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

J.E. Griffin\(^*\), M. F. J. Steel\(^b\)

\(^a\)Institute of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury, CT2 7NF, U.K.
\(^b\)Department of Statistics, University of Warwick, Coventry, CV4 7AL, U.K.

Abstract

Continuous superpositions of Ornstein-Uhlenbeck processes are proposed as a model for asset return volatility. An interesting class of continuous superpositions is defined by a Gamma mixing distribution which can define long memory processes. In contrast, previously studied discrete superpositions cannot generate this behaviour. Efficient Markov chain Monte Carlo methods for Bayesian inference are developed which allow the estimation of such models with leverage effects. The continuous superposition model is applied to both stock index and exchange rate data. The continuous superposition model is compared with a two component superposition on the daily Standard and Poor’s 500 index from 1980 to 2000.

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

\[
d y(t) = \{\mu + \beta \sigma^2(t)\} dt + \sigma(t) \, dB(t),
\]

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

\[
y_n \sim N(\mu \Delta + \beta \sigma^2_n, \sigma^2_n),
\]

\(^*\)Corresponding author. Tel.: +44-1227-823627; fax: +44-1227-827932

Email addresses: J.E.Griffin-28@kent.ac.uk (J.E. Griffin), M.F.Steel@stats.warwick.ac.uk (M. F. J. Steel)

Preprint submitted to Elsevier June 10, 2009
given the discrete or “actual” volatility $\sigma_n^2 = \int_{(n-1)\Delta}^{n\Delta} \sigma^2(u)du$. Most models in the literature assume a stochastic process for the instantaneous volatility based on Brownian motion. Following Barndorff-Nielsen and Shephard [2] 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 [21]).

Other studies that have considered Bayesian inference with similar volatility processes are Roberts et al. [23], Griffin and Steel [17], Gander and Stephens [14, 15] and Frühwirth-Schnatter and Sogner [13].

In this paper, we model the instantaneous volatility with a continuous superposition of Ornstein-Uhlenbeck processes, studied in detail in Barndorff-Nielsen [1] and mentioned in the context of volatility modelling in Barndorff-Nielsen and Shephard [2]. These processes can have long memory with particular choices of mixing distribution, such as a Gamma distribution with a shape parameter smaller than one. Alternatively, long memory can be introduced directly into the asset price equation (1), as in Gander and Stephens [15]. We propose efficient Bayesian inference methods for asset models with risk premium and leverage effects, using such continuous superpositions as volatility processes. We are not aware of other approaches to formal likelihood-based inference with these models. Standard Markov chain Monte Carlo (MCMC) methods usually perform poorly for this class of models 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. [23] as well as a modification of the dependent thinning sampler of Griffin and Steel [17] for inference with these models. Matlab code is freely available (see the online version of this paper, at doi:10.1016/j.csda.2009.????).

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 and jumps in prices is briefly described in Section 4, and Section 5 comments on the priors used. Inference methods based on MCMC samplers are described in some detail in Section 6, which also contains a comparison between two different strategies. Section 7 provides applications to different samples from the S&P 500 stock price index and an exchange rate. Finally, the last section contains some concluding remarks.

2. Ornstein-Uhlenbeck processes

The non-Gaussian Ornstein-Uhlenbeck (OU) process (see Barndorff-Nielsen and Shephard [2]) 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 appearance of $\lambda$ in the timing of the Lévy processes implies that the marginal distribution of $\sigma^2(t)$ does not depend on $\lambda$. 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 \( \lambda \). The Lévy density of \( \sigma^2(t) \) (which will be called \( u \)) can be related to the Lévy density of \( z(t) \) (which will be called \( w \)) by the functional relationship
\[ w(x) = -u(x) - xu'(x). \]

The tail mass integral \( W^+(x) = \int_x^{\infty} w(y) \, dy \) is an important function for the simulation of Lévy processes (see e.g. Ferguson and Klass [11], Rosinski [24]) and it is related to \( u \) by the formula
\[ W^+(x) = xu(x). \]

This allows us to specify the BDLP \( z \) 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 considered in the literature (see Gander and Stephens [14, 15]). A practically interesting example is a Gamma distribution, i.e. \( \sigma^2(t) \sim \text{Ga}(\nu, \gamma) \) with probability density function
\[ p(x) = \frac{\gamma^\nu x^{\nu-1} \exp\{-\gamma x\}}{\Gamma(\nu)}, \]
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 that the resulting Gamma OU process can be represented as a shot-noise process (see Bondesson [4]) in the following way
\[ \sigma^2(t) = \sum_{i=1}^{\infty} \mathbb{I}(\tau_i < t) J_i \exp\{-\lambda(t - \tau_i)\}, \]
where \((\tau, J)\) follow a Poisson process with intensity \( \nu \gamma \exp\{-\gamma J\} \). We can interpret \( J \) as jump sizes and \( \tau \) 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 \( \epsilon \) (see e.g. Gander and Stephens [15] for more details) and the intensity of the resulting Poisson process will be \( \lambda \mathbb{I}(J > \epsilon) w(J) \).

Barndorff-Nielsen and Shephard [2] comment that actual financial time series will often be fit better by a superposition of independent OU processes with different rate parameters \( \lambda \). For example, a two-component model is given by
\[ \sigma^2(t) = \sigma^2_1(t) + \sigma^2_2(t), \]
where each component process is an independent OU process as defined above with a \( \operatorname{Ga}(\nu_j, \gamma) \) 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
\[ \rho(s) = \kappa \exp(-\lambda_1 |s|) + (1 - \kappa) \exp(-\lambda_2 |s|), \]
with the weight \( \kappa = \nu_1/(\nu_1 + \nu_2) \), 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 [2] and Griffin and Steel [17].

In this paper we will focus on OU processes with Gamma marginals. For S&P 500 data (as used here) Gander and Stephens [15] 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 [15]).

3. Continuous superpositions of OU processes

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

\[
\sigma^2(t) = \int \sigma_\lambda^2(t) dF(\lambda),
\]

where \( F \) is a distribution function, which will be termed the mixing distribution. The properties of this model are studied by Barndorff-Nielsen [1]. 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 [1], is the Gamma distribution with shape \( \alpha \) and scale \( \phi \) 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 [1] 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

\[
\sigma^2(t) = \sum_{i=1}^{\infty} 1(\tau_i < t) J_i \exp\{-\lambda_i(t - \tau_i)\},
\]

where \((\tau, J, \lambda)\) follow a Poisson process with intensity \( \nu \lambda f(\lambda) \gamma \exp\{-\gamma J\} \). The model differs from (2) by introducing a jump-specific decay parameter \((\lambda_i)\), 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, $\sigma^2(t) = \int_0^t \sigma^2(u) \, du$ has the form

$$\sigma^2(t) = \sum_{i=1}^{\infty} I(0 < \tau_i < t) \frac{J_i}{\lambda_i} (1 - \exp\{-\lambda_i(t - \tau_i)\}) + \sum_{i=1}^{\infty} I(\tau_i < 0) \frac{J_i}{\lambda_i} \exp\{\lambda_i \tau_i\} (1 - \exp\{-\lambda_i t\}),$$

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

$$\sigma^2_n = \sum_{i=1}^{\infty} I((n-1)\Delta < \tau_i < n\Delta) \frac{J_i}{\lambda_i} (1 - \exp\{-\lambda_i(n\Delta - \tau_i)\}) + \sum_{i=1}^{\infty} I(\tau_i < (n-1)\Delta) \frac{J_i}{\lambda_i} (1 - \exp\{-\lambda_i\Delta\}) \exp\{-\lambda_i((n-1)\Delta - \tau_i)\}. \quad (4)$$

A single OU process $\sigma^2(t)$ has an integrated volatility which can be expressed as $A\sigma^2(0) + B$, where $A$ is a constant depending on the decay parameter $\lambda$ and $B$ is the integrated process restricted to jumps that occur in the region $(0, t)$ and independent of $\sigma^2(0)$. Therefore, calculation of $\sigma^2_n$ 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

$$\sigma^2_n = \sum_{i=1}^{\infty} I((n-1)\Delta < \tau_i < n\Delta) \frac{J_i}{\lambda_i} (1 - \exp\{-\lambda_i(n\Delta - \tau_i)\}) + \sum_{i=1}^{\infty} I(B < \tau_i < (n-1)\Delta) \frac{J_i}{\lambda_i} (1 - \exp\{-\lambda_i\Delta\}) \exp\{-\lambda_i((n-1)\Delta - \tau_i)\}. \quad (5)$$

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 [14] 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 extensions

4.1. Modelling leverage effects

Leverage effects refer to the negative correlation between returns and volatility, which was first considered by Black [3]. 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 [2] who suggest changing the drift in equation (1) so that the asset price follows

$$dy(t) = \{\mu + \beta \sigma^2(t) + \rho(z(t) - E[z(t)])\} \, dt + \sigma(t) \, dB(t),$$
which implies that the returns $y_1, \ldots, y_N$ are modelled by

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

where $\sigma_n^2$ 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 $\rho$ will associate jumps in volatility with negative returns and allow different effects of negative and positive price changes.

4.2. Modelling jumps in prices

The price of financial assets sometimes jumps and we take the approach of Eraker et al. [10] who include an independent compound Poisson process with intensity $\lambda_J$ to represent the jumps in prices and assume that the jump distribution is $N(0, \sigma_J^2)$. The model for the returns can be written as

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

where $N_t$ is a Poisson process of intensity $\lambda_J$ and $\zeta$ represents the jump sizes.

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 $(\nu/\gamma)^{-1} \sim \text{Ga}(1, 0.001)$. Informative priors are chosen for the parameters of the mixing distribution. The shape parameter of the mixing distribution $\alpha$ 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 $1/\xi_0$ (in our applications we take $\xi_0 = 1$). This stops the prior placing mass on very large values of the mean. The parameters of the drift $\mu$, the risk premium $\beta$ and the leverage effect $\rho$ are given independent vague zero-mean prior distributions with a standard deviation of 100. The priors for the jump component are taken from Eraker et al. [10] who also analyse daily data. The prior on $\sigma_J^2$ is $\text{Ga}(5, 20)$, which implies that the prior mean of $\sigma_J^2$ is 5. The prior for jump intensity will be given in terms of $b$, an approximate transformation of $\lambda_J$ as explained at the start of Subsection 6.2, and we adopt a Beta prior on $b$ with parameters 2 and 40 on daily data, denoted as $b \sim \text{Be}(2, 40)$. This implies that the prior predictive probability of a jump in a day is 0.0476.

6. Markov chain Monte Carlo inference methods

The representation of the continuous superposition model (3) is expressed in terms of a Poisson process on $(\tau, \lambda, J)$ with intensity $\nu f(\lambda) \gamma \exp\{-\gamma J\}$. If the mean of $F$ is finite, simulation methods follow more naturally if we interpret $(\tau, \lambda, J)$ as a marked Poisson process where the jump times $\tau$ follow a Poisson process with intensity $\nu \mathbb{E}[\lambda]$ and the jumps sizes, $J$, and decay rates, $\lambda$, 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 $f^*(\lambda) \propto \lambda f(\lambda)$. If the mixing distribution is $Ga(\alpha, \phi)$ then $f^*$ is the density of a $Ga(\alpha + 1, \phi)$ distribution.

MCMC methods for OU processes have been developed by Roberts et al. [23] and Griffin and Steel [17]. Standard MCMC methods are difficult to apply due to slow mixing of $\lambda$ and $\nu$. This problem can be addressed by updating the process $z$ jointly with $\lambda$ or $\nu$. Roberts et al. [23] propose a method of retrospective sampling of $z$ and Griffin and Steel [17] suggest “dependent thinning”. Whereas Roberts et al. [23] suggest the use of a reparameterisation to reduce the correlation between the data and the process, the dependent thinning of Griffin and Steel [17] 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 $J_i I(\tau_i < t < s_i) \exp\{-\lambda_i (t - \tau_i)\}$. The actual volatility $\sigma_n^2$ is now expressed as

$$\sigma_n^2 = \sum_{i=1}^{\infty} E_i J_i \int_{\max \{(n-1)\Delta, \tau_i\}}^{\min \{n\Delta, s_i\}} \exp\{-\lambda_i (u - \tau_i)\} \, du$$

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

where $E_i = I(\tau_i < n\Delta$ 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 $\sigma_n^2$ 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

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

where $\Gamma(\alpha, x) = \frac{1}{\Gamma(\alpha)} \int_{x}^{\infty} u^{\alpha-1} \exp\{-u\} \, du$. 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

We follow Eraker et al. [10] by assuming that only zero or one jump is possible in any time interval (the probability of more jumps in any time interval will be small if \( \lambda_j \) is small) and represent the number of jumps in interval \( ((n-1)\Delta, n\Delta) \) as \( \eta_n \), with \( p(\eta_n = 1) = b \). If we choose a Gamma marginal distribution then \( E[\eta_n] = \frac{n}{b} \xi \Delta \), where \( \xi = \alpha/\phi \) is the mean of the mixing distribution. So the model that we fit can be expressed in the following way for each observation \( n = 1, \ldots, N \)

\[
y_n \sim \mathcal{N}(\mu \Delta + \beta \sigma_n^2 + \rho \left( z_n - \frac{\nu}{\gamma} \xi \Delta \right), \sigma_n^2 + \eta_n \sigma_J^2),
\]

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

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

where

\[
k \sim \text{Pn}(\nu \xi (n\Delta - B)), \quad \tau_1, \tau_2, \ldots, \tau_k \sim \text{U}([B, N\Delta]), \quad \eta_1, \eta_2, \ldots, \eta_N \sim \text{Ber}(b)
\]

\[
J_1, J_2, \ldots, J_k \overset{i.i.d.}{\sim} \text{Ga}(1, \gamma), \quad \lambda_1, \lambda_2, \ldots, \lambda_k \overset{i.i.d.}{\sim} \text{Ga}\left(\alpha + 1, \frac{\alpha}{\xi}\right).
\]

We write \( \text{U}(R) \) for the uniform distribution on the set \( R \), \( \text{Ber}(b) \) to represent a Bernoulli distribution with success probability \( b \) and \( \text{Pn}(a) \) 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

\[
p(y|\psi)p(J|\gamma)p(\lambda|\alpha, \xi)p(\tau|\nu)p(\beta|\rho)p(\xi|\gamma)p(\nu|\nu)p(\alpha)p(p(\beta)p(p(\rho)p(\sigma_J^2)p(p|\eta)b)p(b),
\]

where \( \psi = (J, \lambda, \tau, \mu, \beta, \rho, \xi, \nu, \gamma, \sigma_J^2, \eta), y = (y_1, \ldots, y_N)' \), \( \eta = (\eta_1, \ldots, \eta_N)' \), \( J, \tau \) and \( \lambda \) 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, \( \psi' \) will refer to a proposed value of \( \psi \) 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 \( \epsilon_i \sim \text{N}(0, \sigma_J^2) \), which is accepted with probability

\[
\min\left\{1, \frac{p(y|\psi')p(J'|\gamma)p(\lambda|\alpha, \xi)p(\tau|\nu)p(\beta|\rho)p(\sigma_J^2)p(\eta|b)p(b)}{p(y|\psi)p(J|\gamma)p(\lambda|\alpha, \xi)p(\tau|\nu)p(\beta|\rho)p(\sigma_J^2)p(\eta|b)p(b)}\right\}.
\]

The variance \( \sigma_J^2 \) is chosen to give an average acceptance rate of between 0.2 and 0.3.
6.2.2. Updating $\lambda$

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, $\lambda'_j$, is accepted with probability

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

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 $\tau$

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 $\epsilon_i \sim N(0, \sigma^2)$ 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 $\sigma_\tau$ 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 MCMC [16]. 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 $J', \lambda'$ and $\tau'$ where $J'_i = J_i$, $\lambda'_i = \lambda_i$ and $\tau'_i = \tau_i$ for $1 \leq j \leq k$ and $J'_{k+1} \sim \text{Ga}(1, \gamma)$, $\lambda'_{k+1} \sim \text{Ga}\left(\alpha + 1, \frac{\alpha}{\xi}\right)$ and $\tau'_{k+1} \sim U([B, N \Delta])$. The values $k', J', \lambda'$ and $\tau'$ are accepted with probability

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

If $k' = k - 1$, then a value $j$ is drawn at random from $\{1, 2, \ldots, k\}$ and new vectors $J', \lambda'$ and $\tau'$ 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

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

6.2.5. Updating $\mu$, $\beta$ and $\rho$

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

$$N \left( (\Lambda + X'\Sigma^{-1}X)^{-1} \Sigma^{-1}X' y, (\Lambda + X'\Sigma^{-1}X)^{-1} \right),$$
where $\Sigma$ is a diagonal matrix with elements $\Sigma_{jj} = \sigma_i^2 + \eta_i \sigma_j^2$, $\Lambda$ is the prior precision of $\theta$ and $X$ is a $N \times 3$-dimensional matrix with $j$-th row equal to

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

6.2.6. Updating $\alpha$

The full conditional distribution of $\alpha$ is proportional to

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

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

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

which is the density of a Generalized Inverse Gaussian (GIG) distribution. Efficient methods for the simulation from this distribution are described by Devroye [8]. If we denote by $f$ the target density and $g$ is the rejection envelope, then $\max(f/g) = 0$. 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 $\xi$

The full conditional distribution of $\xi$ 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 $\xi$ through $E[z_n] = \frac{\xi}{\gamma} \Delta$ which is included to model the leverage effect. This is not likely to change a lot with $\xi$ 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 $\nu$

We can directly apply the retrospective method of Roberts et al. [23]. Suppose we propose to move from $\nu$ to $\nu'$ using transition kernel $q(\nu, \nu')$. Then we propose a new process in the following way. If $\nu' > \nu$, then simulate $m \sim \text{Pn}((\nu' - \nu)\xi(N\Delta - B))$ 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 $J_i' \sim \text{Ga}(1, \gamma)$, $\tau_i' \sim U([B, N\Delta])$ and $\lambda_i' \sim \text{Ga} \left( \alpha + 1, \frac{2}{\tau} \right)$. The new process is formed by taking the superposition of the current values $(\tau, J, \lambda)$ with the new values $(\tau', J', \lambda')$. The acceptance probability of $\nu'$ is calculated by taking the ratio

$$
p(\nu | \nu') q(\nu', \nu) \over p(\nu | \psi) q(\nu, \nu')
$$

(5)

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

The dependent thinning method of Griffin and Steel [17] 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 $(\tau', J', \lambda')$ is defined in the following way. If $\nu' > \nu$, they simulate $m \sim \text{Pn}((\nu' - \nu)\xi(N\Delta - B))$ and propose $J_i' = J_i + \log(\frac{\nu}{\nu'}) / \gamma$, $\tau_i' = \tau_i$ and $\lambda_i' = \lambda_i$ for $1 \leq i \leq k$ and $J_i' \sim \text{Ga}(1, \gamma)$ where $J_i' < \log \left( \frac{\nu}{\nu'} \right) / \gamma, \tau_i' \sim U([B, N\Delta]), \lambda_i' \sim \text{Ga} \left( \alpha + 1, \frac{2}{\tau} \right)$ 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 $\lambda_i' = \lambda_i$ 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, $\lambda_i$ differ between jumps. Jumps with smaller values of $\lambda_i$ 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 $\frac{1}{\gamma}$. This suggests, when $\nu' > \nu$ proposing $J_i' = J_i + \frac{2\Delta^*}{\gamma}$ (where $\Delta^*$ 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 $(0, \infty) \times (0, \frac{\Delta^*}{\gamma}) \times (B, N\Delta)$. The process can be simply simulated in the following way: (if the mixing distribution is $\text{Ga} \left( \alpha, \frac{2}{\tau} \right)$)

1. The number of new jumps $m$ is Poisson distributed with mean $\nu' \xi \left[ 1 - \left( 1 + a^* \frac{\xi}{\alpha} \right)^{-(\alpha + 1)} \right](N\Delta - B)$.

2. $\tau_{k+1}', \tau_{k+2}', \ldots, \tau_{k+m}' \sim \text{U}([B, N\Delta])$.

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 $\lambda_i' \sim \text{Ga} \left( \alpha + 1, \frac{2}{\tau} \right)$ and $J_i'$ from an $\text{Ga}(1, \gamma)$ truncated to the region $(0, \frac{\Delta^* \lambda_i'}{\gamma})$. These values are accepted with probability $1 - \exp\{-a^* \lambda_i'\}$.

The acceptance probability of $(\nu', J', \tau', \lambda')$ is given by

$$A = \left( \frac{\nu'}{\nu} \right)^k \exp \left\{ -a^* \sum_{i=1}^{k} \lambda_i \right\} \exp \left\{ (N\Delta - B) \xi \left[ \nu' - \nu \left( 1 + a^* \frac{\xi}{\alpha} \right)^{-(\alpha + 1)} \right] \right\}.$$

where

$$A = \left( \frac{\nu'}{\nu} \right)^k \exp \left\{ -a^* \sum_{i=1}^{k} \lambda_i \right\} \exp \left\{ (N\Delta - B) \xi \left[ \nu' - \nu \left( 1 + a^* \frac{\xi}{\alpha} \right)^{-(\alpha + 1)} \right] \right\}.$$
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. [5] 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 \( a^* \) to guarantee \( A = 1 \) for all parameter values. Therefore we choose the value of \( a^* \) to guarantee that \( \mathbb{E}[A] = 1 \) for all values of the parameters where the expectation is taken with respect to \( \lambda_1, \lambda_2, \ldots, \lambda_k \). Then
\[
\mathbb{E}[A] = \left( \frac{\nu}{\nu'} \right)^k \left( 1 + a^* \xi \right)^{-k(\alpha+1)} \exp \left\{ (N\Delta - B)\xi \left[ \nu - \nu' \left( 1 + a^* \xi \right)^{-\alpha} \right] \right\},
\]
and \( \mathbb{E}[A] = 1 \) when
\[
a^* = \frac{\alpha}{\xi} \left( \left( \frac{\nu}{\nu'} \right)^{1/(\alpha+1)} - 1 \right).
\]

6.2.9. Updating \( \gamma \)

The full conditional distribution of \( \gamma \) 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 \( \gamma \) through \( \mathbb{E}[z_n] = \nu \xi \Delta \) 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.

6.2.10. Updating \( \eta_n \)

The parameters \( \eta_1, \eta_2, \ldots, \eta_N \) are updated from their full conditional distribution. The full conditional distribution of \( \eta_n \) is
\[
p(\eta_n = 0) \propto \frac{1 - b}{\sqrt{\sigma_n^2}} \exp \left\{ -\frac{1}{2} \sigma_n^{-2} (y_n - \mu_n)^2 \right\}, \quad p(\eta_n = 1) \propto \frac{b}{\sqrt{\sigma_n^2 + \sigma_J^2}} \exp \left\{ -\frac{1}{2} \left( \sigma_n^2 + \sigma_J^{-2} \right)^{-1} (y_n - \mu_n)^2 \right\}
\]
where \( \mu_n = \mu \Delta + \beta \sigma_n^2 + \rho \left( z_n - \frac{\xi}{\gamma} \xi \Delta \right) \).

6.2.11. Updating \( b \)

The full conditional distribution of \( b \sim \text{Be} \left( 2 + \sum_{n=1}^{N} \eta_n, 40 + N - \sum_{n=1}^{N} \eta_n \right) \).

6.2.12. Updating \( \sigma_J^2 \)

To update \( \sigma_J^2 \), we first simulate jumps \( \zeta_n \) if \( \eta_n = 1 \) from the distribution \( \mathcal{N} \left( \frac{\sigma_n^2 (y_n - \mu_n)}{\sigma_n^2 + \sigma_J^2}, \frac{1}{\sigma_n^2 + \sigma_J^2} \right) \) then \( \sigma_J^2 \sim \text{Ga} \left( 5 + \frac{1}{2} \sum_{n=1}^{N} \eta_n, 20 + \frac{1}{2} \sum_{\{n|\eta_n=1\}} \zeta_n \right) \).
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 $\nu$ in the context of the first 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 $\nu$ 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. Examples

7.1. Application to a stock index

We consider daily observations of the Standard and Poor’s 500 index of the New York Stock Exchange from September 26, 1980 to June 6, 2000. Thus, the observed $y$ consists of the 5000 data points shown in Figure 2. This is almost the same sample as used in Li et al. [19] and extends the period covered by Griffin and Steel [17].

The parameter estimates for a one-component OU model, two-component discrete superposition and continuous superposition with gamma mixing distribution with and without jumps 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 the continuous models. The posterior median of $E[\lambda]$ changes very little between the four models. However, the width of the 95% credible interval increases as we move from the
Table 1: Posterior estimates of parameters for one and two components and the continuous superposition with the S & P 500 data between September 26, 1980 and June 6, 2000: posterior median with 95% credible interval in brackets

<table>
<thead>
<tr>
<th>parameter</th>
<th>one-component</th>
<th>two-component</th>
<th>continuous</th>
<th>continuous with jumps</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E(\sigma^2(t))$</td>
<td>0.80 (0.67, 0.98)</td>
<td>0.93 (0.65, 1.31)</td>
<td>0.99 (0.55, 1.96)</td>
<td>0.91 (0.47, 1.90)</td>
</tr>
<tr>
<td>SD[$\sigma^2(t)$]</td>
<td>0.45 (0.37, 0.56)</td>
<td>0.64 (0.47, 0.89)</td>
<td>0.70 (0.49, 1.06)</td>
<td>0.76 (0.51, 1.24)</td>
</tr>
<tr>
<td>$w$</td>
<td>0.84 (0.69, 0.92)</td>
<td>0.016 (0.011, 0.022)</td>
<td>0.004 (0.002, 0.006)</td>
<td>0.002 (0.001, 0.005)</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.082 (0.045, 0.150)</td>
<td>0.026 (0.12, 0.59)</td>
<td>0.017 (0.008, 0.037)</td>
<td>0.015 (0.006, 0.037)</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.016 (0.011, 0.022)</td>
<td>0.004 (0.002, 0.006)</td>
<td>0.012 (0.008, 0.037)</td>
<td>0.002 (0.006, 0.037)</td>
</tr>
<tr>
<td>$\sigma^2(t)$</td>
<td>0.006 (-0.034, 0.044)</td>
<td>0.013 (-0.028, 0.053)</td>
<td>0.045 (-0.012, 0.103)</td>
<td>0.034 (-0.026, 0.094)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.054 (-0.003, 0.112)</td>
<td>0.046 (-0.013, 0.106)</td>
<td>0.045 (-0.012, 0.103)</td>
<td>0.034 (-0.026, 0.094)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-4.56 (-6.03, -3.39)</td>
<td>-3.09 (-4.36, -2.08)</td>
<td>-3.75 (-4.00, -1.80)</td>
<td>-2.47 (-3.76, -1.50)</td>
</tr>
<tr>
<td>$b$</td>
<td>0.028 (0.006,0.076)</td>
<td>0.028 (0.006,0.076)</td>
<td>0.028 (0.006,0.076)</td>
<td>0.028 (0.006,0.076)</td>
</tr>
<tr>
<td>$\sigma^2_J$</td>
<td>2.15 (1.28, 4.59)</td>
<td>2.15 (1.28, 4.59)</td>
<td>2.15 (1.28, 4.59)</td>
<td>2.15 (1.28, 4.59)</td>
</tr>
</tbody>
</table>

one-component to the two-component model and to the continuous models, which is in line with the increasing flexibility on $\lambda$. In all models there seems to be strong evidence for a leverage effect. The estimates of $\rho$ also change markedly between the four models. The one-component model has a much larger (in absolute value) posterior median estimate of $\rho$ 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. In particular, the more flexible models can have larger means and variances of $\sigma^2(t)$, which offers increased mass to larger volatilities. Therefore, in these models large movements in the returns can be explained by larger volatilities and we do not require such high amounts of leverage. The continuous model shows clear evidence in favour of long memory. The posterior median estimate of $\alpha$ 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.

The introduction of jumps has a small effect on the inference since $b$ (as defined in the beginning of Subsection 6.2) is estimated to be 0.028. The inference about jumps is shown in Figure 3 and shows that some days have a large probability of including a jump with values around 0.8. The figure also shows the relationship between the returns and the jump probabilities. These show that the largest absolute returns are not indicated as jumps but there are some moderate returns which are assigned to jumps in the price process.

Posterior estimates of the volatility are shown in Figure 4 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 5 which shows posterior median estimates of the volatility between January 22, 1987 and July 22, 1988 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 oth-
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 on January 11, 1988 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 November 27, 1987.

The autocorrelation functions for $\sigma^2(t)$ are illustrated in Figure 6 and show a large difference between the estimates. As we might expect, the continuous 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 the lag 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. We use a discrete time stochastic volatility (SV) model where the volatility is modelled by an AR(1) process using the MCMC methods of Kim et al. [18] as a baseline. 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 [7] for a discussion of the application of particle filtering methods to discrete superposition models). We use the method introduced by Carpenter et al. [6]. The results for the five models are shown in Table 2. All the OU models massively outperform the discrete time SV model. The two-component and continuous superposition models also 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. The continuous model with and without jumps have similar performance and we will concentrate on the continuous model without jumps for the rest of the analysis with these data. Figure 7 shows the difference between the running LPS for the continuous 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 [17] compare models using marginal likelihoods estimated by the modified harmonic mean estimators of Newton and Raftery [20] but we find that these methods give unreliable results for the continuous superposition model. Other methods for evaluating marginal likelihoods, as described in e.g. Diciccio et al. [9] and in Frühwirth-Schnatter [12] in the context of mixture models, might prove more reliable but are not that straightforward to implement for these models.
Another measure of fit is given by the models’ ability to reproduce the empirical properties of the data. Figure 8 shows the empirical 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.

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 generate a closer fit to the empirical variance at the longest frequency. The two em-
Table 3: S & P 500 data between September 26, 1980 and June 6, 2000: Empirical and theoretical moments evaluated at the posterior median of the parameters for the two-component and continuous superposition models for three values of $\Delta$. The bracketed empirical moments exclude the week starting with “Black Monday”.

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

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 [17] 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.

We test the stability of the estimates over time by analysing the Standard and Poors index between April 28, 1989 and February 26, 2009. The samples are overlapping but the new sample excludes “Black Monday” and includes a period of high volatility in the final 18 months of observations. Figure 9 shows the data and the posterior estimates of the volatility $\sigma_n$ and the au-

![Figure 9: The returns of the S & P 500 index between April 28, 1989 and February 26, 2009 (in panel (a)) with posterior estimates of: (b) $\sigma_n$ and (c) the autocorrelation of $\sigma^2(t)$ (solid lines are median values and dotted lines form point-wise 95% credible intervals)](image-url)
Table 4: Posterior estimates of parameters the continuous superposition model with and without jumps for the S & P 500 index between April 28, 1989 and February 26, 2009: posterior median with 95% credible interval in brackets

tocorrelation function of $\sigma^2(t)$ for the continuous superposition model with jumps (the results are very similar for the continuous model without jumps). The results show periods of high volatility between 1997 and 2002 and rapidly increasing volatility from the middle of 2006. Table 4 shows posterior estimates of the parameters of the model. The results are similar to the results for the earlier period. The mean and standard deviation of volatility are estimated slightly higher but the estimate of $\alpha$ is very similar. However, the leverage effect $\rho$ is estimated to be a lot lower.

The run times for the sampler are of the order of 40 hours for 5000 data points as in this subsection using Matlab code (freely available at doi:10.1016/j.csda.2009.??..??) on a 3Ghz Pentium 4 processor. Execution times would be much shorter for compiled languages such as C or C++.

### 7.2. Exchange rate data

As an alternative type of financial data, we consider daily exchange rates of the Swiss franc to the US dollar, as previously analysed by Frühwirth-Schnatter and Sogner [13]. The data cover the period from January 2, 1985 until December 13, 1999 with a total of 3778 observations. We fit a continuous superposition model with jumps to the data and set $\mu = \beta = \rho = 0$ to make our results comparable to the results of Frühwirth-Schnatter and Sogner [13]. The results in Table 5 show some differences with respect to the stock price index data. There is weaker evidence of long memory with the 95% credible interval of $\alpha$ extending beyond 1 (in fact, the posterior probability that $\alpha < 1$ is 0.82) and there is a much higher rate of jumps in price. Figure 10 illustrates that volatility is more stable over time for these data and inference on volatility levels seems roughly comparable to the finite superposition model of Frühwirth-Schnatter and Sogner [13].

Table 5: Posterior estimates of parameters in the continuous superposition with the Swiss franc - US dollar exchange rate between January 2 1985 until December 13 1999: posterior median with 95% credible interval

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Continuous</th>
<th>Continuous with Jumps</th>
</tr>
</thead>
<tbody>
<tr>
<td>$E[\sigma^2(t)]$</td>
<td>0.54 (0.28, 1.18)</td>
<td>0.38 (0.25, 0.68)</td>
</tr>
<tr>
<td>$SD[\sigma^2(t)]$</td>
<td>0.53 (0.18, 2.87)</td>
<td>0.008 (0.003, 0.028)</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.10 (0.050, 0.181)</td>
<td>1.63 (1.13, 2.58)</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>2.21 (1.22, 5.07)</td>
<td>1.23 (0.68, 2.46)</td>
</tr>
<tr>
<td>$\mu$</td>
<td>1.08 (0.75, 1.63)</td>
<td>1.19 (0.81, 1.86)</td>
</tr>
<tr>
<td>$\beta$</td>
<td>0.025 (0.01, 0.052)</td>
<td>0.019 (0.009, 0.038)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>-0.026 (-0.063, 0.013)</td>
<td>-0.029 (-0.069, 0.009)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>0.013 (0.002, 0.046)</td>
<td>2.21 (1.22, 5.07)</td>
</tr>
</tbody>
</table>
Figure 10: Returns of Swiss franc - US Dollar exchange rate between January 2 1985 until December 13 1999 (in panel (a)) with posterior estimates of: (b) $\sigma_n$ and (c) the autocorrelation of $\sigma^2(t)$ (solid lines are median values and dotted lines form point-wise 95% credible intervals).

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, leverage and price jumps, 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. [23] and we propose a new version of the dependent thinning methods of Griffin and Steel [17]. 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.

One interesting aspect of OU models of volatility is that they model data by two separate parts: 1) a deterministic decay component and 2) a underlying Poisson process. More elaborate models can be defined by changing these components. For example, allowing the intensity of the Poisson process to change over time would define models where the marginal distribution of volatility is time-dependent. An interesting difference to standard change point analysis is that the distribution will change smoothly from one distribution to the next. Similar models have been explored in other stochastic volatility modelling frameworks by Perello et al. [22]. Developing models in this direction will be an area of future work.

Acknowledgements

We thank the Guest Editor and two referees for helpful comments, and we are grateful to Sylvia Frühwirth-Schnatter for making the exchange rate data available.
References


