# MULTIVARIATE STOCHASTIC VOLATILITY USING THE HESSIAN METHOD

Abstract. We propose a new method for the analysis of multivariate stochastic volatility models, based on efficient draws of volatility from its conditional posterior distribution. It applies to models with several kinds of cross-sectional dependence. Full VAR autoregression and covariance matrices give cross-sectional volatility dependence. Mean factor structure allows conditional correlations, given states, to vary in time and covary with conditional variances; factors are Student's $t$ with factor-specific degrees of freedom. Given factors, returns have heterogeneous Student's $t$ marginals and a copula completes their joint distribution. We draw each volatility series as a block, one series at a time, using the HESSIAN method of McCausland (2012). Using daily returns data for ten currencies, we show that all features of the model are important.

Key words: Bayesian analysis, Factor models, MCMC, State space models

## 1. Introduction

Multivariate volatility models are powerful tools. Different kinds of static and dynamic cross-sectional dependence among asset returns capture different stylized facts.

1.1. **Stylized facts and their significance.** Asset return volatility varies over time, in response to news and revised expectations of future value. It tends to cluster, so that large price changes tend to be followed by other large changes. There is cross-sectional conditional dependence of volatility across markets and assets, and this dependence is time-varying. Cross-sectional correlations increase in periods of high market volatility, especially in bear markets. The distribution of returns has heavier tails than the normal distribution; this remains true however much one tries to condition on current information. There is an asymmetric relation between price and volatility changes known as the "leverage effect", according to which increases in volatility are associated more with large decreases in price than with large increases. These stylized facts are documented in Cont (2001), for the univariate case, and in Christodoulakis (2007) for the multivariate.

Multivariate volatility models that can capture these empirical regularities have many important applications, especially in modern portfolio management. Learning about the joint distribution of asset returns is a key element for the evaluation and construction of portfolios. Accurate estimation of the conditional dependence in a cross section of returns allows investors to identify opportunities or risks associated with particular portfolios, especially during periods of market stress. Financial crises usually have a strong impact on correlation: as the risk of some assets increases, investors wish to sell other risky assets, which leads to more highly correlated returns. The unfortunate consequence is that diversification is least effective at reducing risk at the very times when that risk is highest.

1.2. **Multivariate volatility models.** Two difficulties arise when we extend volatility models to the multivariate case. First, the conditional variance of returns given states must be positive definite at every point in time. Second, there is a severe trade-off between parsimony

*Date*: Current version: August 6, 2019.

and flexibility, as the number of parameters can increase quickly in the number of assets. Restraining the number of parameters or using informative priors can mitigate the danger of overfitting. Much of the difference between multivariate models reflects a choice about how to do this. This has implications on which stylized facts can be captured by the model.

As with univariate volatility models, there are two main types of multivariate volatility models: observation-driven and parameter-driven. In observation-driven models such as GARCH, volatility is a deterministic function of observed variables, which allows straightforward evaluation of the likelihood function. This advantage has made the GARCH model and its extensions popular for univariate and multivariate problems alike.

In parameter-driven volatility models, known as stochastic volatility (SV) models, volatility is a latent process. These models are more natural discrete time representations of the continuous time models often used in asset pricing and upon which much of modern finance theory is based. See Eraker and Wang (2015), for example, on the use of continuous time models for asset pricing. The recent literature in macroeconomics introducing conditional heteroscedasticity into Dynamic Stochastic General Equilibrium (DSGE) models—see, for example, Justiniano (2008) and Caldara, Fernández-Villaverde, Rubio-Ramírez, and Yao (2012)—heavily favours SV models: agents' behaviour in these models reflects their uncertainty about future volatility, by design. Several papers give empirical evidence in favour of SV models over observation-driven models: Kim, Shephard, and Chib (1998), Jacquier, Polson, and Rossi (1994), Geweke (1994), Carnero, Pena, and Ruiz (2004) and Chan and Grant (2016) in the univariate case and Danielsson (1998) in the multivariate case.

Likelihood evaluation in parameter-driven models, which amounts to high-dimensional integration over latent states, is difficult. But it is not necessary in most Bayesian approaches. It suffices to be able to evaluate the joint density of returns, states and parameters, a known function. Since the introduction of Bayesian Markov chain Monte Carlo (MCMC) methods by Jacquier, Polson, and Rossi (1994) for univariate SV models, inference for these models has become much more feasible.

This paper focuses on (parameter-driven) Multivariate SV (MSV) models. For a review of (observation-driven) multivariate GARCH models, see Bauwens, Laurent, and Rombouts (2006). We propose new MCMC methods for Bayesian analysis of MSV models, based on efficient draws of volatility from its conditional posterior distribution.

We model daily returns, but our methods could be extended to models where there is not only a measurement equation for returns but also for realized measures such as realized volatilities and realized covariances. Such models combine complementary information: daily returns, that are less subject to microstructure noise, with realized measures that incorporate more frequently observed data. Some recent articles of this type following a stochastic volatility approach are Shirota, Omori, Lopes, and Piao (2015), Venter and de Jongh (2014), Shirota, Hizu, and Omori (2014) and Koopman and Scharth (2013). See also Jin and Maheu (2013), following a multivariate GARCH approach; and Liu and Maheu (2015) and Jin and Maheu (2016), Markov switching approaches.

1.3. **Copulas in multivariate volatility models.** A copula is a multivariate distribution with uniform marginals. Sklar (1959) showed that any continuous multivariate distribution can be expressed, uniquely, in terms of its marginals and a copula. For the purposes of specification and estimation, one can decouple the marginals from other features of the joint distribution. Copulas have proven useful for multivariate volatility modelling: all the research and development that went into, and continues to go into univariate volatility models can be carried over to the multivariate case by combining existing univariate models, series by series, using copulas. See Patton (2009) for an overview of the application of copulas in the modelling of financial time series and Kolev, dos Anjos, and de M. Mendez (2006) for a survey and contributions to copula theory.

Typically, estimation is performed in two steps: The first step is point estimation of the parameters of the marginals, series by series, followed by the computation of probability integral transforms, often by way of standardized residuals. The second step is estimation of the parameters of the copulas relating, at each time period, residuals over the cross-section; the copula may or may not be time-varying. Examples include Chollete, Heinen, and Valdesogo (2009), Hafner and Manner (2012), Min and Czado (2010) and Oh and Patton (2013). Two-step estimation is easier than joint estimation, whether one uses some optimization criterion such as maximum likelihood, or Bayesian inference. However, joint estimation shares information across series; all the available data inform local estimation. From a frequentist perspective, this translates to greater estimation efficiency. From a Bayesian perspective, conditioning on more data allows one to learn more about local parameters and latent variables. And, what is more serious, second step estimation cannot be truly Bayesian, due to the elimination of uncertainty in the first step. (No analogous problem arises in the frequentist case if one analyses the sampling variability of the two-step estimator.)

Here, we use Gibbs sampling to simulate the full joint posterior distribution, isolating the copula parameters in a Gibbs block. In this way, we retain some of the advantages copulas offer in decoupling marginals from other features of the joint distribution, without giving up the advantages of joint estimation. However, it does raise some computational difficulties; we identify these below and show how we overcome them.

1.4. **Features of our model and their significance.** We propose a factor model for $p$ observed return series, with $q$ factor series. For each of these $m = p + q$ series, a SV series describes its conditional variance. Our inferential methods, described below, allow for a combination of features that is difficult or intractable using other methods. First, factors do not need to be multivariate Gaussian, nor a mixture of these such as the popular multivariate Student's $t$. The multivariate Student's $t$ is a scale mixture of Gaussians, with all variates scaled by the same draw from the mixing distribution. Thus not only are marginals Student's $t$ with the same degrees of freedom, variates tend to have extreme values at the same time. Second, volatility factors can be statistically dependent: the vector of log volatilities is a first-order Gaussian VAR, with full autoregression and variance matrices. Third, there can be conditional cross-sectional dependence across returns, given factors, which we model using copulas. Copulas allow us to represent a multivariate distribution of innovations in a cross-section in a very flexible way, by decoupling the choice of marginal distributions—which we allow to be different from each other—from the choice of the dependence structure.

Thus our model accounts for cross-sectional dependence in three ways: (1) cross-sectional dependence of log volatilities; (2) mean factor structure, allowing conditional correlations to covary with conditional variances, (3) cross-sectional dependence of returns that remains after accounting for dependence attributable to the common factors.

We allow heavy-tailed conditional return distributions. In our applications, we use Student's $t$ marginals, but this is not essential, as we don't rely on data augmentation to obtain conditional Gaussianity, in contrast to many methods for models with Student's $t$ distributions that exploit the fact that they are Gaussian mixtures. In general, we allow the marginal distribution to vary by asset, which in our application translates to asset-specific degrees of freedom parameters. We also depart from the usual assumption of Gaussian factors and allow Student's t factors, with factor-specific degrees of freedom.

For our empirical example we made some specific choices: mean factors are Student's $t$, the autoregressive coefficient and innovation variance matrices of the volatility vector have a parsimonious—but not diagonal—representation, and we use a Gaussian copula to describe dependence across return innovations. We emphasize, however, that our methods do not rely on these special features.

1.5. **The role of the HESSIAN method in model flexibility.** Most methods for Gibbs-updating the conditional posterior distribution of volatility in SV are based on the method introduced by Kim, Shephard, and Chib (1998). Take the univariate case first. Typically, conditioning on parameters and any other latent variables yields an SV model with a measurement equation of the form $r_t = m_t + \sigma_t e^{\alpha_t/2}\epsilon_t$, $\epsilon_t \sim N(0,1)$, where $m_t$ and $\sigma_t$ do not depend on $\alpha_t$ and are therefore (conditionally) constant. An example of $m_t$ is a jump term. An example of $\sigma_t$ is a mixing variable in a scale mixture model, used to thicken the tail of the conditional return distribution.

Kim, Shephard, and Chib (1998) developed the standard method for the special case $m_t = 0$, $\sigma_t = 1$. Taking the logarithm of the square of both sides of the measurement equation yields an equation linear in $\alpha_t$. The non-Gaussian distribution of $\log \epsilon_t^2$ is approximated by a finite Gaussian mixture, tabulated in advance. The state space is augmented to include mixture component indicators; conditioning on these yields a linear Gaussian model. Chib, Nardari, and Shephard (2002) exploit the fact that the transformation $\tilde{r}_t = (r_t - m_t)/\sigma_t$ yields the model $\tilde{r}_t = e^{\alpha_t/2}\epsilon_t$, amenable to the method in Kim, Shephard, and Chib (1998). They describe simulation methods for models with various $m_t$ and $\sigma_t$, including the examples above. Cogley and Sargent (2005) and Primiceri (2005) develop methods for SV models where $r_t$ is multivariate; here, conditioning on all unknown parameters and all latent variables except volatility yields a model of the form

$$r_t = m_t + A_t^{-1}\Sigma_t^{1/2}\epsilon_t, \quad \Sigma_t = \mathrm{diag}(\exp(\alpha_t)), \quad \epsilon_t \sim N(0, I), \tag{1}$$

where the exponential is taken element-wise, and $m_t$ and $A_t$ do not depend on $\alpha_t$. A similar transformation yields $A_t(y_t - m_t) = \Sigma_t^{1/2}\epsilon_t$; the right hand side consists of independent univariate SV models, each amenable to the method of Kim, Shephard, and Chib (1998). This approach is widely used in multivariate stochastic volatility models, but it requires that the conditional measurement equation be transformable into the form given in (1). To see how this is restrictive, note that dependence across elements of $A_t^{-1}\Sigma_t^{1/2}\epsilon_t$ is incompatible with diagonal $A_t^{-1}$; when $A_t^{-1}$ is not diagonal, the amount of variation of conditional (given $A_t$ and $\Sigma_t$) kurtosis across elements of $A_t^{-1}\Sigma_t^{1/2}\epsilon_t$ is limited, since linear combinations of independent random variables have tails as fat as their fattest tailed components. Later, we will see that our empirical exercise gives evidence both for widely varying conditional (given factors) kurtosis across assets and for conditional correlations across innovations—despite our selecting a number of factors greater than that suggested by a principal components analysis.

Our paper adopts an alternative approach, based on the HESSIAN[1] method described in McCausland (2012). This is a procedure to draw all latent states in univariate state

[1] An acronym for Highly Efficient Simulation Smoothing, In A Nutshell, and based on the Hessian matrix of the log target distribution.

space models as a block, preserving their exact conditional posterior distribution. It is fast and numerically efficient and does not require data augmentation. It is generic, not relying on any particular features of the distribution of $y_t$. The observed $y_t$ can be univariate or multivariate; if multivariate, there can be cross-sectional dependence and its length need not be constant: missing or mixed frequency data is easily accommodated. The conditional distribution of $y_t$ can depend on the history $y_{1:t-1}$, and it can be discrete, continuous, or mixed, series by series. Implementing it for a new model involves providing code to evaluate derivatives of the log conditional density $\log \pi(y_t|\alpha_t, y_{1:t-1})$ with respect to $\alpha_t$, for fixed $y_{1:t}$.

While the HESSIAN method is designed for models with univariate states, we can apply it series by series: the conditional distribution of one state sequence, given the others, parameters and data, is the conditional posterior distribution of states in a suitably defined univariate state space model. Very close approximations to these conditional posterior distributions are used as proposal distributions. We can also draw any volatility series, together with some of its associated parameters, as a single block. Because of strong dependence between volatilities and these parameters, the result is higher numerical efficiency.

To apply the HESSIAN method in this way, we require only that the multivariate state sequence be a Gaussian first-order vector autoregressive process and that the conditional distribution of the observed vector $y_t$, given the state sequence $(\alpha_1, \alpha_2, \ldots)$ and the history $(y_1, \ldots, y_{t-1})$, depends only on $\alpha_t$ and $(y_1, \ldots, y_{t-1})$. This requirement is satisfied for a wide variety of state space models, including MSV models, many of which cannot be transformed to auxiliary mixture models in the way that the models of Cogley and Sargent (2005) and Primiceri (2005) can.

1.6. **Outline.** In Section 2, we describe our multivariate stochastic volatility model. In Section 3, we describe our methods for posterior simulation. In Section 4, we verify the correctness of our proposed algorithm using a test of program correctness similar to that proposed by Geweke (2004). In Section 5, we present a daily exchange rate application. In Section 6, we conclude.

## 2. The Model

This section describes our model and compares it to some other specifications in the literature. We also provide prior distributions. Table 1 describes all of the model's variables. The notation is similar to that in Chib, Nardari, and Shephard (2006).

There are $p$ observed return series, $q$ factor series and $m = p + q$ latent log volatility series. The conditional distribution of the latent factor vector $f_t = (f_{t1}, \ldots, f_{tq})$ and the observed return vector $r_t = (r_{t1}, \ldots, r_{tp})$ given the contemporaneous state vector $\alpha_t$ is given by $r_t = Bf_t + V_t^{1/2}\epsilon_t^{(1)}$ and $f_t = D_t^{1/2}\epsilon_t^{(2)}$, or alternatively

$$y_t = \begin{bmatrix} r_t \\ f_t \end{bmatrix} = \begin{bmatrix} V_t^{1/2} & BD_t^{1/2} \\ 0 & D_t^{1/2} \end{bmatrix} \epsilon_t, \tag{2}$$

where $B$ is a $p \times q$ factor loading matrix, $V_t = \mathrm{diag}(\exp(\alpha_{t1}), \ldots, \exp(\alpha_{tp}))$ and $D_t = \mathrm{diag}(\exp(\alpha_{t,p+1}), \ldots, \exp(\alpha_{t,p+q}))$ are volatility matrices and $\epsilon_t = (\epsilon_t^{(1)\top}, \epsilon_t^{(2)\top})^\top$ is an innovation vector. We could have added a constant or lagged returns to the measurement equation, without causing a problem for the HESSIAN method. Given that we are modelling currency returns, we chose not to include these features.

| Symbol | dimensions | description |
|---|---|---|
| $\bar{\alpha}$ | $m \times 1$ | mean of state $\alpha_t$ |
| $A$ | $m \times m$ | autocorrelation coefficient matrix for $\alpha_t$ |
| $\Sigma$ | $m \times m$ | unconditional variance of $\alpha_t$ |
| $\nu$ | $m \times 1$ | vector of degrees of freedom parameters |
| $B$ | $p \times q$ | factor loading matrix |
| $R$ | $m \times m$ | Gaussian copula parameter |
| $\epsilon_t$ | $m \times 1$ | period $t$ return/factor innovation |
| $\alpha_t$ | $m \times 1$ | period $t$ log volatility state |
| $r_t$ | $p \times 1$ | period $t$ return vector |
| $f_t$ | $q \times 1$ | period $t$ factor |
| $y_t$ | $m \times 1$ | $(r_t^\top, f_t^\top)^\top$ |

TABLE 1. Table of symbols

Given parameters $\bar{\alpha}$, $A$ and $\Sigma$, the latent log volatility process is a stationary Gaussian first order vector autoregression, given by ($\Sigma$ is the unconditional variance)

$$\alpha_1 \sim N(\bar{\alpha}, \Sigma), \qquad \alpha_{t+1}|\alpha_t \sim N((I-A)\bar{\alpha} + A\alpha_t, \Sigma - A\Sigma A^\top). \tag{3}$$

We specify the distribution of $\epsilon_t = (\epsilon_{t1}, \ldots, \epsilon_{tm})$ by providing marginals and a copula. For each $\epsilon_{ti}$, let $F_\epsilon(\epsilon_{ti}|\theta_i)$ be its cumulative distribution function (cdf) and $\pi(\epsilon_{ti}|\theta_i)$ be its density. We will use Student's $t$ marginals with asset-specific degrees of freedom, but these could be replaced by other distributions, with suitable modification of the derivations below. We choose a Gaussian copula with variance

$$R = \begin{bmatrix} R_{11} & 0 \\ 0 & I_q \end{bmatrix},$$

where $R_{11}$, and thus $R$, are correlation matrices. Again, one could replace the Gaussian copula with a non-Gaussian one, with suitable modifications of the derivations below. However, this would be computationally costly; we benefit from the fact that the derivatives of a log Gaussian density are non-zero only up to second order. At the same time, the benefits are not clear: while the Gaussian copula affords little flexibility to capture co-movements in the tails of the return distributions, we are already allowing for these through fat-tailed factors.

We denote by $C_R(u_1, \ldots, u_m) = \Phi_R(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m))$ the Gaussian copula with correlation matrix $R$. Here, $\Phi$ and $\phi$ are the cdf and density of the univariate $N(0,1)$; and $\Phi_R$ and $\phi_R$ are the cdf and density of the $m$-variate $N(0,R)$. Then the density of vector $\epsilon_t$ is the product of the Gaussian copula density and the Student-t marginal densities:

$$\pi_\epsilon(\epsilon_t|\theta) = c_R(F_\epsilon(\epsilon_{t1}|\theta_1), \ldots, F_\epsilon(\epsilon_{tm}|\theta_m)) \prod_{i=1}^m \pi(\epsilon_{ti}|\theta_i), \tag{4}$$

where

$$c_R(u_1, \ldots, u_m) = \frac{\partial^{(m)} C_R(u_1, \ldots, u_m)}{\partial u_1 \cdots \partial u_m} = \frac{\phi_R(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_m))}{\prod_{i=1}^m \phi(\Phi^{-1}(u_i))}.$$

Letting $x_i \equiv \Phi^{-1}(u_i)$, $i = 1, \ldots, m$ and $x \equiv (x_1, \ldots, x_m)$, we can write

$$\log c_R(u_1, \ldots, u_m) = -\frac{1}{2}(\log|R| + \log(2\pi) + x^\top(R^{-1} - I))x. \tag{5}$$

We use the notation $\pi_\epsilon$ here instead of the generic $\pi$ to clarify that it is the density function of $\epsilon_t$. We can now write the conditional density of $y_t$ given $\alpha_t$, $B$, $\nu$ and $R$ as

$$\pi(y_t|\alpha_t, B, \nu, R) = \pi_\epsilon\left(\left.\begin{bmatrix} V_t^{-1/2}(r_t - Bf_t) \\ D_t^{-1/2} f_t \end{bmatrix}\right| \nu, R\right) \prod_{i=1}^{m} \exp(-\alpha_{ti}/2). \tag{6}$$

The following decomposition implies conditional independence relationships in our model:

$$\pi(\bar{\alpha}, A, \Sigma, \nu, B, R, \alpha, f, r) = \pi(\bar{\alpha}, A, \Sigma, \nu)\pi(B)\pi(R) \cdot \pi(\alpha_1|\bar{\alpha}, A, \Sigma) \prod_{t=1}^{n-1} \pi(\alpha_{t+1}|\alpha_t, \bar{\alpha}, A, \Sigma)$$
$$\cdot \prod_{t=1}^{n} \left[\pi(f_t|\alpha_t, \nu)\pi(r_t|f_t, \alpha_t, \nu, B, R)\right].$$

2.1. **Related MSV models.** As mentioned before, different MSV model specifications reflect, to a large extent, different choices on how to balance flexibility and parsimony. In our model, we can restrict the parameters of the marginal distribution of volatility in (3), the parameters of the conditional distribution of returns and factors given volatility, in (2), or both.

Consider first the marginal distribution of volatilities. For the most flexible dynamics, we can specify $A$ and $\Sigma$ in (3) as full matrices. Alternatively, we can impose prior independence among volatilities by specifying diagonal $A$ and $\Sigma$. Intermediate possibilities are possible, including the relatively parsimonious specification in Section 2.2, where $A$ and $\Sigma$ are not diagonal, but are functions of $2m$ free parameters each.

Now consider cross-sectional dependence arising from the conditional distribution of returns given parameters and volatilities, marginal of latent factors. We can write the conditional variance of returns as:

$$\mathrm{Var}[r_t|\alpha_t] = V_t^{1/2} R_{11} V_t^{1/2} + BD_tB^\top. \tag{7}$$

With no factors ($q = 0$) the second term disappears. The conditional variance varies in time, but the conditional correlation $R_{11}$ is constant. Models with constant correlations have been studied by Harvey, Ruiz, and Shephard (1994), Danielsson (1998), Smith and Pitts (2006), Chan, Kohn, and Kirby (2006) and So, Li, and Lam (1997). Other authors, including Yu and Meyer (2006), Philipov and Glickman (2006), Gourieroux (2006), Gourieroux, Jasiak, and Sufana (2004), Carvalho and West (2006) and Asai and McAleer (2009), consider models in which the return innovation correlation is time-varying, which is more realistic. However, as the number of assets increases, the estimation of a separate time varying correlation matrix becomes very challenging. Also, when correlation and volatility are modelled separately, it is more difficult to capture co-movements of correlations and volatility.

Introducing mean factors is another way to introduce time-varying correlations. Here, co-movements of asset returns are driven by a small number of latent common factors, typically modelled as univariate SV processes. Usually, factor MSV models specify $R_{11} = I$, in which case $\mathrm{Var}[r_t|\alpha_t] = V_t + BD_tB^\top$. Mean factor models are parsimonious, they give time varying conditional correlations and they have a natural link with factor models in finance, which hold that the expected return of an asset is a linear function of various factors. In addition, mean factor structure allows the conditional correlations and conditional variances to covary in a way that is broadly consistent with well known stylized facts. Longin and Solnik (2001) and Ang and Chen (2002) document a positive correlation between conditional variances and conditional correlations. Given all these characteristics, factor MSV models have become popular in the literature. The basic model assumed normal returns and constant factor loadings. See, for example, Jacquier, Polson, and Rossi (1995), Pitt and Shephard (1999) and Aguilar and West (2000). Other studies proposed extensions such as jumps in the return equation and heavy-tailed returns (Chib, Nardari, and Shephard (2006)), time varying factor loading matrices and regime-switching factors (Lopes and Carvalho (2007)) or first-order autoregressive factors (Han (2006)). See Chib, Omori, and Asai (2009) for a review and comparison of different MSV models.

If we compare these models to ours, we notice that ours is fairly general and incorporates some other specifications as special cases. In its most general version, without parameter restrictions, the model allows for cross-sectional volatility dependence, time-varying conditional correlations through the specification of a mean factor structure; and cross-sectional conditional return dependence through copulas. The conditional variance matrix of returns in equation (7) is time-varying. The conditional correlation matrix is also time varying, and covaries with the conditional variances.

2.2. **Prior Distributions.** We first describe a prior for a low dimensional specification of $\bar{\alpha}$, $A$, $\Sigma$, $\nu$, and $B$. We parameterize $A$ and $\Sigma$ in the following parsimonious way:

$$A = \mathrm{diag}(\lambda) + \begin{bmatrix} (1/p)\delta\iota_p^\top & 0 \\ 0 & 0 \end{bmatrix}. \quad \Sigma = (\mathrm{diag}(\sigma))^2 + \begin{bmatrix} \beta\beta^\top & 0 \\ 0 & 0 \end{bmatrix}, \tag{8}$$

where $\sigma$ and $\lambda$ are $m \times 1$, $\beta$ and $\delta$ are $p \times 1$ and $\iota_p$ is the $p \times 1$ vector of ones. The log volatility series for the $q$ factors are conditionally independent given $\Sigma$ and $A$. The log volatility vector $\alpha$ has a factor structure—$\Sigma$ is the sum of a positive definite diagonal matrix $(\mathrm{diag}(\sigma))^2$ and a rank-one positive semi-definite matrix $\beta\beta^\top$. The matrix $A$ is determined by the $(p+q)$-vector $\lambda$ and the $p$-vector $\delta$. Writing the conditional mean equation by equation shows that each conditional mean depends linearly on both the same-equation lagged value and the lagged arithmetic average:

$$E[\alpha_{ti}|\alpha_{t-1}, A, \Sigma] = (1 - \lambda_i - \delta_i)\bar{\alpha}_i + \lambda_i \alpha_{t-1,i} + \delta_i \frac{1}{p} \sum_{j=1}^{p} \alpha_{t-1,j}.$$

We organize the parameters associated with each series $i$ (a return for $i \le p$ or a factor for $i > p$) as

$$\theta_i = \begin{cases} \left(\bar{\alpha}_i, \tanh^{-1}(\lambda_i), \tanh^{-1}(\lambda_i + \delta_i), \log\sigma_i, \beta_i/\sigma_i, \log\nu_i\right)^\top, & 1 \le i \le p, \\ \left(\tanh^{-1}(\lambda_i), \log\sigma_i, \log\nu_i\right)^\top, & p+1 \le i \le m, \end{cases}$$

$$B_i = (B_{i1}, \ldots, B_{iq}), \quad i = 1, \ldots, p.$$

and let $\theta = (\theta_1^\top, \ldots, \theta_m^\top)^\top$. The elements of $\theta_i$ and $B_i$ are mutually independent and Gaussian. The $\log(\cdot)$ and $\tanh^{-1}(\cdot)$ functions map parameters defined on $[0, \infty)$ and $(-1, 1)$ to the real line. The posterior distribution of $\theta$ is closer to Gaussian than that of the untransformed parameters; this improves numerical efficiency.

Prior means and standard deviations for the "Getting it right" exercise and for the currency application are shown in Table 2. The "Getting it right" simulation exercise, described below, is meant to formally test the correctness of our posterior simulation methods. For this, we use a relatively tight prior that favours lower autocorrelations. This reduces dependence across variables in this simulation, increasing numerical efficiency and thereby increasing the power of the correctness tests for a given amount of simulation. The prior used for inference is more diffuse, as it intended to cover all plausible regions of the parameter space.

| | Getting it right | | Currency data | |
|---|---|---|---|---|
| Parameter | mean | standard deviation | mean | standard deviation |
| $\bar{\alpha}_i$ | $\ln(0.003^2)$ | $\ln(1.25^2)$ | $\ln(0.003^2)$ | $\ln(3.0^2)$ |
| $\tanh^{-1}(\lambda_i)$ | $\tanh^{-1}(0.96)$ | 0.2 | $\tanh^{-1}(0.98)$ | 0.2 |
| $\tanh^{-1}(\lambda_i+\delta_i)$ | $\tanh^{-1}(0.96)$ | 0.2 | $\tanh^{-1}(0.98)$ | 0.2 |
| $\log\sigma_i$ | $\ln(\ln(1.25))$ | 0.1 | $\ln(\ln(2.0))$ | 0.2 |
| $\beta_i/\sigma_i$ | 0.0 | 0.2 | 0.0 | 0.5 |
| $\log\nu_i$ | $\ln(20.0)$ | $\ln(1.25)$ | $\ln(20.0)$ | $\ln(2.0)$ |
| $B_{ij}$, $j=1,\ldots,q$ | 0.0 | 0.002 | 0.0 | 0.003 |

TABLE 2. Prior means and standard deviations of series-specific parameters.

The prior distribution for the $A$ matrix is in fact a truncated distribution, with truncation to the region where the eigenvalues of $AA^\top$ are in the unit circle. Since the probability, in the untruncated prior, that an eigenvalue lies outside the unit circle is low, the truncated prior is similar to the untruncated one.

The likelihood function is invariant to many parameter transformations. Imposing independence of the factor series, as we do with our prior, rules out some but not all of these transformations. There remain sign and labelling invariance in factors and their associated loadings: we can either multiply both the $j$'th column of $B$ and the $j$'th factor series by -1, or exchange the $j$'th and $k$'th columns of $B$ and the $j$'th and $k$'th factor series without changing the likelihood. The likelihood is also invariant to multiplying $\beta$ by -1. In the empirical exercise, we impose sign and labelling restrictions to break invariance.

We now describe a prior distribution for the correlation matrix R. A common approach is to put a prior on its Cholesky factor $L$. The condition that all diagonal elements of $R$ equal one is equivalent to the condition that all rows of $L$ have unit length.

We modify this approach so that the prior is invariant to the ordering of the series. For intuition, note that for a given $R$, the Cholesky factor $L$ is not the only matrix $V$ with rows of unit length satisfying $VV^\top = R$: for any orthogonal matrix $C$, $V \equiv LC$ is another. We specify a prior on $V$—inducing a prior on $R = VV^\top$—where the rows $v_i$ of $V$ are iid (although exchangeability would suffice for order invariance). The cost is that $V$ has a larger number of non-zero elements than $L$. The elements of $V$ are not identified, but since $VV^\top$ is, this is not a serious concern. The rows of $V$ are points on the unit hypersphere of dimension $p$; seeing this may make it easier to understand the prior and how we draw $R$ from its conditional posterior. For $i = 1, ..., p$, let $\zeta_i \equiv \cos^{-1}(V_{i1})$, the angle between row $v_i$ and $(1, 0, ..., 0)$. We specify a prior for $\zeta_i$ and let $v_i$ be uniformly distributed on the surface of the $(p-1)$-dimensional hypersphere (of radius $\sin\zeta_i$) at an angle $\zeta_i$ away from $(1, 0, ..., 0)$. Thus $\pi(v_i) \propto \pi(\zeta_i)\sin^{p-2}\zeta_i$. In our applications, we use $\zeta_i/\pi \sim \mathrm{Be}(40, 40)$; using simulations we estimate the prior mean and standard deviation of off-diagonal elements of $R$ as 0.000 and 0.326. This is an informative prior shrinking correlations towards zero, which makes up for the large number of parameters. In this way, we favour factor structure over correlated innovations while still allowing for correlations that are not well captured by factors.

## 3. Posterior inference using MCMC

We use a five-block Gibbs sampler to simulate the posterior distribution. Each block is described in one of the following sections.

### 3.1. Draw of $\theta_i, \alpha_i$, $i = 1, \ldots, m$.

We draw $(\theta_i, \alpha_i)$ as a single block. Our proposal of $(\theta_i, \alpha_i)$ consists of a random walk proposal of $\theta_i^*$ followed by a proposal of $\alpha_i^*$ given $\theta_i^*$. We accept $(\theta_i^*, \alpha_i^*)$ with probability

$$\min\left(1, \frac{\pi(\theta_i^*)\pi(\alpha_i^*|\theta_i^*, \theta_{-i}, \alpha_{-i})\pi(y|\alpha_i^*, \alpha_{-i}, \theta_i^*, \theta_{-i}, B, R)}{\pi(\theta_i)\pi(\alpha_i|\theta, \alpha_{-i})\pi(y|\alpha, \theta, B, R)} \cdot \frac{g(\alpha_i^*|\theta_i^*, \theta_{-i}, \alpha_{-i}, B, R)}{g(\alpha_i|\theta, \alpha_{-i}, B, R)}\right),$$

where $g(\alpha_i^*|\theta_i^*, \theta_{-i}, B, R)$ is the conditional proposal density for $\alpha_i^*$ given $\theta_i^*$.

The random walk $(\theta_i^* - \theta_i)$ is Gaussian with mean zero and variance $\Xi$. We obtain $\Xi$ using an adaptive random walk Metropolis algorithm, described in Vihola (2011), during a burn-in period in which $\Xi$ is adjusted after each draw to track a target acceptance probability. We use the final value of $\Xi$ at the end of the burn-in period for all future draws; ending the adaptation ensures that our posterior simulator is truly Markov.

We draw $\alpha_i^*|\theta_i^*, \theta_{-i}, \alpha_{-i}, B, R$ using the HESSIAN method in McCausland (2012). The HESSIAN method uses an approximation $g(\alpha|y)$ of $\pi(\alpha|y)$ for univariate models in which $\alpha \sim N(\bar{\Omega}^{-1}\bar{c}, \bar{\Omega})$, with $\bar{\Omega}$ tridiagonal and $\pi(y|\alpha) = \prod_{t=1}^n \pi(y_t|\alpha_t)$. It requires one to specify the precision $\bar{\Omega}$ and covector $\bar{c}$ and provide routines to compute the first five derivatives of $\log \pi(y_t|\alpha_t)$ with respect to $\alpha_t$. Here states are multivariate, but conditioning on the other volatility series (denoted $\alpha_{-i}$) yields a univariate model for $\alpha_i$ amenable to the HESSIAN method. The conditional density we need to approximate is

$$\pi(\alpha_i|\alpha_{-i}, y) \propto \pi(\alpha_i|\alpha_{-i}) \prod_{t=1}^n \pi(y_t|\alpha_t).$$

In Appendix A, we provide $\bar{\Omega}^{(i)}$ and $\bar{c}^{(i)}$, in terms of $\bar{\Omega}$ and $\bar{c}$, such that $\alpha_i|\alpha_{-i} \sim N((\bar{\Omega}^{(i)})^{-1}\bar{c}^{(i)}, \bar{\Omega}^{(i)})$. $\bar{\Omega}^{(i)}$ is tridiagonal, as required by the HESSIAN method.

We need to compute five derivatives of $\log \pi(y_t|\alpha_{ti}, \alpha_{t,-i})$ with respect to $\alpha_{ti}$ at any point. We do *not* need analytic expressions; instead we use automatic routines to combine derivatives of primitive functions according to Faa di Bruno's rule, a generalization of the chain rule to higher derivatives. Appendix B describes how we compute the required derivatives.

### 3.2. Draw of $(B, f)$.

Here we update $B$ and $f$ in a way that does not change the values of the matrix-vector products $Bf_t$, $t = 1, \ldots, n$. While the block is redundant—we are also updating $B$ and $f$ elsewhere—$B$ and $f$ are less well identified than the product $Bf$ and this block improves posterior mixing of both $B$ and $f$. At the same time, it is computationally cheap: since the $Bf_t$ do not change, we do not need to evaluate $\pi(r|\theta, B, \alpha, f)$.

We first draw a random diagonal $q \times q$ matrix $\Lambda$, where $n\Lambda_{ii} \sim \text{iid}\, \chi^2(n)$. With probability $1/2$, we form proposals $B^* = B\Lambda$, $f_t^* = \Lambda^{-1} f_t$, $t = 1, \ldots, n$ and with complementary probability, we form $B^* = B\Lambda^{-1}$, $f_t^* = \Lambda f_t$, $t = 1, \ldots, n$. The acceptance probabilities are, respectively,

$$\min\left(1, |\Lambda|^{-(n-p)} \frac{\pi(B^*) \prod_{t=1}^n \pi(f_t^*|\alpha, \nu)}{\pi(B) \prod_{t=1}^n \pi(f_t|\alpha, \nu)}\right), \quad \min\left(1, |\Lambda|^{(n-p)} \frac{\pi(B^*) \prod_{t=1}^n \pi(f_t^*|\alpha, \nu)}{\pi(B) \prod_{t=1}^n \pi(f_t|\alpha, \nu)}\right).$$

The factors $|\Lambda|^{-(n-p)}$ and $|\Lambda|^{(n-p)}$ are determinants of the Jacobian matrices for the linear transformations of the $f_t$ and the $p$ rows of $B$. The computational cost of this draw is low, and in the applications, we repeat the update of $(B, f)$ ten times.

3.3. **Draw of $B$.** We draw each row $B_i$ using a Gaussian proposal approximating its conditional posterior distribution. The approximate distribution is what the conditional posterior distribution would be if the Student's $t$ degrees of freedom $\nu_i$ were all infinite and the correlation matrix $R$ were equal to $I$. Thus the proposal distribution is $B_i^* \sim N\left(\bar{\bar{H}}_B^{-1}\bar{\bar{c}}_B, \bar{\bar{H}}_B^{-1}\right)$, where

$$\bar{\bar{H}}_B = \bar{H}_B + \sum_{t=1}^{n} e^{-\alpha_{ti}} f_t f_t^\top \quad \text{and} \quad \bar{\bar{c}}_B = \sum_{t=1}^{n} e^{-\alpha_{ti}} r_t f_t.$$

Here, $\bar{H}_B$ is the diagonal prior precision matrix of any row of $B_i$. Each of its diagonal elements is the prior precision of an element of $B_{ij}$, the reciprocal of the square of the prior standard deviation. Denote the proposal density by $g(B_i^*)$. We accept the proposal $B_i^*$ with probability

$$\min\left(1, \frac{\pi(B_i^*)}{\pi(B_i)} \frac{g(B_i)}{g(B_i^*)} \frac{\pi(y|\theta, \alpha, B_i^*, B_{-i}, R)}{\pi(y|\theta, \alpha, B, R)}\right).$$

3.4. **Draw of $f$.** We draw each $f_t$ using a Gaussian proposal that approximates its conditional posterior distribution. Again, the approximate distribution is what the conditional posterior distribution would be if the $\nu_i$ were all infinite and the correlation matrix $R$ were equal to $I$. Thus the proposal distribution is $f_t^* \sim N\left(\bar{\bar{H}}_f^{-1}\bar{\bar{c}}_f, \bar{\bar{H}}_f^{-1}\right)$, where $\bar{\bar{H}} \equiv B^\top V_t^{-1} B + D_t^{-1}$ and $\bar{\bar{c}} \equiv B^\top V_t^{-1} r_t$. Denote the proposal density by $g(f_t^*)$. We accept the proposal $f_t^*$ with probability

$$\min\left(1, \frac{q(f_t)}{q(f_t^*)} \frac{\pi(r_t, f_t^*|\alpha_t, \theta, B, R)}{\pi(r_t, f_t|\alpha_t, \theta, B, R)}\right).$$

3.5. **Draw of $R$.** We draw rows $v_i$ of $V$ one at a time, using a random walk (on the $p$-dimensional unit hypersphere) proposal. The direction is uniformly distributed and the arc length $\vartheta$, in radians, has a Beta distribution scaled to the interval $[0, \pi]$. In practice, we draw $\vartheta/\pi \sim \mathrm{Be}(1, 199)$ and a $d \sim N(0, I_p)$ that determines the direction, then construct

$$v_i^* = \cos\vartheta \cdot v_i + \sin\vartheta \cdot \frac{d_\perp}{||d_\perp||}, \quad \text{where} \quad d_\perp = d - \frac{v_i d}{||v_i||^2} v_i,$$

which we accept with probability

$$\min\left(1, \frac{\pi(y|\alpha, \theta, B, R^*)\pi(v_i^*)}{\pi(y|\alpha, \theta, B, R)\pi(v_i)}\right).$$

Once the sufficient statistic for drawing $R$ is constructed, the marginal cost of drawing $R$ is low; in our application, we update each $v_i$ ten times. Further repetition does little to improve numerical efficiency.

# 4. Getting it Right

We tested the correctness of our posterior simulators using a simulation strategy similar to that proposed by Geweke (2004). We simulated the joint distribution of parameters, states, factors and data, using a Gibbs sampler consisting of all the blocks in Section 3 and an additional block, described in Appendix C, to update the distribution of returns given parameters, states and factors. A testable implication of the correctness of our posterior simulators is that this sampler has a stationary distribution whose marginal on the parameter subspace agrees with the specified prior distribution of parameters.

We obtain a sample $\{(\theta_1^{(j)}, \ldots, \theta_m^{(j)})\}_{j=1}^J$ of size $J = 10^7$ and construct, for $i = 1, \ldots, m$ and $j = 1, \ldots, J$ the vectors $z^{(i,j)} \equiv L_i^{-1}(\theta_i^{(j)} - \mu_i)$, where $\mu_i$ is the prior mean and $L_i$ is the lower Cholesky factor of the prior variance of $\theta_i$. If the $\theta_i^{(j)}$ are truly multivariate Gaussian with variance $L_i L_i^\top$, the elements of $z^{(i,j)}$ are iid $N(0,1)$. The vectors $z^{(i,j)}$ have length $K_i = 6$ for $i = 1, \ldots, p$ and length $K_i = 3$ for $i = p+1, \ldots, m$. Since the $z^{(i,j)}$, $i = 1, \ldots, m$, are independent, we have $\sum_{i=1}^m z_i^\top z_i \sim \chi^2(6p + 3q)$.

We compute the following sample frequencies for all quantiles $Q = 0.1, 0.3, 0.5, 0.7, 0.9$, return/factor indices $i = 1, \ldots, m$, and parameter indices $k = 1, \ldots, K_i$:

$$\hat{I}_{ik}^{(Q)} = \frac{1}{J} \sum_{j=1}^J 1\left(z_k^{(i,j)} \leq \Phi^{-1}(Q)\right),$$

and report them in Table 3. Each row except the last is associated with a particular $i$ and $k$; each column, with a particular quantile $Q$. We also construct for the same quantiles, the sample frequencies

$$\hat{I}_0^{(Q)} = \frac{1}{J} \sum_{j=1}^J 1\left(\sum_{i=1}^m (z^{(i,j)})^\top z^{(i,j)} \leq F^{-1}(Q)\right),$$

where $F$ is the cdf of the $\chi^2$ distribution with $6p + 3q$ degrees of freedom. We report these in the last line of Table 3.

We should observe sample frequencies close to $Q$. Table 3 shows, with the sample frequencies $\hat{I}_{ik}^{(Q)}$, their estimated numerical errors $s_{ik}^{(Q)}$, obtained using the method of batch means. In all cases, the sample frequencies are very similar to their respective population values.

# 5. An Exchange Rate Application

5.1. **Data.** We analyze daily returns of 10 currencies relative to the US dollar: the Australian Dollar (AUD), Brazilian Real (BRL), Euro (EUR), Japanese Yen (JPY), Mexican Peso (MXN), New Zealand Dollar (NZD), Singapore Dollar (SGD), Swiss Franc (CHF), British Pound (GBP), and Canadian Dollar (CAD). We obtained noon spot rates from the Bank of Canada, from July 8, 2005 to July 8, 2015 inclusive and computed log returns between consecutive weekdays that are not bank holidays, giving 2505 observations for each currency.

Table 4 presents some descriptive statistics. The sample standard deviation varies a lot, with the Brazilian Real, and the Australian and New Zealand Dollars being the most volatile and the Singapore dollar the least. Sample skewness varies considerably in magnitude, with equal numbers of currencies of each sign. All series present excess kurtosis, and this too varies considerably, from 6.1 for the Euro to 46.7 for the Swiss Franc. The first-order sample autocorrelations of squared returns suggest varying levels of volatility persistence. The log

| $i$ | $k$ | $\hat{I}_{i,k}^{(0.1)}$ | $s_{i,k}^{(0.1)}$ | $\hat{I}_{i,k}^{(0.3)}$ | $s_{i,k}^{(0.3)}$ | $\hat{I}_{i,k}^{(0.5)}$ | $s_{i,k}^{(0.5)}$ | $\hat{I}_{i,k}^{(0.7)}$ | $s_{i,k}^{(0.7)}$ | $\hat{I}_{i,k}^{(0.9)}$ | $s_{i,k}^{(0.9)}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0.0999 | 0.00038 | 0.2995 | 0.00067 | 0.4993 | 0.00081 | 0.6996 | 0.00069 | 0.9001 | 0.00041 |
| 1 | 2 | 0.0998 | 0.00035 | 0.3000 | 0.00054 | 0.4999 | 0.00063 | 0.7002 | 0.00063 | 0.9003 | 0.00034 |
| 1 | 3 | 0.1003 | 0.00034 | 0.3004 | 0.00057 | 0.5004 | 0.00059 | 0.7002 | 0.00057 | 0.9000 | 0.00036 |
| 1 | 4 | 0.1001 | 0.00034 | 0.2996 | 0.00057 | 0.4995 | 0.00064 | 0.6998 | 0.00058 | 0.8999 | 0.00038 |
| 1 | 5 | 0.1002 | 0.00037 | 0.3007 | 0.00054 | 0.5013 | 0.00057 | 0.7008 | 0.00052 | 0.9002 | 0.00032 |
| 1 | 6 | 0.1006 | 0.00037 | 0.3007 | 0.00058 | 0.5008 | 0.00063 | 0.7009 | 0.00058 | 0.9005 | 0.00034 |
| 2 | 1 | 0.1003 | 0.00047 | 0.2999 | 0.00084 | 0.4999 | 0.00098 | 0.6998 | 0.00085 | 0.8999 | 0.00048 |
| 2 | 2 | 0.1004 | 0.00035 | 0.3000 | 0.00057 | 0.4996 | 0.00065 | 0.6999 | 0.00052 | 0.8999 | 0.00034 |
| 2 | 3 | 0.0997 | 0.00034 | 0.2996 | 0.00055 | 0.4988 | 0.00060 | 0.6997 | 0.00053 | 0.8996 | 0.00034 |
| 2 | 4 | 0.0997 | 0.00037 | 0.2990 | 0.00060 | 0.4993 | 0.00064 | 0.6994 | 0.00052 | 0.8996 | 0.00032 |
| 2 | 5 | 0.0999 | 0.00035 | 0.3006 | 0.00058 | 0.5005 | 0.00063 | 0.7000 | 0.00051 | 0.9000 | 0.00032 |
| 2 | 6 | 0.1001 | 0.00036 | 0.3002 | 0.00062 | 0.5007 | 0.00065 | 0.7006 | 0.00057 | 0.9002 | 0.00033 |
| 3 | 1 | 0.1000 | 0.00023 | 0.2993 | 0.00053 | 0.4998 | 0.00059 | 0.6997 | 0.00046 | 0.8998 | 0.00030 |
| 3 | 2 | 0.1000 | 0.00029 | 0.3006 | 0.00046 | 0.5002 | 0.00052 | 0.7004 | 0.00048 | 0.9002 | 0.00031 |
| 3 | 3 | 0.1002 | 0.00028 | 0.3004 | 0.00049 | 0.5003 | 0.00055 | 0.7003 | 0.00052 | 0.9004 | 0.00028 |
| 4 | 1 | 0.1002 | 0.00027 | 0.3002 | 0.00052 | 0.4998 | 0.00052 | 0.7002 | 0.00040 | 0.8996 | 0.00026 |
| 4 | 2 | 0.1008 | 0.00029 | 0.3006 | 0.00050 | 0.5004 | 0.00055 | 0.7005 | 0.00052 | 0.9002 | 0.00029 |
| 4 | 3 | 0.0999 | 0.00027 | 0.3003 | 0.00048 | 0.5002 | 0.00054 | 0.7001 | 0.00048 | 0.8996 | 0.00026 |
| | | 0.0995 | 0.00033 | 0.2994 | 0.00053 | 0.4996 | 0.00061 | 0.7000 | 0.00057 | 0.9001 | 0.00036 |

TABLE 3. “Getting it right” sample quantiles and their numerical standard errors

| | Mean | SD | Skewness | Kurtosis | $r_t^2$ autocorr. | Log variance |
|---|---|---|---|---|---|---|
| AUD | 0.05 | 14.42 | -0.638 | 16.0 | 0.27 | -9.40 |
| BRL | -3.10 | 16.21 | 0.015 | 15.5 | 0.40 | -9.16 |
| EUR | -0.75 | 10.07 | 0.208 | 6.1 | 0.05 | -10.12 |
| JPY | -0.72 | 10.48 | 0.166 | 7.0 | 0.10 | -10.03 |
| MXN | -3.86 | 11.33 | -0.819 | 17.7 | 0.54 | -9.88 |
| NZD | 0.03 | 14.60 | -0.320 | 7.8 | 0.12 | -9.37 |
| SGD | 2.31 | 5.65 | -0.034 | 8.1 | 0.11 | -11.27 |
| CHF | 3.20 | 11.90 | 1.646 | 46.7 | 0.13 | -9.78 |
| GBP | -1.23 | 9.75 | -0.167 | 8.3 | 0.12 | -10.18 |
| CAD | -0.41 | 10.10 | 0.066 | 9.1 | 0.13 | -10.11 |

TABLE 4. Descriptive statistics for log returns: annualized mean (%), annualized standard deviation (%), skewness, kurtosis, squared return autocorrelation and log variance.

variance figures, though redundant, allow for easy comparison with the $\bar{\alpha}_i$ parameters, which give mean idiosyncratic log conditional variances.

In Table 5 we show the sample correlation matrix. Correlations vary from -0.16 to 0.84. The strongest negative correlation is for the pair (MXN, JPY) and the strongest positive correlation is for the pair (AUD, NZD).

5.2. **Order selection.** Ideally, we would compute the posterior distribution of $q$, the number of factors, and report results for the value (or values) of $q$ with non-negligible posterior probability. As this would involve computing Bayes factors, and since we can only integrate

| | AUD | BRL | EUR | JPY | MXN | NZD | SGD | CHF | GBP | CAD |
|---|---|---|---|---|---|---|---|---|---|---|
| AUD | 1.00 | 0.59 | 0.60 | -0.07 | 0.60 | 0.84 | 0.65 | 0.37 | 0.59 | 0.68 |
| BRL | 0.59 | 1.00 | 0.39 | -0.15 | 0.66 | 0.50 | 0.46 | 0.19 | 0.38 | 0.50 |
| EUR | 0.60 | 0.39 | 1.00 | 0.21 | 0.39 | 0.59 | 0.65 | 0.70 | 0.66 | 0.53 |
| JPY | -0.07 | -0.15 | 0.21 | 1.00 | -0.16 | -0.04 | 0.18 | 0.36 | 0.07 | -0.08 |
| MXN | 0.60 | 0.66 | 0.39 | -0.16 | 1.00 | 0.52 | 0.50 | 0.19 | 0.39 | 0.53 |
| NZD | 0.84 | 0.50 | 0.59 | -0.04 | 0.52 | 1.00 | 0.61 | 0.38 | 0.58 | 0.62 |
| SGD | 0.65 | 0.46 | 0.65 | 0.18 | 0.50 | 0.61 | 1.00 | 0.47 | 0.53 | 0.55 |
| CHF | 0.37 | 0.19 | 0.70 | 0.36 | 0.19 | 0.38 | 0.47 | 1.00 | 0.45 | 0.32 |
| GBP | 0.59 | 0.38 | 0.66 | 0.07 | 0.39 | 0.58 | 0.53 | 0.45 | 1.00 | 0.52 |
| CAD | 0.68 | 0.50 | 0.53 | -0.08 | 0.53 | 0.62 | 0.55 | 0.32 | 0.52 | 1.00 |

TABLE 5. Sample correlation matrix for Bank of Canada currency panel

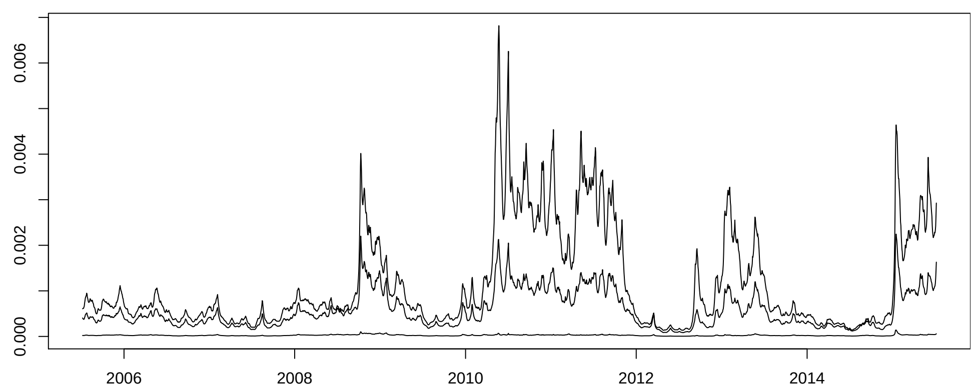


FIGURE 1. EUR idiosyncratic volatility: posterior mean and standard deviation, nse for the mean

out one volatility sequence, this would be difficult if not infeasible. We first performed a static principal components analysis to suggest the number of factors. Horn's Parallel analysis, which recommends retaining factors whose eigenvalues are greater than the 95th percentile of the eigenvalues of iid data, suggests two factors. The first three factors account for fractions 0.549, 0.155, 0.088 (cumulatively, 0.549, 0.704 and 0.792) of total variance. We obtained full results for two and three factors and found that the third factor was important in ways that a principal components analysis misses. Specifically, the volatility of volatility of the third factor is particularly high, and the factor loadings are quite high for nearly all of the currencies: in eight of ten cases, the posterior mean of the third factor loading is larger in absolute value than that of the second; in nine, the posterior mean is more than four posterior standard deviations from zero. These results, described in more detail below, show that in a few highly volatile periods, the third factor accounts for much of the common variation across currencies.

5.3. **Estimation results.** We report results for ten univariate models and the full multivariate model. We use comparable priors in the two models and compare corresponding posterior distributions. Throughout Section 5.3, numerical standard error (nse) and relative numerical efficiency (rne) are computed using the R library coda, which uses a spectral density method.

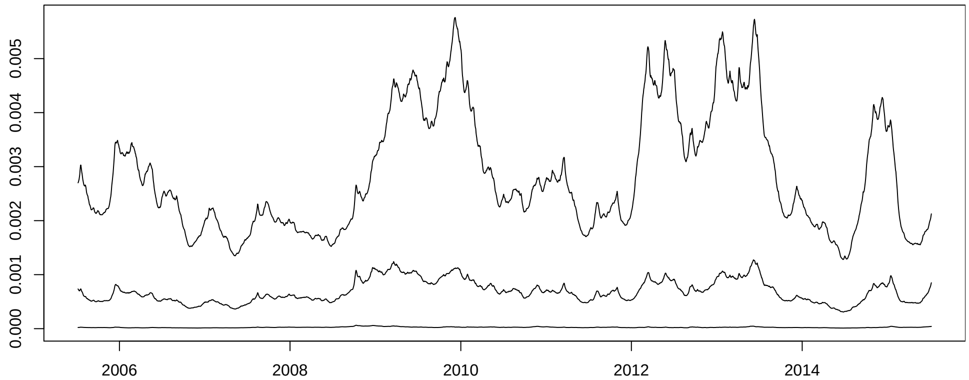


FIGURE 2. JPY idiosyncratic volatility: posterior mean and standard deviation, nse for the mean

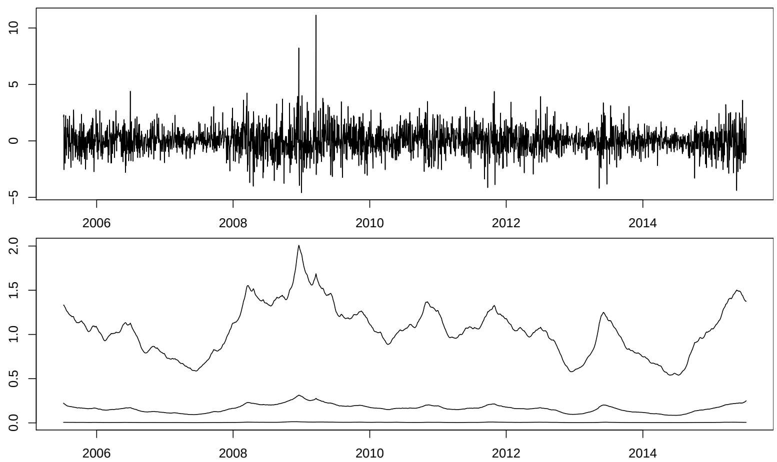


FIGURE 3. First factor series: top panel gives the posterior mean of $f_{t1}$; bottom panel gives the posterior mean and standard deviation of $\alpha_{t,11}$ (the log volatility of $f_{t1}$), and the numerical standard error for the mean of $\alpha_{t,11}$.

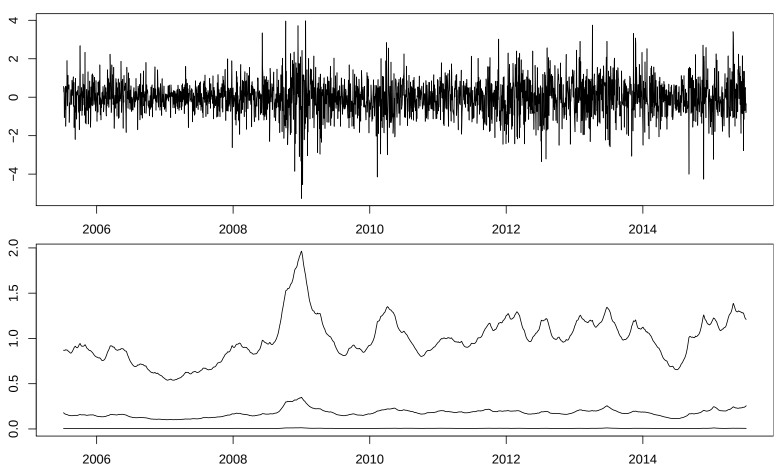


FIGURE 4. Second factor series: top panel gives the posterior mean of $f_{t2}$; bottom panel gives the posterior mean and standard deviation of $\alpha_{t,12}$ (the log volatility of $f_{t2}$), and the numerical standard error for the mean of $\alpha_{t,12}$.

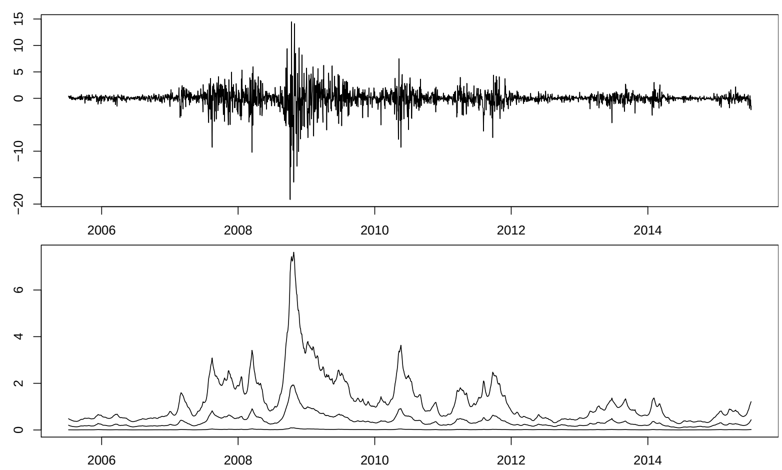


FIGURE 5. Third factor series: top panel gives the posterior mean of $f_{t3}$; bottom panel gives the posterior mean and standard deviation of $\alpha_{t,13}$ (the log volatility of $f_{t3}$), and the numerical standard error for the mean of $\alpha_{t,13}$.

We generated a posterior sample of size 60000 and discarded the first 10000 draws. Then we imposed three identification restrictions, chosen so that the posterior density is low near the boundary of the restricted region. To break sign invariance of loadings and factors, we set the Euro factor loadings to be positive, multiplying columns of $B$ and the corresponding factor series by -1 as needed. To break labelling invariance of loadings and factors, we ordered the factors so that Euro loadings were in descending order, exchanging columns of $B$ and factor series as needed. To break sign invariance of the vector $\beta$, we set the element associated with the Euro to be positive, multiplying $\beta$ by -1 as needed.

Table 6 summarizes the posterior distribution of the mean vector $\bar{\alpha}$ and the autocorrelation matrix $A$ of the log volatility process $\alpha_t$. Recall that $A$ is parameterized, in equation (8), in terms of the vectors $\lambda$ and $\delta$. For each currency $i$, we give the posterior mean and standard deviation of $\bar{\alpha}_i$, the mean log idiosyncratic volatility; $\lambda_i$, the coefficient of the same-series lagged value; and $\delta_i$, the coefficient of the lagged arithmetic average. The first four numeric columns report results for independent SVt models, where $\delta = 0$. The next four numeric columns give results for the full MSV model, where $A$ has non-zero off-diagonal elements. In all cases, the parameter $\bar{\alpha}_i$ giving the mean log idiosyncratic volatility is considerably smaller for the factor model—in many cases, it is much smaller—indicating that the three factors capture a good deal of common variation.

The autoregressive coefficient $\lambda_i$ differs little between models. Except for AUD and MXN, the posterior mean of $\delta_i$ is within 1.5 posterior standard deviations of zero; for AUD and MXN, it is between 1.5 and 2.0. The $\delta_i$ appear to matter little, but their posterior means are negative for all but one currency: individual log-volatilities appear to be slightly repelled from their cross-sectional average.

Table 7 summarizes the posterior distribution of the unconditional variance $\Sigma$ of the idiosyncratic log volatility processes and the degrees of freedom parameters $\nu_i$. Equation (8) parameterizes $\Sigma$ in terms of the vectors $\sigma$ and $\beta$. For each currency $i$, the table reports the posterior mean and standard deviation of $\sigma_i$ and $\beta_i$; diagonal and off diagonal elements of $\Sigma$ are $\Sigma_{ii} = \sigma_i + \beta_i^2$ and $\Sigma_{ij} = \beta_i\beta_j$. Here there are much stronger signs of dependence. The common sign of the posterior means of the $\beta_i$ indicates positive unconditional correlation for all currency pairs. Many of the correlations are well above zero, with some greater than 0.5. The $\nu_i$ have lower posterior means in the MSV model than in the univariate models for most of the currencies, suggesting that while there is a lot of common variation across currencies, there are some large shocks to individual currencies that are not well accounted for by the factors. The posterior distributions of the $\nu_i$ vary widely across currencies: the posterior mean ranges from 4.40 for JPY to 32.11 for BRL.

Table 8 reports posterior means and standard deviations of the parameters governing the three factor series. For the three factors, log-volatility persistence is high and similar to that of the idiosyncratic log-volatilities. The first factor has a lower degrees-of-freedom parameter, indicating high conditional kurtosis. The third factor has an unconditional standard deviation of log-volatility that is considerably higher than the other two.

Table 9 reports posterior means and standard deviations of factor loadings $B_{ij}$. Recall that the indices $j = 1, 2, 3$ are chosen so that the Euro loadings are in descending order. For almost all currencies, the first factor loading is highest in absolute value with high posterior probability. The common sign suggests the factor is related to the US dollar, the numeraire currency. Loadings for the second factor are much lower in absolute value, for most of the currencies. Only EUR, CHF, AUD and NZD have loadings greater than 0.001 in absolute value, with those of the European currencies (EUR, CHF) having the opposite sign of those of the Australasian (AUD, NZD) ones. Loadings for the third factor are remarkably high except for SGD and CHF, but recall the high unconditional variance of the factor's log volatility: the factor is close to zero except during a few periods of high volatility. So while the variance attributable to the third factor is relatively low on average, it is quite high during these highly volatility periods.

Table 10 shows posterior moments of the elements of $R$, the copula correlation matrix. The mean of most, but definitely not all, copula correlations is within two standard deviations from zero and much closer to zero than the sample correlations in Table 5, as we would expect since the factors are capturing much of the cross-sectional dependence. The mean of the MXN-BRL correlation is nearly eight standard deviations away from zero. The MXN-JPY, MXN-SGD and MXN-CAD correlations also have means that lie outside two standard deviations, as does the AUD-NZD correlation. Even these correlations are quite a bit smaller than the sample correlations. Although the three factors capture much of the dependence among currencies, the copula is also clearly capturing some important remaining dependence.

Figures 1 and 2 show the posterior mean and standard deviation of the idiosyncratic volatility, as well as the numerical standard error (nse) for the posterior mean, for the currencies EUR and JPY, over time. We prefer this graphical display to the more usual practice of plotting inter-quantile bands because it makes it easier to see how the variance of log-volatility varies over time. In these and later figures, we see that at the scale used to plot the mean and variance of log-volatility, the nse is barely distinguishable from zero. This is exactly what we are trying to convey; the uncertainty associated with simulation noise (measured by the nse) is much smaller than the posterior uncertainty (measured by the posterior standard deviation). The relatively low nse is attributable to the use of the HESSIAN method. McCausland (2012) documents (Table 3) the high numerical efficiency of this method compared to auxiliary mixture model methods.

Figures 3 through 5 show the posterior mean and standard deviation of the three factors, through time, and the posterior mean and standard deviation of the their respective volatilities. Again, the numerical standard error of the mean (this time for the factors and their volatility) shows that the uncertainty associated with simulation noise is small compared to the posterior uncertainty. Consistent with the posterior distributions of their respective parameters, the first factor series exhibits a large number of outliers. The third factor series exhibits a great deal of variation in volatility; it is very important in volatile spells and unimportant during tranquil ones.

## 6. Conclusions

We have introduced a new approach to posterior simulation for MSV models, using the HESSIAN method, a numerically efficient method for drawing univariate volatility series; it can be applied one series at a time. The method is flexible, allowing model specifications with different types of dependence. It is less model specific than auxiliary mixture methods, and does not require that the model be transformable to a form where volatility sequences are independent and transformed innovations are identically distributed. We tested and failed to reject the hypothesis that our implementation is correct.

We now revisit the features described in Section 1.4, in the light of our empirical results, illustrating their importance. We get time-varying conditional correlations by incorporating factors in the return equation; the factors are independent SV processes with heterogeneous Student's $t$ innovations. Unsurprisingly, we find abundant evidence for factors. More interestingly, we find evidence that factors are fat-tailed to different degrees, justifying the flexibility we allow by not requiring that factors be multivariate Gaussian or any mixture of these such as the multivariate Student's $t$.

We also find evidence that idiosyncratic log-volatility (return volatility remaining after conditioning on factors) features cross-sectional dependence. We allowed non-diagonal autocorrelation ($A$) and unconditional variance ($\Sigma$) of idiosyncratic log-volatility, but in a parsimonious way. The posterior distribution of the coefficients $\delta_i$ of the lagged cross-sectional average points to a tendency for log-volatilities to be repelled away from this average, although the evidence is not strong. The evidence for positive unconditional correlation across log-volatilities is much stronger.

We incorporate copulas to allow conditional return dependence given factors, without giving up heterogeneity in marginals. We saw that the evidence for such heterogeneity (here, in the $\nu_i$ parameters of the Student's $t$) was strong. The correlation matrix defining the Gaussian copula has a somewhat informative prior, designed to shrink correlations towards zero. In this way we favour factor structure but allow for correlations that are not well captured by the factors. It turns out that most correlations have distributions with non-negligible mass on both sides of zero. However, five out of forty-five correlations have a posterior mean more than two posterior standard deviations away from zero, including one with a mean nearly eight standard deviations away. The factors capture much of the common variation in returns, but the copula clearly captures some remaining conditional dependence.

We find three volatility factors to be important, despite the fact that a (static and homoscedastic) principal components analysis favours two. The volatility of the third factor varies widely over time, in such a way that the factor picks up a lot of common variation during some highly volatile periods and stays close to zero during more tranquil periods.

## Appendix A. Computing $\bar{\Omega}^{(i)}$ and $\bar{c}^{(i)}$

Here we compute $\bar{\Omega}^{(i)}$ and $\bar{c}^{(i)}$, the conditional precision and covector of the Gaussian conditional distribution $\alpha_i|\alpha_{-i}$, in terms of $\bar{\Omega}$ and $\bar{c}$, the prior precision and covector of $\alpha$.

$\bar{\Omega}$ is a $nm \times nm$ block band-diagonal matrix. We denote by $\bar{\Omega}_{st}$, $s,t = 1,\ldots,n$, the $m \times m$ submatrix at row $(s-1)m+1$ and column $(t-1)m+1$. The non-zero submatrices are

$$\bar{\Omega}_{11} = \Sigma_0^{-1} + A^\top \Sigma^{-1}, \qquad \bar{\Omega}_{nn} = \Sigma^{-1},$$

$$\bar{\Omega}_{tt} = \Sigma^{-1} + A^\top \Sigma^{-1} A, \quad t = 2,\ldots,n-1,$$

$$\bar{\Omega}_{t+1,t}^\top = \bar{\Omega}_{t,t+1} = -A^\top \Sigma^{-1}, \quad t = 1,\ldots,n-1.$$

The co-vector is a $nm \times 1$ vector stacking $n$ $m \times 1$ subvectors $\bar{c}_t$, given by:

$$\bar{c}_1 = \Sigma_0^{-1}\bar{\alpha} - A^\top \Sigma^{-1}(I-A)\bar{\alpha}, \qquad \bar{c}_n = \Sigma^{-1}(I-A)\bar{\alpha}.$$

$$\bar{c}_t = \Sigma^{-1}(I-A)\bar{\alpha} - A^\top \Sigma^{-1}(I-A)\bar{\alpha}, \quad t = 2,\ldots,n-1.$$

We now derive $\bar{\Omega}^{(i)}$ and $\bar{c}^{(i)}$. We know $\pi(\alpha_i|\alpha_{-i}) \propto \pi(\alpha)$ as a function of $\alpha_i$. Matching coefficients of first and second order terms of $\log \pi(\alpha_i|\alpha_{-i})$ gives the non-zero elements

$$\bar{\Omega}_{tt}^{(i)} = (\bar{\Omega}_{tt})_{ii}, \quad \bar{\Omega}_{t,t+1}^{(i)} = \bar{\Omega}_{t+1,t}^{(i)} = (\bar{\Omega}_{t,t+1})_{ii}.$$

$$\bar{c}_t^{(i)} = (\bar{c}_t)_i - \sum_{j\neq i} \left[ (\bar{\Omega}_{tt})_{ji}\alpha_{tj} + (\bar{\Omega}_{t,t+1})_{ji}\alpha_{t+1,j} + (\bar{\Omega}_{t-1,t})_{ji}\alpha_{t-1,j} \right].$$

## Appendix B. Computing $\log \pi(y_t|\alpha_t, \nu, B, R)$ and derivatives

Using equations (4), (5), and (6), we can write $\log \pi(y_t|\alpha_t, B, \nu, R)$ as

$$\log \pi(y_t|\alpha_t,\nu,B,R) = -\frac{1}{2}\left\{ \log|R| + \log 2\pi + x_t^\top (R^{-1}-I)x_t + \sum_{i=1}^m \left[ \alpha_{ti} + (\nu_i+1)\log\left(1+\frac{\epsilon_{ti}^2}{\nu_i}\right)\right]\right\}$$
$$+ \sum_{i=1}^m \left[ \log\Gamma\left(\frac{\nu_i+1}{2}\right) - \log\Gamma\left(\frac{\nu_i}{2}\right) - \frac{1}{2}\log(\nu_i\pi)\right],$$

where $x_t = (x_{t1},\ldots,x_{tm})$ and for $i = 1,\ldots,m$, $x_{ti} = \Phi^{-1}(u_{ti})$, $u_{ti} = F_\epsilon(\epsilon_{ti}|\nu_i))$, and

$$\epsilon_{ti} = \begin{cases} \exp(-\alpha_{ti}/2)(r_{ti} - \sum_{j=1}^q B_{ij} f_{tj}), & i = 1,\ldots,p, \\ \exp(-\alpha_{ti}/2) f_{t,i-p}, & i = p+1,\ldots,m. \end{cases}$$

We can evaluate $\log \pi(y_t|\alpha_t, B, \nu, R)$ as a function of $\alpha_{ti}$ bottom up, evaluating the $\epsilon_{ti}$ at $\alpha_{ti}$, then the $u_{ti}$ at $\epsilon_{ti}$, then the $x_{ti}$ at $u_{ti}$ then $\log \pi(y_t|\alpha_t, B, \nu, R)$ at $\epsilon_t$ and $x_t$.

We require five derivatives of $\log \pi(y_t|\alpha_t, B, \nu, R)$ with respect to $\alpha_{ti}$, evaluated at $\alpha_{ti}$. Because it is a multi-level compound function of the $\alpha_{ti}$, computing these in closed form would be tedious and error-prone. Instead, we compute any values we need, bottom up, using Faà di Bruno's formula at each step to compute derivatives of a compound function by combining derivatives of its component functions. We compute, in order,

1. five derivatives of $\psi(\alpha_{ti}) \equiv \log \pi_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)$ with respect to $\alpha_{ti}$ at $\alpha_{ti}$ (B.1).
2. five derivatives of $x^\top(R^{-1}-I)x$ with respect to $x_{ti}$ at $x_{ti}$, as described in B.2.
3. five derivatives of $x_{ti}$ with respect to $u_{ti}$ at $u_{ti}$, as described in B.3.
4. five derivatives of $u_{ti}$ with respect to $\alpha_{ti}$ at $\alpha_{ti}$, as described in B.4.
5. five derivatives of $x_{ti}$ with respect to $\alpha_{ti}$ at $\alpha_{ti}$, using the Faà di Bruno formula to combine the derivatives of $x_{ti}$ with respect to $u_{ti}$ at step 3 and the derivatives of $u_{ti}$ with respect to $\alpha_{ti}$ at step 4.

(6) five derivatives of $x^\top(R^{-1}-I)x$ with respect to $\alpha_{ti}$ at $\alpha_{ti}$, using the Faà di Bruno formula to combine the derivatives of $x^\top(R^{-1}-I)x$ with respect to $x_{ti}$ at step 2 and the derivatives of $x_{ti}$ with respect to $\alpha_{ti}$ at step 5.

(7) five derivatives of $\log\pi(y_t|\alpha_t,\theta,B,R)$ with respect to $\alpha_{ti}$ at $\alpha_{ti}$ using the derivatives at steps 1 and 6.

We define $\eta_t=(\eta_{t1},\ldots,\eta_{tm})^\top=((r_t-Bf_t)^\top,f_t^\top)^\top$ to simplify notation below.

B.1. **Derivatives of $\psi(\alpha_{ti})$ with respect to $\alpha_{ti}$.** For the special case of Student's $t$ F,

$$\pi_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|v_i)=\frac{\Gamma(\frac{\nu_i+1}{2})}{\sqrt{\nu_i\pi}\Gamma(\frac{\nu_i}{2})}\left(1+\frac{e^{-\alpha_{ti}}\eta_{ti}^2}{\nu_i}\right)^{-\frac{\nu_i+1}{2}}$$

$$\psi(\alpha_{ti})=\log\left[\frac{\Gamma(\frac{\nu_i+1}{2})}{\sqrt{\nu_i\pi}\Gamma(\frac{\nu_i}{2})}\right]-\frac{\nu_i+1}{2}\log(1+s_{ti})$$

where $s_{ti}\equiv e^{-\alpha_{ti}}\eta_{ti}^2/\nu_i$. Noting that $\partial s_{ti}/\partial\alpha_i=-s_{ti}$, we compute

$$\psi'(\alpha_{ti})=\frac{\nu_i+1}{2}\frac{s_{ti}}{1+s_{ti}},\quad \psi''(\alpha_{ti})=-\frac{\nu_i+1}{2}\frac{s_{ti}}{(1+s_{ti}^2)},$$

$$\psi'''(\alpha_{ti})=\frac{\nu_i+1}{2}\frac{s_{ti}(1-s_{ti})}{(1+s_{ti})^3},\quad \psi^{(4)}(\alpha_{ti})=-\frac{\nu_i+1}{2}\frac{s_{ti}(1-4s_{ti}+s_{ti}^2)}{(1+s_{ti})^4},$$

$$\psi^{(5)}(\alpha_{ti})=\frac{\nu_i+1}{2}\frac{s_{ti}(1-11s_{ti}+11s_{ti}^2-s_{ti}^3)}{(1+s_{ti})^5}.$$

B.2. **Derivatives of $x^\top(I-R^{-1})x$ with respect to $x_{ti}$.** Here we compute partial derivatives of $\log c(u_1,\ldots,u_m)$ with respect to the $u_i$. We can write

$$\begin{aligned}\log c_R(u_1,\ldots,u_m)&=\log\phi_R(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_m))-\sum_{i=1}^m\log\phi(\Phi^{-1}(u_i))\\&=\frac{1}{2}|H|+\frac{1}{2}x^\top(I-R^{-1})x,\end{aligned}$$

where $x=(x_1,\ldots,x_m)=(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_m))$. The gradient and Hessian of $\log(c_R)$ with respect to $u$ are as follows; all third order partial derivatives and higher are zero.

$$\frac{\partial\log c(u)}{\partial x}=(I-R^{-1})x,\quad \frac{\partial\log c(u)}{\partial x\partial x^\top}=I-R^{-1}.$$

B.3. **Derivatives of $x_{ti}$ with respect to $u_{ti}$.** Differentiating $\Phi(x_i)=u_i$ with respect to $u_i$ gives $\phi(x_i)\frac{\partial x_i}{\partial u_i}=1$, and thus

$$\frac{\partial x_i}{\partial u_i}=\frac{1}{\phi(x_i)},\quad \frac{\partial^2 x_i}{\partial u_i}=2\pi e^{x_i^2}x_i,\quad \frac{\partial^3 x_i}{\partial u_i}=(2\pi)^{3/2}e^{3x_i^2/2}(2x_i^2+1),$$

$$\frac{\partial^4 x_i}{\partial u_i}=(2\pi)^2e^{2x_i^2}(6x_i^3+7x_i),\quad \frac{\partial^5 x_i}{\partial u_i}=(2\pi)^{5/2}e^{5x_i^2/2}(24x_i^4+46x_i^2+7).$$

B.4. **Derivatives of** $F_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)$**.** Here we compute five derivatives of $F_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)$ with respect to $\alpha_{ti}$. We write down the derivatives in terms of $\psi(\alpha_{ti}) \equiv \log \pi_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)$:

$$\frac{\partial F_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)}{\partial \alpha_{ti}} = \pi_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)\left(-\frac{1}{2}e^{-\alpha_{ti}/2}\eta_{ti}\right) = -\frac{\eta_{ti}}{2}e^{-0.5\alpha_{ti}+\psi(\alpha_{ti})},$$

$$\frac{\partial^2 F_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)}{\partial \alpha_{ti}^2} = -\frac{\eta_{ti}}{2}e^{-0.5\alpha_{ti}+\psi(\alpha_{ti})}[-0.5+\psi'(\alpha_{ti})]$$

$$\frac{\partial^3 F_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)}{\partial \alpha_{ti}^3} = -\frac{\eta_{ti}}{2}e^{-0.5\alpha_{ti}+\psi(\alpha_{ti})}\left[\psi''(\alpha_{ti}) + (-0.5+\psi'(\alpha_{ti})^2\right]$$

$$\frac{\partial^4 F_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)}{\partial \alpha_{ti}^4} = -\frac{\eta_{ti}}{2}e^{-0.5\alpha_{ti}+\psi(\alpha_{ti})}\left[\psi'''(\alpha_{ti}) + 3(-0.5+\psi'(\alpha_{ti}))\psi''(\alpha_{ti}) + (-0.5+\psi'(\alpha_{ti}))^3\right]$$

$$\begin{aligned}\frac{\partial^5 F_\epsilon(e^{-\alpha_{ti}/2}\eta_{ti}|\theta_i)}{\partial \alpha_{ti}^5} = -\frac{\eta_{ti}}{2}e^{\psi(\alpha_{ti})}\Big[&\psi^{(4)}(\alpha_{ti}) + 4(-0.5+\psi'(\alpha_{ti}))\psi'''(\alpha_{ti}) + 3(\psi''(\alpha_{ti}))^2 \\ &+ 6(-0.5+\psi'(\alpha_{ti}))^2\psi''(\alpha_{ti}) + (-0.5+\psi'(\alpha_{ti}))^4\Big]\end{aligned}$$

## Appendix C. Drawing $r|\alpha, \theta, f, B, R$

Here we draw $r$ from $\pi(r|\alpha, \theta, f, B, R)$. We first compute the Cholesky decomposition $R = LL^\top$ of the correlation matrix $R$. Then for each $t = 1, ..., n$:

(1) Draw $z \sim N(0, I_m)$, set $g = Lz$ so that $g \sim N(0, R)$.
(2) Compute the integral probability transforms $u_i = \Phi(g_i)$, $i = 1, ..., m$.
(3) Transform each $u_i$ to a Student's $t$ with $\nu_i$ degree of freedom: $\tau_i = F_\nu^{-1}(u_i)$, where $F_\nu$ is the cdf of a Student's $t$ with $\nu_i$ degrees of freedom.
(4) Scale each of the $\tau_i$ to form $\epsilon_{ti} = \tau_i \exp(0.5\alpha_{ti})$, construct $r_t = Bf_t + \epsilon_t$.

## Appendix D. Tables of results

| | | Independent SVt | | | | Full Model | | | |
|---|---|---|---|---|---|---|---|---|---|
| Series | | mean | sd | nse | rne | mean | sd | nse | rne |
| AUD | $\bar{\alpha}_i$ | -10.01 | 0.20 | 0.001 | 0.496 | -12.21 | 0.26 | 0.008 | 0.020 |
| | $\lambda_i$ | 0.9888 | 0.0025 | 0.00002 | 0.475 | 0.9868 | 0.0050 | 0.00016 | 0.019 |
| | $\delta_i$ | | | | | -0.0134 | 0.0080 | 0.00029 | 0.015 |
| BRL | $\bar{\alpha}_i$ | -9.83 | 0.19 | 0.001 | 0.502 | -10.33 | 0.21 | 0.005 | 0.036 |
| | $\lambda_i$ | 0.9788 | 0.0043 | 0.00003 | 0.504 | 0.9814 | 0.0055 | 0.00018 | 0.018 |
| | $\delta_i$ | | | | | -0.0097 | 0.0080 | 0.00029 | 0.015 |
| EUR | $\bar{\alpha}_i$ | -10.49 | 0.20 | 0.001 | 0.493 | -14.55 | 0.63 | 0.034 | 0.007 |
| | $\lambda_i$ | 0.9919 | 0.0020 | 0.00001 | 0.445 | 0.9818 | 0.0048 | 0.00020 | 0.012 |
| | $\delta_i$ | | | | | -0.0085 | 0.0117 | 0.00043 | 0.015 |
| JPY | $\bar{\alpha}_i$ | -10.55 | 0.17 | 0.001 | 0.469 | -11.95 | 0.28 | 0.011 | 0.012 |
| | $\lambda_i$ | 0.9867 | 0.0035 | 0.00002 | 0.471 | 0.9872 | 0.0033 | 0.00009 | 0.029 |
| | $\delta_i$ | | | | | -0.0077 | 0.0054 | 0.00017 | 0.021 |
| MXN | $\bar{\alpha}_i$ | -10.57 | 0.20 | 0.001 | 0.463 | -11.13 | 0.28 | 0.008 | 0.026 |
| | $\lambda_i$ | 0.9849 | 0.0033 | 0.00002 | 0.493 | 0.9902 | 0.0032 | 0.00011 | 0.017 |
| | $\delta_i$ | | | | | -0.0112 | 0.0056 | 0.00019 | 0.018 |
| NZD | $\bar{\alpha}_i$ | -9.82 | 0.18 | 0.001 | 0.414 | -11.26 | 0.16 | 0.004 | 0.030 |
| | $\lambda_i$ | 0.9896 | 0.0026 | 0.00002 | 0.463 | 0.9877 | 0.0044 | 0.00011 | 0.031 |
| | $\delta_i$ | | | | | -0.0050 | 0.0049 | 0.00014 | 0.023 |
| SGD | $\bar{\alpha}_i$ | -11.76 | 0.19 | 0.001 | 0.467 | -12.82 | 0.19 | 0.005 | 0.037 |
| | $\lambda_i$ | 0.9875 | 0.0029 | 0.00002 | 0.497 | 0.9842 | 0.0056 | 0.00016 | 0.023 |
| | $\delta_i$ | | | | | -0.0073 | 0.0070 | 0.00025 | 0.015 |
| CHF | $\bar{\alpha}_i$ | -10.43 | 0.18 | 0.001 | 0.455 | -13.11 | 0.36 | 0.012 | 0.018 |
| | $\lambda_i$ | 0.9896 | 0.0027 | 0.00002 | 0.544 | 0.9827 | 0.0047 | 0.00018 | 0.014 |
| | $\delta_i$ | | | | | -0.0013 | 0.0081 | 0.00024 | 0.022 |
| GBP | $\bar{\alpha}_i$ | -10.55 | 0.20 | 0.001 | 0.426 | -11.34 | 0.18 | 0.005 | 0.032 |
| | $\lambda_i$ | 0.9923 | 0.0019 | 0.00001 | 0.490 | 0.9876 | 0.0044 | 0.00013 | 0.021 |
| | $\delta_i$ | | | | | 0.0014 | 0.0050 | 0.00017 | 0.018 |
| CAD | $\bar{\alpha}_i$ | -10.61 | 0.20 | 0.001 | 0.432 | -11.27 | 0.19 | 0.004 | 0.041 |
| | $\lambda_i$ | 0.9903 | 0.0023 | 0.00002 | 0.416 | 0.9901 | 0.0033 | 0.00009 | 0.024 |
| | $\delta_i$ | | | | | -0.0066 | 0.0044 | 0.00015 | 0.018 |

Table 6. Posterior mean and standard deviation, numerical standard error and relative numerical efficiency for the parameters $\bar{\alpha}_i$, $\lambda_i$ and $\delta_i$: at left, independent SVt models; at right, MSV model

| | | Independent SVt | | | | Full Model | | | |
|---|---|---|---|---|---|---|---|---|---|
| Series | | mean | sd | nse | rne | mean | sd | nse | rne |
| AUD | $\sigma_i$ | 0.751 | 0.073 | 0.000 | 0.529 | 0.846 | 0.090 | 0.004 | 0.009 |
| | $\beta_i$ | | | | | 0.630 | 0.175 | 0.007 | 0.011 |
| | $\nu_i$ | 20.04 | 8.24 | 0.07 | 0.262 | 10.14 | 5.36 | 0.24 | 0.010 |
| BRL | $\sigma_i$ | 0.939 | 0.079 | 0.001 | 0.491 | 0.918 | 0.089 | 0.004 | 0.012 |
| | $\beta_i$ | | | | | 0.638 | 0.172 | 0.008 | 0.010 |
| | $\nu_i$ | 38.16 | 21.55 | 0.15 | 0.408 | 32.11 | 17.97 | 0.44 | 0.033 |
| EUR | $\sigma_i$ | 0.625 | 0.067 | 0.000 | 0.495 | 1.526 | 0.207 | 0.011 | 0.007 |
| | $\beta_i$ | | | | | 1.793 | 0.447 | 0.028 | 0.005 |
| | $\nu_i$ | 14.98 | 4.87 | 0.04 | 0.374 | 25.70 | 18.81 | 0.61 | 0.019 |
| JPY | $\sigma_i$ | 0.638 | 0.068 | 0.000 | 0.515 | 0.857 | 0.116 | 0.004 | 0.014 |
| | $\beta_i$ | | | | | 0.132 | 0.174 | 0.007 | 0.013 |
| | $\nu_i$ | 7.73 | 1.21 | 0.01 | 0.515 | 4.40 | 0.63 | 0.03 | 0.012 |
| MXN | $\sigma_i$ | 0.859 | 0.079 | 0.001 | 0.443 | 0.893 | 0.093 | 0.004 | 0.010 |
| | $\beta_i$ | | | | | 0.698 | 0.179 | 0.008 | 0.011 |
| | $\nu_i$ | 33.94 | 18.74 | 0.12 | 0.476 | 16.38 | 6.64 | 0.18 | 0.028 |
| NZD | $\sigma_i$ | 0.639 | 0.068 | 0.000 | 0.489 | 0.555 | 0.071 | 0.002 | 0.019 |
| | $\beta_i$ | | | | | 0.200 | 0.093 | 0.004 | 0.011 |
| | $\nu_i$ | 15.71 | 5.06 | 0.04 | 0.286 | 6.90 | 1.28 | 0.04 | 0.022 |
| SGD | $\sigma_i$ | 0.714 | 0.072 | 0.000 | 0.505 | 0.759 | 0.080 | 0.003 | 0.011 |
| | $\beta_i$ | | | | | 0.487 | 0.139 | 0.005 | 0.014 |
| | $\nu_i$ | 10.74 | 2.41 | 0.02 | 0.438 | 7.11 | 1.21 | 0.03 | 0.030 |
| CHF | $\sigma_i$ | 0.615 | 0.069 | 0.000 | 0.486 | 1.281 | 0.129 | 0.005 | 0.014 |
| | $\beta_i$ | | | | | 1.262 | 0.284 | 0.016 | 0.006 |
| | $\nu_i$ | 7.97 | 1.23 | 0.01 | 0.512 | 6.21 | 1.63 | 0.07 | 0.012 |
| GBP | $\sigma_i$ | 0.613 | 0.066 | 0.000 | 0.514 | 0.627 | 0.070 | 0.003 | 0.015 |
| | $\beta_i$ | | | | | 0.306 | 0.107 | 0.005 | 0.010 |
| | $\nu_i$ | 32.99 | 16.87 | 0.13 | 0.349 | 19.80 | 8.41 | 0.21 | 0.031 |
| CAD | $\sigma_i$ | 0.698 | 0.071 | 0.000 | 0.459 | 0.652 | 0.071 | 0.002 | 0.018 |
| | $\beta_i$ | | | | | 0.282 | 0.106 | 0.004 | 0.012 |
| | $\nu_i$ | 17.35 | 6.63 | 0.06 | 0.234 | 26.31 | 14.41 | 0.39 | 0.027 |

TABLE 7. Posterior mean and standard deviation, numerical standard error and relative numerical efficiency for the parameters $\sigma_i$, $\beta_i S$ and $\nu_i$: at left, independent SVt models; at right, MSV model

| Series | | mean | sd | nse | rne |
|---|---|---|---|---|---|
| $f_1$ | $\lambda_i$ | 0.9904 | 0.0025 | 0.00005 | 0.048 |
| | $\sigma_i$ | 0.592 | 0.066 | 0.001 | 0.060 |
| | $\nu_i$ | 12.58 | 3.76 | 0.10 | 0.031 |
| $f_2$ | $\lambda_i$ | 0.9879 | 0.0037 | 0.00009 | 0.034 |
| | $\sigma_i$ | 0.593 | 0.074 | 0.002 | 0.028 |
| | $\nu_i$ | 23.38 | 12.91 | 0.32 | 0.032 |
| $f_3$ | $\lambda_i$ | 0.9853 | 0.0031 | 0.00007 | 0.040 |
| | $\sigma_i$ | 1.279 | 0.138 | 0.005 | 0.013 |
| | $\nu_i$ | 18.67 | 12.79 | 0.39 | 0.021 |

TABLE 8. Posterior mean and standard deviation, numerical standard error and relative numerical efficiency for the parameters of the factor series, full model with $q = 3$

| Series | | mean | sd | nse | rne |
|---|---|---|---|---|---|
| AUD | $B_{i1}$ | 0.00430 | 0.00041 | 0.00002 | 0.007 |
| | $B_{i2}$ | -0.00170 | 0.00031 | 0.00002 | 0.006 |
| | $B_{i3}$ | 0.00331 | 0.00054 | 0.00003 | 0.005 |
| BRL | $B_{i1}$ | 0.00233 | 0.00030 | 0.00002 | 0.007 |
| | $B_{i2}$ | -0.00023 | 0.00023 | 0.00001 | 0.008 |
| | $B_{i3}$ | 0.00283 | 0.00044 | 0.00003 | 0.006 |
| EUR | $B_{i1}$ | 0.00416 | 0.00036 | 0.00002 | 0.007 |
| | $B_{i2}$ | 0.00236 | 0.00033 | 0.00002 | 0.006 |
| | $B_{i3}$ | 0.00106 | 0.00025 | 0.00002 | 0.005 |
| JPY | $B_{i1}$ | 0.00297 | 0.00032 | 0.00002 | 0.006 |
| | $B_{i2}$ | -0.00058 | 0.00021 | 0.00001 | 0.006 |
| | $B_{i3}$ | -0.00210 | 0.00031 | 0.00002 | 0.006 |
| MXN | $B_{i1}$ | 0.00166 | 0.00022 | 0.00001 | 0.007 |
| | $B_{i2}$ | -0.00051 | 0.00019 | 0.00001 | 0.007 |
| | $B_{i3}$ | 0.00194 | 0.00033 | 0.00002 | 0.006 |
| NZD | $B_{i1}$ | 0.00435 | 0.00041 | 0.00002 | 0.007 |
| | $B_{i2}$ | -0.00146 | 0.00031 | 0.00002 | 0.006 |
| | $B_{i3}$ | 0.00325 | 0.00054 | 0.00003 | 0.005 |
| SGD | $B_{i1}$ | 0.00186 | 0.00016 | 0.00001 | 0.008 |
| | $B_{i2}$ | -0.00019 | 0.00012 | 0.00001 | 0.005 |
| | $B_{i3}$ | 0.00080 | 0.00016 | 0.00001 | 0.005 |
| CHF | $B_{i1}$ | 0.00449 | 0.00037 | 0.00002 | 0.008 |
| | $B_{i2}$ | 0.00232 | 0.00034 | 0.00002 | 0.005 |
| | $B_{i3}$ | 0.00016 | 0.00026 | 0.00002 | 0.005 |
| GBP | $B_{i1}$ | 0.00305 | 0.00026 | 0.00001 | 0.008 |
| | $B_{i2}$ | 0.00069 | 0.00021 | 0.00001 | 0.006 |
| | $B_{i3}$ | 0.00121 | 0.00024 | 0.00001 | 0.006 |
| CAD | $B_{i1}$ | 0.00232 | 0.00025 | 0.00001 | 0.007 |
| | $B_{i2}$ | -0.00058 | 0.00019 | 0.00001 | 0.006 |
| | $B_{i3}$ | 0.00208 | 0.00035 | 0.00002 | 0.006 |

TABLE 9. Posterior mean and standard deviation, numerical standard error and relative numerical efficiency for factor loadings, full model with $q = 3$

| i/j | AUD | BRL | EUR | JPY | MXN | NZD | SGD | CHF | GBP | CAD |
|---|---|---|---|---|---|---|---|---|---|---|
| AUD | 1.000 | | | | | | | | | |
| BRL | -0.017 | 1.000 | | | | | | | | |
| EUR | -0.003 | 0.009 | 1.000 | | | | | | | |
| JPY | -0.003 | -0.006 | -0.021 | 1.000 | | | | | | |
| MXN | -0.024 | 0.165 | 0.019 | -0.052 | 1.000 | | | | | |
| NZD | 0.062 | -0.032 | -0.013 | -0.015 | -0.036 | 1.000 | | | | |
| SGD | -0.019 | 0.032 | 0.030 | 0.043 | 0.076 | -0.032 | 1.000 | | | |
| CHF | 0.005 | -0.011 | -0.013 | 0.040 | -0.034 | 0.021 | -0.026 | 1.000 | | |
| GBP | -0.019 | -0.029 | 0.003 | -0.007 | -0.015 | 0.008 | -0.004 | 0.002 | 1.000 | |
| CAD | -0.012 | 0.036 | 0.003 | -0.030 | 0.072 | -0.025 | 0.007 | -0.008 | 0.018 | 1.000 |
| AUD | 0.000 | | | | | | | | | |
| BRL | 0.023 | 0.000 | | | | | | | | |
| EUR | 0.025 | 0.024 | 0.000 | | | | | | | |
| JPY | 0.024 | 0.023 | 0.025 | 0.000 | | | | | | |
| MXN | 0.023 | 0.019 | 0.024 | 0.024 | 0.000 | | | | | |
| NZD | 0.024 | 0.020 | 0.024 | 0.022 | 0.021 | 0.000 | | | | |
| SGD | 0.023 | 0.020 | 0.024 | 0.022 | 0.020 | 0.021 | 0.000 | | | |
| CHF | 0.024 | 0.021 | 0.025 | 0.025 | 0.022 | 0.023 | 0.022 | 0.000 | | |
| GBP | 0.022 | 0.020 | 0.024 | 0.023 | 0.019 | 0.021 | 0.020 | 0.023 | 0.000 | |
| CAD | 0.023 | 0.020 | 0.024 | 0.023 | 0.020 | 0.021 | 0.020 | 0.021 | 0.020 | 0.000 |
| AUD | 0.000 | | | | | | | | | |
| BRL | 0.001 | 0.000 | | | | | | | | |
| EUR | 0.001 | 0.001 | 0.000 | | | | | | | |
| JPY | 0.001 | 0.001 | 0.001 | 0.000 | | | | | | |
| MXN | 0.001 | 0.000 | 0.001 | 0.001 | 0.000 | | | | | |
| NZD | 0.001 | 0.000 | 0.001 | 0.001 | 0.001 | 0.000 | | | | |
| SGD | 0.001 | 0.000 | 0.001 | 0.001 | 0.001 | 0.000 | 0.000 | | | |
| CHF | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.001 | 0.000 | | |
| GBP | 0.001 | 0.000 | 0.001 | 0.001 | 0.000 | 0.000 | 0.000 | 0.001 | 0.000 | |
| CAD | 0.001 | 0.000 | 0.001 | 0.001 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

TABLE 10. Elementwise posterior mean (upper panel), posterior standard deviation (middle panel) and numerical standard error (lower panel) of correlation matrix $R_{11}$

## References


Aguilar, O., and West, M. (2000). 'Bayesian dynamic factor models and variance matrix discounting for portfolio allocation', *Journal of Business and Economic Statistics*, 18: 338–357.

Ang, A., and Chen, J. (2002). 'Asymmetric correlations of equity portfolios', *Journal of Financial Econometrics*, 63: 443–494.

Asai, and McAleer (2009). 'The structure of dynamic correlations in multivariate stochastic volatility models', *Journal of Econometrics*, 150: 182–192.

Bauwens, L., Laurent, S., and Rombouts, J. V. K. (2006). 'Multivariate GARCH : A survey', *Journal of Applied Econometrics*, 21: 79–109.

Caldara, D., Fernández-Villaverde, J., Rubio-Ramírez, J. F., and Yao, W. (2012). 'Computing DSGE models with recursive preferences and stochastc volatility', *Review of Economic Dynamics*, 15(2): 188–206.

Carnero, M. A., Pena, D., and Ruiz, E. (2004). 'Persistence and kurtosis in GARCH and stochastic volatility models', *Journal of Financial Econometrics*, 2: 319–342.

Carvalho, C. M., and West, M. (2006). 'Dynamic matrix-variate graphical models', *Bayesian Analysis*, 1: 1–29.

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

Chan, J. J. C., and Grant, A. L. (2016). 'Modeling energy price dynamics: GARCH versus stochastic volatility', *Energy Economics*, 54: 182–189.

Chib, S., Nardari, F., and Shephard, N. (2002). 'Markov Chain Monte Carlo Methods for Stochastic Volatility Models', *Journal of Econometrics*, 108: 281–316.

——— (2006). 'Analysis of High Dimensional Multivariate Stochastic Volatility Models', *Journal of Econometrics*, 134: 341–371.

Chib, S., Omori, Y., and Asai, M. (2009). 'Multivariate Stochastic Volatility.', in T. G. Andersen, R. A. Davis, J.-P. Kreib, and T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 365–400. Springer.

Chollete, L., Heinen, A., and Valdesogo, A. (2009). 'Modeling international financial returns with a multivariate regime-switching copula', *Journal of Financial Econometrics*, 7(4): 437–480.

Christodoulakis, G. A. (2007). 'Common volatility and correlation clustering in asset returns', *European Journal of Operational Research*, 182: 1263–1284.

Cogley, T., and Sargent, T. J. (2005). 'Drifts and Volatilities: Monetary Policies and Outcomes in the Post-World War II U.S.', *Review of Economic Dynamics*, 8: 262–302.

Cont, R. (2001). 'Empirical properties of asset returns: stylized facts and statistical issues', *Quantitative Finance*, 1: 223–236.

Danielsson, J. (1998). 'Multivariate stochastic volatility models: Estimation and comparison with VGARCH models.', *Journal of Empirical Finance*, 5: 155–173.

Eraker, B., and Wang, J. (2015). 'A non-linear dynamic model of the variance risk premium', *Journal of Econometrics*, 187: 547–556.

Geweke, J. (1994). 'Bayesian Comparison of Econometric Models', Working Paper, Federal Reserve Bank of Minneapolis Research Department.

Geweke, J. (2004). 'Getting it Right: Joint Distribution Tests of Posterior Simulators', *Journal of the American Statistical Association*, 99: 799–804.

Gourieroux, C. (2006). 'Continous time Wishart process for stochastic risk.', *Econometric Reviews*, 25: 177–217.

Gourieroux, C., Jasiak, J., and Sufana, R. (2004). 'The Wishart autorregresive process of multivariate stochastic volatility.', Discussion Paper, University of Toronto.

Hafner, C. M., and Manner, H. (2012). 'Dynamic stochastic copula models: estimation, inference and applications', *Journal of Applied Econometrics*, 27: 269–295.

Han (2006). 'The economics value of volatility modelling: Asset allocation with a high dimensional dynamic latent factor multivariate stochastic volatility model.', *Review of Financial Studies*, 19: 237–271.

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

Jacquier, E., Polson, N., and Rossi, P. (1994). 'Bayesian Analysis of Stochastic Volatility Models', *Journal of Business and Economic Statistics*, 12(4): 371–388.

Jacquier, E., Polson, N., and Rossi, P. (1995). 'Models and prior distributions for multivariate stochastic volatility.', Working paper, Graduate School of Business, University of Chicago.

Jin, X., and Maheu, J. M. (2013). 'Modelling Realized Covariances and Returns', *Journal of Financial Econometrics*, 11(2): 335–369.

——— (2016). 'Bayesian Semiparametric Modeling of Realized Covariance Matrices', *Journal of Econometrics*, 192(1): 19–39.

Justiniano, Alejandro; Primiceri, G. E. (2008). 'The time-varying volatility of macroeconomic fluctuations', *American Economic Review*, 98(3): 604–641.

Kim, S., Shephard, N., and Chib, S. (1998). 'Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models', *Review of Economic Studies*, 65(3): 361–393.

Kolev, N., dos Anjos, U., and de M. Mendez, B. V. (2006). 'Copulas: a review and recent developments.', *Stochastic Models*, 22: 617–660.

Koopman, S. J., and Scharth, M. (2013). 'The analysis of stochastic volatility in the presence of daily realized measures', *Journal of Financial Econometrics*, 11(1): 76–115.

Liu, J., and Maheu, J. M. (2015). 'Improving Markov switching models using realized variance', *Journal of Applied Econometrics*, Forthcoming.

Longin, F., and Solnik, B. (2001). 'Extreme correlations of international equity markets.', *Journal of Finance*, 56: 649–676.

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.

McCausland, W. J. (2012). 'The HESSIAN Method: Highly Efficient State Smoothing, in a Nutshell', *Journal of Econometrics*, 168: 189–206.

Min, A., and Czado, C. (2010). 'Bayesian inference for multivariate copulas using pair-copula constructions', *Journal of Financial Econometrics*, 8(4): 511–546.

Oh, D. H., and Patton, A. J. (2013). 'Simulated method of moments estimation for copula-based multivariate models', *Journal of the American Statistical Association*, 108(502): 689–700.

Patton, A. J. (2009). 'Copula-Based Models for Financial Time Series.', in T. G. Andersen, R. A. Davis, J.-P. Kreib, and T. Mikosch (eds.), *Handbook of Financial Time Series*, pp. 767–784. Springer.

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

Pitt, M. K., and Shephard, N. (1999). ‘Time varying covariances: a factor stochastic volatility approach with discussion.’, in J. Bernardo, J. Berger, A. Dawid, and A. Smith (eds.), *Bayesian Statistics*, pp. 547–570. Oxford University Press, Oxford.

Primiceri, G. (2005). ‘Time varying structural vector autoregressions and monetary policy’, *Review of Economic Studies*, 72: 821–852.

Shirota, S., Hizu, T., and Omori, Y. (2014). ‘Realized stochastic volatility with leverage and long memory’, *Computational statistics and data analysis*, 76: 618–641.

Shirota, S., Omori, Y., Lopes, H. F., and Piao, H. (2015). ‘Cholesky realized stochastic volatility model’, working paper.

Sklar, A. (1959). ‘Fonctions de répartition à $n$ dimensions et leurs marges.’, *Publications de l’Institut Statistique de l’Université de Paris*, 8: 229–231.

Smith, M., and Pitts, A. (2006). ‘Foreign exchange intervention by the Bank of Japan: Bayesian analyisis using a bivariate stochastic volatility model.’, *Econometric Reviews*, 25: 425–451.

So, M. K. P., Li, W. K., and Lam, K. (1997). ‘Multivariate modelling of the autorregresive random variance process.’, *Journal of Time Series Analysis*, 18: 429–446.

Venter, J. H., and de Jongh, P. J. (2014). ‘Extended stochastic volatility models incorporating realised measures’, *Computational statistics and data analysis*, 76: 687–707.

Vihola, M. (2011). ‘Robust Adaptive Metropolis Algorithm with Coerced Acceptance Rate’, arXiv eprint.

Yu, J., and Meyer, R. (2006). ‘Multivariate stochastic volatility models: Bayesian estimation and model comparison.’, *Econometric Reviews*, 25: 361–384.