J. R. Statist. Soc. B (2001)
63, *Part* 2, *pp.* 167–241

Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics

Ole E. Barndorff-Nielsen

University of Aarhus, Denmark

and Neil Shephard

Nuffield College, Oxford, UK

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, October 18th, 2000*, Professor P. J. Diggle *in the Chair*]

Summary. Non-Gaussian processes of Ornstein–Uhlenbeck (OU) type offer the possibility of capturing important distributional deviations from Gaussianity and for flexible modelling of dependence structures. This paper develops this potential, drawing on and extending powerful results from probability theory for applications in statistical analysis. Their power is illustrated by a sustained application of OU processes within the context of finance and econometrics. We construct continuous time stochastic volatility models for financial assets where the volatility processes are superpositions of positive OU processes, and we study these models in relation to financial data and theory.

Keywords: Background driving Lévy process; Econometrics; Lévy density; Lévy process; Long-range dependence; Option pricing; Ornstein–Uhlenbeck processes; Particle filter; Stochastic volatility; Subordination; Superposition

1. Introduction

1.1. *Motivation*

Non-Gaussian processes of Ornstein–Uhlenbeck (OU) type have considerable potential as building-blocks for stochastic models of observational series from a wide range of fields. They offer the possibility of capturing important distributional deviations from Gaussianity and for flexible modelling of dependence structures. This paper aims at developing this potential, drawing on and extending powerful results from probability theory for applications in statistical analysis. We illustrate their power by a sustained application of OU processes within the context of finance and econometrics. On the basis of well-known (empirical) stylized facts, we construct continuous time stochastic volatility (SV) models for financial assets where the volatility processes are superpositions of positive OU processes, and we study these models in relation to financial data and theory. The study has also required the development of new numerical methods and these are discussed in detail.

The general definition of an OU process $y(t)$ is as the solution of a stochastic differential equation of the form

$$\mathrm{d}y(t) = -\lambda\, y(t)\, \mathrm{d}t + \mathrm{d}z(t) \tag{1}$$

Address for correspondence: Neil Shephard, Nuffield College, Oxford, OX1 1NF, UK.
E-mail: neil.shephard@nuf.ox.ac.uk

 1369–7412/01/63167

where z, with $z(0) = 0$, is a (homogeneous) Lévy process, i.e. a process with independent and stationary increments (see, for example, Rogers and Williams (1994), pages 73–84, Bertoin (1996, 1999), Protter and Talay (1999) and Sato (1999)). Familiar special cases of Lévy processes are Brownian motion and the compound Poisson process. All Lévy processes except for Brownian motion have jumps. As z is used to drive the OU process we shall call $z(t)$ a background driving Lévy process (BDLP) in this context.

Our interest in this paper will be in the existence and properties of stationary solutions to equation (1) in cases where z has no Gaussian component and the increments of z are positive, implying positivity of the process y. We shall write a continuous time stationary and non-negative latent process $\sigma^2(t)$ as representing the changing volatility underlying a financial asset. The simplest OU-based model for $\sigma^2(t)$ will have

$$\mathrm{d}\sigma^2(t) = -\lambda\, \sigma^2(t)\, \mathrm{d}t + \mathrm{d}z(\lambda t), \qquad \lambda > 0. \tag{2}$$

The unusual timing $\mathrm{d}z(\lambda t)$ is deliberately chosen so that it will turn out that whatever the value of λ the marginal distribution of $\sigma^2(t)$ will be unchanged. Hence we separately parameterize the distribution of the volatility and the dynamic structure. The process $z(t)$ has positive increments and no drift. This type of process is often called a subordinator (Bertoin (1996), chapter 3). Correspondingly $\sigma^2(t)$ moves up entirely by jumps and then tails off exponentially. This type of model has been used in storage theory by, for example, Cinlar and Pinsky (1972), Harrison and Resnick (1976) and Brockwell *et al.* (1982). Extensions to the autoregressive moving average (ARMA) case are discussed by Brockwell (2001). However, under the models that we have in mind small jumps are predominant. Although having OU dynamics looks restrictive, we shall show that we can construct more flexible processes by the addition of independent OU processes.

The main advantage of these OU processes is that they offer plenty of analytic tractability which is not available for more standard models such as geometric Gaussian OU processes and constant elasticity of volatility processes. For geometric Gaussian OU processes, $\log\{\sigma^2(t)\}$ is assumed to follow a Gaussian OU process. For constant elasticity of volatility processes

$$\mathrm{d}\sigma^2(t) = -\lambda\{\sigma^2(t) - \zeta\}\, \mathrm{d}t + \delta\, \sigma^2(t)^k\, \mathrm{d}b(t),$$

where $b(t)$ is standard Brownian motion, $k \geqslant \frac{1}{2}$. The former is highlighted by Hull and White (1987) whereas the latter is used extensively by Meddahi and Renault (1996). For example integrated volatility, which in finance is a key measure,

$$\begin{aligned}\sigma^{2*}(t) &= \int_0^t \sigma^2(u)\, \mathrm{d}u \\ &= \lambda^{-1}\{1 - \exp(-\lambda t)\}\, \sigma^2(0) + \lambda^{-1} \int_0^t [1 - \exp\{-\lambda(t-s)\}]\, \mathrm{d}z(\lambda s) \\ &= \lambda^{-1}\{z(\lambda t) - \sigma^2(t) + \sigma^2(0)\},\end{aligned} \tag{3}$$

has a simple structure. (All integrated processes will be denoted by having a superscript asterisk. The main examples are integrated volatility and intensity and the log-price level of a stock.)

A more general class of processes, which is also quite mathematically tractable, is given by

$$\sigma^2(t) = \int_{-\infty}^0 f(s)\, \mathrm{d}z(\lambda t + s),$$

for bounded, positive $f(\cdot)$ and with z as above. To be technically precise, $\{z(t)\}_{t \geqslant 0}$ is assumed to be caglad (non-decreasing with right continuous paths) and $\{z(-t)\}_{t \geqslant 0}$ is an independent copy of $\{-z(t)\}_{t \geqslant 0}$ but modified to be also caglad. Further, $f(\cdot)$ must be a positive function tailing off sufficiently fast to ensure the existence of the integral. In particular if $f(s) = \exp(s)$ we recover the OU processes. Given $f(\cdot)$ such a process is stationary and positive. This type of process is reminiscent of a standard infinite order linear moving average model.

1.2. Stochastic volatility processes

Continuous time models built out of Brownian motion play a crucial role in modern finance, providing the basis of most option pricing, asset allocation and term structure theory currently being used. An example is the so-called Black–Scholes or Samuelson model which models the logarithm of an asset price by the solution to the stochastic differential equation

$$\mathrm{d}x^*(t) = \{\mu + \beta\sigma^2\}\,\mathrm{d}t + \sigma\,\mathrm{d}w(t), \qquad t \in [0, S], \tag{4}$$

where $w(t)$ is standard Brownian motion. (We have used $x^*(t)$ to denote the price level as this is an integrated process.) This means aggregate returns over intervals of length $\Delta > 0$ are

$$y_n = \int_{(n-1)\Delta}^{n\Delta} \mathrm{d}x^*(t) = x^*(n\Delta) - x^*\{(n-1)\Delta\} \tag{5}$$

implying returns are normal and independently distributed with a mean of $\mu\Delta + \beta\sigma^2\Delta$ and a variance of $\Delta\sigma^2$. Unfortunately for moderate to small values of Δ (corresponding to returns measured over 5-minute to 1-day intervals) returns are typically heavy tailed, exhibit volatility clustering (in particular the $|y_n|$ are correlated) and are skew (see the discussion in, for example, Campbell *et al.* (1997), pages 17–21), although for higher values of Δ a central limit theorem seems to hold and so Gaussianity becomes a less poor assumption for $\{y_n\}$ in that case. This means that every single assumption underlying the Black–Scholes model is routinely rejected by the type of data that are usually used in practice.

This common observation, which carries over to the empirical rejection of option pricing models based on this model, has resulted in an enormous effort to develop empirically more reasonable models which can be integrated into finance theory. The most successful of these are the generalized autoregressive conditional heteroscedastic (GARCH) and the diffusion-based SV processes. This very large literature, which was started by Clark (1973), Engle (1982) and Taylor (1982), is reviewed in, for example, Bollerslev *et al.* (1994), Ghysels *et al.* (1996) and Shephard (1996).

Our model will also be of an SV type, based on a more general stochastic differential equation,

$$\mathrm{d}x^*(t) = \{\mu + \beta\,\sigma^2(t)\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}w(t), \tag{6}$$

where $\sigma^2(t)$, the instantaneous volatility, will be assumed to be stationary, latent and stochastically independent of $w(t)$. Even though $\sigma^2(t)$ exhibits jumps $x^*(t)$ is a continuous process for all parameter values. This formulation also makes it clear that in the special case where $\mu = \beta = 0$ an SV process can be thought of as a subordinated Brownian motion. We shall delay our discussion of this well-known connection until Section 6. Instead our earlier sections will focus on our main innovation, which will be to use OU processes to model $\sigma^2(t)$. We do this as it will allow us to gain a much better analytic understanding than conventional diffusion-based SV models do.

SV models in general, by appropriate design of the stochastic process for $\sigma^2(t)$, allow aggregate returns $\{y_n\}$ to be heavy tailed, skewed, to exhibit volatility clustering and to aggregate to Gaussianity as Δ becomes large. To see why this happens, whatever the model for σ^2, it follows that

$$y_n|\sigma_n^2 \sim N(\mu\Delta + \beta\sigma_n^2, \sigma_n^2).$$

where

$$\begin{aligned} \sigma_n^2 &= \sigma^{2*}(n\Delta) - \sigma^{2*}\{(n-1)\Delta\}, \\ \sigma^{2*}(t) &= \int_0^t \sigma^2(u)\,\mathrm{d}u. \end{aligned} \tag{7}$$

So returns are scaled mixtures of normals, where the scaling is typically time dependent, inducing dependence in the returns. Hence this model class can produce empirically reasonable models. For example, if $\sigma^2(t)$ has an inverse Gaussian law then y_n will be approximately a normal inverse Gaussian (NIG) variable. In turn, these models allow us to think about the appropriate implications for the pricing of derivatives written on underlying assets obeying SV processes. We shall do this in Section 5 and Section 6.2.

It is possible to generalize equation (6) to allow for the feed-back of the innovations of the volatility process into the level of the asset price. In particular, we write

$$\mathrm{d}x^*(t) = \{\mu + \beta\,\sigma^2(t)\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}w(t) + \rho\,\mathrm{d}\bar{z}(\lambda t), \tag{8}$$

where $\bar{z}(t) = z(t) - E\{z(t)\}$, the centred version of the BDLP. This allows the model to deal with the so-called leverage type of problem that is associated with the work of Black (1976) and Nelson (1991) which formalizes the observation that for equities a fall in the price is associated with an increase in future volatility. We shall discuss some aspects of this model in Section 4.

1.3. Structure of the paper

This paper has six other sections and an appendix. In Section 2 we discuss the detailed mathematical construction behind the OU processes that we favour, focusing on building appropriate BDLPs. We show that they are sufficiently flexible to allow us to design models to fit marginal features of the distribution of returns as well as to deal separately with the observed dependence structure in the returns. As this section is quite technical, readers whose main interest is in the SV aspect of this paper could skip it on their first reading. Related, more advanced, technical details may be found in our second paper on this topic: Barndorff-Nielsen and Shephard (2000). Section 3 looks at the construction of volatility models by the addition of OU processes. This provides a way of constructing a wide class of dynamics for volatility, including (quasi-)long memory models. In Section 4 we give results for the temporal aggregation of returns from a continuous time SV model. This allows us to relate our linear SV models to the popular GARCH discrete time models associated with the work of Engle (1982). In Section 5 we discuss the empirical fitting of these models by using linear and non-linear methods. We show that it is not straightforward to implement likelihood-based estimation procedures for our models, although various moment-based methods are simple to use. Section 6 discusses various additional issues such as multivariate extensions of the models and the precise connection between SV and subordination, as well as showing formally that SV models do not allow for arbitrage and giving results on the pricing of derivatives written using an SV model. Section 7 concludes. Appendix A collects various proofs and derivations which we have omitted from the main text of the paper.

2. Construction of Ornstein–Uhlenbeck processes

2.1. Definition and existence

Before we discuss the SV models in detail we shall introduce the mathematical basis of the OU processes, showing how they are constructed and how to simulate from them.

The stationary process σ^2 is of OU type if it is representable as

$$\sigma^2(t) = \int_{-\infty}^{0} \exp(s)\,\mathrm{d}z(\lambda t + s) \tag{9}$$

in which case it may also be written as

$$\sigma^2(t) = \exp(-\lambda t)\,\sigma^2(0) + \int_0^t \exp\{-\lambda(t-s)\}\,\mathrm{d}z(\lambda s).$$

Here $z = \{z(t)\colon t \in R\}$ is a (homogeneous) Lévy process and λ is a positive number. When this is the case $\sigma^2(t)$ satisfies the stochastic differential equation (2). The process $z(t)$ is termed the BDLP or subordinator corresponding to the process $\sigma^2(t)$. A simulated example of the paths that the $\sigma^2(t)$ and $z(\lambda t)$ processes follow is given in Fig. 1.

In essence, given a one-dimensional distribution D (not necessarily restricted to the positive half-line) there is a stationary process of OU type (i.e. satisfying a stochastic differential

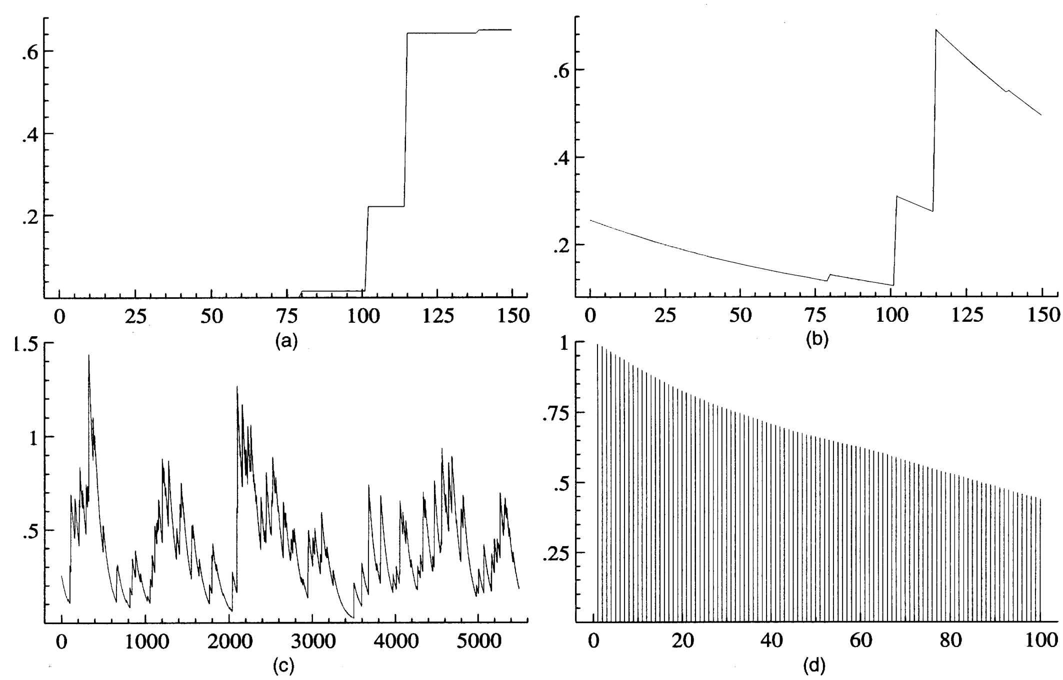

Fig. 1. OU process with $\Gamma(\nu, \alpha)$ marginals (throughout, $\nu = 3$, $\alpha = 8.5$, $\lambda = 0.01$ and $\Delta = 1$): (a) $z(\lambda n\Delta)$ against n (short series, BDLP); (b) $\sigma^2(n\Delta)$ against n (short series, volatility process); (c) $\sigma^2(n\Delta)$ against n (long series, volatility process); (d) empirical autocorrelation function for $\sigma^2(n\Delta)$ (correlogram)

equation of form (1)) whose one-dimensional marginal law is D if and only if D is *self-decomposable*, i.e. if and only if the characteristic function ϕ of D satisfies $\phi(\zeta) = \phi(c\zeta)\,\phi_c(\zeta)$ for all $\zeta \in \mathbf{R}$ and all $c \in (0, 1)$ and for some family of characteristic functions $\{\phi_c\colon c \in (0, 1)\}$. This restriction does, however, still leave a great flexibility in the choice of D. The precise statement of existence is as follows; cf. Wolfe (1982) and Jurek and Vervaat (1983) (see also Barndorff-Nielsen *et al.* (1998)).

Theorem 1. Let ϕ be the characteristic function of a random variable x. If x is self-decomposable, i.e. if

$$\phi(\zeta) = \phi(c\zeta)\,\phi_c(\zeta)$$

for all $\zeta \in \mathbf{R}$ and all $c \in (0, 1)$, then there is a stationary stochastic process $x(t)$ and a Lévy process $z(t)$ such that $x(t) =^{\mathcal{L}} x$ and

$$x(t) = \int_{-\infty}^{t} \exp\{-\lambda(t-s)\}\,\mathrm{d}z(\lambda s) = \int_{-\infty}^{0} \exp(\lambda u)\,\mathrm{d}z\{\lambda(t+u)\} = \int_{-\infty}^{0} \exp(u)\,\mathrm{d}z(\lambda t + u) \tag{10}$$

for all $\lambda > 0$.

Conversely, if $x(t)$ is a stationary stochastic process and $z(t)$ is a Lévy process such that $x(t) =^{\mathcal{L}} x$ and $x(t)$ and $z(t)$ satisfy equation (10) for all $\lambda > 0$ then x is self-decomposable.

If the stationary OU process $\sigma^2(t)$ is square integrable, it has autocorrelation function $r(u) = \exp(-\lambda|u|)$. It will be helpful later to establish the notation that the cumulant-generating functions for $\sigma^2(t)$ and $z(1)$ (if they exist) be written as

$$\acute{k}(\theta) = \log(E[\exp\{-\theta\,\sigma^2(t)\}])$$

and

$$k(\theta) = \log(E[\exp\{-\theta\,z(1)\}])$$

respectively. Indeed they are related by the fundamental equality (Barndorff-Nielsen, 2000)

$$\acute{k}(\theta) = \int_0^\infty k\{\theta\,\exp(-s)\}\,\mathrm{d}s, \tag{11}$$

which can be re-expressed as

$$k(\theta) = \theta\,\acute{k}'(\theta) \tag{12}$$

(where $\acute{k}'(\theta) = \mathrm{d}\acute{k}(\theta)/\mathrm{d}\theta$). It then follows that if we write the cumulants of $\sigma^2(t)$ and $z(1)$ (when they exist) as respectively $\acute{\kappa}_m$ and κ_m ($m = 1, 2, \ldots$) we have that $\kappa_m = m\acute{\kappa}_m$, for $m = 1, 2, \ldots$.

2.2. Lévy densities

Suppose that we choose a probability distribution D on the positive half-line which is self-decomposable. Then, as just discussed, there is a strictly stationary OU process

$$\sigma^2(t) = \exp(-\lambda t)\,\sigma^2(0) + \int_0^t \exp\{-\lambda(t-s)\}\,\mathrm{d}z(\lambda s) \tag{13}$$

such that $\sigma^2(t) \sim D$ and where z is a Lévy process. The increments of z are positive and

$$k(\theta) = \log(E[\exp\{-\theta\,z(1)\}]) = -\int_{0+}^{\infty} \{1 - \exp(-\theta x)\}\,W(\mathrm{d}x), \tag{14}$$

where W is the Lévy measure of the Lévy–Khintchine representation for $z(1)$. We shall generally assume that W has a density w. It is related to the Lévy density u of $\sigma^2(t)$ by the formula

$$w(x) = -u(x) - x\,u'(x) \tag{15}$$

(this presupposes that u is differentiable) and, letting

$$W^{+}(x) = \int_x^{\infty} w(y)\,\mathrm{d}y, \tag{16}$$

we have, moreover,

$$W^{+}(x) = x\,u(x) \tag{17}$$

(Barndorff-Nielsen, 1998a). Finally, we shall denote the inverse function of W^{+} by W^{-1}, i.e.

$$W^{-1}(x) = \inf\{y > 0\colon W^{+}(y) \leqslant x\}.$$

2.3. Models via D

One approach to model building is to write down a specific parametric form for D and then to calculate the implied behaviour of the BDLP. We do this here for the generalized inverse Gaussian (GIG) marginal law $\sigma^2(t) \sim \mathrm{GIG}(\nu, \delta, \gamma)$. (The standard notation for the GIG distribution is $\mathrm{GIG}(\lambda, \delta, \gamma)$; however, the notation λ was not available to us.) The GIG class seems particularly interesting as a plausible model basis for volatility models as special cases have been extensively used (though in different contexts from the present) particularly in various recent papers. See, in particular, Eberlein and Keller (1995), Barndorff-Nielsen (1997, 1998a), Rydberg (1999) and Eberlein (2000). Recall that if $x \sim \mathrm{GIG}(\nu, \delta, \gamma)$ then it has a density

$$\frac{(\gamma/\delta)^{\nu}}{2\,K_{\nu}(\delta\gamma)} x^{\nu-1} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}, \qquad x > 0, \tag{18}$$

where K_{ν} is a modified Bessel function of the third kind. When δ or γ is 0, the norming constant in the formula for the density of the GIG distribution must be interpreted in the limiting sense, using the well-known results that for $x \downarrow 0$ we have

$$K_{\nu}(x) \sim \begin{cases} -\log(x) & \text{if } \nu = 0, \\ 2^{|\nu|-1}\Gamma(|\nu|)x^{-|\nu|} & \text{if } \nu \neq 0. \end{cases}$$

Special cases of the GIG density are

(a) the inverse Gaussian law, where $\nu = -\frac{1}{2}$,
(b) the positive hyperbolic law where $\nu = 1$,
(c) the inverse χ^2-law with df degrees of freedom where $\nu = -\mathrm{df}/2$, $\delta = \surd\mathrm{df}$ and $\gamma = 0$, and
(d) Γ, where $\delta = 0$ and $\nu > 0$.

Of course if $\sigma^2 \sim \mathrm{GIG}(\nu, \delta, \gamma)$ and is independent of $\epsilon \sim N(0, 1)$, then $x = \mu + \beta\sigma^2 + \sigma\epsilon$ is the generalized hyperbolic distribution. If we define $\alpha = \surd(\beta^2 + \gamma^2)$, then the density is

$$\frac{(\gamma/\delta)^{\nu}}{\surd(2\pi)\alpha^{\nu-1/2}\,K_{\nu}(\delta\gamma)}\{\delta^2 + (x-\mu)^2\}^{(\nu-1/2)/2}\,K_{(\nu-1/2)}[\alpha\surd\{\delta^2 + (x-\mu)^2\}]\,\exp\{\beta(x-\mu)\}. \tag{19}$$

Hence a continuous time volatility model built using a volatility model of OU type with GIG marginals will have generalized hyperbolic marginals for instantaneous returns. Special cases of this include the NIG distribution, the hyperbolic and the Student t. These distributions have been studied in the context of finance in Prause (1998) and Raible (1998).

It is known that the GIG(ν, δ, γ) law is self-decomposable (Halgreen, 1979) so stationary OU processes with GIG marginals do exist. The following theorem specifies the Lévy measure.

Theorem 2. The Lévy measure of the GIG distribution is absolutely continuous with density

$$u(x) = x^{-1}\left\{\frac{1}{2}\int_0^\infty \exp\left(-\frac{1}{2}\delta^{-2}x\xi\right) g_\nu(\xi)\,\mathrm{d}\xi + \max(0,\nu)\right\} \exp\left(-\frac{\gamma^2 x}{2}\right) \tag{20}$$

where

$$g_\nu(x) = \frac{2}{x\pi^2}\{J_{|\nu|}^2(\sqrt{x}) + N_{|\nu|}^2(\sqrt{x})\}^{-1}$$

and J_ν and N_ν are Bessel functions.

For a proof see Appendix A.

For the definitions and properties of Bessel functions see, for example, Gradstheyn and Ryzhik (1965), pp. 958–971.

We note that the Bessel functions have simple forms when $|\nu|$ is half odd. We shall now discuss four special cases of this result.

2.3.1. *GIG($-\frac{1}{2}$, δ, γ): inverse Gaussian*

Its law means that $\sigma^2(t) \sim \mathrm{IG}(\delta, \gamma)$ whose density is

$$\frac{\delta}{\sqrt{(2\pi)}}\exp(\delta\gamma)x^{-3/2}\exp\left\{-\frac{1}{2}(\delta^2x^{-1}+\gamma^2x)\right\}, \qquad x>0, \tag{21}$$

where the parameters δ and γ satisfy $\delta > 0$ and $\gamma \geqslant 0$. We find that the upper tail integral (recalling that $W^+(x) = x\,u(x)$) is

$$W^+(x) = \frac{\delta}{\sqrt{(2\pi)}}x^{-1/2}\exp\left(-\frac{1}{2}\gamma^2x\right). \tag{22}$$

2.3.2. *GIG(1, δ, γ): positive hyperbolic distribution*

The density of the positive hyperbolic distribution is

$$\frac{\gamma/\delta}{2\,K_1(\delta\gamma)}\exp\left\{-\frac{1}{2}(\delta^2x^{-1}+\gamma^2x)\right\}, \qquad x>0,$$

where the parameters δ and γ satisfy $\delta > 0$ and $\gamma \geqslant 0$. When the law of $\sigma^2(t)$ is positive hyperbolic we find that the upper tail integral is

$$W^+(x) = \left\{\delta^2\int_0^\infty \exp(-x\xi)\,g_1(2\delta^2\xi)\,\mathrm{d}\xi + 1\right\}\exp\left(-\frac{\gamma^2x}{2}\right). \tag{23}$$

2.3.3. *GIG($-\nu$, δ, 0): reciprocal gamma distribution*

The reciprocal gamma distribution (i.e. the law of the reciprocal of a gamma variate) has density

$$\frac{\alpha^{\nu}}{\Gamma(\nu)} x^{-\nu-1} \exp(-\alpha x^{-1}), \qquad x > 0, \quad \nu > 0, \quad \alpha = \delta^2/2.$$

The corresponding upper tail integral is

$$W^{+}(x) = \tfrac{1}{2} \int_0^{\infty} \exp(-\tfrac{1}{4}\alpha^{-1} x\xi)\, g_{\nu}(\xi)\, \mathrm{d}\xi. \tag{24}$$

2.3.4. GIG($\nu > 0$, 0, γ): gamma distribution

The gamma marginal law has probability

$$\frac{\alpha^{\nu}}{\Gamma(\nu)} x^{\nu-1} \exp(-\alpha x), \qquad x > 0, \quad \alpha = \frac{\gamma^2}{2}.$$

This has the corresponding upper tail integral of the Lévy density $W^{+}(x) = \nu \exp(-\alpha x)$, which has the convenient property that it can be analytically inverted:

$$W^{-1}(x) = \max\left\{0, \ -\frac{1}{\alpha} \log\left(\frac{x}{\nu}\right)\right\}. \tag{25}$$

2.4. Models via the background driving Lévy process

Instead of specifying a model for $\sigma^2(t)$ and working out the density for the BDLP, it is possible to go the other way and to construct the model through the BDLP. Of course there are constraints on valid BDLPs which must be satisfied. Specifically a necessary and sufficient condition for the stochastic differential equation

$$\mathrm{d}x(t) = -\lambda\, x(t)\, \mathrm{d}t + \mathrm{d}z(\lambda t) \tag{26}$$

to have a stationary solution is that $E[\log\{1 + |z(1)|\}] < \infty$ (cf. Wolfe (1982) and Jurek and Mason (1993), theorem 3.6.6).

Lemma 1. Let z be a Lévy process with positive increments and cumulant function

$$\log(E[\exp\{-\theta\, z(1)\}]) = -\int_{0+}^{\infty} \{1 - \exp(-\theta x)\}\, W(\mathrm{d}x),$$

and assume that

$$\int_1^{\infty} \log(x)\, W(\mathrm{d}x) < \infty. \tag{27}$$

Suppose moreover, for simplicity, that the Lévy measure W has a differentiable density w, and define the function u on R_{+} by

$$u(x) = \int_1^{\infty} w(\tau x)\, \mathrm{d}\tau. \tag{28}$$

Then u is the Lévy density of a random variable x of the form

$$x = \int_0^{\infty} \exp(-s)\, \mathrm{d}z(s)$$

and the specification

$$x(t) = \int_{-\infty}^{t} \exp\{-\lambda(t-s)\}\,\mathrm{d}z(s)$$

determines a stationary process $\{x(t)\}_{t\in R}$ with z as its BDLP.

Proof. The proof may be concluded from a more general result given in Jurek and Mason (1993), theorem 3.6.6.

2.4.1. *Example 1*

We give a simple valid construction which allows easy simulation and analytic results for the implied density of $\sigma^2(t)$. Let W be a Lévy measure determined in terms of its tail integral by

$$W^{+}(x) = cx^{-\epsilon}(1+x)^{-\beta}\exp(-\tfrac{1}{2}\gamma^2 x)$$

where c is a positive constant, $0 \leqslant \epsilon < 1$, $0 \leqslant \beta$, $0 \leqslant \gamma$ and $\max(\beta-1, \gamma) > 0$. Then

$$w(x) = c\{\epsilon x^{-1} + \beta(1+x)^{-1} + \tfrac{1}{2}\gamma^2\}x^{-\epsilon}(1+x)^{-\beta}\exp(-\tfrac{1}{2}\gamma^2 x). \tag{29}$$

Hence lemma 1 applies and ensures the existence of an OU process $\sigma^2(t)$ whose BDLP $z(t)$ has w as the Lévy density of $z(1)$. Furthermore, recalling that the Lévy density u of $\sigma^2(t)$ satisfies $x\,u(x) = W^{+}(x)$, we find that

$$u(x) = cx^{-1-\epsilon}(1+x)^{-\beta}\exp(-\tfrac{1}{2}\gamma^2 x).$$

For $\epsilon = \frac{1}{2}$ and $\beta = 0$ we recover the inverse Gaussian law for $\sigma^2(t)$. If $\gamma = 0$, implying $\beta > 1$, then for the moments of $\sigma^2(t)$ we have

$$E\{\sigma^2(t)^{\nu}\} < \infty \qquad \text{if and only if } \nu < \beta + \epsilon.$$

Furthermore, the jth-order cumulant of $\sigma^2(t)$ $(j < \beta+\epsilon)$ is $c\,B(j-\epsilon, \beta+\epsilon-j)$ where $B(x, y)$ denotes the beta function.

The idea of modelling by choice of Lévy density rather than probability density has been introduced into the study of turbulence by Novikov (1994) and Koponen (1995) to capture the distributional characteristics of distributions of velocity differences in high Reynolds number turbulent fluids (where, in fact, NIG laws generally give very good fits; for an example, see Barndorff-Nielsen (1998b)). Related work is discussed in Cont *et al.* (1997) and Mantegna and Stanley (2000).

2.5. *Simulation via series representations*

A crucial feature of our approach will be that we simulate from the volatility process

$$\sigma^2(t) = \exp(-\lambda t)\,\sigma^2(0) + \int_0^t \exp\{-\lambda(t-s)\}\,\mathrm{d}z(\lambda s)$$

to simulate returns from the $x^*(t)$ process and so to analyse data. To be able to do that we shall have to simulate from

$$\exp(-\lambda t)\int_0^{\lambda t} \exp(s)\,\mathrm{d}z(s), \tag{30}$$

rather than from the BDLP $z(s)$ itself. One approach to this is to simulate directly from the

Lévy processes and then to approximate the corresponding integrals. This is difficult owing to the jump character of the processes. Instead we use infinite series representations of these types of integrals. The required results are, in essence, available from work of Marcus (1987) and Rosiński (1991). A self-contained exposition of this result is given in Barndorff-Nielsen and Shephard (2000), whereas recent developments are surveyed in Rosiński (2000); see also Protter and Talay (1999), Ferguson and Klass (1972), Vervaat (1979) and Walker and Damien (2000). The last three papers discuss, in particular, simulation procedures in line with those considered in the present paper but for non-homogeneous Lévy processes satisfying a regularity condition. Again we let W be the Lévy measure of $z(1)$ and W^{-1} denote the inverse of the tail mass function W^{+}. Then the desired result is that

$$\int_0^{\lambda} f(s)\,\mathrm{d}z(s) \overset{\mathcal{L}}{=} \sum_{i=1}^{\infty} W^{-1}(a_i/\lambda)\, f(\lambda r_i). \tag{31}$$

Here the $\{a_i\}$ and $\{r_i\}$ are two independent sequences of random variables with the r_i independent copies of a uniform random variable r on [0, 1] and $a_1 < \ldots < a_i < \ldots$ as the arrival times of a Poisson process with intensity 1.

Our practical experience with using expression (31) is that it is quite quickly converging; however, theory suggests that it must be used carefully. Consider the special case of the inverse Gaussian model; then equation (22) implies $W^{-1}(x)$ will, for large values of x, behave essentially as x^{-2}. This is studied in more detail in Barndorff-Nielsen and Shephard (2000).

2.5.1. *Example 2: gamma–Ornstein–Uhlenbeck ($\Gamma(\nu, \alpha)$ marginals) process*

We need a method to sample from expression (30). We have already noted the expression for $W^{-1}(x)$ in equation (25). Thus, defining $c_1 < c_2 < \ldots$ as the arrival times of a Poisson process with intensity $\nu\lambda t$ and $N(1)$ as the corresponding number of events until time 1, then

$$\begin{aligned}
\exp(-\lambda t) \int_0^{\lambda t} \exp(s)\,\mathrm{d}z(s) &\overset{\mathcal{L}}{=} \exp(-\lambda t) \sum_{i=1}^{\infty} W^{-1}(a_i/\lambda t)\, \exp(\lambda t r_i) \\
&= -\alpha^{-1} \exp(-\lambda t) \sum_{i=1}^{\infty} \mathbf{1}_{]0,\nu[}(a_i/\lambda t)\, \log(a_i/\nu\lambda t)\, \exp(\lambda t r_i) \\
&= \alpha^{-1} \exp(-\lambda t) \sum_{i=1}^{\infty} \mathbf{1}_{]0,1[}(c_i)\, \log(c_i^{-1})\, \exp(\lambda t r_i) \\
&= \alpha^{-1} \exp(-\lambda t) \sum_{i=1}^{N(1)} \log(c_i^{-1})\, \exp(\lambda t r_i).
\end{aligned} \tag{32}$$

To illustrate these results we simulate a regularly spaced OU gamma process $\sigma^2(n\Delta)$ using the above representation for the parameter values $\Delta = 1$, $\nu = 3$, $\lambda = 0.01$ and $\alpha = 8.5$. The results are presented in Fig. 1. There we graph both $z(\lambda n\Delta)$ and $\sigma^2(n\Delta)$ against time using only a small range of values of n, which shows the jumps in the process. Of course the $z(\lambda n\Delta)$ process is a non-decreasing integrated process, whereas the $\sigma^2(n\Delta)$ process is stationary. For the larger series we see that the jumps look less extreme and instead our eyes tend to focus on the large upward movements in the OU process followed by slower declines. The final picture is the corresponding empirical autocorrelation function of the $\sigma^2(n\Delta)$ process. Finally, it is worth noting that the simulation is very fast for OU gamma processes. Over many different parameter values we could produce processes of length of half a million in around 5 s on a modern personal computer using the `Ox` programming language of Doornik (1998).

3. Superposition

Although we have focused on the simplest OU volatility process, our model and technique extend to where volatility follows a weighted sum of independent OU processes with different persistence rates, i.e.

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

where

$$\sum_{j=1}^{m} w_j^+ = 1,$$

with

$$\mathrm{d}\sigma_j^2(t) = -\lambda_j \sigma_j^2(t)\,\mathrm{d}t + \mathrm{d}z_j(\lambda_j t),$$

where the $\{z_j(t)\}$ are independent (not necessarily identically distributed) BDLPs. In such a case we would have a process for the price of the type

$$\mathrm{d}x^*(t) = \{\mu + \beta\sigma^2(t)\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}w(t) + \sum_{j=1}^{m} \rho_j\,\mathrm{d}\bar{z}_j(\lambda_j t),$$

where $\bar{z}_j(t) = z_j(t) - E\{z_j(t)\}$, allowing the leverage effect to be different for the various components of volatility.

By the adding together of independent OU processes with different persistence rates we obtain more general correlation patterns in the volatility structure. This implies an autocorrelation function which is a weighted sum of exponentials

$$r(u) = w_1 \exp(-\lambda_1|u|) + \ldots + w_m \exp(-\lambda_m|u|), \tag{33}$$

where the w_i are positive and sum to 1. Hence some of the components of the volatility may represent short-term variation in the process whereas others represent long-term movements. Alternative, discrete time, empirical models of this are discussed by Engle and Lee (1999), Dacorogna *et al.* (1998) and Barndorff-Nielsen (1998a).

By choosing the weights and damping factors in equation (33) appropriately and letting $m \to \infty$ it is possible to construct tractable volatility models with long-range or quasi-long-range dependence. In particular, Barndorff-Nielsen (2000) shows that there is a limiting model for which

$$r(u) = (1 + \lambda|u|)^{-2(1-H)}$$

with $\lambda > 0$ and $H \in (\frac{1}{2}, 1)$ being the long memory parameter. (Barndorff-Nielsen (2000) constructed this, and more general models, not by a limiting procedure, but in terms of the theory of independently scattered measures and Lévy random fields.) Similar types of arguments have previously been used for real-valued time series models by, for example, Granger (1980) and Cox (1991). Ding and Granger (1996) have studied long memory in volatility using the addition of short memory processes whereas Andersen and Bollerslev (1997a) have used the theory of heterogeneous information arrivals to motivate a long memory volatility model. Finally, Comte and Renault (1998) constructed a long-range dependent SV model by writing the logarithm of the instantaneous volatility as fractional Brownian motion.

It is possible to extend this to multifractal behaviour where

$$r(u) = \sum_{i=1}^{m} w_i(1 + \lambda_i|u|)^{-2(1-H_i)}, \qquad H_i \in (\tfrac{1}{2}, 1), \quad \lambda_i > 0,$$

and where the w_i are positive and sum to 1. These types of continuous time models imply that discrete returns have long memory features.

4. Aggregation results

4.1. Behaviour of $x^*(t)$, the log-price

In this section we shall study the behaviour of integrals, or aggregations, of the instantaneous returns $\mathrm{d}x^*(t)$. There will be two points of focus. First, in this subsection we shall look at the log-price itself $x^*(t)$, recalling that $x^*(0)$ is defined to be 0. The second focus, developed in the next subsection, will be on characterizing the dependence structure of the returns $\{y_n\}$, defined in equation (5) as the change in $x^*(t)$ over non-overlapping intervals of length Δ.

First we shall state some general results for the non-leverage SV models given in equation (6) with arbitrary SV processes; then we shall go on to produce a complete description of the behaviour of $x^*(t)$ in the OU volatility case allowing $\rho \neq 0$. In general we have that if we write (when they exist) ξ, ω^2 and r respectively as the mean, variance and the autocorrelation function of the process $\sigma^2(t)$ then

$$E\{\sigma^{2*}(t)\} = \xi t,$$
$$\mathrm{var}\{\sigma^{2*}(t)\} = 2\omega^2\, r^{**}(t),$$

where

$$r^*(t) = \int_0^t r(u)\,\mathrm{d}u,$$
$$r^{**}(t) = \int_0^t r^*(u)\,\mathrm{d}u \tag{34}$$

(we use $r^{**}(t)$ to denote the double integral over the autocorrelation function). A consequence of this result is that

$$E\{x^*(t)\} = (\mu + \beta\xi)t,$$
$$\mathrm{var}\{x^*(t)\} = t\xi + 2\beta^2\,\omega^2\, r^{**}(t),$$

whereas, when $\mu = \beta = 0$,

$$\mathrm{var}\{x^*(t)^2\} = 6\omega^2\, r^{**}(t) + 2\xi^2 t^2.$$

Further we have that if $\sigma^2(u)$ is ergodic then, as $t \to \infty$,

$$t^{-1}\,\sigma^{2*}(t) = t^{-1} \int_0^t \sigma^2(u)\,\mathrm{d}u \to \xi, \qquad \text{almost surely},$$

implying, for the SV model, that $t^{-1/2}\{x^*(t) - \mu t - \beta\,\sigma^{2*}(t)\}$ is asymptotically normal with mean 0 and variance ξ (i.e. the log-returns tend to normality for long lags—a similar result has been known within the ARCH class since Diebold (1988), pages 12–16). This follows from the subordination interpretation of the SV models discussed in Section 6.1. The convergence of $t^{-1/2}\{x^*(t) - \mu t - \beta\sigma^{2*}(t)\}$ to normality will, however, be slow in the case where the process $\sigma^2(t)$ exhibits long-range dependence.

As $x^*(t)$ is the sum of a continuous local martingale (see Section 6) and a continuous bounded variation process, its quadratic variation (QV) is $\sigma^{2*}(t)$, i.e. we have

$$[x^*](t) = \operatorname*{plim}_{r\to\infty}\left[\sum\{x^*(t_{i+1}^r) - x^*(t_i^r)\}^2\right] = \sigma^{2*}(t) \tag{35}$$

for any sequence of partitions $t_0^r = 0 < t_1^r < \ldots < t_{m_r}^r = t$ with $\sup_i(t_{i+1}^r - t_i^r) \to 0$ for $r \to \infty$. The QV estimation of integrated volatility has recently been highlighted, following the initial draft of this paper and the concurrent independent work of Andersen and Bollerslev (1998a), by Andersen *et al.* (2000) in foreign exchange markets.

When we assume that $\sigma^2(t)$ is an OU process then we can strengthen some of these results to give a complete description of the leveraged $x^*(t)$ process (8) via its cumulant-generating functional. The formula is in terms of the cumulant function k for the BDLP of $\sigma^2(t)$. However, it can easily be recast in terms of the cumulant function $\acute{k}$ for $\sigma^2(t)$; see formulae (11) and (12). Let f denote an 'arbitrary' function; then the log-characteristic-function of $f\cdot x^*$, which we interpret as the stochastic integral $\int_0^\infty f(s)\,\mathrm{d}x^*(s)$ (Protter, 1992), is

$$C\left(\zeta\ddagger\int_0^\infty f\cdot x^*\right) = \lambda\int_0^\infty [k\{J\exp(-\lambda s)\} + k\{H(s)\}]\,\mathrm{d}s + i\zeta(\mu - \lambda\rho\xi)\int_0^\infty f(s)\,\mathrm{d}s \tag{36}$$

where

$$J = \int_0^\infty \{\tfrac{1}{2}\zeta^2 f^2(u) - i\zeta\beta f(u)\}\exp(-\lambda u)\,\mathrm{d}u \tag{37}$$

and

$$H(s) = \int_0^\infty \{\tfrac{1}{2}\zeta^2 f^2(s+u) - i\zeta\beta f(s+u)\}\exp(-\lambda u)\,\mathrm{d}u - i\zeta\rho f(s). \tag{38}$$

The derivation of this result is given in Barndorff-Nielsen and Shephard (2000). It is important to understand the full scope of this expression. It gives a calculus for computing all the cumulants for any weighted sum of the path of the log-price. In other words this is a full description of the whole process.

Expressions for the cumulant functions of the finite dimensional distributions of the x^*-process are directly obtainable from equation (36) by a suitable choice of f. As an illustration, we consider the cumulant function for $x^*(t)$ for an arbitrary value of t. For notational simplicity we suppose that $\mu = \beta = \rho = 0$; an extension to the general case causes no substantial difficulty. Letting $f = \mathbf{1}_{[0,t]}$ we find, after a little algebra,

$$C\{\zeta\ddagger x^*(t)\} = \lambda\int_0^\infty k[\tfrac{1}{2}\zeta^2\lambda^{-1}\{1 - \exp(-\lambda t)\}\exp(-\lambda s)]\,\mathrm{d}s + \lambda\int_0^t k[\tfrac{1}{2}\zeta^2\lambda^{-1}\{1 - \exp(-\lambda s)\}]\,\mathrm{d}s.$$

Note that from this formula the cumulants of $x^*(t)$ are explicitly expressible in terms of the cumulants of $z(1)$ or, alternatively, of $\sigma^2(t)$.

4.1.1. *Example 3*

Suppose that $\sigma^2(t) \sim \mathrm{IG}(\delta, \gamma)$, as in expression (21); then $\acute{k}(\theta) = \delta\gamma\{1 - (1 + 2\theta/\gamma^2)^{1/2}\}$ and so, by formula (12),

$$k(\theta) = \frac{\delta\theta}{\gamma}\left(1 + \frac{2\theta}{\gamma^2}\right)^{-1/2} = \sum_{m=1}^{\infty} \kappa_m (-1)^{m-1} \frac{\theta^m}{m!},$$

where

$$\kappa_m = m \frac{\delta}{\gamma}\left(\frac{2}{\gamma^2}\right)^{m-1} \binom{\frac{1}{2}}{m-1}.$$

Hence, for instance, the variance of $x^*(t)$ is seen to be $\kappa_m(t) = (\delta/\gamma)t$, as could, of course, also have been found by simple direct calculation.

4.2. Dependence of returns

In this subsection we derive the moments of discrete time returns implied by a general continuous time SV model. In particular when μ and β are 0 then, using the definitions given in equations (34),

$$\mathrm{cov}(\sigma_n^2, \sigma_{n+s}^2) = \omega^2 \diamond r^{**}(\Delta s), \tag{39}$$

$$\mathrm{cor}(y_n^2, y_{n+s}^2) = \frac{\diamond r^{**}(\Delta s)}{6\, r^{**}(\Delta) + 2\Delta^2 (\xi/\omega)^2} \tag{40}$$

$$= q^{-1} \Delta^{-2} \diamond r^{**}(\Delta s), \tag{41}$$

where

$$\diamond r^{**}(s) = r^{**}(s + \Delta) - 2\, r^{**}(s) + r^{**}(s - \Delta) \tag{42}$$

and

$$q = 6\Delta^{-2}\, r^{**}(\Delta) + 2(\xi/\omega)^2. \tag{43}$$

4.2.1. Example 4

If $\sigma^2(t) \sim \mathrm{OU}$ with its variance existing then $r(u) = \exp(-\lambda|u|)$, which means that $r^{**}(s) = \lambda^{-2}\{\exp(-\lambda|s|) - 1 + \lambda s\}$ and

$$\diamond r^{**}(\Delta s) = \lambda^{-2}\{1 - \exp(-\lambda\Delta)\}^2 \exp\{-\lambda\Delta(s-1)\}, \qquad s > 0.$$

This implies

$$\begin{aligned} \mathrm{cor}(\sigma_n^2, \sigma_{n+s}^2) &= d \exp\{-\lambda\Delta(s-1)\}, \\ \mathrm{cor}(y_n^2, y_{n+s}^2) &= c \exp\{-\lambda\Delta(s-1)\}, \qquad s > 0, \end{aligned} \tag{44}$$

where

$$1 \geqslant d = \frac{\{1 - \exp(-\lambda\Delta)\}^2}{2\{\exp(-\lambda\Delta) - 1 + \lambda\Delta\}} \tag{45}$$

$$\geqslant c = \frac{\{1 - \exp(-\lambda\Delta)\}^2}{6\{\exp(-\lambda\Delta) - 1 + \lambda\Delta\} + 2(\lambda\Delta)^2(\xi/\omega)^2} \geqslant 0.$$

Equations (44) imply that σ_n^2 and y_n^2 follow constrained ARMA(1, 1) processes with common autoregressive parameters and with the moving average root being stronger for σ_n^2 than for the y_n^2. The ARMA structure implies that y_n is weak GARCH(1, 1) in the sense of Drost and

Nijman (1993) and as emphasized in the work of Meddahi and Renault (1996). Andersen and Bollerslev (1997b), page 137, have fitted GARCH(1, 1) models to (seasonally adjusted) equity and exchange rate returns measured at a variety of values of Δ and found that the above aggregation results broadly describe the fit of the various GARCH models. These simple analytic results generalize to the situation where we add together a weighted sum of uncorrelated OU processes, as was suggested in the previous section on superpositions and long memory models. Finally, as $\Delta \to 0$ so $d \to 1$ and so σ_n^2 behaves like a first-order autoregression.

More abstractly, Sørensen (1999) and Genon-Catalot *et al.* (2000) have independently noted that when $\mu = \beta = 0$ then the return sequence $\{y_n\}$ is α mixing if the instantaneous volatility $\sigma^2(t)$ is α mixing and further that the mixing coefficients for returns are less than or equal to the mixing coefficients for the instantaneous volatility process.

4.3. Leverage case

In the leverage case (8) the calculations are inevitably more specialized. When $\sigma^2(t) \sim \mathrm{OU}$ we can produce very concrete results. In particular

$$E(y_n y_{n+s}) = 0,$$

$$\mathrm{cov}(y_n, y_{n+s}^2) = E(y_n y_{n+s}^2) = \lambda^{-1}\rho\kappa_2\{1 - \exp(-\lambda\Delta)\}\exp\{-\lambda\Delta(s-1)\},$$

$$\mathrm{cov}(y_n^2, y_{n+s}^2) = \left(\frac{\kappa_2}{2\lambda^2} + \rho^2\mu_3\right)\{1 - \exp(-\lambda\Delta)\}^2 \exp\{-\lambda\Delta(s-1)\}.$$

The effect of the leverage term is to allow $\mathrm{cov}(y_n, y_{n+s}^2)$ to be negative if $\rho < 0$. However, in addition both $\mathrm{cov}(y_n, y_{n+s}^2)$ and $\mathrm{cov}(y_n^2, y_{n+s}^2)$ damp down exponentially with the lag length s. Exactly the same dynamic structure was found by Sentana (1995) in his work on the discrete time quadratic ARCH (QARCH) model. Hence we can think of the QARCH model as a kind of discrete time representation of our continuous time leverage model, generalizing the unleveraged result associated with the work of Drost and Nijman (1993) and Drost and Werker (1996).

5. Estimating and testing models

5.1. Olsen high frequency exchange rate data

In this paper we shall study 5-minute return series (recorded using Greenwich Mean Time) for the Deutsche Mark–dollar exchange rate from December 1st, 1986, to November 30th, 1996, constructed from the Olsen and Associates database using the semi-cleaning procedures carefully documented in Andersen *et al.* (2000). It is difficult to go below 5-minute returns without suffering from problems of discreteness which we shall briefly discuss in Section 6. Recent econometric papers on this topic include Russell and Engle (1998) and Rydberg and Shephard (1998). The series is defined using an average of bid and ask quotations. As a result they do not represent returns on transactions; however, the evidence of transaction data (which is not generally available in this quantity) of Goodhart *et al.* (1996) and Danielsson and Payne (1999) suggests that the properties of transaction and quote data, at this frequency, closely match.

The semi-cleaning procedure of Andersen *et al.* (2000) does not remove some heavy intraday effects in the volatility of the series; nor does it take into account the known timing of macroeconomic announcements which influence the volatility in the market. We shall not deal with the latter problem as adjusting for announcements is a challenging and important task in its own right and so we judge this to be beyond the scope of this paper (see Andersen and Bollerslev (1998b)). We have imposed some adjustments ourselves on the intraday effects. These included taking out all data from 10.30 p.m. on Friday until Sunday 11 p.m. each week, as well as bank-holidays. In addition we have estimated a strong intraday volatility effect (see Guillaume *et al.* (1997) for a discussion of this) by running a cubic spline (with 40 degrees of freedom) on the variance of each 5-minute period in active days. After some initial analysis we have set the intraday effect to be the same for Tuesdays, Wednesdays and Thursdays. Further, we have allowed the 5-minute return after the opening of the New York stock exchange to have its own free level as its variance is much higher than the rest of the data. The resulting smoothed estimate of the intraday seasonal component is given in Fig. 2. The most interesting features of this graph is the high volatility of the series on Monday mornings and Friday afternoons and the high level of volatility which generally occurs when the New York market is open.

After full adjustments are taken into account, we are left with a single unbroken time series made up of 684867 5-minute observations. For each observation we standardize it by dividing through by its intraday effect in an attempt to achieve a homogeneous series. We then study the marginal distribution of the resulting standardized series. Fig. 2 gives the log-histogram of returns where we split the returns into four sections of 125000 observations (i.e. each section is just over 2 years of adjusted 5-minute returns). To calibrate the graphs we have drawn the corresponding normal density. The graph indicates that returns are consistently much heavier tailed than is suggested by the normal distribution.

An interesting feature of the log-histograms is that the tails look almost linear, suggesting that we need models for marginal returns over short intervals of the form

$$\text{constant} \times |y|^{\rho_\pm} \exp(-\sigma_\pm |y|)$$

for some ρ_+, $\rho_- \in \mathbf{R}$ and σ_+, $\sigma_- \geqslant 0$. (Granger and Ding (1995) modelled $|y_n|$ as having a marginal distribution which is exponential.) One class of densities which has this property is the NIG densities.

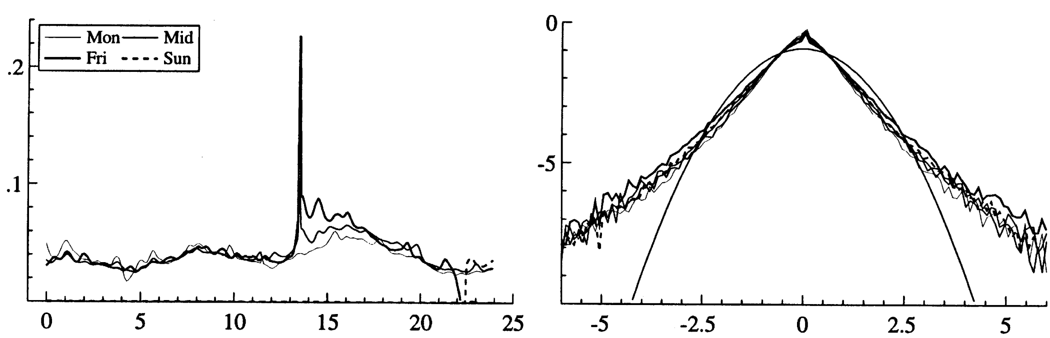

Fig. 2. (a) Estimated intraday pattern of volatility (standard deviations) for each day (in particular Monday, average over Tuesday–Thursday, Friday and Saturday) over 5-minute periods using 10 years of data (the x-axis denotes hours); (b) marginal log-density of returns over 5-minute periods (the data are split into series of length 125000; ··········, corresponding fitted normal log-density)

5.2. *Estimating marginal distribution*

Although the basic data set that we use takes Δ as representing 5 minutes, we can think about returns at other frequencies. In Fig. 3 we show the log-histograms of the fully adjusted returns for a variety of values of Δ. As expected from our discussion in Section 4.1 on aggregation, as Δ lengthens, the marginal log-densities seemingly become more accurately approximated by quadratics, i.e. normal densities. Fig. 3 also shows the fitted log-densities of NIG and Student t type, where the parameters of the fit are chosen by maximizing the corresponding likelihood assuming that the returns are independent and identically distributed (IID). We thus interpret these fits as of quasi-likelihood type.

Table 1 records the quasi-likelihood fits for each of the models, once again showing that the normal distribution is dominated by the other candidates. Here, for simplicity of exposition, we have only fitted symmetric distributions, as exchange rate returns (unlike equity returns) are known to be approximately symmetric. Further μ is taken to be 0, although in theory we should allow it to depend on the difference in interest rates between the two countries. However, in practice the drift is negligible in this case. Further for small values of Δ the NIG outperforms the Student t even though it is clear that the Student t has heavier tails. For larger values of Δ the fits are basically identical. The convergence towards normality as Δ increases is also shown in Table 1 where we compute the average Kullback–Leibler distance (per observation) between the normal density and the other two candidates that we study here.

5.3. *Estimating dependence structure*

We now turn our attention to the time dependence structure in high frequency fully adjusted returns. The correlogram of the series itself shows little activity, but the squares are another matter. We again decided to split our long series into the four shorter series of length 125 000 and have drawn in Fig. 4 the average correlogram which results. Note that the x-axis of the correlogram is marked out in days, not in 5-minute periods. Fig. 4(a) focuses on the short-term dynamics and shows a fast initial decay which then levels out. Fig. 4(b), which averages the correlograms within each day (the raw correlogram is very noisy), looks at longer-term dependence and shows a slow decay with memory lasting many days.

Fig. 4(c) is more unusual. Each day has 288 observations of 5-minute adjusted returns. We have computed the empirical variation within each day

$$s^2_{n,288} = \sum_{j=1}^{288} y^2_{288(n-1)+j}$$

which we know, from equation (35), should be a good estimator of the integrated volatility over a day

$$\{\sigma^{2*}(288n\Delta) - \sigma^{2*}[\{288(n-1)+1\}\Delta]\} = \sigma^2_{n,288}.$$

As a result we call $s^2_{n,288}$ the QV estimator. Having computed the daily $\{s^2_{n,288}\}$ series we have drawn in Fig. 4 the average (over our four series) correlogram (starting at lag 3 to be compatible with the above analysis). QV-type estimators of the integrated volatility process $\{\sigma^2_n\}$ have been used before us in Andersen *et al.* (2000). They studied the empirical correlograms and marginal distributions of the resulting statistics. However, they used unadjusted data. Our theoretical results suggest that the autocorrelation function of the $\{\sigma^2_{n,288}\}$ should be proportional to that for the averaged correlogram for the $\{y^2_n\}$ process given in Fig. 4(b). This seems to be very roughly confirmed here. However, we can see that the dependence among the empirical variance is much stronger than among just the noisy plain squared returns. This

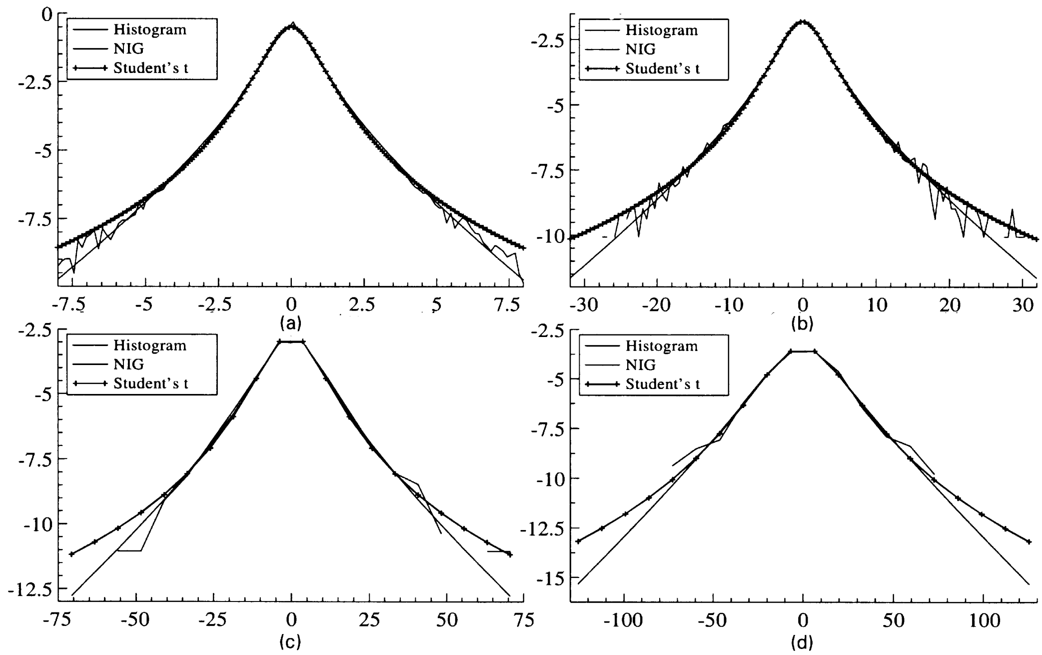

Fig. 3. Log-densities of returns at different levels of temporal aggregation (histograms and estimated (by quasi-maximum likelihood) NIG and Student's t-distributions): (a) 5-minute returns; (b) 70-minute returns; (c) 7-hour returns; (d) 27-hour returns (in (a) and (b) the histograms were computed using 128 bins; in (c) and (d) only 32 bins were used)

Table 1. Fit of the marginal distributions of returns y_n using zero-mean, symmetric distributions†

Model	*Measure of fit (distance from normal)*	*Results for the following values of* Δ*:*			
		1	*16*	*81*	*256*
Student t	Quasi-log-likelihood	−880240	−111090	−29215.	−10884.
	KL distance	34.22	2.048	0.2482	0.03944
	Degrees of freedom	2.954	2.926	3.366	5.154
NIG	Quasi-log-likelihood	−879800	−111060	−29198.	−10886.
	KL distance	34.38	2.059	0.2549	0.03889
	γ, δ	0.709, 0.679	0.193, 2.52	0.0971, 6.65	0.0799, 17.0
Normal	Quasi-log-likelihood	−971860	−116570	−29880.	−10990.

†We use the scaled Student t, NIG (parameters γ and δ) and the normal distribution with unknown variance. $\Delta = 1$ is chosen to represent 5 minutes. Reported are the maxima of the quasi-likelihood functions. The Kullback–Leibler distance KL is the average difference (per data point) between the log-likelihood function and the log-likelihood for the normal. We use it to measure the departure from normality of the returns.

is not a surprise; nor does it indicate that the QV estimator brings any additional statistical information beyond what is available from the autocorrelation function of the high frequency squared returns.

The empirical results suggest that we shall not be able to build satisfactory volatility models from the direct use of OU processes, for these have exponential decays in their autocorrelation functions. Fig. 4(a) has a heavy initial decay which then falls less steeply at longer lags. This immediately points us towards the use of the superposition of a number of OU processes for the continuous time volatility.

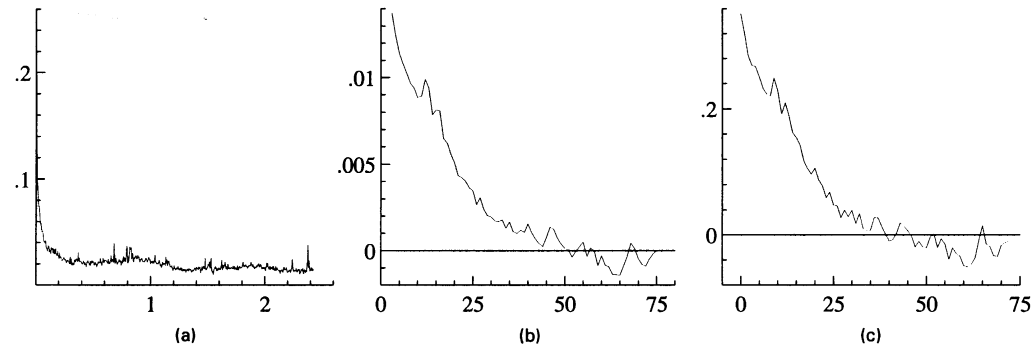

Fig. 4. Averages of five correlograms each with 125000 returns (labels for the lags of the correlograms are for days, not 5-minute periods): (a) first 750 lags, to show the short-term dynamics; (b) next 15000 lags to focus on the long-term pattern; (c) equivalent QV estimator based on squared 5-minute returns measured over a day

In this section we shall assume that the instantaneous volatility process $\{\sigma^2(t)\}$ is made up by the addition of m independent stationary processes $\{\sigma_j^2(t)\}$. For ease of exposition we shall assume

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

$$\sigma_j^2(t) \sim \mathrm{IG}(\delta w_j, \gamma),$$

where $\Sigma_{j=1}^{m} w_j = 1$ and $\{w_j \geqslant 0\}$. The inverse Gaussian assumptions will play no formal role in this analysis as it will be based only on the second-order properties of the model. Then $\sigma^2(t) \sim \mathrm{IG}(\delta, \gamma)$, and so $E\{\sigma^2(t)\} = \zeta = \delta/\gamma$ and $\mathrm{var}\{\sigma^2(t)\} = \omega^2 = \delta/\gamma^3$. The corresponding integrated volatility is

$$\begin{aligned} \sigma_n^2 &= \sum_{j=1}^{m} \sigma_{jn}^2, \\ \sigma_{jn}^2 &= \int_{(n-1)\Delta}^{n\Delta} \sigma_j^2(t)\,\mathrm{d}t. \end{aligned} \tag{46}$$

An implication is that $\mathrm{var}(y_n) = \Delta\xi$. Further, for $s > 0$,

$$\begin{aligned} \mathrm{cov}(y_n^2, y_{n+s}^2) &= \mathrm{cov}(\sigma_n^2, \sigma_{n+s}^2) \qquad (47) \\ &= \sum_{j=1}^{m} w_j \,\mathrm{cov}(\sigma_{jn}^2, \sigma_{jn+s}^2) \\ &= \omega^2 \sum_{j=1}^{m} w_j \diamond r_j^{**}(\Delta s) \\ &= \omega^2 \sum_{j=1}^{m} w_j \lambda_j^{-2} \{1 - \exp(-\lambda_j \Delta)\}^2 \exp\{-\lambda_j \Delta (s-1)\}. \end{aligned}$$

To estimate the parameters of the model we used a fitting procedure which employed a non-linear least squares comparison of the empirical autocovariance function $\{c_s\}$, based on the single time series of length 684000 observations, with the parameterized model given in equation (47). In particular the criterion that we minimized was

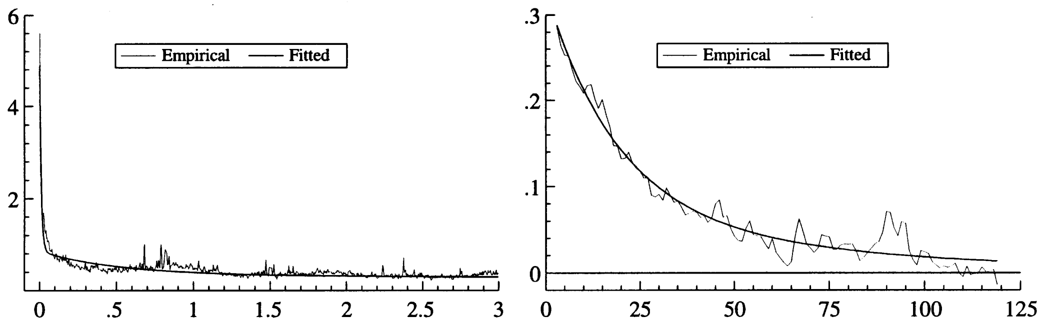

Fig. 5. Fitted and raw autocovariance functions for the single series of 684000 observations (the x-axis is marked in days, not 5-minute periods; the graphed fit uses a superposition of four independent OU processes): (a) short lags, average autocovariances in the day (rather than all the 5-minute correlations); (b) long lags

Table 2. Fit of the autocovariance function using a variety of superpositions of OU processes†

m	w_j				$exp(\lambda_j\Delta)$				ω^2	Ss
1	1.00				0.99988				0.303	430.7
2	0.212	0.788			0.99995	0.99982			0.335	346.1
3	0.017	0.064	0.919		0.99995	0.99982	0.9064		4.13	336.9
4	0.008	0.030	0.061	0.901	0.99995	0.99982	0.9931	0.711	8.75	334.8

†The fit is based on the single series of around 684000 observations. The number of processes is denoted by m. The weights are denoted by w_j, whereas the memory of the components is $\exp(\lambda_j\Delta)$. The variance of the volatility is written as ω and appears in $\text{cov}(y_n^2, y_{n+s}^2)$. Finally, Ss denotes the sum of squares.

$$\text{Ss} = \sum_{s=1}^{3\times 288} \{c_s - \text{cov}(y_n^2, y_{n+s}^2)\}^2 + 288 \sum_{s=3}^{123} \left\{ \frac{1}{288} \sum_{k=1}^{288} c_{288s+k} - \frac{1}{288} \sum_{k=1}^{288} \text{cov}(y_n^2, y_{n+288s+k}^2) \right\}^2.$$

The second term in this expression is slightly non-standard for we are working with the average autocovariances over each day of lags. The raw data are given in Fig. 5, together with the corresponding fit using $m = 4$. The broad picture is a fast initial decay, together with a small amount of correlation at longer lags.

Table 2 shows the fitted parameters for the analysis. It shows the effect of the changing value of m. For small values of m longer-term dependences are focused on, whereas for larger values of m the longer-term dynamics are clarified whereas the short-term dynamics are picked up. The most interesting feature of Table 2 is that a very large percentage of the volatility changing in the process is basically unpredictable. Hence we can think that this is merely a heavy-tailed component of the exchange rate movements. However, around 10% of the volatility movements are largely predictable. It is these effects which are more important when we measure returns at longer time horizons.

5.4. Traditional inference approaches

5.4.1. *Background*

In this subsection we shall discuss likelihood and various moment-based estimators of the parameters indexing the SV models. In addition we shall outline several approaches to estimating the current level of volatility in the series given a sequence of returns.

5.4.2. Likelihood

In principle we would like to use likelihood methods to estimate a fully parametric version of the model. To be concrete we shall work with the IG(δ, γ) OU process with no leverage. Then the likelihood function for $\theta = (\mu, \beta, \delta, \gamma, \lambda)$,

$$\begin{aligned} f(y; \theta) &= \int f(y_1, \ldots, y_T | \sigma_1^2, \ldots, \sigma_T^2; \mu, \beta)\, f(\sigma_1^2, \ldots, \sigma_T^2; \delta, \gamma, \lambda)\, \mathrm{d}\sigma_1^2, \ldots, \mathrm{d}\sigma_T^2 \\ &= \int \left\{ \prod_{n=1}^{T} f(y_n | \sigma_n^2; \mu, \beta) \right\} f(\sigma_1^2, \ldots, \sigma_T^2; \delta, \gamma, \lambda)\, \mathrm{d}\sigma_1^2, \ldots, \mathrm{d}\sigma_T^2 \end{aligned}$$

is, unfortunately, not directly computable (see, for example, Kim *et al.* (1998) and West and Harrison (1997)). We can simulate from $f(\sigma_1^2, \ldots, \sigma_T^2; \delta, \gamma, \lambda)$, by first recalling that

$$\begin{aligned} \sigma_n^2 &= \sigma^{2*}(n\Delta) - \sigma^{2*}\{(n-1)\Delta\} \qquad \text{where } \sigma^{2*}(t) = \lambda^{-1}\{z(\lambda t) - \sigma^2(t) + \sigma^2(0)\}, \qquad (48) \\ &= \lambda^{-1}\langle z(\lambda n\Delta) - \sigma^2(n\Delta) - [z\{\lambda(n-1)\Delta\} - \sigma^2\{(n-1)\Delta\}]\rangle \end{aligned}$$

while noting that

$$\begin{Bmatrix} \sigma^2(n\Delta) \\ z(\lambda n\Delta) \end{Bmatrix} = \begin{pmatrix} \exp(-\lambda\Delta)\, \sigma^2\{(n-1)\Delta\} \\ z\{\lambda(n-1)\Delta\} \end{pmatrix} + \eta_n, \qquad \eta_n = \begin{pmatrix} \exp(-\lambda\Delta) \int_0^{\Delta} \exp(\lambda t)\, \mathrm{d}z(\lambda t) \\ \int_0^{\Delta} \mathrm{d}z(\lambda t) \end{pmatrix}. \qquad (49)$$

Here the $\{\eta_n\}$ are IID and can be simulated by using equation (31) or by other methods.

5.4.2.1. Example 5. Suppose that the $\sigma^2(t)$ is an OU process with $\Gamma(\nu, \alpha)$ marginals. Then the result in expression (32) applies and we have

$$\eta_n = \alpha^{-1} \begin{Bmatrix} \exp(-\lambda\Delta) \sum_{i=1}^{N(1)} \log(c_i^{-1}) \exp(\lambda\Delta r_i) \\ \sum_{i=1}^{N(1)} \log(c_i^{-1}) \end{Bmatrix}, \qquad r_i \overset{\text{IID}}{\sim} U(0, 1),$$

and defining $c_1 < c_2 < \ldots$ as the arrival times of a Poisson process with intensity $\nu\lambda\Delta$ and $N(1)$ as the corresponding number of events up until time 1.

In general we do not know the explicit form of $f(\sigma_1^2, \ldots, \sigma_T^2; \delta, \gamma, \lambda)$, and so we cannot hope to solve for $f(y; \theta)$ analytically or to use an importance sampler to estimate the likelihood function. However, estimating the likelihood function without using an importance sampler is likely to be hopelessly inaccurate. Hence, with currently available techniques, direct likelihood methods are not feasible in our case.

Although the likelihood function is not directly available it may be possible that we could carry out Bayesian inference based on Markov chain Monte Carlo (MCMC) methods (Gilks *et al.*, 1996) to draw samples from $\theta | y$ if we place a prior on θ. This method has proved effective for log-normal SV models (see Jacquier *et al.* (1994) and Kim *et al.* (1998)) using the idea of data augmentation designing an MCMC algorithm for sampling from $\theta, \sigma^2 | y$, where $\sigma^2 = (\sigma_1^2, \ldots, \sigma_T^2)$. A generic scheme for carrying this out is as follows.

Step 1: initialize σ^2 and θ.
Step 2: update σ^2 from $\sigma^2|\theta, y$, by using a Metropolis–Hastings algorithm (one element at a time (e.g. Carlin *et al.* (1992)) or by using a blocking strategy (e.g. Shephard and Pitt (1997))).
Step 3: perform a Metropolis update on $\theta|y, \sigma^2$.
Step 4: go to step 2.

Cycling through steps 2 and 3 is a complete sweep of this sampler. The MCMC sampler will require us to perform many thousands of sweeps to generate samples from $\theta, \sigma^2|y$. Wong (1999) has shown that even in cases where it is possible to produce quite good samplers for drawing from step 2 of this procedure, in effect sampling from $\sigma^2|y, \theta$, the overall performance of the sampler is extraordinarily poor. This is because knowing $\sigma_1^2, \ldots, \sigma_T^2$ basically determines λ in a simple OU model—i.e. when we know the volatility we are overconditioning. The easiest way of thinking about this is to work with a discrete time version of this type of model where

$$\sigma_n^2 = \exp(-\lambda)\sigma_{n-1}^2 + \eta_n,$$

where $\eta_n > 0$ and is IID. Then

$$\exp(-\lambda) \leqslant \min_n(\sigma_n^2/\sigma_{n-1}^2).$$

This suggests that the likelihood function will have a mode very close to $\exp(-\lambda)$. Indeed it can be shown that the maximum likelihood estimator of λ is superconsistent for this type of problem (see Nielsen and Shephard (1999) and the references contained within). Hence the sampler is completely unable to move speedily through the sample space. This is not the case in a log-normal SV model (see Kim *et al.* (1998)). This very unfortunate effect seems inevitable for this type of parameterization.

The above problems can potentially be removed if we reparameterize the MCMC problem to work more directly in terms of the components of the shock terms $\{\eta_n\}$. Recall that they have an infinite series representation (31) which can be used to simulate from them. Each draw in these infinite series is based on the sequences, independent over n, $\{a_{in}\}$ and $\{r_{in}\}$. Here the r_{in} are independent copies of a uniform random variable r on [0, 1] and $a_{1n} < \ldots < a_{in} < \ldots$ are the arrival times of a Poisson process with intensity 1. Suppose that we truncate the sequence after K random variables for each value of n and write $a_{(n)} = (a_{1n}, \ldots, a_{Kn})'$ and $r_{(n)} = (r_{1n}, \ldots, r_{Kn})'$, and $a = (a_{(1)}, \ldots, a_{(T)})$ and $r = (r_{(1)}, \ldots, r_{(T)})$. Then we could perform MCMC-based inference based on sampling from

$$f\{\theta, a, r, \sigma^2(0)|y\} \propto f\{y|\theta, a, r, \sigma^2(0)\}\, f\{\sigma^2(0)|\delta, \gamma\}\, f(a, r).$$

This is straightforward for

$$f\{y|\theta, a, r, \sigma^2(0)\} = \prod_{n=1}^{T} f(y_n|\sigma_n^2),$$

as θ, a, r and $\sigma^2(0)$ determine $\{\sigma_n^2\}$. In principle this would only be an approximation (owing to the truncation of the infinite series representation), as it would be based on K variables; however, if K was chosen as a large number then it is likely to perform well.

So far we have not implemented this strategy as it is computationally burdensome.

5.4.3. Best linear predictors

To simplify the exposition suppose that $\beta = \rho = 0$ (which may be reasonable for exchange rate data). (The extension to the leverage case would write $y_n = \mu\Delta + \bar{z}_n + u_{1n}$ and $y_n^2 = \mu^2\Delta^2 + \sigma_n^2 + E(\bar{z}_n^2) + u_n$.) Then we note that $y_n|\sigma_n^2 \sim N(\mu\Delta, \sigma_n^2)$ and so

$$\begin{pmatrix} y_n \\ y_n^2 \end{pmatrix} = \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 + \sigma_n^2 \end{pmatrix} + u_n, \tag{50}$$

$$\mathrm{var}(u_{1n}) = E(\sigma_n^2) = \xi\Delta,$$

$$\mathrm{cov}(u_{1n}, u_{2n}) = 2\mu\Delta\, E(\sigma_n^2) = 2\mu\Delta^2\xi,$$

$$\begin{aligned} \mathrm{var}(u_{2n}) &= 4\mu^2\Delta^2\, E(\sigma_n^2) + 2\, E(\sigma_n^4) \\ &= 4\mu^2\Delta^3\xi + 2\{2\omega^2\, r^{**}(\Delta) + \xi^2\Delta^2\}, \end{aligned}$$

where u_n is a vector martingale difference sequence. Further (σ_n^2, z_n) is a linear process which is driven by the IID noise $\{\eta_n\}$. It is easy to see that

$$E(\eta_n) = \xi \begin{pmatrix} 1 - \exp(-\lambda\Delta) \\ \lambda\Delta \end{pmatrix},$$

$$\mathrm{var}(\eta_n) = 2\omega^2 \begin{Bmatrix} \frac{1}{2}\{1 - \exp(-2\lambda\Delta)\} & 1 - \exp(-\lambda\Delta) \\ 1 - \exp(-\lambda\Delta) & \lambda\Delta \end{Bmatrix}.$$

These results imply that a linear state space representation of the (y_n, y_n^2) (with uncorrelated $\{u_n\}$ and $\{\eta_n\}$) is

$$\begin{pmatrix} y_n \\ y_n^2 \end{pmatrix} = \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \lambda^{-1} & 0 \end{pmatrix} \alpha_n + u_n,$$

with

$$\begin{aligned} \alpha_{n+1} &= \begin{bmatrix} z\{\lambda(n+1)\Delta\} - z(\lambda n\Delta) + \sigma^2(n\Delta) - \sigma^2\{(n+1)\Delta\} \\ \sigma^2\{(n+1)\Delta\} \end{bmatrix} \\ &= \begin{Bmatrix} 0 & 1 - \exp(-\lambda\Delta) \\ 0 & \exp(-\lambda\Delta) \end{Bmatrix} \alpha_n + \begin{pmatrix} \eta_{2n} - \eta_{1n} \\ \eta_{1n} \end{pmatrix}, \end{aligned}$$

which allows us to use the Kalman filter (see, for example, Harvey (1989)) to provide a best linear (based on y_n and y_n^2) predictor of σ_n^2 and the associated mean-square error. (As σ_n^2 has an ARMA(1, 1) representation the minimal dimension of the state space form is 2. However, it is possible to remove $z\{\lambda(n+1)\Delta\} - z(\lambda n\Delta)$ from the transition equation and to have a single state variable. This would result in correlated measurement and transition noise.) Let us write these quantities as $s_{n|n-1}$ and $p_{n|n-1}$; then it is straightforward in the case that $\mu = 0$ to demonstrate that if $s_{1|0} \geqslant 0$ then $s_{n|n-1}$ is always non-negative and, in a steady state, takes the form of a GARCH(1, 1) recursion in the squares of the data. We should note that these estimates of volatility are really semiparametric, in the sense that they do not rely on any distributional assumptions about the volatility process, only on ξ, ω^2, μ and λ. For related ideas, in the context of discrete time log-normal SV models, see Harvey *et al.* (1994) and

Harvey and Shephard (1996) where a linear state space form is constructed for $\log(y_n^2)$. Estimates based on this representation are known to be inefficient (Jacquier *et al.*, 1994) principally because of the variance caused by inliers (small values of y_n^2). This particular problem does not necessarily carry over to our current treatment.

A simple way of estimating the parameters of this model is to use a (Gaussian) quasi-likelihood based around the output from the Kalman filter (e.g. Harvey (1989)). The asymptotic theory associated with the maximum quasi-likelihood estimator is worked out in Dunsmuir (1979). It will be asymptotically equivalent to an estimator defined via the Whittle likelihood.

The above arguments also generalize to where we sum m independent OU processes (46). Suppose that $E\{\sigma_j^2(t)\} = w_j\xi$ and $\mathrm{var}\{\sigma_j^2(t)\} = w_j\omega^2$. Then we have (σ_{jn}^2, z_{jn}) are independent over j and are again linear processes driven by noise $\{\eta_{jn}\}$. In this set-up

$$E(\eta_{jn}) = w_j\xi\begin{pmatrix} 1-\exp(-\lambda_j\Delta) \\ \lambda_j\Delta \end{pmatrix},$$

$$\mathrm{var}(\eta_{jn}) = 2w_j\omega^2\begin{Bmatrix} \frac{1}{2}\{1-\exp(-2\lambda_j\Delta)\} & 1-\exp(-\lambda_j\Delta) \\ 1-\exp(-\lambda_j\Delta) & \lambda_j\Delta \end{Bmatrix}.$$

The resulting representation has $2m$ state variables. Further, the only change in the measurement equation is that

$$\begin{aligned} E(\sigma_n^4) &= E(\sigma_n^2)^2 + \mathrm{var}(\sigma_n^2) \\ &= 2\omega^2\sum_{j=1}^{m} w_j\, r_j^{**}(\Delta) + \xi^2\Delta^2. \end{aligned}$$

5.4.4. *Particle filter*

The Kalman filter's estimate of σ_n^2 is the best linear estimator $s_{n|n-1}$ but it is not necessarily the efficient $E(\sigma_n^2|\mathcal{F}_{n-1})$, where $\mathcal{F}_{n-1}$ denotes the information that is available at time $(n-1)\Delta$. In this part of the paper we show that this quantity can be recursively computed by using a particle filter (see Pitt and Shephard (1999a) and Doucet *et al.* (2000) for a book length review of this material) and, further, we shall indicate that the linear and efficient estimators are close to one another.

A particle filter is a method for approximately recursively sampling from the filtering distribution $\sigma_n^2|\mathcal{F}_n$ for $n = 1, \ldots, T$. It has the following basic structure (Gordon *et al.*, 1993).

Step 1: assume a sample $\sigma^{2(1)}(n\Delta), \ldots, \sigma^{2(M)}(n\Delta)$ from $\sigma_n^2, \sigma^2(n\Delta)|\mathcal{F}_n$. Set $n = 0$.
Step 2: for each $\{\sigma^{2(m)}(n\Delta)\}$ generate K offspring

$$\{\sigma_{n+1}^{2(m,k)}, \sigma^{2(m,k)}\{(n+1)\Delta\}\}, \qquad k = 1, \ldots, K,$$

using equations (48) and (49). Compute

$$\log(w_{m,k}^*) = -\frac{1}{2}\log(\sigma_{n+1}^{2(m,k)}) - \frac{y_{n+1}^2}{2\sigma_{n+1}^{2(m,k)}}, \qquad k = 1, \ldots, K.$$

Step 3: calculate normalized weights $w_{m,k} \propto w_{m,k}^*$ which sum to 1 over m and k.

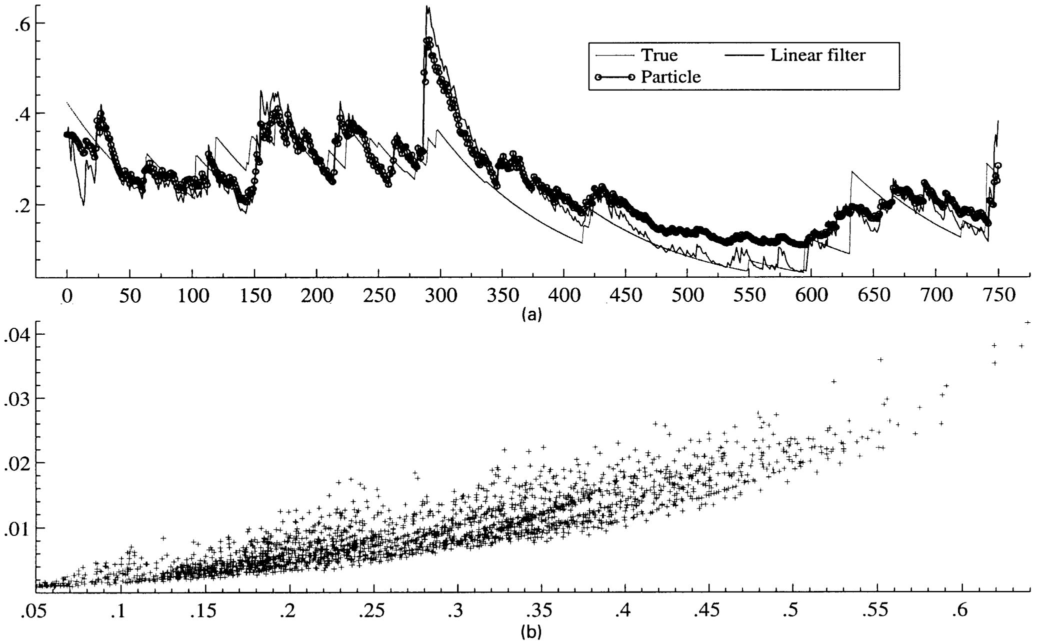

Fig. 6. OU process with $\Gamma(\nu, \alpha)$ marginals (throughout, $\Delta = 1$, $\nu = 3$, $\alpha = 8.5$ and $\lambda = 0.01$): (a) true σ_n^2, the best linear estimator and the particle filter's estimator of $E(\sigma_n^2|\mathcal{F}_{n-1})$ plotted against time; (b) $\mathrm{var}(\sigma_n^2|\mathcal{F}_{n-1})$ plotted against $E(\sigma_n^2|\mathcal{F}_{n-1})$, where both terms are estimated by using the particle filter (to do this we take $T = 2300$)

Step 4: resample, with unequal weights, among the $\{\sigma^{2(m,k)}\{(n+1)\Delta\}, w_{m,k}\}$ to produce a new sample $\sigma^{2(1)}\{(n+1)\Delta\}, \ldots, \sigma^{2(M)}\{(n+1)\Delta\}$. This sample is approximately from $\sigma_{n+1}^2|\mathcal{F}_{n+1}$.
Step 5: go to step 2.

As M grows large so the particle filter becomes more accurate, with the samples truly coming from the required filtering densities. In practice values of M of around 1000–10 000 are effective, whereas we typically take K as 3. Fig. 6 gives an example where we simulate from an OU process for $\{\sigma^2(t)\}$ and then use both the Kalman filter and a particle filter to estimate the unobserved integrated volatility $\{\sigma_n^2\}$ process. Fig. 6(a) shows that both procedures give rough estimates of the true integrated volatility with the major feature being that the two estimates are close together. Extensive work on this aspect suggests that the particle filter is only very marginally more efficient than the best linear estimator.

Fig. 6(b) graphs the particle filters estimate of $\mathrm{var}(\sigma_n^2|\mathcal{F}_n)$ against $E(\sigma_n^2|\mathcal{F}_n)$ and shows that the variance increases with the level of volatility, which is not surprising given the process that generates the integrated volatility but is not reflected in the corresponding calculations based on the Kalman filter.

5.4.5. Estimating equations

Earlier we derived general expressions for the second-order moments of the return sequence $\{y_n\}$. Recently Sørensen (1999) has studied how to use these moments to construct optimal estimating equations for OU-based SV models. These results, together with more general frameworks presented in Sørensen (1999) and Genon-Catalot *et al.* (1998), provide powerful methods for estimating these types of model. However, we are yet to study their effectiveness in practice.

5.4.6. Indirect inference

Equations (48) and (49) can be used to simulate a return sequence $\{y_n\}$ without any form of discretization error. However, it is now clear that this is insufficient for us to conduct straightforward likelihood-based inference, even when we are willing to use MCMC or particle-filter-based methods. This situation is not unfamiliar in econometrics where a new form of inference method, now generally called indirect inference, has been developed by Smith (1993) to deal with such situations (see Gourieroux *et al.* (1993) and Gallant and Tauchen (1996) for clear expositions). The basis of this approach is to use an incorrect 'auxiliary model', such as a GARCH(1, 1) model, as an approximation to the process and then to correct for the approximation by simulation.

To establish the notation write y as the data, θ as the parameters indexing the SV model, $\hat{y}^S(\theta)$ as a simulation of length S from the SV model based on the parameter θ and ψ to be the parameters of the GARCH(1, 1) model. Then indirect inference for θ follows the following approach.

Step 1: find the maximum likelihood estimator of ψ

$$\hat{\psi} = \arg_\psi \max[\log\{L_{\text{GARCH}}(\psi; y)\}]$$

as if the data had been produced by the GARCH model.
Step 2: find $\hat{\theta}$ such that

$$\hat{\psi} = \arg_\psi \max(\log[L_{\text{GARCH}}\{\psi; \hat{y}^S(\hat{\theta})\}]),$$

i.e. change the simulated data until its GARCH version of the maximum likelihood estimator is the same as that which results from the data.

We call $\hat{\theta}$ the indirect estimator of θ and typically base it on very large values of S (many times the sample size T). It is typically consistent and asymptotically normal (e.g. Gourieroux and Monfort (1996)). Of course it is also inefficient.

6. Further issues

6.1. Subordination

The modelling of financial processes by subordination of Brownian motion goes back to Clark (1973). Recent work on this topic includes the variance gamma model of Madan and Seneta (1990) (which is particularly notable as it uses a Lévy process as its subordinator) and that of Ghysels and Jasiak (1994), Conley *et al.* (1997) and Ané and Geman (2000). Subordination of Brownian motion is taken here in a general sense. It means a time transformation by a positive monotonically increasing stochastic process $\tau(t)$ that tends to ∞ for t tending to ∞ and is independent of the Brownian motion b. The resulting process is $b\{\tau(t)\}$.

Now consider models of the type

$$x^*(t) = \int_0^t \sigma(s)\,\mathrm{d}w(s), \tag{51}$$

where the processes σ and w are independent, w being a Brownian motion and σ being positive and predictable and such that $\sigma^{2*} \to \infty$ for $t \to \infty$. It turns out that, in essence, there is equivalence between the model formulation by equation (51) and the model formulation by subordination with an independent subordinator σ^{2*}.

To see this, note first that the process x^* is a continuous local martingale whose quadratic characteristic satisfies $[x^*](t) = \sigma^{2*}(t)$. As is well known, the Dubins–Schwarz theorem (see, for instance, Rogers and Williams (1996), page 64) tells us that, if we define processes γ and b by

$$\gamma(t) = \inf\{u\colon [x^*](u) > t\}$$

and

$$b(t) = x^*\{\gamma(t)\},$$

then b is a Brownian motion and

$$\{x^*(t)\}_{t \geqslant 0} = \{b\{[x^*](t)\}\}_{t \geqslant 0}. \tag{52}$$

To establish the equivalence it remains to prove that the processes b and σ^{2*} are independent. But this is equivalent to showing that

$$E[\exp\{i(f \cdot [x^*] + g \cdot b)\}] = E\{\exp(if \cdot [x^*])\}\, E\{\exp(ig \cdot b)\}. \tag{53}$$

But this is straightforward to show by using iterative expectations by first conditioning on σ.

6.2. Pricing

6.2.1. Non-arbitrage

In this subsection we shall show that our leveraged SV model does not allow arbitrage. In the case of no leverage, $\rho = 0$, non-arbitrage follows essentially from Lipster and Shiryayev (1977), chapter 6, and is well known. The arguments given below combine their technique with the Esscher transformation technique that is well known for Lévy process models. We study the process in parts

$$x^*(t) = x_0^*(t) + \beta\, \sigma^{2*}(t) + \rho\, \bar{z}(\lambda t) \tag{54}$$

where $\bar{z}(t) = z(t) - t\xi$ and

$$x_0^*(t) = \int_0^t \sigma(s)\, \mathrm{d}w(s)$$

with

$$\sigma^2(t) = \exp(-\lambda t) \int_{-\infty}^t \exp(\lambda s)\, \mathrm{d}z(\lambda s).$$

Once again we assume that w and z are independent, and we write $\{\mathcal{F}_t\}_{t \geqslant 0}$ to represent the filtration generated by the pair of processes (w, z). Further, in establishing non-arbitrage only finite time horizons will be considered, i.e. we restrict t to the interval $[0, T]$ for some, arbitrary, $T > 0$.

We must verify the existence of an equivalent martingale measure (EMM) under which the process $\exp\{x^*(t)\}$ is a local martingale. Let P be the original probability measure governing the behaviour of w and z over the time interval $[0, T]$, let $\phi = \beta + \frac{1}{2}$ and let θ' be the solution to

$$\kappa(\rho + \theta') - \kappa(\theta') = \xi\rho, \tag{55}$$

existence of the solution being assumed. Now, define the process $d(t)$ by $d(t) = \exp\{u^*(t)\}$ with

$$u^*(t) = -\phi\, x_0^*(t) - \tfrac{1}{2}\phi^2\, \sigma^{2*}(t) + \theta'\, \bar{z}(\lambda t) - \lambda t\, \bar{\kappa}(\theta') \tag{56}$$

and where $\bar{\kappa}(\theta) = \kappa(\theta) - \xi\theta$ is the cumulant function corresponding to the Lévy process $\bar{z}$, i.e. the cumulant function of $\bar{z}(1)$. Note that equation (55) may be re-expressed as

$$\bar{\kappa}(\rho + \theta') = \bar{\kappa}(\theta'). \tag{57}$$

Furthermore, let P' be the measure given by $\mathrm{d}P' = d(T)\,\mathrm{d}P$.

Proposition 1. Under the above set-up we have

(a) the process $d(t)$ is a mean 1 martingale, and hence P' is a probability measure, and
(b) the price process $\exp\{x^*(t)\}$ is a martingale under P'.

The proof of this result is given in Appendix A.

6.2.1.1. Example 6. Suppose that $z(1) \sim \mathrm{IG}(\delta, \gamma)$. Then

$$\begin{aligned}\kappa(\rho+\theta) - \kappa(\theta) &= \delta\gamma[(1 - 2\theta/\gamma^2)^{1/2} - \{1 - 2(\rho+\theta)/\gamma^2\}^{1/2}] \\ &= 2(\delta/\gamma)\rho[(1 - 2\theta/\gamma^2)^{1/2} + \{1 - 2(\rho+\theta)/\gamma^2\}^{1/2}]^{-1} \\ &= 2\xi\rho[(1 - 2\theta/\gamma^2)^{1/2} + \{1 - 2(\rho+\theta)/\gamma^2\}^{1/2}]^{-1}.\end{aligned}$$

Seeking a solution to equation (55) is therefore equivalent to solving

$$(1 - 2\theta/\gamma^2)^{1/2} + \{1 - 2(\rho+\theta)/\gamma^2\}^{1/2} = 2. \tag{58}$$

Suppose that $\rho \leqslant 0$, which is the econometrically relevant case. Then, as θ increases from $-\infty$ to its upper bound $\gamma^2/2$, the left-hand side of equation (58) decreases monotonically from ∞ to $|\rho|\sqrt{2}/\gamma$. Consequently, equation (58) is solvable if and only if $|\rho| \leqslant \gamma\sqrt{2}$ (which in practice is not a very binding constraint).

6.2.2. Derivatives

The fact that our SV model is arbitrage free means there is at least one EMM with which we can compute derivative prices. An important question is which one do we use? Recently Nicolato and Venardos (2000) and Nicolato (1999) have tackled this problem for our model when $\sigma^2(t) \sim \mathrm{IG}$ in the special case of $\rho = 0$. They have shown that a particularly convenient option price formula results if we choose to price the derivative with the EMM, written Q, which is closest to the physical measure, written P, in a relative entropy sense

$$\int \log(\mathrm{d}Q/\mathrm{d}P)\,\mathrm{d}Q.$$

This way of selecting from a set of EMMs was advocated in Föllmer and Schweizer (1991) using an elegant hedging argument. In particular if we write

$$C\{K, x^*(n\Delta), n\Delta + \Delta\}$$

for the price at time $n\Delta$ of a European call option on $x^*(t)$, with initial value $x^*(n\Delta)$, strike price K and expiration date $n\Delta + \Delta$ we have that

$$\begin{aligned}C\{K, x^*(n\Delta), n\Delta + \Delta\} &= E^Q\{x^*(n\Delta + \Delta) - K\}^+ \\ &= \int_{R_+} \mathrm{BS}\left\{K, x^*(n\Delta), \frac{1}{\Delta}\sigma_{n+1}^2, n\Delta + \Delta\right\} \mathrm{d}P\left\{\frac{1}{\Delta}\sigma_{n+1}^2 | \sigma^2(n\Delta)\right\}\end{aligned}$$

where $\mathrm{BS}\{K, x^*(n\Delta), (1/\Delta)\sigma^2_{n+1}, n\Delta + \Delta\}$ denotes the Black–Scholes price of the option with initial value $x^*(n\Delta)$, strike price K and constant volatility $(1/\Delta)\sigma^2_{n+1}$. This is particularly straightforward for the law of the volatility process is the same under the physical measure and the EMM. This result extends to more general cases as long as the volatility process is independent of the Brownian motion; in particular, it holds under superposition of OU processes.

In practice we can unbiasedly estimate $C\{\cdot\}$ simply by simulation for we can quickly draw many samples from $\sigma^2_{n+1}|\sigma^2(n\Delta)$ by using the series representations developed in Section 2 of this paper. Feasible alternatives to this approach include using either saddlepoint approximations or Fourier inversion methods based on the characteristic function, under Q, of

$$x^*(n\Delta + \Delta)|x^*(n\Delta), \sigma^2(n\Delta).$$

Here we shall derive the cumulant-generating function, whereas Scott (1997) and Carr and Madan (1998) have discussed the computations involved in moving to option prices from this type of function.

The required function is, for the canonical case of $n = 1$ and writing r to denote the riskless interest rate,

$$\begin{aligned} K\{\zeta \ddagger x^*(\Delta)|x^*(0), \sigma^2(0)\} &= \log(E^Q[\exp\{\zeta x^*(\Delta)\}|x^*(0), \sigma^2(0)]) \\ &= \{x^*(0) + r\Delta\}\zeta + K\{(\zeta\beta + \tfrac{1}{2}\zeta^2) \ddagger \sigma^2_\Delta|\sigma^2(0)\}. \end{aligned}$$

(This is a slight abuse of notation for we have previously assumed that $x^*(0) = 0$, which is not our intention here.) Hence the only unsolved problem is to compute the cumulant-generating function of $\sigma^2_\Delta|\sigma^2(0)$.

Recall that

$$\begin{aligned} \sigma^2_\Delta &= \lambda^{-1}\{z(\lambda\Delta) - \sigma^2(\Delta) + \sigma^2(0)\} \\ &= \int_0^\Delta \epsilon(\Delta - s; \lambda)\, \mathrm{d}z(\lambda s) + \epsilon(\Delta; \lambda)\, \sigma^2(0), \end{aligned}$$

where $\epsilon(t; \lambda) = \lambda^{-1}\{1 - \exp(-\lambda t)\}$. Consequently it is sufficient to work with

$$\begin{aligned} \bar{K}\{\theta \ddagger \sigma^2_\Delta|\sigma^2(0)\} &= \log[E\{\exp(-\theta\sigma^2_\Delta)|\sigma^2(0)\}] \\ &= -\theta\, \epsilon(\Delta; \lambda)\, \sigma^2(0) + \bar{K}\left\{\theta \ddagger \lambda^{-1} \int_0^{\lambda\Delta} \{1 - \exp(-\lambda\Delta + u)\}\, \mathrm{d}z(u)\right\} \\ &= -\theta\, \epsilon(\Delta; \lambda)\, \sigma^2(0) + \int_0^{\lambda\Delta} \bar{K}\{\theta\lambda^{-1}\{1 - \exp(-\lambda\Delta + u)\} \ddagger z(1)\}\, \mathrm{d}u \\ &= -\theta\, \epsilon(\Delta; \lambda)\, \sigma^2(0) + \lambda \int_0^\Delta \bar{K}\{\theta\, \epsilon(\Delta - s; \lambda) \ddagger z(1)\}\, \mathrm{d}s \\ &= -\theta\, \epsilon(\Delta; \lambda)\, \sigma^2(0) + \lambda \int_0^\Delta \bar{K}\{\theta\, \epsilon(s; \lambda) \ddagger z(1)\}\, \mathrm{d}s \\ &= -\theta\, \epsilon(\Delta; \lambda)\, \sigma^2(0) + \lambda \int_0^\Delta k\{\theta\, \epsilon(s; \lambda)\}\, \mathrm{d}s \\ &= -\theta\, \epsilon(\Delta; \lambda)\, \sigma^2(0) + \int_0^{1-\exp(-\lambda\Delta)} (1 - u)^{-1}\, k(\lambda^{-1}\theta u)\, \mathrm{d}u. \end{aligned}$$

6.2.2.1. Example 7. Suppose that $z(1) \sim \mathrm{IG}(\delta, \gamma)$, implying $k(\theta) = \delta\gamma - \delta\gamma(1 + 2\gamma^{-2}\theta)^{1/2}$. Then

$$\int_0^{1-\exp(-\lambda\Delta)} (1-u)^{-1}\, k(\lambda^{-1}\theta u)\, \mathrm{d}u = \delta\gamma \int_0^{1-\exp(-\lambda\Delta)} \frac{1-(1+\kappa u)^{1/2}}{1-u}\, \mathrm{d}u$$
$$= \delta\gamma\{\lambda\Delta - I(\kappa, \Delta)\},$$

where $\kappa = 2\gamma^{-2}\lambda^{-1}\theta$ and

$$I(\kappa, \Delta) = \int_0^{1-\exp(-\lambda\Delta)} \frac{(1+\kappa u)^{1/2}}{1-u}\, \mathrm{d}u$$
$$= \lambda\Delta\surd(1+\kappa) + 2\left[1 - b(\kappa) + \surd(1+\kappa)\log\left\{\frac{\surd(1+\kappa)+b(\kappa)}{\surd(1+\kappa)+1}\right\}\right].$$

Here $b(\kappa) = \surd\{1 + \kappa - \kappa\exp(-\lambda\Delta)\}$.

The result that we have the analytic cumulant-generating function, under Q, of $x^*(\Delta)|x^*(0)$, $\sigma^2(0)$ seems important for we can now regard the option pricing problem as being analytically solved for this class of models. In the financial economics literature the only equivalent result for SV models has been found by Heston (1993) and Duffie *et al.* (2000) (see also Stein and Stein (1991)) working with a square-root process

$$\mathrm{d}\sigma^2(t) = -\lambda\{\sigma^2(t) - \zeta\}\,\mathrm{d}t + \delta\,\sigma(t)\,\mathrm{d}b(t).$$

6.3. Trade-by-trade dynamics

Recently vast data sets recording the price, times and volumes of actual market transactions have become routinely available to researchers. It is interesting to try to link empirically plausible models of these trade-by-trade pricing dynamics with our SV models. To enable us to present general results we shall adopt the Rydberg and Shephard (2000) framework for tick-by-tick data. We model the number of trades $N(t)$ up to time t as a Cox process (which is sometimes called a doubly stochastic point process) with random intensity $\delta(t) = \delta\,\sigma^2(t) > 0$. In general we write τ_i as the time of the ith event and so $\tau_{N(t)}$ is the time of the last recorded event when we are standing at calendar time t.

Then a stylized version of the Rydberg–Shephard framework writes the current log-price as

$$x_\delta^*(t) = \mu\tau_{N(t)} + \beta\,\sigma^{2*}(\tau_{N(t)}) + \frac{1}{\surd\delta}\sum_{k=1}^{N(t)} y_k, \tag{59}$$

where for simplicity the $\{y_i\}$ are assumed independent standard normal and

$$\sigma^{2*}(t) = \int_0^t \sigma^2(u)\,\mathrm{d}u.$$

We assume that the Cox process and the $\{y_i\}$ are all completely independent. This model models prices as being discontinuous in time, jumping with the arrivals from the Cox process. Then we have the following result.

Theorem 3. For the price process (59), if the $\{y_i\}$ are assumed independent standard normal, $\sigma^{2*}(t) = \int_0^t \sigma^2(u)\,\mathrm{d}u$ and $N(t)$ is a Cox process with random intensity $\delta(t) = \delta\,\sigma^2(t) > 0$, then

$$\lim_{\delta\uparrow\infty}\{x_\lambda^*(\cdot)\} \xrightarrow{\mathcal{L}} x^*(\cdot),$$

where $x^*(t)$ is given in equation (6).

The proof is given in Appendix A.

This means that the tick-by-tick model will converge to an SV model as the amount of trading becomes large and the average tick size becomes small. We should note that the requirement that the $\{y_i\}$ are independent standard normal can be relaxed to allow general sequences of $\{y_i\}$ which exhibit a central limit theorem for the sample average. This is particularly useful for in practice the $\{y_i\}$ live on a discrete set and have quite complicated dependence structures which are not easy to model (see Rydberg and Shephard (1998, 2000)).

6.4. Vector Ornstein–Uhlenbeck processes

6.4.1. Construction of the process

So far our discussion has dealt with univariate processes. In this subsection we discuss extending this to the case of a vector of OU processes with dependence between the series. We introduce the q-dimensional volatility process

$$\sigma^2(t) = (\sigma_1^2(t), \ldots, \sigma_q^2(t))$$

via the BDLPs $z(t) = (z_1(t), \ldots, z_q(t))$ as follows. The multivariate form of equation (14) is

$$k(\theta) = \log(E[\exp\{-\langle\theta, z(1)\rangle\}]) = -\int_{R_+^q} \{1 - \exp(-\langle\theta, x\rangle)\}\, W(\mathrm{d}x), \tag{60}$$

where $\theta = (\theta_1, \ldots, \theta_q)$, $x = (x_1, \ldots, x_q)$, $R_+ = (0, \infty)$ and $\langle\theta, x\rangle = \Sigma_{i=1}^q \theta_i x_i$, and W is a Lévy measure on R_+^q, i.e. a measure satisfying

$$\int_{R_+^q} \min(1, \langle|x|\rangle)\, W(\mathrm{d}x) < \infty,$$

where $|x|$ is the Euclidean norm. Now let $z = (z_1, \ldots, z_q)$ be a q-dimensional Lévy process with $\log(E[\exp\{-\langle\theta, z(1)\rangle\}])$ as in equation (60). Suppose for simplicity that W has a density w with respect to Lebesgue measure, and let $w_i(x_i)$ be the ith marginal of w, i.e.

$$w_i(x_i) = \int_{R_+^{q-1}} w(x)\, \mathrm{d}x_1 \ldots \mathrm{d}x_{i-1}\, \mathrm{d}x_{i+1} \ldots \mathrm{d}x_q.$$

Imposing the condition

$$\int_1^\infty \log(x_i)\, w_i(x_i)\, \mathrm{d}x_i < \infty$$

we may then, on account of lemma 1, define the stationary process $\sigma_i^2(t)$ by

$$\sigma_i^2(t) = \int_{-\infty}^0 \exp(s)\, \mathrm{d}z_i(\lambda_i t + s).$$

Note that

$$\log(E[\exp\{-\theta_i\, z_i(1)\}]) = -\int_{0+}^\infty \{1 - \exp(-\theta_i x_i)\}\, w_i(x_i)\, \mathrm{d}x_i.$$

The full specification of $\sigma^2 = (\sigma_1^2, \ldots, \sigma_q^2)$ then rests on the choice of w, which we may aim to reflect the dependences among the volatility processes $\sigma_1^2(t), \ldots, \sigma_q^2(t)$.

This approach is at present under development. Here we just present a simple example.

6.4.1.1. Example 8. Let $q = 2$ and let w, defined in polar co-ordinates (r, a), be

$$\tilde{w}(r, a) = g(r; \delta, \gamma)\, b(a; \phi)$$

where $g(r; \delta, \gamma)$ is the Lévy density of the BDLP for the OU IG(δ, γ) process and

$$b(a; \phi) = B(\phi, \phi)^{-1}\left\{\frac{2}{\pi}a\left(1 - \frac{2}{\pi}a\right)\right\}^{\phi-1},$$

ϕ being a positive parameter. In the limit for $\phi \downarrow 0$ we obtain that $z_1(s)$ and $z_2(s)$ are independent BDLP–inverse Gaussian OU processes, whereas for $\phi \uparrow \infty$ the processes $z_1(s)$ and $z_2(s)$ tend to one and the same BDLP–inverse Gaussian OU process. Thus ϕ serves as a dependence parameter.

6.4.2. *Series representations*

Series representations of multivariate Lévy processes are available from the work of Rosiński (1990, 1999). Here we restrict discussion to presenting a result from the simplest type of setting. A fuller account is given in Barndorff-Nielsen and Shephard (2001).

Consider a q-dimensional BDLP process z with density $w(x)$ as in the Section 6.4.1 and let $\tilde{w}(r, a)$ $(a = (a_1, \ldots, a_{q-1}))$ be the representation of w in polar co-ordinates. We assume, for simplicity (and as in example 7), that $\tilde{w}$ factors as

$$\tilde{w}(r, a) = g(r)\, b(a)$$

where g is a one-dimensional Lévy density on R_+ and b is a probability density. Now let

$$G^{-1}(s) = \inf\{r > 0 \colon G^+(r) \leqslant s\},$$

where

$$G^+(r) = \int_r^\infty g(\rho)\, \mathrm{d}\rho.$$

Proposition 2. Let a_j, $j = 1, 2, \ldots$, be the arrival times of a Poisson process with rate 1 and let u_j, $j = 1, 2, \ldots$, be an IID sequence of unit vectors independent of $\{a_j\}$, such that the law of u_j is that determined by the probability density b. Furthermore, for $s \in [0, 1]$ let

$$\tilde{z}(s) = \sum_{j=1}^\infty \mathbf{1}_{[0,s]}(r_j)\, G^{-1}(a_j) u_j \tag{61}$$

where $\{r_j\}_{j\in\mathbf{N}}$ is an IID sequence of random variables uniformly distributed on [0, 1] and independent of the sequences $\{a_j\}_{j\in\mathbf{N}}$ and $\{u_j\}_{j\in\mathbf{N}}$. Then series (61) converges almost surely and

$$\{z(s) \colon 0 \leqslant t \leqslant 1\} \overset{\mathcal{L}}{=} \{\tilde{z}(s) \colon 0 \leqslant t \leqslant 1\}. \tag{62}$$

Furthermore we have the following proposition.

Proposition 3. If f_i, $i = 1, \ldots, d$, are positive and integrable functions on [0, 1] then

$$\int_0^1 f_i(s)\,\mathrm{d}z_i(s) \stackrel{\mathcal{L}}{=} \sum_{j=1}^{\infty} G^{-1}(a_j)\,u_{ij}\,f_i(r_j) \tag{63}$$

for $i = 1, \ldots, d$ and the u_{ij} IID with law determined by b.

6.5. *Multivariate stochastic volatility models*

6.5.1. *Model structure*

A simple q-dimensional version of the SV model for log-prices sets $x^*(t) = \{x_1^*(t), \ldots, x_q^*(t)\}$ with

$$\mathrm{d}x^*(t) = \{\mu + \beta\,\Sigma(t)\}\,\mathrm{d}t + \Sigma(t)^{1/2}\,\mathrm{d}w(t),$$

where $\Sigma(t)$ is a time-varying stochastic covariance matrix and β is a vector of risk premiums. Corresponding to this model structure is the integrated covariance

$$\Sigma^*(t) = \int_0^t \Sigma(u)\,\mathrm{d}u.$$

Then defining $y_n = x^*(n\Delta) - x^*\{(n-1)\Delta\}$ we have that

$$y_n|\Sigma_n^* \sim N(\mu\Delta + \beta\Sigma_n^*,\ \Sigma_n^*),$$

where $\Sigma_n^* = \Sigma^*(n\Delta) - \Sigma^*\{(n-1)\Delta\}$.

We can estimate $\Sigma^*(t)$ by using QV for $x^*(t)$ is a continuous q-dimensional local martingale plus a process which is continuous with bounded variation and so

$$[x^*](t) = \operatorname*{plim}_{r\to\infty} \left[\sum \{x^*(t_{i+1}^r) - x^*(t_i^r)\}\{x^*(t_{i+1}^r) - x^*(t_i^r)\}' \right] = \Sigma^*(t) \tag{64}$$

for any sequence of partitions $t_0^r = 0 < t_1^r < \ldots < t_{m_r}^r = t$ with $\sup_i(t_{i+1}^r - t_i^r) \to 0$ for $r \to \infty$.

6.5.2. *Factor models*

An important problem is to specify a model for $\Sigma^*(t)$. One approach is to do this indirectly via a factor structure

$$\Sigma(u) = \mathrm{diag}(\{\sigma_1^2(u), \ldots, \sigma_q^2(u)\}) + \sigma_{q+1}^2(u)\phi\phi'.$$

Here $\phi = (\beta_1, \ldots, \beta_q)$ are unknown parameters and the $\sigma_1, \sigma_2, \ldots, \sigma_{q+1}$ are mutually independent OU processes which are square integrable and stationary. It has common, but differently scaled, SV model and individual SV models for each series. It generalizes straightforwardly to allow for two or more factors. This style of model is in keeping with the latent factor models of Diebold and Nerlove (1989), King *et al.* (1994), Pitt and Shephard (1999b) and Chib *et al.* (1999). Its motivation is that in financial assets often returns move together, with a few common driving mechanisms. The common factors allow us to pick this up in a straightforward and parsimonious way. This model could be generalized by allowing the volatilities to be dependent by using the multivariate OU-type processes introduced in the previous subsection.

Finally, we should note that generating economically useful models via direct subordination arguments seems difficult even when we have vector OU processes. Let $b(t)$ be a vector of independent Brownian motions; then a multivariate, rotated, subordinated model would be $\beta\,b\{\sigma^2(t)\}$, for some matrix β and $\sigma^2(t)$ a vector of dependent OU processes.

However, such a model has a time invariant correlation matrix of returns, which is unsatisfactory from an economic viewpoint (for example asset allocation theory depends on correlations).

7. Conclusion

Non-Gaussian processes driven by Lévy processes are both mathematically tractable and have important applications. It is possible to build compelling SV models using OU processes to represent volatility. Log-returns from these types of model have many of the properties of familiar discrete time GARCH models. These SV models are empirically reasonable as well as having many appealing features from a theoretical finance perspective. In particular our class of models does not allow arbitrage and gives very simple expressions for standard option pricing problems under SV.

Although the treatment of OU processes that we have presented in this paper is extensive, there are several unresolved issues. A principal difficulty is that exact likelihood inference for SV models in continuous time but with discrete observations seems difficult. We hope that others may be able to solve this problem.

The generalization to the multivariate case is at its infant stage and much work must be carried out to make this a very flexible framework.

More generally, we believe that Lévy-driven processes have great potential for applications in fields other than finance and econometrics, e.g. in turbulence studies. They can also be further developed to a general toolbox for time series analysis. In this connection, we note that, although in the present paper we have concentrated on integrated processes x^*, one can also introduce very tractable stationary processes x driven by Lévy processes and having continuous sample paths, a simple and appealing possibility being the stationary solutions to stochastic differential equations of the form

$$\mathrm{d}x(t) = \{\mu + \beta\,\sigma^2(t) - \lambda\,x(t)\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}w(t) \tag{65}$$

with $\sigma^2(t)$ an OU process as in equation (2). See Barndorff-Nielsen and Shephard (2001) for a discussion of some of the work on this topic and its use in interest rate theory. Another alternative is to produce a positive stationary process by driving equation (65) not by Brownian motion but by another independent Lévy process with positive increments.

Acknowledgements

We are grateful to Peter Brockwell, Gary Chamberlain, Elisa Nicolato, Bent Nielsen, Jan Pedersen, Tina Rydberg, George Tauchen and Jack Wong for various helpful discussions, while the comments of Torben Andersen on a summary of a preliminary version of an earlier paper were very helpful. We owe a large debt to Michel M. Dacorogna at the Olsen Group for allowing us to use Olsen's high frequency exchange rate data in our study and Tim Bollerslev for supplying us with an interpolated version of these data.

Ole E. Barndorff-Nielsen's work is supported by the Centre for Analytical Finance (`http://www.caf.dk`), which is funded by the Danish Social Science Research Council, and by the Centre for Mathematical Physics and Stochastics (`http://www.maphysto.dk`), which is funded by the Danish National Research Foundation. Neil Shephard's research is supported by the UK's Economic and Social Research Council through the grant 'Econometrics of trade-by-trade price dynamics', which is coded R00023839.

Appendix A

A.1. Background

This appendix collects various proofs and results which are not given in the main text of the paper. It will be convenient to use the following notation for the cumulant function of an arbitrary random variable x

$$C(\zeta \ddagger x) = \log[E\{\exp(i\zeta x)\}],$$

while writing

$$\bar{K}\{\theta \ddagger x\} = \log[E\{\exp(-\theta x)\}],$$

in cases where x is positive. Similar notation applies to vector variates

A.2. Generalized inverse Gaussian Lévy density: proof of theorem 2

Let $z \sim \mathrm{GIG}(\nu, \delta, \gamma)$. From Halgreen (1979) we have that if $\nu \leqslant 0$ then

$$\bar{K}\{\theta \ddagger z\} = -\delta^2 \int_{\gamma^2/2}^{\infty} g_\nu\{2\delta^2(y - \gamma^2/2)\} \log(1 + \theta/y)\,\mathrm{d}y.$$

Differentiating both sides of this equation with respect to θ and transforming the integral by setting $\xi = y - \gamma^2/2$ we obtain

$$\begin{aligned}\frac{\partial \bar{K}\{\theta \ddagger z\}}{\partial \theta} &= -\delta^2 \int_0^\infty g_\nu(2\delta^2\xi)(\gamma^2/2 + \theta + \xi)^{-1}\,\mathrm{d}\xi \\ &= -\delta^2 \int_0^\infty g_\nu(2\delta^2\xi) \int_0^\infty \exp\{-(\gamma^2/2 + \theta + \xi)x\}\,\mathrm{d}x\,\mathrm{d}\xi \\ &= -\int_0^\infty \exp(-\theta x) x\, u(x)\,\mathrm{d}x\end{aligned}$$

and this shows that

$$u(x) = \delta^2 x^{-1} \int_0^\infty \exp(-x\xi)\, g_\nu(2\delta^2\xi)\,\mathrm{d}\xi\, \exp(-\gamma^2 x/2)$$

is the Lévy density of z. In this connection see also Pitman and Yor (1981), page 346, where a relationship with Bessel processes is established.

For $\nu > 0$ the expression for u follows from the convolution formula

$$\mathrm{GIG}(\nu, \delta, \gamma) = \mathrm{GIG}(-\nu, \delta, \gamma) * \Gamma(\nu, \gamma^2/2)$$

where $\Gamma(\nu, \phi)$ is the gamma distribution with probability density

$$\frac{\phi^\nu}{\Gamma(\nu)} x^{\nu-1} \exp(-\phi x)$$

and corresponding Lévy density $\nu x^{-1} \exp(-\phi x)$.

A.3. Non-arbitrage: proof of proposition 1

(a) For $0 \leqslant s \leqslant t \leqslant T$ we find that

$$\begin{aligned}E_P\{d(t)|\mathcal{F}_s\} &= E_P[E_P\{d(t)|z, \mathcal{F}_s\}\mathcal{F}_s] \\ &= \exp\{-\lambda t\, \acute{\kappa}(\theta')\}\, E_P(\exp\{\theta'\, \bar{z}(\lambda t) - \tfrac{1}{2}\phi^2\, \sigma^{2*}(t)\}\, E_P[\exp\{-\phi\, x_0^*(t)\}|\sigma, \mathcal{F}_s]|\mathcal{F}_s)\end{aligned}$$

and here

$$E_P[\exp\{-\phi\, x_0^*(t)\}|\sigma, \mathcal{F}_s] = \exp[-\phi\, x_0^*(s) + \tfrac{1}{2}\phi^2\{\sigma^{2*}(t) - \sigma^{2*}(s)\}]$$

so that

$$E_P\{d(t)|\mathcal{F}_s\} = d(s)\ \exp\{-\lambda(t-s)\,\bar{\kappa}(\theta')\}\,E_P(\exp[\theta'\{\bar{z}(\lambda t)-\bar{z}(\lambda s)\}]|\mathcal{F}_s) = d(s).$$

Thus $d(t)$ is a martingale and taking $s=0$ we have that $E_P\{d(t)\} = 1 = E_{P'}(1)$.

(b) Note first that

$$\beta - \tfrac{1}{2}\phi^2 + (1-\phi)^2 = 0. \tag{66}$$

By the martingale property of $d(t)$ we have, for arbitrary $\mathcal{F}_t$ measurable random variables y_t,

$$E_{P'}(y_t|\mathcal{F}_s) = E_P\{y_t\,d(T)/d(s)|\mathcal{F}_s\} = E_P\{y_t\,d(t)/d(s)|\mathcal{F}_s\}. \tag{67}$$

Hence

$$\begin{aligned} E_{P'}[\exp\{x^*(t)\}|\mathcal{F}_s] &= E_P[\exp\{x^*(t)\}\,d(t)/d(s)|\mathcal{F}_s] \\ &= \exp\{x^*(s) - \lambda(t-s)\,\bar{\kappa}(\theta')\}\,E_P(\exp[(\rho+\theta')\{\bar{z}(\lambda t)-\bar{z}(\lambda s)\}]J|\mathcal{F}_s) \end{aligned}$$

where

$$J = \exp[(\beta-\tfrac{1}{2}\phi^2)\{\sigma^{2*}(t)-\sigma^{2*}(s)\}]\,E_P(\exp[(1-\phi)\{x_0^*(t)-x_0^*(s)\}]|\sigma,\mathcal{F}_s).$$

However, by equation (66),

$$J = \exp[\{\beta - \tfrac{1}{2}\phi^2 + (1-\phi)^2\}\{\sigma^{2*}(t)-\sigma^{2*}(s)\}] = 1$$

so that, in view of condition (57),

$$\begin{aligned} E_{P'}[\exp\{x^*(t)\}|\mathcal{F}_s] &= \exp\{x^*(s) - \lambda(t-s)\,\bar{\kappa}(\theta')\}\,E_P(\exp[(\rho+\theta')\{\bar{z}(\lambda t)-\bar{z}(\lambda s)\}]|\mathcal{F}_s) \\ &= \exp[x^*(s) - \lambda(t-s)\{\bar{\kappa}(\rho+\theta')-\bar{\kappa}(\theta')\}] \\ &= \exp\{x^*(s)\}. \end{aligned}$$

A.4. Trade-by-trade dynamics

Lemma 2. Let $N(t)$ be a Cox process with random intensity $\delta(t) = \delta\,\sigma^2(t) > 0$. We write τ_i as the time of the ith event and so $\tau_{N(t)}$ is the time of the last recorded event when we are standing at calendar time t. Then for $\delta\to\infty$ we have that

$$\tau_{N(t)} \xrightarrow{p} t.$$

Proof. It suffices to show that for every $\epsilon > 0$ we have that

$$\Pr(\text{no event in } [t-\epsilon, t]) \to 0 \text{ as } \delta\to\infty.$$

Now, via conditioning on the intensity process we find, for every $\delta_1 > 0$,

$$\begin{aligned} \Pr(\text{no event in } [t-\epsilon,t]) &= E[\Pr\{\text{no event in } [t-\epsilon,t]|\delta(\cdot)\}] \\ &= E\left[\exp\left\{-\int_{t-\epsilon}^t \delta(s)\,\mathrm{d}s\right\}\right] \\ &= E\left[\exp\left\{-\delta\int_{t-\epsilon}^t \sigma^2(s)\,\mathrm{d}s\right\}\right] \\ &= E(\exp[-\delta\{\sigma^{2*}(t)-\sigma^{2*}(t-\epsilon)\}]) \\ &= E(\mathbf{1}_{\{\sigma^{2*}(t)-\sigma^{2*}(t-\epsilon)>\delta_1\}}\exp[-\delta\{\sigma^{2*}(t)-\sigma^{2*}(t-\epsilon)\}]) \\ &\quad + E(\mathbf{1}_{\{\sigma^{2*}(t)-\sigma^{2*}(t-\epsilon)\leqslant\delta_1\}}\exp[-\delta\{\sigma^{2*}(t)-\sigma^{2*}(t-\epsilon)\}]) \\ &\leqslant \Pr\{\sigma^{2*}(t)-\sigma^{2*}(t-\epsilon)\leqslant\delta_1\} + \exp(-\delta_1\delta). \end{aligned}$$

Consequently

$$\lim_{\lambda\uparrow\infty} \sup\{\Pr(\text{no event in } [t-\epsilon, t])\} \leqslant \Pr\{\sigma^{2*}(t) - \sigma^{2*}(t-\epsilon) \leqslant \delta_1\}$$

and since this holds for all $\delta_1 > 0$ the conclusion of lemma 2 follows.

A.4.1. Proof of theorem 3

It is helpful to rewrite the process as

$$x_\delta^*(t) = -\mu(t - \tau_{N(t)}) + \beta\{\sigma^{2*}(t) - \sigma^{2*}(\tau_{N(t)})\} + \beta\,\sigma^{2*}(t) + \mu t + \frac{1}{\surd\delta}\sum_{k=1}^{N(t)} y_k.$$

We obtain from lemma 2 and the continuity of $\sigma^{2*}(t)$ that the limiting behaviour in the distribution of $x_\delta^*(t)$, as $\delta \to \infty$, is the same as that of

$$\bar{x}_\delta^*(t) = \mu t + \beta\,\sigma^{2*}(t) + \frac{1}{\surd\delta}\sum_{k=1}^{N(t)} y_k.$$

Further, for the characteristic function of $\bar{x}_\delta^*(t)$ we find that

$$\begin{aligned} E[\exp\{i\xi\,\bar{x}_\delta^*(t)\}] &= \exp(i\xi t\mu)\,E\left[\exp\{i\xi\beta\,\sigma^{2*}(t)\}\,E\left\{\exp\left(i\xi\frac{1}{\surd\delta}\sum_{k=1}^{N(t)} y_k\right)|\delta(\cdot)\right\}\right] \\ &= \exp(i\xi t\mu)\,E\left\{\exp\{i\xi\beta\,\sigma^{2*}(t)\}\,E\left(\exp\left[i\xi\sqrt{\left\{\frac{N(t)}{\delta}\right\}}\,\bar{y}_{N(t)}\right]|\delta(\cdot)\right)\right\}, \end{aligned}$$

where $\bar{y}_{N(t)} = \surd(1/n)(y_1 + \ldots + y_n)$. Trivially, conditionally on $\delta(\cdot)$ we have that $N(t)/\delta \to \sigma^{2*}(t)$ almost surely as $\delta \to \infty$ and $\bar{y}_{N(t)} \sim N(0, 1)$ exactly. Thus

$$\begin{aligned} \lim_{\delta\uparrow\infty}(E[\exp\{i\xi\,x_\delta^*(t)\}]) &= \lim_{\delta\uparrow\infty}(E[\exp\{i\xi\,\bar{x}_\delta^*(t)\}]) \\ &= \lim_{\delta\uparrow\infty}\{\exp(i\xi t\mu)\,E(\exp[i\xi\{\beta\,\sigma^{2*}(t) + \sigma^*(t)u\}])\}, \end{aligned}$$

where $u \sim N(0, 1)$ and is independent of $\sigma^{2*}(t)$, i.e. the limiting distribution of $x_\delta^*(t)$ is the same as the law of $x^*(t)$. This argument is easily extended to convergence of all finite dimensional distributions of $x_\delta^*(t)$, i.e.

$$x_\delta^*(\cdot) \xrightarrow{\mathcal{L}} x^*(\cdot).$$

References

Andersen, T. G. and Bollerslev, T. (1997a) Heterogeneous information arrivals and return volatility dynamics: uncovering the long-run in high frequency returns. *J. Finan.*, **52**, 975–1005.

——— (1997b) Intraday periodicity and volatility persistence in financial markets. *J. Emp. Finan.*, **4**, 115–158.

——— (1998a) Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev.*, **39**, 885–905.

——— (1998b) Deutsche mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *J. Finan.*, **53**, 219–265.

Andersen, T. G., Bollerslev, T., Diebold, F. X. and Labys, P. (2000) The distribution of exchange rate volatility. *J. Am. Statist. Ass.*, **95**, in the press.

Ané, T. and Geman, H. (2000) Order flow, transaction clock and normality of asset returns. *J. Finan.*, **54**, in the press.

Barndorff-Nielsen, O. E. (1997) Normal inverse Gaussian distributions and stochastic volatility modelling. *Scand. J. Statist.*, **24**, 1–14.

——— (1998a) Processes of normal inverse Gaussian type. *Finan. Stochast.*, **2**, 41–68.

——— (1998b) Probability and statistics: selfdecomposability, finance and turbulence. In *Probability Towards 2000* (eds L. Acccardi and C. C. Heyde), pp. 47–57. New York: Springer.

——— (2000) Superposition of Ornstein-Uhlenbeck type processes. *Theor. Probab. Applic.*, to be published.

Barndorff-Nielsen, O. E., Jensen, J. L. and Sørensen, M. (1998) Some stationary processes in discrete and continuous time. *Adv. Appl. Probab.*, **30**, 989–1007.
Barndorff-Nielsen, O. E. and Shephard, N. (2000) Modelling by Lévy processes for financial econometrics. In *Lévy Processes — Theory and Applications* (eds O. E. Barndorff-Nielsen, T. Mikosch and S. Resnick). Boston: Birkhäuser. To be published.
——— (2001) *Lévy Based Dynamic Models for Financial Economics*. To be published.
Bertoin, J. (1996) *Lévy Processes*. Cambridge: Cambridge University Press.
——— (1999) Subordinators: examples and applications. *Lect. Notes Math.*, **1717**.
Black, F. (1976) Studies of stock price volatility changes. *Proc. Bus. Econ. Statist. Sect. Am. Statist. Ass.*, 177–181.
Bollerslev, T., Engle, R. F. and Nelson, D. B. (1994) ARCH models. In *The Handbook of Econometrics* (eds R. F. Engle and D. McFadden), vol. 4, pp. 2959–3038. Amsterdam: North-Holland.
Brockwell, P. J. (2001) Lévy driven CARMA processes. *Ann. Inst. Statist. Math.*, **52**, 1–18.
Brockwell, P. J., Resnick, S. I. and Tweedie, R. L. (1982) Storage processes with general release rule and additive inputs. *Adv. Appl. Probab.*, **14**, 392–433.
Campbell, J. Y., Lo, A. W. and MacKinlay, A. C. (1997) *The Econometrics of Financial Markets*. Princeton: Princeton University Press.
Carlin, B. P., Polson, N. G. and Stoffer, D. (1992) A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *J. Am. Statist. Ass.*, **87**, 493–500.
Carr, P. and Madan, D. B. (1998) Option valuation using the fast Fourier transform. Robert H Smith School of Business, University of Maryland, Baltimore.
Chib, S., Nardari, F. and Shephard, N. (1999) Analysis of high dimensional multivariate stochastic volatility models. Nuffield College, Oxford.
Cinlar, E. and Pinsky, M. (1972) On dams with additive inputs and a general release rule. *J. Appl. Probab.*, **9**, 422–429.
Clark, P. K. (1973) A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica*, **41**, 135–156.
Comte, F. and Renault, E. (1998) Long memory in continuous-time stochastic volatility models. *Math. Finan.*, **8**, 291–323.
Conley, T. G., Hansen, L. P., Luttmer, E. G. J. and Scheinkman, J. A. (1997) Short-term interest rates as subordinated diffusions. *Rev. Finan. Stud.*, **10**, 525–577.
Cont, R., Potters, M. and Bouchaud, J. P. (1997) Scaling in stock market data: stable laws and beyond. In *Scale Invariance and Beyond* (eds B. Dubrulle, F. Graner and D. Sornette), pp. 75–85. Les Houches.
Cox, D. R. (1991) Long-range dependence, non-linearity and time irreversibility. *J. Time Ser. Anal.*, **12**, 329–335.
Dacorogna, M. M., Muller, U. A., Olsen, R. B. and Pictet, O. V. (1998) Modelling short term volatility with GARCH and HARCH. In *Nonlinear Modelling of High Frequency Financial Time Series* (eds C. Dunis and B. Zhou). Chichester: Wiley.
Danielsson, J. and Payne, R. (1999) Real trading patterns and prices in spot foreign exchange markets. Financial Markets Group, London School of Economics and Political Science, London.
Diebold, F. X. (1988) *Empirical Modelling of Exchange Rate Dynamics*. New York: Springer.
Diebold, F. X., and Nerlove, M. (1989) The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *J. Appl. Econometr.*, **4**, 1–21.
Ding, Z. and Granger, C. W. J. (1996) Modeling volatility persistence of speculative returns: a new approach. *J. Econometr.*, **73**, 185–215.
Doornik, J. A. (1998) *Ox: Object Oriented Matrix Programming, 2.0*. London: Timberlake.
Doucet, A., de Freitas, N. and Gordon, N. (2000) *Sequential Monte Carlo Methods in Practice*. Berlin: Springer. To be published.
Drost, F. C. and Nijman, T. E. (1993) Temporal aggregation of GARCH processes. *Econometrica*, **61**, 909–927.
Drost, F. C. and Werker, B. (1996) Closing the GARCH gap: continuous time GARCH modelling. *J. Econometr.*, **74**, 31–57.
Duffie, D., Pan, J. and Singleton, K. (2000) Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, **68**, 1343–1376.
Dunsmuir, W. (1979) A central limit theorem for parameter estimation in stationary vector time series and its applications to models for a signal observed with noise. *Ann. Statist.*, **7**, 490–506.
Eberlein, E. (2000) Application of generalized hyperbolic Lévy motion to finance. In *Lévy Processes — Theory and Applications* (eds O. E. Barndorff-Nielsen, T. Mikosch and S. Resnick). Boston: Birkhäuser. To be published.
Eberlein, E. and Keller, U. (1995) Hyperbolic distributions in finance. *Bernoulli*, **1**, 281–299.
Engle, R. F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of the United Kingdom inflation. *Econometrica*, **50**, 987–1007.
Engle, R. F. and Lee, G. G. J. (1999) A permanent and transitory component model of stock return volatility. In *Cointegration, Causality, and Forecasting: a Festschrift in Honour of Clive W. J. Granger* (eds R. F. Engle and H. White), ch. 20, pp. 475–497. Oxford: Oxford University Press.
Ferguson, T. S. and Klass, M. J. (1972) A representation of independent increment processes without gaussian components. *Ann. Math. Statist.*, **43**, 1634–1643.

Föllmer, H. and Schweizer, M. (1991) Hedging of contingent claims under incomplete information. In *Applied Stochastic Analysis*, pp. 389–414. New York: Gordon and Breach.

Gallant, A. R. and Tauchen, G. (1996) Which moments to match. *Econometr. Theory*, **12**, 657–681.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998) Limit theorems for discretely observed stochastic volatility models. *Bernoulli*, **4**, 283–303.

——— (2000) Stochastic volatility as hidden Markov models and statistical applications. *Bernoulli*, **6**, in the press.

Ghysels, E., Harvey, A. C. and Renault, E. (1996) Stochastic volatility. In *Statistical Methods in Finance* (eds C. R. Rao and G. S. Maddala), pp. 119–191. Amsterdam: North-Holland.

Ghysels, E. and Jasiak, J. (1994) Stochastic volatility and time deformation: an application of trading volume and leverage effects. Department of Economics, University of Montreal, Montreal.

Gilks, W. R., Richardson, S. and Spiegelhalter, D. J. (eds) (1996) *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.

Goodhart, C., Ito, T. and Payne, R. (1996) One day in June 1993: a study of the working of the Reuters 2000-2 electronic foreign exchange trading system. In *The Microstructure of Foreign Exchange Markets* (eds J. A. Frankel, G. Galli and A. Giovannini), pp. 107–179. Chicago: Chicago University Press.

Gordon, N. J., Salmond, D. J. and Smith, A. F. M. (1993) A novel approach to nonlinear and non-Gaussian Bayesian state estimation. *IEE Proc.* F, **140**, 107–113.

Gourieroux, C. and Monfort, A. (1996) *Simulation Based Econometric Methods*. Oxford: Oxford University Press.

Gourieroux, C., Monfort, A. and Renault, E. (1993) Indirect inference. *J. Appl. Econometr.*, **8**, S85–S118.

Gradstheyn, I. and Ryzhik, I. W. (1965) *Table of Integrals, Series and Products*. New York: Academic Press.

Granger, C. W. J. (1980) Long memory relationships and the aggregation of dynamic models. *J. Econometr.*, **14**, 227–238.

Granger, C. W. J. and Ding, Z. (1995) Some properties of absolute returns, an alternative measure of risk. *Ann. Econ. Statist.*, **40**, 67–91.

Guillaume, D. M., Dacorogna, M. M., Dave, R. R., Muller, U. A., Olsen, R. B. and Pictet, O. V. (1997) From the bird's eye view to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets. *Finan. Stochast.*, **2**, 95–130.

Halgreen, C. (1979) Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Z. Wahrsch. Ver. Geb.*, **47**, 13–17.

Harrison, J. M. and Resnick, S. I. (1976) The stationary distribution and first exit probabilities of a storage process with general release rule. *Math. Oper. Res.*, **1**, 347–358.

Harvey, A. C. (1989) *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.

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

Harvey, A. C. and Shephard, N. (1996) The estimation of an asymmetric stochastic volatility model for asset returns. *J. Bus. Econ. Statist.*, **14**, 429–434.

Heston, S. L. (1993) A closed-form solution for options with stochastic volatility, with applications to bond and currency options. *Rev. Finan. Stud.*, **6**, 327–343.

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

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994) Bayesian analysis of stochastic volatility models (with discussion). *J. Bus. Econ. Statist.*, **12**, 371–417.

Jurek, Z. J. and Mason, J. D. (1993) *Operator-limit Distributions in Probability Theory*. New York: Wiley.

Jurek, Z. J. and Vervaat, W. (1983) An integral representation for selfdecomposable Banach space valued random variables. *Z. Wahrsch. Ver. Geb.*, **62**, 247–262.

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

King, M., Sentana, E. and Wadhwani, S. (1994) Volatility and links between national stock markets. *Econometrica*, **62**, 901–933.

Koponen, I. (1995) Analytic approach to the problem of convergence of truncated Lévy flights towards the gaussian stochastic process. *Phys. Rev.* E, **52**, 1197–1199.

Lipster, R. S. and Shiryayev, A. N. (1977) *Statistics of Random Processes*, vol. 1. New York: Springer.

Madan, D. B. and Seneta, E. (1990) The VG model for share market returns. *J. Bus.*, **63**, 511–524.

Mantegna, R. and Stanley, H. E. (2000) *Introduction to Econophysics: Correlation and Complexity in Finance*. Cambridge: Cambridge University Press.

Marcus, M. B. (1987) ξ-radial processes and random Fourier series. *Mem. Am. Math. Soc.*, **368**.

Meddahi, N. and Renault, E. (1996) Aggregation and marginalization of GARCH and stochastic volatility models. Centre de Recherche en Economie et Statistique, Paris.

Nelson, D. B. (1991) Conditional heteroskedasticity in asset pricing: a new approach. *Econometrica*, **59**, 347–370.

Nicolato, E. (1999) A class of stochastic volatility models for the term structure of interest rates. *PhD Thesis*. Department of Mathematical Sciences, Aarhus University, Aarhus.

Nicolato, E. and Venardos, E. (2000) Derivative pricing in Barndorff-Nielsen and Shephard's OU type stochastic volatility models. Department of Mathematical Sciences, Aarhus University, Aarhus.

Nielsen, B. and Shephard, N. (1999) Likelihood analysis of first order autoregressive models with exponential innovations. Nuffield College, Oxford.

Novikov, E. A. (1994) Infinitely divisible distributions in turbulence. *Phys. Rev.* E, **50**, R3303–R3305.

Pitman, J. and Yor, M. (1981) Bessel processes and infinite divisible laws. *Lect. Notes Math.*, **851**, 285–370.

Pitt, M. K. and Shephard, N. (1999a) Filtering via simulation: auxiliary particle filter. *J. Am. Statist. Ass.*, **94**, 590–599.

——— (1999b) Time varying covariances: a factor stochastic volatility approach (with discussion). In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 547–570. Oxford: Oxford University Press.

Prause, K. (1998) The generalised hyperbolic models: estimation, financial derivatives and risk measurement. *PhD Thesis*. Mathematics Faculty, Freiburg University, Freiburg.

Protter, P. (1992) *Stochastic Integration and Differential Equations*. New York: Springer.

Protter, P. and Talay, D. (1999) The Euler scheme for Lévy driven stochastic differential equations. *Ann. Probab.*, **25**, 393–423.

Raible, S. (1998) Lévy processes in finance: theory, numerics and empirical facts. *PhD Thesis*. Mathematics Faculty, Freiburg University, Freiburg.

Rogers, L. C. G. and Williams, D. (1994) *Diffusions, Markov Processes and Martingales*, vol. 1, *Foundations*. Chichester: Wiley.

——— (1996) *Diffusions, Markov Processes and Martingales*, vol. 2, *Ito Calculus*. Chichester: Wiley.

Rosiński, J. (1990) On series representations of infinitely divisible random vectors. *Ann. Probab.*, **18**, 405–430.

——— (1991) On a class of infinitely divisible processes represented as mixtures of Gaussian processes. In *Stable Processes and Related Topics* (eds S. Cambanis, G. Samorodnitsky and M. S. Taqqu), pp. 27–41. Basel: Birkhäuser.

——— (1999) Series expansion without compensation for infinitely divisible distributions. University of Tennessee, Knoxville.

——— (2000) Series representations of Lévy processes from the perspective of point processes. In *Lévy Processes—Theory and Applications* (eds O. E. Barndorff-Nielsen, T. Mikosch and S. Resnick). Boston: Birkhäuser.

Russell, J. R. and Engle, R. F. (1998) Econometric analysis of discrete-valued, irregularly-spaced financial transactions data using a new autoregressive conditional multinomial model. *2nd Int. Conf. High Frequency Data in Finance, Zurich, Apr.*

Rydberg, T. H. (1999) Generalized hyperbolic diffusions with applications towards finance. *Math. Finan.*, **9**, 183–201.

Rydberg, T. H. and Shephard, N. (1998) Dynamics of trade-by-trade price movements: decomposition and models. *Wrkshp Econometrics and Finance, Isaac Newton Institute, Cambridge, Oct.*

——— (2000) A modelling framework for the prices and times of trades made on the NYSE. In *Nonlinear and Nonstationary Signal Processing* (eds W. J. Fitzgerald, R. L. Smith, A. T. Walden and P. C. Young). Cambridge: Isaac Newton Institute–Cambridge University Press. To be published.

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

Scott, L. (1997) Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: applications of Fourier inversion methods. *Math. Finan.*, **7**, 413–426.

Sentana, E. (1995) Quadratic ARCH models: a potential re-interpretation of ARCH models. *Rev. Econ. Stud.*, **62**, 639–661.

Shephard, N. (1996) Statistical aspects of ARCH and stochastic volatility. In *Time Series Models in Econometrics, Finance and Other Fields* (eds D. R. Cox, D. V. Hinkley and O. E. Barndorff-Nielsen), pp. 1–67. London: Chapman and Hall.

Shephard, N. and Pitt, M. K. (1997) Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, **84**, 653–667.

Smith, A. A. (1993) Estimating nonlinear time series models using simulated vector autoregressions. *J. Appl. Econometr.*, **8**, S63–S84.

Sørensen, M. (1999) Prediction based estimating equations. Department of Theoretical Statistics, University of Copenhagen, Copenhagen.

Stein, E. M. and Stein, J. (1991) Stock price distributions with stochastic volatility: an analytic approach. *Rev. Finan. Stud.*, **4**, 727–752.

Taylor, S. J. (1982) Financial returns modelled by the product of two stochastic processes—a study of the daily sugar prices 1961–75. In *Time Series Analysis: Theory and Practice* (ed. O. D. Anderson), vol. 1, pp. 203–226. Amsterdam: North-Holland.

Vervaat, W. (1979) On a stochastic differential equation and a representation of nonnegative infinitely divisible random variables. *Adv. Appl. Probab.*, **11**, 750–783.

Walker, S. and Damien, P. (2000) Representations of Lévy processes without Gaussian components. *Biometrika*, **87**, 477–483.

West, M. and Harrison, J. (1997) *Bayesian Forecasting and Dynamic Models*, 2nd edn. New York: Springer.

Wolfe, S. J. (1982) On a continuous analogue of the stochastic difference equation $x_n = \rho x_{n-1} + b_n$. *Stoch. Processes Applic.*, **12**, 301–312.

Wong, C. K. (1999) New methods for performing efficient inference for linear and log-normal volatility models. *MPhil Thesis*. University of Oxford, Oxford.

Discussion on the paper by Barndorff-Nielsen and Shephard

S. D. Hodges (*University of Warwick, Coventry*)
We have come a long way since Kendall's (1953) work drew attention to the random walk nature of price series. The current literature on this topic is now very extensive: there is a plethora of models within at least three distinct categories, generalized autoregressive conditional heteroscedastic, stochastic volatility (SV) and implied processes (not previously referred to). Barndorff-Nielsen and Shephard's impressive paper proposes a class of models for representing the observed behaviour of security price processes. The paper belongs clearly within the SV family, to which it makes an important contribution. It brings us closer to our philosopher's stone, not of turning base metals into gold, but of combining realism with tractability.

Empirical regularities
We start with the conventional empirical regularities, which have been well documented by many researchers:

(a) returns are heavy tailed,
(b) exhibit volatility clustering and
(c) are skewed (in some cases).

Not too surprisingly, conditional return distributions become more Gaussian as the horizon is increased, but at a slower rate than the central limit theorem would predict if they were independent and identically distributed. Probably almost any model which combines these features will give a reasonable description of price processes.

Class of models
Barndorff-Nielsen and Shephard have given us not one model but a class of models, indeed a whole modelling approach. It is based on modelling the variance as an Ornstein–Uhlenbeck Lévy process with positive increments: $\sigma^2(t)$ moves up entirely by jumps and then tails off exponentially. This contrasts with the more conventional SV models which allow stochastic movement in both directions and secure non-negativity by scaling the shocks, or applying a transformation. It is also somewhat more realistic, though I have some reservations about the feature that there can be a strong conditional lower bound on the variance.

Using the generalized inverse Gaussian law for $\sigma^2(t)$ permits some simple forms. It would be nice to know more about how these forms differ and the details of implementing them. This approach seems likely to be more popular than the alternative of starting with the background driving Lévy process and then deriving the density of $\sigma^2(t)$ from it.

The availability of efficient means of simulation is particularly important in the context of many of the likely applications in the valuation of derivatives and risk management.

Estimations
The estimations described illustrate the techniques well and contain some interesting features: starting with the data; the use of quotes was probably an advantage; the data cleaning problems arising from bid–ask bounce in transactions' data are more complicated (see Roll (1984) and elsewhere). In less liquid markets there would be no point in working with such high frequency data. It comes as no great surprise that volatility is very unpredictable or that a superposition of Ornstein–Uhlenbeck processes was required to fit the autocorrelation function over the diverse time frames of this data set. In practice, few applications really need such time consistency over different scales. Some data sets seem to show negative correlation beyond about 50 days. A remaining puzzle is whether this is a real phenomenon and, if so, what to do about it.

Although fitting the autocorrelation function satisfactorily will also provide sensible quadratic variation (QV) estimates, since derivatives hedging is often directly exposed to the sample QV over various horizons, there remains a case for the direct use of QV measures (see Hodges and Tompkins (2000)).

It is important to have comparisons between competing models. One such comparison is provided by Tompkins and Hubalek (2000a). They use daily futures prices in various markets to compare four models: a strawman GBM, a conventional diffusion-based SV model (Heston, 1993), a normal inverse Gaussian Lévy return process with a diffusion SV and a Barndorff-Nielsen and Shephard model using a gamma distribution for $\sigma^2(t)$. They conclude that the last two models clearly dominate the others, and prefer the Barndorff-Nielsen–Shephard one for its convenience. It remains to be seen whether future work can discriminate between these.

Valuation
The paper describes one approach to the valuation of derivatives. This remains a continuing research issue. As the authors describe, in an incomplete market there will be many different martingale measures which support the prices of traded assets and give different valuations for untraded derivatives. Although the use of the minimal martingale measure has become quite popular within the mathematical finance community, this is more because of its elegance than its meaning, which remains doubtful. It is close to assuming that unhedgeable residual risks are not priced, and comparable with using least squares to estimate the location of a one-sided frontier.

Other issues
However much you give them, people want more. Further work on multivariate problems would be immediately useful. The authors' applications to term structure theory may also turn out to be significant, though here it will take rather stronger reasons to displace the incumbent models.

Implied process models
Much as I admire the models presented and look forward to working with them, I have reservations about the extent that the practitioner community is likely to embrace them. Derivative practitioners require simple models, and models that exactly calibrate to the prices of traded assets, including derivatives. When pricing exotic options, and hedging with plain options, it is nice to have a model that returns the correct prices of the plain options — even if it is misspecified! Dupire (1992) was the first to describe the principle of this implied process approach, which parallels the evolutionary term structure theory of Heath *et al.* (1992). Subsequent papers, e.g. Rubinstein (1994), Derman and Kani (1997), Britten-Jones and Neuberger (2000) and Skiadopoulos and Hodges (2000), have provided implementations for various deterministic volatility and (rudimentary) SV models. It would be of particular interest if the current approach could be stretched to encompass this idea.

Finally, we must be aware that our price models are simply models and do not contain fundamental truths in the same way as they might in physical situations. Long-term capital management and other market events have shown us that even the most apparently reliable statistical regularities can be brought to an end by a default, a change in an exchange rate regime or a monopolistic squeeze.

This paper makes an important contribution and deserves to be studied carefully. I have great pleasure in proposing the vote of thanks.

Gareth Roberts (*Lancaster University*)
I would like to congratulate the authors on a thought-provoking paper which I am sure will stimulate much further research into non-Gaussian Ornstein–Uhlenbeck (OU) models, and their use in financial modelling. These models seem to offer advantages in tractability and flexibility over many existing methods. I would like to concentrate on two aspects of the paper: the difficulty in performing exact likelihood-based inference and the construction of models of this type which have long-range dependence.

The authors mention two ways in which the non-Gaussian OU process can be specified — in terms of the distribution of the Lévy process increments in a unit time interval or in terms of the invariant distribution of the OU process itself. However, there is a third way which can be used to characterize a large class of such processes. Since the driving Lévy process includes no Brownian or negative increment components, we consider Lévy processes which make jumps according to some parameterized positive jump distribution, at times of a Poisson process. Thus the OU process can be described as

$$X_t = X_0 + \sum_{I=i}^{T_t} J_i \exp\{-\lambda(t-t_i)\}$$

where t_i denotes the time of the ith jump of the Poisson process and the J_i are an independent and identically distributed collection from the increment distribution.

This characterization suggests the parameterization of the OU process in terms of a marked Poisson process where the jump sizes represent the marks of the Poisson process driving the jumps. Using this parameterization, Markov chain Monte Carlo (MCMC) inference for the parameters of the jump distribution, λ and the latent marked Poisson process given the observed discretely observed stock price data (assuming for example that the stock exhibits no drift) is in principle straightforward. The method merely alternates between updating the points process, the jumps and the various parameters.

However, this method's performance varies considerably depending both on the length and the

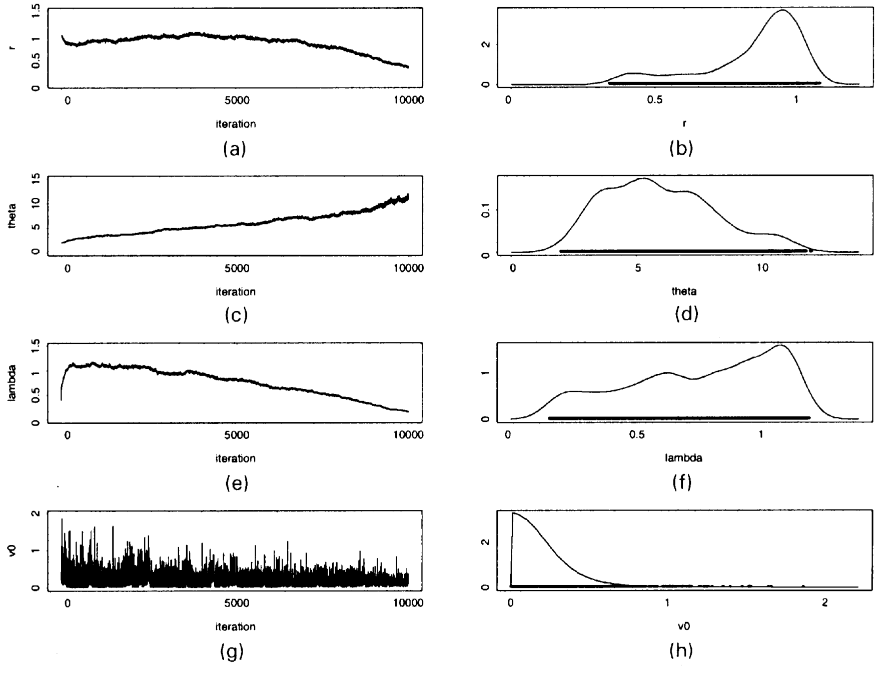

Fig. 7. Centred algorithm, 5000 daily data ($r = 0.1$, $\theta = 10$ and $\lambda = 0.05$): (a) trace of r (10 001 values per trace); (b) kernel density for r (10 001 values); (c) trace of θ (10 001 values per trace); (d) kernel density for θ (10 001 values); (e) trace of λ (10 001 values per trace); (f) kernel density for λ (10 001 values); (g) trace of v_0 (10 001 values per trace); (h) kernel density for v_0 (10 001 values)

density in time of the observed time series. Slow convergence is essentially caused by high serial correlation between the latent variables and the parameters. The problem is particularly acute in the case of sparsely observed time series, or any case in which the time series contains many data. The problems of imputation where the imputed data contain considerably more information than the actual observed data have been noted (see for example Roberts and Sahu (1997)) and are similar to related problems encountered by corresponding EM algorithms (see for example Meng and van Dyk (1997)).

With Omiros Papaspiliopoulos, we have performed a simulation study to investigate the performance of this algorithm in many situations. Fig. 7 demonstrates the behaviour of the MCMC algorithm for a simulated data set in which daily stock price data are simulated with $\lambda = 0.05$, $r = 0.1$ and $\theta = 0.05$. This data set is not even particularly sparse, with 10 data points per shock on average. However, the MCMC traces of the parameters demonstrate hopeless mixing.

A much more robust algorithm can be produced by a modified parameterization, whereby the Poisson process of rate λ on $[0, T]$ is written as a thinning of a unit rate Poisson process on $[0, T] \times [0, \infty)$ where a point x is deemed to be in the Poisson(λ) process if and only if (x, υ) is a point of the unit rate spatial Poisson process for some $0 \leqslant \upsilon \leqslant \lambda$. The new parameterization now records the spatial Poisson process and λ, two independent quantities *a priori* as unknowns. In all latent models of this kind, the data are inherently weak in estimating hyperparameters of the unobserved latent process, so the orthogonality of the prior is approximately maintained in the posterior. Fig. 8 shows output from this improved sampler (which we call *non-centred* since it shows some similarities to the non-centred parameterization used in Gaussian hierarchical models; see for example Roberts and Sahu (1997)).

In the case of the superposition of multiple processes, problems with MCMC mixing are exacerbated by identifiability issues, and it seems statistically dubious to attempt to fit large numbers of these processes, at least by any formal likelihood-based method. The main motivation for using superposition is to construct long-range dependence in the volatility process to conform with empirical observation. An alternative approach is to use heavy-tailed jump distributions in the model described above.

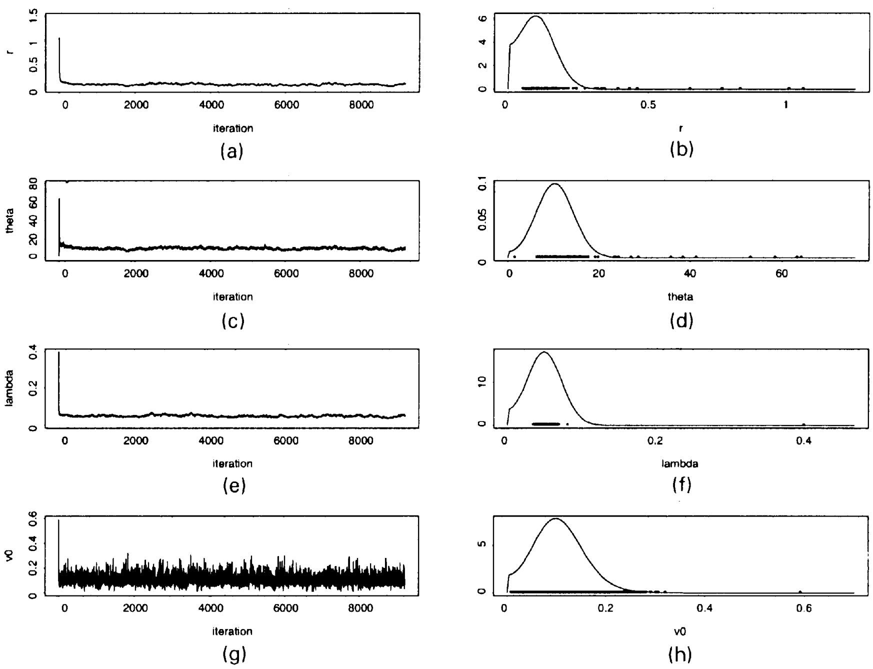

Fig. 8. Non-centred algorithm ($r = 0.1$, $\theta = 10$ and $\lambda = 0.05$): (a) trace of r (9216 values per trace); (b) kernel density for r (9216 values); (c) trace of θ (9216 values per trace); (d) kernel density for θ (9216 values); (e) trace of λ (9216 values per trace); (f) kernel density for λ (9216 values); (g) trace of v_0 (9216 values per trace); (h) kernel density for v_0 (9216 values)

For light-tailed jump distributions with density $h(\cdot)$, the exponential rate of convergence of the shot noise OU process is λ as the authors point out. However, if we set

$$h(x) \sim \frac{1}{x^{1+p}}$$

for large x, then for sufficiently small p the rate of convergence of the process is strictly less than λ. Furthermore, for jump distributions with density

$$h(x) \sim \frac{1}{x \log(x)^{1+p}}$$

for large x we can produce arbitrarily slow polynomial rates of convergence for sufficiently small $p > 0$. (To prove these results, it suffices to use Cheeger's inequality; see for example Lawler and Sokal (1988).) Of course, it is unclear to what extent processes with such heavy-tailed shocks model financial phenomena adequately.

It is a great pleasure to second the vote of thanks to the authors for their paper.

The vote of thanks was passed by acclamation.

Omiros Papaspiliopoulos (*Lancaster University*)
I would like to report on on-going work with Gareth Roberts and Petros Dellaportas inspired by the paper currently under discussion. Our aim is to make likelihood-based inference for the class of models introduced in the paper. In particular, we have looked at the case where the volatility is a shot noise process (jump times arriving as a Poisson process with rate r, jump sizes independent and identically distributed from an exponential distribution with parameter θ and an exponential decay rate λ) and the log-price process is a subordinated Brownian motion. We model the background driving Lévy process

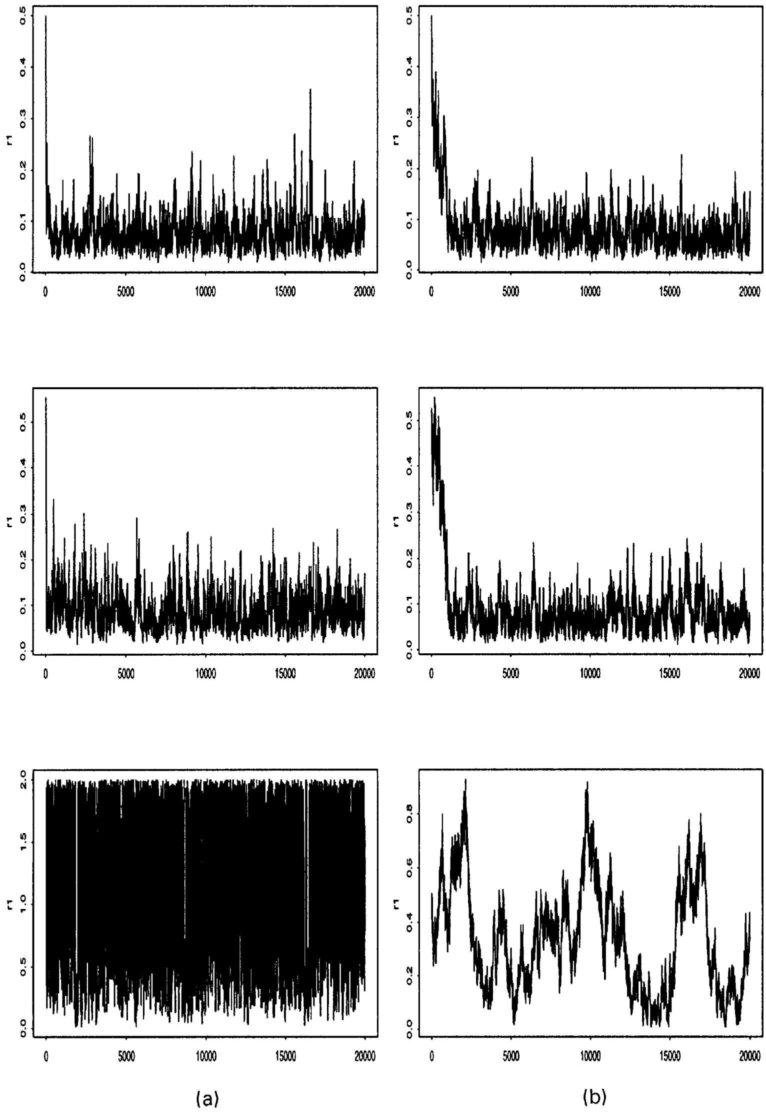

Fig. 9. Comparative behaviour of (a) the centred and (b) the non-centred algorithm under different sampling frequency: data were simulated from a driftless Brownian motion in an interval of 1000 days with shot noise process volatility with parameters $r = 0.1$, $\theta = 5$ and $\lambda = 0.05$; the columns show the MCMC traces (plotted every 100th iteration) for the parameter r for the two algorithms; in the first row the data are sampled daily, in the second every 2 days and in the third every 10 days

as a marked Poisson process and treat it as missing data. Our aim is to sample, using Markov chain Monte Carlo methods, from the joint distribution of the parameters and missing data.

For this we have constructed two Markov chain Monte Carlo algorithms which are feasible to implement and fast to execute. The algorithms are described in more detail in the vote of thanks by Gareth Roberts. The behaviour of the first algorithm, which we call *centred*, relies substantially on the structure of the data, in particular the length of the time series and the sampling frequency. The quantity of information about the parameters contained in the imputed missing data is substantially more than that in the observed data. This induces a high dependence between the parameters and the missing data. Our second algorithm, which we call *non-centred*, reparameterizes the missing data by introducing a marked Poisson process *a priori* independent of the parameter. This is then thinned and scaled using the parameters to yield the 'true' marked Poisson process. By doing so, it breaks down the dependence between the parameters and missing data. The non-centred algorithm is considerably more robust to the structure of the observed data than is the centred version. The computing time per iteration is comparable in the two algorithms but slightly greater in the non-centred than in the centred version. The behaviour of the two algorithms under different sampling frequencies is demonstrated in Fig. 9.

Both algorithms extend naturally to the case of superposition of Ornstein–Uhlenbeck processes. Lately we have implemented the centred algorithm for the case of the superposition of two Ornstein–Uhlenbeck processes. We are currently working on the implementation of the non-centred algorithm.

Enrique Sentana (*Centro de Estudios Monetarios y Financieros, Madrid*)
My comments are centred on several avenues for further research.

Other Ornstein–Uhlenbeck models
The independence assumption on the increments excludes models in which the volatility is either low or high, with the transition probabilities being a function of the current state (see Hamilton (1988) or Ryden *et al.* (1998) for discrete time versions). Given its binary nature, it is straightforward to write this continuous time Markov chain as an Ornstein–Uhlenbeck (OU) process with non-independent and identically distributed innovations. Such processes can be extended by superposition or by allowing for more states.

Diagnostics
The statistical foundations for model diagnostics based on histograms and kernels, higher order sample moments or sample correlograms have not been laid down yet. The issues are trickier for likelihood comparisons of marginal distributions. Finally, the data used are 'seasonally' adjusted before applying any other procedure.

Contemporaneous aggregation
Sterling–euro exchange rates are perfectly determined by sterling–dollar and euro–dollar rates. A similar situation arises with returns on individual stocks and portfolios (see Meddahi and Renault (1996) or Nijman and Sentana (1996)).

Identification
Although the results in Sentana and Fiorentini (2000) suggest that the identification of the underlying components may be easier in multivariate stochastic volatility models than in traditional factor analysis, those components may not be separately identified from a single series without further assumptions. This seems to be in contrast with the fact that we could keep adding unobserved OU components to the variance and retain identifiability.

Indirect inference
Given the dynamics implied by the model for discrete time observations, and the shape of the unconditional distribution for returns, it would be worth trying to estimate the OU stochastic volatility models by indirect inference on the basis of quadratic autoregressive conditional heteroscedastic auxiliary models with leptokurtic conditional distributions (see Fiorentini *et al.* (2000) and Calzolari *et al.* (2000)).

Mean returns
Volatility models for martingale differences are used in option valuation (but see Lo and Wang (1995) and León and Sentana (1997)) and short run value-at-risk assessments, but not in asset pricing or portfolio allocation. It may be possible to extend the work of Merton (1973), Cox *et al.* (1985) and Chamberlain (1988) on asset pricing models based on Brownian motion processes to the models considered here.

Finally, I congratulate the authors for their work. I am looking forward to reading Barndorff-Nielsen and Shephard (2001a).

N. H. Bingham (*Brunel University, Uxbridge*)
I have several comments to make on this interesting and important paper.

The underlying Lévy measure and simulation
The idea (in Section 2.4) of using the Lévy measure W of the underlying Lévy process directly,

particularly for simulation purposes, is a very good one and goes back at least as far as Bondesson (1982). Bondesson's approach is based on *shot noise* representations (or, equivalently, on Campbell's theorem), leading to series expansions, and associated truncation questions, akin to those of Section 2.5 around equation (31). What is needed is that W should be known explicitly.

Curse of dimensionality
The authors discuss (in Sections 6.4 and 6.5) the multidimensional case. In practice, the dimensionality q corresponds to the number of assets in a portfolio, and with unlimited investment opportunities available this may be large. Methods for which the computational complexity does not explode with q are thus at a premium — in other words, the curse of dimensionality limits us. One way to avoid it is via the theory of *elliptically contoured distributions* (Bingham and Kiesel, 2001; Hodgson *et al.*, 2001); this combines well with self-decomposability.

Quadratic variation
It is a far cry from knowing the link (35) to being able to exploit this link in practice. The crux is the continuous time nature of quadratic variation and the discrete time nature of sampling. In addition to the sources cited here, see also Genon-Catalot *et al.* (1999) and Kessler (2000).

Modelling volatility
The stochastic volatility models considered here are elegant, flexible and useful. Nevertheless, we need to step back and to think about the actual nature of volatility in the real world, and of change and uncertainty in it. A healthy scepticism towards any model of volatility is inescapably fostered by the study of 'financial forensics' (the phrase is due to S. A. Ross) — the examination of major financial disasters and lessons to be learned from them. A good example is the long-term capital management debacle of 1998, for a good account of which see Dunbar (1999). Here, every modelling assumption used by the highly sophisticated financial agents involved collapsed under the knock-on effects triggered by Russia's defaulting on its debts. Again, the markets can take fright at any reported remark of the chairman of the Federal Reserve, Mr Alan Greenspan (the FT-SE 100 index dropped 4% in a day after his 'irrational exuberance' comment of December 5th, 1996). I do not know how to model Mr Greenspan.

D. R. Cox (*Nuffield College, Oxford*)
It is a pleasure to congratulate both authors on an extremely impressive paper. A central theme is that of structured changes in variance. In most formulations, including the authors', these are imposed externally, not generated out of the dynamics of the process. Yet from a broad perspective a crucial aspect is: how is variance structure generated and, if it is indicative of some underlying instability, how can the system be modified? I appreciate that this is not the present authors' objective.

A toy representation of such an internal generation is obtained from a non-linear first-order autoregression with Gaussian innovations $\{\epsilon_t\}$, namely $Y_{t+1} = f(Y_t, \epsilon_{t+1})$ (Jones, 1978). If we expand by Taylor's theorem and drop the interesting term in Y_t^2 representing ultimate non-stationarity, then we may represent such a process in bilinear form as

$$Y_{t+1} = \alpha Y_t + (\beta/\sigma_\epsilon) Y_t \epsilon_{t+1} + \epsilon_{t+1},$$

where α and β are dimensionless parameters. Conditions for a stationary distribution with finite moments are known (Tong (1990), page 170). The bilinear form is mathematically but not conceptually equivalent to an assumed form for the conditional variance of Y_{t+1} given the current state.

The linear Markov form of $\mathrm{corr}(Y_{t+h}, Y_t)$ as α^h is retained and, subject to a technical condition of convergence which maybe can be evaded by truncation, $\mathrm{corr}(Y_{t+h}^2, Y_t^2)$ is a mixture of terms in α^h and $(\alpha^2 + \beta^2)^h$. If $\alpha = 0$ the autocorrelation of squares is β^{2h}. I am, of course, not advocating this as a model of any particular series, least of all, perhaps, financial series.

Elisa Nicolato (*University of Aarhus*) **and Emmanouil Venardos** (*Nuffield College, Oxford*)
Ornstein–Uhlenbeck (OU)-type models are appealing to derivatives pricing not only because they capture stylized features of observed time series but also because closed form solutions are available. This is mainly because the distribution of the integrated variance σ^{2*} is known. The following results are discussed extensively in Nicolato and Venardos (2000).

Following Barndorff-Nielsen and Shephard, we start with an OU-type process for the variance under the historical measure P. Then, there exists a family $\mathcal{M}'$ of equivalent martingale measures (EMMs) such that for arbitrary $Q \in \mathcal{M}'$ the associated risk neutral stochastic variance process is again of the OU type, i.e.

$$\mathrm{d}X_t = (r - \tfrac{1}{2}\sigma_t^2)\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t^Q + \rho\,\mathrm{d}Z_{\lambda t}^Q - \lambda\,\kappa(\rho)\,\mathrm{d}t,$$
$$\mathrm{d}\sigma_t^2 = -\lambda\sigma_t^2\,\mathrm{d}t + \mathrm{d}Z_{\lambda t}^Q$$

where $\kappa(\cdot)$ is the cumulant function of Z_1^Q.

This subclass $\mathcal{M}'$ is 'sufficiently large' in the sense that the range of option prices that can be spanned by varying $Q \in \mathcal{M}'$ is the same with the range spanned by the whole set of EMMs. This range of option prices is shown to be the open interval

$$\left(\mathrm{BS}\left\{S_t, \frac{\min(\sigma^{2*})}{T-t}\right\}, S_t\right) \tag{68}$$

where $\mathrm{BS}(y, z)$ is the Black–Scholes option price corresponding to the spot price y and volatility $\surd z$, and where

$$\min(\sigma^{2*}) = \frac{1}{\lambda}[1 - \exp\{-\lambda(T-t)\}]\sigma_t^2.$$

Two representations of the option price are available. Letting the Q stationary law of σ_t^2 be, for example, inverse Gaussian, the cumulant function of the log-price conditional on current information $\ln\{\phi(\theta)\} = \ln[\mathbb{E}_t^Q\{\exp(\mathrm{i}\theta X_T)\}]$ is known analytically and is

$$\ln\{\phi(\theta)\} = \mathrm{i}\theta\{X_t + r(T-t) - \lambda(T-t)\,\kappa(\rho)\} - \frac{\theta^2 + \mathrm{i}\theta}{2\lambda}[1 - \exp\{-\lambda(T-t)\}]\sigma_t^2 + \delta\{\surd f_1 - \surd(\gamma^2 - 2\mathrm{i}\theta\rho)\}$$
$$+ \frac{2\delta\{\mathrm{i}\theta\rho - (\theta^2 + \mathrm{i}\theta)/2\lambda\}}{\lambda\surd f_2}\left[\tan^{-1}\left\{\frac{\surd(\gamma^2 - 2\mathrm{i}\theta\rho)}{\surd f_2}\right\} - \tan^{-1}\left(\frac{\surd f_1}{\surd f_2}\right)\right]$$

where

$$f_1 = \gamma^2 - 2\mathrm{i}\theta\rho + \frac{\theta^2 + \mathrm{i}\theta}{\lambda}[1 - \exp\{-\lambda(T-t)\}]$$

and

$$f_2 = -\gamma^2 + 2\mathrm{i}\theta\rho - \frac{\theta^2 + \mathrm{i}\theta}{\lambda}.$$

The price of a European call $V(S_t, \sigma_t^2, t)$ with strike price K admits the representation

$$V(S_t, \sigma_t^2, t) = -\frac{K\exp\{-r(T-t)\}}{2\pi}\int_{-\infty-\mathrm{i}\omega}^{\infty-\mathrm{i}\omega}\mathrm{Re}\left[\frac{\exp\{-\ln(K)\mathrm{i}\theta\}\,\phi(\theta)}{\theta(\theta+\mathrm{i})}\right]\mathrm{d}\theta \tag{69}$$

for some $\varpi > 1$ in the domain of definition of the moment-generating function of X_T. The relevant inverse Laplace transform can be calculated numerically by using fast and reliable techniques.

Alternatively, the claim's price also admits the representation

$$V(S_t, \sigma_t^2, t) = \mathbb{E}_t^Q\{\mathrm{BS}(\tilde{S}, \tilde{V})\} \tag{70}$$

where

$$\tilde{V} = \frac{\sigma^{2*}}{T-t}$$

and

$$\tilde{S} = S_t\,\exp\{\rho(Z_{\lambda T}^Q - Z_{\lambda t}^Q) - \lambda(T-t)\,\kappa(\rho)\}.$$

The simulation of $\tilde{S}$ and $\tilde{V}$ is readily available from Rosiński expansions and the expectation in equation (70) can be consistently estimated by a sample average across simulations.

The natural question to ask is which $Q \in \mathcal{M}'$ should be used for pricing? Since some parameters are common under the historical and equivalent measure, it is only natural that the selection mechanism of the EMM should incorporate data from both the spot and the derivatives market.

The techniques and results generalize to the superposition of independent OU-type processes for which there is big empirical support.

Frank Critchley (*The Open University, Milton Keynes*)

It is a pleasure to welcome such an original and stimulating paper. My comments all relate to the choice of model or model checking. As n is so large, we might hope to have plenty of empirical information with which to estimate accurately, and then to choose between, alternative models that have been confirmed as well fitting. However, major complications arise due, in particular, to

(a) the lack of an explicit parametric likelihood,
(b) (possibly long-term) dependence in the data,
(c) macroeconomic announcements (Section 5.1) and
(d) the possibility of a change in model over time.

In these regards, we note the authors' comments in Section 5.4.3 on the detrimental effects of inliers, observe that one entry point for the literature on influence in time series is Bruce and Martin (1989) and wonder the following in the context of this paper.

(i) What overall modelling approach is more appropriate here? Is the loss of efficiency due to (a) entirely bad news? Efficiency and 'rigidity' (proneness to observations exerting unduly large influence) are opposite sides of the parametric modelling coin. Especially in view of (b) and (c), could (a) even be viewed positively in so far as it encourages adopting an overall modelling strategy which, while recognizing the strong appeal of the parametric approach, includes appropriate components of semiparametric (as in Section 5.4.3) (conceivably nonparametric) modelling as a guard against unduly influential cases?
(ii) With reference to superposition (Section 3) and its implementation in, especially, Table 2 of Section 5.3, is there a non-identifiability problem between the integer m and possibly zero weights w_j? Is inference about zero weights possible? How can inference on m be made?
(iii) How can appropriate auxiliary models be chosen in indirect inference (Section 5.4.6)? What guiding principles apply? Might a sequential empirical approach be helpful, in which several such models are used in turn with diagnostic feed-back at each stage?
(iv) What is the scope for residual and influence analysis? How tractable are these analyses here? In particular, can an effective form of inlier or outlier detection be developed?
(v) Could (systematic) subsampling, such as the four-way split introduced in Section 5.1, be developed as a general diagnostic tool? Alongside the authors' splits into non-overlapping time intervals, which seem well suited to (d), might other types of split be useful? In particular, could some form of cross-validation be helpfully developed?

Mark H. A. Davis (*Imperial College of Science, Technology and Medicine, London*)

Dealing with stochastic volatility is one of the most difficult and most important problems in financial risk management, and I congratulate the authors on a major contribution to this area.

The authors evaluate their models in terms of the goodness of fit to a long series of FX data. It is also relevant to consider the purposes for which the models will ultimately be used. There are three.

Marking to market

Large portfolios of underlying assets and derivatives must be revalued every day. Some of these will be exchange traded, so the value is just the current market value, whereas the valuation of non-exchange-traded derivatives requires a model. Since, however, model parameters are invariably 'calibrated' to reproduce the known prices of exchange-traded options, details of the model are relatively unimportant — any 'smooth function' will give essentially the same value.

Hedging

In addition to option values, traders need to know hedge parameters such as the famous Black–Scholes 'delta' to rebalance their books. In this context the ultimate test of a model is whether it leads to superior hedge performance. This is very difficult to test: in admittedly limited simulations using historical price data, I and colleagues at Tokyo–Mitsubishi International found it virtually impossible to distinguish between the various popular methods of volatility forecasting on the basis of hedge performance. This may be related to the fact, discussed in Davis (2000), that successful hedging is quite possible with the 'wrong' model.

Estimating value at risk
The value at risk measures the risk of a portfolio by estimating quantiles of the return distribution, and such calculations are a standard part of every bank's risk management process. Of course the essence of the value at risk is accurately capturing the lower tail of the return distribution, and this is where the methods introduced in the present paper could have an enormous practical effect.

Robert Tompkins (*University of Technology, Vienna*)
First, I congratulate the authors on their work. I know it quite well, and it has been an inspiration to our efforts in Vienna.

I have two comments. The first concerns the estimation of parameters for the models. The model itself is elegant and rich, and we have had long discussions about the ability of having the jump processes and the Lévy process combined with stochastic volatility. This allows analytic tractability and pricing in a very rich class of models. However, no matter how rich the class of models, if parameter values cannot be found for them, their applicability is limited. In Vienna (and also in Aarhus and Oxford), research is examining Bayesian Markov chain Monte Carlo methods along the lines suggested in the paper. To date, we have found it difficult to find solutions, but this research is ongoing.

My suggestion is to use another method for parameter estimation, which we have used successfully in Vienna for a variety of stochastic volatility models where inference from maximum likelihood is difficult (including special cases of models suggested in your paper). This approach is the simulated method-of-moments approach (or a moment matching approach) suggested by Duffie and Singleton (1993). Recent work by Andersson (1999) and Andersen *et al.* (1999) has shown that this approach compares well with traditional methods of maximum likelihood or quasi-maximum-likelihood methods for parameter estimation of stochastic volatility models. Regarding your models, Sylvia Frühwirth-Schnatter in Vienna is comparing our approach with the Bayesian Markov chain Monte Carlo approach that you suggest. Preliminary results suggest that both approaches will yield similar parameter estimates. As a starting-point, this might be helpful to use such a simulated method-of-moments approach and then perhaps you could use other methods to refine the parameter estimation problem from that point.

The second issue concerns measuring and estimating the options value, the p-measure and the q-measure. The difficulty in calibrating the model into actual option prices is that it could be very unstable in terms of fitting the parameters. An interesting issue, which was been pointed out in Madan *et al.* (1998), is the comparison between the models parameterized to the p- and the q-processes with comparisons of the differences.

I would encourage two lines of work to be done. One is from q to p, taking and fitting the option prices and working backwards, and the other looking from p to q to understand the nature of the risk neutral adjustment. Obviously, there is an assumption of a particular martingale measure, and the question is how sensitive that is to alternative martingale measure specifications.

The following contributions were received in writing after the meeting.

Fred Espen Benth (*University of Oslo*), **Kenneth Hvistendahl Karlsen** (*University of Bergen*) **and Kristin Reikvam** (*University of Oslo*)
We congratulate the authors on an impressive and inspiring paper.

In our discussion, we would like to draw attention to some applications of the suggested model in mathematical finance that were not mentioned by the authors. A major area in finance is portfolio optimization, which has the purpose of understanding investment behaviour in a stochastic market (see Merton (1971) and Hindy and Huang (1993)). Moreover, ideas and techniques from portfolio optimization theory may be used in pricing derivative contracts in incomplete markets (see Hodges and Neuberger (1989)), taking a completely different approach from the arbitrage theory used by the authors.

To reach realistic conclusions from the analysis of a portfolio optimization problem, we need realistic models for the financial assets constituting the portfolio. From the standpoint of stochastic analysis, the financial assets should be modelled by theoretically tractable processes as well. The standard model for the asset price dynamics is geometric Brownian motion. Many extensions to these dynamics have been suggested and analysed in the context of optimal portfolio theory, but none with striking empirical properties like the present model.

One of the basic techniques for studying portfolio optimization problems is the dynamic programming approach, translating the stochastic control problem to the analysis of a Hamilton–Jacobi–Bellman equation. This approach is only possible if the dynamics of the risky asset are described by a Markov process. The suggested asset price model is a two-dimensional continuous time (diffusion) process with a Markovian structure. Noteworthy, the subordinator driving the volatility dynamics leads to a second-order partial integrodifferential equation.

As the authors say, the logarithmic returns from the assets are approximately normal inverse Gaussian distributed. Simplified asset dynamics assuming independent logarithmic returns are given by the geometric normal inverse Gaussian Lévy process (see Barndorff-Nielsen (1998)). Benth *et al.* (1999) have studied a portfolio optimization problem where the risky asset is modelled by such dynamics.

It is an interesting question to extend the results in Benth *et al.* (1999) to risky assets following the dynamics suggested by the authors. We believe that their class of stochastic volatility models will lead to new and interesting results in these types of financial applications.

P. J. Brockwell and R. A. Davis (*Colorado State University, Fort Collins*)
We congratulate the authors on their innovative and illuminating models (6) and (8), in which the instantaneous volatility $\sigma^2(t)$ is taken to be a Lévy-driven Ornstein–Uhlenbeck process (i.e. a Lévy-driven first-order continuous time autoregressive or CAR(1) process). To extend the range of achievable autocorrelation functions, while preserving the non-negativity of the kernel f in the representation

$$\sigma^2(t) = \int_{-\infty}^{0} f(s)\,\mathrm{d}z(\lambda t + s), \tag{71}$$

the authors consider linear combinations, with positive weights, of independent Lévy-driven CAR(1) processes. To extend this range still further, we suggest the use of second-order Lévy-driven continuous time autoregressive moving average (CARMA(p, q) with $p > q$) processes with non-negative kernel. Provided that the zeros $\lambda_1, \ldots, \lambda_p$ of the polynomial $a(z) = z^p + a_1 z^{p-1} + \ldots + a_p$ all have negative real parts, the Lévy-driven CARMA(p, q) process with autoregressive polynomial $a(z)$ and moving average polynomial $b(z) = b_0 + b_1 z + \ldots + b_{p-1} z^{p-1}$ (with $b_j := 0$ for $j > q$) is defined (see Brockwell (2001)) as $x(t) = (b_0\ b_1 \ldots b_{p-1})\,\mathbf{x}(t)$, where $\mathbf{x}$ is the stationary solution of

$$\mathrm{d}\mathbf{x}(t) = A\mathbf{x}\,\mathrm{d}t + (0\ 0 \ldots 0\ 1)\,\mathrm{d}z(t) \tag{72}$$

and A is the matrix

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -a_p & -a_{p-1} & -a_{p-2} & \ldots & -a_1 \end{pmatrix}.$$

The process $\{x(t)\}$ has the representation, analogous to equation (71),

$$x(t) = \int_{-\infty}^{0} f(s)\,\mathrm{d}z(t + s), \tag{73}$$

where

$$f(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-\mathrm{i}s\lambda) \frac{b(\mathrm{i}\lambda)}{a(\mathrm{i}\lambda)}\,\mathrm{d}\lambda = \sum_{r=1}^{p} \frac{b(\lambda_r) \exp(-\lambda_r s)}{\prod_{j \neq r} (\lambda_r - \lambda_j)}, \tag{74}$$

and the second equality holds only if $\lambda_1, \ldots, \lambda_p$ are all distinct. From equation (74), necessary and sufficient conditions for the non-negativity of a CARMA(2, 1) kernel with λ_1 and λ_2 real are $b_1 \geqslant 0$ and $b_0 \geqslant b_1 \min_i |\lambda_i|$. Of these processes, those whose autocorrelation functions are those of superpositions of CAR(1) processes are those which satisfy the additional constraint, $b_0 \leqslant b_1 \max_i |\lambda_i|$. This simple example demonstrates the existence of a non-empty class of autocorrelations attained by CARMA processes with non-negative kernel which may usefully extend the modelling flexibility of the superpositions considered in the paper.

The authors' models allow considerable flexibility for modelling both the marginal distribution function and the covariance function of the volatility process, overcoming restrictions in frequently used (e.g. generalized autoregressive conditional heteroscedastic) models for financial time series. Their tractability for calculations is also impressive. As Barndorff-Nielsen (2000) shows, there is a limiting form whose autocorrelation function decays slowly like that of a long memory process. It may, however, be difficult to retain tractability and a parsimonious parameterization in using the limiting model.

Although modelling the dependence of the volatility process via the autocorrelation function is an excellent starting-point, it would be interesting to see how well the models proposed capture the dynamical dependence structures in the volatility process beyond those of second order.

Bent Jesper Christensen (*University of Aarhus*)
I would like to congratulate Professor Barndorff-Nielsen and Professor Shephard on this extremely stimulating and important work. Over the years, finance researchers have assembled a list of stylized features of empirical stock return (and related) series $\{y_n\}$, say. Thus, although the serial correlation in y_n is negligible, it is strong in y_n^2 and $|y_n|$ (volatility clustering), to the point of (quasi-)long-range dependence. The marginal distribution of y_n is non-normal, exhibiting skewness and, in particular, excess kurtosis, although y_n approaches normality as the interval Δ over which the return is measured is increased (aggregational Gaussianity). Finally, large negative returns $y_n < 0$ tend to be associated with more marked increases in return volatility than positive returns are (the leverage effect). The Ornstein–Uhlenbeck stochastic volatility (OU–SV) model allows all these stylized features to be incorporated yet retains analytical tractability.

Several points are worth highlighting. Firstly, dynamic properties and marginal distributions are handled separately. Secondly, from Fig. 3 the normal inverse Gaussian marginal return distribution provides a very precise picture of activity in the market-place. Minor discrepancies occur only in the tails, possibly because of extreme events that would in any case be outside the model, and certainly in regions where there is little information. Thirdly, given the repeated findings of generalized autoregressive conditional heteroscedastic GARCH(1, 1) behaviour in the literature, consistency with the OU–SV model (Section 4.2.1) is important. Finally, the only equally explicit SV option pricing formulae rely on Wiener-driven square-root SV processes. The wide possibilities for alternative marginal volatility laws (theorem 1) in the Lévy-driven OU–SV model should become important.

The remaining concerns relate to the fine structure of the model proposed. The leverage effect is captured by letting volatility innovations into the price process. However, the economic argument (increased risk of defaulting in a levered firm) would have causality run in the opposite direction, with drops in price leading to increases in volatility. This suggests the model

$$\mathrm{d}x^*(t) = \{\mu + \beta\,\sigma^2(t)\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}w(t), \tag{75}$$

$$\mathrm{d}[\log\{\sigma^2(t)\}] = -\lambda\,\log\{\sigma^2(t)\}\,\mathrm{d}t + \mathrm{d}z(\lambda t) + \rho\,\mathrm{d}w(t) \tag{76}$$

with $\rho < 0$ as an alternative to equations (1) and (8). The relationship between the alternative models would be of interest. Similarly, Figs 1(b) and 1(c) show that in the OU–SV model the volatility jumps up sharply, then declines gradually. Clearly, the correlogram, the non-linear least squares objective function and related objects are not sufficient. The empirical foundation of this asymmetry between the rises and falls in volatility is an exciting topic for future research. Finally, the multivariate extensions at the end of the paper are of key interest for practical portfolio management, but a related area of importance is the empirical analysis of options. Here, the relationship between implied volatilities from option prices and the integrated realized volatilities is of interest.

Petros Dellaportas (*Athens University of Economics and Business*) **and Emma J. McCoy and David A. Stephens** (*Imperial College of Science, Technology and Medicine, London*)
We would like to congratulate the authors for a very stimulating paper. It is of particular interest to people who operate in discrete time stochastics to investigate whether the ideas of the paper provide a framework for new time series models. We are currently investigating the possibility of deriving discrete time long memory volatility models that are more flexible and tractable than those currently available in the literature (see, for example, Baillie (1996)). Following the ideas of Section 3 and of Barndorff-Nielsen (2000), we have been working with a model in which long memory volatility is achieved by just considering volatility as an autoregressive AR(1) process with coefficient α and $-\log(\alpha)$ distributed as a gamma density. This appears to facilitate a Markov chain Monte Carlo implementation and provides an

alternative to the models produced by Ding and Granger (1996) and Bollerslev and Mikkelsen (1996).

We have two questions: first are the components in the superposed models developed in Section 3 genuinely identifiable from a single sample of data? We suspect not. Second, is the move to continuous time motivated by anything more than a requirement for coherent aggregation properties and, if so, are the benefits of this step outweighed by the estimation and computational difficulties that are encountered? Truncation of series representations of Lévy processes and subsequent approximations of stochastic integrals, in any case, essentially represent a discrete approximation to the continuous model. Is it possible to resolve this problem in a formal decision theoretic framework, where the effect on the decision to be made (option pricing, hedging etc.) is investigated for the model and its discrete approximation?

Francis X. Diebold (*University of Pennsylvania, Philadelphia*)

Barndorff-Nielsen and Shephard have produced a novel and important paper. In contributing to the discussion, I shall focus on the specification of the marginal distribution of volatility, temporal aggregation and the link.

On the marginal distribution of volatility

Separating the marginal distribution of $\sigma^2(t)$ from its dynamic structure is novel and welcome. It forces us to *think* about the marginal distribution of volatility, which is rarely done despite its importance for financial applications. Barndorff-Nielsen and Shephard advocate the use of a generalized inverse Gaussian (GIG) distribution for the marginal. However, recent empirical work finds that the log-normal distribution is a good approximation to the empirical variation (Andersen *et al.*, 2001a, b). Moreover, log-normality is often implicitly assumed in traditional theoretical developments, as for example when working with geometric Gaussian Ornstein–Uhlenbeck volatility processes, presumably in part because log-normality is viewed as realistic. What, then, is the relationship between the GIG and log-normal distributions? Of course all our models are abstractions and approximations, and the 'real world' is neither precisely GIG nor precisely log-normal, but I would at least like to know how good an approximation the GIG distribution could provide if the world were truly log-normal.

On temporal aggregation

Barndorff-Nielsen and Shephard note that, under conditions, returns generated from their models converge to normality under temporal aggregation, which is desirable because convergence to unconditional normality is observed in financial asset return data. In long memory stochastic volatility environments — obtained for example by their clever superposition argument — one of those conditions would have to be $H < 1$ (or $d < \frac{1}{2}$ in the notation of fractionally integrated autoregressive moving average modelling), because if $H > 1$ the variation would have infinite unconditional variance, which presumably would preclude the possibility of Gaussian central limit theorems for temporally aggregated returns. This would case doubt on the estimates of $H > 1$ that sometimes occur in empirical work: financial asset return volatility dynamics may involve long memory ($H > 0$), but $H < 1$ must hold if temporal aggregation to normality is to be respected. The argument parallels that of Diebold and Lopez (1995), who noted that the infinite unconditional variance implied by inverse generalized autoregressive conditional heteroscedastic conditional variance dynamics cannot be consistent with convergence to unconditional normality under temporal aggregation.

On the link

In a sense, log-normality fits *too* well, in that log-normality of the empirical variation appears approximately preserved under temporal aggregation (Andersen *et al.*, 2001b), despite the fact that a sum of log-normal distributions is not log-normal! This needs to be addressed in future research that invokes log-normal volatility. The obvious related question relevant to the present paper is whether GIG volatility is preserved under temporal aggregation, and whether the properties of temporally aggregated GIG volatility match those of the empirical variation of temporally aggregated returns data.

Sylvia Frühwirth-Schnatter (*University of Business Administration and Economics, Vienna*)

Firstly, the authors are to be congratulated on one of the most stimulating papers I have read in recent years. My remark concerns estimation based on the likelihood function and Bayesian methods. The problem is that the conditional distribution $f(y_n|\sigma_n^2, \mu, \beta)$ of the observed returns y_n depends on the unobservable integrated volatility σ_n^2. As is common for such incomplete-data problems, the authors introduce latent variables X,

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

and carry out (approximate) Bayesian inference for the augmented parameter vector. X results from discretizing the continuous time latent processes $z(t)$ and $\sigma^2(t)$ at the observation times $t = \Delta, \ldots, T\Delta$. Although this is a natural choice in the light of equations (3) and (7), the problem is that the distribution $f(X|\theta)$ has no simple analytical form.

I shall show that in the special case of a volatility model of Ornstein–Uhlenbeck type with gamma marginal law Bayesian inference via Markov chain Monte Carlo methods is possible under an alternative parameterization of the latent processes $z(t)$ and $\sigma^2(t)$. For this model the background driving Lévy process is a compound Poisson process with N jumps at times $\tau_1, \ldots, \tau_N$, where the interarrival times $\tau_j - \tau_{j-1}$ are independent and identically distributed (IID) exponential $\mathcal{E}(\nu)$, and the jump sizes $J_1, \ldots, J_N$ are IID exponential $\mathcal{E}(\alpha)$. For this process a natural timing is to take the jump times $t = \tau_1, \tau_2, \ldots, \tau_N$ rather than the observation times to describe $z(t)$ and $\sigma^2(t)$ leading to the definition of the following latent variables:

$$X = (N, \tau_1, \ldots, \tau_N, J_1, \ldots, J_N).$$

Figs 10 and 11 illustrate the differences between these two types of parameterization. Under the new parameterization the prior $f(X|\theta)$ has a very simple analytical form:

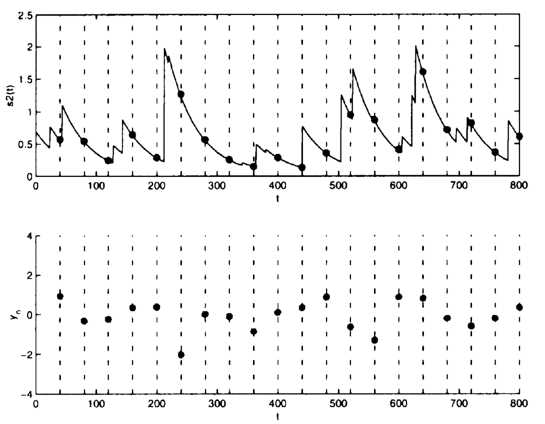

Fig. 10. Choice of X suggested by Barndorff-Nielsen and Shephard

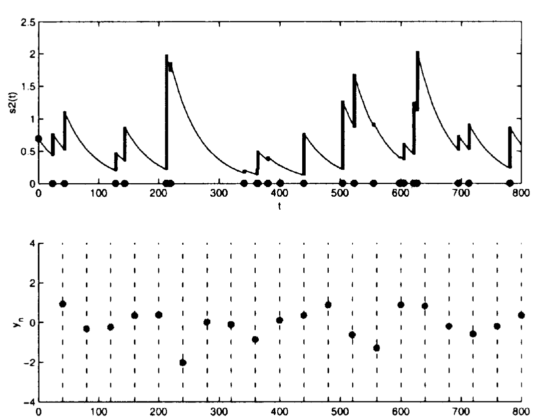

Fig. 11. Choice of X based on the jump times of the hidden Lévy process

$$f(X|\theta) = f(\tau_1, \ldots, \tau_N|N)\, f(J_1, \ldots, J_N|N, \alpha)\, f(N|\nu, \lambda).$$

Here the number of jumps N follows the Poisson distribution $\mathcal{P}(\lambda\nu T\Delta)$, and, conditional on knowing N, the jump sizes $J_1, J_2, \ldots, J_n$ are IID $\mathcal{E}(\alpha)$, whereas the distribution of the jump times $\tau_1, \ldots, \tau_N$ is the distribution of the order statistics of N random variables IID $\mathcal{U}[0, T\Delta]$:

$$f(\tau_1, \ldots, \tau_N|N) = \frac{N!}{(T\Delta)^N}.$$

We are working out the details of a Markov chain Monte Carlo sampler based on these results within the Viennese collaborative research project 'Adaptive information systems and modelling in economics and management science' (Frühwirth-Schnatter and Sögner, 2001). To sample X which is a quantity of variable dimensionality we are currently using the reversible jump Metropolis–Hastings algorithm of Green (1995), whereas an ordinary Metropolis–Hastings algorithm is used for joint sampling of the remaining model parameters.

Valentine Genon-Catalot (*Université de Marne-la-Vallée*) **and Catherine Larédo** (*Institut National de la Recherche Agronomique, Jouy-en-Josas, and Université Paris VI–VII*)
We congratulate the authors for this very stimulating paper. Ornstein–Uhlenbeck (OU)–Lévy processes are among the recent models proposed for modelling stochastic volatility. They have many interesting properties that are put forward and proved here. This motivates the reader to use them in both finance and other applications. For our part, we have studied models where the volatility is a positive diffusion process. We compare below these two classes of models.

State space
Both models can produce processes with positive values.

Stationarity
Conditions are well known for diffusions. They are proved for OU–Lévy processes (with a self-decomposability condition on marginal densities). This restricts the set of possible marginals. The generalized inverse Gaussian processes GIG(ν, δ, γ) discussed in Section 2.3 are also possible marginals for diffusions

$$\mathrm{d}X_t = \mathrm{d}\sigma_t^2 = b(X_t)\,\mathrm{d}t + a(X_t)\,\mathrm{d}B_t:$$

(a) extended Cox–Ingersoll–Ross (CIR) diffusion — $a(x) = x^{1/2}$ and $b(x) = \lambda(\beta - x + \eta/x)$ with $\lambda > 0$, $\eta > 0$;
(b) extended bilinear — $a(x) = x$ and $b(x) = \lambda(\beta - x - \eta x^2)$; $\lambda\eta > 0$, $\lambda\beta > 0$;
(c) bilinear diffusion — $a(x) = x$ and $b(x) = \lambda(\beta - x)$ with $\lambda\beta > 0$, $\lambda > -\frac{1}{2}$;
(d) CIR diffusion — $a(x) = x^{1/2}$ and $b(x) = \lambda(\beta - x)$ with $\beta > 0$, $\lambda\beta > \frac{1}{2}$.

Correlation structure
For OU–Lévy processes the only form is $\exp(-\lambda|u|)$, leading the authors to superimpose independent OU–Lévy processes, thus losing the Markov property. Mean reverting diffusions have the same correlation.

Transition densities
Both models are Markov models. However, in this respect they stand apart. For diffusions, transition densities are quite intractable whereas they are explicit and easy to simulate for OU–Lévy processes. This is a strong advantage.

Integrated volatility (σ_n^2)
The integrated volatility process plays a key role in the likelihood of the observations $(y_1, \ldots, y_T)$. As far as the exact distribution of $(\sigma_1^2, \ldots, \sigma_T^2)$ is concerned, there is no improvement with OU–Lévy processes, but an exact simulation is possible whereas this is impossible with diffusions (except by using Euler schemes). This seems to us very positive, although not used in the paper.

We have very much appreciated the probabilistic contribution, but less the statistical one. Only statistics based on marginal distributions and correlation structure are developed. Therefore, they cannot distinguish between the two classes. Studying the exact likelihood of the observations presents the same complexity in both cases. Clearly, this is a difficult problem that we ourselves have been confronted with (Genon-Catalot *et al.*, 1999, 2000).

Clive W. J. Granger (*University of California at San Diego, La Jolla*)
Financial time series currently face some choices that will eventually become vital elsewhere in economics: Gaussian or not; linear or not; continuous time or discrete; possibly equal interval; time units? The present paper balances between these divides and illustrates the difficulties encountered in moving across the divisions and is to be particularly welcomed for introducing a new and flexible class of distributions to describe the empirical distributions of returns.

Statisticians use the Gaussian distribution as their base-line. Although the present paper emphasizes the theoretical and empirical importance of non-Gaussian distributions, it continues to use normal-based concepts. For example, it considers a 'positive measure of volatility' above equation (2) but uses the usual notation for the variance. Of course the variance is a widely used measure of dispersion, being especially useful for the Gaussian distribution, but other appropriate measures exist for other distributions, such as the expected absolute deviation for the double-exponential distribution, as mentioned later in the paper. Similarly the correlation is not necessarily the best measure of dependence for non-Gaussian distributions and linear forecasts are certainly not the best way to measure the extent of forecastability.

I applaud the attempt to link the continuous time theory with the high frequency (5-minute) discrete time data but feel that the link is still tenuous: the theory does not satisfactorily explain the main features of the data, the distribution shapes shown in Figs 2 and 3 and the 'long memory property' illustrated in Fig. 5, both known with daily return data. A satisfactory explanation of the property can be constructed by using a process with breaks in the mean and this could certainly be embedded in the Ornstein–Uhlenbeck model. Some further thought is required to bridge the continuous to the discrete time gap. It is difficult to think of realistic decisions that can be made in continuous time but not at 5-minute intervals or that the empirical results would change with a smaller interval.

The very large data sets found here make hypothesis testing virtually impossible: any two models will be 'significantly different' and it is very difficult to have a precise null hypothesis that is not strongly rejected. It is extremely unlikely that any model, linear or not, will have perfectly white noise residuals. This whole area is likely to remain an exciting one for statisticians and this paper suggests plenty of interesting topics.

J. E. Griffin and M. F. J. Steel (*University of Kent at Canterbury*)
We take this opportunity to congratulate the authors on an impressive paper, demonstrating the potential of Ornstein–Uhlenbeck (OU) processes in modelling high frequency financial data. The paper is most explicit about the theory of OU processes and the analytic tractability that they provide when used to model volatility processes. Empirical issues are less deeply explored, and, in particular, no formal likelihood-based inference is conducted.

In this comment we shall focus on formal inference with these models, cast in a Bayesian framework, exploring the suggestion provided in Section 5.4.2 in detail. In particular, we use the parameterization in terms of the shocks $\{\eta_n\}$ and assume an OU process for the volatilities $\sigma^2(t)$ with $\Gamma(\nu, \alpha)$ marginals. In this case, the series representation is finite (see equation (32)) and no truncation is required to sample from the volatility process.

We analyse daily changes in the Standard and Poors 500 stock price index over the years 1980 until 1987 ($T = 2023$). The same prefiltered series was used by for example Jacquier *et al.* (1994). Proper, but vague, priors are used throughout. The Markov chain Monte Carlo scheme is simplified by analytically integrating out the parameter ν. A Markov chain Monte Carlo sampler on $(\mu, \beta, \alpha, \lambda, \eta_1, \ldots, \eta_T, \sigma^2(0))$ is then implemented, where values for $\sigma^2(0)$ are drawn from the prior marginal process.

Fig. 12 graphs the posterior mean values for the volatilities σ_n^2, based on taking every 10th value from a chain of 50000 draws after a burn-in of 5000. Comparing these posterior means with the actual data clearly indicates high mean volatility in periods of large fluctuations. Posterior standard deviations for the volatilities are roughly a third of the means.

We feel that this testifies to the feasibility of formal likelihood-based inference in the context of stochastic volatility models based on Lévy-driven OU processes. Extensions to superpositions of OU processes, non-gamma marginals and the inclusion of a leverage effect should be quite feasible.

David Hobson (*University of Bath*)
Barndorff-Nielsen and Shephard propose a novel class of models for stochastic volatility which have the

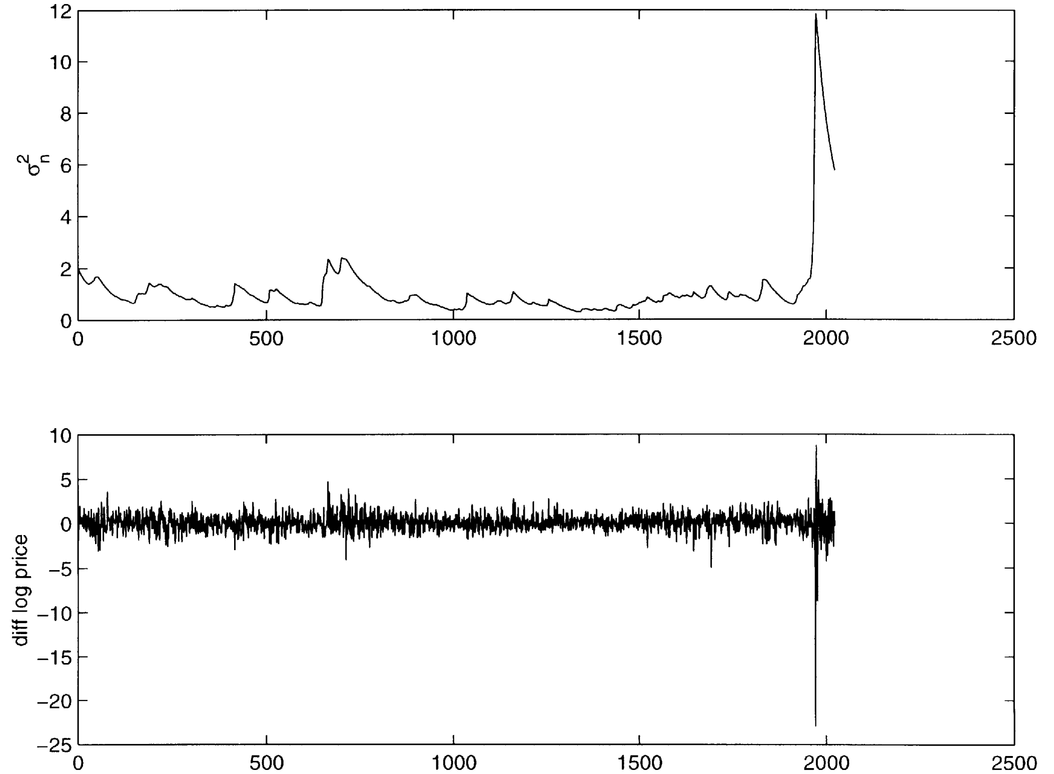

Fig. 12. Posterior volatility means and data

nice interpretation that volatility shocks, which could be thought of as arrivals of new information, happen in discrete packets. These models have the useful property that the asset price is continuous, and that the integrated squared volatility is a tractable random quantity. The integrated squared volatility is a fundamental quantity in the Black–Scholes formula and derivative pricing.

For plain options the purpose of models is not to price derivatives (prices are determined by the market) but instead to explain and predict observed biases. For example two stylized facts about call prices are that the Black–Scholes implied volatilities exhibit 'smiles' and 'skews'.

Stochastic volatility is a potential explanation for the smile effect. In this context the precise model class (generalized autoregressive conditional heteroscedastic, autonomous diffusion or Barndorff-Nielsen and Shephard) is rarely important; once calibrated many models are likely to exhibit qualitatively similar behaviour. The tractability of the authors' model is a highly desirable feature, but the ability to explain smile effects is unlikely to be grounds for the choice of one specification of background-driven Lévy processes above another; the simplest choice may be most appropriate.

Skews in implied volatility can be explained by leverage effects. There are two standard ways to incorporate leverage into models: first by making volatility a decreasing function of the price level and second by using the same sources of uncertainty to drive both the volatility and the price processes. Barndorff-Nielsen and Shephard choose the second approach, but since their volatility is driven by a process with jumps this means that the price process now also has jumps. Thus the inclusion of a leverage term has the undesirable property that it completely changes the character of the price process. Is there any way of incorporating leverage such that tractability is preserved but such that the price remains a continuous process?

Jens Ledet Jensen (*University of Aarhus*)

The authors present a very interesting class of models. In my comments I shall concentrate on questions related to the interpretation of data.

The use of stochastic volatility models raises some questions within the field of inverse problems. The fundamental question is how much do the data tell us about the underlying volatility process? Can we distinguish between the models used here and, say, a diffusion-based model for σ? Even the much simpler problem of determining the marginal distribution of $\sigma(t)$ is quite difficult. Experiences with similar data show that a distribution concentrated in five points, say, gives a fit which is comparable with the use of an inverse Gaussian distribution. Thus an alternative model could be a hidden Markov model with a small number of states. The methods discussed in Hartvig *et al.* (2001) could be used here to investigate the variability in the possible distributions of $\sigma(t)$.

Turning to the second aspect of the modelling, namely the correlation structure, I again wonder how much the data actually tell us. How do we distinguish between the Ornstein–Uhlenbeck models used here with a correlation very close to 1 and a description in terms of a slowly varying trend (or a piecewise linear process)? The latter is somewhat easier to understand intuitively. Can we use the ability to predict future volatility to distinguish two models? Also, more fundamentally, it seems of interest to ask the question whether the volatility is locally constant. In Jensen and Pedersen (1997) an initial attempt is made to model similar data from the point of view of a slowly varying trend. A piecewise constant or piecewise linear process can be modelled easily by a hidden Markov model and the analysis will then proceed using the Kalman filter technique.

Finally, I wonder whether the intraday standardization introduces some extra correlation. Can the intraday variation somehow be explained by covariates?

M. C. Jones (*The Open University, Milton Keynes*)
This is clearly an excellent paper. I have just one small point. Should I be worried that stochastic volatility models involve mixing a normal distribution whose mean is of the dimensionally incorrect form μ plus β times the variance rather than μ plus β times the standard deviation?

A. J. Lawrance (*University of Birmingham*)
The illuminating presentation at the meeting justified the Royal Statistical Society's tradition that it is often good to attend the meeting before reading the paper, not to mention the associated social benefits. Technically, this paper is a *tour de force* which I much admire for its desire and achievement of tractability. The authors refer to financial matters and I would have liked more exemplification; mainly they refer to logarithmic returns and think of their models as a more statistically realistic replacement for geometric Brownian motion. There is mention of options and their pricing, but how many of us make use of these rather esoteric assets? What we may have is a modest or otherwise financial portfolio in a limited number of stock-market assets. The behaviour of their prices over time is our concern and these might be modelled by the authors' non-Gaussian Ornstein–Uhlenbeck models. Then a natural question is to consider whether a form of the Gaussian-based mean–variance portfolio analysis originated by Markowitz might be developed to assist investment decisions. Financial risk would no longer be described by the variance and so ideas of optimality would need to be reformulated in other more probabilistic ways; they would necessarily involve aspects of the behaviour of joint non-Gaussian Ornstein–Uhlenbeck processes. This point just serves to reinforce the authors' comments that there is plenty of room for further development of their work, perhaps in ways which will assist those modestly concerned with 'personal finance' rather than 'mathematical finance'. One small technical comment based on recent experience—I think that the authors should be cautious if statistically using auto-correlations of squares with returns which have not been adjusted to zero mean. It is true that they will be 0 for independent processes but their use as a measure of non-linear dependence includes a level effect.

Anthony W. Ledford (*University of Surrey, Guildford*)
The modelling approach developed by the authors provides a novel extension to current stochastic volatility (SV) methodology and yields a flexible probabilistic framework within which analytic results are often tractable. From a statistical estimation perspective, though, the framework is less tractable and has significant computational difficulties that so far have prevented exact likelihood or inference based on Markov chain Monte Carlo methods from being undertaken routinely. Given this, it is not clear what advantages the suggested framework currently offers for applications. Resolving these estimation issues will be key in widening the appeal of these models and furthering their use by practitioners. Somewhat related to this, diagnostics for both model selection and assessing model adequacy are required, as are extensive analyses of simulated and real data sets.

The underlying assumption that volatility is driven by a Lévy process with positive increments has conceptual appeal and clear interpretability as a model for the effect of fast breaking news. Although discontinuities are allowed at this latent level, the model adopted for most of the paper assumes continuity at the observation level. Is this for convenience or because asset prices are actually continuous? More generally, it is routinely assumed in SV modelling that the observations have conditional Gaussian distributions. Is this assumption worth relaxing? If it is then how is the observation level stochastic differential equation affected?

The log-density plots shown in the paper allow the behaviour in the tail to be examined informally. Statistical extreme value methodology provides additional tools for this and is playing an increasingly important role in both the theory of financial modelling and its application in financial risk control. See, for example, Embrechts *et al.* (1997). In contrast with the approach adopted by the authors, which is to model the observed discrete time process throughout the bulk and tails of its distribution by using an underlying continuous time model, most (but not all; see Leadbetter *et al.* (1993)) extreme value methodology is discrete time based and focuses on tail properties alone to reduce the possibility of bias when inferences relating to extreme events are required. The resulting marginal distributions and dependence measures are different from those which are relevant for describing the overall process. For details, see Leadbetter *et al.* (1983), Davison and Smith (1990), Leadbetter (1983), Ledford and Tawn (1997) and Coles *et al.* (1999). Quantifying the extremal properties of the models proposed remains an area for further research.

N. N. Leonenko (*Cardiff University*)

Ole Barndorff-Nielsen and Neil Shephard have written an excellent paper on Lévy-driven Ornstein–Uhlenbeck (OU)-type processes and their applications in financial econometrics. I have two comments on this stimulating paper.

Construction of stochastic volatility processes with long-range dependence

The idea (Section 3) of using the superposition of non-Gaussian OU processes to construct tractable stochastic volatility (SV) models with long-range dependence (LRD) or quasi-LRD is a very good one. Similar arguments have been used recently by Oppenheim and Viano (1999) and Iglòi and Terdik (1999). An alternative SV continuous time model with LRD and non-Gaussian marginal distributions is discussed in Taqqu (1979). This is referred to as the Gaussian subordination model. The non-Gaussian marginal distributions can be obtained by the use of non-linear transformations of Gaussian processes with LRD. Other classes of SV processes with given marginal distributions and LRD can be constructed. The bivariate densities of these processes have diagonal expansions (see Anh and Leonenko (1999) and the references therein). In particular, there is a class of stochastic processes with χ^2 marginal distributions and LRD. However, a natural mathematical object to describe LRD is the fractional operator (see Rosenblatt (1976), Chambers (1996) and Leonenko (1999)). This is in contrast with the approaches in which LRD is obtained from the noise term (see Comte and Renault (1996)) or by randomization of the regression coefficient in OU processes (see Barndorff-Nielsen (2000)) or via random initial conditions in deterministic partial differential equations (see Woyczynski (1998), Leonenko (1999) and Anh and Leonenko (1999)). Hence, if we can obtain LRD from the fractional derivatives in the fractional version of the Langevin equation (see Anh *et al.* (2000)) or the Langevin equation with delay (see Inoue (1993)), then we may use the noise term to represent other effects such as infinitely divisible distributions (the Lévy noise) or intermittency (see Anh *et al.* (2000)). Another alternative to obtaining quasi-LRD approximations to stationary time series is discussed in section 6.4 of Golyandina *et al.* (2000).

Statistical inference

Barndorff-Nielsen and Shephard have presented an excellent survey of statistical methods for SV models (Section 5). Nevertheless, they do not pay enough attention to estimation in the frequency domain, which has considerable potential for non-Gaussian data if we can use not only the second-order information (the second-order spectral densities) but also the higher order information (the higher order spectral densities). Some results in this direction for discrete time stochastic processes have been obtained by Leonenko *et al.* (1998).

Developing such alternative SV models and the corresponding statistical methods is certainly an interesting area for future research.

Sergei Levendorskii (*Rostov State University of Economics*)

I would like to point out yet another possibility of analytically tractable modelling of non-Gaussian processes having mean reverting features, by either calculating the infinitesimal generator of a Feller process obtained from a general diffusion process by subordination or explicitly defining the infinitesimal generator as a pseudodifferential operator with non-constant symbol; the process itself is constructed by using the representation theorem for analytic semigroups. As an example of the second approach, one can visualize the normal inverse Gaussian Lévy process with state-dependent parameters; and to reproduce mean reverting features it suffices to make the skewness parameter state dependent. Both constructions are studied in Barndorff-Nielsen and Levendorskii (2000). It is shown that processes obtained by the first approach form a subclass of processes obtained by the second approach, and that under realistic assumptions the generators differ little.

Approximate pricing formulae for European options are derived; they are as simple as the corresponding formulae for the normal–inverse Gaussian Lévy process. Similarly, we can try to derive approximate formulae for barrier options and touch-and-out options, i.e. options which pay a fixed amount when the strike price level is crossed from above (down-and-out put option) or below (up-and-out call option); these possibilities are currently under investigation.

Benoit B. Mandelbrot (*Yale University, New Haven*)

The authors aim for two central features of financial price changes that I discovered in the 1960s: far from being Gaussian and independent, price changes have long tails and an infinite span of dependence. Being pleasantly surprised to be asked to comment on this paper, I regret to say that I see little purpose or merit to it.

To model prices, I proposed parsimonious models based on 'dilation invariances', first for long tails and dependence separately, then (Mandelbrot, 1997) by using multifractals to combine both features. As is supposed to be the case, the inputs are sparse, versatile and transparent. The outputs are rich and in part unexpected. Easy computations agree with the facts (quantitatively and also visually) and open new questions of direct concrete importance. The invariances, which of course remain to be justified, involve intrinsic numerical invariants. These are quantities like the fractal dimension, the Hölder exponent and the intrinsic invariants of multifractals. Together, they gave for the first time a quantifiable intrinsic meaning to the loose notion of 'roughness', which occurs widely but until then could not be measured.

A popular counter-proposal formally represents non-Gaussianity by a mixture of Gaussian building-blocks, and non-independence by a mixture of Gauss–Markov (Ornstein–Uhlenbeck) linear building-blocks. This paper proposes a new family of building-blocks: of staggering and unmotivated complication. They involve (via the Lévy measure) an infinite number of parameters: each is individually tunable but nearly all correspond to nothing concrete, old or new. The difficulties in parameter estimation are sketched but a good fit, even if achieved, would bring no clear benefit.

This new proposal also fails to represent at least one essential class of financial prices with which I am familiar. But I do not think that a further generalization is necessary, or that it is useful to put on the record my reactions to diverse details of the paper.

A more general issue arises. I watched many models when they were struggling to take off; all the successful ones started with a light initial load. Among those initially weighted by an unorganized infinite mixture, some involve statistically acceptable representations but not one has left the ground. Parsimony always pays; early in a theory, it is essential.

Nour Meddahi (*Université de Montréal*)

The authors are to be congratulated on a stimulating and a comprehensive paper on volatility modelling. I have some brief comments.

Time series dependences and marginal distribution

One of the major contributions of the paper is the separation between the marginal distribution of the volatility and its dynamic structure. This is very important in volatility modelling because several models in the literature (e.g. generalized autoregressive conditional heteroscedastic) imply that, with the empirical value of the persistence of the volatility, the fourth moment of the returns is not finite. It is important to note that this separation result is general and can be considered Brownian-type stochastic volatility (SV). More precisely, Chen *et al.* (2000) argue that a more parsimonious approach for continuous time modelling is to specify the unconditional distribution of the process and the diffusion parameter. Therefore the separation between the marginal distribution and the dynamics structure holds. It is worthwhile to note that we do not observe the volatility but a noisy version (squared return).

Therefore we cannot estimate directly the marginal density of the volatility. Finally, it is not necessary to assume that the positive noise which explains empirically 90% of the variance has a self-decomposable characteristic function.

Conditional information
The daily integrated volatility $\int_{(n-1)\Delta}^{n\Delta} \sigma^2(u)\,\mathrm{d}u$ is the variance of the daily return given the information $J_n = \sigma[\sigma(\tau\Delta),\ x^*\{(\tau-1)\Delta\},\ \tau \leqslant n]$ (when $(\mu, \beta, \rho) = (0, 0, 0)$). However, in practice, we have only (in the best case) $J_{n-1} = \sigma[\sigma\{(\tau-1)\Delta\};\ x^*\{(\tau-1)\Delta\},\ \tau \leqslant n]$. Therefore the volatility of interest is $\operatorname{var}(y_n|\tilde{J}_{n-1})$ which is an affine function of $\sigma^2(n-1)$. Indeed, this volatility is a filter of the integrated volatility. However, if we are interested in smoothing the volatility or on option pricing (by simulation), then the integrated volatility is the volatility of interest (see Meddahi and Renault (2000a)).

The variance of the variance
A potential limitation of the dynamics of the volatility is that the variance of the variance is constant in continuous time (but not in discrete time). This is in contradiction with the usual SV models (square root or log-normal). The empirical implications are not clear. A potential alternative specification of the volatility σ in equation (4) is to say that $\sigma^2(t) = f\{\tilde{\sigma}^2(t)\}$ where $\tilde{\sigma}^2(t)$ is defined by equation (2) (the paper adopts $f(x) = x$). This maintains the separation property but, in general, the autoregressive moving average (ARMA) structure will not hold. However, if $f(x) = x^i$, $i \in \mathbf{N}$, then the squared returns are ARMA(i, i).

Inference
As shown in Meddahi and Renault (2000b), integrable positive Ornstein–Uhlenbeck are SR–stochastic autoregressive volatility models. Therefore, multiperiod moment restrictions (see Hansen (1985) and Hansen and Singleton (1996)) can be derived for non-linear inference purposes. For instance, we have for $J = 1$:

$$E[y_n^2 - \xi\Delta\{1 - \exp(-\lambda\Delta)\} - \exp(-\lambda\Delta)\,y_{n-1}^2 | y_\tau,\ \tau \leqslant n-2] = 0.$$

In this equation, we can correct the heteroscedasticity of the squared returns.

Michael K. Pitt (*University of Warwick, Coventry*) **and Stephen Walker** (*University of Bath*)
We would like to congratulate the authors on a stimulating paper. We can, following Walker (2000), obtain the error term explicitly in equation (13) of the paper for the gamma, $\Gamma(\nu, \alpha)$, example, illustrated in Fig. 1. Without loss of generality, we restrict ourselves to the $\Gamma(\nu, 1)$ marginal. Using the notation of equation (13) of the paper,

$$\sigma^2(t) = \exp(-\lambda t)\,\sigma^2(0) + \exp(-\lambda t)\,\epsilon(\lambda t),$$

where $\epsilon(\lambda t)$ is an independent mixture random variable. Explicitly,

$$\epsilon(\lambda t) \sim \mathrm{Ga}(z, 1),$$
$$z \sim \mathrm{PoGa}\left\{\nu, \frac{1}{\exp(\lambda t) - 1}\right\},$$

where $z \sim \mathrm{PoGa}(\alpha, \beta)$ means $z|w \sim \mathrm{Po}(w\beta^{-1})$, and $w \sim \mathrm{Ga}(\alpha, 1)$. We obtain $\Pr\{\epsilon(\lambda t) = 0\} = \exp(-\nu\lambda t)$. The conditional density of $\epsilon(\lambda t)$, given that it is greater than 0, is

$$f_{\epsilon|\epsilon>0}(x) = \frac{1}{1 - \exp(-\nu\lambda t)} \sum_{z=1}^{\infty} \mathrm{Ga}(x|z;\ 1)\,\mathrm{PoGa}\left\{z|\nu;\ \frac{1}{\exp(\lambda t) - 1}\right\}.$$

We have found that this is very close to an exponential density when λt is small. The $\Gamma(\nu, 1)$ case shows that Markov chain Monte Carlo techniques may be directly applied as we have an explicit one-step-ahead density for $\sigma_n^2|\sigma_{n-1}^2$, in equation (7) of the paper. Efficient Markov chain Monte Carlo techniques would require that blocks of variances are jointly proposed as moves. The smooth, in the parameters, particle filter of Pitt (2000) can be used to provide maximum likelihood solutions to the more general models considered in this paper. This may provide a unified approach to estimating models of the type considered by Professor Barndorff-Nielsen and Professor Shephard, since it is only required that we can simulate from the transition density $\sigma_n^2|\sigma_{n-1}^2$ and evaluate the measurement density $y_n|\sigma_n^2$: the same

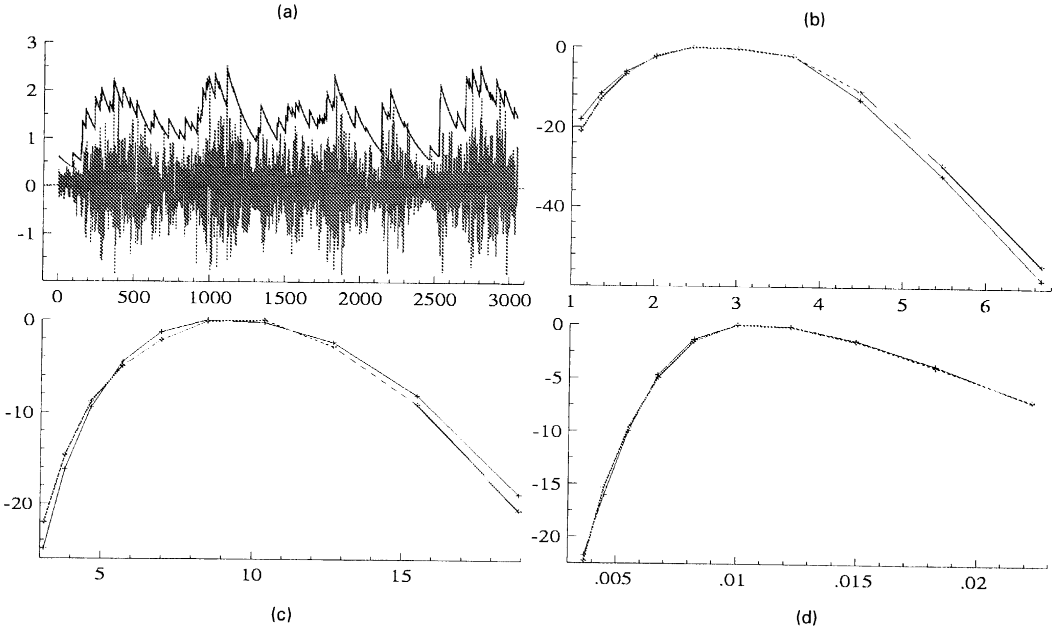

Fig. 13. Profile likelihood for the three-parameter gamma model: (a) simulated data set; (b) log-likelihood ($\nu|\alpha$, λ); (c) log-likelihood ($\alpha|\nu$, λ); (d) log-likelihood ($\lambda|\nu$, α)

requirements for the general auxiliary particle filter of Pitt and Shephard (1999). We can illustrate by simulating a data set of size 3050 where $\sigma_n^2|\sigma_{n-1}^2$ arises from the $\Gamma(\nu, 1)$ process with persistence parameter λ and $y_n|\sigma_n^2 \sim N(0, \sigma_n^2/\alpha)$. We take a unit sampling interval, $\alpha = 8.5$, $\nu = 3$ and $\lambda = 0.01$ (comparing with the authors' example in Fig. 1). In Fig. 13, we display the profile log-likelihood from two simulations of the smooth particle filter. The three parameters are shown together with the actual simulated data set (Fig. 13(a)).

M. B. Priestley (*University of Manchester Institute of Science and Technology*)
The authors have presented an extremely interesting paper, the mathematical aspects of which I found very stimulating. I was particularly pleased to note that the authors develop their models in the more mathematically elegant continuous time format rather than the discrete time format which has dominated time series analysis over the past three decades. I do not, however, have sufficient expert knowledge of the field of financial economics to be able to judge whether their models for stochastic volatility are realistic and how well they compare with alternative models.

The Ornstein–Uhlenbeck (OU) model for $\sigma^2(t)$ (equation (2)) seems rather restrictive since, as the authors point out, $\sigma^2(t)$ is then constrained to decay exponentially between jumps. To obviate this constraint the authors then propose a more general model in which $\sigma^2(t)$ is the sum of a number of independent OU processes. However, this generalization seems rather strange in that it raises the problem of determining the value of m, the number of independent OU processes in the overall model. (In the paper the value of m seems to be chosen in a rather arbitrary fashion.) Since the OU model is essentially a continuous time autoregressive AR(1) scheme (with a non-Gaussian driving process) it would seem more natural to consider instead a generalization to higher order AR schemes — these would lead to processes with essentially the same forum of autocovariance functions as the three generated by adding independent OU processes — of equation (33).

The crucial test of any model is how well it matches the data — as measured, for example, by its predictive accuracy. Although the prediction of the original processes $x^*(t)$ or y_n may not be of great interest within the context of financial economics it would, nevertheless, be interesting to compare the predictive efficiency of the authors' models with alternative models such as a model based on one of the standard non-linear time series models for $\log\{\sigma^2(t)\}$.

Eric Renault (*University of Montreal*)
The paper proposes a new continuous time stochastic volatility (SV) model for financial asset returns but it keeps the general framework termed 'stochastic autoregressive volatility' (SARV) by Andersen (1994): up to extensions by superposition or subordination, the volatility dynamics on any time interval h are captured by the stationary autoregression

$$\sigma^2(t+h) - \sigma^2 = \exp(-\lambda h)\{\sigma^2(t) - \sigma^2\} + \upsilon_h(t+h), \qquad \lambda > 0, \quad E[\upsilon_h(t+h)/\sigma^2(\tau), \tau \leqslant t] = 0. \tag{77}$$

In this context, emphasis is cleverly put on Lévy processes with a self-decomposable marginal law, as a necessary consequence of an assumption of independence between $\upsilon_h(t+h)$ and $\sigma^2(t)$. My comments assess the relative advantages of alternative models in the SARV class.

For a fair comparison, it is first worth noting that some nice properties of the SARV processes are not specific to the subordinators considered in this paper. In particular, general (semiparametric) SARV modelling provides by definition a joint linear prediction of the return and the squared return (in so far as the risk premium is linear with respect to $\sigma^2(t)$) and related characterizations of leverage effects (Meddahi and Renault, 1996).

Although leverage effects are poorly captured by the popular generalized autoregressive conditional heteroscedastic (GARCH) model since it adds to equation (77) the drastic restriction of perfect correlation between $\upsilon_1(t+1)$ and the squared innovation of the return process, it is extreme to replace perfect correlation by an independence assumption between $\upsilon_1(t+1)$ and $\sigma^2(t)$. Actually, it is sensible to allow the conditional variance process to be conditionally heteroscedastic: in the square-root model (Section 6.2.2) the conditional variance of $\sigma^2(t+h)$ is proportional to $\sigma^2(t)$, which is consistent with the linear specification of the risk premium. Up to a linearization, this proportionality is also underpinned by the popular log-normal SV model which, by directly specifying Gaussian autoregressive dynamics for $\log\{\sigma^2(t)\}$, exempts us from having to resort to non-Gaussian Lévy processes and opens the door for more versatile models of superposition (Comte and Renault, 1996).

By forcing $\sigma^2(t)$ to move up entirely by jumps, the subordinator SV model may lack flexibility. For instance, any leverage effect will produce jumps in the stock price process; the two effects cannot be disentangled. Moreover option prices will be strictly increasing functions of the underlying volatility process which features jumps. The implied jump risk merits thinking about implementing the proposed option pricing and hedging theory (see also Stute (2000)).

Overall, the framework of infinitely divisible (ID) probability distributions is well suited to volatility modelling. ID distributions can be seen as limits of compound Poisson distributions, which is appealing for simulation and economics interpretations in terms of flows of the arrival of information. But, besides the model proposed, we may keep more classical volatility models which also provide an ID marginal law of $\sigma^2(t)$: gamma distributions (square-root processes), log-normal or inverse gamma (continuous time limit of a GARCH(1, 1) process).

Jan Rosiński (*University of Tennessee, Knoxville*)
I would like to announce a new result (Rosiński, 2000) on stochastic series representations that was motivated by the authors' work and can improve the simulation of certain Ornstein–Uhlenbeck (OU) processes. The practical use of series representations for simulation (Section 2.5) can be greatly facilitated when the inverse of the tail mass of the Lévy measure has a closed form. However, quite often this is not so; examples include the inverse Gaussian and a more general class of exponentially tempered stable (ETS) laws. An OU process is said to be an ETS OU process if $\sigma^2(t)$ is a positive infinitely divisible random variable with Lévy density of the form

$$u(x) = Ax^{-\alpha-1}\exp(-Bx), \qquad x > 0, \tag{78}$$

where $\alpha \in (0, 1)$ and A, $B > 0$ are parameters. The inverse Gaussian law is the special case of the ETS law with $\alpha = \frac{1}{2}$. The volatility process is given by

$$\sigma^2(t) = \exp(-\lambda t)\,\sigma^2(0) + \exp(-\lambda t)\int_0^{\lambda t} \exp(s)\,\mathrm{d}z(s) \tag{79}$$

where $z(t)$ is the background driving Lévy process with the tail mass of the Lévy measure of $z(1)$ given by

$$W^{+}(x) = Ax^{-\alpha} \exp(-Bx), \qquad x > 0 \tag{80}$$

(Section 2.2). The volatility process is composed of two independent parts. The second part can be simulated by the method discussed in Section 2.5 but the inverse W^{-1} of W^{+} must be found numerically. Because of the large number of terms needed to simulate for each time step, particularly when α is close to 1, the errors due to numerical inversions of W^{+} can accumulate substantially. Another problem is with the simulation of $\sigma^2(0)$, appearing in the first part of equation (79), whose density is known only in a few cases of α, most notably, for $\alpha = \frac{1}{2}$. Simulation of $\sigma^2(0)$ by the series representations of Section 2.5 requires a very large number of numerical inversions of the function

$$U(x) = \int_x^\infty At^{-\alpha-1} \exp(-Bt)\,\mathrm{d}t.$$

These difficulties disappear when we apply alternative series representations based on a random cut-off of jumps of stable processes. Such representations hold for general ETS Lévy processes with Lévy density

$$u(x) = A_{\pm}|x|^{-\alpha-1} \exp(-B_{\pm}|x|), \qquad x \neq 0.$$

As an application we obtain explicit formulae for both parts of $\sigma^2(t)$. Let $\{e_i\}$ be a sequence of independent and identically distributed (IID) exponential random variables with parameter B, independent of the other random sequences defined in Section 2.5. Then, in the notation of Section 2.5,

$$\int_0^{\lambda t} \exp(s)\,\mathrm{d}z(s) \overset{\mathcal{L}}{=} \sum_{i=1}^{\infty} \{(a_i\alpha/A\lambda t)^{-1/\alpha} \wedge e_i\} \exp(\lambda t r_i). \tag{81}$$

Let $\{v_i\}$ be a sequence of IID uniform random variables on [0, 1], independent of all previous random sequences. Put $w_i = e_i v_i^{1/\alpha}$. Then

$$\sigma^2(0) \overset{\mathcal{L}}{=} \sum_{i=1}^{\infty} (a_i\alpha/A)^{-1/\alpha} \wedge w_i. \tag{82}$$

In conclusion, the stochastic series representations can be easily implemented in simulation of OU volatility processes with ETS marginal distributions. In the inverse Gaussian case ($\alpha = \frac{1}{2}$), $\sigma^2(0)$ can be simulated directly but the use of equation (82) improves the previous method.

Ken-iti Sato (*Nagoya University*)
Professor Barndorff-Nielsen and Professor Shephard have made an important contribution to the wide applicability of processes of Ornstein–Uhlenbeck type. It would be interesting to study theoretically the class of processes described by equations (6) or (8). I shall make some comments on subordination and self-decomposability.

The subordination that the authors are using in Section 6 is a much wider concept than the usual subordination in the theory of stochastic processes. The latter means time substitution by independent subordinators; here subordinators mean one-dimensional increasing processes with stationary independent increments starting at 0. This was introduced by Bochner (1949); it transforms Markov processes to Markov processes and Lévy processes to Lévy processes, as is expounded in Sato (1999). But the subordination in the wider sense does not have this property.

Self-decomposability is a concept introduced by Lévy (1937) answering a problem posed by Khintchine. It characterizes the class of limit distributions of normalized sums of independent random variables satisfying the uniform asymptotic negligibility condition. See Loève (1977, 1978). Another name for this class is class L. Many distributional properties (such as unimodality and degree of smoothness) are known in this class. See Yamazato (1978) and Sato and Yamazato (1978). Theorem 1 of this paper, establishing the one-to-one correspondence between self-decomposable distributions and stationary processes of Ornstein–Uhlenbeck type, was proved by Wolfe (1982) and Sato and Yamazato (1983). See Sato and Yamazato (1984) for historical comments.

In view of the importance of self-decomposability expressed by theorem 1, the following remarks should be of interest. In the case of subordination (in the usual sense) of Brownian motion, the inheritance of self-decomposability from subordinators is known. This was proved by Halgreen (1979) and Ismail and Kelker (1979). The same is true for strictly stable Lévy processes in place of Brownian motion. Recently I have extended this fact to Brownian motion with drift (Sato, 2000). But it is an open problem whether an extension to (not strictly) stable Lévy processes is possible or not. A multivariate extension of subordination and its connection to multivariate self-decomposability and stability are discussed in Barndorff-Nielsen *et al.* (2000).

Stephen J. Taylor (*Lancaster University*)
Papers about the prices of financial assets are rare in the journals of the Royal Statistical Society, although their influence is often significant. The empirical study by Kendall (1953) applied primitive computational technology to what are now called low frequency data, to obtain the historically important conclusion that stock price indices move like random walks. The superb and epic paper by Barndorff-Nielsen and Shephard also makes a significant contribution to our understanding of appropriate models for the development through time of asset prices.

The theoretical analysis is impressive. The economic reasons for the existence of stochastic volatility remain unclear, although the occasional announcement of relevant information and the psychology of traders and investors must contribute to any satisfactory explanation. Certainly the background driving Lévy process set-up in this paper, as illustrated by Fig. 1(b), is plausible and permits the jumps to be associated with news announcements. The success of the authors' framework is shown by the generally satisfactory description of observed distributions and autocorrelations, as shown by Figs 3 and 5. The implications of the theoretical models for finance researchers include interesting new ways to value options, which will require assumptions about the risk premium associated with the unpredictable jumps in the volatility process.

The paper makes remarkable progress towards explaining the empirical volatility dependence for all timescales, with results for lags from 5 minutes to 100 days, as shown by Fig. 5. I note, however, that the empirical autocovariance function has a local minimum at about 0.5 days and that the empirical values are well above the fitted curve from 0.70 to 0.85 days (see Fig. 5(a), where it is also not easy to see what happens for lags less than 0.1 days).

Taylor and Xu (1997) also investigated the Deutsche Mark–dollar exchange rate and presented the first application of the quadratic variation estimator used to obtain Fig. 4(c). In that paper we observed that intraday volatility seasonals reflect local time, which is not a simple shift of Greenwich Mean Time because clocks go back and forth. Thus the single spike on Fig. 2(a) is surprising, because the market opening in New York and the announcement of macroeconomic news occur at two possible Greenwich Mean Time times depending on the season, winter or summer. We also showed for one year of data that there is a statistically significant day of the week effect in the quadratic variation statistics. On average these statistics increased monotonically through the week, probably reflecting the timing of macroeconomic news announcements.

Howell Tong (*University of Hong Kong and London School of Economics and Political Science*) **and Hailiang Yang** (*University of Hong Kong*)
We congratulate the authors on a very timely paper. We would like to say a little about using the Esscher transformation, introduced in Gerber and Shiu (1994), to option pricing. The advantages of this approach are that it can deal with both continuous and discrete time models in a unified way and it enables us to obtain a unique price even in the incomplete market case, where it is known that the option price is consistent with the price calculated using the utility maximization framework. Theoretically, we can use Gerber and Shiu's method for any model (including stochastic volatility models) as long as we know the distribution of the underlying asset's return. However, it is not easy to find the distribution of the underlying asset return if the underlying asset price follows a stochastic volatility model such as that in this paper. The paper uses the series representation to simulate the volatility process. The price of derivatives is then obtained. A possible alternative approach to the option pricing problem under stochastic volatility is to use the 'random Esscher transform' introduced in Siu *et al.* (2001). There the random Esscher transform was introduced to deal with the risk measures for portfolios containing derivative securities. Let $F(x, t)$ denote the distribution function of the stock return at time t. To capture the subject view and risk preference, the random distribution of X_t is defined via the random Esscher transform

$$F(x, t; \Theta) = \frac{\int_{(-\infty, x]} \exp(\Theta y)\, F(\mathrm{d}y, t)}{M(\Theta, t)}$$

where $M(\theta, t)$ denotes the moment-generating function of X_t and Θ is a random variable with a prior distribution which represents the subject view and/or risk preference of the individual trader. For the option pricing under stochastic volatility model, we must treat σ rather than μ as a random variable.

Alexander Yu. Veretennikov (*Leeds University*)
I have a comment and a question. In the first sentences of the paper linear models of volatility based on Brownian motion are criticized because they only provide Gaussian processes with light tails.

Comment
I would like to draw attention to the fact that various non-linear Brownian models may be constructed with explicit marginal stationary distributions, with arbitrarily heavy stationary tails, non-Gaussian, and, finally, which mimic 'long-range dependence' despite the fact that they are Markovian. A wide class of such processes can be described by an Itô equation

$$\mathrm{d}v_t = \mathrm{d}B_t + r\,h(v_t)\,\mathrm{d}t, \qquad \text{for some } v_0,$$

where $r > 0$ and $h(v) = -\mathrm{sgn}(v)(a + |v|)^{\alpha}$, $a > 0$, $\alpha \geqslant -1$.

(a) If $-1 < \alpha < 0$ then v_t is ergodic and the stationary marginal density has subexponential tails.
(b) If $\alpha = -1$ and $r > \frac{3}{2}$ then v_t is still ergodic and an explicit expression for the stationary density is available: polynomial weak and β-mixing hold true and this imitates long-range dependence with polynomial decay, and the stationary tails decrease with a polynomial rate depending on r.
(c) If $\alpha \geqslant 0$ then the tails are exponential or lighter.
(d) The Ornstein–Uhlenbeck process is included in the case $\alpha = 1$ (if we take $a = 0$; note that $a > 0$ is essential only if $\alpha < 0$).
(e) Other classes with similar properties can be constructed with changes of function h near zero.

Question
Might this be useful for models of volatility?

S. G. Walker (*University of Bath*)
My comments on this fine paper are reserved for Section 2, and in particular Section 2.5. A general representation of a Lévy process of the kind considered by Barndorff-Nielsen and Shephard has been given by Ferguson and Klass (1972). Indeed, the main result of Section 2, formula (31), can be easily derived from the work of Ferguson and Klass.

In Ferguson and Klass (1972) the Lévy measure is written as $\mathrm{d}N_t(u)$ and

$$-\log(E[\exp\{-\theta\, z(t)\}]) = \int_0^{\infty} \{1 - \exp(-\theta u)\}\,\mathrm{d}N_t(u).$$

In the homogeneous case, $\mathrm{d}N_t(u) = t\,W(\mathrm{d}u)$, where W is as given by Barndorff-Nielsen and Shephard.

The representation of Ferguson and Klass (1972) for $z(t)$ on $(0, \lambda]$ is given by

$$z(t) = \sum_{i=1}^{\infty} J_i\, I\{r_i < n_t(J_i)\},$$

where, as with Barndorff-Nielsen and Shephard, the $\{r_i\}$ are independent and identically distributed from the uniform distribution on [0, 1]. In the homogeneous case, $n_t(u) = t/\lambda$ and $a_i = \lambda\, W(J_i, \infty)$, where, as with Barndorff-Nielsen and Shephard, the $\{a_i\}$ are arrival times of a Poisson process with intensity 1. Here $W(u, \infty) = \int_u^{\infty} W(\mathrm{d}x)$.

Consequently,

$$z(t) = \sum_i W^{-1}(a_i/\lambda)\, I\{t > \lambda r_i\},$$

which leads to the expression (31) of Barndorff-Nielsen and Shephard.

Bas J. M. Werker (*Tilburg University*)
The paper discusses (superpositions of) non-Gaussian Ornstein–Uhlenbeck processes as possible models for the volatility of financial assets. As the authors show, the models thus obtained are both

empirically relevant (Section 5) and analytically tractable (or, at least, easily simulated; Sections 2–4). In finance especially, tractability is important if a realtime implementation is to be achieved. This, then, holds both for the estimation of the model parameters and the pricing of financial derivatives. I would like to focus this discussion on the pricing of derivatives as considered in Section 6.2.

The authors show, in Section 6.2.1, that the leveraged stochastic volatility model does not allow arbitrage. This result is established by showing that there is an equivalent martingale measure. However, this equivalent martingale measure is not unique (as in most stochastic volatility models) whereby derivatives cannot be priced by arbitrage arguments alone. In Section 6.2.2, the authors propose to use a specific measure for derivative pricing. This measure is, in the notation of Section 6.2.2, the equivalent martingale measure Q that is 'closest' to the physical measure P. As the authors note, in case the volatility process σ is independent of the Brownian motion w in model (6), the law of the volatility process σ is the same under P and Q. A similar choice of equivalent martingale measure figures in the Hull and White (1987) model.

Another interpretation of this specific equivalent martingale measure is obtained by noting that it implicitly assumes that the risk in the stochastic volatility does not pay a risk premium, i.e. this risk can be diversified away. Such an assumption is sometimes defended on economic grounds, but it is difficult to maintain empirically. The empirical results in this direction have existed for a long time and are discussed in detail in Guo (1998). Therefore, it seems more reasonable on empirical grounds to consider the appropriate transformation to the equivalent martingale measure, as far as the volatility process is concerned, as an empirical issue. A particularly simple characterization that allows for an estimation of the empirically relevant equivalent martingale measure, using data on option prices, is given in Melenberg and Werker (1999), which also shows how to handle the leveraged case.

Andy Wood (*Nottingham University*)
It is a pleasure to congratulate the authors on a very interesting and stimulating paper which I look forward to studying in greater detail. For the moment, I only have a minor technical question. I was puzzled by the fact that the final sum in expression (32) only contains a finite number of terms almost surely, as a gamma process has an infinite number of jumps on any finite open interval (though 'most' of these jumps will be small). On closer scrutiny, the Lévy density for the gamma process given at the end of Appendix A.2 does not invert as stated in formula (25), which appears to explain the discrepancy. Have I misunderstood something here?

The **authors** replied later, in writing, as follows.

We would like to thank all the contributors to the discussion on our paper. Many of the comments have certainly advanced our understanding of Ornstein–Uhlenbeck (OU) processes and stochastic volatility (SV). We have structured our reply by topic, going through alternative models, inference, Lévy processes, option pricing and other issues.

Alternative models
Several of the discussants have pointed clearly to alternative models which share features, such as second-order properties, with our OU-based volatility models. We mentioned in our paper some diffusion-based alternatives and these are highlighted in the comments by Valentine Genon-Catalot and Catherine Larédo, Eric Renault and Nour Meddahi. These diffusion alternatives are generally non-linear processes with Gaussian increments, with the non-linearity forcing the process to be positive. Our approach is to advocate linear processes with non-Gaussian increments for volatility. Although diffusions have many advantages, only in the Cox–Ingersoll–Ross case (to our knowledge) is it possible to study the cumulant functional of $x^*(t)$, $\sigma^{2*}(t)|\sigma^2(0)$ easily analytically. This is the vital issue in option pricing theory. We think that our models open up a new class of analytic option pricing models. This is studied, following our initial work, by Nicolato and Venardos (2000) and Tompkins and Hubalek (2000b).

Eric Renault points out the work of Andersen on discrete time autoregressive volatility models. It is clear that we should have referenced this important and related work. Of course moving to continuous time changes the model structure very considerably as time aggregation means discrete time increments to integrated volatility do not have an autoregressive structure (although instantaneous volatility does). This point is made forcefully in the work by Meddahi and Renault quoted above. Professor Renault worries that our OU-based model does not allow the conditional variance of volatility to be proportional to the conditional mean. This fear is shared by Nour Meddahi. However, Fig. 6 shows that this is actually the case when we condition on returns, rather than on the unobserved instantaneous volatility.

Peter Brockwell and Richard Davis make an interesting contribution, introducing an autoregressive moving average type of Lévy-based continuous time volatility models. They give conditions on the volatility process so that it is positive. We look forward to thinking about this process in detail. In a sense their comment has answered one of the queries of Maurice Priestley. The other point that Professor Priestley makes is that we should compare the fit of our model with alternative non-linear diffusion-based models. This is surely right, although statistical fit is only one criterion for use. Another, equally important, is that of tractability.

Sir David Cox makes an important point, that we are using a parameter-driven model (Cox, 1981) and so are not really explaining volatility in terms of past data. Instead he suggests an observation-driven model, derived via a Taylor expansion from a general non-linear autoregression. The resulting model is autoregressive conditional heteroscedastic (ARCH) like. Such models are indeed appealing, although the properties of observation-driven models are often difficult to discern. Further, they are often difficult to manipulate when it comes to option pricing theory.

Frank Diebold makes some interesting comments about the marginal distribution of increments to integrated volatility. He argues that his work on realized volatility suggests that it is close to log-normal (LN). The LN distribution is self-decomposable (Bondesson (1992), pages 30 and 59–60; see also Thorin (1977)) and so we could set up an LN–OU process. LN–OU processes have substantially heavier tails than inverse Gaussian (IG)–OU processes, which has some attractions in the context of equity data. We are currently working out the detailed implications of the LN–OU process and hope to report on it in the future. Finally, although IG–OU processes do not temporally aggregate to being IG, calculations suggest that the disagreement is mild (see Barndorff-Nielsen and Shephard (2001b)). We do not yet know whether this is true for LN–OU processes.

Clive Granger points out that the non-normality in our models is built out of a normal distribution. This is true, but the flexibility that is achieved with normal variance–mean mixtures (or, put another way, with subordination of Brownian motion with drift) is extraordinary — allowing us to deal with, for example, the double-exponential distribution favoured by Professor Granger in some of his recent writing. We agree that our linkage with trade-by-trade dynamics is primitive and much work needs to be carried out in this context. Finally, we share his concern about the role of hypothesis testing based on huge data sets.

Benoit Mandelbrot dismisses our models as being extremely complicated. We shall leave it to the reader to decide whether our linear volatility models are more complicated than Professor Mandelbrot's favoured multifractal processes.

Inference

Gareth Roberts and Omiros Papasiliopoulos productively focused on the gamma–OU volatility case, reparameterizing the model into jump times and jump sizes. This approach is also independently introduced by Sylvia Frühwirth-Schnatter. All three of these researchers then design Markov chain Monte Carlo (MCMC) algorithms to sample parameters, jump sizes and times given the returns. This can, of course, be carried out in various ways, with varying degrees of effectiveness. Their discussion studies carefully several approaches. This is clearly an important and productive technique which is, in principle, extendable to the superposition and multivariate cases. Further, the method works with any OU process which has a background driving Lévy process (BDLP) with an integrable Lévy density, for such BDLPs all correspond to compound Poisson processes. This is a wide class of processes. However, it does not include cases, such as the IG–OU process, which do not have an integrable Lévy density, which means that the BDLP has an infinite number of jumps in any finite interval of time, and so some adaptation of the above procedure would be needed.

Professor Griffin and Professor Steel implement an MCMC algorithm via the series representation in the gamma–OU case. We found this very interesting and hope that they will report their results more extensively elsewhere. The comment of Mike Pitt and Stephen Walker was innovative. They suggested a simulation-based approach to estimating the likelihood function for the SV model in the gamma–OU case. This is based on a smooth particle filter which Mike Pitt has been developing. At the moment we do not understand how this approach can be used in cases where the density of σ_n^2, $z(n\lambda\Delta)|\sigma_{n-1}^2$, $z\{(n-1)\lambda\Delta\}$ is unknown (which is the case typically). We hope that Pitt and Walker will report this at length elsewhere. Certainly their comments greatly interested us.

Petros Dellaportas, Emma McCoy and David Stephens have been studying long memory models by the superposition of discrete time AR(1) models. These can then be handled by MCMC algorithms. This approach to long memory is certainly worthy of study. They asked us about the utility of the continuous time modelling. This raises the mathematical difficulty of working in this area, but the choice of Δ is basically in the hands of the econometrician nowadays as prices are mostly recorded in continuous time. Hence basing the analysis in continuous time seems suitable. Further, one of our wishes is to carry out option pricing of these models, which is most easily achieved via continuous time.

Both the above discussants and Enrique Sentana and Frank Critchley asked us about the identification of the superposition of OU processes. It is helpful in thinking about this issue to work with the IG(δ, γ)–OU case, with

$$\sigma^2(t) = \sum_{j=1}^{m} \sigma_j^2(t), \qquad \text{where } \sigma_j^2(t) \sim \mathrm{IG}(\delta w_j, \gamma)\text{–OU},$$

where the weights $\{w_j\}$ are strictly positive and sum to 1, while the corresponding damping values are $\{\lambda_j\}$. To gain statistical identification it is necessary to order either the weights or the damping factors. Under such a set-up the mean, variance and autocorrelation function identify all the parameters in the model and hence this model can be estimated from data. It is this structure which we have recently been using in Barndorff-Nielsen and Shephard (2000) to estimate these models in practice.

Valentine Genon-Catalot and Catherine Larédo express their disappointment that we did not manage to estimate these models off non-second-order information. We share their concern and hope that progress can be made in this area. Our recent work on realized volatility is aimed at improving matters, but there is clearly still much to be carried out.

Enrique Sentana makes a series of points about the statistical basis of our estimation methods. They are well taken and clearly some more work needs to be made in this direction. We have formalized some of these ideas in Barndorff-Nielsen and Shephard (2000). Certainly indirect inference methods may be useful in this context, particularly as generalized ARCH or quadratic ARCH based models seem such obvious auxiliary models in this context.

Bent Jesper Christensen asks us about our leverage model, where he argues for a more traditional log-volatility model with changes in the log-price appearing in the volatility process. Although this model has much merit, it removes the linear structure of the process and so it becomes much less mathematically tractable. Although Professor Christensen is of course correct about the causal story he tells, in terms of observables the two models can produce very similar effects.

David Hobson asks whether we can introduce a leverage effect which allows us to maintain the property that log-prices have continuous sample paths. This would clearly be desirable from a mathematical finance viewpoint. The issues are clearest when $z(t)$ is a compound Poisson process and $\mu = \beta = 0$. Then our model has

$$x^*(t) = \int_0^t \sigma(s)\,\mathrm{d}w(s) + \rho \sum_{j=1}^{N(t)} z_j.$$

We may 'smooth' this by modifying to

$$x^*(t) = \int_0^t \sigma(s)\,\mathrm{d}w(s) + \rho \sum_{j=1}^{N_t} z_j\, h(t - \tau_j)$$

where τ_j is the jth arrival time of the Poisson process $N(t)$ and h is a non-negative continuous function such that $h(s) = 0$ for $s \leqslant 0$, $h(s) > 0$ for $s > 0$ and $h(s) \to 1$ for $s \to \infty$, i.e. we have a shot noise type of behaviour.

Lévy processes

Nick Bingham makes a series of interesting points about Lévy processes. His work with Rudiger Kiesel certainly sounds interesting and we look forward to reading it. Multivariate modelling is challenging and stimulating. His point about quadratic variation is of course true; however, we have recently been studying a finite sample version of it in Barndorff-Nielsen and Shephard (2000). The motivation for it is in dealing with intraday data.

Like Professor Bingham, Professor Benth, Professor Karlsen and Professor Reikvam make very interesting points about multivariate models. Our paper has only scratched the surface of this topic. We know from informal discussions with Professor Benth that he has been thinking about portfolio theory in the context of our models, where the investor is faced with transaction costs. We look forward to being able to read about this work when it is completed. Professor Christensen, Professor Lawrance and Professor Sentana's comments accord with our view that this is a vital topic.

Jan Rosiński's new result on series expansion is highly interesting to us for it removes the requirement to compute the inverse tail mass of the Lévy measure for many problems. In particular it covers the IG case. We have been using this result in Barndorff-Nielsen and Shephard (2001c).

Ken-iti Sato makes some points of historical worth, whereas his new result on self-decomposability and subordination of Brownian motion with drift and work extending subordination to the multivariate case are of particular importance.

Option pricing

Elisa Nicolato and Manos Venardos briefly discuss their work on option pricing for our SV models. This shows that the linear structure of the model means that analytic option pricing results can be found for a wide class of distributions. In particular their result on the leverage case is particularly welcome. This relates also to Robert Tompkins who discusses various estimation methods for these models via option data. This may allow us to have a better understanding of the choice of equivalent martingale measure (EMM).

Stewart Hodges's wide-ranging discussion puts our work in context, and we thank him for this. His comments about our choice of the EMM is of course correct. We hope that we shall eventually be able to understand the choice of the EMM within the context of the choice of utility function. Work along these lines is being carried out by Professor Benth and co-workers at the University of Oslo. We think that this type of research is really important. Finally, Professor Hodges makes some interesting links with the implied process models which have recently been used in the finance literature. It is surely the case that we need stronger links to that approach.

Mark Davis discusses various areas where the option pricing theory based on our model could be used. He argues that these models have their largest potential in the value-at-risk type of calculations. This may be true, although we have yet to study these fields in any detail. However, his wise words are surely helpful in guiding us.

Howell Tong and Hailiang Yang emphasize the importance of the Esscher transformation for option pricing. This is a very convenient tool. However, from an economics viewpoint its choice seems somewhat arbitrary. As we mentioned above, theory based on utility functions would seem a rather sounder object. We hope that such methods will be developed for our models.

Other issues

Stephen Taylor asks about the intraday seasonal component of volatility. His points are, of course, correct and more sophisticated modelling would allow the various effects which he discussed to be taken into account. It is clear that Taylor and Xu (1997) is of importance in this field.

Frank Critchley asks several questions about the estimation of our models. In particular he desires a more formal cross-validation approach to breaking the data set into pieces. Our hope in carrying this out in a simple way was to see whether the model was reasonably stable over time. At the moment our main effort is to think about design effective estimation methods, while we hope that we shall be able to return to issues of outliers and inliers later.

Jens Ledet Jensen wonders whether the use of hidden Markov models may give a simple model structure for these types of problems. In some senses this is true; however, in terms of the properties of integrated volatility our models are quite simple compared with hidden Markov models. It is certainly the case that a slowly moving trend model of the type that he suggests may give a good description of this type of data; however, mean reversion in volatility is now a standard assumption following many years of rigorous empirical testing.

Chris Jones asks us why our volatility models are not of the type

$$\mathrm{d}x^*(t) = \{\mu + \beta\,\sigma(t)\}\,\mathrm{d}t + \sigma(t)\,\mathrm{d}w(t).$$

It is certainly the case that economic theory does not tell us that the risk premium (which relates the mean to the variance) should be of the form that we use, $\mu + \beta\,\sigma^2(t)$, rather than the form which he favours. Our choice was based on mathematical tractability and, more importantly, on the fact that our model structure can alternatively be viewed as being obtained by subordinating Brownian motion with drift by a generalized subordinator—integrated volatility.

It is a great pleasure that Professor Lawrance made a comment on our paper, as it gives us the opportunity to correct an oversight in not quoting his important research on autoregressive models with non-negative errors. This is clearly related to our continuous time work. Lawrance and Lewis (1985) is a good starting-point to read about this work.

Anthony Ledford discusses the extremal behaviour of returns for our SV models. This is an important topic, but it is clear that the tail index of returns y_n is immediately inherited from the tail index of $\sigma^2(t)$. This is one of the advantages of these types of models over discrete time ARCH-type models where these issues are much more involved.

References in the discussion

Andersen, T. G. (1994) Stochastic autoregressive volatility: a framework for volatility modeling. *Math. Finan.*, **4**, 75–102.
Andersen, T. G., Bollerslev, T., Diebold, F. X. and Ebens, H. (2001a) The distribution of stock return volatility. *J. Finan. Econ.*, to be published.
Andersen, T., Bollerslev, T., Diebold, F. X. and Labys, P. (2001b) The distribution of realized exchange rate volatility. *J. Am. Statist. Ass.*, **96**, 42–55.
Andersen, T. G., Chung, H.-J. and Sørensen, B. E. (1999) Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *J. Econometr.*, **91**, 61–87.
Andersson, J. (1999) Maximum likelihood estimation of the normal inverse Gaussian stochastic volatility. *Compr. Summ. Uppsala Dissertns Fac. Socl Sci.*, **82**, 1–11.
Anh, V. V., Heyde, C. C. and Leonenko, N. N. (2000) Dynamic models of long-memory processes driven by Lévy noise with applications to finance and macroeconomics. Submitted to *J. Appl. Probab.*
Anh, V. V. and Leonenko, N. N. (1999) Non-Gaussian scenarios for the heat equation with singular initial conditions. *Stoch. Process. Applic.*, **84**, 91–114.
Baillie, R. T. (1996) Long memory processes and fractional integration in econometrics. *J. Econometr.*, **73**, 5–59.
Barndorff-Nielsen, O. E. (1998) Processes of normal inverse Gaussian type. *Finan. Stochast.*, **2**, 41–68.
——— (2000) Superposition of Ornstein-Uhlenbeck type processes. *Theor. Probab. Applic.*, to be published.
Barndorff-Nielsen, O. E. and Levendorskii, S. (2000) Feller processes of normal inverse Gaussian type. *Research Report 45*. Centre for Mathematical Physics and Stochastics, University of Aarhus, Aarhus.
Barndorff-Nielsen, O. E., Pedersen, J. and Sato, K. (2000) Multivariate subordination, selfdecomposability and stability. *Adv. Appl. Probab.*, to be published.
Barndorff-Nielsen, O. E. and Shephard, N. (2000) Econometric analysis of realised volatility and its use in estimating Lévy based non-Gaussian OU type stochastic volatility models. *Discussion Paper*. Nuffield College, Oxford.
——— (2001a) *Lévy Based Dynamic Models for Financial Economics*. To be published.
——— (2001b) Integrated OU processes. Nuffield College, Oxford.
——— (2001c) Normal modified stable processes. Nuffield College, Oxford.
Benth, F. E., Karlsen, K. H. and Reikvam, K. (1999) Optimal portfolio selection with consumption and nonlinear integro-differential equations with gradient constraint: a viscosity solution approach. *Finan. Stoch.*, to be published.
Bingham, N. H. and Kiesel, R. (2001) Semi-parametric modelling in finance. *Preprint*. Brunel University, Uxbridge.
Bochner, S. (1949) Diffusion equation and stochastic processes. *Proc. Natn. Acad. Sci. USA*, **35**, 368–370.
Bollerslev, T. and Mikkelsen, H. O. (1996) Modeling and pricing long memory in stock market volatility. *J. Econometr.*, **73**, 151–184.
Bondesson, L. (1982) On simulation from infinitely divisible distributions. *Adv. Appl. Probab.*, **14**, 855–869.
——— (1992) *Generalized Gamma Convolutions and Related Classes of Distributions and Densities*. New York: Springer.
Britten-Jones, M. and Neuberger, A. (2000) Option prices, implied price processes and stochastic volatility. *J. Finan.*, **55**, 839–866.
Brockwell, P. J. (2001) Lévy driven CARMA processes. *Ann. Inst. Statist. Math.*, **52**, 1–18.
Bruce, A. G. and Martin, R. D. (1989) Leave-k-out diagnostics for time series (with discussion). *J. R. Statist. Soc.* B, **51**, 363–424.
Calzolari, G., Fiorentini, G. and Sentana, E. (2000) Constrained EMM and indirect inference estimation. *Working Paper 0005*. Centro de Estudios Monetarios y Financieros, Madrid.
Chamberlain, G. (1988) Asset pricing in multiperiod securities markets. *Econometrica*, **56**, 1283–1300.
Chambers, M. J. (1996) The estimation of continuous parameter long-memory time series models. *Econometr. Theory*, **12**, 374–390.
Chen, X., Hansen, L. P. and Scheinkman, J. (2000) Principal components and the long run. University of Chicago, Chicago.
Coles, S. G., Currie, J. E. and Tawn, J. A. (1999) Dependence measures for extreme value analyses. *Extremes*, **2**, 339–365.
Comte, F. and Renault, E. (1996) Long memory continuous time models. *J. Econometr.*, **73**, 101–149.
Cox, D. R. (1981) Statistical analysis of time series: some recent developments. *Scand. J. Statist.*, **8**, 93–115.

Cox, J., Ingersoll, J. and Ross, S. (1995) An intertemporal general equilibrium model of asset prices. *Econometrica*, **53**, 363–384.
Davis, M. (2000) Mathematics of financial markets. In *Mathematics Unlimited* (eds B. Engquist and W. Schmid). Berlin: Springer.
Davison, A. C. and Smith, R. L. (1990) Models for exceedances over high thresholds (with discussion). *J. R. Statist. Soc.* B, **52**, 393–442.
Derman, E. and Kani, I. (1997) Stochastic implied trees. *Quantitative Strategies Research Notes*, Apr. Goldman Sachs, London.
Diebold, F. X. and Lopez, J. (1995) Modeling volatility dynamics. In *Macroeconometrics: Developments, Tensions, and Prospects* (ed. K. Hoover), pp. 427–472. Boston: Kluwer.
Ding, Z. and Granger, C. W. J. (1996) Modeling volatility persistence of speculative returns: a new approach. *J. Econometr.*, **73**, 185–215.
Duffie, D. and Singleton, K. J. (1993) Simulated moments estimation of Markov models of asset prices. *Econometrica*, **50**, 929–952.
Dunbar, N. (1999) *Inventing Money*. New York: Wiley.
Dupire, B. (1992) Arbitrage pricing with stochastic volatility. *Working Paper*. Société Generale, Paris.
Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling Extremal Events*. Berlin: Springer.
Ferguson, T. S. and Klass, M. J. (1972) A representation of independent increment processes without Gaussian processes. *Ann. Math. Statist.*, **43**, 1634–1643.
Fiorentini, G., Sentana, E. and Calzolari, G. (2000) The score of conditionally heteroskedastic dynamic regression models with Student t innovations, and an LM test for multivariate normality. *Working Paper 0007*. Centro de Estudios Monetarios y Financieros, Madrid.
Frühwirth-Schnatter, S. and Sögner, L. (2001) Bayesian estimation of stochastic volatility models based on OU processes with marginal gamma law.
Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1999) Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, **5**, 855–872.
——(2000) Stochastic Volatility models, hidden Markov Models and statistical applications. *Bernoulli*, to be published.
Gerber, H. U. and Shiu, E. S. W. (1994) Option pricing by Esscher transforms. *Trans. Soc. Act.*, **46**, 99–191.
Golyandina, N., Nekrutkin, V. and Zhigljavsky, A. (2000) *Analysis of Structure of Time Series: SSA and Related Methods*. London: Chapman and Hall.
Guo, D. (1998) The risk premium of volatility implicit in currency options. *J. Bus. Econ. Statist.*, **16**, 498–507.
Halgreen, C. (1979) Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Z. Wahrsch Ver. Geb.*, **47**, 13–17.
Hamilton, J. D. (1988) Rational-expectations econometric analysis of changes in regime: an investigation of the term structure of interest rates. *J. Econ. Dynam. Control*, **12**, 385–423.
Hansen, L. P. (1985) A method for calculating bounds in the asymptotic covariance matrix of generalized method of moments estimators. *J. Econometr.*, **30**, 203–238.
Hansen, L. P. and Singleton, K. J. (1996) Efficient estimation of linear asset-pricing models with moving average errors. *J. Bus. Econ. Statist.*, **14**, 53–68.
Hartvig, N. V., Jensen, J. L. and Pedersen, J. (2001) A class of risk neutral densities with heavy tails. *Finan. Stoch.*, **5**, in the press.
Heath, D., Jarrow, R. and Morton, A. (1992) Bond pricing and the term structure of interest rates. *Econometrica*, **60**, 77–105.
Heston, S. L. (1993) A closed-form solution for options with stochastic volatility, with applications to bond and currency options. *Rev. Finan. Stud.*, **6**, 327–343.
Hindy, A. and Huang, C. (1993) Optimal consumption and portfolio rules with durability and local substitution. *Econometrica*, **61**, 85–121.
Hodges, S. D. and Neuberger, A. (1989) Optimal replication of contingent claims under transaction costs. *Rev. Fut. Mrkts*, **8**, 222–239.
Hodges, S. D. and Tompkins, R. (2000) The sampling properties of volatility cones. *FORC Preprint 00/103*. University of Warwick, Coventry.
Hodgson, D. O., Linton, O. and Vorkink, K. (2001) Testing the capital asset pricing model efficiently under elliptical symmetry: a semi-parametric approach. *Preprint*. London School of Economics and Political Science.
Hull, J. and White, A. (1987) The pricing of options on assets with stochastic volatilities. *J. Finan.*, **42**, 281–300.
Igloi, E. and Terdik, G. (1999) Long-range dependence through Gamma-mixed Ornstein-Uhlenbeck processes. *Electron. J. Probab.*, **4**, 1–33.
Inoue, A. (1993) On the equations of stationary processes with divergent diffusion coefficients. *J. Fac. Sci. Univ. Tokyo*, sect IA, **40**, 307–336.
Ismail, M. E. H. and Kelker, D. H. (1979) Special functions, Stieltjes transforms and infinite divisibility. *SIAM J. Math. Anal.*, **10**, 884–901.
Jacquier, E., Polson, N. G. and Rossi, P. E. (1994) Bayesian analysis of stochastic volatility models (with discussion). *J. Bus. Econ. Statist.*, **12**, 371–417.

Jensen, J. L. and Pedersen, J. (1997) A note on models for stock prices. *Research Report 372*. Department of Theoretical Statistics, University of Aarhus, Aarhus.
Jones, D. A. (1978) Nonlinear autoregressive processes. *Proc. R. Soc. Lond.* A, **360**, 71–95.
Kendall, M. G. (1953) The analysis of economic time-series — Part 1: prices. *J. R. Statist. Soc.* A, **116**, 11–25.
Kessler, M. (2000) Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.*, **27**, 65–82.
Lawler, G. F. and Sokal, A. (1988) Bounds on the L_2 spectrum for Markov chains and Markov processes: a generalization of Cheeger's inequality. *Trans. Am. Math. Soc.*, **309**, 557–580.
Lawrance, A. J. and Lewis, P. A. W. (1985) Modelling and residual analysis of nonlinear autoregressive time series in exponential variables (with discussion). *J. R. Statist. Soc.* B, **47**, 165–202.
Leadbetter, M. R. (1983) Extremes and local dependence in stationary sequences. *Z. Wahrsch. Ver. Geb.*, **65**, 291–306.
Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983) *Extremes and Related Properties of Random Sequences and Series*. New York: Springer.
Ledford, A. W. and Tawn, J. A. (1997) Modelling dependence within joint tail regions. *J. R. Statist. Soc.* B, **59**, 475–499.
León, A. and Sentana, E. (1997) Pricing options on assets with predictable white noise returns. *Working Paper 9704*. Centro de Estudios Monetarios y Financieros, Madrid.
Leonenko, N. (1999) *Limit Theorems for Random Fields with Singular Spectrum*. Dordrecht: Kluwer.
Leonenko, N. N., Sikorskii, A. Yu. and Terdik, G. (1998) On spectral and bispectral estimator of the parameter of nongaussian data. *Rand. Oper. Stoch. Eq.*, **6**, no. 2, 159–182; errata, **7**, no. 1, 200.
Lévy, P. (1937) *Théorie de l'Addition des Variables Aléatoires*. Paris: Gauthier-Villars.
Lo, A. W. and Wang, J. (1995) Implementing option pricing models when asset returns are predictable. *J. Finan.*, **50**, 87–129.
Loève, M. (1977) *Probability Theory*, 4th edn, vol. 1. New York: Springer.
——— (1978) *Probability Theory*, 4th edn, vol. 2. New York: Springer.
Madan, D. B., Carr, P. P. and Chang, E. C. (1998) The variance gamma process and option pricing. *Eur. Finan. Rev.*, **2**, 79–105.
Mandelbrot, B. B. (1997) *Fractals and Scaling in Finance: Discontinuity, Concentration, Risk*. New York: Springer.
Meddahi, N. and Renault, E. (1996) Aggregation and marginalization of GARCH and stochastic volatility models. Centre de Recherche en Economie et Statistique, Paris.
——— (2000a) Conditioning information in volatility models. To be published.
——— (2000b) Temporal aggregation of volatility models. *Report CIRANO DP 2000s-22*.
Melenberg, B. and Werker, B. J. M. (1999) A convenient way to characterize equivalent martingale measures in incomplete markets. *Statist. Inf. Stoch. Process.*, **2**, 11–30.
Meng, X.-L. and van Dyk, D. (1997) The EM algorithm — an old folk-song sung to a fast new tune (with discussion). *J. R. Statist. Soc.* B, **59**, 511–567.
Merton, R. (1971) Optimum consumption and portfolio rules in a continuous-time model. *J. Econ. Theory*, **3**, 373–413.
——— (1973) An intertemporal asset pricing model. *Econometrica*, **41**, 867–887.
Nicolato, E. and Venardos, E. (2000) Derivative pricing in Barndorff-Nielsen and Shephard's OU type stochastic volatility models. *Working Paper*. University of Aarhus, Aarhus.
Nijman, T. and Sentana, E. (1996) Marginalization and contemporaneous aggregation in multivariate GARCH processes. *J. Econometr.*, **71**, 71–87.
Oppenheim, G. and Viano, M. C. (1999) Obtaining long-memory by aggregation random coefficients discrete and continuous time series short memory processes. *Publ. IRMA*, **49**, no. 5, 1–16.
Pitt, M. K. (2000) Smooth particle filters for likelihood evaluation and maximisation. *Working Paper*. University of Warwick, Coventry.
Pitt, M. K. and Shephard, N. S. (1999) Filtering via simulation based on auxiliary particle filters. *J. Am. Statist. Ass.*, **94**, 590–599.
Roberts, G. O. and Sahu, S. K. (1997) Updating schemes, correlation structure, blocking and parameterization for the Gibbs sampler. *J. R. Statist. Soc.* B, **59**, 291–317.
Roll, R. (1984) A simple implicit measure of the effective bid-ask spread in an efficient market. *J. Finan.*, **39**, 1127–1139.
Rosenblatt, M. (1976) Fractional integral of stationary processes and the central limit theorem. *J. Appl. Probab.*, **13**, 723–732.
Rosiński, J. (2000) Stochastic series representations of inverse Gaussian and other exponentially tempered stable processes. *Research Report 42*. Centre for Mathematical Physics and Stochastics, University of Aarhus, Aarhus.
Rubinstein, M. (1994) Implied binomial trees. *J. Finan.*, **49**, 771–818.
Ryden, T., Terasvirta, T. and Asbrink, S. (1998) Stylized facts of daily return series and the hidden Markov model. *J. Appl. Econometr.*, **13**, 217–244.
Sato, K. (1999) *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press.

——— (2000) Subordination and selfdecomposability. *Research Report*. Centre for Mathematical Physics and Stochastics, University of Aarhus, Aarhus.
Sato, K. and Yamazato, M. (1978) On distribution functions of class *L*. *Z. Wahrsch. Ver. Geb.*, **43**, 273–308.
——— (1983) Stationary processes of Ornstein–Uhlenbeck type. *Lect. Notes Math.*, **1021**, 541–551.
——— (1984) Operator-selfdecomposable distributions as limit distributions of processes of Ornstein–Uhlenbeck type. *Stoch. Process. Applic.*, **17**, 73–100.
Sentana, E. and Fiorentini, G. (2000) Identification of conditionally heteroskedastic factor models. *J. Econometr.*, to be published.
Siu, T. K., Tong, H. and Yang, H. (2001) Bayesian risk measures for derivatives via random Esscher transform. *Research Report 286*. Department of Statistics and Actuarial Science, University of Hong Kong, Hong Kong.
Skiadopoulos, G. and Hodges, S. D. (2000) Simulating the evolution of the implied distribution. *FORC Preprint 00/110*. University of Warwick, Coventry.
Stute, W. (2000) Jump diffusion processes with shot noise effects and their applications to finance. In *Statistical Modelling: Proc. 15th Int. Wrkshp Statistical Modelling, University of the Basque Country* (eds V. Nunez-Anton and E. Ferreira), pp. 86–94. Bilbao: Universidad del Pais Vasco.
Taqqu, M. S. (1979) Convergence of integrated processes of arbitrary Hermite rank. *Z. Wahrsch. Ver. Geb.*, **50**, 53–83.
Taylor, S. J. and Xu, X. (1997) The incremental volatility information in one million foreign exchange quotations. *J. Emp. Finan.*, **4**, 317–340.
Thorin, O. (1977) On the infinite divisibility of the lognormal distribution. *Scand. Act. J.*, **47**, 121–148.
Tompkins, R. and Hubalek, F. (2000a) Does it matter where jumps comes from?: underlying versus volatility processes. *Working Paper*. Vienna University of Technology, Vienna.
——— (2000b) On closed form solutions for pricing options with jumping volatility. Vienna University of Technology, Vienna.
Tong, H. (1990) *Nonlinear Time Series*. Oxford: Oxford University Press.
Walker, S. G. (2000) A note on the innovation distribution of a gamma distributed autoregressive process. *Scand. J. Statist.*, **27**, 575–576.
Wolfe, S. J. (1982) On a continuous analogue of the stochastic difference equation $x_n = \rho x_{n-1} + b_n$. *Stoch. Processes Applic.*, **12**, 301–312.
Woyczynski, W. A. (1998) Burgers-KPZ turbulence. *Lect. Notes Math.*, **1700**.
Yamazato, M. (1978) Unimodality of infinitely divisible distribution functions of class *L*. *Ann. Probab.*, **6**, 523–531.