Analysis of High Dimensional Multivariate Stochastic Volatility Models

SIDDHARTHA CHIB
John M. Olin School of Business, Washington University, St Louis, MO 63130, USA

FEDERICO NARDARI
John M. Olin School of Business, Washington University, St Louis, MO 63130, USA

NEIL SHEPHARD
Nuffield College, Oxford OX1 1NF, UK

July 2, 1999

Abstract

This paper is concerned with the fitting and comparison of high dimensional multivariate time series models with time varying correlations. The models considered here combine features of the classical factor model with those of the univariate stochastic volatility model. Specifically, a set of unobserved time-dependent factors, along with an associated loading matrix, are used to model the contemporaneous correlation while, conditioned on the factors, the noise in each factor and each series is assumed to follow independent three-parameter univariate stochastic volatility processes. A complete analysis of these models, and its special cases, is developed that encompasses estimation, filtering and model choice. The centerpieces of our estimation algorithm (which relies on MCMC methods) is (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 sampling of the loading matrix (containing typically many hundreds of parameters) is done via a highly tuned Metropolis-Hastings step. The resulting algorithm is completely scalable in terms of series and factors and very simulation-efficient. We also provide methods for estimating the log-likelihood function and the filtered values of the time-varying volatilities and correlations. We pay special attention to the problem of comparing one version of the model with another and for determining the number of factors. For this purpose we use MCMC methods to find the marginal likelihood and associated Bayes factors of each fitted model. In sum, these procedures lead to the first unified and practical likelihood based analysis of truly high dimensional models of stochastic volatility. We apply our methods in detail to two datasets. The first is the return vector on 20 exchange rates against the US Dollar. The second is the return vector on 40 common stocks quoted on the New York Stock Exchange.

Keywords: Bayesian inference, Markov Chain Monte Carlo; Marginal likelihood; Metropolis-Hastings algorithm; Particle filter; Simulation; State space model; Volatility.

1 INTRODUCTION

The analysis of time series measured at relatively high frequencies occupies an important place in modern time series analysis. One primary emphasis of this work is on the proper modeling of the time dependencies that are present in the volatility of the time series which is generally observed in the squares and absolute values of the change in the series. Two classes of models, called the ARCH-GARCH and stochastic volatility (SV) models, have been developed to deal with this aspect of the time series. Reviews of the literature on this topic are given in Bollerslev, Engle, and Nelson (1994), Shephard (1996) and Ghysels, Harvey, and Renault (1996). For the most part, the volatility models deal with univariate time series despite the need for more general models in areas such as risk management and portfolio analysis. Unfortunately, attempts at generalizing the univariate models to the multivariate case, with a view to modeling the cross-covariance pattern in volatility, have been mostly incomplete or unsuccessful, ostensibly due to the proliferation of parameters in these multivariate generalizations. A major aim of this paper is to overcome these problems and demonstrate a unified fitting and inference framework for truly high dimensional models of stochastic volatility.

Previous work on multivariate models of volatility has been reported by Bollerslev, Engle, and Wooldridge (1988), Diebold and Nerlove (1989), Bollerslev (1990), Engle, Ng, and Rothschild (1990), Lin (1992), King, Sentana, and Wadhwani (1994), Engle and Kroner (1995), Braun, Nelson, and Sunier (1995) within the ARCH-GARCH tradition, and by Harvey, Ruiz, and Shephard (1994), Jacquier, Polson, and Rossi (1995), Pitt and Shephard (1999b) and Aguilar and West (1998) within the stochastic volatility framework. Unfortunately, the ARCH-GARCH generalizations are, for large problems, parameter rich and extremely difficult to estimate due to complicated constraints on the parameter space. More tractable versions of multivariate ARCH models (for example, see Bollerslev, Engle, and Nelson (1994, pp. 3002-10)) become so simplified that they are unlikely to satisfactorily approximate the complexities of the data. An example of this is the model by Bollerslev (1990) which assumes that the conditional correlations amongst the series is constant over time (see also Harvey, Ruiz, and Shephard (1994)). By contrast, the versions based on the stochastic volatility paradigm are more parsimonious but none of the papers to date have provided a fitting approach that is both efficient and scalable in the dimension of the problem.

In this paper we concentrate on a multivariate model of volatility that was recently proposed by Kim, Shephard, and Chib (1998) and subsequently studied by Pitt and Shephard (1999b) and Aguilar and West (1998). If we let $y_t = (y_{1t}, ..., y_{pt})$ denote the p vector of observations at time t $(t \leq n)$ then this model is characterized by k unobserved factors $f_t = (f_{1t}, ..., f_{kt})$ and a $p \times k$ loading matrix B, and univariate stochastic volatility processes for the *unobserved variances* of each of the p observation errors and k factors. Specifically, the observations and factors are modeled as being generated by the hierarchical process

$$y_t = Bf_t + \varepsilon_t\,, \qquad \begin{pmatrix} \varepsilon_t \\ f_t \end{pmatrix} |V_t, D_t \sim \mathcal{N}_{k+p}\left\{0, \begin{pmatrix} V_t & 0 \\ 0 & D_t \end{pmatrix}\right\}, \tag{1}$$

with the $\{\varepsilon_t, f_t, |D_t, V_t\}$ being a conditionally independent sequence. The time varying variance matrices V_t and D_t are modeled as

$$\begin{aligned} V_t &= V_t(h_t) = \mathrm{diag}\left\{\exp(h_{1t}), ..., \exp(h_{pt})\right\} : p \times p \\ D_t &= D_t(h_t) = \mathrm{diag}\left\{\exp(h_{N+1t}), ..., \exp(h_{p+k,t})\right\} : k \times k\,, \end{aligned} \tag{2}$$

in terms of *latent variables* $h_t = (h_{1t}, ..., h_{p+kt})$ that each follow an independent three-parameter $(\mu_j, \phi_j, \sigma_j)$ univariate process

$$h_{jt} - \mu_j = \phi_j(h_{jt-1} - \mu_j) + \sigma_j \eta_{jt}\,,\ \eta_{jt} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)\,. \tag{3}$$

The latent variables h_{jt} may be interpreted as latent time-varying *log volatilities.* After the imposition of identifying constraints on B, that $b_{ij} = 0$ for $j > i$ and $b_{ii} = 1$ for $i \leq k$, this model contains $pk - (k^2 + k)/2$ unknown elements in B and $3(p + k)$ unknown parameters in the stochastic volatility processes $\{\varepsilon_t, f_t\}$. Furthermore, the model contains $n(p+k)$ latent volatilities $\{h_t\}$ that appear non-linearly in the specification of V_t and D_t.

One purpose of this paper is to demonstrate a fitting procedure for this model when p and k are both large. In one instance for example we demonstrate the fitting in a problem where $p = 40$ and $k = 10$ leading to approximately five hundred parameters. It is difficult to overemphasize the challenge of fitting this model. First, the likelihood function of the model is intractable although, as we show below, it may be computed with some effort by an intensive simulation-based procedure. The resulting estimate of the likelihood, however, is not smooth in the parameters and cannot therefore be embedded within a gradient based optimization routine. Second, existing attempts to circumvent the computation of the likelihood function through Bayesian Markov chain Monte Carlo (MCMC) methods have not been entirely successful. For example, the MCMC approach developed in Pitt and Shephard (1999b) generates considerable serial correlation in the MCMC output *even when* the dimension of the parameter space is small. In addition, Aguilar and West (1998) develop a slightly different MCMC approach (based on work by Kim, Shephard, and Chib (1998)) for a special case of the model (one in which V_t is a constant diagonal matrix) but despite the simplified model structure and the small dimensions to which the approach is applied, the method still suffers from poor mixing. Despite this background, what we report in this paper is a new method, also based on MCMC simulations, that does not suffer from these problems and leads to a practical fitting strategy that is scalable in terms of p and k and requires little or no user tuning. The centerpieces of the estimation algorithm are (1) a reduced blocking scheme for sampling the free elements of the loading matrix B and the factors and (2) the use of the Chib, Nardari, and Shephard (1998) method for sampling the parameters of the univariate SV process. The sampling of the loading matrix B (containing typically many hundreds of parameters) is done via a highly tuned Metropolis-Hastings step that is based on a local approximation to the current density of the sample, marginalized over the factors. Local tuning of this kind appears initially in Chib and Greenberg (1994) but has not, to our knowledge, been attempted on dimensions as large as ours. We show that the autocorrelation times from the proposed algorithm in a model with about five hundred parameters are substantially less than those from the aforementioned algorithms applied to far simpler and smaller models.

Another inference problem that is taken up in this paper, that has not been addressed by Pitt and Shephard (1999b) and Aguilar and West (1998), relates to the determination of the number of factors k, and the comparison of the general model (which we call the MSV model) with various special cases that reflect other assumptions about the volatility structure of y_t. Because it is difficult to assess a priori if a stochastic volatility component is present in both ε_t and f_t, one would be interested in comparing the MSV model with one in which V_t is constant, or with a model in which D_t is constant, or with a model in which both V_t and D_t are constant, which is the classical factor model. To formally approach this problem we develop a technique based on Chib (1995) to calculate the marginal likelihood of each model. This estimate of the marginal likelihood is studied first in a Monte Carlo experiment in which data on y_t is generated a number of times from a given model and the marginal likelihood is computed across model classes and with different number of factors. The marginal likelihood computation is then further illustrated in two real data examples. In our view, the computation of the marginal likelihood is an essential part of the fitting strategy given the various choices of models and possible factors. It is also desirable because the marginal likelihood contains an automatic penalty against complexity and hinders the use of unnecessarily large models.

Finally, based on the work of Pitt and Shephard (1999b) we also provide methods for estimating the log-likelihood function and the filtered values of the time-varying volatilities $\{h_{jt}\}$. This leads to a complete analysis of the MSV model and its special cases that encompasses estimation, filtering and model choice. In sum, we believe that our procedures comprise the first unified and practical likelihood based analysis of empirically reasonable models of multi-dimensional high frequency time series.

The rest of the paper is organized as follows. In Section 2 we discuss the MCMC fitting of the general model and its special cases. The problem of marginal likelihood computation is addressed in Section 3. A procedure for the sequential filtering of the unknown volatilities and for the calculation of the likelihood function is also described in this section. Section 4 deals with a simulation study on the performance of the marginal likelihood criterion across various simulated data sets. Sections 5-6 are concerned with the empirical analysis of two data sets, a twenty dimensional vector of exchange rates and forty dimensional vector of US equity returns. Section 7 provides extensions of the models and Section 8 concludes.

2 PROPOSED ESTIMATION APPROACH

2.1 Formalization of the approach

Consider the question of estimating the general multivariate stochastic volatility model given in equations (1)-(3). Let the parameters of the model be denoted by $\psi = (\beta, \theta_1, ..., \theta_{p+k})$ where β denotes the free elements of B and θ_j is the vector consisting of $(\phi_j, \mu_j, \sigma_j^2)$. To develop our estimation approach we exploit two facets of the model. The first is that the distribution of y_t marginalized over f_t can be expressed as

$$y_t | h_t, B \sim \mathcal{N}_p(0, \Omega_t) \tag{4}$$

where

$$\Omega_t = \Omega_t(h_t) = V_t(h_t) + B D_t(h_t) B'$$

is a time-varying covariance matrix that depends on the loading matrix B. This form of the density of y_t is used in the MCMC simulation to update β. This expression also leads in principle to the likelihood function of the parameters ψ. If we let $\mathcal{F}_{t-1}$ denote the history of the $\{y_t\}$ process up to time $t-1$, and $p(h_t|\mathcal{F}_{t-1}, \psi)$ the density of h_t conditioned on $(\mathcal{F}_{t-1}, \psi)$, then the likelihood function of ψ given the data $y = (y_1, ..., y_n)$ is given by

$$\begin{aligned} p(y|\psi) &= \prod_{t=1}^{n} \int p(y_t|h_t, B) p(h_t|\mathcal{F}_{t-1}, \psi) dh_t \\ &= \prod_{t=1}^{n} \int \phi_p(y_t|0, \Omega_t) p(h_t|\mathcal{F}_{t-1}, \psi) dh_t \end{aligned} \tag{5}$$

where $\phi_p(.|\mu, \Sigma)$ is the p-variate density of a multivariate normal distribution with mean vector μ and covariance matrix Σ. The problem, alluded to in Section 1, is that neither $p(h_t|\mathcal{F}_{t-1}, \psi)$ nor the integral of $\phi_p(y_t|0, \Omega_t)$ over h_t is available in closed form which complicates the use of any direct method of doing inference on the parameters.

The second fact of the model is that by following the procedure introduced by Kim, Shephard, and Chib (1998) it is possible, given the matrix B and the factors $\{f_t\}$, to convert the MSV model into an independent collection of $(p+k)$ conditionally Gaussian state space models. To show how this is done, let

$$h_{j.} = (h_{j1}, ..., h_{jn}), \ \ (j \leq p+k),$$

denote the log-volatilities of the jth SV process over the sample of n observations. Also define $a_t = Bf_t$, a p vector with components a_{jt} and let

$$z_{jt} = \begin{cases} \ln(y_{jt} - a_{jt})^2 \ , & j \leq p \\ \ln(f_{j-pt}^2) \ , & j \geq p+1 \end{cases} . \tag{6}$$

Note that z_{jt} is a function of the data, the loading matrix B and the factors f_t. Then, from Kim, Shephard, and Chib (1998) it follows that the distribution of z_{jt}, which is h_{jt} plus a log chi-squared random variable with one degrees of freedom, may be approximated rather closely by a seven component mixture of normal distributions. As a consequence of this, the MSV model can be devolved into $(p+k)$ independent conditionally Gaussian state space models

$$\begin{aligned} z_{jt}|s_{jt}, h_{jt} &\sim \mathcal{N}(h_{jt} + m_{s_{tj}}, v_{s_{tj}}^2) \\ h_{jt} - \mu_j &= \phi_j(h_{jt-1} - \mu_j) + \sigma_j \eta_{jt} \ , \ j \leq p+k \end{aligned} \tag{7}$$

where s_{jt} is a discrete component indicator variable with mass function

$$\Pr(s_{jt} = i) = q_i \ , \ i \leq 7, \ t \leq n$$

and $m_{s_{tj}}$, $v_{s_{tj}}^2$ and q_i are parameters that are reported in Table 1. Thus, under this representation, conditioned on $s_{j.} = (s_{j1}, ..., s_{jn})$, each series z_{jt} $(j \leq p+k)$ can be subjected to separate

s_{jt}	q	$m_{s_{jt}}$	$v^2_{s_{jt}}$
1	0.00730	-11.40039	5.79596
2	0.10556	-5.24321	2.61369
3	0.00002	-9.83726	5.17950
4	0.04395	1.50746	0.16735
5	0.34001	-0.65098	0.64009
6	0.24566	0.52478	0.34023
7	0.25750	-2.35859	1.26261

Table 1: *Parameters of seven-component Gaussian mixture to approximate the distribution of* $\log \chi_1^2$.

analysis to determine the updates of the parameters $(\phi_j, \mu_j, \sigma_j^2)$. In particular, note that the joint density of the "observations" $z_{j.} = (z_{j1}, ..., z_{jn})$ conditioned on $(s_{j.}, \theta_j)$ can be expressed as

$$f(z_{j.}|\mathcal{F}^*_{jt-1}, s_{j.}, \theta_j) = \prod_{t=1}^{n} f(z_{jt}|\mathcal{F}^*_{j,t-1}, s_{j.}, \theta_j) \tag{8}$$

where $\mathcal{F}^*_{j,t-1}$ denotes the history of z_{jt} up to time $(t-1)$ and $f(z_{jt}|\mathcal{F}^*_{j,t-1}, s_{j.}, \theta_j)$ is a normal density whose parameters are obtained by the Kalman filter recursions, adapted to the differing components, as indicated by the component vector $s_{j.}$. This representation of the model is what ensures that our approach is scalable in both p and k.

We mention that, in the MCMC context, the use of this approximation is highly recommended because the minor approximation error can be removed easily, at the conclusion of the posterior sampling, by a reweighting procedure, as discussed by Kim, Shephard, and Chib (1998). This strategy of working with a highly efficient approximating model, and then reweighting the posterior sample ex-post, is a powerful and general method of dealing with complicated models in the simulation context.

2.2 Prior distributions

The MSV model provided above is completed through the specification of the distribution $p(\beta, \theta)$ on the parameters. In the context of our computational MCMC estimation scheme the nature of the prior is not important and virtually any distributional form may be adopted in the data analysis. Nonetheless, in the examples below we assume that the parameters are mutually independent and that prior information can be represented by the following distributions. For ϕ_j, we follow Kim, Shephard, and Chib (1998) and formulate our prior in terms of $\phi_j = 2\phi_j^* - 1$, where ϕ_j^* is distributed as Beta with parameters $\left(\phi^{(1)}, \phi^{(2)}\right)$. This implies that the prior on $\phi_j \in (-1, 1)$ is

$$p(\phi_j) = c\left\{0.5(1+\phi_j)\right\}^{\phi^{(1)}-1}\left\{0.5(1-\phi_j)\right\}^{\phi^{(2)}-1}, \quad \phi^{(1)}, \phi^{(2)} > 0.5\,, \tag{9}$$

where

$$c = 0.5\frac{\Gamma\left(\phi^{(1)} + \phi^{(2)}\right)}{\Gamma(\phi^{(1)})\Gamma(\phi^{(2)})}$$

is the normalizing constant. Under this prior, the prior mean is

$$2\phi^{(1)}/(\phi^{(1)} + \phi^{(2)} - 1)\,.$$

Each of these priors imposes stationarity which in our view is important in this context. For σ_j we assume that our prior information can be represented by an inverse gamma density

$$p(\sigma_j) = \frac{\delta_0^{\nu_0}}{\Gamma(\nu_0)}\left(\frac{1}{\sigma_j}\right)^{\nu_0+1}\exp\left(-\delta_0/\sigma_j\right), \quad \sigma_j > 0$$

and for μ_j we assume independent normal priors $\mathcal{N}(\mu|\mu_0, M_0)$. Finally, for each of the elements of B we assume the normal distributions $\mathcal{N}(b_{ij}|b_0, A_0)$. The values of the hyperparameters $(\mu_0, M_0, \nu_0, \delta_0, \phi^{(1)}, \phi^{(2)}, b_0, A_0)$ are specified when the examples are discussed.

2.3 Posterior simulation

We now consider the simulation of the posterior distribution of the parameters

$$\psi = (\beta, \theta_1, ..., \theta_{p+k})$$

by Markov chain Monte Carlo methods. One basic idea is to include the latent $\{h_t\}$ and the latent factors $\{f_t\}$ in the sampling. Another is the use of reduced blocking steps to improve the mixing of the chain. Specifically, consider the sampling of the posterior distribution

$$\pi\left(B, \{f_t\}, \{\theta_j\}, \{h_{j.}\}|y\right) \propto p\left(y|B, \{f_t\}, \{h_{j.}\}\right)p\left(\{f_t\}|\{h_{j.}\}\right)p\left(\{h_{j.}\}|\{\theta_j\}\right)p\left(B, \{\theta_{j.}\}\right)$$

where β as before denotes the free elements of B and y the $T \times N$ matrix of observations. To simulate from this posterior density we utilize essentially two steps: a step in which $(B, \{f_t\})$ in sampled in one block conditioned on $(y, \{h_{j.}\})$ and a second step in which $(\{\theta_{j.}\}, \{h_{j.}\})$ is sampled in one block conditioned on $(y, B, \{f_t\})$. Each step is important for improving the mixing of the MCMC output. To understand the motivation for the first step, which has not been contemplated for such models before, note that B and f_t appear in product form. Thus, sampling B conditioned on $\{f_t\}$ and then sampling $\{f_t\}$ conditioned on B produces unacceptably high serial correlations. It is surprising that this point has been overlooked in prior work. The reason for the second step is based on the work of Chib, Nardari and Shephard (1998) where the benefits of that step are documented in the context of univariate SV models. We note that the implementation of the second step is achieved by devolving the model into $(p+k)$ conditionally Gaussian state space models through the introduction of the latent state variables $(s_{1.}, ..., s_{p+k,.})$ and the transformation of the conditioning elements $(y, B, \{f_t\})$ into the variables $(z_{1.}, ..., z_{p+k,.})$.

Algorithm for sampling MSV **posterior distribution**

1. *Initialize* $\{h_{j.}\}$.
2. *Sample* $[\beta, \{f_t\}|y, \{h_{j.}\}]$ *by drawing*
   (a) β *from* $[\beta|y, \{h_{j.}\}]$
   (b) f_t *from* $[f_t|y_t, B, \{h_{j.}\}]$, $t \leq n$.
3. *Compute* $\{z_{tj}\}$ *and repeat for each* $j \leq p+k$,
   (a) *Sample* $s_{j.}$ *from* $[s_{j.}|z_{j.}, h_{j.}]$
   (b) *Sample* $[\theta_j, h_{j.}]$ *by drawing* θ_j *from* $[\theta_j|z_{j.}, s_{j.}]$ and $h_{j.}$ *from* $[h_{j.}|z_{j.}, s_{j.}, \theta_j]$.
4. *Goto step 2 and repeat.*

2.4 Detailed discussion of the algorithm

The algorithm begins in Step 1 with an initialization only of $\{h_{j.}\}$. In Step 2 we sample $(\beta, \{f_t\})$ jointly by the method of composition from the distribution $\pi(\beta, \{f_t\}|y, \{h_{j.}\})$. This is done by drawing β from $\pi(\beta|y, \{h_{j.}\})$, marginalized over $\{f_t\}$, and then sampling from $\{f_t\}$ from $\pi(\{f_t\}|y, B, \{h_{j.}\})$. The first of these two sampling steps involves the density

$$\begin{aligned}\pi(\beta|y, \{h_{j.}\}) &\propto p(\beta)\prod_{t=1}^{n} p(y_t|B, h_t) \\ &\propto p(\beta)\prod_{t=1}^{n} \phi_p(y_t|0, V_t + BD_tB'),\end{aligned}$$

which is not easy to simulate given that the dimension of β is typically very large (of the order of many hundreds). To confront this problem we follow Chib and Greenberg (1998) and employ a tailored Metropolis-Hastings step that is based on a multivariate-t proposal $f_T(\beta|m, \Sigma, \nu)$ with ν degrees of freedom. The mean m of the t-proposal is taken to be the approximate mode of $l = \log\{\prod_{t=1}^{n} \phi_p(y_t|0, V_t + BD_tB')\}$ and the dispersion matrix Σ is specified as minus the inverse of the second derivative matrix of l. Write

$$\begin{aligned}l &= \sum_{t=1}^{n} \log \phi_p(y_t|0, \Omega_t) \\ &= const - \frac{1}{2}\sum_{t=1}^{n} \log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{n} y_t'\Omega_t^{-1}y_t\,,\end{aligned}$$

where $\Omega_t = V_t + BD_tB'$. If we let the ij element of B be denoted by b_{ij} then

$$\begin{aligned}\frac{\partial l}{\partial b_{ij}} &= \frac{1}{2}\sum_{t=1}^{n}\left\{y_t'\Omega_t^{-1}\frac{\partial\Omega_t}{\partial b_{ij}}\Omega_t^{-1}y_t - tr\left(\Omega_t^{-1}\frac{\partial\Omega_t}{\partial b_{ij}}\right)\right\} \\ &= \frac{1}{2}\sum_{t=1}^{n}\left[y_t'\Omega_t^{-1}\left(\frac{\partial B}{\partial b_{ij}}D_tB' + BD_t\frac{\partial B'}{\partial b_{ij}}\right)\Omega_t^{-1}y_t - tr\left\{\Omega_t^{-1}\left(\frac{\partial B}{\partial b_{ij}}D_tB' + BD_t\frac{\partial B'}{\partial b_{ij}}\right)\right\}\right],\end{aligned}$$

using, for example, Magnus and Neudecker (1988, p. 151). So if we write $s_t = \Omega_t^{-1}y_t$ and $E_t = \Omega_t^{-1}BD_t$ then this simplifies to

$$\frac{\partial l}{\partial b_{ij}} = \sum_{t=1}^{n}\left[s_t'\frac{\partial B}{\partial b_{ij}}D_tB's_t - tr\left\{E_t\frac{\partial B'}{\partial b_{ij}}\right\}\right],$$

where the term that dominates the computation is Ω_t^{-1}. However (e.g. Harvey (1981, Ch. 4)),

$$\begin{aligned} \Omega_t^{-1} &= \left(V_t + BD_tB'\right)^{-1} \\ &= V_t^{-1} - V_t^{-1}B\left(D_t^{-1} + B'V_t^{-1}B\right)^{-1} B'V_t^{-1} \end{aligned}$$

which is far simpler and much faster to evaluate. These derivatives are used to implement a sequence of Newton-Raphson iterations in order to find m and Σ. Now define

$$\alpha(\beta, \beta^*|y, \{h_{j.}\}) = \min\left\{1, \frac{p(\beta^*)\prod_{t=1}^n \phi_p(y_t|0, V_t + B^*D_tB^{*\prime})}{p(\beta)\prod_{t=1}^n \phi_p(y_t|0, V_t + BD_tB')}\frac{f_T(\beta|m, \Sigma, \nu)}{f_T(\beta^*|m, \Sigma, \nu)}\right\}, \quad (10)$$

then the M-H algorithm (see Chib and Greenberg (1995)) is implemented by drawing a value β^* from $f_T(\cdot|m, \Sigma, \nu)$ and accepting the proposal value with probability $\alpha(\beta, \beta^*|y, \{h_{j.}\})$, where β is the current value. If the proposal value is rejected, then β is retained. We would like to emphasize that our reduced conditional sampling step of B is a centerpiece of this algorithm which in conjunction with the tailored sampling step (which is redone after every sweep as the conditioning elements change) leads to an algorithm that appears not to suffer from any dimensionality problem. As we will show below, we have been able to implement the above Metropolis-Hastings step even when the dimension of B is around three hundred.

The next part of Step 2 is concerned with the sampling of $\{f_t\}$ from $\pi(\{f_t\}|B, h, y)$. This step is quite simple because the $\{f_t\}$ are conditionally independent and each can be sampled from $f_t|y_t, h_t, B,\ t \leq n$, where

$$\pi(f_t|y_t, h_t, B) \propto p(f_t|h_t)p(y_t|f_t, h_t, B)$$

with

$$f_t|h_t \sim N(0, D_t), \qquad \text{and} \qquad y_t|f_t, h_t, B \sim N(Bf_t, V_t)\,.$$

As a result $f_t|y_t, \theta, h_t, B \sim N(\hat{f}_t, F_t)$, where

$$F_t = \left(B'V_t^{-1}B + D_t^{-1}\right)^{-1}, \qquad \hat{f}_t = F_tB'V_t^{-1}y_t.$$

In Step 3, having sampled $\{f_t\}$, we compute $\{z_{jt}\}$ according to (6). We can then write

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

which implies that the log-volatilities, mixture indicators and series specific parameters can be sampled series by series. Specifically, for each j one samples $(s_{j.}, \theta_j, h_{j.})$ by the univariate SV sampler of Chib, Nardari, and Shephard (1998), which extended and clarified Kim, Shephard, and Chib (1998). This sampling is done by first sampling $s_{j.}$ from $[s_{j.}|z_{j.}, h_{j.}]$. This is straightforward as

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

where $p(s_{jt}|z_{jt}, h_{jt})$ is a mass function with seven points of support given by

$$p(s_{jt}|z_{jt}, h_{jt}) \propto p(s_{jt}) p_N(z_{jt}|h_{jt} + m_{s_{tj}}, v_{s_{tj}}^2).$$

This mass function can be evaluated for each of the permissible values of s_{jt} and then sampled. Next, one samples $[\theta_j, h_{j.}]$ by drawing θ_j from $[\theta_j|z_{j.}, s_{j.}]$ followed by a draw of $h_{j.}$ from $[h_{j.}|z_{j.}, s_{j.}, \theta_j]$. The sampling of the θ_j is from the density

$$\pi(\theta_j|z_{j.}, s_{j.}) \propto p(\theta_j) p(z_{j.}|s_{j.}, \theta_j)$$

where

$$p(z_{j.}|s_{j.}, \theta_j) = p(z_{j1}|s_{j.}, \theta_j) \prod_{t=2}^{T} p(z_{jt}|\mathcal{F}_{jt-1}^*, s_{j.}, \theta_j), \tag{11}$$

is computed from the output of the Kalman filter recursions (see Harvey (1989)) and $\mathcal{F}_{jt}^*$ denotes the history of the $\{z_{jt}\}$ process up until time t. To sample $\pi(\theta_j|z_{j.}, s_{j.})$ one utilizes the Metropolis-Hastings algorithm with a multivariate t proposal (see Chib, Nardari, and Shephard (1998) for further details). Finally, we sample $h_{j.}$ from $[h_{j.}|z_{j.}, s_{j.}, \theta_j]$ which is a multivariate normal density. This can be sampled exactly and efficiently by the simulation smoother of de Jong and Shephard (1995) as discussed in this context by Kim, Shephard, and Chib (1998).

The algorithm, therefore, has two key steps: one involving the simulation of β and the other the sampling of $\{\theta_j\}$. Both steps are crucial to providing a procedure that is reliable, scalable and efficient.

2.5 Special cases of the model

We now explain how the general approach developed above can be modified for some special cases of the model.

Classical Factor model (CF).

This model is defined by the set-up

$$y_t|B, f_t \sim \mathcal{N}_p(Bf_t, V), \quad f_t \overset{i.i.d.}{\sim} \mathcal{N}_k(0, I)$$

where V is a diagonal matrix. Thus, the variances of both the errors and factors are constant. Lopes and West (1999) is a recent paper that deals with the MCMC analysis of this model but the fitting algorithm in that paper does not use reduced blocking to sample the loading matrix B. Due to the reasons mentioned above reduced blocking is important even in this more simplified setting. Let the components of V and D, namely $\{v_{jj}\}$ and $\{d_{jj}\}$, follow an inverse gamma prior distribution and let $h_{jt} = 0$ for all j and t. Also let $\theta = (v_{11}, ..., v_{pp}, d_{11}, ..., d_{kk})$. Then an efficient algorithm to fit this model has the following simple structure.

1. *Initialize* $\theta = (v_{11}, ..., v_{pp}, d_{11}, ..., d_{kk})$.
2. *Sample* $[\beta, \{f_t\}|y]$ *by drawing*
   (a) β *from* $[\beta|y, \theta]$ *by the M-H algorithm with proposal value and probability of move defined similarly to (10).*
   (b) f_t *from* $[f_t|y_t, B, \theta]$, *an updated normal distribution,* $t \leq n$.
3. *Sample* v_{jj} *by drawing* v_{jj} *from* $[v_{jj}|y_{j\cdot} - a_{j\cdot}]$, $j \leq p$, *an updated inverse gamma distribution.*

4. *Sample* d_{jj} *by drawing* d_{jj} *from* $[D_{jj}|f_{j.}]$, $j \leq k$, *an updated inverse gamma distribution.*

5. *Goto step 2 and repeat.*

Stochastic Volatility in Factor (SVF) model.

Under this model,

$$\begin{aligned} y_t|B, f_t &\sim \mathcal{N}_p(Bf_t, V), \quad f_t|h_t \sim \mathcal{N}_k(0, D_t) \\ h_{jt} - \mu_j &= \phi_j(h_{jt-1} - \mu_j) + \sigma_j \eta_{jt} \ , \ j \leq k \end{aligned}$$

where the variances of the y_t are constant. This model is similar to that of Diebold and Nerlove (1989) where an ARCH process was specified to generate volatility clustering. It has also been studied by Jacquier, Polson, and Rossi (1995) and at more length by Aguilar and West (1998). Under this simplified model, if $k < p$ then it becomes possible to find linear combinations (portfolios in financial applications) which are homoskedastic as

$$B_{\perp} y_t | h_t, B \sim N(0, B_{\perp} V B_{\perp}'), \qquad \text{where} \qquad B_{\perp} B = 0.$$

This reduction is not possible in the general MSV model. This type of transformation of a vector process, which remove a major feature of the original model (in this case stochastic volatility), was the focus of the common features paper by Engle and Kozicki (1993). To fit this model, we set $\{h_{jt} = 0, j \leq p\}$ and condense Step 3 of the general algorithm to run only over the indices $j = p+1$ to $j = p+k$. This delivers the following algorithm.

1. *Initialize* $\{h_{j.}, j > p\}$ and $v_{11}, ..., v_{pp}$.

2. *Sample* $[\beta, \{f_t\}|y, \{h_{j.}, j > p\}, v_{11}, ..., v_{pp}]$ *by drawing*

   (a) β *from* $[\beta|y, \{h_{j.}, j > p\}, v_{11}, ..., v_{pp}]$ *by the M-H algorithm with proposal value and probability of move defined similarly to (10).*

   (b) f_t *from* $[f_t|y_t, B, \{h_{j.}, j > p\}, v_{11}, ..., v_{pp}]$, *an updated normal distribution,* $t \leq n$.

3. *Compute* $\{z_{tj}\}$ *and repeat for each* $j > p$,

   (a) *Sample* $s_{j.}$ *from* $[s_{j.}|z_{j.}, h_{j.}]$

   (b) *Sample* $[\theta_j, h_{j.}]$ *by drawing* θ_j *from* $[\theta_j|z_{j.}, s_{j.}]$ and $h_{j.}$ *from* $[h_{j.}|z_{j.}, s_{j.}, \theta_j]$.

4. *Sample* v_{jj} *by drawing* v_{jj} *from* $[v_{jj}|y_{j.} - a_{j.}]$, $j \leq p$, *an updated inverse gamma distribution.*

5. *Goto step 2 and repeat.*

Stochastic volatility in Error (SVE) model.

In this model,

$$\begin{aligned} y_t|B, f_t, h_t &\sim \mathcal{N}_p(Bf_t, V_t), \quad f_t \overset{i.i.d.}{\sim} \mathcal{N}_k(0, I) \\ h_{jt} - \mu_j &= \phi_j(h_{jt-1} - \mu_j) + \sigma_j \eta_{jt} \ , \ j \leq p. \end{aligned}$$

This has factors which are independent normals, while the variances of the y_t process each follow SV processes. To fit this model, we set $\{h_{jt} = 0, j > p\}$ and condense Step 3 of the general algorithm to run only over the indices $j = 1$ to $j = p$. Finally, the θ's for $j > p$, which only consist of the parameters d_{11} to d_{kk}, are sampled from an inverse gamma distribution, conditioned on the factors and the loading matrix B. Since the algorithm is very similar to the one for the SVF model the details are suppressed.

3 MODEL COMPARISON

3.1 Bayes factors

To date there has been no formal study of the empirical validity of multivariate volatility models in relation to alternative multivariate and univariate specifications. This may be because the extant problems of estimating such models has overshadowed all other inferential questions. We now rectify this deficiency by considering the computation of the marginal likelihood of the general multivariate SV model. The marginal likelihood of the other models follows as a special case of the technique that we develop. We assume that the reader is familiar with the notion of the marginal likelihood and of its role in Bayesian model comparisons. The approach we adopt here for the computation of the marginal likelihood and Bayes factors is based on the method developed by Chib (1995). Let $\psi = (\beta, \theta)$, then from the basic marginal likelihood identity in Chib (1995), the log of the Bayes factor for comparing non-nested models $\mathcal{M}_1$ to $\mathcal{M}_2$ is written as

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

where $p(y|\mathcal{M}, \psi_j^*)$ is the likelihood function under $\mathcal{M}_j$, $p(\psi_j^*|\mathcal{M}_j)$ and $\pi(\psi_j^*|y, \mathcal{M}_j)$ are the corresponding prior and posterior densities, each evaluated at the point ψ_j^*. The choice of the points ψ_1^* and ψ_2^* is arbitrary but for computational accuracy we take these to be the posterior means that are estimated from the MCMC output. Suppressing the model subscript for convenience, there are two key quantities that must be computed in this case for each model under study: the posterior ordinate $\pi(\psi^*|y, \mathcal{M})$ and the likelihood ordinate $p(y|\mathcal{M}, \psi^*)$. Because the dimension of the vector ψ is typically in the hundreds, direct calculation of $\pi(\psi^*|y, \mathcal{M})$ and the likelihood ordinate $p(y|\mathcal{M}, \psi^*)$ is not possible. We therefore need to develop estimators for each of the quantities.

3.2 Posterior ordinate

Chib (1995) and Chib and Greenberg (1998) present an effective way of estimating posterior ordinates that proceeds in terms of a marginal/conditional decomposition of the posterior density and the output from the original and subsequent "reduced MCMC runs." Let

$$\begin{aligned} \pi(\psi^*|y, \mathcal{M}) &= \pi(\beta^*, \psi_1^*, ..., \psi_{p+k}^*|y, \mathcal{M}) \\ &= \pi(\beta^*|y, \mathcal{M})\pi(\theta_1^*, ..., \theta_{p+k}^*|y, \mathcal{M}, \beta^*) \end{aligned}$$

and consider the estimation of each of the terms 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 use the approximation that $\pi(\beta^*|y, \mathcal{M})$ is the ordinate of a normal density with mean vector and covariance matrix obtained from the MCMC output. To estimate the conditional ordinate $\pi(\theta_1^*, ..., \theta_{p+k}^*|y, \mathcal{M}, \beta^*)$ we adopt the strategy of Chib and Greenberg (1998) and 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 $\pi(\theta_1^*, \theta_2^*|y, \mathcal{M}, \beta^*)$ we fix β at β^* and run our MCMC algorithm given above. The desired ordinate is then estimated by the kernel method applied to the output on θ_1 and θ_2 from this run. Next, to estimate $\pi(\theta_3^*, \theta_4^*|y, \mathcal{M}, \beta^*, \theta_1^*, \theta_2^*)$ we fix β at β^* and θ_1 and θ_2 at θ_1^*, θ_2^* and proceed with the second reduced MCMC run. The output on (θ_3, θ_4) from this run is then used to estimate the second reduced conditional ordinate. The process is repeated in sequence, in each case with additional parameters being held fixed. Then, the sum of the log of each of these estimates is taken to the estimate of the log posterior ordinate.

3.3 Likelihood ordinate via particle filtering

Now consider the estimation of the likelihood ordinate $p(y|\mathcal{M}, \psi^*) = \prod_{t=1}^n p(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$ where

$$p(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) = \int \phi_p(y_t|0, \Omega_t)p(h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)dh_t$$

is a normal density marginalized over the distribution of $h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*$ and $\mathcal{F}_t = (y_1, ..., y_t)$ is the data up to time t. Suppose that a sample of $h_{t-1}^{(g)}$ $(g \leq M)$ is available from the filtered distribution $h_{t-1}|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*$ for each t. Then, the sequence of one-step ahead densities $f(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$ can be obtained by the following procedure.

Estimation of the likelihood function

1. *Set $t = 1$, initialize ψ^* and obtain a sample of $h_{t-1}^{(g)}$ $(g \leq M)$.*

2. *For each value of* $h_{j,t-1}^{(g)}$, *sample* $h_{j,t}$ *from the Gaussian evolution equation*

$$h_{jt}|h_{j,t-1}^{(g)}, \psi^* \sim \mathcal{N}\left(\mu_j^* + \phi_j^*(h_{j,t-1}^{(g)} - \mu_j^*), \sigma^2\right) \ , \ j \leq p+k$$

3. *Estimate the one-step ahead density as*

$$\widehat{f}(y_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) = \frac{1}{M}\sum_{j=1}^{M} \phi_p\left(y_t|0, \Omega_t(h_t^{(g)})\right)$$

where

$$\Omega_t(h_t^{(g)}) = V_t^{(g)} + BD_t^{(g)}B',$$

and $V_t^{(g)}$ *and* $D_t^{(g)}$ *are the diagonal matrices with typical term given by* $\exp(h_{j,t}^{(g)})$.

4. *Apply the filtering procedure described below to obtain a sample* $h_t^{(1)}, ..., h_t^{(M)}$ *from the filtered distribution* $h_t|\mathcal{M}, \mathcal{F}_t, \psi^*$.

5. *Increment* t *to* $t+1$ *and goto Step 2.*

In this procedure Steps 2 and 3 are both straightforward. The key question is the sampling of $h_t^1, ..., h_t^M$ from the filtered distribution. This may be done using what is called the *particle filter* in the literature (see Pitt and Shephard (1999a)). From Bayes theorem,

$$p(h_t|\mathcal{M}, \mathcal{F}_t, \psi^*) \propto p(y_t|\mathcal{M}, h_t, \psi^*)p(h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) \tag{13}$$

where $p(y_t|h_t, \psi^*) = \phi_p(y_t|0, \Omega_t)$,

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

and $p(h_t|\mathcal{M}, h_{t-1}, \psi) = \prod_{j=1}^{p+k} f_N\left(h_{tj}|\mu_j^* + \phi_j^*(h_{j,t-1} - \mu_j^*), \sigma^2\right)$. The latter integral can be estimated from the sample $h_{t-1}^{(1)}, ..., h_{t-1}^{(M)}$ leading to the approximations

$$\begin{aligned} p(h_t|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*) &\simeq \frac{1}{M}\sum_{g=1}^{M} p(h_t|\mathcal{M}, h_{t-1}^{(g)}, \psi^*) \\ p(h_t|\mathcal{M}, \mathcal{F}_t, \psi^*) &\dot{\propto} f(y_t|h_t, \psi^*)\frac{1}{M}\sum_{g=1}^{M} f(h_t|\mathcal{M}, h_{t-1}^{(g)}, \psi^*). \end{aligned}$$

We now show how one can sample h_t from the latter density. The procedure requires a first stage in which proposal values $h_t^{*(1)}, ..., h_t^{*(R)}$ are created. These values are then reweighted to produce draws $\{h_t^{(1)}, ..., h_t^{(M)}\}$ that correspond to draws from $p(y_t|\mathcal{M}, h_t, \psi^*) \sum_{g=1}^{M} p(h_t|\mathcal{M}, h_{t-1}^{(g)}, \psi)$. Typically one may take R to be five or ten times larger than M. In all the examples below, we let $M = 2000$ and $R = 10,000$ though one may select even higher values if computer resources are available. We now summarize the steps involved for the filter in period t.

Auxiliary particle filter for multivariate SV model

1. *Given values* $\{h_{t-1}^{(1)}, ..., h_{t-1}^{(M)}\}$ from $(h_{t-1}|\mathcal{M}, \mathcal{F}_{t-1}, \psi^*)$ *calculate*

$$\hat{h}_{j,t}^{*(g)} = \mu_j^* + \phi_j^*(h_{j,t-1}^{(g)} - \mu_j^*) \ , \ j \leq p+k$$

$$w_g = \phi_p(y_t|0, \Omega_t(\hat{h}_t^{*(g)}), \psi^*) \ , \ g = 1, ..., M \ ,$$

   and sample R *times the integers* $1, 2, ..., M$ *with probability proportional to* $\{w_j\}$. *Let the sampled indexes be* $k_1, ..., k_R$ *and associate these with* $\hat{h}_t^{*(k_1)}, ..., \hat{h}_t^{*(k_R)}$.

2. *For each value of* k_j *from Step 1, simulate the values* $\{h_t^{*(1)}, ..., h_t^{*(R)}\}$ *from the volatility process as*

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

   where $\eta_{j,t}^{(g)} \sim N(0,1)$.

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

$$\frac{\phi_p\left(y_t|0, \Omega_t(h_t^{*(g)})\right)}{\phi_p\left(y_t|0, \Omega_t(\hat{h}_t^{*k_g})\right)}, \ g = 1, ..., R \ ,$$

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

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,

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

the s-step ahead return vector or

$$\frac{1}{s}\sum_{j=1}^{s} \Omega^*(h_{t+1}, ..., h_{t+s})|\mathcal{F}_t, \theta^*, \mathcal{M}, \quad \text{where} \quad \Omega^*(h_{t+1}, ..., h_{t+s}) = \sum_{j=1}^{s} \Omega(h_{t+j}),$$

the associated average covariance over that period of length s. Given the output from the particle filter of particles $h_t^1, ..., h_t^M$, we can simply propagate each of these particles forward using their autoregressive structure (3) to get samples from $f(h_{t+1}, ..., h_{t+s}|\mathcal{F}_t, \theta^*, \mathcal{M})$. These can be used in a variety of ways, for example noting that

$$\sum_{j=1}^{s} y_{t+j}|\mathcal{F}_t, \theta^*, \mathcal{M}, h_{t+1}, ..., h_{t+s} \sim N\left\{0, \Omega^*(h_{t+1}, ..., h_{t+s})\right\},$$

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

4 SIMULATION STUDY

Before turning to the real data examples, we utilize simulated data to assess the performance of the marginal likelihood estimated by the Chib method from the MCMC and particle filter output. We focus on the ability of this criterion to find the correct model across model types and, within a given model class, to identify the correct number of factors. The computation of the marginal likelihood proceeds as described in Section 4.2 through 4.4.

In the simulation design, data sets are generated from two of the models described earlier in the paper: the MSV model with three Factors and the SVF model with three Factors. Each simulated data set consists of ten series of 1200 observations each, similar to the sample size in the real data examples below. Specifically, in the design with MSV as the true model, the model parameters are randomly generated as follows. The elements of the factor loading matrix B are sampled from a normal distribution with mean 0.90 and variance 0.2. In addition, the μ's are sampled from $\mathcal{N}(-3, 4)$, the ϕ's from a scaled beta distribution with mean 0.93 and variance 0.03 and the σ's from $\mathcal{IG}(2.5, 0.5)$. Furthermore, when SVF is used as the true model, B and the factor θ's are generated as above whereas the elements of the V matrix are generated from $\mathcal{IG}(2.5, 0.5)$. The actual parameter values generated by these procedures are available from the authors upon request.

We generate a total of 25 data sets from each of the two models. All the four models considered in the paper (MSV, SVE, SVF and CF) are then fit to these data sets, each with between 2 and 4 factors. Thus, for each design twelve models are each fit twenty five times. For the prior distributions the hyperparameters are specified as follows: $\mu_0 = 0, M_0 = 25, \nu_0 = 3, \delta_0 = 1, \phi^{(1)} = 20, \phi^{(2)} = 1.5, b_0 = 0, A_0 = 25$. These choices imply a prior mean of 0.86 for the ϕ_j's and a prior mean of 0.5 for the σ_j's and/or for the diagonal elements of the V matrix in the cases where V is not time-varying.

The marginal likelihood for each model for each simulated data set is based on $G = 5000$ MCMC iterations (beyond a burn-in of 500 iterations) followed by reduced runs of 2500 iterations. In calculating the likelihood ordinate the particle filter algorithm is based on $M = 2000$ and $R = 10000$.

Model	Factors		
	2	3	4
CF	0.0	0.0	0.0
SVF	0.0	0.0	0.0
SVE	0.0	0.0	0.0
MSV	8.0	84.0	0.0

Table 2: Comparison of marginal likelihoods on simulated data from the MSV 3-factor Model. Each entry in the table is the percentage of times the corresponding model receives support over all other models. In some cases there are ties and then we do not allocate the result to a particular cell. See the main text for details.

To describe the sampling performance of the marginal likelihood we calculate the frequency with which each possible k-factor model ($k = 2, 3, 4$) is picked over the other models. Table 7 reports the results for the case when the MSV model with three factors is used as the true data generating process. A first important result is that the MSV model (regardless of the number of factors) is always picked over the other models. Within the MSV model class, we find that the (true) 3-factor model is picked over the 2 and 4 factor versions 84% of the times while 8% of the times the 2-factor MSV model is picked. The rest of the times the two factor and three factor models are tied. These assessments are based on the Jeffreys' scale. Interestingly, the 4-factor MSV model always receives less support than the more parsimonious versions.

In Table 8 we present the results from the simulation when the 3-factor SVF model is the used to generate the data. In this case, a model from the SVF model class is supported in each experiment. Within the SVF class, the 3-factor model is correctly selected as the best model 88% of the times and it is found to be equivalent to the 2-factor model in 8% of the cases.

Model	Factors		
	2	3	4
CF	0.0	0.0	0.0
SVF	4.0	88.0	0.0
SVE	0.0	0.0	0.0
MSV	0.0	0.0	0.0

Table 3: Comparison of marginal likelihoods on simulated data from the SVF 3-factor Model. Each entry in the table is the percentage of times the corresponding model receives support over all other models. In some cases there are ties and then we do not allocate the result to a particular cell. See the main text for details.

In conclusion, the results of this section support the use of the marginal likelihood as a model choice criterion for selecting both the appropriate model class within the multivariate stochastic volatility specifications considered in this study and for identifying the correct number of common factors within each class.

5 DATA DESCRIPTION

We continue our study of the methods developed in this paper by analyzing two fairly large data sets.. The first is a data set on daily foreign exchange rates (recorded at noon EST each day and taken from Datastream International) for the period from 12/06/91 to 11/21/96, yielding 1250 time series observations. If we let r_t denote the exchange rate at time t, our response variable is the zero-mean daily continuously compounded return $y_t = \log(r_t/r_{t-1}) - \frac{1}{n}\sum \log(r_t/r_{t-1})$, which for simplicity ignores the small effect arising from the difference between the daily riskless interest rate in the two countries. Information about the data being used is given in Table 4, along with some basic summary statistics (mean, variance and Box-Ljung statistic on y_t and $|y_t|$ using 48 lags). All the currencies are quite heavily traded which implies that there is little autocorrelation amongst the y_t, but it can be seen from the Table that almost all series display volatility clustering. In addition, there is considerable comovement in the series. To display this comovement, Table 4 also gives the regression coefficient of the return on the three most important currencies DM, Pound and Yen. Finally, the correlation matrix of the returns, given in Table 5, contains large entries with the exception of the correlations involving the Canadian dollar, the Australian dollar and the Hungarian Forint.

Our second data set looks at the daily returns from the CRSP data base on 40 major stocks listed on the New York Stock Exchange. The return sequences include dividend payments and are adjusted for stock splits. The period of study is 11/9/1992 to 10/20/1997, yielding 1250 observations. The assets are listed in Table 6 together with some basic summary statistics. We also give in Table 6 the regression coefficient for returns on the average return of the other 39 stocks. These reflect the overall sensitivity of the stock to market conditions. We see that some stocks are very sensitive to this measure, while others seem not to follow the general movements of the index.

Table 6 shows that some stocks have a very moderate amount of linear predictability as the

Currency	Mean	Stdev	LB(48) y_t	LB(48) $\|y_t\|$	β_{pd}	β_{dm}	β_{yen}
British Pound	0.006	0.663	90.450	1134.000	1.000	0.691	0.372
French Franc	-0.004	0.675	45.590	448.400	0.737	0.888	0.526
German Mark	-0.004	0.698	82.110	397.900	0.764	1.000	0.555
Japanese Yen	-0.011	0.696	75.960	175.800	0.410	0.553	1.000
Swiss Franc	-0.007	0.789	75.060	272.600	0.834	1.038	0.634
Italian Lira	0.018	0.723	73.450	996.700	0.669	0.691	0.352
Canadian Dollar	0.013	0.292	52.920	238.700	-0.002	-0.040	-0.077
New Zealand Dollar	-0.020	0.391	73.180	68.710	0.109	0.070	0.059
Spanish Peseta	0.018	0.718	62.230	603.700	0.730	0.820	0.426
Australian Dollar	-0.003	0.497	49.290	165.900	0.040	-0.049	-0.109
Norwegian Krone	0.002	0.694	79.870	712.500	0.740	0.867	0.505
Netherlands Guilder	-0.004	0.723	76.100	483.200	0.751	0.969	0.569
Swedish Krona	0.011	0.770	83.880	535.100	0.690	0.739	0.370
Danish Krone	-0.005	0.673	62.480	448.400	0.728	0.873	0.511
Finnish Markka	0.005	0.867	80.400	385.500	0.785	0.883	0.504
Greek Drachma	0.000	0.705	105.500	760.000	0.797	0.799	0.421
Irish Punt	0.022	0.697	69.980	453.400	0.720	0.862	0.517
Belgian Franc	-0.004	0.724	54.440	344.900	0.735	0.915	0.562
Austrian Schilling	-0.004	0.711	54.680	344.700	0.644	0.784	0.453
Hungarian Forint	0.058	0.821	152.800	171.100	-0.049	0.044	0.032

Table 4: Summary Statistics for the twenty Exchange Rate series considered in this study. LB(48) is the Ljung-Box statistics computed using 48 lags. This is carried out on the return series, y_t and its absolute values $|y_t|$. Both the LB(48) statistics have null distributions of χ^2_{48}. β_{pd}, β_{dm} and β_{yen} are regression coefficients of the return series on the Pound, Deutsche Mark and Yen return, respectively

.73	1																	
.73	.92	1																
.39	.54	.55	1															
.70	.87	.92	.56	1														
.61	.67	.67	.34	.60	1													
-.01	-.08	-.10	-.18	-.10	.02	1												
.18	.13	.12	.11	.15	.10	.05	1											
.68	.79	.80	.41	.72	.66	-.04	.12	1										
.05	-.07	-.07	-.15	-.06	.01	.18	.51	-.05	1									
.71	.86	.87	.51	.81	.65	-.08	.14	.77	-.04	1								
.69	.89	.93	.55	.87	.64	-.11	.13	.76	-.07	.86	1							
.59	.68	.67	.34	.61	.64	.01	.09	.66	.00	.67	.65	1						
.72	.91	.91	.53	.84	.67	-.07	.14	.81	-.04	.90	.89	.70	1					
.60	.70	.71	.41	.65	.57	-.01	.12	.64	-.02	.72	.69	.64	.73	1				
.75	.77	.79	.42	.73	.64	-.03	.16	.72	.03	.77	.77	.63	.79	.62	1			
.69	.85	.86	.52	.82	.62	-.11	.13	.71	-.05	.82	.84	.60	.83	.64	.73	1		
.67	.88	.88	.54	.83	.62	-.07	.14	.76	-.07	.86	.87	.65	.91	.70	.76	.81	1	
.60	.75	.77	.44	.72	.51	-.04	.10	.66	-.02	.73	.75	.53	.77	.59	.70	.68	.79	1
-.04	.03	.04	.03	.03	.02	.00	-.04	.03	-.07	.04	.03	.01	.04	.00	.00	.01	.07	.09

Table 5: Correlation Matrix of the 20 Exchange Rate series considered in this study.

			LB(48)						LB(48)		
Stock	Mean	Stdev	y_t	$\|y_t\|$	β	Stock	Mean	Stdev	y_t	$\|y_t\|$	β
IBM	.00064	.0148	63.8	358.14	0.743	GT	.00081	.0153	49.6	54.82	0.805
JNJ	.00088	.0156	67.0	184.66	1.028	C	.00081	.0146	67.8	263.73	0.952
MCD	.00068	.0163	65.5	129.23	0.865	ALD	.00044	.0176	59.1	60.35	0.654
UA	.00125	.0166	68.6	118.37	1.029	JPM	.00066	.0119	90.9	398.76	0.645
GE	.00076	.0142	64.1	477.38	0.787	DIS	.00085	.0165	42.1	43.70	0.906
AXP	.00041	.0123	63.3	204.70	0.625	HWP	.00112	.0126	58.7	373.79	0.909
J	.00052	.0154	54.7	83.24	1.003	DD	.00072	.0168	39.2	60.29	0.814
XON	.00134	.0163	51.0	143.55	1.134	MO	.00106	.0145	29.8	152.43	0.724
GM	.00063	.0140	45.6	108.90	0.567	AA	.00057	.0149	50.1	81.17	0.624
EK	.00089	.0157	42.6	193.78	0.830	UK	.00133	.0213	54.8	131.01	0.998
CTR	.00054	.0190	60.7	81.55	0.893	MRR	.00007	.0215	45.8	54.82	0.610
PG	.00093	.0144	88.8	1143.3	0.834	MRK	.00107	.0198	52.0	263.73	0.783
S	.00073	.0136	45.7	50.06	0.679	MTC	.00034	.0153	55.9	60.35	0.557
KO	.00112	.0174	52.0	113.38	1.017	RAL	.00088	.0154	62.7	398.76	0.918
T	.00074	.0128	67.1	103.20	0.679	BUD	.00035	.0140	64.7	43.70	0.570
IP	.00092	.0192	51.1	70.53	1.085	GS	.00064	.0135	40.3	373.79	0.476
WMT	.00168	.0176	76.7	218.99	1.098	F	.00085	.0164	50.4	60.29	0.445
SD	.00101	.0136	47.1	161.07	0.842	BCC	.00047	.0130	65.1	152.43	0.313
CHW	.00123	.0176	52.3	66.98	0.848	FNC	.00075	.0116	65.8	81.17	0.135
BA	.00042	.0286	72.2	65.14	1.257	TWX	.00123	.0166	60.6	131.01	0.932

Table 6: Summary Statistics for the 40 stock return series considered in this study. LB(48) is the Ljung-Box statistics computed using 48 lags. This is carried out on the return series, y_t and its absolute values $|y_t|$. Both the LB(48) statistics have null distributions of χ^2_{48}. β is the regression coefficient of each return series on the equally weighted average of the other 39 return series

LB(48) on the y_t is sometimes slightly outside what is plausible for a χ^2_{48}. This is not surprising as stocks are well known to often exhibit mild autocorrelation due to, for example, discontinuous trading (see, for example, Scholes and Williams (1977) and Lo and MacKinlay (1988)).

More striking is the high level of autocorrelation in the absolute values of the returns. Around two-thirds of the individual stocks have LB statistics which are very high, while the others are all at least moderate. This suggests volatility clustering is a common feature of the individual stocks although it seems more moderate than in the exchange rate data. Our model with common effects across stocks should be helpful in explaining these features.

6 EMPIRICAL RESULTS

6.1 Preliminaries

In this section we report the results from fitting the four alternative multivariate stochastic volatility models to the two data sets described above. In all cases the MCMC simulation is implemented for an initial burn-in period of 1000 sweeps followed by a subsequent set of 10000 sweeps. In presenting the results, we describe the efficacy of the proposed procedures in terms of the inefficiency factors or autocorrelation times and provides marginal posterior densities and marginal likelihoods for the fitted models. The inefficiency factor, or the autocorrelation time, of each posterior estimate (computed as a sample average over the simulated values) is defined as the spectrum at zero, $1 + 2\sum_{j=1}^{\infty} \rho_j$, where $\{\rho_j\}$ denotes the autocorrelation function of the simulated values. The inefficiency factor is estimated as

$$1 + \frac{2N}{N-1} \sum_{j=1}^{B_N} K\left(\frac{j}{B_N}\right) \hat{\rho}_j$$

where $\hat{\rho}_j$ is an estimate of the autocorrelation at lag j and $K(\cdot)$ is the Parzen kernel based on the first B_N autocorrelations. The inefficiency factor is equal to the square of the numerical standard error divided by the variance of the posterior estimate under (hypothetical) i.i.d. sampling from the posterior.

The prior distributions used in this analysis are identical to those adopted in the simulation experiments in Section 5 except that $\nu_0 = 5$, and $\delta_0 = 0.4$, implying a mean of 0.1 for the prior distribution on σ_j's and/or for the diagonal elements of the V matrix in the cases where V is not time-varying.

6.2 Exchange rate data

6.2.1 Estimation results

The general factor SV model (MSV) as well as the three special cases described earlier in Section 4.1 are fit to the twenty exchange rate series in Table 4. In each case, the model is fit sequentially increasing the number of common factors until there is no improvement in the model

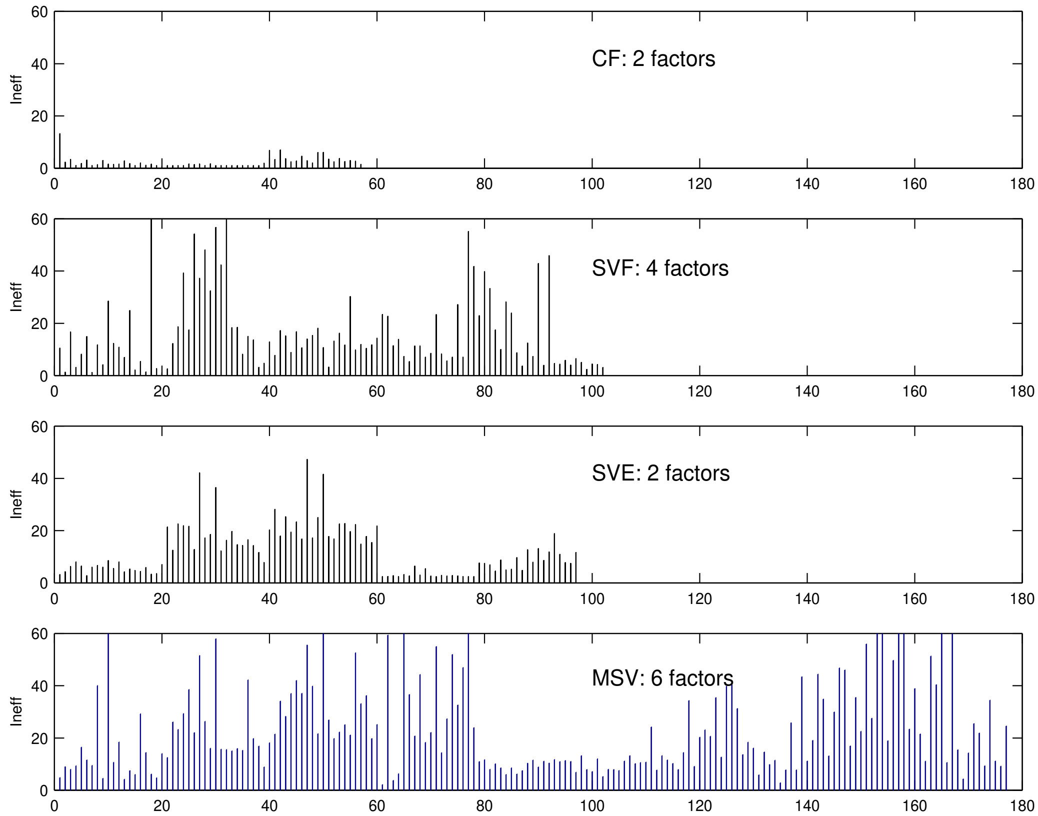

Figure 1: Inefficiency factors for best fitting models (selected from the marginal likelihoods in Table 7) in each of the model classes CF, SVF, SVE and MSV. See text for further details.

performance. Consider first the diagnostics on the mixing properties of the MCMC sampling scheme. Figure 1 reports the inefficiency measures for different number of factors within each model class. In the first panel of the figure, bars 1-20 are the diagonal elements of V; 21-57 are free elements of B by column. In the second panel, bars 1-20 are diagonal elements of V; 21-24 are the μ's, 25-28 are ϕ's and 29-32 are σ's, each in the factor volatility model; 33-102 are the free elements of B by column. In the third panel, bars 1-20 are the μ's, 21-40 are ϕ's and 41-60 are σ's, each in the error volatility model; 61-97 are the free elements of B by column. Finally, in the fourth panel, bars 1-20 are the μ's, 21-40 are ϕ's and 41-60 are σ's, each in the error volatility model; 61-66 are μ's, 67-72 are ϕ's and 73-78 are σ's, each in the factor volatility model; 79-177 are the free elements of B by column.

¿From this figure we can observe that the MCMC algorithm for each of the models mixes extremely well when fairly complicated and high dimensional models are fit. In the CF 2-factor model, all inefficiency factors are below fifteen while in the SVF 4-factor model, most inefficiency factors are below thirty with some values around 50. In the SVE 2-factor model that contains one hundred and fifty parameters, about 95% of the inefficiency factors are below 35. Finally, in the MSV 6-factor model that contains one hundred and ninety three parameters, only a few inefficiency factors exceed 40. In each case, it is worth noting that there is no material difference in the inefficiency factors of the parameters in the factor loading matrix (right-hand side of the bar charts) and those of the volatility parameters.

For comparative purposes we wish to emphasize that the procedure proposed by Pitt and Shephard (1999b), which is applied to only five series, yields inefficiency factors above 350 for the B parameters and above 500 for some of the volatility parameters with just one common factor. All in all, the efficiency gains of our proposed MCMC algorithms appear to be huge. Two other indicators (not reported but available upon request) support this conclusion. First, the autocorrelation function of the simulated draws decays quite rapidly, reaching values around 0.1 by lag 30 in almost all cases. Second, the acceptance rates of the Metropolis-Hastings step are extremely high, ranging from 70% in the worse cases to 85-95% in the average cases. These figures clearly illustrate the value of tailoring in the formulation of the proposal density.

Next we turn our attention to the relative support of each model from the data. For this purpose, we compute the model marginal likelihoods computed using (12). They are reported in Table 7. A first observation is that the extensions from the CF to richer structures are supported by the data. Most support is received by the MSV model class, regardless of the number of factors. Consider now the effect of increasing the number of factors within each model class. In both the CF and SVE model classes, two factors seem to be adequate. In the case of the SVF model and MSV models, four and six factors, respectively, yield the best fitting models.

Figure 2 provides the posterior box-plots, computed from the MCMC output, of factor

Model	Factors					
	2	3	4	5	6	7
CF	-6062.72	-6520.44	-6691.66	-7150.12	-7306.94	-8569.91
SVF	-5575.86	-5431.04	-5373.98	-5408.16	-5462.28	-5717.01
SVE	-2414.29	-2932.32	-3148.57	-3522.34	-3689.55	-4183.60
MSV	-1785.65	-1684.51	-1650.40	-1593.81	-1563.56	-1563.99

Table 7: Log (base 10) marginal likelihoods for multivariate models with the twenty exchange rates data.

loading matrix B. The first factor has a loading that is close to one on all exchange rates except for the ones marked 3, 6, 7, 9 and 19, which correspond to the Yen, Canadian dollar, New Zealand Dollar and the Hungarian Forint. This is not surprising as the first rate is normalized at one for the British Pound and so can be thought of as a European effect. The Australian Dollar and the Hungarian Forint are exceptions and seem to be in the wrong categories. The sixth factor is approximately zero except for two currencies (the Australian Dollar and the New Zealand Dollar). This is perhaps a consequence of the close connection between these two economies.

6.2.2 Filtered estimates of 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 correlation using contemporaneous information. We could derive the correlations through the filtered time varying covariance matrix

$$\Omega_{t|t-1} = E\left\{\Omega(h_t)|\mathcal{F}_{t-1}, \mathcal{M}\right\} = E\left(V_t|\mathcal{F}_{t-1}, \mathcal{M}\right) + B\left\{E\left(D_t|\mathcal{F}_{t-1}, \mathcal{M}\right)\right\}B'.$$

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

$$R_{t|t-1} = E\left\{\Psi(h_t)|\mathcal{F}_{t-1}, \mathcal{M}\right\},$$

where

$$\Psi(h_t) = diag\left\{\Omega(h_t)\right\}^{-1/2}\Omega(h_t)diag\left\{\Omega(h_t)\right\}^{-1/2},$$

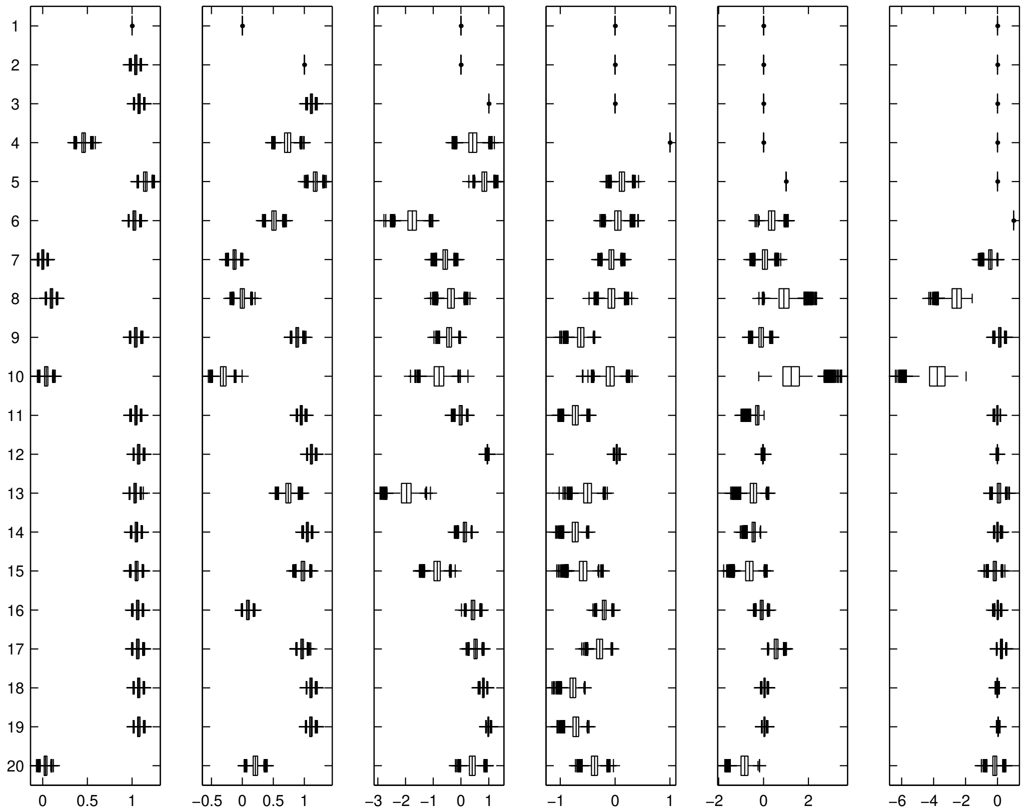

Figure 2: Exchange rate data: posterior boxplots of the ninety nine unrestricted factor loadings in the MSV six factor model arranged by each column of the B matrix. The constrained loadings are also included in each column.

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

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

where $h_t^{(1)},...,h_t^{(M)}$ are the output from the particle filter for the density $f(h_t|\mathcal{F}_{t-1},\mathcal{M})$. One can also study the correlation between assets over s-periods in the future

$$Cor\left(\sum_{j=1}^{s} y_{t+j}, \sum_{j=1}^{s} y'_{t+j}|\mathcal{F}_t,\theta^*,\mathcal{M}\right) = E\left\{\Psi(h_{t+1},...,h_{t+s})|\mathcal{F}_t,\theta^*,\mathcal{M}\right\},$$

where $\Psi(h_{t+1},...,h_{t+s})$ equals

$$diag\left\{\Omega^*(h_{t+1},...,h_{t+s})\right\}^{-1/2}\Omega^*(h_{t+1},...,h_{t+s})diag\left\{\Omega^*(h_{t+1},...,h_{t+s})\right\}^{-1/2}.$$

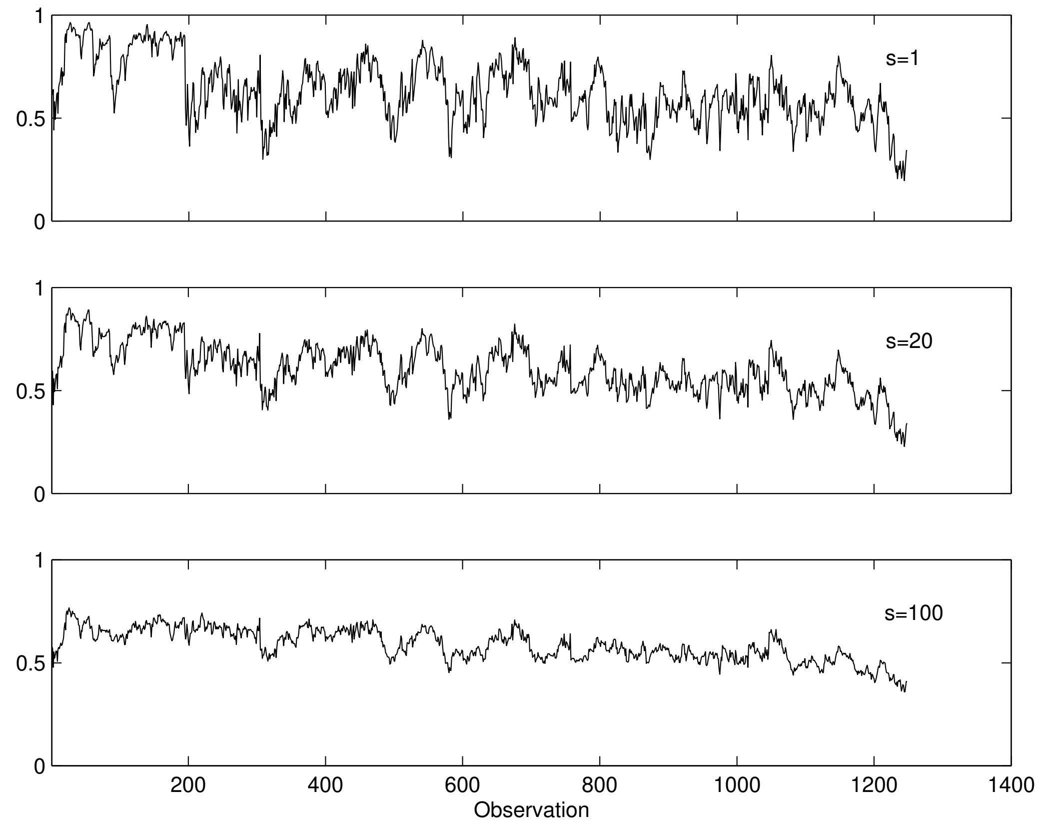

Figure 3: Foreign exchange data: filtered correlation from MSV six factor model between British Pound and Italian Lira for $s = 1, 20, 100$.

Given the output from the particle filter of particles $h_t^1, ..., h_t^M$, one can propagate each of these particles forward using their autoregressive structure (3) to sample $f(h_{t+1}, ..., h_{t+s}|\mathcal{F}_t, \theta^*, \mathcal{M})$. This then allows one to estimate the correlation matrix as

$$\frac{1}{M}\sum_{g=1}^{M}\Psi(h_{t+1}^{(g)}, ..., h_{t+s}^{(g)})$$

Clearly as $s \to \infty$ the correlation will converge to the unconditional correlation structure of the model so long as each of the SV processes is stationary.

Figure 3 shows the time-varying s-period conditional correlation of the Pound with the Italian Lira. We have graphed the correlations for $s = 1, 20, 100$ throughout the sample. This shows the correlation is most jagged for the short term forecast, whereas for the 100 steps ahead the outcomes are more smooth.

6.3 Equity return data

6.3.1 Estimation results

To further show how the proposed approach can be easily and efficiently applied to large dimensional problems, the MSV model is fit to 40 stock return series with the number of common factors ranging from 2 to 10. These choices imply that the smallest model contains 203 parameters while the largest contains 495 parameters. In addition, we compare the SVF model to the MSV model, for factors varying from two to ten.

In figure 4 we report the inefficiency factors for the MSV model. In all four panels, bars 1-40 are μ's, 41-80 are ϕ's and 81-120 are σ's, each in the error volatility model. In the first panel, bars 121-122 are the μ's, 123-124 are ϕ's and 125-126 are σ's, each in the factor volatility model; bars 127-203 are the free elements of B by column. In the second panel, bars 121-124 are the μ 's, 125-128 are ϕ's and 129-132 are σ's, each in the factor volatility model; 133-282 are the free elements of B by column. In the third panel, bars 121-126 are the μ's, 127-132 are ϕ's and 133-138 are σ's, each in the factor volatility model; 139-358 are free elements of B by column. Finally, in the fourth panel, bars 121-128 are the μ's, 129-136 are ϕ's and 137-144 are σ's, each in the factor volatility model; 145-428 are the free elements of B by column.

As can be seen from this figure, for the volatility parameters (left-hand side of the bar charts), most of the inefficiency measures are between ten and forty, with some occasional jumps between sixty and eighty. The mixing of the algorithm for those parameters seems to remain fairly stable as the number of common factors in the model grows. As for the parameters of the factor loading matrix (right-hand side of the bar charts), the simulation diagnostics are even better. In particular, notice how the inefficiency factors get lower for better fitting models (from Table 8 below). For example, in the model with six common factors and 219 free parameters in the B matrix, there are less than ten inefficiency factors above forty with the rest mostly below twenty.

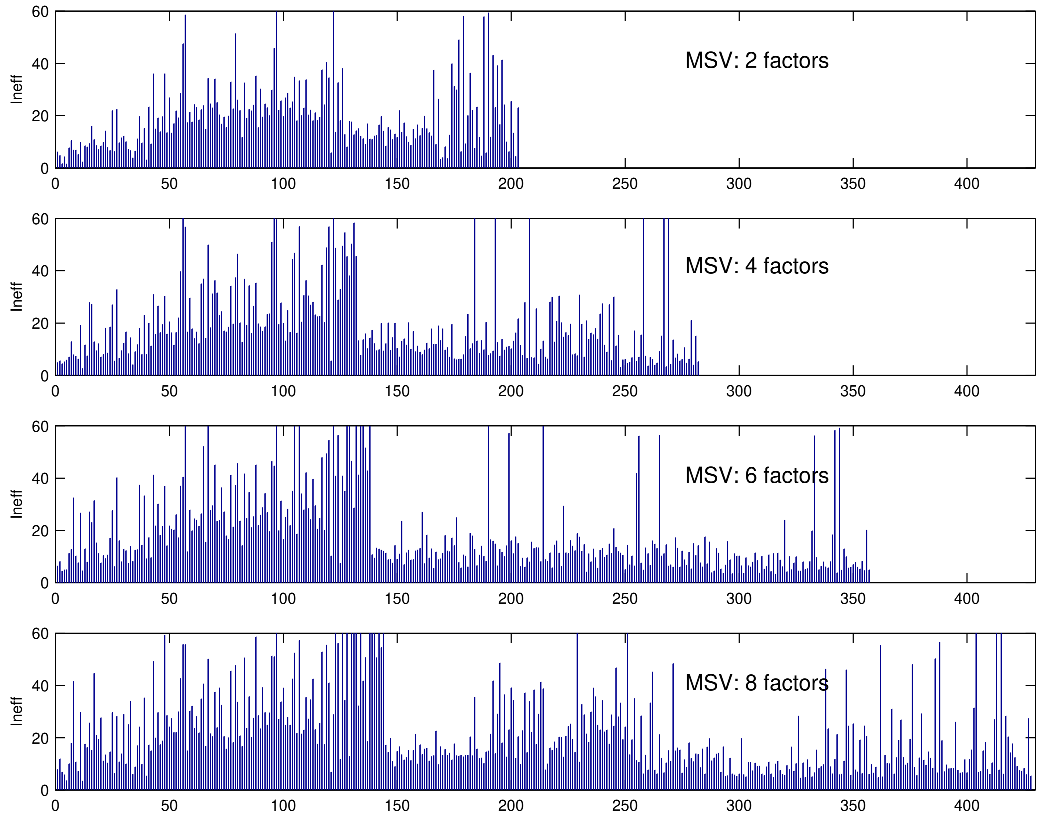

Figure 4: Inefficiency factors for MSV models with two, four, six and eight factors in the forty stocks data series.

This reflects the remarkable efficiency of the MCMC algorithm developed in this paper.

Next, consider the model comparison measures in Table 8 . One can see that additional factors improve the model fit but the improvement gradually tapers off. We obtain the result that the best model overall is the 8-factor MSV model. This comparison also highlights the interesting fact that the MSV specification provides a better fit to the data than the simpler special cases of this model.

Figure 5 contains the posterior box plots of the B matrix arranged by column for the best fitting eight factor MSV model. This figure provides a useful posterior summary of the two hundred and eighty four parameters in the loading matrix. Although there are differences in these posterior distributions the spread of the distributions is relatively small.

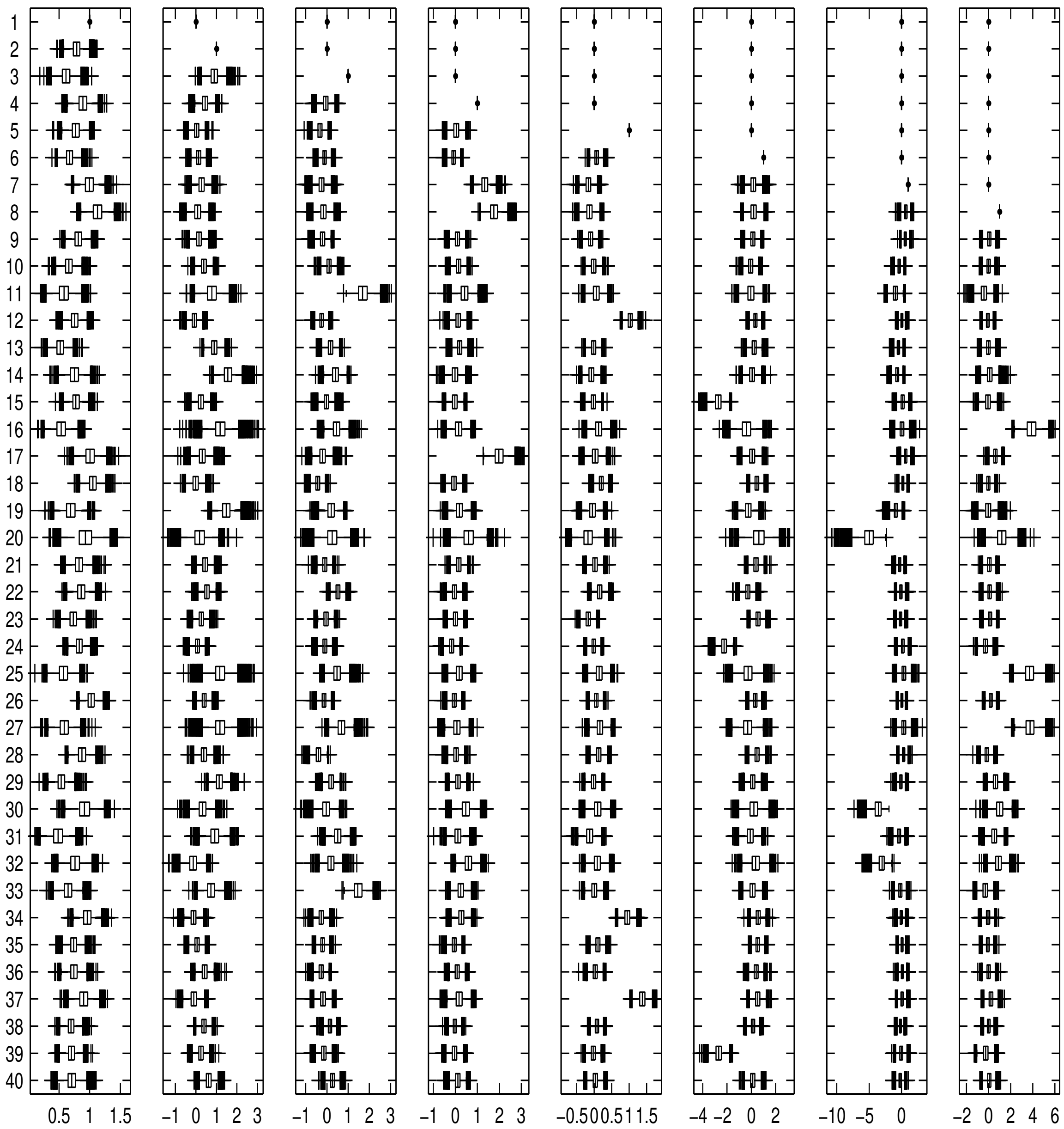

Figure 5: Stock return data: posterior boxplots of the two hundred and eighty four unrestricted factor loadings in the MSV eight factor model arranged by each column of the B matrix. The constrained loadings are also included in each column.

Model	Factors				
	2	4	6	8	10
SVF	-38236	-37930	-37806	-37375	-38335
MSV	-37510	-37274	-37162	-37074	-37163

Table 8: Log (base ten) marginal likelihoods of SVF and MSV models with different number of factors with the forty stocks data.

6.3.2 Filtered Correlations

In the exchange rate example we compute the s-period conditional correlations between two series of rates. Here we can look at the daily correlation between asset returns or between the asset and a market index. The correlation pattern clearly changes over time, with the correlations in both series being highest towards the end of the sample.

To illustrate, Figure 6 displays the correlations between IBM and McDonald's (MCD) and between IBM and United Airlines (UA) for every time period in the sample.

7 MODEL EXTENSIONS

7.1 Heavy tails

It has been noted by several authors that the distribution of returns on speculative assets is quite heavy tailed and that the normality assumption for returns is unreasonable. This is one reason for adopting the stochastic volatility structure since it leads to heavy tails, even under the Gaussian assumption on $\{v_{jt}\}$. Nevertheless, Kim, Shephard, and Chib (1998) and Chib, Nardari, and Shephard (1998) have found through a formal marginal likelihood/Bayes factor comparison that in the univariate case the student-t assumption on $\{v_{jt}\}$ is generally preferred to the Gaussian assumption. Such generalizations and comparisons can also be considered in the multivariate context by modifying the MCMC algorithms given in this paper along the lines of Chib, Nardari, and Shephard (1998). We do not take up this issue further in this paper partly because the required modifications are straightforward and partly because it is our view that the multivariate SV model, which contains additional volatility processes from the factor

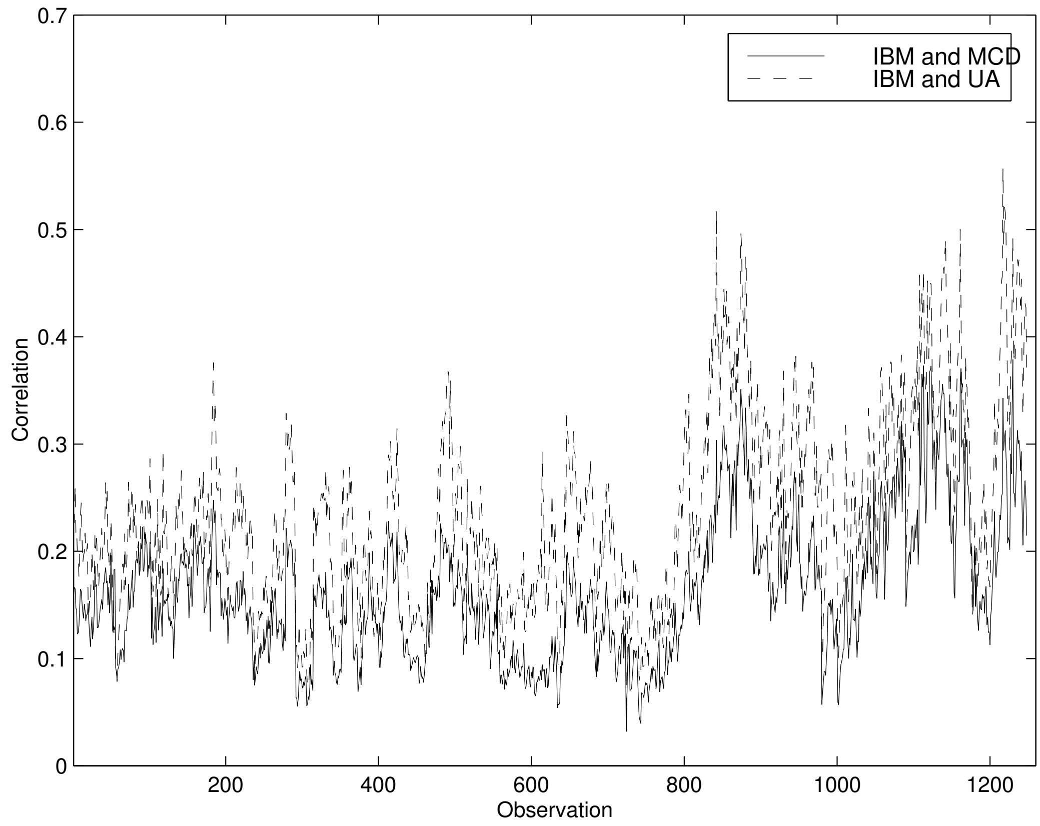

Figure 6: Stock return data: filtered correlation from MSV eight factor model between IBM and MCD; and between IBM and UA.

structure, leads to even thicker tails than is possible in the univariate case, thus mitigating the need for student-t innovations. We accept, of course, that this is ultimately an empirical issue that should be formally examined in subsequent work.

7.2 Leverage

A potential empirical failure of the FSV model for the equity data is the lack of a leverage term (see, for example, Nelson (1991)) in the FSV model. This is typically unimportant for exchange rates. We can include such an effect in a number of ways. One would be to correlated $\{u_{jt}, \eta_{jt}\}$ for each j, allowing the strength of the correlation to vary over j. Another would be to allow lagged values of $\{v_{jt}\}$ to appear in the volatility process. The latter was followed in

Chib, Nardari, and Shephard (1998) and this could be used in this context without difficulty.

7.3 Prior construction

In this paper we have assumed that the $\{\theta_j\}$ $(j \leq N+K)$ are exchangeable with each θ_j drawn from a common distribution $\pi(\theta_j) = \pi(\mu_j)\pi(\phi_j)\pi(\sigma_j^2)$. This prior construction is likely to be poorly behaved if the time series dimension T is small, but the number of cross-sections N is very large. This structure, which is familiar in panel data problems, suggests an alternative hierarchical prior structure where $\{\theta_j|\Theta\}$ are independently drawn. In turn, $\theta_j|\Theta$ now depends upon the common parameters Θ, which itself should be given a prior distribution. This implies a joint density on $\{\theta_1, ..., \theta_{N+K}\}$ which has dependence across the θ_j.

8 CONCLUSION

Our paper has set out methods for implementing a model structure suggested in Kim, Shephard, and Chib (1998) for analyzing high dimensional returns to holding financial assets such as equities and exchange rates. We have designed a tuned MCMC algorithm which delivers output which has little dependence over sweeps. Further, we have written out a filtering method which allows us to compute the likelihood function of the model. We have also extensively discussed the computation of the marginal likelihood and Bayes factors and illustrated this computation with both simulated and real data.

We conclude by emphasizing that the framework developed in this paper should allow one to more fruitfully analyze high frequency multivariate data. We anticipate that our results will lead to important applications in several areas, including portfolio analysis and risk exposure measurement.

References

Aguilar, O. and M. West (1998). Bayesian dynamic factor models and variance matrix discounting for portfolio allocation. Unpublished paper: 98-03 ISDS, Duke University.

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

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

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

Braun, P. A., D. B. Nelson, and A. M. Sunier (1995). Good news, bad news, volatility, and betas. *J. Finance 50*, 1575–1603.

Chib, S. (1995). Marginal likelihood from Gibbs output. *J. American Statistical Association 90*, 1313–21.

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

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

Chib, S. and E. Greenberg (1998). Analysis of multivariate Probit models. *Biometrika 85*, 347–61.

Chib, S., F. Nardari, and N. Shephard (1998). Markov chain Monte Carlo methods for generalized stochastic volatility models. Unpublished paper: Nuffield College, Oxford.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–50.

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

Engle, R. F. and S. Kozicki (1993). Testing for common features (with discussion). *J. Business and Economic Statist. 11*, 369–395.

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

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

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

Harvey, A. C. (1981). *Time Series Models* (1 ed.). New York: Philip Allan.

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

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Rev. Economic Studies 61*, 247–64.

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

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Economic Studies 65*, 361–93.

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

Lin, W.-L. (1992). Alternative estimators for factor GARCH models — a Monte Carlo comparison. *J. Applied Econometrics 7*, 259–279.

Lo, A. and C. MacKinlay (1988). Stock prices do not follow random walks: evidence from a simple specification test. *Review of Financial Studies 1*, 41–66.

Lopes, H. F. and M. West (1999). Model uncertainty in factor analysis. Unpublished paper: Duke University.

Magnus, J. R. and H. Neudecker (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* New York: Wiley.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica 59*, 347–370.

Pitt, M. K. and N. Shephard (1999a). Filtering via simulation based on auxiliary particle filters. *J. American Statistical Association 94*, 590–99.

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

Scholes, M. and J. Williams (1977). Estimating betas from nonsynchronous trading. *Journal of Financial Economics 5*, 309–27.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, and O. E. Barndorff-Nielson (Eds.), *Time Series Models in Econometrics, Finance and Other Fields*, pp. 1–67. London: Chapman & Hall.