doi:10.1111/j.1467-9892.2008.00584.x

FITTING STOCHASTIC VOLATILITY MODELS IN THE PRESENCE OF IRREGULAR SAMPLING VIA PARTICLE METHODS AND THE EM ALGORITHM

By Jeongeun Kim and David S. Stoffer

University of Maryland, Baltimore and University of Pittsburgh

First Version received April 2006

Abstract. Stochastic volatility (SV) models have become increasingly popular for explaining the behaviour of financial variables such as stock prices and exchange rates, and their popularity has resulted in several different proposed approaches to estimating the parameters of the model. An important feature of financial data, which is commonly ignored, is the occurrence of irregular sampling because of holidays or unexpected events. We present a method that can handle the estimation problem of SV models when the sampling is somewhat irregular. The basic idea of our approach is to combine the expectation-maximization (EM) algorithm with particle filters and smoothers in order to estimate parameters of the model. In addition, we expand the scope of application of SV models by adopting a normal mixture, with unknown parameters, for the observational error term rather than assuming a log-chi-squared distribution. We address the problems by using state–space models and imputation. Finally, we present simulation studies and real data analyses to establish the viability of the proposed method.

Keywords. EM algorithm; financial time series; missing data; mixtures of normals; particle filtering; particle smoothing; state–space model; stochastic volatility.

1. INTRODUCTION

The stochastic volatility (SV) model proposed by Taylor (1982) has become increasingly popular for explaining the behaviour of financial time series, and its popularity has resulted in several different approaches to parameter estimation. An important feature of financial data, which is commonly ignored, is the existence of irregular sampling. For example, when the holiday schedules of the financial markets are different, we can experience days when some market prices are missing. Although it may be rare, stretches of data may not be available for various reasons such as unexpected political events, or natural or man-made disasters. Furthermore, inconsistent, unlikely or impossible values may have been input into the data file. Often, missing observations in financial series have been handled by *ad hoc* methods such as ignoring the problem, averaging or aggregation. We present a method to handle the problem of parameter estimation for SV models when the sampling is somewhat irregular. In addition, various authors have argued that financial data often have heavier tails than can be captured by the standard SV model (see, e.g. Shephard, 1996, pp. 40–1; Kim *et al.*, 1998, pp. 385–7; Bai *et al.*, 2003).

0143-9782/08/05 811–833 JOURNAL OF TIME SERIES ANALYSIS Vol. 29, No. 5

To expand the scope of applications and to deal with the problem of heavy tails, we model the observational errors using a normal mixture distribution with unknown parameters for the observation errors as was done in Shumway and Stoffer (2006, Sect. 6.10) and Stoffer and Wall (2004). We present the details of this robust model in Section 2. For model fitting in the presence of missing or irregular observations, we combine the expectation-maximization (EM) algorithm with particle filters and smoothers to estimate the parameters of the model. In the EM framework, we must calculate the expected likelihood. The expected likelihood, however, cannot be calculated directly, and hence we adopt particle filters and smoothers. Our technique is detailed in Section 3, and a general discussion of particle filtering and smoothing is presented in Appendix A. The convergence of our method is discussed in Appendix B. To solve the irregular data problem, we use an imputation technique based on a property of the state–space model; this technique is detailed in Section 4. Finally, to demonstrate the viability of our methods, we perform several numerical exercises and data analyses in Section 5.

In the standard SV model framework, the data are returns, r, that are generated from a probability model $f(r|h)$, where h is a vector of volatilities, and this unobserved vector h has a probabilistic structure $f(h|\theta)$, where θ is a vector of parameters (see Jacquier *et al.*, 1994 or Shephard, 1996 for details). In the standard form of the model, volatility is modelled as an AR process,

$$h_t = \phi h_{t-1} + w_t, \tag{1}$$

and the returns are given by

$$r_t = \beta \exp\left(\frac{h_t}{2}\right)\epsilon_t, \tag{2}$$

where $w_t \overset{\text{i.i.d.}}{\sim} N(0, Q), h_0 \sim N(\mu_0, \sigma_0^2)$, $\epsilon_t \overset{\text{i.i.d.}}{\sim} N(0,1)$ and $\{w_t\}$ and $\{\epsilon_t\}$ are independent processes. We note again that the assumption that ϵ_t in eqn (2) is Gaussian, which was most likely inherited from the autoregressive conditionally heteroscedastic (ARCH) model, is often criticized as an assumption that is not valid for most financial data. This assumption will be dropped in Section 2.

Equation (2) is typically linearized by taking the logarithm of the squared returns, which results in the equation

$$y_t = \alpha + h_t + v_t, \tag{3}$$

where $y_t = \log(r_t^2)$, $\alpha = \log(\beta^2) + E(\log(\epsilon_t^2))$, $v_t = \log(\epsilon_t^2) - E(\log \epsilon_t^2)$. Note that $\epsilon_t^2 \sim \chi_1^2$, so that v_t has a centred log-χ_1^2 distribution.

Equations (1) and (3) represent the standard univariate SV model and together they form a linear, non-Gaussian, state–space model for which (3) is the observation equation and (1) is the state equation. If the observational error process, v_t, were Gaussian, the model could be fit using standard techniques (e.g. Shumway and Stoffer, 2006, Ch. 6). The fact that the model is non-Gaussian, however, leads to several complications. Various approaches for parameter estimation for SV models have been proposed. Some of the approaches are

efficient and others are not. The lack of analytic likelihoods makes the estimation problem difficult, and so either likelihood approximation methods or numerical methods have been considered. Likelihood approximation methods use simpler, easy to handle, likelihood functions that are similar to the true likelihood of the model, whereas numerical methods use Monte Carlo techniques to approximate the true model likelihood. In general, likelihood approximation methods are easier to perform than numerically intensive methods. There is a possibility, however, that the estimation results might be inaccurate because the approximations may be poor at times. Numerically intensive methods have recently become popular because of relatively cheap computing costs. Their strong point is that they use an exact likelihood, but it may take a longer time to obtain stable parameter estimates because of the complexity of the likelihood.

Simulation methods for classical inference applied to SV models were discussed in Danielson (1994) and Sandmann and Koopman (1998). Mellino and Turnbull (1990) used a generalized method of moments (GMM) approach. Harvey *et al.* (1994) proposed a quasi-maximum likelihood approach that approximates the SV model to a linear Gaussian model, and used the well-developed estimation method for the linear Gaussian models. Durbin and Koopman (2000) used the idea of linearization of general state–space models and matched terms in the likelihood or posterior of a linearized model to those of a linear Gaussian model. As a result, the estimation techniques used for the linear Gaussian model can be applied to general state–space models. Liesenfeld and Richard (2003) used a maximum likelihood approach based upon efficient importance sampling. Shumway and Stoffer (2006, Sect. 6.10) and Stoffer and Wall (2004) used an approximate likelihood in a normal mixtures setting. Their basic idea, which was inspired by Kim *et al.* (1998), was to approximate the observation error, v_t, in eqn (3) by a mixture of normals with unknown parameters, and then use resampling techniques to obtain the sampling distribution of the parameter estimates. A Bayesian approach, based on Markov chain Monte Carlo (MCMC) methods, was taken by Jacquier *et al.* (1994). Chib *et al.* (2002) and Kim *et al.* (1998) also adopted these methods. The basic idea was to get a random sample from the posterior density of parameters, given the data. In these approaches, parameters are assumed to have some prior density, while in the classical analysis, parameters are fixed and unknown, as in our study.

Recently, particle methods (sequential Monte Carlo or sequential importance sampling) have been applied to the SV model. Because particle filters are designed to get the samples of hidden states, given parameters and data, in order to estimate parameters, it will be necessary to either adopt other methods for the parameter estimation, or to modify the particle filtering method to embrace the parameters as a part of the hidden states. A standard approach consists of setting a prior distribution on the unknown parameters and then using that to write an extended state–space model; that is, the parameters are considered hidden states. Estimation can be accomplished by applying a filtering algorithm as was done in Kitagawa and Sato (2001). Doucet and Tadic (2003) combined particle filtering methods and gradient algorithms.

Missing (or irregular) data problems have been addressed in Bayesian, MCMC approaches, by sampling from the conditional distribution of the missing data, given all other parameters and state variables (see, e.g. Kim *et al.*, 1998). In such a setting, missing observations are considered unknown parameters, so if there are many missing observations or the sampling is highly irregular, these methods will be less effective. In our case, we deal with the fixed parameter problem and adopt the EM algorithm to handle irregular sampling. Our method is easy to understand and apply to the missing (or irregular) data case, because the estimation results can be obtained by a slight modification of the algorithm for the case of regular sampling.

2. A MIXTURE MODEL

We now consider a modification of the standard SV model given by eqns (1) and (3), wherein it is assumed that the observational noise process, v_t, is a mixture of two normals with unknown parameters. As previously indicated, the model was first proposed in Shumway and Stoffer (2006, Sect. 6.10) and was used again in Stoffer and Wall (2004). The idea, however, was inspired by Kim *et al.* (1998), who used mixtures of normals to approximate the log of a chi-squared distribution.

The model first presented in Shumway and Stoffer (2006, Sect. 6.10) retains the state equation for the volatility as

$$h_t = \phi h_{t-1} + w_t, \tag{1}$$

but the observation eqn (3) is changed to

$$y_t = \alpha + x_t + v_t \quad \text{and} \quad v_t = I_t z_{t1} + (1 - I_t) z_{t0} - \mu\pi, \tag{4}$$

with $z_{t0} \overset{\text{i.i.d.}}{\sim} N(0, R_0)$, $z_{t1} \overset{\text{i.i.d.}}{\sim} N(\mu, R_1)$, and I_t is an indicator variable, $I_t \overset{\text{i.i.d.}}{\sim} B(\pi)$, where π is an unknown mixing probability; i.e. $\Pr(I_t = 1) = \pi = 1 - \Pr(I_t = 0)$. In this model, the standard SV model $\log \chi_1^2$ distributional assumption on v_t has been replaced by a mixture of normals distributional assumption; the term $-\mu\pi$, is added to make v_t a zero-mean variable. The observational noise is constructed using two normals: the $N(0, R_0)$ term is used to account for most of the noise, whereas the $N(\mu, R_1)$ term is used to account for the lower tail behaviour of the noise. Consequently, μ will be negative and $R_1 \geq R_0 \geq c > 0$, where c is a constant; i.e. R_0 is bounded away from zero. This restriction is enough to ensure a global maximizer of the likelihood function (see Hathaway, 1985 for details).

Equation (4) is equivalent to

$$y_t = h_t + v_t \quad \text{and} \quad v_t = I_t z_{t1} + (1 - I_t) z_{t0} \tag{5}$$

with $z_{t0} \overset{\text{i.i.d.}}{\sim} N(m_0, R_0)$, $z_{t1} \overset{\text{i.i.d.}}{\sim} N(m_1, R_1)$, $m_0 = \alpha - \mu\pi$, $m_1 = \alpha + (1 - \pi)\mu$, where α and μ are unknown parameters, and $I_t \overset{\text{i.i.d.}}{\sim} B(\pi)$. Although the original version of the model was written as (1)–(4), we find it easier to work with the model written as it is in (1)–(5).

The main advantage of the mixture SV model, (1)–(5), over the standard SV model, (1)–(3), is its flexibility. In the standard model, the distribution of v_t contains no unknown parameters, and hence no room is left for deviation from the assumed model. If the observational error distribution is very different from the assumed $\log(\chi_1^2)$ distribution, the resulting parameter estimation is inconsistent. An example is shown in Section 5.1. The normal mixtures distribution, however, allows for flexibility. In the modified model, v_t has its own parameter set, $\{m_0, m_1, R_0, R_1, \pi\}$, which must be estimated along with the state parameters, $\{\phi, Q\}$. Therefore, this modification can give better parameter estimates for $\{\phi, Q\}$ because it uses an observational error distribution that reflects the data structure, rather than the predetermined observational error of the standard model.

The mixture assumption for the observational noise used in Kim *et al.* (1998) was primarily a device to approximate the $\log(\chi_1^2)$ distribution; their mixture had no unknown parameters and hence did not use data that fit the errors. Shumway and Stoffer (2006, Sect. 6.10) and subsequently Stoffer and Wall (2004) realized that the $\log(\chi_1^2)$ assumption was too restrictive and hence they used the model as given in eqn (1)–(4). Since then, a number of authors have picked up on the idea, and we mention a few. Durham (2007) proposed a mixture model that is identical to the one presented in Shumway and Stoffer (2006, Sect. 6.10) and subsequently Stoffer and Wall (2004); in that article, Durham's estimation is based on the simulated maximum likelihood approach via importance sampling, which is similar to Liesenfeld and Richard (2003). Our goal in this article is not so much to establish another Monte Carlo method for estimating the parameters of an SV model, but rather to establish a method for estimation when the sampling is irregular.

3. PARAMETER ESTIMATION FOR THE MIXTURE SV MODEL

In this section, we consider parameter estimation for the SV mixture model when the data are sampled at regular intervals and there are no missing observations. The proof of the convergence of the algorithm presented in this section is given in Appendix B. After developing a proper method for this case, we discuss its modification in the irregular data case in Section 4. Although there are alternatives to fitting the mixture SV model to data, e.g. Stoffer and Wall (2004), the main advantage of the EM algorithm is its ability to handle irregular sampling and missing data.

As previously stated, the assumption that R_0/R_1 is bounded away from zero is enough to ensure a global maximizer of the likelihood function. In essence, we are preforming restricted maximum likelihood estimation. There is, however, no need to worry about this restriction in the iterative procedure because the two normal distributions are well separated and problems will occur only if the sample size is unduly small or if the model is misspecified in such a way that the observational

noise is truly Gaussian and not skewed left. These problems are not likely to occur in the analysis of returns; moreover, an investigator could detect such problems prior to using our proposed algorithm. Examples of the separation of the two normal components in the mixture can been seen in the examples in Shumway and Stoffer (2006, Sect. 6.10).

The basic strategy for the mixture model, (1)–(5), is to apply the EM algorithm to the complete data, $\{h_0,h_1,\ldots,h_n; I_1,\ldots,I_n; y_1,\ldots,y_n\}$, where $\{h_0, h_1,\ldots,h_n; I_1,\ldots,I_n\}$ are considered missing. The observations, $Y_n = \{y_1,\ldots,y_n\}$, are considered the incomplete data, and in general, we define $Y_t = \{y_1,\ldots,y_t\}$.

The result of using the filtering step will be particle filters that are random samples from $f(h_t,I_t|Y_t)$. Because the standard SV model, (1)–(3), falls into the category of general state–space models, the algorithm explained in Appendix A can be applied directly. The difference between the standard model and the mixture model is that $\{I_t\}$ is also sampled by assuming that $\{I_t\}$ is another state variable. The state equation of this model can be re-expressed as follows:

$$\begin{pmatrix} h_t \\ I_t \end{pmatrix} = \begin{pmatrix} \phi & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} h_{t-1} \\ I_{t-1} \end{pmatrix} + \begin{pmatrix} w_t \\ B_t \end{pmatrix}, \tag{6}$$

where $w_t \overset{\text{i.i.d.}}{\sim} N(0,Q)$ and $B_t \overset{\text{i.i.d.}}{\sim} B(\pi)$.

3.1. *Filtering*

The following is the algorithm for the filtering step, from which we will obtain M samples from $f(h_t, I_t|Y_t)$ for each t.

(1) Generate $f_0^{(j)} \sim N(\mu_0,\sigma_0^2)$.
(2) For $t = 1,\ldots,n$:
  (a) Generate random numbers $w_t^{(j)} \sim N(0,Q)$ and $B_t^{(j)} \sim B(\pi)$, for $j = 1,\ldots,M$.
  (b) Compute $p_t^{(j)} = \phi f_{t-1}^{(j)} + w_t^{(j)}$ and $\tilde{p}_t^{(j)} = B_t^{(j)}$.
  (c) Compute

$$w_t^{(j)} = f(y_t|p_t^{(j)},\tilde{p}_t^{(j)}) \propto \frac{1}{R_t^{*(j)}} \exp\left(\frac{(y_t - p_t^{(j)} - \mu_t^{*(j)})^2}{2R_t^{*(j)}}\right),$$

  where $\mu_t^{*(j)} = \tilde{p}_t^{(j)} m_1 + (1 - \tilde{p}_t^{(j)}) m_0$ and $R_t^{*(j)} = \tilde{p}_t^{(j)} R_1 + (1 - \tilde{p}_t^{(j)}) R_0$.
  (d) Generate $[f_t^{(j)}, \tilde{f}_t^{(j)}]$ by resampling with weights, $w_t^{(j)}$.

3.2. *Smoothing*

In the smoothing step, we get particle smoothers that are needed to get the expected likelihood in the expectation step of the EM algorithm. The following is

the algorithm to get particle smoothers, which is a slight modification of Godsill *et al.* (2004).

Suppose that equally weighted particles $\{(f_t^{(j)}, \tilde{f}_t^{(j)}) : j = 1, \ldots, M\}$ from $f(h_t, I_t | Y_t)$ are available for $t = 1, \ldots, n$ from the filtering step.

(1) Choose $[s_n^{(j)}, \tilde{s}_n^{(j)}] = [f_n^{(i)}, \tilde{f}_n^{(i)}]$ with probability $1/M$.
(2) For $n - 1$ to 0,
    (a) Calculate

$$w_{t|t+1}^{(i)} \propto f(s_{t+1}^{(j)}, \tilde{s}_{t+1}^{(j)} | f_t^{(i)}, \tilde{f}_t^{(i)}) \propto \exp\left(-\frac{(s_{t+1}^{(j)} - \phi f_t^{(i)})^2}{2Q}\right) \pi^{\tilde{s}_{t+1}^{(j)}} (1-\pi)^{1-\tilde{s}_{t+1}^{(j)}}$$

    for each i.
    (b) Choose $[s_t^{(j)}, \tilde{s}_t^{(j)}] = [f_t^{(i)}, \tilde{f}_t^{(i)}]$ with probability $w_{t|t+1}^{(i)}$
(3) Repeat (1)–(2) for $j = 1, \ldots, M$.

At the end of this smoothing step, we have $\{s_0^{(j)}, \ldots, s_n^{(j)}, \tilde{s}_1^{(j)}, \ldots, \tilde{s}_n^{(j)}; j = 1, \ldots, M\}$, which are random samples from $f(h_0, \ldots, h_n, I_1, \ldots, I_n | Y_n)$. These values can then be used for the estimation step, which we discuss next.

3.3. *Estimation*

In the estimation step, we update the parameters via the EM algorithm. To apply the EM algorithm, we must calculate the complete data likelihood, $L(\theta)$, and the expected likelihood given the data, Y_n, and the current value of the parameters, say θ'.

The complete data likelihood is given by

$$\begin{aligned} L(\theta) &= f_\theta(h_0) \prod_{t=1}^{n} f_\theta(h_t | h_{t-1}) \prod_{t=1}^{n} f_\theta(I_t) \prod_{t=1}^{n} f_\theta(y_t | h_t, I_t) \\ &= \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(h_0 - \mu_0)^2}{2\sigma_0^2}\right) \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi Q}} \exp\left(-\frac{(h_t - \phi h_{t-1})^2}{2Q}\right) \\ &\quad \times \prod_{t=1}^{n} \pi^{I_t} (1-\pi)^{1-I_t} \prod_{t=1}^{n} \frac{1}{\sqrt{2\pi R_t^*}} \exp\left(-\frac{(y_t - h_t - \mu_t^*)^2}{2R_t^*}\right), \end{aligned}$$

where $R_t^* = I_t R_1 + (1 - I_t) R_0, \mu_t^* = I_t m_1 + (1 - I_t) m_0$.

The expected conditional complete data likelihood is given by

$$\begin{aligned}\mathcal{Q}(\theta|\theta') &= E(-2\log L(\theta)|Y_n,\theta') \\ &= \log\sigma_0^2 + \frac{(h_0^n-\mu_0)^2+p_0^n}{\sigma_0^2} \\ &\quad + \sum_{t=1}^{n}\left[\log Q + \frac{(h_t^n-\phi h_{t-1}^n)^2 + P_t^n + \phi^2 P_{t-1}^n - 2\phi P_{t,t-1}^n}{Q}\right] \\ &\quad - 2\left[\log\pi\sum_{t=1}^{n}\pi_t^n + \log(1-\pi)\left(n-\sum_{t=1}^{n}\pi_t^n\right)\right] \\ &\quad + \sum_{i=0}^{1}\left[\sum_{t=1}^{n}\pi_{ti}^n\log R_i\right] + \sum_{i=0}^{1}\left[\frac{1}{R_i}\sum_{t=1}^{n}\{E(I_{ti}(y_t-h_t-m_i)^2|Y_n)\}\right], \quad (7)\end{aligned}$$

where $h_t^n = E[h_t \mid Y_n,\theta'], P_{t,s}^n = E[(h_t-h_t^n)(h_s-h_s^n)\mid Y_n,\theta'], P_t^n = P_{t,t}^n$, $\pi_{t1}^n = \pi_t^n = E[I_t \mid Y_n,\theta']$, $\pi_{t0}^n = 1-\pi_t^n$, $I_{t1} = I_t$, $I_{t0} = 1-I_t$.

By minimizing eqn (7), we get the following estimates.

$$\hat{\phi} = \frac{S_{10}}{S_{00}}, \quad \hat{Q} = \frac{1}{n}\left(S_{11} - \frac{S_{10}^2}{S_{00}}\right), \quad \hat{\pi} = \frac{\sum_{t=1}^n \pi_t^n}{n},$$

$$\hat{m_0} = \frac{\sum_{t=1}^n E[(1-I_t)(y_t-h_t)|Y_n,\theta']}{n-\sum_{t=1}^n \pi_t^n}, \quad \hat{m_1} = \frac{\sum_{t=1}^n E[I_t(y_t-h_t)|Y_n,\theta']}{\sum_{t=1}^n \pi_t^n},$$

$$\hat{R_0} = \frac{\sum E[(1-I_t)((y_t-h_t-\hat{m}_0)^2|Y_n,\theta')]}{n-\sum_{t=1}^n \pi_t^n}, \quad \hat{R_1} = \frac{\sum E[(I_t)((y_t-h_t-\hat{m}_1)^2)|Y_n,\theta']}{\sum_{t=1}^n \pi_t^n},$$

where

$$S_{00} = \sum_{t=1}^{n}\left((h_{t-1}^n)^2 + P_{t-1}^n\right), \quad S_{11} = \sum_{t=1}^{n}\left((h_t^n)^2 + P_t^n\right) \text{ and } S_{10} = \sum_{t=1}^{n}(h_{t-1}^n h_t^n + P_{t,t-1}^n).$$

Because the smoothing step returns the particle smoothers, we can obtain the EM estimates by plugging the sample means of the functions of particle smoothers for the expectations. For example,

$$\frac{\hat{h}_t^n = \sum_{j=1}^M s_t^{(j)}}{M}.$$

3.4. *Initial parameter selection*

A common criticism of the EM algorithm is that convergence can be quite slow (see Mclachlan and Krishnan, 1997). In order to save computing time, it is essential to start with good initial parameters. Anderson *et al.* (1969) suggested consistent estimates of the parameters of a linear system based on the idea of the method of moments. We apply their idea to get initial parameters.

For the mixture SV model, there are seven parameters that need initial values. To apply the method of moments, seven equations should be solved, and so it is a complex job. Therefore, we assume that we can choose reasonable values for the initial parameters for the normal mixtures and applied the method of moments for three other parameters. In many cases involving univariate data, the choice of the starting values will not be critical for estimating the parameters of normal mixtures (see Everitt and Hand, 1981).

(1) Pick arbitrary initial parameters for $m_1 - m_0 = k_1, R_0^{(0)} = k_2, R_1^{(0)} = k_3, \pi^{(0)} = k_4$. As a general procedure, we have found choosing $k_1 = -3$, $k_2 = k_3 = 4$, and $k_4 = 0.5$ reasonable starting values; these values are used in the examples of Section 5.
(2) Set $m_0^{(0)} = \bar{y} - \pi^{(0)}(m_1 - m_0) = \bar{y} - k_4 k_1$ and $m_1^{(0)} = k_1 + m_0^{(0)}$.
(3) Set

$$\phi^{(0)} = \frac{\hat{\gamma}_y(2)}{\hat{\gamma}_y(1)} \quad \text{where} \quad \hat{\gamma}_y(h) = n^{-1} \sum_{t=1}^{n-h} (y_{t+h} - \bar{y})(y_t - \bar{y}).$$

(4) Set

$$Q^{(0)} = n^{-1} \sum_{t=2}^{n} \{(y_t - \bar{y}) - \phi(y_{t-1} - \bar{y})\}^2 - \hat{\sigma}_v^2(1 - \hat{\phi}^2).$$

Here, $\hat{\sigma}_v^2 = (1 - k_4)(k_1^2 k_4 + k_2) + k_3 k_4$; alternatively, one may set $\hat{\sigma}_v^2 = 5$, which is approximately the variance of the log-χ_1^2 distribution.

3.5. *Relative likelihood*

Next, we introduce the relative likelihood that we use to assess convergence. It is known that the likelihood of the observed data increases at every iteration of the EM algorithm. But, because our *E*-step uses particles to calculate the expected likelihood, the monotone likelihood property is not guaranteed. Hence, it is essential to monitor the behaviour of the relative likelihood. Generically, let y represent the incomplete data and (x, y) represent the complete data. The relative likelihood is the ratio of the likelihoods at two adjacent iterations, and the relative likelihood at the ith iteration, $f_{\theta^{(i)}}(y)/f_{\theta^{(i-1)}}(y)$, can be calculated by using the complete likelihood as follows:

$$\frac{f_{\theta^{(i)}}(y)}{f_{\theta^{(i-1)}}(y)} = \frac{f_{\theta^{(i)}}(x,y)}{f_{\theta^{(i)}}(x \mid y)} \cdot \frac{f_{\theta^{(i-1)}}(x \mid y)}{f_{\theta^{(i-1)}}(x,y)}.$$

Multiplying by $f_{\theta^{(i)}}(x \mid y)$ and integrating out x, we get

$$\frac{f_{\theta^{(i)}}(y)}{f_{\theta^{(i-1)}}(y)} = E_{\theta^{(i-1)}}\left[\frac{f_{\theta^{(i)}}(x,y)}{f_{\theta^{(i-1)}}(x,y)} \middle| y\right].$$

Hence, the change in the log-likelihood is

$$\Delta l_y(\theta^{(i-1)}, \theta^{(i)}) = \log f_{\theta^{(i)}}(y) - \log f_{\theta^{(i-1)}}(y) = \log \frac{f_{\theta^{(i)}}(y)}{f_{\theta^{(i-1)}}(y)}$$
$$= \log E_{\theta^{(i-1)}}\left[\frac{f_{\theta^{(i)}}(x,y)}{f_{\theta^{(i-1)}}(x,y)}\middle| y\right]. \tag{8}$$

If we have samples $\{x_j\}_{j=1}^{M}$ from $f_{\theta^{(i-1)}}(x|y)$, then $\Delta l_y(\theta^{(i-1)})$ can be estimated by

$$\hat{\Delta} l_y(\theta^{(i-1)}, \theta^{(i)}) = \log\left(\frac{1}{M}\sum_{j=1}^{M}\frac{f_{\theta^{(i)}}(x_j,y)}{f_{\theta^{(i-1)}}(x_j,y)}\right). \tag{9}$$

This is a slight modification of Chan and Ledolter (1995).

3.6. *Stopping rule*

Theoretically, the suggested algorithm converges when the particle size, M, and the number of iterations, N, are large. Practically, it is not possible to use infinitely large M and N. The choice of N is important because the estimates from the procedure when stopped too early may not be reliable. Moreover, it is waste of time and resources to run the procedure longer than necessary. Many numerical procedures involving iterative solutions compare estimates from contiguous iterations; if the two are close enough, the process is considered to have converged, and is stopped. Equivalently, the relative likelihood can be considered as a measure in assessing convergence. In particular, a small relative likelihood signifies that a process may be approaching convergence. We recommend that the procedure be stopped if the relative likelihood is less than the predetermined tolerance level ϵ. Regarding the selection of M, Tanner (1996) mentioned that it is wise to start with a small M, increasing it as the current approximation moves closer to the maximum likelihood estimate in the Markov chain EM setting. We apply Tanner's method to save computing time.

3.7. *Standard errors*

Standard errors can be estimated from the inverse of observed information. Again, we generically denote the incomplete data as y and the complete data as (x, y). According to Louis (1982), because $f(y|\theta)$ is hard to handle in the EM setting, the observed information is calculated using the complete likelihood. The observed information of the data, y, is

$$-\frac{\partial^2 \log f(y|\theta)}{\partial\theta\partial\theta'} = E\left[-\frac{\partial^2}{\partial\theta\partial\theta'}\log f(x,y|\theta)\middle| y\right] - E\left[\left(\frac{\partial}{\partial\theta}\log f(x,y|\theta)\right)\left(\frac{\partial}{\partial\theta'}\log f(x,y|\theta)\right)\middle| y\right] + E\left[\frac{\partial}{\partial\theta}\log f(x,y|\theta)\middle| y\right] E\left[\frac{\partial}{\partial\theta'}\log f(x,y|\theta)\ \middle|\ y\right]. \quad (10)$$

Therefore, if x_i are sampled from $f(x|y)$, the observed information can be estimated by

$$-\frac{\partial^2 \widehat{\log f}(y|\theta)}{\partial\theta\partial\theta'} = \frac{1}{M}\sum_{i=1}^{M} -\frac{\partial^2}{\partial\theta\partial\theta'}\log f(x_i,y|\theta) - \frac{1}{M}\sum_{i=1}^{M}\left(\frac{\partial}{\partial\theta}\log f(x_i,y|\theta)\right)\left(\frac{\partial}{\partial\theta'}\log f(x_i,y|\theta)\right) + \left(\frac{1}{M}\sum_{i=1}^{M}\frac{\partial}{\partial\theta}\log f(x_i,y|\theta)\right)\left(\frac{1}{M}\sum_{i=1}^{M}\frac{\partial}{\partial\theta'}\log f(x_i,y|\theta)\right). \quad (11)$$

Once we obtain the observed information matrix, the variance–covariance matrix can be obtained by taking the inverse. However, when we applied eqn (11) to simulated and real data sets, we met a practical problem; the information matrices may not be positive-definite. We suggest using a 5% trimmed mean instead of the sample mean for the elements of the second term in eqn (11) to solve this problem. (For details and justification, see Appendix C of Kim, 2005.)

4. IRREGULAR OR MISSING DATA

In this section, we consider the case in which the sampling period, Δt, may be irregular. In particular, we assume that the irregular sampling is of the type where, for the most part, $\Delta t = 1$ unit of time, but where it may be possible that $\Delta t = k$ units of time, where $k = 2, 3, \ldots$, is relatively small. In this case, we write the data as $y_1, \ldots, y_n$, where it is possible that certain values may not be observed. The fundamental idea is to make the data complete by filling in unobserved values, and then use the method presented in Section 3.

To cover the possibility of irregular data, we write the model as

$$h_t = \phi h_{t-1} + w_t, \quad y_t = a_t h_t + v_t, \quad (12)$$

where $a_t = 1$ if there is an observation at time t, and $a_t = 0$ otherwise. As before, we have $w_t \overset{\text{i.i.d.}}{\sim} N(0, Q)$, $v_t \overset{\text{i.i.d.}}{\sim} I_t N(m_1, R_1) + (1 - I_t)N(m_0, R_0)$, $I_t \overset{\text{i.i.d.}}{\sim} B(\pi)$, and $h_0 \sim N(\mu_0, \sigma_0^2)$.

The only difference in the algorithm for fitting the model in the irregular data case is in the data-completion step. In the data-completion step, we fill in values

with those generated from the model and then proceed as in the case where no observations are missed. When $a_t = 0$ in eqn (12), the observation equation can be simplified as $y_t = v_t$, which has no relation to the state variable, h_t. Therefore, when y_t is not observed, it can be directly generated from the observation equation if the parameters are given. With the addition of a data-completion step, we possess a data set that is regularly observed, and the method proposed in Section 3 can be applied with only slight modification. The following algorithm can be used to fit the model in the irregular data case.

Select initial parameters $\theta^{(0)} = \{\phi^{(0)}, Q^{(0)}, m_0^{(0)}, m_1^{(0)}, R_0^{(0)}, R_1^{(0)}, \pi^{(0)}\}$. For $i = 1, 2, \ldots,$:

1. *Data completion*: If an observation y_t is unobserved ($a_t = 0$), generate a random y_t from the normal mixture distribution:

$$y_t \sim I_t N(m_1^{(i-1)}, R_1^{(i-1)}) + (1 - I_t) N(m_0^{(i-1)}, R_0^{(i-1)}),$$

   where $I_t \sim B(\pi^{(i-1)})$. This step uses single imputation, although it is possible to use multiple imputation here.
2. *Filtering*: Obtain the particle filters $(f_t^{(j)}, \tilde{f}_t^{(j)})$ from $f(h_t, I_t | Y_t, \theta^{(i-1)})$, $j = 1, \ldots, M$. The only modification to the filtering step of Section 3 is that, in part (c), the weights are now

$$w_t^{(j)} = f(y_t | p_t^{(j)}, \tilde{p}_t^{(j)}) \propto \exp\left((y_t - a_t p_t^{(j)} - \mu_t^{*(j)})^2 / 2R_t^{*(j)}\right) / R_t^{*(j)}.$$

3. *Smoothing*: Obtain the particle smoothers $\{s_0^{(j)}, \ldots, s_n^{(j)}, \tilde{s}_1^{(j)}, \ldots, \tilde{s}_n^{(j)}\}$ from $f(h_0, \ldots, h_n, I_1, \ldots, I_n | Y_n, \theta^{(i-1)})$, $j = 1, \ldots, M$. The steps here are identical to the smoothing steps detailed in Section 3.
4. *Estimation*: Update the estimates, $\theta^{(i)} = \{\phi^{(i)}, Q^{(i)}, m_0^{(i)}, m_1^{(i)}, R_0^{(i)}, R_1^{(i)}, \pi^{(i)}\}$ by maximizing the expected likelihood. For this step, everything is the same as the estimation step in Section 3, except that h_t is replaced by $a_t h_t$ in $\hat{m}_0, \hat{m}_1, \hat{R}_0$ and $\hat{R}_1$.
5. Repeat steps 1–4 until the stopping criterion is met.

5. SIMULATION STUDY AND DATA ANALYSIS

We apply the proposed method to two simulated data sets and two real data sets. In the simulation studies, one data set is generated from the standard version of the SV model, (1)–(3), which has the logarithm of chi-squared distribution as its observation noise. Another data set is generated from the mixture model, (1)–(5). For the data analyses, we consider two types of pound/dollar exchange rates to check the performance of the proposed method for real data. In all examples, we used the following initial values for the parameters, as specified in Section 3: $k_1 = -3$, $k_2 = k_3 = 4$, $k_4 = 0.5$.

We note here that the fitting algorithm discussed in Section 3 may easily be applied to fitting the standard SV model, (1)–(3). In this case, $\{h_1, \ldots, h_n; y_1, \ldots, y_n\}$ are the complete data, and $Y_n = \{y_1, \ldots, y_n\}$, are the incomplete data. If the expected likelihood of the complete data given Y_n is available, parameter estimates can be obtained by maximizing it. To get an expected likelihood, we must calculate quantities such as $h_t^n = E(h_t|Y_n)$ and $P_t^n = E\{(h_t - h_t^n)^2 \mid Y_n\}$. These quantities can be calculated by using the particle filtering and smoothing algorithms given in Appendix A as follows: Let the initial parameters be $\theta^{(0)}$. For iteration $i = 1, 2, \ldots$:

Filtering step: obtain the particle filters via Algorithm A.1;
Smoothing step: obtain the particle smoothers via Algorithm A.2;
Estimation step: obtain the estimated parameters, $\theta^{(i)}$, by maximizing the expected likelihood, which is calculated using the particle smoothers;

repeat until the stopping rule is met. We will use this method for fitting the standard SV model in this section.

When fitting the mixture SV model, (1)–(5), we will use the method presented in Section 3. When comparing results for the mixture SV model with the results for the standard SV model, we will sometimes use the following equivalent representation of the observation equation given in eqn (4), that is,

$$y_t = \alpha + h_t + v_t$$

where $v_t = I_t z_{t1} + (1 - I_t) z_{t0} - \mu\pi$, $z_{t0} \overset{\text{i.i.d.}}{\sim} N(0, R_0)$, $z_{t1} \overset{\text{i.i.d.}}{\sim} N(\mu, R_1)$, and $I_t \overset{\text{i.i.d.}}{\sim} B(\pi)$.

5.1. *Simulation studies*

In the simulation studies, we generated two sets of data as follows:

Simulation A: Data were generated from the standard SV model, (1)–(3), with the true parameter set of $(\phi, Q, \alpha) = (0.9, 1, -3)$.
Simulation B: Data were generated from the mixture SV model, (1)–(5), with the true parameter set of $(\phi, Q, m_0, m_1, R_0, R_1, \pi) = (0.8, 1.5, -4, -7, 3, 5, 0.5)$.

We use the second data set to observe the behaviour of the estimation procedure when there is a departure from the log-χ_1^2 observational error assumption. In each case we used a sample size of $n = 1000$; to guarantee the processes had reached stability, we generated more than 1000 observations and discarded the initial values. For the missing data cases, we randomly removed observations and then applied the method in Section 4 to examine the performance of the proposed method for missing data cases.

Table I shows the results of the estimation for Simulation A. The processes were stopped when the value of relative likelihood was less than ϵ, which is

TABLE I

ESTIMATION RESULTS FOR SIMULATION A

True	ϕ 0.9	Q 1	m_0	m_1	R_0	R_1	π	α -3
[1]	0.8783	1.4059						−2.5659
	(0.0184)	(0.1425)						(0.1109)
[2]	0.9077	1.0180	−1.3622	−4.1971	1.8067	9.1332	0.3160	−2.2579
	(0.0151)	(0.0950)	(0.1049)	(0.3013)	(0.2614)	(1.0564)	(0.0359)	
[3]	0.8730	0.9679	−1.7672	−4.9413	2.8191	10.8303	0.2705	−2.6259
	(0.0224)	(0.1511)	(0.2169)	(0.5323)	(0.3613)	(1.5125)	(0.0566)	
[4]	0.8608	0.9064	−1.5988	−4.6787	2.8207	11.6589	0.3017	−2.5281
	(0.0272)	(0.1756)	(0.3191)	(0.4729)	(0.4806)	(1.5701)	(0.0538)	

[1] *Standard SV model:* no missing observations, $M = 1000$, $\epsilon = 0.001$. [2] *Mixture SV model:* no missing observations, $M = 1000$, $\epsilon = 0.001$. [3] *Mixture SV model:* 10% missing, $M = 4000$, $\epsilon = 0.001$. [4] *Mixture SV model:* 20% missing, $M = 2000$, $\epsilon = 0.01$. Standard errors are shown in parenthesis, M is the number of particles and ϵ is tolerance which assesses convergence.

specified in the table. It can be said that the estimation procedure based on the assumption of the standard SV model works well in the sense that the estimates are close to the true parameters (see [1] in Table I). Furthermore, the results show that the method based on the mixture model gives good estimates even if the true observation noise is not a normal mixture distribution (see [2] in Table I). Next, we randomly removed 10% of the data and applied the method presented in Section 4; the results are listed in [3] of Table I. The results listed in [4] of Table I apply to the case where 20% of the data are missing. Although, as the rate of missing data increases, it gets harder to achieve a certain tolerance and a bigger number of particles is needed, we can see that our proposed method handles missing data cases well. Estimated standard errors of the parameter estimates are also presented in Table I.

We then performed a similar simulation study using the data from Simulation B, and the results are presented in Table II. The results in [1] of Table II show

TABLE II

ESTIMATION RESULTS FOR SIMULATION B

True	ϕ 0.8	Q 1.5	m_0 -4	m_1 -7	R_0 3	R_1 5	π 0.5	α
[1]	0.5833	4.3719						5.7427
	(0.0347)	(0.3183)						(0.1156)
[2]	0.7654	1.8131	−4.3240	−7.2851	3.0971	4.9751	0.4761	
	(0.0303)	(0.2188)	(0.1611)	(0.2361)	(0.4034)	(0.5950)	(0.0408)	
[3]	0.7231	1.8238	−4.3254	−7.3412	3.3379	5.2886	0.4810	
	(0.0355)	(0.2113)	(0.0903)	(0.2830)	(0.6774)	(0.8861)	(0.0117)	
[4]	0.7629	1.3329	−4.1378	−7.1377	3.2313	5.2576	0.4837	
	(0.1824)	(1.6103)	(1.0013)	(0.5510)	(0.5597)	(2.6681)	(0.0559)	

[1] *Standard SV model:* no missing observations, $M = 1000$, $\epsilon = 0.001$. [2] *Mixture SV model:* no missing observations, $M = 2000$, $\epsilon = 0.001$. [3] *Mixture SV model:* 10% missing, $M = 2000$, $\epsilon = 0.01$. [4] *Mixture SV model:* 20% missing, $M = 2000$, $\epsilon = 0.05$. Standard errors are shown in parenthesis, M is the number of particles and ϵ is tolerance which assesses convergence.

that when the model is the standard SV model, applying the techniques based on that model may lead to poor estimates; i.e. when the $\log(\chi_1^2)$ assumption is not met, the result from the method based on that assumption is not reliable. When the algorithm based on the mixture SV model is applied, the estimates are close to the true parameters (see [2] of Table II. Moreover, the method based on the normal mixture SV model works well in both cases (Simulation A and Simulation B). From the results in [3] and [4] of Table II, we can then see that our proposed method handles the missing data cases well.

5.2. *Pound and dollar daily exchange rates*

In this section we present the analysis of two data sets. The first example, which was used in Harvey *et al*. (1994), is an analysis of the pound–dollar daily exchange rates from 1 October 1981 to 28 June 1985. We analysed these data to compare the performance of the mixture SV model, (1)–(4), along with our proposed estimation method to that of previously suggested methods.

Table III compares the estimates based on the particle–EM procedures we have described under the standard SV model, (1)–(3), and under the mixture SV model, (1)–(4), to the analyses presented in Doucet and Tadic (2003) and in Durbin and Koopman (2000). In particular, Doucet and Tadic (2003) used a batch ML algorithm for 1000 iterations with 10,000 particles, whereas Durbin and Koopman (2000) estimated parameters by approximating the likelihood to the linear Gaussian model. The results of the estimation for each different approach shown in Table III are similar. These results further validate the use of the mixture SV model, (1)–(4), and the techniques we proposed to fit the model to data. Finally, we note that the particle–EM method is time consuming; for example, with the mixture SV model and $M = 500$, the algorithm converged in 145 iterations; the filtering step took about 48 seconds to complete on average, the smoothing step took about 145 seconds to complete on average, and the estimation step took about half a second to complete on average. Clearly, the smoothing step is the slowest part of the algorithm, but we note that the calculations were performed using Matlab on a PC with a Pentium 4 CPU (3.20 GHz) and 512 MB of RAM.

In a second analysis, we consider another pound–dollar daily exchange rate data set to study the performance of the proposed method when there is irregular

TABLE III

ESTIMATION RESULTS FOR THE POUND/DOLLAR EXCHANGE RATES

Method	ϕ	α	Q
Doucet and Tadic	0.968	−2.1737	0.0353
Durbin and Koopman	0.973	−2.1863	0.0299
Standard SV Model (1)–(3)	0.976	−2.2320	0.0255
Mixture SV Model (1)–(4)	0.978	−2.2393	0.0228

sampling. These data are British pound and US dollar exchange rates taken from Franses and van Dijk (2000), and expressed as the number of units of foreign currency per US dollar. We focused on the period of 1 January 1 1996 to 31 December 1998, in which case there are 784 daily observations and 29 of them are missing. If e_t represents the daily exchange rate, and $r_t = \log(e_t/e_{t-1})$ denotes the log return exchange rate, then our observations are defined as $y_t = \log\{(r_t - \bar{r})^2\}$ where $\bar{r}$ is the sample average return. After the transformation was used to obtain the logged squared returns, y_t, there were 57 missing observations.

Figure 1 shows the plots of the data. The method presented in Section 4 was used to fit the mixture SV model, (1)–(5). We started with 500 particles and increased them until we achieved the tolerance 0.01. Figure 2 shows the trajectory of parameter estimates and the history of the relative likelihood. The final estimates, along with their standard errors, were

$$\hat{\phi} = 0.896\,(0.027),\ \hat{Q} = 0.070\,(0.010),\ \hat{m}_0 = -11.502\,(0.099),\ \hat{m}_1 = -14.065\,(0.267),$$
$$\hat{R}_0 = 1.817\,(0.161),\ \hat{R}_1 = 5.420\,(0.695),\ \hat{\pi} = 0.320\,(0.034).$$

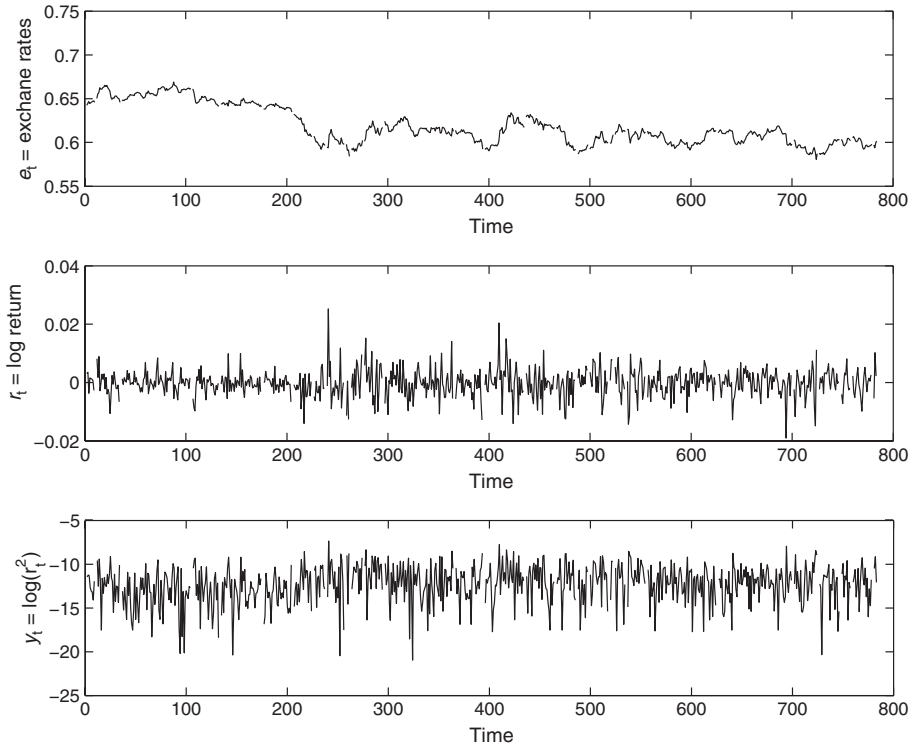

FIGURE 1. Pound/dollar daily exchange rates with missing values; the exchange rates, e_t (top), the log returns, r_t (middle) and the transformed log returns, y_t (bottom).

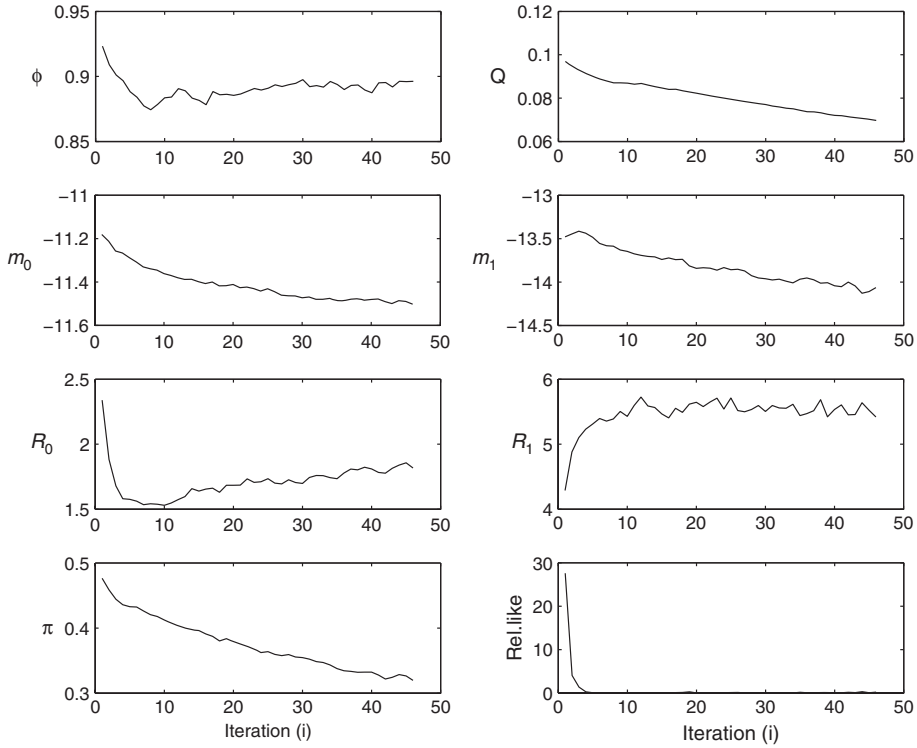

FIGURE 2. Estimation results for the pound/dollar exchange rates with missing values.

It is not easy to compare our results with those in Franses and van Dijk (2000) because, to avoid missing data problems, they analysed weekly data by using the exchange rate on Wednesday unless a Wednesday exchange rate was not available. In this case, they used the exchange rate on Tuesday, or on Thursday if Tuesday's data were also missing.

6. DISCUSSION

In this article, we presented a mixture SV model and a strategy for fitting the model that combines the EM algorithm and particle methods. In addition, we presented a modification that allows the analysis of irregularly sampled data. In addition, we showed how to use our estimation procedure when fitting standard SV models. Simply speaking, we changed the observation error from the log-χ^2_1 distribution to a more comprehensive distribution, a mixture distribution with two normal components. This change to the standard model allows for a more robust fit. Finally, we presented simulation studies and real data analyses to exhibit the viability of our proposed method.

APPENDIX A: PARTICLE FILTERS AND SMOOTHERS

In this appendix, we introduce the general concepts of particle filtering and smoothing for state–space models. The main idea of particle methods is to represent the probability distribution of Z, which is usually hard to obtain directly, using M particles and associated weights, $\{z^{(j)}, w^{(j)}\}_{j=1}^{M}$, so that the empirical distribution of particles, $\sum_{j=1}^{M} w^{(j)} I(Z \leq z^{(j)}) / \sum_{j=1}^{M} w^{(j)}$, replaces the cumulative density function (cdf) of Z, $F(z)$, where $I(A)$ is an indicator function whose value is 1 if A is true, and 0 otherwise.

Particle filters and smoothers are sequential Monte Carlo methods grounded in particle representations, and they can be considered as generalizations of well-known Kalman filters and smoothers for general state–space models. Particle filters and smoothers provide particles with associated weights to approximate the conditional density $f(h_t|Y_s)$, where $Y_s = \{y_1, \ldots, y_s\}$. This set of particles makes it possible to approximate anything related to their true density.

The basic strategy used to get particles of the desired density is based on sequential importance sampling and resampling. Sequential importance sampling (SIS) is a Monte Carlo method that forms the basis for most particle methods. The problem with SIS is degeneracy: after a few iterations, most of the particles have very small weights and they contribute little to the desired density $p(h_t|Y_t)$. Resampling is designed to solve this problem by removing particles with small weights, concentrating, instead, on particles with large weights. A generic particle filter draws $\{h_t^{(j)}\}_{j=1}^{M}$ using an SIS particle filter, and resamples $\{h_t^{(j*)}\}_{j=1}^{M}$ when degeneracy has occurred. Arulampalam *et al.* (2002) and Doucet *et al.* (2001) are good sources of information on particle filters and their several variants.

A.1. PARTICLE FILTERING ALGORITHM

The particle filtering algorithm returns $\{f_t^{(j)}, w_t^{(j)}\}_{j=1}^{M}$, for $t = 1, \ldots, n$, where $f_t^{(j)}$ represents jth realization of particle filter and $w_t^{(j)}$ represents an associated weight, i.e.

$$f(h_t|Y_t) \approx \sum_{j=1}^{M} w_t^{(j)} \delta(h_t - f_t^{(j)}).$$

Kitagawa (1996) and Kitagawa and Sato (2001) suggested an algorithm for filtering in general state–space models: $h_t = F_t(h_{t-1}, w_t)$, $y_t = H_t(h_t, v_t)$, where F_t and H_t are known functions that may depend on parameters θ; here, w_t and v_t are independent white-noise processes and $h_0 \sim p_0(x)$, $w_t \sim q(w)$. The following is the summary of the algorithm.

Algorithm A.1: Monte Carlo filter for general state–space models

(1) Generate a random number $f_0^{(j)} \sim p_0(x)$ for $j = 1, \ldots, M$.
(2) Repeat the following steps for $t = 1, \ldots, n$, where n is the length of data.
  (a) Generate a random number $w_t^{(j)} \sim q(w)$, for $j = 1, \ldots, M$.
  (b) Compute $p_t^{(j)} = F_t(f_{t-1}^{(j)}, w_t^{(j)})$, for $j = 1, \ldots, M$.
  (c) Compute $w_t^{(j)} = p(y_t \,|\, p_t^{(j)})$, for $j = 1, \ldots, M$.
  (d) Generate $f_t^{(j)}$, for $j = 1, \ldots, M$ by resampling $p_t^{(1)}, \ldots, p_t^{(M)}$, with weights $\{w_t\}$.

Here, M is the number of particles and $p(\cdot|\cdot)$ refers to a generic conditional density. This algorithm is the same as that of the sequential importance resampling (SIR) filter, which can be easily derived from the SIS filter. The SIR filter chooses the conditional density of h_t, given h_{t-1}, $f(h_t|h_{t-1})$ as a sequential importance density, $q(h_t|h_{t-1}, Y_t)$, and applies the resampling step at every time index.

A.2. PARTICLE SMOOTHING ALGORITHM

A primary goal of particle smoothing algorithms is to get particle smoothers, $\{s_t^{(j)}\}_{j=1}^M$, with associated weights, $\{w_t^{(j)}\}_{j=1}^M$, where

$$f(h_t|Y_n) \approx \sum_{j=1}^{M} w_t^{(j)} \delta(h_t - s_t^{(j)}).$$

Theoretically, the trajectories of states $\{s_1, \ldots, s_n\}$ from the $f(h_1, \ldots, h_n|Y_n)$ can be obtained by an usual particle filtering method (e.g. given in Section A.1). However, in practice, because the diversity of the paths of the particles is reduced by the resampling step, smoothed estimates degenerate. In other words, only a small number of particles appear repeatedly in earlier time points if n gets bigger. Kitagawa and Sato (2001) suggested resampling only part of the data instead of sampling the whole path from 0 to t at time point t in order to prevent degeneracy.

Godsill *et al.* (2004) suggested a new smoothing method called 'particle smoother using backwards simulation'. Unlike other smoothing methods, this method is free from degeneracy and concerns the entire trajectory of states $\{h_0, \ldots, h_n\}$ from the joint density $f(h_0, \ldots, h_n|Y_n)$, not just the individual marginal smoothing density, $f(h_t|Y_n)$. This particle smoother using backward simulation assumes that the filtering has already been performed so that the particles and associated weights, $\{f_t^{(j)}, w_t^{(j)}\}_{j=1}^M$, can approximate the filtering density, $f(h_t|Y_t)$, by $\sum w_t^{(j)} \delta(h_t - f_t^{(j)}) / \sum w_t^{(j)}$. The following is the algorithm from Godsill *et al.* (2004); by repeating this algorithm M times, we get M trajectories of the states, given the data.

Algorithm A.2: particle smoother using backward simulation

Suppose weighted particles $\{(f_t^{(j)}, w_t^{(j)})\}_{j=1}^M$ are available for $t = 1, \ldots, n$.

For $j = 1, \ldots, M$:

(1) Choose $s_n^{(j)} = f_n^{(i)}$ with probability $w_n^{(i)}$.

(2) For $n - 1$ to 1,

(a) Calculate $w_{t|t+1}^{(i)} \propto w_t^{(i)} f(s_{t+1}^{(j)} | f_t^{(i)})$ for each i.

(b) Choose $s_t^{(j)} = f_t^{(i)}$ with probability $w_{t|t+1}^{(i)}$.

(3) $s_{1:n}^{(j)} = \{s_1^{(j)}, \ldots, s_n^{(j)}\}$ is an approximate realization from $p(h_1, \ldots, h_n|Y_n)$.

APPENDIX B: PROOF OF CONVERGENCE OF THE PROPOSED ALGORITHMS

To establish the convergence properties of the suggested algorithm presented in Section 3, we adopt the approach of Chan and Ledolter (1995). For completeness, we restate their theorem in terms of our problem. To ease the notation, we let $X = (x,y)$ represent the complete data and y represent the incomplete data.

THEOREM (CHAN AND LEDOLTER, 1995). *Suppose that assumptions A1–A4 stated below are satisfied. Let θ^* be an isolated local maximizer of $l_y(\theta)$, the observed likelihood. Then there exists a neighbourhood, say V_1, of θ^* such that for starting values of the algorithm presented in Section 3 inside V_1 and for all $\epsilon_1 > 0$, there exists a k_0 such that $P(|\theta_k^{(M)} - \theta * | < \epsilon_1$ for some $k \leq k_0) \to 1$ as $M \to \infty$.*

ASSUMPTIONS

A1: For all θ', $l_X(\theta') = q(Z,\theta')$, where Z is a measurable vector function of X, q is linear in Z and $l_X(\theta')$ is the complete likelihood at the parameter θ'.
A2: $q(Z,\cdot)$ attains the unique global maximum at $\mathcal{M}(Z)$ and $\mathcal{M}$ is continuous in Z.
A3: The convergence, in probability, of $\bar{Z}_m(\theta) \to E_\theta(Z)$ as $M \to \infty$ is uniform over compact subsets of Ω where Ω is a parameter space, $\bar{Z}_m(\theta)$ is the sample mean of Z values computed from the particle sample $X_1(\theta), \ldots, X_m(\theta)$.
A4: $l_y(\theta)$ is continuous in θ.

We establish the convergence of the algorithm presented in Section 3, which pertains to the parameter estimation for the mixture SV model, by showing that all four assumptions of the theorem are met. We note that the convergence of the algorithm for the standard SV model can be established in a similar manner.

The log-likelihood of the complete data for the mixture SV model is

$$\begin{aligned} l_X(\theta) = -\frac{1}{2}\Bigg[& Z_1 \cdot \frac{1}{Q} + Z_2 \cdot \frac{\phi^2}{Q} - 2Z_3 \cdot \frac{\phi}{Q} \\ & + Z_4 \cdot \left(-2\log\frac{\pi}{1-\pi} + \log\frac{R_1}{R_0} + \frac{m_1^2}{R_1} - \frac{m_0^2}{R_0}\right) + Z_5 \cdot \left(\frac{1}{R_1} - \frac{1}{R_0}\right) \\ & + Z_6 \cdot \left(-2\frac{m_1}{R_1} + 2\frac{m_0}{R_0}\right) + Z_7 \cdot \frac{1}{R_0} - Z_8 \cdot 2\frac{m_0}{R_0} + C(\theta)\Bigg] \\ = {} & q(Z,\theta) \end{aligned}$$

where

$$\begin{aligned} Z = (Z_1, \ldots, Z_8) = \Big(& \sum h_t^2, \sum h_{t-1}^2, \sum h_t h_{t-1}, \sum I_t, \sum (y_t - h_t)^2 I_t, \\ & \sum (y_t - h_t) I_t, \sum (y_t - h_t)^2, \sum (y_t - h_t) \Big) \end{aligned}$$

and $C(\theta) = n\log Q - 2n\log(1-\pi) + n\log R_0 + n(m_0^2/R_0)$. Since Z is a measurable vector function of X and q is linear in Z, the assumption A1 is satisfied. Moreover, $q(Z,\cdot)$ attains the unique global maximum at $\mathcal{M}(Z)$, where

$$\begin{aligned} \mathcal{M}(Z) = & \left(\underset{\phi}{\arg\max}\, q, \underset{Q}{\arg\max}\, q, \underset{\pi}{\arg\max}\, q, \underset{m_1}{\arg\max}\, q, \underset{m_0}{\arg\max}\, q, \underset{R_1}{\arg\max}\, q, \underset{R_0}{\arg\max}\, q \right) \\ = & \Bigg(\frac{Z_3}{Z_2}, \frac{Z_1 + Z_2\phi^2 - 2\phi Z_3}{n}, \frac{Z_4}{n}, \frac{Z_6}{Z_4}, \frac{Z_8 - Z_6}{n - Z_4}, \frac{m_1^2 Z_4 + Z_5 - 2m_1 Z_6}{Z_4}, \\ & \frac{m_0^2(n - Z_4) + (Z_7 - Z_5) - 2m_0(Z_8 - Z_6)}{n - Z_4} \Bigg) \end{aligned}$$

and $\mathcal{M}(Z)$ is continuous in Z. Furthermore, since

$$l_y(\theta) = \int \int l_X(\theta)\, \mathrm{d}h_1 \cdots \mathrm{d}h_n\, \mathrm{d}I_1 \cdots \mathrm{d}I_n,$$

$l_y(\theta)$ is continuous in θ. Hence, the suggested algorithm satisfies A2 and A4.

To check A3, we use the convergence result of the particle smoothing algorithm in Godsill *et al.* (2004). By Theorem A2 of Godsill *et al.* (2004), for all $t \in (1, \ldots, n)$, there exists $c_{t|n}(\theta)$ independent of M such that for any bounded, Borel-measurable function ϕ on R^n,

$$E\left[\left(\int \phi(h_1, \ldots, h_n)\, \mathrm{d}\pi^M_{1:n|n}(h_1, \ldots, h_n; \theta) - \int \phi(h_1, \ldots, h_n)\, \mathrm{d}\pi_{1:n|n}(h_1, \ldots, h_n; \theta)\right)^2\right] \leq c_{1|n}(\theta) \frac{\|\phi\|^2}{M}. \tag{B.1}$$

$$\int \phi(h_1, \ldots, h_n)\, \mathrm{d}\pi^M_{1:n|n}(h_1, \ldots, h_n; \theta) \xrightarrow{P} \int \phi(h_1, \ldots, h_n)\, \mathrm{d}\pi_{1:n|n}(h_1, \ldots, h_n; \theta). \tag{B.2}$$

Since eqn (B.1) is equivalent to (B.2) and we can assume squared returns, and hence the y_t, are bounded by a big number (practically, we use $y_t - \bar{y}$ instead of y_t to prevent the logarithm from being $-\infty$), then since Z is bounded we get $\bar{Z}_M(\theta) \xrightarrow{P} E_\theta(Z)$ (as $M \to \infty$).

Now, we need to show that for every $\epsilon > 0$ there is an integer N such that $M \geq N$ implies $P(|\bar{Z}_M(\theta) - E_\theta(Z)| \geq \epsilon) < \epsilon$ for all $\theta \in E$ where E is a compact subset of Ω to establish the uniform convergence. Since $\bar{Z}_M$ and $E_\theta(Z)$ are continuous in θ, for any $\epsilon_1 > 0$ there exists $\delta_1 > 0$ and $\delta_2 > 0$ such that $|\theta_1 - \theta_2| < \delta_1$ implies $|\bar{Z}_M(\theta_1) - \bar{Z}_M(\theta_2)| < \epsilon_1$ and $|\theta_1 - \theta_2| < \delta_2$ implies $|E_{\theta_1}(Z) - E_{\theta_2}(Z)| < \epsilon_1$. Letus consider an open covering $\{O_\alpha\}$ of E such that $O_\alpha = \{\theta : |\theta - \theta_\alpha| < \delta\}$ where δ is the minimum of δ_1 and δ_2. Then by compactness of E we can choose the finite subset of $\{O_\alpha\}$, $\{O_{\alpha'}\}, \alpha' = 1, \ldots, K$ which covers E.

Then for any θ in E we can find O_β such that $\theta \in O_\beta$ and $\beta \in \{1, \ldots, K\}$. Since $|\theta - \theta_\beta| < \delta$, we get $|\bar{Z}_M(\theta) - \bar{Z}_M(\theta_\beta)| < \epsilon_1$ and $|E_\theta(Z) - E_{\theta_\beta}(Z)| < \epsilon_1$. Therefore,

$$\begin{aligned} &P(|\bar{Z}_M(\theta) - E_\theta(Z)| \geq \epsilon) \\ &\quad \leq P(|\bar{Z}_M(\theta) - \bar{Z}_M(\theta_{\alpha'})| + |\bar{Z}_M(\theta_{\alpha'}) - E_{\theta\alpha'}(Z)| + |E_{\theta\alpha'}(Z) - E_\theta(Z)| \geq \epsilon) \\ &\quad \leq P(|\bar{Z}_M(\theta) - \bar{Z}_M(\theta_{\alpha'})| \geq \epsilon/3) + P(|\bar{z}_M(\theta_{\alpha'}) - E_{\theta\alpha'}(Z)| \geq \epsilon/3) \\ &\qquad + P(|E_{\theta\alpha'}(Z) - E_\theta(Z)| \geq \epsilon/3). \end{aligned}$$

The first and third terms of the above expression are less than $\epsilon/3$ if we let $\epsilon_1 = \epsilon/3$, and the second term is less than $\epsilon/3$ by pointwise convergence for $M \geq N_{\alpha'}$. Thus, letting $N = \max N_{\alpha'}$, we get that for every $\theta \in E$, $P(|\bar{Z}_M - E_\theta(Z| \geq \epsilon) \leq \epsilon$ for $m \geq N$. Hence, because the algorithm satisfies all four assumptions of the theorem, the convergence follows.

ACKNOWLEDGEMENTS

The work of J. Kim and D. S. Stoffer was supported, in part, by a grant from the National Science Foundation.

NOTE

Corresponding author: David S. Stoffer, Department of Statistics, University of Pittsburgh, Pittsburgh, PA 15260, USA. E-mail: stoffer@pitt.edu

REFERENCES

ANDERSON, W. N., KLEINDORFER, G. B., KLEINDORFER, P. R. and WOODROOFE, M. B. (1969) Consistent estimates of the parameters of a linear system. *The Annals of Mathematical Statistics* 40, 2064–75.

ARULAMPALAM, M. S., MASKELL S., GORDON, N. and CLAPP, T. (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing* 50, 174–87.

BAI, X., RUSSELL, J. R. and TIAO, G. C. (2003) Kurtosis of GARCH and Stochastic Volatility Models with Non-normal Innovations. *Journal of Econometrics* 114, 349–60.

CHAN K. S. and LEDOLTER, J. (1995) Monte Carlo EM estimation for time series models involving counts. *Journal of the American Statistical Association* 90, 242–52.

CHIB, S., NARDARI, F. SHEPHARD, N. (2002) Markov chain Monte Carlo methods for stochastic volatility models. *Journal of Econometrics* 108, 281–316.

DANIELSON, J. (1994) Stochastic volatility in asset prices: estimation with simulated maximum likelihood. *Journal of Econometrics* 61, 375–400.

DOUCET, A. and TADIC, B. B. (2003) Parameter estimation in general state-space models using particle methods. *Annals of Institute of Statistical Mathematics* 55, 409–22.

DOUCET, A., DE FREITAS, N. and GORDON, N. (2001) *Sequential Monte Carlo Methods in Practice*. New York: Springer.

DURBIN, J. and KOOPMAN, S. J. (2000) Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Society, Series* B, 62: 3–56.

DURHAM, G. B. (2007) SV mixture models with application to s&p 500 index returns. *Journal of Financial Economics* 85, 822–56.

EVERITT, B. S. and HAND, D. J. (1981) *Finite Mixture Distributions*. New York: Chapman and Hall.

FRANSES, P. H. and VAN DIJK, D. (2000) *Nonlinear Time Series Models in Empirical Finance*. New York: Cambridge University Press.

GODSILL, S., DOUCET, A. and WEST, M. (2004) Monte Carlo smoothing for non-linear time series. *Journal of the American Statistical Association* 199, 156–68.

HARVEY, A. C., RUIZ, E. and SHEPHARD, N. (1994) Multivariate stochastic variance models. *Review of Economic Studies* 61, 247–64.

HATHAWAY, R. J. (1985) A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *The Annals of Statistics* 13, 795–800.

JACQUIER, E., POLSON, N. G. and ROSSI, P. E. (1994) Bayesian analysis of stochastic volatility models. *Journal of Business and Economic Statistics* 12, 371–417.

KIM, J. (2005) *Parameter Estimation in Stochastic Volatility Models with Missing Data Using Particle Methods and the EM Algorithm*. PhD thesis. University of Pittsburgh, Pittsburgh, PA.

KIM, S., SHEPHARD, N. and CHIB, S. (1998) Stochastic volatility: likelihood inference and comparison with ARCH models. *The Review of Economic Studies* 65, 361–93.

KITAGAWA, G. (1996) Monte Carlo filter and smoother for non-Gaussian nonlinear state-space models. *Journal of Computational and Graphical Statistics* 5, 1–25.

KITAGAWA, G. and SATO, S. (2001) Monte Carlo smoothing and self-organising state-space model. In *Sequential Monte Carlo Methods in Practice* (eds A. DOUCET, N. DE FREITAS and N. GORDON). New York: Springer-Verlag, pp. 177–95.

LIESENFELD, R. and RICHARD, J.-F. (2003) Univariate and multivariate stochastic volatility models: estimation and diagnostics. *Journal of Empirical Finance*, 10, 505–31.

LOUIS, T. A. (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society, Series* B 44, 226–33.

MCLACHLAN, G. J. and KRISHNAN, T. (1997) *The EM Algorithm and Extensions*. New York: John Wiley and Sons, Inc.

MELLINO, A. and TURNBULL, S. (1990) Pricing foreign currency options with stochastic volatility. *Journal of Econometrics* 45, 7–39.

SANDMANN, G. and KOOPMAN, S. (1998) Estimation of stochastic volatility models via monte carlo maximum likelihood. *Journal of Econometrics* 87, 271–301.

SHEPHARD, N. (1996) Statistical aspects of ARCH and stochastic volatility. In Time Series Models: In Econometrics, Finance and Other Fields (eds D. R. COX, D. V. HINKLEY and O. E. BARNDORFF-NIELSEN). London: Chapman and Hall, pp. 1–67.

SHUMWAY, R. H. and STOFFER, D. S. (2006) *Time Series Analysis and its Applications: With R Examples*. New York: Springer.

STOFFER, D. and WALL, K. (2004) Resampling in state space models. In *State Space and Unobserved Component Models: Theory and Applications* (eds A. HARVEY, S. J. KOOPMAN, and N. SHEPHARD). Cambridge: Cambridge University Press, pp. 171–202.

TANNER, M. A. (1996) *Tools for Statistical Inference*. New York: Springer.

TAYLOR, S. J. (1982) Financial returns modelled by the product of two stochastic processes–a study of daily sugar prices, 1961–79. In *Time Series Analysis: theory and Practice*, Volume 1 (ed. O. D. ANDERSON). New York: Elsevier Science Publishing Co., Amsterdam: North-Holland Publishing Co., pp. 203–26.