Stochastic Volatility Modelling with General Marginal Distributions: Inference, Prediction and Model Selection For Option Pricing.

Matthew P. S. Gander
David A. Stephens†
Department of Mathematics, Imperial College London

Summary. We compare option pricing results for stochastic volatility models where the underlying volatility process has a specific marginal form. The marginals that we consider are the *Generalised Inverse Gaussian* and *Tempered Stable* distributions; these general cases have not been implemented previously. We use the continuous time stochastic volatility model proposed in Barndorff-Nielsen and Shephard (2001b), where the volatility follows the Ornstein-Uhlenbeck equation driven by a background driving Lévy process. Model developments that incorporate long range dependence are also considered, and their merit and practical relevance discussed. We find that the *Generalised Inverse Gaussian* and *Inverse Gamma* marginal distributions accurately fit real data. Furthermore, we find evidence to suggest that, for many real data sets, an asset equation with long range dependence is not required. Inference is carried out in a Bayesian framework, with computation using Markov chain Monte Carlo (MCMC). We develop a general MCMC algorithm that appears to match the performance of other algorithms that have proved successful in the case of a Gamma marginal model.

1. Introduction

In this paper, we extend the Markov chain Monte Carlo (MCMC) algorithms of Roberts et al. (2004) and Griffin and Steel (2003), to fit the stochastic volatility (SV) models of Barndorff-Nielsen and Shephard (2001b) with marginal distributions other than the Gamma, which is the only marginal model yet implemented. In addition, the asset equation is generalised to incorporate long-range dependence (long-memory) using an approximation to fractional Brownian motion (fBm). These models are fitted to observed share values of 14 companies on the NYSE. The empirical performance of the different models is then compared using predictive densities and option pricing results.

Most option pricing is based on the standard Black-Scholes model (Black and Scholes (1973)). This model does not fit some observed properties of financial data and there have been many generalisations of the model to try to explain these. SV models are one generalisation, where the volatility is allowed to vary over time. For a review of recent SV models see Carr et al. (2003) and Schoutens (2003).

A new and popular class of continuous time SV models was proposed in Barndorff-Nielsen and Shephard (2001b) (referred to as the BNS model). For these models, the

†*Address for correspondence:* Department of Mathematics, Imperial College London, SW7 2AZ, London, UK.
E-mail: d.stephens@imperial.ac.uk

volatility follows an Ornstein-Uhlenbeck (OU) equation, with increments driven by a background driving Lévy process (BDLP). The volatility process is stationary, with jumps in volatility caused by jumps in the Lévy process. Between jumps, the volatility decays exponentially at a rate determined by one of the parameters in the OU equation. These models can therefore explain volatility clustering.

For the BNS model, the marginal distribution of the volatility is completely specified by the type of BDLP which drives the OU equation, and Barndorff-Nielsen and Shephard (2001b) derived the relationship between the marginal distribution and BDLP. This allows the practitioner to pick the marginal distribution of the volatility (under certain mild constraints) and simulate from it using the implied BDLP that must be used to obtain the chosen marginal distribution. Using stochastic (as opposed to constant) volatility, the thickness of the tails of the log-returns is increased.

Although the SV models introduced in Barndorff-Nielsen and Shephard (2001b) are attractive, MCMC inference for these models is quite involved. When a *Gamma* marginal distribution is used for the volatility, the MCMC inference is somewhat more straightforward, though still non-trivial. Roberts et al. (2004) and Griffin and Steel (2003) have performed the MCMC inference for a *Gamma* marginal distribution and have found the model to fit observed data well. Inferences for any other marginal distributions have, thus far, not been obtained; neither, therefore, has the adequacy of the Gamma marginal model, or the superiority of competing marginal models, been assessed for real financial data. This is the principal objective of this paper; our interest is to carry out the MCMC inference using other marginal distributions and to compare their empirical performance for option pricing and out of sample fit on daily data of a variety of shares on the NYSE.

A further generalisation of the Black-Scholes model is considered, where the Brownian motion is replaced by a multivariate normal approximation to fractional Brownian motion (fBm). This model can induce long-memory in the returns. We utilize this model for inferential purposes, in order to assess whether a long range dependence structure is warranted, rather than to facilitate option pricing, and thus problems associated with arbitrage are avoided.

1.1. Recent related work

Following the pioneering paper of Barndorff-Nielsen and Shephard (2001b), Griffin and Steel (2003) and Roberts et al. (2004) examined a simple case of the models we investigate. In recent work, Nicolato and Venardos (2003) derive the set of equivalent martingale measures (i.e. the set of all risk neutral measures) and closed-form option prices for simple derivatives, when the volatility follows the BNS SV model with a *Gamma* or *Inverse Gaussian* marginal distribution. They fit these BNS SV models to option prices on S&P 500 data and compare this with the fit of affine jump diffusion (AJD) models, using the results of Duffie et al. (2000). Nicolato and Venardos (2003) find little difference between the performance of the *Gamma* and *Inverse Gaussian* BNS SV models and that the AJD models provide a slightly better fit to the market option prices. They argue this is partly because the AJD models have more parameters and suggest using a superposition of BNS SV models so that the number of parameters of the models are similar.

The use of the Gamma marginal model appears to be motivated by computational tractability, rather than by any theoretical or practical reasoning. The observed returns are leptokurtic when considered marginally, and this empirical evidence may be modelled using a conditional normal error structure with stochastic variance at each time point generated

from some specified mixing distribution. Although any scale mixture of normals has higher kurtosis than the normal, an *Inverse Gamma* marginal model for the volatility is a more natural than the *Gamma*, as this induces a *Student-t* marginal model for the returns.

The contribution of this paper is to fit the BNS SV models to observed log-returns of a variety of real data sets and to assess which marginal distributions should be used to predict future asset movement most accurately. This assessment is performed using empirical option pricing results and predictive densities and both of these techniques naturally penalize over-parameterized models, so it is not necessary to only compare models with the same number of parameters. Unlike Nicolato and Venardos (2003), we find the mean and variance of the of the volatility to be consistent across marginal distributions and some marginals to perform better than others.

1.2. Plan of this paper

Section 2 introduces the models, whilst Section 3 recalls some important properties of the BNS SV models and how to simulate from them. These results are used in Section 4, which describes how MCMC can be used to estimate the unknown parameters of the models. Section 5 tests the performance of the models on real data and Section 6 concludes.

2. Generalisations of the Black-Scholes equation

The standard Black-Scholes equation with stochastic volatility, $\sigma^2(t)$, to model the movement of an underlying, $S(t)$, is

$$dS(t) = \mu S(t)\, dt + \sigma(t)\, S(t)\, dW(t), \tag{1}$$

where μ is the drift and, in the absence of arbitrage, must be equal to the risk free interest rate (see Hull (2000)) and $W(t)$ is Brownian motion, so $dW(t) \sim N(0, dt)$. For simplicity, we shall assume throughout that the interest rate is constant (see James and Webber (2000) for a review of interest rate models). Continue assuming that we have T equally spaced observed log returns, $y_1, \ldots, y_T$, each separated by Δ days.

The generalisations considered are using the BNS SV model for $\sigma^2(t)$ and replacing the Brownian motion by an approximate fractional Brownian motion. These generalisations are now introduced.

2.1. The BNS Ornstein-Uhlenbeck Stochastic Volatility model

A thorough description of the BNS SV model is given in Barndorff-Nielsen and Shephard (2001b); we summarise their approach here.

The returns of financial series are often rescaled so that they are of a reasonable size and so it is attractive for volatility to have a *self-decomposable* distribution, as the marginal distribution is altered in a predictable way by rescaling. Wolfe (1982) proved that $\sigma^2(t)$ has a *self-decomposable* distribution if and only if it can be written as

$$\sigma^2(t) = \int_{-\infty}^{0} \exp(s)\, dz(\lambda t + s),$$

where λ is any positive constant and $z(t)$ is a homogeneous Lévy process (see for example Bertoin (1994) and Sato (1999)), referred to as the background driving Lévy process

(BDLP). It follows that

$$\sigma^2(t) = \sigma^2(0)\, e^{-\lambda t} + e^{-\lambda t} \int_0^t e^{\lambda s} dz(\lambda s) \tag{2}$$

or, equivalently,

$$d\sigma^2(t) = -\lambda \sigma^2(t) + dz(\lambda t). \tag{3}$$

This SV model is described in Barndorff-Nielsen and Shephard (2001b); the volatility follows the OU equation (3). Such models have short-memory as the correlation between volatilities decays exponentially at a rate determined by λ.

The BDLP is constant apart from where it has positive jumps. Thus $\sigma^2(t)$ jumps when the BDLP jumps and decays exponentially in-between jumps (where $dz(\lambda t) = 0$). Therefore, λ controls both the rate at which jumps occur in $\sigma^2(t)$ and the rate at which the volatility decays in-between jumps.

2.2. Integrated Volatility and the Discrete Time Likelihood

The integrated volatility process, $\left\{\sigma_t^{2*}\right\}$, related to $\left\{\sigma_t^2\right\}$ is defined as

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

This is an important quantity for pricing European options (see Hull and White (1987)) and, for the BNS SV model, it can be shown that

$$\sigma^{2*}(t) = \frac{1}{\lambda}\left\{z(\lambda t) - \sigma^2(t) + \sigma^2(0)\right\}.$$

This relatively simple form for the integrated volatility is an attractive feature of the model. The discretely observed or actual volatility is

$$\sigma_i^2 = \sigma^{2*}(i\Delta) - \sigma^{2*}((i-1)\Delta). \tag{4}$$

Barndorff-Nielsen and Shephard (2001b) have shown that

$$corr\left(\sigma_i^2, \sigma_{i+s}^2\right) = d(\lambda\Delta)\, e^{-\lambda\Delta(s-1)},$$

where $d(\lambda\Delta)$ is independent of s and $0 < d(\lambda\Delta) < 1$. From equation (1), the log of the underlying, $x(t)$, satisfies

$$dx(t) = \left\{\mu - \frac{\sigma^2(t)}{2}\right\} dt + \sigma(t)\, dW(t) \tag{5}$$

and, if inference about μ and σ_i^2 is required, the likelihood for $y_1, \ldots, y_T$ is given by noting that

$$y_i \sim N\left(\left(\mu - \frac{\sigma_i^2}{2}\right)\Delta, \sigma_i^2\Delta\right). \tag{6}$$

2.3. Fractional Brownian Motion (fBm)

A commonly cited "stylised feature" of observed financial data is that series have long-memory, even though results in the literature are mixed. Some sources suggest that long-memory is present in the square of the log returns, such as Ding et al. (1993) and Bollerslev and Mikkelsen (1996), whilst others do not (see for example Krämer et al. (2002)). On the other hand, Barkoulas et al. (2000) find evidence of long-memory in the returns of the Greek Stock market. The model in (1) can be generalised to allow for long-memory, should the observed data require it, using two constructions. We may either induce long-memory via the volatility process, or leave the volatility process unaltered (so it has short-memory) but alter the share equation (1) itself. Roberts et al. (2004) and Griffin and Steel (2003) attempt to induce long-memory using superposition of volatility processes, each with their own BDLP and correlation parameter, as suggested in Barndorff-Nielsen and Shephard (2001b). However, although a superposition of finite numbers of volatility processes allows different BDLPs to describe short-range and longer-range dependencies in the volatility, the resulting process is still short-memory. In any case, estimating λ accurately is not easy (even with the single volatility process).

We feel that such a superposition, although theoretically appealing, is of limited practical use if a real attempt at long-memory modelling is to be made. Additionally, as more volatility processes are used, identifying the parameters of the component processes would become problematic. Therefore, we shall instead try to induce long-memory via equation (1) directly.

Equation (1) is driven by Brownian motion, which has independent increments, as it is driven by white noise. Fractional Brownian motion (fBm) is a generalisation of Brownian motion, which can have correlated increments and these increments are called fractional Gaussian noise (fGn). The strength of this correlation is determined by the Hurst parameter, $0 < H < 1$. When $0 < H < 0.5$ there is negative correlation and when $0.5 < H < 1.0$ there is positive correlation in the fGn and the fBm has long-memory. When $H = 0.5$, standard Brownian motion is recovered. For further details on fBm see Samorodnitsky and Taqqu (1994).

2.4. fBM and arbitrage

When the asset equation is driven by fBm, if pathwise integration is used for option pricing, for $H \neq 0.5$, there are arbitrage opportunities (see Rogers (1997) and Dai and Heyde (1996)). This problem is not difficult to bypass. A method for constructing a stochastic process with the same long range dependence behaviour as fBm that does not lead to arbitrage was given in Rogers (1997). For Gaussian approximations based on such processes that are arbitrarily close to fBm, fitting model parameters and pricing options give identical results to using the fBm model. In addition, Cheridito (2003) showed that if trading is restricted to time points at least a fixed time interval apart, arbitrage may be avoided even with the original fBm model. Thus fBm-type models should not be rejected for arbitrage reasons alone. Cajueiro and Fajardo (2003) use fBm to drive the asset equation with constant volatility. They estimate $H = 0.59$, for options on Brazilian stocks, by fitting model option prices to market prices. This suggests that the stocks have long-memory and that Brownian motion with constant volatility is not capable of accurately modelling the stock movement.

2.5. A Multivariate Normal Likelihood Approximation

Inference about model parameters using MCMC requires evaluation of the likelihood given a particular set of parameters of the model. For fGn, the likelihood is not readily available and so, instead of replacing the white noise with fGn, it is replaced by a multivariate normal (MVN) approximation to fGn.

For fGn with Hurst parameter, H, the correlation between two increments at discretely observed times i and j is

$$\Sigma_{i,j} = \frac{1}{2}\left\{|j-i+1|^{2H} - 2|j-i|^{2H} + |j-i-1|^{2H}\right\}$$

see Beran (1994), pg 74. Equation (5) is then replaced by

$$dx(t) = \mu dt - \frac{\sigma^2(t)}{2}(dt)^{2H} + \sigma(t)\, dW_{MVN}(t),$$

where W_{MVN} is the approximation to fBm. The likelihood is

$$f_Y(y) = \left(\prod_{i=1}^{T} \frac{1}{\sigma_i\sqrt{\Delta}}\right) f_G(g), \tag{7}$$

where

$$g_i = \frac{y_i - \left(\mu\Delta - \frac{\sigma_i^2}{2}(\Delta)^{2H}\right)}{\sigma_i\sqrt{\Delta}}$$

and $G \sim MVN(0, \Sigma)$. In the case that $H = 0.5$, the approximate fBm recovers the original Brownian motion model.

Evaluating the likelihood in equation (7) requires the computation of $|\Sigma|$ and Σ^{-1}. For general matrices, these are $O\left(n^3\right)$ calculations and even for relatively small T, calculation is not feasible. As Σ is a Toeplitz matrix, there are $O\left(n^2\right)$ algorithms available (see Ammar (1996) and Golub and Van Loan (1996)). For the data sets investigated, $T = 1000$ and computation is feasible for the $O\left(n^2\right)$ algorithms.

2.5.1. Leverage

In practice, negative log returns often generate a larger volatility than positive log returns of similar magnitude (this is referred to as leverage). It is generally thought that leverage is most significant in share data (see for example Meyer and Yu (2000)), though there is evidence that leverage is also present in FX data (see for example McKenzie (2002)), as well as evidence that it is not present for FX data (see for example Jacquier et al. (2001)). The BNS SV model of Section 2.1 can be generalised to incorporate the leverage effect, using the model proposed in Barndorff-Nielsen and Shephard (2001a). This has been implemented in Griffin and Steel (2003). Although we have also implemented this generalisation, our focus here is on assessing which marginals are most suitable for stochastic volatility models and all reported results are without leverage. We found a performance improvement for share data with the leverage model and no advantage for FX data.

3. Properties of the BNS SV model

Before estimating the parameters of the models using MCMC, we need to be able to sample from the discretely observed volatility (see equation (4)), given our chosen *self-decomposable* (and therefore *infinitely divisible*) marginal distribution for $\sigma^2(t)$. To do this, the relationship between the distribution of $\sigma^2(t)$ and the BDLP must be known, as well as how to sample from stochastic integrals with respect to the Lévy process. The required results are available from Barndorff-Nielsen and Shephard (2001b) and Barndorff-Nielsen and Shephard (2000) and these facilitate the MCMC implementation in Section 4. These results are now given.

3.1. Relationship between the BDLP and marginal distribution of $\sigma^2(t)$

For any *self-decomposable* distribution, there is a unique BDLP, $z(t)$, that will generate the required marginal for the volatility in equation (3). The relationship between the marginal and the BDLP has been derived in Barndorff-Nielsen and Shephard (2001b) and this relationship is between the Lévy measures of the marginal distribution of the volatility and $z(1)$. The following famous Theorem can be found, for example, in Bertoin (1994).

THEOREM 1. *Levy-Khintchine Formula*
A distribution, with density $f_X(x)$ and characteristic exponent $\Psi(\lambda)$, is infinitely divisible *if and only if there exists some $a \in \mathbb{R}^d$, a positive semi-definite quadratic Q on $\mathbb{R}^d$ and some measure $u(x)$ on $\mathbb{R}^d/\{0\}$ such that $\forall \lambda \in \mathbb{R}^d$*

$$\Psi(\lambda) = ia.\lambda + \frac{1}{2}Q(\lambda) + \int_{\mathbb{R}^d} \left[1 - e^{i\lambda.x} + i\lambda.x\mathbf{1}_{\{|x|<1\}}\right] u(x)\, dx \tag{8}$$

and

$$\int_{-\infty}^{\infty} \left(1 + |x|^2\right) u(x)\, dx < \infty.$$

Equation (8) is called the ***Lévy-Khintchine formula*** *, $u(x)$ is the* ***Lévy measure*** *of $f_X(x)$ and $Q(\lambda)$ is the* ***Gaussian coefficient****.*

If a *self-decomposable* marginal distribution for $\sigma^2(t)$ is chosen, with Lévy measure $u(x)$, and if $z(1)$ has Lévy measure $w(x)$, Barndorff-Nielsen and Shephard (2000) have shown that if $\sigma^2(t)$ follows the OU equation (3), then

$$w(x) = -u(x) - x\frac{du(x)}{dx} \tag{9}$$

and, if the *infinitely divisible* marginal distribution for $\sigma^2(t)$ is chosen, the BDLP is specified by equation (9).

Using the same notation as Barndorff-Nielsen and Shephard (2001b), define the Tail Mass function as

$$W_p^+(x) = \int_x^{\infty} w(y)\, dy = xu(x) \tag{10}$$

and the Inverse Tail Mass function as

$$W_p^{-1}(x) = \inf\left[y > 0 : W_p^+(y) \leq x\right], \tag{11}$$

where p are the parameters specifying the exact marginal distribution of $\sigma^2(t)$. These are both monotonic decreasing functions.

The last result needed, before we can sample from σ_i^2, is how to sample from stochastic integrals with respect to the BDLP, of the form given in equation (2).

3.2. Series representation of Stochastic integrals with respect to the BDLP

Griffin and Steel (2003) have shown that the discretely observed volatility can be written as

$$\sigma_i^2 = \frac{1}{\lambda}\left\{\eta_{i,2} - \eta_{i,1} + \left(1 - e^{-\lambda\Delta}\right)\sigma^2\left((i-1)\Delta\right)\right\}, \tag{12}$$

where

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

and

$$\eta_i = \left\{\begin{array}{cc} e^{-\lambda\Delta} & \int_0^{\Delta} e^{\lambda t} dz(\lambda t) \\ & \int_0^{\Delta} dz(\lambda t) \end{array}\right\} = \left\{\begin{array}{cc} e^{-\lambda\Delta} & \int_0^{\lambda\Delta} e^{t} dz(t) \\ & \int_0^{\lambda\Delta} dz(t) \end{array}\right\} \tag{13}$$

is a vector of random jumps, equal to a stochastic integral with respect to the BDLP, $z(t)$.

Barndorff-Nielsen and Shephard (2000) proved that if $f(s) \geq 0$ for $0 < s < \Delta$ and, if $f(s)$ is integrable with respect to $dz(s)$, then

$$\int_0^{\Delta} f(s)\, dz(s) \stackrel{\mathfrak{L}}{=} \sum_{i=1}^{\infty} W_p^{-1}\left(a_i/\Delta\right) f\left(\Delta e_i\right), \tag{14}$$

where $W_p^{-1}()$ is the Inverse Tail Mass function as defined in equation (11), a_i are the arrival times of a Poisson process of intensity 1 and e_i are independent standard uniform variates (also independent of a_i). Note that $W_p^{-1}\left(a_i/\Delta\right) \geq 0$ is a decreasing function and that, if it is non-zero for large a_i, the integral can be approximated by truncating the infinite series at some point. This is similar to the truncation scheme used to sample from Lévy processes in Walker and Damien (2000). We consider using GIG and TS distributions for the volatility and the only special case of these distributions, where the terms $W_p^{-1}\left(a_i/\Delta\right)$ are zero for sufficiently large a_i, is the *Gamma* distribution.

Assume that η_i is truncated by discarding all Poisson points which are greater than a_c (so the same truncation scheme is used for each element of the random shock vector). Let n_i be the number of Poisson points which are less than a_c for the i^{th} entry of the random shock vector (i.e. the number of Poisson points which contribute to η_i). The approximation to equation (13) is then

$$\eta_i \stackrel{\mathfrak{L}}{=} \left\{\begin{array}{cc} e^{-\lambda\Delta} & \sum_{j=1}^{n_i} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) e^{\lambda\Delta r_{i,j}} \\ & \\ & \sum_{j=1}^{n_i} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) \end{array}\right\}, \tag{15}$$

where $a_{i,j}$ and $r_{i,j}$ are Poisson points and uniforms as described previously. Continue with the notation

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n_1} \\ \vdots & \ddots & \vdots \\ a_{T,1} & \dots & a_{T,n_T} \end{pmatrix} \qquad R = \begin{pmatrix} r_{1,1} & \dots & r_{1,n_1} \\ \vdots & \ddots & \vdots \\ r_{T,1} & \dots & r_{T,n_T} \end{pmatrix}.$$

The method to sample from σ_i^2 is as follows: we select a *self-decomposable* distribution for $\sigma^2(t)$, and find the Lévy measure of this distribution and then the Lévy measure of $z(1)$, using equation (9). We then truncate the Poisson point process at a_c and use equations (11), (12) and (15) to generate σ_i^2. It is not always obvious how to find the Lévy measure in (1) and this is addressed in Section 4. Details on how a_c is chosen are given in Appendix A.1.1 and A.2.

4. Different marginal distributions for $\sigma^2(t)$

To fit the models described in Sections 2.1 and 2.3, the unknown parameters of the models must be estimated. Typically, in finance literature, these are estimated by minimizing the sum of the squared difference between model and market option prices. Nicolato and Venardos (2003) fit the BNS SV models for *Gamma* and *Inverse Gaussian* marginal distributions using this technique. Instead, we fit the models to observed log returns and use MCMC to estimate the parameters. Roberts et al. (2004) and Griffin and Steel (2003) have performed this inference when the volatility has a *Gamma* marginal distribution. Although easy to implement, the *Gamma* model is not especially plausible and other distributions are considered. Initially these distributions, along with their Lévy measures and Inverse Tail Mass functions, are introduced.

4.1. Different marginal distributions and their Inverse Tail Mass functions

The marginals considered are the *Generalised Inverse Gaussian* (GIG) and *Tempered Stable* (TS) distributions and special cases of these. These are *infinitely divisible* distributions on $\mathbb{R}^+$ and so are suitable choices for the marginal distribution of $\sigma^2(t)$ in equation (3). Many standard distributions are special cases of the GIG and TS distributions.

The TS distribution is less well known than the GIG and its density is difficult to interpret as it is only available as an infinite series. However, it is a flexible distribution and the implementation is more straightforward than for the GIG as its Lévy Measure is simple, leading to a straightforward Inverse Tail Mass function (see equation (11)). Details on the TS distribution can be found in Tweedie (1984) and Barndorff-Nielsen and Shephard (2001c).

4.1.1. Generalised Inverse Gaussian: $GIG(\gamma, \nu, \alpha)$

If $X \sim GIG(\gamma, \nu, \alpha)$, for $\gamma \in \mathbb{R}$ and $\nu, \alpha > 0$, the density is

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

where K_ν is a modified Bessel function of the third kind. The Lévy measure of X is then

$$u(x) = \frac{1}{x}\left\{\frac{1}{2}\int_0^\infty \exp\left(-\frac{x\xi}{2\nu^2}\right) g_\gamma(\xi)\, d\xi + \max(0,\gamma)\right\} \exp\left(-\frac{\alpha^2 x}{2}\right), \tag{16}$$

where

$$g_\gamma(x) = \frac{2}{x\pi^2}\left\{J^2_{|\gamma|}\left(\sqrt{x}\right) + N^2_{|\gamma|}\left(\sqrt{x}\right)\right\}^{-1}$$

and $J_{|\nu|}$ and $N_{|\nu|}$ are Bessel functions of the first and second kind respectively (see Barndorff-Nielsen and Shephard (2001b) for proof).

Using equation (10), the Tail Mass function is

$$W^+_{\gamma,\nu,\alpha}(x) = \left\{\frac{1}{2}\int_0^\infty \exp\left(-\frac{x\xi}{2\nu^2}\right) g_\gamma(\xi)\, d\xi + \max(0,\gamma)\right\} \exp\left(-\frac{\alpha^2 x}{2}\right) \tag{17}$$

and equation (11) implies the Inverse Tail Mass function is

$$W^{-1}_{\gamma,\nu,\alpha}(x) = z,$$

where z satisfies

$$x = \left\{\frac{1}{2}\int_0^\infty \exp\left(-\frac{z\xi}{2\nu^2}\right) g_\gamma(\xi)\, d\xi + \max(0,\gamma)\right\} \exp\left(-\frac{\alpha^2 z}{2}\right). \tag{18}$$

Computation of the Inverse Tail Mass function in general for the GIG distribution is feasible numerically. The value of x for a given z can then be found using a look up table and binary search. The integral was split into two parts and Gaussian Quadrature was used to evaluate the integral on a finite domain (that includes the origin). Gauss-Laguerre integration was used to evaluate the remaining integral on the infinite domain (see Atkinson (1988) for details on these numerical algorithms). The $GIG(\gamma,\nu,\alpha)$ marginal was implemented, as well as three standard distributions which are special cases.

Gamma: $Ga(\nu,\alpha)$ If $X \sim GIG\left(\nu, 0, \sqrt{2\alpha}\right)$, for $\nu,\alpha > 0$, then $X \sim Ga(\nu,\alpha)$ and the density is

$$f_X(x) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\alpha x}, \quad for \quad x > 0.$$

Using equation (10), the Tail Mass function is

$$W^+_{\nu,\alpha}(x) = \nu e^{-\alpha x}$$

and equation (11) implies the Inverse Tail Mass function is

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

It is unusual to be able to write $W_p^{-1}(x)$ in such a simple analytic form. Note that only when $x < \nu$ is $W_p^{-1}(x) \neq 0$. This is the only case of the $GIG(\gamma,\nu,\alpha)$ and $TS(\kappa,\nu,\alpha)$ distributions where the Inverse Tail Mass function is zero for all sufficiently large x and the summation in equation (15) need not be truncated. For all the other marginals considered, the infinite sum which constructs η_i must be truncated. Details on this truncation can be found in Appendix A.1.1 and A.2.

Positive Hyperbolic: $RPH(\nu,\alpha)$ If $X \sim GIG(1,\nu,\alpha)$, for $\nu,\alpha>0$, then $X \sim RPH(\nu,\alpha)$ and the density is

$$f_X(x) = \frac{\alpha}{2\nu K_1(\nu\alpha)} \exp\left\{-\frac{1}{2}\left(\frac{\nu^2}{x}+\alpha^2 x\right)\right\}, \quad for \quad x>0.$$

The Inverse Tail Mass function is available as a special case of equation (18) and can be evaluated using a similar method to the one used for the GIG marginal distribution.

Inverse Gamma: $IGa(\nu,\alpha)$ If $X \sim GIG\left(-\nu,\sqrt{2\alpha},0\right)$, for $\nu,\alpha>0$, then $X \sim IGa(\nu,\alpha)$ (i.e. the density of the reciprocal of a $Ga(\nu,\alpha)$ random variable) and the density is

$$f_X(x) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{-\nu-1} e^{-\alpha/x}, \quad for \quad x>0.$$

The Inverse Tail Mass function is available as a special case of equation (18) and can be evaluated using a similar method to the one used for the GIG marginal distribution.

4.1.2. Tempered Stable: $TS(\kappa,\nu,\alpha)$
If $X \sim TS(\kappa,\nu,\alpha)$, for $0<\kappa<1$ and $\nu,\alpha>0$, the density is

$$f_X(x) = e^{\nu\alpha} f_{Y|\kappa,\nu}(x) \exp\left(-\frac{\alpha^{1/\kappa}}{2}x\right), \quad for \quad x>0,$$

where

$$f_{Y|\kappa,\nu}(x) = \frac{\nu^{-1/\kappa}}{2\pi} \sum_{j=1}^{\infty} \frac{(-1)^{j-1}}{j!} \sin(j\kappa\pi)\,\Gamma(j\kappa+1)\,2^{jk+1}\left(x\nu^{-1/\kappa}\right)^{-j\kappa-1}, \quad for \quad x>0,$$

is the density function of the positive κ-stable law (see Feller (1971) and Barndorff-Nielsen and Shephard (2001c)). If $\kappa=0.5$ the *Inverse Gaussian* distribution is recovered.

The Lévy measure of X is then

$$u(x) = Ax^{-B-1}e^{-Cx}, \tag{19}$$

where $A=\nu\kappa 2^\kappa/\Gamma(1-\kappa)$, $B=\kappa$ and $C=\alpha^{1/\kappa}/2$ (see Barndorff-Nielsen and Shephard (2001c)). For this Lévy measure the Inverse Tail Mass function is

$$W^{-1}_{\kappa,\nu,\alpha}(x) = \left(\frac{A}{x}\right)^{1/B} \exp\left[-L_W\left(\frac{C}{B}\left(\frac{A}{x}\right)^{1/B}\right)\right], \tag{20}$$

where L_W is the Lambert-W function which satisfies

$$L_W(x) * \exp[L_W(x)] = x$$

and is a standard function available numerically. For further details on L_W, see Jeffrey et al. (1996). For the *Tempered Stable* distribution, an alternative series representation to equation (14) has been suggested in Rosiński (2000). This series representation avoids the

calculation of $W^{-1}_{\kappa,\nu,\alpha}(x)$, though the convergence of the series is slower. When implementing the MCMC for the *Tempered Stable* marginal, for large κ, many terms in the summation are required before the answer is sufficiently accurate to truncate. For this reason, the alternative representation is not implemented and the Barndorff-Nielsen and Shephard (2000) series representation is used. Additionally, from an MCMC viewpoint, this representation has fewer random terms in it, reducing the dimension of the problem on which the MCMC must be performed. A comparison of the two representations is now given and graphs of typical sizes of the terms for each series are shown in Figure 1.

Consider $\eta_{i,1}$; For the Barndorff-Nielsen and Shephard (2000) series representation,

$$\begin{aligned}\eta_{i,1} &= e^{-\lambda\Delta}\sum_{j=1}^{\infty} W^{-1}_{\kappa,\nu,\alpha}\left(\frac{a_{i,j}}{\lambda\Delta}\right)e^{\lambda\Delta r_{i,j}} \\ &= e^{-\lambda\Delta}\sum_{j=1}^{\infty}\left(\frac{A\lambda\Delta}{a_{i,j}}\right)^{1/B}\exp\left[-L_W\left(\frac{C}{B}\left(\frac{A\lambda\Delta}{a_{i,j}}\right)^{1/B}\right)\right]e^{\lambda\Delta r_{i,j}},\end{aligned}$$

whilst for the Rosiński (2000) series representation,

$$\eta_{i,1} = e^{-\lambda\Delta}\sum_{j=1}^{\infty}\min\left\{\left(\frac{a_{i,j}B}{A\lambda\Delta}\right)^{-1/B}, e_i v_i^{1/B}\right\}e^{\lambda\Delta r_{i,j}},$$

where $e_i \overset{iid}{\sim} \exp\left(\frac{1}{C}\right)$, $v_i, r_{i,j} \overset{iid}{\sim} U(0,1)$ and $a_{i,j}$ and $r_{i,j}$ are the same as in the Barndorff-Nielsen and Shephard (2000) representation.

Graphs of the log of the average of the terms for a $TS(\kappa,1,1)$ marginal, using the two representations given above, when $\lambda = \Delta = 1$, can be seen in Figure 1. The terms $e^{-\lambda\Delta}$ and $e^{\lambda\Delta r_{i,j}}$ are not included in these graphs as they are common to both series. Additional details on the Rosiński (2000) series can be found in Barndorff-Nielsen and Shephard (2001c), where our results for the Rosiński (2000) representation (dashed line) can also be verified. Averages were taken over 1,000,000 samples.

Inverse Gaussian: $IG(\nu,\alpha)$ If $X \sim TS\left(\frac{1}{2},\nu,\alpha\right)$, for $\nu,\alpha>0$, then $X \sim IG(\nu,\alpha)$ and the density is

$$f_X(x) = \frac{\nu e^{\nu\alpha}}{\sqrt{2\pi}}x^{-3/2}\exp\left\{-\frac{1}{2}\left(\frac{\nu^2}{x}+\alpha^2 x\right)\right\}, \quad for \quad x>0. \tag{21}$$

From equation (20), the Inverse Tail Mass function, defined in equation (11), is

$$W^{-1}_{\nu,\alpha}(x) = \frac{1}{\alpha^2}L_W\left(\frac{\nu^2\alpha^2}{2\pi x^2}\right).$$

Note that, although the $IG(\nu,\alpha)$ is a special case of both the $GIG(\gamma,\nu,\alpha)$ and $TS(\kappa,\nu,\alpha)$ distributions (when $\gamma=-0.5$ and $\kappa=0.5$), it is more straightforward to evaluate the Inverse Tail Mass function given in equation (20) than the one given in (18).

4.1.3. Properties of the six marginal distributions

The first four moments of the distributions we study are readily calculable. Such quantities are useful for comparing the performance of the six different marginals in capturing observed

baehaviour; we can inspect the mean/variance/kurtosis of the different distributions, to ensure that the prior/posterior marginal models for the volatility processes are in some sense similar. For the *Inverse Gamma* distribution, the mean, variance and kurtosis are finite when $\nu > 1$, $\nu > 2$ and $\nu > 4$ respectively, whilst for the *Tempered Stable*, *Inverse Gaussian*, *Gamma* and *Positive Hyperbolic* marginals, the mean, variance and kurtosis are always finite for $\nu, \alpha > 0$. For the six different marginal distributions for $\sigma^2(t)$, the kurtosis of the log returns is greater than three and so the tails of the log returns will be heavier than in the standard Black-Scholes equation with constant volatility.

We are able to pick the marginal distribution for $\sigma^2(t)$ from a rich class of distributions and sample from the discretely observed volatility, σ_i^2. An outline of the MCMC algorithm to estimate the parameters is given in Appendix A.1. The methods used to analyze the empirical performance of the models are now described.

4.2. Comparison with Alternative MCMC approaches

More complicated MCMC algorithms are considered in Roberts et al. (2004). The two algorithms presented there are tested on six training data sets and the ACF plots of posterior parameters are examined to compare the performance of each algorithm.‡ For these data sets, our algorithm and their preferred algorithm perform similarly for four of the six data sets. For the remaining two (large) data sets, our algorithm gives a more attractive ACF on one data set and theirs is favourable on the other. For real data sets, our algorithm usually produces samples that have a lower autocorrelation. This is illustrated in Figure 9, where the performance of their hybrid algorithm and ours is compared on six real data sets. In this assessment, the outputs from the algorithms were thinned so that the algorithms ran for the same length of time. Our conclusion is that the more complicated algorithms do not appear to offer significant improvements over our more straightforward approach.

5. Performance of the different marginals

The performances of the different marginal distributions are now compared on real data. The data sets are various shares of companies on the NYSE. The models are compared using predictive densities, empirical option pricing results and posterior parameter samples of parameters of the *GIG* distribution. Each data set is size 1000, with the last observation on 1^{st} December 2003. Options all expire in 20 days.

5.1. Model Selection: Predictive densities

Assume we have observed log returns, $y = \{y_1, \ldots, y_T\}$ (on which the model is fitted) and unseen data $Y' = \{y_{T+1}, \ldots, y_{T+t}\}$, which is not used to fit the model (and so can be viewed as a random variable). Let θ be a vector of the non-latent parameters specifying the model. The posterior predictive density is an average of predictions over the posterior distribution $p(\theta|y)$. That is

$$p\left(Y'|y\right) = \int p\left(Y'|\theta\right) p\left(\theta|y\right) d\theta.$$

For the models under investigation, it is not possible to calculate this posterior analytically, so B samples from the posterior distribution are taken (using the MCMC method described

‡We are very grateful to Dr Omiros Papaspiliopoulos for providing his code.

in Appendix A.1) and denoted by $\theta_1, \ldots, \theta_B$. The estimate for the predictive density is then

$$\widehat{p}\left(Y^{'}|y\right) = \frac{1}{B}\sum_{i=1}^{B} p\left(Y^{'}|\theta_i, y\right), \tag{22}$$

where $p\left(Y^{'}|\theta_i, y\right)$ is estimated by averaging the likelihood function given in equation (7) over many volatilities generated from the non-latent parameters θ_i. This method is similar to those described in Pitt and Shephard (1999) and Vrontos et al. (2003). The predictive density given θ_i is thus estimated using equation (22), with θ_i replaced by σ_j, a volatility generated using θ_i. Models, which are fitted to y and have large $\widehat{p}\left(Y^{'}|y\right)$, explain the unseen data, $Y^{'}$, well. This gives a way of comparing the out of sample fit of the different models.

5.2. Option Pricing

For the option pricing in this section, the asset equation with short range dependence only is used, and thus arbitrage opportunities are avoided. Pricing is performed under a risk neutral measure as described by Nicolato and Venardos (2003). All options are European, expiring in 20 days. The 27 different options, which are not described individually, are all popular standard options, such as European call, binary call, Parisian, Asian, knock in/out options etc.

To calculate the fair price of an option which expires at time t, given T observed data points, the technique used is as follows:-

(1) Perform MCMC on the data set of size T until convergence, so we are sampling from the posterior of $\mu, \lambda, \gamma, \kappa, \nu, \alpha$ and H (A, R and $\sigma^2(0\Delta)$ are latent parameters).

(2) Simulate $\sigma_1^2, ..., \sigma_t^2 | \lambda, \gamma, \kappa, \nu, \alpha$ from equation (12) by generating A, R and $\sigma^2(0\Delta)$ direct from their priors given $\gamma, \kappa, \nu, \alpha$.

(3) Perform Monte Carlo integration in t dimensions, simulating the asset forwards (using μ and H), taking the average discounted payoff, discounting using the constant interest rate, μ.

(4) go to (2) until enough volatilities have been used so that the expected discounted payoff given $\mu, \lambda, \gamma, \kappa, \nu, \alpha$ is sufficiently accurate.

(5) go to (1) and take another sample from the posterior of $\mu, \lambda, \gamma, \kappa, \nu, \alpha$ and average the estimates from (4) . Repeat this until this estimate is sufficiently accurate.

The rescaled sum of squared errors between the expected discounted and actual discounted payoff of the options are then examined to compare the performance of the different models. Histograms of the expected discounted payoff from (4), for real data sets, are given in Figures 2 and 3.

The fair price of an option is the expected discounted payoff, so prices of the algorithm are indifferent to risk and risk neutral. Nicolato and Venardos (2003) derive the set of equivalent martingale measures (i.e. the set of all risk neutral measures) when the volatility follows the BNS SV model with *Gamma* or *Inverse Gaussian* marginals. To test the

empirical performance of the BNS SV models for risk neutral pricing, only one risk neutral measure is required and the algorithm described above is used. As options expire in 20 days and typically μ is small, the discounting in (3) only slightly alters the option price.

5.3. *Testing of the option pricing algorithm*

Consider the standard Black-Scholes model (with constant volatility) for two of the simplest standard options: the European call and put (sometimes referred to as a vanilla call and put as they are standard calls and puts). The European call gives the owner the option to purchase the asset at a price E at time t. The European Put gives the owner the option to sell the asset at a price E at time t.

For known constant volatility, σ, and constant interest rate, r, if $V_C(t,E)$ is the fair price (at the present time) of a European call which expires at time t on asset S and $V_P(t,E)$ is the fair price (at the present time) of a European Put on the same asset S, then (see Hull (2000))

$$V_C(t,E) - V_P(t,E) = S(0) - Ee^{-rt}.$$

For this test, let $t = 20$, $S(0) = 100$, $E = 97$ and $r = 0.000133681$ (the daily interest rate corresponding to a rate of 5 per year). The fair price of the call-put is £3.26. For $\sigma = 0.03$, the fair price of the call and put are £7.05 and £3.79 respectively. These can be calculated by numerical solution of the Black-Scholes equation for the appropriate boundary conditions.

To test the correct implementation of the option pricing algorithm, training data were generated from the Black-Scholes model with constant volatility and inference is made on the non-latent parameters controlling the stochastic volatility model using MCMC. Volatilities were then generated using these non-latent parameters and the expected discounted payoff, given these simulated volatilities, is computed using Monte Carlo integration, discounting the payoff using the estimated interest rate, μ. Samples from the MCMC were taken after a burn in period of 10,000 iterations, thinning by taking every 250^{th} sample. Here we present a limited summary of the MCMC analysis, concentrating on the option pricing results only; For each marginal distribution, the convergence of the method to the correct price can be seen in Figures 4, 5 and 6. The thick line is the expected result knowing the correct constant value for σ^2.

5.4. *Posterior samples of parameter of the* GIG *marginal distribution*

We inspect the posterior distributions of parameters of the *GIG* distribution to give a greater understanding of which distributions are suitable for the marginal distribution of the volatility. Table A.2 lists special cases of the $GIG(\gamma,\nu,\alpha)$ distribution and Figure 7 shows the posterior distribution of γ and ν for the Heinz and Host Marriott data sets. This figure demonstrates that the *Ga* distribution is not supported for either data set - this result is typical for the real data sets investigated - and that for the Heinz data set the only special case of the *GIG* distribution which is supported is the *IGa*.

5.5. *Results*

The results for the predictive densities and option pricing for real data are now summarized. For predictive densities, the entries in the tables are the median and 95% credible intervals on the log scale. For predictive densities over 20 unseen data points, $B = 1000$ volatilities

were used to estimate $\widehat{p}(Y'|\theta_i, y)$ in equation (22) and for 80 unseen data points, $B = 10,000$ volatilities were used. For option pricing, the sum of the rescaled squared error between expected discounted and actual discounted payoff are reported.

5.5.1. Predictive densities over 20 unseen data points and option pricing results
Predictive density and option pricing results are given in Tables 2 and 4. For the predictive densities, for the three parameter distributions, the GIG has a larger log predictive density than the TS and so provides a better out of sample fit to the unseen data. For the two parameter distributions, the Ga has the largest predictive density, though this is not as large as for the GIG. Boxplots demonstrating the large GIG predictive density are given in Figure 8. For option pricing, for the three parameter distributions, we find the GIG to predict the payoff more accurately than the TS and so again the GIG outperforms the TS. For the two parameter distributions, we find the IGa marginal to predict the expected discounted payoff of the options most accurately, when using a squared error loss. The IGa also has a smaller loss than the GIG distribution and demonstrates that if the user is interested in option pricing, then generalisation to the more complicated GIG density is not warranted and the IGa distribution should be used instead.

5.5.2. Long-memory modelling: Predictive densities over 80 unseen data points
Results for the long-memory model are given in Table 5. For all data sets, the posterior of H only has support near 0.5, indicating that the data exhibit weak long-memory behaviour. As a result, the predictive densities were unable to distinguish consistently between the short and long-memory models. Coca-Cola Co has the strongest support for long-memory, as the posterior for the Hurst parameter contains larger values than for the other data sets. Figure 10 shows posterior histograms of the Hurst parameter for Coca-Cola Co and British Airways PLC and suggests that for Coca-Cola Co, the long-memory model is required, as $H = 0.5$ is not supported. For British Airways PLC there is support for $H = 0.5$, suggesting that the long-memory model is not necessary. For all the other data sets, the posteriors for H are similar to the British Airways PLC share and this suggests that the approximate fBm generalisation is not necessary and the short-memory model is preferred.

6. Conclusions

We performed MCMC for the BNS OU stochastic volatility models when the volatility has a *Generalised Inverse Gaussian* (GIG) or *Tempered Stable* (TS) marginal distribution. *Gamma*, *Positive Hyperbolic*, *Inverse Gamma* and *Inverse Gaussian* marginals are special cases of the GIG and TS and these are implemented individually. To allow for long-memory in the share equation, the Brownian motion is replaced by an approximation to fractional Brownian motion. The models were tested empirically on real data using predictive densities and squared-error loss between expected and actual discounted option payoffs.

For the different marginal distributions, we found the *Generalised Inverse Gaussian* marginals to perform best for predicting unseen data and the *Inverse Gamma* marginal to perform best for option pricing; the log returns are then approximately *Student-t* distributed. If the focus of the user is option pricing, the generalisation to the more flexible *Generalised Inverse Gaussian* is unnecessary, as the *Inverse Gamma* provides more accurate

predictions. For the long-memory model, we are unable to identify any performance difference empirically, using predictive densities, between the generalised and original models. However, the posterior densities of the Hurst parameter often have a considerable support for $H = 0.5$, indicating that there is no evidence for the long-memory model, and that the generalisation to fBm is not necessary for the data sets examined.

A. Appendix

A.1. MCMC algorithm

The MCMC was tested thoroughly on simulated data to ensure its correct implementation, before the real data were studied. Several specific points are worthy of particular discussion.

A.1.1. Truncation of the random shock vector

To use the approximation for the random shock vector given in equation (15), the critical value at which the Poisson points are truncated, a_c, must be chosen. If only terms from the Poisson Point process which contribute at least z_{tol} to the summation are included, then

$$W_p^{-1}\left(\frac{a_c}{\lambda\Delta}\right) = z_{tol}$$

and so

$$a_c = \lambda\Delta W_p^{+}\left(z_{tol}\right),$$

which can be evaluated using the Tail Mass functions given in Section 4.1.

For the *Gamma* distribution, we choose $z_{tol} = 0$ as the summation only has a finite number of non-zero terms. For the *Positive Hyperbolic* and *Inverse Gamma* distributions, we typically choose $z_{tol} = 0.001$. For the *Tempered Stable* and *Inverse Gaussian* distributions, a more advanced truncation scheme is used, which uses knowledge of the asymptotic behaviour of the Inverse Tail Mass function. This truncation method can be found in Appendix A.2. When sampling from the model, these truncations generate marginal distributions with the correct mean and variance for a wide range of parameter values, as well as the correct correlation structure for the volatility process.

Whenever a move in a non-latent parameter alters a_c, in order to maintain a consistent level of accuracy, Poisson points and uniforms must be added/removed at the same time as this non-latent parameter move. Such moves require Reverse Jump MCMC (see for example Green (1995)).

A.1.2. Treating $\sigma^2(0\Delta)$ as an unknown parameter

Equation (12) requires $\sigma^2(0\Delta)$ to be known. For all distributions other than the *Tempered Stable*, this is treated as a latent parameter, with a prior the same as the marginal distribution used for $\sigma^2(t)$. This idea was suggested by J.E. Griffin in his comment to Barndorff-Nielsen and Shephard (2001b).

For the *Tempered Stable* marginal, it is difficult to evaluate the density, particularly for small arguments (techniques to evaluate the *Tempered Stable* density can be found in Nolan (1997)) and so a different representation for $\sigma^2(0\Delta)$ is used. Rosiński (2000) has shown

that, when a $TS(\kappa,\nu,\alpha)$ marginal is used for $\sigma^2(t)$, the distribution of $\sigma^2(0\Delta)$ is given by

$$\sigma^2(0\Delta) \stackrel{\mathfrak{L}}{=} \sum_{i=1}^{\infty} \min\left\{\left(\frac{a_i\kappa}{B}\right)^{-1/\kappa}, e_i v_i^{1/\kappa}\right\}, \tag{23}$$

where a_i are the arrival times of a Poisson process with intensity 1, $e_i \overset{iid}{\sim} \exp\left(\frac{1}{C}\right)$ and $v_i \overset{iid}{\sim} U(0,1)$ (all independent of each other) and

$$B = \left(\frac{\nu\kappa 2^{\kappa}}{\Gamma(1-\kappa)}\right) \qquad C = \frac{\alpha^{1/\kappa}}{2}.$$

This representation is used in Barndorff-Nielsen and Shephard (2001c). To be consistent with the representation used in equation (15), we use the equivalent representation

$$\sigma^2(0\Delta) \stackrel{\mathfrak{L}}{=} \sum_{i=1}^{\infty} W^{-1}_{\kappa,\nu,\alpha}(a_{0,j}), \tag{24}$$

avoiding evaluation of the *Tempered Stable* density.

A.1.3. Priors

The joint prior is of the form

$$p\left(\sigma^2(0\Delta), \mu, \lambda, \gamma, \kappa, \nu, \alpha, H, A, R\right) = p\left(\sigma^2(0\Delta)\right) p(\mu)\, p(\lambda)\, p(\gamma)\, p(\kappa)\, p(\nu)\, p(\alpha)\, p(H)\, p(A)\, p(R).$$

The priors used for real data are given in Table 1. The quantity $\sigma^2(0\Delta)$ is not treated as an additional parameter for the *Tempered Stable* distribution as the representation in equation (24) is used instead.

For testing purposes, the hyperparameter in the Gamma prior for λ, l_p is taken to be 0.001, although it has been argued by Griffin and Steel (2003) that for observed data, $l_p = 1$ is more appropriate and this is used on real data in Section 5.5. Typically n_p and a_p are chosen to be small (say 0.001) so that the priors are reasonably flat. These two priors are not for the original *GIG* parameters but are priors for the parameters of the specific marginal.

For the $TS(\kappa,\nu,\alpha)$ distribution, in the limit as κ tends to 1, the volatility becomes constant. We wish to largely exclude this possibility in the analysis of real data series. Thus an informative prior for κ is used that has most of its weight away from 1 (for training data, it thought legitimate to use a $U(0,1)$ prior, as some constant volatility simulated series were studied). For the *Inverse Gamma* marginal, a $Ga(1,n_p)$ prior is used for $(\nu-4)$ so the mean, variance and kurtosis of the log returns are finite. For the Hurst parameter, H, a $U(0.5,1)$ prior is used (rather than $U(0,1)$), as we do not want negative correlation between the increments in the fGn approximation.

A.1.4. Proposals

The interest rate, μ, is on the entire real line and the proposal is

$$\mu' \sim N(\mu, c),$$

where c is tuned to give suitable acceptance rates. The parameter ν is on $\mathbb{R}^+$ and the proposal is

$$\nu^{'} \sim Ga\left(c, c/\nu\right),$$

where again c is tuned to give suitable acceptance rates. Similar proposals are used for $\sigma^2\left(0\Delta\right), \lambda$ and α, which are also on $\mathbb{R}^+$. κ and H are positive parameters on a finite range. These are transformed onto $\mathbb{R}^+$ and a *Gamma* local move is performed but on the transformed parameter, with appropriate amendment of the acceptance probability.

Whenever a new parameter is proposed, which alters the truncation point of the Poisson points, a_c, we propose to add (or remove) Poisson points and uniforms to keep a consistent accuracy throughout the simulation. The addition of Poisson points and uniforms are proposed direct from their priors, whilst the removal of points is deterministic (because of the nature of reverse jump MCMC). The acceptance probability for a move in ν is then given by

$$\min\left[1, \frac{l\left(y_i|\left(\sigma_i^2\right)^{'}\right)}{l\left(y_i|\sigma_i^2\right)} \frac{p\left(\nu^{'}\right)}{p\left(\nu\right)} \frac{q\left(\nu^{'} \rightarrow \nu\right)}{q\left(\nu \rightarrow \nu^{'}\right)}\right],$$

with similar acceptance probabilities for other parameters.

After proposing a move for each parameter, we perform a sweep through all the Poisson points and uniforms, proposing to update each row of A and R in order, again, with proposals direct from their priors. For these moves, as proposals are direct from their priors, the acceptance probability is

$$\min\left[1, \frac{l\left(y_i|\left(\sigma_i^2\right)^{'}\right)}{l\left(y_i|\sigma_i^2\right)}\right].$$

The order of the updates is $\kappa, \gamma, \nu, \alpha, \lambda, H, \sigma^2\left(0\right), \mu$ and then the first row of A and R are updated together, then the second row until all rows have had a proposed update. When the *Tempered Stable* marginal is used, the $\sigma^2\left(0\right)$ update is replaced by an update of the Poisson points which specify $\sigma^2\left(0\right)$ in equation (24).

A.2. Improved Truncation for the Tempered Stable *marginal*

Assume an initial truncation has been made as suggested in Appendix A.1.1. Now consider the error term, R_i, for $\eta_{i,2}$ (note that the terms for $\eta_{i,1}$ are less than the terms for $\eta_{i,2}$). Recall equation (15)

$$\eta_{i,2} = \sum_{j=1}^{\infty} W_{\nu,\alpha}^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right).$$

At present, let the truncation of the Poisson point process be at $a_{c,1}$, so

$$\eta_{i,2} = \sum_{j=1}^{n_i} W_{\nu,\alpha}^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) + R_i,$$

where n_j is the number of Poisson Points occurring before $a_{c,1}$ and

$$R_i = \sum_{j=n_i+1}^{\infty} W_{\nu,\alpha}^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right).$$

Assume that for the R_i summation, all the $a_{i,j}$ are large and asymptotic assumptions can be made.

The central idea is then to approximate R_i analytically and pick a new truncation point (if the accuracy is not sufficient) which makes R_i small. This requires knowledge of the exact form of the Inverse Tail Mass function.

For the *Tempered Stable* marginal, the Lévy measure is given by equation (19) and using equations (10) and (11), for large x we have

$$W_{\nu,\alpha}^{-1}(x) = z \approx B^{1/C} \frac{1}{x^{1/C}}.$$

Consider truncating $a_{i,j}$ at $d > a_{c,1}$ and dropping the ordering of the Poisson Points. The error, R_i, will be approximately

$$R_i \approx (B\lambda\Delta)^{1/C} \sum_{j=1}^{n_{i,2}} \frac{1}{u_{i,j}^{1/C}},$$

where $n_{i,2} \sim Po(d - a_{c,1})$ is the number of the Poisson Points in $(a_{c,1}, d)$ and $u_{i,j} \overset{iid}{\sim} U(a_{c,1}, d)$ is also independent of $n_{i,2}$. Taking expectations for constant $n_{i,2}$, we have

$$\begin{aligned} E[R_i|n_{i,2}] &\approx n_{i,2} (B\lambda\Delta)^{1/C} E\left[u_{i,j}^{-1/c}\right] \\ &= n_{i,2} (B\lambda\Delta)^{1/C} \left(\frac{C}{1-C}\right) \left(\frac{a^{\frac{C-1}{C}} - d^{\frac{C}{1-C}}}{d - a_{c,1}}\right). \end{aligned}$$

Taking the expectation with respect to $n_{i,2}$ gives

$$E[R_i] = (B\lambda\Delta)^{1/C} \left(\frac{C}{1-C}\right) \left(a^{\frac{C-1}{C}} - d^{\frac{C}{1-C}}\right).$$

Noting that $0 < B = \kappa < 1$ and letting $d \to \infty$ gives

$$E[R_i] \approx \left(\frac{C}{1-C}\right) (B\lambda\Delta)^{1/C} a_{c,1}^{\frac{C-1}{C}},$$

so consider using the new truncation point

$$a_{c,2} = \left\{E[R_i] \left(\frac{1-C}{C}\right)\right\}^{\frac{C}{C-1}} (B\lambda\Delta)^{\frac{1}{1-C}},$$

where $E[R_i]$ is now the expected error in the summation that we would like (for our simulations we chose $E[R_i] = 0.001$). The maximum of $a_{c,1}$ and $a_{c,2}$ can then be used as previously.

For the *Inverse Gaussian*(ν, α) distribution ($TS\left(\frac{1}{2}, \nu, \alpha\right)$), $B = \nu/\sqrt{2\pi}$ and $C = \frac{1}{2}$ and

$$a_{c,2} = \frac{(\lambda\nu\Delta)^2}{2\pi} \frac{1}{E[R_i]}.$$

References

Ammar, G. S. (1996). Classical foundations of algorithms for solving positive definite Toeplitz equations. *Calcolo 33*, 99–113.

Atkinson, K. E. (1988). *An Introduction to Numerical Analysis*. Wiley.

Barkoulas, J. T., C. F. Baum, and N. Travlos (2000). Long memory in the Greek stock market. *Applied Financial Economics 10*, 177–185.

Barndorff-Nielsen, O. E. and N. Shephard (2000). Modelling by Lévy processes for financial econometrics. In O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes - Theory and Applications*, pp. 283–318. Boston: Birkhäuser.

Barndorff-Nielsen, O. E. and N. Shephard (2001a). Incorporation of a leverage effect in a stochastic volatility model. *Unpublished paper*.

Barndorff-Nielsen, O. E. and N. Shephard (2001b). Non-Gaussian Ornstein-Uhlenbeck based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B 63*, 167–241.

Barndorff-Nielsen, O. E. and N. Shephard (2001c). Normal modified stable processes. *Theory of Probability and Mathematical Statistics 65*, 1–19.

Beran, J. (1994). *Statistics for Long-Memory Processes*. Chapman and Hall.

Bertoin, J. (1994). *Lévy Processes*. Chapman and Hall.

Black, F. and M. S. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*, 637–654.

Bollerslev, T. and H. O. A. Mikkelsen (1996). Modelling and pricing long-memory in stock market volatility. *Journal of Econometrics 73*, 151–184.

Cajueiro, D. O. and J. Fajardo (2003). Volatility estimation and option pricing with fractional Brownian motion. *Unpublished paper*.

Carr, P., H. Geman, D. P. Madan, and M. Yor (2003). Stochastic volatility for Lévy processes. *Mathematical Finance 13*, 345–382.

Cheridito, P. (2003). Arbitrage in fractional Brownian motion models. *Finance and Stochastics 7*, 533–553.

Dai, W. and C. C. Heyde (1996). Itô's formula with respect to fractional Brownian motion and its application. *Journal of Applied Mathematics and Stochastic Analysis 9*, 439–448.

Ding, Z., C. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*, 83–106.

Duffie, D., J. Pan, and K. Singleton (2000). Transofrm analysis and analysis pricing for affine jump diffusions. *Econometrica 68*, 1343–1376.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications*. Wiley.

Golub, G. H. and C. F. Van Loan (1996). *Matrix Computations*. John Hopkins University Press.

Green, P. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika 82*, 711–732.

Griffin, J. E. and M. F. J. Steel (2003). Inference with non-Gaussian Ornstein-Uhlenbeck processes for stochastic volatility. *Unpublished paper*.

Hull, J. and A. White (1987). The pricing of options on assets with stochastic volatility. *The Journal of Finance 3*, 281–300.

Hull, J. C. (2000). *Options, Futures and Other Derivatives*. Prentice Hall.

Jacquier, E., N. G. Polsen, and P. E. Rossi (2001). Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. *Unpublished paper*.

James, J. and N. Webber (2000). *Interest Rate Modelling*. Wiley.

Jeffrey, G. H., D. J. Hare, and D. E. G. Corless (1996). Unwinding the branches of the Lambert W function. *The Mathematical Scientist 21*, 1–7.

Krämer, W., P. Sibbertsen, and C. Kleiber (2002). Long memory versus structural change in financial time series. *Allgemeines Statistisches Archiv 86*, 83–96.

McKenzie, M. (2002). The economics of exchange rate volatility asymmetry. *International Journal of Finance and Economics 7*, 247–260.

Meyer, R. and J. Yu (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal 3*, 198–215.

Nicolato, E. and E. Venardos (2003). Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Mathematical Finance 13*, 445–466.

Nolan, J. P. (1997). Numerical calculation of stable densities and distribution functions. *Communications in Statistics - Stochastic Models 13*, 759–774.

Pitt, M. K. and N. Shephard (1999). Time varying covariance: A factor stochastic volatility approach. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Eds.), *Proceedings of the Sixth Valencia International Meeting on Bayesian Statistics*, pp. 547–570. Oxford: Oxford University Press.

Roberts, G., O. Papaspiliopoulos, and P. Dellaportas (2004). Bayesian inference for non-Gaussian Ornstein-Uhlenbeck stochastic volatility processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66*, 369–393.

Rogers, L. C. G. (1997). Arbitrage with fractional Brownian motion. *Mathematical Finance 7*, 95–105.

Rosiński, J. (2000). Stochastic series representations of inverse Gaussian and other exponentially tempered stable processes. *Research Report 42*. Centre for mathematical Physics and Stochastics, University of Aarhus, Aarhus.

Samorodnitsky, G. and M. S. Taqqu (1994). *Stable Non-Gaussian Processes: Stochastic Models with Infinite Variance*. Chapman and Hall.

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

Schoutens, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives*. Wiley.

Tweedie, M. (1984). An index which distinguishes between some important exponential families. In J. Ghosh and J. Roy (Eds.), *Proceedings of the Indian Statistical Institute Golden Jubilee Conference on Statistics: Applications and New Directions*, pp. 579–604.

Vrontos, D., P. Dellaportas, and N. Politis (2003). Inference for some multivariate ARCH and GARCH models. *Journal of Forecasting 22*, 427–446.

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

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

Table 1. Priors for the MCMC. ‡ M is the marginal distribution that is used for $\sigma^2(t)$.

$p\left(\sigma^2(0\Delta)\,|\nu,\alpha\right)$	=	$M(\nu,\alpha)$ ‡
$p(\mu)$	=	$N(0, 0.0004^2)$
$p(\lambda)$	=	$Ga(1, l_p)$
$p(\gamma)$	=	$N\left(0, 5^2\right)$
$p(\kappa)$	=	$Beta(1, 15)$
$p(\nu)$	=	$Ga(1, n_p)$
$p(\alpha)$	=	$Ga(1, a_p)$
$p(H)$	=	$U(0.5, 1)$

Table 2. Median and 95% credible intervals of predictive densities for GIG, TS and IG, Ga, RPH and IGa distributions.

Share	$GIG(\gamma,\nu,\alpha)$	$TS(\kappa,\nu,\alpha)$	$IG(\nu,\alpha)$
British Airways PLC	-33.6 (-34.1, -33.0)	-34.4 (-38.4, -33.0)	-33.9 (-34.8, -33.0)
Citigroup Inc	-28.8 (-29.8, -28.0)	-34.9 (-38.6, -32.6)	-29.5 (-31.0, -28.1)
Coca-Cola Co	-33.0 (-33.8, -32.3)	-38.0 (-40.7, -36.6)	-35.0 (-36.2, -33.9)
General Motors Corp	-45.6 (-46.2, -45.0)	-46.3 (-47.2, -45.8)	-46.0 (-46.4, -45.6)
HJ Heinz Co	-24.9 (-27.0, -24.0)	-34.8 (-42.6, -29.2)	-27.8 (-31.0, -25.1)
Host Marriott Corp	-36.2 (-37.0, -35.2)	-36.5 (-37.4, -35.6)	-35.9 (-36.7, -35.2)
JP Morgan Chase & Co	-26.0 (-28.6, -24.4)	-25.9 (-28.3, -24.4)	-25.9 (-27.5, -24.4)
Kellogg Co	-35.9 (-36.4, -35.5)	-37.8 (-39.9, -36.6)	-36.5 (-37.4, -35.8)
McDonald's Corp	-48.1 (-48.5, -47.6)	-48.9 (-49.5, -48.3)	-48.5 (-49.2, -48.0)
Microsoft	-24.9 (-27.1, -23.1)	-26.3 (-32.2, -23.8)	-24.1 (-25.8, -22.6)
Procter & Gamble Co	-28.2 (-28.8, -27.7)	-28.0 (-28.7, -27.4)	-28.0 (-28.8, -27.3)
S&P 500	-32.8 (-33.4, -32.3)	-37.4 (-41.9, -35.5)	-33.6 (-34.9, -32.5)
Textron Inc	-42.6 (-42.9, -42.2)	-42.7 (-43.3, -42.2)	-42.6 (-43.0, -42.2)
Time Warner Inc	-31.8 (-32.2, -31.4)	-37.4 (-44.3, -32.8)	-32.3 (-33.8, -31.3)
	$Ga(\nu,\alpha)$	$RPH(\nu,\alpha)$	$IGa(\nu,\alpha)$
British Airways PLC	-34.0 (-34.8, -33.3)	-34.0 (-34.9, -33.3)	-34.7 (-35.5, -34.1)
Citigroup Inc	-28.8 (-30.0, -28.0)	-29.2 (-30.8, -28.0)	-31.3 (-32.5, -29.9)
Coca-Cola Co	-34.9 (-36.3, -33.9)	-34.9 (-36.1, -34.0)	-35.8 (-37.2, -34.7)
General Motors Corp	-45.9 (-46.3, -45.5)	-46.0 (-46.4, -45.5)	-46.0 (-46.4, -45.6)
HJ Heinz Co	-26.1 (-28.6, -24.5)	-26.3 (-30.2, -23.8)	-30.4 (-32.7, -27.9)
Host Marriott Corp	-36.1 (-36.9, -35.4)	-36.2 (-37.1, -35.5)	-37.6 (-38.2, -37.1)
JP Morgan Chase & Co	-24.4 (-26.1, -23.0)	-25.2 (-27.0, -23.7)	-29.1 (-30.5, -27.7)
Kellogg Co	-36.3 (-37.0, -35.7)	-36.5 (-37.5, -35.8)	-37.1 (-38.0, -36.4)
McDonald's Corp	-48.8 (-49.7, -48.3)	-48.7 (-49.7, -48.2)	-50.0 (-51.3, -48.7)
Microsoft	-22.4 (-24.4, -21.2)	-23.4 (-25.7, -21.8)	-27.8 (-29.2, -26.3)
Procter & Gamble Co	-28.6 (-29.2, -28.1)	-29.2 (-30.5, -28.4)	-29.2 (-30.2, -28.4)
S&P 500	-33.6 (-34.9, -32.5)	-33.5 (-35.0, -32.4)	-34.4 (-35.7, -33.2)
Textron Inc	-42.6 (-43.0, -42.1)	-42.6 (-43.0, -42.2)	-42.5 (-42.8, -42.2)
Time Warner Inc	-32.1 (-33.2, -31.5)	-32.4 (-35.1, -31.5)	-33.3 (-34.8, -32.1)

Table 3. Special cases of the GIG distribution, with conventional two parameter representation.

Form of $GIG(\gamma,\nu,\alpha)$ distribution	Standard two parameter family
$GIG\left(\nu, 0, \sqrt{2\alpha}\right)$	$Ga(\nu,\alpha)$
$GIG(1,\nu,\alpha)$	$RPH(\nu,\alpha)$
$GIG\left(-\nu, \sqrt{2\alpha}, 0\right)$	$IGa(\nu,\alpha)$
$GIG\left(-\frac{1}{2},\nu,\alpha\right)$	$IG(\nu,\alpha)$

Table 4. Summaries of option payoff errors for different marginal distributions; rescaled squared errors.

Option Pricing	$GIG(\gamma,\nu,\alpha)$	$TS(\kappa,\nu,\alpha)$	$IG(\nu,\alpha)$	$Ga(\nu,\alpha)$	$RPH(\nu,\alpha)$	$IGa(\nu,\alpha)$
British Airways PLC	1.02	1.26	1.01	1.02	1.01	1.00
Citigroup Inc	1.00	0.99	1.00	1.00	1.00	1.00
Coca-Cola Co	1.04	1.04	1.07	1.05	1.05	1.00
General Motors Corp	1.01	0.99	1.01	1.01	1.00	1.00
HJ Heinz Co	1.01	0.97	1.03	1.03	1.05	1.00
Host Marriott Corp	1.01	1.24	1.04	1.00	1.00	1.00
JP Morgan Chase & Co	1.00	0.99	1.00	1.00	1.00	1.00
Kellogg Co	1.04	0.99	1.04	1.05	1.04	1.00
McDonald's Corp	1.02	1.11	1.00	0.99	0.99	1.00
Microsoft	1.01	1.14	1.04	1.03	1.02	1.00
Procter & Gamble Co	0.99	1.01	1.01	0.99	1.03	1.00
S&P 500	1.00	1.01	1.00	0.99	0.99	1.00
Textron Inc	1.00	0.98	1.00	0.99	1.00	1.00
Time Warner Inc	1.00	0.99	1.00	0.99	1.00	1.00
Sum	14.16	14.70	14.26	14.15	14.19	14.00

Table 5. Median and 95% credible intervals of predictive densities over 80 unseen data points of models with Brownian and approximate fractional Brownian motion.

Predictive Density	Brownian Motion	Approximate fBm	PS for H
British Airways PLC	-168.5 (-170.0, -166.1)	-168.2 (-170.5, -166.1)	0.512 (0.501, 0.548)
Citigroup Inc	-112.1 (-115.9, -109.6)	-112.1 (-116.4, -109.7)	0.505 (0.501, 0.523)
Coca-Cola Co	-137.6 (-141.7, -135.3)	-138.3 (-143.2, -135.5)	0.523 (0.502, 0.563)
General Motors Corp	-160.4 (-161.9, -158.9)	-160.2 (-161.6, -158.5)	0.512 (0.500, 0.542)
HJ Heinz Co	-131.6 (-135.5, -129.1)	-131.6 (-135.7, -129.1)	0.503 (0.500, 0.517)
Host Marriott Corp	-143.0 (-145.6, -140.7)	-143.1 (-145.3, -140.8)	0.512 (0.503, 0.551)
JP Morgan Chase & Co	-105.1 (-112.3, -99.2)	-105.0 (-111.8, -99.3)	0.510 (0.501, 0.540)
Kellogg Co	-124.8 (-127.9, -122.9)	-125.7 (-130.0, -123.6)	0.504 (0.501, 0.514)
McDonald's Corp	-164.0 (-167.6, -161.3)	-164.5 (-169.3, -161.9)	0.510 (0.502, 0.535)
Microsoft	-93.9 (-101.9, -86.7)	-92.3 (-100.5, -86.0)	0.506 (0.501, 0.529)
Procter & Gamble Co	-113.9 (-115.1, -112.8)	-115.1 (-117.4, -113.0)	0.508 (0.500, 0.528)
S&P 500	-138.7 (-141.7, -137.1)	-138.5 (-141.1, -137.0)	0.507 (0.500, 0.530)
Textron Inc	-150.9 (-152.8, -149.6)	-151.0 (-152.9, -149.6)	0.518 (0.501, 0.554)
Time Warner Inc	-115.5 (-119.0, -114.0)	-115.1 (-118.7, -113.5)	0.510 (0.502, 0.535)

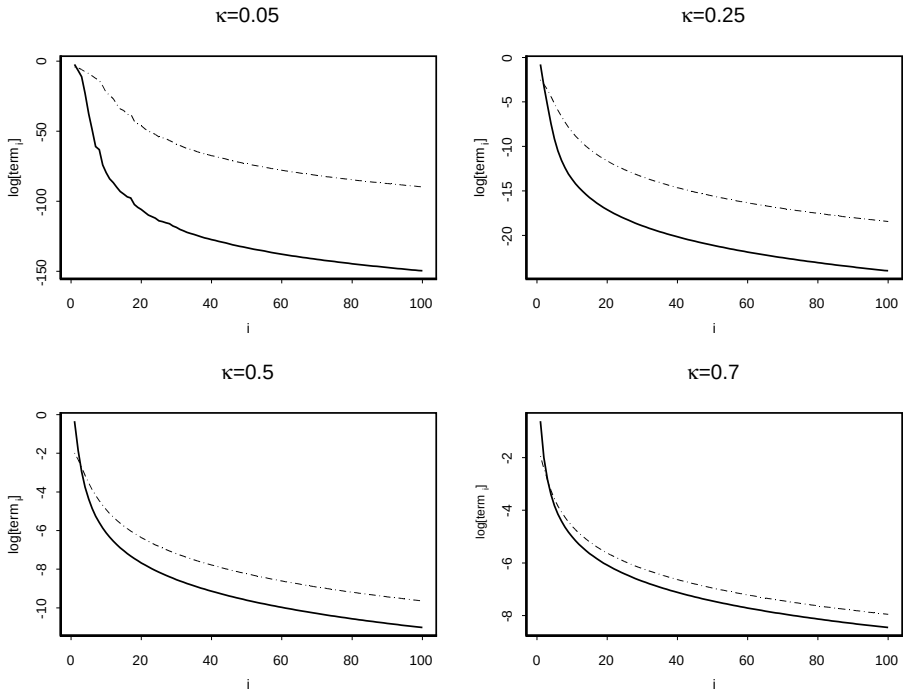

Fig. 1. Graphs of the log of the individual terms for the Barndorff-Nielsen and Shephard (2000) and Rosiński (2000) series representation for the $TS(\kappa, 1, 1)$ distribution.

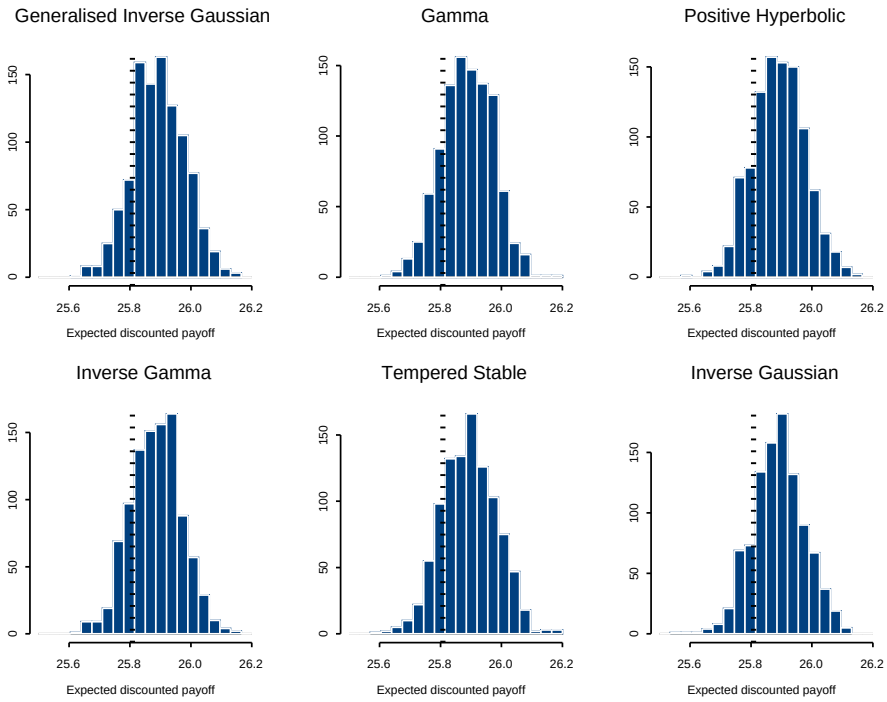

Fig. 2. Histograms of the expected discounted payoff for different posterior samples from the MCMC, for an arithmetic Asian option on Microsoft shares. The dashed line is the actual discounted payoff.

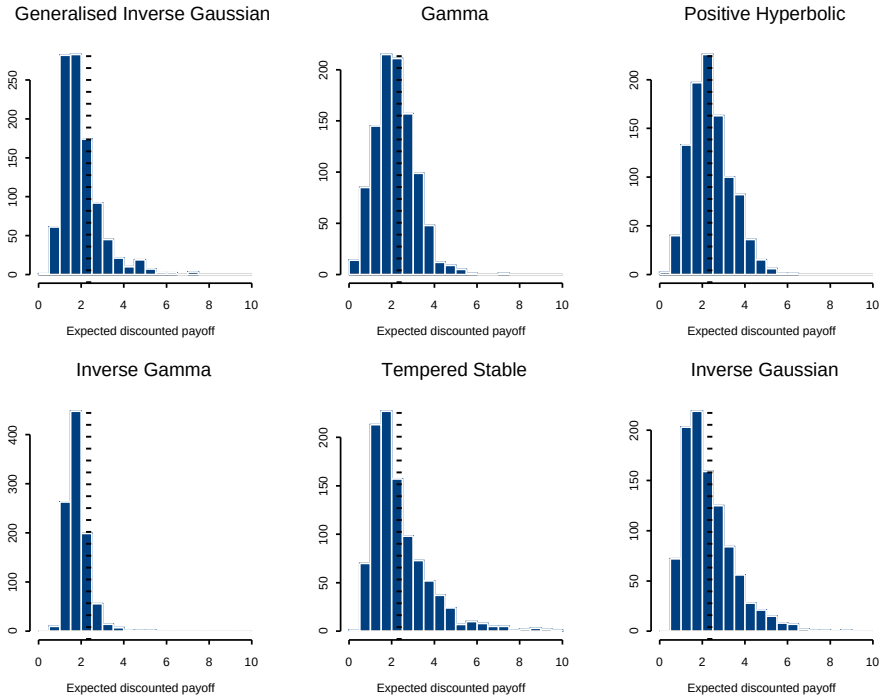

Fig. 3. Histograms of the expected discounted payoff for different posterior samples from the MCMC, for knock in option with vanilla call payoff on Procter & Gamble Co shares. The dashed line is the actual discounted payoff.

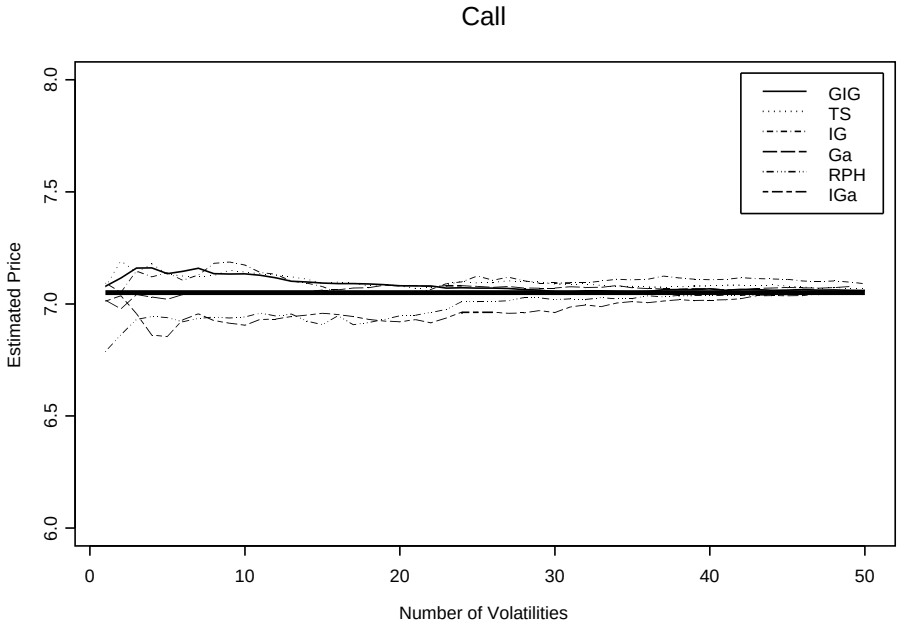

Fig. 4. Graphs of the estimated fair price of vanilla call for constant volatility, $\sigma = 0.03$.

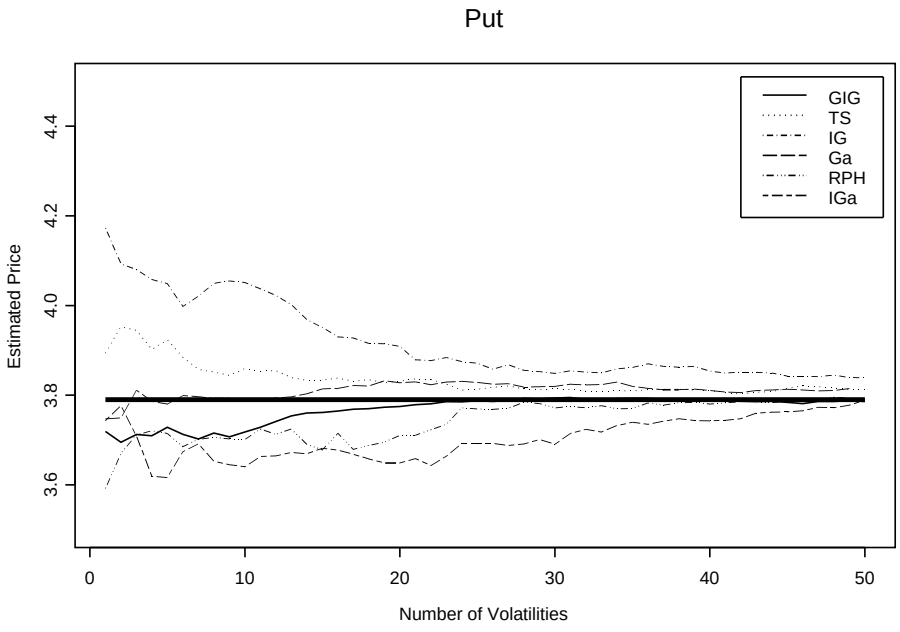

Fig. 5. Graphs of the estimated fair price of vanilla put for constant volatility, $\sigma = 0.03$.

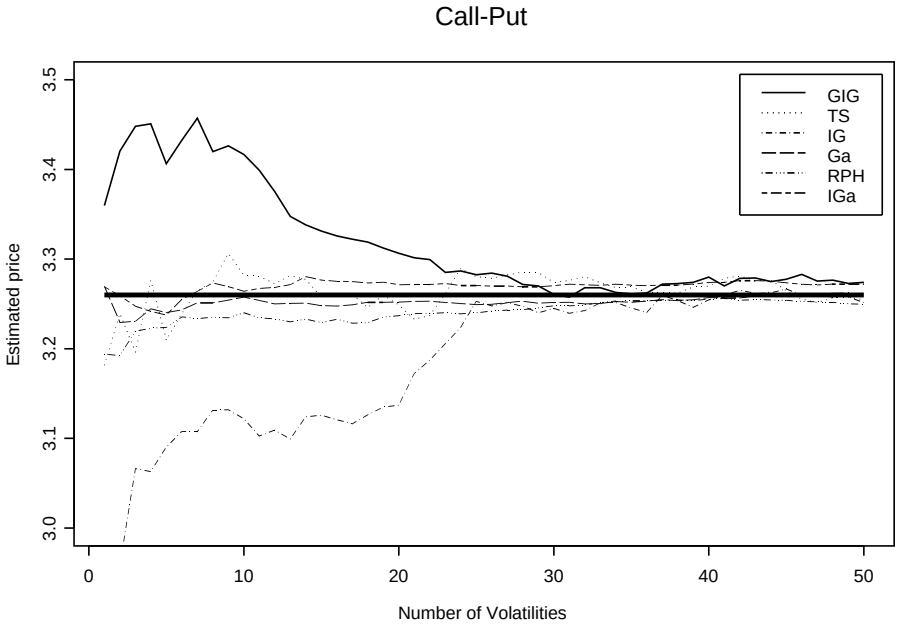

Fig. 6. Graphs of the estimated fair price of vanilla call-put for constant volatility, $\sigma = 0.03$.

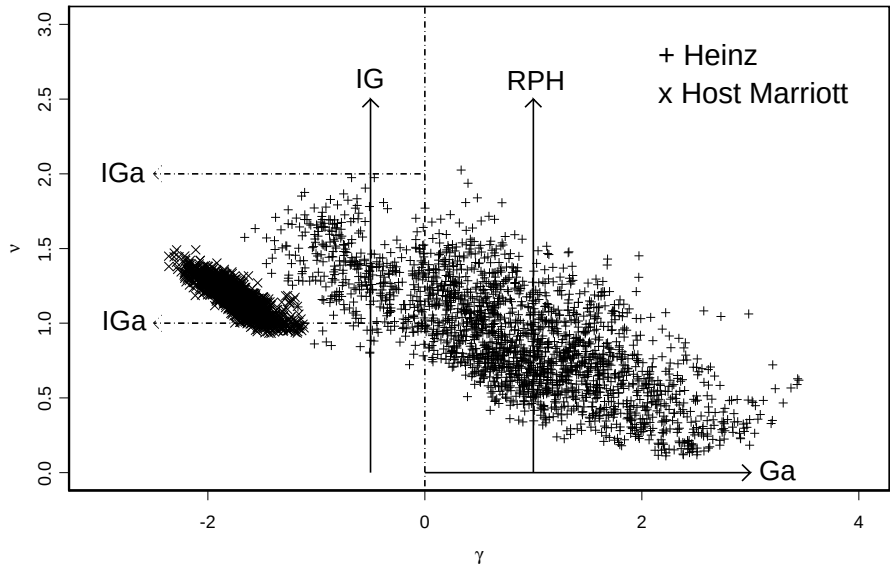

Fig. 7. Posterior samples in the (γ, ν) plane for the Heinz and Host Marriott data sets.

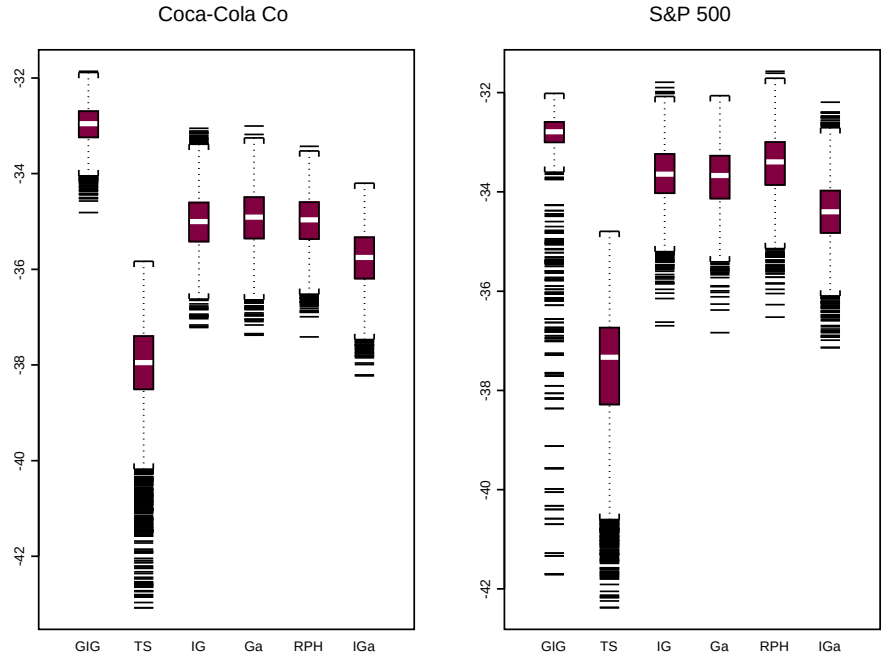

Fig. 8. Boxplots of samples from the predictive densities of the six different marginals on Coca-Cola Co and S&P 500 data.

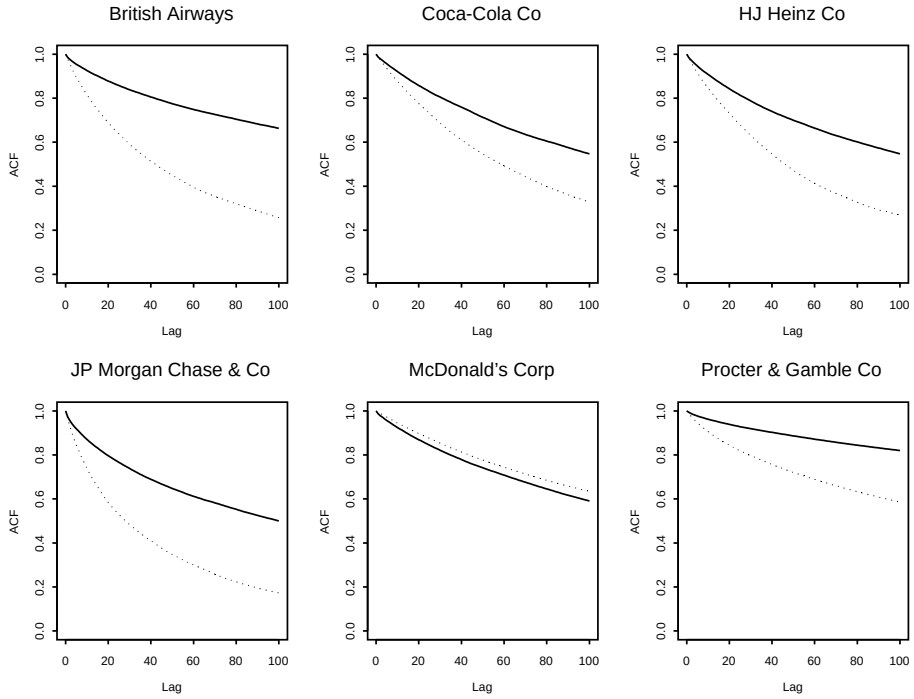

Fig. 9. ACF($\nu\alpha$) plots of the Hybrid MCMC algorithm of Roberts et al. (2004) (solid line) and our algorithm (dashed line).

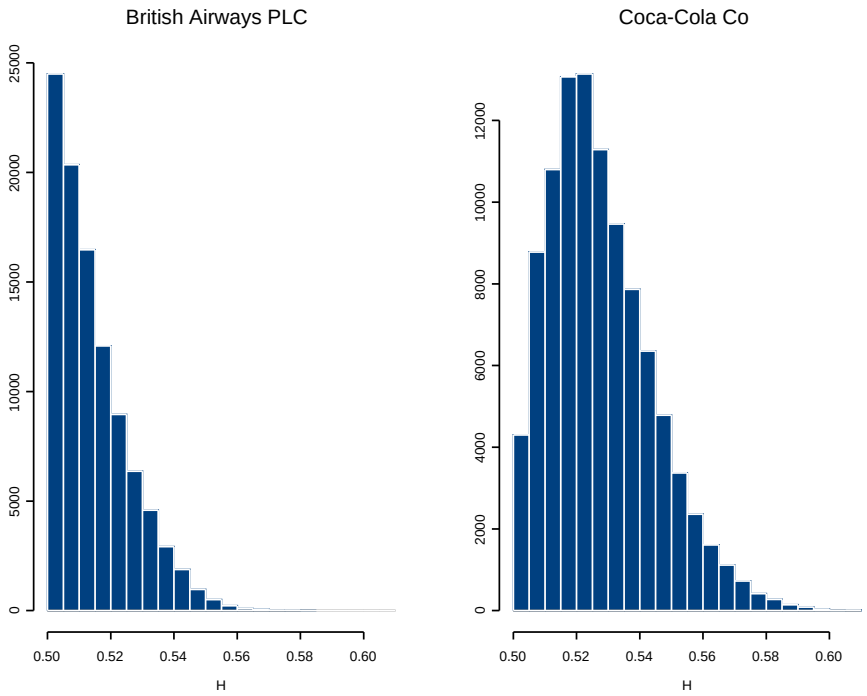

Fig. 10. Posterior histograms of the Hurst parameter for British Airways PLC and Coca-Cola Co.