Bayesian inference with stochastic volatility models using continuous superpositions of non-Gaussian Ornstein-Uhlenbeck processes

Griffin, Jim and Steel, Mark F.J.

University of Warwick, Statistics Department

13 October 2008

Bayesian inference with stochastic volatility models using continuous superpositions of non-Gaussian Ornstein-Uhlenbeck processes

J.E. Griffin*
Institute of Mathematics, Statistics and Actuarial Science,
University of Kent

M. F. J. Steel
Department of Statistics, University of Warwick

Abstract

This paper discusses Bayesian inference for stochastic volatility models based on continuous superpositions of Ornstein-Uhlenbeck processes. These processes represent an alternative to the previously considered discrete superpositions. An interesting class of continuous superpositions is defined by a Gamma mixing distribution which can define long memory processes. We develop efficient Markov chain Monte Carlo methods which allow the estimation of such models with leverage effects. This model is compared with a two-component superposition on the daily Standard and Poor's 500 index from 1980 to 2000.

*Corresponding author: Jim Griffin, Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K. Tel.: +44-1227-823627; Fax: +44-1227-827932; Email: J.E.Griffin-28@kent.ac.uk.

1 Introduction

Continuous-time stochastic volatility models have been shown to be successful in modelling the behaviour of financial time series, such as stock prices and exchange rates. They also have nice properties. Volatility can be consistently modelled at different observational frequencies and option pricing formulae can sometimes be derived. Such models have a separate stochastic process driving the instantaneous latent volatility of the observables. Let us assume we are modelling the log of an asset price or index $y(t)$, where t indicates time. A common modelling specification is through the stochastic differential equation

$$d\,y(t) = \{\mu + \beta\,\sigma^2(t)\}\,dt + \sigma(t)\,dB(t), \tag{1}$$

where $\sigma^2(t)$ is the instantaneous volatility, independently distributed from $B(t)$, which is Brownian motion, and μ and β are drift and risk premium parameters, respectively. Aggregate returns over a time interval of length Δ, say, defined as $y_n = \int_{(n-1)\Delta}^{n\Delta} \mathrm{d}y(t)$ and observed at times $n = 1, \ldots, N$ are then Normally distributed as

$$y_n \sim \mathrm{N}(\mu\Delta + \beta\sigma_n^2, \sigma_n^2),$$

given the discrete or "actual" volatility $\sigma_n^2 = \int_{(n-1)\Delta}^{n\Delta} \sigma^2(u)\mathrm{d}u$. Most models in the literature assume a stochastic process for the instantaneous volatility based on Brownian motion. Following Barndorff-Nielsen and Shephard (2001) we consider, instead, a non-Gaussian Ornstein-Uhlenbeck volatility process. One advantage of the latter is that it often facilitates analytic option pricing (see Nicolato and Venardos, 2003).

Other studies that have considered Bayesian inference with similar volatility processes are Roberts *et al.* (2004), Griffin and Steel (2006), Gander and Stephens (2007a,b) and Frühwirth-Schnatter and Söger (2008).

In this paper, we model the instantaneous volatility with a continuous superposition of Ornstein-Uhlenbeck processes. Using, for example, a Gamma mixing distribution this can lead to long memory properties. Alternatively, long memory can be introduced directly into the asset price equation (1), as in Gander and Stephens (2007a). In addition, we propose efficient inference methods for asset models with risk premium and leverage effects, using such continuous superpositions as volatility processes. Standard Markov chain Monte Carlo (MCMC) methods usually perform poorly for this class of mod-

els and specific techniques need to be developed. We use an easily controlled approximation to the volatility process to reduce the computational cost, and we implement the retrospective method of Roberts *et al.* (2004) as well as a modification of the dependent thinning sampler of Griffin and Steel (2006) for inference with these models.

The paper is organised as follows: Section 2 describes the specification of the Ornstein-Uhlenbeck volatility processes, and Section 3 defines continuous superpositions of such processes. An asset price model extended to include leverage effects is briefly described in Section 4, and the following section comments on the priors used. Inference methods based on Markov chain Monte Carlo samplers are described in some detail in Section 6. An application to the S&P 500 stock price index is provided in Section 7. Finally, the last section contains some concluding remarks.

2 Ornstein-Uhlenbeck processes

The non-Gaussian Ornstein-Uhlenbeck (OU) process (Barndorff-Nielsen and Shephard, 2001) for modelling stochastic volatility is defined by the stochastic differential equation

$$d\sigma^2(t) = -\lambda\sigma^2(t) + dz(\lambda t).$$

where $\lambda > 0$ is a scalar decay parameter and $z(t)$ is a non-decreasing Lévy process (*i.e.* a subordinator), called the background driving Lévy process or BDLP. The equation has the strong solution

$$\sigma^2(t) = \exp\{-\lambda t\}\sigma^2(0) + \int_0^t \exp\{-\lambda(t-s)\}dz(\lambda s).$$

This form implies that the autocorrelation function of $\sigma^2(t)$ will decay exponentially with rate λ. The appearance of λ in the timing of the Lévy processes implies that the marginal distribution of $\sigma^2(t)$ does not depend on λ. There is also a simple functional relationship between the Lévy density of $\sigma^2(t)$ (which will be called u) and the Lévy density of $z(t)$ (which will be called w).

$$w(x) = -u(x) - xu'(x).$$

This allows us to define the marginal distribution of $\sigma^2(t)$ and to derive the appropriate BDLP. The tail mass integral $W^+(x) = \int_x^\infty w(y)\,dy$ and its inverse

W^{-1} are important functions for the simulation of Lévy processes (see *e.g.* Ferguson and Klass, 1972, Rosinski, 2001). There is also a functional relationship between W^+ and u

$$W^+(x) = xu(x).$$

This allows us to specify z by choosing the marginal distribution of $\sigma^2(t)$ and deriving the Lévy density of z. Several classes of marginal distribution for $\sigma^2(t)$ have been consider in the literature (see Gander and Stephens, 2007a,b). A practically interesting example is a Gamma distribution, *i.e.* $\sigma^2(t) \sim \text{Ga}(\nu, \gamma)$ with probability density function

$$p(x) = \frac{\gamma^\nu}{\Gamma(\nu)} x^{\nu-1} \exp\{-\gamma x\}$$

which has Lévy density

$$u(x) = \nu x^{-1} \exp\{-\gamma x\}$$

and so $w(x) = \gamma\nu \exp\{-\gamma x\}$ and $W^+(x) = \nu \exp\{-\gamma x\}$. The Lévy density has finite integral which implies that the BDLP is a compound Poisson process and the resulting Gamma OU process can be represented as a shot-noise process (Bondesson, 1988) in the following way

$$\sigma^2(t) = \sum_{i=1}^{\infty} \text{I}(\tau_i < t) J_i \exp\{-\lambda(t - \tau_i)\} \tag{2}$$

where (τ, J) follow a Poisson process with intensity $\nu\lambda\gamma \exp\{-\gamma J\}$. We can interpret J as jump sizes and τ as jump times. For many other choices of the marginal process, the BDLP will be a Lévy process with infinite rather than finite activity. For infinite activity processes we can use a series representation of the Lévy process with jumps truncated to be above a small value ϵ (see *e.g.* Gander and Stephens, 2007a, for more details) and the intensity of the resulting Poisson process will be $\lambda \text{I}(J > \epsilon)\, w(J)$.

Barndorff-Nielsen and Shephard (2001) comment that actual financial time series will often be fit better by combining independent OU processes with different rate parameters λ. For example, a two-component model is given by

$$\sigma^2(t) = \sigma_1^2(t) + \sigma_2^2(t),$$

where each component process is an independent OU process as defined above with a $\text{Ga}(\nu_j, \gamma)$ distribution and rate parameter $\lambda_j, j = 1, 2$. Then we retain a

Gamma marginal distribution for $\sigma^2(t)$ but the autocorrelation function takes the more general form

$$\rho(s) = w \exp(-\lambda_1|s|)) + (1-w)\exp(-\lambda_2|s|),$$

with the weight $w = \nu_1/\{\nu_1+\nu_2\}$, the ratio of the component variances. The use of such superpositions was shown to behave better than single component OU processes in Barndorff-Nielsen and Shephard (2001) and Griffin and Steel (2006).

In this paper we will focus on OU processes with Gamma marginals. For S&P 500 data (as used here) Gander and Stephens (2007a) provide some evidence that the choice of a Gamma marginal is sensible, but extensions to OU processes with other marginal distributions can be considered (for example, using the methods of Gander and Stephens, 2007a).

3 Continuous superpositions of OU processes

Let $\sigma^2_\lambda(t)$ be a non-Gaussian OU process with decay parameter λ. The continuous superposition model assumes that the instantaneous volatility process is given by

$$\sigma^2(t) = \int \sigma^2_\lambda(t)\, dF(\lambda),$$

where F is a distribution function, which will be termed the mixing distribution. The properties of this model are studied by Barndorff-Nielsen (2001). The autocorrelation function of the superposed process is

$$\rho(s) = \int \exp\{-\lambda|s|\}\, dF(\lambda).$$

Different choices of F lead to different shapes of autocorrelation function. As a special case, we can find the two-component model of the previous section by simply using a two-point discrete distribution for F. Generally, the form of the autocorrelation function is given by the moment generating function of F which provides a simple way to find the autocorrelation function for any chosen distribution and also for finding the F that leads to a particular autocorrelation function. An interesting choice, suggested in Barndorff-Nielsen (2001), is the Gamma distribution with shape α and scale ϕ which leads to an autocorrelation function of the form

$$\rho(s) = \left(1 + \frac{|s|}{\phi}\right)^{-\alpha},$$

corresponding to long memory with Hurst coefficient $1 - \frac{\alpha}{2}$ if $\alpha < 1$.

We can use the results of Barndorff-Nielsen (2001) to derive a useful representation of the process which extends the shot-noise representation of the OU process given in equation (2). If F has a density f, the continuous superposition model can be represented as

$$\sigma^2(t) = \sum_{i=1}^{\infty} \mathrm{I}(\tau_i < t) J_i \exp\{-\lambda_i(t - \tau_i)\}, \tag{3}$$

where (τ, J, λ) follow a Poisson process with intensity $\nu\lambda f(\lambda)\gamma \exp\{-\gamma J\}$. The model differs from (2) by introducting a jump-specific decay parameter (λ_i), which allows for extra flexibility. In statistical terms, we have moved from the decay parameter being a fixed effect to a random effect. The integrated volatility, $\sigma^{2\star}(t) = \int_0^t \sigma^2(u)\, du$ has the form

$$\begin{aligned}\sigma^{2\star}(t) = &\sum_{i=1}^{\infty} \mathrm{I}(0 < \tau_i < t) \frac{J_i}{\lambda_i} \left(1 - \exp\{-\lambda_i(t - \tau_i)\}\right) \\ &+ \sum_{i=1}^{\infty} \mathrm{I}(\tau_i < 0) \frac{J_i}{\lambda_i} \exp\{\lambda_i \tau_i\} \left(1 - \exp\{-\lambda_i t\}\right),\end{aligned}$$

and the increment (the actual volatility as in Section 1) $\sigma_n^2 = \sigma^{2\star}(n\Delta) - \sigma^{2\star}((n-1)\Delta)$ has the form

$$\begin{aligned}\sigma_n^2 = &\sum_{i=1}^{\infty} \mathrm{I}((n-1)\Delta < \tau_i < n\Delta) \frac{J_i}{\lambda_i} \left(1 - \exp\{-\lambda_i(n\Delta - \tau_i)\}\right) \\ &+ \sum_{i=1}^{\infty} \mathrm{I}(\tau_i < (n-1)\Delta) \frac{J_i}{\lambda_i} \left(1 - \exp\{-\lambda_i \Delta\}\right) \exp\{-\lambda_i((n-1)\Delta - \tau_i)\}.\end{aligned} \tag{4}$$

A single OU process $\sigma^2(t)$ has an integrated volality which can be expressed as $A\sigma^2(0) + B_t$ where A is a constant depending on the decay parameter λ and B_t is the integrated process restricted to jumps that occur in the region $(0, t)$. Therefore, calculation of σ_n^2 only depends on jumps before time 0 through $\sigma^2(0)$. This is not true for the continuous superpositions. The contribution of jumps before time 0 is given by the second sum in equation (4). The number of terms is infinite even if the BDLP is a finite activity Lévy process (as in the Gamma case) and so some truncation of this sum will be needed. We will

use the approximation

$$\sigma_n^2 = \sum_{i=1}^{\infty} \mathrm{I}((n-1)\Delta < \tau_i < n\Delta) \frac{J_i}{\lambda_i} \left(1 - \exp\{-\lambda_i(n\Delta - \tau_i)\}\right)$$
$$+ \sum_{i=1}^{\infty} \mathrm{I}(B < \tau_i < (n-1)\Delta) \frac{J_i}{\lambda_i} \left(1 - \exp\{-\lambda_i \Delta\}\right) \exp\{-\lambda_i((n-1)\Delta - \tau_i)\}.$$

where B is chosen to be smaller than zero and such that we avoid a noticeable truncation error in the actual volatility. The same effect was noted by Gander and Stephens (2007b) for their models, who explain why the form of the OU process leads to a simplified form of the integrated volatility.

To be specific, we will focus on Gamma mixing in the rest of the paper, but the ideas in the sampler are essentially unchanged if we use a different mixing distribution.

4 Modelling leverage effects

Leverage effects refer to the negative correlation between returns and volatility, which was first considered by Black (1976). Models with a return process and volatility process driven by Brownian motions usually include leverage by introducing correlation between these processes. An analogous approach for volatility models driven by a non-Gaussian Lévy process was introduced by Barndorff-Nielsen and Shephard (2001) who suggest changing the drift in equation (1) so that the asset price follows

$$d\,y(t) = \{\mu + \beta\,\sigma^2(t) + \rho(z(t) - \mathrm{E}[z(t)])\}\,dt + \sigma(t)\,dB(t),$$

which implies that the returns $y_1, \ldots, y_N$ are modelled by

$$y_n \sim \mathrm{N}\left(\mu\Delta + \beta\sigma_n^2 + \rho(z_n - \mathrm{E}[z_n]), \sigma_n^2\right),$$

where σ_n^2 is the actual volatility as defined in Section 1 and $z_n = \int_{(n-1)\Delta}^{n\Delta} z(s)\,ds$. Negative values for the leverage coefficient ρ will associate drops in returns with jumps in volatility and allow different effects of negative and positive price changes.

5 Priors

We parameterise both the Gamma marginal distribution and Gamma mixing distribution in terms of their mean and shape parameter. Independent priors

are placed on all these parameters. The parameters of the marginal distribution of $\sigma^2(t)$ are given vague priors: $\nu \sim \text{Ga}(1, 0.001)$ and $\nu/\gamma \sim \text{IG}(1, 0.001)$. Informative priors are chosen for the parameters of the mixing distribution. The shape parameter of the mixing distribution α is given an inverted Gamma prior with shape parameter 1 and mean $1/\log 2$ which places half of its mass on long memory processes and half of its mass on short memory processes. The choice of an inverted Gamma distribution places some mass at values much larger than 1. The mean parameter $\xi = \alpha/\phi$ is given an exponential distribution with mean $1/\xi_0$. This stops the prior placing mass on very large values of the mean. The parameters of the drift μ, the risk premium β and the leverage effect ρ are given independent vague zero-mean prior distributions with a standard deviation of 100.

6 Markov chain Monte Carlo inference methods

The representation of the continuous superposition model (3) is expressed in terms of a Poisson process on (τ, λ, J) with intensity $\nu\lambda f(\lambda)\gamma \exp\{-\gamma J\}$. If the mean of F is finite, it is useful for simulation methods to think of the process as a marked Poisson process. In that case, the jump times τ follow a Poisson process with intensity $\nu\text{E}[\lambda]$ and the jumps sizes, J, and decay rates, λ, are marks. The jumps are independent and exponentially distributed with mean $1/\gamma$ and the decay rates are independent and their distribution has the density $f^\star(\lambda) \propto \lambda f(\lambda)$. If the mixing distribution is $\text{Ga}(\alpha, \phi)$ then $f^\star$ is the density of a $\text{Ga}(\alpha + 1, \phi)$ distribution.

Markov chain Monte Carlo (MCMC) methods for OU processes have been developed by Roberts *et al* (2004) and Griffin and Steel (2006). Standard MCMC methods are difficult to apply due to slow mixing of λ and ν. This problem can be addressed by jointly updating the process z jointly with λ or ν. Roberts *et al.* (2004) propose a method of retrospective sampling of z and Griffin and Steel (2006) suggest "dependent thinning". Whereas Roberts *et al.* (2004) suggest the use of a reparameterisation to reduce the correlation between the data and the process, the dependent thinning of Griffin and Steel (2006) restricts the changes in the proposed process to relatively small jumps. We will consider extending both of these methods to the continuous superposition case.

6.1 An approximate process

The superposed process can be approximated in the following way. The effect of the i-th jump at time t is $J_i \exp\{-\lambda_i(t-\tau_i)\}$ and the contribution to the integrated volatility after time s_i is

$$\int_{s_i}^{\infty} J_i \exp\{-\lambda_i(t-\tau_i)\}\, dt = \frac{J_i}{\lambda_i} \exp\{-\lambda_i(s_i-\tau_i)\}.$$

Choosing a small fraction d (where $0 < d < 1$) allows us to define an approximate version of the effect of the i-th jump. In particular, let

$$s_i = \tau_i - \frac{1}{\lambda_i} \log d$$

and define an approximation to the effect of the jump to be $J_i \mathrm{I}(\tau_i < t < s_i) \exp\{-\lambda_i(t-\tau_i)\}$. The actual volatility σ_n^2 is now expressed as

$$\begin{aligned}\sigma_n^2 &= \sum_{i=1}^{\infty} E_i\, J_i \int_{\max\{(n-1)\Delta,\tau_i\}}^{\min\{n\Delta,s_i\}} \exp\{-\lambda_i(u-\tau_i)\}\, du \\ &= \sum_{i=1}^{\infty} E_i\, J_i \left[\exp\{-\lambda_i(\max((n-1)\Delta-\tau_i,0)\} - \exp\{-\lambda_i(\min(n\Delta-\tau_i, s_i-\tau_i)\}\right],\end{aligned}$$

where $E_i = \mathrm{I}(\tau_i < n\Delta \text{ and } (n-1)\Delta < s_i)$. In this manner, we only disregard a fraction $0 < d < 1$ of the total effect of each jump. Usually d is taken to be very small. The main point of the approximation is to avoid lots of computational effort by carrying along very small residual effects of jumps. More formally, to calculate σ_n^2 the expected number of elements in the sum (*i.e.* the number of jumps) is $(n\Delta - B)\nu\alpha/\phi$ whereas using the truncation the expected number is

$$-\nu \log d + \nu \left[\xi \Gamma\left(\alpha+1, -\frac{\alpha}{\xi}\frac{\log d}{n\Delta - B}\right) + \log d \Gamma\left(\alpha, -\frac{\alpha}{\xi}\frac{\log d}{n\Delta - B}\right)\right]$$

where $\Gamma(\alpha, x) = \frac{1}{\Gamma(\alpha)} \int_0^x u^{\alpha-1} \exp\{-u\}\, du$. For small d and large n, the expected number will be bounded above by $-\nu \log d$. It follows that calculation of the log-likelihood involves $O(n^2)$ terms without truncation but $O(n)$ terms with truncation. This difference in computational complexity makes an important difference to speed of execution when we have a long time series.

6.2 Sampling algorithms

If we choose a Gamma marginal distribution then $\mathrm{E}[z_n] = \frac{\nu}{\gamma}\xi\Delta$, where $\xi = \alpha/\phi$ is the mean of the mixing distribution. The model that we fit can then be

expressed in the following way for each observation $n = 1, \ldots, N$

$$y_n \sim \mathrm{N}\left(\mu\Delta + \beta\sigma_n^2 + \rho\left(z_n - \frac{\nu}{\gamma}\xi\Delta\right), \sigma_n^2\right),$$

$$\text{with } z_n = \sum_{i=1}^{\infty} \mathrm{I}((n-1)\Delta < \tau_i < n\Delta) J_i \text{ and}$$

$$\sigma_n^2 = \sum_{i=1}^{\infty} E_i J_i \left[\exp\{-\lambda_i(\max((n-1)\Delta - \tau_i, 0)\} - \exp\{-\lambda_i(\min(n\Delta - \tau_i, s_i - \tau_i)\}\right],$$

where

$$k \sim \mathrm{Pn}\left(\nu\xi(n\Delta - B)\right), \qquad \tau_1, \tau_2, \ldots, \tau_k \sim \mathrm{U}([B, N\Delta])$$

$$J_1, J_2, \ldots, J_k \overset{i.i.d.}{\sim} \mathrm{Ga}(1, \gamma), \qquad \lambda_1, \lambda_2, \ldots \lambda_k \overset{i.i.d.}{\sim} \mathrm{Ga}\left(\alpha + 1, \frac{\alpha}{\xi}\right).$$

We write $\mathrm{U}(R)$ for the uniform distribution on the set R and $\mathrm{Pn}(a)$ denotes the Poisson distribution with mean a. The prior distributions for all model parameters are given in Section 5. The posterior distribution of the parameters is proportional to

$$p(y|\psi)p(J|\gamma)p(\lambda|\alpha,\xi)p(\tau)p(k|\xi,\nu)p(\gamma|\nu)p(\nu)p(\alpha)p(\xi)p(\mu)p(\beta)p(\rho),$$

where $\psi = (J, \lambda, \tau, \mu, \beta, \rho, \xi, \nu, \gamma)$, $y = (y_1, \ldots, y_N)'$, J, τ and λ are the sets of $(J_1, \ldots, J_k)$, $(\tau_1, \ldots, \tau_k)$ and $(\lambda_1, \ldots, \lambda_k)$, respectively, for all observations and k is the total number of jumps within the finite interval $[B, N\Delta]$, which will cover the observation times of the data. The following subsections describe simulation methods necessary to build a Gibbs sampler for this posterior distribution. Throughout these sections, ψ' will refer to a proposed value of ψ where all parameters apart from those being updated are kept at their current values.

6.2.1 Updating J

The parameters $J_1, J_2, \ldots, J_k$ are updated with a single-site Metropolis-Hastings random walk sampler on the log scale, *i.e.* we propose $J_i' = J_i \exp\{\epsilon_i\}$ where $\epsilon_i \sim \mathrm{N}(0, \sigma_J^2)$, which is accepted with probability

$$\min\left\{1, \frac{p(y|\psi')J_i' \exp\{-\gamma J_i'\}}{p(y|\psi)J_i \exp\{-\gamma J_i\}}\right\}.$$

The variance σ_J^2 is chosen to give an average acceptance rate of between 0.2 and 0.3.

6.2.2 Updating λ

The parameters $\lambda_1, \lambda_2, \ldots, \lambda_k$ are updated one-at-a-time using a Metropolis-Hastings random walk on the log scale. The proposed value, λ'_j, is accepted with probability

$$\min\left\{1, \frac{p(y|\psi')\lambda'_j{}^{\alpha+1}\exp\{-\lambda'_j\alpha/\xi\}}{p(y|\psi)\lambda_j{}^{\alpha+1}\exp\{-\lambda_j\alpha/\xi\}}\right\}$$

The variance of the increments of the random walk is chosen to give an average acceptance rate of between 0.2 and 0.3.

6.2.3 Updating τ

The parameters $\tau_1, \tau_2, \ldots, \tau_k$ are updated using a Metropolis-Hastings random walk. A new $\tau'_i = \tau_i + \epsilon_i$ with $\epsilon_i \sim \mathrm{N}(0, \sigma^2_\tau)$ is proposed, which is rejected if $\tau'_i > N\Delta$ or $\tau'_i < B$. Otherwise, the proposed value is accepted with probability

$$\min\left\{1, \frac{p(y|\psi')}{p(y|\psi)}\right\}.$$

The variance σ_τ is tuned to have an acceptance rate of between 0.35 to 0.4. We choose a higher rate than would be standard to avoid poor mixing of larger jumps. Smaller jumps will usually be easier to move than large jumps. A standard value of the average acceptance rate would be lead to a much smaller acceptance rate for large jumps.

6.2.4 Updating k

Updating the parameter k involves a change of dimension of the parameter space and uses Reversible Jump Markov chain Monte Carlo (Green, 1995). The sampler has two moves: $k' = k + 1$ or $k' = k - 1$, which are both proposed with probability 1/2. If $k' = k + 1$, we propose new vectors J', λ' and τ' where $J'_i = J_i$, $\lambda'_i = \lambda_i$ and $\tau'_i = \tau_i$ for $1 \leq j \leq k$ and $J'_{k+1} \sim \mathrm{Ga}(1, \gamma)$, $\lambda'_{k+1} \sim \mathrm{Ga}\left(\alpha + 1, \frac{\alpha}{\xi}\right)$ and $\tau'_{k+1} \sim \mathrm{U}([B, N\Delta])$. The values k', J', λ' and τ' are accepted with probability

$$\min\left\{1, \frac{p(y|\psi')}{p(y|\psi)}\frac{(N\Delta - B)\nu\xi}{k'}\right\}.$$

If $k' = k - 1$, then a value j is drawn at random from $\{1, 2, \ldots, k\}$ and new vectors J', λ' and τ' are proposed where $J'_i = J_i$, $\lambda'_i = \lambda_i$ and $\tau'_i = \tau_i$ for

$1 \leq i < j$ and $J_i' = J_{i+1}$, $\lambda_i' = \lambda_{i+1}$ and $\tau_i' = \tau_{i+1}$ for $j \leq i \leq k'$. The proposed values are accepted with probability

$$\min\left\{1, \frac{p(y|\psi')}{p(y|\psi)} \frac{k}{(N\Delta - B)\nu\xi}\right\}.$$

6.2.5 Updating μ, β and ρ

Let $\theta = (\mu, \beta, \rho)'$ then the full conditional distribution of θ is

$$\mathrm{N}\left((\Lambda + X'\Sigma^{-1}X)^{-1}\Sigma^{-1}X'y, (\Lambda + X'\Sigma^{-1}X)^{-1}\right)$$

where Σ is a diagonal matrix with elements $\Sigma_{jj} = \sigma_j^2$, Λ is the prior precision of θ and X is a $N \times 3$-dimensional matrix with j-th row equal to

$$\left(\Delta,\ \sigma_j^2,\ \sum_{i=1}^{k} \mathrm{I}((j-1)\Delta < \tau_i < j\Delta)J_i - \frac{\nu}{\gamma}\xi\Delta\right).$$

6.2.6 Updating α

The full conditional distribution of α is proportional to

$$p(\lambda|\alpha,\xi)p(\alpha) \propto \alpha^{-2}\exp\left\{-\frac{\log 2}{\alpha}\right\}\frac{\alpha^{k\alpha}}{\xi^{k\alpha}\left(\Gamma(\alpha)\right)^k}\left(\prod_{i=1}^{k}\lambda_i\right)^{\alpha}\exp\left\{-\frac{\alpha}{\xi}\sum_{i=1}^{k}\lambda_i\right\}.$$

A rejection envelope can be defined using a simplification of Stirling's approximation $\Gamma(\alpha) \approx \exp\{-\alpha\}\alpha^{\alpha-1/2}(2\pi)^{-1/2}$. Putting this formula into the full conditional above gives the following envelope

$$\alpha^{k/2-2}\exp\left\{-\alpha\left[\frac{\sum_{i=1}^{k}\lambda_i}{\xi} - k + k\log\xi - \sum_{i=1}^{k}\log\lambda_i\right] - \frac{\log 2}{\alpha}\right\}.$$

which is the density of a Generalized Inverse Gaussian (GIG) distribution. Efficient methods for the simulation from this distribution are described by Devroye (1986). If we denote by f the target density and g is the rejection envelope, then $\max(f/g) = 0.4^k$. For large k the rejection sampler may have too small a chance of proposing a value in a reasonable amount of time. Therefore, we suggest that if there are more than 100 rejections, we update using a standard random walk Metropolis-Hastings sampler tuned to obtain an acceptance rate of around 25%.

6.2.7 Updating ξ

The full conditional distribution of ξ is proportional to

$$p(y|\psi)\xi^{-k\alpha}\exp\left\{-\frac{\alpha}{\xi}\sum_{i=1}^{k}\lambda_i-\xi(\xi_0+\nu(N\Delta-B))\right\}.$$

The likelihood $p(y|\psi)$ only depend on ξ through $\mathrm{E}[z_n]=\frac{\nu}{\gamma}\xi\Delta$ which is included to model the leverage effect. This is not likely to change a lot with ξ and so we use a Metropolis-Hastings independence sampler where values are proposed from a density proportional to

$$\xi^{-k\alpha}\exp\left\{-\frac{\alpha}{\xi}\sum_{i=1}^{k}\lambda_i-\xi(\xi_0+\nu(N\Delta-B))\right\}.$$

This density is proportional to that of a GIG distribution.

6.2.8 Updating ν

We can directly apply the retrospective method of Roberts *et al* (2004). Suppose we propose to move from ν to ν' using transition kernel $q(\nu,\nu')$. Then we propose a new process in the following way. If $\nu'>\nu$, then simulate $m\sim\mathrm{Pn}((\nu'-\nu)\xi(N\Delta-B))$ and simulate $J_1',J_2',\ldots,J_m'$, $\tau_1',\tau_2',\ldots,\tau_m'$ and $\lambda_1',\lambda_2',\ldots,\lambda_m'$ where $J_i'\sim\mathrm{Ga}(1,\gamma)$, $\tau_i'\sim\mathrm{U}([B,N\Delta])$ and $\lambda_i'\sim\mathrm{Ga}\left(\alpha+1,\frac{\alpha}{\xi}\right)$. The new process is formed by taking the superposition of the current values (τ,J,λ) with the new values (τ',J',λ'). The acceptance probability of ν' is calculated by taking the ratio

$$\frac{p(y|\psi')q(\nu',\nu)}{p(y|\psi)q(\nu,\nu')} \tag{5}$$

If $\nu'<\nu$ then we form the new process (τ',J',λ') by thinning the current states (τ,J,λ) with thinning probability ν'/ν. The new value is accepted with the probability in (5).

The dependent thinning method of Griffin and Steel (2006) is similar to the retrospective method but uses a different method for proposing the new process. Direct application of their method to the continuous superposition model would imply that the proposed process (τ',J',λ') is defined in the following way. If $\nu'>\nu$, they simulate $m\sim\mathrm{Pn}((\nu'-\nu)\xi(N\Delta-B))$ and propose $J_i'=J_i+\log(\frac{\nu'}{\nu})/\gamma$, $\tau_i'=\tau_i$ and $\lambda_i'=\lambda$ for $1\leq i\leq k$ and $J_i'\sim\mathrm{Ga}(1,\gamma)$ where $J_i'<\log\left(\frac{\nu'}{\nu}\right)\Big/\gamma$, $\tau_i'\sim\mathrm{U}([B,N\Delta])$, $\lambda_i'\sim\mathrm{Ga}\left(\alpha+1,\frac{\alpha}{\xi}\right)$ for $k+1\leq i\leq k+m$.

If $\nu' < \nu$, the proposed process is $J_i' = J_i - \log(\frac{\nu'}{\nu})/\gamma$, $\tau_i' = \tau_i$ and $\lambda_i' = \lambda$ for $1 \leq i \leq k$ where $(J_i', \tau_i', \lambda_i')$ is only included if $J_i' > 0$. Once again the acceptance probability is given by equation (5). In continuous superposition models this method leads to poor mixing in the chain. It is not hard to see the reason. Unlike the discrete superposition, λ_i differ between jumps. Jumps with smaller values of λ_i will have longer-lasting effects on the integrated volatility and more effect on the change in the likelihood value between the current and proposed states. This problem can be addressed by allowing the change in jumps sizes to depend on their effect on the integrated volatility (and consequently the decay rates). The contribution of a jump to the integrated volatility is $\frac{J_i}{\lambda_i}$. This suggests, when $\nu' > \nu$ proposing $J_i' = J_i + \frac{a^\star \lambda_i}{\gamma}$ (where $a^\star$ is chosen as explained below) for $1 \leq i \leq k$ and drawing the new jumps from a Poisson process with intensity $f(\lambda, J, \tau) = \nu' \lambda f(\lambda) \exp\{-\gamma J\}$ truncated to the region $(0, \infty) \times (0, \frac{a^\star \lambda}{\gamma}) \times (B, N\Delta)$. The process can be simply simulated in the following way: $\left(\text{if the mixing distribution is Ga}\left(\alpha, \frac{\alpha}{\xi}\right)\right)$

1. The number of new jumps m is Poisson distributed with mean
$$\nu' \xi \left[1 - \left(1 + a^\star \frac{\xi}{\alpha}\right)^{-(\alpha+1)}\right](N\Delta - B).$$
2. $\tau_{k+1}', \tau_{k+2}', \ldots, \tau_{k+m}' \overset{i.i.d.}{\sim} \mathrm{U}([B, N\Delta])$
3. $(J_{k+1}', \lambda_{k+1}'), (J_{k+2}', \lambda_{k+2}') \ldots, (J_{k+m}', \lambda_{k+m}')$ are independent and can be simulated using a rejection sampler. The rejection envelope generates $\lambda_i' \sim \text{Ga}\left(\alpha + 1, \frac{\alpha}{\xi}\right)$ and J_i' from an $\text{Ga}(1, \gamma)$ truncated to the region $\left(0, \frac{a^\star \lambda_i'}{\gamma}\right)$. These values are accepted with probability $1 - \exp\{-a^\star \lambda_i'\}$.

The acceptance probability of $(\nu', J', \tau', \lambda')$ is given by the minimum of 1 and
$$A \frac{p(y|\psi')q(\nu', \nu)}{p(y|\psi)q(\nu, \nu')}.$$
where
$$A = \left(\frac{\nu'}{\nu}\right)^k \exp\left\{-a^\star \sum_{i=1}^k \lambda_i\right\} \exp\left\{(N\Delta - B)\xi \left[\nu - \nu' \left(1 + a^\star \frac{\xi}{\alpha}\right)^{-(\alpha+1)}\right]\right\}.$$
In the retrospective sampler and the original dependent thinning method $A = 1$. This is a useful value since it replicates the acceptance probability of a Metropolis-Hastings random walk sampler (see Brooks *et al.*, 2003, for a discussion of this point for general reversible jump MCMC algorithms). In this

case, it is hard to find a simple method for choosing $a^\star$ to guarantee $A = 1$ for all parameter values. Therefore we choose the value of $a^\star$ to guarantee that $\mathrm{E}[A] = 1$ for all values of the parameters where the expectation is taken with respect to $\lambda_1, \lambda_2, \ldots, \lambda_k$. Then

$$\mathrm{E}[A] = \left(\frac{\nu'}{\nu}\right)^k \left(1 + a^\star \frac{\xi}{\alpha}\right)^{-k(\alpha+1)} \exp\left\{(N\Delta - B)\xi\left[\nu - \nu'\left(1 + a^\star \frac{\xi}{\alpha}\right)^{-(\alpha+1)}\right]\right\},$$

and $\mathrm{E}[A] = 1$ when

$$a^\star = \frac{\alpha}{\xi}\left(\left(\frac{\nu'}{\nu}\right)^{1/(\alpha+1)} - 1\right).$$

6.2.9 Updating γ

The full conditional distribution of γ is proportional to

$$p(y|\psi)p(\gamma|\nu)\gamma^{k-1}\exp\left\{-\gamma\sum_{i=1}^{k}\lambda_i\right\}.$$

Once again, the likelihood $p(y|\psi)$ only depends on γ through $\mathrm{E}[z_n] = \frac{\nu}{\gamma}\xi\Delta$ and we use a Metropolis-Hastings independence sampler with proposal density proportional to

$$\gamma^{k-\alpha-2}\exp\left\{-\gamma\sum_{i=1}^{k}\lambda_i - \beta\nu/\gamma\right\}.$$

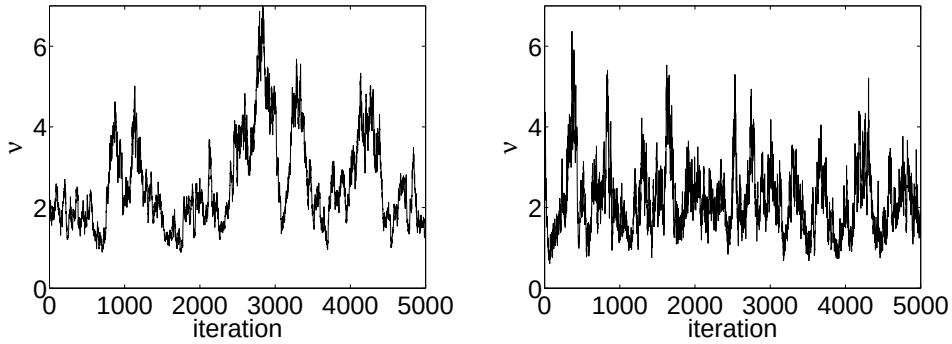
This proposal distribution is a GIG distribution.

6.3 Comparison of samplers

Both the retrospective sampler and the revised dependent thinning explained in Subsection 6.2.8 are used in the algorithm. Figure 1 shows trace plots of the draws for ν in the context of the application described in the next section. The simulation show 5000 values derived from a chain of length 250,000 with thinning to retain every 50th value. The execution times for both samplers are similar and the acceptance rate for ν in both samplers was chosen to be close to 0.25. It seems clear that the dependent thinning sampler mixes better in this case.

7 Application to a stock price index

We consider daily observations of the Standard and Poor's 500 index of the New York Stock Exchange from June 6, 1980 to June 6, 2000. Thus, the ob-



(a) (b)

Figure 1: Trace plots of the draws for ν using (a) the retrospective sampler and (b) the revised dependent thinning explained in Subsection 6.2.8

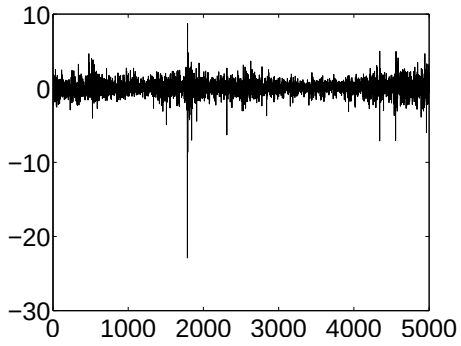


Figure 2: Returns of the Standard and Poor's index between June 6, 1980 and June 6, 2000

served y consists of the 5000 data points shown in Figure 2. This is almost the same sample as used in Li *et al.* (2008) and extends the period covered by Griffin and Steel (2006).

The parameter estimates for a one-component OU model, two-component discrete superposition and continuous superposition with gamma mixing distribution are presented in Table 1. There are several trends in the results. Firstly, the posterior median estimates of the expectation and standard deviation of $\sigma^2(t)$ increase as we move from one component to two components and to the continuous model. The posterior median of $\mathrm{E}[\lambda]$ changes very little between the three models. However, the width of the 95% credible interval increases as we move from the one-component to the two-component model and to the continuous model, which is in line with the increasing flexibility on λ. In all models there seems to be strong evidence for a leverage effect. The estimates of ρ also markedly change between the three models.

	one-component	two-component	continuous
E$[\sigma^2(t)]$	0.80 (0.67, 0.98)	0.93 (0.65, 1.31)	0.99 (0.55, 1.96)
SD$[\sigma^2(t)]$	0.45 (0.37, 0.56)	0.64 (0.47, 0.89)	0.70 (0.49, 1.06)
w		0.84 (0.69, 0.92)	
λ_1	0.016 (0.011, 0.022)	0.004 (0.002, 0.006)	
λ_2		0.082 (0.045, 0.150)	
α			0.26 (0.12, 0.59)
E$[\lambda]$	0.016 (0.011, 0.022)	0.017 (0.011, 0.027)	0.017 (0.008, 0.037)
μ	0.006 (-0.034, 0.044)	0.013 (-0.028, 0.053)	0.012 (-0.028, 0.051)
β	0.054 (-0.003, 0.112)	0.046 (-0.013, 0.106)	0.045 (-0.012, 0.103)
ρ	-4.56 (-6.03, -3.39)	-3.09 (-4.36, -2.08)	-2.75 (-4.00, -1.80)

Table 1: Posterior estimates of parameters in the Gamma model for one and two components and the continuous superposition: posterior median with 95% credible interval in brackets

The one-component model has a much larger (in absolute value) posterior median estimate of ρ than the other two models. In fact, this estimate does not fall within the 95% credible intervals for the two-component and continuous models. These results show that there is a trade-off in these models between estimates of the distribution of $\sigma^2(t)$ and the leverage effect. The more flexible models can have larger means and variances of $\sigma^2(t)$, which offers increased mass to larger volatilities. Therefore large movement in the returns can be explained by larger volatilities and lower amounts of leverage. The continuous model shows clear evidence in favour of long memory. The posterior median estimate of α is 0.26 with a 95% credible interval which is far away from 1. The posterior probability for long memory ($\alpha < 1$) is virtually one in this model.

Posterior estimates of the volatility are shown in Figure 3 with both posterior median and 95% credible intervals included. The inferences from the two-component and continuous superposition models are similar. However, there are clear differences between the inference from these models and the one-component models. The one-component models tends to have a smaller range of volatility estimates and the inflexibility of the dynamics leads to an oversmoothed estimate. The differences between the three models are perhaps most marked at times of rapid change in volatility. Such a period is illustrated in Figure 4 which shows posterior median estimates for observations 1600 to

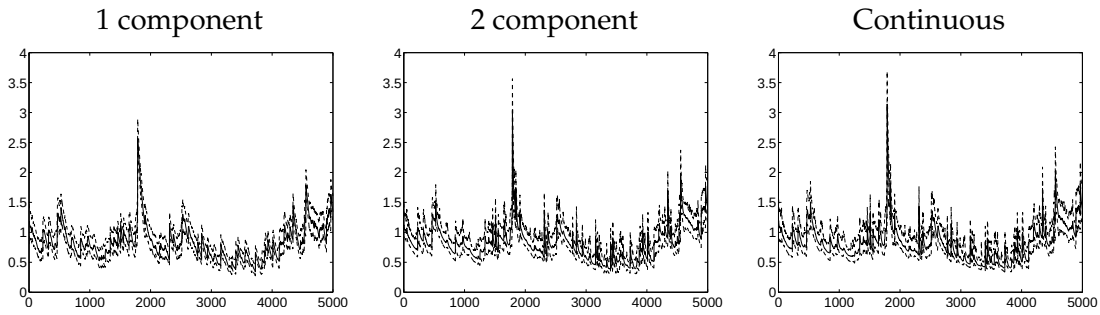


Figure 3: Posterior smoothed estimates of σ_n for the three models (solid lines are median values and dotted lines form point-wise 95% credible intervals)

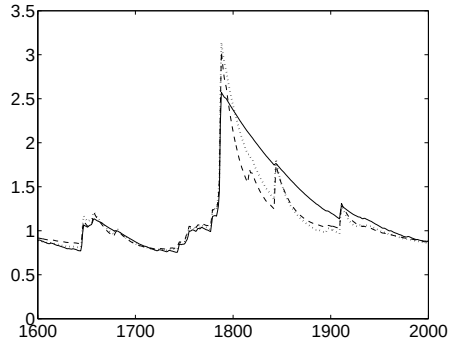


Figure 4: Posterior smoothed median estimates of σ_n for observations 1600 to 2000 for the three different models: one-component (solid line), two-component (dashed line) and continuous (dotted line)

2000 covering the "Black Monday" crash on October 19, 1987. In this period there are once again clear differences between the one-component model and the others. The one-component model has a smaller jump on that day with a slower decay. The differences between the continuous and two-component models are restricted to the size of the jump and the period directly after the jump. The additional jump at about 1840 leads to estimates from the two models which are virtually indistinguishable. However, the jump with the continuous model is larger and decays more slowly than the two-component model. In fact, the two-component model introduces a second jump at about observation 1815.

The autocorrelation functions for $\sigma^2(t)$ are illustrated in Figure 5 and show a large difference between the estimates. As we might expect, the continuous

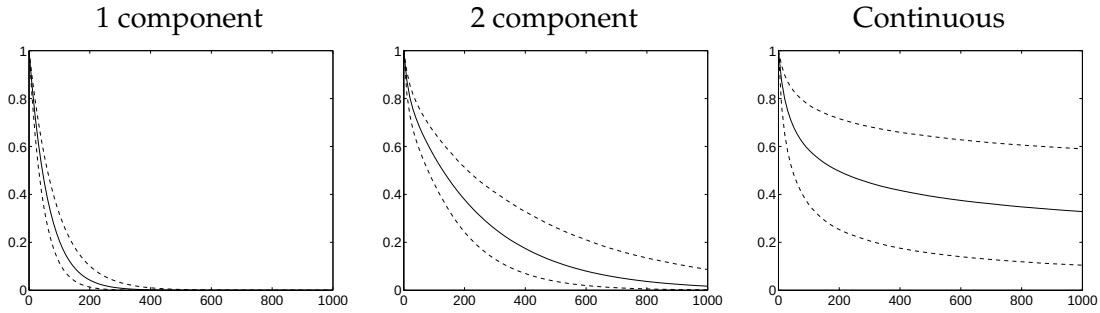


Figure 5: Posterior estimates of the autocorrelation of $\sigma^2(t)$ for the three models (solid lines are median values and dotted lines form point-wise 95% credible intervals)

superposition leads to a much slower decay than is possible with the two-component superposition. However, the shape of the autocorrelation function seems to be different between the two models even for small values of t with the continuous superposition giving more posterior mass to large autocorrelations.

We compare the models using one-step-ahead out-of-sample predictions with the parameter values fixed at their posterior median value. In particular, we calculate the log predictive score as

$$LPS = -\frac{1}{N}\sum_{n=1}^{N} \log p(y_n|y_1, \ldots, y_{n-1}, \hat{\theta})$$

where $\hat{\theta}$ refers to the posterior median of the parameters. The necessary conditional distributions can be simply implemented using particle filtering methods (see Creal, 2008 for a discussion of the application of particle filtering methods to discrete superposition models). We use the method introduced by Carpenter *et al* (1999). The results for the three models are shown in Table 2. They show that both the two-component and continuous superposition models outperform the one-component model. However, the difference between the two-component and continuous model is much less marked, making it difficult to distinguish between them in terms of predictive performance.

one-component	0.349
two-component	0.342
continuous	0.341

Table 2: Log predictive scores for the three models

Figure 6 shows the difference between the running LPS for the continuous

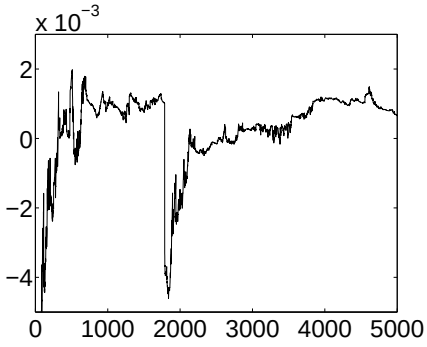


Figure 6: The difference in running LPS for the continuous and two-component superposition models

superposition and two-component superposition models. The graph clearly shows that the two-component superposition model predicts well at the moment of the crash but at most other times the continuous superposition is outperforming it in terms of predictive performance. Griffin and Steel (2006) compare models using marginal likelihoods estimated by the modified harmonic mean estimators of Newton and Raftery (1994) but we find that these methods give unreliable results for the continuous superposition model.

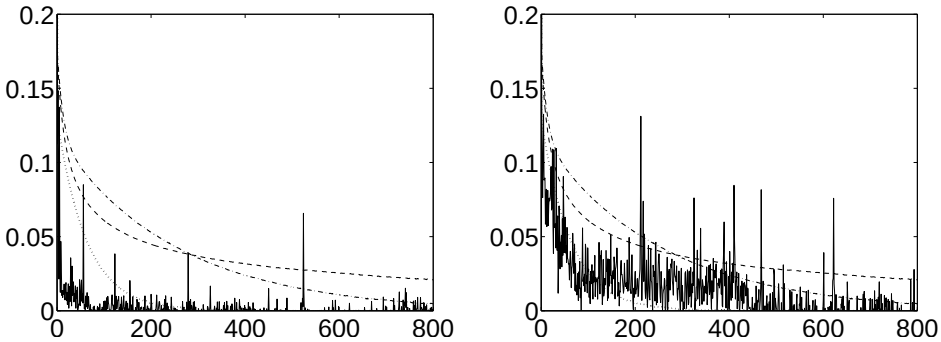


Figure 7: The autocorrelation of the squared returns for (a) all data and (b) all data excluding the week starting with "Black Monday". The empirical autocorrelation function is the solid line overplotted with the median theoretical autocorrelation associated with: the continuous superposition (dashed line), 1 component model (dotted line) and 2 component model (dash-dot line)

Another measure of the fit of the models is given by the models' ability to reproduce the empirical properties of the data. Figure 7 shows the empiri-

cal autocorrelation function of the squared returns and its theoretical median value for the different models. The theoretical autocorrelations are calculated using a Monte Carlo method at the posterior median estimates of the parameters. The large movements in the market on and following "Black Monday" have a large effect on the estimated autocorrelation and a second empirical autocorrelation function excluding that week is also shown (panel (b)). Perhaps not surprisingly, the theoretical models seem to more closely resemble this second estimate. The one-component model can capture the rapid decay at short lags but cannot generate the persistent autocorrelations at longer lags. The other models generate larger autocorrelations at shorter lags but can also generate more persistence. The autocorrelation for the two-component model seems to decay to zero around lag 800 where the long memory process still retains a relatively large autocorrelation.

		Mean	Variance	Skewness	Kurtosis
$\Delta = 1$	Empirical	0.05 (0.05)	1.04 (0.91)	-2.53 (-0.47)	59.63 (8.74)
	2 component	0.06	1.06	-0.46	6.93
	Continuous.	0.04	0.84	-0.67	9.06
$\Delta = 5$	Empirical	0.24 (0.26)	4.77 (4.48)	-0.83 (-0.29)	10.32 (5.95)
	2 component	0.28	5.27	-0.19	4.92
	Continuous	0.22	4.16	-0.30	5.96
$\Delta = 20$	Empirical	0.97 (1.03)	17.35 (16.76)	-0.59 (-0.15)	6.42 (3.64)
	2 component	1.12	20.95	-0.04	4.42
	Continuous	0.88	16.47	-0.11	5.12

Table 3: Empirical and theoretical moments for the two-component and continuous superposition models for three values of Δ. The bracketed empirical moments exclude the week starting with "Black Monday".

Table 3 shows the theoretical moments calculated at the posterior median estimates of the parameters at different frequencies for the two-component and continuous superposition models. We would hope that models could generate values of the moments that are consistent with the empirical estimates. Both models can generate values close to the empirical mean at all time frequencies. The two-component model generates a larger variance than the continuous superposition model. At longer time frequencies, this leads to a value above the empirical value whereas the continuous model is able to gen-

erate a closer fit to the empirical variance at the longest frequency. The two empirical estimates of the higher moments are very sensitive to the effect of the week beginning with "Black Monday" with the bracketed values showing much smaller (in absolute value) estimates. At the shortest time interval (highest frequency) the two-component model can generate a level of skewness closer to the bracketed empirical estimate than the continuous. At longer time intervals, the skewness in the two-component model decreases at a much faster rate than the empirical values. In contrast, the continuous model generates more skewness at the higher frequencies ($\Delta = 5$ and $\Delta = 20$). In the case of kurtosis the continuous model has theoretical moments much closer to the empirical values than the two-component model for $\Delta = 1$ and $\Delta = 5$.

The fact that the continuous model does not convincingly beat the two-component model in terms of fit to the data is perhaps not surprising in view of the fact that on a subset of these data, Griffin and Steel (2006) find a two-component model to behave best among superposition models with a finite number of components. The continuous model fits roughly the same, but provides the extra possibility to accommodate long memory (strongly supported by the data) and does not require the specific selection of a finite number of components.

8 Conclusion

We have examined models for stochastic volatility based on continuous superpositions of OU processes driven by pure jump Lévy processes. Such models are interesting as they can generate long memory, which corresponds to a simple parametric restriction on the shape parameter of the mixing distribution if we use a Gamma mixture of the rate parameters. In addition, there is no need to choose a finite number of components in a superposition model. The model will naturally be adaptive to the data.

In the context of an asset returns model with risk premium and leverage, we propose efficient MCMC methods for Bayesian inference. In order to propose efficient updates of the process, we implement both the retrospective method of Roberts *et al.* (2004) and we propose a new version of the dependent thinning methods of Griffin and Steel (2006). Applied to an S&P 500 returns series, it appears the latter method mixes better. Comparison of a competitive two-component model and the continuous superposition model reveals that

both have relative strengths and weaknesses leading to roughly similar overall fits, but the data do strongly support the presence of long memory in the continuous superposition model.

References

Barndorff-Nielsen, O. E. (2001): "Superposition of Ornstein-Uhlenbeck type processes," *Theory of Probability and its Applications*, 45, 175-194.

Barndorff-Nielsen, O. E. and N. Shephard (2001): "Non-Gaussian OU based models and some of their uses in financial economics," *Journal of the Royal Statistical Society B*, 63, 167-241 (with discussion).

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

Bondesson, L. (1988): "Shot-Noise Processes and Distributions," *Encyclopedia of Statistical Science*, Vol 8. Wiley: New York.

Brooks, S. P., P. Giudici and G. O. Roberts (2003): "Efficient construction of reversible jump Markov chain Monte Carlo proposal distributions," *Journal of the Royal Statistical Society B*, 65, 3-55.

Carpenter, J., P. Clifford and P. Fearnhead (1999): "An improved particle filter for non-linear problems," *IEE proceedings - Radar, Sonar and Navigation*, 146, 2-7.

Creal, D. D. (2008): "Analysis of filtering and smoothing algorithms for Lévy-driven stochastic volatility models," *Computational Statistics and Data Analysis*, 52, 2863-2876.

Devroye, L. (1986): "Non-Uniform Random Variate Generation," Springer-Verlag: New York.

Ferguson, T. and Klass, M. J. (1972): "A representation of independent increment processes without Gaussian components. *Annals of Mathematical Statistics*, 43, 1634-1643.

Frühwirth-Schnatter, S. and L. Sögner (2008): "Bayesian Estimation of Stochastic Volatility Models based on OU processes with Marginal Gamma Law," *The Annals of the Institute of Statistical Mathematics*, forthcoming.

Gander, M. P. S. and D. A. Stephens (2007a): "Stochastic Volatility Modelling with General Marginal Distributions: Inference, Prediction and Model Selection," *Journal of Statistical Planning and Inference*, 137, 3068-3081.

Gander, M. P. S. and D. A. Stephens (2007b): "Simulation and inference for stochastic volatility models driven by Lévy processes," *Biometrika*, 94, 627-646.

Green, P. J. (1995): "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, 82, 711-732.

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

Li, H., M. T. Wells and C. Yu (2008): "A Bayesian analysis of returns dynamics with Lévy jumps," *Review of Financial Studies*, 21, 2345 - 2378.

Newton, M. A. and A. E. Raftery (1994): "Approximate Bayesian inference by the weighted likelihood bootstrap," *Journal of the Royal Statistical Society B*, 56, 3-48.

Nicolato, E. and E. Venardos (2003): "Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type," *Mathematical Finance*, 13, 445-466.

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

Rosinski, J. (2001): "Series representations of Lévy processes from the perspective of point process," in *Lévy processes - Theory and Applications* eds.: O. E. Barndorff-Nielsen, T. Mikosch and S. Resnick. Birkhäuser: Boston.