Ann Inst Stat Math (2009) 61:159–179
DOI 10.1007/s10463-007-0130-8

Bayesian estimation of stochastic volatility models based on OU processes with marginal Gamma law

Sylvia Frühwirth-Schnatter · Leopold Sögner

Received: 19 December 2005 / Revised: 3 July 2006 / Published online: 28 April 2007

Abstract This paper discusses practical Bayesian estimation of stochastic volatility models based on OU processes with marginal Gamma laws. Estimation is based on a parameterization which is derived from the Rosiński representation, and has the advantage of being a non-centered parameterization. The parameterization is based on a marked point process, living on the positive real line, with uniformly distributed marks. We define a Markov chain Monte Carlo (MCMC) scheme which enables multiple updates of the latent point process, and generalizes single updating algorithm used earlier. At each MCMC draw more than one point is added or deleted from the latent point process. This is particularly useful for high intensity processes. Furthermore, the article deals with superposition models, where it discuss how the identifiability problem inherent in the superposition model may be avoided by the use of a Markov prior. Finally, applications to simulated data as well as exchange rate data are discussed.

Keywords Data augmentation · Identification · Marked point processes · Markov chain Monte Carlo

S. Frühwirth-Schnatter (✉)
Department of Applied Statistics and Econometrics,
Johannes Kepler Universität Linz,
Altenbergerstraße 69, 4040 Linz, Austria
e-mail: sylvia.fruehwirth-schnatter@jku.at

L. Sögner
Department of Management Science, Vienna University of Technology,
Theresianumgasse 27, 1040 Vienna, Austria
e-mail: soegner@ibab.tuwien.ac.at

1 Introduction

Continuous time models are nowadays widely used in modern mathematical finance, providing the basis for option pricing, asset allocation and term structure theory. A classical example is the so called Black–Scholes model (Black and Scholes 1973) which characterizes the log of an asset price $x^\star(t)$ as the solution of the stochastic differential equation

$$\mathrm{d}x^\star(t) = \{\mu + \beta\sigma^2\}\mathrm{d}t + \sigma\,\mathrm{d}W(t), \tag{1}$$

where $W(t)$ is a standard Brownian motion. Implying that aggregate returns are normally distributed with constant variance, well-known stylized features of financial time series such as heavy tails, skewness and volatility clustering are not captured by this model. To improve model (1) stochastic volatility models have been introduced:

$$\mathrm{d}x^\star(t) = \{\mu + \beta\sigma^2(t)\}\mathrm{d}t + \sigma(t)\mathrm{d}W(t), \tag{2}$$

where the volatility $\sigma^2(t)$ is allowed to change over time. Various assumptions have been made concerning the stochastic nature of the volatility process, most of them based on diffusion type models, e.g. square root processes (Hull and White 1987) or Ornstein-Uhlenbeck (OU) processes for the log volatility (Andersen and Lund 1997). More recently, Barndorff-Nielsen and Shephard (2001) suggested to use non-Gaussian OU processes driven by a Lévy process to model the stochastic volatility $\sigma^2(t)$:

$$\mathrm{d}\sigma^2(t) = -\lambda\sigma^2(t)\mathrm{d}t + \mathrm{d}z(\lambda t), \tag{3}$$

where $z(\lambda t)$ is a Lévy process with independent, strictly positive increments consisting entirely of jumps. $z(\lambda t)$ is called the back-ground-driving Levy process, BDLP for short. As shown by Barndorff-Nielsen and Shephard (2001), this model, which will be called subsequently the BNS stochastic volatility model, implies volatility clusters as well as heavy tails.

The focus of the present paper is to discuss Bayesian estimation of the BNS stochastic volatility model with marginal Gamma law, which results as that special case of model (2) and (3) where the marginal distribution of $\sigma^2(t)$ is a Gamma distribution:

$$\sigma^2(t) \sim \mathcal{G}(\alpha, \delta)\,. \tag{4}$$

If the underlying asset is an integrated variable, for instance the log of foreign exchange rates, a discrete time series $(y_1, \ldots, y_T)$ of aggregated returns $y_n = x^\star(t_n) - x^\star(t_{n-1})$ could be used to estimate the unknown parameter θ driving the distributional law of the underlying asset process, like $\theta = (\alpha, \delta, \lambda)$ for a process with marginal Gamma law.

Parameter estimation for stochastic volatility models is in general known to be a difficult problem, see e.g. the excellent review in Shephard (1996), and the BNS stochastic volatility model is no exception to this rule. The problem stems from the fact

that the conditional distribution of the aggregated returns y_n, although being normal, depends on the latent processes $z(\lambda t)$ and $\sigma^2(t)$:

$$y_n|\sigma_n^2 \sim \mathcal{N}\left(\mu\Delta + \beta\sigma_n^2, \sigma_n^2\right), \tag{5}$$

where $\Delta = (t_n - t_{n-1})$ and σ_n^2 may be expressed in the following way (Barndorff-Nielsen and Shephard 2001):

$$\sigma_n^2 = \frac{1}{\lambda}\left[z(\lambda t_n) - z(\lambda t_{n-1}) - (\sigma^2(t_n) - \sigma^2(t_{n-1}))\right]. \tag{6}$$

As a consequence the one-step ahead distribution $p(y_n|y_1, \ldots, y_{n-1}, \theta)$ of y_n given observations up to t_{n-1} has no explicit form. Straightforward estimation of θ by maximizing the likelihood $p(Y|\theta) = \prod_{n=1}^{T} p(y_n|y_1, \ldots, y_{n-1}, \theta)$, where $Y = (y_1, \ldots, y_T)$, is therefore infeasible. It is common for such incomplete data problems to introduce latent variables X and carry out Bayesian inference for the augmented parameter vector (X, θ) through Markov chain Monte Carlo (MCMC) methods by sampling X from the conditional posterior density $p(X|\theta, Y)$ and sampling θ from $p(\theta|X, Y)$. Both densities are proportional to the augmented posterior density $p(X, \theta|Y)$,

$$p(X, \theta|Y) \propto p(Y|X, \theta)p(X|\theta)p(\theta),$$

where the " complete data" likelihood $p(Y|X, \theta)$ is easily obtained from (5) as the product of T densities from a normal distribution:

$$p(Y|X, \theta) = \prod_{n=1}^{T} f_{\mathcal{N}}(y_n; \mu\Delta + \beta\sigma_n^2, \sigma_n^2). \tag{7}$$

$p(X|\theta)$ is the "prior density" of the distribution of the latent variables X under θ, whereas $p(\theta)$ is the prior of θ.

Several authors have considered Bayesian estimation of the BNS stochastic volatility model with marginal Gamma law using data augmentation and MCMC methods (Barndorff-Nielsen and Shephard 2001; Roberts et al. 2004; Griffin and Steel 2006). Barndorff-Nielsen and Shephard (2001) consider the following set of latent variables as missing data:

$$X = \left(z(\lambda\Delta), \ldots, z(\lambda T\Delta), \sigma^2(\Delta), \ldots, \sigma^2(T\Delta)\right),$$

which is based on using the timing of the data, namely $t = \Delta, \ldots, T\Delta$, to discretize the latent processes $\sigma^2(t)$ and $z(\lambda t)$. This is a natural choice in the light of Eq. (6), and allows to cast the model into a state space form with the state vector consisting of the components $\sigma^2(t_n)$, $\sigma^2(t_{n-1})$, $z(\lambda t_n)$ and $z(\lambda t_{n-1})$. The problem with this choice, however, is that the prior distribution $p(X|\theta)$ has no simple analytical form. Nevertheless, as demonstrated by Griffin and Steel (2006), it is possible to run MCMC based on a Metropolis-Hastings move by using the prior $p(X|\theta)$ as a proposal density, as

long as it possible to *sample* X from the prior. Since the ratio of prior and proposal cancels from the acceptance probability, it is not necessary to evaluate the functional value of the prior $p(X|\theta)$. This strategy, however, limits the class of possible moves considerably.

Data augmentation based on a latent process X with a simple probabilistic structure was first suggested in various discussions of the paper by Barndorff-Nielsen and Shephard (2001). Roberts (2001), Papaspiliopoulos (2001), and Frühwirth-Schnatter (2001a) all pointed out that for BNS stochastic volatility model with marginal Gamma law, a timing different from the observation times may be used to discretize the latent processes $z(\lambda t)$ and $\sigma^2(t)$. A finite dimensional exact representation of these latent processes emerges from the interval representation of a Poisson process, and consequently exact data augmentation is feasible. Detailed investigations reported in Roberts et al. (2004), however, revealed that this parameterization which is called the centered one, is prone to mixing problems for models where α is large. Furthermore, the sampler tends to be trapped, if started far out in the tails of the posterior distribution. Furthermore, Roberts et al. (2004) and Papaspiliopoulos et al. (2003) show that choices of X where the prior distribution is independent of the unknown model parameters seem to be preferable. They introduce a so-called non-centered parameterization where the latent process X is defined from a marked point process living on a stripe in the real plane with exponentially distributed marks.

In the present paper we introduce in Sect. 2 an alternative non-centered parameterization for data augmentation and MCMC estimation, which is based on a marked Poisson process living on the real line with uniformly distributed marks. It is a non-centered parameterization derived from the series representation of the back-ground-driving Levy process introduced in Rosiński (2001). MCMC estimation will be discussed in detail in Sect. 3, where we introduce a new method of multiple updating of the marked Poisson process, which generalizes the single update algorithm of Geyer and Møller (1994). Section 4 deals with superposition models, where we discuss how the identifiability problem inherent in the superposition model may be avoided by the use of a Markov prior. In Sect. 5, we discuss applications to exchange rate data, whereas Sect. 6 concludes.

2 Data augmentation for OU processes with marginal Gamma law

For Bayesian estimation based on data augmentation and MCMC, we need a representation of the BDLP $z(\lambda t)$ and the volatility process $\sigma^2(t)$ in terms of random variables with a simple probabilistic structure. In Sect. 2.1 we review data augmentation based on the interval representation. In Sect. 2.2 we discuss the Rosiński series representation, which is the basis for our data augmentation scheme.

2.1 The interval representation

For the BNS stochastic volatility model with marginal Gamma law a convenient representation is the so-called interval representation (Cox and Isham 1988), which expresses the BDLP and the volatility process on the interval $[0, T\Delta]$ in terms of jump

times and jump sizes, using the fact that the BDLP is a compound Poisson process with exponentially distributed jumps. The inter arrival times between two jumps of the BDLP follow the exponential distribution $\mathcal{E}(\alpha)$. Note that the timing of the BDLP $z(\lambda t)$, introduced by Barndorff-Nielsen and Shephard (2001), is different from the timing of the volatility process $\sigma^2(t)$, leading to a marginal distribution that is independent of λ. Consequently, the volatility process $\sigma^2(t)$ jumps at $\tau_1, \ldots, \tau_j, \ldots$, with the inter arrival times $\tau_j - \tau_{j-1}$ being iid exponential $\mathcal{E}(\lambda\alpha)$, whereas the BDLP jumps at $\lambda\tau_1, \ldots, \lambda\tau_j, \ldots$. To each jump time τ_j and $\lambda\tau_j$, respectively, corresponds a jump size J_j from the $\mathcal{E}(\delta)$ distribution.

Let $N = \#\{\tau_j | \tau_j \leq T\Delta\}$ be the number of jumps occurring till $T\Delta$. Then the sequences $(\tau_1, \ldots, \tau_N)$ of jump times and $(J_1, \ldots, J_N)$ of jump sizes are a complete description of the latent processes $\sigma^2(t)$ and $z(\lambda t)$ on the interval $[0, T\Delta]$:

$$z(\lambda t) = \sum_{j=1}^{N_t} J_j, \tag{8}$$

$$\sigma^2(t) = \exp(-\lambda t)\sigma^2(0) + \sum_{j=1}^{N_t} \exp\left\{-\lambda(t - \tau_j)\right\} J_j, \tag{9}$$

where $N_t = \#\{\tau_j | \tau_j \leq t\}$. According to representation (9) the volatility at time t is an exponentially weighted sum of past "shocks" J_j with the weight being the smaller the more time passed since the shock occurred at time τ_j.

As mentioned in the introduction, data augmentation could be based on choosing the jump times and jump sizes as missing data, $X = \{(\tau_j, J_j), j = 1, \ldots, N\}$, a choice which leads to a prior $p(X|\theta)$ of closed form:

$$\log p(X|\theta) \propto N \log \delta - \delta \sum_{j=1}^{N} J_j + N \log(\lambda\alpha) - \lambda\alpha T\Delta.$$

Roberts et al. (2004), however, revealed that this parameterization is prone to mixing problems for models where α is large.

2.2 Data augmentation based on the Rosiński representation

In this section, we consider a data augmentation method, which is based on a marked Poisson process X, living on the positive real line $(0, \infty)$ with uniformly distributed marks:

$$X = \{(a_j, r_j), j = 1, 2, \ldots\}, \tag{10}$$

with a_j being the arrival times of a homogeneous Poisson process with unit intensity, while the marks r_j are independent of a_j and uniformly distributed on $[0, 1]$. We define

$$N = \#\{a_j | a_j \leq \lambda\alpha T\Delta\} \tag{11}$$

as the number of all point (a_j, r_j) in X which fulfill the condition

$$a_j \leq \lambda\alpha T\Delta. \tag{12}$$

The prior $p(X|\theta)$ of X is given by:

$$\log p(X|\theta) \propto \log\Gamma(N+1) + N\log\lambda\alpha T\Delta - \lambda\alpha T\Delta. \tag{13}$$

To translate X into the jump times and the jump sizes of the BDLP $z(\lambda t)$ and the volatility process $\sigma^2(t)$, we will use the Rosiński representation (Rosiński 2001).

2.2.1 The Rosiński series representation

A useful representation of the processes $z(\lambda t)$ and $\sigma^2(t)$ is given by the Rosiński series (Rosiński 2001), as mentioned first by Barndorff-Nielsen and Shephard (2001, p. 176), and applied by Griffin and Steel (2006) for practical Bayesian estimation. Both Barndorff-Nielsen and Shephard (2001) and Griffin and Steel (2006) considered a separate Rosiński representation of the processes $\sigma^2(t)$ and $z(\lambda t)$ for each observation interval $[(n-1)\Delta, n\Delta]$. The goal of this subsection is to show that a single Rosiński type representation of the whole processes $\sigma^2(t)$ and $z(\lambda t)$ on the entire observation interval $t \in [0, T\Delta]$ is available.

Let W be the Lévy measure of $z(1)$ and W^{-1} denote the inverse of the tail mass function W^+. Then the integral of a positive, integrable function $f(s)$ with respect to $z(s)$ is representable in law as:

$$\int_0^L f(s)\mathrm{d}z(s) \stackrel{\mathcal{L}}{=} \sum_{j=1}^{\infty} W^{-1}(a_j/L) f(Lr_j), \tag{14}$$

where $\{(a_j, r_j)\}$ are the points of the process X defined in (10). For the Gamma marginal law (4) an explicit expression for $W^{-1}(x)$ is available as:

$$W^{-1}(x) = \max\left\{0, -\frac{1}{\delta}\log\left(\frac{x}{\alpha}\right)\right\}. \tag{15}$$

From (14) and (15) we obtain:

$$\int_0^{\lambda T\Delta} f(s)\mathrm{d}z(s) \stackrel{\mathcal{L}}{=} \sum_{j=1}^{\infty} \max\left\{0, -\frac{1}{\delta}\log\left(\frac{a_j}{\alpha\lambda T\Delta}\right)\right\} f(\lambda T\Delta r_j). \tag{16}$$

As a_j is an increasing sequence and $-\log(a_j)$ is decreasing in a_j, all components in (16) with $a_j > \alpha\lambda T\Delta$ are equal to 0. Therefore, only those points (a_j, r_j) contribute

to the integral which fulfill condition (12). Integral (16) may be represented by a finite sum:

$$\int_0^{\lambda T\Delta} f(s)\mathrm{d}z(s) \stackrel{\mathcal{L}}{=} \sum_{j=1}^{N} -\frac{1}{\delta}\log\left(\frac{a_j}{\alpha\lambda T\Delta}\right) f(\lambda T\Delta r_j), \tag{17}$$

with N defined by (11). To represent $z(\lambda t)$ on $[0, T\Delta]$ we use

$$z(\lambda t) = \int_0^{\lambda T\Delta} f(s)\mathrm{d}z(s), \tag{18}$$

where $f(s) = I_{\{s\le\lambda t\}}(s)$. Applying (17) we obtain the following finite Rosiński representations for $z(\lambda t)$:

$$z(\lambda t) = \sum_{j=1}^{N} -\frac{1}{\delta}\log\left(\frac{a_j}{\alpha\lambda T\Delta}\right) I_{\{T\Delta r_j\le t\}}(r_j). \tag{19}$$

Note that the points (a_j, r_j) in (19) are ordered with respect to the first coordinate a_j. Now we reorder the points (a_j, r_j) according to r_j. Let $r_{(1)}, \ldots, r_{(N)}$ denote the order statistics of $r_1, \ldots, r_N$, let $a_{(1)}, \ldots, a_{(N)}$ be the corresponding arrival times. Then

$$z(\lambda t) = \sum_{j=1}^{N_t} -\frac{1}{\delta}\log\left(\frac{a_{(j)}}{\alpha\lambda T\Delta}\right), \tag{20}$$

where $N_t = \#\{r_{(j)} | T\Delta r_{(j)} \le t\}$. Similarly, to represent $\sigma^2(t)$ on $[0, T\Delta]$ we use:

$$\sigma^2(t) = \exp(-\lambda t)\sigma^2(0) + \int_0^{T\Delta} f(s)\mathrm{d}z(\lambda s) = \exp(-\lambda t)\sigma^2(0) + \int_0^{\lambda T\Delta} f(s'/\lambda)\mathrm{d}z(s'),$$

where $f(s) = \exp\{-\lambda(t-s)\}I_{\{s\le t\}}(s)$. Applying (17) and reordering as above we obtain the following finite Rosiński representation of the volatility process:

$$\sigma^2(t) = \exp(-\lambda t)\sigma^2(0) - \sum_{j=1}^{N_t} \exp\{-\lambda(t - T\Delta r_{(j)})\}\frac{1}{\delta}\log\left(\frac{a_{(j)}}{\alpha\lambda T\Delta}\right). \tag{21}$$

2.2.2 Data augmentation

To implement data augmentation based on the marked Poisson process X defined in (10), a relation has to established between the points in X and the jump times and the jump sizes of the volatility process. By comparing the Rosiński representation

(20) and (21) with the interval representation (8) and (9) we find how the arrival times $\{a_j, j = 1, 2, \ldots\}$ of the latent process X could be used to determine the number N of jumps in the volatility process on the interval $[0, T\Delta]$ as well as their sizes $\{J_j, j = 1, \ldots, N\}$, whereas the marks $\{r_j, j = 1, \ldots, N\}$ could be used to determine the arrival times $\{\tau_j, j = 1, \ldots, N\}$ of these jumps.

More precisely, let N be defined by (11), let $r_{(1)}, \ldots, r_{(N)}$ denote the order statistics of the marks $r_1, \ldots, r_N$, and let $a_{(1)}, \ldots, a_{(N)}$ be the corresponding sequence of arrival times. The ordered marks $r_{(j)}$ of X translate into the jump times τ_j, whereas the corresponding arrival times $a_{(j)}$ translate into the jump sizes J_j of the volatility process in the following way:

$$J_j = -\frac{1}{\delta} \log\left(\frac{a_{(j)}}{\lambda\alpha T\Delta}\right), \tag{22}$$

$$\tau_j = T\Delta r_{(j)}. \tag{23}$$

To complete the definition of the missing data, we have to add the unobservable volatility $\sigma^2(0)$ at time $t = 0$ as an additional component. As Roberts et al. (2004), we parameterize $\sigma^2(0) = X_0/\delta$ in terms of a standard volatility X_0. An equivalent transformation allows us to express the volatility process $\sigma^2(t)$ and the BDLP $z(\lambda t)$ being independent of δ for arbitrary t:

$$\sigma^2(t) = \tilde{\sigma}^2(t)/\delta\ , \quad z(\lambda t) = \tilde{z}(\lambda t)/\delta\ .$$

Consequently, σ_n^2 in (6) factorizes as $\sigma_n^2 = \tilde{\sigma}_n^2(\alpha, \lambda)/\delta$ with

$$\tilde{\sigma}_n^2(\alpha, \lambda) = \frac{1}{\lambda}\left[\tilde{z}(\lambda t_n) - \tilde{z}(\lambda t_{n-1}) - (\tilde{\sigma}^2(t_n) - \tilde{\sigma}^2(t_{n-1}))\right]. \tag{24}$$

3 MCMC estimation through data augmentation based on the Rosiński representation

In this section we will discuss Bayesian estimation of the BNS stochastic volatility model with marginal Gamma law using MCMC methods and data augmentation based on the Rosiński representation as introduced in Sect. 2.2. We will discuss only that case where the parameters μ and β appearing in (2) are fixed. For returns from exchange rates, for instance, μ as well as β may be set to 0. This leaves the parameters α, δ and λ (or some transformations involving only these parameters) to be estimated from the data. Subsequently, we consider a parameterization in terms of the parameters α and δ of the marginal Gamma law and the autocorrelation $\rho = e^{-\Delta\lambda}$ at lag 1.

We assume that all components of $\theta = (\alpha, \delta, \rho)$ are independent apriori, and use the following priors, which are closely related to the ones considered by Roberts et al. (2004) and Griffin and Steel (2006): $\alpha \sim \mathcal{G}(a_0, A_0)$, $\delta \sim \mathcal{G}(d_0, D_0)$, $\rho \sim \mathcal{B}(b_0, B_0)$. Concerning the initial volatility X_0, we assume that $X_0 \sim \mathcal{G}(a_0, 1)$ with a_0 being a fixed or random hyperparameter. In this choice we differ from the work of Roberts et al. (2004) and Griffin and Steel (2006), who assume that $\sigma^2(0)$ arises from the stationary distribution, $\sigma^2(0) \sim \mathcal{G}(\alpha, \gamma)$, and consequently $X_0 \sim \mathcal{G}(\alpha, 1)$. We found that this

stationary prior introduced quite a bias for highly persistent volatility processes, and that a non-stationary prior is preferable.

MCMC estimation iterates between the three following sampling steps:

1. sample the latent marked Poisson process X from the conditional posterior $p(X|\theta, X_0, Y)$,
2. sample the initial volatility X_0 from the conditional posterior $p(X_0|\theta, X, Y)$,
3. sample θ from the conditional posterior $p(\theta|X, X_0, Y)$.

We will provide more details in the following sections.

3.1 Sampling the missing data

Sampling realizations of the latent process $X = \{(a_j, r_j), j = 1, 2, \ldots\}$ from $p(X|X_0, \theta, Y)$ is crucial for efficient estimation of the model. Conditional on holding θ fixed, the missing data X consist of two parts X_1 and X_2 with the following properties. The first part, defined by $X_1 = \{(a_j, r_j)|a_j \leq \lambda\alpha T\Delta\}$, translates through the Rosiński representation (20) and (21) into the jump times and jump sizes in the volatility representation (9). We will use the notation $\mathcal{V}(X_1, \theta)$ to refer to the corresponding jump times and jump sizes. The second part, defined by $X_2 = \{(a_j, r_j)|a_j > \lambda\alpha T\Delta\}$, does not influence the likelihood $p(Y|X, X_0, \theta)$, when holding θ fix:

$$p(Y|X, X_0, \theta) = p(Y|\mathcal{V}(X_1, \theta), X_0, \theta).$$

Therefore the posterior of X takes the form:

$$p(X|X_0, \theta, Y) = p(X_1|\theta, X_0, Y)p(X_2|\theta),$$

where

$$p(X_1|\theta, X_0, Y) \propto p(Y|\mathcal{V}(X_1, \theta), X_0, Y)p(X_1|\theta).$$

Hence X_1 and X_2 are independent conditional on θ, and sampling of X is carried out in two blocks:

(a1) sample X_1 from $p(X_1|\theta, X_0, Y)$,
(a2) sample X_2 from $p(X_2|\theta)$.

As the conditional posterior of X_2 is equal to the prior $p(X_2|\theta)$ of X_2 given θ, MCMC sampling of X_2 is straightforward. We sample $X_2 = \{(a_j, r_j), j = N+1, N+2, \ldots\}$ as $a_{N+j} = \lambda\alpha T\Delta + \tilde{a}_j$, where $\tilde{a}_j$ are the arrival times of a Poisson process with intensity 1, and $r_{N+j} \sim U[0, 1]$, $j = 1, 2, \ldots$.

When sampling X_1 from $p(X_1|\theta, X_0, Y)$, we choose with probability 0.5 one of the following moves: a global death with immigration move where the number of jumps changes, and a local multiple displacement move, where the number of jumps remains unchanged and the likelihood changes only locally. Both moves are generalizations of the corresponding single update moves suggested by Roberts et al. (2004). To improve mixing, we conclude each draw by a Metropolis-Hastings move, which updates all jump times in a similar way as in Roberts et al. (2004).

3.1.1 Death with immigration move

This is a global move where the number N of jumps changes. At each move, each pair (a_j, r_j), $j = 1, \ldots, N$, is deleted with a certain probability p_d. The total number of deleted points is denoted by N_d. New points $(a_j^{\text{new}}, r_j^{\text{new}})$, $j = 1, 2, \ldots, N_b$, are added with a certain birth rate p_b, by proposing the number of new points N_b from a Poisson distribution with intensity $\lambda\alpha T \Delta p_b$. The new number of jumps is equal to $N^{\text{new}} = N + N_b - N_d$.

The new arrival times a_j^{new}, $j = 1, 2, \ldots, N_b$, are proposed as the order statistics of N_b random numbers being uniformly distributed on $[0, \lambda\alpha T \Delta]$. An equivalent way of proposing a_j^{new}, $j = 1, 2, \ldots, N_b$, jointly with N_b, is to sample a_j^{new} as the arrival times of a Poisson process with intensity p_b, which is then truncated at $\lambda\alpha T \Delta$. In any case, the new marks r_j^{new}, $j = 1, 2, \ldots, N_b$, are proposed as iid draws from a uniform distribution. Note that this implies following proposal density $q(N^{\text{new}}|N)$ for the new number of jumps N^{new}, given that the current number of jumps is equal to N:

$$q(N^{\text{new}}|N) = \mathrm{e}^{-\lambda\alpha T \Delta p_b} \sum_{i=0}^{\min(N, N^{\text{new}})} \frac{(\lambda\alpha T \Delta p_b)^{N^{\text{new}}-i}}{\Gamma(N^{\text{new}} - i + 1)} \binom{N}{i} (1 - p_d)^i p_d^{N-i},$$

whereas the new arrival times and the new marks are sampled from the prior. Let X_1^{new} denote the collection of all pairs (a_j, r_j) that were not deleted in the first step and all pairs $(a_j^{\text{new}}, r_j^{\text{new}})$ that were added in the second step. The acceptance rate for a move from X_1 to X_1^{new} is given by $\min(1, r_X)$ with

$$r_X = \frac{p(Y|\mathcal{V}(X_1^{\text{new}}, \theta), X_0, \theta) q(N|N^{\text{new}})}{p(Y|\mathcal{V}(X_1, \theta), X_0, \theta) q(N^{\text{new}}|N)} (\lambda\alpha T \Delta p_b)^{(N^{\text{new}}-N)} \frac{\Gamma(N^{\text{new}} + 1)}{\Gamma(N + 1)}. \quad (25)$$

We would like to mention that the algorithm of Geyer and Møller (1994) which is applied in Roberts et al. (2004) results as that special case of the methods described in this subsection, where with probability 0.5 one selects either $(N_d = 1, N_b = 0)$ or $(N_d = 0, N_b = 1)$.

Our algorithm is more flexible, as it allows deleting and adding of more than one point. There are two parameters for tuning this move, namely the death rate p_d and the birth rate p_b. Increasing any of these rates will decrease the acceptance rate of the move. We observed a certain sensitivity to choosing appropriate rates for high intensity processes. To achieve more robustness, we found it useful to work with random rates p_d and p_b, rather than with fixed ones. At each MCMC draw, we sample p_d and p_b independently from a uniform distribution over $[0, p_{\max}]$.

3.1.2 Local multiple displacement move

Next we consider a local move within a subset of the whole space that leaves the likelihood unchanged outside this subset. The move usually operates on more than one point simultaneously, and is a direct extension of the single local displacement move suggested in Roberts et al. (2004). Rather than moving directly the points of X_1, we

operate for this move on the jump times and jump sizes of the volatility process given by $\mathcal{V} = \{(\tau_j, J_j), j = 1, \ldots, N\}$. The latent process $\mathcal{V}$ is then updated by modifying all jump times and jump sizes within a block $[\tau_a, \tau_b]$ by the following local multiple displacement move.

We displace all jumps times within the block $[\tau_a, \tau_b]$, i.e. $\tau_a < \tau_j < \tau_b$, randomly: $\tau_j^{\text{new}} \sim \mathcal{U}[\tau_a, \tau_b]$ and define the new jump sizes by:

$$J_j^{\text{new}} = \mathrm{e}^{-\lambda(\tau_j^{\text{new}} - \tau_j)} J_j.$$

All jump times and jump sizes outside the block $[\tau_a, \tau_b]$ remain unchanged: $\tau_j^{\text{new}} = \tau_j$, $J_j^{\text{new}} = J_j$. This choice leaves the volatility $\sigma^2(t)$ unchanged outside $[\tau_a, \tau_b]$:

$$\sigma^2(t) = \sum_{j=1}^{N_t} J_j^{\text{new}} \mathrm{e}^{-\lambda(t-\tau_j^{\text{new}})} = \sum_{j=1}^{N_t} J_j \mathrm{e}^{-\lambda(t-\tau_j)},$$

for all $t < \tau_a$ and $t > \tau_b$. The acceptance rate of a move from $\mathcal{V}$ to $\mathcal{V}^{\text{new}} = \{(\tau_j^{\text{new}}, J_j^{\text{new}}), j = 1, \ldots, N\}$ is given by $\min(1, r_X)$ with

$$r_X = \frac{p(Y|\mathcal{V}^{\text{new}}, X_0, \theta)}{p(Y|\mathcal{V}, X_0, \theta)} \exp\left\{-\delta \sum (J_j^{\text{new}} - J_j) - \lambda \sum (\tau_j^{\text{new}} - \tau_j)\right\}. \quad (26)$$

If the new jump times and jump sizes are accepted, they are matched back to X_1 through the appropriate transformation, based on the Rosiński representation:

$$a_j = \exp(-\delta J_{(j)}^{\text{new}}) \lambda \alpha T \Delta, \quad (27)$$

$$r_j = \frac{\tau_{(j)}^{\text{new}}}{T\Delta}, \quad (28)$$

where $J_{(1)}^{\text{new}}, \ldots, J_{(N)}^{\text{new}}$ denote the order statistics of the jump sizes $J_1^{\text{new}}, \ldots, J_N^{\text{new}}$ and the same ordering is applied to $\tau_1^{\text{new}}, \ldots, \tau_N^{\text{new}}$ to obtain $\tau_{(1)}^{\text{new}}, \ldots, \tau_{(N)}^{\text{new}}$.

It remains to discuss how to choose the block $[\tau_a, \tau_b]$. In order to include multiple jumps, the whole path is divided at each scan of the MCMC sampler into blocks, with the block length as well as the number of blocks being random. To this aim we first select the average block length to be B, and define the boundaries between the blocks as the arrival times of a Poisson process with intensity $1/B$, running till ΔT. Obviously the first block starts at 0, whereas the last block ends at ΔT. Note that the actual number of blocks changes, with the expected number of blocks being equal to T/B. Then we select one block randomly and update all jump times and jump sizes within this block simultaneously by the multiple displacement move described above. This move is tuned by selecting the average block length B (measured in multiples of Δ), where an increase of the average block length B will reduce the acceptance rate.

3.1.3 Blocked Metropolis-Hastings-update of the jump sizes

To improve mixing, we implemented a blocked Metropolis-Hastings update of all jump sizes, conditional on the current number of jumps and current jump times, which is a slight modification of the update suggested by Roberts et al. (2004). For each jump size J_j, a new jump size J_j^{new} is proposed according to log-normal random walk Metropolis Hastings algorithm:

$$\log J_j^{\text{new}} = \log J_j + \xi_j, \quad \xi_j \sim \mathcal{N}\left(0, c_\xi^2\right).$$

The acceptance rate of a move from $\mathcal{V} = \{(\tau_j, J_j), j = 1, \ldots, N\}$ to $\mathcal{V}^{\text{new}} = \{(\tau_j, J_j^{\text{new}}), j = 1, \ldots, N\}$ is given by $\min(1, r_X)$ with

$$r_X = \frac{p(Y|\mathcal{V}^{\text{new}}, X_0, \theta)}{p(Y|\mathcal{V}, X_0, \theta)} \exp\left\{\sum_{j=1}^{N} \xi_j - \delta \sum_{j=1}^{N} (J_j^{\text{new}} - J_j)\right\}. \tag{29}$$

We select the variance c_ξ^2 of the proposal in such a way that the expected prior ratio is not too small, i.e. equal to a value not to far from 1, say $\gamma = 0.8$:

$$c_\xi^2 = 2\log(1 - \log(\gamma)/N).$$

3.2 Sampling the model parameter θ

MCMC estimation requires sampling θ conditional on X and X_0 from $p(\theta|X, X_0, Y)$. As in Roberts et al. (2004), the parameters (α, δ) are sampled jointly, conditional on holding λ fixed, whereas λ is sampled in a single-move manner conditional on knowing (α, δ).

We use a Metropolis-Hastings-step with a normal random walk proposal for $\log \alpha$, by proposing a new value α^{new} from $\log \alpha^{\text{new}} \sim \mathcal{N}(\log \alpha, c_a)$). Then δ^{new} is sampled from the conditional posterior density $p(\delta|X, X_0, Y, \alpha^{\text{new}}, \lambda)$ which is the density of the Gamma-distribution $\mathcal{G}(d_T, D_T(\alpha^{\text{new}}, \lambda))$, with $d_T = d_0 + T/2$ and

$$D_T(\alpha, \lambda) = D_0 + \frac{1}{2} \sum_{n=1}^{T} \frac{y_n^2}{\tilde{\sigma}_n^2(\alpha, \lambda)}, \tag{30}$$

where $\tilde{\sigma}_n^2(\alpha, \lambda)$ is defined in (24). $(\alpha^{\text{new}}, \delta^{\text{new}})$ is accepted with probability $\min\{1, r_\theta\}$ where $r_\theta = w(\alpha^{\text{new}})/w(\alpha)$ and

$$w(\alpha) = \frac{p(\alpha)\alpha}{\prod_{n=1}^{T} \sqrt{\tilde{\sigma}_n^2(\alpha, \lambda)} D_T(\alpha, \lambda)^{d_T}}. \tag{31}$$

We use a Metropolis-Hastings-step with a normal random walk proposal for updating λ, by proposing $\lambda^{\text{new}} \sim \mathcal{N}(\lambda, c_l)$. This corresponds to a log random walk proposal

for $\rho^{\text{new}} = \exp(-\Delta\lambda^{\text{new}})$. The new parameter $\theta^{\text{new}} = (\alpha, \delta, \rho^{\text{new}})$ is accepted with probability $\min\{1, r_\rho\}$ where

$$r_\rho = \frac{p(Y|X, X_0, \theta^{\text{new}})p(X|\theta^{\text{new}})p(\rho^{\text{new}})\rho^{\text{new}}}{p(Y|X, X_0, \theta)p(X|\theta)p(\rho)\rho}. \tag{32}$$

c_a and c_l influence the possible relative changes of α^{new} and ρ^{new}, compared to α and ρ. Choosing $c_a = 0.1$ and $c_l = 0.003$, as we will do in Sect. 5, implies that α^{new} lies with probability 0.95 within the range of ± 62 percent of α, whereas ρ^{new} lies with the same probability within the range of ± 10 percent of ρ.

3.3 Updating the initial volatility

Finally we update the initial volatility X_0, by proposing X_0^{new} from a proposal density $q(X_0^{\text{new}}|X_0)$. The acceptance probability of X_0^{new} is equal to $\min\{1, r_0\}$ where:

$$r_0 = \frac{p(Y|X_0^{\text{new}}, X, \theta)p(X_0^{\text{new}})q(X_0|X_0^{\text{new}})}{p(Y|X_0, X, \theta)p(X_0)q(X_0^{\text{new}}|X_0)}.$$

If the lag 1 autocorrelation $\rho = \mathrm{e}^{-\lambda\Delta}$ is smaller than 0.5, we propose X_0^{new} from the prior, $q(X_0^{\text{new}}|X_0) = p(X_0^{\text{new}})$, in which case r_0 simplifies to the ratio of the two likelihoods. Otherwise we propose X_0^{new} from a log random walk proposal, $\log X_0^{\text{new}} \sim \mathcal{N}(\log X_0, c_0)$, in which case the ratio of the proposal densities is equal to X_0^{new}/X_0.

3.4 A simulation study

In order to access the performance of the Rosiński parameterization with multiple updates in comparison to the non-centered parameterization introduced by Roberts et al. (2004), we repeated some of the simulation experiments studied in that paper. For all simulations $T = 500$, $\Delta = 1$, and $\delta = 10$, whereas four different scenarios are obtained by combining $\lambda = 0.03$ and $\lambda = 0.1$ with $\alpha = 2/3$ and $\alpha = 2$.

We use the same priors as Roberts et al. (2004), namely a $\mathcal{G}(1, 0.01)$-prior for α and δ, and a $\mathcal{G}(1, 1)$-prior for λ, which is equal to the $\mathcal{B}(1, 1)$-prior on $\rho = \mathrm{e}^{-\lambda}$. Finally, we choose the same starting values, namely $(\alpha, \delta, \lambda) = (10, 10, 1)$. We ran both algorithms for 1 million iterations, removed the first 50,000 draws as burn-in, and stored every 100th iteration.

To monitor converge, we found it useful not only to consider the posterior draws of α, δ, λ, and transformations of these parameters such as the mean α/δ, the variance α/δ^2 and the lag 1 autocorrelation ρ of the latent volatility process, but also further statistics involving the latent processes $z(\lambda t)$ and $\sigma^2(t)$. Two particularly useful quantities are the number $N = \#\{\tau_j|\tau_j \le T\Delta\}$ of jumps and $z(\lambda T\Delta) = \sum_{j=1}^{N} J_j$, the total jump mass of the BDLP, both of which are directly available at each MCMC draw.

Figures 1 and 2 show for all four simulated series the estimated autocorrelation function of N and $z(\lambda T\Delta)$, obtained after removing the burn-in sample and thinning as

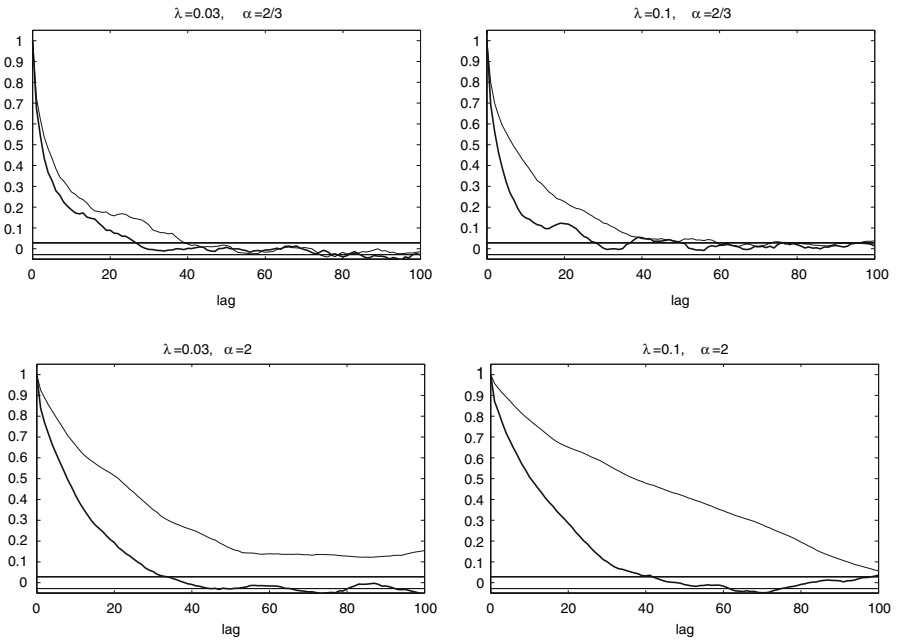

Fig. 1 Simulation study: autocorrelation function of the number N of jumps under data augmentation based on the Rosiński representation (*thick line*) and the non-centered parameterization of Roberts et al. (2004) (*thin line*), obtained from 1 million MCMC draws, after a burn-in of 50,000 draws, thinning one every hundred; true value of λ equal to 0.03 (*left hand side*) and 0.1 (*right hand side*), true value of α equal to 2/3 (*top*) and 2 (*bottom*)

explained above. We find that the mixing behavior of the latent processes, measured by these two statistics, is better under data augmentation based on the Rosiński representation, which is combined with the multiple updating algorithm for the latent process X introduced in Sect. 3.1, than under the non-centered parameterization Roberts et al. (2004). The better mixing properties under the Rosiński representation result from the prior correlation, introduced through the latent process a_j, between increasing or decreasing N and adding or deleting small jumps, only. This allows larger steps through the parameter space in comparison to the non-centered parameterization of Roberts et al. (2004), where new jumps are sampled from the exponential prior.

4 Superposition models

The model defined above implies that the log of the autocorrelation function of the squared returns decays exponentially which is not in accordance with the empirical autocorrelation of returns from financial time series. To obtain a model with a more flexible autocorrelation structure, Barndorff-Nielsen and Shephard (2001) suggested to combine various independent OU-processes to form a so-called superposition model. The log price $x^\star(t)$ is modelled as solution of the SDE (2), where $\sigma^2(t)$ is the sum of independent OU process with marginal Gamma law:

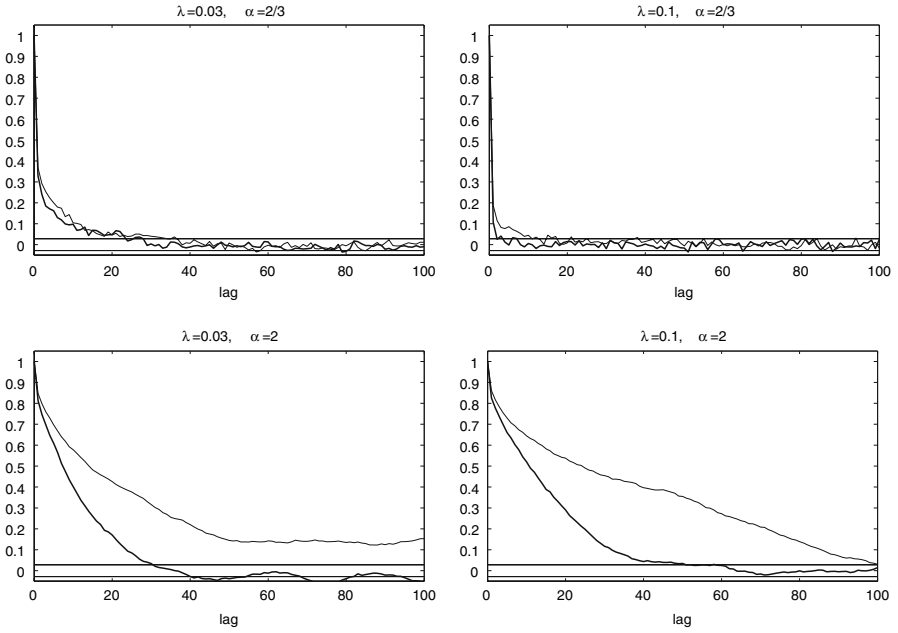

Fig. 2 Simulation study: autocorrelation function of the total jump mass $z(\lambda T)$ under data augmentation based on the Rosiński representation (*thick line*) and the non-centered parameterization of Roberts et al. (2004) (*thin line*), obtained from 1 million MCMC draws, after a burn-in of 50,000 draws, thinning one every hundred; true value of λ equal to 0.03 (*left hand side*) and 0.1 (*right hand side*), true value of α equal to 0.03 (*top*) and 0.1 (*bottom*)

$$\sigma^2(t) = \sum_{i=1}^{k} \sigma_i^2(t)$$

$$\mathrm{d}\sigma_i^2(t) = -\lambda_i \sigma_i^2(t)\mathrm{d}t + \mathrm{d}z_i(\lambda_i t). \tag{33}$$

$z_i(\lambda t)$ is a Lévy process with independent, strictly positive increments consisting entirely of jumps and the marginal law of $\sigma_i^2(t)$ is a $\mathcal{G}(\alpha_i, \delta_i)$-distribution.

If all $\delta_i \equiv \delta$, then the marginal law of $\sigma^2(t)$ is a $\mathcal{G}\left(\sum_i \alpha_i, \delta\right)$-distribution, otherwise it is a finite mixture of Gamma distributions. If all $\alpha_i \equiv \alpha$, whereas the δ_is are different, then it is possible to write the model as a weighted sum of independent OU process with identical marginal Gamma law:

$$\sigma^2(t) = \sum_{i=1}^{k} w_i \sigma_i^2(t),$$

where $\sigma_i^2(t) \sim \mathcal{G}(\alpha, \delta)$, $1/\delta = \sum_i 1/\delta_i$, and $w_i = \delta/\delta_i$. Obviously, $\sum_i w_i = 1$.

The superposition model is identified only up to relabelling the indices of the components, as the observed returns follow a normal distribution as in (6) with

$$\sigma_n^2 = \sum_{i=1}^{k} \frac{1}{\lambda_i} \left[z_i(\lambda_i t_n) - z_i(\lambda_i t_{n-1}) - (\sigma_i^2(t_n) - \sigma_i^2(t_{n-1})) \right] \tag{34}$$

being invariant to relabelling the components of the superposition model. For $k = 2$, for instance, the likelihood function $p(Y|\theta)$ is the same for any pair $\theta = (\alpha_1, \delta_1, \rho_1, \alpha_2, \delta_2, \rho_2)$ and $\theta^\star = (\alpha_2, \delta_2, \rho_2, \alpha_1, \delta_1, \rho_1)$, where $\rho_i = \mathrm{e}^{-\Delta\lambda_i}$, and consequently exhibit two equivalent modal regions, with each mode corresponding to one way of labelling the components. In general, when fitting a superposition model of k processes to data generated from such a model, the likelihood function has $k!$ modes.

The consequences this kind of unidentifiability has on MCMC estimation within a Bayesian approach is well-known for mixture models, see for instance Stephens (2000) and Frühwirth-Schnatter (2001b), and similar issues are relevant for the superposition of OU processes. The posterior distribution $p(\alpha_1, \delta_1, \rho_1, \ldots, \alpha_k, \delta_k, \rho_k|Y)$ inherits the multimodality of the likelihood function, if the latter is combined with a prior that is invariant to relabelling the components, for instance with the independence prior

$$p(\alpha_1, \delta_1, \rho_1, \ldots, \alpha_k, \delta_k, \rho_k) = \prod_{i=1}^{k} p(\alpha_i) p(\delta_i) p(\rho_i), \tag{35}$$

where for each OU process the priors of α_i, δ_i and ρ_i are chosen as in Sect. 3. When sampling from this posterior, the MCMC sampler will stick at one of the modes, if the modes of the posterior are well-separated, in which case we may use the MCMC draws for parameter estimation by taking ergodic averages. If the modes are not well-separated, label switching between the modes is likely to take place, which renders parameter estimation based on ergodic averages useless. Whether the modes are well-separated or not, depends on the data as well as on the differences between the true parameters in relation to the information contained in the data, and is not under the control of the investigator.

Asymmetric prior distributions can often help to make the posterior distribution unimodal, if they are introduced for a parameter that is different between the different processes. In the financial applications of the superposition model reported in Barndorff-Nielsen and Shephard (2001) and Roberts et al. (2004), the processes differ mainly in λ_i. Thus we replace the independence prior (35) by an asymmetric prior which takes the form:

$$p(\alpha_1, \delta_1, \rho_1, \ldots, \alpha_k, \delta_k, \rho_k) = p(\rho_1, \ldots, \rho_k) \prod_{i=1}^{k} p(\alpha_i) p(\delta_i), \tag{36}$$

where

$$p(\rho_1, \ldots, \rho_k) = p(\rho_1) \prod_{i=2}^{k} p(\rho_i|\rho_{i-1}). \tag{37}$$

We choose a marginal $\mathcal{B}(b_1, B_1)$-prior for ρ_1, whereas

$$\rho_i|\rho_{i-1} \sim \rho_{i-1}\mathcal{B}(b_i, B_i)\,. \tag{38}$$

This prior generalizes the Markov prior, suggested by Roeder and Wasserman (1997) for univariate mixtures of normal distributions, to superposition models of OU processes. The Markov prior (38) induces an ordering on the decay factors given by

$$\mathrm{e}^{-\lambda_1\Delta} > \cdots > \mathrm{e}^{-\lambda_k\Delta},$$

and leads to a posterior density that is no longer invariant to relabelling the components. For $b_i = B_i = 1, i = 1, \ldots, k$, for instance, a uniform prior over $\lambda_1 < \cdots < \lambda_k$ results.

For MCMC estimation of the superposition model we introduce a latent process X_i and an initial standardized volatility $X_{0,i}$ for each OU-process. Each component X_i and each initial standardized volatility $X_{0,i}$ is sampled separately, conditional on knowing the remaining components. Various blocking strategies could be used for updating the parameters. For $k = 2$, for instance, when $\delta_i \equiv \delta$, we sample α_1, α_2 and δ jointly, by proposing α_1^{new} and α_2^{new} independently from a log random walk proposal, whereas δ is sampled from the conditional posterior $p(\delta|X_1, X_2, X_{0,1}, X_{0,2}, \alpha_1^{\text{new}}, \alpha_2^{\text{new}}, \lambda_1, \lambda_2)$, which remains a Gamma distribution. λ_1 and λ_2 are sampled in a single move manner, using a Metropolis-Hastings update with a random walk proposal.

5 Application to real data

We fitted the BNS stochastic volatility model to a series of daily exchange rates of the Swiss franc to the US dollar, provided by Robert Tompkins. The data cover the period from January 2, 1985 to December 13, 1999, thus $T = 3778$ and $\Delta = 1$. The daily aggregate returns are derived after scaling the log-prices by a factor 100, hence returns are measured in percentage terms.

First we consider a single OU-process with marginal Gamma law. Computations are based on following priors: $\alpha \sim \mathcal{G}(1, 1)$, $\delta \sim \mathcal{G}(1, 0.01)$, $\rho = \mathrm{e}^{-\lambda\Delta} \sim \mathcal{B}(1, 1)$, and $X_0 \sim \mathcal{G}(1, 1)$. We started MCMC estimation from two different starting values, namely $(\alpha, \delta, \lambda) = (1, 1, 0.3)$ and $(\alpha, \delta, \lambda) = (1, 2, 0.15)$. Starting values for the jump times and the jump sizes were simulated from the prior $p(X|\alpha, \delta, \lambda)$. We tuned the moves on the latent process X by choosing $p_{\max} = 0.03$, and $B = 40$, whereas the variances in the proposals for α, λ and X_0 are equal to $c_a = 0.1$, $c_l = 0.003$ and $c_0 = 2$.

We experienced some difficulties with the algorithm of Roberts et al. (2004) when fitting a single OU process to this time series, which failed to converge within 1 million iterations from the second starting value. For both starting values, the non-centered parameterization based on the Rosiński representation converged quickly to the same modal region of the posterior, see Fig. 3 for the MCMC draws obtained from the second starting value.

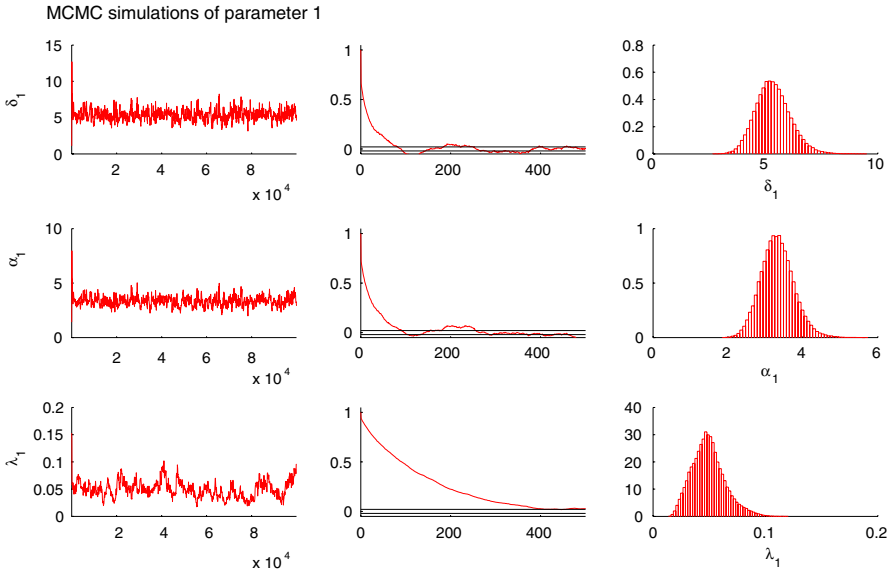

Fig. 3 10,000 posterior draws of δ, α, and λ obtained from fitting a single OU model to the Swiss franc/US dollar series (original chain of 1 million thinned one every hundred); autocorrelation and histograms obtained after removing the initial 500 draws

Table 1 Posterior parameter summaries obtained from fitting a single OU model to the Swiss franc/US dollar series

Parameter	Posterior mean	Posterior median	Standard deviation
δ	5.34	5.3	0.768
α	3.32	3.3	0.439
λ	0.0493	0.0483	0.0146
α/δ	0.624	0.623	0.0396
α/δ^2	0.12	0.118	0.0219
$\lambda\alpha$	0.164	0.16	0.053
$e^{-\lambda}$	0.952	0.953	0.0139

Concerning the mixing behavior, the MCMC chain mixes acceptably for the parameters α and δ, governing the marginal law, see Fig. 3, as well as for the corresponding mean α/δ and the corresponding variance α/δ^2 (figures not reported). We observe in general a poorer mixing behavior for λ, see again Fig. 3, and consequently for the intensity $\alpha\lambda$ and the lag 1 autocorrelation $e^{-\lambda}$ (figures not reported). Posterior summaries for various parameters are reported in Table 1.

We may combine the Rosiński representation with single updating as in Geyer and Møller (1994) and Roberts et al. (2004) rather than multiple updates, however, our experience is that this leads to slower convergence, if the sampler is started far out in the tails of the posterior distribution.

Next we considered the superposition of two OU-process with marginal Gamma laws with $\delta_1 = \delta_2 = \delta$. Estimation is based on the priors $\delta \sim \mathcal{G}(1, 0.01)$, $\alpha_i \sim \mathcal{G}(1, 1)$, $\rho_i \sim \mathcal{B}(1, 1)$, and $X_{0,i} \sim \mathcal{G}(1, 1)$, for $i = 1, 2$. To check convergence, we

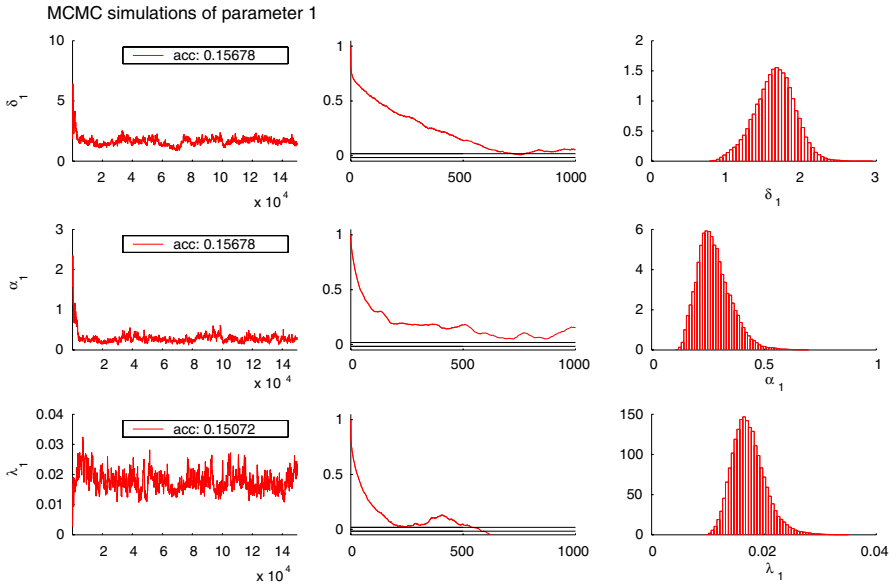

Fig. 4 15,000 posterior draws of δ, α_1 and λ_1 obtained from fitting a superposition OU model to the Swiss franc/US dollar series (original chain of 1.5 million thinned one every hundred); autocorrelation and histograms obtained after removing the initial 2,000 draws

considered again two different starting values, namely $(\delta, \alpha_1, \alpha_2, \lambda_1, \lambda_2) = (1, 1, 0.1, 0.01, 10)$ and $(\delta, \alpha_1, \alpha_2, \lambda_1, \lambda_2) = (1, 1, 1, 0.1, 5)$. Again, the non-centered parameterization based on the Rosiński representation converged rather quickly to the same modal region of the posterior from both starting values, however mixing is much slower for the superposition model than for the single OU-model, see Fig. 4.

Posterior summaries for the parameters are given in Table 2. We find that the two processes differ mainly in the memory parameter λ_1 and λ_2, whereas the difference in the volatility of volatility parameters α_1 and α_2 is not very pronounced. The second process is a high activity process with 2.3 expected jumps per day, see also Fig. 5 showing the posterior mean of each volatility process, $E\sigma_i^2(t)|Y$, $i = 1, 2$, as well as the observed squared returns y_n^2.

6 Concluding remarks

In the present paper practical Bayesian estimation of the BNS stochastic volatility model with marginal Gamma laws has been studied. A parameterization of an OU process with marginal Gamma law based on the Rosiński representation (Rosiński 2001) has been considered, which has the advantage of being a non-centered parameterization in the sense of Roberts et al. (2004). As in Roberts et al. (2004), the parameterization is based on a marked Poisson process, however the process lives on the positive real line, rather than on the two-dimensional space $[0, T] \times (0, \infty)$. The arrival times a_j of the Poisson process are used twice to define both the number of jumps and the jump sizes. This parameterization has the advantage of automatic

Table 2 Posterior parameter summaries obtained from fitting a superposition OU model to the Swiss franc/US dollar series

Parameter	Posterior mean	Posterior median	Standard deviation
δ	1.65	1.66	0.267
α_1	0.279	0.269	0.0751
λ_1	0.0173	0.017	0.003
α_2	0.642	0.643	0.1
λ_2	3.66	3.53	0.835
α_1/δ	0.171	0.167	0.0422
α_1/δ^2	0.108	0.102	0.0377
$\lambda_1\alpha_1$	0.00482	0.00462	0.00149
$e^{-\lambda_1}$	0.983	0.983	0.00294
α_2/δ	0.391	0.392	0.0253
α_2/δ^2	0.246	0.237	0.0514
$\lambda_2\alpha_2$	2.29	2.26	0.364
$e^{-\lambda_2}$	0.0343	0.0292	0.0235

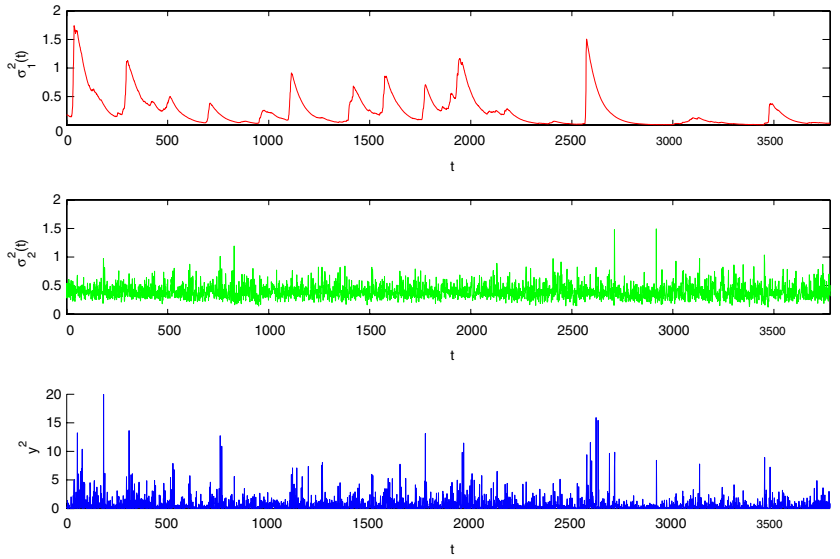

Fig. 5 Posterior mean of the each volatility processes, $E\sigma_1^2(t)|Y$ (*top*), $E\sigma_2^2(t)|Y$ (*middle*) in comparison to the observed squared returns y_n^2 (*bottom*) obtained from fitting a superposition OU model to the Swiss franc/US dollar series

thinning (deleting small jumps) when decreasing N, and of adding small jumps, when increasing N. In the later case, a proposal with new jumps is more likely to be accepted than the proposal used in Roberts et al. (2004), where new jumps are sampled from the exponential prior. Simulation experiments as well the analysis of real exchange rate confirmed this expectation.

Finally, we developed an MCMC scheme which enables multiple updates of the latent point process X. At each MCMC draw we add and delete more than one point from our latent point process, which is particularly useful for high intensity processes.

Acknowledgments This work was supported by the Austrian Science Foundation (FWF) under grant SFB 010 ('Adaptive Information Systems and Modelling in Economics and Management Science'). We would like to thank Omiros Papaspiliopoulos, Gareth Roberts, Friedrich Hubalek, Robert Tompkins, and, in particular, Neil Shephard for many extremely helpful comments and suggestions concerning this work. Finally, we thank Christoph Pamminger for his support with running the MCMC simulations.

References

Andersen, T. G., Lund, J. (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, *77*, 343–377.

Barndorff-Nielsen, O. E., Shephard, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *Journal of Royal Statistical Society, Series B*, *63*, 167–241.

Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *8*, 1637–1659.

Cox, D. R., Isham, V. (1988). *Point processes*. London: Chapman & Hall.

Frühwirth-Schnatter, S. (2001a). Discussion of the paper by O. Barndorff-Nielsen and N. Shephard on "Non-Gaussian OU based models and some of their uses in financial economics". *Journal of Royal Statistical Society, Series B*, *63*, 220–222.

Frühwirth-Schnatter, S. (2001b). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, *96*, 194–209.

Geyer, C., Møller, J. (1994). Simulation procedures and likelihood inference for spatial point processes. *Scandinavian Journal of Statistics*, *21*, 359–373.

Griffin, J. E., Steel, M. F. J. (2006). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. *Journal of Econometrics*, *134*, 605–644.

Hull, J., White, A. (1987). The pricing of options on assets with stochastic volatilities. *The Journal of Finance*, *42*, 281–300.

Papaspiliopoulos, O. (2001). Discussion of the paper by O. Barndorff-Nielsen and N. Shephard on "Non-Gaussian OU based models and some of their uses in financial economics". *Journal of Royal Statistical Society, Series B*, *63*, 211–213.

Papaspiliopoulos, O., Roberts, G., Skold, M. (2003). Non-centered parameterizations for hierarchical models and data augmentation. In J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. Dawid, D. Heckerman, A. F. M. Smith, M. West (Eds.), *Bayesian statistics 7*, (pp. 307–326).

Roberts, G. O. (2001). Discussion of the paper by O. Barndorff-Nielsen and N. Shephard on "Non-Gaussian OU based models and some of their uses in financial economics". *Journal of Royal Statistical Society, Series B*, *63*, 209–211.

Roberts, G. O., Papaspiliopoulos, O., Dellaportas, P. (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *Journal of Royal Statistical Society, Series B*, *66*, 369–393.

Roeder, K., Wasserman, L. L. (1997). Practical Bayesian density estimation using mixtures of normals. *Journal of the American Statistical Association*, *92*, 894–902.

Rosiński, J. (2001). Series representations of Lévy processes from the perspective of point processes. In O. E. Barndorff-Nielsen, T. Mikosch, S. I. Resnick (Eds.), *Lévy processes: theory and applications*. Birkhäuser.

Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. In D. R. Cox, D. V. Hinkley, O. E. Barndorff-Nielsen (Eds.), *Time series models: in econometrics, finance and other fields*, (pp. 1–67). London: Chapman & Hall.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of Royal Statistical Society, Series B*, *62*, 795–809.