MCMC Estimation of Multiscale Stochastic Volatility Models

German Molina, Chuan-Hsiang Han and Jean-Pierre Fouque

Technical Report #2003-6
June 3, 2003

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Statistical and Applied Mathematical Sciences Institute
PO Box 14006
Research Triangle Park, NC 27709-4006
www.samsi.info

MCMC Estimation of Multiscale Stochastic Volatility Models

German Molina*, Chuan-Hsiang Han † and Jean-Pierre Fouque‡

June 2, 2003

Abstract

In this paper we propose to use Monte Carlo Markov Chain methods to estimate the parameters of Stochastic Volatility Models with several factors varying at different time scales. The originality of our approach, in contrast with classical one-factor models is the identification of well-separated time scales and the number of these. This is tested with simulated data as well as foreign exchange data.

Key words and phrases. Time scales in volatility, Bayesian estimation, MCMC, Foreign exchange.

1. Introduction

Modeling volatility and estimating volatility models from historical data has been extensively studied during the last 15 years. (Nelson, 1991) proposed an EGARCH model by using the maximum likelihood estimation. (Ruiz, 1994, Harvey *et al.*, 1994) proposed the Gaussian quasi-maximum likelihood estimation approach. Several papers are devoted to the Generalized Method of Moments (GMM), for example (Melino and Turnbull, 1990) and (Duffie and Singleton, 1993). We refer to (Bates, 1996) for a review on these directions. There is also a numerous amount of literature based on likelihood estimation through numerical integration, as in (Fridman and Harris, 1998) or Monte Carlo integration, as in (Sandmann and Koopman, 1998, Jacquier *et al.*, 1994, Kim *et al.*, 1998).

From the point of view of derivative pricing there is a huge literature on continuous time stochastic volatility modeling (see for instance (Fouque *et al.*, 2000a) and references therein). Volatility

*Department of Statistics, Duke University, Durham, NC 27708-0251, german@stat.duke.edu.

†Department of Mathematics, NC State University, Raleigh, NC 27695-8205, chan2@unity.ncsu.edu.

‡Department of Mathematics, NC State University, Raleigh, NC 27695-8205, fouque@math.ncsu.edu. Work partially supported by NSF grant DMS-0071744.

time scales can be retrieved from modern financial data at the level of minute by minute, daily, or monthly. Using low-frequency data, for example daily data over several years, allows a scale on the order of months to be estimated (Alizadeh *et al.*, 2002). Similarly using high-frequency intraday data, for example every 5 minutes over one year, a scale on the order of days emerges, as in (Fouque *et al.*, 2003), where variogram-based techniques are used to analyze the time scale content of the volatility.

Recently, two-factor stochastic volatility models have been shown to produce the observed kurtosis, fat-tailed return distribution and long memory effect. This is well-documented in (LeBaron, 2001), where actually three-factor models at three different time scales are used. (Alizadeh *et al.*, 2002) used ranged-based estimation to indicate the existence of two volatility factors including one highly persistent factor and one quickly mean-reverting factor. (Chernov *et al.*, 2002) used the Efficient Method of Moments (EMM) to calibrate multiple stochastic volatility factors and jump components. One of main results they found is that two factors are necessary for log-linear models.

In this paper we consider a class of stochastic volatility models where the volatility is the exponential of a sum of mean reverting diffusion processes (more precisely Ornstein-Uhlenbeck processes). Each of these mean reverting processes varies on well-separated time scales. We then rewrite these models in discrete time and use Markov Chain Monte Carlo (MCMC) techniques to estimate their parameters (Section 2). Our study is illustrated on simulated data (Section 3) where we address the issue of the choice of the number of factors and show that we can effectively determine the number of significant factors (i.e. with well-separated time scales). We also present (Section 4) an application of our technique to a set of foreign exchange data.

2. Multiscale Modeling and MCMC Estimation

2.1. Continuous Time Model

We first introduce our K factor volatilily model in continuous time. In this class of models the volatility is simply the exponential of the sum of K mean reverting O-U processes;

$$\begin{aligned} dS_t &= \kappa S_t dt + \sigma_t S_t dW_t^{(0)}, \\ \log(\sigma_t^2) &= \sum_{j=1}^{K} Y_t^{(j)} \\ dY_t^{(j)} &= \alpha_j(\mu_j - Y_t^{(j)})dt + \beta_j dW_t^{(j)},\ j = 1, \ldots, K \end{aligned} \tag{2.1}$$

here S is the asset price, κ is the constant rate of return and $W^{(0)}, W^{(1)}, \ldots, W^{(K)}$ are (possibly correlated) Brownian Motions. Each of the factors Y's is a mean reverting O-U process, where

μ_j denotes the long-run mean, α_j is the rate of mean reversion, and β_j is the "volatility of volatility" of the $Y^{(j)}$th factor. The typical time scale of this factor is given by $1/\alpha_j$. In our model these time scales are well-separated and $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_K$, so that the Kth factor corresponds to the shortest time scale and the first factor corresponds to the longest time scale. We also assume that the variances $(\beta_j^2/2\alpha_j)$ of the long-run distributions of these factors are of the same order. In particular this implies that the β_j's are also correspondingly ordered $\beta_1 < \beta_2 < \cdots < \beta_K$

2.2. Discretization of the Model

The goal of this section is to present a discretized version of this stochastic volatility model (SVM) given in (2.1). Given a fixed time increment $\Delta > 0$, we use the following Euler discretization of the asset process (S_t) and volatility $(Y_t^{(j)})$, $j = 1, \ldots, K$ at the times $t_k = k\Delta$:

$$\begin{aligned} S_{t_{k+1}} - S_{t_k} &= \kappa S_{t_k} \Delta + \sigma_{t_k} S_{t_k} \sqrt{\Delta}\, \epsilon_{t_k}, \\ \log(\sigma_{t_k}^2) &= \sum_{j=1}^{K} Y_{t_k}^{(j)} \\ Y_{t_k}^{(j)} - Y_{t_{k-1}}^{(j)} &= \alpha_j(\mu_j - Y_{t_{k-1}}^{(j)})\Delta + \beta_j \sqrt{\Delta}\, v_{t_k}^{(j)}, \end{aligned}$$

where (ϵ_{t_k}) and $(v_{t_k}^{(j)})$ are (possibly correlated) sequences of iid standard normal random variables. In this paper we restrict ourselves to the case where the sequence of ϵ's is independent of the $v^{(j)}$'s, which, themselves can be correlated, as made explicit in formula (2.5)

We regroup terms within the equations above such that

$$\begin{aligned} \frac{1}{\sqrt{\Delta}}\left(\frac{S_{t_{k+1}} - S_{t_k}}{S_{t_k}} - \kappa\Delta\right) &= \sigma_{t_k}\, \epsilon_{t_k}, \\ \log(\sigma_{t_k}^2) &= \sum_{j=1}^{K} Y_{t_k}^{(j)} \\ Y_{t_k}^{(j)} - \mu_j &= (1 - \alpha_j\Delta)(Y_{t_{k-1}}^{(j)} - \mu_j) + \beta_j \sqrt{\Delta}\, v_{t_k}, \end{aligned}$$

is obtained.

To study the hidden volatility process, we introduce the returns

$$y_{t_k} = \frac{1}{\sqrt{\Delta}}\left(\frac{S_{t_{k+1}} - S_{t_k}}{S_{t_k}} - \kappa\Delta\right),$$

the discrete driving (vector of) volatilities

$$\mathbf{h_{t_k}} = \mathbf{Y_{t_k}} - \boldsymbol{\mu},$$

, where $\mathbf{Y_{t_k}}$ (resp $\boldsymbol{\mu}$) denotes the vector $(y_{t_k}^{(1)}, \ldots, y_{t_k}^{(K)})'$ (resp $(\mu_1, \ldots, \mu_K)'$). The autoregressive parameters will be denoted by

$$\phi_j = 1 - \alpha_j \Delta, \ j = 1, \ldots, K$$

and the standard deviation parameters will be defined by

$$\sigma_v^{(j)} = \beta_j \sqrt{\Delta}, \ j = 1, \ldots, K$$

With an abuse of notation, and since the observations are equally spaced, we will use t instead of t_k/Δ for the discrete time indices, so that $t \in \{0, 1, 2, \cdots, T\}$. To summarize, our model is a K-dimensional AR(1) for the log(volatilities), and can be written in state space form as:

$$y_t = e^{(\mathbf{1'h_t} + \mathbf{1'}\boldsymbol{\mu})/2} \epsilon_t, \quad (2.2)$$

$$\mathbf{h_t} = \mathbf{\Phi h_{t-1}} + \mathbf{\Sigma}^{1/2} \mathbf{v_t}.\text{where } t = 1, \ldots, T \quad (2.3)$$

where $\mathbf{v_t}$ is the vector $(v_t^{(1)}, \ldots, v_t^{(K)})$ and $\mathbf{1}$ is a (K-dimensional) vector of ones.
The K-dimensional autoregressive (diagonal) matrix with elements ϕ_j is denoted by $\mathbf{\Phi}$ and the covariance matrix is denoted by $\mathbf{\Sigma}$. We show in the next section that, without loss of generality, this covariance matrix can be chosen diagonal for the case of K=2 positively correlated factors. In the following we assume that the covariance matrix $\mathbf{\Sigma}$ is diagonal, in part for practical reasons since the correlation parameter between the factors is not (uniquely) identifiable through the sum of these factors and also because our model focuses on the identification of several independent volatility factors contributing to the distribution of returns.
In the next section we make precise the parametrization of our discrete model, including the initial state and random variable notation.

2.3. Discrete time model specification

The simplest one-factor stochastic volatility model can be defined in a hierarchical way as in 2.4. We take the notation in (Kim *et al.*, 1998) and denote h_t as the log-volatility at time t, which follows an AR(1) process which reverts to its long-term mean μ according to the mean-reversion parameter ϕ and with a volatility of volatility σ_v^2. The initial state h_0 is assumed to come from its invariant distribution.

$$\begin{aligned}
y_t &= e^{(h_t+\mu)/2}\epsilon_t, \\
h_t &= \phi h_{t-1} + \sigma_v v_t, \\
h_0 &\sim N_1\left(h_0 \left|0, \sigma_v^2/(1-\phi^2)\right.\right) \\
\begin{bmatrix} \epsilon_t \\ v_t \end{bmatrix} &\sim N_2\left(\begin{bmatrix} \epsilon_t \\ v_t \end{bmatrix}\middle| 0, \mathbf{I_2}\right) \\
(\mu, \phi, \sigma_v) &\sim \pi(\mu, \phi, \sigma_v),
\end{aligned} \tag{2.4}$$

Where h_t is the detrended (log)volatility series and μ is the mean (log)volatility level. To complete the joint distribution a prior density $\pi(\mu, \phi, \sigma_v)$ on the remaining parameters must also be specified.

For simplicity, the mean return is assumed equal to zero. A simple modification of the procedure here presented would allow us to consider a non-zero mean in the returns.

This model, although extremely flexible, has been subject of several modifications. These have gone from varying the distributional assumptions of the returns (Jacquier *et al.*, 1998) to jumps in returns or in the volatility process (ter Horst, 2003).

Factor models extend the simple SVM to the multivariate case, having each of the return series its own factor (Harvey *et al.*, 1995) or reducing the number of factors that affect a series to those that have the bigger impact (Aguilar and West, 2000).

Our contribution to the extension of this literature doesn't focus on the modelization of extreme events (fat tails in the distributions and/or jumps) but instead focuses on the assumption of one single volatility process driving the series, or, as expressed in the previous chapter, on the assumption of having the volatility process driving the returns defined on a common time scale. We will assume for simplicity that the volatility processes are independent ($\mathbf{\Phi}$ and $\mathbf{\Sigma}$ in 2.5 are diagonal). We show in the appendix that, for the two-volatility-process (positively correlated) case, this is equivalent to having two independent factors both for estimation and for prediction. Indeed the covariance is not identifiable, since the model is overparametrized. We ignore the correlation between the returns and the volatilities (leverage) and focus on the modeling of multiple scales.

The one-factor model can be extended to the multiple scale case as

$$\begin{aligned}
y_t &= e^{(\mathbf{1'h_t}+\mu)/2}\epsilon_t \\
\mathbf{h_t} &= \mathbf{\Phi h_{t-1}} + \mathbf{\Sigma}^{1/2}\mathbf{v_t} \\
\mathbf{h_0} &\sim N_K\left(\mathbf{h_0}\,|\mathbf{0},\mathbf{\Omega}\right) \\
\begin{bmatrix} \epsilon_t \\ \mathbf{v_t} \end{bmatrix} &\sim N_{K+1}\left(\begin{bmatrix} \epsilon_t \\ \mathbf{v_t} \end{bmatrix}\middle| \mathbf{0},\mathbf{I_{K+1}}\right) \\
(\boldsymbol{\mu},\mathbf{\Sigma},\mathbf{\Phi}) &\sim \pi(\boldsymbol{\mu},\mathbf{\Sigma},\mathbf{\Phi})
\end{aligned} \tag{2.5}$$

Where $\mathbf{\Sigma}$ is the (K×K) diagonal matrix of conditional variances $\{\sigma_k^2\}_{k=1}^K$ of the (log)volatility processes and $\mathbf{\Omega} = \mathbf{\Phi\Omega\Phi} + \mathbf{\Sigma}$ is the implied (stationary) marginal variance for the initial state (log)volatility vector $\mathbf{h_0}$ and $\mathbf{\Phi}$ is a K×K diagonal matrix with ordered mean reverting parameters $1 > \phi_1 > \phi_2 > \cdots > \phi_K > -1$.
Since $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ are diagonal matrices, $\mathbf{\Omega}$ is a diagonal matrix with elements $\omega_i = \sigma_i^2/(1-\phi_i^2)$ along the diagonal.
Notice that we do not have a vector $\boldsymbol{\mu}$ in the observation equation 2.5, but instead a single parameter μ. This is because only the linear combination (sum) is identifiable. Therefore we define the mean (log)volatility of the series $\mu = \mathbf{1'}\boldsymbol{\mu}$

2.4. Prior specification

We specify independent priors for each of the (sets of) parameters, which is the product measure $\pi(\mu,\mathbf{\Phi},\mathbf{\Sigma}) = \pi_1(\mathbf{\Phi})\pi_2(\mu)\pi_3(\mathbf{\Sigma})$ We set a uniform prior between -1 and 1 for each of the K mean reverting parameters. In order to guarantee identifiability of the different series, we specify the (proper) prior on the ordered parameters, such that $\pi_1(\mathbf{\Phi}) \propto 1_{1>\phi_1>\phi_2>\cdots>\phi_K>-1}$.
Usually in financial data we will have prior information about both μ and $\mathbf{\Sigma}$. We could in principle set an informative normal prior for μ, $\pi_2(\mu) = N_1(\mu|m_0, v_0)$ with fixed hyperparameters m_0 and v_0.

Similarly an informative Inverse Wishart for $\mathbf{\Sigma}$ could be specified. Any such Inverse Wishart could be transformed into inverse gammas for each of the equivalent independent volatility components. We impose an ordering as well in these volatility components based on the explanations in previous sections. Since the σ's are proportional to the β's, this induces an ordering in our volatility components. In practice it helps the MCMC to have a faster convergence.
A noninformative prior (flat over the real line) for μ is obtained by letting $v_0 \to \infty$, which will be the case we will use in the analysis.
And $\pi_3(\mathbf{\Sigma}) \propto \prod_{k=1}^K IG(\sigma_k^2|a_k,b_k)1_{\sigma_1^2<\sigma_2^2<\cdots<\sigma_K^2}$ for $k = 1\ldots K$ with fixed a_k, b_k. We specify

values for these hyperparameters such that the prior is flat. The results seem to be invariant to the choice of even flatter priors.

In practice imposing ordering restrictions in the prior, as we have done both for $\mathbf{\Phi}$ and $\mathbf{\Sigma}_{\boldsymbol{v}}$, is equivalent to rejecting draws that don't comply with those restrictions.

2.5. Estimation

Since the volatility parameters $\mathbf{h_t}$ come nonlinearly in the likelihood, a mixture approximation to the likelihood as detailed in (Kim *et al.*, 1998) has become the natural linearization scheme. We square and take logarithm of the mean corrected returns y_t such that equation 2.5 becomes

$$\log(y_t^2) = \mathbf{1}'\mathbf{h_t} + \mu + \underbrace{\log(\epsilon_t^2)}_{\eta_t} \tag{2.6}$$

Where $\eta_t \sim \log -\chi_1^2$ is approximated as a mixture of seven normals with (known) means m_i, variances v_i and weights q_i. We introduce vectors of indicator variables $\boldsymbol{\delta_t} = \{\delta_{1t}, \ldots, \delta_{7t}\}$ pointing at the component of the mixture with prior probabilities equal to the weights as explained in (Kim *et al.*, 1998).

Then $[\eta_t|\delta_{it} = 1] \sim N_1(\eta_t|m_i - 1.2704, v_i)$

For simplicity of notation, we will denote $\mathbf{y}^* = \{\log(y_1^2), \ldots, \log(y_t^2)\}$ and $\mathbf{h} = \{\mathbf{h_0}, \mathbf{h_1}, \ldots, \mathbf{h_T}\}$

The new state-space representation can be written as follows:

$$\begin{aligned}
[y_t^*|\{\delta_{it}\}_{i=1}^7, \mathbf{h_t}, \mu] &\sim N_1\left(y_t^* \middle| \mathbf{1}'\mathbf{h_t} + \mu - 1.2704 + \sum_{i=1}^7 \delta_{it} m_i, \sum_{i=1}^7 \delta_{it} v_i\right) \\
[\mathbf{h_t}|\mathbf{h_{t-1}}, \mathbf{\Phi}, \mathbf{\Sigma}] &\sim N_K(\mathbf{h_t}\,|\mathbf{\Phi}\mathbf{h_{t-1}}, \mathbf{\Sigma}) \\
[\mathbf{h_0}|\mathbf{\Phi}, \mathbf{\Sigma}] &\sim N_K(\mathbf{h_0}\,|\mathbf{0}, \mathbf{\Omega}) \\
[\boldsymbol{\mu}, \mathbf{\Phi}, \mathbf{\Sigma}] &\sim \pi(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Phi}) \\
[\boldsymbol{\delta}] &\sim \prod_{t=1}^T MN(\boldsymbol{\delta_t}\,|1, q_1, \ldots, q_7)
\end{aligned} \tag{2.7}$$

We can use a Gibbs sampling scheme to do inference on the parameters by following the steps

1. Initialize $\boldsymbol{\delta}, \mu, \mathbf{\Phi}, \mathbf{\Sigma}$

2. Sample $\mathbf{h} \sim [\mathbf{h}|\mathbf{y}^*, \mu, \mathbf{\Sigma}, \boldsymbol{\delta}, \mathbf{\Phi}]$

3. Sample $\boldsymbol{\delta} \sim [\boldsymbol{\delta}|\mathbf{y}^*, \mu, \mathbf{h}]$

4. Sample $\mu \sim [\mu|\mathbf{y}^*, \mathbf{h}, \boldsymbol{\delta}]$

5. Sample $\mathbf{\Phi} \sim [\mathbf{\Phi}|\mathbf{\Sigma}, \mathbf{h}]$

6. Sample $\mathbf{\Sigma} \sim [\mathbf{\Sigma}|\mathbf{h}, \mathbf{\Phi}]$

Details of the full conditionals are given in the appendix. In principle one could use improved sampling versions by integrating out some parameters from the full conditionals (Kim *et al.*, 1998). However, we will not focus here on the efficiency of the sampling algorithm but on testing the existence of more than one volatility time scale.

3. Simulation study

In this section we test the validity of the algorithm under different scenarios. We test the behaviour of the algorithm when the estimating model is the one generating the data and also when we try to estimate using the wrong model (wrong number of scales). We also test this performance of the algorithm under different sample sizes and parameter settings.
Across all experiments, we set the true value of $\mu = -1.5$, which is a value similar to the posterior mean of μ using real data. We will denote as K^* the true number of volatility scales that generated the data, while K will be the value with which we do the estimation.

- We generate the data under three possible parameter settings:

  Setting A: K^*=1, $\phi_1 = 0.95$, $\sigma_1 = 0.26$

  Setting B: K^*=2, $\phi_1 = 0.98$, $\sigma_1 = 0.20$, $\phi_2 = 0.60$, $\sigma_2 = 0.80$

  Setting C: K^*=2, $\phi_1 = 0.98$, $\sigma_1 = 0.20$, $\phi_2 = 0.30$, $\sigma_2 = 0.80$

- Number of scales used in estimation: K=1 or K=2

- Sample sizes: T=1500 or T=3000

Any combination of the above forms an experiment. As an example (as we can see in the Appendix B, a possible experiment comprises generating 3000 data points with the parameter setting A (one volatility scale) and estimating it with K=2 scales. (4th row in the setting A table in that Appendix). The results in Appendix B are based on 100 Monte Carlo experiments for each of the Settings and for each of the experiments. The results reported are means of the posterior medians (bold), means of the posterior standard deviations (in parenthesis to the right). Then below we have the standard deviation of the posterior medians and in parenthesis the standard deviation of the (100) standard deviations.

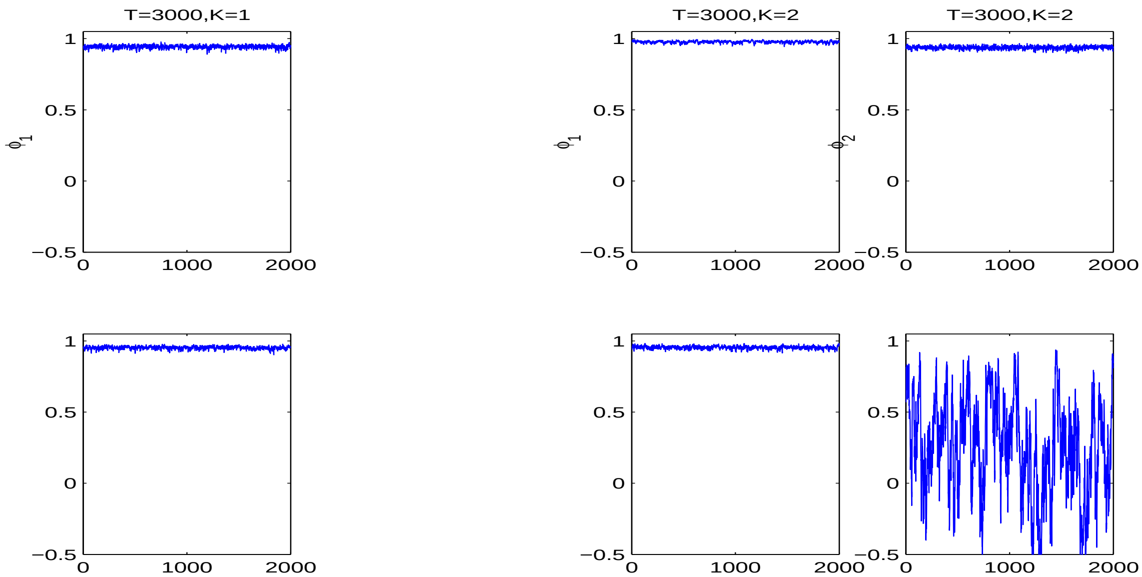

Figure 1: Two typical experiments (horizontal) with the estimation of ϕ_k under K=1 (1st column) and K=2 (2nd and 3rd columns) when K^*=1 ($\phi_1 = 0.95$).

3.1. Results from the simulation study

The results in the table in the appendix show the effect of overparameterization ($K > K^*$) and underparameterization ($K < K^*$). It also shows the performance of the algorithm under different parameter settings even when the number of scales is the true one.
Figure 1 shows the posterior distribution for ϕ_k, $k = 1, \ldots, K$ when the data was generated with one volatility scale ($K^* = 1$). After 50,000 iterations of burn-in, we record (with a thinning of 25) the following 50,000 iterations. We take two of those typical runs. We can see that the estimation of ϕ_1 under K=1 is very stable around the true value (0.95). On the other hand, the estimation of ϕ_2 is very unstable, since the data was generated with only one scale. This graph shows the pattern we should expect when we overparametrize the model. We either get confounding estimates of ϕ_1 and ϕ_2 (the volatility series splits into two parts) or ϕ_2 moves all over the support. We will see similar behaviour in the data analyzed in Section 4, where both patterns show up.

Figure 3 shows the posterior distribution for ϕ_k, $k = 1, \ldots, K$ when the data was generated with 2 volatility scales (K^*=2 and $\phi_2 = 0.6$). The plots represent the same as before, but now the data has been generated with two scales. The model with K=1 is underparametrized, leading to an apparent convergence. The value to which it converges is between the two true

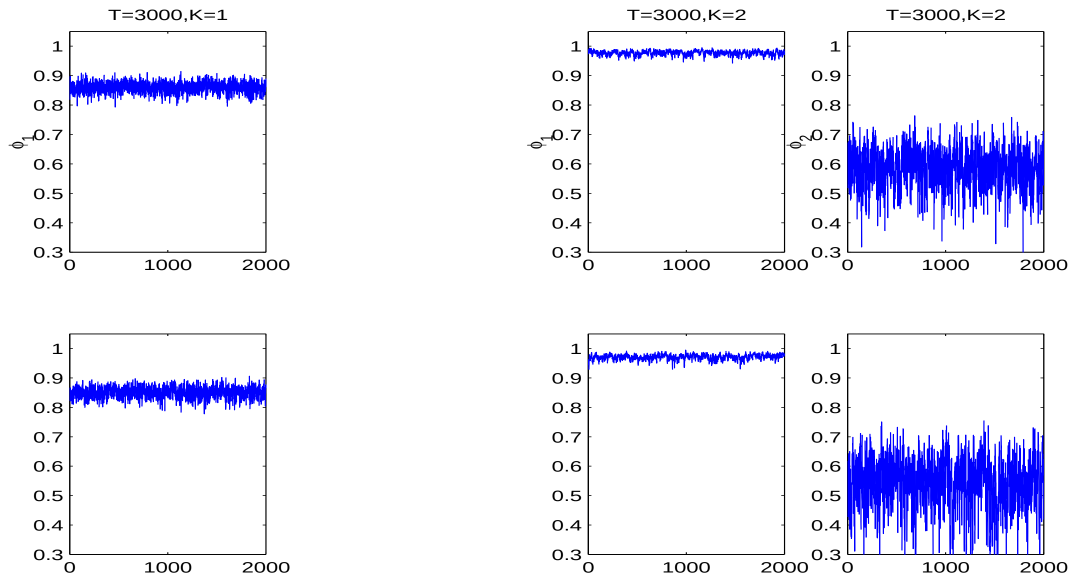

Figure 2: Two typical experiments (horizontal) with the estimation of ϕ_k under K=1 (1st column) and K=2 (2nd and 3rd columns) when K^*=2 ($\phi_1 = 0.98, \phi_2 = 0.6$).

values that generated the data. The convergence is clearly fine for the case when we estimate with K=2. Finally we can see in **??** two typical estimates of a 2-factor SVM when we try to estimate it with three factors. The second factor tends to oscillate between the other two. This might sometime give a wrong impression of convergence. In any case, the excessive variability in the second factor (first three plots) or the periods of the chain where the factors are confounded (last three plots) are enough indication of overparametrization.

4. Empirical applications

We study the volatility in several financial time series. Since most time series could be argued to have a positive trend, we construct the detrended series as in (Kim *et al.*, 1998). In principle we could introduce mean processes (constant or time-dependent), but we refrain from doing so to allow simpler comparisons with previous literature and to just focus on the volatility modeling.

4.1. Foreign Exchange data

To illustrate the algorithm we study the daily exchange rate between the U.K. Sterling and the U.S. Dollar between 10/09/1986 and 08/09/1996. We transform the spot prices into detrended returns and perform the analysis by running the algorithm with one, two and three factors. The

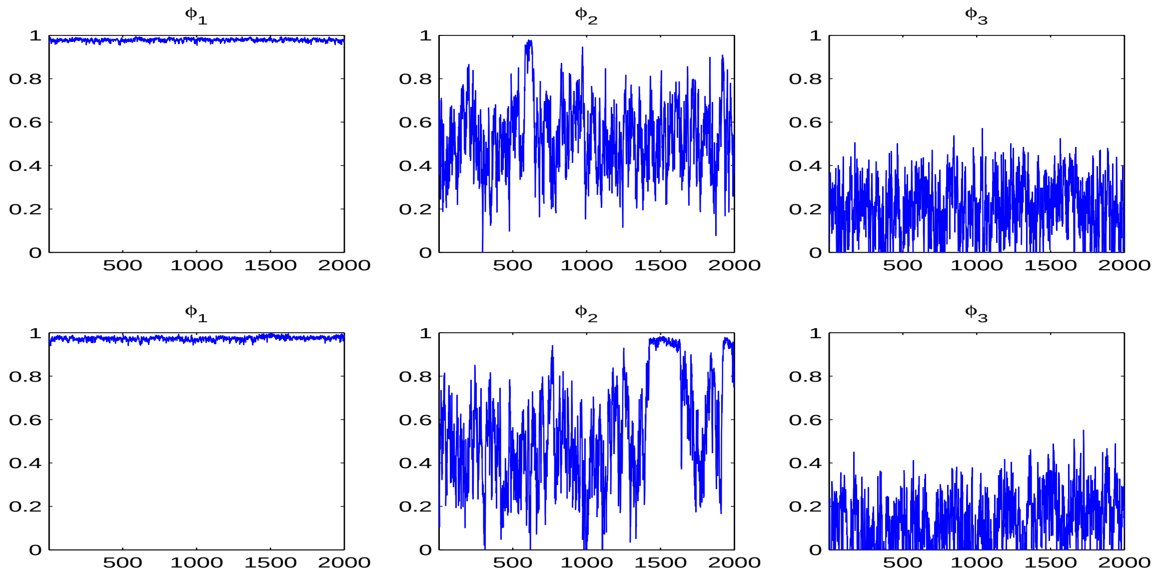

Figure 3: Two typical experiments (horizontal) where the estimation of $\phi_k, k = 1, \dots, K$ was done for K=3 and the data was generated with K^*=2 ($\phi_1 = 0.98, \phi_2 = 0.3$).

results are shown in Table 1. All results are posterior medians (posterior standard deviations). Notice that for K=3, since the first and third factor seem to capture the true factors in the data, the second one (given the restrictions) will wonder between these two. This indicates that merely looking at posterior means/medians is not enough to determine the existence of further factors and visualization of the behaviour of the parameter estimates is necessary.

	ϕ_1	ϕ_2	ϕ_3	μ	σ_1	σ_2	σ_3
GBP/USD K=1	**0.916**	-	-	**-1.220**	**0.134**	-	
	(0.057)	-	-	(0.106)	(0.130)	-	
GBP/USD K=2	**0.988**	**0.149**	-	**-1.470**	**0.013**	**0.947**	
	(0.006)	(0.090)	-	(0.255)	(0.007)	(0.132)	-
GBP/USD K=3	**0.993**	**0.896**	**0.119**	**-1.486**	**0.004**	**0.026**	**0.883**
	(0.006)	(0.297)	(0.098)	(0.411)	(0.007)	(0.104)	(0.167)

Table 1: GBP/USD results

4.2. Results from the Foreign Exchange data analysis

From figure 5 we can see that, for the GBP/USD series, using one scale gives reasonably stable results. There is a heavy left tail in the posterior distribution of ϕ_1 under K=1. This could be

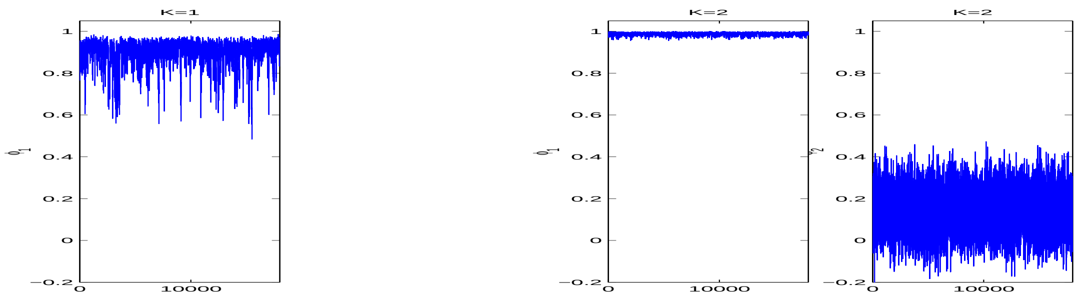

Figure 4: Traceplots after burn-in of $\phi_k, k = 1, \ldots, K$ for the GBP/USD series K=1 and K=2

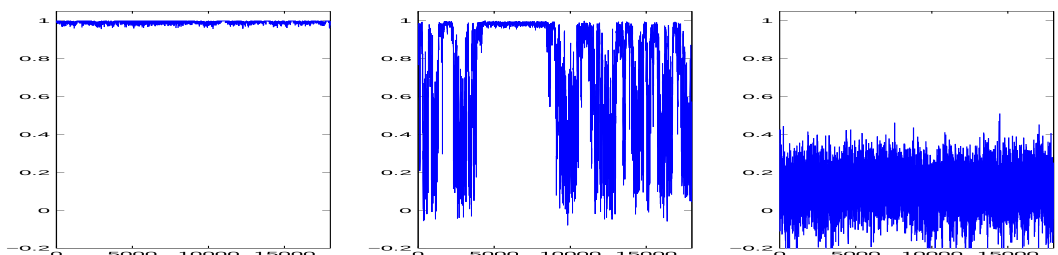

Figure 5: Traceplots after burn-in of $\phi_k, k = 1, \ldots, K$ for the GBP/USD series K=3

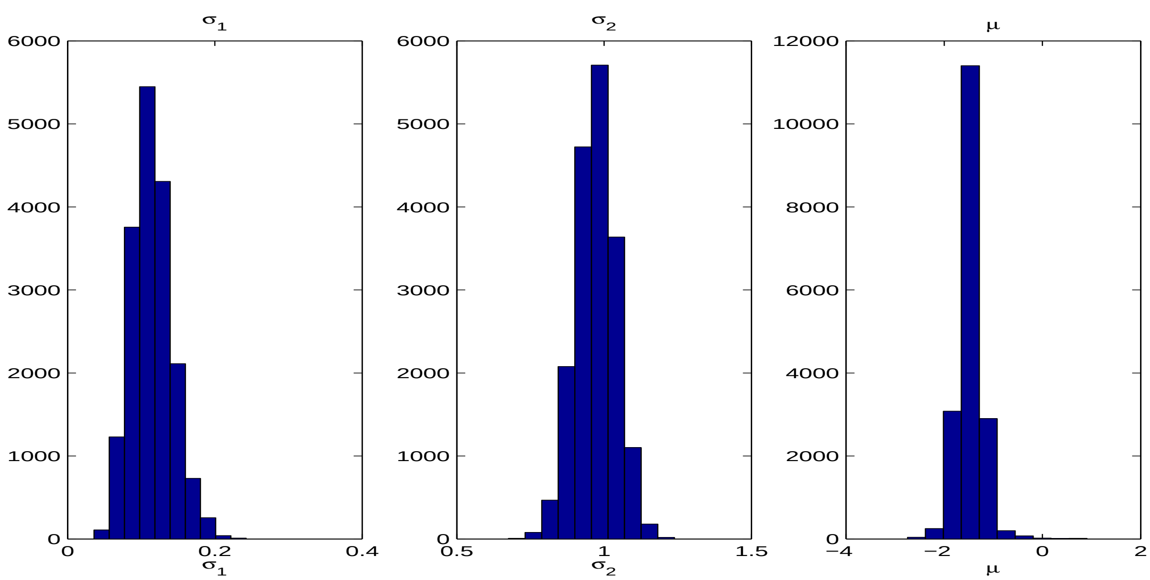

Figure 6: Histograms of σ_1, σ_2 and μ for the GBP/USD series for K=2

an indication of the existence of a second volatility scale at a faster scale. In this case the unique volatility scale would try to pick up both fast and slow mean reverting volatility processes, achieving an intermediate result in terms of its estimation.
If we run the estimation with K=2 we can see that there is a second volatility factor running at a very fast mean reverting rate. The identification of this second factor allows us to identify that the first factor was moving at a lower frequency than previously thought (the posterior median of the mean reverting parameter for K=2 for the slower scale is 0.988 under K=2, but it was 0.916 under K=1).
Running the algorithm with a third volatility process brings identifiability problems for the second one, showing that apparently there is not a third factor in the series and that the middle factor moves (gets confounded) between the first and the third factors. Simple visualization seems enough to identify the adequate number of factors.
Figure 7 shows the posterior distribution of the mean (log)volatilities for each time point t and for each k when K=2, together with the data.
Notice that the short time scale process captures the extreme events, while the long scale captures the trends and is much less influenced by extreme events. There is a peak the returns between the 1900th and the 2000th observation. This peak occurs during a period of relative stability (the absolute value of the returns are clearly more stable than in previous periods). We can see that the slow mean reverting volatility process remain in low values and is not affected by this peak, while the fast mean reverting one exhibits a peak coinciding with that day. We do not address the question of whether this should be considered a jump in volatility or a second volatility process. However, the estimates in Table 1 seem to show the existence of a clear second volatility factor at a (much) faster scale.

One can also construct easily any functional of the parameters in the model and do inference on it. Indeed, an interesting quantity to look at is the variance of the long-run distributions of the factors. For this dataset these long-run variances (as defined in Section 2.1) are of the same order when K=2 (0.5597 for the slow mean-reverting process and 0.5589 for the fast mean reverting one), while, when estimating them with K=1, we get a variance of 0.8215. Given that the factors are quite separated, this result gives a very intuitive way to construct joint priors on the mean reverting parameters and the variance parameters. It also confirms the adequacy of the ordering assumption introduced in Section 2.1.

5. Conclusions

This problem raises questions about the true number of scales we should include and whether the data can inform us about that. Interesting questions from the model discrimination point of view, including model selection and model averaging problems. These questions, although interesting,

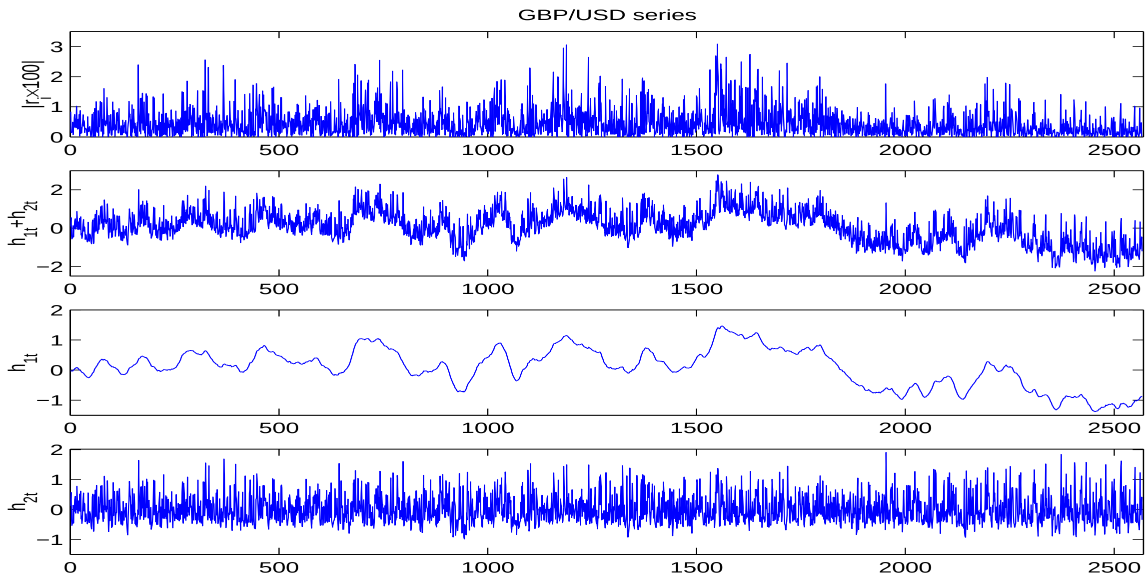

Figure 7: GBP/USD series and posterior means of the volatility processes under K=2

can be easily identified in most cases from a simple analysis of the posterior distribution under different models, since overparameterized models will bring clear signs of lack of convergence, while underparameterized ones will bring an incorrect appearence of convergence. We would, therefore, recommend to run the algorithm always with at least 2 scales. Then, if there is lack of convergence, reduce to K=1. If, to the contrary, there is apparent convergence for K=2, we should test K=3. More than 3 time scales is in general unreasonable, especially for practical purposes, since identifying more than three scales would require them to be well-separated and the dataset to be large enough to identify all of them.

Acknowledgements

This material was based upon work supported by the National Science Foundation under Agreement No. DMS-0112069. We would like to thank Prof. Mike West for the data used in this paper as well as the members of the Stochastic Computation research group at the Statistical and Applied Mathematical Sciences Institute for their valuable comments.

Appendix A: Proof of equivalence between positively correlated and independent factor modeling for the two-factor case

In most of the cases in practice we will debate between one volatility scale or two. In the case where they are positively correlated, the problem simplifies to one of two independent volatility processes.

Let $y_t \sim N_1\left(y_t \middle| 0, e^{\mathbf{1'h_t}+\mu}\right)$. Define an iid sequence of stochastic (degenerate) vectors $\{\boldsymbol{\lambda_t} \sim N_2(\boldsymbol{\lambda_t}|\mathbf{0},\mathbf{R})\}_{t=1}^{T}$, where $\mathbf{R}$ is a positive semi-definite 2 by 2 covariance matrix such that $0 < r_{11} = r_{22} = -r_{12}$ (see Chapter 5 of West and Harrison (1999) for a similar example). We can see that $\mathbf{1'}\boldsymbol{\lambda_t} \equiv 0 \; \forall t$. Then we can define the 'equivalent' uncorrelated volatility vector $\mathbf{h_t^*} = \mathbf{h_t} - \boldsymbol{\lambda_t}$ with conditional diagonal covariance matrix $\boldsymbol{\Sigma}$. As we see in 5.8, the likelihood remains unchanged, while we can see in 5.9 that the observation equations are independent.

$$y_t \sim N_1\left(y_t \middle| \mathbf{0}, e^{\mathbf{1'h_t}+\mu}\right) = N_1\left(y_t \middle| \mathbf{0}, e^{\mathbf{1'}(\mathbf{h_t^*}+\boldsymbol{\lambda_t})+\mu}\right) = N_1\left(y_t \middle| \mathbf{0}, e^{\mathbf{1'h_t^*}+\mu}\right) \tag{5.8}$$

$$\begin{aligned} \mathbf{h_t} &\sim N_1\left(\mathbf{h_t} \middle| \boldsymbol{\Phi}\mathbf{h_{t-1}}, \boldsymbol{\Sigma^*}\right) = \\ \mathbf{h_t^*} + \boldsymbol{\lambda_t} &\sim N_2\left(\mathbf{h_t^*} + \boldsymbol{\lambda_t} \middle| \boldsymbol{\Phi}(\mathbf{h_{t-1}^*} + \boldsymbol{\lambda_{t-1}}), \boldsymbol{\Sigma^*}\right) = \\ \mathbf{h_t^*} &\sim N_2\left(\mathbf{h_t^*} \middle| \boldsymbol{\Phi}\mathbf{h_{t-1}^*}, \underbrace{\mathbf{R} + \boldsymbol{\Phi}\mathbf{R}\boldsymbol{\Phi} + \boldsymbol{\Sigma^*}}_{\boldsymbol{\Sigma}}\right) \end{aligned} \tag{5.9}$$

If $\Sigma_{12}^* > 0, \phi_1, \phi_2 \neq 0$, $\exists\ \mathbf{R} : \mathbf{R} + \boldsymbol{\Phi}\mathbf{R}\boldsymbol{\Phi} + \boldsymbol{\Sigma^*} = \boldsymbol{\Sigma}$ for some diagonal positive definite matrix $\boldsymbol{\Sigma}$ with elements σ_i^2. Matching the elements gives us that $r_{11} = r_{22} = \Sigma_{12}^*/(1 - \phi_1\phi_2) = -r_{12}$. The variance of these independent volatility processes will be one-to-one versions of the original ones, since $\sigma_i^2 = \Sigma_{12}^*(1 + \phi_1^2)/(1 - \phi_1\phi_2)$. Therefore any two positively correlated volatility processes are equivalent to two independent ones, making nonidentifiable a parameter like Σ_{12}^*.

The implied invariant distribution (marginal) for the initial state would just be a normal centered at zero and with diagonal covariance matrix $\Omega = \boldsymbol{\Phi}\Omega\boldsymbol{\Phi} + \boldsymbol{\Sigma}$.

Appendix B: Results from the Monte Carlo experiment

	ϕ_1	ϕ_2	μ	σ_1	σ_2
Setting A K^*=1	**0.95**	**-**	**-1.5**	**0.26**	**-**
T=1500 K=1	**0.945** (0.016) 0.016 (0.004)	- -	**-1.503** (0.144) 0.143 (0.026)	**0.265** (0.036) 0.036 (0.005)	- -
T=3000 K=1	**0.945** (0.011) 0.011 (0.002)	- -	**-1.501** (0.098) 0.100 (0.014)	**0.268** (0.026) 0.025 (0.003)	- -
T=1500 K=2	**0.962** (0.018) 0.012 (0.007)	**0.661** (0.312) 0.363 (0.169)	**-1.538** (0.153) 0.141 (0.042)	**0.118** (0.063) 0.090 (0.032)	**0.292** (0.053) 0.041 (0.014)
T=3000 K=2	**0.964** (0.013) 0.009 (0.004)	**0.607** (0.291) 0.376 (0.161)	**-1.501** (0.101) 0.010 (0.018)	**0.146** (0.047) 0.089 (0.029)	**0.278** (0.041) 0.024 (0.011)
Setting B K^*=2	**0.98**	**0.60**	**-1.5**	**0.20**	**0.80**
T=1500 K=1	**0.855** (0.027) 0.047 (0.007)	- -	**-1.442** (0.142) 0.259 (0.032)	**0.688** (0.064) 0.089 (0.006)	- -
T=3000 K=1	**0.870** (0.018) 0.030 (0.003)	- -	**-1.461** (0.104) 0.215 (0.015)	**0.669** (0.045) 0.060 (0.003)	- -
T=1500 K=2	**0.969** (0.018) 0.017 (0.011)	**0.530** (0.140) 0.148 (0.057)	**-1.522** (0.315) 0.236 (0.176)	**0.238** (0.076) 0.067 (0.030)	**0.780** (0.081) 0.088 (0.009)
T=3000 K=2	**0.977** (0.008) 0.009 (0.003)	**0.554** (0.076) 0.069 (0.020)	**-1.491** (0.208) 0.171 (0.065)	**0.217** (0.042) 0.046 (0.011)	**0.819** (0.054) 0.052 (0.004)
Setting C K^*=2	**0.98**	**0.30**	**-1.5**	**0.20**	**0.80**
T=1500 K=1	**0.894** (0.026) 0.052 (0.010)	- -	**-1.417** (0.157) 0.262 (0.044)	**0.521** (0.065) 0.116 (0.012)	- -
T=3000 K=1	**0.905** (0.017) 0.036 (0.005)	- -	**-1.406** (0.112) 0.214 (0.022)	**0.505** (0.046) 0.075 (0.006)	- -
T=1500 K=2	**0.974** (0.012) 0.012 (0.005)	**0.228** (0.160) 0.148 (0.041)	**-1.516** (0.281) 0.252 (0.097)	**0.217** (0.045) 0.044 (0.015)	**0.777** (0.079) 0.079 (0.008)
T=3000 K=2	**0.979** (0.007) 0.007 (0.002)	**0.276** (0.104) 0.093 (0.015)	**-1.496** (0.205) 0.179 (0.059)	**0.205** (0.028) 0.027 (0.005)	**0.802** (0.053) 0.053 (0.003)

Appendix C: Details of the MCMC/sampling

Sampling **h** After approximating the likelihood through the mixture of normals and conditionally on the mixture component, we can (jointly) sample the vector of volatilities, including the initial state, using its (multivariate) DLM structure by running the Forward Filtering/Backward Smoothing algorithm (West and Harrison, 1999)

Sampling $\boldsymbol{\delta}$ Each indicator vector $\boldsymbol{\delta_t}$ comes from a multinomial distribution with 1 outcome and with probabilities proportional to $q_i^* = q_i f_N\left(\log(y_t^2)|m_i + \mu + \mathbf{1'h_t} - 1.2704, v_i\right)$, where m_i, v_i, q_i, $i = 1, \ldots, 7$ are the prior means, variances and weights of the mixture distributions and f_N denotes the normal density function. Therefore $\boldsymbol{\delta_t} \sim MN(\boldsymbol{\delta_t}|1, q_1^*, \ldots, q_7^*)$

Sampling μ Under a normal prior $\pi(\mu) \sim N(m_0, v_0)$ the full conditonal for μ is also normal, $[\mu|\ldots] \sim N_1(\mu|m_0^*, v_0^*)$, where $m_0^* = (1/v_0^*) \times \sum_{t=1}^T (\log(y_t^2) - \mathbf{1'h_t} - a_t + 1.2704)/b_t$ and $v_0^* = (v_0^{-1} + \sum_{t=1}^T b_t^{-1})^{-1}$ and a_t, b_t are the mixture component means and variances, $b_t = \sum_{i=1}^7 \delta_{it} v_i$ and $a_t = \sum_{i=1}^7 \delta_{it} m_i$

Sampling $\boldsymbol{\Phi}$ Imposing stationarity and ordering a priori can be done with a uniform prior on (-1,1) such that $\phi_1 > \phi_2 > \cdots > \phi_K$. Then we can sample $\phi_k^* \sim N(\phi_k^*|m_k, v_k)$, where $m_k = (1/v)\sum_{t=2}^T h_{k,t}h_{k,t-1}$ and $v_k = \sum_{t=1}^{T-1} h_{k,t}^2$ until the proposed set of values follows the ordering conditions. Then, if we define $\boldsymbol{\Omega^*} = \boldsymbol{\Phi\Omega^*\Phi} + \boldsymbol{\Sigma}$, we will set $\boldsymbol{\Phi} = \boldsymbol{\Phi^*}$ with probability $= \min(1, \exp\{\log[f_N(\mathbf{h_0}|0, \boldsymbol{\Omega^*})] - \log[f_N(\mathbf{h_0}|0, \boldsymbol{\Omega})]\})$

Sampling $\boldsymbol{\Sigma}$ We can update the components of $\boldsymbol{\Sigma}$ either all at once or one at a time. Under a prior of the form $\pi(\boldsymbol{\Sigma}) \propto \prod_{k=1}^K IG(\sigma_k^2|a_k, b_k) 1_{\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_K^2}$ (a_k, b_k can denote previous information), we can sample $\sigma_k^2 \sim IG(\sigma_k^2|\alpha_k/2, \beta_k/2) 1_{\sigma_1^2 < \sigma_2^2 < \cdots < \sigma_K^2}$, where $\alpha_k = T+1+2a_k$ and $\beta_k = \sum_{t=1}^T [h_{k,t}^2 - \phi_k h_{k,t-1}]^2 + h_{k,0}^2(1 - \phi_k^2) + 2b_k$

References

[1] Aguilar O. and West, M., (2000) "Bayesian Dynamic Factor modles and variance matrix discounting for portfolio allocation", Journal of Business and Economic Statistics, 18, pp. 338-357.

[2] Alizadeh, S., Brandt M. and Diebold F. (2002) "Range-based estimation of stochastic volatility models." Journal of Finance, 57, pp 1047-1091.

[3] Bates D. (1996) "Testing option pricing models, in Statistical Methods in Finance," Vol. 14 of Handbook of Statistics, eds. G.Maddala and C. Rao, North Holland, Amsterdam (1996) Chap. 20, 567-611.

[4] Chernov M., Gallant R., Ghysels E., and Tauchen G. (2001) "Alternative models for stock price dynamics." Preprint, Columbia University.

[5] Duffie, D. and Singleton, K. (1993) “Simulated moments estimation of Markov models of asset prices,” Econometrica, 61, pp 929-952.

[6] Fouque, J.-P., Papanicolaou G., and Sircar R. (2000a) “Derivatives in Financial Markets with Stochastic Volatility”, Cambridge University Press.

[7] Fouque, J.-P., Papanicolaou G., and Sircar R. (2000b) “International Journal of Theoretical and Applied Finance,” 3 (1), pp 101-142.

[8] Fouque, J.-P., Papanicolaou G., Sircar R., and Solna K. (2003) “Short time-scales in $S\&P$ 500 volatility,” Journal of Computational Finance, Vol 6(4), summer 2003.

[9] Fridman, M. and Harris, L. (1998) “A maximum likelihood approach for non-Gaussian stochastic volatility models,” Journal of Business and Economic statistics, 16, pp. 284-291.

[10] Harvey, A., Ruiz, E. and Shephard N. (1994) “Multivariate stochastic variance models,” Review of Economic Studies, 61, pp 247-264.

[11] ter Horst, E. (2003), “A Levy generalization of compound Poisson Processes in Finance: Theory and applications”, Dissertation.

[12] Jacquier, E., Polson, N. G. and Rossi, P. E. (1994) “Bayesian analysis of stochastic volatility models,” Journal of Business and Economic statistics, 12, pp 413-417.

[13] Jacquier, E., Polson, N.G. and Rossi, P.E., (1998) “Stochastic Volatility: Univariate and Multivariate extensions”, CIRANO working paper.

[14] Kim, S., Shephard, N. and Chib, S. (1998) “Stochastic Volatility: Likelihood Inference and Comparison with ARCH models”, The Review of Economic Studies, 65, pp 361-393.

[15] LeBaron. “Stochastic volatility as a simple generator of apparent financial power laws and long memory”. Quantitative Finance, Vol 1, No 6: 621-31. November 2001.

[16] Melino A. and Turnbull S. (1990) “Pricing foreign currency options with stochastic volatility,” J. Econometrics, 45, pp 239-265.

[17] Nelson, D. (1991) “Conditional heteroskedasticity in asset returns: A new approach,” Econometrica, 59(2), pp 347-370.

[18] Ruiz, E. (1994) “Quasi-maximum likelihood estimation of stochastic volatility models,” Journal of Econometrics, 63, pp 289-306.

[19] Sandmann, G. and Koopman, S. J. (1998) “Estimation of stochastic volatility models via Monte Carlo maximum likelihood,” Journal of Econometrics, 87, pp 271-301.

[20] West, M., and Harrison, J. (1999), “Bayesian Forecasting and Dynamic Models”, Springer-Verlag.