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

Ole E. Barndorff-Nielsen
*The Centre for Mathematical Physics and Stochastics (MaPhySto), University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark.**
oebn@mi.aau.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; 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Authors' note: This paper supersedes our previously circulated but unpublished papers "Aggregation and model construction for volatility models" and "Incorporation of a leverage effect in a stochastic volatility model". We are grateful to a number of people for their comments on earlier versions. In particular we would like to thank Gary Chamberlain, Jan Pedersen, Tina Rydberg and Jack Wong for various discussions, while the comments of Torben Andersen on a summary of a preliminary version of our aggregation paper were particularly helpful.

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

*MaPhySto is funded by the Danish National Research Council

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

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

where $\grave{z}$, with $\grave{z}(0) = 0$, is a (homogeneous) Lévy process, i.e. a process with independent and stationary increments (see, for example, Rogers and Williams (1994, pp. 73–84)). 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 $\grave{z}$ is used to drive the OU process we will call $\grave{z}(t)$ 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 the increments of $\grave{z}$ 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

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

The unusual timing $\mathrm{d}\grave{z}(\lambda t)$ 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. As $\grave{z}(t)$ has strictly positive increments, it has no Brownian motion component and so consists entirely of jumps. It follows that $\sigma^2(t)$ must also exhibit jumps. However, under the models we have in mind minute 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 Ornstein-Uhlenbeck processes. For example integrated volatility, which in finance is a key measure,

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

has a simple structure.

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

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

for bounded $\{f(s) > 0\}$ and with $\grave{z}$ as above[1]. Any such process is always stationary and positive. This type of process is reminiscent of a standard infinite order linear moving average model.

1.2 Stochastic volatility processes

Continuous time models built out of Brownian motion play a crucial role in modern mathematical finance, providing the basis of most option pricing, asset allocation and term structure theory

[1] To be technically precise: $\{\grave{z}(t)\}_{t\geq 0}$ is assumed to be caglad and $\{\grave{z}(-t)\}_{t\geq 0}$ is an independent copy of $\{-\grave{z}(t)\}_{t\geq 0}$ but modified to be also caglad.

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. 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 not a 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,

$$\mathrm{d}x^*(t) = \left\{\mu + \beta\sigma^2(t)\right\}\mathrm{d}t + \sigma(t)\mathrm{d}w(t), \quad t \in [0, S], \tag{6}$$

where $\sigma^2(t)$, the instantaneous volatility, is going to be assumed to be stationary, latent and stochastically independent of $w(t)$. If $\beta = 0$ then $x^*(t)$ is a continuous process, otherwise it has jumps due to the Lévy nature of $\sigma^2(t)$. 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 mixture 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.

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

$$\mathrm{d}x^*(t) = \left\{\mu + \beta\sigma^2(t)\right\}\mathrm{d}t + \sigma(t)\mathrm{d}w(t) + \rho\mathrm{d}\bar{z}(\lambda t), \quad t \in [0, S], \tag{8}$$

where $\bar{z}(t) = \grave{z}(t) - \mathrm{E}\grave{z}(t)$, 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 two 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. Section three 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 four 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 five we discuss the empirical fitting of these models using linear and non-linear methods. Section six discusses various additional issue 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 seven 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

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

in which case it may also be written as

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

Here $\grave{z} = \{\grave{z}(t) : t \in R\}$ 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 $\grave{z}(t)$ is termed the *background driving Lévy process* (BDLP) corresponding to the process $\sigma^2(t)$.

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*, 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 $\grave{z}(t)$ such that $x(t) \overset{\mathcal{L}}{=} x$ and

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

for all $\lambda > 0$.

Conversely, if $x(t)$ is a stationary stochastic process and $\grave{z}(t)$ is a Lévy process such that $x(t) \overset{\mathcal{L}}{=} x$ and $x(t)$ and $\grave{z}(t)$ 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 $\grave{z}(1)$ (if they exist) be written as

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

respectively. Indeed they are related by the fundamental equality (Barndorff-Nielsen (1999))

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

which can be reexpressed as

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

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

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

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

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

where W is the Lévy measure of the Lévy-Khintchine representation for $\grave{z}(1)$. 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 $\sigma^2(t) \sim GIG(\overline{\lambda}, \delta, \gamma)$. 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) and Rydberg (1999). Recall that if $x \sim GIG(\overline{\lambda}, \delta, \gamma)$ then it has a density

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

where $K_{\overline{\lambda}}$ 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

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

Special cases of the GIG density are: (i) the inverse Gaussian law, where $\overline{\lambda} = -\frac{1}{2}$, (ii) the positive hyperbolic law where $\overline{\lambda} = 1$, (iii) and the inverse chi-squared (inverse gamma) law with ν degrees of freedom which occurs when $\gamma = 0$, $\overline{\lambda} = -\nu/2$ and $\delta = \sqrt{\nu}$, (iv) gamma, where $\delta = 0$ and $\overline{\lambda} > 0$. Of course if $\sigma^2 \sim GIG(\overline{\lambda}, \delta, \gamma)$ and is independent of $\varepsilon \sim N(0,1)$, then $\sigma\varepsilon$ is the generalized hyperbolic with density

$$\frac{(\gamma/\delta)^{\overline{\lambda}}}{\sqrt{2\pi}\gamma^{\overline{\lambda}-1/2} K_{\overline{\lambda}}(\delta\gamma)} \left(\sqrt{\delta^2 + x^2}\right)^{\overline{\lambda}-1/2} K_{\overline{\lambda}-1/2}\left(\gamma\sqrt{\delta^2 + x^2}\right). \tag{19}$$

Hence a continuous time volatility model built using a volatility model of OU type with GIG marginals will have generalized hyperbolic marginals for instantaneous returns. Special cases of this include the normal inverse Gaussian distribution, the hyperbolic and the Student t.

It is known that the $GIG(\overline{\lambda}, \delta, \gamma)$ 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

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

where

$$g_{\lambda}(x) = \left[(\pi^2/2)x\left\{J^2_{|\overline{\lambda}|}(\sqrt{x}) + N^2_{|\overline{\lambda}|}(\sqrt{x})\right\}\right]^{-1}$$

and $J_{\overline{\lambda}}$ and $N_{\overline{\lambda}}$ are Bessel functions.

□

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

PROOF See Appendix.

We note that the Bessel functions have simple forms when $|\overline{\lambda}|$ 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

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

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

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

- $GIG(1,\delta,\gamma)$: *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

$$W^+(x) = \left\{\delta^2 \int_0^\infty e^{-x\xi} g_1(2\delta^2\xi) \mathrm{d}\xi + 1\right\} \exp\left(-\gamma^2 x/2\right). \tag{22}$$

- $GIG(\overline{\lambda} < 0, 1, 0)$: *Reciprocal gamma distribution.* The reciprocal gamma distribution (i.e. the law of the reciprocal of a gamma variate) has density

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

The corresponding upper tail integral is

$$W^+(x) = \frac{1}{2}\int_0^\infty \exp\left(-\frac{1}{2}x\xi\right) g_\nu(\xi)\mathrm{d}\xi, \qquad \text{where} \qquad g_\nu(x) = \frac{2}{\pi^2 x}\left\{J_\nu^2(\sqrt{x}) + N_\nu^2(\sqrt{x})\right\}^{-1}, \tag{23}$$

where J_ν and N_ν are Bessel functions. Recall these functions have simple forms when ν is a half integer.

- $GIG(\overline{\lambda} > 0, 0, 1)$: *Gamma distribution.* The gamma marginal law has probability

$$\frac{1}{\Gamma\left(\overline{\lambda}\right)} x^{\overline{\lambda}-1} \exp\left(-x\right), \quad x > 0.$$

This has the corresponding upper tail integral of the Lévy density $W^+(x) = \overline{\lambda} e^{-x}$, which has the convenient property that it can be analytically inverted. In particular

$$W^{-1}(x) = \max\left\{0, -\log\left(\frac{x}{\overline{\lambda}}\right)\right\}.$$

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 condition for the stochastic differential equation

$$\mathrm{d}x(t) = -\lambda x(t)\mathrm{d}t + \mathrm{d}\check{z}(\lambda t) \tag{24}$$

to have a stationary solution is that (cf. Wolfe (1982)) $\mathrm{E}\left[\max\{0, \log|x|\}\right] < \infty$).

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

$$\int_1^\infty \log(x) W(dx) < \infty. \tag{25}$$

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

$$u(x) = \int_1^\infty w(\tau x)\mathrm{d}\tau. \tag{26}$$

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 hence 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 See Appendix.

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, $0 \le \varepsilon < 1$, $0 \le \beta$, $0 \le \gamma$ and $\max\{(\beta-1), \gamma\} > 0$. Then

$$w(x) = c\{\varepsilon x^{-1} + \beta(1+x)^{-1} + \frac{1}{2}\gamma^2\}x^{-\varepsilon}(1+x)^{-\beta}\exp\left(-\frac{1}{2}\gamma^2 x\right). \tag{27}$$

Hence Lemma 2.1 applies and ensures the existence of an OU process $\sigma^2(t)$ whose BDLP $\check{z}(t)$ has w as the Lévy density of $\check{z}(1)$. 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 Series representations of some stochastic processes

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

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

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

$$e^{-\lambda t}\int_0^{\lambda t} e^s \mathrm{d}\dot{z}(s), \tag{28}$$

rather than the BDLP $\dot{z}(s)$ itself. Carrying out direct simulation of the Lévy processes and of integrals with respect to such processes is not really practical due to the jump character of the processes. Instead in this section and the appendix we briefly review some series representations of positive Lévy processes and of a related type of integral which do allow us to perform simulation. They are, in essence, available from work of Marcus (1987) and Rosinski (1991).

Again we let W be the Lévy measure of $\dot{z}(1)$ and W^{-1} denote the inverse of the tail mass function W^+.

Proposition 2.1 Let $\{a_i\}$ and $\{r_i\}$ be 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. The Lévy process $\dot{z}$ is representable in law, on the time interval $[0,1]$, as

$$\{\dot{z}(s) : 0 \leq s \leq 1\} \stackrel{\mathcal{L}}{=} \{\tilde{z}(s) : 0 \leq s \leq 1\} \tag{29}$$

where

$$\tilde{z}(s) = \sum_{i=1}^{\infty} W^{-1}(a_i)\mathbf{1}_{[0,s]}(r_i). \tag{30}$$

□

A derivation of this key result is presented in the appendix.

Corollary 2.1 Let the process $\dot{z}(t)$ be as above and let f be a positive and integrable function on $[0,1]$. Then

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

□

A heuristic verification of this result is immediate from Proposition 2.1, as follows

$$\int_0^1 f(s)\mathrm{d}\dot{z}(s) \stackrel{\mathcal{L}}{=} \int_0^1 f(s)\mathrm{d}\tilde{z}(s) \tag{32}$$

and

$$\begin{aligned} \int_0^1 f(s)\mathrm{d}\tilde{z}(s) &= \sum_{i=1}^{\infty} W^{-1}(a_i)\int_0^1 f(s)\mathrm{d}\mathbf{1}_{[0,s]}(r_i) \\ &= \sum_{i=1}^{\infty} W^{-1}(a_i)f(r_i). \end{aligned} \tag{33}$$

For a rigorous proof, see Appendix.

The formula (31) is easily extended to the case where the upper limit in the integral is an arbitrary positive number. It will turn out to be convenient to write this number as λ, then

$$\int_0^\lambda f(s)\mathrm{d}\grave{z}(s) \overset{\mathcal{L}}{=} \sum_{i=1}^{\infty} W^{-1}(a_i/\lambda) f(\lambda r_i). \tag{34}$$

To see this, note first that

$$\int_0^\lambda f(s)\mathrm{d}\grave{z}(s) = \int_0^1 f(\lambda s)\mathrm{d}z_\lambda(s), \tag{35}$$

where $z_\lambda(s) = \grave{z}(\lambda s)$. Therefore, by Corollary 2.1,

$$\int_0^\lambda f(s)\mathrm{d}\grave{z}(s) \overset{\mathcal{L}}{=} \sum_{i=1}^{\infty} W_\lambda^{-1}(a_i) f(\lambda r_i), \tag{36}$$

where W_λ^{-1} is the inverse tail integral of the Lévy density of $z_\lambda(1) = \grave{z}(\lambda)$. The result now follows on noting that, for any $t > 0$, the Lévy density of $\grave{z}(t)$ is $tw(x)$ where, as before, $w(x)$ denotes the Lévy density of $\grave{z}(1)$.

For the class of GIG based models we have studied in detail, (20) implies $u(x)$, and hence $W^{-1}(x)$, has an exponential tail when γ is not zero. This means (36) is going to converge rapidly, which concords with our practical experience in using it.

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

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

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

$$r(u) = w_1 \exp(-\lambda_1 |u|) + ... + w_m \exp(-\lambda_m |u|), \tag{37}$$

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 (1992), Dacorogna, Muller, Olsen, and Pictet (1997) and Barndorff-Nielsen (1998).

By choosing the weights and damping factors in (37) 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 (1999) 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. 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

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

where

$$r^*(t) = \int_0^t r(u)\mathrm{d}u \quad \text{and} \quad R^*(t) = \int_0^t r^*(u)\mathrm{d}u. \tag{38}$$

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). This follows from the subordination interpretation of the SV models discussed in Section 6.1. The convergence of $t^{-1/2}x^*(t)$ to normality will, however, be slow in the case where the process $\sigma^2(t)$ exhibits long range dependence.

If $\mu = \beta = 0$ then $x^*(t)$ is a continuous local martingale (see section 6) and its quadratic variation is $\sigma^{*2}(t)$, i.e. we have

$$[x^*](t) = \mathrm{p}\!-\!\lim_{r\to\infty} \sum \{x^*(t_{i+1}^r) - x^*(t_i^r)\}^2 = \sigma^{2*}(t) \tag{39}$$

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 $r \to \infty$. Note also that, since $\sigma^{2*}(t)$ is continuous, the two terms in (39) constitute the Doob-Meyer decomposition of $x^*(t)$. The quadratic variation of volatility models has recently been highlighted by Andersen and Bollerslev (1998) in order to approximately estimate integrated volatility in foreign exchange markets. It is clear that for any finite r such a quadratic variation estimator will be an unbiased estimator of the integrated volatility, with its variance falling as $r \to \infty$. However, in practice this style of continuous time model will be a poor approximation to the sample paths of the price process when we look at very fine time intervals, which tend to have some form of discreteness. Hence there is a bias/variance trade-off. One way of studying this problem would be to use the tick-by-tick models we study in section 6 of this paper.

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 $\grave{k}$ 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

$$\mathrm{C}\left\{\zeta \ddagger \int_0^\infty f(s)\mathrm{d}x^*(s)\right\} = \log \mathrm{E}\left[\exp\left\{i\zeta \int_0^\infty f(s)\mathrm{d}x^*(s)\right\}\right] = \int_0^\infty M(s)\mathrm{d}s \qquad (40)$$

where

$$M(s) = \grave{k}\left[\int_0^\infty \{N(s,u)e^{-\lambda u}\mathrm{d}u + i\zeta\rho f(\lambda^{-1}s)\}\mathrm{d}u\right] - i\zeta(\mu + \lambda\rho\xi)f(s)$$

with

$$N(s,u) = \frac{1}{2}\zeta^2\{f^2(u)e^{-s} + f^2(\lambda^{-1}s+u)\} - i\zeta\beta\{f(u)e^{-s} + f(\lambda^{-1}s+u)\}.$$

The derivation of this result is given in the Appendix. 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 (40) 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,

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

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

Example 2 *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),*

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

where

$$\grave{\kappa}_m = m\,(\delta/\gamma)\left(2/\gamma^2\right)^{m-1}\binom{1/2}{m-1}.$$

Hence, for instance, the variance of $x^*(t)$ is seen to be $\grave{\kappa}_m(t) = (\delta/\gamma)\,t$, 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 (38),

$$\text{cor}\{y_n^2, y_{n+s}^2\} = q^{-1}\Delta^{-2}\diamondsuit R^*(\Delta s), \tag{41}$$

where

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

and

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

If $\sigma^2(t) \sim OU$ with its variance existing then $r(u) = \exp(-\lambda|u|)$. This implies

$$\text{cor}\{y_n^2, y_{n+s}^2\} = ce^{-\lambda\Delta(s-1)}, \tag{44}$$

where

$$c = q^{-1}(\lambda\Delta)^{-2}(1-e^{-\lambda\Delta})^2 \qquad \text{and} \qquad q = 6(\lambda\Delta)^{-2}(e^{-\lambda\Delta} - 1 + \lambda\Delta) + 2(\xi/\omega)^2. \tag{45}$$

Note that $0 < c < 1$ and that (44) implies that y_n^2 follows a constrained ARMA(1,1) process. This 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 (1997, 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.

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

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

$$\text{Cov}\left(y_n, y_{n+s}^2\right) = \text{E}\{y_n y_{n+s}^2\} = \rho\grave{\kappa}_2(1-e^{-\lambda\Delta})^2 \exp\{-\lambda\Delta(s-1)\}$$

$$\text{Cov}(y_n^2, y_{n+s}^2) = \left(\frac{\grave{\kappa}_2}{2\lambda^2} + \rho^2\grave{\mu}_3\right)(1-e^{-\lambda\Delta})^2 \exp\{-\lambda\Delta(s-1)\}.$$

The effect of the leverage is to allow $\text{E}\left(y_n y_{n+s}^2\right)$ and $\text{Cov}(y_n^2, y_{n+s}^2)$ to damp down exponentially with the lag length s. We should note that exactly the same dynamic structure was found by Sentana (1991) in his work on the discrete time quadratic ARCH model (QARCH). Hence we can think of the QARCH model as a 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 Background

The empirical fitting of SV models has two main goals: the estimation of the unknown parameters using discrete returns $y_1, ..., y_T$ and the sequential estimation, or filtering, of the current state of the unobserved volatility $\left\{\sigma^2(\Delta n), \check{z}(\lambda\Delta n)\right\}$ using contemporaneously available data $y_1, ..., y_n$ in order to be able to forecast future integrated volatility using (3). In this section we will give a complete solution to the filtering task, while suggesting a number of, not entirely satisfactory, estimators of the parameters. Our treatment follows quite closely a recent paper on this problem by Shephard and Wong (1999) who give a more detailed treatment of this topic.

Parameter estimation for SV models is known to be a difficult problem (see the reviews in Shephard (1996) and Ghysels, Harvey, and Renault (1996)). Our preferred method of analysis would be via the likelihood function, which equals

$$f(y; \mu, \beta, \rho, \lambda, \psi) = \int f(y_1, ..., y_T | \sigma_1^2, ..., \sigma_T^2; \mu, \beta, \rho) f(\sigma_1^2, ..., \sigma_T^2; \lambda, \psi) \mathrm{d}\sigma_1^2, ..., \mathrm{d}\sigma_T^2,$$

where ψ is the vector of parameters indexing the density of the BDLP. In general we can evaluate

$$f(y_1, ..., y_T | \sigma_1^2, ..., \sigma_T^2; \mu, \beta, \rho) = \prod_{n=1}^{T} f(y_n | \sigma_n^2; \mu, \beta, \rho),$$

but do not know the explicit form of $f(\sigma_1^2, ..., \sigma_T^2; \lambda, \psi)$, and so we cannot hope to solve for $f(y; \mu, \beta, \rho, \lambda, \psi)$ or use an importance sampler to estimate the likelihood function. The results in Section 2.5 mean that we can easily simulate from $f(\sigma_1^2, ..., \sigma_T^2; \lambda, \psi)$, but 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.

Section 4 delivers a great number of moments which we could use to estimate the model. In particular for OU processes we can find a simple expression for the spectrum of squared returns which can be used in a Whittle likelihood to estimate the continuous time model using $y_1, ..., y_T$. We have not performed extensive experiments to assess the effectiveness of such an estimator, nor alternative estimating equation style solutions. Instead, this section will focus on three inferential themes. First, we use the linear structure of the model to estimate parameters and perform filtering. This will be based on Kalman filtering (KF) and will be suboptimal in both cases, but will have close links to the GARCH model. Second, we use a particle filter to estimate via simulation the optimal filtering quantities, $\mathrm{E}(\sigma_{n+1}^2 | \mathcal{F}_n)$, and compare them to the simpler best linear ones achieved by the KF. Third, we design a Markov chain Monte Carlo (MCMC) technique to perform Bayesian inference on the parameters of the model. This last method is computationally burdensome.

In order to focus on the crucial aspects of the problem we will first deal with a discrete approximation to the SV model, before briefly discussing the continuous time generalisation which raises few new issues but is more complicated. In particular we will show that two of our three inferential approaches generalise immediately to the continuous case, while the third partially does. Our discrete time approximation to the continuous time OU based SV model is built using an Euler style approximation, with σ_n^2 taken as $\Delta\sigma^2(n\Delta)$,

$$y_n = \mu\Delta + \beta\sigma_n^2 + \rho\left(\eta_n - \xi\Delta\right) + \varepsilon_n\sigma_n, \qquad n = 1, ..., T = [S/\Delta], \tag{46}$$

with $\{\varepsilon_n\}$ being a sequence of independent standard normal variables and

$$\sigma_{n+1}^2 = e^{-\Delta\lambda}\sigma_n^2 + \left(1 - e^{-\Delta\lambda}\right)\eta_n, \qquad \text{and} \qquad E(\eta_n) = E(\sigma_n^2) = \xi\Delta. \tag{47}$$

Throughout the

$$\eta_n = \Delta \left[\frac{\dot{z}\left\{\lambda\Delta\left(n+1\right)\right\} - \dot{z}\left(\lambda\Delta n\right)}{\lambda\Delta} \right] \stackrel{\mathcal{L}}{=} \lambda^{-1}\dot{z}\left(\lambda\Delta\right),$$

are independent and identically distributed strictly positive random variables, stochastically independent from $\{\varepsilon_n\}$. Typically η_n will have a very skew density and so volatility will rise much more sharply than it decreases. For ease of exposition we will set $\Delta = 1$ throughout this section. We will assume that we can evaluate the density of η_n and simulate from it.

5.2 Linear estimator

In the case where $\mu = \beta = \rho = 0$ we can transform the discrete time model (46) and (47) into a linear state space representation (the more general case of when we do not constrain μ, β, ρ is dealt with in Shephard and Wong (1999)). In particular

$$y_n^2 = \sigma_n^2 + u_n, \qquad \text{where} \qquad u_n = \sigma_n^2\left(\varepsilon_n^2 - 1\right). \tag{48}$$

We notice that $\{u_n\}$ is a zero mean white noise process with

$$\sigma_u^2 = \text{Var}(u_n) = \text{Var}(\varepsilon_n^2)\text{E}(\sigma_n^4) = 3\left\{\frac{1-e^{-\lambda}}{1+e^{-\lambda}}\text{Var}(\eta_n) + \xi^2\right\}$$

so long as $\text{Var}(\eta_n) < \infty$. Further, this equation can be combined with

$$\sigma_{n+1}^2 = \left(1-e^{-\lambda}\right)\xi + e^{-\lambda}\sigma_n^2 + v_n, \qquad \text{where} \qquad v_n = \left(1-e^{-\lambda}\right)\{\eta_n - \xi\}, \tag{49}$$

to yield a linear state space form (see, for example, Harvey (1981))[2]. Hence the best linear estimator, written $a_{n+1|n}$, of σ_{n+1}^2 using $y_1^2, ..., y_n^2$ is given by the KF recursion

$$a_{n+1|n} = e^{-\lambda}\left\{a_{n|n-1} + \frac{p_{n|n-1}}{p_{n|n-1}+1}\left(y_n^2 - a_{n|n-1}\right)\right\} + \left(1-e^{-\lambda}\right)\xi,$$

and its associated mean square error is $\sigma_u^2 p_{n+1|n}$, where

$$p_{n+1|n} = \frac{e^{-2\lambda}}{p_{n|n-1}+1} + \frac{\sigma_v^2}{\sigma_u^2}, \qquad \text{where} \qquad \sigma_v^2 = \text{Var}(v_n).$$

As $\lambda > 0$, it follows that $p_{n+1|n}$ converges to a steady state p, implying for moderately large n

$$a_{n+1|n} = e^{-\lambda}\frac{p}{p+1}y_n^2 + e^{-\lambda}\frac{1}{p+1}a_{n|n-1} + \left(1-e^{-\lambda}\right)\xi,$$

which coincides with the GARCH filter of the conditional volatility.

Sensible initial values for the KF are $a_{1|0} = \xi$ and $\sigma_u^2 p_{1|0} = \text{Var}(\sigma_n^2)$. A simple way of estimating the parameters of this model is to use a quasi-likelihood based around the output from the KF

$$\log L_Q = -\frac{1}{2}\sum_{n=1}^{T}\log p_{n|n-1} - \frac{1}{2}\sum_{n=1}^{T}\frac{1}{p_{n|n-1}}\left(y_n^2 - a_{n|n-1}\right)^2. \tag{50}$$

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.

[2]An interesting alternative is to allow $\text{Var}(\varepsilon_n^2)$ to be a free parameter, instead of imposing it to be 3, to be estimated from the data at the same time as λ, ξ and $\text{Var}(\eta_n)$.

To illustrate these methods, and ones we develop later in the paper, we analyze the daily observations of weekday close exchange rates $\{r_n\}$ for the Sterling, Deutschmark (DM), Japanese Yen and Swiss Franc (ChF) exchange rates, all against the US Dollar from 1/10/81 to 28/6/85. The sample size is $T = 946$. This data set has been previously analysed using a log-normal SV model by Harvey, Ruiz, and Shephard (1994) and Kim, Shephard, and Chib (1998). The mean-corrected returns will be computed as

$$y_n = 100 \times \left\{ (\log r_n - \log r_{n-1}) - \frac{1}{n} \sum_{i=1}^{T} (\log r_i - \log r_{i-1}) \right\}. \qquad (51)$$

	Constrained QML			
Series	$e^{-\lambda}$	ξ	γ	Quasi-L
Sterling	.984 (.960,.993)	.534 (.0005,.562)	.00538 (.00000,.0144)	-1566.6
DM	.970 (.941,.987)	.401 (.213,.417)	.0125 (.00208,.0272)	-1207.7
Yen	.992 (.975,.997)	.323 (.142,.394)	.00316 (.000306,.0141)	-971.50
ChF	.786 (.704,.899)	.345 (.110,.425)	.00752 (.000547,.0189)	-1827.5
	Unconstrained QML (estimating $Var(\varepsilon_t^2)$)			
Series	$e^{-\lambda}$	ξ	γ	Quasi-L
Sterling	.985 (.925,.992)	.559 (.325,.858)	.00580 (.00384,.0637)	-1566.6
DM	.969 (.876,.987)	.503 (.397,.656)	.0222 (.0130,.146)	-1206.8
Yen	.992 (.909,.997)	.344 (.220,.522)	.00368 (.00229,.0997)	-971.48
ChF	.584 (.186,.817)	.656 (.471,.887)	.0265 (.0125,.0711)	-1759.2

Table 1: *Constrained and unconstrained (treating $Var(\varepsilon_n)$ as unknown) QML estimates of the parameters of the linear SV model, reporting results for the implied inverse Gaussian model. The 90 percent confidence intervals are constructed by a parametric bootstrap using an inverse Gaussian assumption on η_n. We used 250 replications.*

For each of these series we used the QML method given in equation (50) to estimate the parameters of the model. These are given in Table 1, together with 90 per cent confidence intervals constructed by a simple parametric bootstrap using an inverse Gaussian assumption on $\{\eta_n\}$ and 250 replications. For ease of later comparison we have mapped the parameters into those for an inverse Gaussian model for η_n, so that the $\mathrm{E}(\eta_n) = \xi$ and $\mathrm{Var}(\eta_n) = \xi^3/\gamma$. These show the Yen series has the most persistence, but the lowest level of average volatility. The ChF has very little long term volatility clustering. This is picked up by the estimator, and is also very clear from the correlogram of $|y_n|$ given in Table 2. The memory in this series seems to last only a very few lags.

Table 2 also shows the estimated values of $\sqrt{a_{n|n-1}}$ graphed with $|y_n|$. It shows quite a jagged picture with occasional sharp up-moves in the series. The ChF is an extreme series for its' filtered volatility is very jagged. The normed series, $\left\{|y_n|/\sqrt{a_{n|n-1}}\right\}$, seems satisfactory for each of these series. The model basically takes out the entire structure of the volatility clustering.

5.3 How good is the linear method?

It is only possible to measure the effectiveness of the KF methods by making assumptions about the distribution of $\{\eta_n\}$ and then comparing the efficient estimator of σ_n^2, that is $\mathrm{E}(\sigma_n^2|\mathcal{F}_{n-1})$, with the best linear one computed by the KF, $a_{n|n-1}$. We employ the ASIR particle filtering of Pitt and Shephard (1999a) to carry out this task.

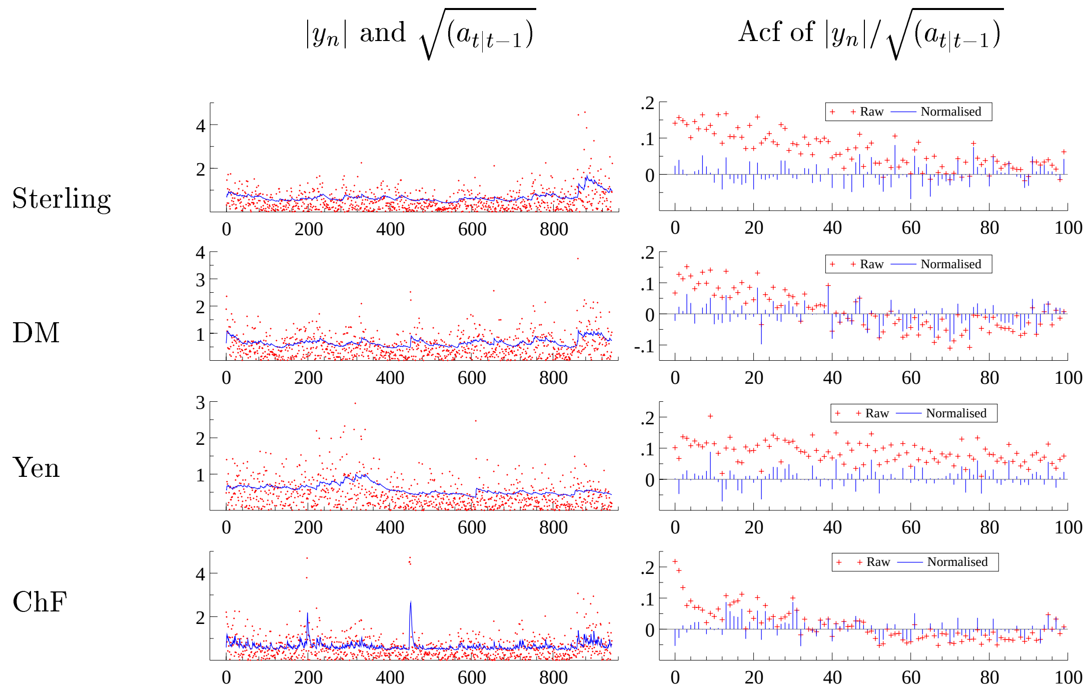

Table 2: *QML estimation of the filtered volatility and raw and normalised correlograms for* $|y_n|$.

We use the notation $f(\sigma^2_{n+1}|\sigma^2_n)$ to denote the Markov evolution of the unobserved log-volatilities over time of the discrete time model. The particle filter has the following basic structure. The density of $\sigma^2_n|\mathcal{F}_n$ is approximated by a sample $\sigma^2_{1,n}, ..., \sigma^2_{M,n}$. The particle filter regenerates these points into an approximate sample from $\sigma^2_{n+1}|\mathcal{F}_{n+1}$ by sampling from

$$\widehat{f}(\sigma^2_{n+1}|\mathcal{F}_{n+1}) \propto f\left(y_{n+1}|\sigma^2_{n+1}\right)\sum_{k=1}^{M} f(\sigma^2_{n+1}|\sigma^2_{k,n}). \tag{52}$$

This is carried out by sampling k with probability proportional to $f\left(y_{n+1}|\mu^k_{n+1}\right)$, where $\mu^k_{n+1} = \mathrm{E}(\sigma^2_{n+1}|\sigma^2_{k,n})$, to produce a sample we write as $k^1, ..., k^R$. We then draw $\sigma^2_{j,n+1} \sim \sigma^2_{n+1}|\sigma^2_{k^j,n}$, for $j = 1, ..., R$. The resulting population of particles are given weights proportional to

$$w_j = \frac{f\left(y_{n+1}|\sigma^2_{j,n+1}\right)}{f\left(y_{n+1}|\sigma^2_{k^j,n+1}\right)}, \qquad \pi_j = \frac{w_j}{\sum_{i=1}^R w_i}, \qquad j = 1, ..., R.$$

We resample this population with probabilities $\{\pi_j\}$ to produce a sample of size M, $\sigma^2_{1,n+1}$, ..., $\sigma^2_{M,n+1}$. This sample is approximately from $\sigma^2_{n+1}|\mathcal{F}_{n+1}$. In this way we update the sample at each time step through the entire sample, $n = 1, 2, ..., T$. We can estimate $\mathrm{E}\left(\sigma^2_{n+1}|\mathcal{F}_n\right)$ by

$$e^{-\lambda}\frac{1}{M}\sum_{j=1}^{M}\sigma^2_{j,n} + \left(1 - e^{-\lambda}\right)\xi.$$

In practice when we applied the ASIR particle filter in this paper we have taken $M = 10,000$ and $R = 3M$, unless explicitly noted.

To compare the KF solution and the computationally fully efficient method we conducted a simulation experiment, based on assuming the $\{\eta_n\}$ to be $IG(\xi, \gamma)$ with the parameters initially calibrated using the estimated parameters from the exchange rate returns. We simulated the inverse Gaussian SV model for a sequence of T observations one hundred times, each time recording the true volatility. We then compared the KF best linear estimator of the volatility, $a_{n|n-1}$, against particle filter based estimates (varying M and taking $R = 3M$) of $\mathrm{E}(\sigma_n^2|\mathcal{F}_{n-1})$ using a mean square error criterium. It is, a priori, not clear which is better although we know that as $M \to \infty$ the particle filter will achieve the lowest possible mean square error.

Figure 1 gives the results for the first of our one hundred simulations based on the Sterling series. It shows the path of the true volatility as well as the data and the linear and particle filter estimate of the efficient estimate of the volatility. We can see the linear and efficient methods are very close with $M = 1000$, with the difference being negligible. This result will be reinforced by the results from our wider simulation. The picture also gives the impression that the filtered volatility considerably lags the true volatility. This is due to the noisy nature of the volatility process and provides an incentive for us to try to use higher frequency data.

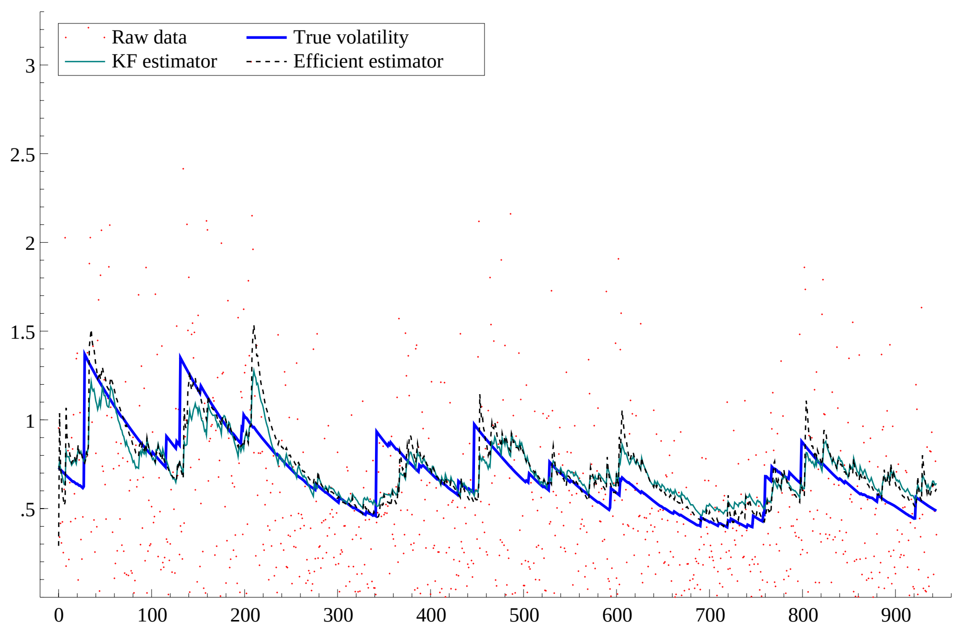

Figure 1: *Top graph has raw* $|y_n|$ *and true* σ_n, *as well as* $\sqrt{(a_{t|t-1})}$ *and* $E(\sigma_n|\mathcal{F}_{n-1})$. *Bottom graph has the associated correlograms of the* $|y_n|$ *and the normalised version using the linear volatility estimator and the efficient version.*

Table 3 gives the results from our 100 simulations, repeating the exercise for all four exchange rate series. It computes the average mean square error of the linear estimator and the particle filter for various values of M. Particle filter theory says that as $M \to \infty$ the particle filter will become optimal, but obviously the computational expense of this method increases with M linearly as well. Hence an important point is to ask if the potential increase in efficiency is actually worthwhile. Table 3 suggests this is not the case. If we only care about estimating the volatility accurately then the KF solution is remarkably close to being efficient. Only when $M > 1000$ is there evidence that the particle filter can achieve the precision of the KF.

		Particle filter: M				
	KF	25	100	250	1000	5000
Sterling	0.0868	0.160	0.115	0.101	0.0838	0.0793
DM	0.0398	0.0548	0.0455	0.0424	0.0426	0.0408
Yen	0.0136	0.0303	0.0217	0.0172	0.0135	0.0134
ChF	0.523	0.594	0.544	0.534	0.518	0.496

Table 3: *Mean square error for the various estimates of the volatility, computed using the Kalman filter and the particle filter. The particle filter depends upon M which is varied in this Table. The results are the average mean square errors taken over 100 independent replications.*

5.4 Bayesian approach

5.4.1 Methodology

In this subsection we will set $\mu = \beta = \rho = 0$, leaving a more general treatment to Shephard and Wong (1999). The aim of the analysis will be to understand the density of $\psi|y$ where $\psi = \left(\sigma_1^2, ..., \sigma_T^2, \theta\right)$ with $\theta = (\xi, \gamma, \lambda)$. We do this by using Markov chain Monte Carlo (MCMC) methods to draw samples from this joint density (see, for example Gilks, Richardson, and Spiegelhalter (1996)).

The most immediately straightforward algorithm samples each of the volatilities. It has the following detailed structure:

1. Initialize $\sigma_1^2, ..., \sigma_T^2$ and θ.
2. Sample σ_n^2 from $\sigma_n^2|\psi\backslash\sigma_n^2, y$, $n = 1, ..., T$.
3. Perform a Metropolis update on $\theta|y, \sigma_1^2, ..., \sigma_n^2$.
4. Goto 2.

Cycling through 2 to 3 is a complete sweep of this (single move) sampler. The MCMC sampler will require us to perform many thousands of sweeps to generate samples from $\theta, \sigma_1^2, ..., \sigma_T^2|y$.

Shephard and Wong (1999) has shown that even though it is possible to produce quite good samplers for drawing from step 2 of this procedure, in effect sampling from $\sigma_1^2, ..., \sigma_T^2|y, \theta$, the overall performance of the sampler is very poor as knowing $\sigma_1^2, ..., \sigma_T^2$ basically determines λ — that is we are over conditioning. To see this note that $e^{-\lambda} \leq \min_n \sigma_n^2/\sigma_{n-1}^2$. Hence the sampler is unable to move speedily through the sample space. This very unfortunate effect seems inevitable for this type of parameterisation.

The above problems are removed if we focus on a rotation of ψ. We design a method to sample from $\psi^*|y$ where

$$\psi^* = \sigma_1^2, \eta_2, \eta_3, ..., \eta_T, \theta.$$

The problem that we will now have is that we will lose some of the computationally convenient conditional structure of the model. As a result the new sample has the following form

1. Initialize $\sigma_1^2, \eta_2, ..., \eta_T$ and θ.
2. Draw $k \sim Poisson(\vartheta)$. Replace

$$\eta^o = (\eta_2, \eta_3, ..., \eta_{n-1}, \eta_n, \eta_{n+1}, ..., \eta_{n+k}, \eta_{n+k+1}, ..., \eta_T)$$

with

$$\eta^w = \left(\eta_2, \eta_3, ..., \eta_{n-1}, \eta_n^w, \eta_{n+1}^w, ..., \eta_{n+k}^w, \eta_{n+k+1}, ..., \eta_T\right).$$

The suggested replacements $\eta_n^w, \eta_{n+1}^w, ..., \eta_{n+k}^w$ are i.i.d. draws from $\{\eta_{t+j}|\theta, j = 0, 1, ..., k\}$. Then we accept the proposed move with probability

$$\min\left\{1, \frac{f(y_{n+1}, ..., y_T|\eta^w, \sigma_1^2, \theta)}{f(y_{n+1}, ..., y_T|\eta^o, \sigma_1^2, \theta)}\right\}.$$

We repeat this for $n = 1, ..., T$.

3. Perform a Metropolis update on $\theta|y, \sigma_1^2, \eta_2, \eta_3, ..., \eta_T$.

4. Goto 2.

Cycling through 2 to 3 is a complete sweep of this block sampler, with blocks of average size of $\vartheta + 1$. We have experimented with different selections for ϑ, but have used 100 throughout the work reported here. In general this sampler has the deeply unsatisfactory property that a single sweep takes, potentially, $O\left(T^2\right)$ computations due to step 2. However, as $\lambda > 0$ so the effect of changing the $\{\eta_n\}$ is guaranteed to die out making this method in practice $O(T)$. However, the results we present in a moment for our exchange series each took around one day of computer time on a fast PC.

Sampling from $\theta|y, \sigma_1^2, \eta_2, \eta_3, ..., \eta_T$ is not particularly difficult for

$$f(y, \sigma_1^2, \eta_2, \eta_3, ..., \eta_T|\theta) = f(y|\sigma_1^2, \eta_2, \eta_3, ..., \eta_T, \lambda) f(\sigma_1^2, \eta_2, \eta_3, ..., \eta_T|\xi, \gamma),$$

has a simple structure. In particular, sampling from $\lambda|y, \sigma_1^2, \eta_2, \eta_3, ..., \eta_T, \xi, \gamma$ is straightforward now. This is discussed in Shephard and Wong (1999).

We will assume that under the joint prior λ, ξ, γ are independent, with the following structure.

1. We let $e^{-\lambda} \sim Beta\left\{\lambda^{(1)}, \lambda^{(2)}\right\}$, following earlier work by Kim, Shephard, and Chib (1998). This structure has the advantage that it enforces stationarity on the volatility model (which we require). In our work we have set $\lambda^{(1)} = 40$ and $\lambda^{(1)} = 3$. Alternative priors could also be used (e.g. Chib and Greenberg (1994) and Marriott and Smith (1992)).

2. The average level of volatility $\xi \sim Gamma(0.6, 0.6/0.5)$. This means the prior mean is 0.6 and variance is 2. The former number is taken from previous work on volatility, while the later means this prior is quite flat.

3. Initial work on a number of speculative assets suggested γ should be in the range of 0.002 and 0.1. In our work we will allow $\gamma \sim Gamma(4, 4 \times 50)$, which gives a mean of 0.02 and a standard deviation of 0.01.

Our experiments suggest that our analysis is most sensitive to our prior on λ, which is also the hardest to specify as its meaning is less immediate than the other parameters.

5.4.2 Empirical results

Initially we will focus on the daily returns on the Sterling/Dollar exchange rate. We used 100,000 sweeps of our sampler, discarding the first 5,000 sweeps as burn-in. The latter number was determined by some previous experiments on simulated and real datasets. The sample paths of the draws for the parameters are given in Figure 4, together with associated histograms and correlograms of the sweeps. Broadly the posteriors for ξ and γ are reasonably symmetric, while the posterior on $e^{-\lambda}$ is skewed. The correlograms indicate very little autocorrelation in the $\xi|y$ sequence, while $\gamma|y$ has some correlation at 300 lags. The most dependent draws are for $e^{-\lambda}$, but even then the performance is not really poor. Overall these results are not unsatisfactory

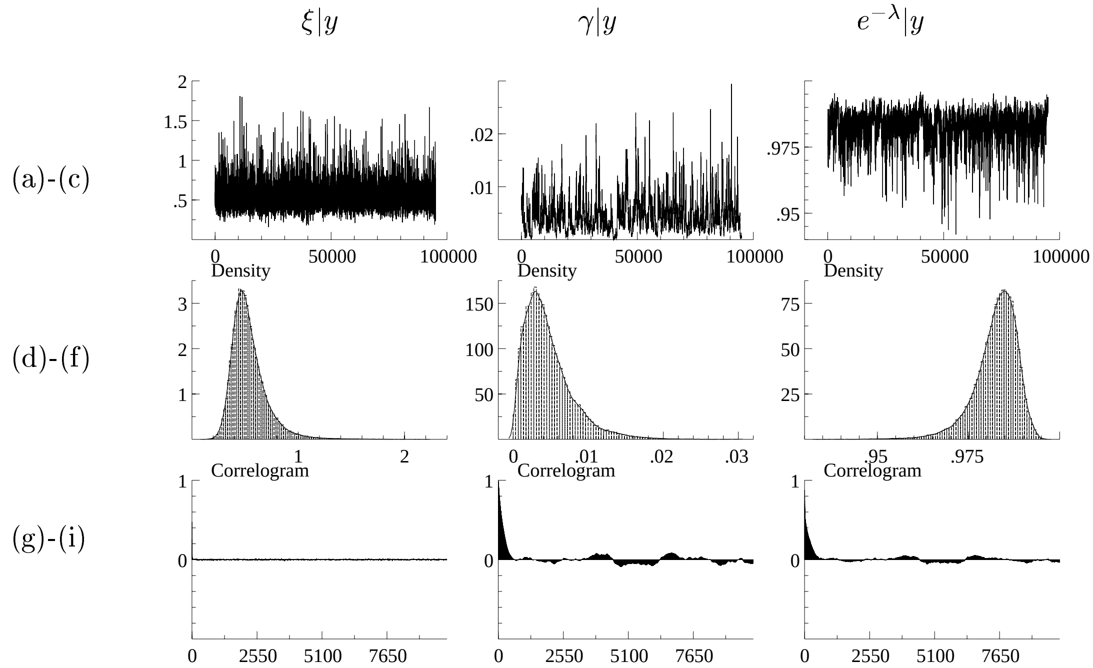

Table 4: *MH disturbance sampler for the Sterling series. Graphs (a)-(c): simulations against iteration. Graphs (d)-(f): histograms of marginal distribution. Graphs (g)-(i): corresponding correlograms for simulation. In total 102,000 iterations were drawn, discarding the first 2,000.*

for such a complicated model and are broadly inline with all but one of the samplers reported in Kim, Shephard, and Chib (1998) for the non-normal SV model.

Table 5 gives some information on the fit of the model. It gives the estimated log-likelihood at the mean of the posterior distribution for the parameters — these are considerably higher than the Gaussian GARCH model likelihoods recorded for this series in Kim, Shephard, and Chib (1998). The main focus of the table is on the diagnostics which gives a standardised third and fourth moment of the $r_n = \Phi^{-1}(\widehat{v}_n)$, where $\widehat{v}_n$ is an estimate made by the particle filter of the one-step ahead distribution function of the $n - th$ observation, $F(y_n|\mathcal{F}_{n-1})$, which should be uniform if the model is fitting well. Overall the kurtosis of the model is satisfactory, but the innovations are slightly positively skew with a longer right hand tail. These observations are not very clear from the associated QQ plot which is given in Figure 2. In this picture the darkest line is a 45 degree line, while we also give the QQ plots of the original data and the transformed version $\widehat{v}_n$

The Box-Ljung statistic of the transformed innovations $r_n = \Phi^{-1}(2\left|\widehat{v}_n - \frac{1}{2}\right|)$ is very good for this series, while the correlogram of the absolute values of the raw data and innovations is given in Figure 2. These innovations, suggested by Kim, Shephard, and Chib (1998), focus on misspecification in the volatility in the model. The correlograms of these functions of the data show how well the model deals with the short lags in the volatility clustering in the data, but at longer lags there are some quite large spikes in the correlogram.

We now report results for the three other exchange rates against the US Dollar. We use exactly the same algorithm and prior distributions in this analysis. In Table 5 we give results for each of the three series, while the results are summarised as follows.

- *DM against the US Dollar.* The correlogram and QQ plots of the innovations and raw

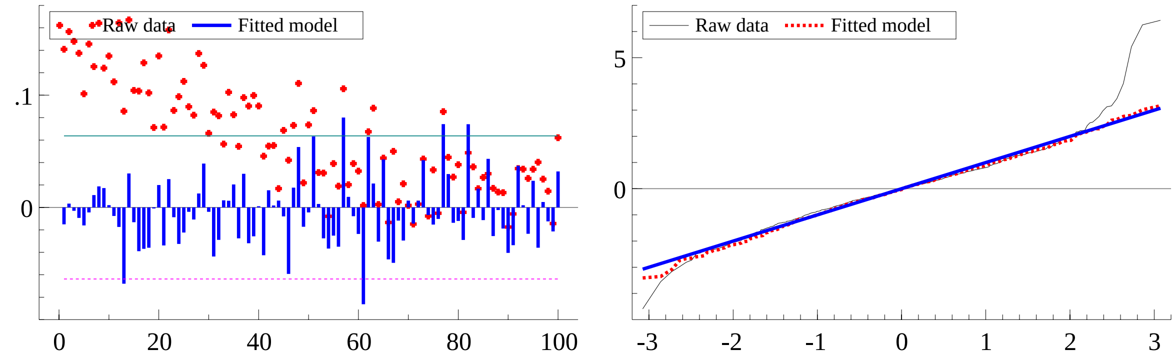

Figure 2: *Correlogram and QQ plot for raw and fitted Sterling series.*

data are given in Figure 3. The QQ plot and the summary statistics on normality given in Table 5 suggest the distributional fit of the model is good, while the BL statistic is a slight cause for concern. As we are using 30 lags an overall statistic of 39 is not very large, there are some quite large spikes in the correlogram. However, the correlogram at longer lags does not indicate any particular failure in the fit of the model. Overall this data set seems to provide some support for the model structure.

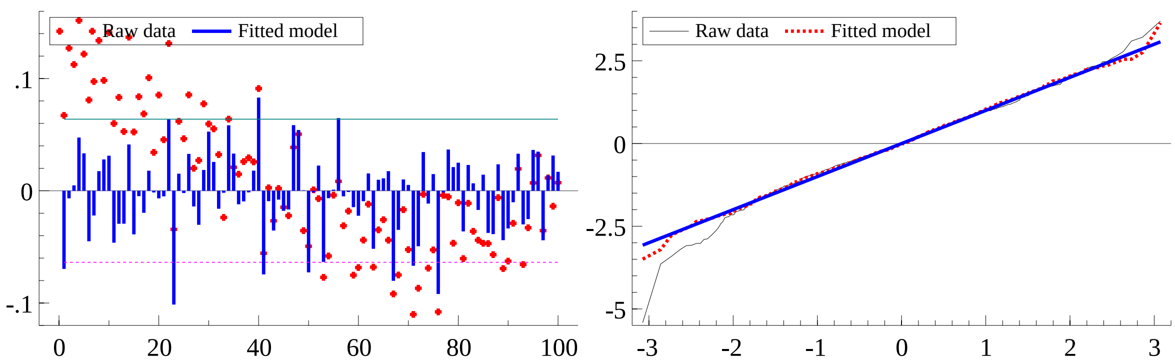

Figure 3: *Correlogram and QQ plot for raw and fitted DM series.*

- *Yen against the US Dollar.* Table 5 reports results from the fitting of the Yen series, giving a medium level of persistence to volatility shocks. Overall the normality statistic is slightly high, while the BL statistic is not overly worrying. The more detailed diagnostics are given in Figure 4, which suggests some failure in the left hand tail of the fitted innovations from the model and a reasonably well behaved correlogram. Overall the model is again reasonably satisfactory.

- *ChF against the US Dollar.* The raw data of the ChF has quite heavy tails which are mostly dealt with very well by the model. Both the overall normality and BL statistics seem very satisfactory. The detailed diagnostic Figure 5 again suggests the model transforms unusual behaving returns data into innovations which are perfectly satisfactory.

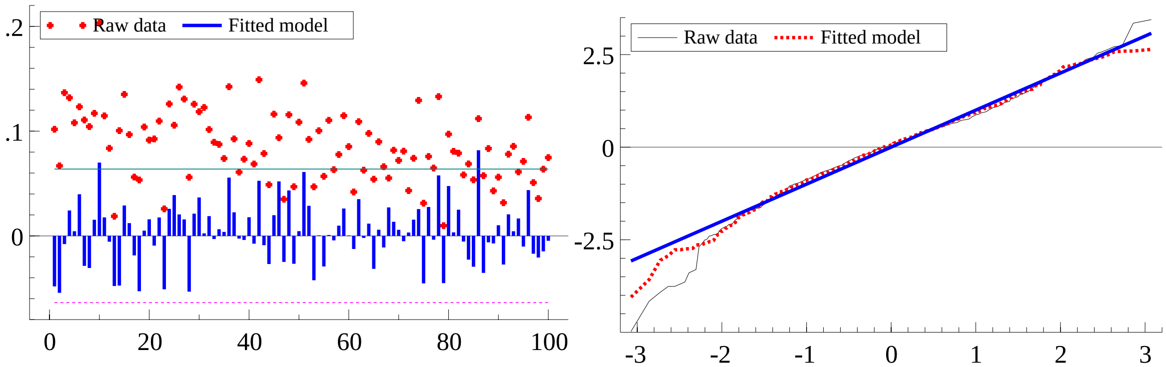

Figure 4: *Correlogram and QQ plot for raw and fitted Yen series*

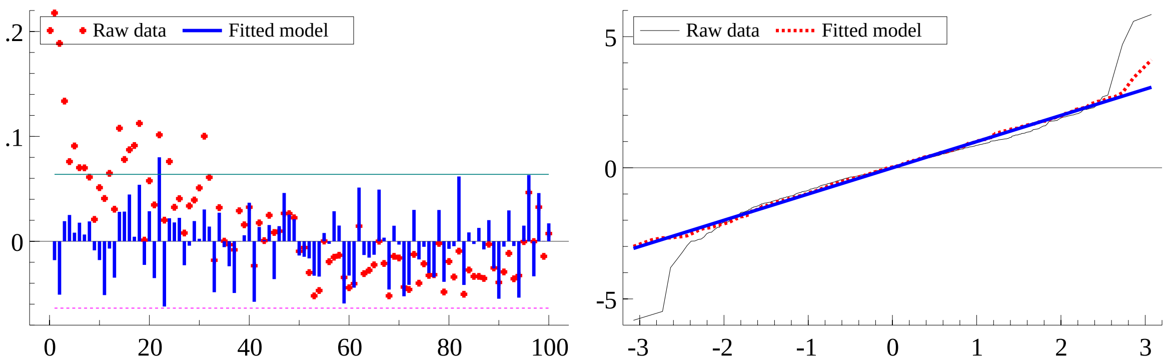

Figure 5: *Correlogram and QQ plot for raw and fitted ChF series*

5.5 Continuous time versions

5.5.1 Linear approach

All of the above arguments go through when we think of the linear SV model in continuous time rather than the discrete time version given in (46) and (47). This is because when $\mu = \beta = \rho = 0$ then if $\{\varepsilon_n\}$ are i.i.d. standard normal then we can write

$$y_n^2 = \sigma_n^2 + u_n, \qquad \text{where} \qquad u_n = \sigma_n^2 \left(\varepsilon_t^2 - 1\right),$$

and σ_n^2 is given in (7). Again let $\Delta = 1$, then we know that

$$\sigma_{n+1}^2 = \lambda^{-1} \left\langle \left[\grave{z}\left\{\lambda\left(n+1\right)\right\} - \grave{z}\left(\lambda n\right) + \sigma^2(n) - \sigma^2\left\{(n+1)\right\}\right] \right\rangle.$$

Hence this time our state, in the state space representation, will have to remember the four quantities $\grave{z}\left\{\lambda\left(n+1\right)\right\}$, $\grave{z}\left(\lambda n\right)$, $\sigma^2(n)$ and $\sigma^2\left\{(n+1)\right\}$, in order to construct σ_{n+1}^2. However, we will now see that these quantities are themselves Markov. First, using (9), we have that

$$\begin{aligned} s_{n+1} &= \left[\begin{array}{c} \sigma^2\left\{(n+1)\right\} \\ z\left\{\lambda\left(n+1\right)\right\} \end{array}\right] = \grave{\kappa}_1 \left(\begin{array}{c} 1 - e^{-\lambda} \\ \lambda \end{array}\right) + \left(\begin{array}{cc} e^{-\lambda} & 0 \\ 0 & 1 \end{array}\right) s_n + v_n \\ &= m + T^* s_n + v_n, \end{aligned}$$

where

$$v_n = \left\{ \begin{array}{c} e^{-(n+1)} \int_n^{n+1} e^{t\lambda} \mathrm{d}\overline{z}(\lambda t) \\ \int_n^{n+1} \mathrm{d}\overline{z}(\lambda t) \end{array} \right\}, \qquad \text{recalling that} \qquad \overline{z}(t) = \grave{z}(t) - \mathrm{E}\grave{z}(t), \tag{53}$$

is a zero mean, i.i.d., vector. Further

$$\mathrm{Var}\,(v_n) = \int_{\Delta n}^{\Delta(n+1)} \left\{ \begin{array}{cc} e^{-2(n+1)} e^{2t\lambda} & e^{-(n+1)} e^{t\lambda} \\ e^{-(n+1)} e^{t\lambda} & \lambda \end{array} \right\} \mathrm{d}\,(\grave{\kappa}_2 \lambda t) = \grave{\kappa}_2 \left\{ \begin{array}{cc} \frac{1}{2}\left(1 - e^{-2\lambda}\right) & \left(1 - e^{-\lambda}\right) \\ \left(1 - e^{-\lambda}\right) & \lambda \end{array} \right\}.$$

So if we write

$$\alpha_{n+1} = \begin{pmatrix} s_{n+1} \\ s_n \end{pmatrix} = \begin{pmatrix} m \\ 0 \end{pmatrix} + \begin{pmatrix} T^* & 0 \\ I & 0 \end{pmatrix} \alpha_n + \begin{pmatrix} v_n \\ 0 \end{pmatrix},$$

and

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

this model can be placed into a linear state space form and again the KF provides the best linear estimator, written $a_{n+1|n}$, of σ_{n+1}^2 using $y_1^2, ..., y_n^2$ as well as a quasi-likelihood function via (50).

5.5.2 Particle filter approach

The particle filter approach is straightforward to generalise to the continuous time case having established in the previous subsection that $\{\alpha_n\}$ is Markov. We now use samples to represent the density of $f(\alpha_n|\mathcal{F}_n)$ and then act as before realising that α_n determines σ_n^2 and so $f(y_n|\sigma_n^2)$. The only new task is to simulate from the i.i.d. $\{v_n\}$ given in (53). But precisely this simulation problem is solved by equation (34) which gives quickly converging series representations for such integrals. In practice we have found very little difference between the filter based on the discrete and continuous time solutions in practical SV problems.

5.5.3 Bayesian estimation

The MCMC based discrete time Bayesian analysis we presented above does not entirely generalise to the continuous time model. If we define $\psi^* = \sigma_1^2, v_2, v_3, ..., v_T$, then we can sample from $\psi^*|y$ as we know how to sample from the i.i.d. $\{v_n\}$ and this is all we need for the MCMC method outlined above. However, generalising this argument to learn about the parameters in the model seems difficult as this requires knowledge of $f(\alpha_{n+1}|\alpha_n)$ which we do not have. Hence the construction of methods to carry out an exact likelihood treatment of these models is still unresolved.

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 (1997). 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

$$x^*(t) = \int_0^t \sigma(s) dw(s), \tag{54}$$

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 (54) 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

$$\{x^*(t)\}_{t\geq 0} \overset{\mathcal{L}}{=} \{b([x^*](t))\}_{t\geq 0} \tag{55}$$

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

$$\mathrm{E}\{e^{i(f\bullet[x^*]+g\bullet b)}\} = \mathrm{E}\{e^{if\bullet[x^*]}\}\mathrm{E}\{e^{ig\bullet b}\}. \tag{56}$$

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

6.2 Pricing

6.2.1 Non-arbitrage

We show non-arbitrage under the model

$$x^*(t) = x_0^*(t) + \beta\sigma^{2*}(t) + \rho\bar{z}(\lambda t) \tag{57}$$

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

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

Once again we assume w and $\grave{z}$ are independent, while we write $\{\mathcal{F}_t\}_{t\geq 0}$ to represent the filtration generated by the pair of processes $(w, \grave{z})$. 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 $\grave{z}$ over the time interval $[0, T]$, let $\phi = \beta + \frac{1}{2}$, and let θ' be the solution to the equation

$$\grave{\kappa}(\rho + \theta') - \grave{\kappa}(\theta') = \xi\rho \tag{58}$$

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

$$u^*(t) = -\phi x_0^*(t) - \frac{1}{2}\phi^2\sigma^{2*}(t) + \theta'\bar{z}(\lambda t) - \lambda t\bar{\kappa}(\theta') \tag{59}$$

and where $\bar{\kappa}(\theta) = \grave{\kappa}(\theta) - \xi\theta$ is the cumulant function corresponding to the Lévy process $\bar{z}$, i.e. the cumulant function of $\bar{z}(1)$. Note that equation (58) may be reexpressed as

$$\bar{\kappa}(\rho + \theta') = \bar{\kappa}(\theta') \tag{60}$$

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

(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 $\grave{z}(1) \sim IG(\delta, \gamma)$. Then

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

Seeking a solution to (58) is therefore equivalent to solving

$$\{1 - 2\theta/\gamma^2\}^{1/2} + \{1 - 2(\rho+\theta)/\gamma^2\}^{1/2} = 2 \tag{61}$$

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 (61) decreases monotonically from ∞ to $|\rho|\sqrt{2}/\gamma$. Consequently, (61) is solvable if and only if $|\rho| \leq \sqrt{2}\gamma$.

□

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 (1999) 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 $C\{K, x^*(S), S+T\}$ for the price at time S of a European call option on $x^*(t)$, with initial value $x^*(S)$, with strike price K and expiration date $T+S$ we have that

$$\begin{aligned}
C\{K, x^*(S), S+T\} &= E^Q\{x^*(S+T) - K\}^+ \\
&= \int_{R_+} BS\{K, x^*(S), \sigma^2_{S,T+S}, S+T\}\, dP(\sigma^2_{S,T+S}|\mathcal{F}_S)
\end{aligned}$$

where $\mathcal{F}_S$ is the available information at time S, $BS\left\{K, x^*(S), \sigma^2_{S,T+S}, S+T\right\}$ which denotes the Black-Scholes price of the option with initial value $x^*(S)$, strike price K and constant volatility $\sigma^2_{S,T+S}$, where

$$\sigma^2_{S,T+S} = \frac{1}{T}\left\{\sigma^{2*}(S+T) - \sigma^{2*}(S)\right\}.$$

In practice $\mathcal{F}_S$ will provide an estimate of $\grave{z}(\lambda S), \sigma^2(S)$, while we can unbiasedly estimate $C\{K, x^*(S), S+T\}$ simply by simulation for we can quickly draw many samples from $\sigma^2_{S,T+S}|\grave{z}(\lambda S), \sigma^2(S)$ using the series representations developed in Section 2 of this paper.

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 (1998b) 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 $\lambda(t) = \lambda\sigma^2(t) > 0$. 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 we model the current log-price as

$$x_\lambda^*(t) = \mu\tau_{N(t)} + \beta\sigma^{2*}\left\{\tau_{N(t)}\right\} + \frac{1}{\sqrt{\lambda}}\sum_{k=1}^{N(t)} y_k, \tag{62}$$

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 (62), if $\{y_i\}$ are assumed independent standard normal and $\sigma^{2*}(t) = \int_0^t \sigma^2(u)\mathrm{d}u$, $N(t)$ is a Cox process with random intensity $\lambda(t) = \lambda\sigma^2(t) > 0$, then

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

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 (1998b) and Rydberg and Shephard (1998a)).

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 $\sigma^2(t) = \left\{\sigma_1^2(t), ..., \sigma_q^2(t)\right\}$ via $q-$dimensional BDLPs $\grave{z}(t) = \{\grave{z}_1(t), ..., \grave{z}_q(t)\}$ as follows. The multivariate form of (14) is

$$\grave{k}(\theta) = \log \mathrm{E}\left[\exp\left\{-\langle\theta, \grave{z}(1)\rangle\right\}\right] = -\int_{R_+^q}\left(1 - e^{-\langle\theta,x\rangle}\right) W(\mathrm{d}x), \tag{63}$$

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 $\grave{z} = (\grave{z}_1, ..., \grave{z}_q)$ be a $q-$dimensional Lévy process with $\log \mathrm{E}\left[\exp\left\{-\langle\theta, \grave{z}(1)\rangle\right\}\right]$ as in (63). 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

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

Note that

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

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 3 *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 $\grave{z}_1(s)$ and $\grave{z}_2(s)$ are independent BDLP/IG-OU processes, while for $\phi \uparrow \infty$ the processes $\grave{z}_1(s)$ and $\grave{z}_2(s)$ 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 will be given elsewhere.

Consider a $q-$dimensional BDLP process $\grave{z}$ 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 3), 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

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

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 $\left\{a_j^*\right\}$, such that the law of u_j is that determined by the probability density b. Furthermore, for $s \in [0, 1]$ let

$$\tilde{z}(s) = \sum_{j=1}^\infty \mathbf{1}_{[0,s]}(r_j) G^{-1}(a_j^*) u_j \tag{64}$$

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 (64) converges a.s. and

$$\{\check{z}(s) : 0 \leq t \leq 1\} \stackrel{\mathcal{L}}{=} \{\tilde{z}(s) : 0 \leq t \leq 1\} \tag{65}$$

□

Furthermore we have

Proposition 6.2 If f_i, $i = 1, ..., d$, are positive and integrable functions on $[0,1]$ then

$$\int_0^1 f_i(s)\mathrm{d}\check{z}_i(s) \stackrel{\mathcal{L}}{=} \sum_{j=1}^{\infty} G^{-1}(a_j^*)u_{ij}f_i(r_j) \tag{66}$$

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

□

6.5 Multivariate SV models

A simple q-dimensional version of the SV model sets $x^*(t) = \left\{x_1^*(t), ..., x_q^*(t)\right\}$ where

$$x_i^*(t) = \beta_i \int_0^t \sigma_{q+1}(u)\mathrm{d}w_{q+1}(u) + \int_0^t \sigma_i(u)\mathrm{d}w_i(u).$$

Here $\beta_1, ..., \beta_q$ are unknown parameters and $\sigma_1, \sigma_2, ..., \sigma_{q+1}$ and $w_1, w_2, ..., w_{q+1}$ are $2(q+1)$ processes such that $\sigma_1, \sigma_2, ..., \sigma_{q+1}$ are mutually independent OU processes which are square integrable and stationary while $w_1, w_2, ..., w_{q+1}$ are mutually independent Wiener processes. We also assume that $\sigma_1, \sigma_2, ..., \sigma_{q+1}$ are independent of $w_1, w_2, ..., w_{q+1}$. The process $x^*(t)$ is a continuous q-dimensional local martingale. It constitutes a factor style model with a 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 $\sigma^2(t)$ is 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 as 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 ´cumulative´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

$$\mathrm{d}x(t) = -\lambda x(t)\mathrm{d}t + \sigma(t)\mathrm{d}w(t)$$

with $\sigma^2(t)$ an OU process as in (2).

8 Appendix

8.1 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

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

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

8.2 GIG Lévy density

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

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

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

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

and this shows that

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

is the Lévy density of z.

For $\overline{\lambda} > 0$ the expression for u follows from the convolution formula

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

where $\Gamma(\overline{\lambda}, \phi)$ is the gamma distribution with probability density

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

and corresponding Lévy density $\overline{\lambda} x^{-1} e^{-\phi x}$.

□

8.3 BDLP criterion

Proof of Lemma 2.1 By the definition of u we have

$$\begin{aligned}
\int_{0+}^{\infty} \min\left\{1, x^2\right\} u(x)\mathrm{d}x &= \int_1^{\infty}\int_{0+}^{\infty} \min\left\{1, x^2\right\} w(\tau x)\mathrm{d}x\mathrm{d}\tau \\
&= \int_1^{\infty}\int_{0+}^{\infty} \min\left\{1, \tau^{-2}y^2\right\} \tau^{-1} w(y)\mathrm{d}y\mathrm{d}\tau \\
&= \int_{0+}^{1} y^2 w(y)\mathrm{d}y \int_1^{\infty} \tau^{-3}\mathrm{d}\tau + \int_1^{\infty} w(y)\left(\int_1^{y} \tau^{-1}\mathrm{d}\tau + y^2\int_y^{\infty}\tau^{-3}\mathrm{d}\tau\right)\mathrm{d}y \\
&= \frac{1}{2}\int_{0+}^{1} y^2 w(\mathrm{d}y) + \int_1^{\infty} \log y w(y)\mathrm{d}y + \frac{1}{2}\int_1^{\infty} w(y)\mathrm{d}y.
\end{aligned}$$

In the latter expression the first and third integrals are finite since W is a Lévy measure and the second integral is finite by assumption (25). Hence u is a Lévy density, and differentiation of (26) shows that u and w are related as in (15), implying the validity of the remaining part of the conclusion of the Lemma.

□

8.4 Series representation of OU processes

To establish Proposition 2.1 and its Corollary we shall use

Lemma 8.1 Consider the infinite sum

$$\mathbf{s} = \sum_{i=1}^{\infty} W^{-1}(a_i)h_i \tag{67}$$

where $\{a_i\}$ and $\{h_i\}$ are two independent sequences of random variables with $a_1 < ... < a_i < ...$ as the arrival times of a Poisson process with intensity 1 while h_i, $i = 1, 2, ...$ are independent copies of a positive random variable h. Then

$$\bar{\mathrm{K}}\{\theta \ddagger \mathbf{s}\} = \mathrm{E}\left\{\bar{\mathrm{K}}\{\theta h \ddagger \check{z}\}\right\} \tag{68}$$

□

PROOF Adapted from Marcus (1987). Let $k_\tau = \#\{a_i < \tau\}$ and define

$$\mathbf{s}_\tau = \sum_{i=1}^{k_\tau} W^{-1}(a_i)h_i \tag{69}$$

Then k_τ follows the Poisson law with mean τ and, conditionally on $k_\tau = k$, we have, with $\mathcal{L}$ denoting probabilistic law,

$$\mathcal{L}\{(a_1, ..., a_k) \mid k_\tau = k\} = \mathcal{L}\{\tau(u_{(1)}, ..., u_{(k)})\} \tag{70}$$

where $u_{(1)}, ..., u_{(k)}$ are the order statistics of a random sample of size k from the uniform distribution on $[0, 1]$. Hence

$$\mathcal{L}\{s_\tau \mid k_\tau = k\} = \mathcal{L}\left\{\sum_{i=1}^{k} W^{-1}(\tau u_{(i)})h_i\right\} \tag{71}$$

or, equivalently,

$$\mathcal{L}\{s_\tau \mid k_\tau = k\} = \mathcal{L}\left\{\sum_{i=1}^{k} W^{-1}(\tau u_i)h_i\right\} \tag{72}$$

Therefore, for u a uniform random variable on $[0, 1]$ and writing $\overline{L}\{\theta \ddagger x\}$ for the Laplace transform of a positive random variable, we find that

$$\begin{aligned} \bar{\mathrm{L}}\{\theta \ddagger \mathbf{s}_\tau\} &= e^{-\tau}\sum_{k=0}^{\infty}\frac{\tau^k}{k!}\mathrm{E}\exp\left\{-\theta\sum_{i=1}^{k} W^{-1}(\tau u_i)h_i\right\} \\ &= \exp\left\langle\tau\left[\bar{\mathrm{L}}\{\theta \ddagger W^{-1}(\tau u)h\} - 1\right]\right\rangle \end{aligned} \tag{73}$$

As the next step we must consider the distribution of the random variable $y_\tau = W^{-1}(\tau u)h$. We find

$$P\{y_\tau \le x\} = P\{u \ge \tau^{-1}W^{+}(h^{-1}x)\} \tag{74}$$

or, equivalently,

$$P\{y_\tau \le x\} = \begin{cases} 1 - \tau^{-1}\mathrm{E}\{W^{+}(h^{-1}x)\} & \text{if} \quad W^{+}(h^{-1}x) \le \tau \\ 0 & \text{if} \quad W^{+}(h^{-1}x) > \tau \end{cases} \tag{75}$$

It follows that

$$\begin{aligned} \bar{\mathrm{L}}\{\theta \ddagger W^{-1}(\tau u)h\} &= \tau^{-1}\mathrm{E}\left\{-\int_{hW^{-1}(\tau)}^{\infty} e^{-\theta x}\mathrm{d}W^{+}(h^{-1}x)\right\} \\ &= \mathrm{E}\left\{\tau^{-1}\int_{hW^{-1}(\tau)}^{\infty}(1 - e^{-\theta x})\mathrm{d}W^{+}(h^{-1}x)\right\} + 1 \end{aligned} \tag{76}$$

Hence

$$\bar{\mathrm{K}}\{\theta \ddagger \mathbf{s}_\tau\} = \mathrm{E}\left\{\int_{hW^{-1}(\tau)}^{\infty}(1 - e^{-\theta x})\mathrm{d}W^{+}(h^{-1}x)\right\} \tag{77}$$

Now, letting $\tau \to \infty$ we obtain

$$\begin{aligned} \bar{\mathrm{K}}\{\theta \ddagger \mathbf{s}\} &= \mathrm{E}\left\{\int_{0}^{\infty}(1 - e^{-\theta x})\mathrm{d}W^{+}(h^{-1}x)\right\} \\ &= \mathrm{E}\left\{\int_{0}^{\infty}(1 - e^{-\theta h x})\mathrm{d}W^{+}(x)\right\} \\ &= \mathrm{E}\left\{\bar{\mathrm{K}}\{\theta h \ddagger z\}\right\} \end{aligned} \tag{78}$$

□

Proof of Proposition 2.1 From the lemma above, putting $h = \mathbf{1}_{[0,s]}(r)$ we find

$$\bar{\mathrm{K}}\{\theta \ddagger \tilde{z}(s)\} = s\bar{\mathrm{K}}\{\theta \ddagger \grave{z}(1)\} = \bar{\mathrm{K}}\{\theta \ddagger \grave{z}(s)\} \tag{79}$$

showing that for any $s \in [0,1]$ that $\grave{z}(s) \overset{\mathcal{L}}{=} \tilde{z}(s)$. A minor extension of this argument then establishes the full conclusion of the corollary.

□

Proof of Corollary 2.1 A key formula, holding for arbitrary Lévy processes z, states that

$$\mathrm{C}\{\zeta \ddagger f \bullet z\} = \int_0^\infty \mathrm{C}\{\zeta f(s) \ddagger z(1)\}\mathrm{d}s. \tag{80}$$

It follows that

$$\bar{\mathrm{K}}\left\{\theta \ddagger \int_0^1 f(s)\mathrm{d}\grave{z}(s)\right\} = \int_0^1 \bar{\mathrm{K}}\{\theta f(s) \ddagger \grave{z}(1)\}\mathrm{d}s = \bar{\mathrm{K}}\left\{\theta \ddagger \sum_{i=1}^{\infty} W^{-1}(a_i)f(r_i)\right\} \tag{81}$$

where the second equality is a consequence of Lemma 2.1.

□

8.5 Cumulants of integrals of the $x^*(t)$ process

In this subsection we derive formula (40). The distributional properties of x^* are embodied in the class of integrals of the form

$$f \bullet x^* = \int_0^\infty f(t)\mathrm{d}x^*(t)$$

where f is a deterministic real function. We determine the cumulant function of such integrals (when they exist). In particular, by suitable choice of f one obtains the cumulant functions of the multivariate marginal distributions of x^*. Noting that x^* is a local martingale, we interpret $f \bullet x^*$ as a stochastic integral (as defined, for instance, in Protter (1992)).

Since

$$\mathrm{d}x^*(t) = \sigma(t)\mathrm{d}w(t) + \mu + \beta\sigma^2(t)\mathrm{d}t + \rho\mathrm{d}\grave{z}(\lambda t) - \rho\xi\lambda\mathrm{d}t$$

we have

$$f \bullet x^* = (f\sigma) \bullet w + \beta \int_0^\infty f(t)\sigma^2(t)\mathrm{d}t + \rho f(\lambda^{-1}\cdot) \bullet \grave{z} + (\mu - \rho\xi\lambda)\int_0^\infty f(t)\mathrm{d}t$$

Now,

$$\mathrm{E}\{\exp(i\zeta(f\sigma) \bullet w)|\grave{z}(\cdot)\} = \exp\{-\frac{1}{2}\zeta^2 \int_0^\infty f^2(t)\sigma^2(t)\mathrm{d}t\}$$

and hence

$$\mathrm{E}\{\exp(i\zeta f \bullet x^*)\} = \mathrm{E}\{\exp G[f]\}\exp\left(i\zeta(\mu - \rho\xi\lambda)\int_0^\infty f(t)\mathrm{d}t\right)$$

where

$$G[f] = \int_0^\infty \{-\frac{1}{2}\zeta^2 f^2(t) + i\zeta\beta f(t)\}\sigma^2(t)\mathrm{d}t + i\zeta\rho\int_0^\infty f(\lambda^{-1}t)\mathrm{d}\grave{z}(t))$$

Furthermore, using the representation

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

we find, for an arbitrary function h

$$\int_0^\infty h(t)\sigma^2(t)\mathrm{d}t = I_0 + I_1$$

where

$$\begin{aligned} I_0 &= \int_0^\infty h(t)e^{-\lambda t}\int_{-\infty}^0 e^{\lambda s}\mathrm{d}\grave{z}(\lambda s)\mathrm{d}t \\ &= (\int_0^\infty e^{-\lambda t}h(t)\mathrm{d}t)\int_{-\infty}^0 e^{s}\mathrm{d}\grave{z}(s) \qquad (82) \\ &= (\int_0^\infty e^{-\lambda t}h(t)\mathrm{d}t)\int_0^\infty e^{-s}\mathrm{d}\grave{z}(s) \qquad (83) \end{aligned}$$

and

$$\begin{aligned} I_1 &= \int_0^\infty \int_s^\infty e^{-\lambda(t-s)}h(t)\mathrm{d}t\mathrm{d}\grave{z}(\lambda s) \\ &= \int_0^\infty \int_0^\infty e^{-\lambda u}h(s+u)\mathrm{d}u\mathrm{d}\grave{z}(\lambda s) \qquad (84) \\ &= \int_0^\infty \int_0^\infty e^{-\lambda u}h(\lambda^{-1}s+u)\mathrm{d}u\mathrm{d}\grave{z}(s) \qquad (85) \end{aligned}$$

It follows that

$$\mathrm{E}\{\exp(i\zeta f \bullet x^*)\} = \mathrm{E}\left\{\exp\left(\int_0^\infty H(s)\mathrm{d}\grave{z}(s)\right)\right\}\exp\left\{i\zeta(\mu-\rho\xi\lambda)\int_0^\infty f(t)\mathrm{d}t\right\} \qquad (86)$$

where

$$H(s) = -\int_0^\infty N(s,u)e^{-\lambda u}\mathrm{d}u + i\zeta\rho f(\lambda^{-1}s)$$

where

$$N(s,u) = \frac{1}{2}\zeta^2\{f^2(u)e^{-s} + f^2(\lambda^{-1}s+u)\} - i\zeta\beta\{f(u)e^{-s} + f(\lambda^{-1}s+u)\}$$

A key formula, holding for arbitrary Lévy processes z, states that

$$\mathrm{C}\{\zeta \ddagger f \bullet z\} = \int_0^\infty \mathrm{C}\{\zeta f(s) \ddagger z(1)\}\mathrm{d}s$$

Applying formula (80) to (86) we obtain formula (40).

8.6 Non-arbitrage

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

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

and here

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

so that

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

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

(ii) Note first that

$$\beta - \frac{1}{2}\phi^2 + (1-\phi)^2 = 0 \tag{87}$$

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

$$\mathrm{E}_{P'}\{y_t|\mathcal{F}_s\} = \mathrm{E}_P\{y_t d(T)/d(s)|\mathcal{F}_s\} = \mathrm{E}_P\{y_t d(t)/d(s)|\mathcal{F}_s\} \tag{88}$$

Hence

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

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 (87),

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

so that, in view of condition (60),

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

□

8.7 Trade-by-trade dynamics

Lemma 8.2 Let $N(t)$ be a Cox process with random intensity $\lambda(t) = \lambda\sigma^2(t) > 0$. 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 $\lambda \to \infty$ we have that $\tau_{N(t)} \xrightarrow{p} t$.
Proof: It suffices to show that for every $\varepsilon > 0$ we have that

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

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

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

Consequently

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

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

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

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

We obtain from Lemma 8.2 and the continuity of $\sigma^{2*}(.)$ that the limiting behaviour in the distribution of $x_\lambda^*(t)$, as $\lambda \to \infty$, is the same as that of

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

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

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

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

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

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

References

Andersen, T. G. and T. Bollerslev (1997). Intraday periodicity and volatility persistence in financial markets. *J. Empirical Finance 4*, 115–58.

Andersen, T. G. and T. Bollerslev (1998). Answering the skeptics: yes, standard volatility models do provide accurate forecasts. Unpublished paper: Kellogg School of Management, Northwestern University.

Ané, T. and H. Geman (1997). Transaction clock and asset price dynamics: Clark's model revisited. Unpublished paper.

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

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

Barndorff-Nielsen, O. E. (1999). Superposition of Ornstein-Uhlenbeck type processes. Unpublished paper: Department of Mathematics, Aarhus University.

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.

Black, F. (1976). Studies of stock price volatility changes. *Proceedings of the Business and Economic Statistics Section*, 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. North-Holland.

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

Chib, S. and E. Greenberg (1994). Bayes inference for regression models with ARMA(p,q) errors. *J. Econometrics 64*, 183–206.

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

Clark, P. K. (1973). A subordinated stochastic process model with fixed variance for speculative prices. *Econometrica 41*, 135–156.

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.

Dacorogna, M. M., U. A. Muller, R. B. Olsen, and O. V. Pictet (1997). Modelling short term volatility with GARCH and HARCH. In *Nonlinear Modelling of High Frequency Financial Time Series.* Chichester: Wiley.

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

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. *J. Econometrics 74*, 31–57.

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. and U. Keller (1995). Hyperbolic distributions in finance. *Bernoulli 1*, 281–99.

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 (1992). A permanent and transitory component model of stock return volatility. Unpublished paper: Department of Economics, University of California at San Diego.

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.

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.

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

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

Harvey, A. C. (1981). *Time Series Models* (1 ed.). New York: Philip Allan.

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

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

Kim, S., N. Shephard, and S. Chib (1998). Stochastic volatility: likelihood inference and comparison with ARCH models. *Rev. Economic Studies 65*, 361–93.

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

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

Marriott, J. M. and A. F. M. Smith (1992). Reparameterization aspects of numerical Bayesian methodology for autoregressivce moving-average models. *J. Time Series Analysis 13*, 327–43.

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 (1999). Derivative pricing in stochastic volatility models of Ornstein-Uhlenbeck type. Unpublished paper: Dept. of Mathematics, Aarhus University.

Pitt, M. K. and N. Shephard (1999a). Filtering via simulation based on auxiliary particle filters. *J. American Statistical Association 94*. Forthcoming.

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*. Oxford: Oxford University Press. Forthcoming.

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

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*, 405–30.

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: Birkhauser.

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

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

Rydberg, T. H. and N. Shephard (1998a). 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 (1998b). A modelling framework for the prices and times of trades made on the NYSE. Working paper: Nuffield College, Oxford. Presented at Workshop on Mathematical Finance, University of Bremen, Germany, February.

Sentana, E. (1991). Quadratic ARCH models: a potential re-intepretation of ARCH models. *Rev. Economic Studies 62*, 639–61.

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 C. K. Wong (1999). Empirical analysis of linear stochastic volatility models. Unpublished paper: Nuffield College, Oxford.

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*. Amsterdam: North-Holland.

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

			Sterling			
	Mean	MC S.E.	Ineff	Covariance & *Correlation*		
$\xi\|y$	0.5361	0.00111	4.3	0.0270	*-0.0399*	*0.0439*
$\gamma\|y$	0.004640	0.000181	300	-2.11e-005	1.04e-005	*-0.708*
$e^{-\lambda}\|y$	0.9825	0.000262	189	4.23e-005	-1.34e-005	3.44e-005
Model	Diagnostics					
	Log-L		Skew	Kurtosis	Norm	BL
Linear SV	-924.26		2.32	0.633	5.80	16.0
NID			7.58	30.5	988	1348
			DM			
	Mean	MC S.E.	Ineff	Covariance & *Correlation*		
$\xi\|y$	0.4712	0.000663	4.48	0.00931	*-0.0353*	*0.00389*
$\gamma\|y$	0.01099	0.000326	241	-2.21e-005	4.20e-005	*-0.694*
$e^{-\lambda}\|y$	0.9692	0.000335	145	3.22e-006	-3.86e-005	7.36e-005
Model	Diagnostics					
	Log-L		Skew	Kurtosis	Norm	BL
Linear SV	-944.95		0.246	0.335	0.173	39.3
NID			-3.49	9.11	95.2	2016
			Yen			
	Mean	MC S.E.	Ineff	Covariance & *Correlation*		
$\xi\|y$	0.3789	0.00159	13.6	0.0177	*-0.113*	*0.0295*
$\gamma\|y$	0.002214	6.78e-005	251	-1.97e-005	1.73e-006	*-0.580*
$e^{-\lambda}\|y$	0.9829	0.000154	111	1.77e-005	-3.45e-006	2.04e-005
Model	Diagnostics					
	Log-L		Skew	Kurtosis	Norm	BL
Linear SV	-790.59		2.25	1.22	6.54	35.5
NID			-6.43	12.3	193	2099
			ChF			
	Mean	MC S.E.	Ineff	Covariance & *Correlation*		
$\xi\|y$	0.5953	0.000545	3.10	0.00909	*0.00231*	*-0.0482*
$\gamma\|y$	0.02050	0.000337	158	1.81e-006	6.81e-005	*-0.650*
$e^{-\lambda}\|y$	0.9505	0.000348	79.4	-5.54e-005	-6.46e-005	0.000145
Model	Diagnostics					
	Log-L		Skew	Kurtosis	Norm	BL
Linear SV	-1047.6		0.614	0.568	0.699	31.2
NID			-3.31	35.6	1275	877

Table 5: *Daily returns for various exchange rates against the Dollar. The Monte Carlo S.E. of simulation is computed using Parzen kernel with a window of 100, 1,000 and 1,000 respectively. Italics are correlations rather than covariances of the posterior. Log-L denotes estimated log-likelihood function (using a particle filter) evaluated at the posterior mean.*