficiency may vary greatly. To illustrate this, we show box plots of the inefficiency factors stemming from repeated data-generating processes in the left panel of Figure 2. Note the enormous range for the standard sampler; depending on the data, IFs of 5000 or more are not uncommon, while at the same time IFs of around 100 can be observed. However, independently of the actual data, interweaving attenuates this effect drastically and increases efficiency uniformly. The right panel of Figure 2 shows pairwise scatter plots of these IFs. Note that shallow interweaving yields efficiency improvements which are more or less independent of the actual data (around five-fold for all datasets), whereas deep interweaving IFs appear less clearly correlated.

To provide a more complete picture, we list the inefficiency factors for all elements of $\boldsymbol{\Lambda}$ for the various algorithms in Table 1, averaged over all 100 runs. Note that shallow interweaving permits efficiency gains of around two- to eight-fold as opposed to the standard sampler, whereas deep interweaving delivers gains up to about 400-fold.

It comes as no surprise that sampling (in)efficiencies of draws for the volatility parameters μ_i, $i \in \{1, \ldots, m\}$ as well as ϕ_i and σ_i, $i \in \{1, \ldots, m + r\}$ are not affected substantially by this interweaving strategy, thus they are not reported here. It is however worth noting that the inefficiency of factor $\boldsymbol{f}_{j,\bullet}$ and factor log-variance draws $\boldsymbol{h}_{m+j,\bullet}$, $j \in \{1, \ldots, r\}$ may be influenced by bad mixing of $\boldsymbol{\Lambda}$. For illustration, IFs are reported for the final factors f_{1T} and f_{2T} and their log-variances $h_{m+1,T}$ and $h_{m+2,T}$ in Table 2.

To conclude the simulation exercise, we investigate the predictive performance of under- and overfitting models through cumulative log predictive Bayes factors in Figure 3; see Kastner (2017) for computational details. It stands out that the biggest predictive gain over a model that ignores contemporaneous correlations comes from introducing the first factor, that is, allowing for co-volatility through one common factor. Then, as expected, the second factor bumps the predictive score to its maximum. After that, it remains (almost) constant for three and more factors, irrespective of whether the lower triangular restriction is enforced or not. This points out that underfitting models are severly worse in terms of prediction while overfitting models hardly suffer from the extra parameters introduced.

Note that, similar to a scree plot in principal component analysis, Figure 3 can also be used as a graphical tool for finding the

Table 1. Average IFs for factor loadings matrix $\boldsymbol{\Lambda}$.

	(a) No interweaving			(b) Shallow interweaving			(c) Deep interweaving	
	1	2		1	2		1	2
1	2901.87		1	462.07		1	8.56	
2	2630.89	999.88	2	434.09	186.39	2	10.81	8.69
3	2931.02	233.08	3	460.59	80.57	3	8.48	10.92
4	2936.83	673.08	4	457.71	141.30	4	8.55	9.00
5	2909.98	876.57	5	451.84	164.85	5	8.79	8.46
6	2772.16	934.87	6	437.95	174.34	6	9.33	8.25
7	2303.80	958.02	7	406.73	178.91	7	10.38	8.19
8	1463.70	968.13	8	337.78	181.54	8	12.36	8.17
9	605.16	974.56	9	215.53	183.30	9	16.07	8.14
10	113.16	976.10	10	67.89	184.30	10	22.07	8.18

Table 2. Average IFs for the final factors f_{jT} and their log-variances $h_{m+j,T}$, $j \in \{1, 2\}$.

	$f_{1,1000}$	$f_{2,1000}$	$h_{11,1000}$	$h_{12,1000}$
No interweaving	121.54	53.30	232.23	21.59
Shallow interweaving	48.98	20.42	127.24	16.27
Deep interweaving	3.79	3.76	5.44	5.85

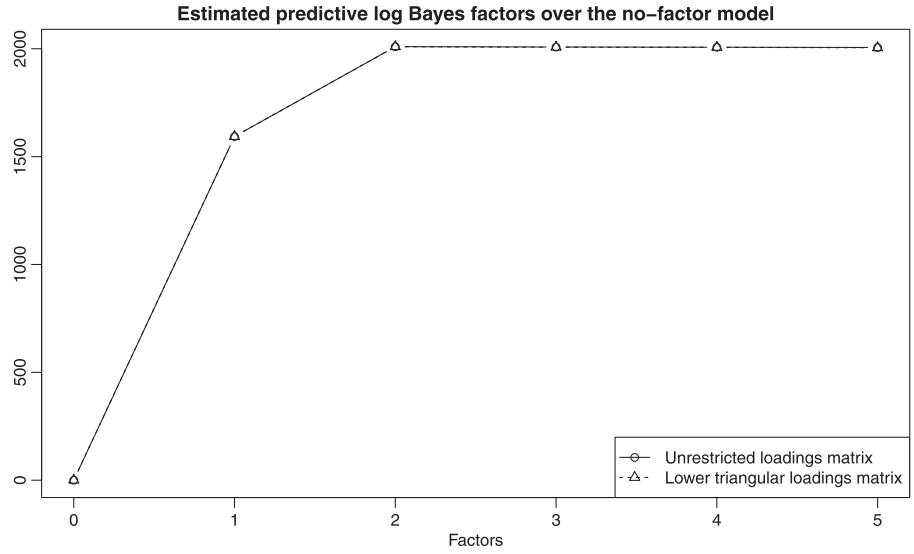

Figure 3. Log predictive Bayes factors in favor of the *r*-factor model over the no-factor model. The first 500 returns in the dataset are treated as prior information and log one-day-ahead predictive likelihoods are accumulated over the following 500 days. Circles connected with a solid line indicate values obtained with completely unrestricted loadings matrices; triangles connected with dashed lines indicate values where the loadings matrix is restricted to be lower triangular.

appropriate number of factors. For this exercise, we clearly find the true number of two factors.

5. Application to Exchange Rate Data

In this section, we analyze exchange rates with respect to EUR. Data were obtained from the European Central Bank's Statistical Data Warehouse and ranges from April 1, 2005 to August 6, 2015. It contains $m = 26$ (all which were available for this time frame) daily exchange rates on 2650 days listed in Table 3. For further analysis, we thus use $T = 2649$ demeaned log returns. The data are displayed in Figure C.1 in Appendix C. Common "stylized facts" of financial time series are clearly visible; note, for example, the obvious volatility clustering during 2008 and 2009 and again throughout late 2014 and early 2015. To put the robustness of our sampler to the test, we use the data as-is, that is, without excluding series containing extreme outliers such as the CHF spike on January 14, 2015 or the near collapse of RUB around December 16, 2014.

5.1. Model Specification

For selecting the number of factors in this application, it is important to keep in mind the primary purpose of the analysis. Different sampling strategies are applied, depending on whether identification of $\mathbf{\Lambda}$ is of no concern (e.g., for covariance matrix prediction only), or whether identification is instrumental for understanding the unobserved, underlying factors.

For the first case, we experimented with fitting unrestricted models to the exchange rates data. This implies that the method is completely invariant to series ordering and there are no model-implied "leading factors" as is usually the case. With respect to selecting the number of factors, we found that higher-order models without any restriction on the factor loadings matrix yield higher marginal likelihoods and are thus recommended. Figure 4 illustrates this via log predictive Bayes factors.

If identification is warranted, an important step is the appropriate ordering of the variables, before the usual lower triangular structure is imposed on the factor loadings matrix to guarantee mathematical identifiability, as outlined in Section 2.2. This, however, makes inference on the factor loadings matrix dependent on the appropriate ordering of the variables. We exemplify this by the predictive Bayes factors for models, where the component series are ordered alphabetically and the lower triangular structure is imposed on the first three series appearing in Table 3, namely AUD, CAD, and CHF. As shown in Figure 4, for a given number of factors *r*, the log predictive Bayes factors of the constrained models are consistently smaller than for the unrestricted models, indicating that a purely mathematical identifiability constraint may be in conflict with the data.

Also for the constrained models, the log predictive Bayes factors are ever increasing for the exchange rate data. However,

Table 3. Currency abbreviations.

AUD	Australia dollar	CAD	Canada dollar	CHF	Switzerland franc
CNY	China yuan renminbi	CZK	Czech R. koruna	DKK	Denmark krone
GBP	UK pound	HKD	Hong Kong dollar	HRK	Croatia kuna
HUF	Hungary forint	IDR	Indonesia rupiah	JPY	Japan yen
KRW	South Korea won	MYR	Malaysia ringgit	NOK	Norway krone
NZD	New Zealand dollar	PHP	Philippines peso	PLN	Poland zloty
RON	Romania fourth leu	RUB	Russia ruble	SEK	Sweden krona
SGD	Singapore dollar	THB	Thailand baht	TRY	Turkey lira
USD	US dollar	ZAR	South Africa rand		

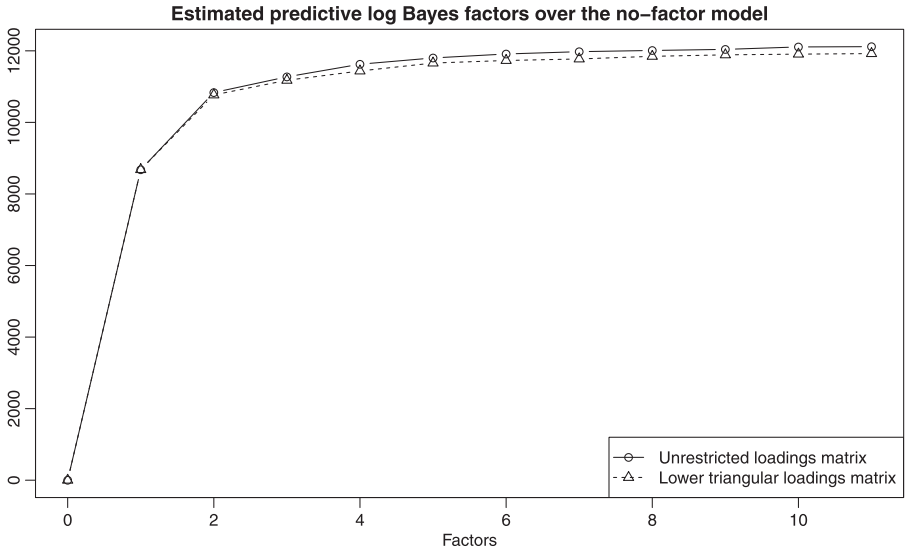

Figure 4. Log predictive Bayes factors in favor of the r-factor model over the no-factor model. The first 1000 returns in the dataset are treated as prior information and log one-day-ahead predictive likelihoods are accumulated over the following 1000 days. Circles connected with a solid line indicate values obtained with completely unrestricted loadings matrices; triangles connected with dashed lines indicate values, where the loadings matrix is restricted to be lower triangular. The component series are ordered alphabetically.

it stands out that the relative gain per additional factor are highest for few factors, flattening out quickly. Furthermore, draws of the factor loading matrix in these higher-order models are difficult to identify, in particular, if the lower triangular constraint is in conflict with the data and spurious factors, that is, factors which are significantly loaded on by only few series, are present (see Frühwirth-Schnatter and Lopes 2017, for a detailed discussion of this issue in the static factor context). Thus, to keep presentation feasible and to avoid spurious factors, we restrict ourselves to a model with $r = 4$ factors for the following in-depth discussion.

One way to find constraints that are not in conflict with the data is to post-process the MCMC draws of an unrestricted sampler with $r = 4$ factors, a method that has been applied in Conti et al. (2014) and Aßmann, Boysen-Hogrefe, and Pape (2016). Note that rather than reordering the variables before imposing a lower triangular constraint, we can choose three (out of the 26) currencies, and impose the $r(r-1)/2 = 6$ zero restrictions on the corresponding factor loadings, see, for example, Dunn (1973). While the choice of these currencies is not unique, inference is robust to specific choices, as long as the corresponding currencies serve as "leaders" for specific factors. This is exemplified by Figure C.3 which displays the posterior median of the MCMC draws of the factor loadings obtained from an unrestricted sampler with $r = 4$. To solve column switching, the columns of $\mathbf{\Lambda}$ are rearranged by the size of their maximum median loading. According to Figure C.3 in Appendix C, USD is a definite candidate to lead factor one, PLN leads a second factor, and AUD leads a third factor. Alternatively, identification could be based on any other currency strongly loading on factor 1 (such as HKD or CNY), in combination with HUF (instead of PLN) and NZD (instead of AUD).

Prior hyperparameters are the same as for the simulation study in Section 4. A sensitivity analysis shows that none of the hyperparameter choices turn out to be very influential in this particular application with the exception of the prior factor loadings variances B_Λ. These, however, are only important for the absolute scaling of the factors and do not notably influence the relative loadings sizes or predictive Bayes factors. We run each sampler for 550,000 iterations, discard the first 50,000 draws as burn-in, leaving 500,000 which we use for posterior inference. Even after this substantial amount of iterations, it is not clear that the sampler without interweaving has properly converged; we therefore omit its presentation. IFs from the interwoven samplers are presented in Table C.1 in Appendix C.

Finally, we identify the signs of the loadings in the post-processing phase by investigating the MCMC draws. For each factor, the series whose posterior absolute loadings distribution is furthest away from zero is assigned a positive sign, the other loadings are aligned thereafter, see also Section A.4 in Appendix A.

5.2. Posterior Factor Volatilities and Their Loadings

We begin by discussing the log-variances of the latent factors, visualized in Figure 5, alongside the corresponding factor loadings whose posterior distribution is depicted in Figure 6 and whose posterior means are listed in Table 4.

The first factor can clearly be interpreted as the USD-driven one, as the pegged triplet USD, CNY, and HKD loads very highly on this factor, alongside many other currencies. Its volatility is generally very smooth, rising in the aftermath of the 2008 financial crisis and going down again after 2009; a second increase can be seen in the second half of 2014, possibly in connection with the Greek government-debt crisis. Factor 2's log-variance appears slightly less persistent and more volatile, it is driven by ZAR, the only African currency in the sample, alongside Eastern Europe's/Southwestern Asia's HUF, PLN and TRY. Interestingly, JPY loads negatively on this factor.

The third factor shows a similar overall pattern as the first. The highest loading series for this factor are AUD and NZD, emphasizing the Trans–Tasman relations. Other commodity currencies such as ZAR and CAD also load highly on this factor.

Factor 4 is clearly driven by the currencies of the Tiger Cub economies such as MYR, KRW, PHP, and SGD.

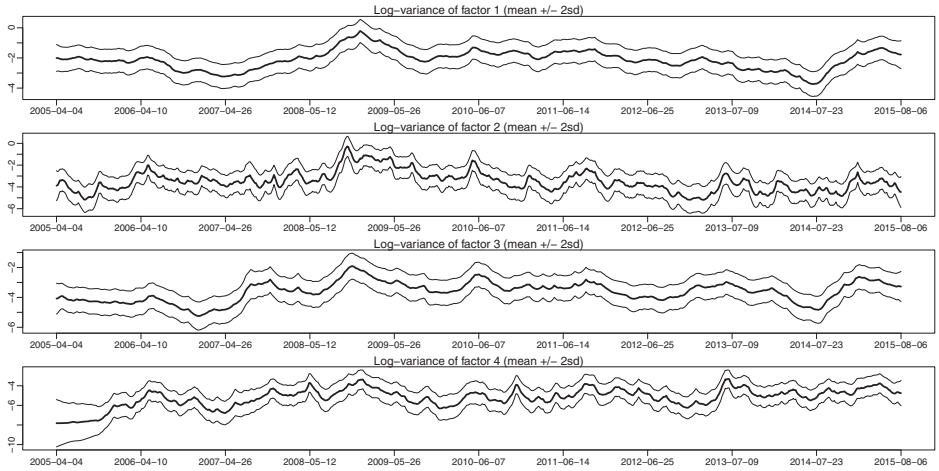

Figure 5. Marginal posteriors of the factor log-variances $h_{m+j,t}$, $j = 1, \ldots, 4$ (mean $\pm$ 2 $\times$ sd).

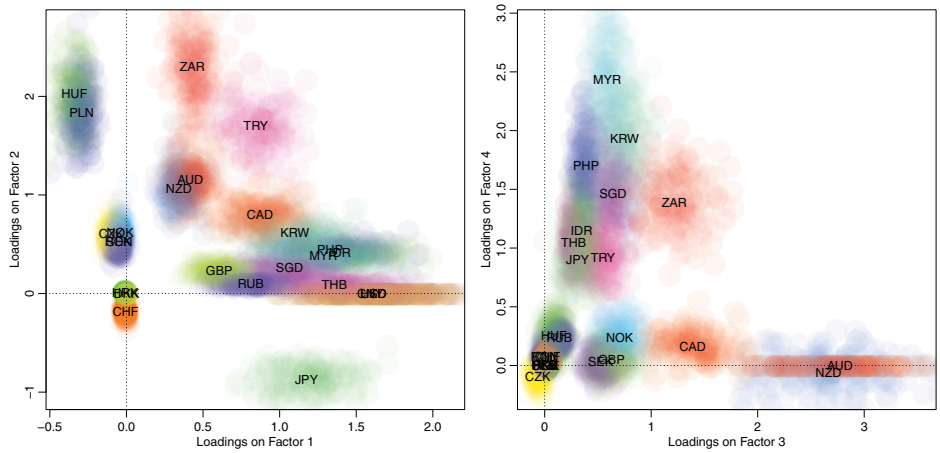

Figure 6. Marginal posterior distribution of the factor loadings, visualized through arbitrarily colored scatterplots of MCMC draws.

Table 4. Posterior means of $p(\mathbf{\Lambda}|\boldsymbol{y})$, in alphabetical order. Blank entries signify that the respective marginal distribution is not bound away from zero with at least 99% posterior probability. Starred entries are those which have been set to zero a priori.

	$\mathbf{\Lambda}_{\bullet,1}$	$\mathbf{\Lambda}_{\bullet,2}$	$\mathbf{\Lambda}_{\bullet,3}$	$\mathbf{\Lambda}_{\bullet,4}$		$\mathbf{\Lambda}_{\bullet,1}$	$\mathbf{\Lambda}_{\bullet,2}$	$\mathbf{\Lambda}_{\bullet,3}$	$\mathbf{\Lambda}_{\bullet,4}$
AUD	0.418	1.156	2.772	*	MYR	1.285	0.391	0.587	2.439
CAD	0.873	0.805	1.389		NOK		0.619	0.704	
CHF		−0.184			NZD	0.342	1.066	2.665	
CNY	1.592			0.076	PHP	1.330	0.449	0.389	1.702
CZK	− 0.099	0.605			PLN	− 0.292	1.835	*	*
DKK	0.002				RON	− 0.051	0.530		
GBP	0.605	0.230	0.627		RUB	0.813	0.104	0.138	0.237
HKD	1.611		0.003	0.005	SEK	− 0.049	0.529	0.527	
HRK					SGD	1.065	0.260	0.642	1.463
HUF	− 0.339	2.028			THB	1.358	0.092	0.273	1.049
IDR	1.395	0.419	0.347	1.153	TRY	0.845	1.702	0.549	0.920
JPY	1.176	− 0.875	0.310	0.904	USD	1.614	*	*	*
KRW	1.100	0.617	0.750	1.935	ZAR	0.431	2.303	1.219	1.390

5.3. Posterior Volatilities and Correlations

In order not to overload the graphical displays used to visualize the results of the analysis, we display the results for a two-year period only for the rest of this section. More specifically, we look at the years 2008 and 2009, covering the most volatile span during the financial crisis. Irrespectively of that, the full dataset has been used for estimation and other time spans could be displayed analogously.

We start out by visualizing the marginal posterior means of univariate volatilities for all 26 currencies in Figure 7 from the last day of 2007 until the last day of 2009. Series such as DKK or HRK are (very) closely pegged to EUR and unsurprisingly show very low volatility throughout the crisis. Other European currencies (CHF[1], RON, SEK,

[1] It is interesting to note that the Swiss franc stays comparably stable from a EUR perspective throughout 2007–2009. Very differently during Summer 2011, where

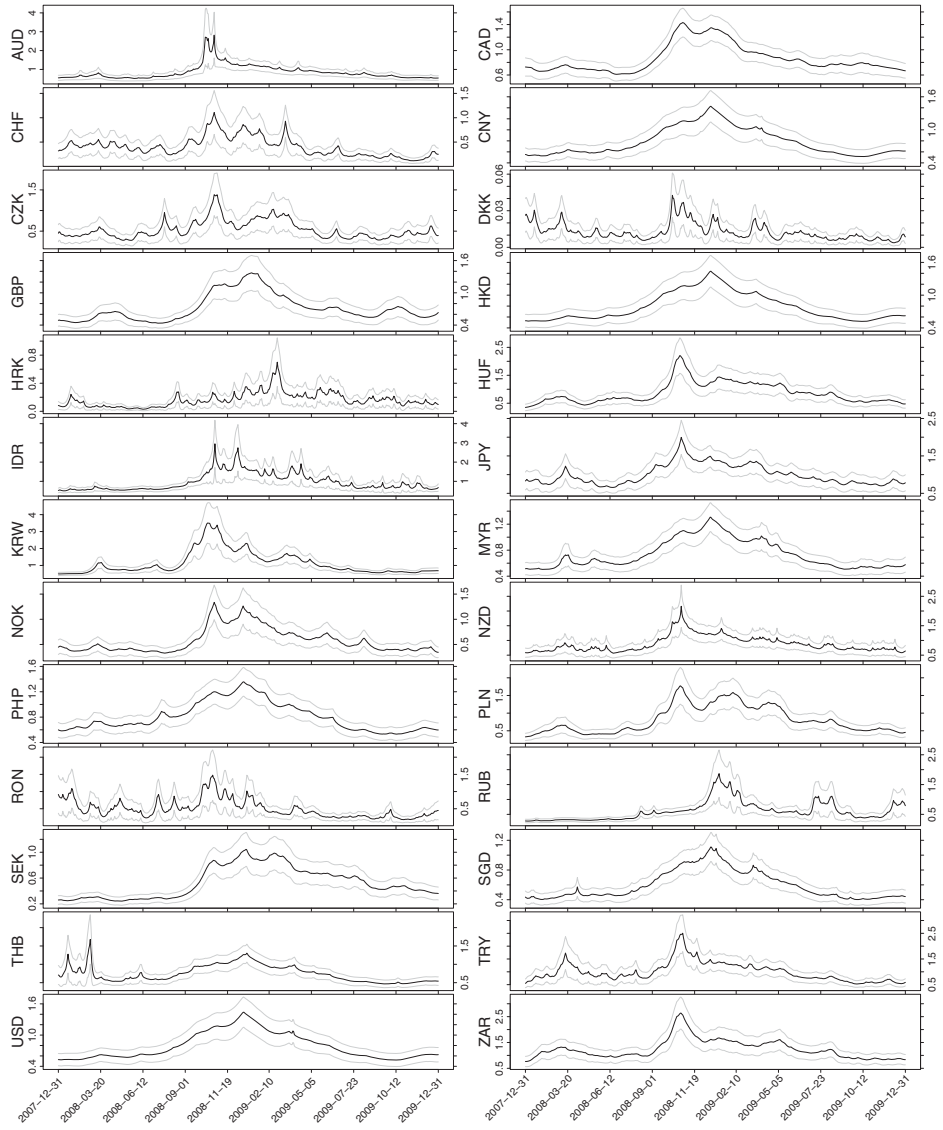

Figure 7. Posterior volatilities of exchange rate log returns with respect to EUR, from the last day of 2007 until the last day of 2009 (mean $\pm$ 2 $\times$ sd).

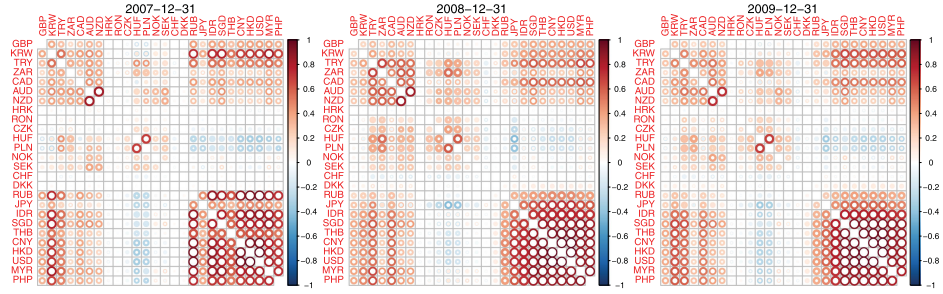

Figure 8. Posterior correlation matrices on the last trading days of 2006, 2007, and 2008. For each element, the size of the outer/inner circle is determined by the posterior mean plus/minus two posterior standard deviations, thereby indicating posterior uncertainty. Color and opacity are determined by the posterior mean. The remaining days are visualized in a video to be found online at *https://vimeo.com/212887492*.

CZK) follow suit. Tiger Cub economies such as PHP, HKD, THB, and MYR align very closely with USD and CNY. The most volatile currencies during this period are KRW, ZAR, IDR, JPY, and also TRY, followed by NZD, AUD, and CAD. Overall, it stands out that even though some series-specific ups and downs can be spotted, a common trend is clearly visible.

Next, implied correlation matrices are displayed in Figure 8, exemplified for the last day of 2007, 2008, and 2009. Additionally to displaying the mean posterior pairwise correlations (at the given dates) via color and shading, these plots visualize posterior uncertainty; the outer and inner circles' sizes

CHF shows atypical and very high volatility until the Swiss Central Bank sets the minimum exchange rate at CHF 1.20 per EUR 1 on September 6.

correspond to posterior mean $\pm$ 2 standard deviations, respectively. The images were generated using the R package `corrplot` (Wei 2016); its option `hclust` (hierarchical clustering) was used for ordering the series to emphasize the blocks of currencies.

To further illustrate variability over time, we determine the posterior means of the pairwise time-varying correlations of USD against the other currencies which are plotted in Figure C.2 in Appendix C. As was to be expected, correlations of CNY and HKD with USD are almost always very close to one; IDR, THB, SGD, MYR, and PHP show rather high correlation throughout. The correlation between USD and RUB on the other hand falls from around 0.9 in early 2008 to around 0.4 in late 2009, whereas THB moves in the opposite direction; its correlation with USD is around 0.5 at the beginning of the time window and increases quickly to around 0.9. Eastern European non-euro currencies, in particular PLN and HUF, appear to be slightly negatively correlation with USD throughout the entire period.

6. Conclusion and Outlook

Estimating time-varying (dynamic) covariance and correlation matrices of financial and economic time series constitutes a current and active area of research. One of the main challenges thereby is the curse of dimensionality, that is, the fact that the number of elements of these matrices grows quadratically with the number of observed series. We address this issue by imposing a low-dimensional latent factor structure where the factors are allowed to exhibit stochastic volatility and thereby govern co-movement of volatility over time. To conduct reliable statistical inference, we propose novel Bayesian MCMC algorithms which exploit the model-inherent identifiability constraints. By interweaving different (but mathematically equivalent) parameterizations, the proposed strategies substantially improve mixing of draws obtained from the posterior distribution, in particular for the factor loadings matrix. The method proposed is fully automatic in the sense that the end-user is not required to manually adjust any tuning parameters.

In an extensive case study discussing exchange rates with respect to EUR we show that the algorithm plays well with real-world data that exhibits a fair degree of outliers (e.g., CHF, RUB) which are captured through the idiosyncratic stochastic volatility components. The model structure allows for a covariance decomposition in four interpretable factors (USD/CNY driven, Eastern Europe, commodity currencies, Tiger Cub economies). These, alongside the idiosyncratic volatilities, drive the dynamics of the joint correlation structure. The pairwise correlations with USD range from "almost perfect" (CNY, HKD) over "hardly existent" (CHF, HRK) to "slightly negative" (PLN, HUF) with a varying and time-dependent degree of variability.

Concerning extensions of the model, we point out that due to the modular nature of MCMC, all the ideas of this article can be straightforwardly generalized to models that independently model the mean, be it through a simple nonzero mean vector, a local level model, external regressors, or via (vector) autoregressive processes. For models where the level of the returns explicitly depends on the (co-)volatilities (*volatility-in-mean*-type-effects, see, for example, Chan 2017) or the returns are assumed to be correlated with the (co-)volatilities (*leverage*-type-effects, see, for example, Ishihara and Omori 2017), more involved estimation methods are required; this unfortunately places their discussion outside the scope of this article. Nevertheless, due to the growing number of successful applications of interweaving methods in different contexts, there is good reason to hope for similar effects when they are used for these type of factor SV model extensions.

Supplementary Materials

Web Appendix containing (A) details of the various sampling steps of Algorithm 1 and details concerning the R package `factorstochvol`, (B) the data-generating parameter values for the simulation study in Section 4, and (C) further results for the exchange rate data discussed in Section 5. Available at https://arxiv.org/abs/1602.08154. (.pdf file)

Video displaying the time-varying conditional correlation matrix distribution for the full dataset (see Figure 8). Available at *https://vimeo.com/212887492*. (.avi file)

R-package `factorstochvol`, version 0.8.3, containing code to run the samplers described in the article. Available at *https://cran.r-project.org/package=factorstochvol*. (GNU zipped tar file)

Replication code for the results in the article, including the exchange rate data. (GNU zipped tar file)

ORCID

Gregor Kastner http://orcid.org/0000-0002-8237-8271

References

Aguilar, O., and West, M. (2000), "Bayesian Dynamic Factor Models and Portfolio Allocation," *Journal of Business & Economic Statistics*, 18, 338–357. [905,906]

Aßmann, C., Boysen-Hogrefe, J., and Pape, M. (2016), "Bayesian Analysis of Static and Dynamic Factor Models: An Ex-post Approach Toward the Rotation Problem," *Journal of Econometrics*, 192, 190–206. [913]

Bitto, A., and Frühwirth-Schnatter, S. (2016), Achieving Shrinkage in a Time-Varying Parameter Model Framework. arxiv:1611.01310v1. [908]

Chan, J. (2017), "The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling," *Journal of Business & Economic Statistics*, 35, 17–28. [916]

Chib, S., Nardari, F., and Shephard, N. (2006), "Analysis of High-Dimensional Multivariate Stochastic Volatility Models," *Journal of Econometrics*, 134, 341–371. [905,906,907]

Conti, G., Frühwirth-Schnatter, S., Heckman, J. J., and Piatek, R. (2014), "Bayesian Exploratory Factor Analysis," *Journal of Econometrics*, 183, 31–57. [913]

Dunn, J. E. (1973), "A Note on Sufficiency Condition for Uniqueness of a Restricted Factor Matrix," *Psychometrika*, 38, 141–143. [913]

Frühwirth-Schnatter, S., and Lopes, H. F. (2017), "Parsimonious Bayesian Factor Analysis when the Number of Factors is Unknown," Technical report. [906,908,913]

Frühwirth-Schnatter, S., and Wagner, H. (2010), "Stochastic Model Specification Search for Gaussian and Partial Non-Gaussian State-Space Models," *Journal of Econometrics*, 154, 85–100. [908]

Geweke, J., and Zhou, G. (1996), "Measuring the Pricing Error of the Arbitrage Pricing Theory," *The Review of Financial Studies*, 9, 557–587. [906]

Ghosh, J., and Dunson, D. B. (2009), "Default Prior Distributions and Efficient Posterior Computation in Bayesian Factor Analysis," *Journal of Computational and Graphical Statistics*, 18, 306–320. [908]

Han, Y. (2006), "Asset Allocation with a High-Dimensional Latent Factor Stochastic Volatility Model," *Review of Financial Studies*, 19, 237–271. [905,906]

Harvey, A. C., Ruiz, E., and Shephard, N. (1994), "Multivariate Stochastic Variance Models," *Review of Economic Studies*, 61, 247–264. [905]

Hörmann, W., and Leydold, J. (2013), "Generating Generalized Inverse Gaussian Random Variates," *Statistics and Computing*, 24, 1–11. [910]

Ishihara, T., and Omori, Y. (2017), "Portfolio Optimization using Dynamic Factor and Stochastic Volatility: Evidence on Fat-Tailed Error and Leverage," *The Japanese Economic Review*, 68, 63–94. [916]

Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business & Economic Statistics*, 12, 371–389. [905]

Jungbacker, B., and Koopman, S. J. (2006), "Monte Carlo Likelihood Estimation for Three Multivariate Stochastic Volatility Models," *Econometric Reviews*, 25, 385–408. [905]

Kastner, G. (2016a), "Dealing with Stochastic Volatility in Time Series using the R Package stochvol," *Journal of Statistical Software*, 69, 1–30. [907]

——— (2016b), `factorstochvol`: Bayesian Estimation of (Sparse) Latent Factor Stochastic Volatility Models, R package version 0.8.3. [908]

——— (2017), "Sparse Bayesian Time-Varying Covariance Estimation in Many Dimensions," arXiv:1608.08468. [911]

Kastner, G., and Frühwirth-Schnatter, S. (2014), "Ancillarity-Sufficiency Interweaving Strategy (ASIS) for Boosting MCMC Estimation of Stochastic Volatility Models," *Computational Statistics and Data Analysis*, 76, 408–423. [905,907,908,910]

Kim, S., Shephard, N., and Chib, S. (1998), "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65, 361–393. [905,907]

Leydold, J., and Hörmann, W. (2015), `GIGrvg`: Random Variate Generator for the GIG Distribution, R package version 0.4. [910]

Liesenfeld, R., and Richard, J.-F. (2006), "Classical and Bayesian Analysis of Univariate and Multivariate Stochastic Volatility Models," *Econometric Reviews*, 25, 335–360. [905]

Lintner, J. (1965), "The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets," *The Review of Economics and Statistics*, 47, 13–37. [905]

Lopes, H. F., and Carvalho, C. M. (2007), "Factor Stochastic Volatility with Time Varying Loadings and Markov Switching Regimes," *Journal of Statistical Planning and Inference*, 137, 3082–3091. [906]

Lopes, H. F., and West, M. (2004), "Bayesian Model Assessment in Factor Analysis," *Statistica Sinica*, 14, 41–67. [907]

Nakajima, J., and West, M. (2013), "Dynamic Factor Volatility Modeling: A Bayesian Latent Threshold Approach," *Journal of Financial Econometrics*, 11, 116–153. [906]

Nardari, F., and Scruggs, J. T. (2007), "Bayesian Analysis of Linear Factor Models with Latent Factors, Multivariate Stochastic Volatility, and APT Pricing Restrictions," *Journal of Financial and Quantitative Analysis*, 42, 857–891. [905]

Omori, Y., Chib, S., Shephard, N., and Nakajima, J. (2007), "Stochastic Volatility with Leverage: Fast and Efficient Likelihood Inference," *Journal of Econometrics*, 140, 425–449. [905]

Papaspiliopoulos, O., Roberts, G., and Sköld, M. (2007), "A General Framework for the Parameterization of Hierarchical Models," *Statistical Science*, 22, 59–73. [908]

Pitt, M. K., and Shephard, N. (1999), "Time-Varying Covariances: A Factor Stochastic Volatility Approach," in *Bayesian Statistics 6—Proceedings of the Sixth Valencia International Meeting*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford: Oxford University Press, pp. 547–570. [905]

Plummer, M., Best, N., Cowles, K., and Vines, K. (2006), "CODA: Convergence Diagnosis and Output Analysis for MCMC," *R News*, 6, 7–11. [911]

Ross, S. A. (1976), "The Arbitrage Theory of Capital Asset Pricing," *Journal of Economic Theory*, 13, 341–360. [905]

Sentana, E., and Fiorentini, G. (2001), "Identification, Estimation and Testing of Conditionally Heteroskedastic Factor Models," *Journal of Econometrics*, 102, 143–164. [906]

Sharpe, W. F. (1964), "Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk," *The Journal of Finance*, 19, 425–442. [905]

Shephard, N. (1996), "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models in Econometrics, Finance and Other Fields* eds. D. R. Cox, O. E. Barndorff-Nielsen, and D. V. Hinkley, London: Chapman & Hall, pp. 1–67. [905]

Simpson, M. (2015), "Application of Interweaving in DLMs to an Exchange and Specialization Experiment," in *Bayesian Statistics from Methods to Models and Applications—Research from BAYSM 2014, vol. 126 of Springer Proceedings in Mathematics & Statistics*, eds. A. Frühwirth-Schnatter Bitto, G. Kastner, and A. Posekany, New York: Springer, pp. 75–90. [908]

Simpson, M., Niemi, J., and Roy, V. (2017), "Interweaving Markov Chain Monte Carlo Strategies for Efficient Estimation of Dynamic Linear Models," *Journal of Computational and Graphical Statistics*, 26, 152–159. [908]

van Dyk, D., and Meng, X.-L. (2001), "The Art of Data Augmentation," *Journal of Computational and Graphical Statistics*, 10, 1–50. [908]

Wei, T. (2016), `corrplot`: Visualization of a Correlation Matrix, R package version 0.77. [916]

Yu, Y., and Meng, X.-L. (2011), "To Center or not to Center: That is not the Question—An Ancillarity-Suffiency Interweaving Strategy (ASIS) for Boosting MCMC Efficiency," *Journal of Computational and Graphical Statistics*, 20, 531–570. [905,908]

Zhou, X., Nakajima, J., and West, M. (2014), "Bayesian Forecasting and Portfolio Decisions using Dynamic Dependent Sparse Factor Models," *International Journal of Forecasting*, 30, 963–980. [905,906]