Inference With Non-Gaussian Ornstein-Uhlenbeck Processes for Stochastic Volatility*

J.E. Griffin and M.F.J. Steel†

March 4, 2003

Institute of Mathematics and Statistics
University of Kent at Canterbury, CT2 7NF, UK

Abstract

Continuous-time stochastic volatility models are becoming an increasingly popular way to describe moderate and high-frequency financial data. Recently, Barndorff-Nielsen and Shephard (2001a) proposed a class of models where the volatility behaves according to an Ornstein-Uhlenbeck process, driven by a positive Lévy process without Gaussian component. These models introduce discontinuities, or jumps, into the volatility process. They also consider superpositions of such processes and we extend that to the inclusion of a jump component in the returns. In addition, we allow for leverage effects and we introduce separate risk pricing for the volatility components. We design and implement practically relevant inference methods for such models, within the Bayesian paradigm. The algorithm is based on Markov chain Monte Carlo (MCMC) methods and we use a series representation of Lévy processes. MCMC methods for such models are complicated by the fact that parameter changes will often induce a change in the distribution of the representation of the process and the associated problem of overconditioning. We avoid this problem by dependent thinning methods. An application to stock price data shows the models perform very well, even in the face of data with rapid changes, especially if a superposition of processes with different risk premiums and a leverage effect is used.

JEL classification: C11; C22; G12

KEYWORDS: Bayesian inference, Leverage effect, Lévy process, Markov chain Monte Carlo, Risk premium, Return jumps, Stock price, Superposition, Volatility jumps

*We acknowledge very useful comments by Neil Shephard and Gareth Roberts.
†Email: M.F.Steel@ukc.ac.uk

1 Introduction

In this paper we provide an exact Bayesian inference method for certain Non-Gaussian models based on Ornstein-Uhlenbeck processes, recently introduced into the finance literature by Barndorff-Nielsen and Shephard (2001a). These processes allow option prices for empirically reasonable models to be computed analytically (see Nicolato and Venardos, 2003). The class of models considered here are continuous-time stochastic volatility models, which have separate stochastic specifications for the observables and the latent volatility process. Thus, they differ from the more usual continuous-time models in finance, based on a single stochastic differential equation, such as those pioneered by Black and Scholes (1973) and Merton (1973) for option pricing and by Vasicek (1977) and Cox, Ingersol and Ross (1985) for modelling the term structure of interest rates. See Sundaresan (2000) for a comprehensive survey. In order to capture the empirical behaviour of financial time series, the more recent literature has typically moved towards multi-factor models, often by allowing the volatility to be driven by an independent random process, such as in Heston (1993). In addition, jumps in the price process were introduced in an effort to improve the empirical performance of these models (as in *e.g.* Bates, 2000 and Pan, 2002). Such stochastic volatility models are often complicated and do not allow for analytic solutions for pricing derivatives. An exception is provided by Duffie, Pan and Singleton (2000), who derive closed-form expressions for securities prices in the context of affine jump diffusion processes. Chernov, Gallant, Ghysels and Tauchen (2003) provide a recent survey of stochastic volatility models and discuss their impact on option pricing.

Various methods for likelihood-based statistical inference with continuous-time models driven by Brownian motion have recently been proposed: simulation-based maximum likelihood methods integrate out unobserved states at intermediate points between observations. Key references are Pedersen (1995), Aït-Sahalia (2002) and Durham and Gallant (2002), who also provide an insightful survey and comparison of various existing methods. Bollerslev and Zhou (2002) and Barndorff-Nielsen and Shephard (2002) use realized volatility, *i.e.* they exploit the existence of high-frequency intraday data to directly estimate the moments of the integrated volatility. Yu and Phillips (2001) propose a time-change transformation to induce Gaussian behaviour in the context of non-Gaussian interest rate models. Bayesian methods were proposed by Eraker (2001), Elerian, Chib and Shephard (2001) and Jones (2001), who use Markov chain Monte Carlo (MCMC) methods based on data augmentation with simulated high-frequency observations. Roberts and Stramer (2001) point out that such data augmentation schemes can lead to poor convergence properties in MCMC algorithms and propose a reparameterization of the missing data.

Here we develop a Bayesian inference procedure for continuous-time stochastic volatility models where the unobserved volatility has a non-Gaussian driving process. We base ourselves on the type of models introduced by Barndorff-Nielsen and Shephard (2001a), which we extend by adding a jump component process in the returns. In addition, we introduce different risk premia for different volatility components. We also allow for component-specific leverage effects. Barndorff-Nielsen and Shephard (2001a, 2002) suggest a linear form for the leverage effects. In order to increase flexibility of the models, we suggest including powers of the relevant processes.

Even allowing for these modifications, this class of models still lead to a simple model for the observ-

able returns, which makes statistical inference quite straightforward conditional on the volatility process. As a consequence, we can base inference on simulating the actual volatility process, rather than on simulating intermediate observations. We simulate the volatility process through a series representation of Lévy processes and use MCMC methods to design the algorithm. Inference for the models considered here is complicated by the fact that the parameters and the latent volatility process can be highly correlated *a posteriori*. The usual Gibbs updating of the parameters conditioned on the process will lead to extremely slowly mixing algorithms. This problem is generally referred to as overconditioning and must be addressed by updating the process jointly with the parameters. If the parameters are updated using a Metropolis-Hastings sampler, the proposed process must be in accordance with both the parameters and the data in order to obtain reasonable acceptance probabilities. We introduce an approach called dependent thinning where the changes in the processes are restricted to relatively small jumps. Roberts, Papaspiliopoulos and Dellaportas (2001) take an alternative approach. They suggest a reparameterisation which reduces the correlation between the data and the process. The reparameterised process is then proposed only in accordance with the parameters.

We also consider the issue of prior elicitation and conduct an empirical application on a set of observations of the S&P 500 stock price index. This illustrates the feasibility of our approach and it reveals the empirical support for the various models considered in this particular context.

2 The Model

The usual continuous-time stochastic volatility model in finance is driven by Brownian motion with an instantaneous volatility process, $\sigma^2(t)$. The logarithm of an asset price, say[1] $y^\star$, is often assumed to be determined by the following stochastic differential equation:

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

where $B(t)$ is Brownian motion, the drift parameter μ and the risk premium β are defined on $\Re$, and $\sigma^2(t)$ is a latent instantaneous volatility which is assumed to be stationary and independent from $B(t)$. This will serve as our initial model. The implication of (1) is that aggregate returns over a time interval of length Δ, say, given by

$$y_i = \int_{(i-1)\Delta}^{i\Delta} \mathrm{d}y^\star(t) = y^\star(i\Delta) - y^\star\{(i-1)\Delta\} \tag{2}$$

and observed at time $i = 1, \ldots, T$, will have independent Normal distributions

$$y_i \sim \mathrm{N}(\mu\Delta + \beta\sigma_i^2, \sigma_i^2), \tag{3}$$

where $\sigma_i^2 = \int_{(i-1)\Delta}^{i\Delta} \sigma^2(u)\, du$, which is called the actual volatility in Barndorff-Nielsen and Shephard (2002). Since $\sigma^2(t)$ is a stochastic process, the marginal distribution of y_i will be a location-scale

[1]We follow the notational convention of Barndorff-Nielsen and Shephard (2001a) to indicate integrated processes with a superscript $\star$

mixture of normals. The latter aims at capturing the empirical evidence of volatility clustering, skewness and thick tails, which become thinner for large Δ. Clearly, as Δ becomes large, σ_i^2 will constitute a large area under the process of instantaneous volatility and, given the ergodicity of the latter, σ_i^2 will eventually tend to a constant $\times\ \Delta$ and the marginal distribution of y will tend to a Normal. Note that the distribution in (3) is exact, and will be used to construct the likelihood by integrating over the volatilities σ_i^2. The flexibility needed to adequately model financial series will be provided by a combination of the underlying stochastic process on the volatilities and more complicated specifications of the drift function in (1), leading to a more general mean function in (3).

Possible choices for the volatility process are geometric Brownian motion on the instantaneous volatility as in Hull and White (1987) and a Gaussian Ornstein-Uhlenbeck process on $\log \sigma^2(t)$ as in Melino and Turnbull (1990), whereas Meddahi and Renault (2003) focus on a constant elasticity of volatility process. All these processes are driven by Brownian motion. Here we shall follow the approach outlined in Barndorff-Nielsen and Shephard (2001a), who use, instead, a non-Gaussian Ornstein-Uhlenbeck volatility process, which leads to an easy and exact treatment of integrated volatility. The latter is an important quantity for deriving analytic option pricing solutions (see *e.g.* Pastorello, Renault and Touzi, 2000).

2.1 Ornstein-Uhlenbeck Volatility Processes

We consider the stochastic volatility process suggested by Barndorff-Nielsen and Shephard (2001a), which is given by the solution to the following differential equation:

$$\mathrm{d}\sigma^2(t) = -\lambda\sigma^2(t)\mathrm{d}t + \mathrm{d}z(\lambda t), \tag{4}$$

where $\lambda \in \Re_+$ and $z(t)$ is a purely non-Gaussian Lévy process (see *e.g.* Sato, 1999) with positive increments (jumps) and with $z(0) = 0$. The resulting process in (4) is an Ornstein-Uhlenbeck process restricted to positive values. The time index of the Lévy process in (4) is chosen so that the marginal distribution of $\sigma^2(t)$ is unchanged by the parameter λ, which controls both the rate at which jumps in volatility occur and the rate at which the process decays.

Barndorff-Nielsen and Shephard (2001a) show that any self-decomposable[2] marginal distribution for $\sigma^2(t)$ can be described by the stationary Ornstein-Uhlenbeck process in (4) in combination with some Lévy process $z(t)$. The Lévy measure of the Lévy-Khintchine representation of $z(t)$ is denoted by W, which has no atom at zero and satisfies

$$\int_{\Re_+} \min\{1, x^2\} W(\mathrm{d}x) < \infty, \tag{5}$$

and is assumed throughout to admit a density w. The relationship between the Lévy density w of $z(t)$ and the Lévy density u of $\sigma^2(t)$ is given by

$$w(x) = -u(x) - xu'(x), \tag{6}$$

[2]A random variable has a self-decomposable distribution if its characteristic function ϕ satisfies $\phi(\zeta) = \phi(c\zeta)\phi_c(\zeta)$ for all $\zeta \in \Re$ and all $c \in (0,1)$ and for some family of characteristic functions $\{\phi_c : c \in (0,1)\}$.

where u is assumed to be differentiable with derivative u'.

The jumps in $z(t)$, the so-called "background driving Lévy process", will induce discontinuities[3] in the instantaneous volatility process for $\sigma^2(t)$, but the integrated volatility $\sigma^{2\star}(t) = \int_0^t \sigma^2(u)\mathrm{d}u$ always has continuous sample paths. Thus, the resulting process for $y^\star(t)$ in (1) (the log asset price) also has continuous sample paths.

The analytical tractability of the process in (4) is illustrated by the following simple expression for integrated volatility

$$\sigma^{2\star}(t) = \lambda^{-1}\{z(\lambda t) - \sigma^2(t) + \sigma^2(0)\}. \tag{7}$$

Barndorff-Nielsen and Shephard (2001a) show that these models are arbitrage-free and lead to simple analytical expressions for option pricing (see also Nicolato and Venardos, 2003). Similar results are quite rare in the literature; Duffie *et al.* (2000), which provides pricing solutions for affine jump diffusions, is a notable exception.

Assuming the marginal distribution of $\sigma^2(t)$ has a finite variance, the autocorrelation function of the process is given by $r(s) = \exp(-\lambda|s|)$, just like the constant elasticity of volatility processes. In the case where β in (1) is zero (no risk premium) we can also derive the dependence structure of the discrete-time squared returns, which is given by

$$\mathrm{cor}(y_i^2, y_{i+s}^2) = c\exp\{-\lambda\Delta(s-1)\}, \quad s > 0, \tag{8}$$

where $c \in [0, 1]$ depends on λ, Δ and the marginal mean and variance of the instantaneous volatility process (see Barndorff-Nielsen and Shephard, 2001a, Example 4).

2.2 Superposition of Processes

The relative simplicity of the class of models outlined above implies some limitations. In particular, the implied dependence structure of squared returns follows the simple ARMA(1,1) pattern outlined in (8). Barndorff-Nielsen and Shephard (2001a) propose to capture more realistic dependence structures by combining independent Ornstein-Uhlenbeck processes with different rate parameters λ. A superposition of processes $\sigma_1^2(t), \ldots, \sigma_m^2(t)$ defines a new process $\sigma^2(t)$ through

$$\sigma^2(t) = \sum_{j=1}^{m} \sigma_j^2(t), \tag{9}$$

where each component process $\sigma_j^2(t)$ is generated by (4) with its own specific λ_j and driving process $z_j(t)$ while the latter are independent across components. Many analytical results from the simple model ($m = 1$) carry over to this more general superposition. In particular, the integrated volatility for this

[3]Clearly, the instantaneous volatility is subject to positive jumps and decays exponentially in between jumps.

process, $\sigma^{2\star}(t)$, is

$$\begin{aligned}\sigma^{2\star}(t) &= \int_0^t \sigma^2(u)du = \sum_{j=1}^m \sigma_j^{2\star}(t) \\ &= \sum_{j=1}^m \lambda_j^{-1}\{z_j(\lambda_j t) - \sigma_j^2(t) + \sigma_j^2(0)\},\end{aligned} \tag{10}$$

and the discrete time (or actual) volatility σ_i^2 can be derived as

$$\begin{aligned}\sigma_i^2 &= \sigma^{2\star}(i\Delta) - \sigma^{2\star}\{(i-1)\Delta\} \\ &= \sum_{j=1}^m \sigma_{ij}^2.\end{aligned} \tag{11}$$

In addition, assuming that each of the processes $\sigma_j^2(t)$ has mean ξ_j, variance ω_j^2 and autocorrelation function $r_j(s) = \exp(-\lambda_j|s|)$, the autocorrelation function of $\sigma^2(t)$ constructed from the superposition in (9) takes the simple form

$$r(s) = \sum_{j=1}^m w_j \exp(-\lambda_j|s|) \tag{12}$$

where the positive weights w_j add up to one and are given by

$$w_j = \frac{\omega_j^2}{\sum_{j=1}^m \omega_j^2}.$$

This induces the same finite mixture behaviour for the autocorrelations of the squared returns, which in the absence of a risk premium is given by

$$\operatorname{cor}\left(y_i^2, y_{i+s}^2\right) = \sum_{j=1}^m c_j \exp\{-\lambda_j\Delta(s-1)\}, \quad s > 0, \tag{13}$$

where

$$c_j = \frac{\frac{w_j}{\lambda_j^2}\{1-\exp(-\lambda_j\Delta)\}^2}{6\sum_{j=1}^m \frac{w_j}{\lambda_j^2}\{\exp(-\lambda_j\Delta) + \lambda_j\Delta - 1\} + 2\Delta^2\frac{(\sum_{j=1}^m \xi_j)^2}{\sum_{j=1}^m \omega_j^2}}. \tag{14}$$

This more general dynamic behaviour is focussed specifically at capturing the frequently encountered phenomenon of rapid initial decline in the empirical autocorrelation function of squared returns combined with a slow decay at higher lags. Superposition of processes with quite distinct values for the rate parameters λ_j is promising in this regard, and has been used successfully in Barndorff-Nielsen and Shephard (2001a, 2002), using a quasi-likelihood procedure to fit the model. The latter papers use a restricted parameterization where each component process has the same mean/variance ratio as the overall process $\sigma^2(t)$, which has mean ξ and variance ω^2. This implies that $\xi_j = w_j\xi$, $\omega_j^2 = w_j\omega^2$, $j = 1, \ldots, m$. One possible interpretation of the weights in this restricted parameterisation is as a discrete mixing distribution for λ which takes the form $p(\lambda) = \sum_{j=1}^m w_j\delta_{\lambda_j}$, where δ_x is the Dirac measure at x. Barndorff-Nielsen (2001) discusses many properties of superpositions of Ornstein-Uhlenbeck processes and shows that a superposition with a continuous mixing distribution can induce long memory

processes. Analytical solutions to derivative pricing are also available for superpositions, as discussed in Nicolato and Venardos (2003). Finally, superpositions of volatility processes are a natural and immediate way of capturing the behaviour observed for equities and interest rates by Alizadeh, Brandt and Diebold (2002) and Andersen, Benzoni and Lund (2002), who find evidence of one highly persistent factor and one rapidly moving factor in volatility. The presence of two volatility factors is also favoured by Chernov *et al.* (2003) in order to simultaneously capture observed tail thickness and volatility persistence.

2.3 Pure Jump Components and Leverage Effects

From an economic point of view, there has been considerable interest in using jump processes to model the price, *e.g.* Bates (2000) and Duffie *et al.* (2000) consider jump-diffusion models and Carr, Geman, Madan and Yor (2002) discuss models where the price process follows a non-Gaussian Lévy process. Barndorff-Nielsen, Nicolato and Shephard (2002) also discuss the possibility of including a Lévy component in only the drift function and Duffie *et al.* (2000) explicitly allow for jumps in both the price and the volatility processes. We consider both the inclusion of a pure jump component in the drift of the returns process and the incorporation of a leverage effect, the latter extending the specifications described in Barndorff-Nielsen and Shephard (2001a). From a statistical point of view, such model extensions have often proved important in fitting observed financial time series, see *e.g.* Andersen *et al.* (2002) and Eraker, Johannes and Polson (2003). In particular, the latter paper finds strong empirical evidence for jumps in volatility, even after accounting for jumps in returns.

2.3.1 A Jump Component in the Returns

In this subsection, we consider adding a pure jump Lévy process with positive increments, which will be called the jump component in the sequel, to the drift of the returns process. The model described in (1) with the instantaneous volatility process introduced above can be considered as a subordination of Brownian motion, in particular,

$$y^\star(t) \stackrel{d}{=} \mu\, t + \beta \sum_{j=1}^{m} \sigma_j^{2\star}(t) + B_1 \left\{ \sum_{j=1}^{m} \sigma_j^{2\star}(t) \right\}.$$

where B_1 is Brownian motion. To model a component without persistence, we consider adding an independent process, $\upsilon(t)$, to the drift of the return series. We restrict our attention to the class of processes which are time-changed Brownian motion with a Lévy process for the time-change so that $\upsilon(t) \stackrel{d}{=} B_2\{\zeta(t)\}$, where B_2 is Brownian motion and $\zeta(t)$ is a Lévy process. The stochastic specification is completed by allowing the process to have its own drift term $\tilde{\beta}\,\zeta(t)$ which will allow the risk to be priced and leads to the expression

$$y^\star(t) \stackrel{d}{=} \mu\, t + \beta \sum_{j=1}^{m} \sigma_j^{2\star}(t) + \tilde{\beta}\,\zeta(t) + B_2\left\{\zeta(t)\right\} + B_1 \left\{ \sum_{j=1}^{m} \sigma_j^{2\star}(t) \right\}$$

where B_1 and B_2 are independent. Alternatively, if we assume $\tilde{\beta} = \beta$, we can write

$$y^\star(t) \stackrel{d}{=} \mu t + \beta\, \sigma^{2\star}(t) + B\{\sigma^{2\star}(t)\},$$

as before, where now

$$\sigma^{2\star}(t) = \zeta(t) + \sum_{j=1}^{m} \sigma_j^{2\star}(t), \tag{15}$$

which extends (10). We assume that the distribution of $\zeta(1)$ has mean $w_{m+1}\xi$ and variance $w_{m+1}\omega^2$ and now $\sum_{j=1}^{m+1} w_j = 1$ where all weights w_j are positive and $\sigma_j^{2\star}(t)$ for $j = 1, \ldots, m$ are defined as before. The m weights in the autocorrelation function of $\sigma^2(t)$ as in (12) now no longer add up to one, and $\sum_{j=1}^{m} w_j$ is the variance of the volatility process over the total variance of the process $\sigma^2(t)$. The autocorrelation function for the squared returns is given by (13) and (14), but with the index of the sums in the second term of the denominator in (14) now running from 1 to $m + 1$.

Carr *et al.* (2002) consider subordinating a Lévy process by a Cox, Ingersoll and Ross (1985) process. As noted by Barndorff-Nielsen *et al.* (2002) this has some connection to the approach described above. In contrast to Carr *et al.* (2002), our subordinator contains both persistent and non-persistent components.

2.3.2 Incorporating a Leverage Effect

Following the work of Black (1976), negative movements in stock returns are frequently associated with larger volatility increases than positive movements. Such asymmetric behaviour is often observed for returns series, and has motivated for example the EGARCH model of Nelson (1991). Economic interpretations are provided by Campbell and Kyle (1993) and Shin (2003). Barndorff-Nielsen and Shephard (2001a,b) propose to introduce a leverage effect in the model by replacing (1) by the following stochastic differential equation:

$$\mathrm{d}y^\star(t) = \{\mu + \beta\sigma^2(t)\}\mathrm{d}t + \rho\mathrm{d}\bar{z}(\lambda t) + \sigma(t)\mathrm{d}B(t), \tag{16}$$

where $\bar{z}(t) = z(t) - \mathrm{E}\{z(t)\}$ is the centred version of the background driving Lévy process in (4). As a consequence, the aggregate returns will now be given by

$$p(y_i|\sigma_i^2, z_i) = \mathrm{N}(y_i|\mu\Delta + \beta\sigma_i^2 + \rho z_i, \sigma_i^2) \tag{17}$$

with $z_i = \int_{(i-1)\Delta}^{i\Delta} \mathrm{d}z(\lambda t)\mathrm{d}t - \Delta\mathrm{E}\{z(1)\}$. Negative values for the leverage coefficient relate positive volatility changes with drops in returns. Note that the addition of the (jump) leverage term in the log price process also induces discontinuous sample paths. In the case of a superposition of m Ornstein-Uhlenbeck processes as in (9) or (15) we can include up to m leverage terms in the returns equation, each with their own coefficient, ρ_j, which multiplies $z_{i\,j} = \int_{(i-1)\Delta}^{i\Delta} \mathrm{d}z_j(\lambda_j t)\mathrm{d}t - \Delta\mathrm{E}\{z_j(1)\}$, $j = 1, \ldots, m$.

As there is no compelling argument for linearity of the leverage term, we will also explore more general leverage terms. A benefit of this class of models is the ease with which models can be extended to capture this effect. Generally, we could assume that the leverage effect enters the drift as a centred

version of some continuous function of the background driving Lévy process, $f(z(t))$, that could be approximated by a Taylor series up to order p, modifying (17) to

$$p(y_i|\sigma_i^2, z_i) = \mathrm{N}\left(y_i \middle| \mu\Delta + \beta\sigma_i^2 + \sum_{k=1}^{p}\sum_{j=1}^{m} \rho_j^{(k)} z_{i\,j}^{(k)}, \sigma_i^2\right) \tag{18}$$

where $z_{i\,j}^{(k)} = \int_{(i-1)\Delta}^{i\Delta} \mathrm{d}z_j^{(k)}(\lambda_j t)\mathrm{d}t - \Delta\mathrm{E}\{z_j^{(k)}(1)\}$, $j = 1, \ldots, m$. The notation $z_{i\,j}^{(k)}$ is used to emphasise that these are integrals of powers of a Lévy process rather than powers of integrals of a Lévy process.

2.4 Component-Specific Risk Premia

Having a superposition of volatility components allows for a volatility process which is driven by both quite persistent and rare shocks as well as very frequent and short-lived movements. It is not at all clear that both components should lead to the same risk premium in the price equation in (1). In fact, one may well argue in favour of quite different market reactions to these volatility components. Bates (2000) considers a two-factor geometric jump-diffusion model for future option prices and assigns separate risk premia to each factor. Pan (2002) builds on the Bates model and distinguishes the premia for "Brownian" return risks and jump risks. Using a volatility process with $\tilde{m}$ components as in (9) ($\tilde{m} = m$) or (15) ($\tilde{m} = m + 1$ and defining $\zeta(i\Delta) - \zeta\{(i-1)\Delta\}$ by $\sigma_{i\,m+1}^2 = \int_{(i-1)\Delta}^{i\Delta} \sigma_{i\,m+1}^2(t)dt$ for notational convenience), we then obtain the following log price process:

$$\mathrm{d}y^\star(t) = \left\{\mu + \sum_{j=1}^{\tilde{m}} \beta_j \sigma_{i\,j}^2(t)\right\}\mathrm{d}t + \sigma(t)\mathrm{d}B(t). \tag{19}$$

Combining this idea with the leverage effect of Subsection 2.3 gives us the most general form that we shall consider for the returns

$$p(y_i|\sigma_i^2, z_{i\,1}, \ldots, z_{i\,m}) = \mathrm{N}\left(y_i \middle| \mu\Delta + \sum_{j=1}^{\tilde{m}} \beta_j \sigma_{i\,j}^2 + \sum_{k=1}^{p}\sum_{j=1}^{m} \rho_j^{(k)} z_{i\,j}^{(k)}, \sigma_i^2\right), \tag{20}$$

where some of the risk premia or leverage coefficients can be constrained to be zero or equal to each-other. For example, by taking $\beta_j = \beta_l$ we are effectively pricing the superposition of the volatility components j and l for the purpose of risk.

The three general model extensions described in the previous and the current subsection are important modelling features of our approach, which allow for substantial extra flexibility and will be examined in greater detail in the context of our application. Whereas Barndorff-Nielsen and Shephard (2001a) provides a quasi-likelihood fit of the model with superpositions as in (9) (and zero values for μ and β), we stress that the inference method proposed here can actually deal with the incorporation of a pure jump process in returns as well as the much more general drift function in (20) and provides us with exact likelihood-based inference in the context of these important model extensions.[4]

[4]Chernov *et al.* (2003) comment on the statistical problems posed by jump components.

3 The Sampling Algorithm

MCMC methods for inference with discrete-time stochastic volatility models were introduced in *e.g.* Jacquier, Polson and Rossi (1994) and Kim, Shephard and Chib (1998). The latter papers also provide references to the MCMC literature and general descriptions of some underlying algorithms. This section introduces an MCMC sampling algorithm to make inference in the class of continuous-time models described above. The discussion centres on background driving Lévy processes which have finite intensity, in particular that of the Gamma-OU process, the OU process which has a Gamma marginal distribution. The algorithm is easily adapted to superpositions of volatility processes and can simply be extended to include jump components, leverage effects and separate risk premia. The volatility process is parameterized by λ and θ which are the parameters of the marginal distribution of $\sigma^2(t)$. The likelihood of the parameters $L(\lambda,\theta) = \int \prod_{i=1}^{T} p(y_i|\sigma_i^2)p(\sigma_1^2,\ldots,\sigma_T^2|\lambda,\theta)\,\mathrm{d}\sigma_1^2\,\ldots\,\mathrm{d}\sigma_T^2$ will not be available in closed form. We propose augmenting the parameter vector λ,θ with the actual volatility process $\sigma_1^2,\ldots,\sigma_T^2$. This data augmentation strategy was mentioned in Barndorff-Nielsen and Shephard (2001a), but not implemented for reasons of computational complexity. The structure of the model implies that conditional on these volatilies the observations are Gaussian and independent.

From (7) we immediately obtain a convenient expression for the discretely observed process σ_i^2 as

$$\begin{aligned}\sigma_i^2 &= \sigma^{2\star}(i\Delta) - \sigma^{2\star}((i-1)\Delta) \\ &= \lambda^{-1}\{z(\lambda i\Delta) - \sigma^2(i\Delta) - z(\lambda(i-1)\Delta) + \sigma^2((i-1)\Delta)\}. \end{aligned} \tag{21}$$

In addition, we can write the latent volatility process (from (4)) and the driving Lévy process in terms of a vector of random shocks η_i as

$$\left\{ \begin{array}{c} \sigma^2(i\Delta) \\ z(\lambda i\Delta) \end{array} \right\} = \left(\begin{array}{c} \exp\{-\lambda\Delta\}\sigma^2((i-1)\Delta) \\ z(\lambda(i-1)\Delta) \end{array} \right) + \eta_i, \tag{22}$$

where

$$\eta_i \stackrel{d}{=} \left\{ \begin{array}{c} \exp\{-\lambda\Delta\}\int_0^{\Delta}\exp\{\lambda t\}\mathrm{d}z(\lambda t) \\ \int_0^{\Delta}\mathrm{d}z(\lambda t) \end{array} \right\}. \tag{23}$$

We use the recursions in (22) to construct realizations for the volatility process in (21), starting from the shocks in (23). In order to draw the latter shocks, we will use a series representation of the integrals involved.

3.1 A Representation of Lévy Jump Processes

The simulation of Lévy processes without Gaussian components is much facilitated by the approach of Ferguson and Klass (1972), who propose a useful series representation. The inverse tail mass function W^{-1} of a Lévy process with Lévy measure W is defined as

$$W^{-1}(x) = \inf\{y > 0 : W^{+}(y) \leq x\}, \tag{24}$$

where $W^+(x)$ is the tail mass function[5]

$$W^+(x) = \int_x^\infty w(y)\mathrm{d}y. \tag{25}$$

It can be shown that the integral of some positive (and integrable) function f with respect to a positive Lévy jump process has the same distribution as the following infinite sum:

$$\int_0^\Delta f(s)\mathrm{d}z(s) \stackrel{d}{=} \sum_{j=1}^\infty W^{-1}(a_j/\Delta)f(\Delta r_j), \tag{26}$$

where $\{a_j\}$ and $\{r_j\}$ are independent with r_j independent uniform random variables on (0,1) and $a_1 < a_2 < \ldots$ the arrival times of a Poisson process of intensity 1. This representation is discussed in some detail by Barndorff-Nielsen and Shephard (2001b). The main idea was introduced for nonhomogeneous processes in Ferguson and Klass (1972).

Applying (26), we can now evaluate the integrals in (23) as

$$\eta_i \stackrel{d}{=} \left\{ \begin{array}{c} \exp\{-\lambda\Delta\}\sum_{j=1}^\infty W^{-1}(\frac{a_j}{\lambda\Delta})\exp\{\lambda\Delta r_j\} \\ \sum_{j=1}^\infty W^{-1}(\frac{a_j}{\lambda\Delta}) \end{array} \right\}. \tag{27}$$

Thus, simply sampling $\{a_j\}$ and $\{r_j\}$ will allow us to construct sample paths for the volatility process and the Lévy driving process, which can then be used to construct a sequence for the actual volatilities $\sigma_i^2, i = 1, \ldots, T$ by (21). The only remaining problem then lies in the fact that we have an infinite sum in our representations through (26). We propose to deal with that issue by only taking into account the jumps greater than a certain size, *i.e.* by restricting ourselves to the jumps for which $W^{-1}(\cdot) \geq \varepsilon > 0$, so we make the following approximation

$$\sum_{j=1}^\infty W^{-1}\left(\frac{a_j}{\lambda\Delta}\right) f(\lambda\Delta r_j) \approx \sum_{\{j \,|\, a_j/\lambda\Delta < W^+(\varepsilon)\}} W^{-1}\left(\frac{a_j}{\lambda\Delta}\right) f(\lambda\Delta r_j). \tag{28}$$

Thus, we only include the jumps corresponding to Poisson arrival times in the region $(0, \lambda\Delta W^+(\varepsilon))$. This region is well-defined because $W^+(\cdot)$ is a monotone non-increasing function. From the integrability condition in (5) we know that $W^+(\varepsilon)$ is always finite, so we only include a finite number of jumps. Hansen, Ickstadt and Wolpert (2000) mention a similar approach.

3.2 The Gamma-OU process

This paper will consider only the Gamma-OU process, the instance of the Ornstein-Uhlenbeck process which has a Gamma marginal distribution, i.e. $\sigma^2(t) \sim \mathrm{Gamma}(\nu, \alpha)$ where $\nu > 0$ is the shape parameter and α is the usual precision parameter. To use the series representation, the tail mass function is

$$W^+(x) = \nu\exp(-\alpha x)$$

[5]For the Ornstein-Uhlenbeck class of processes described in Subsection 2.1 the tail mass function has the simple formula $W^+(x) = xu(x)$, where u is the Lévy density of $\sigma^2(t)$ (see Barndorff-Nielsen, 1998).

which admits an analytical solution for the inverse function

$$W^{-1}(x) = \max\left\{0, -\frac{1}{\alpha}\log\left(\frac{x}{\nu}\right)\right\}. \tag{29}$$

and we will not need any truncation, since the sum in (26) will then reduce to a finite one. It is immediate from (29) that only Poisson arrivals in $(0, \lambda\Delta\nu)$ will contribute to this sum. If we also include a jump component in the returns as in Subsection 2.3, the Gamma marginal distribution for $\sigma^2(t)$ is maintained by adding a Normal-Gamma process to the drift term (see *e.g.* Bibby and Sørensen, 2003). In this case the time-change is a Gamma process which has tail mass function

$$W^{+}(x) = \int_x^{\infty} \nu x^{-1}\exp(-\alpha x)\mathrm{d}x. \tag{30}$$

The representation of the latter process does require truncation as in (28). To use the extended model in (18), we need to calculate $\mathrm{E}\{z^{(k)}(1)\}$ which in the case of a Gamma-OU process is given by the expression $\mathrm{E}\{z^{(k)}(1)\} = \lambda\Delta\nu\mathrm{E}(v^k)/\alpha^k$ where $v \sim \text{Exponential}(1)$.

3.3 Sampling the Process

The form of the shocks for a Gamma-OU process can be derived by substituting (29) in (27):

$$\eta_i \stackrel{d}{=} \alpha^{-1}\begin{pmatrix} \exp\{-\lambda\Delta\}\sum_{j=1}^{n}\log(\frac{\lambda\nu\Delta}{a_j})\exp\{\lambda\Delta r_j\} \\ \sum_{j=1}^{n}\log(\frac{\lambda\nu\Delta}{a_j}) \end{pmatrix},$$

where $a \times r$ is a Poisson process on $(0, \lambda\nu\Delta) \times (0, 1)$ with intensity 1 and n is the number of points in this Poisson process. We first consider the following reparameterisation $a = (a_1, \ldots, a_n) \longrightarrow (\phi, \varphi_1, \ldots, \varphi_n)$ where $\phi = \sum_{j=1}^{n}\log(\frac{\lambda\nu\Delta}{a_j})$ and $\varphi_j\phi = \log(\frac{\lambda\nu\Delta}{a_j})$. Then $\phi \sim \text{Ga}(n, 1)$, a Gamma distribution with mean n and $\varphi = (\varphi_1, \ldots, \varphi_n) \sim \text{Di}_{n-1}(1, \ldots, 1)$, a Dirichlet distribution. Thus, given the parameters λ and θ, any realization of the discretely observed (or actual) volatility process σ_i^2 will be fully determined by its associated values of $(\phi, \varphi, r, n)_i$. In the sequel, we denote the subset $(\phi, \varphi, r)_i$ by Ψ_i and the number of jumps for observation i by n_i. Grouping these random variables for all observations, we will define $\Psi = (\Psi_1, \ldots, \Psi_T)$ and $N = (n_1, \ldots, n_T)$. The basic strategy will be an MCMC sampler on $(\Psi, N, \lambda, \theta)$, alternately updating the process and the parameters.

The process conditional on the parameters is sampled using a Metropolis-Hastings sampler with three possible proposals: tailored independent proposal, random walk step or jump moved. In each sweep of the Gibbs sampler, a preset number of time points are selected to be updated by a move chosen at random. The independence and random walk proposals have a proposal density of the form $p(\Psi_i'|\theta, \lambda, n_i')\, q(n_i, n_i')$. The acceptance rate using Reversible Jump MCMC (Green, 1995) can be calculated for these moves (see Appendix A.1) as

$$\min\left\{1, \frac{p(y|\Psi_{(i)}', N_{(i)}', \lambda, \theta)}{p(y|\Psi, N, \lambda, \theta)}\frac{p(n_i'|\lambda, \theta)}{p(n_i|\lambda, \theta)}\frac{q(n_i', n_i)}{q(n_i, n_i')}\right\},$$

where $\Psi_{(i)}' = \{\Psi_1, \ldots, \Psi_{i-1}, \Psi_i', \Psi_{i+1}, \ldots, \Psi_T\}$, $N_{(i)}' = \{n_1, \ldots, n_{i-1}, n_i', n_{i+1}, \ldots, n_T\}$. We, first, consider the two possible transition densities: $q_1(n_i, n_i')$, which is a random walk proposal, and $q_2(n_i')$,

a tailored independent proposal using an approximation of $p(n_i|y_i, \sigma^2((i-1)\Delta), \lambda, \theta)$. The latter distribution captures the idea of a "filtered" distribution of the number of jumps, by conditioning on the observed return for that period and the previous value of the instantaneous volatility process, which summarizes the past in building σ_i^2 in (21). In all cases the parameters ϕ, φ, r for observation i (*i.e.* Ψ_i) are drawn from their prior distribution conditional on n_i. The random walk has three possible transitions $n_i' = n_i + 1$, $n_i' = n_i$, and $n_i' = n_i - 1$ (provided $n_i > 0$), leading to a simple, robust sampler. The tailored proposal $q_2(n_i')$ will generally have better convergence properties in cases where the data are in agreement with the model, but may have problems with "unusual" observations. As stated above, it is based on the distribution described by $p(n_i|y_i, \sigma^2((i-1)\Delta, \lambda, \theta))$, which is proportional to

$$\int p(y_i|\sigma^2((i-1)\Delta), \Psi_i, n_i, \lambda, \theta)\, p(\Psi_i|n_i)\, \mathrm{d}\Psi_i\; p(n_i)$$

where the first factor equals $p(y_i|\sigma_i^2)$. This integral cannot be calculated analytically but is here approximated by $p(y_i|\mathrm{E}_{\Psi_i}(\sigma_i^2|n_i))$. The distribution is then approximated pointwise in n_i. For a superposition of processes, this approach can be extended using the proposal distribution

$$q(n_i^{(1)}, \Psi_i^{(1)}, \ldots, n_i^{(m)}, \Psi_i^{(m)}) = \prod_{j=1}^{m} p(\Psi_i^{(j)}|n_i^{(j)})\, q(n_i^{(j)}|n_i^{(1)}, \ldots, n_i^{(j-1)}).$$

where the superscripts index the component processes of the superposition. In other words each $n_i^{(j)}$ is drawn in turn, conditionally on the draws for the previous components. The proposal distribution $q(n_i^{(j)}|n_i^{(1)}, \ldots, n_i^{(j-1)})$ is an approximation similar to the approximation discussed above. If we define $\mathcal{S}_j = \{\sigma^{2\,(j)}((i-1)\Delta), \lambda^{(j)}, \theta^{(j)}\}$, we can write

$$p(n_i^{(j)}|n_i^{(1)}, \ldots, n_i^{(j-1)}, y_i, \mathcal{S}_1, \ldots, \mathcal{S}_m) \propto$$
$$\int p(y_i|n_i^{(1)}, \ldots, n_i^{(j)}, \mathcal{S}_1, \ldots, \mathcal{S}_m) p(\Psi_i^{(j)}|n_i^{(j)})\, \mathrm{d}\Psi_i^{(j)}\; p(n_i^{(j)}).$$

The integral is approximated by

$$p(y_i|\mathrm{E}_{\Psi_i^{(1)}|n_i^{(1)}}(\sigma_i^{2\,(1)}), \ldots, \mathrm{E}_{\Psi_i^{(j-1)}|n_i^{(j-1)}}(\sigma_i^{2\,(j-1)}), \mathrm{E}_{\Psi_i^{(j)}|n_i^{(j)}}(\sigma_i^{2\,(j)}), \mathrm{E}_{\Psi_i^{(j+1)}}(\sigma_i^{2\,(j+1)}),$$
$$\ldots, \mathrm{E}_{\Psi_i^{(m)}}(\sigma_i^{2\,(m)})),$$

which is possible because both $\mathrm{E}_{\Psi_i^{(j)}|n_i^{(j)}}(\sigma_i^{2\,(j)})$ and $\mathrm{E}_{\Psi_i^{(j)}}(\sigma_i^{2\,(j)})$ can be evaluated (see Appendix A.2).

Figure 1 illustrates the superior performance of the independence sampler for regular (*i.e.* well-behaved) problems with a reasonable number of jumps.

The third possible proposal involves moving jumps between observations in such a way that only the volatility process between these two jumps is changed. A similar proposal was introduced by Roberts *et al.* (2001) and helps the mixing of the chain when the exact location of a large jump is poorly defined. Before the proposal is applied assume that the jumps are associated with $(i_1, r_1), \ldots, (i_k, r_k)$, where $k = \sum_{i=1}^{T} n_i$ and the pairs (i_l, r_l) denote the discrete timing of the lth jump and its value of r. The proposal selects one jump at random and proposes to move it in time from time period i_j to i_j' in such a way that the process $\sigma^2(t)$ outside of $\min(i_j, i_j')$ and $\max(i_j, i_j')+1$ is unchanged. The form of the model

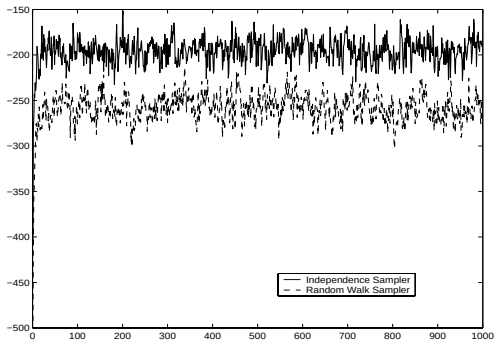


Figure 1: The log posterior density of the process given the parameters for the independent proposal and the random walk proposal using simulated data with parameters $\lambda = 1$, $\nu = 1$, $\alpha = 1$ and $T = 1000$.

keeps the process before $\min(i_j, i'_j)$ the same. Letting $h = \max(i_j, i'_j) + 1$ and $\kappa_l \equiv \alpha^{-1}\log(\lambda\nu\Delta/a_l)$ be the jump associated with (i_l, r_l), we can express $\sigma^2(h\Delta)$ as

$$\sigma^2(h\Delta) = \exp\{-h\lambda\Delta\}\sigma^2(0) + \sum_{l=1}^{k} I(i_l \leq h)\exp\{-(h - i_l + 1 - r_l)\lambda\Delta\}\kappa_l.$$

We maintain the value of $\sigma^2(h\Delta)$ by taking

$$\kappa'_j = \exp\{-(i'_j + r'_j - i_j - r_j)\lambda\Delta\}\kappa_j.$$

Mapping $(i_j, r_j, \kappa_j) \longrightarrow (i'_j, r'_j, \kappa'_j)$ involves a Jacobian which reduces to $\mathrm{d}\kappa'_j/\mathrm{d}\kappa_j = \exp\{-(i'_j + r'_j - i_j - r_j)\lambda\Delta\}$. The acceptance probability for this Metropolis-Hastings move is

$$\min\left\{1, \frac{p(y|\Psi', N', \lambda, \theta)}{p(y|\Psi, N, \lambda, \theta)} \frac{q\left((i'_j, r'_j), (i_j, r_j)\right)}{q\left((i_j, r_j), (i'_j, r'_j)\right)} \left|\frac{\mathrm{d}\kappa'_j}{\mathrm{d}\kappa_j}\right|\right\},$$

where Ψ', N' describe the volatility process after the proposed move of a jump.

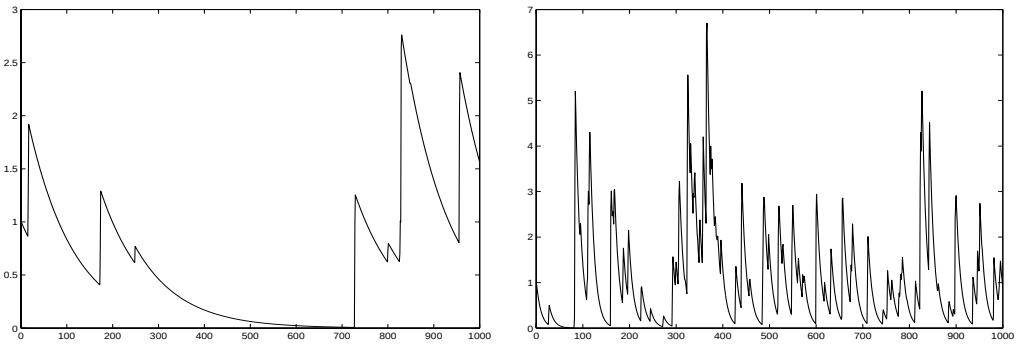
3.4 Sampling the Parameters

The problem of sampling the parameters is more complicated. Roberts *et al.* (2001) note that there is a potential problem with overconditioning if a simple Gibbs sampling scheme is used to sample the parameters. Consider updating the parameter λ conditional on Ψ: the full conditional posterior density for λ is proportional to

$$p(y|\Psi, N, \lambda, \theta)\lambda^k \exp\{-T\lambda\nu\Delta\}p(\lambda),$$

where $y = (y_1, \ldots, y_T)$. The second and third factors of this expression are derived from the Poisson process driving the series approximation to the Lévy process. This part is proportional to a Gamma distribution with a mean of $(k+1)/T\nu\Delta$ and a variance $T\nu\Delta$ times smaller than the mean. Therefore, this distribution is largely determined by the value of k (the total number of jumps in the process) which in turn was largely determined by the current value of λ. This situation is commonly known as

overconditioning. To combat this problem the number of jumps must change with the value of λ. The parameter λ determines the rate at which shocks occur, but also changes the slope (rate of decay) of the volatility process and the shock size. Figure 2 shows realisations of processes with the same marginal



(a) (b)

Figure 2: Influence of parameter λ: (a) $\lambda = 0.01$, (b) $\lambda = 0.1$. We have taken $\nu = 1$, $\alpha = 1$, $\Delta = 1$.

Gamma distribution, but different values of λ. It is clear that to move between processes with the same marginal distribution we must add more jumps when λ increases to counteract the increased slope. One approach to solving this problem is to change the process jointly with the parameters, to ensure that the new process is not at great odds with the data. This will be implemented through the Reversible Jump MCMC sampler.

In particular, if a new value λ' is chosen such that $\lambda' < \lambda$, the current value, then it is natural to consider a thinning of the Poisson process driving the shocks. We use a dependent thinning of the process which is motivated by the following observation. The inverse of the tail probability function $W^{-1}(\cdot)$ is decreasing. Consequently, the early arrival times in the Poisson process (small a_j) make the largest contribution to the value of the shock. Therefore, removing an appropriate number of late arrival times will have less of an effect on the volatility process. In particular, the arrival times, $\{a_i\}_{i=1}^{\infty}$, of the Poisson process restricted to $(0, \lambda'\Delta\nu)$ will be a Poisson process with the correct intensity. Figure 3 illustrates this point. Graph (b) shows the underlying Poisson processes. The points above the line are deleted in the thinning. The main features of the original process remain, the largest jump removed is around observation number 330 and is also close to the cut-off point. For Gamma marginals[6], we use the following updating scheme:

- If $\lambda' > \lambda$ new points of the Poisson process are drawn in the region $(\lambda\nu\Delta, \lambda'\nu\Delta)$.
- If $\lambda' < \lambda$ the arrival times in the region $(\lambda'\nu\Delta, \lambda\nu\Delta)$ are deleted.

The acceptance rate for this move is the minimum of 1 and (see Appendix A.3)

$$\frac{p(y|\sigma_1^{2\prime}, \ldots, \sigma_T^{2\prime})}{p(y|\sigma_1^2, \ldots, \sigma_T^2)} \frac{p(\lambda')}{p(\lambda)} \frac{q(\lambda', \lambda)}{q(\lambda, \lambda')}$$

[6]For other marginal distributions of $\sigma^2(t)$ or for the pure jump component, we replace ν by $W^+(\varepsilon)$.

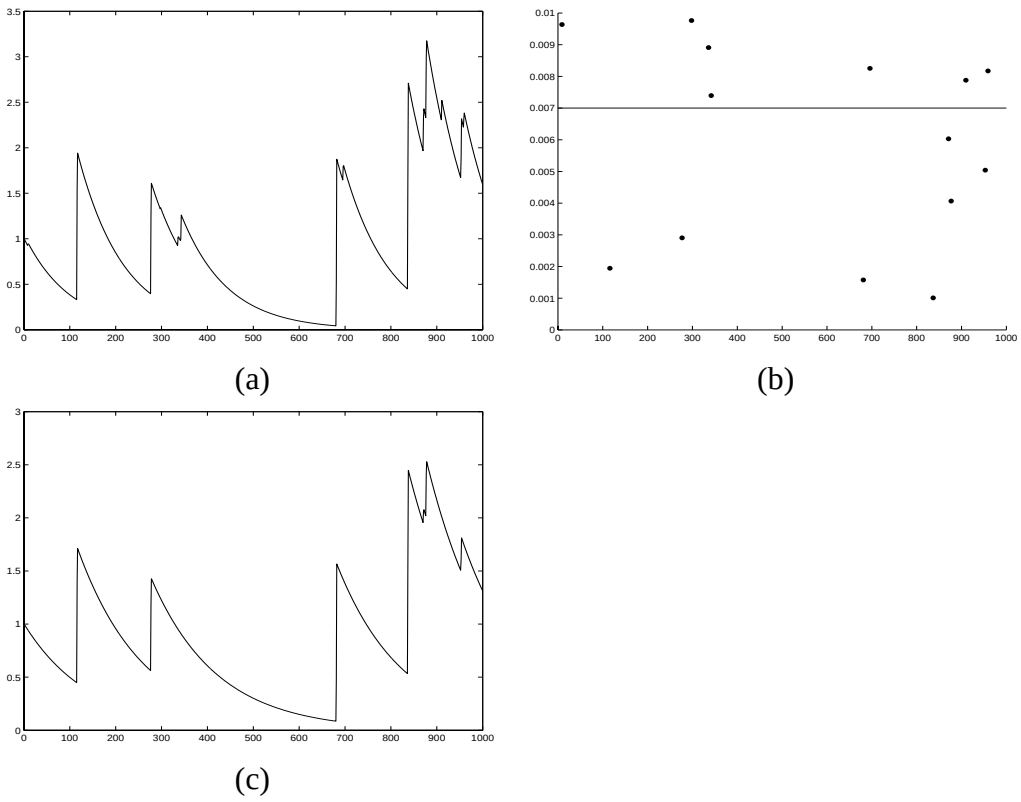


Figure 3: Dependent Thinning from $\lambda = 0.01$ to $\lambda' = 0.007$: (a) Original Process, (b) Arrival times of the underlying Poisson process at each discrete time, (c) Proposed process. Parameter values: $\nu = 1$, $\alpha = 1$, $\Delta = 1$. Note that the realized values of n_i are either 0 or 1 and the vertical axis in graph (b) denotes a.

The need to carefully introduce and remove points is illustrated by the following example. If a process $\{c_i\}$ is defined by $c_i = a_i/\lambda\nu\Delta$ then $\{c_i\}$ is a Poisson process on $(0, 1)$ with intensity $\lambda\nu\Delta$. For $\lambda' < \lambda$, a new process could be proposed by independently thinning the current process with probability λ'/λ. For $\lambda' > \lambda$, a reverse move would add points uniformly on $(0, 1)$. Figure 4(b) illustrates the problems associated with such an approach. This method tends to introduce large jumps (for example around observation numbers 1300 and 1900), which make it very unlikely that the proposed process will be supported by the data and accepted. For comparison, no thinning leads to a process which tends to be below the previous process (see graph (c)).

The parameters of the marginal volatility distribution, θ, are also updated using dependent thinning if the interval used in the series approximation is affected by the parameter. Otherwise simple random walk Metropolis-Hastings is used. For Gamma marginals, the shape parameter ν affects the interval but the scale parameter α does not. The parameterisation of the weights introduced by Barndorff-Nielsen and Shephard (2001a) is used throughout the paper. The sampler works on the processes $\sigma_j^2(t)$. In this case, the weights w_j affect the interval, so dependent thinning is used. In the Gamma case it will prove numerically more efficient to parameterize the marginal distributions in terms of the mean ν/α and the scale parameter, α. This reduces the posterior correlation between the parameters.

We have evaluated the performance of our algorithm and the effect of various priors on the basis of

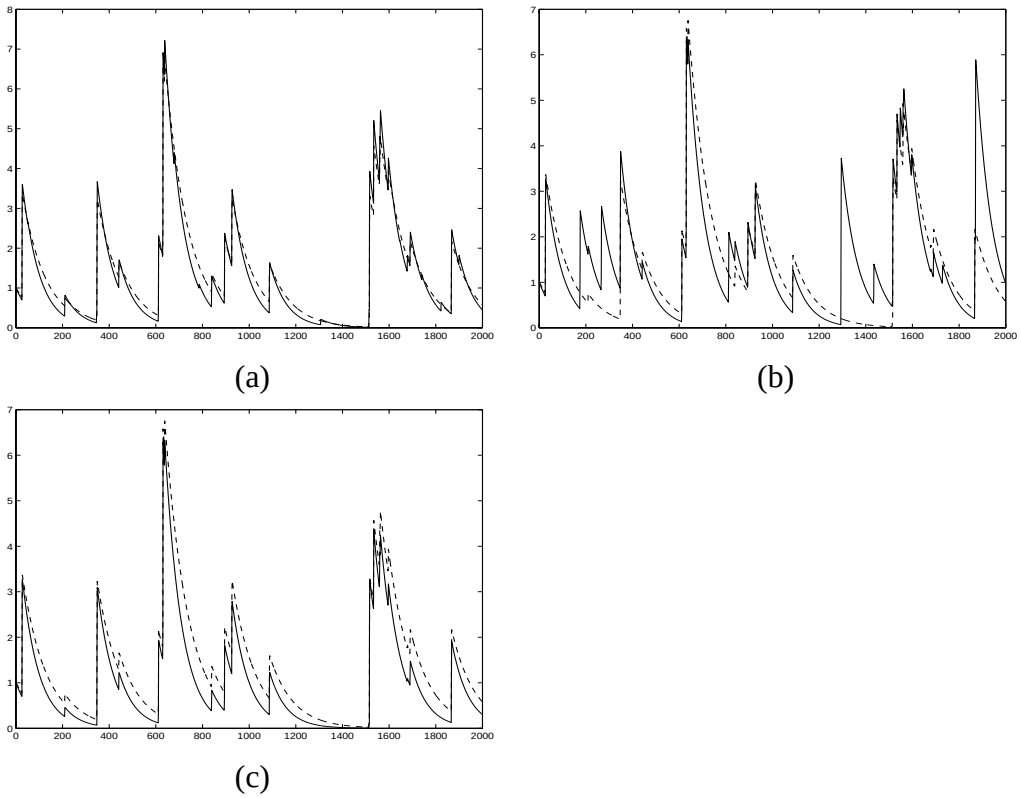


Figure 4: Three types of thinning: (a) Dependent thinning, (b) Independent thinning, (c) No thinning. The dotted line is the original process and the solid line is the proposed process with parameter values: $\nu = 1$, $\alpha = 1$, $\Delta = 1$, $\lambda = 0.01$, $\lambda' = 0.014$.

simulated data. Appendix A.4 presents some of these results, indicating that the algorithm works well (especially in view of the complexity of the model). Some of our findings on the effect of the prior are mentioned in the next section.

4 The Prior Distribution

We shall consider proper, but mostly weakly informative, priors for the parameters in our sampling model. This will ensure the existence of a posterior distribution. In the context of our sampling model in (16) combined with (4), we have explicitly introduced the drift parameters μ, β, ρ in (16) and the rate parameter λ in (4). In addition, we require parameters to describe the marginal distribution of $\sigma^2(t)$ (or, equivalently, the Lévy measure). For the superposition in (9) we need a prior on the weights w_j, and each component now has its own rate λ_j and possibly its own leverage and risk premium parameters $(\beta_j, \rho_j^{(k)}), k = 1, \ldots, p$. In addition, if we include a jump process as well, we need to specify a prior on the associated weight w_{m+1}. Throughout, we will adopt a product structure prior where only the weights will be prior dependent.

In our experience, the most critical aspect of the prior is that on λ (or λ_j). In order to elicit this prior,

we first state that μ and β_j are *a priori* independent Normally distributed with mean zero and variance 90. This means that the prior is centered over the case where the correlation of squared returns is the simple expression in (8) from which we see that for $s > 0$

$$\frac{\text{cor}(y_i^2, y_{i+s+1}^2)}{\text{cor}(y_i^2, y_{i+s}^2)} = \exp(-\lambda\Delta) \equiv \tau. \tag{31}$$

Thus, we can think of λ as governing how quickly subsequent correlations of squared returns (with Δ as the observational period) decay. If we assume a Gamma(a, b) prior on λ, the implied prior on τ is

$$p(\tau) = \left(\frac{b}{\Delta}\right)^a [\Gamma(a)]^{-1} (\log \tau^{-1})^{a-1} \tau^{b/\Delta - 1}. \tag{32}$$

In particular, choosing $a = 1$ and $b = \Delta$ will induce a Uniform prior over the range (0,1) for this ratio of squared returns correlations. If, instead, we would take b much smaller than Δ, we would impose strong prior beliefs that the correlation declines very rapidly. This does not seem in keeping with generally held prior ideas on most financial time series for small or moderate values of Δ and would make it hard to distinguish our dynamic process from a nonpersistent process. For our applications with daily observations (where we take $\Delta = 1$), we find that a Gamma(1,1) density is a reasonable choice for $p(\lambda)$ or $p(\lambda_j)$ in case $m > 1$.

If we have a superposition of processes, the weights are given a Dirichlet prior distribution of appropriate dimensions with all parameters equal to 1. In the applications, we only use Gamma marginal distributions for $\sigma^2(t)$ with the restricted parameterization mentioned in Subsections 2.2 and 2.3. The two parameters of the Gamma marginal distribution of $\sigma^2(t)$ are given independent vague but proper Exponential prior distributions with precision parameter 0.001. Finally, for the parameters $\beta_j^{(k)}, j = 1, \ldots, m, k = 1, \ldots, p$ an independent Normal prior distribution with mean zero and variance 900 is used.

5 Dealing With Model Uncertainty

We can consider models that differ in the number of components used in the superposition (9), in the inclusion of a jump component in (15), or in the specification of leverage effects and risk premia in (20). In addition, we could also consider differences in the marginal distribution of the volatility process $\sigma^2(t)$, although that is not done in the following application. It would seem feasible to use Reversible Jump MCMC to explore the posterior distribution over the number of components, but it proves to be quite difficult to implement (it is hard to merge and split entire latent processes). A simple solution is to compute the marginal likelihood values[7] for each model from the sampler. We investigated various ways of computing this on the basis of the MCMC output. Throughout, we will present results based on the modified harmonic mean estimator $\hat{p}_4$ of Newton and Raftery (1994). An alternative approach would be to use the method by Chib (1995) as implemented in Elerian *et al.* (2001).

[7]The marginal likelihood is the sampling density integrated out with the prior.

The ratio of the marginal likelihoods of two competing models is generally called the Bayes factor and the product of Bayes factor and prior odds (the ratio of prior model probabilities) gives the posterior odds. Thus, if we take prior odds equal to unity, the Bayes factor is the ratio of the posterior model probabilities.

Of course, posterior probabilities can be used to select one of the contending models, namely the one with highest posterior probability. A related use of these posterior model probabilities is their role as model weights in producing inferences averaged over all models under consideration. This Bayesian model averaging provides a formal way of incorporating model uncertainty into the analysis. For example, it provides a natural approach to predictive inference (inference on the observables, such as forecasts).

6 Application to Stock Price Data

We use data from the Standard and Poor 500 stock price index, in particular, $T = 2023$ daily returns from January 2, 1980 until December 30, 1987. Gallant, Rossi and Tauchen (1992) use the same series and prefilter it in order to remove systematic effects from both mean and variance. The same filtered data was used in Jacquier *et al.* (1994) in a discrete stochastic volatility model. We use the raw data, which are displayed in Figure 5. The October 19, 1987 crash is immediately obvious from the plot.

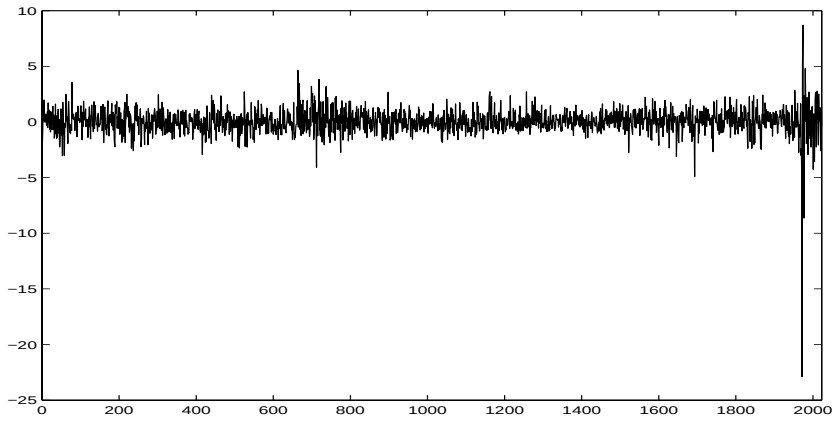


Figure 5: Daily returns from the S&P 500 stock price index; 2023 observations from January 2, 1980 until December 30, 1987.

Sample measures of skewness and kurtosis of the returns are -4.1 and 90.3, respectively.[8] We adopt the prior distribution described in Section 4 and the results for this dataset are typically based on Markov chains of length 215,000, where we discard the first 15,000 drawings and thin the remaining observations by recording every 10^{th} value. Throughout, we will use volatility processes with Gamma marginal distributions.

[8]Throughout, skewness is measured as $\mu_3/\mu_2^{3/2}$, and kurtosis as μ_4/μ_2^2, where μ_i is the ith central moment.

First, we consider the basic returns model in (3) in combination with the simplest Ornstein-Uhlenbeck volatility process in (4). Posterior results on the actual volatilities $\sigma_i, i = 1, \ldots, T$ are given in Figure 6(a)[9] in the form of the medians (joined with a drawn line) and the 2.5 and 97.5 percentiles (joined with dashed lines). Behaviour is as expected: high volatility values for periods with large price changes. The median volatility corresponding to October 19, 1987 is 3.3%. This is almost three times the long-run marginal mean of the volatility process (given by ν/α which has a posterior median of 1.17), testifying to an ability to account for sudden volatility shocks. Table 1 presents some posterior results on the model parameters as well as on the marginal moments of the instantaneous volatility process $\sigma^2(t)$. The

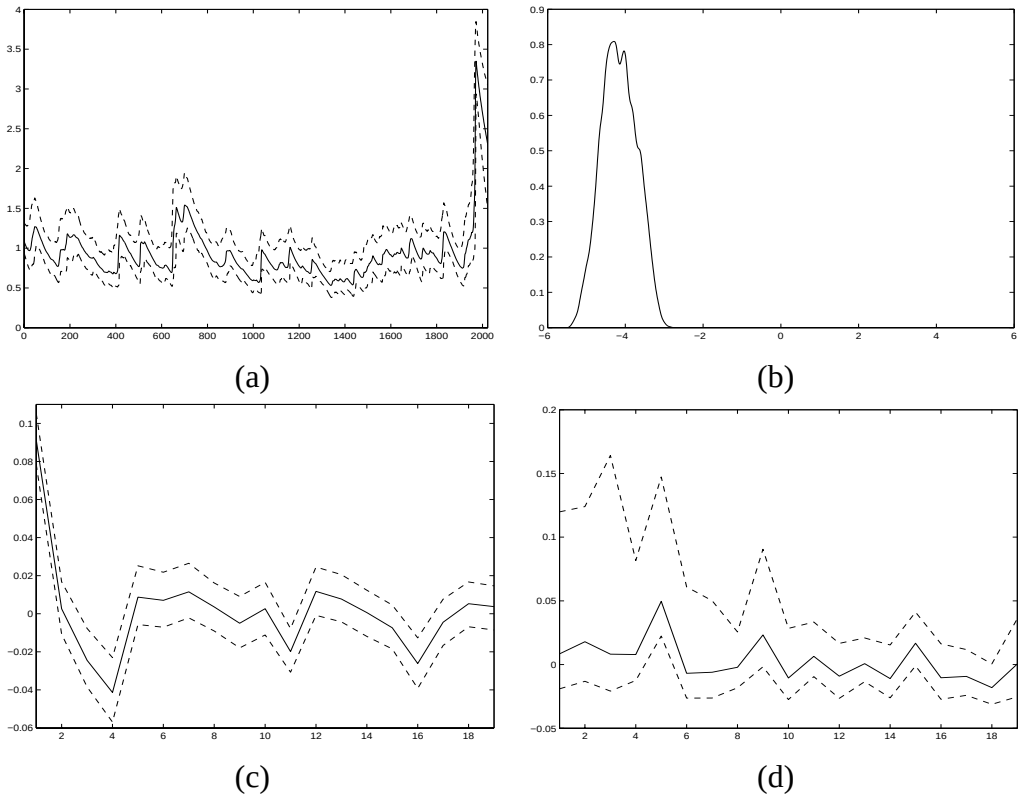


Figure 6: Posterior results for model with one component ($m = 1$) for the S&P 500 data: (a) σ_i, (b) $\ln(\lambda)$, (c) ACF of scaled residuals, (d) ACF of squared scaled residuals. In graphs (a), (c) and (d) drawn lines are posterior medians, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

posterior distribution of $\ln(\lambda)$ in Figure 6(b) shows that the mass is spread in the region from -5.5 to -3, corresponding to the interval (0.004, 0.05) for λ. The autocorrelation functions of the scaled residuals $(y_i - \mu\Delta - \beta\sigma_i^2)/\sigma_i$ and their squares are displayed in Figures 6(c) and (d); both suggest the model does an adequate job at capturing the empirical dynamics. However, if we consider summary statistics[10] of the predictive returns distribution implied by this model, displayed in Table 2, we see that the model

[9]These percentiles are smoothed. Filtered results are presented in Figure 18, as discussed in the sequel.

[10]These are computed on the basis of 20,000,000 realisations of the process in a Monte Carlo scheme. For each of the 20,000 samples drawn from the posterior distribution, 1000 realisations of the returns process were sampled.

does not manage to reproduce the high levels of negative skewness and kurtosis observed in the daily data ($\Delta = 1$). The latter table also contains the implied properties at other observation frequencies, namely $\Delta = 5$ (weekly), $\Delta = 21$ (monthly) and $\Delta = 63$ (three-monthly). Whereas the data were not used for estimating the model at these frequencies, it is of interest to assess the implied properties at various Δ of the model estimated using daily data. Clearly, as Δ increases, the simple one-component model gets closer to the observed returns distributions which displays roughly Gaussian behaviour for large Δ.

Table 1: Posterior medians, 95% credible intervals (in parentheses) and marginal likelihoods for basic model

	1 Component	2 Components	3 Components
μ	0.07 (-0.00, 0.14)	0.06 (-0.02, 0.13)	0.06 (-0.01, 0.13)
β	-0.02 (-0.10, 0.05)	-0.02 (-0.10, 0.07)	-0.02 (-0.10, 0.06)
E$(\sigma^2(t))$	1.17 (0.87,1.72)	0.99 (0.82, 1.24)	0.93 (0.81, 1.19)
St.Dev.$(\sigma^2(t))$	0.81 (0.58,1.26)	1.07 (0.82, 1.43)	1.00 (0.79, 1.26)
Log Marg.Lik.	-2659	-2521	-2521

Table 2: Empirical and predictive properties of the returns distribution

	Observed data				One-component basic model			
	$\Delta = 1$	$\Delta = 5$	$\Delta = 21$	$\Delta = 63$	$\Delta = 1$	$\Delta = 5$	$\Delta = 21$	$\Delta = 63$
Mean	0.0	0.2	0.9	2.6	0.0	0.2	0.8	2.4
Variance	1.3	5.9	23.7	76.2	1.2	6.0	25.9	82.1
Skewness	-4.1	-1.0	-0.8	-0.8	0.0	-0.1	-0.2	-0.2
Kurtosis	90.3	10.8	6.1	4.1	4.7	4.7	4.6	4.6

In order to add flexibility to the dynamic properties of the model, we try a superposition of processes as in (9). Throughout, we use the parameterization described in Subsection 2.2, *i.e.* $\xi_j = w_j \xi$ and $\omega_j^2 = w_j \omega^2, j = 1, \ldots, m$. If we take a superposition of two processes ($m = 2$), we obtain the medians and credible intervals for the discrete time volatility σ_i graphed in Figure 7(a).

For superpositions, the inference about λ will be presented in terms of the discrete mixing distribution discussed in Section 2.2. Figure 7(b) displays the posterior density function of this mixing distribution for $\ln(\lambda)$ which indicates that the relative weights assigned to each component are roughly equal, and that one component has much larger values of λ than the other. This results in the volatility behaviour in Figure 7(a), where the large jumps (with relatively slow decay) correspond to the process with small λ (displayed in Figure 7(c)) and the many smaller (and spiky) jumps originate from the process with the higher rate and decay parameter λ (depicted in Figure 7(d)). Thus, there are two clearly distinguished components, one for large persistent shocks and one for fast-moving "noise". The more flexible model now allows volatility to increase further in order to accommodate the crash: median volatility now gets up to 3.72%, but the possibility of having shocks whose effect on the volatility is short-lived allows

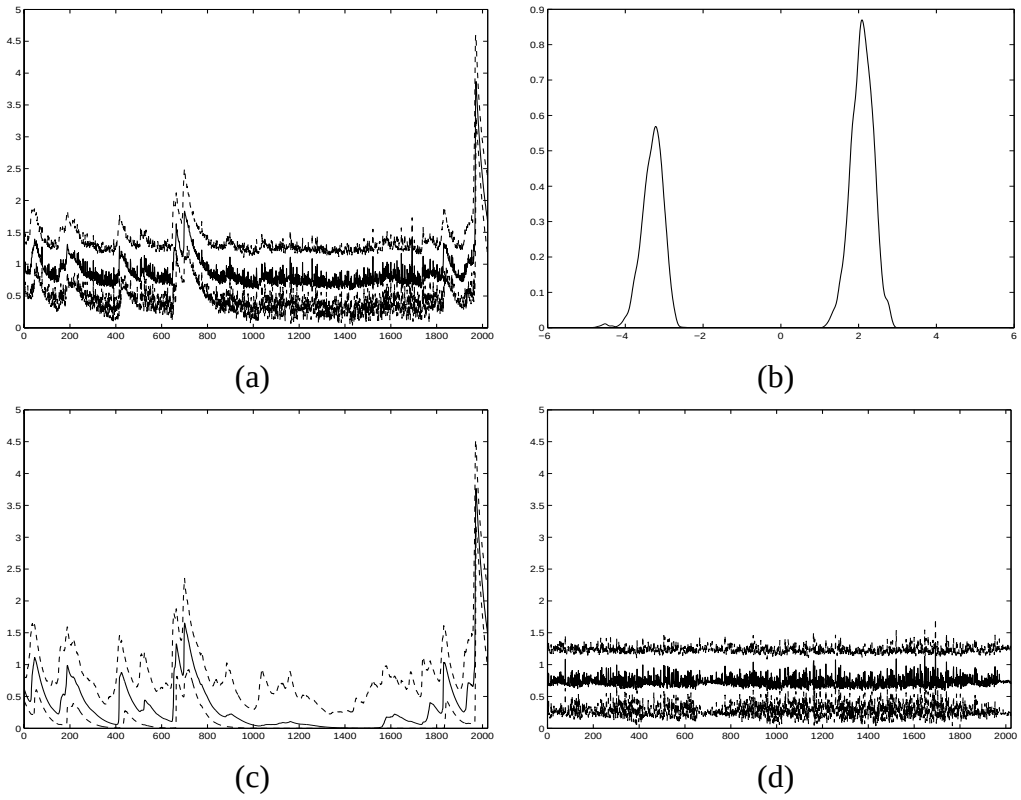
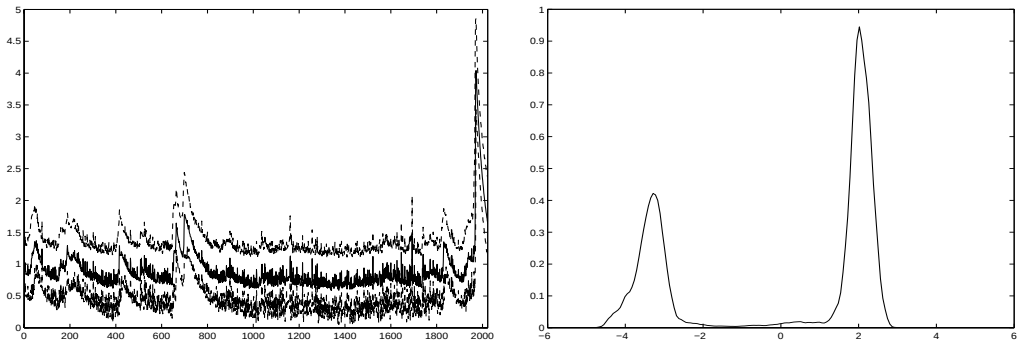


Figure 7: Posterior results for model with two components ($m = 2$): (a) σ_i, (b) $w_1\delta_{\ln(\lambda_1)} + w_2\delta_{\ln(\lambda_2)}$, (c) σ_{i1}, (d) σ_{i2}. In graphs (a), (c) and (d) drawn lines are posterior medians, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

the process to settle down at fairly low values between observations 1000 and 1800. Accordingly, the marginal mean volatility (see Table 1) is smaller, while the standard deviation has risen. Also, the marginal moments of the volatility process are more precisely estimated. In line with the expectations, the two component model enjoys much more support from the data than the one-component model. This can be assessed from the log marginal likelihood values, computed as in Section 5 and presented in Table 1. Moving to a superposition of three components ($m = 3$) leads to a virtually indistinguishable volatility process, displayed in Figure 8(a) and no improvement in marginal likelihood (Table 1). Indeed, as shown in Figure 8(b), the extra component corresponds to an intermediate λ, but has little weight. Thus, we shall mostly focus on the model with two components in the sequel. Note that all basic models lead to a risk premium β with most posterior mass on negative values, in contrast to expectations.

We now examine the inclusion of a pure jump component in the returns, as described in Subsection 2.3. The series representation in (26) will require truncation for this Gamma process, as explained in Subsection 3.1. We used $\varepsilon = 0.001$ in (28), which led to reliable results. Figure 9 indicates the sum of the weights of the m persistent components (*i.e.* $\sum_{j=1}^{m} w_j$) with $m = 1$ and 2, which suggests that inclusion of a jump component is not warranted by the data, once two persistent components are present in the volatility process. As a consequence, we shall retain the model with two persistent components as the main model when exploring further model additions.

A component-specific leverage effect, as in Subsection 2.3, is now introduced into the returns equa-



(a) (b)

Figure 8: Posterior results for model with three component ($m = 3$): (a) σ_i, (b) $w_1\delta_{\ln(\lambda_1)} + w_2\delta_{\ln(\lambda_2)} + w_3\delta_{\ln(\lambda_3)}$. In graph (a) the drawn lines is the posterior median, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

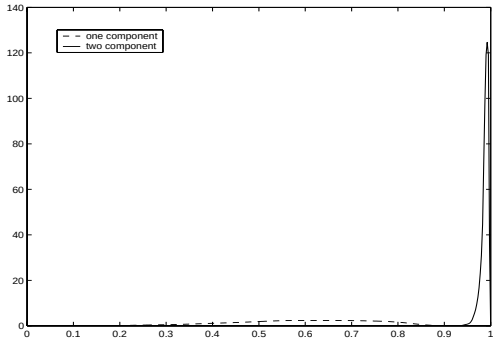


Figure 9: Posterior weight on the m persistent components in the presence of a pure jump component, $m = 1, 2$.

tion. Figure 10(a)-(d) presents the results on the volatility process (with two components) in the same format as Figure 7. The main differences with the non-leveraged model are a smaller weight on the fast-moving component and more persistence (smaller λ-values), especially for the slow-moving component. Figures 10(e) and (f) show that a number of large drops in returns are taken up by the leveraged mean function, so that the volatility does not need to reach such high peak values as before. Accordingly, the marginal standard deviation of the volatility is much lower, as displayed in Table 3. The risk premium coefficient is now more in line with expectations in the sense that most of its mass is on positive values. Leverage on the first (slow-moving) component, is concentrated on negative values, as expected; the fast-moving component, though, gets a mostly positive leverage coefficient. The marginal likelihood is much improved with respect to the basic models. Table 3 also presents evidence on two other models with roughly comparable support from the data. To implement leverage in the three-component model, we have added the first two components (with the smallest λ's) and $\rho_1^{(1)}$ is the associated (linear) leverage coefficient. The coefficient ρ_2 relates to the fast-moving (third) component. As before, the three-component model really adds a component with intermediate λ-values, but with very little weight attached to it. In fact, as illustrated by Figure 11, the most noticeable differences in σ_t (with respect to the two-component linear leverage model) are in the autumn of 1986 and after the October 1987 crash.

This is where the intermediate component kicks in and makes the volatility die down more rapidly in the case with $m = 3$. The two-component model with quadratic leverage on the first component required a prior negativity restriction on the coefficient $\rho_1^{(1)}$ to avoid a perverse sign of the effect. The inclusion of the quadratic leverage term reduces the weight on the fast-moving component even further and requires even smaller peak volatility levels (the median value of σ_i corresponding to the October 1987 crash is now only 2.2%).

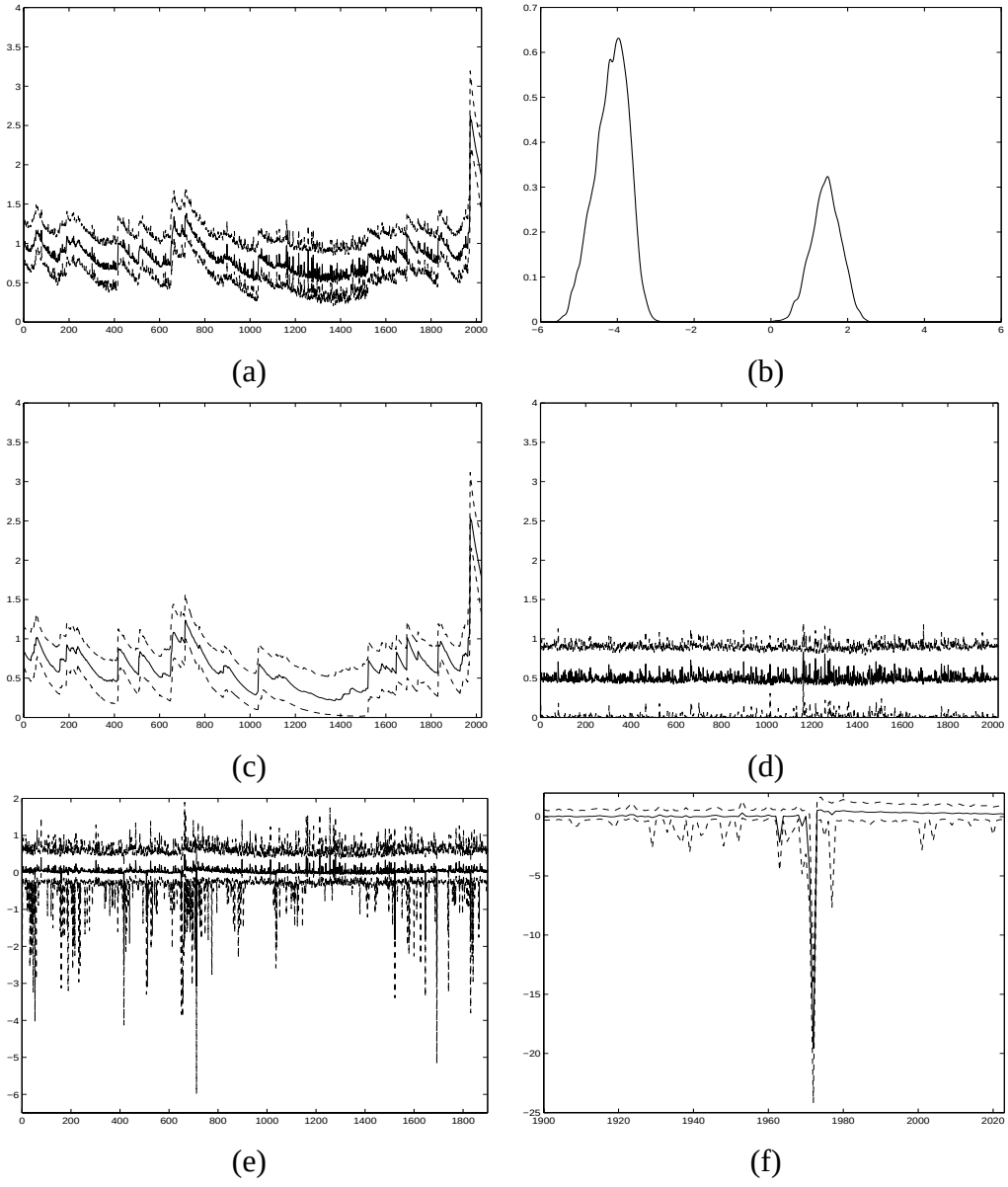


Figure 10: Posterior results for model with two components ($m = 2$) and linear leverage: (a) σ_i, (b) $w_1\delta_{\ln(\lambda_1)} + w_2\delta_{\ln(\lambda_2)}$, (c) σ_{i1}, (d) σ_{i2}, (e) mean function until July 7, 1987, (f) mean function as of July 8, 1987. In graphs (a), (c), (d), (e) and (f) drawn lines are posterior medians, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

Figure 12 indicates the difference between the linear and quadratic leverage effects in these models:

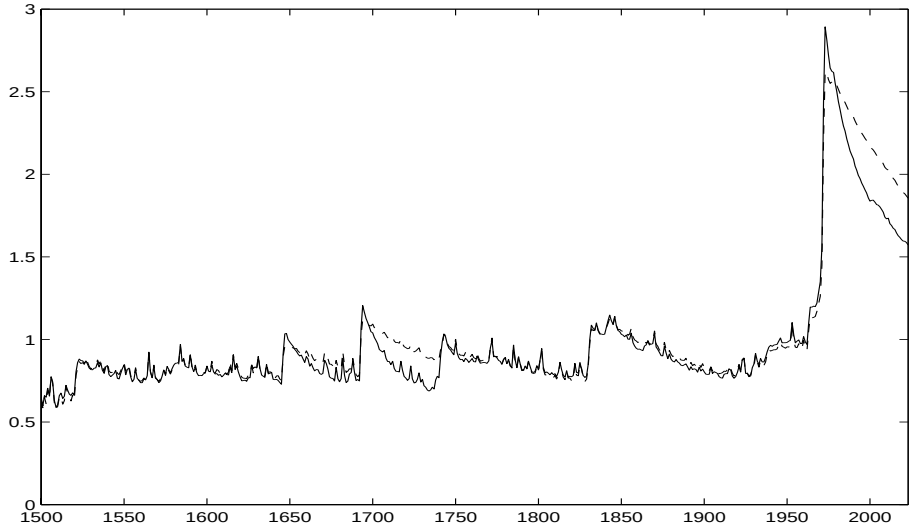


Figure 11: The median value of σ_i for 2 and 3 component-models with linear leverage: dashed line $m = 2$, drawn line $m = 3$.

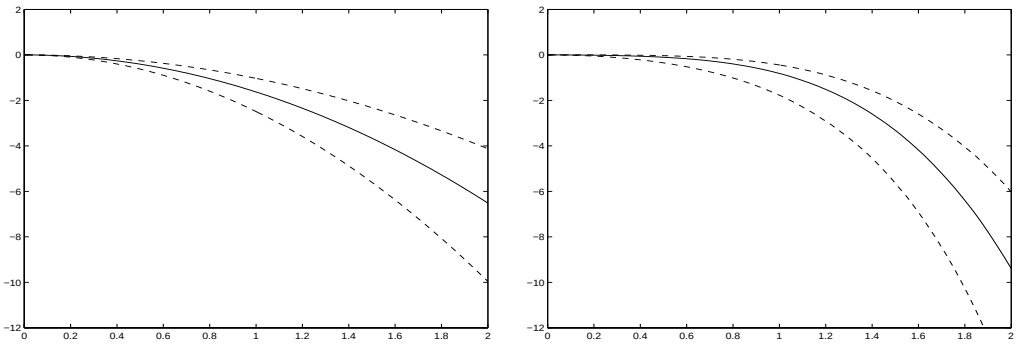
Table 3: Posterior medians, 95% credible intervals (in parentheses) and marginal likelihoods for leveraged model

	2 Components	3 Components (joint)	2 Components (quadr.)
μ	-0.02 (-0.11, 0.07)	-0.02 (-0.11, 0.07)	-0.05 (-0.15, 0.04)
β	0.07 (-0.04, 0.19)	0.07 (-0.04, 0.20)	0.10 (-0.02, 0.23)
$\rho_1^{(1)}$	-4.59 (-7.43, -2.62)	-3.51 (-5.89, -1.82)	-1.01 (-4.05, -0.04)
$\rho_1^{(2)}$			-6.40 (-18.47, -1.94)
ρ_2	0.23 (0.01, 1.01)	0.24 (0.03, 0.96)	0.39 (0.02, 2.76)
E$(\sigma^2(t))$	0.98 (0.77, 1.36)	1.03 (0.77, 1.44)	0.97 (0.73, 1.29)
St.Dev.$(\sigma^2(t))$	0.59 (0.43, 0.84)	0.67 (0.49, 1.02)	0.52 (0.73,1.29)
Log Marg.Lik.	-2465	-2470	-2473

the quadratic specification has little effect for small volatility changes, but the effect becomes more pronounced for large increases in volatility. If quadratic leverage is introduced in the three-component model (on the sum of the two most persistent components) the intermediate component becomes even less important and the marginal likelihood indicates the data favour this model substantially less than the two-component model with quadratic leverage.

Key characteristics of the predictive returns distribution for the two-component models with linear and quadratic leverage (on the first component) are presented in Table 4. In comparison with Table 2, it is clear that including linear leverage improves the potential of the model to capture the skewness and kurtosis of the data, whereas the model with a quadratic leverage effect closely captures the moments of the observed data, for all observational frequencies examined. Thus, we shall focus on the two-component model with a quadratic leverage effect on the first component in the sequel.

Finally, the introduction of component-specific risk premia is explored. Table 5 collects some results,



(a) (b)

Figure 12: The leverage effect for the first component in the two component model with: (a) linear leverage, (b) quadratic leverage. The number on the horizontal axis is the change in $\sigma_1(t)$. Drawn lines are posterior medians, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

Table 4: Predictive properties of the returns distribution for leveraged models ($m = 2$)

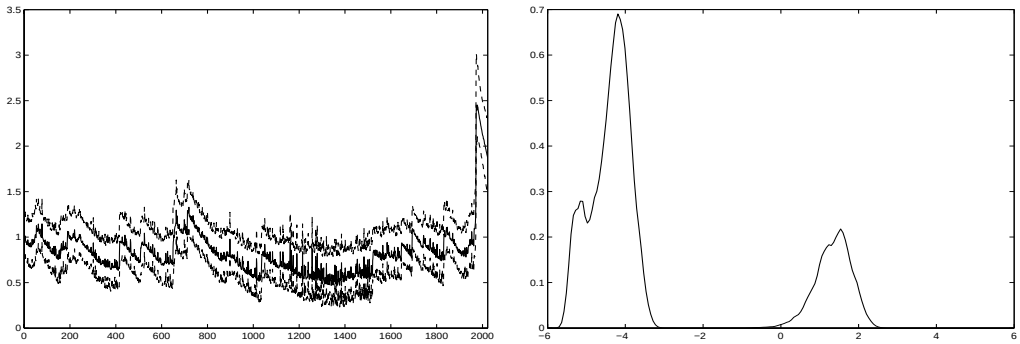
	Linear leverage				Quadratic leverage			
	$\Delta = 1$	$\Delta = 5$	$\Delta = 21$	$\Delta = 63$	$\Delta = 1$	$\Delta = 5$	$\Delta = 21$	$\Delta = 63$
Mean	0.0	0.2	1.0	3.0	0.0	0.2	1.0	2.9
Variance	1.2	6.1	26.1	79.4	1.4	7.0	29.5	87.6
Skewness	-0.6	-0.3	-0.2	-0.1	-3.6	-1.6	-0.7	-0.2
Kurtosis	8.5	5.7	7.0	5.0	105.2	25.6	9.5	5.2

indicating that the data heavily favour separate risk premia. The posterior results for the full model suggest that a model with no leverage on the second component is worth investigating, but such a restriction is not favoured by the data as shown in Table 5. In the presence of separate risk premia, we no longer need to impose any sign restrictions on the leverage coefficients. Figure 13 presents the posterior results for volatility and the λ's for the most favoured model corresponding to Table 5, which has a quadratic leverage effect on the first component, a linear leverage effect on the second component and separate risk premia for each component. In the sequel, we shall focus on this model and, for simplicity, we shall also denote it as the "retained" model. The effect of risk premia and leverage is summarized in Figure 14 where we plot the risk premium as a function of the cumulative distribution function (cdf) of the second component of the volatility as well as the leverage effect as a function of the change in $\sigma_1(t)$. In comparison with the quadratic leverage of the common risk premia model in Figure 12(b), the effect in Figure 14(b) stays close to zero for longer (note that there is no restriction on the sign of the effect in this specification), and then remains smaller in absolute value and better determined for large volatility shocks.

There are a number of ways in which the separate risk premia can be interpreted. For example, in our case with $m = 2$, we could define a function $w_i^A = \sigma_{i1}^2/\sigma_i^2$ which can be interpreted as a measure of market predictability, given that σ_{i1}^2 is a highly persistent (and thus predictable) process. For a given

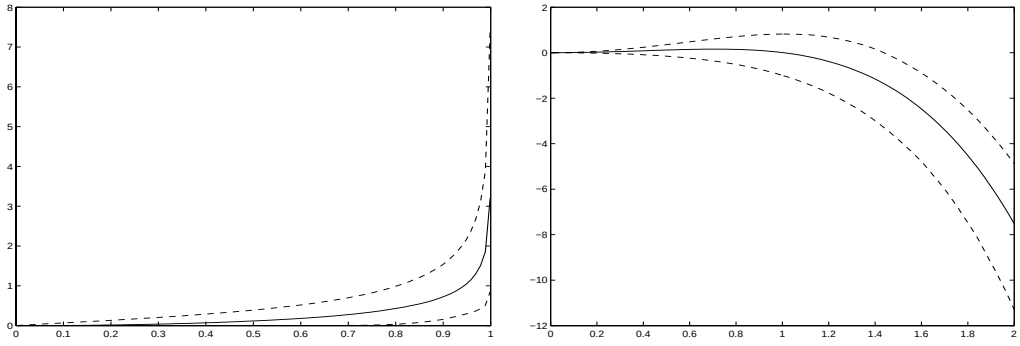
Table 5: Posterior medians, 95% credible intervals (in parentheses) and marginal likelihood for leveraged model with component-specific risk premia

μ	-0.28 (-0.61, -0.07)	-0.30 (-0.55, -0.10)
β_1	0.08 (-0.05, 0.21)	0.07 (-0.05, 0.20)
β_2	2.92 (0.50, 10.78)	2.89 (0.78, 8.90)
$\rho_1^{(1)}$	3.05 (-1.83, 9.10)	2.71 (-1.87, 8.79)
$\rho_1^{(2)}$	-15.27 (-34.27, -4.09)	-13.61 (-32.80, -4.82)
ρ_2	0.06 (-0.71, 1.02)	
E$(\sigma^2(t))$	0.85 (0.67, 1.11)	0.87 (0.70, 1.11)
SD$(\sigma^2(t))$	0.42 (0.32, 0.62)	0.44 (0.33, 0.60)
Log Marg.Lik.	-2433	-2443



(a) (b)

Figure 13: Posterior results for model with two components ($m = 2$), quadratic leverage and separate risk premia: (a) σ_i, (b) $w_1\delta_{\ln(\lambda_1)} + w_2\delta_{\ln(\lambda_2)}$. In graph (a) the drawn line is the posterior median, dashed lines are 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles.

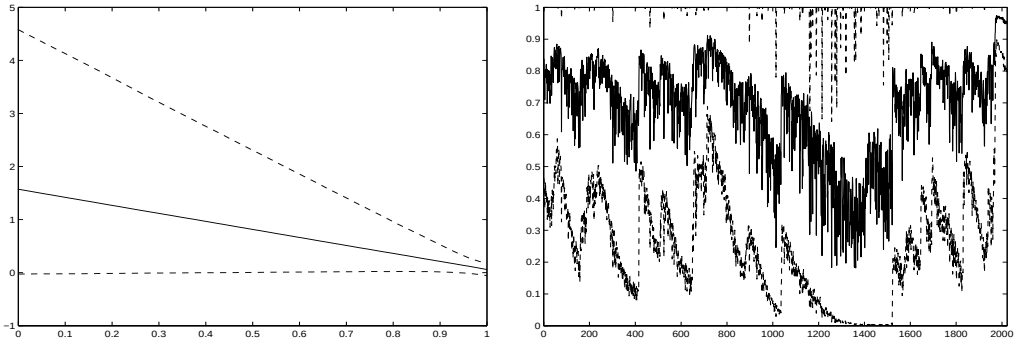


(a) (b)

Figure 14: Posterior results for model with two components ($m = 2$), quadratic leverage and separate risk premia: (a) risk premium on second component as a function of cdf of $\sigma_2^2(t)$, (b) leverage effect as a function of change in $\sigma_1(t)$. In both graphs drawn lines are posterior medians and dashed lines are 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles.

level of a volatility, w_i^A measures how much of that volatility is due to first component. The mean of a daily return, assuming no jumps, is then $\mu\Delta + \beta(w_i^A)\sigma_i^2$ where $\beta\left(w_i^A\right) = \beta_1 w_i^A + \beta_2\left(1 - w_i^A\right)$. Periods of low volatility are generally associated with small values of w^A; however, this relatively small volatility will be quite unpredictable and thus carry a higher risk premium.

The function $\beta\left(w^A\right)$ is illustrated in Figure 16 for the retained model. For comparison, Figure 15 plots this function for the counterpart of the retained model with linear leverage (*i.e.* $\rho_1^{(2)} = 0$). The posterior distribution of $w_1^A, \ldots, w_T^A$ is also shown in the two figures to give an indication of the plausible range of values. In both cases, $\beta\left(w^A\right)$ is close to zero, and well-determined, for large values of w^A. The median risk premium decreases with w^A which measures predictability. Also, for smaller values of w^A, the credible intervals become much larger. This is partly due to the leverage effect which is harder to distinguish from the risk premium for a rapidly varying volatility process. From Figures 15(b) and 16(b) we can also deduce that the weight on the fast-moving component is generally reduced by the introduction of quadratic leverage (as we found in the cases with common risk premium). Overall, the main difference between the common risk premium models and the models with component-specific risk premia lies in the premium of the second (fast-moving) component, which usually accounts for the smallest volatility contribution but is also the hardest to predict.

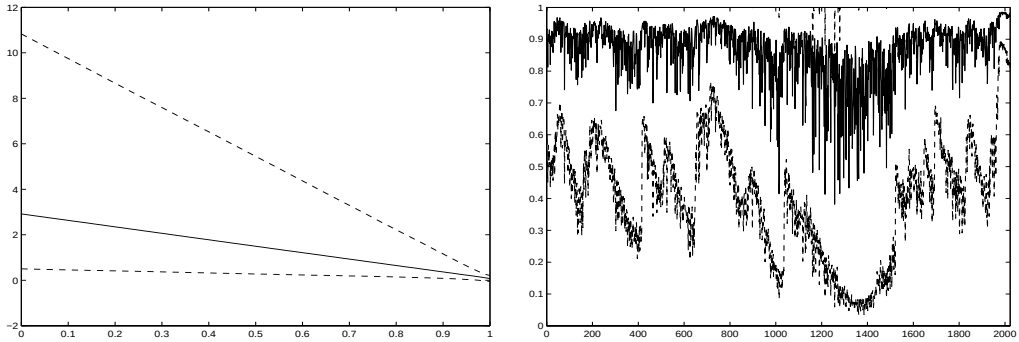


(a) (b)

Figure 15: Linear leverage model: (a) $\beta\left(w^A\right)$ as a function of w^A, (b) $w_i^A, i = 1, \ldots, T$. Drawn lines are posterior medians, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

Table 6 presents summary statistics of the predictive returns distribution for the retained model. The moments at the daily observation frequency ($\Delta = 1$, see Table 2) are closely matched by the model, which captures the higher moments well, and even tends to somewhat overestimate the kurtosis. Of course, the predictive moments presented here reflect two separate influences: the intrinsic properties of the sampling model given a set of parameter values, and the uncertainty regarding the parameter values. Whereas the latter uncertainty will not greatly affect the predictive mean, higher moments may contain an important parameter uncertainty component. In order to separate out these two aspects, Table 7 displays some posterior percentiles of the predictive moments of interest for $\Delta = 1$, computed by ranking the sampling moments obtained by simulation[11], conditionally upon the drawn posterior values

[11]For each of the 20,000 retained sets of parameter values, we simulate 100,000 volatility processes and derive the conditional



(a) (b)

Figure 16: Retained model: (a) $\beta\left(w^A\right)$ as a function of w^A, (b) $w_i^A, i = 1, \ldots, T$. Drawn lines are posterior medians, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

of the parameters. Clearly, parameter uncertainty mainly affects the higher moments, where it can have a substantial effect: for example, the median conditional kurtosis for the retained model is now 96.8 (close to the value observed for the data), while the predictive kurtosis (with parameter uncertainty integrated out) is 155.9 (see Table 6).

Table 6: Predictive properties of the returns distribution for two-component models with quadratic leverage and component-specific risk premia

	$\Delta = 1$	$\Delta = 5$	$\Delta = 21$	$\Delta = 63$
Mean	0.0	0.2	1.0	2.9
Variance	1.3	6.7	28.2	85.3
Skewness	-4.3	-1.7	-0.8	-0.4
Kurtosis	155.9	27.5	9.2	5.3

Table 7: Posterior medians and 95% credible intervals of the sampling moments of the returns distribution, $\Delta = 1$

	1 Component	2 Comp.; lin. lev.	2 Comp.; quadr. lev.	Retained model
Mean	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	0.0 (0.0, 0.1)	-0.1 (-0.2, 0.1)
Variance	1.2 (0.8, 1.7)	1.2 (0.9, 1.6)	1.4 (1.0, 1.9)	1.3 (1.1, 1.7)
Skewness	0.0 (-0.2, 0.1)	-0.5 (-1.0, -0.2)	-3.0 (-7.3, -1.2)	-3.6 (-8.5, -1.8)
Kurtosis	4.4 (4.0, 5.6)	7.5 (5.0, 13.1)	67.7 (20.9, 334.0)	96.8 (35.4, 481.9)

Besides the predictive moments, we can also consider the entire predictive distribution at various observational frequencies Δ. Figure 17 displays the log predictive densities of selected models, and illustrates the progression towards more (negative) skewness and kurtosis (mostly driven by the left tail).

predictive moments.

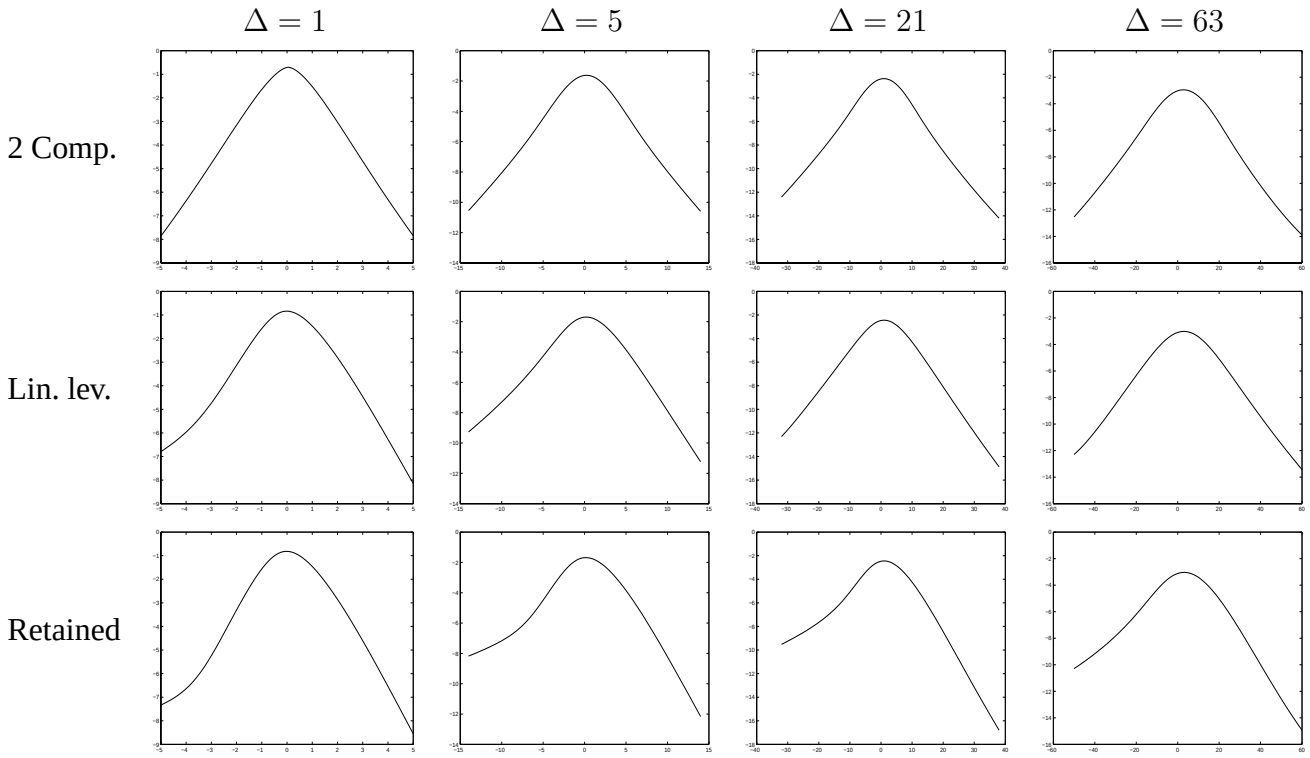


Figure 17: Log Predictive Densities: Columns correspond to various observational frequencies, and rows to models. The first row is for the basic two-component model; row 2 relates to the two-component model with linear leverage; row 3 corresponds to the retained two-component model with quadratic leverage and separate risk premia.

Finally, we consider some filtered results.[12] Figure 18 presents filtered medians and credible intervals for σ_i and for the predictive returns, as well as standardized predictive residuals (one-step-ahead forecast errors). The simple two-component model leads to large volatility spikes and, hence, considerable variation in the width of the predictive credible intervals. Allowing for leverage reduces the volatility surges and makes the one-step ahead predictive distribution a bit more concentrated. The retained model with separate risk premia leads to a further reduction in volatility (both peaks and mean levels) and results in a predictive spread which only substantially increases after the October 1987 crash. Predictive residuals for this model are quite satisfactory, except for the big crash (as could be expected).

7 Conclusions

We have proposed a simulation-based method for conducting Bayesian inference with the stochastic volatility models introduced in Barndorff-Nielsen and Shephard (2001a). These models combine a sub-

[12] As in Elerian *et al.* (2001), filtering was conducted conditionally upon the posterior mean values for the parameters. Taking means rather than medians ensures that the weights w_j add up to one. A simple particle filter, similar to that described in Barndorff-Nielsen and Shephard (2001a) was used. The resampling step was performed using the method described in Carpenter, Clifford and Fearnhead (1999).

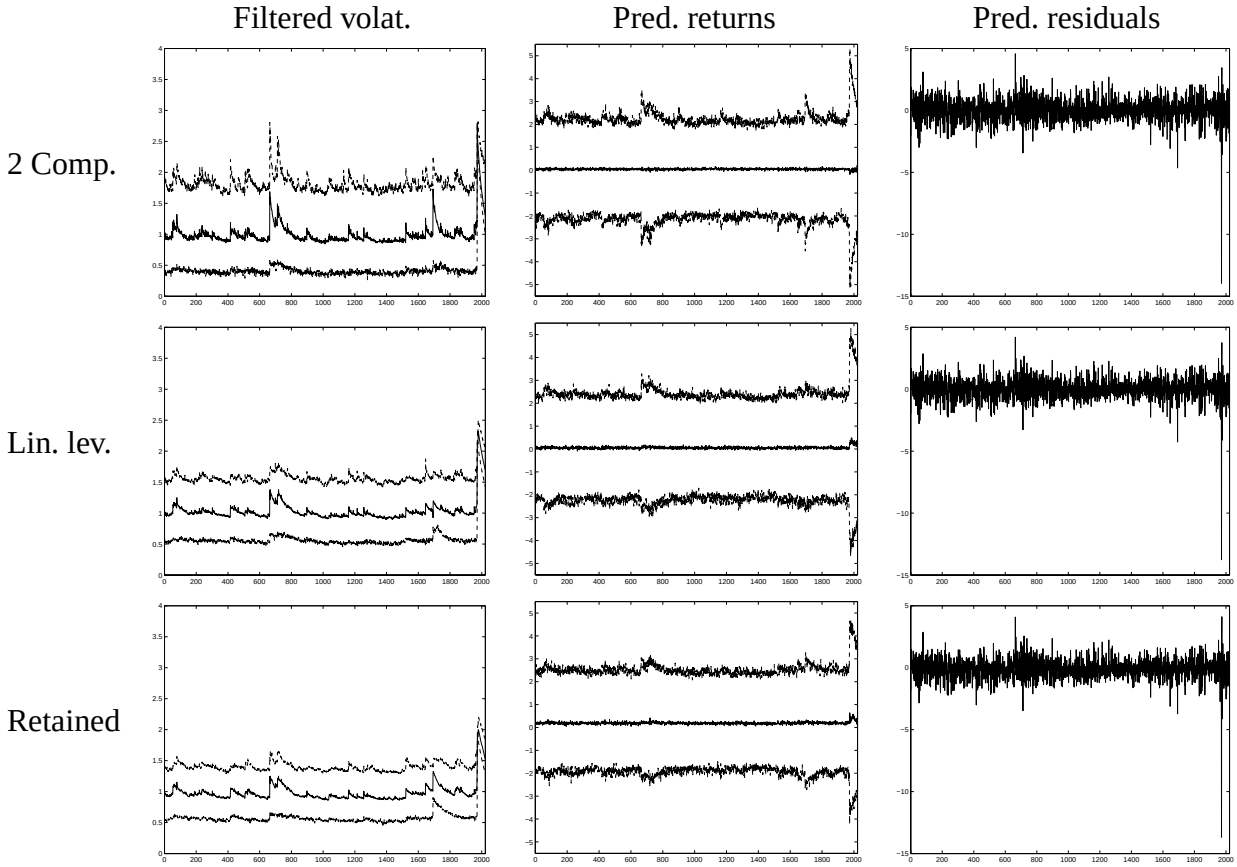


Figure 18: Filtered results: Columns show the filtered volatility series, the filtered predictive returns and the standardized one-step ahead predictive residuals for selected models. The first row corresponds to the basic two-component model; row 2 is for the two-component model with linear leverage; the final row corresponds to the retained two-component model with quadratic leverage and separate risk premia. In graphs for the filtered volatility and returns drawn lines are posterior medians, dashed lines are 2.5^{th} and 97.5^{th} percentiles.

stantial amount of flexibility with analytical tractability. This flexibility is partly induced by the volatility process, which is an Ornstein-Uhlenbeck process driven by a non-Gaussian Lévy process with positive jumps or a superposition of such Ornstein-Uhlenbeck processes. We add the possibility to include a jump component in the returns process. In addition, we extend the model with component-specific leverage effects (possibly using powers of processes) and risk premia, adding flexibility in the modelling of the mean function. Our inference method exploits the fact that, given the volatilities, we have a simple Gaussian model for the returns. Thus, we augment the parameter space with the volatilies and construct a joint sampler for the volatilities and the parameters in the model. We use a series representation to generate realizations of the volatility process. In the case where volatility has a marginal Gamma distribution and we have no jump component, this representation is finite, but in general it is not and, thus, requires truncation in practice. We propose a way to implement this truncation. The sampling of the volatility process given the parameters is conducted via Reversible Jump MCMC with three possible proposal distributions, and the parameters given the volatility process are sampled using dependent thinning of the

process.

An application to daily S&P 500 stock returns (over the years 1980-1987) illustrates the flexibility of the model and its ability to deal with very large shocks, especially when superpositions of processes are used for the volatility. In accordance with many other empirical studies, we find that a two-factor model for the volatility works well. In particular, we adopt a superposition of two persistent components, where one component is slow-moving and accounts for large, infrequent shocks and the other component is fast mean-reverting and behaves more like "noise". As we introduce leverage into the model, the relative weight of the fast-moving component decreases and the model is now able to account for the observed skewness and kurtosis of the data, especially with quadratic leverage. Component-specific risk premia improve the model fit even further. The best model in terms of marginal likelihood has a quadratic leverage effect on the first component, a linear leverage effect on the second component and risk premia for each component separately.

Improvements in terms of Bayes factors associated with the various model additions are quite large[13]. Going from the basic one-component model to the two-component model is favoured with a log Bayes factor of 138 (or a Bayes factor of $\exp(138) = 8.6 \cdot 10^{59}$), the inclusion of a leverage effect in the two-component model leads to a log Bayes factor of 56 (Bayes factor is $2.1 \cdot 10^{24}$), and the introduction of separate risk premia gives a further improvement of 32 in log marginal likelihood (a Bayes factor of $7.9 \cdot 10^{13}$). Overall, this leads to a massive Bayes factor of $\exp(226) = 1.4 \cdot 10^{98}$ in favour of our retained model with respect to the basic one-component model[14].

A Details of the Sampling Algorithm

Green (1995) introduces a method for Metropolis-Hastings sampling in problems with a variable dimension state space. The method relies on defining moves between state spaces of different dimension. Think of modelling an observable y given a parameter θ that can vary in dimension. To move between spaces of dimension m and m', we introduce random variables u and u' of dimensions t and t' so that $m + t = m' + t'$ and define a one-to-one transformation $(\theta, u) = (\theta', u')$. If we propose a move from dimension m to m' with probability $j(m, m')$, the acceptance rate for the transition from (m, θ) to (m', θ') is

$$\min\left\{1,\ \frac{p(y|\theta')p(\theta')}{p(y|\theta)p(\theta)}\ \frac{j(m',m)p(u')}{j(m,m')p(u)}\ \left|\frac{\partial(\theta',u')}{\partial(\theta,u)}\right|\right\}, \tag{33}$$

where, in practice, moves are usually set up such that either t or t' is zero.

[13] Given these large differences between models, Bayesian model averaging, as described in Section 5 would not lead to predictive results that are noticeably different from those on the basis of the best model.

[14] This has to be seen in relation to sample size, and implies an average Bayes factor for each observation conditionally upon the past of $\exp(226/T) = 1.12$.

A.1 Acceptance rate for updating the process

The acceptance rate for the sampler of the process given the parameters described in Subsection 3.3 can be calculated as follows. Consider proposing a new realisation of the Poisson process for observation i, Ψ_i', with a new number of jumps n_i' through

$$j(n_i, n_i') = q(n_i, n_i'), \;\; j(n_i', n_i) = q(n_i', n_i),$$

$$p(u) = p(\Psi_i'|\theta, \lambda, n_i') \text{ and } p(u') = p(\Psi_i|\theta, \lambda, n_i).$$

Since the new process is independent of the old process, the Jacobian is 1, and the acceptance probability in (33) reduces to

$$\min\left\{1, \frac{p(y|\Psi_{(i)}', N_{(i)}', \lambda, \theta)p(n_i'|\lambda, \theta)q(n_i', n_i)}{p(y|\Psi, N, \lambda, \theta)p(n_i|\lambda, \theta)q(n_i, n_i')}\right\},$$

where $\Psi_{(i)}'$ and $N_{(i)}'$ are defined as in the main text.

A.2 Details of independent proposals

The independence Metropolis-Hastings sampler in Subsection 3.3 uses a number of conditional expectations. Using that

$$\mathrm{E}(\eta_i|\lambda, \theta, \sigma^2((i-1)\Delta), n_i) = \begin{pmatrix} \alpha^{-1}n_i(\lambda\Delta)^{-1}(1-\exp\{-\lambda\Delta\}) \\ \alpha^{-1}n_i \end{pmatrix},$$

$$\mathrm{E}(\eta_i|\lambda, \theta, \sigma^2((i-1)\Delta)) = \begin{pmatrix} \alpha^{-1}\nu(1-\exp\{-\lambda\Delta\}) \\ \alpha^{-1}\lambda\nu\Delta \end{pmatrix},$$

and that we have defined

$$\sigma_i^2 = \lambda^{-1}\left(\eta_{i,2} - \eta_{i,1} + [1-\exp\{-\lambda\Delta\}]\sigma^2((i-1)\Delta)\right),$$

we obtain

$$\mathrm{E}(\sigma_i^2|\lambda, \theta, \sigma^2((i-1)\Delta), n_i) =$$
$$\lambda^{-1}\left(n_i\alpha^{-1}(1-(\lambda\Delta)^{-1}[1-\exp\{-\lambda\Delta\}]) + (1-\exp\{-\lambda\Delta\})\sigma^2((i-1)\Delta)\right),$$

$$\mathrm{E}(\sigma_i^2|\lambda, \theta, \sigma^2((i-1)\Delta)) =$$
$$\lambda^{-1}\left(\alpha^{-1}\lambda\nu\Delta - \alpha^{-1}\nu[1-\exp\{-\lambda\Delta\}] + (1-\exp\{-\lambda\Delta\})\sigma^2((i-1)\Delta)\right).$$

A.3 Acceptance rate for Dependent Thinning

The idea of dependent thinning for sampling the parameters is described in Subsection 3.4. Consider proposing a new value $\lambda' > \lambda$, the current value of λ. The set of original Poisson processes

is $\Psi = \{\Psi_1, \ldots, \Psi_T\}$ with $N = (n_1, \ldots, n_T)$ jumps and the proposed Poisson processes are $\Psi' = \{\Psi'_1, \ldots, \Psi'_T\}$ with $N' = (n_1, \ldots, n_T)$ jumps. The acceptance rate is

$$\min\left\{1, \frac{p(y|\Psi', N', \lambda', \theta)p(\Psi', N'|\lambda', \theta)p(\lambda'|\theta)}{p(y|\Psi, N, \lambda, \theta)p(\Psi, N|\lambda, \theta)p(\lambda|\theta)} \frac{j((N', \lambda'), (N, \lambda))}{j((N, \lambda), (N', \lambda'))p(u)} \left|\frac{\partial(\Psi')}{\partial(\Psi, u)}\right|\right\},$$

$$\text{with } j((N, \lambda), (N', \lambda')) = q(\lambda, \lambda'),\ j((N', \lambda'), (N, \lambda)) = q(\lambda', \lambda),$$

$$\text{and } p(u) = p(\{\Psi'_1 - \Psi_1\}, \ldots, \{\Psi'_T - \Psi_T\}|N', \lambda', \lambda, \theta).$$

Since $\{\Psi' - \Psi\}$ is independent of Ψ we have that

$$p(\Psi', N'|\lambda', \theta) = p(\{\Psi' - \Psi\}|N', \lambda', \lambda, \theta)p(\Psi, N|\lambda, \theta)$$

and the Jacobian is 1. As a consequence, the acceptance rate is

$$\min\left\{1, \frac{p(y|\Psi', N', \lambda', \theta)p(\lambda'|\theta)}{p(y|\Psi, N, \lambda, \theta)p(\lambda|\theta)} \frac{q(\lambda', \lambda)}{q(\lambda, \lambda')}\right\}.$$

A move to $\lambda' < \lambda$ will induce a loss in dimension, which will be implemented through a deterministic change to the process, ensuring a reversible step. To see that this sampler is reversible consider first moving from current state λ' to λ which involves a potential increase in dimension at every observation. If the reverse move λ to λ' follows then those new points, and only those points, will be deleted.

A.4 Simulated Data

The sampler was tested on a data set of $T = 2000$ values generated using the model in (3) with a zero mean function and a one-component stochastic volatility process. Thus, we focus only on the part which is really challenging for the sampler. The true values for the parameters were chosen to be "typical" values for stock price data, namely $\lambda = 0.01$, $\nu = 1$ and $\alpha = 1$. The results of running the MCMC sampler are presented in Figure 19 using a burnin of 15,000 iterations and retaining every tenth value for inference. The posterior distribution for all the parameters cover the true values. The parameters of the gamma marginal distribution place relatively little posterior mass around the true values. However, the posterior medians for the mean and the variance of the marginal distribution are 1.07 and 1.01 respectively, close to the unitary values implied by the true model. The fact that all posteriors have a tendency to favour values larger than the true values is in line with the prior, which has mean 1 for λ and 1000 for ν and α. The autocorrelation function for λ dies away after less than 300 lags but the mixing properties for ν are less impressive. However, given the complexity of the model, the behaviour is quite satisfactory, in our view.

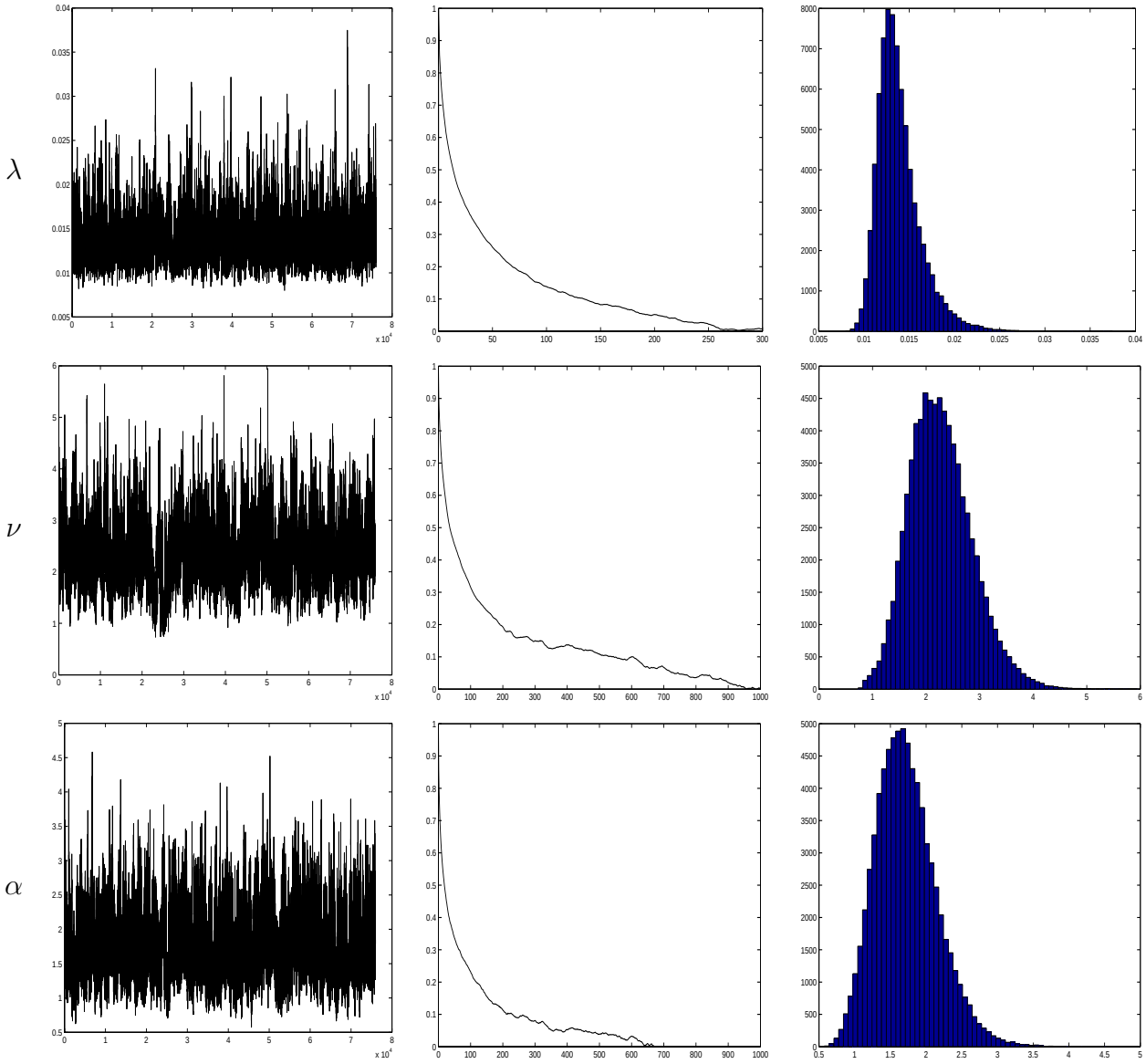


Figure 19: MCMC output for simulated data showing from left to right: the trace, the autocorrelation function and a histogram of drawn values; rows correspond to the parameters λ, ν and α.

References

Aït-Sahalia, Y. (2002): "Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach," *Econometrica*, 70, 223-262.

Alizadeh, S., M.W. Brandt, and F.X. Diebold (2002): "Range-Based Estimation of Stochastic Volatility Models," *Journal of Finance*, 57, 1047-1091.

Andersen, T.G., L. Benzoni, and J. Lund (2002): "An Empirical Investigation of Continuous Time Equity Return Models," *Journal of Finance*, 57, 1239-1284.

Barndorff-Nielsen, O.E. (1998): "Processes of Normal Inverse Gaussian Type," *Finance and Stochastics*, 2, 41-68.

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

Barndorff-Nielsen, O.E., Nicolato, E., and N. Shephard (2002): "Some Recent Developments in Stochastic Volatility Modelling," *Quantitative Finance*, 2, 11-23.

Barndorff-Nielsen, O.E., and N. Shephard (2001a): "Non-Gaussian OU Based Models and Some of Their Uses in Financial Economics," *Journal of the Royal Statistical Society, B (Statistical Methodology)*, 63, 167-241 (with discussion).

Barndorff-Nielsen, O.E., and N. Shephard (2001b): "Modelling by Lévy Processess for Financial Econometrics, " in *Lévy Processes – Theory and Applications*, ed. by O.E. Barndorff-Nielsen, T. Mikosch and S. Resnick, Boston: Birkhauser, 283-318.

Barndorff-Nielsen, O.E., and N. Shephard (2002): "Econometric Analysis of Realised Volatility and its Use in Estimating Stochastic Volatility Models," *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 64, 253-280.

Bates, D. (2000): "Post-'87 Crash Fears in S&P 500 Futures Options," *Journal of Econometrics*, 94, 181-238.

Bibby, B.M. and M. Sørensen (2003): "Hyperbolic processes in Finance," in S.T. Rachev (ed.): *Handbook of Heavy Tailed Distributions in Finance*, North Holland.

Bollerslev, T. and H. Zhou (2002): "Estimating Stochastic Volatility Diffusion Using Conditional Moments of Integrated Volatility", *Journal of Econometrics*, 109, 33-65.

Black, F. (1976): "Studies of Stock Price Volatility Changes," *Proceedings of the Business and Economic Statistics Section*, American Statistical Association, pp. 177-181.

Black, F. and M. Scholes (1973): "The Pricing of Options and Corporate Liabilities," *Journal of Political Economy*, 81, 637-654.

Campbell, J.Y. and A.S. Kyle (1993): "Smart Money, Noise Trading and Stock Price Behaviour," *Review of Economic Studies*, 60, 1-34.

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.

Carr, P., H. Geman, D.P. Madan and M. Yor (2002): "Stochastic Volatility for Lévy Processes," *Mathematical Finance*, 12, forthcoming.

Chernov, M., A.R. Gallant, E. Ghysels and G. Tauchen (2003): "Alternative Models for Stock Price Dynamics," *Journal of Econometrics*, forthcoming.

Chib, S. (1995): "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90, 1313-1321.

Cox, J.C., J.E. Ingersoll, and S.A. Ross (1985): "A Theory on the Term Structure of Interest Rates," *Econometrica*, 53, 385-407.

Duffie, D., J. Pan, and K. Singleton (2000): "Transform Analysis and Asset Pricing for Affine Jump Diffusions," *Econometrica*, 68, 1343-1376.

Durham, G.B., and A.R. Gallant (2002): "Numerical Techniques for Maximum Likelihood Estimation of Continuous-Time Diffusion Processes," *Journal of Business and Economic Statistics*, 20, 297-338 (with discussion).

Elerian, O., S. Chib, and N. Shephard (2001): "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69, 959-993.

Eraker, B. (2001): "MCMC Analysis of Diffusion Models With Applications to Finance," *Journal of Business and Economic Statistics*, 19, 177-191.

Eraker, B., Johannes, M., and Polson, N. (2003): "The Impact of Jumps in Volatility and Returns," *Journal of Finance*, 57, forthcoming.

Ferguson, T.S., and M.J. Klass (1972): "A Representation of Independent Increment Processes Without Gaussian Components," *Annals of Mathematical Statistics*, 43, 1634-1643.

Gallant, A.R., P.E. Rossi, and G. Tauchen (1992): "Stock Prices and Volume," *Review of Financial Studies*, 5, 199-242.

Green, P.J. (1995): "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination, *Biometrika*, 82, 711-732.

Hansen, M.B., K. Ickstadt, and R.L. Wolpert (2000): "Inverse Problems and a Lévy Process Solution," ISDS Discussion Paper 00-30, Duke University.

Heston, S.L. (1993): "A Closed Form Solution for Options with Stochastic Volatility, with Applications to Bond and Currency Options," *Review of Financial Studies*, 6, 327-343.

Hull, J., and A. White (1987): "The Pricing of Options on Assets With Stochastic Volatilities," *Journal of Finance*, 42, 281-300.

Jacquier, E., N.G. Polson, and P.E. Rossi (1994): "Bayesian Analysis of Stochastic Volatility Models," *Journal of Business and Economic Statistics*, 12, 371-417 (with discussion).

Jones, C. (2001): "Nonlinear Mean Reversion in the Short-Term Interest Rate," mimeo, Simon School of Business, University of Rochester.

Kim, S., N. Shephard and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models", *Review of Economic Studies*, 65, 361-393.

Meddahi, N., and E. Renault (2003): "Temporal Aggregation of Volatility Models", *Journal of Econometrics*, forthcoming.

Melino, A., and S. Turnbull (1990): "Pricing Foreign Currency Options With Stochastic Volatility," *Journal of Econometrics*, 45, 239-265.

Merton, R.C. (1973): "Theory of Rational Option Pricing," *Bell Journal of Economics*, 4, 141-183.

Nelson, D.B. (1991): "Conditional Heteroskedasticity in Asset Pricing: A New Approach," *Econometrica*, 59, 347-370.

Newton, M. and A. Raftery (1994): "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society, B*, 56, 3-48 (with discussion).

Nicolato, E., and E. Venardos (2003): "Option Pricing in Stochastic Volatility Models of the Ornstein-Uhlenbeck Type," *Mathematical Finance*, 13, forthcoming.

Pan, J. (2002): "The Jump-Risk Premia Implicit in Options: Evidence From an Integrated Time-Series Study," *Journal of Financial Economics*, 63, 3-50.

Pastorello, S., E. Renault, and N. Touzi (2000): "Statistical Inference for Random-Variance Option Pricing," *Journal of Business and Economic Statistics*, 18, 358-367.

Pedersen, A.R. (1995): "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations on Discrete Observations," *Scandinavian Journal of Statistics*, 27, 55-71.

Roberts, G.O., O. Papaspiliopoulos, and P. Dellaportas (2001): "Bayesian Inference for Non-Gaussian Ornstein-Uhlenbeck Stochastic Volatility Processes," mimeo, Department of Mathematics and Statistics, University of Lancaster.

Roberts, G.O., and O. Stramer (2001): "On Inference for Partially Observed Nonlinear Diffusion Models Using the Metropolis-Hastings Algorithm," *Biometrika*, 88, 603-621.

Sato, K. (1999): "*Lévy Processes and Infinitely Divisible Distributions*," Cambridge University Press: Cambridge.

Shin, H.S. (2003): "Disclosures and Asset Returns," *Econometrica*, 71, 105-133.

Sundaresan, S.M. (2000): "Continuous-Time Methods in Finance: A Review and an Assessment," *Journal of Finance*, 55, 1569-1633.

Vasicek, O. (1977): "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, 5, 177-188.

Yu, J., and P.C.B. Phillips (2001): "A Gaussian Approach for Continuous Time Models of the Short-Term Interest Rate," *Econometrics Journal*, 4, 211-225.