Non-Gaussian OU based models and some of their uses in financial economics

OLE E. BARNDORFF-NIELSEN
Centre for Mathematical Physics and Stochastics (MaPhySto), University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark.
oebn@imf.au.dk

NEIL SHEPHARD
Nuffield College, Oxford OX1 1NF, UK
neil.shephard@nuf.ox.ac.uk

Abstract

Non-Gaussian processes of Ornstein-Uhlenbeck type, or *OU processes* for short, 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; OU process; Particle filter; Stochastic volatility; Subordination; Superposition.

Authors' note: This paper supersedes our previously circulated but unpublished papers "Aggregation and model construction for volatility models" and "Incorporation of a leverage effect in a stochastic volatility model".

1 Introduction

1.1 Motivation

Non-Gaussian processes of Ornstein-Uhlenbeck type, or *OU processes* as we shall call them, 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. Based on well-known (empirical) stylized facts, 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. The study has also required the development of new numerical methods and these are discussed in some detail.

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

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

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, pp. 73–84), Bertoin (1996), Bertoin (1999), Protter and Talay (1999) and Sato (1999)). Familiar special cases of Lévy processes are Brownian motion and the compound Poisson process. Lévy's theorem tells us that all Lévy processes except for Brownian motion have jumps. As z is used to drive the OU process we will 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 (1) in cases where z has no Gaussian component and the increments of z are positive, implying positivity of the process y. We will write a continuous time stationary and nonnegative 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

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

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 parameterise the distribution of the volatility and the dynamic structure. The $z(t)$ has positive increments and no drift. This type of process is often called a subordinator (Bertoin (1996, Ch. 3)). Correspondingly $\sigma^2(t)$ moves up entirely by jumps and then tails off exponentially[1]. However, under the models we have in mind small jumps are predominant. Although having OU dynamics looks restrictive, we will show we can construct more complicated processes by the addition of independent OU processes.

The main advantage of these OU processes is that they offer a great deal of analytic tractability which is not available for more standard models such as geometric Gaussian OU processes and constant elasticity of volatility processes[2]. For example integrated[3] volatility, which in finance is a key measure,

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

has a simple structure.

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[4]. 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]This type of model has been used in storage theory by, for example, Cinlar and Pinsky (1972), Harrison and Resnick (1976) and Brockwell, Resnick, and Tweedie (1982).

[2]For geometric Gaussian OU processes, $\log\sigma^2(t)$ is assumed to follow a Gaussian OU process. For constant elasticity of volatility processes

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

where $b(t)$ is standard Brownian motion, $k \geq 1/2$. The former is highlighted by Hull and White (1987) while the latter is used extensively by Meddahi and Renault (1996).

[3]All integrated processes will be denoted by having a superscript $*$. The main examples are integrated volatility and intensity and the log-price level of a stock.

[4]To be technically precise: $\{z(t)\}_{t\geq 0}$ is assumed to be caglad and $\{z(-t)\}_{t\geq 0}$ is an independent copy of $\{-z(t)\}_{t\geq 0}$ but modified to be also caglad. Further, $f(\cdot)$ has to be a positive function tailing off sufficiently fast to ensure the existence of the integral. In particular if $f(s) = e^s$ we recover the OU processes.

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 log of an asset price by the solution to the stochastic differential equation

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

where $w(t)$ is standard Brownian motion[5]. This means aggregate returns over intervals of length $\Delta > 0$, are

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

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 one 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, Lo, and MacKinlay (1997, pp. 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 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 generalised autoregressive conditional heteroskedastic (GARCH) and the diffusion based stochastic volatility (SV) processes. This very large literature, which was started by Clark (1973), Engle (1982) and Taylor (1982), is reviewed in, for example, Bollerslev, Engle, and Nelson (1994), Ghysels, Harvey, and Renault (1996) and Shephard (1996).

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

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

where $\sigma^2(t)$, the instantaneous volatility, is going to 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 will delay our discussion of this well known connection until Section 6 of this paper. 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.

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, exhibit volatility clustering and aggregate to Gaussianity as Δ gets 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

$$\sigma_n^2 = \sigma^{2*}(n\Delta) - \sigma^{2*}\left\{(n-1)\Delta\right\}, \quad \text{and} \quad \sigma^{2*}(t) = \int_0^t \sigma^2(u)\mathrm{d}u. \tag{7}$$

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, allowing us to think about the appropriate implications for the pricing of derivatives written on underlying assets obeying SV processes. We will do this in Section 5 and Subsection 6.2 of the paper.

[5] We have used $x^*(t)$ to denote the price level as this is an integrated process.

It is possible to generalise (6) to allow for the feedback of the innovations of the volatility process into the level of the asset price. In particular, we write

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

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

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 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 separately deal 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 this section on their first reading of the paper. 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 using linear and non-linear methods. Section 6 discusses various additional issues such as multivariate extensions of the models, 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 a SV model. Section 7 concludes. The Appendix collects various proofs and derivations we have omitted from the main text of the paper.

2 Construction of OU processes

2.1 Definition and existence

Before we discuss the SV models in detail we will 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 Ornstein-Uhlenbeck type if it is representable as

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

in which case it may also be written as

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

Here $z = \{z(t) : t \in R\}$ 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 *background driving Lévy process* (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 follows is given in Figure 1.

In essence, given a one-dimensional distribution D (not necessarily restricted to the positive halfline) there exists a stationary process of Ornstein-Uhlenbeck type (i.e. satisfying a stochastic differential equation of form (1)) whose one-dimensional marginal law is D if and only if D is *selfdecomposable*,

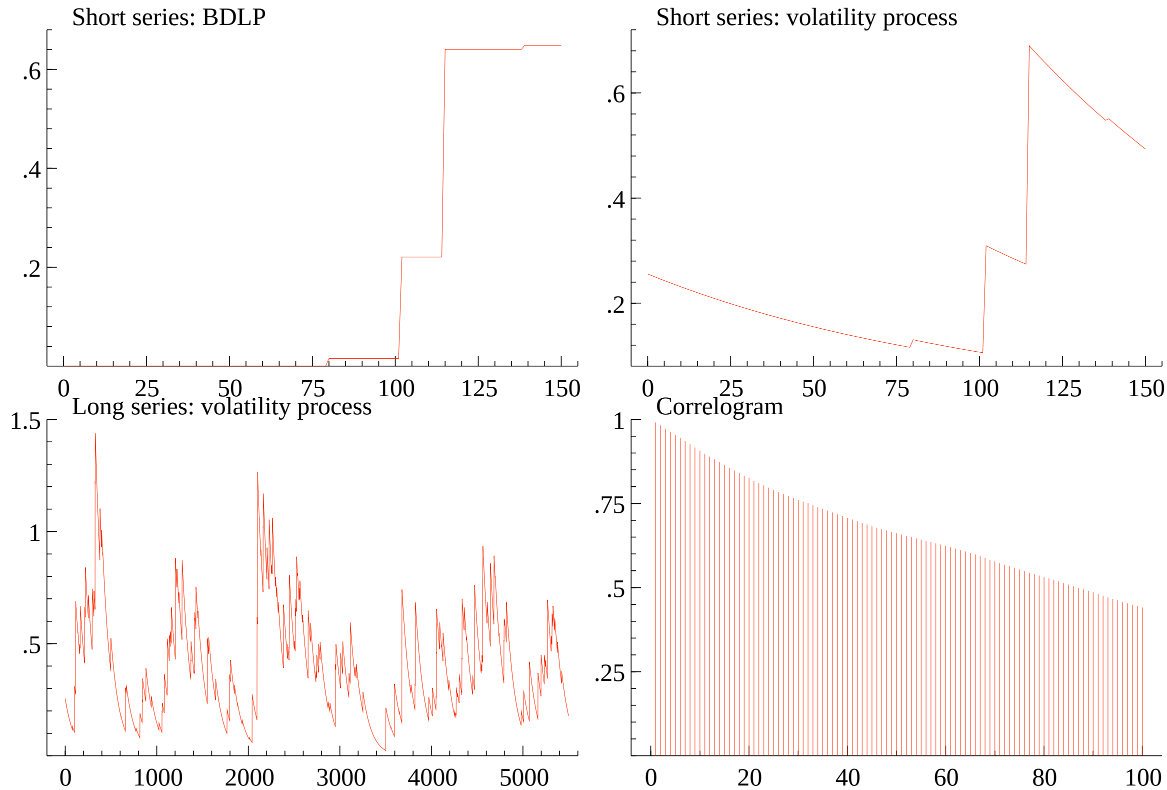

Figure 1: OU process with $\Gamma(\nu,\alpha)$ marginals. *Throughout,* $\nu = 3$, $\alpha = 8.5$, $\lambda = 0.01$ *and* $\Delta = 1$. *Top left: plot of* $z(\lambda n\Delta)$ *against* n. *Top right: plot of* $\sigma^2(n\Delta)$ *against* n. *Same graph but for longer series in bottom left. Bottom right: as a numerical check we also present the empirical autocorrelation function for* $\sigma^2(n\Delta)$.

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 $\{\phi_c : c \in (0,1)\}$. 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, Jensen, and Sørensen (1998)).

Theorem 2.1 Let ϕ be the characteristic function of a random variable x. If x is selfdecomposable, 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 exists a stationary stochastic process $x(t)$ and a Lévy process $z(t)$ such that $x(t) \overset{\mathcal{L}}{=} x$ and

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

for all $\lambda > 0$.

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

If the stationary OU process $\sigma^2(t)$ is square integrable, it has autocorrelation function $r(u) = \exp\left(-\lambda\left|u\right|\right)$. 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

$$\acute{k}(\theta) = \log \mathrm{E}\left[\exp\left\{-\theta\sigma^2(t)\right\}\right] \qquad \text{and} \qquad k(\theta) = \log \mathrm{E}\left[\exp\left\{-\theta z(1)\right\}\right],$$

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

$$\acute{k}(\theta) = \int_{0}^{\infty} k(\theta e^{-s})\mathrm{d}s, \tag{11}$$

which can be reexpressed 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 $(m = 1, 2, ...)$ we have that $\kappa_m = m\acute{\kappa}_m$, for $m = 1, 2, ...$.

2.2 Lévy densities

Suppose we choose a probability distribution D on the positive halfline which is self-decomposable. Then, as just discussed, there exists a strictly stationary Ornstein-Uhlenbeck process

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

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

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

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

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

(this presupposes that u is differentiable) and, letting

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

we have, moreover

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

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

$$W^{-1}(x) = \inf\left\{y > 0 : W^+(y) \leq x\right\}.$$

2.3 Models via D

One approach to model building is to write down a specific parametric form for D and then calculate the implied behaviour of the BDLP. We do this here for the generalized inverse Gaussian (GIG) marginal law $\sigma^2(t) \sim GIG(\nu, \delta, \gamma)$[6]. 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) in various recent papers. See, in particular, Eberlein and Keller (1995), Barndorff-Nielsen (1997), Barndorff-Nielsen (1998), Rydberg (1999) and Eberlein (2000). Recall that if $x \sim GIG(\nu, \delta, \gamma)$ then it has a density

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

where K_ν is a modified Bessel function of the third kind. Note that when δ or γ are 0, the norming constant in the formula for the density of the generalized inverse Gaussian distribution has to be interpreted in the limiting sense, using the well-known results that for $x \downarrow 0$ we have

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

[6]The standard notation for the generalised inverse Gaussian distribution is $GIG(\lambda, \delta, \gamma)$, however the notation λ was not available to us.

Special cases of the GIG density are: (i) the inverse Gaussian law, where $\nu = -\frac{1}{2}$, (ii) the positive hyperbolic law where $\nu = 1$, (iii) and inverse chi-squared law with df degrees of freedom where $\nu = -df/2$, $\delta = \sqrt{df}$ and $\gamma = 0$, (iv) gamma, where $\delta = 0$ and $\nu > 0$. Of course if $\sigma^2 \sim GIG(\nu, \delta, \gamma)$ and is independent of $\varepsilon \sim N(0,1)$, then $x = \mu + \beta\sigma^2 + \sigma\varepsilon$ is the generalized hyperbolic distribution. If we define $\alpha = \sqrt{\beta^2 + \gamma^2}$, then the density is

$$\frac{(\gamma/\delta)^{\nu}}{\sqrt{2\pi}\alpha^{\left(\nu-\frac{1}{2}\right)}K_{\nu}\left(\delta\gamma\right)}\left\{\delta^2 + (x-\mu)^2\right\}^{\frac{1}{2}\left(\nu-\frac{1}{2}\right)}K_{\left(\nu-\frac{1}{2}\right)}\left(\alpha\sqrt{\delta^2 + (x-\mu)^2}\right)\exp\left\{\beta\left(x-\mu\right)\right\}. \quad (19)$$

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 normal inverse Gaussian distribution, the hyperbolic and the Student t.

It is known that the $GIG(\nu, \delta, \gamma)$ law is self-decomposable (Halgreen (1979)) so that stationary OU processes with GIG marginals do exist. The following theorem specifies the Lévy measure.

Theorem 2.2 The Lévy measure of the generalized inverse Gaussian distribution is absolutely continuous with density

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

where

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

and J_{ν} and N_{ν} are Bessel functions.

□

PROOF See Appendix.

For the definitions and properties of Bessel functions see, for example, Gradstheyn and Ryzhik (1965, pp. 958-71).

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

- $GIG(-\frac{1}{2}, \delta, \gamma)$: *Inverse Gaussian*. Its marginal law means $\sigma^2(t) \sim IG(\delta, \gamma)$ whose density is

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

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

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

- $GIG(1, \delta, \gamma)$: *Positive hyperbolic distribution*. The density of the positive hyperbolic distribution is

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

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

$$W^+(x) = \left\{\delta^2\int_0^{\infty} e^{-x\xi}g_1(2\delta^2\xi)\mathrm{d}\xi + \lambda\right\}\exp\left(-\gamma^2 x/2\right). \quad (23)$$

- $GIG(-\nu,\delta,0)$: *Reciprocal gamma distribution.* The reciprocal gamma distribution (i.e. the law of the reciprocal of a gamma variate) has density

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

  The corresponding upper tail integral is

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

- $GIG(\nu>0,0,\gamma)$: *Gamma distribution.* The gamma marginal law has probability

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

  This has the corresponding upper tail integral of the Lévy density $W^{+}(x)=\nu e^{-\alpha x}$, 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 BDLP

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 construct the model through the BDLP. Of course there are constraints on valid BDLPs which must be satisfied. We note in passing that 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 $\mathrm{E}\left[\log\left\{1+|z(1)|\right\}\right]<\infty$ (cf. Wolfe (1982) and Jurek and Mason (1993, Theorem 3.6.6)).

Lemma 2.1 Let z be a Lévy process with positive increments and cumulant function

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

and assume that

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

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

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

and the specification

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

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

PROOF This may be concluded from a more general result given in Jurek and Mason (1993, Theorem 3.6.6).

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*

$$W^+(x) = cx^{-\varepsilon}(1+x)^{-\beta}\exp\left(-\frac{1}{2}\gamma^2 x\right)$$

where c *is a positive constant,* $0 \leq \varepsilon < 1$, $0 \leq \beta$, $0 \leq \gamma$ *and* $\max\{(\beta-1), \gamma\} > 0$. *Then*

$$w(x) = c\{\varepsilon x^{-1} + \beta(1+x)^{-1} + \frac{1}{2}\gamma^2\}x^{-\varepsilon}(1+x)^{-\beta}\exp\left(-\frac{1}{2}\gamma^2 x\right). \tag{29}$$

Hence Lemma 2.1 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* $xu(x) = W^+(x)$, *we find*

$$u(x) = cx^{-1-\varepsilon}(1+x)^{-\beta}\exp\left(-\frac{1}{2}\gamma^2 x\right).$$

Note that for $\varepsilon = \frac{1}{2}$ *and* $\beta = 0$ *we recover the* IG *law for* $\sigma^2(t)$. *If* $\gamma = 0$, *implying* $\beta > 1$, *then for the moments of* $\sigma^2(t)$ *we have*

$$\mathrm{E}\left[\left\{\sigma^2(t)\right\}^\nu\right] < \infty \quad \textit{if and only if} \quad \nu < \beta + \varepsilon.$$

Furthermore, the j*-th order cumulant of* $\sigma^2(t)$ *(*$j < \beta+\varepsilon$*) is* $cB(j-\varepsilon, \beta+\varepsilon-j)$ *where* $B(x,y)$ *denotes the beta function.*

2.5 Simulation via series representations

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

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

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

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

rather than the BDLP $z(s)$ itself. One approach to this is to directly simulate from the Lévy processes (e.g. through Wolpert and Ickstadt (1998) and Protter and Talay (1999)) and then approximate the corresponding integrals. This is difficult due 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 Rosinski (1991). A self-contained exposition of this result is given in Barndorff-Nielsen and Shephard (2000), while recent developments are surveyed in Rosinski (2000). 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 results is that

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

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

Our practical experience with using (31) is that it is quite quickly converging, however theory suggests that it has to be used carefully. Consider the special case of the IG model, then (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).

Example 2 *OU gamma ($\Gamma(\upsilon,\alpha)$ marginals) process. We need a method to sample from (30). We have already noted the expression for $W^{-1}(x)$ in (25). Thus, defining $c_1 < c_2 < ...$ as the arrival times of a Poisson process with intensity $\nu\lambda t$ and $N(1)$ as the corresponding number of events up until time 1, then*

$$e^{-\lambda t}\int_0^{\lambda t} e^s \mathrm{d}z(s) \stackrel{\mathcal{L}}{=} e^{-\lambda\Delta}\sum_{i=1}^{\infty} W^{-1}(a_i/\lambda t)e^{\lambda t r_i} \tag{32}$$

$$= -\alpha^{-1}e^{-\lambda t}\sum_{i=1}^{\infty}\mathbf{1}_{]0,\nu[}(a_i/\lambda t)\log(a_i/\nu\lambda t)e^{\lambda t r_i}$$

$$= \alpha^{-1}e^{-\lambda t}\sum_{i=1}^{\infty}\mathbf{1}_{]0,1[}(c_i)\log(c_i^{-1})e^{\lambda t r_i}$$

$$= \alpha^{-1}e^{-\lambda t}\sum_{i=1}^{N(1)}\log(c_i^{-1})e^{\lambda t r_i}.$$

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 Figure 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, while the $\sigma^2(n\Delta)$ is stationary. For the larger series we see the jumps look less extreme and instead our eyes tend to focus on the large up movements in 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 were able to produce processes of length of half a million in around 5 seconds on a modern PC 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 Ornstein-Uhlenbeck processes with different persistence rates. That is

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

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

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

where $\bar{z}_j(\lambda_j t) = z_j(\lambda_j t) - \mathrm{E}\{z_j(\lambda_j t)\}$, 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

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

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 while others represent long term movements. Alternative empirical models of this, written directly in discrete time, are discussed by Engle and Lee (1999), Dacorogna, Muller, Olsen, and Pictet (1998) and Barndorff-Nielsen (1998).

By choosing the weights and damping factors in (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 there exists 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[7]. 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 while 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 log of the instantaneous volatility as fractional Brownian motion.

It is possible to extend this to multifractal behaviour where

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

and where the w_i are positive and sum to one. 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 will 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 will look at the log-price itself $x^*(t)$, recalling that $x^*(0)$ is defined to be zero. The second focus, developed in the next subsection, will be on characterising the dependence structure of the returns $\{y_n\}$, defined in (5) as the change in $x^*(t)$ over non-overlapping intervals of length Δ.

First we will state some general results for the non-leverage SV models given in (6) with arbitrary stationary volatility processes, then we will 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

$$\mathrm{E}\left\{\sigma^{2*}(t)\right\} = \xi t, \quad \mathrm{Var}\{\sigma^{2*}(t)\} = 2\omega^2 r^{**}(t),$$

where[8]

$$r^*(t) = \int_0^t r(u)\mathrm{d}u \quad \text{and} \quad r^{**}(t) = \int_0^t r^*(u)\mathrm{d}u. \tag{34}$$

A consequence of the above result is that

$$\mathrm{E}\left\{x^*(t)\right\} = (\mu + \beta\xi)\, t \qquad \text{and} \qquad \mathrm{Var}\{x^*(t)\} = t\xi + 2\beta^2\omega^2 r^{**}(t),$$

[7] 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.

[8] We use $r^{**}(t)$ to denote the double integral over the autocorrelation function.

while, 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 $t \to \infty$,

$$t^{-1}\sigma^{2*}(t) = t^{-1}\int_0^t \sigma^2(u)\mathrm{d}u \overset{a.s.}{\to} \xi,$$

implying, for the SV model, that $t^{-1/2}\left\{x^*(t) - \mu t - \beta\sigma^{2*}(t)\right\}$ is asymptotically normal with mean 0 and variance ξ (i.e. the log returns tend to normality for long lags — a similar result is known within the ARCH class since Diebold (1988, pp. 12-16)). This follows from the subordination interpretation of the SV models discussed in Section 6.1. The convergence of $t^{-1/2}\left\{x^*(t) - \mu t - \beta\sigma^{2*}(t)\right\}$ 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 is $\sigma^{*2}(t)$, i.e. we have

$$[x^*](t) = \mathrm{p}-\lim_{r\to\infty} \sum \{x^*(t^r_{i+1}) - x^*(t^r_i)\}^2 = \sigma^{2*}(t) \tag{35}$$

for any sequence of partitions $t^r_0 = 0 < t^r_1 < ... < t^r_{m_r} = t$ with $\sup_i\{t^r_{i+1} - t^r_i\} \to 0$ for $r \to \infty$. The quadratic variation estimation of integrated volatility has recently been highlighted, following the initial draft of this paper and the concurrent independent work of Andersen and Bollerslev (1998), by Andersen, Bollerslev, Diebold, and Labys (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. Note, however, that it can easily be recast in terms of the cumulant function $\acute{k}$ for $\sigma^2(t)$, cf. formulae (11) and (12). Let f denote an 'arbitrary' function then the log of the characteristic function of $f \bullet x^*$, which we interpret as the stochastic integral $\int_0^\infty f(s)\mathrm{d}x^*(s)$ (Protter (1992)), is

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

where

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

and

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

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 (36) by 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 $\mu = \beta = \rho = 0$; extension to the general case causes no substantial difficulty. Letting $f = \mathbf{1}_{[0,t]}$ we find, after a bit of algebra,

$$\begin{aligned}\mathrm{C}\{\zeta \ddagger x^*(t)\} &= \lambda \int_0^t k\left\{\frac{1}{2}\zeta^2\lambda^{-1}(1-e^{-\lambda t})e^{-\lambda s}\right\}\mathrm{d}s \\ &\quad + \lambda\int_t^\infty k\left\{\frac{1}{2}\zeta^2\lambda^{-1}(1-e^{-\lambda t})e^{-\lambda s}\right\}\mathrm{d}s.\end{aligned}$$

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

Example 3 *Suppose $\sigma^2(t) \sim IG(\delta, \gamma)$, as in (21), then $\acute{k}(\theta) = \delta\gamma\left\{1 - (1 + 2\theta/\gamma^2)^{1/2}\right\}$ and so, by formula (12),*

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

where

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

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 zero then, using the definitions given in (34),

$$\begin{aligned}\mathrm{Cov}\{\sigma_n^2, \sigma_{n+s}^2\} &= \omega^2 \diamondsuit r^{**}(\Delta s), &(39)\\ \mathrm{cor}\{y_n^2, y_{n+s}^2\} &= \frac{\diamondsuit r^{**}(\Delta s)}{6r^{**}(\Delta) + 2\Delta^2(\xi/\omega)^2} &(40)\\ &= q^{-1}\Delta^{-2}\diamondsuit r^{**}(\Delta s), &(41)\end{aligned}$$

where

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

and

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

Example 4 *If $\sigma^2(t) \sim OU$ with its variance existing then $r(u) = \exp(-\lambda|u|)$, which means that $r^{**}(s) = \lambda^{-2}\left\{e^{-\lambda|s|} - 1 + \lambda s\right\}$ and*

$$\diamondsuit r^{**}(\Delta s) = \lambda^{-2}(1 - e^{-\lambda\Delta})^2 e^{-\lambda\Delta(s-1)}, \quad s > 0.$$

This implies

$$\mathrm{cor}\{\sigma_n^2, \sigma_{n+s}^2\} = de^{-\lambda\Delta(s-1)}, \quad \mathrm{cor}\{y_n^2, y_{n+s}^2\} = ce^{-\lambda\Delta(s-1)}, \quad s > 0 \tag{44}$$

where

$$\begin{aligned} 1 \quad \geq \quad & d = \frac{(1 - e^{-\lambda\Delta})^2}{2\left\{e^{-\lambda\Delta} - 1 + \lambda\Delta\right\}} &(45)\\ \geq \quad & c = \frac{(1 - e^{-\lambda\Delta})^2}{6\left\{e^{-\lambda\Delta} - 1 + \lambda\Delta\right\} + 2(\lambda\Delta)^2(\xi/\omega)^2} \geq 0. \end{aligned}$$

Note that (44) implies 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 emphasised in the work of Meddahi and Renault (1996). Andersen and Bollerslev (1997b, p. 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 generalise to the situation where we add together a weighted sum of uncorrelated Ornstein-Uhlenbeck 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 with no moving average component.

More abstractly, Sørensen (1999) and Genon-Catalot, Jeantheau, and Larédo (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 specialised. When $\sigma^2(t) \sim OU$ we are able to produce very concrete results. In particular

$$\mathrm{E}\{y_n y_{n+s}\} = 0,$$

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

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

The effect of the leverage term is to allow $\mathrm{Cov}\left(y_n y_{n+s}^2\right)$ to be negative if $\rho < 0$. However, in addition both $\mathrm{Cov}\left(y_n y_{n+s}^2\right)$ and $\mathrm{Cov}(y_n^2, y_{n+s}^2)$ damp down exponentially with the lag length s. We should note that exactly the same dynamic structure was found by Sentana (1995) in his work on the discrete time quadratic ARCH model (QARCH). Hence we can think of the QARCH model as a kind of discrete time representation of our continuous time leverage model, generalising 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 will study five minute[9] return series (recorded using Greenwich Mean Time) for the DM/$ exchange rate from 1/12/86 to 30/11/96 constructed from the Olsen and Associates database using the semi-cleaning procedures carefully documented in Andersen, Bollerslev, Diebold, and Labys (2000). It should be noted that 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, Ito, and Payne (1996) and Danielsson and Payne (1999) suggests the properties of transaction and quote data, at this frequency, closely match.

The semi-cleaned Andersen, Bollerslev, Diebold, and Labys (2000) data does not remove some heavy intra-day effects in the volatility of the series. As a result we imposed some adjustments ourselves. These included taking out all data from 10.30pm Friday until Sunday 11pm each week, as well as bank holidays. In addition we have estimated a strong intra-day volatility effect (see Guillaume, Dacorogna, Dave, Muller, Olsen, and Pictet (1997) for a discussion of this) by running a cubic spline (with 40 degrees of freedom) on the variance of each five minute period in active day. After some initial analysis we have set the intra-day 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 intra-day seasonal component is given in Figure 2. The most interesting features of this graph is the high volatility of the series on Monday mornings, 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 684,867 five minute observations. For each observation we standardise it by dividing through by its intra-day effect in an attempt to achieve a homogeneous series. We then study the marginal distribution

[9] It is difficult to go below 5 minute returns without suffering from problems of discreteness which we will briefly discuss in Section 6. Recent econometric papers on this topic include Russell and Engle (1998) and Rydberg and Shephard (1998).

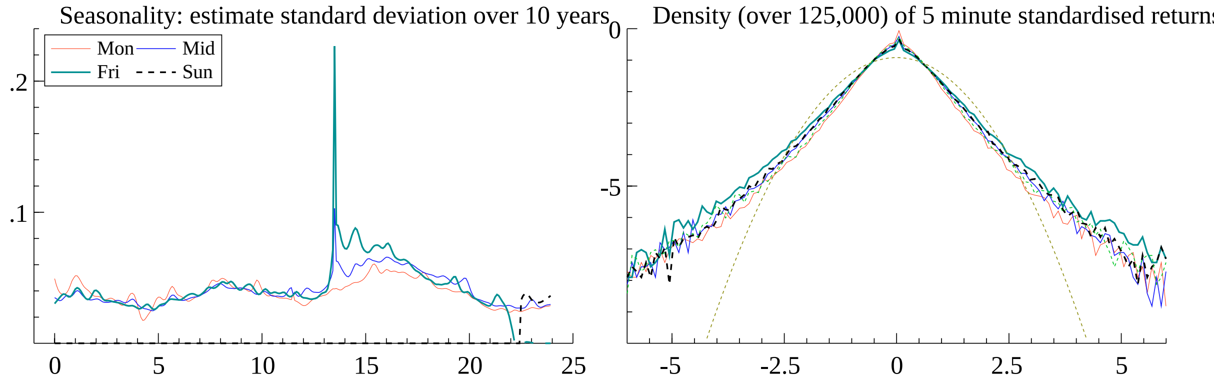

Figure 2: *Left: Estimated intra-day pattern of volatility (standard deviations) for each day (in particular Monday, average over Tuesday through Thursday, Friday and Sunday) over 5 minute periods using 10 years of data. X-axis denotes hours. Right: marginal log-density of returns over 5 minute period — data split into series of length 125,000. Dotted line is corresponding fitted normal log-density.*

of the resulting standardised series. Figure 2 gives the log of the histogram of returns where we split the returns into four sections of 125,000 observations (that is each section is just over two years of adjusted five 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[10], suggesting we need models for extreme marginal returns over short intervals of the form

$$const. \left|y\right|^{\rho_\pm} \exp(-\sigma_\pm \left|y\right|)$$

for some $\rho_+, \rho_- \in \mathbf{R}$ and $\sigma_+, \sigma_- \geq 0$. One class of densities which has this property are the normal inverse Gaussians.

5.2 Estimating marginal distribution

Although the basic dataset we use takes Δ as representing five minutes, we can think about returns at other frequencies. In Figure 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, that is normal densities. The Figure also shows the fitted log-densities of normal inverse Gaussian and Student t type, where the parameters of the fit are chosen by maximising the corresponding likelihood assuming the returns are i.i.d.. We thus interpret these fits as of quasi-likelihood type.

Table 1 records the quasi-likelihood fits for each of the models[11], once again showing that the normal distribution is dominated by the other candidates. Further for small values of Δ the normal inverse Gaussian out-performs the Student t even though it is clear that the Student t has heavier tails. For larger values of Δ the fit is basically identical. The convergence towards normality as Δ increases is also shown in the Table where we compute the average Kullback-Liebler distance (per observation) between the normal density and the other two candidates we study here.

[10] Granger and Ding (1995) model $|y_n|$ as having a marginal distribution which is exponential.

[11] 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 zero, although in theory we should allow it to depend upon the difference in interest rates between the two countries. However, in practice the drift in neglible is this case.

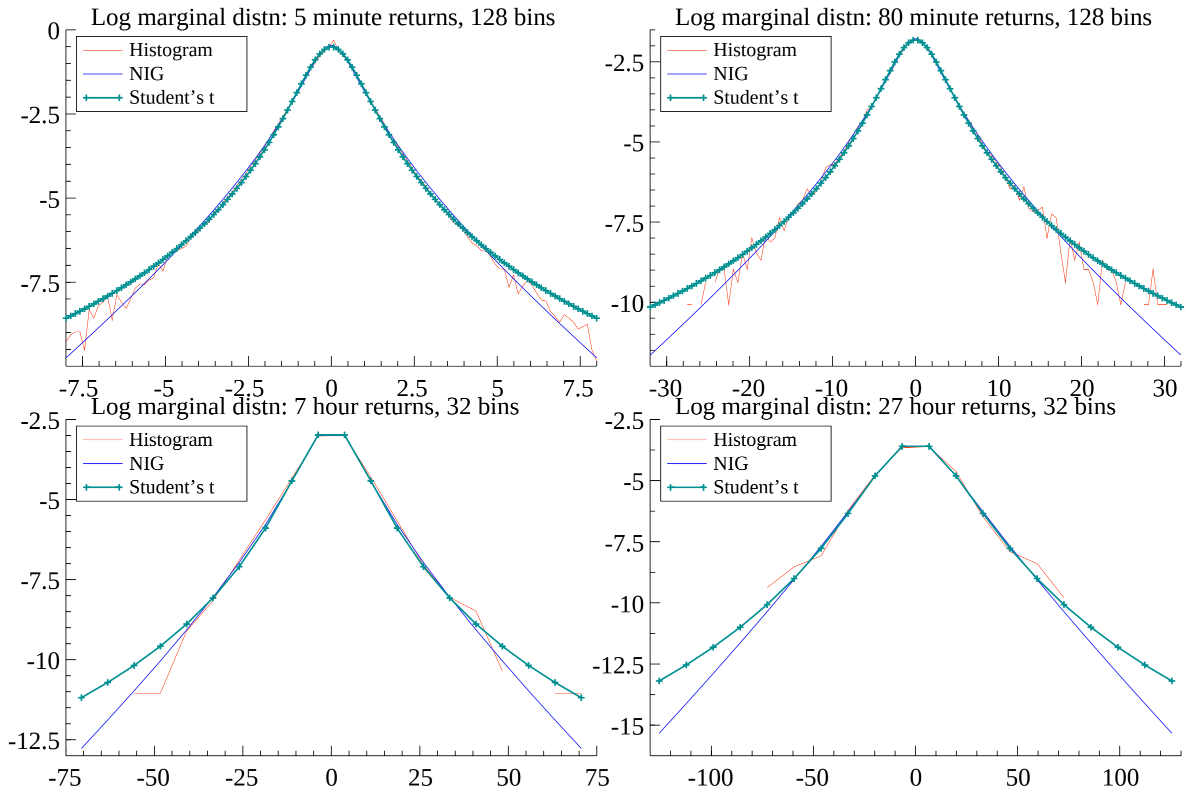

Figure 3: *Log-densities of returns at different levels of temporal aggregation. Plotted are histograms, estimated (by quasi-ML) NIG and Student's t distributions. Top left: 5 minute returns. Top right: 70 minute returns. Bottom left: 7 hours. Bottom right: 27 hours. Top graphs histograms computed using 128 bins, bottom graphs have only 32.*

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 Figure 4 the average correlogram which results. Note the x-axis of the correlogram is marked out in days, not in 5 minute periods. The left hand graph focuses on the short term dynamics and shows a fast initial decay which then levels out. The middle graph, 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.

The right hand side graph of Figure 4 is more unusual. Each day has 288 observations of 5 minute adjusted returns. We have computed the empirical variation within each day

$$s_{n,288}^2 = \sum_{j=1}^{288} y_{288(n-1)+j}^2$$

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

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

As a result we call $s_{n,288}^2$ the QV estimator. Having computed the daily $\left\{s_{n,288}^2\right\}$ series we have drawn in Figure 4 the average (over our four series) correlogram (starting at lag 3 to be compatible with above

Model	Measure of fit (distance from normal)	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.

Table 1: *Fit of the marginal distributions of returns y_n using zero meaned, symmetric distributions. We use the scaled Student t, normal inverse Gaussian (parameters γ and δ) and the normal distribution with unknown variance. $\Delta = 1$ is chosen to represent five minutes. Reported is the maxima of the quasi-likelihood functions. KL (Kullback-Liebler) distance 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.*

analysis)[12]. 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 $\{y_n^2\}$ process given in the middle picture. This seems to be very roughly confirmed here. However, we can see that the dependence amongst the empirical variance is much stronger than amongst just the noisy plain squared returns. This 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.

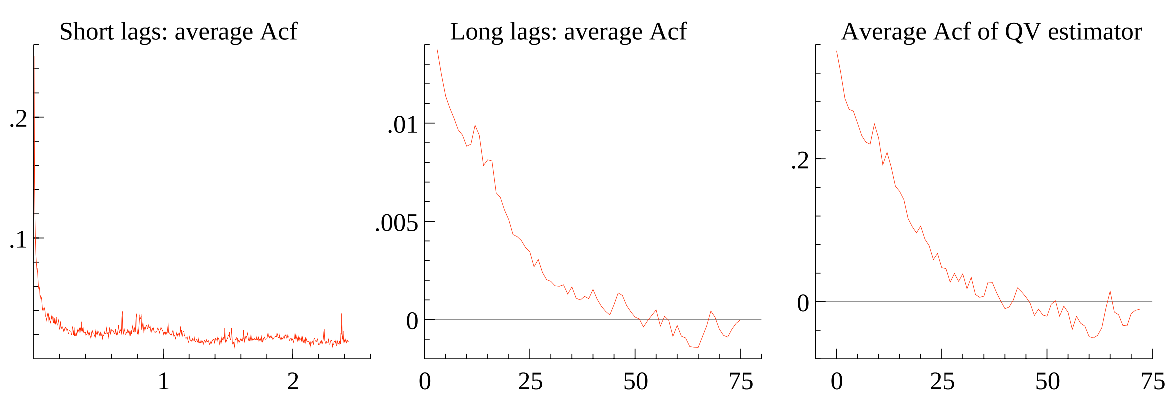

Figure 4: *Averaged of 5 correlograms each with 125,000 returns. Labels for the lags of correlogram are written using days, not 5 minute periods. Left: first 750 lags, to show short term dynamics. Middle: next 15000 lags to focus on long term pattern. Right: equivalent quadratic variation estimator based on squared 5 minute returns measured over a day.*

The empirical results suggest that we will not be able to build satisfactory volatility models from the direct use of OU processes, for these have exponential decays in their autocorrelation functions. The left hand graph of Figure 4 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

[12] Quadratic variation type estimators of the integrated volatility process $\{\sigma^2_n\}$, have been used before us in Andersen, Bollerslev, Diebold, and Labys (2000). They study the empirical correlograms and marginal distributions of the resulting statistics. However, in their paper they use unadjusted data.

continuous time volatility.

In this section we will assume the instantaneous volatility process $\{\sigma^2(t)\}$ is made up by the addition of m independent stationary processes $\left\{\sigma_j^2(t)\right\}$. For ease of expositions we will assume[13]

$$\sigma^2(t) = \sum_{j=1}^{m} \sigma_j^2(t), \quad \sigma_j^2(t) \sim IG(\delta w_j, \gamma), \quad \text{where} \quad \sum_{j=1}^{m} w_j = 1 \quad \text{and} \quad \{w_j \geq 0\}.$$

Then $\sigma^2(t) \sim IG(\delta, \gamma)$, and so $\mathrm{E}(\sigma^2(t)) = \zeta = \delta/\gamma$ and $\mathrm{Var}(\sigma^2(t)) = \omega^2 = \delta/\gamma^3$. The corresponding integrated volatility is

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

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

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

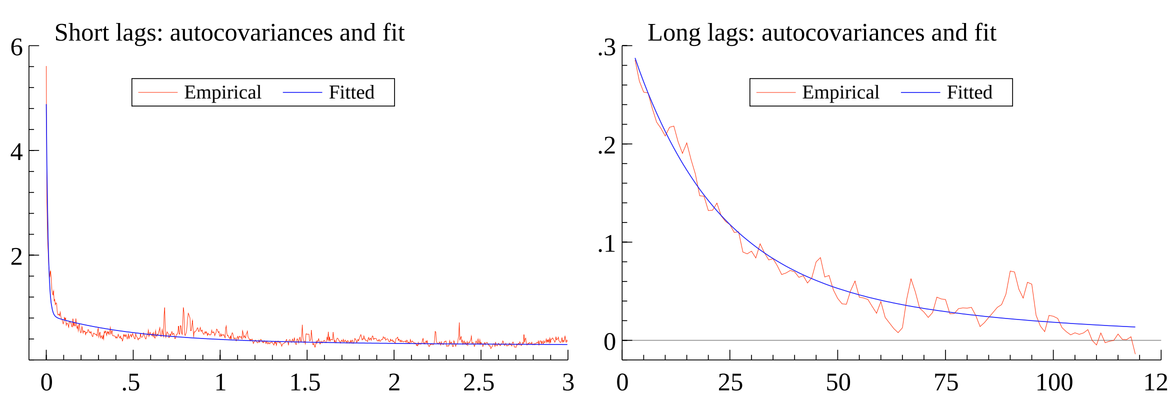

Figure 5: *Fitted and raw autocovariance functions for the single series of 684,000 observations. The x-axis is marked in days not 5 minute periods. The graphed fit uses a superposition of four independent OU processes. Left hand graph draws the average autocovariance in the day, rather than graphing all the 5 minute correlations.*

In order 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

[13]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.

series of length 684,000 observations, with the parameterised one given in (47). In particular the criterion we minimised was

$$
\begin{aligned}
S &= \sum_{s=1}^{3\times 288}\left\{c_s-\mathrm{Cov}(y_n^2,y_{n+s}^2)\right\}^2 \\
&\quad +288\sum_{s=3}^{123}\left\{\frac{1}{288}\sum_{k=1}^{288}c_{288s+k}-\frac{1}{288}\sum_{k=1}^{288}\mathrm{Cov}(y_n^2,y_{n+288s+k}^2)\right\}^2 .
\end{aligned}
$$

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 is given in Figure 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.

m	w_j				$\exp(\lambda_j\Delta)$				ω^2	S
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

Table 2: *Fit of the autocovariance function using a variety of superpositions of OU processes. The fit is based on the single series of around 684,000 observations. The number of processes is denoted by* m. *The weights are denoted by* w_j, *while the memory of the components is* $\exp(\lambda_j\Delta)$. *The variance of the volatility is written as* ω and appears in $\mathrm{Cov}(y_n^2,y_{n+s}^2)$. *Finally,* S *denotes sum of squares given above.*

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 dependencies are focused on, while for larger values of m the longer term dynamics are clarified while the short term dynamics are picked up. The most interesting feature of the table 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 ten percent 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 Tradition inference approaches

5.4.1 Likelihood

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

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

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

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

while noting that

$$\left\{ \begin{array}{c} \sigma^2(n\Delta) \\ z(\lambda n\Delta) \end{array} \right\} = \left(\begin{array}{c} e^{-\lambda\Delta}\sigma^2\left\{(n-1)\,\Delta\right\} \\ z\left\{\lambda\,(n-1)\,\Delta\right\} \end{array} \right) + \eta_n, \quad \eta_n \stackrel{L}{=} \left(\begin{array}{c} e^{-\lambda\Delta}\int_0^\Delta e^{\lambda t}\mathrm{d}z(\lambda t) \\ \int_0^\Delta \mathrm{d}z(\lambda t) \end{array} \right). \tag{49}$$

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

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

$$\eta_n \stackrel{L}{=} \alpha^{-1} \left\{ \begin{array}{c} e^{-\lambda\Delta}\sum_{i=1}^{N(1)}\log(c_i^{-1})e^{\lambda\Delta r_i} \\ \sum_{i=1}^{N(1)}\log(c_i^{-1}) \end{array} \right\}, \quad r_i \stackrel{i.i.d.}{\sim} U(0,1),$$

and defining $c_1 < c_2 < ...$ 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 $f(\sigma_1^2,...,\sigma_T^2;\delta,\gamma,\lambda)$, and so we cannot hope to solve for $f(y;\theta)$ analytically or 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 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, Richardson, and Spiegelhalter (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, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1998)) using the idea of data augmentation designing a MCMC for sampling from $\theta,\sigma^2|y$, where $\sigma^2 = \left(\sigma_1^2,...,\sigma_T^2\right)$. A generic scheme for carrying this out is given below:

1. Initialize σ^2 and θ.
2. Update σ^2 from $\sigma^2|\theta,y$, using a Metropolis-Hastings algorithm (one element at a time (e.g. Carlin, Polson, and Stoffer (1992)) or using a blocking strategy (e.g. Shephard and Pitt (1997))).
3. Perform a Metropolis update on $\theta|y,\sigma^2$.
4. Goto 2.

Cycling through 2 to 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 $\sigma_1^2,...,\sigma_T^2$ basically determines λ in a simple OU model — that is when we know the volatility we are over-conditioning[14]. 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, Shephard, and Chib (1998)). This very unfortunate effect seems inevitable for this type of parameterisation.

The above problems can potentially be removed if we reparameterise the MCMC problem to work more directly in terms of the components of the shock terms $\{\eta_n\}$. Recall they have an infinite series

[14]The easiest way of thinking about this is to work with a discrete time version of this type of model where

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

where $\eta_n > 0$ and is i.i.d.. Then

$$e^{-\lambda} \leq \min_n \sigma_n^2/\sigma_{n-1}^2.$$

This suggests the likelihood function will have a mode very close to $e^{-\lambda}$. 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).

representation (31) which can be used to simulate from them. Each draw in these infinite series are based on the sequences, independent over n, $\{a_{in}\}$ and $\{r_{in}\}$. Here the $r_{in}'s$ are independent copies of a uniform random variable r on $[0,1]$ and $a_{1n} < ... < a_{in} < ...$ are the arrival times of a Poisson process with intensity 1. Suppose we truncate the sequence after K random variables for each value of n and write $a_{(n)} = (a_{1n}, ..., a_{Kn})'$ and $r_{(n)} = (r_{1n}, ..., r_{Kn})'$, and $a = (a_{(1)}, ..., a_{(T)})$ and $r = \left(r_{(1)}, ..., r_{(T)}\right)$. Then we could perform MCMC based inference based upon 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 $\theta, a, r, \sigma^2(0)$ determine $\left\{\sigma_n^2\right\}$. In principle this would only be an approximation (due to the truncation of the infinite series representation), as it would be based upon K variables, however if K was chosen as a large number then it is likely to perform well.

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

5.4.2 Best linear predictors

In order to simplify the exposition suppose that $\beta = \rho = 0$ (which may be reasonable for exchange rate data)[15]. Then we note that $y_n|\sigma_n^2 \sim N(\mu\Delta, \sigma_n^2)$ and so

$$\begin{aligned} \begin{pmatrix} y_n \\ y_n^2 \end{pmatrix} &= \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 + \sigma_n^2 \end{pmatrix} + u_n, \quad \text{where} \quad u_n \sim MD, \\ \text{Var}(u_{1n}) &= \text{E}\left(\sigma_n^2\right) = \xi\Delta \\ \text{Cov}(u_{1n}, u_{2n}) &= 2\mu\Delta\text{E}\left(\sigma_n^2\right) = 2\mu\Delta^2\xi. \\ \text{Var}(u_{2n}) &= 4\mu^2\Delta^2\text{E}\left(\sigma_n^2\right) + 2\text{E}\left(\sigma_n^4\right) \\ &= 4\mu^2\Delta^3\xi + 2\left\{2\omega^2 r^{**}(\Delta) + \xi^2\Delta^2\right\}. \end{aligned} \tag{50}$$

Further $\left(\sigma_n^2, z_n\right)$ is a linear process which is driven by the i.i.d. noise $\{\eta_n\}$. It is easy to see that

$$\text{E}(\eta_n) = \xi \begin{pmatrix} 1 - e^{-\lambda\Delta} \\ \lambda\Delta \end{pmatrix}, \quad \text{Var}\left(\eta_n\right) = 2\omega^2 \begin{Bmatrix} \frac{1}{2}\left(1 - e^{-2\lambda\Delta}\right) & \left(1 - e^{-\lambda\Delta}\right) \\ \left(1 - e^{-\lambda\Delta}\right) & \lambda\Delta \end{Bmatrix}.$$

These results imply that a linear state space representation of the $\left(y_n, y_n^2\right)$ (with uncorrelated $\{u_n\}$ and $\{\eta_n\}$)[16] is

$$\begin{aligned} \begin{pmatrix} y_n \\ y_n^2 \end{pmatrix} &= \begin{pmatrix} \mu\Delta \\ \mu^2\Delta^2 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ \lambda^{-1} & 0 \end{pmatrix} \alpha_n + u_n, \qquad \text{with} \\ \alpha_{n+1} &= \begin{bmatrix} \{z(\lambda\left(n+1\right)\Delta) - z\left(\lambda n\Delta\right)\} + \sigma^2\left(n\Delta\right) - \sigma^2\left\{\left(n+1\right)\Delta\right\} \\ \sigma^2\left\{\left(n+1\right)\Delta\right\} \end{bmatrix} \\ &= \begin{Bmatrix} 0 & \left(1 - e^{-\lambda\Delta}\right) \\ 0 & e^{-\lambda\Delta} \end{Bmatrix} \alpha_n + \begin{pmatrix} \eta_{2n} - \eta_{1n} \\ \eta_{1n} \end{pmatrix}. \end{aligned}$$

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. Let us write these quantities

[15] The extension to the leverage case would write $y_n = \mu\Delta + \overline{z}_n + u_{1n}$ and $y_n^2 = \mu^2\Delta^2 + \sigma_n^2 + \text{E}(\overline{z}_n^2) + u_n$.

[16] As σ_n^2 has an ARMA(1,1) representation the minimial dimension of the state space form is two. However, it is possible to remove $z(\lambda\left(n+1\right)\Delta) - z\left(\lambda n\Delta\right)$ from the transition equation and have a single state variable. This would result in correlated measurement and transition noise.

as $s_{n|n-1}$ and $p_{n|n-1}$, then it is straightforward in the case that $\mu = 0$ to demonstrate that if $s_{1|0} \geq 0$ then $s_{n|n-1}$ is always non-negative and, in 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 semi-parametric, in the sense that they do not rely on any distributional assumptions about the volatility process only on ξ, ω^2, μ and λ[17].

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 generalise to where we sum m independent OU processes (46). Suppose $\mathrm{E}(\sigma_j^2(t)) = w_j\xi$ and $\mathrm{Var}(\sigma_j^2(t)) = w_j\omega^2$. Then we have $\left(\sigma_{jn}^2, z_{jn}\right)$ are independent over j and are again linear processes driven by noise $\left\{\eta_{jn}\right\}$. In this setup

$$\mathrm{E}(\eta_{jn}) = w_j\xi \left(\begin{array}{c} 1 - e^{-\lambda_j\Delta} \\ \lambda_j\Delta \end{array} \right), \quad \mathrm{Var}\left(\eta_{jn}\right) = 2w_j\omega^2 \left\{ \begin{array}{cc} \frac{1}{2}\left(1 - e^{-2\lambda_j\Delta}\right) & \left(1 - e^{-\lambda_j\Delta}\right) \\ \left(1 - e^{-\lambda_j\Delta}\right) & \lambda_j\Delta \end{array} \right\}.$$

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

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

5.4.3 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 $\mathrm{E}(\sigma_n^2|\mathcal{F}_{n-1})$, where $\mathcal{F}_{n-1}$ denotes the information available at time $(n-1)\,\Delta$. In this part of the paper we show this quantity can be recursively computed using a particle filter (see Pitt and Shephard (1999a) and Doucet, de Freitas, and Gordon (2000) for a book-length review of this material) and, further, we will 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 of $\sigma_n^2|\mathcal{F}_n$ for $n = 1, ..., T$. It has the following basic structure

`Basic particle filter (Gordon, Salmond, and Smith (1993))`

1. Assume a sample $\sigma^{2(1)}(n\Delta), ..., \sigma^{2(M)}(n\Delta)$ from $\sigma_n^2, \sigma^2(n\Delta)|\mathcal{F}_n$. Set $n = 0$.
2. For each $\left\{\sigma^{2(m)}(n\Delta)\right\}$ generate K offspring

$$\left\{\sigma_{n+1}^{2(m,k)}, \sigma^{2(m,k)}((n+1)\,\Delta)\right\}, \quad k = 1, ..., K,$$

   using (48) and (49). Compute

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

3. Calculate normalised weights $w_{m,k} \propto w_{m,k}^*$ which sums to one over m and k.

[17] For related ideas, in the context of discrete time log-normal SV models, see Harvey, Ruiz, and Shephard (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, Polson, and Rossi (1994)) principally due to the variance caused by inliers (small values of y_n^2). This particular problem does not necessarily carry over to our current treatment.

4. Resample, with unequal weights, amongst the $\{\sigma^{2(m,k)}((n+1)\Delta), w_{m,k}\}$ to produce a new sample $\sigma^{2(1)}((n+1)\Delta), ..., \sigma^{2(M)}((n+1)\Delta)$. This sample is approximately from $\sigma^2_{n+1}|\mathcal{F}_{n+1}$

5. Goto 2.

As M gets 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 1,000 to 10,000 are effective, while we typically take K as 3. Figure 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 $\{\sigma^2_n\}$ process. The top of the Figure 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.

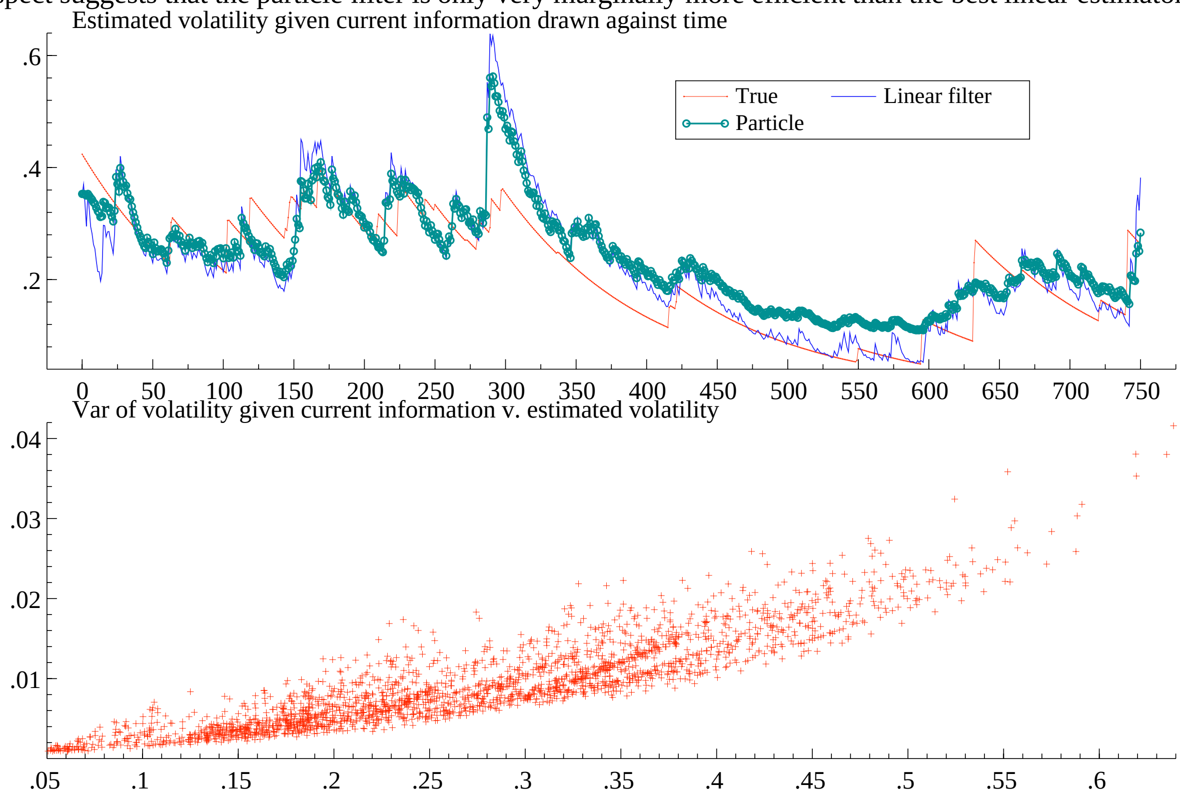

Figure 6: OU process with $\Gamma(\nu, \alpha)$ marginals. *Throughout,* $\Delta = 1$, $\nu = 3$, $\alpha = 8.5$, $\Delta = 0.01$. *Top: against time we plot the true* σ^2_n, *the best linear estimator and the particle filter's estimator of* $\mathrm{E}(\sigma^2_n|\mathcal{F}_{n-1})$. *Bottom: against* $\mathrm{E}(\sigma^2_n|\mathcal{F}_{n-1})$ *we plot* $\mathrm{Var}(\sigma^2_n|\mathcal{F}_{n-1})$ *where both terms are estimated using the particle filter. To do this we take* $T = 2,300$.

The bottom of Figure 6 graphs the particle filters estimate of $\mathrm{Var}(\sigma^2_n|\mathcal{F}_n)$ against $\mathrm{E}(\sigma^2_n|\mathcal{F}_n)$. The graph 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.

5.4.4 Estimating equations

Earlier we derived general expressions for the second order moments of the return sequence $\{y_n\}$. In a recent paper 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, Jeantheau, and Laredo (1998), provide powerful methods for estimating these types of models. However, we are yet to study their effectiveness in practice.

5.4.5 Indirect inference

Equations (48) and (49) can be used to simulate a return sequence $\{y_n\}$ without any form of discretisation error. However, it is now clear that this is insufficient for us to conduct straightforward likelihood based inference, even when we are prepared 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, Monfort, and Renault (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 correct for the approximation by simulation.

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

```
Indirect inference:  auxiliary model is GARCH
```

1. Find the MLE of ψ
$$\widehat{\psi} = \arg_{\psi} \max \log L_{GARCH}(\psi; y),$$
as if the data had been produced by the GARCH model.

2. Find $\widehat{\theta}$ such that
$$\widehat{\psi} = \arg_{\psi} \max \log L_{GARCH}(\psi; \widehat{y}^S(\widehat{\theta})).$$
That is change the simulated data until its GARCH version of the MLE is the same as that which results from the data.

We call $\widehat{\theta}$ 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 the paper by Clark (1973). Recent work on this topic includes Ghysels and Jasiak (1994), Conley, Hansen, Luttmer, and Scheinkman (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 infinity for t tending to infinity and is independent of the Brownian motion b. The resulting process is $b\{\tau(t)\}$.

Now consider models of the type

$$x^*(t) = \int_0^t \sigma(s) dw(s), \qquad (51)$$

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

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, p. 64)) tells us that, if we define processes γ and b by

$$\gamma(t) = \inf\{u : [x^*](u) > t\} \qquad \text{and} \qquad b(t) = x^*(\gamma(t))$$

then b is a Brownian motion and

$$\{x^*(t)\}_{t\geq 0} \stackrel{\mathcal{L}}{=} \{b([x^*](t))\}_{t\geq 0} \tag{52}$$

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

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

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

6.2 Pricing

6.2.1 Non-arbitrage

In this subsection we will show that our leveraged SV model does not allow arbitrage[18]. 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

$$x_0^*(t) = \int_0^t \sigma(s)dw(s) \qquad \text{with} \qquad \sigma^2(t) = e^{-\lambda t}\int_{-\infty}^t e^{\lambda s}\mathrm{d}z(s).$$

Once again we assume w and z are independent, while we write $\{\mathcal{F}_t\}_{t\geq 0}$ 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 have to verify the existence of an equivalent martingale measure 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 the equation

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

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

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

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 reexpressed 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 6.1 Under the above setup we have

[18]In the case of no leverage, $\rho = 0$, non-arbitrage follows essentially from Lipster and Shiryayev (1977, Ch. 6) and is well known. The arguments given below combines their technique with the Esscher transformation technique well known for Lévy process models.

(i) the process $d(t)$ is a mean 1 martingale, and hence P' is a probability measure

(ii) the price process $\exp\{x^*(t)\}$ is a martingale under P'.

□

The proof of this result is given in the Appendix.

Example Suppose $z(1) \sim IG(\delta, \gamma)$. 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 (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 $\rho \leq 0$, which is the econometrically relevant case. Then, as θ increases from $-\infty$ to its upper bound $\gamma^2/2$ the left hand side of (58) decreases monotonically from ∞ to $|\rho|\sqrt{2}/\gamma$. Consequently, (58) is solvable if and only if $|\rho| \leq \sqrt{2}\gamma$ (which in practice is not a very binding constraint).
□

6.2.2 Derivatives

The fact that our SV model is arbitrage-free means there exists at least one equivalent martingale measure (EMM) with which we can compute derivative prices. An important question is which one to use? In a recent paper Nicolato and Prause (2000) have tackled this problem for our model when $\sigma^2(t) \sim IG$ 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 EMM 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

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

where $BS\left\{K, x^*(n\Delta), \frac{1}{\Delta}\sigma^2_{n+1}, n\Delta+\Delta\right\}$ denotes the Black-Scholes price of the option with initial value $x^*(n\Delta)$, strike price K and constant volatility $\frac{1}{\Delta}\sigma^2_{n+1}$. 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 $\sigma^2_{n+1}|\sigma^2(n\Delta)$ 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 will derive the cumulant generating function, while Scott (1997) and Carr and Madan (1998) discuss 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[19],

$$\begin{aligned} \mathrm{K}\left\{\zeta \ddagger x^*(\Delta)\right\} &= \log \mathrm{E}^Q\left[\exp\left\{\zeta x^*(\Delta)\right\}|x^*(0), \sigma^2(0)\right] \\ &= \left\{x^*(0) + r\Delta\right\}\zeta + \mathrm{K}\left\{\left(\zeta\beta + \frac{1}{2}\zeta^2\right) \ddagger \sigma_1^2|\sigma^2(0)\right\}. \end{aligned}$$

Hence the only unsolved problem is to compute the cumulant generating function of $\sigma_1^2|\sigma^2(0)$.

Recall

$$\begin{aligned} \sigma_1^2 &= \lambda^{-1}\left\{z(\lambda\Delta) - \sigma^2(\Delta) + \sigma^2(0)\right\} \\ &= \int_0^\Delta \varepsilon(\Delta - s;\lambda)\mathrm{d}z(\lambda s) + \varepsilon(\Delta;\lambda)\sigma^2(0), \quad \text{where} \quad \varepsilon(t;\lambda) = \lambda^{-1}(1 - e^{-\lambda t}). \end{aligned}$$

Consequently it is sufficient to work with

$$\begin{aligned} \overline{\mathrm{K}}\left\{\theta \ddagger \sigma_1^2|\sigma^2(0)\right\} &= \log \mathrm{E}\left(e^{-\theta\sigma_1^2}|\sigma^2(0)\right) \\ &= -\theta\varepsilon(\Delta;\lambda)\sigma^2(0) + \overline{\mathrm{K}}\{\theta \ddagger \lambda^{-1}\int_0^{\lambda\Delta}(1 - e^{-\lambda\Delta+u})\mathrm{d}z(u)\} \\ &= -\theta\varepsilon(\Delta;\lambda)\sigma^2(0) + \int_0^{\lambda\Delta}\overline{\mathrm{K}}\{\theta\lambda^{-1}(1 - e^{-\lambda\Delta+u}) \ddagger z(1)\}\mathrm{d}u \\ &= -\theta\varepsilon(\Delta;\lambda)\sigma^2(0) + \lambda^{-1}\int_0^\Delta \overline{\mathrm{K}}\{\theta\varepsilon(\Delta - s;\lambda) \ddagger z(1)\}\mathrm{d}s \\ &= -\theta\varepsilon(\Delta;\lambda)\sigma^2(0) + \lambda^{-1}\int_0^\Delta \overline{\mathrm{K}}\{\theta\varepsilon(s;\lambda) \ddagger z(1)\}\mathrm{d}s \\ &= -\theta\varepsilon(\Delta;\lambda)\sigma^2(0) + \lambda^{-1}\int_0^\Delta k(\theta\varepsilon(s;\lambda))\mathrm{d}s \\ &= -\theta\varepsilon(\Delta;\lambda)\sigma^2(0) + \lambda^{-2}\int_0^{1-e^{-\lambda\Delta}}(1-u)^{-1}k(\lambda^{-1}\theta u)\mathrm{d}u. \end{aligned}$$

Example 6 *Suppose $z(1) \sim IG(\delta, \gamma)$, implying $k(\theta) = \delta\gamma - \delta\gamma(1 - 2\gamma^{-2}\theta)^{1/2}$. Then*

$$\begin{aligned} \int_0^{1-e^{-\lambda\Delta}}(1-u)^{-1}k(\lambda^{-1}\zeta u)\mathrm{d}u &= \delta\gamma\int_0^{1-e^{-\lambda\Delta}}\frac{1-(1+\varkappa u)^{1/2}}{1-u}\mathrm{d}u \\ &= \delta\gamma\left\{\lambda\Delta - I(\varkappa,\Delta)\right\}, \end{aligned}$$

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

$$\begin{aligned} I(\varkappa,\Delta) &= \int_0^{1-e^{-\lambda\Delta}}\frac{(1+\varkappa u)^{1/2}}{1-u}\mathrm{d}u \\ &= \lambda\Delta\sqrt{1+\varkappa} + 2\left\{\left[1 - b(\varkappa) + \sqrt{1+\varkappa}\log\frac{\{\sqrt{1+\varkappa}+b(\varkappa)\}}{\{\sqrt{1+\varkappa}+1\}}\right]\right\}. \end{aligned}$$

Here $b(\varkappa) = \sqrt{1+\varkappa-\varkappa e^{-\lambda\Delta}}$.

[19]This is a slight abuse of notation for we have previously assumed $x^*(0) = 0$, which is not our intention here.

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, Pan, and Singleton (2000) (see also Stein and Stein (1991)) working with a square root process

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

6.3 Trade-by-trade dynamics

Recently vast datasets 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 will 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 $i - th$ event and so $\tau_{N(t)}$ is the time of the last recorded event when we are standing at calender time t.

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

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

where for sake of simplicity the $\{y_i\}$ are assumed independent standard normal and $\sigma^{2*}(t) = \int_0^t \sigma^2(u)\mathrm{d}u$. We assume 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 6.1 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

$$\lim_{\delta\uparrow\infty} x_\lambda^*(\cdot) \stackrel{\mathcal{L}}{\rightarrow} x^*(\cdot).$$

Proof: Given in the Appendix.

This means that the tick-by-tick model will converge to a stochastic volatility model as the amount of trading gets 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 lattice and have quite complicated dependence structures which are not easy to model (see Rydberg and Shephard (2000) and Rydberg and Shephard (1998)).

6.4 Vector OU 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

$$\sigma^2(t) = \left(\sigma_1^2(t), ..., \sigma_q^2(t)\right) \quad \text{via the BDLPs} \quad z(t) = (z_1(t), ..., z_q(t))$$

as follows. The multivariate form of (14) is

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

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

$$\int_{R_+^q} \min\left\{1, \langle \xi, x \rangle^2\right\} W(\mathrm{d}x) < \infty, \qquad \text{for all } \xi \in R_+^q.$$

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

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

Imposing the condition

$$\int_1^\infty \log(x_i)\, w_i(x_i) dx_i < \infty$$

we may then, on account of Lemma 2.1, define the stationary process $\sigma_i^2(t)$ by

$$\sigma_i^2(t) = \int_{-\infty}^0 e^s dz_i(\lambda_i t + s).$$

Note that

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

The full specification of σ^2 then rests on the choice of w, which we may aim to reflect the dependencies amongst the volatility processes $\sigma_1^2(t), ..., \sigma_q^2(t)$.

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

Example 7 *Let $q = 2$ and let w, defined in polar coordinates (r, a), be*

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

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/IG-OU processes, while for $\phi \uparrow \infty$ the processes $z_1(s)$ and $z_2(s)$ tend to one and the same BDLP/IG-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 Rosinski (1990) and Rosinski (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 (2000).

Consider a $q-$dimensional BDLP process z with density $w(x)$ as in the subsection directly above and let $\widetilde{w}(r, a)$ ($a = (a_1, ..., q_{q-1})$) be the representation of w in polar coordinates. We assume, for simplicity (and as in Example 7), that $\widetilde{w}$ factors as $\widetilde{w}(r, a) = g(r)b(a)$ where g is a one-dimensional Lévy density on R_+ and b is a probability density. Now let

$$G^{-1}(s) = \inf\left\{r > 0 : G^+(r) \le s\right\}, \qquad \text{where} \qquad G^+(r) = \int_r^\infty g(\rho) d\rho.$$

Proposition 6.1 Let a_j, $j = 1, 2, ...$ be the arrival times of a Poisson process with rate 1 and let u_j, $j = 1, 2, ...$ be an i.i.d. 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 i.i.d. 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 the series (61) converges a.s. and

$$\{z(s) : 0 \le t \le 1\} \stackrel{\mathcal{L}}{=} \{\tilde{z}(s) : 0 \le t \le 1\} \tag{62}$$

□

Furthermore we have

Proposition 6.2 If f_i, $i = 1, ..., d$, 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 $i = 1, ..., d$ and the u_{ij} i.i.d. with law determined by b.

□

6.5 Multivariate SV models

6.5.1 Model structure

A simple q-dimensional version of the SV model for log-prices sets $x^*(t) = \left\{x_1^*(t), ..., x_q^*(t)\right\}$ 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 $\Sigma^*(t)$ using quadratic variation for $x^*(t)$ is a continuous $q-$dimensional local martingale plus a process which is continuous with bounded variation and so

$$[x^*](t) = \text{p}-\lim_{r \to \infty} \sum \{x^*(t_{i+1}^r) - x^*(t_i^r)\}\{x^*(t_{i+1}^r) - x^*(t_i^r)\}' = \Sigma^*(t) \tag{64}$$

for any sequence of partitions $t_0^r = 0 < t_1^r < ... < t_{m_r}^r = t$ 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

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

Here $\phi = \left(\beta_1, ..., \beta_q\right)$ are unknown parameters and the $\sigma_1, \sigma_2, ..., \sigma_{q+1}$ are mutually independent OU processes which are square integrable and stationary. It has common, but differently scaled, stochastic volatility model and individual stochastic volatility 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, Sentana, and Wadhwani (1994), Pitt and Shephard (1999b) and Chib, Nardari, and Shephard (1999). Its motivation is that in financial assets it is often the case that 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 generalised by allowing the volatilities to be dependent 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 $\beta b\left(\sigma^2(t)\right)$, 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 (e.g. 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 models 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 stochastic volatility.

Although the treatment of OU processes we have presented in this paper is extensive, there are a number of unresolved issues. A principle 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 generalisation to the multivariate case is at its infant stage and much work has to be carried out in order to make this a very flexible framework.

More generally, we believe that Lévy driven processes have great potential for applications to fields other than finance and econometrics, for instance to turbulence studies. It can also be further developed to a general toolbox for time series analysis. In this connection, we note that while 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

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

with $\sigma^2(t)$ an OU process as in (2). See Barndorff-Nielsen and Shephard (2000) 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 (65) not by Brownian motion but by another independent Lévy process with positive increments.

8 Acknowledgments

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 our aggregation 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 this data.

Ole E. Barndorff-Nielsen's work is supported by CAF (`www.caf.dk`), which is funded by the Danish Social Science Research Council, and by MaPhySto (`www.maphysto.dk`), which is funded by the Danish National Research Foundation. Neil Shephard's research is supported by the UK's ESRC through the grant "Econometrics of trade-by-trade price dynamics," which is coded R00023839.

9 Appendix

9.1 Background

This Appendix collects various proofs and results 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

$$C(\zeta \ddagger x) = \log \mathrm{E}\left(e^{i\zeta x}\right), \qquad \text{while writing} \qquad \overline{\mathrm{K}}\left\{\theta \ddagger x\right\} = \log \mathrm{E}\left(e^{-\theta x}\right),$$

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

9.2 GIG Lévy density

Proof of Theorem 2.2 Let $z \sim GIG(\nu, \delta, \gamma)$. From Halgreen (1979) we have that if $\nu \leq 0$ 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

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

and this shows that

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

is the Lévy density of z.

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

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

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

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

and corresponding Lévy density $\nu x^{-1} e^{-\phi x}$.
□

9.3 Non-arbitrage

Proof of Proposition 6.1 (i) For $0 \leq s \leq t \leq T$ we find

$$\begin{aligned} \mathrm{E}_P\{d(t)|\mathcal{F}_s\} &= \mathrm{E}_P\{\mathrm{E}_P\{d(t)|z,\mathcal{F}_s\}\mathcal{F}_s\} \\ &= e^{-\lambda t\acute{\kappa}(\theta')}\mathrm{E}_P\left\{\exp\left\{\theta'\bar{z}(\lambda t)-\frac{1}{2}\phi^2\sigma^{2*}(t)\right\}\mathrm{E}_P\{\exp\{-\phi x_0^*(t)\}|\sigma,\mathcal{F}_s\}|\mathcal{F}_s\right\} \end{aligned}$$

and here

$$\mathrm{E}_P\left\{e^{-\phi x_0^*(t)}|\sigma,\mathcal{F}_s\right\} = \exp\left[-\phi x_0^*(s)+\frac{1}{2}\phi^2\{\sigma^{2*}(t)-\sigma^{2*}(s)\}\right]$$

so that

$$\mathrm{E}_P\{d(t)|\mathcal{F}_s\} = d(s)\exp\left\{-\lambda(t-s)\bar{\kappa}(\theta')\right\}\mathrm{E}_P\{\exp\left\{\theta'\{\bar{z}(\lambda t)-\bar{z}(\lambda s)\}\right\}|\mathcal{F}_s\} = d(s)$$

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

(ii) Note first that

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

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

$$\mathrm{E}_{P'}\{y_t|\mathcal{F}_s\} = \mathrm{E}_P\{y_t d(T)/d(s)|\mathcal{F}_s\} = \mathrm{E}_P\{y_t d(t)/d(s)|\mathcal{F}_s\} \tag{67}$$

Hence

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

where

$$J = e^{\{\beta-\frac{1}{2}\phi^2\}\{\sigma^{2*}(t)-\sigma^{2*}(s)\}}\mathrm{E}_P\{e^{(1-\phi)(x_0^*(t)-x_0^*(s))}|\sigma,\mathcal{F}_s\}$$

However, by (66),

$$J = e^{\{\beta-\frac{1}{2}\phi^2+(1-\phi)^2\}\{\sigma^{2*}(t)-\sigma^{2*}(s)\}} = 1$$

so that, in view of condition (57),

$$\begin{aligned} \mathrm{E}_{P'}\{\exp\{x^*(t)\}|\mathcal{F}_s\} &= \exp\{x^*(s)-\lambda(t-s)\bar{\kappa}(\theta')\}\mathrm{E}_P\left[\exp\left\{(\rho+\theta')\{\bar{z}(\lambda t)-\bar{z}(\lambda s)\}\right\}|\mathcal{F}_s\right] \\ &= \exp\left[x^*(s)-\lambda(t-s)\left\{\bar{\kappa}(\rho+\theta')-\bar{\kappa}(\theta')\right\}\right] \\ &= \exp\{x^*(s)\} \end{aligned}$$

□

9.4 Trade-by-trade dynamics

Lemma 9.2 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 $i-th$ event and so $\tau_{N(t)}$ is the time of the last recorded event when we are standing at calender time t. Then for $\delta \to \infty$ we have that $\tau_{N(t)} \xrightarrow{p} t$.

Proof: It suffices to show that for every $\varepsilon > 0$ we have that

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

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

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

Consequently

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

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

Proof of Theorem 6.1 It is helpful to rewrite the process as

$$x_\delta^*(t) = -\mu\left\{t-\tau_{N(t)}\right\}+\beta\left[\sigma^{2*}(t)-\sigma^{2*}\left\{\tau_{N(t)}\right\}\right]+\beta\sigma^{2*}(t)+\mu t+\frac{1}{\sqrt{\delta}}\sum_{k=1}^{N(t)}y_k.$$

We obtain from Lemma 9.2 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

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

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

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

where $\overline{y}_{N(t)} = \sqrt{\frac{1}{n}}\left(y_1+...+y_n\right)$. Trivially, conditionally on $\delta(\cdot)$ we have that $N(t)/\delta\overset{a.s.}{\to}\sigma^{2*}(t)$ as $\delta\to\infty$ and $\overline{y}_{N(t)}\sim N(0,1)$ exactly. Thus

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

where $u\sim N(0,1)$ and is independent of $\sigma^{2*}(t)$. That is, 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. $x_\delta^*(\cdot)\overset{\mathcal{L}}{\to}x^*(\cdot)$.
□

References

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

Andersen, T. G. and T. Bollerslev (1997b). Intraday periodicity and volatility persistence in financial markets. *Journal of Empirical Finance 4*, 115–158.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. *International Economic Review 39*, 885–905.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2000). The distribution of exchange rate volatility. *Journal of the American Statistical Association 95*. Forthcoming.

Ané, T. and H. Geman (2000). Order flow, transaction clock and normality of asset returns. *Journal of Finance 54*. Forthcoming.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics 24*, 1–14.

Barndorff-Nielsen, O. E. (1998). Processes of normal inverse Gaussian type. *Finance and Stochastics 2*, 41–68.

Barndorff-Nielsen, O. E. (2000). Superposition of Ornstein-Uhlenbeck type processes. *Theory of Probability and Its Applications*. Forthcoming.

Barndorff-Nielsen, O. E., J. L. Jensen, and M. Sørensen (1998). Some stationary processes in discrete and continuous time. *Advances in Applied Probability 30*, 989–1007.

Barndorff-Nielsen, O. E. and N. Shephard (2000). Modelling by Lévy processes for financial econometrics. In O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes – Theory and Applications*. Boston: Birkhäuser. Forthcoming.

Bertoin, J. (1996). *Lévy Processes*. Cambridge: Cambridge University Press.

Bertoin, J. (1999). Subordinators: examples and applications. In P. Bernard (Ed.), *Lectures on Probability Theory and Statistics. Lecture Notes in Mathematics 1717*. Berlin: Springer-Verlag.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the Business and Economic Statistics Section, American Statistical Association*, 177–181.

Bollerslev, T., R. F. Engle, and D. B. Nelson (1994). ARCH models. In R. F. Engle and D. McFadden (Eds.), *The Handbook of Econometrics, Volume 4*, pp. 2959–3038. Amsterdam: North-Holland.

Brockwell, P. J., S. I. Resnick, and R. L. Tweedie (1982). Storage processes with general release rule and additive inputs. *Advances in Applied Probability 14*, 392–433.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton, New Jersey: Princeton University Press.

Carlin, B. P., N. G. Polson, and D. Stoffer (1992). A Monte Carlo approach to nonnormal and nonlinear state-space modelling. *Journal of the American Statistical Association 87*, 493–500.

Carr, P. and D. B. Madan (1998). Option valuation using the fast Fourier transform. Unpublished paper: Robert H Smith School of Business, University of Maryland.

Chib, S., F. Nardari, and N. Shephard (1999). Analysis of high dimensional multivariate stochastic volatility models. Unpublished paper: Nuffield College, Oxford.

Cinlar, E. and M. Pinsky (1972). On dams with additive inputs and a general release rule. *Journal of Applied Probability 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 E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*, 291–323.

Conley, T. G., L. P. Hansen, E. G. J. Luttmer, and J. A. Scheinkman (1997). Short-term interest rates as subordinated diffusions. *Review of Financial Studies 10*, 525–77.

Cox, D. R. (1991). Long-range dependence, non-linearity and time irreversibility. *Journal of Time Series Analysis 12*, 329–35.

Dacorogna, M. M., U. A. Muller, R. B. Olsen, and O. V. Pictet (1998). Modelling short term volatility with GARCH and HARCH. In C. Dunis and B. Zhou (Eds.), *Nonlinear Modelling of High Frequency Financial Time Series*. Chichester: Wiley.

Danielsson, J. and R. Payne (1999). Real trading patterns and prices in spot foreign exchange markets. Unpublished paper, Financial Markets Group, London School of Economics.

Diebold, F. X. (1988). *Empirical Modelling of Exchange Rate Dynamics*. New York: Springer-Verlag.

Diebold, F. X. and M. Nerlove (1989). The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. *Journal of Applied Econometrics 4*, 1–21.

Ding, Z. and C. W. J. Granger (1996). Modeling volatility persistence of speculative returns: A new approach. *Journal of Econometrics 73*, 185–215.

Doornik, J. A. (1998). *Ox: Object Oriented Matrix Programming, 2.0*. London: Timberlake Consultants Press.

Doucet, A., N. de Freitas, and N. Gordon (2000). *Sequential Monte Carlo Methods in Practice*. Berlin: Springer-Verlag. Forthcoming.

Drost, F. C. and T. E. Nijman (1993). Temporal aggregation of GARCH processes. *Econometrica 61*, 909–927.

Drost, F. C. and B. Werker (1996). Closing the GARCH gap: continuous time GARCH modelling. *Journal of Econometrics 74*, 31–57.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affince jump-diffusions. *Econometrica 68*. Forthcoming.

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. *Annals of Statistics 7*, 490–506.

Eberlein, E. (2000). Application of generalized hyperbolic Lévy motion to finance. In O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes – Theory and Applications*. Boston: Birkhäuser. Forthcoming.

Eberlein, E. and U. Keller (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 G. G. J. Lee (1999). A permanent and transitory component model of stock return volatility. In R. F. Engle and H. White (Eds.), *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, Chapter 20, pp. 475–497. Oxford: Oxford University Press.

Föllmer, H. and M. Schweizer (1991). Hedging of contingent claims under incomplete information. In *Applied Stochastic Analysis, Volume 5 of Stochastics Monographs,* pp. 389–414. New York: Gordon and Breach.

Gallant, A. R. and G. Tauchen (1996). Which moments to match. *Econometric Theory 12*, 657–81.

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

Genon-Catalot, V., T. Jeantheau, and C. Larédo (2000). Stochastic volatility as hidden Markov models and statistical applications. *Bernoulli 6*. Forthcoming.

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

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

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

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

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

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

Gourieroux, C., A. Monfort, and E. Renault (1993). Indirect inference. *Journal of Applied Econometrics 8*, S85–S118.

Gradstheyn, I. and I. W. Ryzhik (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. *Journal of Econometrics 14*, 227–238.

Granger, C. W. J. and Z. Ding (1995). Some properties of absolute returns, an alternative measure of risk. *Annals d'Economie et de Statistique 40*, 67–91.

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

Halgreen, C. (1979). Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 47*, 13–17.

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

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

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–64.

Harvey, A. C. and N. Shephard (1996). The estimation of an asymmetric stochastic volatility model for asset returns. *Journal of Business and Economic Statistics 14*, 429–434.

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

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

Jacquier, E., N. G. Polson, and P. E. Rossi (1994). Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economic Statistics 12*, 371–417.

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

Jurek, Z. J. and W. Vervaat (1983). An integral representation for selfdecomposable Banach space valued random variables. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 62*, 247–262.

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Review of Economic Studies 65*, 361–393.

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

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

Marcus, M. B. (1987). *ξ-Radial Processes and Random Fourier Series*, Volume 368. Memoirs of the American Mathematical Society.

Meddahi, N. and E. Renault (1996). Aggregation and marginalization of GARCH and stochastic volatility models. Unpublished paper: CREST-INSEE.

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

Nicolato, E. and K. Prause (2000). Derivative pricing in stochastic volatility models of Ornstein-Uhlenbeck type. Unpublished paper: Dept. of Mathematics Sciences, Aarhus University.

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

Pitt, M. K. and N. Shephard (1999a). Filtering via simulation: auxiliary particle filter. *Journal of the American Statistical Association 94*, 590–599.

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

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

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

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

Rogers, L. C. G. and D. Williams (1996). *Diffusions, Markov Processes and Martingales. Volume 2, Ito Calculus*. Chichester: Wiley.

Rosinski, J. (1990). On series representations of infinitely divisible random vectors. *Annals of Probability 18*, 405–430.

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

Rosinski, J. (1999). Series expansion without compensation for infinitely divisible distributions. Unpublished paper.

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

Russell, J. R. and R. F. Engle (1998). Econometric analysis of discrete-valued, irregularly-spaced financial transactions data using a new autoregressive conditional multinomial models. Unpublished paper: Graduate School of Business, University of Chicago. Presented at Second international conference on high frequency data in finance, Zurich, Switzerland, April.

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

Rydberg, T. H. and N. Shephard (1998). Dynamics of trade-by-trade price movements: decomposition and models. Working paper, Nuffield College, Oxford. Presented at Workshop on Econometrics and Finance, Issac Newton Institute, Cambridge University, October 1998.

Rydberg, T. H. and N. Shephard (2000). A modelling framework for the prices and times of trades made on the NYSE. In W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young (Eds.), *Nonlinear and Nonstationary Signal Processing*. Cambridge: Isaac Newton Institute and Cambridge University Press. Forthcoming.

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. *Mathematical Finance 7*, 413–26.

Sentana, E. (1995). Quadratic ARCH models: a potential re-intepretation of ARCH models. *Review of Economic Studies 62*, 639–661.

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

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–67.

Smith, A. A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics 8*, S63–S84.

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

Stein, E. M. and J. Stein (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies 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 O. D. Anderson (Ed.), *Time Series Analysis: Theory and Practice, 1*, pp. 203–226. Amsterdam: North-Holland.

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

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

Wolpert, R. L. and K. Ickstadt (1998). Poisson/gamma random field models for spatial statistics. *Biometrika 85*, 251–267.

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