J. R. Statist. Soc. B (2004)
66, *Part* 2, *pp.* 369–393

Bayesian inference for non-Gaussian Ornstein–Uhlenbeck stochastic volatility processes

Gareth O. Roberts and Omiros Papaspiliopoulos

Lancaster University, UK

and Petros Dellaportas

Athens University of Economics and Business, Greece

[Received August 2001. Final revision November 2003]

Summary. We develop Markov chain Monte Carlo methodology for Bayesian inference for non-Gaussian Ornstein–Uhlenbeck stochastic volatility processes. The approach introduced involves expressing the unobserved stochastic volatility process in terms of a suitable marked Poisson process. We introduce two specific classes of Metropolis–Hastings algorithms which correspond to different ways of jointly parameterizing the marked point process and the model parameters. The performance of the methods is investigated for different types of simulated data. The approach is extended to consider the case where the volatility process is expressed as a superposition of Ornstein–Uhlenbeck processes. We apply our methodology to the US dollar–Deutschmark exchange rate.

Keywords: Data augmentation; Lévy processes; Marked point processes; Markov chain Monte Carlo methods; Non-centred parameterizations; Stochastic volatility

1. Introduction

Recently, Barndorff-Nielsen and Shephard (2001) introduced a general class of continuous time stochastic volatility (SV) models based on non-Gaussian Ornstein–Uhlenbeck (OU) processes. They assumed the logarithm of an asset price to be the solution of a linear diffusion process, as is common in many continuous time asset pricing models, but they modelled the volatility process as a stationary, latent OU process, driven by a non-Gaussian Lévy process. This specification forces the volatility to move up entirely by jumps and then to tail off exponentially. The jump times and corresponding sizes are controlled by the increment distribution of the driving Lévy process. In Barndorff-Nielsen and Shephard (2001), more general models were considered that allow the volatility to be the addition of independent OU processes and the log-price process to incorporate the so-called leverage effect (Black, 1976). These models are attractive since they combine a large amount of analytic tractability, e.g. for pricing options and deriving aggregation results, with sufficient flexibility to allow consistency with stylized facts observed in financial markets.

The aim of this paper is to provide a general methodology for performing Bayesian inference for a wide class of models that were presented in Barndorff-Nielsen and Shephard (2001). Our approach was briefly introduced in our discussion on Barndorff-Nielsen and Shephard (2001) (Roberts, 2001; Papaspiliopoulos, 2001). The methodology is directly applicable whenever the

Address for correspondence: Gareth O. Roberts, Department of Mathematics and Statistics, Lancaster University, Lancaster, LA1 4YF, UK.
E-mail: g.o.roberts@lancaster.ac.uk

 1369–7412/04/66369

OU process is driven by a compound Poisson process. The jump times of such a process form a Poisson process with finite rate and the corresponding jump sizes are independent and identically distributed positive random variables. To illustrate our methodology, we assume that the jump sizes are generated from an exponential distribution. The resulting OU process has a gamma marginal distribution and is known as the shot noise process.

We propose a latent structure model formulation based on marked point processes, and we develop suitable Markov chain Monte Carlo (MCMC) algorithms for Bayesian inference, using data augmentation. We construct two different MCMC algorithms and show by an extensive simulation study that the performance of the first, based on a standard hierarchical parameterization for the model, is not robust to different parameter values and data sampling frequencies. However, the second algorithm is considerably more robust and produces satisfactorily mixing chains in a wide range of cases. We extend our model by letting the volatility be the sum of two independent shot noise processes and use this model to fit exchange rates data.

The methodology that we introduce is related to the notion of non-centring, as described in Papaspiliopoulos (2003) and Papaspiliopoulos *et al.* (2003), and can be applied in contexts outside this paper. In particular, it can be used for hierarchical models which involve latent Lévy processes (see Wolpert and Ickstadt (1998) for an example in spatial statistics). Moreover, we propose some model checking methods which could be useful to assess model fit in arbitrary hierarchical models.

This paper is organized as follows. In Sections 2.1 and 2.2 we describe the non-Gaussian OU SV model and discuss some of its characteristics. Section 3.1 formulates likelihood-based inference as a missing data problem by using marked Poisson processes. In Sections 3.2 and 3.4 we develop two hierarchical parameterizations of the model. The implementation of the MCMC algorithms under each parameterization is briefly described, with details deferred to appendices, and Section 4 compares their convergence properties for different types of simulated data. Section 3.3 places the proposed parameterizations inside the framework of the non-centred parameterizations of Papaspiliopoulos *et al.* (2003). Section 5 addresses modelling issues, such as the sensitivity of the posterior distributions of the parameters under different prior specifications, and develops some exploratory model checking methods. In Section 6 we fit the models to a series of US dollar–Deutschmark exchange rates.

2. Data and model

2.1. Barndorff-Nielsen and Shephard model

The simplest model suggested by Barndorff-Nielsen and Shephard (2001) assumes that the logarithm of an asset price, $x(t)$, follows a driftless diffusion process

$$\mathrm{d}x(t) = v(t)^{1/2}\,\mathrm{d}B(t), \qquad t \in [0, T], \tag{1}$$

where $x(0) = 0$ almost surely, $B(\cdot)$ is standard Brownian motion and $v(t)$ is the instantaneous variance of $x(t)$. The volatility $v(t)$ is modelled as a stationary non-Gaussian OU process with decay rate $\mu > 0$ which is driven by a homogeneous Lévy process $z(\cdot)$ with positive increments and $z(0) = 0$ almost surely:

$$\mathrm{d}v(t) = -\mu\, v(t)\,\mathrm{d}t + \mathrm{d}z(t). \tag{2}$$

Lévy processes are stochastic processes with independent and stationary increments (Sato, 1999).

Asset price returns exhibit conventional stylized behaviour, such as having excess kurtosis and skewness, tails much heavier than exponential, serial dependence (but not autocorrelation) and volatility clustering. Since the formulation given by equations (1) and (2) models the aggregate returns as scaled mixtures of normals, it captures the excess kurtosis and heavy tails that have been empirically observed. Specific choices for the stationary distribution of $v(t)$ in equation (2) (e.g. inside the generalized inverse Gaussian family) lead to empirically attractive marginal distributions for the returns; see Section 2 of Barndorff-Nielsen and Shephard (2001) for details. Moreover, returns that are generated by this model converge to normality under temporal aggregation, which again is an empirically desirable feature. The serial dependence and the volatility clustering are achieved through the persistence parameter μ in equation (2).

The above continuous time model formulation competes with the quite popular discrete time SV modelling framework, in which the log-volatility follows a latent stochastic process. The usual assumption is that this process is Gaussian, but there are suggestions to use a heavier-tailed distribution such as a Student t (Harvey *et al.*, 1994). Good reviews of SV models have been given by Kim *et al.* (1998), Shephard (1996) and Ghysels *et al.* (1996).

An important special case of Lévy processes is the compound Poisson process

$$z(t) = \sum_{j=1}^{\infty} \varepsilon_j \mathbf{1}(c_j < t), \tag{3}$$

where $\mathbf{1}(\cdot)$ denotes the indicator function. In equation (3), $0 < c_1 < c_2 < \ldots$ are the arrivals of a Poisson process with finite rate, λ say, and the ε_js are identically distributed positive random variables, which are independent of each other and of the Poisson process. When $\lambda < \infty$, the sum in equation (3) involves only a finite number of terms for any $t < \infty$. The methodology that is presented in this paper is directly applicable whenever the background driving Lévy process of equation (2) is a compound Poisson process.

Theorem 1 of Barndorff-Nielsen and Shephard (2001) shows that the marginal distribution of a stationary OU process satisfying equation (2) is self-decomposable (see page 90 of Sato (1999) for a definition). In this paper, we shall focus on the case where this is a gamma distribution $\mathrm{Ga}(\nu, \theta)$, with mean ν/θ and variance ν/θ^2. Then, it follows that the driving Lévy process is a compound Poisson process with exponentially distributed jump sizes, $\varepsilon_j \sim \mathrm{Ex}(\theta)$, and the volatility process is a shot noise Markov process; see for example page 88 of Cox and Isham (1980). Specifically, we have the representation

$$v(t) = \exp(-\mu t)\, v(0) + \sum_{j=1}^{\infty} \exp\{-\mu(t - c_j)\} \varepsilon_j \, \mathbf{1}(c_j < t) \tag{4}$$

where $0 < c_1 < c_2 < \ldots$ denote the arrival times of the homogeneous Poisson process with finite rate $\lambda = \nu\mu$. We also assume that $v(0) \sim \mathrm{Ga}(\nu, \theta)$. The parameters of the OU process are the Poisson rate λ, the parameter of the jump sizes θ and the decay rate μ. Simulated paths for the background driving Lévy process $z(\cdot)$ and the corresponding volatility and log-price processes are presented in Fig. 1. Note that the volatility jumps up sharply and then tails off exponentially. This can be thought of as new information arriving in packets.

A big advantage of this class of SV models is the analytic tractability of quantities of particular interest in theory and applications, such as the integrated volatility process

$$v^*(u, t) = \int_u^t v(s)\, \mathrm{d}s, \qquad 0 \leqslant u \leqslant t.$$

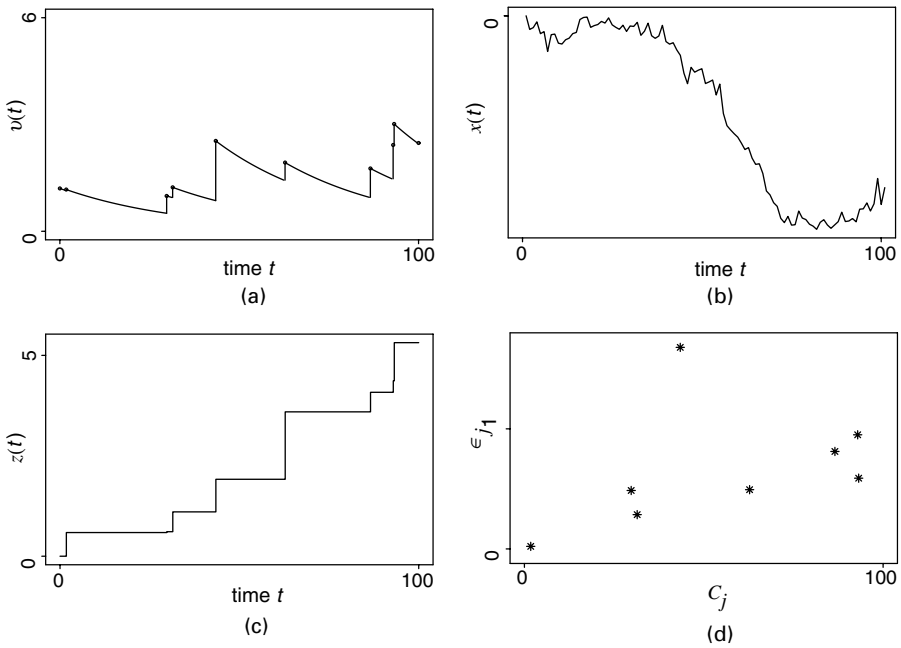

Fig. 1. Simulations from (a) the volatility process $v(\cdot)$, (b) the log-price process $x(\cdot)$, (c) the underlying Lévy process and (d) the marked Poisson process Ψ: the values that were used in the simulation are $\lambda = 0.1$, $\mu = 0.03$, $\theta = 1$ and $T = 100$

In our model equations (2)–(4) imply that

$$v^*(0,t) = \frac{1}{\mu}\sum_{j=1}^{\infty}\varepsilon_j \mathbf{1}(c_j < t) - \frac{1}{\mu}\{v(t) - v(0)\}$$
$$= \frac{1}{\mu}\left(\sum_{j=1}^{\infty}[1-\exp\{-\mu(t-c_j)\}]\varepsilon_j \mathbf{1}(c_j < t) - v(0)\{\exp(-\mu t) - 1\}\right). \tag{5}$$

Let $X = \{x(t_0), x(t_1), \ldots, x(t_n)\}$ be a collection of observations from process (1) at time points $0 = t_0 < t_1 < \ldots < t_n = T$, where $T > 0$. Typically, the time points are equally spaced and, in the examples of this paper, at daily frequency. Let

$$\begin{aligned} v_i^* &= v^*(0,t_i) - v^*(0,t_{i-1}), \\ y_i &= x(t_i) - x(t_{i-1}), \qquad i = 1,\ldots,n; \end{aligned} \tag{6}$$

the terms log-return and actual volatility are typically used for y_i and v_i^* respectively (Barndorff-Nielsen and Shephard, 2002). Then, according to equation (1),

$$\begin{aligned} y_i | v_i^* &\sim N(0, v_i^*), \qquad i = 1,\ldots,n, \\ \operatorname{corr}(y_i, y_k) &= 0, \qquad \text{for all } i \neq k, \end{aligned} \tag{7}$$

and therefore the conditional density of the data X given the integrated volatility process is

$$f(X|v^*) = \prod_{i=1}^{n} \frac{1}{\sqrt{(2\pi v_i^*)}} \exp\left(-\frac{y_i^2}{2v_i^*}\right). \tag{8}$$

The autocorrelation function of $v(\cdot)$ is

$$r(t) = \operatorname{corr}\{v(0), v(t)\} = \exp(-\mu t), \qquad t > 0,$$

which implies an exponentially decaying autocorrelation for the squared log-returns series $\{y_i^2, i = 1, \ldots, n\}$ (Barndorff-Nielsen and Shephard, 2001). Empirical studies using daily and high frequency data have shown that the autocorrelation in the $\{y_i^2\}$ series decays rapidly in the first lags, and then dies out slowly (see also Andersen *et al.* (2001) and Ding and Granger (1996)). Therefore, the model that is presented in this section cannot capture this dependence structure. A way to bypass this model deficiency is described in the next section.

2.2. A model based on superposition of Ornstein–Uhlenbeck processes

More complicated dependence structures can be imposed by modelling the volatility process as the superposition of a number of independent OU processes,

$$v(t) = \sum_{i=1}^{m} v_i(t) \tag{9}$$

with

$$\mathrm{d}v_i(t) = -\mu_i \, v_i(t) \, \mathrm{d}t + \mathrm{d}z_i(t) \tag{10}$$

where $z_i(\cdot)$ are independent Lévy processes. The autocorrelation function of the volatility is now

$$r(t) = \sum_{j=1}^{m} w_j \exp(-\mu_j t) \qquad \text{where } w_j = \mathrm{var}\{v_j(t)\} \Big/ \sum_{i=1}^{m} \mathrm{var}\{v_i(t)\}; \tag{11}$$

see Section 3 of Barndorff-Nielsen and Shephard (2001) for details.

The empirical findings of Barndorff-Nielsen and Shephard (2002) suggest that $m = 2$ is adequate for modelling daily financial data, so we shall only consider this case in this paper. We shall assume that $v_i(\cdot)$ has a $\mathrm{Ga}(\nu_i, \theta)$ stationary distribution; thus the marginal distribution of the SV process $v(t)$ is the same as in the single-OU model. This specification has also computational advantages (Section 3.5). Therefore

$$v_i(t) = \exp(-\mu_i t) \, v_i(0) + \sum_{j=1}^{\infty} \exp\{-\mu_i (t - c_{ij})\} \varepsilon_{ij} \, \mathbf{1}(c_{ij} < t), \qquad i = 1, 2, \tag{12}$$

where $0 < c_{i1} < c_{i2} < \ldots$ are the arrival times of a Poisson process in $[0, t]$ with rate $\lambda_i = \nu_i \mu_i$, $\varepsilon_{ij} \sim \mathrm{Ex}(\theta)$ and $v_i(0) \sim \mathrm{Ga}(\nu_i, \theta)$. The integrated volatility process satisfies

$$v^*(0, t) = v_1^*(0, t) + v_2^*(0, t) \tag{13}$$

where $v_1^*(0, t)$ and $v_2^*(0, t)$ can be derived from equations (12) and (5) in an obvious way.

This modelling approach allows us to capture both short-term variation, represented by an OU process with high decay rate μ_1, and long-term movements in the volatility, modelled as an OU process with much smaller decay rate μ_2. Therefore we shall typically take $\mu_2 < \mu_1$.

3. Data augmentation and parameterization

This section develops a missing data methodology to tackle the inference problem for the models of the previous section. To ease the exposition, the single-OU model is initially considered, and Section 3.5 generalizes the ideas for superposition of OU processes.

3.1. Augmentation based on marked Poisson processes

The volatility process is unobserved. Instead, we observe the logarithmic price process $X = \{x(t_0), \ldots, x(t_n)\}$ at times $0 = t_0 < t_1 < \ldots < t_n = T$. From equation (5) it follows that the

integrated volatility process is entirely known (and can be computed without any discretization error) given μ, $v(0)$, the Poisson process arrival times c_j and the corresponding jump sizes ε_j.

We introduce the stochastic process Ψ on $S = [0, T] \times (0, \infty)$ which contains the points $\{(c_j, \varepsilon_j)\}$ (Fig. 1(d)). The compound Poisson process $z(\cdot)$ is explicitly represented in terms of Ψ in equation (3).

Ψ can be seen either as a point process on S or equivalently as a marked point process with points c_j on $[0, T]$ and marks ε_j on $(0, \infty)$. The colouring theorem (section 5.1 of Kingman (1993)) shows that Ψ is a Poisson process on S with mean measure

$$\Lambda(\mathrm{d}c \times \mathrm{d}\varepsilon) = \lambda\theta \exp(-\theta\varepsilon)\, \mathrm{d}c\, \mathrm{d}\varepsilon; \tag{14}$$

this is the product measure of the mean measure of the Poisson process $\{c_j\}$ and the probability measure of the ε_j.

Bayesian inference requires the use of the likelihood function

$$f(X|\lambda, \theta, \mu) = \int f\{X|\mu, v(0), \Psi\}\, \pi\{v(0)|\lambda, \theta, \mu\}\, \pi(\mathrm{d}\Psi|\lambda, \theta)\, \mathrm{d}v(0), \tag{15}$$

where $f\{X|\mu, v(0), \Psi\} = f(X|v^*)$ is defined in equation (8). The integration in equation (15) is analytically intractable, so the likelihood function is not available in closed form. Nevertheless, the fact that we can compute $f\{X|\mu, v(0), \Psi\}$ for any given values for Ψ and $v(0)$ suggests treating them as missing data and making use of data augmentation methodologies; see for example Tanner and Wong (1987) and Smith and Roberts (1993). Here, we use the term 'missing data' in the general sense of Meng (2000), which includes hidden states or processes, latent or auxiliary variables and more broadly unobserved or unobservable variables.

In what follows we shall be interested in the probability density function of Ψ with respect to some suitable reference distribution. As such we can choose the measure $\mathcal{Q}$ of a Poisson process on S with mean measure $\exp(-\varepsilon)\, \mathrm{d}c\, \mathrm{d}\varepsilon$. Actually, we could choose the measure of any Poisson process on S with mean measure that is independent of the parameters λ and θ and dominates $\Lambda(\cdot)$. Let $\Psi(S)$ be the number of points of Ψ on S, where by construction $\Psi(S) < \infty$. Then the probability measure of Ψ is

$$\pi(\mathrm{d}\Psi \mid \lambda, \theta) = \exp\{-(\lambda - 1)T\}(\lambda\theta)^{\Psi(S)} \exp\left\{-(\theta - 1) \sum_{j=1}^{\Psi(S)} \varepsilon_j\right\} \mathcal{Q}(\mathrm{d}\Psi) \tag{16}$$

where the sum is replaced by 0 if $\Psi(S) = 0$, and it can be verified that this is the probability measure of a Poisson process on S with mean measure $\Lambda(\cdot)$. It is now clear that the density of Ψ (with respect to $\mathcal{Q}$) is

$$\pi(\Psi \mid \lambda, \theta) = \exp\{-(\lambda - 1)T\}(\lambda\theta)^{\Psi(S)} \exp\left\{-(\theta - 1) \sum_{j=1}^{\Psi(S)} \varepsilon_j\right\}. \tag{17}$$

The (λ, θ, μ) parameterization is naturally suggested by the above construction. Nevertheless, our experience suggests that the mean volatility $\lambda/\theta\mu$ is well estimated from the data, inducing high posterior dependence among the parameters and hence slow MCMC mixing. Therefore, for our MCMC algorithms we adopt the transformation $(\lambda, \theta, \mu) \to (\nu, \theta, \mu) = (\lambda/\mu, \theta, \mu)$ that was suggested by Barndorff-Nielsen and Shephard (2001). In this parameterization, θ and ν describe the stationary properties of the process, whereas μ describes the rate of decay, and hence the transient properties of the process. Therefore we would expect μ not to be excessively correlated with ν and θ. Furthermore, the interpretation in terms of stationary and transient properties makes prior elicitation easier under the new parameterization.

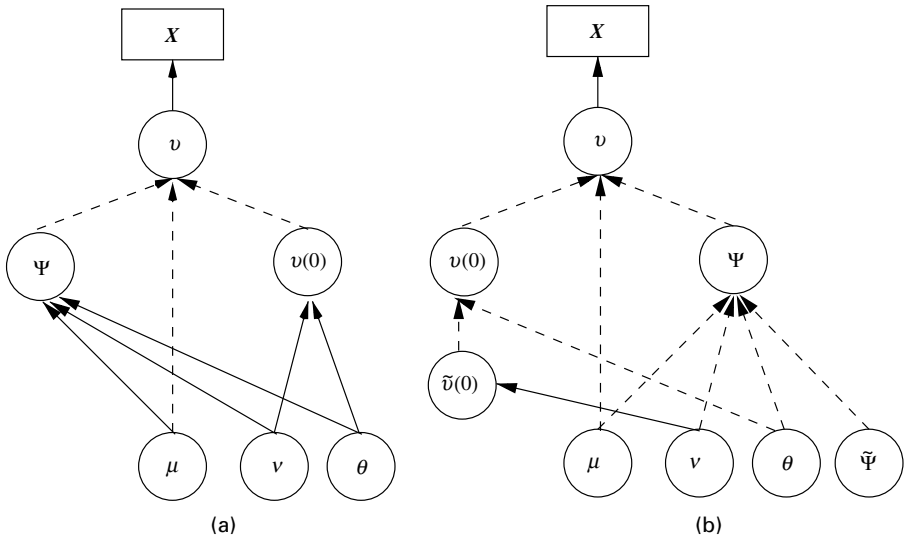

Fig. 2. Graphical model of (a) the CP and (b) the NCP

3.2. Centred parameterization

A straightforward hierarchical parameterization of the augmented model, which we shall term the *centred parameterization* (CP; see Section 3.3) can be constructed by writing the posterior distribution of the parameters and the missing data as

$$\begin{aligned}\pi\{\nu,\theta,\mu,\Psi,v(0)|X\} &\propto f\{X|\Psi,v(0),\mu\}\,\pi\{v(0)|\nu,\theta\}\,\pi(\Psi|\nu,\theta,\mu)\,\pi(\nu,\theta,\mu)\\ &\propto f\{X|\Psi,v(0),\mu\}\,\frac{\theta^{\nu}}{\Gamma(\nu)}v(0)^{\nu-1}\exp\{-\theta\,v(0)-(\nu\mu-1)T\}\\ &\quad\times(\nu\mu\theta)^{\Psi(S)}\,\exp\left\{-(\theta-1)\sum_{j=1}^{\Psi(S)}\varepsilon_j\right\}\pi(\nu,\theta,\mu),\end{aligned}\tag{18}$$

where $\pi(\nu,\theta,\mu)$ is the joint prior density and $\Gamma(\cdot)$ is the gamma function. The dependence structure between the parameters and the missing data is presented as a directed acyclic graphical model in Fig. 2(a). A key feature of this parameterization is that the parameters ν and θ are conditionally independent of the data given μ and the missing data Ψ and $v(0)$.

To complete our model specification we need to define the prior density $\pi(\nu,\theta,\mu)$. We choose a conditional conjugate $\mathrm{Ga}(\alpha_\theta,\beta_\theta)$ prior for θ, a $\mathrm{Ga}(\alpha_\mu,\beta_\mu)$ prior for μ and also a $\mathrm{Ga}(\alpha_\nu,\beta_\nu)$ prior for ν. A discussion of the choice of hyperparameters is deferred to Section 5.

We use a single-component updating MCMC algorithm (which replaces direct conditional simulations in a Gibbs sampler by Metropolis–Hastings steps) to obtain samples from distribution (18). The algorithm, which we call the *centred algorithm* (CA) is described in detail in Appendix A.

3.3. Alternatives to the centred parameterization

The parameterization that is shown in Fig. 2(a) is statistically natural, representing the hierarchy in which the parameters (ν,θ,μ) are used to construct the latent process Ψ, which together with μ and $v(0)$ determine the distribution of the observed data. Since the missing data Ψ and $v(0)$ are hierarchically *centred* between the observed data X and the parameters

(ν, θ), we call the parameterization in Fig. 2(a) the CP. This terminology appeared originally in the context of normal hierarchical models (Gelfand *et al.*, 1995), where it relates to the traditional analysis-of-variance use of the term. As a generalization, Papaspiliopoulos *et al.* (2003) characterized as centred the parameterizations of two-stage hierarchical models, where the parameters are conditionally independent from the observed data given the missing data. CPs from different contexts share certain statistical properties which in turn lead to similar algorithmic characteristics, so it is instructive to consider CPs as a natural family of parameterizations for statistical models. In some simple contexts, CAs can be fully analysed (Roberts and Sahu, 1997), and the conclusions provide important guidelines for more general models.

CPs for hierarchical models are computationally convenient when used in conjunction with the Gibbs sampler or more general single-component MCMC updating techniques. The inherent conditional independence when combined with conjugacy results in relatively easy simulation steps. However, apart from simplicity in implementation, another important aspect of any MCMC algorithm is the speed of convergence. Algorithms which converge slowly produce samples from the posterior distribution which exhibit strong serial correlation, so posterior estimates have high variance. Furthermore, they might fail to visit the tails of the target distribution, thus underestimating the posterior uncertainty. Roberts (2003) has given more detail and precise statements regarding the link between convergence of MCMC algorithms and estimation.

The convergence of the Gibbs sampler (and more generally of the single-component MCMC updating algorithms) critically relies on the dependence between the updated components (Liu *et al.*, 1994; Roberts and Sahu, 1997); strong dependence leads to slow convergence. Therefore, the parameterization of a hierarchical model has a substantial effect on the convergence rate of the MCMC algorithm which is used to sample from the posterior distribution of the parameters and the missing data. The CP implies that when the parameters are updated, as part of an iteration of the Gibbs sampler, the missing data are treated as if they were actually observed and the observed data are ignored, owing to conditional independence. Hence, the sampler will converge slowly when the expected information about the parameters that is contained in the missing data is considerably greater than that contained in the observed data. Informally, in this case the conditional distribution of the parameters given the missing data is much less dispersed than the marginal posterior distribution. Therefore, each iteration of the Gibbs sampler makes very small steps in the parameter space, compared with the actual spread of the target distribution, and convergence is slow. The situation is reversed when the missing data are well identified. Then, the conditional distribution of the parameters given the missing data is close to the marginal posterior distribution, and the Gibbs sampler generates essentially independent draws from the posterior distribution of the parameters.

In the hidden Markov process case, of which the SV models of this paper are special cases, problems caused by dependence between the parameters and the latent process itself can be acute. This is particularly problematic for long time series where the prior structure will be bound by 'ergodicity constraints' which link long-term empirical properties of the latent process with their stationary expectations, which are just functions of the parameters. For our model, an example is

$$\frac{1}{T}\int_0^T v(s)\,\mathrm{d}s \approx \frac{\lambda}{\theta\mu}.$$

Thus, unless the data are sufficiently informative to identify $v(\cdot)$, extremely high posterior correlation will exist between $\int_0^T v(s)\,\mathrm{d}s$ and $\lambda/\theta\mu$, or equivalently between Ψ and λ, leading to poor convergence of the CA of Section 3.2.

One way to improve the efficiency of the single-component updating MCMC algorithm is to reparameterize the target posterior distribution to reduce the dependence between the updated components. We have argued that, under a CP, the prior dependence between the parameters and the missing data can induce strong posterior dependence. Therefore, Papaspiliopoulos *et al.* (2003) proposed a reparameterization of the hierarchical model, which 'orthogonalizes' the prior structure, in the sense that the missing data are transformed so that they are *a priori* independent of the parameters. When the augmented information about the parameters is much higher than the marginal, the transformed missing data and the parameters will be weakly dependent *a posteriori* and the single-component updating MCMC algorithm will converge fast. These parameterizations are called *non-centred parameterizations* (NCPs); see Papaspiliopoulos (2003) and Papaspiliopoulos *et al.* (2003) for a general description and discussion of this methodology. Such a parameterization will be constructed in the next section for the OU models of Sections 2.1 and 2.2. For these models the observed data are more informative about Ψ

(a) the higher the number of data points per Poisson jump (small λ compared with data frequency),
(b) the higher the persistence in the volatility (small μ) and
(c) the smaller the variance of the volatility, when the mean is kept fixed.

Observations (b) and (c) are empirically supported by the simulation study in Barndorff-Nielsen and Shephard (2002). Moreover, they are consistent with convergence rate results for Gaussian models; see Roberts and Sahu (1997), Pitt and Shephard (1999) and Papaspiliopoulos *et al.* (2003) for a recent review. The intuition behind (b) is that many consecutive log-returns will have essentially the same underlying volatility. In contrast the information about λ contained in Ψ increases with T and decreases with $\Psi(S)$.

3.4. Reparameterization and non-centring

We propose a non-centred reparameterization $(\lambda, \theta, \Psi) \to (\lambda, \theta, \tilde{\Psi})$ so that $\tilde{\Psi}$ is *a priori* independent of the parameters. The conditional independence structure can now be written as

$$\pi\{\nu, \theta, \mu, \tilde{\Psi}, v(0)|X\} \propto f\{X|\tilde{\Psi}, v(0), \nu, \mu, \theta\}\, \pi\{v(0)|\nu, \theta\}\, \pi(\tilde{\Psi})\, \pi(\nu, \theta, \mu) \tag{19}$$

and is depicted as a graphical model in Fig. 2(b).

Our transformation constructs $\tilde{\Psi}$ on a higher dimensional space than that of Ψ. Let $\tilde{\Psi}$ be a marked Poisson process with points $\{(c_j, m_j)\}$ on $[0, T] \times (0, \infty)$ and marks $\tilde{\varepsilon}_j$ on $(0, \infty)$ independent of each other and independent of the points; we take

$$\tilde{\Lambda}(\mathrm{d}c \times \mathrm{d}m \times \mathrm{d}\tilde{\varepsilon}) = \exp(-\tilde{\varepsilon})\, \mathrm{d}c\, \mathrm{d}m\, \mathrm{d}\tilde{\varepsilon} \tag{20}$$

to be the mean measure of $\tilde{\Psi}$. In this setting, the points $\{(c_j, m_j)\}$ form a homogeneous Poisson process with unit intensity on $[0, T] \times (0, \infty)$ and $\tilde{\varepsilon}_j \sim \mathrm{Ex}(1)$. The restriction of $\tilde{\Psi}$ to $[0, T] \times (0, \lambda) \times (0, \infty)$, i.e. all points of $\tilde{\Psi}$ with $m_j < \lambda$, will still be a Poisson process with mean measure the restriction of $\tilde{\Lambda}(\cdot)$ to this set (page 17 of Kingman (1993)). The mapping theorem (page 18 of Kingman (1993)) can be used to show that the projection of these points to $[0, T] \times (0, \infty)$ is a Poisson process with mean measure

$$\int_0^{\lambda} \tilde{\Lambda}(\mathrm{d}c \times \mathrm{d}m \times \mathrm{d}\tilde{\varepsilon}) = \lambda \exp(-\tilde{\varepsilon})\, \mathrm{d}c\, \mathrm{d}\tilde{\varepsilon}. \tag{21}$$

Therefore, we might retrieve (Ψ, λ, θ) from $(\tilde{\Psi}, \lambda, \theta)$ as follows (see also Fig. 3(a)).

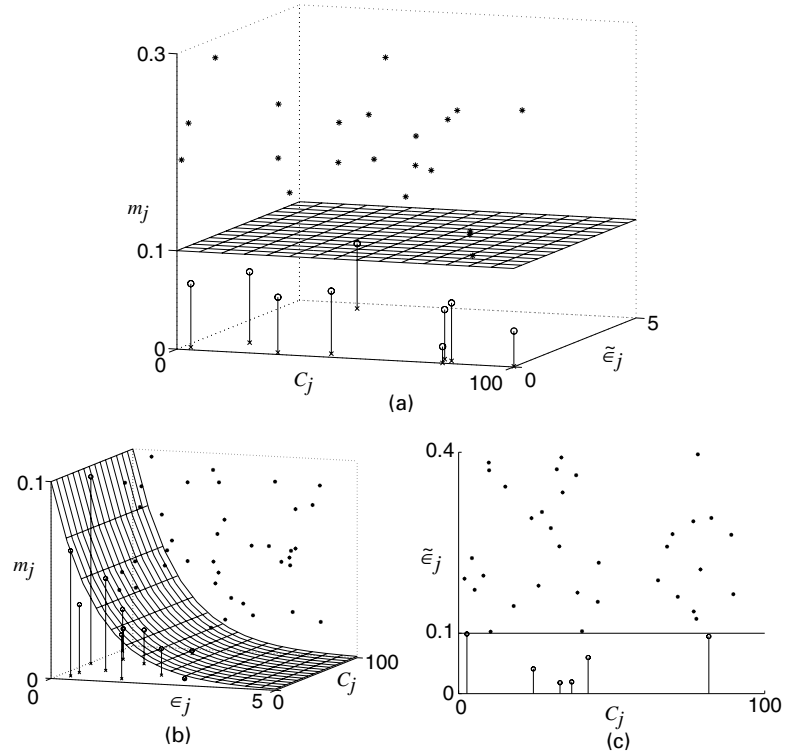

Fig. 3. Different NCPs of (λ, θ, Ψ) described in detail in Sections 3.4 and 3.6 ($\lambda = 0.1$, $\theta = 1$, $S = [0, T] \times (0, \infty)$ and $T = 100$): (a) $\tilde{\Psi}$ is a Poisson process on $[0, T] \times (0, \infty) \times (0, \infty)$ with mean measure $\exp(-\tilde{\varepsilon})\, \mathrm{d}c\, \mathrm{d}m\, \mathrm{d}\tilde{\varepsilon}$—choose all $(c_j, m_j, \tilde{\varepsilon}_j) \in \tilde{\Psi}$ with $m_j \leqslant \lambda$ (∘), project them to S and set $\varepsilon_j = \tilde{\varepsilon}_j/\theta$; (b) $\tilde{\Psi}$ is a unit rate Poisson process on $S \times (0, \infty)$—choose all $(c, \varepsilon_j, m_j) \in \tilde{\Psi}$ with $m_j < \lambda\theta \exp(-\theta\varepsilon_j)$ and project them to S; (c) $\tilde{\Psi}$ is a unit rate Poisson process on S—choose all $(c_j, \tilde{\varepsilon}_j) \in \tilde{\Psi}$ with $\tilde{\varepsilon}_j < \lambda$ and set $\varepsilon_j = -\log(\tilde{\varepsilon}_j/\lambda)/\theta$

Step 1: select all points from $\tilde{\Psi}$ for which $m_j < \lambda$.
Step 2: project these points to $[0, T] \times (0, \infty)$.
Step 3: transform $\{(c_j, \tilde{\varepsilon}_j)\}$ to $\{(c_j, \varepsilon_j)\}$ where $\varepsilon_j = \tilde{\varepsilon}_j/\theta$; Ψ consists of the transformed points.

This transformation can be seen as a simulation method to generate realizations from a Poisson process with some parameterized mean measure, using independent random components. Therefore, it is also a representation of the Poisson process, although not a very useful one for theoretical purposes.

A further transformation, that turns out to be very useful because it results in a conditional conjugate Gibbs step for θ, and makes θ *a priori* independent of ν, is from $v(0)$ to $\tilde{v}(0) = \theta\, v(0)$. The posterior density (19) can then be written as

$$\pi\{\nu, \theta, \mu, \tilde{\Psi}, \tilde{v}(0) | X\} \propto f\{X | \tilde{\Psi}, \tilde{v}(0), \nu, \theta, \mu\} \frac{1}{\Gamma(\nu)} \tilde{v}(0)^{\nu-1} \exp\{-\tilde{v}(0)\} \pi(\tilde{\Psi})\, \pi(\nu, \theta, \mu), \tag{22}$$

where we choose the prior $\pi(\nu, \theta, \mu)$ that was suggested in Section 3.2. The corresponding MCMC algorithm, which we call the *non-centred algorithm* (NCA), is described in Appendix B.

3.5. A non-centred parameterization for the superposition of Ornstein–Uhlenbeck processes

We construct an NCP for the model that was presented in Section 2.2. The parameters of interest are $\nu_1,\ \nu_2,\ \mu_1,\ \mu_2$ and θ and we treat as missing data the initial volatilities $v_1(0)$ and $v_2(0)$ and the marked Poisson process Ψ_i on S with points $\{(c_{ij}, \varepsilon_{ij})\}$, $i = 1, 2$.

The NCP works with a marked Poisson process $\tilde{\Psi}$ with points $\{(c_j, m_j)\}$ on $[0, T] \times (-\infty, \infty)$ and marks $\tilde{\varepsilon}_j$ on $(0, \infty)$ independent of each other and independent of the points. We take

$$\tilde{\Lambda}(\mathrm{d}c \times \mathrm{d}m \times \mathrm{d}\tilde{\varepsilon}) = \exp(-\tilde{\varepsilon})\ \mathrm{d}c\ \mathrm{d}m\ \mathrm{d}\tilde{\varepsilon}$$

to be the mean measure of $\tilde{\Psi}$. The marked Poisson process that was introduced in Section 3.4 is the same as $\tilde{\Psi}$ defined above, but restricted on the subset $[0, T] \times (0, \infty) \times (0, \infty)$. Transformation of $(\lambda_1, \lambda_2, \theta, \tilde{\Psi})$ to $(\lambda_1, \lambda_2, \theta, \Psi_1, \Psi_2)$ might be done as follows.

Step 1: select all points from $\tilde{\Psi}$ for which $-\lambda_2 < m_j < \lambda_1$; denote those with positive m_j as $\{(c_{1j}, m_{1j}, \tilde{\varepsilon}_{1j})\}$ and the rest as $\{(c_{2j}, m_{2j}, \tilde{\varepsilon}_{2j})\}$.
Step 2: project all points to $[0, T] \times (0, \infty)$.
Step 3: Ψ_1 consists of the points $\{(c_{1j}, \varepsilon_{1j})\}$, where $\varepsilon_{1j} = \tilde{\varepsilon}_{1j}/\theta$, and Ψ_2 consists of the points $\{(c_{2j}, \varepsilon_{2j})\}$, where $\varepsilon_{2j} = \tilde{\varepsilon}_{2j}/\theta$.

We also transform $v_i(0)$ to $\tilde{v}_i(0) = v_i(0)/\theta,\ i = 1, 2$, and the resulting posterior density is

$$\begin{aligned}\pi\{\nu_1, \nu_2, \mu_1, \mu_2, \theta, \tilde{\Psi}, \tilde{v}_1(0), \tilde{v}_2(0) | X\} \\ \propto f\{X | \tilde{\Psi}, \tilde{v}_1(0), \tilde{v}_2(0), \nu_1, \nu_2, \mu_1, \mu_2, \theta\} \frac{\tilde{v}_1(0)^{\nu_1 - 1}}{\Gamma(\nu_1)} \frac{\tilde{v}_2(0)^{\nu_2 - 1}}{\Gamma(\nu_2)} \\ \times \exp\{-\tilde{v}_1(0) - \tilde{v}_2(0)\}\ \pi(\tilde{\Psi})\ \pi(\nu_1, \nu_2, \mu_1, \mu_2, \theta)\ \mathbf{1}(\mu_1 > \mu_2). \qquad (23)\end{aligned}$$

We assign a $\mathrm{Ga}(\alpha_{\nu_j}, \beta_{\nu_j})$ prior to ν_j, $j = 1, 2$, and a $\mathrm{Ga}(\alpha_\theta, \beta_\theta)$ prior to θ and construct the joint prior of (μ_1, μ_2) by imposing a $\mathrm{Ga}(\alpha_{\mu_2}, \beta_{\mu_2})$ prior to μ_2 and a $\mathrm{Ga}(\alpha_{\mu_1}, \beta_{\mu_1})$ prior to $\mu_1 - \mu_2$ given μ_2. Therefore, prior elicitation on (μ_1, μ_2) can be done by specifying the marginal prior distribution of μ_2 and by modelling how much larger μ_1 is expected to be than μ_2.

The corresponding MCMC algorithm is similar to the algorithm that was used for the single-OU model; details can be found in Appendices B and C.

3.6. Alternative non-centred parameterizations

It turns out that there is more than one NCP for the OU models that are considered in this paper. They all correspond to the same graphical model in Fig. 2(b) but assign different priors on $\tilde{\Psi}$ and consequently result in different transformations $(\lambda, \theta, \tilde{\Psi}) \to \Psi$. We shall sketch two of them for the single-OU model, which are also based on thinning ideas. The first takes $\tilde{\Psi}$ to be a homogeneous Poisson process on $S \times (0, \infty)$, and Ψ consists of the projection on S of all points $(c, \varepsilon, m) \in \tilde{\Psi}$ such that $m < \lambda\theta \exp(-\theta\varepsilon)$ (see Fig. 3(b)). The second takes $\tilde{\Psi}$ on the same space as Ψ but treats the latter as a marked point process with points the ordered jump sizes marked with the corresponding jump times. The technique is essentially based on the inversion method for simulating from the increments of the Poisson process that is formed by the ordered jump sizes. We take $\tilde{\Psi}$ to be a unit rate Poisson process on S, then choose all $(c, \tilde{\varepsilon}) \in \tilde{\Psi}$ such that $\tilde{\varepsilon} < \lambda$ (see Fig. 3(c)) and then transform (using an inversion method) $\varepsilon = -\log(\tilde{\varepsilon}/\lambda)/\theta$. This is the idea behind the representation of Ferguson and Klass (1972), which can be seen as a non-centring technique, and the corresponding parameterization has been adopted by Griffin and Steel (2002). The main difference from the other methods that we have discussed is that,

when we update λ, we either remove or add the smallest jump rather than a randomly chosen jump.

The MCMC algorithms corresponding to both of these parameterizations iterate the steps given in Appendix B, but they differ in the implementation of step 1 and the transformation steps 2 and 5. In particular, the blocking scheme that is suggested in the Appendix B is not possible in any of the alternative algorithms, since the conditional distribution of θ is not available; see Appendix B for details. We shall not go into any more detail and we shall not consider either of these two alternatives in this paper until Section 7.

4. Simulation study

To assess the performance of the CA and the NCA, we applied them to a varied collection of data simulated from the OU model of Section 2.1. The results for six experiments are presented here, representing different types of dynamic structure and stationary moments to the SV process, and different time series lengths. Our purpose is to obtain an understanding of the kind of data for which each algorithm is more suitable and consequently to provide guidelines on when each should be preferred.

All experiments are done on the assumption of daily data, since having finer data would be equivalent to making μ smaller while fixing ν. Furthermore, variation in θ is also not necessary since its effect can be removed by scaling the observed data. A different simulation design is considered in section 3 of Barndorff-Nielsen and Shephard (2002), where the mean volatility ν/θ is kept fixed, and θ varies, to assess how different estimators (nonparametric and model based) of the integrated volatility perform, using high frequency data.

Table 1 summarizes the parameter values that are used in simulating the six series. The first four use shorter (length 500) time series. The first two of these have the same memory decay ($\mu = 0.03$) but different stationary distributions for the volatility process. The situation is similar in series 3 and 4, but where $\mu = 0.1$, so the volatility processes are less persistent. Series 5 and 6 consist of longer sequences corresponding to the parameter values in series 1 and 4 respectively.

The prior $\pi(\nu, \theta, \mu)$ was chosen to be flat around the true parameter values, since an extremely informative prior could potentially mask problems of convergence in either the CA or the NCA (Roberts and Sahu, 1997; Papaspiliopoulos *et al.*, 2003). In particular, in all the experiments a $\mathrm{Ga}(1, 0.1)$ prior was chosen for θ and ν and a $\mathrm{Ga}(1, 1)$ prior for μ. The problem of choosing priors for the parameters and the latent process is considered in detail in Section 5.

For each of the six series each algorithm was run for 6 million iterations, with the first 50 000 iterations removed as a burn-in period and subsequent parameter values stored every 100th

Table 1. Information about the simulated data sets

Data set	θ	ν	μ	*Length of series*	*Volatility stationary law*
1	10	$\frac{2}{3}$	0.03	500	$\mathrm{Ga}(\frac{2}{3}, 10)$
2	10	2	0.03	500	$\mathrm{Ga}(2, 10)$
3	10	$\frac{2}{3}$	0.1	500	$\mathrm{Ga}(\frac{2}{3}, 10)$
4	10	2	0.1	500	$\mathrm{Ga}(2, 10)$
5	10	$\frac{2}{3}$	0.03	2000	$\mathrm{Ga}(\frac{2}{3}, 10)$
6	10	2	0.1	2000	$\mathrm{Ga}(2, 10)$

iteration. The parameters were started at their true values in the CA but at their prior means in the NCA, for reasons which will be clarified later in this section. The computing time for the two algorithms is comparable: 1000 iterations (while chains were in stationarity) for series 1 took 2.75 s of central processor unit time for the CA and 4.47 s for the NCA on a Pentium III 450 MHz processor. All programs were coded in Fortran.

Fig. 4 shows the estimated autocorrelation function for the MCMC time series (remaining after removing the burn-in samples and thinning) corresponding to $\lambda = \nu\mu$ for the CA and the NCA, for each of the data sets in Table 1. The integrated autocorrelation corresponding to a real function of the Markov chain is proportional to the Monte Carlo error of the ergodic average estimate of the stationary mean of this function (Geyer, 1992). Therefore,

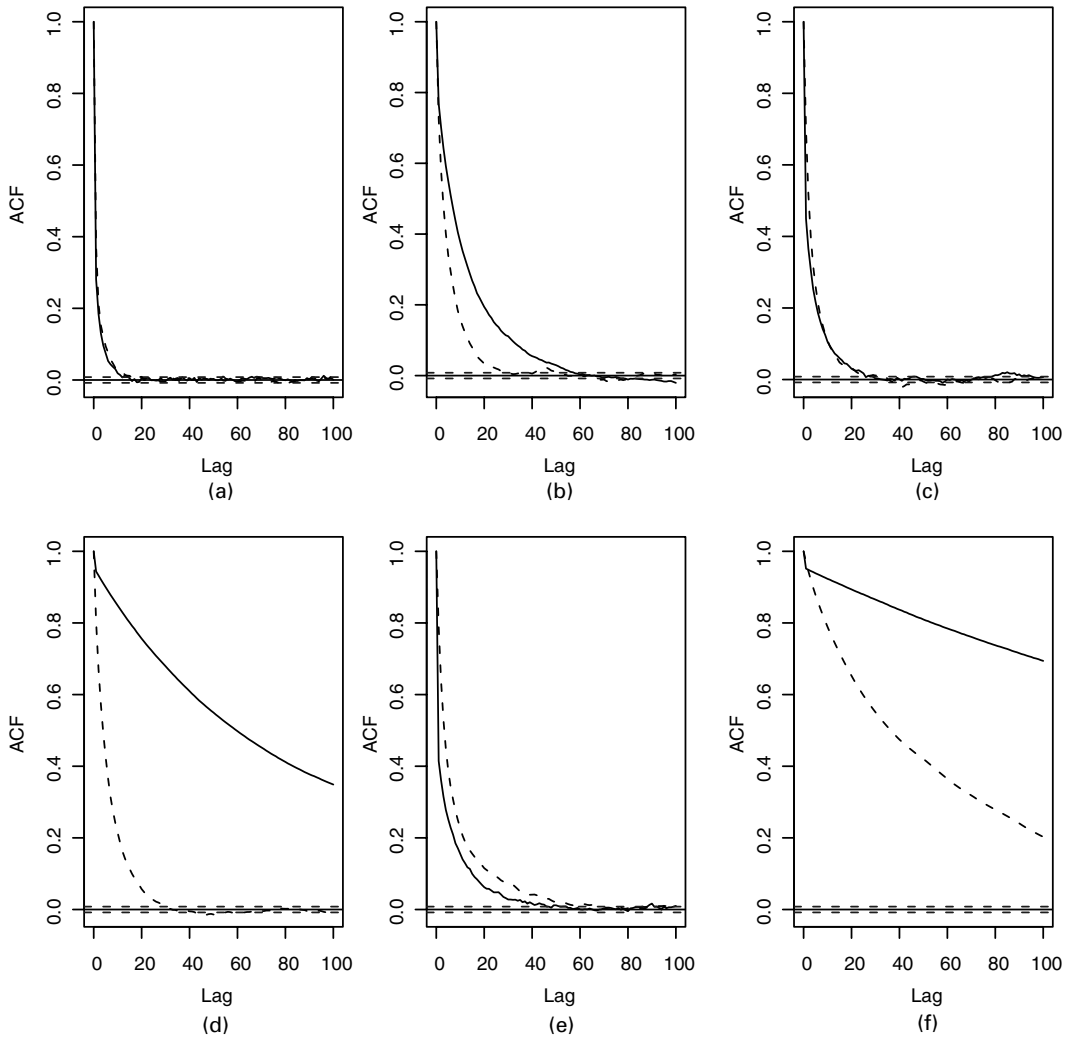

Fig. 4. Estimates of the autocorrelation function of the marginal chain corresponding to $\lambda = \nu\mu$ for the CA (———) and the NCA (- - - - - - -) for each of the simulated data sets of Table 1 (the chains have been thinned taking one every 100 iterations, with a burn-in of 50 000 iterations): (a) series 1, $\nu = \frac{2}{3}$, $\mu = 0.03$; (b) series 2, $\nu = 2$, $\mu = 0.03$; (c) series 3, $\nu = \frac{2}{3}$, $\mu = 0.1$; (d) series 4, $\nu = 2$, $\mu = 0.1$; (e) series 5, $\nu = \frac{2}{3}$, $\mu = 0.03$; (f) series 6, $\nu = 2$, $\mu = 0.1$

when the estimated autocorrelations remain non-negligible for large numbers of iterations, this gives an indication of a slowly mixing chain, which produces highly variable estimates. Since it is impossible to calculate exact convergence rates, we choose the more informal way of examining autocorrelation plots to compare the efficiency of the two competing algorithms. Autocorrelation functions of different parameters give similar results to those given in Fig. 4.

An initial inspection of the results reveals considerable robustness of the NCA to a variety of data sets, as opposed to the CA which seems to be doing badly in some cases. Owing to the complexity of the problem a detailed explanation of the observed behaviour of the algorithms is a daunting task. For example, if the underlying Poisson rate is small, then, although the data are very informative about Ψ (Section 3.3), prior dependence between Ψ and λ increases and it is not clear whether the CA or the NCA should be used. Our experience suggests that the CA should be favoured for high frequency data; otherwise, the NCA appears to be more efficient. Note, however, that the computer algorithms for both are quite similar since they differ only in the step of updating the parameters given the missing data (see the details in Appendix B), and it is not difficult to code.

The NCA is much more robust to starting values than the CA. We performed a run of both algorithms on series 1 with all parameters started from their prior means and the difference between the two approaches was striking. The chain corresponding to λ in the NCA returned immediately from the tails to the modal area, whereas in the CA it took more than 10^6 iterations for the algorithm to reach near the mode. It appears that the posterior of the $\tilde{\Psi}$ depends less on λ as $\lambda \to \infty$, allowing λ to return from the tails to the mode very quickly. We have found that for many non-Gaussian hierarchical models the dependence between the missing data and the parameters in the tails is very different from that in the mode (unlike Gaussian models where it is constant) and the convergence behaviour of the CA can be qualitatively very different from that of the NCA. For a more thorough analysis of this type of behaviour in a simpler context though, see Roberts and Papaspiliopoulos (2003).

The slow return of λ to the mode in the CA is partly due to its posterior dependence with Ψ and partly because a single birth–death step is used to update Ψ. In this way, the updating step for Ψ is less efficient for larger $\Psi(S)$ values. Nevertheless, if it were possible to simulate Ψ directly from its conditional posterior distribution and thus to speed up the CA, it would still be much slower than the NCA. The sensitivity to starting values is a serious problem, largely because unstable excursion behaviour indicates problems of the sampler in exploring tail areas and therefore underestimating uncertainty (Roberts, 2003).

Fig. 5 shows how the chain corresponding to Ψ evolves towards its posterior distribution under series 1. The current configuration of the jump times is plotted against every 1000th iteration for the first 60 000 iterations. The degree of darkness of the points within each configuration reflects their relative jump sizes on a four-colour grey scale, with black corresponding to the largest jumps. The configuration of the jump times that was used in simulating series 1 is given at the far right-hand side. Such plots provide reassurance that the MCMC algorithm moves around the point process space, but they can also be used to identify shocks in the volatility when the model is fitted to real data.

We have tried the NCA for the superposition of OU processes on simulated data and we have obtained satisfactory results, which for brevity are not reproduced here. The parameters are well identified and the mixing of the algorithm is good, although slower than for the single-OU model. We are currently investigating methods which speed up the convergence of the algorithm; see Papaspiliopoulos (2003) for a discussion of this issue and for suggestions for improvements.

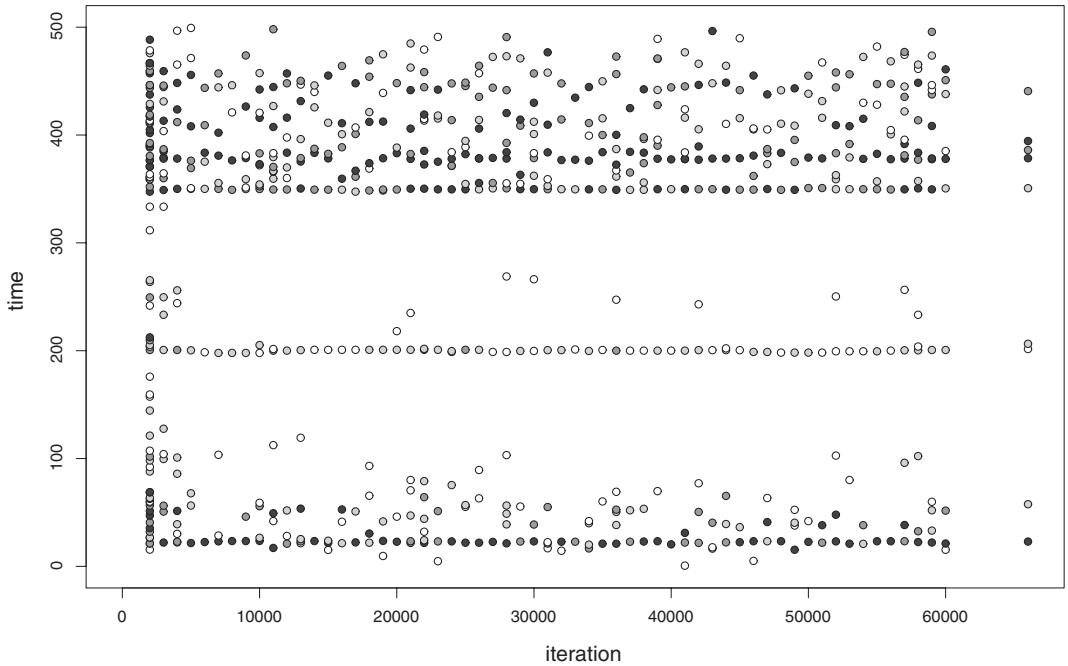

Fig. 5. MCMC output for the point process Ψ for series 1, for the NCA: the configuration of the jump times is being plotted against every 1000th iteration for the first 60 000 iterations of the algorithm; the degree of darkness of the points within each configuration reflects their relative jump sizes; on the far right-hand side the configuration of the jump times that is used in the simulation of the data is plotted

5. Prior sensitivity analysis and diagnostics

To assess the sensitivity of the posterior distributions of the parameters to prior specification, we reanalysed series 1 and 4 of Table 1 assuming that the parameters were *a priori* exponentially distributed with mean given by their true values. The results (which are not reproduced here) indicate reasonable robustness of the posteriors to very different priors.

If we keep ν and θ fixed and let $\mu \to \infty$, the likelihood does not converge to 0; instead it corresponds to a model where the log-returns are independent with the appropriate mixture of Gaussian distributions. Therefore, it is necessary to choose a proper prior for μ. This necessity is well known in latent variable models and has attracted the interest of many researchers, especially in the area of finite mixture models (Diebolt and Robert, 1994; Roeder and Wasserman, 1997). In this case, since even the use of arbitrary vague proper priors is not adequate, 'default' priors have been proposed that are, somehow, data dependent (Richardson and Green, 1997; Robert, 1996). However, it can be argued (section 2.5 of Griffin and Steel (2002)) that a Ga(1, 1) density expresses weak prior information about the volatility correlation decay.

When both the single-OU and the two-OU models are applied to a data set, a sensible prior specification for the latter is the following. We parameterize in terms of $(\nu, w_2, \theta, \mu_1, \mu_2)$, where $\nu = \nu_1 + \nu_2$ and $w_2 = \nu_2/(\nu_1 + \nu_2)$ (see equation (11)). Under this setting (ν, θ) are the parameters of the stationary distribution of the volatility in both models so we can specify a common prior. It is also reasonable to assume that $w_2 \sim \text{Un}[0, 1]$. The prior on (μ_1, μ_2) can be specified according to the recommendation that is given in Section 3.5.

It is interesting to investigate the strength of prior assumptions regarding the latent structure. For this, we have developed some simple graphical diagnostics which use the MCMC output of Ψ to assess the Poisson and exponential assumptions about the jump times and the jump sizes respectively, which the OU models in Sections 2.1 and 2.2 make. Our diagnostics are based on ergodic properties of the model, so they are informative when λT is quite large.

Let $\tilde{C} = \{\lambda c_1, \lambda c_2, \ldots\}$ and $\tilde{E} = \{\tilde{\varepsilon}_1, \tilde{\varepsilon}_2, \ldots\}$ where $\tilde{\varepsilon}_j = \theta \varepsilon_j$ and c_j and ε_j are defined in Section 2.1. Using MCMC samples from the posterior distribution of $\tilde{C}$ and $\tilde{E}$, we graphically examine whether they are consistent with the prior assumptions, namely that $\lambda(c_j - c_{j-1}) \sim \mathrm{Ex}(1)$ and that $\tilde{\varepsilon}_j \sim \mathrm{Ex}(1)$. Our diagnostic plots rely on the property that if $\varepsilon \sim \mathrm{Ex}(1)$ then $-\log P(\varepsilon > t) = t$, $t > 0$. Thus, systematic deviations from a straight line can be inspected by plotting an empirical estimate of $-\log\{P(\varepsilon > t)\}$ against t. If the model is the 'correct' model, then owing to the ergodic properties of the time series model we would expect the sample of the 'residuals' in $\tilde{C}$ and $\tilde{E}$ to be drawn from the prior distribution, since we implicitly average over the empirical distribution of the observed data. We aim at two things: first, to apply our diagnostics to real data as a method of assessing model fit; second, to apply them to data simulated from a different model and to examine whether the misspecification is reflected in the posterior distribution of the parameters and the missing data.

In Fig. 6 the diagnostics are tried for various simulated time series in $[0, 2000]$. Data set 6 (Table 1) is used in Fig. 6(a). The data that were used in Figs 6(b) and 6(c) have been simulated by using the same underlying OU process for both but with a different sampling frequency. The OU process has memory parameter $\mu = 0.1$ and it is driven by a compound Poisson process as in equation (3) where $\lambda = 0.2$, but the jump sizes have been simulated from a $\mathrm{Ga}(0.1, 1)$ prior. One and 100 data points per day were simulated in Figs 6(b) and 6(c) respectively. For daily data, the diagnostics provide weak evidence about misspecification and we have found that the evidence varies considerably across different data sets simulated for the same realization of the OU process. Thus, for this observation frequency some aspects of the model cannot be well identified. However, if high frequency data are used, the diagnostics strongly indicate model misspecification. The posterior means for the parameters (ν, θ, μ) are (1.97,9.86,0.09) in Fig. 6(a), (0.59,2.43,0.089) in Fig. 6(b) and (0.68,2.7,0.098) in Fig. 6(c).

A careful analysis of the applicability and the interpretation of these diagnostic tools will be reported elsewhere.

6. A real data example

We fitted the models of Sections 2.1 and 2.2 to the series of US dollar–Deutschmark exchange rates. The data were obtained from JP Morgan and are daily closing prices that span the period from January 1st, 1986, to January 1st, 1996 (2614 data points in total). We have scaled the original log-prices by a multiplicative factor of $1000^{1/2}$. This transformation affects only the parameter that controls the distribution of the jump sizes, i.e. θ in the models that are considered in this paper. This FX market has been studied in detail by Andersen *et al.* (2001) and by Barndorff-Nielsen and Shephard (2002) for a similar period of time (December 1st, 1986–November 30th, 1996). However, they used the Olsen high frequency data.

For the single-OU model, a $\mathrm{Ga}(1, 0.01)$ prior is chosen for θ, a $\mathrm{Ga}(1, 0.1)$ for ν and a $\mathrm{Ga}(1, 1)$ for μ. For the two-OU model, we follow Section 5 and parameterize in terms of $\nu = \nu_1 + \nu_2$ and $w_2 = \nu_2/(\nu_1 + \nu_2)$. We choose a $\mathrm{Ga}(1, 0.1)$ prior for ν, a $\mathrm{Un}[0, 1]$ for w_2, a $\mathrm{Ga}(1, 0.01)$ prior for θ, a $\mathrm{Ga}(1, 1)$ for μ_2 and a $\mathrm{Ga}(1, 0.01)$ for $\mu_1 - \mu_2$ (conditionally on μ_2).

Tables 2 and 3 provide some posterior summaries for the parameters from fitting the single-OU and the two-OU models respectively. In particular, we report summaries for the

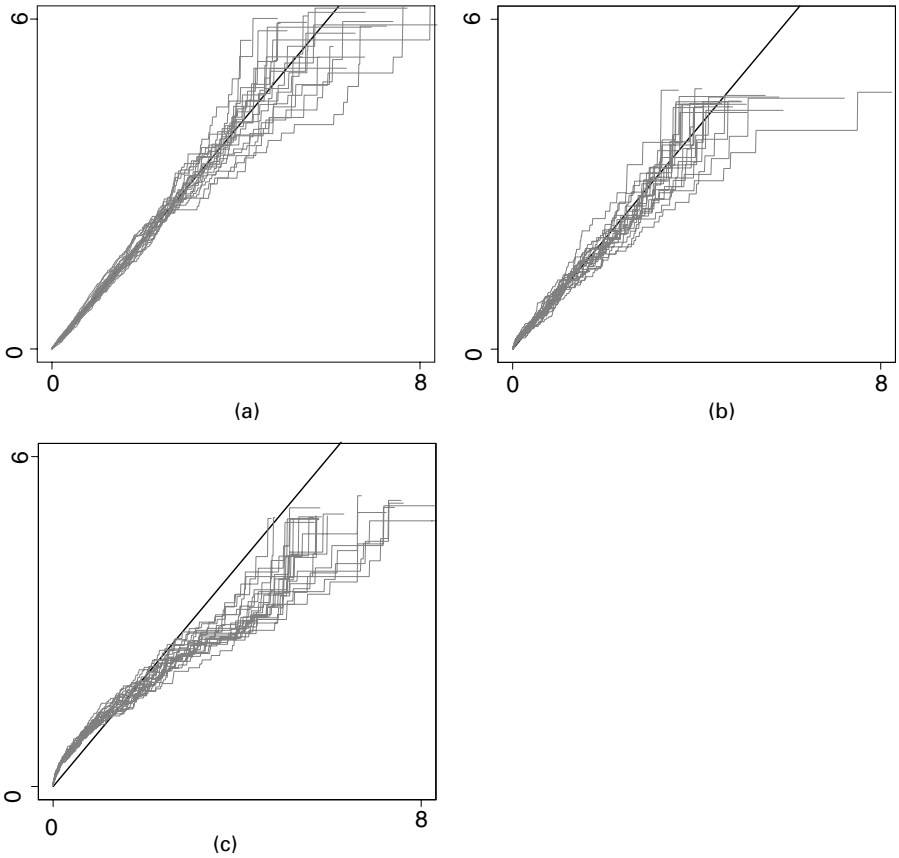

Fig. 6. Model diagnostic plots of the empirical estimate of $-\log\{P(\tilde{\varepsilon} > t)\}$ against t for 20 draws from the posterior distribution of $\tilde{E}$ defined in Section 5 (systematic deviations from the 45° line indicate inadequacy of the model): (a) model fitted to series 6 of Table 1 (daily data); (b), (c) same underlying OU process used for daily data and high frequency data respectively, with jump sizes simulated from a gamma distribution

Table 2. Posterior parameter summaries for the US dollar–Deutschmark series under the single-OU model

Parameter	*Mean*	*Median*	*Standard deviation*
ξ	5.08×10^{-2}	5.07×10^{-2}	3.49×10^{-3}
ω^2	8.55×10^{-4}	8.39×10^{-4}	1.72×10^{-4}
μ	7.1×10^{-2}	6.93×10^{-2}	1.83×10^{-2}

mean $\xi = \nu/\theta$ and the variance $\omega^2 = \nu/\theta^2$ of the volatility process, the memory parameters, μ, μ_1, μ_2 and w_2. Our results can be contrasted with those that were obtained by Barndorff-Nielsen and Shephard (2002), who fitted superposition-based OU SV models to the realized volatility time series constructed from high frequency data, based solely on second-order characteristics by using quasi-likelihood methods. Under this framework, OU and constant elasticity of variance models are indistinguishable. Moreover, Barndorff-Nielsen and Shephard (2002) made no specific assumptions about the form of the Lévy process in equation (2), and consequently about

Table 3. Posterior parameter summaries for the US dollar–Deutschmark series under the two-OU model

Parameter	*Mean*	*Median*	*Standard deviation*
ξ	5.19×10^{-2}	5.16×10^{-2}	4.55×10^{-3}
ω^2	2.43×10^{-3}	2.32×10^{-3}	6.7×10^{-4}
μ_1	4.012	3.789	1.43
μ_2	5.43×10^{-2}	5.18×10^{-2}	1.16×10^{-2}
w_2	0.467	0.469	8.27×10^{-2}

the stationary distribution of the volatility process. However, the results that were summarized in Table 3 of Barndorff-Nielsen and Shephard (2002) are in agreement with ours. There, μ_1 was estimated as 3.74 and μ_2 as 0.043; w_2 was estimated around 0.2 whereas our estimated posterior median is much larger, around 0.47. Therefore, Barndorff-Nielsen and Shephard (2002) estimated a much sharper initial drop in the volatility autocorrelation than we do.

Fig. 7 contains a collection of plots which assist in assessing the model adequacy and fit. Recall the definitions of v_i^* and y_i in equation (6). Fig. 7(a) shows y_i^2 as points and the posterior median of v_i^* under the single-OU (broken curve) and the two-OU (full curve) models, for $i = 1000, \ldots, 1200$. Fig. 7(b) shows this smoothing for the whole period of 2614 days. Fig. 7(c) applies the diagnostics of Section 5 to the jump sizes from the single-OU model. Our diagnostics, applied also to the two-OU model but not reproduced here, indicate no significant model inadequacy. Fig. 7(d) draws a kernel estimate of the log-density of the predictive distribution of $\log(v_i^*)$. It is interesting to compare this with similar estimates that were plotted in Figs 4(b) and 5(b) of Barndorff-Nielsen and Shephard (2002). See also Fig. 1 of Andersen *et al.* (2001) for some nonparametric estimates of the distribution of v_i^*. Finally, Fig. 7(e) plots the posterior median of the autocorrelation function of the series $\{v_i^*\}$ from lag 1 onwards. Barndorff-Nielsen and Shephard (2002) showed that for the single-OU model this is

$$\operatorname{corr}(v_i^*, v_{i+s}^*) = \frac{\{1-\exp(-\mu)\}^2}{2\{\exp(-\mu)-1+\mu\}} \exp\{-\mu(s-1)\}, \qquad s \geqslant 1,$$

which can be easily extended for the superposition of OU processes by using equation (11). Fig. 7(e) shows that the two-OU model results in faster initial and slower subsequent autocorrelation decay than the single-OU model.

All results were obtained by using the NCAs of Sections 3.4 and 3.5. The mixing of the algorithm for the superposition of the OU processes is not very satisfactory, and certainly much worse than when used for simulated data; we are currently trying to improve it.

7. Discussion

In this paper we have introduced two different MCMC algorithms, the CA and the NCA, which correspond to two different hierarchical parameterizations, and which sample from posterior distributions for non-Gaussian OU SV models. We have discussed the situations under which the NCAs are expected to outperform the CAs, and we have given some ideas that are specific to the context of OU models for choosing between them.

Apart from the reparameterization that is thoroughly presented in Section 3.4, Section 3.6 suggested two more NCPs for these models. The MCMC implementation of all three is very

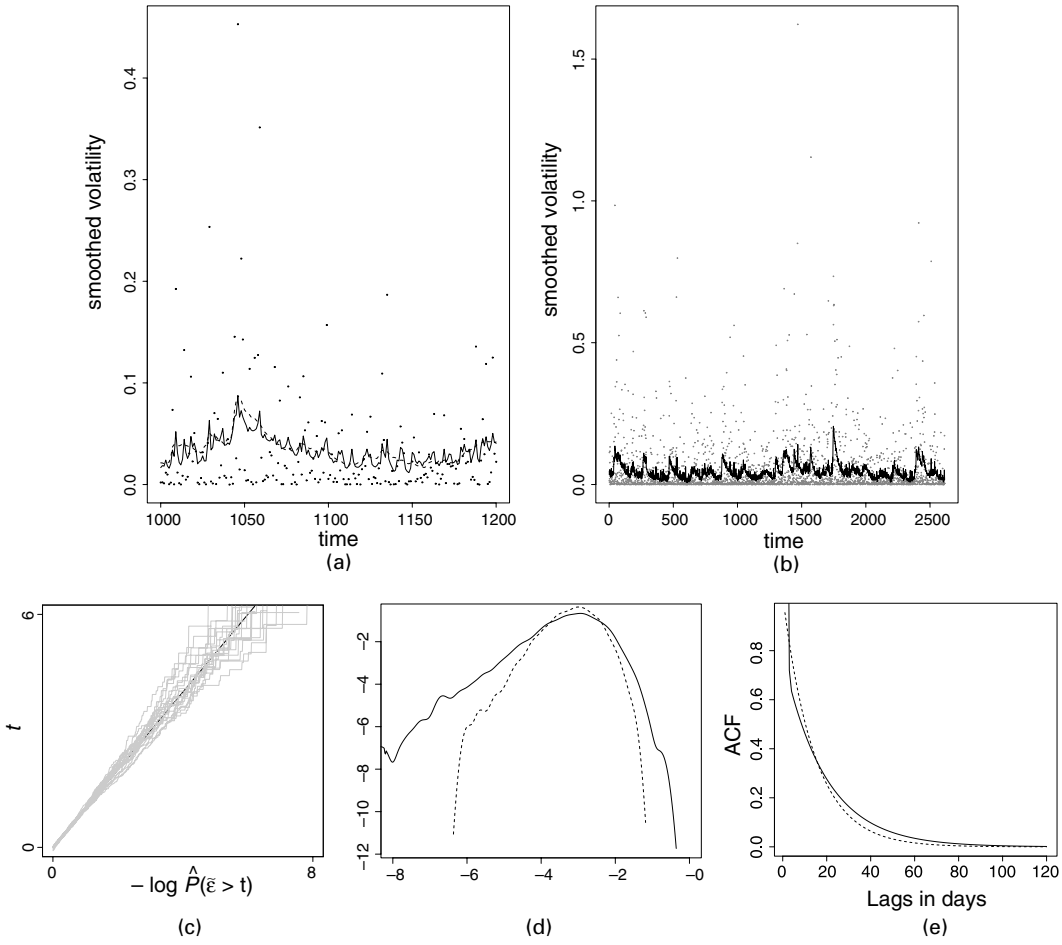

Fig. 7. Results from fitting the OU models of Sections 2.1 and 2.2 to the US dollar–Deutschmark data, described in Section 6 (·····, single-OU model; ———, two-OU model): (a) pointwise posterior median v_i^*, short period of time for the models; (b) pointwise posterior median v_i^*, long period of time for the two-OU model (●, y_i^2); (c) diagnostic of Section 5 applied to the jump sizes from the single-OU model; (d) kernel estimates of the log-predictive density of $\log(v_i^*)$; (e) posterior median of the autocorrelation function of the series $\{v_i^*\}$

similar and iterates the steps that are given in the Appendix B. Our experience suggests that, when the birth–death step is used in step 4 to update Ψ, they all have similar convergence properties, although the algorithm of Section 3.4 appears to be better, since the parameters can be updated efficiently according to the blocking scheme that is described in Appendix B. Nevertheless, we have noticed that, when multiple updates of Ψ are performed for every update of the parameters, the NCA that is based on the Ferguson–Klass representation is much more efficient than the others. This is explained by the fact that, under this scheme, when we update the parameters we do local changes in Ψ, since we either add or remove small jumps ε_j. Therefore big steps in parameter space can be achieved with this scheme. However, the mixing of Ψ is slow, since only small moves in the point process space are attempted, either when updating the parameters given $\tilde{\Psi}$ in step 1, or when updating Ψ itself in step 4. Multiple updates of Ψ are computationally extremely expensive and should be avoided. We have experimented extensively

with more sophisticated approaches for updating Ψ, without significant success. However, if more efficient methods for this updating step could be found, the NCA that is based on the Ferguson–Klass representation would become very attractive.

We have briefly provided a graphical diagnostic tool for the assessment of model fit. This method ought to be of more generic interest in complex stochastic modelling, though a more thorough investigation of its applicability to general hierarchical models is left for future research.

Our approach can be readily applied to arbitrary compound processes, where the Poisson rate is finite. Typically, inference for infinite rate processes will involve some form of approximation. For instance, the likelihood might be approximately computed by using a suitable truncation of the Lévy process (such as that suggested by the Ferguson–Klass representation: Barndorff-Nielsen and Shephard (2001); Griffin and Steel (2002)). However, all infinite rate processes are weak limits of compound Poisson processes, and therefore models from the simpler class can fit data that are generated from the more general class arbitrarily well. Therefore, we feel that the simple compound Poisson process approach is a sensible methodology even where interest lies in more general Lévy process models. Moreover, it is beneficial that the approximation itself represents a valid statistical model, rather than simply a computational device.

Although we have focused on estimation of volatility parameters, our approach can be easily adapted to consider drift terms. For instance it is straightforward to fit a more general model which allows the inclusion of a risk premium term in the driving stochastic differential equation. It is also possible to incorporate leverage terms, as suggested for example in Section 1.2 of Barndorff-Nielsen and Shephard (2001).

Acknowledgements

All the authors acknowledge the support of the training and mobility of researchers network FMRX-CT960095 on 'Spatial and computational statistics'. The second author thanks the Onassis Foundation, Lancaster University and the Engineering and Physical Sciences Research Council under grant GR/M62723/01 for support. We are indebted to Ole Barndorff-Nielsen, Neil Shephard and Sylvia Frühwirth-Schnatter for encouragement and advice. Thanks are due to colleagues in the Accounting and Finance Department at Lancaster University, particularly Stephen Taylor who facilitated access to the exchange rate data set that is used in Section 6 via JP Morgan. We also thank the Joint Editor, an Associate Editor and two referees for many instructive comments and suggestions.

Appendix A: Markov chain Monte Carlo algorithm of Section 3.2

A.1. The centred algorithm

Step 1: update (ν, θ) according to $\pi\{\nu, \theta|\Psi, v(0), \mu\}$.
Step 2: update μ according to $\pi\{\mu|X, \Psi, v(0), \nu, \theta\}$.
Step 3: update $v(0)$ according to $\pi\{v(0)|X, \Psi, \nu, \mu, \theta\}$.
Step 4: update Ψ according to $\pi\{\Psi|X, v(0), \nu, \mu, \theta\}$.
Step 5: go to step 1.

The densities in steps 1–4 are derived up to proportionality from equation (18). Direct Gibbs steps are not available for steps 2 and 3 but we can easily perform a random-walk Metropolis–Hastings updating scheme on the logarithmic scale for the parameters that are involved.

Step 1 exploits the fact that the full conditional distribution of θ is known,

$$\theta|\cdot \sim \mathrm{Ga}\Big\{\Psi(S) + \nu + \alpha_\theta,\ \sum_{j=1}^{\Psi(S)} \varepsilon_j + v(0) + \beta_\theta\Big\}, \tag{24}$$

and consequently we can derive

$$\pi\{\nu|v(0),\Psi,\mu\}\propto\frac{\Gamma\{\nu+\alpha_\theta+\Psi(S)\}}{\Gamma(\nu)}\left\{\frac{v(0)}{\beta_\theta+v(0)+\sum_{j=1}^{\Psi(S)}\varepsilon_j}\right\}^{\nu}\nu^{\Psi(S)+\alpha_\nu-1}\exp\{-(\beta_\nu+\mu T)\nu\}. \quad (25)$$

Thus, we first update ν by using a Metropolis–Hastings step with target (25), and then, conditionally on ν, we simulate θ directly from distribution (24).

Updating the conditional distribution of Ψ in step 4 is more involved since it requires a Metropolis–Hastings step which operates on a point process space. We use the general methodology that was presented in Geyer and Møller (1994) together with some ideas that are specific to our context. At each iteration, we randomly choose between two types of move. The first is a dimension-changing move that proposes either to add or to remove a point from the current configuration of the point process. The second selects one of the existing points at random and attempts to displace it. We describe these steps in detail below. To improve the mixing, at each iteration we simulate from the conditional distribution of the jump sizes given the jump times, using a blocked Metropolis–Hastings step where the variance of the proposal distribution is inversely proportional to the current number of jumps.

Formally, let the current configuration of points be $\psi=\{(c_1,\varepsilon_1),\ldots,(c_m,\varepsilon_m)\}$.

A.2. Birth-and-death move

With probability q we propose to move to the configuration $\psi\cup\{(c,\varepsilon)\}$ by generating c uniformly in $[0,T]$ and ε from an $\mathrm{Ex}(\theta)$ distribution. With probability $1-q$ we propose to move to the configuration $\psi-\{(c,\varepsilon)\}$ by choosing to remove $\{(c,\varepsilon)\}$ uniformly among the existing points of ψ. Then, the Metropolis–Hastings acceptance ratio for the birth-or-death move is

$$\alpha[\psi,\psi\cup\{(c,\varepsilon)\}]=\min\{1,r\{\psi,(c,\varepsilon)\}\}$$

for a birth move and

$$\alpha[\psi\cup\{(c,\varepsilon)\},\psi]=\min\{1,1/r\{\psi,(c,\varepsilon)\}\}$$

for a death move, where

$$\begin{aligned}r\{\psi,(c,\varepsilon)\}&=\frac{f[X|\mu,v(0),\psi\cup\{(c,\varepsilon)\}]\,\pi[\psi\cup\{(c,\varepsilon)\}|\lambda,\theta]}{f\{X|\mu,v(0),\psi\}\,\pi(\psi|\lambda,\theta)}\frac{1-q}{q}\frac{(m+1)^{-1}}{b(c,\varepsilon)}\\&=\frac{f[X|\mu,v(0),\psi\cup\{(c,\varepsilon)\}]}{f\{X|\mu,v(0),\psi\}}\frac{\lambda T}{m+1}\frac{1-q}{q}.\end{aligned}$$

A.3. Displacement move

We construct the displacement transition kernel as a mixture of m Metropolis–Hastings transition kernels; the ith kernel is reversible with respect to the conditional posterior distribution of the ith point $\{(c_i,\varepsilon_i)\}$ given $\psi-\{(c_i,\varepsilon_i)\}$. This can be seen as a distribution on S with Lebesgue density

$$\pi[c,\varepsilon|X,\lambda,\theta,\mu,v(0),\psi-\{(c_i,\varepsilon_i)\}]\propto f[X|\mu,v(0),\psi-\{(c_i,\varepsilon_i)\}\cup\{(c,\varepsilon)\}]\theta\exp(-\theta\varepsilon).$$

Each kernel is chosen with equal probability and it is adequate to describe how the ith kernel is constructed. We propose to move from ψ to $\psi-\{(c_i,\varepsilon_i)\}\cup\{(c,\varepsilon)\}$ and we suggest two strategies to generate $\{(c,\varepsilon)\}$.

The first strategy uses independence sampling from the proposal density

$$q(c,\varepsilon)=T^{-1}\theta\,\exp(-\theta\varepsilon),\qquad\varepsilon>0,\quad 0<c<T.$$

The calculation of the acceptance ratio is straightforward, but we shall not use this strategy in any of our examples.

The second is a strategy that achieves local change of the volatility process (see Fig. 8), and it has been used in our MCMC programs throughout the paper. Assume, without loss of generality, that $c_1<c_2<\ldots<c_m$. We generate c uniformly in $[c_{i-1},c_{i+1}]$ with $c_0=0$ and $c_{m+1}=T$, and we define the transformation $(\varepsilon_i,c_i,c)\to(\varepsilon,c,c_i)=(\varepsilon_i\exp\{-\mu(c-c_i)\},c,c_i)$. This transformation is invertible and its Jacobian is $\exp\{\mu(c-c_i)\}$. Thus, the resulting Metropolis–Hastings acceptance ratio (which can be derived by using

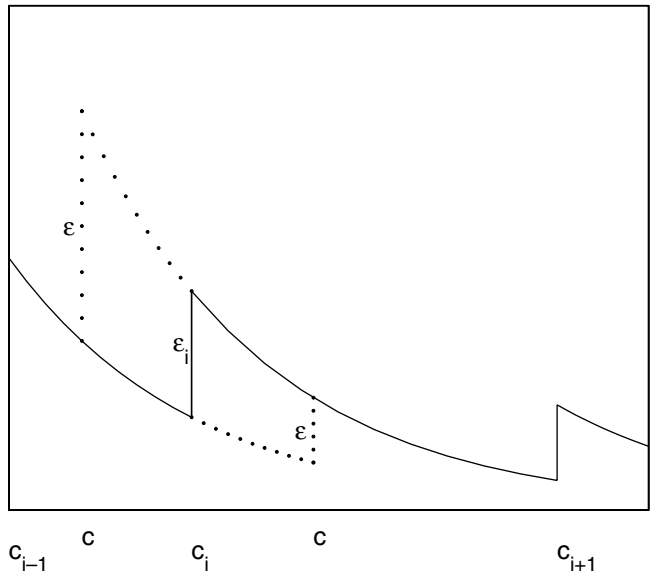

Fig. 8. Local displacement of the point $\{(c_j, \varepsilon_j)\}$: the new jump time c is generated uniformly in (c_{j-1}, c_{j+1}) and the new jump size ε is set so that the volatility process changes only between time points c and c_j; the new jump time c can lie on either side of c_j (both options are shown)

results from Tierney (1998)) is

$$\min\left(1, \frac{f[X|\mu, v(0), \psi - \{(c_i, \varepsilon_i)\} \cup \{(c, \varepsilon)\}]}{f\{X|\mu, v(0), \psi\}} \exp\{\theta(\varepsilon_i - \varepsilon) - \mu(c - c_i)\}\right).$$

Apart from the well-documented advantages of local moves in reversible-jump-type algorithms (see for example section 4 in Dellaportas *et al.* (2002)), this move gives increased computational efficiency since it requires the evaluation of only a small part of the likelihood function (corresponding to the time between the current and proposed jump times).

Appendix B: Markov chain Monte Carlo algorithm of Section 3.4

B.1. The non-centred algorithm

Step 1: update (ν, θ, μ) according to $\pi\{\nu, \theta, \mu | X, \tilde{\Psi}, \tilde{v}(0)\}$.
Step 2: transform $v(0) = \tilde{v}(0)/\theta$ and $(\lambda, \theta, \tilde{\Psi}) \to \Psi$.
Step 3: update $v(0)$ according to $\pi\{v(0)|X, \Psi, \nu, \mu, \theta\}$.
Step 4: update Ψ according to $\pi\{\Psi|X, v(0), \nu, \mu, \theta\}$.
Step 5: transform $\tilde{v}(0) = \theta\, v(0)$ and stochastically transform $(\lambda, \theta, \Psi) \to \tilde{\Psi}$.
Step 6: go to step 1.

Steps 3 and 4 are exactly the same as the corresponding steps of the CA. The reason for transforming Ψ to $\tilde{\Psi}$ is to reduce the dependence between missing data and parameters. However, it is much more natural to construct Metropolis–Hastings steps for Ψ, which has an actual physical interpretation. This is why we prefer to update $\tilde{\Psi}$ by using the updates of Ψ. Moreover, this set-up makes the CA and the NCA more comparable. Note that, if direct simulation from the conditional in step 4 were feasible, then the last two steps of the NCA would produce a sample from the conditional distribution $\tilde{\Psi}|X, v(0), \nu, \theta, \mu$.

The NCA differs from the CA in step 1, where the parameters are updated conditionally on the missing data, and in the two transformation steps 2 and 5. Step 2 is totally deterministic, $\tilde{v}(0) \to v(0)$ is trivial and

$(\lambda, \theta, \tilde{\Psi}) \to \Psi$ is described in Section 3.4 and is illustrated in Fig. 3. The transformation $(\lambda, \theta, \Psi) \to \tilde{\Psi}$ in step 5 is stochastic; however, it is not necessary, as the next section illustrates.

B.2. Step 1: updating the parameters

We first establish the conditional joint density of ν and μ given the observed and missing data, but marginalized with respect to θ. Note that $v^*(t_{i-1}, t_i) = g_i/\theta$ with

$$g_i = \frac{\sum_{t_{i-1}}^{t_i} \tilde{\varepsilon}_j}{\mu} - \frac{1}{\mu}\left[\{\exp(-\mu t_i) - \exp(-\mu t_{i-1})\}\{\tilde{v}(0) + \sum_{0}^{t_{i-1}} \exp(\mu c_j)\tilde{\varepsilon}_j\} + \sum_{t_{i-1}}^{t_i} \exp\{-\mu(t_i - c_j)\}\tilde{\varepsilon}_j\right],$$

where $\{(c_j, \tilde{\varepsilon}_j)\}$ are the resulting points from steps 1 and 2 of the transformation that is described in Section 3.3, and the sums Σ_s^t, with $s \leqslant t$, are interpreted as sums over all (if any) $s \leqslant c_j < t$. Hence, g_i is known given $\tilde{v}(0)$, $\tilde{\Psi}$, ν and μ. Therefore

$$f\{X|\nu, \mu, \theta, \tilde{\Psi}, \tilde{v}(0)\} = d\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0)\}\theta^{n/2} \exp[-\theta k\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0), X\}],$$

where

$$\begin{aligned} k\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0), X\} &= \sum_{i=1}^{n} \frac{\{x(t_i) - x(t_{i-1})\}^2}{2g_i}, \\ d\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0)\} &= \prod_{i=1}^{n} g_i^{-1/2}. \end{aligned} \tag{26}$$

This implies that by choosing a $\mathrm{Ga}(\alpha_\theta, \beta_\theta)$ prior on θ its full conditional posterior distribution is a $\mathrm{Ga}[n/2 + \alpha_\theta, k\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0), X\} + \beta_\theta]$ distribution. Furthermore, by integrating θ out we deduce that

$$\pi\{\nu, \mu|X, \tilde{\Psi}, \tilde{v}(0)\} \propto \frac{\tilde{v}(0)^\nu}{\Gamma(\nu)} d\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0)\}[k\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0), X\} + \beta_\theta]^{-(n/2+\alpha_\theta)} \pi(\nu)\, \pi(\mu).$$

We shall update (ν, θ, μ) by first updating (ν, μ) using a Metropolis–Hastings step on the logarithmic scale with target density $\pi\{\nu, \mu|X, \tilde{\Psi}, \tilde{v}(0)\}$, and then by simulating θ directly from its full conditional. When $\tilde{\Psi}$ is available, updating the parameters is quite straightforward. When new values for ν and μ have been proposed, we compute the proposed value of λ. Then, we perform the first two steps that are described in the transformation of Section 3.4, calculate the proposed values for k and d in expression (26) and compute the value of the target density for the proposed values.

It remains to be clarified how to transform $(\lambda, \theta, \Psi) \to \tilde{\Psi}$. For a given value of λ, $\tilde{\Psi}$ can be written as the union of two point processes, $\tilde{\Psi} = \tilde{\Psi}_{(0,\lambda)} \cup \tilde{\Psi}_{[\lambda,\infty)}$, where $\tilde{\Psi}_{(0,\lambda)}$ contains all points of $\tilde{\Psi}$ with $0 < m_j < \lambda$. $\tilde{\Psi}_{(0,\lambda)}$ and $\tilde{\Psi}_{[\lambda,\infty)}$ are *a posteriori* independent, since they are *a priori* independent, as restrictions of a Poisson process on disjoint sets (Kingman, 1993), and $\tilde{\Psi}_{[\lambda,\infty)}$ is independent of the data. Therefore, we simulate from the full conditional posterior distribution of $\tilde{\Psi}$ by first simulating $\tilde{\Psi}_{(0,\lambda)}$ and $\tilde{\Psi}_{[\lambda,\infty)}$ and then taking their union.

$\tilde{\Psi}_{[\lambda,\infty)}$ is a marked Poisson process on $[0, T] \times (\lambda, \infty) \times (0, \infty)$ with mean measure (20). Conditionally on being less than λ, the marks m_j are independent of each other and of the points $\{(c_j, \tilde{\varepsilon}_j)\}$, distributed as $\mathrm{Un}[0, \lambda]$. Therefore, conditionally on Ψ, $\tilde{\Psi}_{(0,\lambda)}$ is the set containing the points $\{(c_j, m_j, \tilde{\varepsilon}_j)\}$, where $\tilde{\varepsilon}_j = \theta\varepsilon_j$, $(c_j, \varepsilon_j) \in \Psi$ and the m_js are simulated independently as $\mathrm{Un}[0, \lambda]$. This completes the random transformation of step 5 of the NCA.

Note, however, that this transformation is not necessary and can be incorporated in step 1, when updating (ν, μ) by using a Metropolis–Hastings step. Suppose that at this step we have only Ψ; (ν_0, μ_0) and (ν_1, μ_1) denote the current and the proposed state of (ν, μ) respectively, and $\lambda_0 = \nu_0\mu_0$ and $\lambda_1 = \nu_1\mu_1$. It can be easily checked that the algorithm that is described below is equivalent to first randomly transforming $(\lambda, \theta, \Psi) \to \tilde{\Psi}$ and then updating (ν, μ) given $\tilde{\Psi}$. However, it circumvents the problem of storing $\tilde{\Psi}$, which has an infinite number of points on $[0, T] \times [\lambda, \infty) \times (0, \infty)$ for every λ.

If $\lambda_1 < \lambda_0$ then
  delete each point $(c_j, \varepsilon_j) \in \Psi$ with probability $1 - \lambda_1/\lambda_0$;
  rescale $\tilde{\varepsilon}_j = \theta\varepsilon_j$ for each accepted point.
Otherwise

generate a Poisson process on S with mean measure $(\lambda_1 - \lambda_0)\theta \exp(-\theta\varepsilon)\, \mathrm{d}c\, \mathrm{d}\varepsilon$;
take the union of this process and Ψ;
rescale $\tilde{\varepsilon}_j = \theta\varepsilon_j$ for each point (c_j, ε_j) in the union.

Use the points of the resulting random set to compute k and d in expression (26).

Appendix C: Markov chain Monte Carlo algorithm of Section 3.5

We use a single-component updating Metropolis–Hastings algorithm which is largely based on the MCMC algorithm that is described in Appendix B. Note that

$$\theta|\cdot \sim \mathrm{Ga}[n/2 + \alpha_\theta, k\{\nu, \mu, \tilde{\Psi}, \tilde{v}(0), X\} + \beta_\theta],$$

where k is computed as in expression (26). We update (ν_1, ν_2, θ) in one block, and the rest of the parameters are updated individually.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2001) The distribution of realised exchange rate volatility. *J. Am. Statist. Ass.*, **96**, 42–55.

Barndorff-Nielsen, O. E. and Shephard, N. (2001) Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). *J. R. Statist. Soc.* B, **63**, 167–241.

Barndorff-Nielsen, O. E. and Shephard, N. (2002) Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Statist. Soc.* B, **64**, 253–280.

Black, F. (1976) Studies of stock price volatility changes. *Proc. Bus. Econ. Statist. Sect. Am. Statist. Ass.*, 177–181.

Cox, D. R. and Isham, V. (1980) *Point Processes*. New York: Chapman and Hall.

Dellaportas, P., Forster, J. J. and Ntzoufras, I. (2002) On Bayesian model and variable selection using MCMC. *Statist. Comput.*, **12**, 27–36.

Diebolt, J. and Robert, C. P. (1994) Estimation of finite mixture distributions through Bayesian sampling. *J. R. Statist. Soc.* B, **56**, 363–375.

Ding, Z. and Granger, C. W. J. (1996) Modelling volatility persistence of speculative returns: a new approach. *J. Econometr.*, **73**, 185–215.

Ferguson, T. S. and Klass, M. J. (1972) A representation of independent increment processes without Gaussian components. *Ann. Math. Statist.*, **43**, 1634–1643.

Gelfand, A. E., Sahu, S. K. and Carlin, B. P. (1995) Efficient parametrization for normal linear mixed models. *Biometrika*, **82**, 479–488.

Geyer, C. J. (1992) Practical Markov chain Monte Carlo (with discussion). *Statist. Sci.*, **7**, 473–483.

Geyer, C. J. and Møller, J. (1994) Simulation procedures and likelihood inference for spatial point processes. *Scand. J. Statist.*, **21**, 359–373.

Ghysels, E., Harvey, A. C. and Renault, E. (1996) Stochastic volatility. In *Statistical Methods in Finance* (eds C. Rao and G. Maddala). Amsterdam: North-Holland.

Griffin, J. E. and Steel, M. F. J. (2002) Inference with non-Gaussian Ornstein–Uhlenbeck processes for stochastic volatility. *Mimeo*. Institute of Mathematics and Statistics, University of Kent at Canterbury, Canterbury.

Harvey, A. C., Ruiz, E. and Shephard, N. (1994) Multivariate stochastic variance models. *Rev. Econ. Stud.*, **61**, 247–264.

Kim, S., Shephard, N. and Chib, S. (1998) Stochastic Volatility: likelihood inference and comparison with ARCH models. *Rev. Econ. Stud.*, **65**, 361–393.

Kingman, J. F. C. (1993) *Poisson Processes*. Oxford: Clarendon.

Liu, J. S., Wong, W. H. and Kong, A. (1994) Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, **81**, 27–40.

Meng, X. (2000) Missing data: dial M for ??? *J. Am. Statist. Ass.*, **95**, 1325–1330.

Papaspiliopoulos, O. (2001) Discussion on 'Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics' (by O. E. Barndorff-Nielsen and N. Shephard). *J. R. Statist. Soc.* B, **63**, 211–213.

Papaspiliopoulos, O. (2003) Non-centered parameterisations for hierarchical models and data augmentation. *PhD Dissertation*. Department of Mathematics and Statistics, Lancaster University, Lancaster.

Papaspiliopoulos, O., Roberts, G. O. and Sköld, M. (2003) Non-centered parameterisations for hierarchical models and data augmentation. In *Bayesian Statistics 7* (eds J. Bernardo, M. Bayarri, J. Berger, A. Dawid, D. Heckerman, A. Smith and M. West), pp. 307–327. Oxford: Oxford University Press.

Pitt, M. K. and Shephard, N. (1999) Analytic convergence rates and parameterisation issues for the Gibbs sampler applied to state space models. *J. Time Ser. Anal.*, **20**, 63–85.

Richardson, S. and Green, P. J. (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion). *J. R. Statist. Soc.* B, **59**, 731–792; correction, **60** (1998), 661.

Robert, C. P. (1996) Mixtures of distributions: inference and estimation. In *Markov chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson, and D. J. Spiegelhalter), pp. 46–57. London: Chapman and Hall.
Roberts, G. O. (2001) Discussion on 'Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics' (by O. E. Barndorff-Nielsen and N. Shephard). *J. R. Statist. Soc.* B, **63**, 209–211.
Roberts, G. O. (2003) Linking theory and practice of MCMC (with discussion). In *Highly Structured Stochastic Systems* (eds P. Green, N. Hjørt and S. Richardson), pp. 145–171. Oxford: Oxford University Press.
Roberts, G. O. and Papaspiliopoulos, O. (2003) Parameterisation, robustness and MCMC in Bayesian hierarchical models. To be published.
Roberts, G. O. and Sahu, S. K. (1997) Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. R. Statist. Soc.* B, **59**, 291–317.
Roeder, K. and Wasserman, L. (1997) Practical Bayesian density estimation using mixtures of normals. *J. Am. Statist. Ass.*, **92**, 894–902.
Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions*. New York: Cambridge University Press.
Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields* (eds D. Cox, O. Barndorff-Nielsen and D. Hinkley), pp. 1–67. London: Chapman and Hall.
Smith, A. F. M. and Roberts, G. O. (1993) Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods. *J. R. Statist. Soc.* B, **55**, 3–23.
Tanner, M. A. and Wong, W. H. (1987) The calculation of posterior distributions by data augmentation. *J. Am. Statist. Ass.*, **82**, 528–540.
Tierney, L. (1998) A note on Metropolis–Hastings kernels for general state-spaces. *Ann. Appl. Probab.*, **8**, 1–9.
Wolpert, R. and Ickstadt, K. (1998) Poisson/gamma random field models for spatial statistics. *Biometrika*, **85**, 251–267.