A Multivariate Stochastic Volatility Model

Anastasios Plataniotis and Petros Dellaportas
tasosp@aueb.gr petros@aueb.gr
Department of Statistics, Athens University of Economics and Business, Greece

Summary: We introduce a broad class of multivariate stochastic volatility models where transformed eigenvalues and Givens rotation angles are assumed to be AR(1) processes. This decomposition retains the required positive definite structure of a fully time-varying covariance matrix. We provide detailed guidelines for a nested Laplace estimation procedure that exploits an efficient numerical maximization strategy. The methodology is implemented for a real financial dataset with 100 stocks and its forecasting power is compared to other existing multivariate volatility models. There is evidence that our models are superior in terms of both fitting and predicting to other specifications that model time-varying covariance matrices.

Some key words: Givens angles, MCMC, Nested Laplace Approximation, Spectral decomposition

1 Introduction

We aim to predict a sequence of high dimensional $N \times N$ covariance matrices $\{\Sigma_t\}$, $t = 1, \ldots, T$ of an N-dimensional zero mean, normally distributed, time series vector of returns r_t. This is an important problem in modern finance that has received a lot of attention; see for example Tsay (2005) . The major challenge emanates from the fact that Σ_t needs to be positive-definite and its number of parameters grows quadratically in N.

Our departure is the one-dimensional stochastic volatility model introduced by Taylor (1986) [17] which allows the variance of the observations to be an unobserved random process. The extension to the multivariate case is not immediate since the latent path of volatilities is converted to a path of vectors with entries the elements of a covariance matrix. Meyer and Yu (2006) discuss thoroughly some of the dominant models that have appeared in the multivariate volatility modeling literature.

The first key idea of our approach is to provide a class of multivariate stochastic volatility (MSV) models by modeling a series of univariate latent autoregressive models for the components of the Spectral decomposition of Σ_t. By invert transforming back to Σ_t the positive definiteness is immediately ensured. A further complication is that the elements of the orthonormal eigenvector matrices cannot be modeled as independent processes since they must obey the usual restrictions of orthogonality and unit length. We resolve this problem by transforming the elements of the eigenvector matrix to Givens rotation angles. Thus, for a N-dimensional vector of responses, we construct a MSV model with $N(N+1)/2$ latent autoregressive paths, N of those being the eigenvalues and $N(N-1)/2$ being the Givens angles.

Although the above model formulation facilitates the construction of latent processes ensuring the positive definiteness of Σ_t for every time t, the estimation process remains a difficult task since it requires the estimation of $O(N^2)$ latent paths. Although for small N, say less than 5, this may be computationally feasible via an MCMC algorithm, it is certainly completely prohibitive when N represents a portfolio of 100 stocks. For such cases, we adopt the recently proposed methodology of nested Laplace approximations introduced by Rue, Martino and Chopin (2008) . The required maximizations needed to obtain the required posterior summaries of interest are achieved through the following steps.

First, we perform a linear data transformation based on the spectral decomposition of the sample covariance matrix of series r_t, $t = 1, \ldots, T$. An immediate consequence of the fact that the transformed series are nearly uncorrelated is that when our model is applied to the new series the new Givens angles that describe the orientation of the new eigenvectors will be close to zero. This remark is useful for the efficient application of nested Laplace approximations for two reasons. First, when Givens angles are zero the parameters of our model are orthogonal and therefore their maximum likelihood estimates may be assumed to be independent. This ensures that the maximizations required for Laplace approximations can be achieved sequentially but non-iteratively. Second, this orthogonality allows us to break down the nested Laplace approximations for an N-dimensional problem to a series of nested Laplace approximations for 2-dimensional problems. This element of our algorithm not only greatly improves the efficiency of the algorithm but it has a great practical potential in real financial applications, since it allows the treatment of missing data in a natural fashion.

We first tested our algorithm in a small 5-dimensional problem. This allowed comparison of nested Laplace approximations with MCMC estimates. Having assured ourselves that our approximation strategy works well, we proceeded to a more realistic, 100-dimensional empirical study. There, we analyzed stocks of SP100 index and we produced parameter estimates and predictions of a multivariate stochastic volatility model with $N = 100$ stock returns. Other issues such as derivation of parsimonious modeling structures through the use of Bayes factors have been also investigated. We compared our results with the constant conditional correlation model of Bollerslev (1990) , the dynamic conditional correlation model of Engle (2002) , the orthogonal GARCH model of Alexander (2001) , one of the multivariate stochastic volatility models of Harvey et al. (1994) , the dynamic conditional correlation model of Meyer & Yu (2006) and an orthogonal GARCH model in which the GARCH part has been replaced with a stochastic volatility process. For comparative purposes, in many of these models we performed both the suggested in the literature frequentist estimation procedure together with the Bayesian analysis analogue. This required the use of some new MCMC algorithms. Our comparative empirical analysis gave evidence that our proposed model specification outperforms a large collection of existing time-varying covariance matrix modeling perspectives both in fitting and predicting.

The outline of the paper is as follows. In Section 2 we present a short description of the competing models in the literature and we introduce our model. In Section 3 we present the inference implementation via the nested Laplace approximation which is tested extensively in a small dataset in Section 4. Section 5 presents a large empirical study and Section 6 concludes with a small discussion.

2 A Multivariate Stochastic Volatility Model

Assume that we deal with $N \times 1$ vector of returns $r_t = \big(r_{1t}, \ldots, r_{Nt}\big)'$, $t = 1, \ldots, T$, with corresponding covariance matrices Σ_t. Assume further that r_t are second-order stationary so $E(\Sigma_t) = \Sigma$ exists. Historically, there have been many efforts to produce parsimonious structures that capture the dynamic behavior of Σ_t. A recent detailed review can be found in Meyer and Yu (2006) and Asai, McAleer and Yu (2006) .

The sample covariance matrix of the returns S can be decomposed to its spectral decomposition $S = PQP'$ with P and Q denoting the eigenvector and eigenvalues matrices respectively. Let $Z = PQ^{1/2}$ and apply the transformation

$$y_t = Zr_t \tag{1}$$

We propose modeling the transformed returns as $y_t \sim N(0, V_t)$ so that $V_t = Z\Sigma_t Z'$. By decomposing V_t into its spectral decomposition we obtain $V_t = U_t \Lambda_t U_t^T$ and therefore its $N(N+1)/2$ time-changing entries are decomposed to N parameters in the eigenvalue matrices Λ_t and $\widetilde{N} := N(N-1)/2$ parameters in the eigenvector matrices U_t. Furthermore, U_t is an orthonormal matrix so it can be written as a product of Givens rotation matrices

$$U_t = \prod_{j=1}^{\widetilde{N}} G_j(\omega_{jt}) \tag{2}$$

where the elements of each Givens matrix $G_j(\omega_{jt})$ are given by

$$G_j[k,l] = \begin{cases} \cos(\omega_{jt}), & \text{if } k = l = m_1(j) \text{ or } k = l = m_2(j) \\ \sin(\omega_{jt}), & \text{if } k = m_1(j), l = m_2(j) \\ -\sin(\omega_{jt}), & \text{if } k = m_2(j), l = m_1(j) \\ 1, & \text{if } k = l \neq m_1(j), \text{ and } k = l \neq m_2(j) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

and $m_1(j) < m_2(j)$ are integer values in $\{1, \dots, N\}$ satisfying $j = m_2(j) - m_1(j) + (m_1(j) - 1)(N - m_1(j)/2)$; see, for example, Hoffman et al. (1972) . To ensure uniqueness of (2) we must impose ordering of the eigenvalues and $\omega_{jt} \in (-\frac{\pi}{2}, \frac{\pi}{2})$, $j = 1, \dots, \widetilde{N}$. The main advantage of this parameterization is that Givens angles do not need to obey the usual orthonormality constraints of the elements of U_t.

The importance of transformation (1) is that although S may be a poor estimate of Σ, the y_t will till be nearly orthogonal and therefore the Givens angles that describe the rotation to orthogonality will be close to zero. This observation is of crucial importance in our estimation algorithm in Section (3).

Before we introduce our model we transform the rotation angles to near-normality via

$$\delta_{jt} = \log\left(\frac{\frac{\pi}{2}+\omega_{jt}}{\frac{\pi}{2}-\omega_{jt}}\right),\ j=1,\dots,\tilde{N},\ t=1,\dots,T, \tag{4}$$

and set h_{it}, $i=1,\dots,N$, as the log-eigenvalues of V_t. Then, our proposed MSV model is written as

$$\begin{aligned}
h_{i,t+1} &= h_{i0} + \phi_i^h \cdot (h_{it} - h_{i0}) + \sigma_i^h \cdot \eta_{it}^h,\ i=1,\dots,N,\ t=1,\dots,T-1 && (5)\\
\delta_{j,t+1} &= \delta_{j0} + \phi_j^\delta \cdot (\delta_{jt} - \delta_{j0}) + \sigma_j^\delta \cdot \eta_{jt}^\delta,\ j=1,\dots,\tilde{N},\ t=1,\dots,T-1 && (6)\\
h_{i1} &\sim N\left(h_{i0}, \frac{(\sigma_i^h)^2}{1-(\phi_i^h)^2}\right) && (7)\\
\delta_{j1} &\sim N\left(\delta_{j0}, \frac{(\sigma_j^\delta)^2}{1-(\phi_j^\delta)^2}\right) && (8)
\end{aligned}$$

where $\left|\phi_i^h\right| < 1$ and $\left|\phi_j^\delta\right| < 1$ denote persistence of each AR process, σ_i^h and σ_j^δ are corresponding volatilities and $\eta_{it}^h, \eta_{jt}^\delta \sim N(0,1)$ independently. Thus, the parameter vectors that need to be estimated are the transformed Givens angles and eigenvalues δ_t, h_t, for $t-1,\dots,T$, and the latent paths parameters

$$\theta_h = (\phi_1^h, \dots, \phi_N^h, h_{10}, \dots, h_{N0}, \sigma_{1\eta}^h, \dots, \sigma_{N\eta}^h)$$

and

$$\theta_\delta = (\phi_1^\delta, \dots, \phi_{\tilde{N}}^\delta, \delta_{10}, \dots, \delta_{\tilde{N}0}, \sigma_{1\eta}^\delta, \dots, \sigma_{\tilde{N}\eta}^\delta)$$

related to transformed eigenvalues and Givens angles respectively.

3 Estimation

Estimation of the models (5)-(8) via MCMC is based on the algorithm of Kim, Shephard and Chib (1998) where each of the latent paths in (5) and (6) conditional on every other path can be viewed as the parameter driven path of a univariate stochastic volatility model. Since there is typically high persistence in the angles and σ_j^δ is small, the MCMC mixing is poor, resulting in an autocorrelation function that decays to zero at about 400 iterations. For a five-dimensional example which has fifteen latent paths 18 hours were needed for 170.000 iterations in a standard PC with a Pentium 2.8GhZ processor, for a program written in Ox version 6.10; see Doornik (2007) .

A recent Bayesian implementation strategy that is appropriate for univariate stochastic volatility models is the nested Laplace approximation introduced by Rue, Martino and Chopin (2008) . We present here how this methodology can be adopted for our model (5)-(8) even when N is large. There are two key elements in our proposed algorithm. First, we exploit the fact that because of (1) the Fisher information matrix is close to diagonal. To see this, we examine each element (k, ℓ) of the Fisher information matrix of the full likelihood function $L(y|\psi)$ conditional on the vector of the corresponding values $\psi_t = (\delta_t, h_t, \theta_\delta, \theta_h)$ corresponding to the covariance matrices $(\Sigma_1, \Sigma_2, \ldots, \Sigma_T)$, as the average,

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2}{\partial\psi_t(k)\partial\psi_t(\ell)}\log L(y|\psi)\Big|_{\psi=\psi_t}.$$

Here we write y for all data and denote the $k-th$ element of the parameter vector ψ_t as $\psi_t(k)$. Since the joint process (ψ_t, y_t) is ergodic, by the law of large numbers the above average converges to

$$E[\frac{\partial^2}{\partial\psi(k)\partial\psi(\ell)}\log L(y|\psi)]. \tag{9}$$

For the part of ψ that contains only the vectors of eigenvalues and Givens angles, $\psi_t = (\delta_t, h_t)$, the resulting part in (9) is the usual Fisher information matrix obtained in Daniels and Kass (1999) for the static case where y_t is an independent observation from a multivariate normal density $N(0, \Sigma)$. Daniels and Kass (1999) have shown that when the Givens angles, and consequently the transformed angles via (4), are zero, the Fisher information matrix is diagonal. The immediate consequences of this is that the parameters δ_t and h_t are orthogonal in the sense of Cox and Reid (1987) , so the maximum likelihood values $\hat{\delta}_t$ and $\hat{h}_t$ are asymptotically independent and, most importantly for our work, there are important simplifications in their numerical calculation. As Cox and Reid (1987) note, if two parameters δ and h are orthogonal, then the maximum likelihood estimate of δ conditional on h varies only slowly with h. Now recall that due to the transformation (1), the Givens angles are close to zero so these non-diagonal elements would be close to zero. Therefore, a computational feasible strategy when trying to maximize ψ is to maximize the corresponding latent paths of Givens angles and eigenvalues sequentially. Moreover, if the latent paths maximum likelihood estimates are independent, their corresponding θ_δ, θ_h maximum likelihood estimates of parameter vectors will also be independent. This asymptotic independence is a key ingredient in the nested Laplace approximation strategy of next Section.

Another computational result that we exploit from the transformation (1) is the following. Assume that we deal with the N-variate MSV model and without loss of generality assume that the data are ordered such that the first two stocks (y_{1t}, y_{2t}) correspond to the eigenvalue processes h_1, h_2 and to the Givens angle process δ_1. Denote by h_{-12} and δ_{-1} all other latent processes. Now note that the log-likelihood function is written as

$$
\begin{aligned}
\log L(y|h_1, h_2, h^-, \delta_1, \delta^-) &= \sum_{t=1}^{T} \left(-\frac{N}{2} \log |V_t| - \frac{1}{2} y_t' V_t^{-1} y_t \right) + constant \\
&= \sum_{t=1}^{T} \left(-\frac{N}{2} \sum_{i=1}^{N} h_{it} - \frac{1}{2} y' G_{1t} \cdots G_{\tilde{N}t} \Lambda_t^{-1} G_{\tilde{N}t} \cdots G_{1t} y_t \right) + constant \\
&= \sum_{t=1}^{T} \left(-\frac{N}{2} \sum_{i=1}^{N} h_{it} - \frac{1}{2} y_t' G_{1t} G_{-1t} \Lambda_t^{-1} G_{-1t}' G_{1t}' y_t \right) + constant
\end{aligned}
$$

The matrix G_{-1t} is a product of $\tilde{N} - 1$ matrices that correspond to the vector of Givens angles processes δ_{-1} . Consequently, when the Givens angles are zero, G_{-1t} is the identity matrix. Therefore, by assuming that all δ_{-12} latent paths are zero,

$$
\log L(y|h_1, h_2, h^-, \delta_1) = \sum_{t=1}^{T} \left(-\frac{N}{2} \sum_{i=1}^{N} h_{it} - \frac{1}{2} y_t' G_{1t} \Lambda_t^{-1} G_{1t}' y_t \right) + constant
$$

which, after some simple algebra, becomes

$$
\log L(y|h_1, h_2, h_{-12}, \delta_1, \delta_{-12}) \approx \log L(y_1, y_2|h_1, h_2, \delta_1) + f(h_{-12}, y_{-12}) + constant
$$

where f is a function that depends only on h_{-12}, y_{-12}. The last approximation holds only because of the transformation (1) which results to Givens angles being close to zero and suggests that the maximization of each Givens angle path δ_1 can be achieved by maximizing δ_1 in the bivariate stochastic volatility model specified by data (y_{1t}, y_{2t}).

This last observation is not only useful since it brakes down a multivariate inference problem to a series of bivariate, more computationally feasible problems. It also provides a crucial practical advantage when some stocks have missing data: their inference affects only the bivariate models they are involved, thus allowing the exploitation of stock data with different time horizons. This practical issue is normally a big obstacle in the widespread use of other multivariate time varying volatility models.

3.1 Nested Laplace Approximation

Consider the bivariate MSV model with data $y_t = (y_{1t}, y_{2t})'$, obtained from transformation (1), and three latent processes h_{1t}, h_{2t} and δ_{1t} corresponding to two eigenvalues and one Givens angle. Denote the vector of all processes as $x = (h_{1t}, h_{2t}, \delta_{1t})$ the corresponding process parameters as $\theta = (\phi_{h_1}, \sigma_{h_1}, \phi_{h_2}, \sigma_{h_2}, \phi_{\delta_1}, \sigma_{\delta_1})$. This setup is a generalization of the univariate stochastic volatility model analyzed by Rue, Martino and Chopin (2008) , so we follow exactly their methodology for the bivariate case. In particular, the elements of θ are transformed so that they take values in the real line and normal and gamma priors are adopted with the parameters used in Rue, Martino and Chopin (2008) .

Denoting with y all data, the posterior density can be written as

$$\pi(x, \theta|y) \propto \pi(\theta)\pi(x|\theta) \prod_{t \in T} \pi(y_t|x_t, \theta) \propto \pi(\theta) \left|Q(\theta)\right|^{1/2} \exp\Big(-\frac{1}{2} x^T Q(\theta) x + \sum_{t \in T} \log \pi(y_t|x_t, \theta)\Big)$$

where $Q(\theta)$ is the precision matrix of the zero-mean normal density $\pi(x|\theta, y)$. Since the dimension of θ is still small, and given that sequential maximizations of latent paths is possible, the basic steps of the estimating procedure we propose are as follows. First, for a given θ, we obtain the mode $x^*(\theta)$ for the full conditional density $\pi(x|\theta, y)$ using the nested Laplace approximation. This is achieved by first sequentially maximizing the two eigenvalue paths assuming zero angles for all t and then maximizing the angle path conditional on the maximized eigenvalue paths. Then, a normal approximation $\tilde{\pi}_G(x|\theta, y)$ to $\pi(x|\theta, y)$ is obtained, and the marginal posterior of θ, $\pi(\theta|y)$, is approximated by

$$\tilde{\pi}(\theta|y) \propto \left. \frac{\pi(x,\theta,y)}{\tilde{\pi}_G(x|\theta,y)} \right|_{x=x^*(\theta)}.$$

We then obtain numerical estimates for the mode and Hessian of $\tilde{\pi}(\theta|y)$ and explore it by using a set of about eighty abscissa. The curse of dimensionality in the number of abscissa is avoided by noting that the main mass of the six-dimensional density lies in the two-dimensional subspaces of $(\phi_{h_1},\sigma_{h_1})$,$(\phi_{h_2},\sigma_{h_2})$, and $(\phi_{\delta_1},\sigma_{\delta_1})$. Therefore only lattices on these sub-spaces are applied. The above algorithm is performed for all $\tilde{N}$ pairs of stocks and all corresponding $\tilde{N}$ maximized paths $\hat{\delta}_t$. Then, the matrix

$$\hat{G}_t = \hat{G}_{1t}\hat{G}_{2t}\cdots\hat{G}_{\tilde{N}t}$$

is constructed so that a new transformation

$$w_t = \hat{G}_t y_t$$

creates a series of N independent univariate stochastic volatility models with log-volatility paths h_t, that are readily estimated using the exact univariate nested Laplace algorithm of Rue, Martino and Chopin (2008) .

The total computational effort consists of a nested Laplace estimation of $\tilde{N}$ bivariate and N univariate stochastic volatility models. For a program written in Ox version 6.10, see Doornik (2007) , these calculations take about 20 minutes for $N = 5$ and 18 hours for $N = 100$ in a standard PC with a Pentium 2.8GhZ processor. Therefore the reduction in terms of computing time is huge relative to the loss of accuracy if some exists.

3.2 Bayes Factors

The integrations achieved through the nested Laplace approximations allow the easy calculation of marginal likelihoods and therefore Bayes factors. Competing models to (5)-(8) that may be useful in the context of large-N portfolio problems are those that offer parsimony by reducing the number of latent paths. Although standard multivariate statistics literature is concerned with ways with which parameter space is reduced by removing smaller eigenvalues, in our finance context this is not appropriate since small eigenvalues correspond to less volatile stocks which are the most liquid portfolio ingredients. Therefore, we suggest considering competing models in which some Givens angles are not modeled as latent processes but assumed to be constant over time. For every angle δ_j, $j = 1, \dots, \tilde{N}$, the competing model of (6) is the model in which δ_j is a parameter with prior density taken as led as latent processes but assumed to be constant over time. For every angle δ_j, $j = 1, \dots, \tilde{N}$, the competing model of (6) is the model in which δ_j is a parameter with prior density taken as

$$\delta_j \sim N(0, \tau). \tag{10}$$

To avoid Lindley's paradox the prior variance τ should be chosen carefully; we follow Kass and Wasserman (1996) and obtain τ so that the prior approximates the unit information prior. A simple way to obtain its value is to set the unit information prior obtained from the unconditional covariance matrix S,

$$\tau = \frac{q_k q_l}{(q_k - q_l)^2},$$

where k, l are related to the angle j as in (3) and q_k, q_l are the respective sample eigenvalues obtained from the inverted value of the Fisher information matrix when the estimated value of the angle is zero as in Daniels (1999) .

For each model that adopts (10) instead of (6), the estimation procedure of Section 3.1 is repeated and all related estimates of posterior densities are calculated. Then, marginal likelihoods and Bayes factors are calculated by integrating out θ in $\tilde{\pi}(\theta|y)$.

3.3 Prediction

We focus here in the cases where interest lies in predicting Σ_{t+1} using information up to time t. Since $\Sigma_t = ZV_tZ'$, it is adequate to calculate

$$\pi(V_{t+1}|y) = |J|\pi(h_{t+1}, \delta_{t+1}|y) = |J| \int \pi(h_{t+1}, \delta_{t+1}|h_t, \delta_t, \theta)\pi(h_t, \delta_t|\theta, y)d\delta_t dh_t d\theta \qquad (11)$$

where the Jacobian J is given by

$$J = \frac{\prod_{k=1}^{N} \prod_{\ell=1}^{N} \cos(\omega_{k\ell(t+1)})^{k-\ell-1}}{\prod_{k<\ell} \prod_{\ell=1}^{N} (\lambda_{k(t+1)} - \lambda_{\ell(t+1)})};$$

see Daniels (1999) . Marginalization over θ is achieved exactly as in Rue, Martino and Chopin (2008) by exploiting the abscissas used to obtain $\pi(x|y)$. Then, we obtain the mode $\hat{h}_t^i, \hat{\delta}_t^i$ of $\pi(h_t, \delta_t|y)$ that are used as plug-in estimates to approximate the mode of $\pi(h_{t+1}, \delta_{t+1}|\hat{h}_t^i, \hat{\delta}_t^i, y)$ and consequently of $\hat{V}_{t+1}$.

4 Empirical Studies

4.1 Competing Models

We describe here the formulation and inference details of the models used for comparison in our empirical studies. An early attempt to model time varying covariance matrices is the constant conditional correlation (CCC) model of Bollerslev (1990) , in which return shocks are allowed to be correlated constantly across time, so $\Sigma_t = B_t R B_t$ with B_t being the diagonal matrix with elements B_{it}, $i = 1, \ldots, N$, and R denoting the constant correlation matrix. By setting $b_{it} = \log B_{it}$ and assuming an autoregressive process for b_{it} we obtain a subclass of the constant conditional correlation stochastic volatility model introduced by Harvey *et al.*(1994) . Alexander (2001) proposed the orthogonal GARCH (O-GARCH) model in which only the eigenvalues are time-varying whereas the eigenvectors are static. Its stochastic volatility analogue just replaces the eigenvalues GARCH univariate processes with a stochastic volatility. A recent model that has appeared in the literature was the dynamic conditional correlation (DCC) model of Engle, (2002) where the correlation matrix is assumed to be time changing resembling a GARCH-type process. Meyer and Yu (2006) proposed a stochastic volatility analogue to the DCC-GARCH family of models where the univariate volatility elements follow a stochastic volatility process.

Most of the above models can be generalized by altering the distributional form of the returns. A multivariate t-distributed CCC GARCH model (Heavy-GARCH) has been proposed by Bollerslev (1990) and a multivariate t-distributed DCC GARCH model has been discussed in Fiorentini (2003) . When incorporating these models we generalize them slightly by using different degrees of freedom of each t-distribution in each dimension.

4.2 Comparison Against a Proxy

We perform model comparisons based on both model fitness and model prediction. Since true covariance matrices are not known, these two issues are viewed as covariance matrix estimation comparisons against both in-sample and out-of-sample proxies. Such a proxy is borrowed from recent developments in the analysis of realized covariation (Andersen, Bollerslev and Lange (1999) ; Barndorff-Nielsen and Shephard (2004)) that suggest that a good estimate is the realized covariation matrix, calculated for a given day as the cumulative cross-products of intraday returns over each day. For our applications we use five minutes intra-day data. If an element of a covariance matrix σ_{ij} is estimated by a model with $\hat{\sigma}_{ij}$ and its corresponding proxy estimate is σ^*_{ij}, we use the following three discrepancy measures to test how competing models perform:

$$
\begin{aligned}
\text{Mean Absolute Deviation: MAD} &= N^{-2}\sum_{i,j} E|\sigma^*_{ij} - \sigma_{ij}| \\
\text{Root Mean Square Error : RMSE} &= \left[N^{-2}\sum_{i,j} E\left(\sigma^*_{ij} - \sigma_{ij}\right)^2 \right]^{1/2} \\
\text{Mean Determinant Loss : MDL} &= E\left[|\Sigma^*\Sigma^{-1} - I| \right]
\end{aligned}
$$

4.3 FTSE 100 Data

We obtained (source: BLOOMBERG) 100 daily returns of the index FTSE 100, recorded from 12/05/2008 up to 22/10/2009. We also obtained five-minute intraday data from the same returns for the following five days 23-29/10/2009. By selecting the alphabetically first stocks of the SP100 index we performed a comparison study by predicting covariance matrices with all competing models. The advantage of using a small group of stocks is that we can compare our nested Laplace estimates with the MCMC estimates in the MSV model we propose. Figure 1 presents some characteristics of the results of our empirical study based on a DCC model estimated with MCMC and MSV estimated with both MCMC and nested Laplace approximations. Figure 2 shows a typical 95% posterior region of estimated correlation between two stocks with MCMC and nested

Laplace. Clearly nested Laplace underestimates the posterior region. However, the time-varying correlations are rather close together, see Figure 3.

In Table 1 we report the ratio of the RMSE of several models over the RMSE of the saturated integrated Laplace model. Similar results based on MAD and MDL that are not included here to save space, can be found in Plataniotis (2011) . We performed the test with three sets of time periods, 169, 269 and 369 days. We observe that the MSV model outperforms both in terms of in-sample as well as out-of-sample fit most of the models in the lower dimension (5D) examples. In particular, the MCMC estimated MSV model does not differ significantly from the Laplace estimated MSV model. They both outperform other competing model specifications. Especially when we have a smaller number of observations (169) the Laplace model seems to be more robust than its competitors. In terms of out-of-sample forecasts, the results obtained via the nested Laplace approximation are consistently better. Note that the DCC model estimated via MCMC outperforms the usual maximum likelihood estimation results both in terms of in- and out-of-sample estimates in all cases.

The same experiment was replicated with all 100 stocks but without the inefficient MCMC algorithms. The results are reported in Table 2. The O-SV model and the CCC-SV model using the stochastic volatility tend to better fit in-sample the data than the MSV model, however when it comes to out-of-date forecasting, the MSV model clearly outperforms all its competitors. As for the Bayes factors, we see that there is significant reduction in the parameter space since most of the times we only use up to 90% of the latent angle paths.

We find these results to be very promising, especially because our simple rotation based on the sample eigenvector matrix that requires no significant extra computational cost performs well in terms of forecasting. We conjecture that in portfolios containing financial products more diverse than the observed returns the constant correlation model or the simple Orthogonal GARCH will fail to capture the empirical dynamics of all series. Moreover, we emphasize that one of the great advantages of our method is its ease of implementation in high dimensions. The only models that can be routinely applied in high dimensions are the DCC estimated via maximum likelihood and the Orthogonal GARCH models.

5 Conclusions

Our empirical work shows the great promise of the proposed MSV model in providing parsimonious models with good forecasting behavior for time-varying covariance matrices in finance. The adoption of nested Laplace approximations together with the initial data transformation provides an efficient way to obtain posterior summaries of interest when the dimension of the problem increases.

Acknowledgments

We are grateful to Ioannis Kontoyiannis for many interesting conversations related to this work.

References

[1] C. Alexander. *Market Models: A Guide to Financial Data Analysis.* Wiley, New York, 2001.

5 Dimensions

	In-sample			Out-of-sample		
Model	169 Days	269 Days	369 Days	169 Days	269 Days	369 Days
MSV-MCMC	1.028	1.009	0.993	1.038	1.046	1.011
DCC-MCMC	1.064	1.065	1.043	1.054	1.043	0.987
O-SV	1.029	1.010	0.994	1.038	1.049	1.034
DCC-SV	1.055	1.020	1.065	1.046	1.062	1.170
CCC-SV	1.023	1.008	0.994	1.032	1.062	1.023
O-GARCH	1.052	1.058	1.039	1.201	1.143	1.021
CCC-GARCH	1.049	1.061	1.041	1.136	1.157	1.018
CCC-GARCH-t	1.049	1.059	1.039	1.127	1.148	1.002
DCC-GARCH	1.081	1.075	1.106	1.166	1.157	1.152
DCC-GARCH-t	1.079	1.066	1.094	1.150	1.148	1.131
MSV-NL-1	1.012	1.016	1.013	1.001	0.985	1.008
MSV-NL-2	1.011	1.015	1.013	1.001	0.987	1.008

Table 1: Ratio of the RMSE of various models over the RMSE of the saturated MSV model estimated via the Gaussian approximation of the Laplace methodology for the five dimensional example. In-sample RMSE is estimated over the entire set of observations and out-of-sample is averaged over the 5 day ahead forecasts. *MSV*: Our Multivariate Stochastic Volatility model; *DCC*: dynamic conditional correlation structure assumed; *O*: Orthogonal; *CCC*: constant conditional correlation; *SV*: univariate stochastic volatility structure assumed; *MCMC*: indicates that this model is estimated via MCMC methodology; *t*: indicates that the assumed distribution of returns is the multivariate t. *MSV-NL-1*: Our proposed Multivariate Stochastic volatility model estimated with Nested Laplace approximation; *MSV-NL-2*: Our proposed Multivariate Stochastic volatility model estimated with Nested Laplace approximation with some angles being constant across time.

100 Dimensions						
	In-sample			Out-of-sample		
Model	169 Days	269 Days	369 Days	169 Days	269 Days	369 Days
O-SV	0.978	0.982	0.982	1.053	1.039	1.051
O-GARCH	1.000	1.009	1.009	1.086	1.069	1.057
CCC-SV	0.968	0.972	0.973	1.044	1.026	1.045
CCC-GARCH	1.001	1.016	1.015	1.066	1.046	1.042
CCC-GARCH-t	1.001	1.011	1.012	1.067	1.046	1.036
DCC-SV	1.020	1.032	0.973	1.050	1.036	1.045
DCC-GARCH	1.053	1.080	1.087	1.072	1.057	1.056
DCC-GARCH-t	1.018	1.032	1.012	1.067	1.046	1.036
MSV-NL-1	0.999	0.999	1.000	1.005	1.020	0.995
MSV-NL-2	0.998	1.002	0.996	1.000	1.021	0.992

Table 2: Ratio of the RMSE of various models over the RMSE of the saturated MSV model estimated via the Gaussian approximation of the Laplace methodology for the one hundred dimensional example. In-sample RMSE is estimated over the entire set of observations and out-of-sample is averaged over the 5 day ahead forecasts. *MSV*: Our Multivariate Stochastic Volatility model; *DCC*: dynamic conditional correlation structure assumed; *O*: Orthogonal; *CCC*: constant conditional correlation; *SV*: univariate stochastic volatility structure assumed; *MCMC*: indicates that this model is estimated via MCMC methodology; *t*: indicates that the assumed distribution of returns is the multivariate t. *MSV-NL-1*: Our proposed Multivariate Stochastic volatility model estimated with Nested Laplace approximation; *MSV-NL-2*: Our proposed Multivariate Stochastic volatility model estimated with Nested Laplace approximation with some angles being constant across time.

	5 Dimensions			100 Dimensions		
Measure	169 Days	269 Days	369 Days	169 Days	269 Days	369 Days
Number	8	9	10	4419	4384	4454
Total	10	10	10	4950	4950	4950

Table 3: Bayes Factors of the MSV model rotating angles. *Total* indicates the number of Given's angles used in the saturated model, *Number* indicates the number of Given's angles that remain time changing in the reduced model.

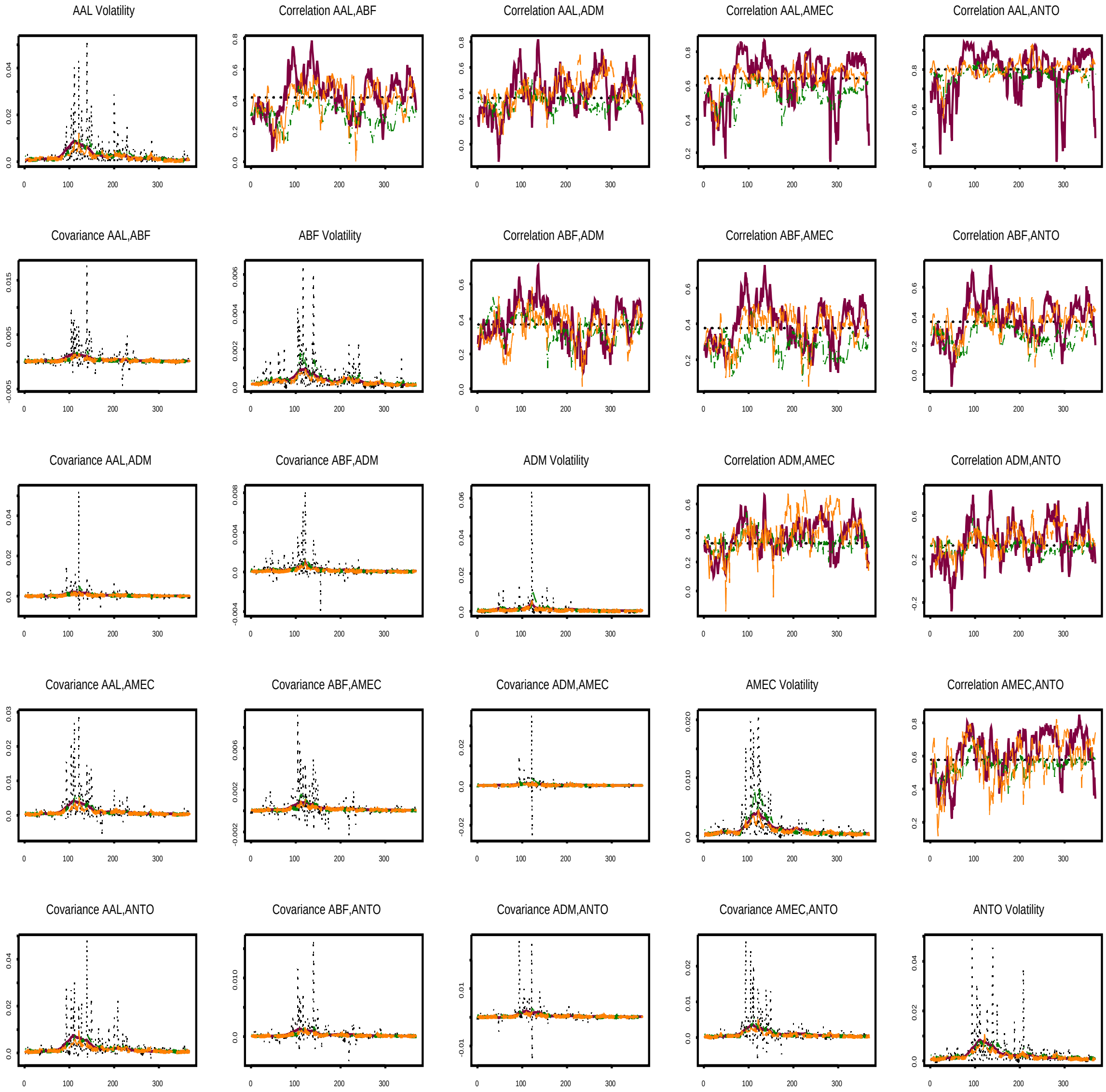

Figure 1: Plots for 5-dimensional returns. Diagonal plots show volatility estimates. Upper diagonal plots show correlations. Lower diagonal plots show covariances. *dotted*: Squared returns; *dashed*: MSV via MCMC; *solid*: DCC via MCMC; *thick solid*: MSV via Laplace

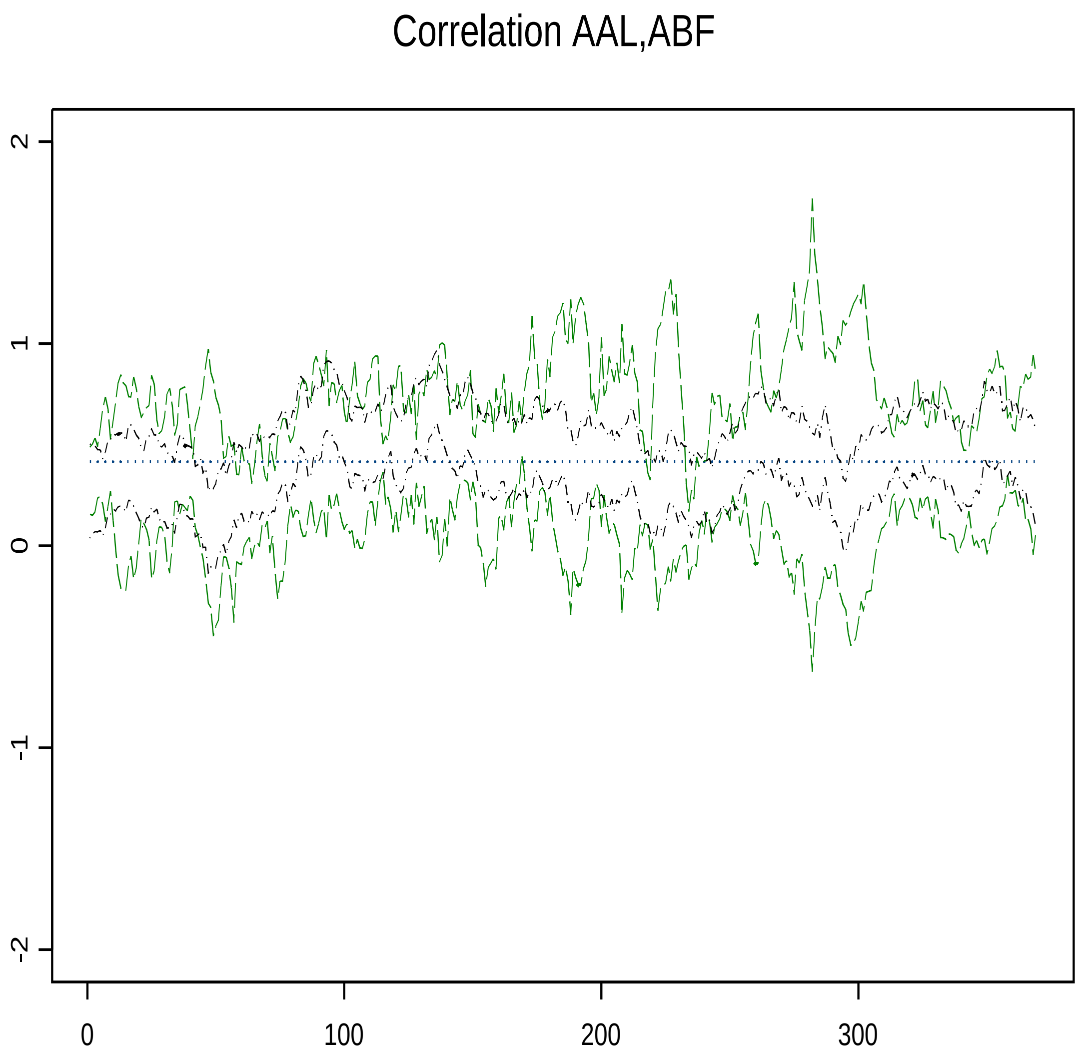

Figure 2: 95% Confidence Intervals using MCMC sampler and Laplace approximation. *dotted*: Sample Correlation; *dashed*: C.I. for MSV via MCMC; *solid*: C.I. for MSV via Laplace

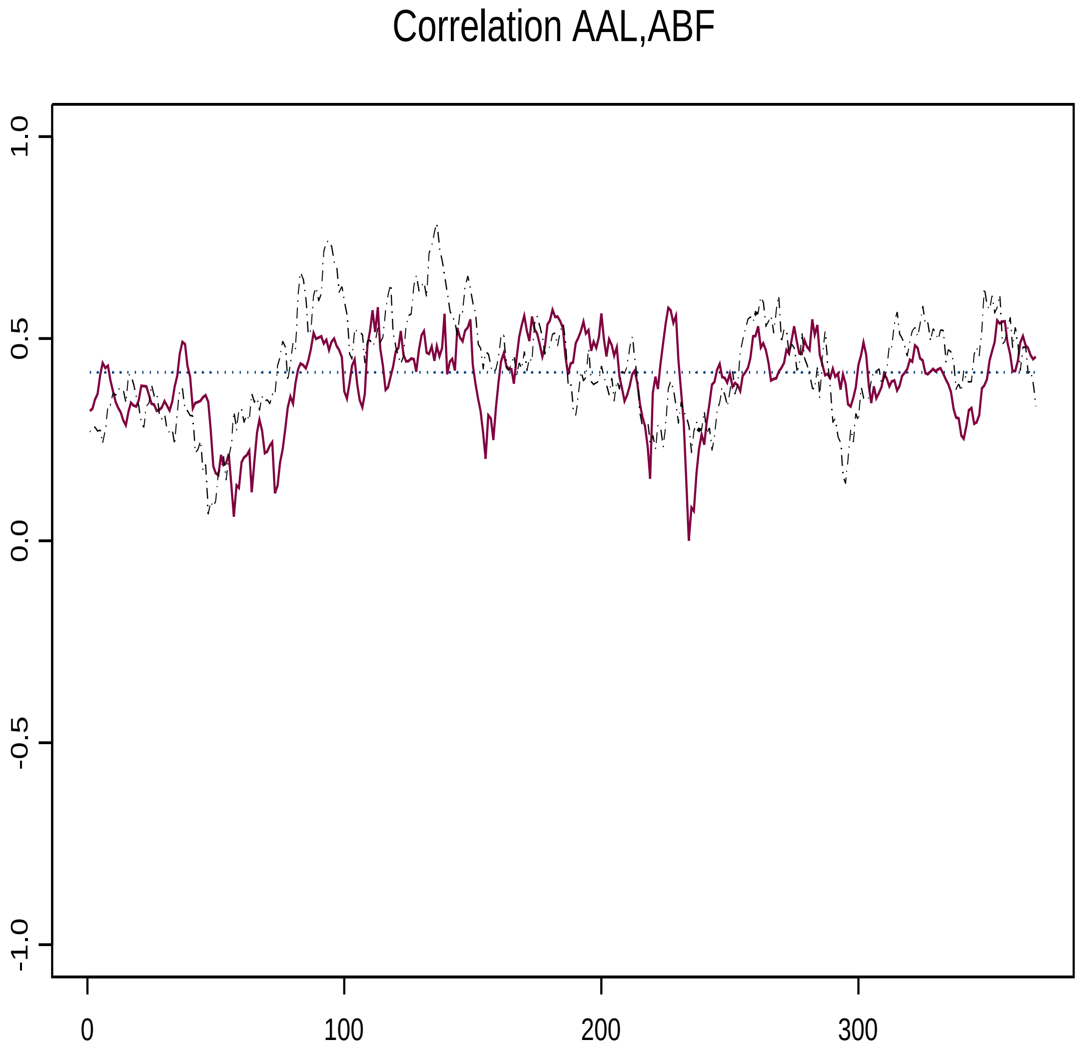

Figure 3: Correlation estimates for AAL,ABF. *dotted*: Sample Correlation; *dashed-dotted*: MSV via MCMC; *dashed*: MSV via Laplace.

[2] Plataniotis Anastasios. *High Dimensional Time-Varying Covariance Matrices With Applications in Finance*. PhD, Athens University of Economics and Business, 2011.

[3] T.G. Andersen, T. Bollerslev, and S. Lange. Forecasting financial market volatility: Sample frequency vis-à-vis forecast horizon. *Journal of Empirical Finance*, 6:457–477, 1999.

[4] M. Asai, M. McAleer, and Yu J. Multivariate stochastic volatility: A review. *Econometric Reviews*, 25:145–175, 2006.

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

[6] T. Bollerslev. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics*, 72:498–505, 1990.

[7] D.R. Cox and Reid. Parameter orthogonality and approximate conditional inference,(with discussion). *Journal of Royal Statistical Society series B*, pages –, 1987.

[8] M.J. Daniels and R.E. Kass. Nonconjugate bayesian estimation of covariance matrices and its use in hierarchical models. *Journal of the American Statistical Association*, 94:1254–1263, 1999.

[9] J.A Doornik. *An Introduction to OxMetrics 5*. London: Timberlake Consultants Press, 2007.

[10] R.F. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economics Statistics*, 20:339–350, 2002.

[11] G. Fiorentini and E. Sentana. Maximum likelihood estimation and inference in multivariate conditionally heteroscedastic dynamic regression models with student-t innovations. *Journal of Business and Economic Statistics*, 21:532–546, 2003.

[12] Ruiz E. Harvey, A.C. and N. Shephard. Multivariate stochastic variance models. *The Review of Economic Studies*, 61:247–264, 1994.

[13] D.K. Hoffman, Raffenetti R.C., and K. Ruedenberg. Generalization of euler angles to n dimensional orthogonal matrices. *Journal of Mathematical Physics*, 13:528–533, 1972.

[14] R. E. Kass and L. Wasserman. Formal rules of selecting prior distributions: A review and annotated bibliography. *Journal of the American Statistical Association*, 91:343–370, 1996.

[15] S. Kim, N. Shephard, and S. Chib. Stochastic volatility: Likelihood inference and comparison with arch models. *The Review of Economic Studies*, 65:361–393, 1998.

[16] H. Rue, S. Martino, and N. Chopin. Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations, discussion paper. *Journal of Royal Statistical Society series B*, pages –, 2008.

[17] S.J. Taylor. *Modelling Financial Time Series.* John Wiley, Chichester, 1986.

[18] R. Tsay. *Analysis of Financial Time Series.* John Wiley, New York, 2005.

[19] J. Yu and R. Meyer. Multivariate stochastic volatility models: Bayesian estimation and model comparison. *Econometric Reviews*, 25:361–384, 2006.