Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Econometrics 134 (2006) 341–371
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

Analysis of high dimensional multivariate stochastic volatility models

Siddhartha Chib[a], Federico Nardari[b], Neil Shephard[c,*]

[a]John M. Olin School of Business, Washington University, St Louis, MO 63130, USA
[b]W.P. Carey School of Business, Arizona State University, Tempe, AZ 85287, USA
[c]Nuffield College, University of Oxford, Oxford OX1 1NF, UK

Available online 22 August 2005

Abstract

This paper is concerned with the Bayesian estimation and comparison of flexible, high dimensional multivariate time series models with time varying correlations. The model proposed and considered here combines features of the classical factor model with that of the heavy tailed univariate stochastic volatility model. A unified analysis of the model, and its special cases, is developed that encompasses estimation, filtering and model choice. The centerpieces of the estimation algorithm (which relies on MCMC methods) are: (1) a reduced blocking scheme for sampling the free elements of the loading matrix and the factors and (2) a special method for sampling the parameters of the univariate SV process. The resulting algorithm is scalable in terms of series and factors and simulation-efficient. Methods for estimating the log-likelihood function and the filtered values of the time-varying volatilities and correlations are also provided. The performance and effectiveness of the inferential methods are extensively tested using simulated data where models up to 50 dimensions and 688 parameters are fit and studied. The performance of our model, in relation to various multivariate GARCH models, is also evaluated using a real data set of weekly returns on a set of 10 international stock indices. We consider the performance along two dimensions: the ability to correctly estimate the conditional covariance matrix of future returns and the unconditional and

*Corresponding author.
E-mail address: neil.shephard@nuf.ox.ac.uk (N. Shephard).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2005.06.026

conditional coverage of the 5% and 1% value-at-risk (VaR) measures of four pre-defined portfolios.

Keywords: Bayesian inference; Markov chain Monte Carlo; Marginal likelihood; Metropolis-Hastings algorithm; Particle filter; Simulation; State space model; Stochastic jumps; Student-*t* distribution; Volatility; Value-at-Risk; Risk measurement; Forecasting; Time-varying correlations

1. Introduction

Two classes of models, ARCH and stochastic volatility (SV), have emerged as the dominant approaches for modeling financial volatility (Bollerslev et al., 1994; Ghysels et al., 1996). For the most part, the literature has dealt with univariate processes despite the need for multivariate models in areas such as asset pricing, portfolio analysis, and risk management. Although some multivariate models of volatility have been proposed, inference is restricted to specifications involving only a few variables, largely because of the proliferation of parameters in high dimensions. As a consequence, multivariate models of time-varying volatility and correlations have found limited applications to problems of practical relevance in finance. A major aim of this paper is to overcome these difficulties and demonstrate a unified Bayesian fitting and inference framework for truly high dimensional multivariate SV models.

In previous work within the ARCH tradition, multivariate models of volatility have been discussed by Bollerslev et al. (1988), Diebold and Nerlove (1989), Engle et al. (1990) and King et al. (1994). Unfortunately, these generalizations are parameter rich and difficult to estimate due to complicated constraints on the parameter space. More tractable versions of multivariate ARCH models (Bollerslev et al., 1994, pp. 3002–3010) are not generally capable of modeling the complexities of the data (e.g. Bollerslev, 1990 assumes that the conditional correlations amongst the series are constant over time). Engle and Sheppard (2001) have tried to overcome this problem but only two parameters index the time-varying multivariate correlation matrix.

The model proposed and considered here is a generalization of the univariate stochastic volatility model. It combines features of the classical factor model with those of the heavy tailed univariate stochastic volatility model. Models along these lines have been discussed by Harvey et al. (1994), Jacquier et al. (1995), Kim et al. (1998), Pitt and Shephard (1999b), and Aguilar and West (2000) but the models in these papers are rather special (for example, not all model heavy tailed errors and none include jumps). In addition, the estimation approaches developed in these papers (which simulate the posterior distribution by Markov chain Monte Carlo (MCMC) methods) are not scalable in the dimension of the model. This is primarily due to two aspects of the previous algorithms. The first is related to the one-at-a time sampling of the latent volatilities, which is known to produce poor mixing even in univariate models (Kim et al., 1998). The second is related to the sampling of the latent factors and the associated parameters in the loading matrix. Specifically, given the remaining unknowns in the model, the loading matrix is sampled conditioned on the factors, which is followed by the sampling of the latent factors conditioned on the loading matrix. This feature also tends to produce poor mixing.

In this paper, we propose a general multivariate SV model with heavy tailed errors and jumps. We develop a unified analysis of the model, and its special cases, that encompasses estimation, filtering and model choice. The centerpieces of our MCMC based estimation algorithm are: (1) a joint sampling of the latent SV processes and (2) the sampling of the free elements of the loading matrix marginalized over the factors. The performance and effectiveness of our estimation method is tested in a large-scale study where models up to 50 dimensions and 688 parameters are fit and studied. The methods are shown to be scalable in terms of series and factors and simulation-efficient. Multivariate volatility models of this size and complexity have never before been estimated successfully.

The rest of the paper is organized as follows. In Section 2, we present the model and our approach for simulating the posterior distribution by MCMC methods. The problem of model comparisons is taken up in Section 3 where a method for estimating the log-likelihood function and the filtered values of the time-varying volatilities and correlations is also provided. In Section 4, we provide a detailed simulation study of the performance of our estimation and model choice procedures. In Section 5 we apply our approach to international equity market data. We assess the forecasting performance of the MSV specification in relation to alternative models including multivariate GARCH models. We consider the performance along two dimensions: the ability to correctly estimate the conditional covariance matrix of future returns and the unconditional and conditional coverage of the 5% and 1% value-at-risk (VaR) measures of four pre-defined portfolios. We conclude with some brief remarks in Section 6.

2. Model and estimation

2.1. Model

We begin by specifying a new and flexible multivariate SV model that permits both series-specific jumps at each time, and student-t innovations with unknown degrees of freedom. Let $y_t = (y_{1t}, \ldots, y_{pt})'$ denote the p observations at time t ($t \leqslant n$) and suppose that conditioned on k unobserved factors $f_t = (f_{1t}, \ldots, f_{kt})'$ and p independent Bernoulli "jump" random variables q_t, we have

$$y_t = Bf_t + K_t q_t + u_t, \tag{1}$$

where B is a matrix of unknown parameters (subject to the identifying restrictions $b_{ij} = 0$ for $j > i$ and $b_{ii} = 1$ for $i \leqslant k$), $K_t = diag\{k_{1t}, \ldots, k_{pt}\}$ are the jump sizes, and u_t is a vector of innovations. Assume that each element q_{jt} of q_t takes the value one with probability κ_j and the value zero with probability $1 - \kappa_j$, and that each element u_{jt} of u_t follows an independent student-t distribution with degrees of freedom $v_j > 2$, which we express in hierarchical form as

$$u_{jt} = \lambda_{jt}^{-1/2}\varepsilon_{jt}, \quad \lambda_{jt} \overset{\text{i.i.d.}}{\sim} gamma\left(\frac{v_j}{2}, \frac{v_j}{2}\right), \quad t = 1, 2, \ldots, n, \tag{2}$$

where

$$\begin{pmatrix} \varepsilon_t \\ f_t \end{pmatrix} | V_t, D_t, K_t, q_t \sim \mathrm{N}_{p+k}\left\{0, \begin{pmatrix} V_t & 0 \\ 0 & D_t \end{pmatrix}\right\}$$

are conditionally independent Gaussian random vectors. The time-varying variance matrices V_t and D_t are taken to depend upon unobserved random variables (log-volatilities) $h_t = (h_{1t}, \ldots, h_{p+k,t})$ in the form

$$V_t = V_t(h_t) = diag\{\exp(h_{1t}), \ldots, \exp(h_{pt})\} : p \times p,$$

$$D_t = D_t(h_t) = diag\{\exp(h_{p+1,t}), \ldots, \exp(h_{p+k,t})\} : k \times k, \tag{3}$$

where each h_{jt} follows an independent three-parameter $(\mu_j, \phi_j, \sigma_j)$ stochastic volatility process

$$h_{jt} - \mu_j = \phi_j(h_{jt-1} - \mu_j) + \sigma_j\eta_{jt}, \eta_{jt} \overset{\text{i.i.d.}}{\sim} \mathrm{N}(0, 1). \tag{4}$$

Our model specification is completed by assuming that the variables $\zeta_{jt} = \ln(1 + k_{jt})$, $j \leqslant p$, are distributed as $\mathrm{N}(-0.5\delta_j^2, \delta_j^2)$, where $\delta = (\delta_1, \ldots, \delta_p)$ are unknown parameters. This assumption is similar to that made by Andersen et al. (2002) in a different context and models the belief that the expected value of k_{jt} is zero.

To understand the size of this model in terms of parameters and latent variables, let β denote the free elements of B after imposing the identifying restrictions. Then there are $pk - (k^2 + k)/2$ elements in β, $3(p + k)$ parameters $\theta_j = (\phi_j, \mu_j, \sigma_j^2)$, $j \leqslant p$, in the autoregressive process of $\{h_{jt}\}$, p degrees of freedom $v = (v_1, \ldots, v_p)$, p jump intensities $\kappa = (\kappa_1, \ldots, \kappa_p)$, and p jump variances $\delta = (\delta_1, \ldots, \delta_p)$. If we let $\psi = (\beta, \theta_1, \ldots, \theta_{p+k}, v, \delta, \kappa)$ denote the entire list of parameters, then the dimension of ψ is 688 when $p = 50$ and $k = 8$, as in one of our models below. Furthermore, the model contains $n(p + k)$ latent volatilities $\{h_t\}$ that appear non-linearly in the specification of V_t and D_t, $2np$ latent variables $\{q_t\}$ and $\{k_t\}$ associated with the jump component, and np scaling variables $\{\lambda_t\}$.

In the sequel, we refer to our model as the multivariate stochastic volatility jump model with student-t errors, or MSVJt for short. We use the acronyms MSVt to denote the model without jumps, MSVJ to denote the model with jumps and Gaussian errors, and MSV to denote the model with no jumps and Gaussian errors. We compare and contrast all four models in our empirical exercises.

2.2. Preliminaries

If we let $\mathscr{F}_{t-1}$ denote the history of the $\{y_t\}$ process up to time $t - 1$, and $p(h_t, \lambda_t, K_t, q_t | \mathscr{F}_{t-1}, \psi)$ the density of the latent variables $(h_t, \lambda_t, K_t, q_t)$ conditioned on

$(\mathscr{F}_{t-1}, \psi)$, then the likelihood function of ψ given the data $y = (y_1, \ldots, y_n)$ is

$$\begin{aligned} p(y|\psi) &= \prod_{t=1}^{n} \int p(y_t|h_t, \lambda_t, K_t, q_t, B) p(h_t, \lambda_t, K_t, q_t|\mathscr{F}_{t-1}, \psi)\, \mathrm{d}h_t\, \mathrm{d}\lambda_t\, \mathrm{d}K_t\, \mathrm{d}q_t \\ &= \prod_{t=1}^{n} \int \mathrm{N}_p(y_t|K_t q_t, \Omega_t) p(h_t, \lambda_t, K_t, q_t|\mathscr{F}_{t-1}, \psi)\, \mathrm{d}h_t\, \mathrm{d}\lambda_t\, \mathrm{d}K_t\, \mathrm{d}q_t, \end{aligned} \tag{5}$$

where $\mathrm{N}_p(\cdot|\cdot,\cdot)$ is the multivariate normal density function, $K_t q_t$ is the mean of y_t marginalized over f_t,

$$V_t^* = V_t(h_t) \odot diag(\lambda_{1t}^{-1}, \ldots, \lambda_{pt}^{-1}) \quad \text{and} \quad \Omega_t = V_t^* + BD_t(h_t)B'$$

is the marginalized variance of y_t depending on the latent variables h_t, λ_t and on the loading matrix B, and the symbol $\odot$ denotes element-by-element multiplication. It is not difficult to see that neither $p(h_t, \lambda_t, K_t, q_t|\mathscr{F}_{t-1}, \psi)$ nor the integral of $\mathrm{N}_p(y_t|K_t q_t, \Omega_t)$ over $(h_t, \lambda_t, K_t, q_t)$ are available in closed form.

We utilize MCMC methods to develop a practical Bayesian estimation approach for this model; Chib and Greenberg (1996) and Chib (2001) provide extensive reviews of these methods. In the MCMC approach, the posterior distribution is sampled by simulation methods and the draws generated from the simulation are used to summarize the posterior distribution. The simulation is conducted by devising, and simulating, the transition density of an irreducible, aperiodic Markov chain whose invariant distribution is the target posterior distribution. In order to implement this approach, one basic idea is to avoid the direct use of the likelihood function (which is, of course, rather complicated and difficult to compute) and to focus on the posterior distribution of the parameters and the latent variables

$$\pi(\beta, \{f_t\}, \{\theta_j\}, \{h_{j\cdot}\}, \{v_j\}, \{\lambda_{j\cdot}\}, \{\delta_j\}, \{\kappa_j\}, \{\zeta_{j\cdot}\}, \{q_{j\cdot}\}|y), \tag{6}$$

where the notation $z_{j\cdot}$ is used to denote the collection $(z_{j1}, \ldots, z_{jn})$. This distribution is quite high-dimensional but as we show in the rest of the paper it can be sampled efficiently by MCMC methods provided the Markov chain is carefully constructed. Efficiency in this context refers to the serial correlations in the sampled output and is measured, for each parameter in turn, by the inefficiency factor which is, intuitively speaking, one plus twice the sum of all the serial correlations.

One issue of particular importance is the type of ''blocking'' that is used. A single block MCMC algorithm proceeds by moving ψ in one simultaneous move from the current point to the next point in the chain. Since this is infeasible given the high-dimension of ψ, the Markov chain is constructed by a divide and conquer strategy wherein blocks of parameters are updated conditioned on the values of the remaining blocks; a single, complete transition of the Markov chain occurs when all the blocks have been thus revised. By carefully managing the blocking structure we show that efficiency of the simulation scheme can be increased by orders of magnitude; without use of our refinements, the inefficiency factors often exceed 1000, and with our refinement they range between 20 and 50, even in our highest dimensional model. The practical ramifications of this are significant. If in one case one needed perhaps quarter-million samples from the posterior distribution, in the other 10,000 would suffice. For models as large as we are interested in fitting, the value of this improvement cannot be overstated.

2.3. Proposed MCMC algorithm

One key step in our algorithm is the sampling of β and the factors $\{f_t\}$ in one block, conditioned on $(y, \{h_{j.}\}, \{\lambda_{j.}\}, \{\zeta_{j.}\}, \{q_{j.}\})$. Because β and f_t appear in product form, the obvious approach of sampling β conditioned on $\{f_t\}$ and then sampling $\{f_t\}$ conditioned on β is less effective, which we demonstrate in Section 4. The next step of the algorithm is also interesting because given $(y, B, \{f_t\}, \{\lambda_{j.}\}, \{\zeta_{j.}\}, \{q_{j.}\})$, and the conditional independence of the errors in (3), the model can be devolved into $(p+k)$ conditionally Gaussian state space models. This implies that the log-volatilities and series specific parameters can be sampled series-by-series. This is one reason that our approach is scalable in both p and k.

Sampling of β: To begin, consider then the sampling of β from the density

$$\begin{aligned}\pi(\beta|y, \{h_{j.}\}, \{\zeta_{j.}\}, \{q_{j.}\}, \{\lambda_{j.}\}) &\propto p(\beta)\prod_{t=1}^{n} p(y_t|B, h_t, \zeta_t, q_t, \lambda_t)\\ &\propto p(\beta)\prod_{t=1}^{n} \mathrm{N}_p(y_t|K_t q_t, \Omega_t),\end{aligned}$$

where $p(\beta)$ is the normal prior density defined above. To sample this density, which is typically quite high-dimensional, we use the Metropolis-Hastings (M-H) algorithm (Chib and Greenberg, 1995). We follow Chib and Greenberg (1994) and take the proposal density to be multivariate-t, $T(\beta|m, \Sigma, v)$, where m is the approximate mode of $l = \log\{\prod_{t=1}^{n} \mathrm{N}_p(y_t|K_t q_t, \Omega_t)\}$, and Σ is minus the inverse of the second derivative matrix of l; the degrees of freedom v is set arbitrarily at 15. If we let the ijth free element of B be denoted by b_{ij} and define $\widetilde{y}_t = y_t - K_t q_t$, we have that

$$\begin{aligned}l = \sum_{t=1}^{n} \log\, \phi_p(y_t|K_t q_t, \Omega_t) &= const - \frac{1}{2}\sum_{t=1}^{n}\log|\Omega_t|\\ &- \frac{1}{2}\sum_{t=1}^{n}(y_t - K_t q_t)'\Omega_t^{-1}(y_t - K_t q_t)\end{aligned}$$

and

$$\begin{aligned}\frac{\partial l}{\partial b_{ij}} &= \frac{1}{2}\sum_{t=1}^{n}\left\{\widetilde{y}_t{}'\Omega_t^{-1}\frac{\partial \Omega_t}{\partial b_{ij}}\Omega_t^{-1}\widetilde{y}_t - tr\left(\Omega_t^{-1}\frac{\partial \Omega_t}{\partial b_{ij}}\right)\right\}\\ &= \sum_{t=1}^{n}\left\{s_t'\frac{\partial B}{\partial b_{ij}}D_t B' s_t - tr\left(\mathrm{E}_t\frac{\partial B'}{\partial b_{ij}}\right)\right\},\end{aligned}$$

where $s_t = \Omega_t^{-1}\widetilde{y}_t$, $\mathrm{E}_t = \Omega_t^{-1}BD_t$, and $\Omega_t^{-1} = (V_t^*)^{-1} - (V_t^*)^{-1}B\{D_t^{-1} + B'(V_t^*)^{-1}B\}^{-1}B'(V_t^*)^{-1}$. With these derivatives, (m, Σ) can be found by a sequence of Newton–Raphson iterations. Then the M-H step for sampling β is implemented by drawing a value β^* from the multivariate-t distribution, namely $T(m, \Sigma, v)$, and accepting the proposal value with probability

$$\alpha(\beta, \beta^*|\tilde{y}, \{h_{j.}\}, \{\lambda_{j.}\}) = \min\left\{1, \frac{p(\beta^*)\prod_{t=1}^{n} \mathrm{N}_p(\tilde{y}_t|0, V_t^* + B^* D_t B^{*\prime})}{p(\beta)\prod_{t=1}^{n} \mathrm{N}_p(\tilde{y}_t|0, V_t^* + BD_t B')} \frac{T(\beta|m, \Sigma, v)}{T(\beta^*|m, \Sigma, v)}\right\}, \tag{7}$$

where β is the current value. If the proposal value is rejected, the next item of the chain is taken to be the current value β.

Sampling of $\{f_t\}$: The joint sampling of β and the factors is completed by sampling $\{f_t\}$ from the distribution $\{f_t\}|\tilde{y}, B, h, \lambda$. This step is simple because the latter distribution breaks up into the product of the distributions $f_t|\tilde{y}_t, h_t, \lambda_t, B$. By standard Bayesian calculations, one can derive that the latter distribution is Gaussian with mean $\hat{f}_t = F_t B'(V_t^*)^{-1}\tilde{y}_t$ and variance $F_t = (B'(V_t^*)^{-1}B + D_t^{-1})^{-1}$.

Sampling of $\{\theta_j\}$ *and* $\{h_{j.}\}$: In the next step of the algorithm, given $(y, B, \{f_t\}, \{\lambda_{j.}\}, \{\zeta_{j.}\}, \{q_{j.}\})$, and the conditional independence of the errors in (3), we exploit the fact that the model separates into $(p+k)$ conditionally Gaussian state space models. Let $\alpha_t = Bf_t$, a p vector with components α_{jt}, and let

$$z_{jt} = \begin{cases} \ln(y_{jt} - \alpha_{jt} - (\exp(\zeta_{jt}) - 1)q_{jt} + c)^2 + \ln(\lambda_{jt}), & j \leqslant p. \\ \ln(f_{j-p,t}^2), & j \geqslant p+1, \end{cases}$$

where c is an "offset" constant that is set to 10^{-6}. Then from Kim et al. (1998) it follows that the $p+k$ state space models can be subjected to an independent analysis for sampling the $\{\theta_j\}$ and $\{h_{j.}\}$. In particular, the distribution of z_{jt}, which is h_{jt} plus a log chi-squared random variable with one degree of freedom, may be approximated closely by a seven component mixture of normal distributions, allowing us to express the MSVJt model as

$$\begin{aligned} &z_{jt}|s_{jt}, h_{jt} \sim \mathrm{N}(h_{jt} + m_{s_{tj}}, v_{s_{tj}}^2), \\ &h_{jt} - \mu_j = \phi_j(h_{jt-1} - \mu_j) + \sigma_j \eta_{jt}, \quad j \leqslant p+k, \end{aligned} \tag{8}$$

where s_{jt} is a discrete component indicator variable with mass function $\Pr(s_{jt} = i) = q_i$, $i \leqslant 7$, $t \leqslant n$, and $m_{s_{tj}}$, $v_{s_{tj}}^2$ and q_i are parameters that are reported in Chib et al. (2002). Thus, under this representation, conditioned on the transformed observations we have that

$$p(\{s_{j.}\}, \theta, \{h_{j.}\}|z) = \prod_{j=1}^{k+p} p(s_{j.}, \theta_j, h_{j.}|z_{j.}),$$

which implies that the mixture indicators, log-volatilities and series specific parameters can be sampled series by series.

Now, for each j, one can sample $(s_{j.}, \theta_j, h_{j.})$ by the univariate SV algorithm given by Chib et al. (2002). Briefly, $s_{j.}$ is sampled straightforwardly from

$$p(s_{j.}|z_{j.}, h_{j.}) = \prod_{t=1}^{n} p(s_{jt}|z_{jt}, h_{jt}),$$

where $p(s_{jt}|z_{jt}, h_{jt}) \propto p(s_{jt})\phi(z_{jt}|h_{jt} + m_{s_{tj}}, v^2_{s_{tj}})$ is a mass function with seven points of support. Next, θ_j is sampled by the M-H algorithm from the density $\pi(\theta_j|z_{j.}, s_{j.}) \propto p(\theta_j)p(z_{j.}|s_{j.}, \theta_j)$ where

$$p(z_{j.}|s_{j.}, \theta_j) = p(z_{j1}|s_{j.}, \theta_j)\prod_{t=2}^{T} p(z_{jt}|\mathscr{F}^*_{jt-1}, s_{j.}, \theta_j) \tag{9}$$

and $p(z_{jt}|\mathscr{F}^*_{j,t-1}, s_{j.}, \theta_j)$ is a normal density whose parameters are obtained by the Kalman filter recursions, adapted to the differing components, as indicated by the component vector $s_{j.}$ Finally, $h_{j.}$ is sampled from $[h_{j.}|z_{j.}, s_{j.}, \theta_j]$ by the simulation smoother algorithm of de Jong and Shephard (1995).

Sampling of $\{v_j\}$, $\{q_{j.}\}$ and $\{\lambda_{j.}\}$: In the remaining steps, the degrees of freedom parameters, jump parameters and associated latent variables are sampled independently for each time series. Significant improvements in simulation efficiency are achieved by sampling v_j marginalized over $\lambda_{j.}$ from the multinational distribution

$$\Pr(v_j|y_{j.}, h_{j.}, B, f, q_{j.}, \zeta_{j.}) \propto \Pr(v_j)\prod_{t=1}^{n} T(y_{jt}|\alpha_{jt} + \{\exp(\zeta_{jt}) - 1\}q_{jt}, \exp(h_{jt}), v_j). \tag{10}$$

Next, the jump indicators $\{q_{j.}\}$ are sampled from the two-point discrete distribution

$$\Pr(q_{jt} = 1|\, y_{j.}, h_{j.}, B, f, v_j, \zeta_{j.}, \kappa_j) \propto \kappa_j T(y_{jt}|\alpha_{jt} + \{\exp(\zeta_{jt}) - 1\}, \exp(h_{jt}), v_j),$$

$$\Pr(q_{jt} = 0|\, y_{j.}, h_{j.}, B, f, v_j, \zeta_{j.}, \kappa_j) \propto (1 - \kappa_j) T(y_{jt}|\alpha_{jt}, \exp(h_{jt}), v_j),$$

followed by the components of the vector $\lambda_{j.}$ from the density

$$\lambda_{jt}|y_{jt}, h_{jt}, B, f, v_j, q_{jt}, \psi_{jt} \sim gamma\left(\frac{v_j + 1}{2}, \frac{v_j + (y_{jt} - \alpha_{jt} - (\exp(\zeta_{jt} - 1)q_{jt}))^2}{2\exp(h_{jt})}\right).$$

Sampling of $\{\delta_j\}$ and $\{\zeta_{j.}\}$: Finally, we sample the parameters δ_j and $\zeta_{j.}$. For simulation efficiency reasons, these two parameters must also be sampled in one block. To see how this is possible, note that if k_{jt} is small, as is true in financial applications with high frequency returns that are measured in decimals, $\exp(\zeta_{jt})$ may be closely approximated by $1 + \zeta_{jt}$, implying that $k_{jt}q_{jt}$ equals $\zeta_{jt}q_{jt}$ and ζ_{jt} can be marginalized out. This permits the sampling of δ_j from the density

$$\pi(\delta_j)\prod_{t=1}^{n} \mathrm{N}(\alpha_{jt} - 0.5\delta_j^2 q_{jt}, \delta_j^2 q_{jt}^2 + \exp(h_{jt})\lambda_{jt}^{-1}) \tag{11}$$

by the M-H algorithm. Once δ_j is sampled, the vectors $\zeta_{j.}$ are sampled, bearing in mind that their posterior distribution is updated only when q_{jt} is one. Therefore, when q_{jt} is zero, we sample ζ_{jt} from $\mathrm{N}(-0.5\delta_j^2, \delta_j^2)$, otherwise we sample from the distribution $\mathrm{N}(\Psi_{jt}(-0.5 + \exp(-h_{jt})\lambda_{jt}y_{jt}), \Psi_{jt})$, where $\Psi_{jt} = (\delta_j^{-2} + \exp(-h_{jt})\lambda_{jt})^{-1}$. The algorithm is completed by sampling the components of the vector κ independently from $\kappa_j|q_{j.} \sim beta(u_{0j} + n_{1j}, u_{1j} + n_{0j})$, where n_{0j} is the count of $q_{jt} = 0$ and $n_{1j} = n - n_{0j}$ is the count of $q_{jt} = 1$.

A complete cycle through these various distributions completes one transition of our Markov chain. These steps are then repeated G times, where G is a large number, and the values beyond a suitable burn-in of say a 1000 cycles, are used for the purpose of summarizing the posterior distribution.

3. Model comparison

In this section, we show how the MSVJt model can be compared with alternative multivariate and univariate specifications. We do this comparison based on the marginal likelihood of each model and the associated Bayes factors (ratios of marginal likelihoods). Because of the dimensions involved, computation of the marginal likelihood presents several challenges. Nonetheless, our study of the problem has revealed that the method of Chib (1995) and Chib and Jeliazkov (2001) is feasible in this context, and quite effective in picking the true model, as we demonstrate in our simulation exercises.

The starting point of the Chib method is the basic marginal likelihood identity under which the log of the Bayes factor for comparing non-nested models $\mathcal{M}_1$ to $\mathcal{M}_2$ can be written as

$$\begin{aligned}
&\log p(y|\mathcal{M}_1) - \log p(y|\mathcal{M}_2) \\
&\quad = \log p(y|\mathcal{M}_1, \psi_1^*) + \log p(\psi_1^*|\mathcal{M}_1) - \log \pi(\psi_1^*|\mathcal{M}_1, y) \\
&\qquad - \{\log p(y|\mathcal{M}_2, \psi_2^*) + \log p(\psi_2^*|\mathcal{M}_2) - \log \pi(\psi_2^*|\mathcal{M}_2, y)\}, \qquad (12)
\end{aligned}$$

where $p(y|\mathcal{M}_j, \psi_j^*)$ is the likelihood function under $\mathcal{M}_j$, $p(\psi_j^*|\mathcal{M}_j)$ and $\pi(\psi_j^*|\mathcal{M}_j, y)$ are the corresponding prior and posterior densities, evaluated at some specified point ψ_j^*, say the posterior mean. The next step is to estimate the likelihood and posterior ordinates by some efficient method.

3.1. Posterior ordinate

To estimate the posterior ordinate we use a marginal/conditional decomposition and the output of the original and subsequent "reduced MCMC runs". To explain this technique, let

$$\begin{aligned}
\pi(\psi^*|\mathcal{M}, y) &= \pi(\beta^*, v^*, \theta^*, \delta^*, \kappa^*|\mathcal{M}, y) \\
&= \pi(\beta^*|\mathcal{M}, y)\pi(v^*|\mathcal{M}, y, \beta^*)\pi(\theta^*|\mathcal{M}, y, \beta^*, v^*) \\
&\quad \times \pi(\delta^*|\mathcal{M}, y, \beta^*, \theta^*, v^*)\pi(\kappa^*|\mathcal{M}, y, \beta^*, \theta^*, v^*, \delta^*)
\end{aligned}$$

and consider the estimation of each of the terms starting in the second line of this decomposition. It turns out that for the sample sizes in our applications, the marginal posterior densities of the factor loadings are very concentrated around the mean and close to normal. For simplicity, therefore, we approximate $\pi(\beta^*|\mathcal{M}, y)$ by the ordinate of a normal density with mean vector and covariance matrix obtained from the full MCMC run.

Second, to estimate the p-dimensional conditional ordinate $\pi(v^*|\mathcal{M}, y, \beta^*)$ we fix β at β^* and continue our MCMC simulation for another G iterations. Within this run, the ordinates of the conditional mass function $\Pr(v_j|y_{j\cdot}, h_{j\cdot}, B^*, f, q_{j\cdot}, \zeta_{j\cdot})$ are averaged and the resulting modal probability is taken as the estimate of $\pi(v^*|\mathcal{M}, y, \beta^*)$.

Third, to estimate the conditional ordinate $\pi(\theta_1^*, \ldots, \theta_{p+k}^*|y, \mathcal{M}, \beta^*, v^*)$ we group the θ_j's in groups of two (each of dimension six) and produce output from the appropriate reduced MCMC runs to estimate the resulting ordinates. Specifically, to estimate $\pi(\theta_1^*, \theta_2^*|y, \mathcal{M}, \beta^*, v^*)$ we fix β at β^*, v at v^* and run the MCMC algorithm given above. The desired ordinate is then estimated by the kernel smoothing method applied to the output on θ_1 and θ_2 from this run. The process is repeated in sequence, in each case with additional parameters held fixed.

Next, we estimate the ordinate $\pi(\delta^*|\mathcal{M}, y, \beta^*, v^*, \theta^*)$ by applying a result given in Chib and Jeliazkov (2001). Specifically, it can be shown that the ordinate

$$\pi(\delta^*|\mathcal{M}, y, \beta^*, v^*, \theta^*) = \int \prod_{j=1}^{p} \pi(\delta_j^*|\mathcal{M}, y_{j\cdot}, \beta^*, v^*, \theta^*, h_{j\cdot}, f, q_{j\cdot}, \lambda_{j\cdot}) \times \mathrm{d}\pi(\{h_{j\cdot}\}, f, \{q_{j\cdot}\}, \{\lambda_{j\cdot}\}|\mathcal{M}, y, \beta^*, v^*, \theta^*),$$

where π denotes generically the distribution of the enclosed random vectors, can be expressed as

$$\frac{\mathrm{E}_1 \prod_{j=1}^{p} \alpha(\delta_j, \delta_j^*|\mathcal{M}, y_{j\cdot}, \beta^*, v^*, \theta^*, h_{j\cdot}, f, q_{j\cdot}, \lambda_{j\cdot}) q(\delta_j^*|\mathcal{M}, y_{j\cdot}, \beta^*, v^*, \theta^*, h_{j\cdot}, f, q_{j\cdot}, \lambda_{j\cdot})}{\mathrm{E}_2 \prod_{j=1}^{p} \alpha(\delta_j^*, \delta_j|\mathcal{M}, y_{j\cdot}, \beta^*, v^*, \theta^*, h_{j\cdot}, f, q_{j\cdot}, \lambda_{j\cdot})}, \tag{13}$$

where α is the probability of move in the M-H step for δ_j, q is the student-t proposal density in that step, E_1 is the expectation with respect to $\pi(\{h_{j\cdot}\}, f, \{q_{j\cdot}\}, \{\lambda_{j\cdot}\}| \mathcal{M}, y, \beta^*, v^*, \theta^*)$ and E_2 is the expectation with respect to

$$\pi(\{h_{j\cdot}\}, f, \{q_{j\cdot}\}, \{\lambda_{j\cdot}\}|\mathcal{M}, y, \beta^*, v^*, \theta^*, \delta^*) \prod_{j=1}^{p} q(\delta_j|\mathcal{M}, y, \beta^*, v^*, \theta^*, \delta^*, h_{j\cdot}, f, q_{j\cdot}, \lambda_{j\cdot}).$$

The first of these expectations can be computed from the output of a reduced MCMC run in which β, v, and θ are fixed at their starred values. The second expectation can be computed from the output of an additional reduced run in which δ is also fixed; for each draw of $\{h_{j\cdot}\}, f, \{q_{j\cdot}\}, \{\lambda_{j\cdot}\}$ in this reduced run, δ_j is drawn from the proposal density and these combined draws are used to average the probability of move in the denominator of (13).

Finally, to estimate the κ^* conditional ordinate, the parameters $(\beta, v, \theta, \delta)$ are fixed and the quantities $\{q_t, \kappa\}$ are drawn in a reduced MCMC run. The required ordinate then follows by averaging the beta density of κ.

3.2. Filtering and likelihood evaluation

We now discuss a simulation-based approach, called the auxiliary particle filtering method (see Pitt and Shephard (1999a) and the book length review of Doucet et al.

(2001)), to estimate the likelihood ordinate $\log f(y_1, \ldots, y_n|\mathcal{M}, \psi^*) = \sum_{t=1}^{n} \log f(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$, where

$$f(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) = \int \mathrm{N}_p(y_t|K_t q_t, \Omega_t) p(\lambda_t, K_t, q_t|\mathcal{M}, \psi^*) p(h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) \times \mathrm{d}h_t \, \mathrm{d}\lambda_t \, \mathrm{d}K_t \, \mathrm{d}q_t$$

is the one-step-ahead predictive density of y_t,

$$p(h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) = \int p(h_t|\mathcal{M}, h_{t-1}, \psi^*) p(h_{t-1}|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) \, \mathrm{d}h_{t-1}$$

is the one-step-ahead predictive density of h_t, $p(h_t|\mathcal{M}, h_{t-1}, \psi^*) = \prod_{j=1}^{p+k} \mathrm{N}(h_{tj}|\mu_j^* + \phi_j^*(h_{j,t-1} - \mu_j^*), \sigma^2)$ is the product of the Markov transition densities and $p(h_{t-1}|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$ is the posterior distribution of h_{t-1} given $\mathcal{F}_{t-1}$ (the filtered distribution).

We now use a sequential Monte Carlo filtering procedure to efficiently estimate the one-step-ahead predictive density of y_t given above. In this procedure, introduced for stochastic volatility models by Kim et al. (1998), samples (particles) from the preceding filtered distribution (e.g., $p(h_{t-1}|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$) are propagated forward to produce samples from the subsequent filtered distribution (namely, $p(h_t|\mathcal{M}, \mathcal{F}_t, \psi^*)$). Suppose then that we have a sample $h_{t-1}^{(g)}$ ($g \leqslant M$) from the filtered distribution $h_{t-1}|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*$. Based on this sample, we can approximate the one-step-ahead predictive density of h_t as

$$p(h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) \simeq \frac{1}{M} \sum_{g=1}^{M} p(h_t|\mathcal{M}, h_{t-1}^{(g)}, \psi^*).$$

Under this approximation, the posterior density of the latent variables at time t is available as

$$\begin{aligned} &p(\lambda_t, K_t, q_t, h_t|\mathcal{M}, \mathcal{F}_t, \psi^*) \\ &\quad \propto \mathrm{N}_p(y_t|K_t q_t, \Omega_t) p(\lambda_t, K_t, q_t|\mathcal{M}, \psi^*) p(h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) \\ &\quad \dot{\propto} \mathrm{N}_p(y_t|K_t q_t, \Omega_t) p(\lambda_t, K_t, q_t|\psi^*) \frac{1}{M} \sum_{g=1}^{M} f(h_t|\mathcal{M}, h_{t-1}^{(g)}, \psi^*) \end{aligned} \tag{14}$$

and the objective is to sample this density. This sampling is carried out as follows. In the first stage, proposal values $h_t^{*(1)}, \ldots, h_t^{*(R)}$ are created. These values are then resampled to produce the draws $\{h_t^{(1)}, \ldots, h_t^{(M)}\}$ that correspond to draws from (14). We have found that R should be five or ten times larger than M to ensure efficient propagation of the particles.

Auxiliary particle filter for multivariate SV model

1. Given values $\{h_{t-1}^{(1)}, \ldots, h_{t-1}^{(M)}\}$ from $(h_{t-1}|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$ calculate $\hat{h}_t^{*(g)} = \mathrm{E}(h_t^{(g)}|h_{t-1}^{(g)})$ and

$$w_g = \mathrm{N}_p(y_t|0, \Omega_t(\hat{h}_t^{*(g)}, 1, B^*), \psi^*), \quad g = 1, \ldots, M$$

and sample R times the integers $1, 2, \ldots, M$ with probability $\overline{w}_t^g = w_g / \sum_{j=1}^M w_j$. Let the sampled indexes be $k_1, \ldots, k_R$ and associate these with $\hat{h}_t^{*(k_1)}, \ldots, \hat{h}_t^{*(k_R)}$.

2. For each value of k_g from Step 1, simulate the values $\{h_t^{*(1)}, \ldots, h_t^{*(R)}\}$ from

$$h_{j,t}^{*(g)} = \mu_j^* + \phi_j^*(h_{j,t-1}^{(k_g)} - \mu_j^*) + \sigma_j^* \eta_{j,t}^{(g)}, \quad g = 1, \ldots, R,$$

where $\eta_{j,t}^{(g)} \sim \mathrm{N}(0,1)$. Likewise draw $\lambda_t^{(g)}$, $K_t^{(g)}$, $q_t^{(g)}$ from their prior $p(\lambda_t, K_t, q_t|\psi^*)$, where $K_t^{(g)} = diag\left\{k_{1t}^{(g)}, \ldots, k_{pt}^{(g)}\right\}$ and $\zeta_{jt}^{(g)} = \ln(1 + k_{jt}^{(g)})$ is drawn from $\mathrm{N}(-0.5\delta_j^{*2}, \delta_j^{*2})$.

3. Resample the values $\{h_t^{*(1)}, \ldots, h_t^{*(R)}\}$ M times with replacement using probabilities proportional to

$$w_g^* = \frac{\mathrm{N}_p(y_t|K_t^{(g)} q_t^{(g)}, \Omega_t(h_t^{*(g)}, \lambda_t^{(g)}, B^*))}{\mathrm{N}_p(y_t|0, \Omega_t(\hat{h}_t^{*(k_g)}, 1, B^*))}, \quad g = 1, \ldots, R$$

to produce the desired filtered sample $\{h_t^{(1)}, \ldots, h_t^{(M)}\}$ from $(h_t|\mathcal{M}, \mathcal{F}_t, \psi^*)$.

As discussed by Pitt (2001), the weights produced in the above algorithm provide a simulation-consistent estimate of the likelihood contribution. In particular,

$$\widehat{f}(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) = \left(\frac{1}{M}\sum_{g=1}^M w_g\right)\left(\frac{1}{R}\sum_{g=1}^R w_g^*\right),$$

which can be shown to converge to $f(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$ in probability as M and R go to infinity. These estimates are obtained for each t and combined to produce our estimate of the likelihood ordinate $\log f(y_1, \ldots, y_n|\mathcal{M}, \psi^*)$.

3.2.1. Forecasting

The output from a particle filter can also be used to perform forecasting. That is we can estimate by simulation the density or moments of, for example,

$$\sum_{j=1}^s y_{t+j}|\mathcal{M}, \mathcal{F}_t, \psi^*,$$

the s-step ahead return vector or

$$\frac{1}{s}\sum_{j=1}^s \Omega^*(h_{t+1}, \ldots, h_{t+s})|\mathcal{M}, \mathcal{F}_t, \psi^*,$$

where

$$\Omega^*(h_{t+1}, \ldots, h_{t+s}) = \sum_{j=1}^s V_{t+j}^* + B^* D_{t+j}(h_{t+j}) B^{*\prime}$$

is the associated average covariance over that period of length s. Recall that V_{t+j}^* depends upon λ_{t+j}. Given the output $\{h_t^1, \ldots, h_t^M\}$ from the particle filter, we can simply propagate each of these particles forward using their autoregressive structure (4) to get samples from $f(h_{t+1},\ldots,h_{t+s}|\mathcal{M},\mathcal{F}_t,\psi^*)$. These can be used in a variety of ways, for example noting that

$$\sum_{j=1}^{s} y_{t+j}|\mathcal{M},\mathcal{F}_t,\theta^*,\{h_{t+1}\},\{\lambda_t,K_t,q_t\}\sim \mathrm{N}\left\{\sum_{j=1}^{s} K_{t+j}q_{t+j},\Omega^*(h_{t+1},\ldots,h_{t+s})\right\}$$

yields an easy way of using a Rao–Blackwell argument to estimate the higher order moments of returns over s periods.

3.2.2. Filtered correlations

In many financial decisions it is important to know the correlations between the returns on holding various risky assets. Here we briefly describe the time series evolution of the correlations for the above series, estimating the correlations using contemporaneous information. We could derive the correlations through the filtered time varying covariance matrix

$$\begin{aligned}\Omega^*_{t|t-1} &= \mathrm{E}\{\Omega(h_t,\lambda_t)+K_tq_tq_t'K_t|\mathcal{M},\mathcal{F}_{t-1}\}\\ &= \mathrm{E}(V_t|\mathcal{M},\mathcal{F}_{t-1})+B\{\mathrm{E}(D_t|\mathcal{M},\mathcal{F}_{t-1})\}B'+\mathrm{E}\{K_tq_tq_t'K_t|\mathcal{M}\}.\end{aligned}$$

However, that would give a biased estimator of the correlation. Instead we work with

$$R_{t|t-1} = \mathrm{E}\{\Psi(h_t,\lambda_t,K_t,q_t)|\mathcal{M},\mathcal{F}_{t-1}\},$$

where

$$\begin{aligned}\Psi(h_t,\lambda_t,K_t,q_t) &= diag\{\Omega^*(h_t,\lambda_t,K_t,q_t)\}^{-1/2}\Omega^*(h_t,\lambda_t,K_t,q_t)\\ &\quad\times diag\{\Omega^*(h_t,\lambda_t,K_t,q_t)\}^{-1/2}\end{aligned}$$

is the conditional correlation matrix for $y_t|h_t,\lambda_t,K_t,q_t$. By construction $R_{t|t-1}$ is the minimum mean square error estimator of the correlation. This can be easily estimated by

$$\frac{1}{M}\sum_{g=1}^{M}\Psi(h_t^{(g)},\lambda_t^{(g)},K_t^{(g)},q_t^{(g)}),$$

where $h_t^{(1)},\ldots,h_t^{(M)}$ is the output from the particle filter from the density $f(h_t|\mathcal{M},\mathcal{F}_{t-1},\psi^*)$ and $\{\lambda_t^{(g)},K_t^{(g)},q_t^{(g)}\}$ are i.i.d. draws from $p(\lambda_t|\mathcal{M},\psi^*)$.

4. Simulation study

We now provide evidence on the effectiveness of our methods when applied to models of up to 50 dimensions. We report on the simulation efficiency of the fitting method, the estimation accuracy, robustness to changes in the prior, and on the reliability of the model selection method.

4.1. Calibration

In simulating the data we aim to replicate the dynamics of commonly observed financial time series. In particular, our simulation designs mimic the features of daily equity returns (measured in decimals). For the calibration, we first fit the MSVJt model with 8 factors to daily returns on 40 major stocks listed on the New York Stock Exchange. The specifics of this data set and detailed estimation results were part of a previous version of the paper and can be obtained from the authors. Second, for each parameter we compute the average and the standard deviation of its posterior mean across stocks. We then generate one set of model parameters from distributions whose mean closely matches the average of the posterior means. For instance, the average posterior mean for μ_j across 40 stocks is -9.017 and the standard deviation of the posterior means is 0.429. We generate the 40 μ_js from a normal distribution with mean equal to 9 and standard deviation equal to 1. The details for the other parameters appear in the following subsections. Finally, given a set of model parameters we simulate the data series.

4.2. Prior distribution

In the experiments we assume that the parameters are a priori mutually independent. To select the hyperparameters we use the equity data as a training sample: for each parameter we set the prior mean close to the average of the posterior means across stocks and the prior standard deviation close to an integer multiple of the cross-sectional standard deviation of the posterior means. The exception is the prior for v which is set to a uniform discrete distribution covering the range of the estimated vs. This leads to the following prior distributions that will be used throughout, unless otherwise noted. Free elements of $B : b_{ij} \sim \mathrm{N}(1, 9)$; $\mu : \mu_i \sim \mathrm{N}(-9, 25)$; $\phi : \phi_j^* \sim beta(a, b)$, where $\phi_j = 2\phi_j^* - 1$, so that the prior mean of ϕ_j is 0.86 and standard deviation is 0.11; $\sigma : \sigma_j \sim IG(c/2, d/2)$ with mean of 0.25 and standard deviation of 0.4; $v : v_j$ is discrete uniform over the grid $(5, 8, 11, 14, 17, 20, 30, 60)$; $\log(\delta) : \log(\delta_j) \sim \mathrm{N}(-3.07, 0.148)$, implying a mean of 0.05 and standard deviation of 0.02 on δ_j; and $\kappa : \kappa_j \sim beta(2, 100)$. Since δ affects directly the variability of the jump size, under the selected prior daily jumps in returns are expected to lie within the $\pm 10\%$ range. The hyperparameters in the prior of κ imply a mean jump probability of 1.96% per observation and a standard deviation of 1.36%. This translates into jumps that are expected to occur about 50 observations apart (four or five jumps per year with daily data).

4.3. Starting values

Our algorithm in Section 2.3 is initialized with values for the following random variables:

$$\{h_{j.}\}_{j=1}^{p+k},\ \{\psi_{j.}\}_{j=1}^{p},\ \{q_{j.}\}_{j=1}^{p},\ \{\lambda_{j.}\}_{j=1}^{p}.$$

We set the $\{\psi_{j.}\}$ and $\{q_{j.}\}$ vectors equal to zero (the no jumps case), the $\{\lambda_{j.}\}$ equal to one (the conditionally Gaussian errors case) and the latent volatilities to -10.

4.4. Simulation efficiency

A key feature of our estimation method is the sampling of B marginalized over the factors. Whereas it is simpler to condition on the factors, as done by Geweke and Zhou (1996), Pitt and Shephard (1999b), Aguilar and West (2000) and Jacquier et al. (1995) in the context of static and dynamic factor models, the sampled output is far less well behaved. To show this, we generate eight data sets, labeled D1–D8, from different models and with different number of assets, factors and time series observations, and evaluate the alternative samplers in terms of the realized inefficiency factors. The inefficiency factor is the inverse of the numerical efficiency measure in Geweke (1992) and is computed from the MCMC output as the square of the numerical standard error divided by the variance of the posterior estimate under (hypothetical) i.i.d. sampling.

We draw the parameters of the models from the following distributions: the free elements of b_{ij} are from $\mathrm{N}(0.9, 1)$; μ_j from $\mathrm{N}(-9, 1)$; ϕ_j from a scaled beta with mean 0.95 and variance 0.03, σ_j, v_j, $\log \delta_j$ and κ_j from their prior distributions. The specifics of each data set are shown in Table 1. It should be noted that the models are quite high-dimensional; the smallest has 142 parameters and the largest has 688.

For each data set, we employ the marginalized sampling procedure and two other methods where the elements of B are sampled either by column or by row, conditioned on the factors. For the algorithm proposed in this paper we run the MCMC sampler for 11,000 iterations, collecting the last 10,000 for inferential purposes. For the other two methods, expecting a drop in simulation efficiency, we collect 50,000 draws after discarding the first 5000. We compare the three methods, as they relate to the sampling of B, in terms of the relative inefficiency factors (the ratio of inefficiency factors). As can be seen from Table 2, in models with four factors (D1–D6) our procedure is between 20 and 40 times more efficient than the other two methods. In models with eight factors (D7 and D8), our method is about 80 times more efficient. Furthermore, the efficiency of our method does not erode as the dimensionality and complexity of the model is increased whereas the other methods become even less efficient. The performance gains from sampling B in the

Table 1
Features of simulated data sets

Data set	Model	p	k	n	Parms	Data set	Model	p	k	n	Parms
D1	MSV	20	4	2000	142	D2	MSV	50	4	2000	352
D3	MSV	20	4	1000	142	D4	MSV	50	4	1000	352
D5	MSV	20	4	5000	142	D6	MSV	50	4	5000	352
D7	MSV	40	8	2000	428	D8	MSVJt	50	8	2000	688

Parms denotes the number of parameters.

Table 2
Summary output for inefficiency factors

Sampling B	Mean	S.D	Low	Upp	Max	Min	Mean	S.D	Low	Upp	Max	Min
	D1						D2					
Row/Marg	34.5	17.3	24.6	46.0	71.0	1.8	29.7	17.0	17.9	39.4	91.4	6.2
Col/Marg	37.9	21.1	22.1	50.3	83.9	1.3	33.4	19.5	18.8	45.4	106	7.4
Col/Row	0.8	0.6	0.5	1.7	2.3	0.2	1.2	0.6	0.8	1.5	3.8	0.5
	D3						D4					
Row/Marg	36.4	29.5	14.1	46.0	113	4.0	41.7	23.7	26.8	51.2	132	2.9
Col/Marg	27.5	15.1	16.5	41.4	59.3	3.6	15.9	9.8	7.6	20.4	45.7	2.0
Col/Row	0.9	0.3	0.7	1.2	1.7	0.5	0.4	0.2	0.3	0.6	1.1	0.1
	D5						D6					
Row/Marg	24.0	27.3	4.6	31.8	167	2.3	88.0	49.3	42.7	131	185	11.8
Col/Marg	14.8	16.2	4.0	18.3	101	1.9	62.3	45.5	26.4	91.0	231	2.3
Col/Row	0.7	0.2	0.5	0.9	1.8	0.3	0.9	1.1	0.4	1.0	9.0	0.1
	D7						D8					
Row/Marg	62.3	36.0	33.4	87.1	161	9.6	76.9	54.7	25.9	119	279	3.3
Col/Marg	89.7	54.9	44.4	126	238	6.1	84.6	56.2	29.2	128	294	5.1
Col/Row	1.5	0.8	1.0	1.9	6.5	0.3	1.3	0.4	1.0	1.7	2.4	0.3

The table summarizes the distribution of relative inefficiency factors for the estimated factor loadings. Row denotes sampling by row, Col sampling by column and Marg sampling marginalized over the factors. Results are reported for different simulated data sets and for alternative sampling schemes for the factor loading matrix B. Low denotes the 25th percentile, Upp denotes the 75th percentile.

way we suggest are worth the computational burden because substantially smaller Monte Carlo samples are needed to achieve a given level of numerical accuracy. On average, our procedure is 5–6 times slower in terms of CPU time per MCMC iteration than the alternative non-marginalized methods. In Section 4.7 below we provide additional details on the computational burden.

We next consider the specifics of our MCMC scheme as they relate to the sampling of v and δ. We generate an additional data set, D9, from the MSVJt model with 50 series, 4 factors and 2000 observations per series and we employ our method along with several alternatives where one or more of the reduced blocking steps in the generation of B, v and δ are switched off. Efficiency factors from these runs are reported in Table 3. Two patterns are noticeable. First, the reduced blocking scheme leads to much better mixing for both v and δ. On average, our proposed method is 40–50 times more efficient than the alternatives. Second, these performance gains are realized even when B is sampled conditioned on the factors.

4.5. Parameter estimates and factor extraction

In this section, we first show the ability of the proposed algorithm to correctly estimate the large number of parameters and latent variables in the model. Second, we assess the robustness of the algorithm to changes in the prior. We contrast the

Table 3
Summary output for inefficiency factors

Sampling	Mean	S.D	Low	Upp	Mean	S.D	Low	Upp	Mean	S.D	Low	Upp
	B				δ				v			
s1/s2	108.4	49.6	66.5	147.8	1.0	0.3	0.9	1.2	1.1	0.3	0.9	1.3
s3/s1	1.0	0.2	1.0	1.0	54.8	27.2	37.1	65.0	41.1	13.2	24.1	49.9
s3/s2	106.4	49.6	63.3	139.1	54.2	25.2	34.5	64.4	43.1	13.0	26.7	51.3

The table summarizes the distribution of relative inefficiency factors for the estimated factor loadings (B), degrees of freedom parameters (v) and jump intensity parameters (δ). Results are reported for a data set of 50 series and 2000 observations per series and for alternative sampling schemes for B, v and δ. Specifically, s1: B non-marginalized, v marginalized, δ marginalized. s2: all marginalized. s3: all non-marginalized. Low denotes the 25th percentile, Upp denotes the 75th percentile.

Table 4
Summary output for simulated data

Sampling	B	μ_a	ϕ_a	σ_a	δ	κ	μ_f	ϕ_f	σ_f	v
*B*marg Prior1	.97	.99	.92	.92	.95	.92	.98	.91	.86	.82
*B*bycol Prior1	.84	.99	.90	.88	.95	.92	.88	.77	.38	.82
*B*marg Prior2	.95	.99	.90	.88	.98	.92	.95	.94	.84	.82
*B*bycol Prior2	.84	.99	.90	.87	.98	.92	.88	.77	.39	.81
*B*marg Prior3	.98	.95	.98	.96	.87	.35	.98	.99	.99	.80
*B*marg Prior4	.98	.97	.98	.97	.96	.30	.99	.99	.99	.83
*B*marg Prior5	.99	.94	.98	.96	.96	.12	.98	.99	.99	.83

Entries are the correlation coefficients between the true parameter values and MCMC estimates. The latter are the average of posterior means across 40 samples with $n = 1250$. Bmarg denotes the sampling of B marginalized over the latent factors, Bbycol denotes the sampling of B conditioning on the factors and done by column. Prior1, Prior2, Prior3, Prior4 and Prior5 are defined in the main text.

results from our proposed method with those where B is sampled by columns, conditioned on the factors.

In these experiments, the artificial data sets are generated from the MSVJt model with 40 series and eight factors. Each simulated series has 1250 observations, equivalent to about 5 years of daily data. We use the same mechanism described in the previous section to generate one set of true parameters. From these parameter values we then generate a total of 40 data sets and we fit the 8 factor MSVJt model to each of them. Due to the differences in the simulation efficiency, the preferred MCMC algorithm is run for 10,000 iterations while the non-marginalized MCMC algorithm is run for 100,000 iterations. We initially use the same priors reported in Section 4.2, defined collectively as Prior1.

Table 4 contains correlations between the true values and the parameter estimates for the alternative procedures and priors. The estimates are obtained as the grand averages of the posterior means across simulated samples.

Consider first the estimates for the factor loading matrix, which in this case has 284 free parameters. The correlation between the true values and the grand averages across samples is substantially higher for the more efficient procedure: 97.28% vs. 83.88%. The bar graph in Fig. 1 shows that the proposed approach yields accurate estimates of the B matrix (elements for only four factors are plotted). Second, the estimates of the volatility parameters for the factors (not reported) are noticeably more accurate for the preferred algorithm. Third, the estimates of the parameters in the volatility evolution equations are also less accurate from the non-reduced blocking scheme. The log-volatility levels, denoted by μ_a, are closely identified by both procedures; somewhat larger deviations are recorded for the ϕs and the σs: however, the correlations of the estimates with the true values are quite high, of the order of 90%. Next, consider the jump parameters, δ and κ. Without providing a graph we mention that the average of the posterior means across the different data sets are slightly closer to the true values for δ (correlation = 95%) than κ (correlation of 92%). In both cases the standard deviations across samples are quite small compared to their respective means. For the jump parameters we do not find meaningful differences across sampling schemes. The performance of both algorithms is relatively less satisfactory for the degrees of freedom parameters of the student-t distributions. The correlation with the true values is only 82%. This could be due to the large overall dimension of the parameter space combined with a relatively limited sample size used in the estimation.

Finally, consider the relationship between the true and estimated factors. Fig. 2 displays the correlations across samples for the common factors: the estimates for these latent variables are obtained by averaging across the MCMC draws for each

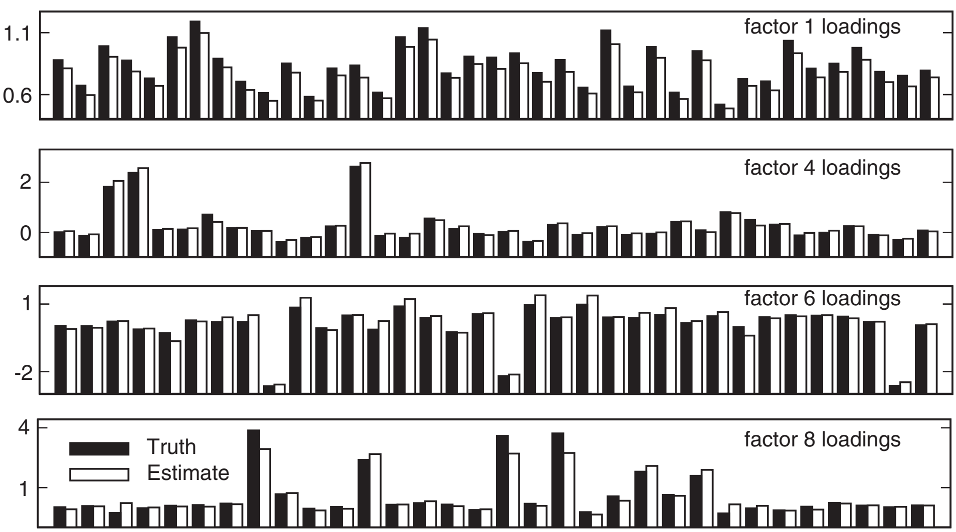

Fig. 1. True values vs. posterior estimates for the factor loadings. Each panel displays the loadings on a different factor (only factor 1, 4, 6 and 8 are reported). The posterior quantities are the average of posterior means across 40 samples with $n = 1250$.

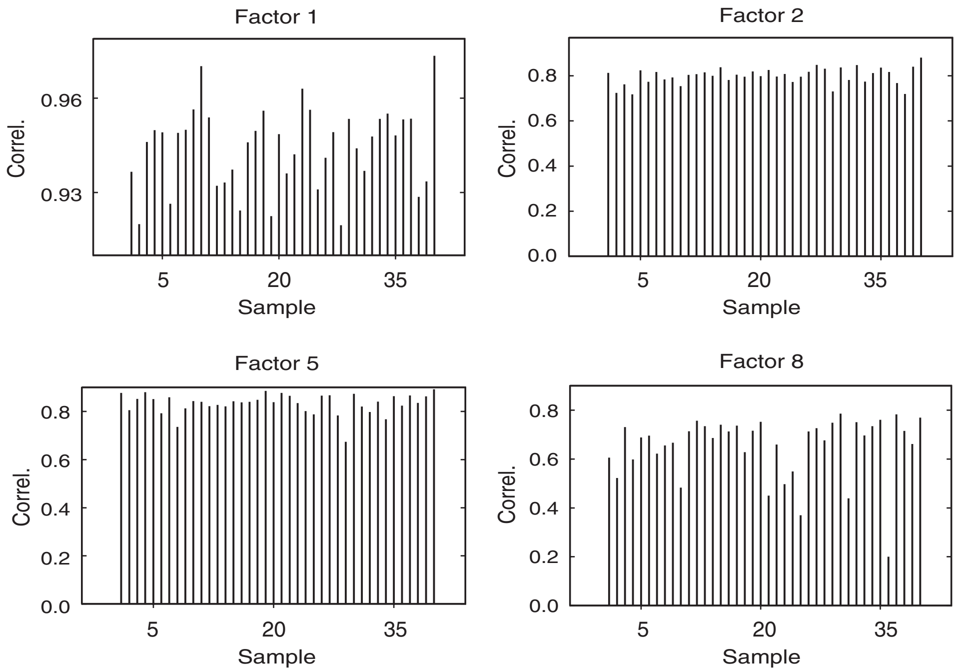

Fig. 2. Correlations between true and estimated factors across simulated samples. For each data set the estimates are obtained by averaging the draws of the MCMC sampler. The results are based on 40 simulated data sets of size 1250 each.

sample. We report the summaries for factor 1, 2, 5 and 8. In all cases the latent series are estimated well with correlations with the true values ranging between 70% and 95%. The precision is high for the first factor, decreasing somewhat for the other factors. These experiments show that the suggested estimation procedure yields reliable inferences for both the model parameters and the latent dynamic factors. Relying on the non-marginalized schemes to update the factor loadings leads to significant biases. These biases arise not only in the estimates of the loading parameters but also in those of the factor volatilities.

4.5.1. Alternative prior distributions

We now repeat the estimation with different choices of hyperparameters. We begin with a more diffuse independent N(0,1000) prior on b_{ij}. This prior is labeled Prior2. The posterior mean of the parameters under this prior are reported in the third and fourth row of Table 4. Both posterior sampling procedures appear to be robust to this change in the prior as the correlations between the true and simulated values are almost unaltered. It remains true, however, that the marginalized sampling scheme is better at estimating the factor loadings and the factor volatility parameters. Next, we investigate our ability to learn about the jump components of the model by varying the prior on the jump parameters δ and κ. The first prior, labeled Prior3, implies infrequent but large jumps whereas the second, labeled Prior4, assigns a higher probability to the occurrence of relatively more frequent but significantly smaller jumps. Specifically, in Prior3 we impose a mean of 10% and a standard deviation of 2% on δ_j and we let $\kappa_j \sim beta(2, 400)$. In this instance jumps are expected to occur only once or twice per year at the daily frequency. In Prior4, $\log(\delta_j) \sim \mathrm{N}(-4.023, 0.223)$, implying a mean of 0.02 and standard deviation of 0.01 on δ_j; and $\kappa_j \sim beta(9, 200)$. The latter choice implies that a jump is expected almost every month. For the third alternative prior, labeled Prior5, we choose a mean and a standard deviation of 10% on δ_j whereas $\kappa_j \sim beta(1, 1)$, a uniform density on $(0, 1)$.

The results from rerunning the estimation over the same 50 simulated data sets are shown in the last two rows of Table 4. We can make two broad conclusions. First, that when the sample is small, the posterior distribution of the jump intensity κ_j is affected by the prior but that the posterior distribution of δ_j is less affected. The experiments confirm the need, indicated by previous studies, for a long time series for robust estimation of the jump intensity. In unreported results we find that to reach a correlation between true values and posterior means in the 0.8–0.9 range one needs a sample size of approximately 7500 daily observations. Our second conclusion is that the prior of the jump parameters does not materially affect the posterior distribution of the remaining parameters.

4.6. Performance of the marginal likelihood criterion

In this section, we utilize simulated data to assess the performance of the marginal likelihood and Bayes factor criterion in identifying the correct model across model types and, within a given model class, the correct number of factors. In the simulation design, data sets are generated from the MSVt model with three factors. Each simulated data set contains 30 series of 2000 observations each. The model parameters in the true model are randomly generated as in Section 4.4. We generate a total of 50 data sets from the true model. The MSVJ, MSVt and MSVJt models are then fitted to these data sets, each with 2, 3 and 4 factors. Thus, nine models are each estimated 50 times under the prior distributions and hyperparameters reported in Section 4.2. The marginal likelihood of each model in each simulated data set is calculated from $G = 10,000$ MCMC iterations (beyond a burn-in of 1000 iterations) followed by reduced runs of 10,000 iterations. Finally, the two parameters of the particle filter algorithm, namely M and R, are set to 20,000 and 200,000, respectively.

4.6.1. Stability

First, we investigate the stability of the posterior ordinate estimate. We randomly pick 5 of our 50 simulated data sets and compute estimates of the posterior ordinate for various values of G, the number of reduced-run iterations. In particular, we let G take the values 5000, 10,000, 20,000 and 50,000. The posterior ordinates from each of the five data sets are then averaged. Although the data are generated from the MSVt model, we do this calculation with the MSVJt model which is a larger model. The estimated values are shown in Table 5.

Table 5
Natural log-posterior ordinate estimates for different simulation sizes. G denotes the number of reduced MCMC draws

Data	Simulation size G			
	5000	10,000	25,000	50,000
D2	329.74	329.73	329.80	329.97
D10	325.40	327.18	327.94	327.99
D30	319.11	323.35	323.87	323.71
D40	318.19	320.86	320.14	320.19
D50	346.97	348.87	348.40	348.92

Results are based on five simulated data sets.

Table 6
Frequency distribution (percentage) of Bayes factors across 50 simulated replications

	True model: MSVt 3f				
	1–3.2	3.2–10	10–100	>100	Total >10
MSVt3f/MSV2f	0	0	0	100	100
MSVt3f/MSV3f	0	0	0	100	100
MSVt3f/MSV4f	0	0	2	98	100
MSVt3f/MSVt2f	0	0	0	100	100
MSVt3f/MSVt4f	0	4	20	60	84
MSVt4f/MSVt2f	0	0	4	96	100
MSVt3f/MSVJt2f	0	0	0	100	100
MSVt3f/MSVJt3f	0	0	2	94	96
MSVt3f/MSVJt4f	0	0	10	78	88

The ranges for Bayes factor values correspond to the Jeffreys' scale.

The table values indicate that the estimates converge when the number of reduced runs is at least 10,000.

4.6.2. Model comparison

We conclude our experiments by examining the performance of the marginal likelihood criterion in selecting the true model. This is done via a sampling experiment in which we count the frequency with which each possible K-factor model ($K = 2, 3, 4$) is picked over the other models, based on the estimated marginal likelihoods. Table 6 reports the relevant results in which we compare the support for the true MSVt model with 3 factors against the other specifications.

According to the Jeffreys' scale, the evidence in favor of the true model is always decisive versus the basic MSV model as well as versus MSVt 2f and it is at least substantial against MSVt 4f in 84% of the cases. When compared to the more highly parameterized MSVJt model, MSVt 3f is still selected as the best model 100% of the times against MSVJt 2f, 98% of the times against MSVJt 3f and 88% of the times against MSVJt 4f. In all these cases the support in favor of the true model is strong or decisive. In summary, the simulation evidence provides a convincing validation of the Bayes factor criterion along two dimensions: the identification of the correct number of common factors and in the selection of the appropriate model specification.

4.7. Computational requirements

For the MSVJt model with 20 series and 4 factors fit to 2000 observations per series, our MCMC algorithm coded in C and running on a Linux 3.4 megahertz Pentium 4 computer consumes about 3 h of CPU time to generate 11,000 MCMC draws for the full MCMC run. The filtering algorithm, implemented with $M = 20,000$ and $R = 200,000$, requires about 5 h. The reduced run for computing the conditional posterior ordinate of θ is conducted with 5000 iterations for each pair of assets and factors. For the 12 pairs used in our benchmark example the required CPU time is slightly less than 4 h. The reduced run for δ needs 2.5 h for 5000 MCMC iterations whereas the reduced run for κ only requires 5–6 min for the same number of MCMC draws. The reported CPU times are essentially linear in the number of time series observations. The computing times required by the different portions of the procedure are affected differently, however, as one varies the number of series, p, included in the model. For instance, doubling the series from 20 to 40 more than triples the time for the full run and for the reduced run for θ, doubles the time for filtering, does not materially change the burden from the reduced runs for δ and for κ.

5. Application to equity returns

In this section, we apply our models to historical stock return data to compare the empirical performance of our specifications in relation to those from alternative models of time-varying covariances and correlations, including multivariate GARCH models. We consider the performance along two dimensions: the ability to correctly estimate the conditional covariance matrix of future returns and the unconditional and conditional coverage of the 5% and 1% Value at Risk (VaR) measures of four pre-defined portfolios.

5.1. Data

The data for the experiments are a set of international weekly stock index returns for the following 10 countries: Australia, France, Germany, Hong Kong, Italy, Japan, Singapore, Switzerland, United Kingdom and United States. Specifically, the data are the weekly Total Market Index series as provided by Datastream International. For each country, we compute the continuously compounded weekly returns (Wednesday to Wednesday) denominated in US dollars. We thus assume the perspective of a US-based investor who does not hedge currency risk. The sample covers the period from January 2, 1973 through September 30, 2003 for a total of 1605 observations. The mean is subtracted from each series. It should be noted that we deliberately restrict ourselves to 10 series because the multivariate GARCH models that we fit cannot be estimated for dimensions much higher than 10. In fact, even with 10 series, the estimation of the various multivariate GARCH models (done with FinMetrics, a commercial package) was quite difficult with the exception of the DCC model.

5.2. Competing models

As alternatives to the MSV models proposed above we consider six different specifications. These choices are motivated by the popularity of the models in the academic literature (see, for example, Ledoit et al., 2003; Lopez and Walter, 2001) as well as in industry practice. Denote the conditional covariance matrix at time t by Σ_t. The following specifications are used:

1. The BEKK GARCH(1,1) model

$$\Sigma_t = A_0 A_0' + A_1(u_{t-1}u_{t-1}')A_1' + B_1 \Sigma_{t-1} B_1',$$

where A_0 is a lower triangular matrix and all the parameter matrices are $p \times p$. In this model, first proposed by Engle and Kroner (1995), the conditional covariance matrix is positive semi-definite by construction.

2. The matrix–diagonal vector (MD-VECH) GARCH(1,1) model

$$\Sigma_t = A_0 + A_1 A_1' \odot (u_{t-1}u_{t-1}') + B_1 B_1' \odot \Sigma_{t-1}, \tag{15}$$

where A_0, A_1 and B_1 are all lower triangular matrices. This specification appears in Ding (1994) and Bollerslev et al. (1994). Although less general than the unrestricted VECH model of Bollerslev et al. (1988), this parametrization insures a positive semi-definite conditional covariance matrix, a property that does not hold for the more general VECH model.

3. The constant conditional correlation (CCC) GARCH(1,1) model. In this model, proposed by Bollerslev (1990), the time-varying covariance matrix is decomposed as follows:

$$\Sigma_t = \Delta_t R \Delta_t, \tag{16}$$

where R is a constant correlation matrix and Δ_t is a diagonal matrix containing the conditional standard deviations of the disturbances, each following a univariate GARCH(1,1) process.

4. The dynamic conditional correlation (DCC) GARCH(1,1) model. This extension of the CCC model was first presented by Engle (2002) and further analyzed by Engle and Sheppard (2001). In the DCC model the matrix R in (16) is allowed to be time-varying. We choose the parametrization

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1}, \tag{17}$$

where $Q_t = S(1-\alpha-\beta) + \alpha(\varepsilon_{t-1}\varepsilon'_{t-1}) + \beta Q_{t-1}$, S is the unconditional covariance of the standardized residuals obtained from fitting univariate GARCH(1,1) models to each return series, and Q_t^* is a diagonal matrix composed of the square root of the diagonal elements of Q_t.

5. The rolling window (RollWin) estimator. This is the multivariate equivalent of an historical volatility measure for univariate data. The matrix forecast for each time t is set equal to the sample covariance matrix conditional on information up to time $t-1$. We adopt a window of 104 observations, corresponding to 2 years of weekly returns, as this appears to be a fairly common choice in practice.
6. The exponentially weighted moving average (EWMA) model:

$$\Sigma_t = (1-\lambda)(u_{t-1}u'_{t-1}) + \lambda\Sigma_{t-1}.$$

This formulation is the result of applying exponentially decaying weights to lagged cross-products of residuals. The calibrated parameter λ is typically set to 0.94 for weekly data: this is the choice we adopt in this study as well. The approach is often used in risk measurement systems and its popularity is, at least partially, due to the fact that it is the method used and commercialized by RiskMetrics.

The MSV and multivariate GARCH models are estimated on a moving 20-year window (1095 weekly returns). For the MSV specification we consider models with 1–4 factors, with and without jumps and fat-tailed errors, and we select the best model in terms of in-sample fit by using Bayes factors. This model specification is then used for 1 year (52 weeks) in the forecasting exercise. At the end of each year we once again determine the best fitting MSV model. Notice that although the model specification is kept constant for a year, the posterior distribution of the parameters is updated every week. The multivariate GARCH models are estimated by quasi maximum likelihood assuming normal and student-t distributions for the error terms. The optimization of the quasi likelihood function is done through the built-in routines in the FinMetrics module of SPLUS for all specifications except the DCC model. For this model the estimation is carried out following the two-step procedure proposed by Engle (2002) and assuming conditionally normal errors. The forecasting experiments are started with the first week of 1994. Every model is re-estimated weekly and a new forecast is generated for the following 1, 2 and 4 weeks using the updated parameter estimates. The associated VaR measures are computed at the 1-week horizon. Overall, for each of the seven models specifications 509 forecasts at three different horizons and 509 1-week VaR measures are produced and evaluated.

5.3. Covariance matrix forecast

We compare the forecast from each model to a baseline variance covariance matrix that is taken as the true covariance matrix. The latter in unobservable but a proxy for it can be constructed as suggested by Andersen et al. (2001), Barndorff-Nielsen and Shephard (2002), Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) and implemented, among others, by Ledoit et al. (2003). In particular, the true covariance matrix for the period ending at week t, denoted by $\Sigma_{t,k}$, is computed as the cumulative cross-product matrix of daily return residuals over the forecast horizon of length k. The covariance forecast for the MSV models is computed as described in Section 3.2.1. For the GARCH models, given the parameter estimates, the k-step forecasts are obtained by recursion. For example, for the MD-VECH model it can be shown that

$$\mathrm{E}_t(\Sigma_{t+k}) = A_0 + (A_1 A_1^{'} + B_1 B_1^{'}) \odot \mathrm{E}_t(\Sigma_{t+k-1}),$$

where E_t indicates expectation conditioning on time t information. For the Roll Win and EWMA models, the multi-step ahead forecasts are computed by multiplying the 1-week forecast by the number of weeks. Denote by $\sigma_{i,j,t,k}$ the i,jth element of $\Sigma_{t,k}$ and by $\hat{\sigma}_{i,j,t,k}$ its forecast under any of the above models. Following Andersen et al. (2001) and Ledoit et al. (2003), we assess the predictive accuracy of a given model in terms of the root-mean-square error (RMSE) and mean absolute deviation (MAD) measures, which are defined as

$$RMSE_k = \left[\frac{1}{p^2}\sum_{i,j} \mathrm{E}(\hat{\sigma}_{i,j,t,k} - \sigma_{i,j,t,k})^2\right]^{1/2} \tag{18}$$

and

$$MAD_k = \frac{1}{p^2}\sum_{i,j} \mathrm{E}|\hat{\sigma}_{i,j,t,k} - \sigma_{i,j,t,k}|, \tag{19}$$

respectively.

5.4. *Value-at-risk*

It is not our objective to survey and rank the numerous approaches that have been proposed for the calculation of VaR. In the present study we focus instead on the relative performance of alternative methods within the variance–covariance approach.

Several different portfolios can be constructed from the same universe of p assets, corresponding to different trading positions. In this case it is common practice to compute the individual portfolios' VaR by using a single estimate of the $p \times p$ covariance matrix. This is the approach adopted, for example, by Ledoit et al. (2003). In the present study we consider the following four geographically identified portfolios:

- A World portfolio: all 10 countries, equally weighted.
- A European portfolio: France, Germany, Italy, Switzerland and United Kingdom, equally weighted.
- A Pacific Rim portfolio: Australia, Hong Kong, Japan and Singapore, equally weighted.
- A US only portfolio.

Denote by y_{pt} the realized return on a given portfolio of assets at time t. Given the vector of portfolio weights w, and the estimate of the conditional variance, $\Sigma_{t,k}$, the predicted portfolio variance is $\hat{\sigma}_{p,t,k} = w'\Sigma_{t,k}w$. The VaR at the 1% and 5% level is computed for each portfolio using the predicted portfolio variance as

$$VaR_{p,t-1,k}(\alpha) = \sqrt{\hat{\sigma}_{p,t,k}}F^{-1}(\alpha)$$

where $F^{-1}(\alpha)$ is the αth percentile of the cumulative one-step-ahead distribution assumed for portfolio returns. In addition to this commonly used method for calculating the VaR, our Bayesian approach offers a second and, potentially, more appealing alternative. Using the structure of the particle filter as shown in Section 3.2.1, one can draw directly from the predictive densities of the individual asset returns, compute the portfolio return distribution and then calculate the quantile (left-tail) of interest. For the MSV models we calculate the VaR in this way.

The accuracy of the VaR estimates is investigated using both unconditional and conditional coverage tests along the lines of Lopez and Walter (2001). Define the indicator variable I_t as

$$I_t = \begin{cases} 1 & \text{if } y_{p,t} < VaR_{p,t-1}, \\ 0 & \text{if } y_{p,t} \geqslant VaR_{p,t-1}. \end{cases}$$

The cases $I_t = 1$ are exceptions or hits. For the VaR estimates to be well behaved, the I_t series must exhibit both serial independence and correct coverage (i.e., its expected value must equal the nominal VaR level of α). To test correct coverage, let T denote the total number of out-of-sample observations for which VaR is computed, let γ denote the number of VaR exceptions over T observations, and let $\hat{\alpha}$ be the ratio γ/T. Then the hypothesis $\hat{\alpha} = \alpha$ is tested with the statistic

$$LR_{\text{uc}} = 2\{\log[\hat{\alpha}^{\gamma}(1-\hat{\alpha})^{T-\gamma}] - \log[\alpha^{\gamma}(1-\alpha)^{T-\gamma}]\},$$

which is distributed asymptotically as $\chi^2(1)$. In the presence of heteroschedasticity in portfolio returns, a test for correct coverage given the information set at a given point in time is likely to be a more relevant measure of VaR accuracy. Christoffersen (1998) derives a test of conditional coverage by jointly testing for correct unconditional coverage and independence in the hit rate series. The independence hypothesis is tested against the alternative of first-order Markov dependence. Define T_{ij} as the number of observations in state j after having been in state i in the previous period, $\pi_{01} = T_{01}/(T_{00} + T_{01})$ and $\pi_{11} = T_{11}/(T_{10} + T_{11})$. Under the alternative hypothesis the likelihood function is $L_A = (1-\pi_{01})^{T_{00}}\pi_{01}^{T_{01}}(1-\pi_{11})^{T_{10}}\pi_{11}^{T_{11}}$. Under the null of independence, the likelihood is instead $L_0 = (1-\pi)^{T_{00}+T_{10}}\pi^{T_{01}+T_{11}}$, where $\pi = (T_{01} + T_{00})/T$ and $\pi_{01} = \pi_{11} = \pi$. The test statistic for independence is

$$LR_{\text{ind}} = 2(\log L_A - \log L_0),$$

which is also distributed asymptotically as $\chi^2(1)$. A likelihood ratio statistic can be used to jointly test the two hypotheses and to test for correct conditional coverage. Specifically, the test is based on the statistic $LR_{\text{cc}} = LR_{\text{uc}} + LR_{\text{ind}}$, which is asymptotically distributed as $\chi^2(2)$.

5.5. Empirical results

Table 7 contains the results on the forecasting accuracy of the different models over three time horizons and two evaluation criteria. In terms of MAD, the performance of the MSV model is very satisfactory, especially at the shorter horizons, where MSV outperforms all other specifications. The next best performers are the multivariate GARCH models that allow for time-varying correlations. Among them, the DCC model yields the most accurate forecast at the 2 and 4 weeks horizon, whereas the BEKK model does better at the 1-week forecast. There is little material difference in the accuracy of the BEKK and the MD VECH model, with the MD VECH model being marginally better. The forecasting ability of these models is not improved by the inclusion of fat-tailed errors. The CCC model does always worse than the other GARCH models, which indicates the importance of modeling time-varying correlations. Finally, the simpler EWMA and Roll Win approaches are significantly outperformed, especially at the 1 and 2 weeks horizons.

In terms of the RMSE criteria, the distinctions among the various models is less clear. The MSV model still somewhat outperforms the other models at the short horizons, whereas at the longer horizons the simpler EWMA and RollWin models perform as well if not better. We may mention, however, that Ledoit et al. (2003) has shown that the MAD criterion may be superior to the RMSE since it tends to be less influenced by outliers.

Tables 8 and 9 report the summaries of the VaR estimation and testing results. In terms of VaR exceptions, MSV performs convincingly both at the 1% and at the 5% confidence level. Across portfolios its performance is uniformly never worse than any other model. In particular, for the World and for the US portfolio MSV provides an increased level of accuracy with respect to the other models. For the World portfolio, only the BEKK-*t* model compares favorably to the MSV model. For the World

Table 7
Evaluation of forecasting accuracy

Model	MAD			RMSE		
	1-week	2-week	4-week	1-week	2-week	4-week
MSV	3.45	5.57	11.90	7.76	13.15	24.72
BEKK	3.63	5.73	12.03	7.77	13.20	24.72
BEKK-*t*	3.70	5.67	11.88	7.81	13.13	24.58
MD VECH	3.67	5.68	11.93	7.80	13.10	24.54
MD VECH-*t*	3.67	5.68	11.93	7.80	13.10	24.55
CCC	3.93	6.70	13.10	7.98	13.45	24.80
CCC-*t*	3.98	6.74	13.01	7.99	13.41	24.70
DCC	3.72	5.60	11.75	7.77	12.85	24.34
EWMA	3.94	6.81	12.83	8.14	13.05	22.29
RollWin	4.02	7.00	12.23	8.30	13.32	22.30

Reported are the mean absolute deviation and root-mean-square-error from Eqs. (19) and (18), respectively. Model descriptions can be found in the main text.

Table 8
Evaluation of VaR estimates

Model	Hit rate		p-values					
			LR_{uc}		LR_{ind}		LR_{cc}	
	1%	5%	1%	5%	1%	5%	1%	5%
World portfolio								
MSV	0.010	0.049	0.35	0.22	0.53	0.27	0.33	0.01
MD VECH	0.016	0.061	0.23	0.27	0.61	0.15	0.43	0.19
MD VECH-t	0.008	0.065	0.61	0.14	0.80	0.02	0.85	0.02
BEKK	0.016	0.065	0.23	0.14	0.61	0.22	0.43	0.16
BEKK-t	0.010	0.059	0.97	0.37	0.75	0.12	0.95	0.20
CCC	0.024	0.070	0.13	0.28	0.19	0.01	0.04	0.02
CCC-t	0.019	0.066	0.11	0.21	0.10	0.01	0.08	0.03
DCC	0.014	0.059	0.42	0.37	0.66	0.12	0.66	0.20
EWMA	0.014	0.059	0.42	0.37	0.66	0.01	0.66	0.02
RollWin	0.022	0.069	0.02	0.07	0.02	0.03	0.00	0.02
European portfolio								
MSV	0.012	0.049	0.69	0.37	0.70	0.12	0.86	0.20
MD VECH	0.016	0.053	0.23	0.75	0.11	0.22	0.13	0.45
MD VECH-t	0.008	0.055	0.61	0.61	0.80	0.26	0.85	0.47
BEKK	0.014	0.047	0.42	0.77	0.66	0.12	0.66	0.28
BEKK-t	0.010	0.049	0.97	0.93	0.75	0.15	0.95	0.35
CCC	0.018	0.062	0.03	0.54	0.11	0.04	0.02	0.01
CCC-t	0.016	0.060	0.06	0.58	0.20	0.08	0.06	0.04
DCC	0.016	0.047	0.23	0.77	0.11	0.02	0.13	0.07
EWMA	0.014	0.047	0.42	0.77	0.66	0.02	0.66	0.07
RollWin	0.026	0.055	0.00	0.61	0.41	0.26	0.01	0.47

Figures refer to weekly VaR forecasts for different models at the 1% and 5% level. Reported are the average hit rate and the p-values for the unconditional coverage test (LR_{uc}), independence test (LR_{ind}), and conditional coverage test (LR_{cc}). Tests and model descriptions can be found in the main text.

portfolio the alternative specifications (with the exception of the EWMA model) tend to underestimate the VaR whereas for the US Portfolio they tend to be too conservative. For the Pacific Rim and European portfolios the differences in model performance becomes somewhat blurred. The CCC specification appears to perform the worst. In terms of the unconditional coverage test LR_{uc}, the performance of most models is adequate with the possible exception of the RollWin and CCC models and some of the GARCH models on the US portfolio. In sum, our MSV model performs favorably in relation to these models.

6. Conclusion

In this paper, we have proposed and analyzed a new multivariate model with time varying correlations. The model contains several features (for example fat tails and

Table 9
Evaluation of VaR estimates

Model	Hit rate		*p*-values					
			LR_{uc}		LR_{ind}		LR_{cc}	
	1%	5%	1%	5%	1%	5%	1%	5%
Pacific rim portfolio								
MSV	0.012	0.051	0.22	0.27	0.49	0.45	0.26	0.41
MD VECH	0.016	0.051	0.23	0.91	0.61	0.75	0.43	0.95
MD VECH-*t*	0.010	0.065	0.97	0.14	0.75	0.36	0.95	0.22
BEKK	0.018	0.057	0.12	0.48	0.57	0.56	0.25	0.66
BEKK-*t*	0.012	0.057	0.69	0.48	0.70	0.56	0.86	0.66
CCC	0.022	0.061	0.02	0.19	0.04	0.33	0.00	0.11
CCC-*t*	0.022	0.061	0.04	0.14	0.05	0.41	0.03	0.19
DCC	0.012	0.055	0.69	0.61	0.71	0.62	0.86	0.78
EWMA	0.012	0.059	0.69	0.37	0.70	0.86	0.86	0.66
RollWin	0.022	0.051	0.02	0.91	0.02	0.57	0.00	0.84
US portfolio								
MSV	0.011	0.053	0.42	0.75	0.66	0.64	0.66	0.85
MD VECH	0.010	0.031	0.97	0.04	0.75	0.52	0.95	0.10
MD VECH-*t*	0.008	0.035	0.61	0.11	0.80	0.66	0.85	0.25
BEKK	0.012	0.033	0.69	0.07	0.70	0.59	0.86	0.16
BEKK-*t*	0.010	0.031	0.97	0.04	0.75	0.52	0.95	0.10
CCC	0.018	0.040	0.41	0.04	0.39	0.30	0.33	0.03
CCC-*t*	0.016	0.039	0.48	0.03	0.28	0.23	0.25	0.04
DCC	0.010	0.045	0.97	0.61	0.75	0.38	0.95	0.60
EWMA	0.016	0.061	0.23	0.27	0.61	0.42	0.43	0.40
RollWin	0.014	0.055	0.42	0.61	0.66	0.26	0.66	0.47

Figures refer to weekly VaR forecasts for different models at the 1% and 5% level. Reported are the average hit rate and the *p*-values for the unconditional coverage test (LR_{uc}), independence test (LR_{ind}), and conditional coverage test (LR_{cc}). Tests and model descriptions can be found in the main text.

jump components) that are particularly relevant in the modeling of financial time series. Our fitting approach, which relies on tuned MCMC methods, was shown to be scalable in terms of both the multivariate dimension and the number of factors. This leads us to believe that this is first viable estimation approach for high-dimensional stochastic volatility models. In the paper we also provide a method for finding the marginal likelihood of the model. This criterion is useful in comparing the general model with various special cases, say defined by the presence or absence of jumps and fat-tails, and in identifying the correct number of pervasive factors. A large-scale simulation study shows that our estimation and inference procedures are both accurate and reliable. An extensive application to international equity return data shows that our proposed model performs favorably in relation to various alternative models in terms of the estimation of the predictive covariance and VaR.

Acknowledgements

The authors thank CINECA (Bologna, Italy) and Brick Network (Saint Louis, MO) for providing computing facilities. Neil Shephard's research is supported by the UK's ESRC through the grant ''High frequency financial econometrics based upon power variation''. The authors also thank the editor, associate editor and three anonymous referees for their helpful suggestions and comments.

References

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

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2001. The distribution of exchange rate volatility. Journal of the American Statistical Association 96, 42–55 (correction published in 2003, vol. 98, p. 501).

Andersen, T.G., Benzoni, L., Lund, J., 2002. An empirical investigation of continuous-time equity return models. Journal of Finance 57, 1239–1284.

Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P., 2003. Modeling and forecasting realized volatility. Econometrica 71, 579–625.

Barndorff-Nielsen, O.E., Shephard, N., 2002. Econometric analysis of realised volatility and its use in estimating stochastic volatility models. Journal of the Royal Statistical Society Series B 64, 253–280.

Barndorff-Nielsen, O.E., Shephard, N., 2004. Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. Econometrica 72, 885–925.

Bollerslev, T., 1990. Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. Review of Economics and Statistics 72, 498–505.

Bollerslev, T., Engle, R.F., Wooldridge, J.M., 1988. A capital asset pricing model with time varying covariances. Journal of Political Economy 96, 116–131.

Bollerslev, T., Engle, R.F., Nelson, D.B., 1994. ARCH models. In: Engle, R.F., McFadden, D. (Eds.), The Handbook of Econometrics, vol. 4. North-Holland, Amsterdam, pp. 2959–3038.

Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–21.

Chib, S., 2001. Markov chain Monte Carlo methods: computation and inference. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, Amsterdam, pp. 3569–3649.

Chib, S., Greenberg, E., 1994. Bayes inference for regression models with ARMA (p, q) errors. Journal of Econometrics 64, 183–206.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings algorithm. The American Statistican 49,327–35.

Chib., S., Greenberg, E., 1996. Markov chain Monte Carlo simulation methods in econometrics. Econometric Theory 12, 409–31.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis-Hastings output. Journal of the American Statistical Association 96, 270–281.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for generalized stochastic volatility models. Journal of Econometrics 108, 281–316.

Christoffersen, P., 1998. Evaluating interval forecasts. International Economic Review 39, 841–862.

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

Diebold, F.X., Nerlove, M., 1989. The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. Journal of Applied Econometrics 4, 1–21.

Ding, Z., 1994. Time series analysis of speculative returns. Unpublished Ph.D. Thesis, Department of Economics, University of California, San Diego.

Doucet, A., de Freitas, N., Gordon, N.J. (Eds.), 2001. Sequential Monte Carlo Methods in Practice. Springer, New York.

Engle, R.F., 2002. Dynamic conditional correlation—a simple class of multivariate garch models. Journal of Business and Economic Statistics 20, 339–350.

Engle, R.F., Kroner, K.F., 1995. Multivariate simultaneous generalized ARCH. Econometric Theory 11, 122–150.

Engle, R.F., Sheppard, K., 2001. Theoretical and empirical properties of dynamic conditional correlation multivariate GARCH. Unpublished paper, UCSD.

Engle, R.F., Ng, V.K., Rothschild, M., 1990. Asset pricing with a factor ARCH covariance structure: empirical estimates for treasury bills. Journal of Econometrics 45, 213–238.

Geweke, J., 1992. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. Computing Science and Statistics: Proceedings of the 23rd Symposium, 571–578.

Geweke, J.F., Zhou, G., 1996. Measuring the pricing error of the arbitrage pricing theory. Review of Financial Studies 9, 557–587.

Ghysels, E., Harvey, A.C., Renault, E., 1996. Stochastic volatility. In: Rao, C.R., Maddala, G.S. (Eds.), Statistical Methods in Finance. North-Holland, Amsterdam, pp. 119–191.

Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247–264.

Jacquier, E., Polson, N.G., Rossi, P.E., 1995. Models and prior distributions for multivariate stochastic volatility. Unpublished paper, GSB, University of Chicago.

Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.

King, M., Sentana, E., Wadhwani, S., 1994. Volatility and links between national stock markets. Econometrica 62, 901–933.

Ledoit, O., Santa-Clara, P., Wolf, M., 2003. Flexible multivariate garch modeling with an application to international stock markets. The Review of Economics and Statistics 85, 735–747.

Lopez, J.A., Walter, C.A., 2001. Evaluating covariance matrix forecasts in a value-at-risk framework. The Journal of Risk 3, 69–97.

Pitt, M.K., 2001. Smooth particle filters for likelihood maximisation. Unpublished paper, Department of Economics, Warwick University.

Pitt, M.K., Shephard, N., 1999a. Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association 94, 590–599.

Pitt, M.K., Shephard, N., 1999b. Time varying covariances: a factor stochastic volatility approach with discussion. In: Bernardo, J.M., Berger, J.O., Dawid, A.P., Smith, A.F.M. (Eds.), Bayesian Statistics, vol. 6. Oxford University Press, Oxford, pp. 547–570.