Inference for Stochastic Volatility Models based on Lévy Processes

by

Matthew Peter Sandford Gander

A thesis submitted for the degree of *Doctor of Philosophy of the University of London* and for the *Diploma of Imperial College*

Department of Mathematics
Section of Statistics
Imperial College London
University Of London
180 Queen's Gate, London SW7 2BZ, England

November 2004

Abstract

The standard Black-Scholes model is a continuous time model to predict asset movement. For the standard model, the volatility is constant but frequently this model is generalised to allow for stochastic volatility (SV). As the Black-Scholes model is a continuous time model, it is attractive to have a continuous time stochastic volatility model and recently there has been a lot of research into such models.

One of the most popular models was proposed by Barndorff-Nielsen and Shephard (2001b) (BNS), where the volatility follows an Ornstein-Uhlenbeck (OU) equation and is driven by a background driving Lévy process (BDLP). The correlation in the volatility decays exponentially and so the model is able to explain the volatility clustering present in many financial time series. This model is studied in detail, with assets following the Black-Scholes equation with the BNS SV model.

Inference for the BNS SV models is not trivial, particularly when Markov chain Monte Carlo (MCMC) is used. This has been implemented in Roberts et al. (2004) and Griffin and Steel (2003) where a *Gamma* marginal distribution for the volatility is used. Their focus is on the difficult MCMC implementation and the performance of different proposals, mainly using training data generated from the model itself. In this thesis, the four main new contributions to the Black-Scholes equation with volatility following the BNS SV model are as follows:-

(1) We perform the MCMC inference for *Generalised Inverse Gaussian* and *Tempered Stable* marginal distributions, as well as the special cases, the *Gamma*, *Positive Hyperbolic*, *Inverse Gamma* and *Inverse Gaussian* distributions.

(2) Griffin and Steel (2003) consider the superposition of several BDLPs to give quasi long-memory in the volatility process. This is computationally problematic and so we allow the volatility process to be non-stationary by allowing one of the parameters, which controls the correlation in the volatility process, to vary over time. This allows the correlation of the volatility to be non-stationary and further volatility clustering.

(3) The standard Black-Scholes equation is driven by Brownian motion and a generalisation of this allowing for long-memory in the share equation itself (as opposed to the volatility equation), which is based on an approximation to fractional Brownian motion, is considered and implemented.

(4) We introduce simulation methods and inference for a new class of continuous time SV models, with a more flexible correlation structure than the BNS SV model.

For each of (1), (2) and (3), our focus is on the empirical performance of different models and whether such generalisations improve prediction of future asset movement. The models are tested using daily Foreign Exchange rate and share data for various different countries and companies.

Acknowledgements

I would most like to thank my supervisor, Dr. D. Stephens, for all his help and suggestions throughout my PhD; both academic and non-academic.

I would like to thank Prof. N. Shephard, Prof. R. Wolpert, Prof. P. Dellaportas and Dr. J. E. Griffin for their advice and ideas on Lévy processes and various other technical problems I ran into during my PhD. Additionally I would like to thank Dr. G. Moore for his help on numerical algorithms and, in particular, Gaussian Quadrature.

I would also like to thank Dr. N. Heard for his coding and statistics help, as well as my office mates, Adrien, Ajay, Cheng, Kitty and Ross for making Imperial an interesting and fun place to work.

I am also grateful to P. Barrow and Dr. G. Pruess for their coding help, as the implementation of many of the MCMC methods can be very expensive computationally.

I thank my two examiners, Prof. N. Shephard and Dr. E. McCoy for suggesting useful alterations to the thesis.

Lastly I would like to thank my parents, Peter and Carole, for their continued support throughout my life at university.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The main practical focus of this thesis is on the generalisation of the Black-Scholes equation to predict asset movement in statistical finance. The main methodological focus is the attempt to characterise different financial processes. The performance of different models is tested empirically on real data. Before introducing these models, the standard Black-Scholes equation is introduced and some of the characteristics, which it is unable to explain in real financial data, are given. A brief discussion of how the Black-Scholes equation could be generalised is given in Section 1.3.2. For such generalisations, inference will be performed using Markov chain Monte Carlo (MCMC) and this is described in Section 1.4.1. Options can then be priced using Monte Carlo integration, which is documented in Section 1.4.2. Section 1.5 gives an overview of the remaining chapters.

Initially a glossary of important definitions that will be used in this thesis, along with some useful theorems, is given.

1.1 Definitions

Many of the following definitions can be found in Bertoin (1994), Grimmett and Stirzaker (2001), Lukacs (1970) and Rogers and Williams (2000).

Definition 1 *The **Characteristic function**, $\phi(\underline{t})$, of a distribution on $\mathbb{R}^d$, with density $f_{\underline{X}}(\underline{x})$, is defined by*

$$\phi(\underline{t}) = \int_{\mathbb{R}^d} e^{i\underline{t}.\underline{x}} f_{\underline{X}}(\underline{x})\, d\underline{x}, \qquad \underline{t} \in \mathbb{R}^d.$$

Definition 2 *A distribution, with characteristic function $\phi(\underline{t})$, is said to be **infinitely divisible** if $\forall n \in \mathbb{Z}^+$ there exists a distribution with characteristic function $\phi_n(\underline{t})$ such*

that

$$\phi\left(\underline{t}\right) = \left\{\phi_n\left(\underline{t}\right)\right\}^{1/n}.$$

Definition 3 *If a distribution is infinitely divisible then the* ***characteristic exponent****,* $\Psi\left(\underline{t}\right)$*, is continuous and is defined by*

$$\phi\left(\underline{t}\right) = \exp\left\{-\Psi\left(\underline{t}\right)\right\}.$$

This is the definition of Bertoin (1994).

Definition 4 *If* $\phi\left(\underline{t}\right)$ *is the characteristic function of a random variable* $\underline{X}$*, then* $\underline{X}$ *is* ***self-decomposable*** *if for all* $c \in (0,1)$ *there is some family of characteristic functions* $\{\phi_c : c \in (0,1)\}$ *such that*

$$\phi\left(\underline{t}\right) = \phi\left(c\underline{t}\right)\phi_c\left(\underline{t}\right).$$

All self-decomposable distributions are infinitely divisible (see Barndorff-Nielsen and Thorbjørnsen (2002)).

Definition 5 *Let* $X_1, \ldots, X_n$ *be observations of a discrete time random process with common mean,* $E\left[X_i\right] = \mu$*, and variance,* $V\left[X_i\right] = \sigma^2 < \infty$ *and define*

$$\rho\left(i,j\right) = \frac{E\left[\left(X_i - \mu\right)\left(X_j - \mu\right)\right]}{\sigma^2}.$$

If

$$\rho\left(i,j\right) = \rho\left(i-j\right) = \rho\left(j-i\right)$$

then X_t *is a* ***weakly stationary process****. The correlation between two observations* X_i *and* X_j *is a function of* $|i-j|$ *(the distance between the observations).*

Definition 6 *For a stationary process,* X_t*, if* $\exists\ \alpha \in (0,1)$ *and a constant* $c \in \mathbb{R}^+$ *such that*

$$\lim_{k\to\infty} \frac{\rho\left(k\right)}{ck^{-\alpha}} = 1$$

then X_t *is a stationary process with* ***long-memory*** *or* ***long range dependence****. A process which has* ***short-memory*** *or* ***short range dependence*** *is a process which does not have long-memory.*

Definition 7 *The* ***Hurst Parameter****,* H*, is defined as* $H = 1 - \frac{\alpha}{2}$*, where* α *is defined in Definition 6. We then have long-memory for* $\frac{1}{2} < H < 1$*.*

Definition 8 *$\{Y_t\}$ is an **integrated process** of order 0, written $Y_t \sim I(0)$, if Y_t is stationary. $\{Y_t\}$ is an integrated process of order 1, written $Y_t \sim I(1)$, if it can be represented as*

$$Y_t = Y_{t-1} + u_t,$$

where u_t is a stationary time series.

Definition 9 *A function, $f(x)$, is said to have **bounded variation** on $[a,b]$ if there exists a finite M such that*

$$|f(x_1) - f(a)| + |f(x_2) - f(x_1)| + \cdots + |f(b) - f(x_n)| \leq M$$

for all $a < x_1 < x_2 < \cdots < x_n < b$.

Definition 10 *A real-valued stochastic process $\{B_t : t \in \mathbb{R}^+\}$ is a **Brownian motion** if the following hold:-*

(i) $B_0(\omega) = 0$, $\forall \omega$.

(ii) the map $t \longmapsto B_t(\omega)$ *is a continuous function of* $t \in \mathbb{R}^+$, $\forall \omega$.

(iii) $\forall t, h \in \mathbb{R}^+$, $B_{t+h} - B_t \sim N(0,h)$ *and is independent of* $\{B_u : 0 \leq u \leq t\}$.

Definition 11 *A real valued process $(X_t)_{t \in T}$ indexed by some set T is said to be a **Gaussian process** if, for any $t_1, \ldots, t_n \in T$, the law of $(X(t_1), \ldots, X(t_n))$ is a multivariate Gaussian.*

Definition 12 *$\left\{B_t^H : t \in \mathbb{R}\right\}$ is **fractional Brownian motion** (fBm), with index $H \in (0,1)$, if it is a zero mean Gaussian process with covariance function*

$$\rho(t,s) = \frac{1}{2}\left(|t|^{2H} + |s|^{2H} - |t-s|^{2H}\right)$$

and, when $H = 0.5$, we have $\rho(t,s) = \min(t,s)$ and standard Brownian motion is recovered.

Definition 13 *If $W(t)$ is Brownian motion with $W(1) = 0$ then $\{W(t) : 0 \leq t \leq 1\}$ is called a **Brownian bridge** (i.e. it is Brownian motion over a unit time interval, where the end points of the process are both fixed at 0).*

Definition 14 *A stochastic process $z(t)$, with $z(0)=0$, is a **Lévy process** if for every $s,t\geq 0$, the increment $z(t+s)-z(t)$ is independent of the process $\{z(u),0\leq u\leq t\}$. If these increments have the same distribution as $z(s)$ (so the increments are stationary), then the process is a **homogeneous Lévy process** and if the increments do not have the same distribution as $z(s)$ we have an **non-homogeneous Lévy process**.*

Definition 15 *A **subordinator** is a Lévy process which is positive. i.e.*

$$z(t)\geq 0 \qquad \forall t\geq 0.$$

Subordinators are non-decreasing.

Definition 16 *$X(t)$ is an **Ornstein-Uhlenbeck process** if it is a solution to the Ornstein-Uhlenbeck equation*

$$dX(t)=\alpha(\mu-X(t))\,dt+\sigma dz(t),$$

where $\alpha,\mu,\sigma\in\mathbb{R}$ and $z(t)$ is a Lévy process.

1.2 Theorems

Theorem 1 *The random variable x is self-decomposable if and only if there is a stationary stochastic process, $x(t)$, and a Lévy process, $z(t)$, such that $x(t)\overset{\mathfrak{L}}{=}x$ (so the stochastic process $x(t)$ has the same marginal distribution as x) and*

$$\begin{aligned} x(t) &= \int_{-\infty}^{t}\exp\{-\lambda(t-s)\}\,dz(\lambda s) \\ &= \int_{-\infty}^{0}\exp(\lambda s)\,dz\{\lambda(t+s)\} \\ &= \int_{-\infty}^{0}\exp(s)\,dz\{\lambda t+s\} \end{aligned}$$

for all $\lambda>0$ (see Wolfe (1982) for details).

Theorem 2 *A distribution, with density $f_{\underline{X}}(\underline{x})$ and characteristic exponent $\Psi(\underline{\lambda})$, is infinitely divisible if and only if there exists some $\underline{a}\in\mathbb{R}^d$, a positive semi-definite quadratic Q on $\mathbb{R}^d$ and some measure $u(x)$ on $\mathbb{R}^d/\{0\}$ such that $\forall\underline{t}\in\mathbb{R}^d$*

$$\Psi(\underline{t})=i\underline{a}.\underline{t}+\frac{1}{2}Q(\underline{\lambda})+\int_{\mathbb{R}^d}\left(1-e^{i\underline{t}.\underline{x}}+i\underline{t}.\underline{x}\mathbf{1}_{|\underline{x}|<1}\right)u(\underline{x})\,d\underline{x} \tag{1.1}$$

and

$$\int_{\mathbb{R}^d}\left(1+|\underline{x}|^2\right)u(\underline{x})\,d\underline{x}<\infty.$$

Equation (1.1) is called the ***Lévy-Khintchine formula****, $u(\underline{x})$ is the* ***Lévy measure*** *of $f_{\underline{X}}(\underline{x})$ and $Q(\underline{t})$ is the* ***Gaussian coefficient****. $\{\underline{a}, Q(\underline{t}), u(\underline{x})\}$ is called the* ***Lévy triplet*** *of the distribution $f_{\underline{X}}(\underline{x})$. See Sato (1999) for a detailed proof of this theorem.*

If $\underline{x} \in \mathbb{R}_+^d$ then $\underline{a} \geq 0$, $Q(\underline{t}) = \underline{0}$ and

$$\int_{\mathbb{R}^d} i\underline{t}.\underline{x}\mathbf{1}_{|\underline{x}|<1} u(\underline{x}) < \infty,$$

so the $i\underline{t}.\underline{x}\mathbf{1}_{|\underline{x}|<1}$ term can be absorbed into $\underline{a}$.

Bertoin (1994) has proved that in the one dimensional case

$$\lim_{|t|\to\infty} \frac{\Psi(t)}{t^2} = \frac{Q}{2} \geq 0. \tag{1.2}$$

If the Lévy process has bounded variation (see Definition 9), then $Q = 0$ and equation (1.1) can be rewritten as

$$\Psi(\underline{t}) = i\underline{d}.\underline{t} + \int_{\mathbb{R}^d} \left[1 - e^{i\underline{t}.\underline{x}}\right] u(\underline{x})\, d\underline{x}$$

and in one dimension with bounded variation

$$\lim_{|t|\to\infty} \frac{\Psi(t)}{t} = id. \tag{1.3}$$

If $f_{\underline{X}}(\underline{x})$ is a one dimensional density on $\mathbb{R}^+$, Theorem 2 can be rewritten as the following theorem (see Rogers and Williams (2000), pg 78).

Theorem 3 *A one dimensional distribution D on $\mathbb{R}^+$, with probability density function (pdf) $f_X(x)$, is infinitely divisible if and only if, $\forall t \in \mathbb{R}^+$, there is a representation*

$$\int_0^\infty e^{-tx} f_X(x)\, dx = \exp\left\{-ct - \int_0^\infty \left(1 - e^{-tx}\right) u(x)\, dx\right\}$$

for some $c \geq 0$ and measure $u(x)$ on $(0, \infty)$ satisfying

$$\int_0^\infty \min(1, x)\, u(x)\, dx < \infty,$$

where $u(x)$ is called the Lévy measure or Lévy density of D. In the literature, the Lévy measure or density are sometimes defined as $u(x)\, dx$ which can be abbreviated to $u(dx)$.

For examples of derivations of Lévy measures of standard distributions on $\mathbb{R}^+$ see Barndorff-Nielsen and Shephard (2001b) and Section 2.2.

The Lévy-Khintchine formula (equation (1.1)) can be expressed in various different forms. One such form can be obtained from the canonical representation of an infinitely divisible distribution as follows. This is only given in one dimension (which is where our interest is focused).

Theorem 4 *(Feller (1971), pg 563) A univariate distribution, with characteristic function $\Psi(t)$, is infinitely divisible if and only if there exists some $b \in \mathbb{R}$, such that for all $\lambda \in \mathbb{R}$, $\Psi(\lambda)$ can be expressed in the form*

$$\Psi(t) = \int_{-\infty}^{\infty} \left\{ \frac{1 + it\sin(x) - e^{itx}}{x^2} \right\} M(x)\, dx + ibt,$$

where $M(x)$ is a measure such that

$$\int_{\alpha}^{\beta} M(x)\, dx$$

is finite for all finite α and β and the following two integrals converge for all $x > 0$

$$M^{+}(x) = \int_{x-}^{\infty} \frac{M(y)}{y^2} dy$$

and

$$M^{-}(-x) = \int_{-\infty}^{-x+} \frac{M(y)}{y^2} dy.$$

$M(x)$ is the canonical measure.

The relationship between the Lévy and canonical measure is $u(x) = M(x)\, x^{-2}$ for $x \neq 0$. The canonical measure, $M(x)$, is defined at the origin, unlike the Lévy measure, $u(x)$, with which we shall be mainly concerned. Note that in Feller (1971) the convention used is $\Psi(t) = \log(\phi(t))$ but we use $\Psi(t) = -\log(\phi(t))$.

The Lévy measure, $u(x)$, appears in another canonical representation (Feller (1971), pg 564)

$$\Psi(t) = it\lambda + \frac{\sigma^2}{2}t^2 + \lim_{\delta \to 0} \int_{|x|>\delta} \left\{1 - e^{itx} - it\sin(x)\right\} u(x)\, dx.$$

Here $u(x)$ is not defined at zero and everywhere else is again equal to $M(x)\, x^{-2}$. Further equivalent representations for the Lévy measure can be found in Sato (1999).

1.3 Black-Scholes equation

The Black-Scholes equation, to model the movement of an asset S, is

$$dS(t) = \mu S(t)\, dt + \sigma S(t)\, dW(t), \tag{1.4}$$

where μ (drift) and σ (volatility) are constants and $W(t)$ is Brownian motion, so $dW(t) \sim N(0, dt)$ (see Definition 10). In a risk neutral world, in the absence of arbitrage, the drift must be equal to the risk free interest rate, r (see Hull (2000)). The returns are defined by $dS(t)$. Much modern option pricing is based on variations of this model.

Equation (1.4) has two components; a deterministic part and a random part. It is a sensible model for asset movement for the following reasons:-

(1) The deterministic part ($dS(t) = \mu S(t)\, dt$) forces $S(t)$ to increase (or decrease) exponentially at a rate determined by μ. If μ is equal to the risk free interest rate, then without the Brownian motion, $S(t)$ will move as if the equivalent amount of money had been invested in a bank. This is important as it prevents arbitrage opportunities (so there is no expected difference in investing in the share or putting the money in a bank).

(2) The random part ($dS(t) = \sigma S(t)\, dW(t)$) gives continuous sample paths for $S(t)$, as well as forcing the share to move up or down with equal probability. As the Brownian motion is multiplied by $S(t)$, we have a geometric solution, which forces $S(t)$ to remain positive.

(3) It is reasonable to assume that the variance of the return over the time period $(t, t + dt)$ is proportional to dt (i.e. the length of the time interval). This is true for equation (1.4) as $dW(t) \sim N(0, dt)$.

Black and Scholes (1973) have derived a differential equation, which the price of any option must satisfy, if it is assumed to follow equation (1.4). As we consider generalising equation (1.4), option prices for the generalised models need not obey the differential equation derived in Black and Scholes (1973).

We shall be interested in capturing certain properties of the log returns, so let $x(t) = \log[S(t)]$. Then $x(t)$ is the log of the asset and equation (1.4) implies it satisfies

$$dx(t) = \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW(t). \tag{1.5}$$

This is simple to derive and is included in Appendix A.1. In practice, the dt term can often be integrated out.

The log returns, $\underline{y} = \{y_1, \ldots, y_T\}$, each separated by Δ days, are

$$y_i = \int_{(i-1)\Delta}^{i\Delta} dx(t).$$

Define the discretely observed volatility as

$$\sigma_i^2 = \int_{(i-1)\Delta}^{i\Delta} \sigma^2(u)\, du, \tag{1.6}$$

which is $\sigma^2\Delta$ for $\sigma^2(u) = \sigma^2$, and let

$$a_i = \frac{y_i - \left(\mu\Delta - \frac{\sigma_i^2}{2}\right)}{\sigma_i},$$

then $a_i \overset{iid}{\sim} N(0,1)$ and, if inference about μ and σ^2 is required, the likelihood for $\underline{y}$ is given by

$$f_{\underline{Y}}(\underline{y}) = \prod_{i=1}^{T} \frac{1}{\sigma\sqrt{\Delta}} f_{A_i}(a_i).$$

1.3.1 Weaknesses in the Black-Scholes formulation

The standard Black-Scholes model works well for predicting asset movement over long time periods where the *normal* tail behaviour is realistic. For small time periods, it is unable to describe some observed properties of financial data. These properties include:-

(1) Log returns have heavier tails than *normal* (see Fama (1965)), particularly for high frequency data. Equation (1.5) gives Gaussian tail behaviour.

(2) There is high correlation in the absolute value of the log returns. The Black-Scholes model with constant volatility generates uncorrelated log returns (and any function of the log returns), as the volatility is constant and W is Brownian motion (which has independent increments). An example of this property in real data is given in the first graph of Figure 1.1, which is the autocorrelation function (ACF) of the absolute value of the log returns of the US Dollar vs Turkish Lira exchange rate from 6^{th} March 1993 to 1^{st} December 2003.

(3) There is volatility clustering - the asset has periods of high and low activity. Equation (1.5) gives constant volatility. An example of this property for real data is given in the second graph of Figure 1.1, which is a plot of the log returns of the Turkish Lira from 6^{th} March 1993 to 1^{st} December 2003.

1.3.2 Extensions to the standard Black-Scholes equation

We try to amend equation (1.5) to capture some observed properties of financial data which were given in Section 1.3.1. The exact details of the generalisations can be found in Chapters 3, 4 and 6 and Appendix D and are outlined here.

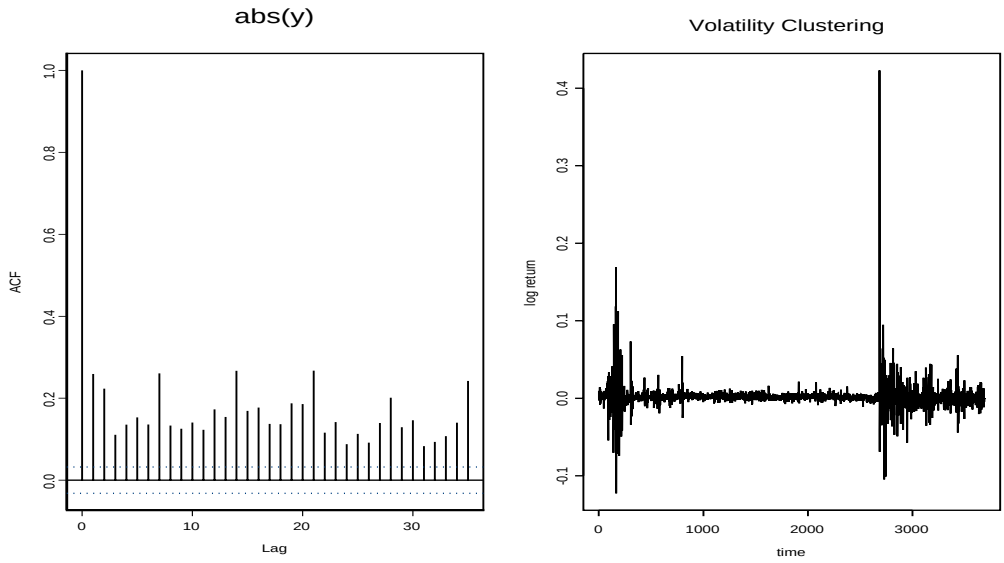


Figure 1.1: Two graphs which demonstrate correlation in the absolute value of the log returns and volatility clustering for the US Dollar vs Turkish Lira exchange rate.

To increase the weights of the tails of the log returns, allow the volatility to vary over time, so

$$dx\left(t\right)=\left(\mu-\frac{\sigma^{2}\left(t\right)}{2}\right)dt+\sigma\left(t\right)dW\left(t\right). \tag{1.7}$$

If we are able to specify the marginal distribution of the volatility, ignoring the dt term (which can often be integrated out), the kurtosis of the log returns, Y, will then be (see Appendix A.2)

$$K_{Y}=3\frac{E\left[\sigma^{4}\left(t\right)\right]}{\left(E\left[\sigma^{2}\left(t\right)\right]\right)^{2}}.$$

The kurtosis will then be greater than three and can be controlled by the form of the marginal distribution imposed on $\sigma^{2}\left(t\right)$.

To try to induce correlation in the absolute value of the log returns, as well as volatility clustering, we would like a volatility model which generates correlated $\sigma^{2}\left(t\right)$, where the strength of this correlation can be controlled.

The model proposed in Barndorff-Nielsen and Shephard (2001b) (referred to as the BNS SV model) is a very flexible model and satisfies all the previously mentioned requirements. Here the volatility follows the Ornstein-Uhlenbeck equation

$$d\sigma^{2}\left(t\right)=-\lambda\sigma^{2}\left(t\right)dt+dz\left(\lambda t\right), \tag{1.8}$$

where $z\left(t\right)$, called the background driving Lévy process (BDLP), is a subordinator with

$z(0) = 0$ (see for example Sato (1999) and Rogers and Williams (2000) for an introduction to Lévy processes) and λ is a positive constant controlling the strength of the correlation in the volatility process. The timing of the subordinator ensures λ does not alter the marginal distribution of $\sigma^2(t)$. Further details on this process are given in Section 3.2 and technical details are given in Appendix D.

The Lévy process is constant apart from when it has positive jumps. A realisation from a Lévy process is given in Figure 1.2.

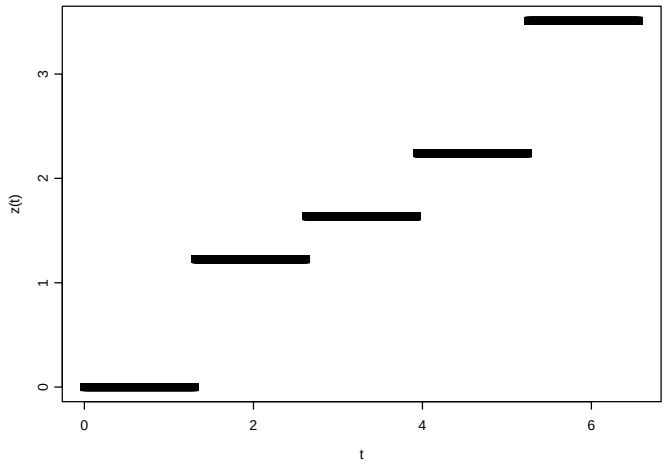


Figure 1.2: Graph of a typical homogeneous Lévy process, with $Ga(1,1)$ jumps, which might drive the Ornstein-Uhlenbeck equation. The jump times are the arrival times of a Poisson process of unit intensity.

Later on, to add more flexibility to the model, correlated increments are used to drive the asset equation (1.7). This is introduced in Section 3.5 and will induce correlation in the absolute value of the log returns. This reduces the amount of correlation that the volatility model must capture.

In Section 3.3, equation (1.8) is generalised by allowing the correlation parameter, λ, to vary over time. This gives the volatility process a richer correlation structure.

1.4 Parameter estimation

The stochastic volatility model used in Barndorff-Nielsen and Shephard (2001b) has latent parameters which specify the BDLP. It is impossible to write out the likelihood function

for the volatility given the other non-latent parameters (such as λ in equation (1.8) and parameters controlling the exact form of the marginal distribution of $\sigma^2(t)$). The only likelihood available directly is $p(y_i|\sigma_i, \mu)$, where the volatility, σ_i, is not observed but is specified by the latent and non-latent parameters. Various likelihood based methods to estimate the non-latent parameters are discussed in Barndorff-Nielsen and Shephard (2001b). We shall use MCMC: an overview of this is now given.

1.4.1 Markov chain Monte Carlo

Let $\underline{\theta} = \{\theta_1, \ldots, \theta_n\}$ be n unknown parameters and let $p(\underline{\theta})$ be our prior belief about the distribution of θ before data, $\underline{y} = \{y_1, \ldots, y_T\}$, are observed. MCMC is an algorithm which samples from the posterior distribution, $p\left(\underline{\theta}|\underline{y}\right)$, of the parameters given the data, $\underline{y}$, and the prior distribution, $p(\underline{\theta})$. This is used in Chapters 4 and 5 to make inference about the unknown parameters of the financial models.

Markov chain Monte Carlo is an iterative procedure, where a Markov chain is constructed with a stationary or target distribution equal to the posterior distribution, $p\left(\underline{\theta}|\underline{y}\right)$, which we are interested in sampling from. Let s_t be the state of the chain at time t. The chain is Markov because the state at time $t+1$, s_{t+1}, is only dependent on the state at time t, s_t, and no previous states. After the chain is run for a sufficiently long period of time, the algorithm samples from the stationary distribution, though these samples are correlated. The key step is to ensure that moves are proposed from s_t to s_{t+1} which guarantee the required stationary distribution (i.e. the posterior distribution of the unknown parameters, $\underline{\theta}$, in light of observed data, $\underline{y}$). There are two standard techniques to do this. The first method to be developed was the *Metropolis-Hastings* (MH) sampler (see Metropolis et al. (1953) and Hastings (1970)) and is described in Section 1.4.1. The second standard method is the *Gibbs* sampler (see Geman and Geman (1994)), where parameters are updated in order, from their full posterior conditional distributions. Although the Gibbs sampler is not strictly used in our MCMC; parameters are updated in order, though not from their full posterior conditional distributions (a Metropolis-Gibbs hybrid). An outline of the Metropolis-Hastings algorithm is now given (the Metropolis-Gibbs hybrid is similar but with parameters updated in order). MCMC produces correlated samples from the posterior distribution and if this correlation is high, the samples can be thinned before they are used, to reduce the correlation.

Metropolis-Hastings algorithm

Choose any starting state, s_0, for the unknown parameters, $\underline{\theta}$ and set $i = 1$. Given the current state, s, propose a new state, $s^{'}$, with probability $q\left(s \rightarrow s^{'}\right)$. Set $s_i = s^{'}$ with probability

$$\alpha\left(s, s^{'}\right) = \min\left[1, \frac{p\left(s^{'}|y\right)}{p\left(s|y\right)} \frac{q\left(s^{'} \rightarrow s\right)}{q\left(s \rightarrow s^{'}\right)}\right],$$

otherwise set $s_i = s$. Increment i and repeat.

Running this chain for a sufficiently long period of time samples dependent variates from the stationary distribution of the chain, $p\left(\underline{\theta}|\underline{y}\right)$.

MCMC for variable dimension problems (reverse jump MCMC)

The MCMC algorithm described above can be used when the number of unknown parameters is fixed. For some of the problems that are later considered, the number of unknown parameters is also an unknown. Reverse jump MCMC (see Green (1995)), abbreviated RJMCMC, can be used to jump between dimensions and is similar in many ways to the standard MH algorithm. Note that when we are not proposing to change the dimension of the problem (even if the problem is variable dimension), the standard MH algorithm can be used to update parameters (though this will not give samples from the posterior distribution, $p\left(\underline{\theta}|\underline{y}\right)$, without the inclusion of a reverse jump move).

Let the chain be in state $s = \{s_1, \ldots, s_{d_1}\}$ of dimension d_1 and suppose a move to a higher dimension state, $s^{'} = \left\{s_1^{'}, \ldots, s_{d_1+d_2}^{'}\right\}$, of dimension $d_1 + d_2$, is proposed. Let $u = \{u_1, \ldots, u_{d_2}\}$ be the d_2 random variables that are generated and used in conjunction with s to generate $s^{'}$, so $s^{'} = g\left(s, u\right)$ (this function must be invertible).

Define the Jacobian transformation matrix, J, as

$$J_{i,j} = \left|\frac{\partial s_i^{'}}{\partial s_j}\right|, \quad \text{for} \quad 1 \leq j \leq d_1$$

and

$$J_{i,j} = \left|\frac{\partial s_i^{'}}{\partial u_{j-d_1}}\right|, \quad \text{for} \quad d_1 + 1 \leq j \leq d_2.$$

Accept state $s^{'}$ with probability

$$\alpha\left(s, s^{'}\right) = \min\left[1, \frac{p\left(s^{'}|y\right)}{p\left(s|y\right)} \frac{q\left(s^{'} \rightarrow s\right)}{q\left(s \rightarrow s^{'}\right)} |J|\right].$$

The acceptance probability for the jump down in dimension from state $s^{'}$ to s is

$$\alpha\left(s^{'}, s\right) = \min\left[1, \frac{p\left(s|y\right)}{p\left(s^{'}|y\right)} \frac{q\left(s \rightarrow s^{'}\right)}{q\left(s^{'} \rightarrow s\right)} \frac{1}{|J|}\right].$$

We frequently choose proposals where the Jacobian is the identity matrix, so the update is very similar to the MH update for a fixed dimension move. Running this chain for a sufficiently long period of time samples dependent variates from the stationary distribution of the chain, $p\left(\underline{\theta}|\underline{y}\right)$.

1.4.2 Monte Carlo integration

Deterministic numerical algorithms provide an efficient method to accurately evaluate low dimensional integrals. For higher dimensional integrals, Monte Carlo (MC) integration tends to provide more accurate estimates. For numerical integration, there are often deterministic bounds on the accuracy of the answer, whilst for Monte Carlo integration (using pseudo random numbers) the error for the integration is stochastic.

In Chapter 5, the expected payoffs of options are evaluated using MC integration. Here MCMC is used to sample from the posterior distribution of some unknown parameters, given observed financial data. These samples can be used to approximate the posterior distribution. We would like to evaluate the expected payoff of an option, given the model enforced on the volatility and the posterior samples of the parameters. To do this, we can directly simulate the share forward (this is called forward sampling) and take the average payoff. This is a form of MC integration which is briefly outlined below.

Let

$$E_p\left[f\left(x\right)\right] = \int f\left(x\right) p\left(x\right) dx$$

be an integral which is not available in analytical form and so must be evaluated computationally. To do this, generate a large sample, $x_1, \ldots, x_N$, from $p\left(x\right)$ (which might be our posterior sample), and estimate the integral as

$$E_p\left[f\left(x\right)\right] \approx \frac{1}{N} \sum_{j=1}^{N} f\left(x_j\right).$$

The estimate converges almost surely under standard regularity conditions.

The standard MC approach can be modified in the following ways. Importance sampling can be used to try to decrease the variance of the MC estimate (so it is more accurate) but, for our problem, we can sample directly from the share and, as a result, it is not easy to use such methods to improve the rate of convergence of the option pricing.

Quasi-random sampling (see Tezuka (1995) and Birge (1995)) is perhaps the easiest way to improve the rate of convergence of the Monte Carlo integration and uses random numbers that have a low discrepancy. Stratified sampling is a similar idea and can be used to improve the efficiency of the method (see for example Ribeiro and Webber (2003)) and these techniques can dramatically improve the rate of convergence (though are quite involved). Antithetic variables (which use random numbers that are negatively correlated) can also improve the rate of convergence. More details on these techniques, as well as other MC speed up methods can be found in Robert and Casella (2002).

Computing high-dimensional integrals becomes much more difficult very quickly as the dimension increases. For this reason, speed up techniques only allow the evaluation of slightly higher-dimensional integrals than before they were implemented and so they are not implemented.

1.5 Plan of thesis

Chapter 2 shows how to sample from Lévy processes and how these are related to the Lévy measure of the process. The Lévy measures of some standard distributions are derived for both Lévy processes with positive jumps - subordinators - (which is the focus of this thesis in later sections) as well as some processes with negative jumps. The technique to sample from four of the Lévy processes is discussed in detail and illustrations of these processes are given.

Chapter 3 considers whether a model to predict share movement should have long-memory. Real data sets are tested for long-memory in the log, absolute value and square of the log returns. A new and popular stochastic volatility model in continuous time is described, as well as an alteration to the original Black-Scholes equation to allow for leverage in the model. The results from the long-memory tests motivate a further generalisation of the standard Black-Scholes model.

The exact MCMC implementation details, for the models described in Chapter 3, are given in **Chapter 4**. The MCMC implementation is extensively tested on four training data sets and one real data set and the results are summarised. The chapter concludes by comparing the efficiency of our algorithm with the Hybrid algorithm of Roberts et al. (2004).

Chapter 5 describes how to compare the performance of the different models, using predictive densities and option pricing methods. First an introduction to predictive densities and option pricing is given, followed by the results of these tests, which are also discussed.

Chapter 6 introduces a new class of stochastic volatility models, based on the work of Wolpert and Taqqu (2004). This chapter focuses on how to simulate from these models and inference, and demonstrates their flexible correlation structure.

Chapter 7 details future work and extensions.

The **Appendix** is split into four parts:-

(A) Analytical Results.

(B) Numerical Algorithms.

(C) Simulation Results.

(D) Theory behind solutions of the Ornstein-Uhlenbeck equation.

Chapter 2

Lévy measures and Lévy processes

Our main use of Lévy processes is to drive the BNS SV model. This chapter introduces the basic principles of Lévy processes and focuses on how to sample from them. Initially the characteristics of Lévy processes are discussed, before the Lévy measures of some standard distributions are given, and simulation is discussed.

2.1 A note on characteristic functions and Lévy processes

If $X \sim N\left(\mu, \sigma^2\right)$ then the characteristic exponent is

$$\Psi_X(t) = -i\mu t + \frac{t^2\sigma^2}{2}.$$

Recall the Lévy-Khintchine formula in one dimension (equation (1.1)); the characteristic exponent of any infinitely divisible distribution can be written as

$$\Psi(t) = iat + \frac{1}{2}Qt^2 + \int_{\mathbb{R}} \left(1 - e^{itx} + itx\mathbf{1}_{|x|<1}\right) u(x)\, dx.$$

This is made up of the characteristic exponent of a $Normal(-a, Q)$ variate and an integral with respect to the Lévy measure, $u(x)$.

If we want to sample from a (possibly non-homogeneous) Lévy process which has a known distribution at time $t = \infty$, or a homogeneous Lévy process which has a known distribution at some time point (which is the focus in Section 3.2 and Chapter 4), then the process has three parts:-

(1) A deterministic drift part (from the iat term).

(2) A Brownian motion part (from the $\frac{1}{2}Qt^2$ term).

(3) A pure jumps part (from the remaining integral).

A rigorous description of this decomposition of Lévy processes can be found in Bertoin (1994). The BNS SV model is driven by a pure jumps Lévy process and this is why the focus of this chapter is simulating from pure jumps Lévy processes. Initially homogeneous Lévy processes (so the process has independent and stationary increments - see Definition 14) are considered but in Sections 2.3 and 3.3 non-homogeneous Lévy processes are investigated.

The most widely used Lévy process is Brownian motion and this only has a Brownian motion part (2). The simplest way to generate a Lévy process is to pick the distribution of the jumps, $z(t+s)-z(t)$, and simulate $z(i\Delta t)$ at discrete time intervals $\Delta t, 2\Delta t, 3\Delta t, \ldots$ for suitably small Δt. For Brownian motion (with variance σ^2), $z(1) \sim N\left(0, \sigma^2\right)$ and the jumps satisfy

$$z(t+\Delta t) - z(t) \sim N\left(0, \Delta t \sigma^2\right).$$

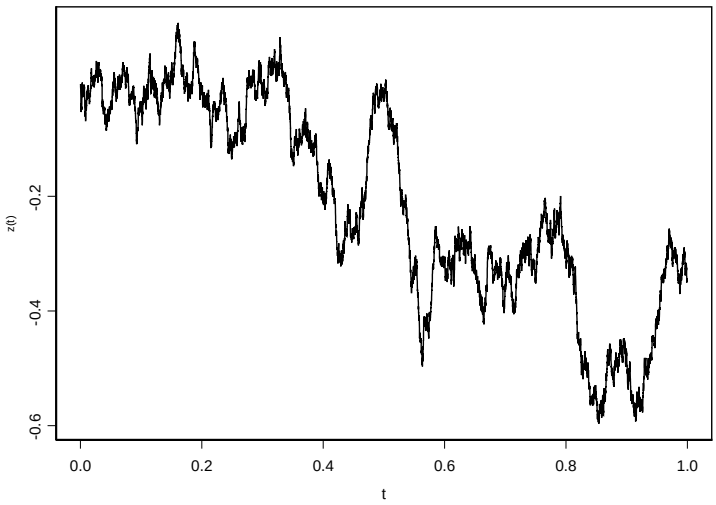


Figure 2.1: Simulated Brownian motion with $\sigma = 0.5$ and $\Delta t = 0.0001$.

This Lévy process is continuous almost surely and is the only such Lévy process (see Feller (1971)).

Another method to generate Lévy processes is given in Section 2.3, which uses the Lévy measure of the distribution from which we want to sample. The Lévy measures of some standard distributions are now given.

2.2 Lévy measures of some standard distributions

From Theorem 2, only infinitely divisible distributions have Lévy measures, so only standard distributions which are infinitely divisible are considered. The first two distributions are perhaps the most important, as they are very flexible distributions with support on $\mathbb{R}^+$ and are used as the marginal distribution of the stochastic volatility in Section 3.2.

An observation on the Lévy measures of shifted random variables (which take all values on $\mathbb{R}$) is now given, as it will be useful in deriving some of the Lévy measures.

2.2.1 Shifted random variables

Let X be an infinitely divisible random variable, which can take all values on $\mathbb{R}$ and has density, $f_X(x)$, characteristic function, $\phi_X(t)$, characteristic exponent, $\Psi_X(t)$, and Lévy measure, $u(x)$. Consider the random variable $Y = X + \mu$, where μ is some constant. The Lévy measure of Y, u_Y, therefore has the same domain as the Lévy measure of X, u_X. Then

$$\phi_Y(t) = e^{it\mu}\phi_X(t)$$

and so

$$\Psi_Y(t) = -it\mu + \Psi_X(t).$$

The $-it\mu$ term can be absorbed into "a" in the Lévy-Khintchine formula (equation (1.1)) and can therefore be ignored. Then

$$u_Y(x) = u_X(x), \qquad \text{for } x \in \mathbb{R}/\{0\}$$

and the Lévy measure is unaltered by the shift.

2.2.2 Convolution of distributions

Let X and Y be independent random variables (which are both infinitely divisible and so have Lévy measures) with densities $f_X(x)$ and $f_Y(y)$. Define the random variable Z as the convolution of X and Y, so $Z = X + Y$, then

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y)\, dy = \int_{-\infty}^{\infty} f_X(y) f_Y(z-y)\, dy.$$

Using the same notation as previously, this gives

$$\phi_Z(t) = \phi_X(t)\phi_Y(t)$$

and hence

$$\Psi_Z(t) = \Psi_X(t) + \Psi_Y(t),$$

so the Lévy measure of Z is

$$u_Z(x) = u_X(x) + u_Y(x).$$

This allows the creation of new distributions and their corresponding Lévy measures by the convolution of two random variables whose Lévy measures are already known. The convolution of two random variables from the same infinitely divisible distribution often gives a random variable with the same distribution (e.g. the convolution of two *Cauchy* random variables gives another *Cauchy* random variable so no "new" Lévy measures are generated); sometimes new distributions can be formed, though often it is difficult to find $f_Z(z)$ in closed form.

2.2.3 Generalised Inverse Gaussian distribution: $GIG(\gamma, \nu, \alpha)$

Many standard distributions on $\mathbb{R}^+$ are special cases of the *GIG* distribution, such as the *Gamma, Positive Hyperbolic, Inverse Gamma* and *Inverse Gaussian* distributions (which will be abbreviated to *Ga*, *RPH*, *IGa* and *IG* respectively). It has been shown in Halgren (1979) that the *Generalised Inverse Gaussian* distribution is self-decomposable and so is infinitely divisible and has a Lévy measure (see Theorem 2).

If $X \sim GIG(\gamma, \nu, \alpha)$, then

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

where $K_\gamma()$ is a modified Bessel function of the third kind.

The Lévy measure of X is

$$u(x) = \frac{1}{x}\left[\frac{1}{2}\int_0^\infty e^{-x\xi/2\nu^2} g_\gamma(\xi)\, d\xi + \max(0, \gamma)\right] e^{-\alpha^2 x/2}, \tag{2.1}$$

where

$$g_\gamma(x) = \frac{2}{x\pi^2}\left\{J^2_{|\gamma|}\left(\sqrt{x}\right) + N^2_{|\gamma|}\left(\sqrt{x}\right)\right\}^{-1} \tag{2.2}$$

and $J_{|\gamma|}$ and $N_{|\gamma|}$ are Bessel functions of the first and second kind respectively. This Lévy measure is derived in Barndorff-Nielsen and Shephard (2001b).

A special case of the $GIG(\nu, \delta, \gamma)$ distribution is the *Inverse Gamma* distribution and this Lévy process is simulated in Section 2.3.3, using the Lévy measure given in equation (2.1). The *GIG* distribution and its four special cases are used as marginal distributions for the BNS SV models in Chapters 3, 4 and 5. These models require the Lévy measure of the marginal distribution of the volatility and this is available from equation (2.1).

2.2.4 Tempered Stable distribution: $TS(\kappa, \nu, \alpha)$

Special cases of the *Tempered Stable* distribution include the *Inverse Gaussian, Lévy* and *Stable* distributions.

If $X \sim TS(\kappa, \nu, \alpha)$, then for $0 < \kappa < 1$ and $\nu, \alpha > 0$, the density is

$$f_X(x) = e^{\nu\alpha} f_{X|\kappa,\nu}(x) \exp\left\{-\frac{\alpha^{1/\kappa}}{2}x\right\}, \qquad \text{for} \qquad x > 0, \tag{2.3}$$

where

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

is the density function of the positive $\kappa - stable$ law (see Feller (1971) and Barndorff-Nielsen and Shephard (2001c)). Different parameterisations can be used for the *Stable* distribution and the relationship between the representation used in Feller (1971) and Barndorff-Nielsen and Shephard (2001c) is given in Appendix A.3. Further details on the *Tempered Stable* distribution can be found in Tweedie (1984).

The Lévy measure of X is

$$u(x) = \left\{\frac{\nu\kappa 2^{\kappa}}{\Gamma(1-\kappa)}\right\} x^{-1-\kappa} \exp\left\{-\frac{\alpha^{1/\kappa}}{2}x\right\}$$

(see Barndorff-Nielsen and Shephard (2001c)).

2.2.5 Generalised Asymmetric Laplace distribution: $GAL(\alpha, \beta, \mu)$

If $X \sim GAL(\alpha, \beta, \mu)$, for $\alpha, \beta \in \mathbb{R}^+$ and $\mu \in \mathbb{R}$, the density is

$$f_X(x) = \begin{cases} \dfrac{\alpha\beta}{\alpha+\beta} e^{\beta(x-\mu)}, & \text{for} \quad -\infty < x \leq \mu \\ \\ \dfrac{\alpha\beta}{\alpha+\beta} e^{-\alpha(x-\mu)}, & \text{for} \quad \mu < x \leq \infty \end{cases}$$

and the characteristic function is

$$\phi(t) = \frac{\alpha\beta\,(-1)^{t\mu/\pi}}{(\alpha - it)(\beta + it)}.$$

Using equation (1.2), $Q = 0$, so we have bounded variation and equation (1.3) can be used to give $a = \mu$. Consider the case $\mu = 0$; the density is made up of two parts, one for $x \leq 0$ and one for $x > 0$.

For $X > 0$, X has the form of a $Ga(1, \alpha)$ distribution (ignoring normalising constants). Then $E[X] = 1/\alpha$ and the characteristic function is

$$\phi(t) = \frac{\alpha}{\alpha - it}$$

and its Lévy measure is given by equation (2.1). That is

$$u(x) = \begin{cases} x^{-1}e^{-\alpha x} & x > 0 \\ \\ 0 & x < 0 \end{cases}.$$

For $X \leq 0$, $-X$ has the form of a $Ga(1, \beta)$ distribution (ignoring normalising constants). This has characteristic function

$$\phi(t) = \frac{\beta}{\beta + it}$$

and Lévy measure

$$u(x) = \begin{cases} 0 & x > 0 \\ \\ -x^{-1}e^{\beta x} & x < 0 \end{cases}.$$

From the definition of the Lévy measure given in equation (1.1), the Lévy measure of the $GAL(\alpha, \beta, 0)$ distribution is

$$u(x) = \begin{cases} x^{-1}e^{-\alpha x} & x > 0 \\ \\ -x^{-1}e^{\beta x} & x < 0 \end{cases}.$$

This Lévy measure has two parameters which control the decay in the positive and negative planes and is similar to that used in the CGMY model (see Carr et al. (2002)).

If $X \sim GAL(\alpha, \beta, 0)$, then $X + \mu \sim GAL(\alpha, \beta, \mu)$ with characteristic function

$$\phi_{X+\mu}(t) = \frac{\alpha\beta(-1)^{t\mu/\pi}}{(\alpha - it)(\beta + it)} = \phi_X(t)(-1)^{t\mu/\pi}.$$

As both X and $X + \mu$ are on $\mathbb{R}$, the Lévy measure is independent of μ and is the same as the Lévy measure of the $GAL(\alpha, \beta, 0)$ distribution (see Section 2.2.1).

2.2.6 Generalised Cauchy distribution

For $\mu \in \mathbb{R}$ and $\theta \in \mathbb{R}^+$, the density function of the *Generalised Cauchy* distribution is

$$f_X(x) = \frac{\theta}{\pi\left(\theta^2 + (x - \mu)^2\right)}, \qquad \text{for} \qquad -\infty < x < \infty$$

and the characteristic exponent is

$$\Psi(t) = -i\mu t + \theta|t|,$$

so the *Generalised Cauchy* process has bounded variation and $Q = 0$ (see equation (1.1)).

The Lévy measure, $u(x)$, will be an even function, as $f_X(x)$ is even when $\mu = 0$. When the exponential part of equation (1.1) is written as trigonometric functions, the integral of $(1 - \cos(tx))\, u(x)$ must be proportional to $|t|$. We therefore seek a solution of the form $u(x) = cx^{-2}$, for some constant, c, (this corresponds to constant canonical measure, $M(x)$, in Theorem 4). Then

$$\begin{aligned}\Psi(t) &= iat + c\int_{-\infty}^{\infty} \frac{1-\cos(tx)}{x^2}dx + ci\int_{-\infty}^{\infty} \frac{tx1_{|x|<1} - \sin(tx)}{x^2}dx \\ &= iat + c\int_{-\infty}^{\infty} \frac{1-\cos(tx)}{x^2}dx + ci\int_{-1}^{1} \frac{tx - \sin(tx)}{x^2}dx - ci\int_{|x|>1} \frac{\sin(tx)}{x^2}dx\end{aligned}$$

The second and third integrands are odd. The second integrand has a removable singularity at $x = 0$ and the third integrand does not have any singularities in the range of integration. Therefore both of these integrals are 0. This gives

$$\Psi(t) = iat + c\int_{-\infty}^{\infty} \frac{1-\cos(tx)}{x^2}dx = iat + c\pi|t|$$

(details in Appendix A.4).

Picking $a = -\mu$ and $c = \theta/\pi$ gives

$$\Psi(t) = -i\mu t + \theta|t|,$$

which is the characteristic exponent of the *Generalised Cauchy* distribution. Therefore the Lévy measure of the *Generalised Cauchy* distribution is

$$u(x) = \frac{\theta}{\pi x^2}.$$

Note that the location parameter, μ, does not appear in the Lévy measure (which was known from Section 2.2.1).

2.2.7 Hyperbolic Cosine (Cosh) distribution

For $a, \mu \in \mathbb{R}$, the density function of the *Hyperbolic Cosine* distribution is

$$f_W(w) = \frac{1}{\pi|a|\cosh\left(\frac{w-\mu}{a}\right)}, \quad \text{for} \quad -\infty < w < \infty.$$

The derivation of the Lévy measure is quite involved and so is included in Appendix A.5. The Lévy measure is

$$u(x) = \frac{sign(a)}{x\left(e^{x/a} - e^{-x/a}\right)}.$$

2.2.8 Hyperbolic Sine (Sinh) distribution

The density function of the *Hyperbolic Sine* distribution is

$$f_X(x) = \frac{2}{\pi^2}\frac{x}{\sinh(x)}, \quad \text{for} \quad -\infty < x < \infty.$$

It can be shown (see Feller (1971), pg 503) that this has characteristic function

$$\phi_1(t) = \int_{-\infty}^{\infty} \frac{2}{\pi^2}\frac{xe^{itx}}{\sinh(x)}dx = \left\{\cosh^2\left(\frac{\pi t}{2}\right)\right\}^{-1}.$$

This is the characteristic function for the random variable $Z = X + Y$, where X and Y are independent and identically distributed *Cosh* random variables (see Appendix A.6). Therefore the *Sinh* distribution is infinitely divisible and has Lévy measure

$$u(x) = 2sign(a)\frac{1}{x\left(e^{x/a} - e^{-x/a}\right)} = \frac{sign(a)}{x\sinh(x/a)}$$

(see Section 2.2.2).

Letting $W = Y - \mu$, for $\mu \in \mathbb{R}$, gives

$$f_W(w) = \frac{2sign(a)}{\pi^2 a^2}\frac{(w-\mu)}{\sinh\left(\frac{w-\mu}{a}\right)}, \quad \text{for} \quad -\infty < w < \infty$$

and it is known from Section 2.2.1 that the Lévy measure will be unaltered by the shift.

2.2.9 Normal distribution: $N(\mu, \sigma^2)$

For $\mu \in \mathbb{R}$ and $\sigma \in \mathbb{R}^+$, the density of the *Normal* distribution is

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad \text{for} \quad -\infty < x < \infty.$$

It easily follows that

$$E\left[e^{itX}\right] = e^{i\mu t - \sigma^2 t^2/2}$$

and the Lévy measure is $u(x) = 0$ (and is not defined at the origin). Note that the location parameter, μ, does not appear in the Lévy measure.

The canonical measure is, perhaps, more interesting than the Lévy measure. The following derivation is from Feller (1971). Using Theorem 4 we have

$$\Psi(t) = \int_{-\infty}^{\infty}\left\{\frac{1 + it\sin(x) - e^{itx}}{x^2}\right\}M(x)\,dx + ibt. \tag{2.4}$$

Let $g(x)$ be the integrand of equation (2.4) and try $M(x) = c\delta(x)$, then (see Appendix A.7)

$$\int_{-\infty}^{\infty} \left\{ \frac{1 + it\sin(x) - e^{itx}}{x^2} \right\} M(x)\, dx = \frac{ct^2}{2}.$$

This requires us to define

$$g(0) = \lim_{x \to 0} g(x).$$

Choosing $b = -\mu$ and $c = \sigma^2$ in equation (2.4) then gives

$$\Psi(t) = \frac{\sigma^2 t^2}{2} - i\mu t,$$

which is the characteristic function of the $N\left(\mu, \sigma^2\right)$ distribution and so the canonical measure is

$$M(x) = \sigma^2 \delta(x).$$

2.2.10 Student-t distribution

For $n \in \mathbb{R}^+$, the density of the *Student-t* distribution with n degrees of freedom is

$$f_X(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \frac{1}{\sqrt{n\pi}} \left(1 + \frac{x^2}{n}\right)^{-\left(\frac{n+1}{2}\right)}, \quad \text{for} \quad -\infty < x < \infty$$

and the characteristic function is

$$\phi(t) = \frac{2^{(3-n)/2}}{\Gamma(n-1)/2} |t|^{(n-1)/2} K_{(1-n)/2}(|t|). \tag{2.5}$$

This can be calculated by noting that if the two independent random variables V, Y are distributed

$$\begin{aligned} V &\sim IGa\left(\frac{n}{2}, \frac{n}{2}\right) \\ Y &\sim N(0, V), \end{aligned}$$

then Y is distributed *Student-t* with n degrees of freedom. This is illustrated in Figure 2.5. The characteristic function can then be found by taking iterated expectations.

The characteristic function is not available in closed form for general degrees of freedom, n, other than in the forms of a modified Bessel function of the third kind given in equation (2.5). The Lévy measure is only known for the case $n = 1$, when the *Cauchy* distribution is recovered (see Section 2.2.6). Although the Lévy measure could be computed numerically by inversion of the characteristic function, this approach is computationally intensive and rather unusable. We adopt a different approach to sample from a Lévy process which has $X_\infty \sim$*Student-t*(n).

As the *Student-t* distribution is a scale mixture of a *normal* random variate and the root of an $IGa\left(\frac{n}{2},\frac{n}{2}\right)$ random variate, we can sample from a Lévy process which has *Student-t* marginal distributions (in a similar way to the techniques used in Section 2.3), even though the Lévy measure is not available. To sample from the *Student-t* distribution process, either generate an *Inverse Gamma* Lévy process, take the root of the process and multiply by a $N(0,1)$ variate (this is referred to as Method1), or generate Brownian motion up to time $t=1$ and multiply this by the root of an *Inverse Gamma* variate (this is referred to as Method2). Algorithms to sample from Brownian motion and the $IGa\left(\frac{n}{2},\frac{n}{2}\right)$ Lévy process are discussed earlier in this chapter. The real *Student-t* Lévy process should have both positive and negative non-infinitesimal jumps. The construction using Method1 or Method2 cannot generate this, as Brownian motion only has infinitesimal jumps and the *Inverse Gamma* Lévy process only has positive jumps (though they are non-infinitesimal). The distribution of the end point of this mixture will have a *Student-t* distribution and this is similar to the Lévy processes described in Section 2.3, though clearly this is not the genuine *Student-t* Lévy process.

2.2.11 Convolution of Cosh variates

The general formula for the density of the convolution of n independent *Cosh* variates has been derived in Baten (1934) and is

$$f_Z(z)=\begin{cases}\dfrac{4^{n-1}}{(2n-1)!2\theta^2}\dfrac{z}{\sinh(\pi z/2\theta)}\displaystyle\prod_{j=1}^{n}\left(\frac{z^2}{4\theta^2}+j^2\right) & n \text{ even}\\ \dfrac{2^{2n-1}}{(2n)!\theta^2\cosh(\pi z/2\theta)}\displaystyle\prod_{j=1}^{n}\left(\frac{z^2}{4\theta^2}+\left(j-\frac{1}{2}\right)^2\right) & n \text{ odd}\end{cases}.$$

The corresponding Lévy measure is then simple to calculate given the known Lévy measure of the *Cosh* distribution and is

$$u(x)=\frac{sign(\theta)\,n}{2x\sinh(2\theta x/\pi)}$$

(see Section 2.2.2).

2.2.12 Tables of densities and Lévy measures

Densities on $\mathbb{R}^+$

Distribution	Density, $f_X(x)$	Lévy measure, $u(x)$
$GIG(\gamma,\nu,\alpha)$	$c\exp\left\{-\frac{1}{2}\left(\nu^2x^{-1}+\alpha^2x\right)\right\}$	$\frac{e^{-\alpha^2x/2}}{x}\left[\frac{1}{2}\int_0^\infty e^{-x\xi/2\nu^2}g_\gamma(\xi)\,d\xi+m(\gamma)\right]$
$TS(\kappa,\nu,\alpha)$	$\exp\left(\nu\alpha-\frac{\alpha^{1/\kappa}}{2}x\right)f_{Y\mid\kappa,\nu}(x)$	$\left(\frac{\nu\kappa2^\kappa}{\Gamma(1-\kappa)}\right)x^{-1-\kappa}\exp\left(-\frac{\alpha^{1/\kappa}}{2}x\right)$

Table 2.1: Table of Lévy measures of standard distributions on $\mathbb{R}^+$.

where

$$c=\frac{(\alpha/\nu)^\gamma}{2K_\gamma(\nu\alpha)}x^{\gamma-1},$$

$m(\gamma)=\max(0,\gamma)$ and $f_{Y\mid\kappa,\nu}(x)$ is the positive $\kappa-$*stable* density (see Feller (1971) and Barndorff-Nielsen and Shephard (2001c)).

Densities on $\mathbb{R}$

Distribution	Density, $f_X(x)$	Lévy measure, $u(x)$
$GAL(\alpha, \beta, \mu)$	$\frac{\alpha\beta}{\alpha+\beta} e^{\beta x} \quad -\infty < x \leq 0$ $\frac{\alpha\beta}{\alpha+\beta} e^{-\alpha x} \quad 0 < x \leq \infty$	$-x^{-1} e^{-\beta x} \quad x > 0$ $x^{-1} e^{-\alpha x} \quad x < 0$
Generalised Cauchy	$\frac{\theta}{\pi \left\{ \theta^2 + (x-\mu)^2 \right\}}$	$\frac{\theta}{\pi x^2}$
Hyperbolic Cosine	$\frac{1}{\pi \lvert a \rvert \cosh\left(\frac{x-\mu}{a}\right)}$	$\frac{sign(a)}{2x \sinh\left(\frac{x}{a}\right)}$
Hyperbolic Sine	$\frac{2 sign(a)}{\pi^2 a^2} \frac{(x-\mu)}{\sinh\left(\frac{x-\mu}{a}\right)}$	$\frac{sign(a)}{x \sinh\left(\frac{x}{a}\right)}$
n Hyperbolic Cosines (n even) (n odd)	$\frac{4^{n-1}}{(2n-1)! 2\theta^2} \frac{x}{\sinh\left(\frac{\pi x}{2\theta}\right)} \prod_{j=1}^{n} \left(\frac{x^2}{4\theta^2} + j^2 \right)$ $\frac{2^{2n-1}}{(2n)! \theta^2 \cosh\left(\frac{\pi x}{2\theta}\right)} \prod_{j=1}^{n} \left(\frac{x^2}{4\theta^2} + \left(j - \frac{1}{2}\right)^2 \right)$	$\frac{sign(\theta) n}{2x \sinh\left(\frac{2\theta x}{\pi}\right)}$
Normal, $N(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{ \frac{(x-\mu)^2}{2\sigma^2} \right\}$	0

Table 2.2: Table of Lévy measures of standard distributions on $\mathbb{R}$.

2.3 Sampling from Lévy processes: the Ferguson and Klass (1972) representation

In this section we study the simulation of Lévy processes to enhance our understanding of them. Later on we sample from stochastic integrals with respect to Lévy processes, using series representations similar to the ones introduced in this section.

Let $\theta \in \mathbb{R}$ and X_θ be a random variable with density $f_{X_\theta}(x_\theta)$ and Lévy measure $u_\theta(x)$. Using the same notation as before, the Lévy representation is (see equation (1.1))

$$\Psi_\theta(t) = ia_\theta t + \frac{Q_\theta}{2}t^2 + \int_{-\infty}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)u_\theta(x)\,dx.$$

Define

$$dN_{l,\theta}(x) = \begin{cases} u_\theta(x)\,dx & x \leq 0 \\ \\ 0dx & x \geq 0 \end{cases}$$

and

$$dN_{u,\theta}(x) = \begin{cases} 0dx & x \leq 0 \\ \\ u_\theta(x)\,dx & x \geq 0 \end{cases},$$

where $u_\theta(x)$ is the Lévy measure and $dN_{l,\theta}(x)$ and $dN_{u,\theta}(x)$ must satisfy certain constraints (see Walker and Damien (2000)), one of which is $dN_{l,\theta}(x), dN_{u,\theta}(x) \geq 0$.

The representation given in Ferguson and Klass (1972) is

$$\begin{aligned}\Psi_\theta(t) = ia_\theta t + \frac{Q_\theta}{2}t^2 \quad &+ \quad \int_{-\infty}^{0}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)dN_{l,\theta}(x) \\ &+ \quad \int_{0}^{\infty}\left(e^{itx} - 1 - \frac{itx}{1+x^2}\right)dN_{u,\theta}(x).\end{aligned} \tag{2.6}$$

Using the same notation as Walker and Damien (2000) let $M(x)$ be the decreasing function

$$M(x) = \int_x^\infty dN_{u,\infty}(z)$$

and let

$$\begin{aligned}\tau_1 &\sim Exp(1) \\ \\ \tau_i - \tau_{i-1} &\overset{iid}{\sim} Exp(1) \qquad i = 2, 3, \ldots.\end{aligned}$$

Define the jumps, J_i, as

$$\begin{aligned}J_i &= 0, \qquad \text{if} \quad \tau_i > M(0) \\ \\ \tau_i &= M(J_i), \qquad \text{if} \quad \tau_i \leq M(0).\end{aligned}$$

As the τ_i are increasing, the J_i are decreasing. This is a fundamental property in the simulation algorithm, as it facilitates truncation of the sum in equation (2.7).

In general, $J_i|\tau_i$ may need to be calculated numerically. In order to generate small samples from each different Lévy process (on modern computers), it generally does not matter how efficient the numerical method is and, for this reason, the numerical techniques

used are not given in this section. Detailed descriptions of the numerical methods used to calculate the jumps are given when the efficiency of generation is more important (such as in Section 4.3 when MCMC is used to make inference about the Lévy process).

Further, let

$$n_\theta(x) = \frac{dN_{u,\theta}(x)}{dN_{u,\infty}(x)},$$

which behaves like a cumulative density for $\theta|x$ and $\theta \in \mathbb{R}^+$, as it is increasing and tends to 1.

The Ferguson and Klass (1972) series representation of a Lévy process with positive jumps is

$$X_\theta \stackrel{\mathfrak{L}}{=} \sum_{i=1}^{\infty} J_i I\{u_i \leq n_\theta(J_i)\}, \tag{2.7}$$

where $u_i \overset{iid}{\sim} U(0,1)$ for $i = 1, 2, \ldots$. Note that the same uniforms are used for all the X_θ and that the series may need to be truncated to simulate from X_θ.

To generate from a Lévy process with positive and negative jumps, the process is split into two Lévy processes, one with positive jumps, one with negative and the processes are added together (as indicated in equation (2.6)). The *Cauchy* and *Cosh* processes are implemented in Sections 2.3.1 and 2.3.2 using this technique.

Note that

$$X_\infty \stackrel{\mathfrak{L}}{=} \sum_{i=1}^{\infty} J_i I\{u_i \leq n_\infty(J_i)\} = \sum_{i=1}^{\infty} J_i$$

and, from Ferguson and Klass (1972) and Walker and Damien (2000), we have

$$-\log\left(E\left[e^{-tX_\theta}\right]\right) = \int_0^\infty \left(1 - e^{-tz}\right) dN_{u,\theta}(z),$$

with a similar result for the negative jumps. As a result, X_θ has Lévy measure $u_\theta(x)$ from the Lévy-Khintchine formula (equation (1.1)) and therefore has density $f_{X_\theta}(x_\theta)$. This allows us to test if the algorithm to sample from a Lévy process has been implemented correctly, comparing the density of X_∞ and its theoretical distribution with histograms and QQplots.

In all of our examples, we choose $dN_\theta(x) = \alpha(\theta)\, u(x)\, dx$, so

$$n_\theta(x) = \frac{dN_\theta(x)}{dN_\infty(x)} = \frac{\alpha(\theta)}{\alpha(\infty)},$$

which is independent of x and then

$$X_\theta \stackrel{\mathfrak{L}}{=} \sum_{i=1}^{\infty} J_i I\left\{u_i \leq \frac{\alpha(\theta)}{\alpha(\infty)}\right\}.$$

The processes chosen for this were the *Cauchy*, *Cosh*, *Inverse Gamma* and *Student-t* distribution processes. Our main focus later on will be Lévy processes with only positive jumps, so the *Cauchy* and *Cosh* processes will not be used outside this chapter. The *Inverse Gamma* process was included as it will be of particular interest later, when it is used as a marginal distribution for volatility.

For our simulations, we choose $\alpha(\theta) = \theta/(1+\theta)$. The results for all four processes are given in Section 2.3.5. Algorithms to simulate from four Lévy processes are now discussed in detail. The algorithms' correct implementation is tested by comparing the distributions of X_∞ and its theoretical distribution.

2.3.1 Cauchy process

For $\theta_1 \in \mathbb{R}^+$ and $\mu \in \mathbb{R}$, the *Cauchy* distribution has density

$$f_X(x) = \frac{\theta_1}{\pi\left(\theta_1^2 + (x-\mu)^2\right)}$$

and the Lévy measure is

$$u(x) = \frac{\theta_1}{\pi x^2}.$$

Therefore

$$dN_\theta(x) = \alpha(\theta)\frac{\theta_1}{\pi x^2}dx$$

and

$$M(x) = \frac{\theta_1}{\pi x}.$$

The integration for $M(x)$ can be performed analytically for the *Cauchy* process.

As $M(0) = \infty$, we have an infinite Lévy measure and so the process will have an infinite number of (mostly small) jumps. Therefore $\tau_i \leq M(0)$ for all i and J_i is given by

$$J_i = \frac{\theta_1}{\pi\tau_i}.$$

This means that $J_i \neq 0 \ \forall \ \tau_i$ and the sum must be truncated at some point.

Generating sufficiently many uniforms gives

$$X_\infty \overset{\mathfrak{L}}{=} \sum_{i=1}^{\infty} J_i I\{u_i \leq n_\infty(J_i)\} = \frac{\theta_1}{\pi}\sum_{i=1}^{\infty}\frac{1}{\tau_i}.$$

For testing purposes, a standard *Cauchy* process ($\theta_1 = 1, \mu = 0$) was simulated (see Figure 2.2).

2.3.2 Cosh process

For $\theta_1 \in \mathbb{R}$, the convolution of $2n$ *Cosh* random variables has density

$$f_X(x) = \frac{4^{n-1}}{(2n-1)!2\theta_1^2} \frac{x}{\sinh(\pi x/2\theta_1)} \prod_{j=1}^{n-1} \left(\frac{x^2}{4\theta_1^2} + j^2 \right)$$

(see Baten (1934)). The Lévy measure is

$$u(x) = \frac{2n}{x \sinh(2\theta_1 x/\pi)}.$$

For $\theta_1 \in \mathbb{R}$, the convolution of $2n+1$ *Cosh* random variables has density

$$f_X(x) = \frac{2^{2n-1}}{(2n)!\theta_1} \frac{1}{\cosh\left(\frac{\pi x}{2\theta_1}\right)} \prod_{j=1}^{n} \left\{ \frac{x^2}{4\theta_1^2} + \left(j - \frac{1}{2}\right)^2 \right\}$$

(see Baten (1934)). The Lévy measure is

$$u(x) = \frac{(2n+1)}{2x \sinh(2\theta_1 x/\pi)},$$

so

$$dN_\theta(x) = \alpha(\theta) \frac{k}{2x \sinh(2\theta_1 x/\pi)} dx,$$

where k can be chosen to be any integer and is specified by the number of *Cosh* random variables that are convoluted. In this example, we pick $k = 1$. Then

$$\begin{aligned} M(x) &= \frac{k}{2} \int_x^\infty \frac{1}{z \sinh(2\theta_1 z/\pi)} dz \\ &= \frac{k}{2} \int_{2\theta_1 x/\pi}^\infty \frac{1}{w \sinh(w)} dw \\ &= \frac{k}{2} \int_0^{\pi/2\theta_1 x} \frac{1}{y \sinh(1/y)} dy \end{aligned}$$

and again, as $M(0) = \infty$, we have an infinite Lévy measure. Therefore $\tau_i \leq M(0)$ and J_i satisfies

$$\tau_i = \frac{k}{2} \int_0^{\pi/2\theta_1 J_i} \frac{1}{y \sinh(1/y)} dy$$

and must be evaluated numerically.

Once the J_i can be calculated, generating sufficiently many uniforms gives

$$X_t \stackrel{\mathfrak{L}}{=} \sum_{i=1}^\infty J_i I\{u_i \leq \alpha(t)\}.$$

For testing purposes, a standard *Cosh* process ($\theta_1 = 1, \mu = 0$) was simulated (see Figure 2.3). For the QQplot for the *Cosh* process, it is easy to generate from the standard *Cosh* distribution using cdf inversion. Details for this are given in Appendix C.1.

2.3.3 Inverse Gamma process

The *Inverse Gamma* process is of particular interest as we are interested in using this for the marginal distribution for our stochastic volatility and comparing its performance against other marginal distributions. When an *Inverse Gamma* distribution is used as a marginal distribution for stochastic volatility in equation (1.8), the returns will be heavy tailed, approximately *Student-t* distributed.

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

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

and from equation (2.1), the Lévy measure is

$$u\left(x\right) = \frac{1}{\pi^2 x} \int_0^\infty \frac{e^{-xy/4\alpha}}{y\left\{J_\nu^2\left(\sqrt{y}\right) + N_\nu^2\left(\sqrt{y}\right)\right\}} dy.$$

Using the same form for $dN_\theta\left(z\right)$ as before,

$$M\left(x\right) = \frac{1}{\pi^2} \int_x^\infty \frac{1}{z} \left[\int_0^\infty \frac{e^{-zy/4\alpha}}{y\left\{J_\nu^2\left(\sqrt{y}\right) + N_\nu^2\left(\sqrt{y}\right)\right\}} dy\right] dz \tag{2.8}$$

and, as $M\left(0\right) = \infty$, we have an infinite Lévy measure, so $J_i \neq 0$ for all finite τ_i. Rearranging $\tau_i = M\left(J_i\right)$ gives

$$\tau_i \pi^2 = \int_0^\infty \frac{\mathrm{Ei}\left(1, J_i y/4\alpha\right)}{y\left\{J_\nu^2\left(\sqrt{y}\right) + N_\nu^2\left(\sqrt{y}\right)\right\}} dy, \tag{2.9}$$

where the exponential integral, $\mathrm{Ei}\left(1, x\right)$, is defined by

$$\mathrm{Ei}\left(1, x\right) = \int_1^\infty e^{-xt} t^{-1} dt.$$

Equation (2.9) is preferable to (2.8) as it is a single integral involving standard functions and can be computed efficiently. Equation (2.8) can be solved using a look up table and binary search, where the integral is evaluated using Gaussian Quadrature (see Appendix B.1). For testing purposes, an $IGa\left(1.3, 1.445\right)$ process was simulated (see Figure 2.4).

2.3.4 Student-t distribution process

The *Student-t* distribution process is the fourth simulation example. This is not strictly the *Student-t* Lévy process, as it is constructed as the product of a Lévy process (the

Normal Lévy process or the *Inverse Gamma* Lévy process) and a random variable (an *Inverse Gamma* or *Normal* random variable respectively), as described in Section 2.2.10. The purpose of simulating from this process is to show the difference in jumps between the two *Student-t* distribution processes and to show that when the root of an *Inverse Gamma* variable is multiplied by a *Normal* random variable, the product is a *Student-t* distribution. This is of interest because, when an *Inverse Gamma* marginal distribution is used for the stochastic volatility process in Chapter 4, the log returns will then be approximately *Student-t* distributed.

2.3.5 Graphs of four Lévy processes

Three graphs for the *Cauchy*, *Cosh* and *Inverse Gamma* Lévy process are given. The first graphs are individual simulations of each process showing how the processes jump. The second graphs are histograms of X_∞ and the third graphs are QQplots of X_∞ against the corresponding distribution that X_∞ should be theoretically, to verify the process has been generated correctly. For the *Student-t* distribution process, results for Method1 are in row 1 and results for Method2 are in row 2 of Figure 2.5. Two graphs are given for each simulation method; the first column shows how the processes jump and the second column is a QQplot of end points of the process and a *Student-t(3)* distribution. As we chose to sample from the *Student-t(3)* process, the parameters for the *Inverse Gamma* Lévy process and *Inverse Gamma* random variable are 1.5.

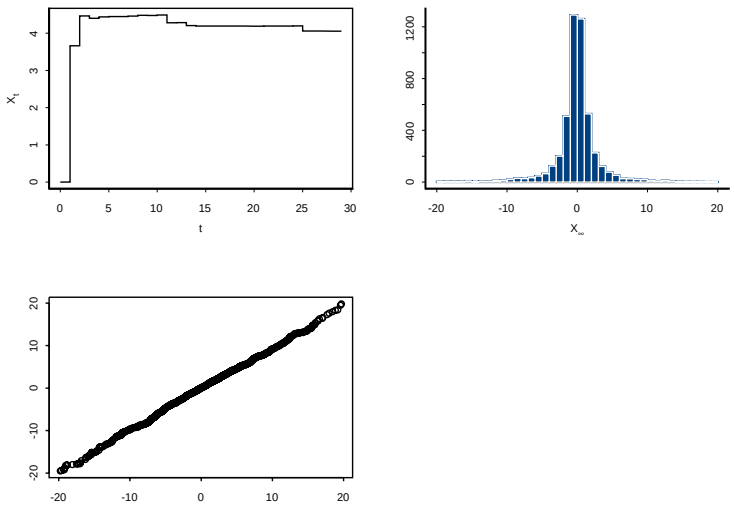


Figure 2.2: Path of a standard *Cauchy* process, histogram of the end points and QQplot of the end points against a *Cauchy* random variable.

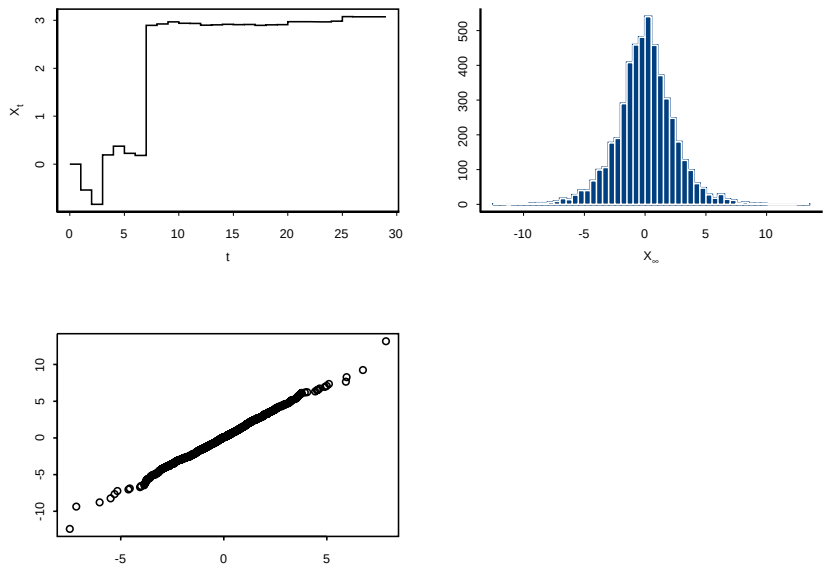


Figure 2.3: Path of a standard *Cosh* process, histogram of the end points and QQplot of the end points against a *Cosh* random variable.

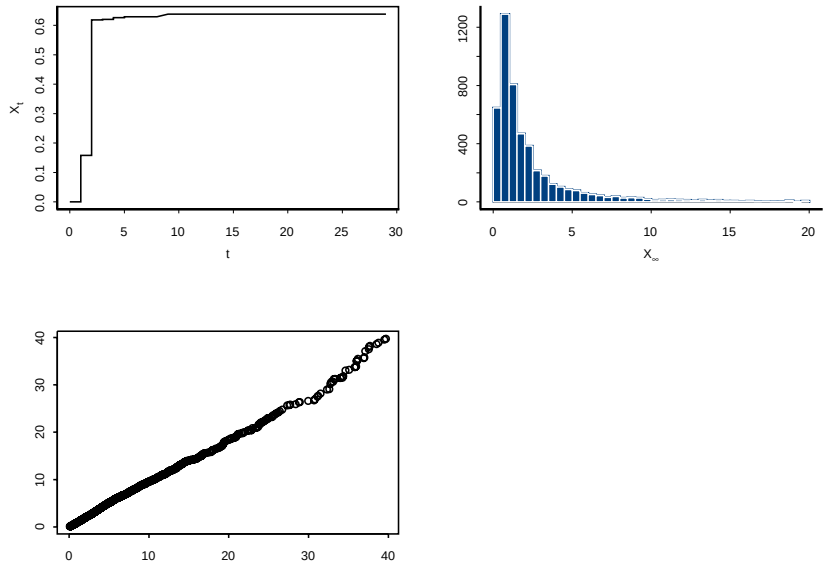


Figure 2.4: Path of an $IGa(1.3, 1.445)$ process, histogram of the end points and QQplot of the end points against an $IGa(1.3, 1.445)$ random variable.

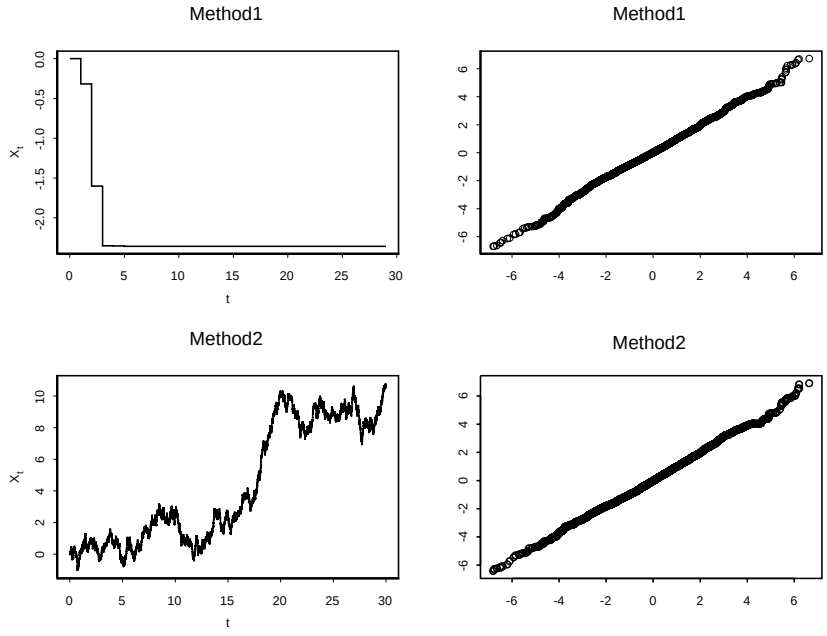


Figure 2.5: Paths of the two *Student-t* distribution processes and QQplots of the end points against a *Student-t(3)* random variable.

The first graph for each process shows how the different processes jump: jumps in X_t tend to get smaller as t increases (apart from Method2 of the *Student-t* distribution process). The histograms and QQplots suggest the algorithms have been implemented correctly. For the *Student-t* distribution process, results are as expected: Method1 has non-infinitesimal jumps in one direction and Method2 has infinitesimal jumps up and down. For both methods, the end point of the process is distributed *Student-t(3)*.

Chapter 3

Applications in Finance

The standard Black-Scholes equation (with constant volatility) was introduced in Section 1.3, as most option pricing is based on some variation of this model. Examples were given to illustrate the standard Black-Scholes model's inadequacy in explaining certain properties of financial data, to help motivate generalisations of the model, which were briefly introduced in Section 1.3.2. One further "stylised feature" of financial data is now illustrated before further details on the generalisations of the Black-Scholes model are given.

The main original contributions from this chapter are:-

(1) We recall the BNS SV model and simulate from this when the volatility has an *Inverse Gamma* marginal distribution.

(2) We allow the correlation parameter of the BNS SV models to vary over time, so that the volatility process has a more flexible correlation structure.

(3) We consider a multivariate normal approximation to fractional Brownian motion and use this to drive the asset equation.

Continue assuming that we have T equally spaced observed log returns, $y_1, \ldots, y_T$, each separated by Δ days. Typically, observations are daily, so $\Delta = 1$.

3.1 Long-memory in financial data

Some of the shortcomings of the standard Black-Scholes equation are well accepted (e.g. kurtosis of the log returns is too small for observed data). However, when it comes to long-

memory in financial data, results in the literature are mixed. It is generally accepted that long-memory is not present in the returns series, though there are still sources suggesting otherwise (see for example Barkoulas et al. (2000), where evidence of long-memory in the returns of the Greek Stock market is found). The long-memory tests are applied to the log-returns series to facilitate comparison with the MCMC results of Secion 5.3.4. Results for the square of the log returns series are more consistent but there are still discrepancies. For example, Ding et al. (1993) and Bollerslev and Mikkelsen (1996) find evidence of long-memory in the square of the log returns, while Krämer et al. (2002) find less conclusive evidence of long-memory in the square of the returns (for different financial time series). Some standard long-memory tests are implemented on the data sets which we investigate later on because of the mixed results in the literature. The results of these tests motivate a further generalisation of the model to allow for long-memory. Tests for kurtosis and volatility clustering are not included as the results for such tests on financial data are widely accepted.

Whilst there are many methods to test for long-memory, there is not an accepted standard test and the results of different tests are not always consistent with each other (see for example Tolvi (2003)). The two tests used here are the R/S statistic, which was introduced by Hurst (1951) and then modified by Lo (1991), and the GPH test, introduced by Geweke and Porter-Hudak (1983). The Hurst parameter (which controls the strength of the long-memory) will be estimated using three different methods.

There are various other long-memory tests but the focus here is to test whether a long-memory model is needed to effectively model financial data, rather than to assess the merits of different tests. All the algorithms used are available in the *S+FinMetrics* add on for *Splus*.

Although most previous long-memory testing has been on individual market indices (see for example Breidt et al. (1998)), the tests are applied to the Foreign Exchange (FX) rates and shares to which the models are later calibrated. Before their application, the tests and estimating methods are briefly introduced .

Recalling the definition of long-memory given in Definition 6, we would like to test if $0.5 < H < 1$ in the equation

$$\lim_{k\to\infty} \frac{\rho(k)}{ck^{2H-1}} = 1,$$

where $c \in \mathbb{R}^+$ and $\rho(k)$ is the correlation between the stationary process of interest at lag k.

Long-memory is only concerned with the asymptotic decay of the correlation, so correlation in the process could decay very rapidly initially yet still have long-memory. Estimating the correlation at large lags is difficult and this is why it is hard to test for

long-memory and why there are many long-memory tests. For the Ornstein-Uhlenbeck stochastic volatility model, the correlation of the log returns decays exponentially. Although this is not a long-memory process, for small λ in equation (1.8), the long-memory tests often conclude that there is significant evidence for long-memory unless the data set is sufficiently large.

3.1.1 Tests for long-memory

The algorithms of the two most popular tests for long-memory are described below. For this section, denote the sample mean (of observed log returns data y_i, $i = 1, \ldots, T$) by $\overline{y}$ and standard deviation by s_T.

R/S statistic

The R/S (rescaled range) statistic calculates the maximum distance between $\sum_{j=1}^{k} (y_i - \overline{y})$ and rescales this by the sample standard deviation. The R/S statistic is

$$Q_T = \frac{1}{s_T} \left\{ \max_{1 \le k \le T} \sum_{j=1}^{k} (y_i - \overline{y}) - \min_{1 \le k \le T} \sum_{j=1}^{k} (y_i - \overline{y}) \right\}.$$

If $y_i \overset{iid}{\sim} N\left(\mu, \sigma^2\right)$, then $Q_T/\sqrt{T}$ converges weakly to V, where V is the range of a Brownian bridge (see Definition 13) on the unit interval.

Lo (1991) observed that this statistic can incorrectly conclude that series have long-memory, when in fact they do not. He suggested a new statistic, $\widetilde{Q_T}$, using a modified standard deviation

$$\widehat{\sigma}_T(q) = \widehat{\gamma}_0 + 2 \sum_{j=1}^{q} \left(1 - \frac{j}{1+q}\right) \widehat{\gamma}_j,$$

where $q < T$ and

$$\widehat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-|k|} \left(X_t - \overline{X}_n\right) \left(X_{t+|k|} - \overline{X}_n\right).$$

$\widehat{\sigma}_T^2(q)$ is the Newey-West estimate of the long run variance for bandwidth q (see Newey and West (1994)). For our tests, $q = \left[4\left(T/100\right)^{1/4}\right]$, which is the default bandwidth choice in *Splus*. For $q = 0$ the original R/S statistic is recovered.

GPH test

Define the fractionally integrated process, y_t, as

$$u_t = \left\{ \sum_{k=0}^{\infty} \left(\begin{array}{c} k \\ d \end{array} \right) (-1)^k L^k \right\} (y_t - \overline{y}), \tag{3.1}$$

where L is the lag operator defined as

$$\begin{aligned} L^0 y_t &= 1 \\ L^k y_t &= y_{t-k} \qquad k = t, \ldots, 1 \text{ and } t = 1, \ldots, T \end{aligned}$$

and u_t is a stationary short-memory process with $E[u_t] = 0$.

Definition 17 *The **spectral density**, $f(\omega)$, of a stationary process, y_t, is*

$$f(\omega) = \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} \rho(k) e^{ik\omega},$$

where $\rho(k)$ is the correlation at lag k.

The spectral density of y_t, at Fourier frequency ω, is

$$f(\omega) = \left\{ 4 \sin^2 \left(\frac{\omega}{2} \right) \right\}^{-d} f_u(\omega),$$

where $f_u(\omega)$ is the spectral density of u_t and d is the fractional difference parameter (so $d = H - \frac{1}{2}$). Then

$$\log\{f(\omega_j)\} = -\log\{f_u(\omega_j)\} - d \log\left\{ 4 \sin^2 \left(\frac{\omega_j}{2} \right) \right\}, \quad \text{for} \quad j = 1, 2, \ldots, n_f(T) = \frac{T}{2} + 1$$

and for small ω_j, $f_u(\omega_j)$ is approximately constant, so

$$\log\{f(\omega_j)\} = -\beta - d \log\left\{ 4 \sin^2 \left(\frac{\omega_j}{2} \right) \right\} + e_j, \quad \text{for} \quad j = 1, 2, \ldots, n_f(T).$$

Definition 18 *For an equally spaced time series $X_1, \ldots, X_n$, the periodogram is a plot at frequency $\omega_j = 2\pi j/n$ and is given by*

$$I(\omega_j) = \frac{1}{2\pi n} \left| \sum_{t=1}^{n} \left(X_t - \overline{X}_n \right) e^{it\omega_j} \right|^2 = \frac{1}{2\pi} \sum_{k=-(n-1)}^{n-1} \widehat{\gamma}_k e^{ik\omega_j} \qquad j = 1, \ldots, \left[\frac{n-1}{2} \right].$$

Using a periodogram estimate of $f(\omega_j)$ and for large T, if $n_f(T) = T^\alpha$ (for $0 < \alpha < 1$), the least squares estimate, $\widehat{d}$, is distributed

$$\widehat{d} \sim N\left(d, \frac{\pi^2}{6\sum_{j=1}^{n_f}\left(U_j - \overline{U}\right)^2}\right),$$

where $U_j = \log\left[4\sin^2(\omega_j/2)\right]$.

This is then used in a hypothesis test, where the null hypothesis is y_t does not have long-memory (so $d = 0$). If the null is rejected, the test concludes y_t has long-memory. This test was developed by Geweke and Porter-Hudak (1983).

3.1.2 Estimating the Hurst parameter

Three methods are described to estimate the Hurst parameter, H, so that the strength of the long-memory in different financial data series can be compared.

R/S method

Mandelbrot (1975) showed that for a short-memory process, where Hurst parameter $H = 0.5$, or for a long-memory process, with Hurst parameter $0.5 < H < 1$, the R/S statistic converges in distribution to a random variable at rate T^H. For large T, the log-log plot of the R/S statistic against sample size should approximately be a straight line with gradient H.

Periodogram method

It can be shown that $f(\omega) \sim c_f |\omega_j|^{1-2H}$ as $|\omega_j| \to 0$. As the spectral density can be estimated by a periodogram, the log-log plot of the periodogram against ω will approximately be a straight line with gradient equal to $1 - 2H$ for small ω_j, typically taken to be the Fourier frequencies $\omega_j = 2\pi j/n$.

Whittle's method

The parameters in equation (3.1) (one of which is d) can be estimated by minimising

$$Q(\theta) = \int_{-\pi}^{\pi} \frac{I(\omega)}{f(\theta, \omega)} d\omega,$$

where θ is a vector of the unknown parameters, $I(\omega)$ is the periodogram of y_t and $f(\theta, \omega)$ is the spectral density (see Beran (1994)). The estimate for H is then $\widehat{H} = \widehat{d} + 1/2$, where $\widehat{d}$ is the estimate for d.

3.1.3 Results of long-memory tests on observed financial data

Foreign Exchange rate data

The tests will be applied to the Foreign Exchange rate data for the US Dollar against various currencies. The data are daily log returns from 6^{th} March 1993 through to 1^{st} December 2003 (excluding when markets were closed). This gives 3653 observations.

US Dollar against	y_i				
	Statistics		Estimates		
	R/S test	GPH test	$H_{R/S}$	H_{pg}	H_W
Australian Dollar	1.8927*	-0.1217	0.60	0.49	0.51
Brazilian Real	1.1167	0.2986	0.52	0.65	0.50
British Pound	1.2149	-1.07	0.55	0.40	0.52
Canadian Dollar	1.3511	-0.1524	0.58	0.53	0.50
Colombian Peso	1.2532	1.3724	0.59	0.60	0.52
Danish Krone	1.8614	1.3143	0.61	0.55	0.51
Euro	1.608	0.5654	0.56	0.52	0.49
Iceland Krona	1.9321*	2.3471*	0.58	0.59	0.54
Japanese Yen	1.6	0.8551	0.60	0.51	0.56
Moroccan Dirham	1.7258	1.4983	0.58	0.48	0.46
Singapore Dollar	1.3602	0.6452	0.61	0.57	0.56
Thai Baht	2.1578**	2.3457*	0.63	0.70	0.53

Table 3.1: Long-memory tests on the log returns of FX data.

US Dollar against	$\|y_i\|$				
	Statistics		Estimates		
	R/S test	GPH test	$H_{R/S}$	H_{pg}	H_W
Australian Dollar	5.1768**	4.9634**	0.66	0.74	0.62
Brazilian Real	4.6917**	5.554**	0.78	1.05	0.71
British Pound	2.7259**	2.6517**	0.69	0.66	0.58
Canadian Dollar	4.7594**	6.7123**	0.68	0.90	0.63
Colombian Peso	2.4478**	2.3767*	0.70	0.79	0.67
Danish Krone	3.2208**	4.6043**	0.68	0.80	0.56
Euro	2.4177**	3.5981**	0.61	0.65	0.55
Iceland Krona	5.2893**	3.9752**	0.59	0.67	0.64
Japanese Yen	3.5464**	4.839**	0.75	0.82	0.63
Moroccan Dirham	3.4952**	4.05**	0.71	0.80	0.62
Singapore Dollar	4.7487**	5.8275**	0.72	0.78	0.71
Thai Baht	5.1327**	6.9568**	0.78	1.02	0.72

Table 3.2: Long-memory tests on the absolute value of the log returns of FX data.

US Dollar against	y_i^2				
	Statistics		Estimates		
	R/S test	GPH test	$H_{R/S}$	H_{pg}	H_W
Australian Dollar	3.9526**	3.8624**	0.67	0.69	0.60
Brazilian Real	2.4255**	2.6609**	0.74	0.94	0.65
British Pound	2.2811**	1.6355	0.66	0.63	0.58
Canadian Dollar	3.8436**	5.9681**	0.68	0.88	0.64
Colombian Peso	1.728	1.4461	0.65	0.70	0.62
Danish Krone	3.2325**	3.9258**	0.67	0.75	0.57
Euro	2.4971**	4.0746**	0.62	0.62	0.56
Iceland Krona	3.9911**	4.6667**	0.60	0.58	0.56
Japanese Yen	2.4926**	4.0993**	0.68	0.71	0.67
Moroccan Dirham	2.287**	2.2124*	0.66	0.68	0.59
Singapore Dollar	3.8127**	4.6522**	0.65	0.71	0.65
Thai Baht	3.0116**	2.3882*	0.76	0.82	0.57

Table 3.3: Long-memory tests on the square of the log returns of FX data.

* : significant at 5% level

** : significant at 1% level

The results suggest that, for FX data, y rarely has significant evidence of long-memory, while $|y|$ and y^2 often have significant evidence of long-memory. For the data sets examined, the long-memory is stronger for $|y|$ than y^2. Most of the time, the two tests agree on whether a particular series has long or short-memory.

Share data

The tests will now be applied to open prices of shares on the New York Stock Exchange (NYSE). The data are daily log returns from 6^{th} March 1989 through to 1^{st} December 2003 (excluding when the market is closed), giving 3733 log returns. For data sets where the company had not been a quoted corporation (equivalent to public limited company in the UK) from 6^{th} March 1989, the data were drawn from the time that it became a publicly quoted company until 1^{st} December 2003.

Company	y_i				
	Statistics		Estimates		
	R/S test	GPH test	$H_{R/S}$	H_{pg}	H_W
British Airways PLC	1.5032	-1.4812	0.51	0.49	0.50
Citigroup Inc	0.8589	-1.5244	0.51	0.42	0.45
Coca-Cola Co	0.6148**	-1.7592	0.47	0.47	0.49
General Motors Corp	1.1258	-1.1585	0.56	0.51	0.52
HJ Heinz Co	0.9276	-1.3009	0.52	0.48	0.42
Host Marriott Corp	1.5211	0.5507	0.58	0.56	0.52
JP Morgan Chase & Co	1.5296	0.4815	0.53	0.51	0.48
Kellogg Co	1.1359	0.5839	0.57	0.50	0.45
McDonald's Corp	1.1647	-0.0398	0.51	0.49	0.47
Microsoft	0.8726	-1.7711	0.49	0.47	0.46
Procter & Gamble Co	0.7686*	-1.0472	0.53	0.42	0.48
S&P 500 INDEX	1.5479	-0.2718	0.54	0.51	0.48
Textron Inc	1.4322	-0.6426	0.54	0.45	0.48
Time Warner Inc	1.1507	-1.82	0.45	0.36	0.46

Table 3.4: Long-memory tests on the log returns of share data.

Company	$\|y_i\|$				
	Statistics		Estimates		
	R/S test	GPH test	$H_{R/S}$	H_{pg}	H_W
British Airways PLC	4.9843**	5.7948**	0.75	0.90	0.69
Citigroup Inc	3.1348**	3.3731**	0.72	0.78	0.62
Coca-Cola Co	2.2878**	2.9215**	0.64	0.64	0.55
General Motors Corp	3.0993**	3.3416**	0.61	0.68	0.58
HJ Heinz Co	2.0911*	3.4386**	0.64	0.65	0.59
Host Marriott Corp	1.9605*	3.3599**	0.80	0.75	0.69
JP Morgan Chase & Co	3.7773**	3.9799**	0.71	0.81	0.64
Kellogg Co	3.412**	4.0863**	0.65	0.73	0.60
McDonald's Corp	2.912**	3.0813**	0.60	0.62	0.57
Microsoft	2.0196*	1.7344	0.60	0.68	0.56
Procter & Gamble Co	2.8967**	3.799**	0.67	0.65	0.57
S&P 500 INDEX	5.8177**	4.4547**	0.75	0.90	0.67
Textron	3.8589**	2.3312*	0.70	0.71	0.61
Time Warner Inc	2.3416**	3.0394**	0.61	0.69	0.60

Table 3.5: Long-memory tests on the absolute value of the log returns of share data.

Company	y_i^2				
	Statistics		Estimates		
	R/S test	GPH test	$H_{R/S}$	H_{pg}	H_W
British Airways PLC	2.5437**	2.7662**	0.69	0.68	0.58
Citigroup Inc	1.5707	0.6638	0.61	0.62	0.52
Coca-Cola Co	0.9151	-0.2979	0.59	0.49	0.48
HJ Heinz Co	0.9418	0.2074	0.58	0.50	0.50
Host Marriott Corp	1.3874	1.271	0.73	0.63	0.59
General Motors Corp	1.0038	2.025*	0.57	0.49	0.50
JP Morgan Chase & Co	1.687	0.5348	0.64	0.50	0.51
Kellogg Co	0.7761*	-0.3505	0.57	0.49	0.50
McDonald's Corp	0.7226*	0.1964	0.60	0.48	0.50
Microsoft	0.7815*	-1.2662	0.50	0.45	0.50
Procter & Gamble Co	0.7964*	-0.9585	0.59	0.45	0.50
S&P 500 INDEX	4.5929**	3.3338**	0.70	0.83	0.67
Textron	1.1368	0.0821	0.67	0.50	0.50
Time Warner Inc	1.7552	0.2615	0.54	0.48	0.51

Table 3.6: Long-memory tests on the square of the log returns of share data.

* : significant at 5% level

** : significant at 1% level

The results suggest that, for share data, y and y^2 rarely have significant evidence of long-memory, while $|y|$ almost always have significant evidence of long-memory. Most of the time, the two tests agree on whether a particular series has long or short-memory.

Although results are mixed, in many of the data sets there is evidence that the data has long-memory in either the absolute value of the log returns and/or the square of the log returns. This is a "stylised feature" that financial data are often said to have, though results in the literature are mixed as well. We are unable to conclude that the data does not require a model with long-memory and so a generalisation of the model is considered which has long-memory. Empirical testing of the models in Chapter 5 will try to assess which data sets require the generalised long-memory model.

3.2 Stochastic volatility: The Ornstein-Uhlenbeck model

Recall equation (1.7),

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

where $W(t)$ is Brownian motion and $\sigma(t)$ is the stochastic volatility, assumed to be independent of $W(t)$. Again, let σ_i be the discretely observed volatility at time $i\Delta$ (defined in equation (1.6)) and $\underline{y} = \{y_1, \ldots, y_T\}$ be the observed log returns and

$$a_i = \frac{y_i - \left(\mu\Delta - \frac{\sigma_i^2}{2}\right)}{\sigma_i},$$

so $A_i \overset{iid}{\sim} N(0,1)$ and, if inference about μ and σ_i is required, the likelihood for $\underline{y}$ is

$$f_{\underline{Y}}(\underline{y}) = \prod_{i=1}^{T} \frac{1}{\sigma_i} f_{A_i}(a_i).$$

For the volatility process, assume that $\sigma^2(t)$ is a stationary non-Gaussian Ornstein-Uhlenbeck (OU) process (as given in equation (1.8)), with

$$d\sigma^2(t) = -\lambda\sigma^2(t)\, dt + dz(\lambda t),$$

where $z(t)$, called the Background driving Lévy process (BDLP), is a non-Gaussian Lévy process with $z(0) = 0$ and λ is a positive constant. This is a popular model suggested in Barndorff-Nielsen and Shephard (2001b).

Recalling what a typical Lévy process, $z(t)$, might look like (see Figure 1.2 and Section 2.3.5) we observe that $dz(t)$ is zero everywhere apart from where the BDLP "jumps". If $z(t)$ does not jump in the time interval $(t, t+\delta t)$, then

$$\sigma^2(t+\delta t) = e^{-\lambda \delta t}\sigma^2(t)$$

and $\sigma^2(t)$ decreases exponentially at a rate determined by λ.

Positive jumps in $z(t)$ cause positive jumps in $\sigma^2(t)$ (times of these jumps can be viewed as times at which new information arrives to the markets, causing the volatility to increase). These jumps occur more frequently as λ increases. Therefore λ determines both the rate at which jumps in $\sigma^2(t)$ occur and the rate of decay in-between these jumps. The form of equation (1.8) is such that the marginal distribution of σ^2 is unaltered by the parameter λ (see Barndorff-Nielsen and Shephard (2001b) and Appendix D).

The aggregate returns over the time $(i-1)\Delta < t < i\Delta$ is

$$y_i = \int_{(i-1)\Delta}^{i\Delta} dx(t),$$

where $x(t) = \log[S(t)]$ is the log asset. Barndorff-Nielsen and Shephard (2001b) have proved if the Ornstein-Uhlenbeck process of equation (1.8) is used as an SV model for the Black-Scholes equation (1.7), then the square of the log returns have correlation structure

$$corr\left(y_i^2, y_{i+s}^2\right) = ce^{-\lambda\Delta(s-1)} \tag{3.2}$$

and the discretely observed volatility (see equation (1.6)) has correlation structure

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

where

$$1 \geq d = \frac{\left\{1-e^{-\lambda\Delta}\right\}^2}{2\left\{e^{-\lambda\Delta}-1+\lambda\Delta\right\}} \geq c = \frac{\left\{1-e^{-\lambda\Delta}\right\}^2}{6\left\{e^{-\lambda\Delta}-1+\lambda\Delta\right\}+2(\lambda\Delta)^2(\zeta/\omega)^2} \geq 0,$$

for $\zeta = E\left[\sigma^2(t)\right]$ and $\omega^2 = Var\left[\sigma^2(t)\right]$. Therefore λ determines the correlation of both the volatility and the aggregate returns and both processes will have short-memory.

Figures 3.1 and 3.2 demonstrate this correlation structure by looking at the autocorrelation function (ACF) of simulations from the BNS SV model. Figure 3.1 shows graphs of volatility processes with $IGa(5,2)$ marginals and Figure 3.2 shows ACF plots of two volatility processes with different λ values. This volatility process should have a mean of approximately 0.5 and a variance of $1/12 = 0.083$.

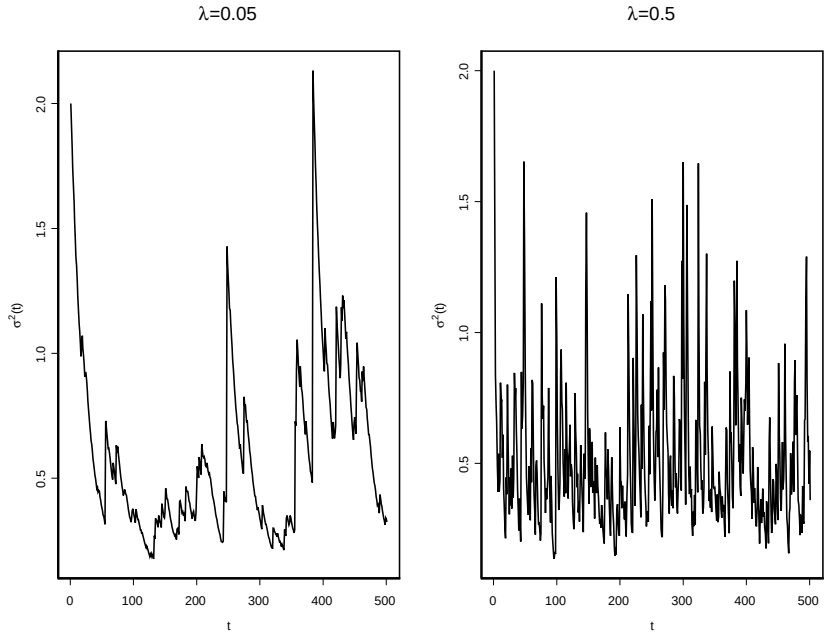


Figure 3.1: Plots of $\sigma^2(t)$ for an $IGa(5,2)$ marginal distribution for $\lambda = 0.05$ and $\lambda = 0.5$.

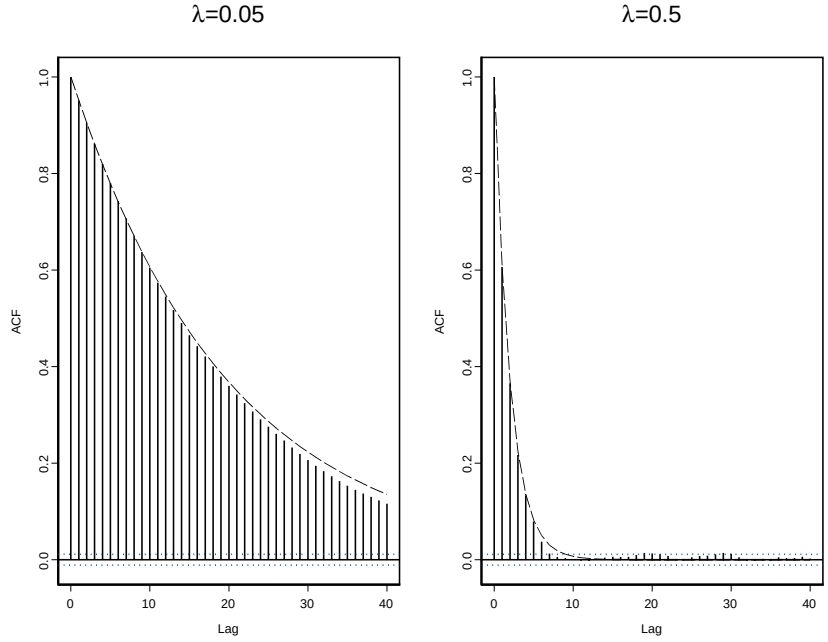


Figure 3.2: ACF plots of $\sigma^2(t)$ for an $IGa(5,2)$ marginal distribution for $\lambda = 0.05$ and $\lambda = 0.5$ and their theoretical correlation function.

	$\lambda = 0.05$	$\lambda = 0.5$
$E\left[\sigma^2(t)\right]$	0.499	0.495
$V\left[\sigma^2(t)\right]$	0.092	0.080

Table 3.7: Mean and variance of volatility processes with $IGa(5, 2)$ marginals and different correlation parameters.

The means and variances of each process are as expected. In practice, for a finite sample, the mean and variance tend to be closer to the theoretical values when λ is large, as the $\sigma^2(t)$ is less correlated.

Figure 3.3 is a QQplot of the two volatility processes given in Figure 3.1 (but for $T = 30,000$) and illustrates that λ does not alter the marginal distribution of $\sigma^2(t)$ (this is a result of Theorem 1. Further details can be found in Appendix D).

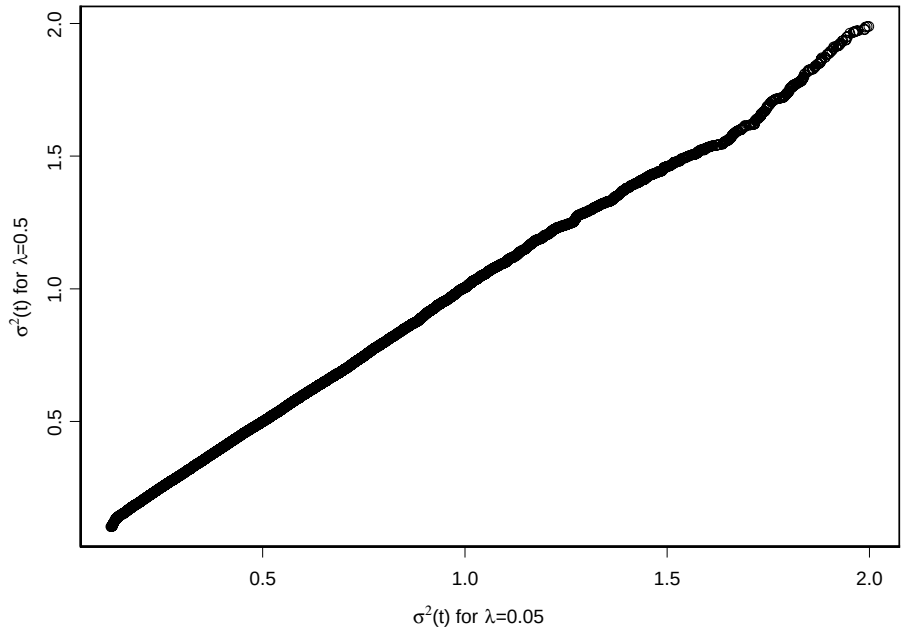


Figure 3.3: QQplot of two volatility processes with $IGa(5, 2)$ marginals but with different λ parameters.

The marginal distribution of $\sigma^2(t)$ is entirely specified by the type of BDLP which drives the Ornstein-Uhlenbeck process in equation (1.8). The relationship between $\sigma^2(t)$ and $z(t)$ is through their Lévy measures and is derived in Barndorff-Nielsen and Shephard

(2000). This means that the practitioner can either pick the form of the BDLP or the marginal distribution of the volatility (which should be on $\mathbb{R}^+$). We shall consider picking the form of the marginal distribution of $\sigma^2(t)$, as it is easier to interpret how this influences the volatility process than choosing the BDLP.

By using different marginal distributions for $\sigma^2(t)$, the weight of the tails of $\sigma(t)\,dW(t)$ can be controlled (e.g. if $\sigma^2 \sim IGa(\nu, \delta)$ then $\sigma(t)\,dW(t)$ has a *Student-t* distribution, which has much fatter tails than in the standard Black-Scholes framework with constant volatility - see for example Section 2.3.5). The model also has volatility clustering because of the correlation in the volatility process.

From equation (1.7), it can be seen that $y_i^2 \approx \sigma_i^2$. The results in Table 3.3 suggest a model with long-memory in the volatility process might be appropriate. Griffin and Steel (2003) consider a superposition of Lévy processes each with their own individual decay parameter, λ_i and BDLP. That is

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

where the weights satisfy

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

The autocorrelation function is then

$$w_1 \exp(-\lambda_1 |s|) + \cdots + w_m \exp(-\lambda_m |s|)$$

(see Barndorff-Nielsen and Shephard (2001b)). This allows the volatility process, $\sigma^2(t)$, to have quasi long-memory. We do not implement this, as, although the results from Section 3.1.3 suggest that, for the data sets the models are applied to, both $|y|$ and y^2 might have long-memory, it was clear that the superposition of processes was over-parameterised and long-memory could be incorporated equally well with fewer parameters (see Sections 3.3 and 3.5). Another variation on the BNS SV models can be found in Barndorff-Nielsen (2001), where a volatility process with long-memory is created by allowing λ to vary for each data point, whilst having that same distribution at each time point. Chapter 6 considers other continuous time SV models, driven by Lévy processes, which have a more flexible correlation structure than the BNS SV models and do not require superposition to give a long-memory model.

The superposition of Lévy processes does allow the appealing feature of having different λ's so that the BDLPs can cause jumps in the volatility at different rates. A similar idea to this is introduced in Section 3.3, by allowing λ to change over time. This keeps the same marginal distribution of volatility over the entire series and uses just one BDLP to drive the Ornstein-Uhlenbeck equation at any individual time point.

3.3 Stochastic λ in the Ornstein-Uhlenbeck model

For the BNS SV model, the correlation of the square of the log returns decays exponentially (see equation (3.2)). However, for financial data, the correlation structure of y_i^2 is often complex and this is illustrated with a real data set. The open value of McDonald's Corp shares on the NYSE from 2^{nd} Feb 1970 to 25^{th} June 2004 has 8678 log returns and a complex correlation structure, changing over time. Figure 3.4 shows two ACF plots of the square of the log returns, split at the 3769^{th} observation.

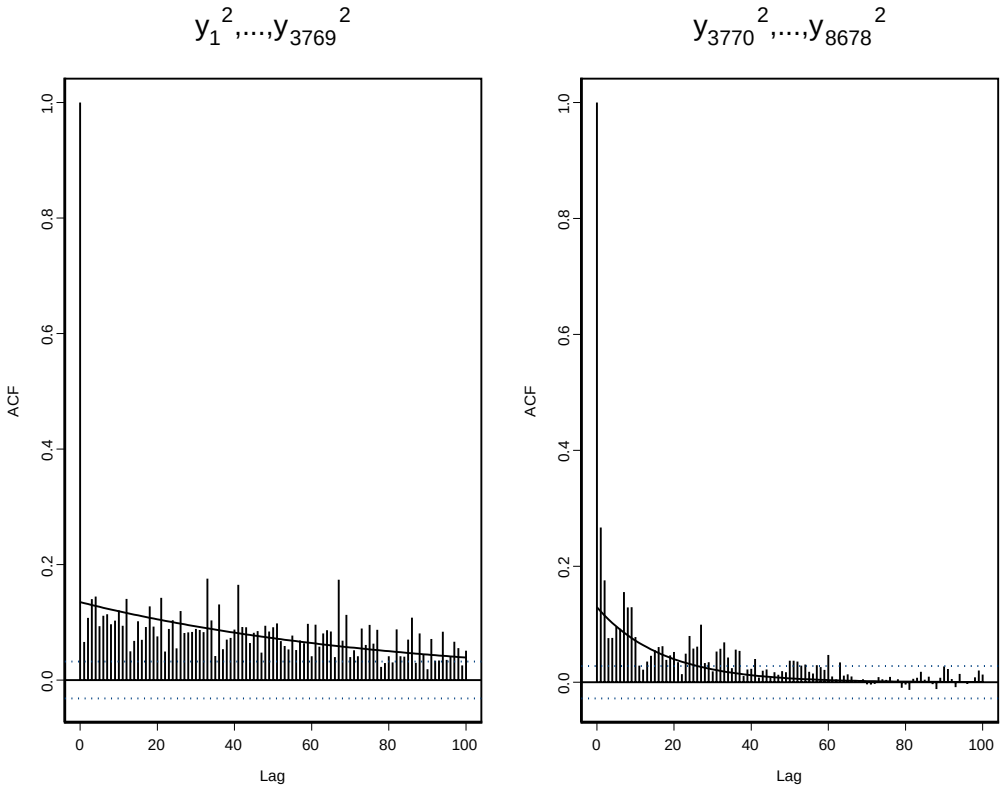


Figure 3.4: ACF plot of the square of the log returns of the McDonald's Corp data set, split at time 3769.

Figure (3.4) shows that the decay in correlation is very different over the two time periods. The lines are possible BNS SV exponential decays which could be used in each series. This motivates generalising the BNS OU SV model to allow λ to vary over time.

The BDLP, $z(\lambda t)$, in equation (1.8), is a homogeneous Lévy process, as the size and rate of jumps are stationary. By allowing λ to vary stochastically over time, the rate at which jumps in $z(\lambda t)$ occur is no longer stationary and the Lévy process is non-homogeneous.

The theorem of Wolfe (1982) requires $\lambda > 0$ in equation (1.8) and so any stochastic process for λ must be strictly positive. For the BNS SV model, the volatility jumps when new information arrives to the market and then decays exponentially. It is less obvious how the arrival of information to the markets should effect the stochastic λ process, so it is desirable to allow λ to jump both up and down, whilst still remaining strictly positive. Let $\varsigma_t = \log(\lambda_t)$ be a pure jumps Lévy process (and so has no Brownian motion part, see for example Bertoin (1994)). The model for ς_t is as follows:-

(1) Jumps occur according to a Poisson process with intensity $r/T\Delta$ $(r > 0)$.

(2) The jumps are independent and identically distributed (and independent of the Poisson process), with distribution $g()$. These jumps need to be both positive and negative to allow λ_t to increase and decrease.

ς_t will have jumps occurring in a similar way to the *Cosh* and *Cauchy* process given in Figures 2.2 and 2.3, though jumps in ς_t will not necessarily decrease in size as time increases.

Let $\{\Lambda_{t\Delta}\}$ be the stochastic process for λ_t and λ_0 be the value of λ at time zero (this is constant from the viewpoint of the stochastic process for λ) and let N_2 be the number of jumps for $\lambda_{t\Delta}$ occurring in $(0, t\Delta)$. Then $N_2 \sim Po(rt/T)$ and $E[\Lambda_{t\Delta}|N_2 = 0] = \lambda_0$.

Consider the case when $g() = N(\omega, \varepsilon^2)$. Then for $n \neq 0$,

$$\Lambda_{t\Delta}|N_2 = n \sim LN\left(n\omega + \log(\lambda_0), n\varepsilon^2\right),$$

where LN is the *log-normal* distribution. Recall if $X \sim LN(\mu, \sigma^2)$,

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{y}\exp\left\{-\frac{1}{2}\left(\frac{\log(y) - \mu}{\sigma}\right)^2\right\}$$

and $E[X] = \exp\left\{\sigma^2/2 + \mu\right\}$. Therefore, for $n \neq 0$,

$$E[\Lambda_{t\Delta}|N_2 = n, \Lambda_0 = \lambda_0] = \exp\left\{\frac{n\varepsilon^2}{2} + n\omega + \log(\lambda_0)\right\}.$$

This gives a quasi *log-normal* Lévy process. The *log-normal* distribution is a popular marginal distribution for some discrete time SV models, though this distribution is not as easy to use as a marginal in the BNS SV models (this is briefly discussed in Barndorff-Nielsen and Shephard (2001b)). This quasi *log-normal* Lévy process is used for the stochastic λ process as it is a straightforward way to introduce both positive and negative jumps and we are unable to use the *log-normal* distribution as a marginal for the distribution of $\sigma^2(t)$.

3.3.1 Properties of $\{\Lambda_t\}$

Let $c = rt/T$ and $d = \varepsilon^2/2 + \omega$. As the Poisson process and the jumps are independent,

$$\begin{aligned} E\left[\Lambda_{t\Delta}|\Lambda_0 = \lambda_0\right] &= \sum_{i=0}^{\infty} P\left(N_2 = i\right) E\left[\Lambda_{t\Delta}|N_2 = i\right] \\ &= \lambda_0 e^{-c} + e^{-c} \sum_{i=1}^{\infty} \frac{(c)^i}{i!} \exp\left\{\frac{i\varepsilon^2}{2} + i\omega + \log\left(\lambda_0\right)\right\} \\ &= \lambda_0 e^{-c} \left\{1 + \sum_{i=1}^{\infty} \frac{(c)^i}{i!} e^{id}\right\} \\ &= e^{-rt/T} \lambda_0^j \sum_{n=0}^{\infty} \frac{(rt/T)^n}{n!} \exp\left[n\left\{\frac{\varepsilon^2}{2} j\left(j-1\right)\right\}\right] \\ &= \lambda_0 \left[\exp\left\{\frac{rt}{T}\left(\exp\left(\frac{\varepsilon^2}{2} + \omega\right) - 1\right)\right\}\right]. \end{aligned}$$

To ensure the mean is stationary (i.e. $E\left[\Lambda_{t\Delta}|\Lambda_0 = \lambda_0\right] = \lambda_0$), pick $\omega = -\varepsilon^2/2$. Then only two further parameters are introduced into the model, ε^2 and r.

$$\begin{aligned} f_{\Lambda_{t\Delta}|\Lambda_0=\lambda_0}\left(\lambda_0\right) &= P\left(N_2 = 0\right) + \sum_{j=1}^{\infty} f_{\Lambda_{t\Delta}|\Lambda_0=\lambda_0, N_2=j}\left(\lambda_0\right) P\left(N_2 = j\right) \\ &= e^{-rt/T}\left[1 + c\sum_{j=1}^{\infty} \frac{(rt/T)^j}{j!} \frac{1}{\sqrt{j}} \exp\left\{-\frac{1}{2}\frac{\left(\log\left(\lambda_0\right) - \mu\right)^2}{j\varepsilon^2}\right\}\right], \quad (3.3) \end{aligned}$$

where

$$c = \frac{1}{\lambda_0} \frac{1}{\varepsilon\sqrt{2\pi}}.$$

Then for $\lambda_{t\Delta} \neq \lambda_0$,

$$\begin{aligned} f_{\Lambda_{t\Delta}|\Lambda_0=\lambda_0}\left(\lambda_{t\Delta}\right) &= \sum_{j=1}^{\infty} f_{\Lambda_{t\Delta}|\Lambda_0=\lambda_0, N_2=j}\left(\lambda_{t\Delta}\right) P\left(N_2 = j\right) \\ &= e^{-rt/T} \frac{1}{\lambda_{t\Delta}} \frac{1}{\varepsilon\sqrt{2\pi}} \sum_{j=1}^{\infty} \frac{(rt/T)^j}{j!\sqrt{j}} \exp\left\{-\frac{1}{2}\frac{\left(\log\left(\lambda_{t\Delta}\right) - \mu\right)^2}{j\varepsilon^2}\right\}. \quad (3.4) \end{aligned}$$

Equations (3.3) and (3.4) cannot be simplified further.

As $\Lambda_{t\Delta}|N_2 = n_2, \Lambda_0 = \lambda_0 \sim LN\left(-n_2\varepsilon^2/2 + \log(\lambda_0), n_2\varepsilon^2\right)$, $E\left[\Lambda_{t\Delta}^j\right]$ is

$$\begin{aligned} E\left[\Lambda_{t\Delta}^j\right] &= e^{-rt/T}\sum_{n=0}^{\infty}\frac{(rt/T)^n}{n!}\int_0^{\infty}x^j f_{\Lambda_{t\Delta}|N_2}(x|n)\,dx \\ &= e^{-rt/T}\sum_{n=0}^{\infty}\frac{(rt/T)^n}{n!}\exp\left\{j\log(\lambda_0) - nj\frac{\varepsilon^2}{2} + nj^2\frac{\varepsilon^2}{2}\right\} \\ &= e^{-rt/T}\lambda_0^j\sum_{n=0}^{\infty}\frac{(rt/T)^n}{n!}\exp\left\{n\frac{\varepsilon^2}{2}j(j-1)\right\} \\ &= e^{-rt/T}\lambda_0^j\exp\left[\left(\frac{rt}{T}\right)\exp\left\{\frac{\varepsilon^2}{2}j(j-1)\right\}\right] \end{aligned}$$

and

$$V[\Lambda_{t\Delta}] = \lambda_0^2\left[\exp\left\{\frac{rt}{T}\left(e^{\varepsilon^2}-1\right)\right\} - 1\right]. \tag{3.5}$$

For constant r, as ε increases, $V[\Lambda_{t\Delta}]$ grows rapidly and so an informative, pragmatically chosen prior for $\varepsilon^2|r$ is used to control the variance of $\Lambda_{T\Delta}$ in MCMC simulations.

Using the series expansion for the characteristic function of $\Lambda_{t\Delta}$,

$$\begin{aligned} E\left[e^{ij\Lambda_{t\Delta}}\right] &= \sum_{k=0}^{\infty}\frac{(ij)^k}{k!}E\left[\Lambda_{t\Delta}^k\right] \\ &= e^{-rt/T}\sum_{k=0}^{\infty}\frac{(ij)^k}{k!}\lambda_0^k\exp\left[\left(\frac{rt}{T}\right)\exp\left\{\frac{\varepsilon^2}{2}k(k-1)\right\}\right], \end{aligned}$$

which cannot be simplified further.

Covariance and correlation of $\{\Lambda_t\}$

Let $s \in \mathbb{Z}^+$, then $corr\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}|\lambda_0\right)$ is specified by $cov\left(\Lambda_t, \Lambda_{t+s}|\lambda_0\right)$, as $V[\Lambda_t|\lambda_0]$ is already known from equation (3.5). Then

$$cov\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}|\lambda_0\right) = E\left[\left(\Lambda_{t\Delta} - \lambda_0\right)\left(\Lambda_{(t+s)\Delta} - \lambda_0\right)|\lambda_0\right],$$

as $E[\Lambda_{t\Delta}|\lambda_0] = \lambda_0$ by construction of the stochastic process for λ_t.

Let K be the number of jumps of the stochastic λ process in $(0,t)$, so $K \sim Po(rt/T)$. Further, let J be the number of jumps in the time $(t, t+s)$, so $J \sim Po(rs/T)$ (because the number of events occurring in any two non-overlapping intervals are independent for a Poisson process). The covariance is

$$cov\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}\right) = \sum_{k=0}^{\infty}P(K=k)\sum_{j=0}^{\infty}P(J=j)\,E\left[\left(\Lambda_{t\Delta} - \lambda_0\right)\left(\Lambda_{(t+s)\Delta} - \lambda_0\right)|\lambda_0, k, j\right],$$

as the Poisson process is independent of the jumps in λ.

Note that if $k = 0$ then $\Lambda_{t\Delta} = \lambda_0$ and this does not contribute to the summation. So,

$$cov\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}\right) = \sum_{k=1}^{\infty} P\left(K = k\right) \sum_{j=0}^{\infty} P\left(J = j\right) E\left[\left(\Lambda_{t\Delta} - \lambda_0\right)\left(\Lambda_{(t+s)\Delta} - \lambda_0\right)|\lambda_0, k, j\right].$$

Also, if $j = 0$ then $\Lambda_{(t+s)\Delta} = \Lambda_{t\Delta}$ and

$$E\left[\left(\Lambda_{t\Delta} - \lambda_0\right)\left(\Lambda_{(t+s)\Delta} - \lambda_0\right)|\lambda_0, K = k, J = 0\right] = V\left[\Lambda_{t\Delta}|\lambda_0, K = k\right],$$

so

$$cov\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}\right) = \sum_{k=1}^{\infty} P\left(K = k\right) \left\{ V\left[\Lambda_{t\Delta}|\lambda_0, K = k\right] + \sum_{j=1}^{\infty} P\left(J = j\right) E_{\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}|} \right\},$$

where

$$E_{\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}|} = E\left[\left(\Lambda_{t\Delta} - \lambda_0\right)\left(\Lambda_{(t+s)\Delta} - \lambda_0\right)|\lambda_0, K = k, J = j\right]$$

denotes the conditional covariance given k and j. Hence

$$cov\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}\right) = V\left[\Lambda_{t\Delta}\right] + \sum_{k=1}^{\infty} P\left(K = k\right) \sum_{j=1}^{\infty} P\left(J = j\right) E_{\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}|} \tag{3.6}$$

and the first term of this is already known from equation (3.5). The expectation term is equal to

$$E_{\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}|} = E\left[\left(\lambda_0 Z - \lambda_0\right)\left(\lambda_0 ZY - \lambda_0\right)|\lambda_0, K = k, J = j\right],$$

where Z and Y are independently distributed random variables with distributions $Z \sim LN\left(-k\varepsilon^2/2, k\varepsilon^2\right)$ and $Y \sim LN\left(-j\varepsilon^2/2, j\varepsilon^2\right)$. The expectation part is then

$$\lambda_0^2 E\left[Z^2 Y - Z - ZY + 1\right] = \lambda_0^2 \left\{\exp\left(k\varepsilon^2\right) - 1\right\}.$$

Substituting this in equation (3.6) gives

$$\begin{aligned}
cov\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}\right) &= V\left[\Lambda_{t\Delta}\right] + \lambda_0^2 \sum_{k=1}^{\infty} P\left(K = k\right) \left\{\exp\left(k\varepsilon^2\right) - 1\right\} \sum_{j=1}^{\infty} P\left(J = j\right) \\
&= V\left[\Lambda_{t\Delta}\right] + \lambda_0^2 \left\{1 - \exp\left(-\frac{rs}{T}\right)\right\} \sum_{k=1}^{\infty} P\left(K = k\right) \left\{\exp\left(k\varepsilon^2\right) - 1\right\} \\
&= V\left[\Lambda_{t\Delta}\right] + \lambda_0^2 \left\{1 - \exp\left(-\frac{rs}{T}\right)\right\} \left\{\exp\left[\frac{rt}{T}\left(e^{\varepsilon^2} - 1\right)\right] - 1\right\} \\
&= V\left[\Lambda_{t\Delta}\right] \left\{2 - \exp\left(-\frac{rs}{T}\right)\right\} \lambda_0^2,
\end{aligned}$$

so

$$corr\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}|\lambda_0\right) = \frac{V\left[\Lambda_{t\Delta}\right] \left\{2 - \exp\left(-\frac{rs}{T}\right)\right\}}{\sqrt{V\left[\Lambda_{t\Delta}\right] V\left[\Lambda_{(t+s)\Delta}\right]}}$$

and the λ_0 terms cancel. Writing $V\left[\Lambda_{(t+s)\Delta}\right] = cV\left[\Lambda_{t\Delta}\right]$, solving for c and substituting into the above equation gives

$$corr\left(\Lambda_{t\Delta}, \Lambda_{(t+s)\Delta}\right) = \left\{2 - \exp\left(-\frac{rs}{T}\right)\right\} \sqrt{\frac{\exp\left(\frac{rt}{T}d\right) - 1}{\exp\left(\frac{r\left(t+s\right)}{T}d\right) - 1}},$$

where $d = \exp\left(\varepsilon^2\right) - 1$.

3.3.2 The $p\left(\lambda_t|\lambda_0, \varepsilon^2, N_2\right)$ prior

The model above is fitted using MCMC and, although priors for most of the models are given in Section 4.3.4, the priors for the parameters of the stochastic λ process are described here, as they are based on calculations of this section. The joint prior is

$$p\left(\lambda_t, \varepsilon^2, r\right) = p\left(\lambda_0\right) p\left(\varepsilon^2|r\right) p\left(r\right) p\left(N_2|r\right) p\left(\lambda_t|\lambda_0, \varepsilon^2, N_2\right),$$

where

$$\begin{aligned} p\left(\lambda_0\right) &= Ga\left(1,1\right) \\ p\left(r\right) &= Ga\left(1, r_p\right) \\ p\left(\varepsilon^2|r\right) &= Ga\left(1, \frac{1}{\log\left(1+\log\left(2\right)/r\right)}\right) \\ p\left(N_2|r\right) &= Po\left(r\right) \end{aligned}$$

and

$$p\left(\lambda_t|\lambda_0, \varepsilon^2, N_2\right) = \frac{(T-N_2-1)!}{(T-1)!} p\left(\lambda_{1\Delta}|\lambda_0, \varepsilon^2, J_u\right) \prod_{j=1}^{T-2} p\left(\lambda_{(j+1)\Delta}|\lambda_{j\Delta}, \varepsilon^2, J_u\right),$$

where J_u are the jump times of the stochastic λ process. The prior for $p\left(\lambda_0\right)$ is discussed in Section 4.3.4. The prior for r prevents the process from jumping too much. The prior expected number of jumps in our time series is controlled by r_p. The prior for $\varepsilon^2|r$ has a prior mean for ε^2 which gives $V\left[\Lambda_{T\Delta}\right] = \lambda_0^2$ (see equation (3.5)) and so this prior controls the variance of the end point of the stochastic process for λ_t. The factorial terms are from the different permutations of the times that the jumps in λ can occur.

Abusing notation, let $\Lambda_{j\Delta}$ be the j^{th} point at which the λ process jumps. Then

$\Lambda_{(j+1)\Delta} = \Lambda_{j\Delta} e^{X_{j+1}}$, where $X_{j+1} \overset{iid}{\sim} N\left(-\varepsilon^2/2, \varepsilon^2\right)$ and

$$\begin{aligned} F_{\Lambda_{(j+1)\Delta}|\Lambda_{j\Delta}}\left(\lambda_{(j+1)\Delta}|\Lambda_{j\Delta} = \lambda_{j\Delta}, J_u\right) &= p\left(\Lambda_{(j+1)\Delta} \leq \lambda_{(j+1)\Delta}|\Lambda_{j\Delta} = \lambda_{j\Delta}, J_u\right) \\ &= p\left(\lambda_{j\Delta} e^{X_{j+1}} \leq \lambda_{(j+1)\Delta}\right) \\ &= F_{X_{j+1}}\left(\log\left(\frac{\lambda_{(j+1)\Delta}}{\lambda_j}\right)\right) \end{aligned}$$

and

$$f_{\Lambda_{j+1}|\Lambda_j}\left(\lambda_{(j+1)\Delta}|\Lambda_{j\Delta} = \lambda_{j\Delta}, J_u\right) = \frac{1}{\lambda_{(j+1)\Delta}} \frac{1}{\varepsilon\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left\{\frac{\log\left(\frac{\lambda_{(j+1)\Delta}}{\lambda_{j\Delta}}\right) + \frac{\varepsilon^2}{2}}{\varepsilon}\right\}^2\right],$$

so

$$\begin{aligned} p\left(\lambda_t|\lambda_0, \varepsilon^2, N_2\right) &= \frac{(T - N_2 - 1)!}{(T-1)!}\left(\frac{1}{\varepsilon\sqrt{2\pi}}\right)^{N_2}\left(\prod_{j=0}^{N_2-1}\frac{1}{\lambda_{(j+1)\Delta}}\right) \\ &\quad \times \exp\left[-\frac{1}{2}\sum_{j=0}^{N_2-1}\left\{\frac{\log\left(\frac{\lambda_{(j+1)\Delta}}{\lambda_{j\Delta}}\right) + \frac{\varepsilon^2}{2}}{\varepsilon}\right\}^2\right]. \end{aligned} \tag{3.7}$$

This section introduced the theory behind the stochastic λ process, allowing the correlation parameter, λ, to vary over time. The size of the jumps are controlled by ε^2 and the number of jumps by r. The empirical performance of this model is investigated in Section 5.3.3.

3.4 Incorporating 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 3.2 is generalised to incorporate the leverage effect, using the model proposed in Barndorff-Nielsen and Shephard (2001a). This has also been implemented in Griffin and Steel (2003). Equation (1.7) is replaced by

$$dx(t) = \left\{\mu - \frac{\sigma^2(t)}{2}\right\}dt + \sigma(t)\, dW(t) + \rho d\overline{z}(\lambda t), \tag{3.8}$$

where ρ is the leverage parameter and $\overline{z}$ is a "centred" BDLP defined as

$$\overline{z}(t) = z(t) - E[z(t)]$$

(see Barndorff-Nielsen and Shephard (2001a)). The likelihood for the log returns y_i is

$$\frac{y_i - \left(\mu\Delta - \frac{\sigma_i^2}{2}\right) - \rho z_i}{\sigma_i} \overset{iid}{\sim} N(0,1), \quad \text{for} \quad i = 1, \ldots, T, \tag{3.9}$$

where z_i is a "centred" version of the BDLP and is equal to

$$\begin{aligned} z_i &= \int_{(i-1)\Delta}^{i\Delta} dz(\lambda t) - E[z(\Delta)] \\ &= \frac{1}{\lambda}\{z(\lambda i\Delta) - z(\lambda(i-1)\Delta)\} - \Delta E[z(1)]. \end{aligned}$$

As

$$\int_{(i-1)\Delta}^{i\Delta} dz(\lambda t) \overset{\mathfrak{L}}{=} z(\lambda\Delta),$$

we have $E[z_i] = 0$.

If $\rho < 0$ and there is a jump in the volatility (so $z_i > 0$), the likelihood will be greatest for $y_i < 0$ and this is equivalent to a fall in the share value $S(t)$ (i.e. falls in the share value tend to generate larger volatility values than rises).

It is important to note that ρ can take any value on the real line (or negative real line to guarantee leverage), unlike the leverage parameter(s) found in many discrete time models, where ρ is a correlation parameter between two (*normal*) variates. This can make ρ difficult to interpret quantitatively in the BNS SV model. A relationship between ρ and a leverage parameter of a popular discrete time model is given in Appendix A.8.

One further potential problem with this leverage parameter, ρ, is that the amount of leverage induced will depend on the variation in z_i, which is determined by the variance of the marginal distribution of the volatility. For a given data set, the different marginal distributions tend to give similar marginal variances and so it is possible to compare the leverage parameter for different marginals on the same data set. If two models have different marginal variances (such as for different data sets), it is difficult to compare the leverage parameters of each model.

An alternative leverage parameter, which has zero mean and unit variance, is discussed in Appendix A.9.

3.5 Fractional Brownian motion

Recalling equation (3.8), the most general model considered so far is

$$dx(t) = \left\{\mu - \frac{\sigma^2(t)}{2}\right\} dt + \sigma(t)\, dW(t) + \rho d\overline{z}(\lambda t).$$

The variance of the log returns is the product of the stochastic volatility multiplied by the independent *normal* variates (from the Brownian motion). Correlation in the log returns must be induced from correlation in the volatility process.

Section 3.1.3 suggests financial data might require a long-memory model. There are two possible approaches:-

(1) Alter the volatility process in Section 3.2 so that it has long-memory.

(2) Leave the volatility process unaltered (so it has short-memory) but alter the share equation itself (3.8).

Griffin and Steel (2003) induce quasi long-memory via (1), using a superposition of BNS SV processes (each with their own BDLP and correlation parameter) as described in Section 3.2. This is not implemented because of fears of over-parameterisation in the model and identifiability problems of the parameters. Although a superposition of volatility processes allows different BDLPs to describe short-term and long-term movements in the volatility specifically (because one process doesn't have to describe all observed data itself), estimating λ accurately is not easy (even with the single volatility process). It was felt that, as more volatility processes are used, identifying these parameters would become problematic. A (finite) superposition of processes may generate a process which looks like it has long-memory but the asymptotic behaviour of the correlation in the volatility will decay exponentially, at the slowest rate of the individual processes, and so still have short-memory. This makes it difficult to interpret quantitatively how strong the long-memory is for a given data set and estimated parameters. Instead, long-memory is induced via (2).

If the increments of the Brownian motion, $W(t)$, were positively correlated, this could reduce the amount of correlation required from the volatility process, as the correlation in the log returns could be partly explained by the correlation in the increments process for the share itself.

Equation (1.4) 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).

3.5.1 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 by Rogers (1997). For Gaussian approximations based on such processes that are arbitrarily close to fBm, fitting model parameters and pricing options gives 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. We utilise 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.

3.5.2 Properties of fractional Brownian motion

Cajueiro and Barbbachan (2003) use fBm to drive the constant volatility asset equation. They estimate $H = 0.59$, for options on Brazilian stocks, by fitting model and market option 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. There are few examples in the literature of MCMC being used to estimate the Hurst parameter, especially in finance.

The correlated increments in fBm are fractional Gaussian noise (fGn). For Brownian motion, the increments are independent and identically distributed. It is the increments, $dW(t)$, which contribute to the likelihood in equation (3.9). Graphs are now given of fBm for different Hurst parameters, H. The first graphs are sample paths of the fBm and the second graphs are ACF plots of the corresponding fGn which is driving the fBm, along with the theoretical correlation function given in equation (3.10).

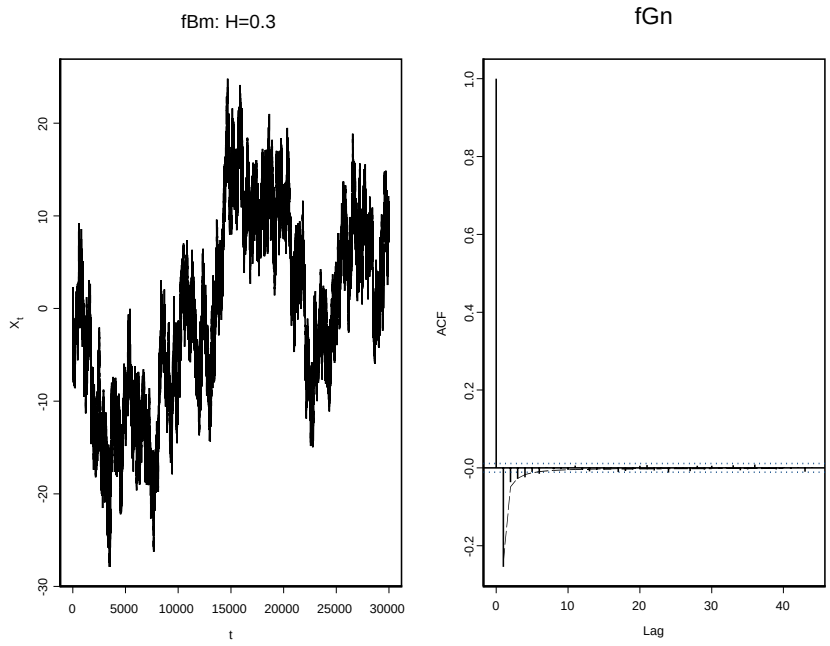


Figure 3.5: Plot of fractional Brownian motion for a Hurst parameter $H = 0.3$ and ACF plot of the corresponding fractional Gaussian noise and theoretical correlation function.

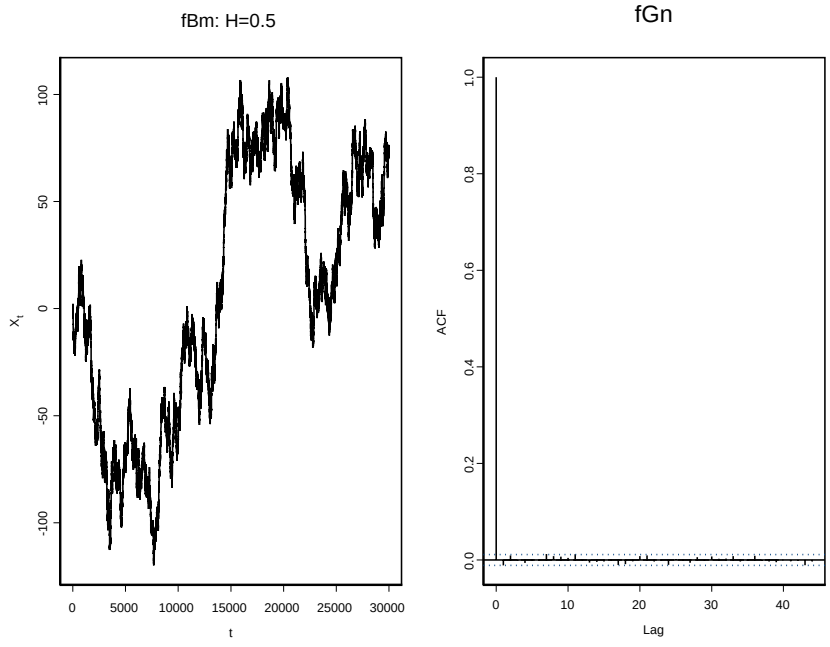


Figure 3.6: Plot of fractional Brownian motion for a Hurst parameter $H = 0.5$ and ACF plot of the corresponding fractional Gaussian noise and theoretical correlation function.

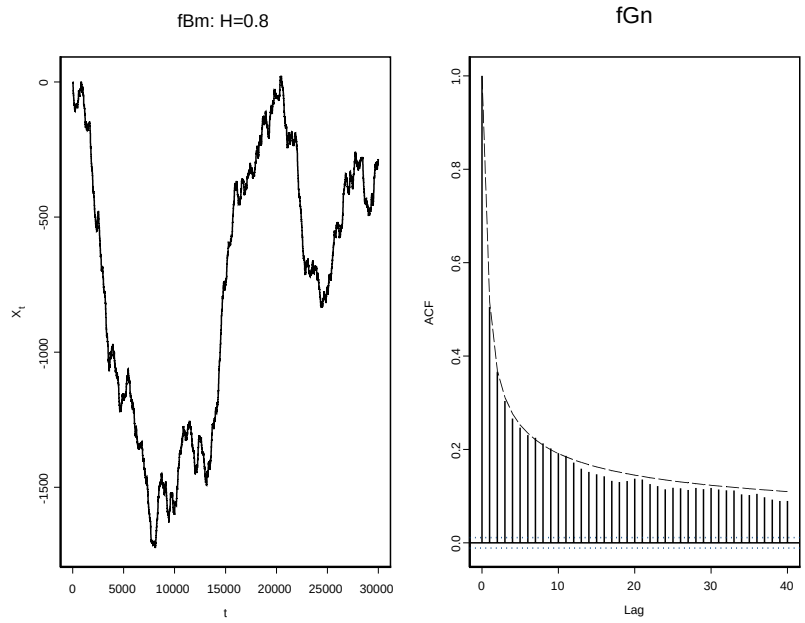


Figure 3.7: Plot of fractional Brownian motion for a Hurst parameter $H = 0.8$ and ACF plot of the corresponding fractional Gaussian noise and theoretical correlation function.

For $0 < H < 0.5$, there is negative correlation in the fGn and the fBm is noisy. For $H = 0.5$, the increments are independent as it is standard Brownian motion. For $0.5 < H < 1$, there is positive correlation in the increments and the process appears less noisy than for $H \leq 0.5$.

3.5.3 Inference for the long-memory model

Using MCMC, to make inference about model parameters, requires evaluation of the likelihood of the data, given a particular set of parameters of the model. Unfortunately for fGn, the likelihood of the log returns is not available analytically and would have to be evaluated numerically, possibly using Monte Carlo integration. This is impractical for large data sets as it requires the frequent evaluation of a high-dimensional integral in the MCMC.

To overcome this problem, the fGn is approximated by a multivariate normal (MVN) distribution, with variance/covariance matrix Σ. For fGn, with Hurst parameter H, the correlation between two increments at discretely observed times i and j is given by (and hence equal to $\Sigma_{i,j}$)

$$\Sigma_{i,j} = \frac{1}{2}\left\{|j-i+1|^{2H} - 2\,|j-i|^{2H} + |j-i-1|^{2H}\right\} \tag{3.10}$$

see Beran (1994), pg 74. Equation (3.8) is then replaced by

$$dx(t) = \mu dt - \frac{\sigma^2(t)}{2}(dt)^{2H} + \sigma(t)\, dW_{MVN}(t) + \rho d\overline{z}(\lambda t), \tag{3.11}$$

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

$$f_{\underline{Y}}(\underline{y}) = \left(\prod_{i=1}^{T} \frac{1}{\sigma_i}\right) f_{\underline{G}}(\underline{g}), \tag{3.12}$$

where

$$g_i = \frac{y_i - \left(\mu\Delta - \frac{\sigma_i^2}{2}\Delta^{2H-1} + \rho z_i\right)}{\sigma_i}$$

and $\underline{G} \sim MVN(0, \Sigma)$ (see Hu and Øksendal (2003)). In the case that $H = 0.5$, the approximate fBm recovers the original Brownian motion model.

Recall if $\underline{G} = (G_1, \ldots, G_n)^T \sim MVN(0, \Sigma)$ then

$$f_{\underline{G}}(\underline{g}) = \frac{1}{(2\pi)^{n/2}\sqrt{|\Sigma|}} \exp\left\{-\frac{1}{2}\underline{g}^T\Sigma^{-1}\underline{g}\right\}. \tag{3.13}$$

i.e. to evaluate the likelihood, the inverse of the variance/covariance matrix, Σ^{-1}, must be calculated, as well as the determinant, $|\Sigma|$.

In general, matrix inversion requires $O(n^3)$ calculations. Cholesky decomposition can be used to take advantage of the symmetry that is present in variance/covariance matrices (i.e. $\Sigma_{i,j} = \Sigma_{j,i}$) but this only doubles the speed of inversion and does not alter the order of the method. For T large (say >1000), it would not be feasible to evaluate likelihoods for MVN densities using this algorithm.

However, Σ is not only symmetric, it is Toeplitz (i.e. $\Sigma_{i,j} = \Sigma_{i+1,j+1}$: for any given south-east diagonal all the elements are the same). That is

$$\Sigma = \begin{pmatrix} 1 & r_1 & \ldots & r_{T-1} \\ r_1 & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & r_1 \\ r_{T-1} & \ldots & r_1 & 1 \end{pmatrix}.$$

An overview of different algorithms which can be used to calculate Σ^{-1} and $|\Sigma|$ is given in Ammar (1996). These algorithms solve

$$\Sigma A = v, \tag{3.14}$$

where Σ is a Toeplitz matrix of size $n \times n$, A is an unknown vector (which we want to solve) and v is some known vector. Typically, our data sets will be of size $n = 1000$.

A brief outline of different algorithms which can be used to calculate Σ^{-1} and $|\Sigma|$, for Toeplitz matrices, is now given.

Algorithms which calculate Σ^{-1}, $|\Sigma|$ and $\Sigma^{-1}v$ for a Toeplitz matrix, Σ, in $O\left(n^2\right)$ calculations are referred to as *fast* algorithms. Levinson's algorithm (see Golub and Van Loan (1996)) was used to compute these quantities in Chapters 4 and 5. This has an inversion phase (where the inverse is computed) and a solution phase (where $\Sigma^{-1}v$ is calculated). For Levinson's method, both of these are $O\left(n^2\right)$ algorithms, though there are complex algorithms to decrease this operation count to $O\left(n\log^2(n)\right)$ and $O\left(n\log(n)\right)$ algorithms respectively. These methods require very large n before they become faster than the simpler *fast* algorithms. Fast Fourier Transform based algorithms are able to compute the second phase of Levinson's algorithm in $O\left(n\log(n)\right)$ operations and are called *superfast* algorithms. These are particularly advantageous when n is very large (and preferably a power of 2) and equation (3.14) must be solved for many different v, all with the same Toeplitz matrix, Σ. In our problem we will need to solve for at most two different v (corresponding to new and old states in the MCMC) whilst using the same Toeplitz matrix. For the Levinson algorithm it is easy to compute all three quantities of interest at the same time (so three separate complex algorithms to find Σ^{-1}, $\Sigma^{-1}v$ and $|\Sigma|$ are not required).

Ammar and Gragg (1988) argued (theoretically) that one *superfast* solver should be as fast as the second phase of Levinson's algorithm for $n = 256$ (and faster thereafter) for positive definite Toeplitz matrices. In our problem, the Toeplitz matrix is symmetric and this can easily be used to give a 50% speed improvement in the Levinson algorithm (which will increase the crossover point to perhaps $n = 512$). For these reasons the *superfast* algorithms were not implemented and Levinson's algorithm for symmetric Toeplitz matrices was used.

Additionally, sampling from the model, requires samples from a $MVN\left(0,\Sigma\right)$ distribution. The standard method that shall be used for this is the upper Cholesky decomposition of Σ multiplied by a vector of independent standard normals. Therefore, the Cholesky decomposition of the symmetric Toeplitz matrix Σ must also be calculated. An example of this is where samples from the model are generated by forward simulation of the asset for option pricing and, in this thesis, this tends to be for options over a small period of time ($T \approx 20$ say). For small T, little is gained by using Cholesky decomposition algorithms which take advantage of the Toeplitz structure of Σ. Should we wish to generate from the model for T large, an $O\left(n^2\right)$ algorithm should be used to find the Cholesky decomposition and this can be found in Golub and Van Loan (1996).

From now on, even if the long-memory model is not implemented, the likelihood for the long-memory model is given. For the short-memory model, this corresponds to a Hurst parameter of 0.5, so $\Sigma = I_T$ in equation (3.13).

Chapter 4

MCMC inference and testing of the models

Chapter 3 introduced several generalisations of the Black-Scholes equation for the movement of an underlying. The unknown parameters from these models can be estimated from different data sets using MCMC. The implementation of the MCMC can be quite involved and the algorithmic details of the MCMC are given in this chapter, which require theory/results from Barndorff-Nielsen and Shephard (2001b) and Appendix D. At the end of this chapter, the MCMC implementation is tested on various training and real data sets.

Although MCMC has been used for many years, the use of Lévy processes in the Ornstein-Uhlenbeck equation is a new idea and the MCMC implementation is difficult. For this reason, there are only two thorough documented implementations of MCMC for these models, namely Roberts et al. (2004) and Griffin and Steel (2003) (although it should be noted that the implementation of Roberts et al. (2004) is not for an asset following the stochastic volatility Black-Scholes equation (1.7)), though Frühwirth-Schnatter and Sögner (2001) have performed similar inference. For these papers, the implementation is simplified because the volatility is restricted to a *Gamma* marginal distribution, which facilitates the MCMC inference (see Section 4.2.1). We extend and generalise the work in these papers.

To the best of our knowledge, no one else has performed the MCMC inference for any marginal distribution other than the *Gamma* or combined this with the generalisations introduced in Chapter 3.

The MCMC algorithm is described before the model generalisations are extensively tested. This chapter tests the following aspects of the MCMC algorithm:-

(1) The correct implementation of the different marginal distributions is tested by generating a volatility process with unit mean and variance and correlation parameter $\lambda = 0.05$ and checking that the posterior supports these values.

(2) The correct implementation of the leverage parameter is tested by generating training data with a leverage parameter $\rho = -3.0$ for each marginal distribution and checking that the posterior supports this.

(3) The correct implementation of the Hurst parameter is tested by generating training data with a Hurst parameter $H = 0.6$ for each marginal distribution and checking that the posterior supports this.

(4) The correct implementation of the stochastic λ process is tested by generating three training data sets, some with λ varying over time, and we check the posterior contains a sensible number of jumps and λ values at each time point, when compared to the λ_i which generated the training data.

4.1 Sampling scheme

The solution to the Ornstein-Uhlenbeck equation (1.8) is

$$\sigma^2(t) = e^{-\lambda t}\sigma^2(0) + e^{-\lambda t}\int_0^t e^{\lambda s} dz(\lambda s) \tag{4.1}$$

and the integrated volatility is

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

which is an important quantity for pricing European options (see Hull and White (1987)). For the SV model proposed in Barndorff-Nielsen and Shephard (2001b), it can be shown that

$$\sigma^{2*}(t) = \frac{1}{\lambda}\left\{z(\lambda t) - \sigma^2(t) + \sigma^2(0)\right\}. \tag{4.2}$$

This relatively simple form for the integrated volatility is an attractive feature of the BNS SV model. If σ_i^2 is the discretely observed volatility, then

$$\sigma_i^2 = \sigma^{2*}(i\Delta) - \sigma^{2*}((i-1)\Delta) \tag{4.3}$$

and it can be shown (see Griffin and Steel (2003), Barndorff-Nielsen (1998) and Appendix D) that

$$\begin{aligned}
\sigma_i^2 &= \sigma^{2*}(i\Delta) - \sigma^{2*}((i-1)\Delta) \\
&= \frac{1}{\lambda}\left\{z(\lambda i\Delta) - \sigma^2(i\Delta) - z(\lambda(i-1)\Delta) + \sigma^2((i-1)\Delta)\right\} \\
&= \frac{1}{\lambda}\left\{\eta_{i,2} - \eta_{i,1} + \left(1 - e^{-\lambda\Delta}\right)\sigma^2((i-1)\Delta)\right\},
\end{aligned} \tag{4.4}$$

where

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

and

$$\eta_i = \left\{ \begin{array}{c} e^{-\lambda\Delta}\int_0^{\Delta} e^{\lambda t} dz(\lambda t) \\ \int_0^{\Delta} dz(\lambda t) \end{array} \right\} = \left\{ \begin{array}{c} e^{-\lambda\Delta}\int_0^{\lambda\Delta} e^{t} dz(t) \\ \int_0^{\lambda\Delta} dz(t) \end{array} \right\} \tag{4.5}$$

is a vector of random jumps and will be referred to as the random shock vector.

If $u(x)$ is the Lévy measure of the marginal distribution of $\sigma^2(t)$ and $w(x)$ is the Lévy measure of $z(1)$, then it has been shown that (Barndorff-Nielsen (1998))

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

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

$$W_p^+(x) = \int_x^{\infty} w(y)\, dy = xu(x) \tag{4.6}$$

and the Inverse Tail Mass function as

$$W_p^{-1}(x) = \inf\left\{ y > 0 : W_p^+(y) \leq x \right\}, \tag{4.7}$$

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

Barndorff-Nielsen and Shephard (2000) proved (the proof is included in Appendix D) that, for our BDLP, $z(t)$, if $f(s) \geq 0$ for $0 < s < \Delta$ and, if $f(s)$ is integrable with respect to $dz(s)$, then

$$\int_0^{\Delta} f(s)\, dz(s) \stackrel{\mathfrak{L}}{=} \sum_{j=1}^{\infty} W_p^{-1}\left(\frac{a_j}{\Delta}\right) f(\Delta r_j), \tag{4.8}$$

where $W_p^{-1}()$ is the Inverse Tail Mass function as defined in equation (4.7), a_i are the arrival times of a Poisson process of intensity 1 and r_i are independent standard uniform variates (also independent of a_i). Note that $W_p^{-1}(a_i/\Delta) \geq 0$ 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 marginal distributions for the volatility and the only special case of these distributions, where the terms $W_p^{-1}(a_i/\Delta)$ are zero for sufficiently large a_i, is the *Gamma* distribution (see Appendix A.10).

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 (4.5) is then

$$\eta_i \stackrel{\mathfrak{L}}{=} \left\{ \begin{array}{c} e^{-\lambda\Delta} \sum_{j=1}^{n_i} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) e^{\lambda\Delta r_{i,j}} \\ \sum_{j=1}^{n_i} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) \end{array} \right\}, \tag{4.9}$$

where $a_{i,j}$ and $r_{i,j}$ are the arrival times of a Poisson point process and uniforms as described previously. Continue with the notation

$$A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n_1} \\ \vdots & \ddots & \vdots \\ a_{T,1} & \dots & a_{T,n_T} \end{pmatrix} \qquad R = \begin{pmatrix} r_{1,1} & \dots & r_{1,n_1} \\ \vdots & \ddots & \vdots \\ r_{T,1} & \dots & r_{T,n_T} \end{pmatrix}. \tag{4.10}$$

4.2 Different marginal distributions for $\sigma^2(t)$ and their Inverse Tail Mass functions

There are two possible ways to proceed when implementing the BNS OU SV models:-

(1) Pick the form of the BDLP, $z(1)$, which specifies the marginal distribution of $\sigma^2(t)$.

(2) Pick the marginal distribution of $\sigma^2(t)$, which specifies the BDLP.

Barndorff-Nielsen and Shephard (2001b) consider both (1) and (2). Roberts et al. (2004) and Griffin and Steel (2003) consider (2) where a *Gamma* marginal distribution is chosen. We pick the marginal of $\sigma^2(t)$, as it is more obvious how this relates to the overall volatility process and easier to understand how the marginal distribution of the volatility controls the tail behaviour of the log returns. In addition to the *Gamma* distribution, *Generalised Inverse Gaussian* (*GIG*) and *Tempered Stable* (*TS*) distributions are implemented, as well as their special cases, the *Positive Hyperbolic*, *Inverse Gamma, Tempered Stable* and *Inverse Gaussian* distributions. These are self-decomposable distributions on $\mathbb{R}^+$ and so are suitable choices for the marginal distribution of $\sigma^2(t)$ in equation (1.8). These distributions can give very different volatility processes and properties of the log returns. A graph demonstrating this is given in Figure C.4, where the 95% credible intervals for the volatility of four of the marginal distributions are plotted, when the BNS SV model is applied to Standard and Poor's 500 INDEX (S&P 500) data. The credible intervals are noticeably different, particularly in the tails of the volatility 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 easier than for the GIG as its Lévy Measure is simple, leading to a straightforward Inverse Tail Mass function (see equations (4.7) and (4.8)). Details on the TS distribution can be found in Tweedie (1984) and Barndorff-Nielsen and Shephard (2001c).

Ideally, the *log-normal* distribution would also have been implemented as this is a popular choice in some discrete time models. Although it is a self-decomposable distribution (see Bondesson (1992)) on $\mathbb{R}^+$ (and so is a valid choice of marginal distribution), the model structure requires the Lévy measure of the *log-normal* distribution and this is not known in closed form. The characteristic function is useful in calculating the Lévy measure and, for the *log-normal* distribution, the characteristic function is not available in analytical form either and is only available as an infinite series (see Leipnik (1991)). It might be possible to evaluate the Lévy measure numerically and from this obtain a numerical value for the Inverse Tail Mass function, though this is likely to be very difficult and slow to implement, even compared to the $GIG(\gamma, \nu, \alpha)$ distribution.

If we wanted a *log-normal* marginal distribution, we could make the log of the volatility have a *normal* distribution and then exponentiate, similar to the *Student-t* distribution process of Section 2.2.10. However, the *normal* Lévy process is Brownian motion and does not have any non-infinitesimal jumps. Therefore the exponential of such a process will also not have non-infinitesimal positive and negative jumps and this model will have very different properties to the BNS SV models. For these reasons a continuous time SV model with *log-normal* marginal distribution was not implemented.

4.2.1 Generalised Inverse Gaussian distribution: $GIG(\gamma, \nu, \alpha)$

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

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

where K_ν is a modified Bessel function of the third kind.

The Lévy measure of X is then (see equation (2.1))

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

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 (4.6), the Tail Mass function is

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

and equation (4.7) implies the Inverse Tail Mass function is

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

where z satisfies

$$x = \left\{ \frac{1}{2} \int_0^\infty \exp\left(-\frac{z\xi}{2\nu^2}\right) g_\gamma(\xi)\, d\xi + \max(0,\gamma) \right\} \exp\left(-\frac{\alpha^2 z}{2}\right). \tag{4.11}$$

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 (including the origin). Gauss-Laguerre integration was used to evaluate the remaining integral on the infinite domain (see Atkinson (1988) and Appendix B.1 for details on these numerical algorithms). The $GIG(\gamma,\nu,\alpha)$ marginal was implemented, as well as three standard distributions which are special cases.

a) Gamma distribution: $Ga(\nu,\alpha)$

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

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

Using equation (4.6), the Tail Mass function is

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

and equation (4.7) 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 (4.9) need not be truncated (a proof of this is given in Appendix A.10). For all the other marginals considered, the infinite sum which constructs η_i must be truncated. Details on this truncation can be found in Section 4.3.5 and Appendix B.2.

b) Positive Hyperbolic distribution: $RPH(\nu, \alpha)$

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

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

The Inverse Tail Mass function is available as a special case of equation (4.11) and can be evaluated using a similar method to the one used for the GIG marginal distribution.

c) Inverse Gamma distribution: $IGa(\nu, \alpha)$

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

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

The Inverse Tail Mass function is available as a special case of equation (4.11) and can be evaluated using a similar method to the one used for the GIG marginal distribution.

4.2.2 Tempered Stable distribution: $TS(\kappa, \nu, \alpha)$

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

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

where

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

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

$$u(x) = Ax^{-B-1} e^{-Cx}, \tag{4.12}$$

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

$$W^{-1}_{\kappa,\nu,\alpha}(x) = \left(\frac{A}{x}\right)^{1/B} \exp\left[-LambertW\left(\frac{C}{B}\left(\frac{A}{x}\right)^{1/B}\right)\right], \tag{4.13}$$

where the *LambertW* function satisfies

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

and is a standard function available numerically. For further details on the *LambertW* function see Jeffrey et al. (1996).

For the *Tempered Stable* distribution, an alternative series representation to equation (4.8) 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 4.1.

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

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

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 4.1. 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.

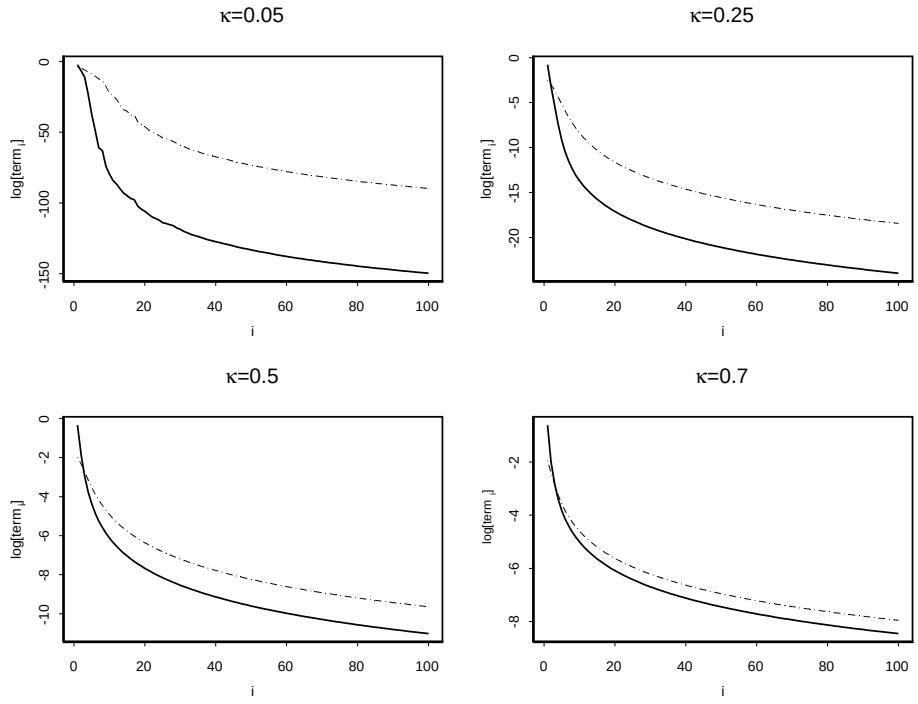


Figure 4.1: Graphs of the log of the individual terms for the Barndorff-Nielsen and Shephard (2000) and Rosiński (2000) series representation for the $TS(\kappa, 1, 1)$ distribution.

Figure 4.1 shows that a higher proportion of the mass of the sum is in the first few terms for the Barndorff-Nielsen and Shephard (2000) series, though each term is more expensive to compute (because of the Inverse Tail Mass function given in equation (4.13)). As $\kappa \to 1$, the difference between the two representations becomes smaller though there is still a noticeable advantage for $\kappa = 0.7$.

a) Inverse Gaussian distribution: $IG(\nu, \alpha)$

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

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

From equation (4.13) the Inverse Tail Mass function, defined in equation (4.7), is

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

Note that, although the $IG(\nu,\alpha)$ distribution is a special case of the both the $GIG(\gamma,\nu,\alpha)$ and $TS(\kappa,\nu,\alpha)$ distributions (when $\gamma=-0.5$ and $\kappa=0.5$), it is easier to evaluate the Inverse Tail Mass function given in equation (4.13) than the one given in (4.11).

4.2.3 Properties of the six marginal distributions

The mean and variance of $\sigma^2(t)$ for the six different marginal distributions are given in Table 4.1, along with the kurtosis of the distribution of the log returns. This is useful for comparing the behaviour of the marginals by matching up the mean/variance of the distributions so that the volatility processes are in some sense similar. Let $B_\gamma = K_\gamma(\nu\alpha)$, then the mean and variance of the distributions are given in Table 4.1.

$f_{\Sigma^2}(\sigma^2)$	Mean	Variance	Kurtosis of log returns
$GIG(\gamma,\nu,\alpha)$	$\nu B_{1+\gamma}/\alpha B_\gamma$	see equation (4.14)	see equation (4.16)
$TS(\kappa,\nu,\alpha)$	$2\nu\kappa\alpha^{1-1/\kappa}$	$4\kappa(1-\kappa)\nu\alpha^{1-2/\kappa}$	$3\frac{\kappa\nu\alpha+1-\kappa}{\kappa\nu\alpha}$
$IG(\nu,\alpha)$	ν/α	ν/α^3	$3\frac{1+\nu\alpha}{\nu\alpha}$
$Ga(\nu,\alpha)$	ν/α	ν/α^2	$3\frac{1+\nu}{\nu}$
$RPH(\nu,\alpha)$	$\frac{1}{\alpha^2}(2+\nu\alpha B_0/B_1)$	see equation (4.15)	see equation (4.17)
$IGa(\nu,\alpha)$	$\frac{\alpha}{\nu-1}$	$\frac{\alpha^2}{(\nu-1)^2(\nu-2)}$	$3\frac{\nu-1}{\nu-2}$

Table 4.1: Mean and variance of the six different marginal distributions and the kurtosis of the log returns.

For the *Inverse Gamma* distribution, the mean, variance and kurtosis are correct when $\nu>1$ and $\nu>2$, whilst for the *Tempered Stable*, *Inverse Gaussian*, *Gamma* and *Positive Hyperbolic* marginals, the mean, variance and kurtosis are always correct for valid parameters $\nu,\alpha>0$.

The mean and variance for the *Tempered Stable* distribution can be found in Appendix A.2 and Schoutens (2003). There is a slight error in the skewness and kurtosis in Schoutens (2003), though once the correct values are known, in principle, calculating the kurtosis of the log returns is straightforwards by calculating the non-centralised moments for the *Tempered Stable* distribution. As there is little cancelling/factoring in the algebra, the

kurtosis of the *Generalised Inverse Gaussian* and *Positive Hyperbolic* distributions were not included in the table.

For the *Generalised Inverse Gaussian* distribution, the variance is

$$\frac{\nu}{\alpha^3 B_\gamma^2}\left\{\nu\alpha\left(B_\gamma^2 - B_{1+\gamma}^2\right) + 2B_\gamma B_{1+\gamma}\left(1+\gamma\right)\right\} \tag{4.14}$$

and for the *Positive Hyperbolic* distribution, it is

$$\frac{1}{\alpha^4}\left\{4 + \nu^2\alpha^2 - \nu^2\alpha^2\left(B_0/B_1\right)^2\right\}. \tag{4.15}$$

For the *Generalised Inverse Gaussian* distribution, the kurtosis of the log returns is

$$3\frac{B_{|\gamma|}\left\{cB_{|\gamma|} + 2\left(1+\gamma\right)B_{|\gamma+1|}\right\}}{cB_{|\gamma+1|}^2}, \tag{4.16}$$

whilst for the *Positive Hyperbolic* distribution, it is

$$3B_1\frac{\left\{c^2B_1 + 4cB_0 + 8B_1\right\}}{\left\{cB_0 + 2B_1\right\}^2}, \tag{4.17}$$

where $c = \nu\alpha$.

For all six different marginal distributions for $\sigma^2\left(t\right)$, 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.

Now we are able to pick the marginal distribution for $\sigma^2\left(t\right)$ 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 now given.

4.3 MCMC algorithm

We try to keep the MCMC algorithm as general as possible and do not focus on proposals or priors tailored for any specific marginal distribution for $\sigma^2\left(t\right)$. Also, we shall not concentrate on trying to find 'optimal' proposals as, in practice, for option pricing most of the computer time is spent on the Monte Carlo integration (simulating the asset forward) rather than the MCMC itself i.e. the chain can easily be run for a longer period of time and/or thinned more and overall it will not significantly increase the amount of time required to price most options numerically.

The 'generic' proposals used here are not significantly outperformed by those considered in Roberts et al. (2004) and Griffin and Steel (2003) when applied to training data.

This is because the correlation between the parameters can be very complex, even when using training data and this makes constructing efficient proposals difficult. For real data, the correlation between the parameters can become even more unpredictable and good proposals are harder to obtain. The proposals used in Griffin and Steel (2003) were designed for training data, where the correlation between parameters is more predictable. A comparison of our algorithm and the preferred method of Roberts et al. (2004) is given in Section 4.4.6.

4.3.1 Treating $\sigma^2(0\Delta)$ as an unknown parameter

Equation (4.4) requires $\sigma^2(0\Delta)$ to be known, which Griffin and Steel (2003) define as some constant. This can cause problems in the MCMC, depending on the properties of the data to which the model is applied and the constant that is chosen. An example of this is shown in Appendix C.2, which shows that setting $\sigma^2(0\Delta)$ to be some constant can cause non-stationarity in the mean and variance of $\sigma^2(t)$. In their comment to Barndorff-Nielsen and Shephard (2001b), J.E. Griffin and M.F.J. Steel said they used $\sigma^2(0\Delta)$ drawn from the prior marginal process. However, the graphs of $\sigma^2(t)$ in Griffin and Steel (2003), pg 15, 16, 17, have $\sigma^2(0\Delta) = 1$. Assuming they rescaled their data so that the marginal mean was one, simulations of $\sigma^2(0\Delta)$ should be concentrated near $\sigma^2(0\Delta) = 1$, though not actually equal to one, so they do not appear to have implemented this prior.

Roberts et al. (2004) consider $\sigma^2(0\Delta)$ as missing data. Whilst this does not lead to the same problems as defining it to be some constant, it reduces the amount of available data and it is not obvious how much data should be used to ensure that the instantaneous volatility has converged to its correct value. The amount of data that must be discarded is highly influenced by the value of λ in equation (1.8) because of the correlation structure of $\sigma^2(t)$.

For the *Tempered Stable* marginal, Rosiński (2000) has derived the distribution of $\sigma^2(0\Delta)|\kappa, \nu, \alpha$ in terms of an infinite series, which avoids the need to define it as some constant or to treat it as missing data. Details of this series representation are given in Section 4.3.2 and Appendix C.2.

For all other marginals, we let $\sigma^2(0\Delta)$ be a latent parameter in the MCMC, with a prior the same as the marginal distribution of $\sigma^2(t)$. Further details on this can be found in Appendix C.2.

4.3.2 The distribution of $\sigma^2(0\Delta)$ for the Tempered Stable marginal

Equation (4.4) requires $\sigma^2(0\Delta)$ to be known. For all distributions other than the *Tempered Stable*, it is treated as a latent parameter, with a prior the same as the marginal distribution used for $\sigma^2(t)$, as described in Section 4.3.1.

The prior for $\sigma^2(0\Delta)$ requires evaluation of the density function of the marginal distribution of $\sigma^2(t)$. For the *Tempered Stable* marginal, it is difficult to evaluate the density (techniques to evaluate the *Tempered Stable* density can be found in Nolan (1997)), particularly for small arguments (see Appendix B.3) and so a different representation for $\sigma^2(0\Delta)$ is used. Rosiński (2000) has shown that, when $\sigma^2(t)$ has a $TS(\kappa,\nu,\alpha)$ marginal distribution, the distribution of $\sigma^2(0\Delta)$ satisfies

$$\sigma^2(0\Delta) \overset{\mathfrak{L}}{=} \sum_{i=1}^{\infty} \min\left\{\left(\frac{a_i\kappa}{A}\right)^{-1/\kappa}, e_i v_i^{1/\kappa}\right\}, \tag{4.18}$$

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

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

This is the representation used in Barndorff-Nielsen and Shephard (2001c).

Although the series in equation (4.18) avoids the calculation of the Inverse Tail Mass function, the series converges more slowly than the Barndorff-Nielsen and Shephard (2000) series representation (see Section 4.2.2). The equivalent Barndorff-Nielsen and Shephard (2000) series representation for equation (4.18) is

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

and, for a *Tempered Stable* marginal, this gives

$$\sigma^2(0\Delta) \overset{\mathfrak{L}}{=} \sum_{i=1}^{\infty} \left(\frac{A}{a_{0,j}}\right)^{1/\kappa} \exp\left[-LambertW\left\{\frac{C}{\kappa}\left(\frac{A}{a_{0,j}}\right)^{1/\kappa}\right\}\right]. \tag{4.19}$$

To be consistent with the representation used in equation (4.9), the equivalent representation of equation (4.19) is used, avoiding evaluation of the *Tempered Stable* density.

4.3.3 Treating the interest rate, μ, as an unknown parameter

Interest rate modelling is a very complicated problem in its own right; there are many models which are used for interest rate prediction and there is no universally accepted

method (see James and Webber (2000) for a review of interest rate models). Should a complicated model for the interest rate be used, this would make the MCMC more complex. For option pricing, the interest rate will alter how the asset is simulated forward, as well as how much the expected payoff of the option must be discounted by, to give the fair price.

The models in this thesis were constructed to try to improve the Black-Scholes model over short periods of time. The options priced in Chapter 5 are over 20 days, where it is realistic to assume that the interest rate is approximately constant and so can be treated as a further unknown parameter in the MCMC. This is the approach taken in Griffin and Steel (2003), and is the one adopted here.

4.3.4 Priors

For each marginal distribution, there are two parameters, ν and α (and additionally one further parameter for the *Generalised Inverse Gaussian* and *Tempered Stable* distributions), which specify the exact form of the distribution of $\sigma^2(t)$. The λ parameter controls the rate of exponential decay in the volatility as well as the rate at which the jumps from the BDLP contribute to $\sigma^2(t)$. The interest rate, μ, is treated as a further unknown parameter in the MCMC, even though it could be argued that the interest rate will be known for share options (provided the options start and end near to the last data point). For FX data, μ would typically be unknown and for this reason it was treated as a further unknown, for share data, so options on FX rates and shares are priced using the same algorithm. The leverage parameter, ρ, and Hurst parameter, H, are the remaining non-latent parameters of the models. The priors for the parameters controlling the stochastic λ process are as described in Section 3.3.

For the latent parameters, the instantaneous volatility at time zero is given a prior which is the same as the marginal distribution of $\sigma^2(t)$. The priors for rows of A and R (defined in Section (4.10)) are the arrival times of Poisson processes of intensity 1 and standard uniform distributions (all independent).

The joint prior is

$$\begin{aligned} p\left(\sigma^2(0\Delta), \rho, \mu, \lambda, \gamma, \kappa, \nu, \alpha, H, A, R\right) &= p\left(\sigma^2(0\Delta)\,|\nu, \alpha\right) p(\rho)\, p(\mu)\, p(\lambda)\, p(\gamma)\, p(\kappa) \\ &\quad \times p(\nu)\, p(\alpha)\, p(A)\, p(R)\, p(H). \end{aligned}$$

Reparameterise by letting

$$\begin{aligned} x_1 &= \alpha^{c_1}\nu^{c_2} \\ x_2 &= \alpha^{c_3}\nu^{c_4}, \end{aligned}$$

where $c_1, \ldots, c_4$ are constants chosen to reduce the correlation in the chain. This transformation must be invertible, which puts some constraints on the possible values of $c_1, \ldots, c_4$. The implied prior for x_1, x_2 is (see Appendix C.3)

$$p(x_1, x_2) \propto p(\alpha(x_1, x_2)) p(\nu(x_1, x_2)) \{\alpha(x_1, x_2)\}^{1-c_1-c_3} \{\nu(x_1, x_2)\}^{1-c_2-c_4}.$$

For the original parameters, the priors used throughout are

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

Table 4.2: Priors for the MCMC.

Note that $\sigma^2(0\Delta)$ is not treated as an additional parameter with a Metropolis-Hastings update for the *Tempered Stable* distribution and is instead updated by performing a Metropolis-Hastings update on the parameters determining $\sigma^2(0\Delta)$ in equation (4.19).

The term r_p (from the stochastic λ process) was chosen to be 1.0 whilst m_p was taken as 0.0004 (so one standard deviation corresponds to approximately 15% interest per year). Typically n_p and a_p are chosen to be small (say 0.001), so that the priors are reasonably flat. Note that the latter two are not priors for the original GIG parameters but are priors for the parameters of the specific marginal.

A discussion of the κ prior is given in Appendix C.4. For the *Inverse Gamma* marginal, a $Ga(1, n_p)$ prior for $(\nu - 2)$ is used, so the mean, variance and kurtosis of the log returns are finite (see Section 4.2.3).

These priors are not consistent across each marginal. Consistent priors could be chosen so that the mean and variance of the volatility process (and skewness for three parameter distributions) are the same in the prior, though this can be difficult because of the complex nature some of the means and variances. It turns out that the posterior is not strongly dependent on the prior. By using relatively flat priors for parameters controlling the marginal distribution, we can easily compare our results with those of Roberts et al. (2004) and Griffin and Steel (2003).

[1] M is the marginal distribution used for $\sigma^2(t)$.

Griffin and Steel (2003) suggest using a $Ga(1,1)$ prior for λ. Little difference was found between this prior and a less informative $Ga(1,0.001)$ prior, apart from on training data which had been generated using constant volatility (see Section 5.2.3). The $Ga(1,1)$ prior kept the system smaller in size, whilst not affecting the simulated volatility process σ_i^2 significantly. For training data, the flatter prior for λ is used and for real data, the informative prior, $Ga(1,1)$, is used. A $Ga(1,250)$ prior is used for μ when the data sets are training data sets generated with $\mu > 0$ or shares (which should have a positive drift, as it should be equal to the interest rate of the country in which the shares are traded). For this prior, the mean is 4.3% interest per year and the probability that the interest rate is less than 28% is 0.9.

4.3.5 Reverse jump and truncation of the random shock vector

The infinite series representation of the random shock vector is

$$\eta_i \overset{\mathfrak{L}}{=} \left\{ \begin{array}{c} e^{-\lambda\Delta} \sum_{j=1}^{\infty} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) e^{\lambda\Delta r_{i,j}} \\ \sum_{j=1}^{\infty} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) \end{array} \right\}, \tag{4.20}$$

where p are the parameters specifying the marginal distribution for $\sigma^2(t)$. As $\eta_{i,2} > \eta_{i,1} > 0$, we focus on suitable truncation for $\eta_{i,2}$, as $\eta_{i,1}$ should be accurate if $\eta_{i,2}$ is.

Griffin and Steel (2003) suggest truncating after the terms $W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right)$ fall below a certain level. Then

$$\eta_i \overset{\mathfrak{L}}{=} \left\{ \begin{array}{c} e^{-\lambda\Delta} \sum_{j=1}^{\infty} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) e^{\lambda\Delta r_{i,j}} \\ \sum_{j=1}^{\infty} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) \end{array} \right\} \approx \left\{ \begin{array}{c} e^{-\lambda\Delta} \sum_{j=1}^{n_i} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) e^{\lambda\Delta r_{i,j}} \\ \sum_{j=1}^{n_i} W_p^{-1}\left(\frac{a_{i,j}}{\lambda\Delta}\right) \end{array} \right\}, \tag{4.21}$$

where n_i is the number of Poisson points which are less than some critical value, a_c.

If only terms from the Poisson point process which contribute at least z_{tol} to the summation are included, then the critical a, a_c, at which the Poisson points are truncated will satisfy

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

so

$$a_c = \lambda\Delta W_p^{+}(z_{tol}). \tag{4.22}$$

Note that if λ varies over time (as in Section 3.3), then there is a vector of a_c's and an alteration in any of the elements of p alters all of the critical "a" values in a similar

way (i.e. each critical "a" value is proportional to the value of λ in that segment, so if a change in p increases one critical "a" value, it increases all of them). This is convenient for reverse jump MCMC as it is easy to calculate how λ moves alter the a_c's, provided a simple truncation scheme as suggested above and in Griffin and Steel (2003) is used. If a more advanced variable truncation scheme is used (as in Appendix B.2), each individual segment of a_c's must be calculated without knowledge of the previous value, as $a_c|z_{tol}, \nu, \alpha$ is no longer proportional to λ for all parameter values.

For the *Gamma* marginal, the series does not need to be truncated, as only a finite number of the terms have non-zero contribution. From equation (4.22)

$$a_c = \lambda \Delta \nu e^{-\alpha z_{tol}}$$

and, setting $z_{tol} = 0$ (no approximation/truncation), gives $a_c = \lambda \nu \Delta$.

For the $GIG(\gamma, \nu, \alpha)$ marginal

$$a_c = \lambda \Delta \left\{ \frac{1}{2} \int_0^\infty \exp\left(-\frac{z_{tol}\xi}{2\nu^2}\right) g_\gamma(\xi)\, d\xi + \max(0, \gamma) \right\} \exp\left(-\frac{\alpha^2 z_{tol}}{2}\right).$$

For the $RPH(\nu, \alpha)$ marginal

$$a_c = \frac{\lambda \Delta}{\pi^2} \exp\left(-\frac{\alpha^2 x}{2}\right) \int_0^\infty \frac{\exp\left(-\frac{z_{tol} y}{2\nu^2}\right)}{y\left\{J_1^2\left(\sqrt{y}\right) + N_1^2\left(\sqrt{y}\right)\right\}} dy + 1.$$

For the $IGa(\nu, \alpha)$ marginal

$$a_c = \frac{\lambda \Delta}{\pi^2} \int_0^\infty \frac{\exp\left(-\frac{z_{tol} y}{2\alpha^2}\right) dy}{y\left\{J_{|\nu|}^2\left(\sqrt{y}\right) + N_{|\nu|}^2\left(\sqrt{y}\right)\right\}} dy.$$

For the $TS(\kappa, \nu, \alpha)$ marginal

$$a_c = \frac{\lambda \nu \Delta \kappa 2^\kappa}{\Gamma(1-\kappa)} z_{tol}^{-\kappa} \exp\left(-\frac{\alpha^{1/\kappa} z_{tol}}{2}\right).$$

For the $IG(\nu, \alpha)$ marginal

$$a_c = \frac{\lambda \nu \Delta}{\sqrt{2\pi z_{tol}}} \exp\left(-\frac{\alpha^2 z_{tol}}{2}\right).$$

In practice, this approximation works well for the six distributions and most parameter values for $z_{tol} = 0.001$, provided the data are rescaled so that $E\left[y_i^2\right] = 1$. Details on rescaling can be found in Section 5.2.1. This can be verified by sampling from the model and testing the marginal distribution and correlation structure of the instantaneous volatility process, such as in Figure 3.2. For the *Generalised Inverse Gaussian, Positive Hyperbolic* and *Inverse Gamma* distributions it is difficult to improve the truncation scheme

because of the form of $W_p^{-1}(x)$ in equation (4.11). For the *Tempered Stable* and *Inverse Gaussian* marginals, $W_p^{-1}(x)$ is simpler and an improved truncation scheme is employed, using knowledge of the asymptotic behaviour of $W_p^{-1}(x)$. For a $TS(\kappa, \nu, \alpha)$ marginal and for large x, it can be shown that

$$W_p^{-1}(x) \approx 2\left(\frac{\nu\kappa}{\Gamma(1-k)}\right)^{1/\kappa}\left(\frac{1}{x}\right)^{1/\kappa}.$$

As κ and ν determine the rate of decay of the terms, a more advanced truncation scheme is considered, which depends on the parameters κ, ν, α (see Appendix B.2). This is also used for the *Inverse Gaussian* marginal ($TS(\frac{1}{2}, \nu, \alpha)$).

Whenever a move in x_1, x_2 or λ is performed, in order to maintain a constant level of precision in the approximation of the infinite sum, we add (or remove) a's and r's (using RJMCMC). The proposals for the new a's and r's are direct from their priors given the change in a_c. The proposal to remove a's and r's is deterministic, with probability specified by the upwards dimension move, because of the construction of reverse jump MCMC. The acceptance probability for a move on x_1 is then

$$\min\left[1, \frac{p\left(y|\sigma'^2\right)}{p\left(y|\sigma^2\right)}\frac{p\left(x_1', x_2\right)}{p\left(x_1, x_2\right)}\frac{q\left(x_1' \to x_1\right)}{q\left(x_1 \to x_1'\right)}\right]$$

(see Appendix C.5). Acceptance probabilities for moves on x_2 and λ are similar.

4.3.6 Proposals

The MCMC algorithm is as follows

(1) γ, κ move (reverse jump) (if a *Generalised Inverse Gaussian* or *Tempered Stable* marginal is used).

(2) x_1 move (reverse jump).

(3) x_2 move (reverse jump).

(4) λ or λ_0 move (reverse jump) (if we have a stochastic λ process as in Section 3.3).

(5) ε^2 move (fixed dimension).

(6) r move (fixed dimension).

(7) N_2 move (reverse jump).

(8) Update the jump times of the stochastic λ process (reverse jump).

(9) Update the jump sizes of the stochastic λ process (reverse jump).

(10) H move (fixed dimension).

(11) $\sigma^2(0\Delta)$ move (fixed dimension) (provided the marginal for the volatility is not *Tempered Stable*).

(12) μ move (fixed dimension).

(13) ρ move (fixed dimension).

(14) Joint A and R move (fixed dimension). If the marginal for the volatility is *Tempered Stable*, this will update $\sigma^2(0\Delta)$ using equation (4.19).

For (2), the proposal is from a $Ga(c_1, c_1/x_1)$ density. Similar moves are used for (3), (4), (5), (6), (9), and (11) as these parameters are also restricted to $\mathbb{R}^+$. For parameters on the entire real line (γ, μ and ρ) a symmetric random walk proposal is used. A symmetric random walk on the positive integers is used for the N_2 move in (7). For parameters on a finite domain (κ and H) a similar *Gamma* local proposal is used, as was used for x_1, but on a function of the parameter (see Appendix C.6). The different constants were chosen to give acceptance rates of approximately 35%.

For the A and R update (14), if $n_i = 0$ then no update can be performed on the i^{th} row of A or R, as the dimension is kept fixed. Otherwise, a move on each row is proposed for A and R in turn, again with proposals direct from the priors given n_i. As proposals are direct from their priors, the acceptance probability is

$$\min\left[1, \frac{p\left(y|\sigma'^2\right)}{p\left(y|\sigma^2\right)}\right].$$

One potential problem with this move is that it is not a "local" move, so acceptance might be poor. It also requires the generation of σ_j^2 for each $j \geq i$ and the likelihood must be calculated for values at which σ_j^2 has altered, which can be expensive. The most extreme example of this is for the *Gamma* marginal, as the Inverse Tail Mass function is easy to compute and a lot of time (approximately 40%) is spent generating the volatility process and evaluating the likelihood. As the Inverse Tail Mass function becomes more difficult to compute (such as for the *Generalised Inverse Gaussian* marginal), the proportion of time spent evaluating the likelihood decreases and this is less of a concern.

Roberts et al. (2004) use a clever update for several points of A and R, which does not alter the likelihood and avoids the problem described above. This update is aided by the simple Inverse Tail Mass function from a *Gamma* marginal. Similar techniques can be implemented for other marginals but numerical errors can cause problems using this

update and so it is not implemented. Even without this proposal, our algorithm performs similarly to the preferred algorithm of Roberts et al. (2004) (see Section 4.4.6). Further details on the MCMC algorithm can be found in Appendix C.5.

4.4 Testing the algorithm on simulated data

Tests are now performed to verify correct implementation of the MCMC. Test1, Test2 and Test3 are tests on training data of size $T = 1500$. For these tests, the results for the *Gamma* marginal are similar to those of Griffin and Steel (2003). Priors with higher variance than described in Section 4.2 are used for λ and κ, to test the correct implementation of the MCMC. The prior for λ is $Ga(1, 0.001)$ and a $U(0, 1)$ prior is used for κ. This is not sensible for real data, as it supports $\kappa = 0, 1$ (which give marginals concentrated at one point - constant volatility). This uniform prior does not cause problems for training data generated from the $TS\left(\frac{1}{2}, 1, 1\right)$ distribution, as $\kappa = \frac{1}{2}$ is well identified and the MCMC does not visit $\kappa = 0, 1$. On real data (such as Test5 and Chapter 5), with the uniform prior, the MCMC sometimes visits large κ values (which gives nearly constant volatility $\sigma^2(t) = 2\nu$). This causes truncation problems in the infinite sum of equation (4.9). For this flat prior, the MCMC does not visit $\kappa = 0$ and so, for real data, a more informative prior, $Beta(1, 15)$, is used for κ to prevent the truncation errors of the infinite sum. Further details on the κ prior can be found in Appendix C.4. As described in Griffin and Steel (2003) and Section 4.3.4, an informative $Ga(1, 1)$ is used for λ for real data. Test4 is designed to test the correct implementation of the algorithm for the stochastic λ process and Test5 is a test on a real data set (the S&P 500 data from 2^{nd} January 1980 to 30^{th} December 1987). This data set has been used in Griffin and Steel (2003), where a *Gamma* marginal was used with inference via MCMC. This data set has also been used in Gallant et al. (1992) and Jacquier et al. (1994), though for different volatility models. Nicolato and Venardos (2003) use the same stochastic volatility model with *Gamma* and *Inverse Gaussian* marginals, making inference by minimising the squared difference between 87 model-produced and market option prices on the S&P 500 on 2^{nd} November 1993. i.e. their inference is based on the option prices on a given date rather than the log returns themselves (note that the S&P 500 data set used in Nicolato and Venardos (2003) are market option prices on 2^{nd} November 1997, whilst our data set ends on the 30^{th} December 1987). This makes it more difficult to compare results quantitatively. In Test6 the efficiency of our MCMC algorithm is investigated and compared with the Hybrid algorithm proposed in Roberts et al. (2004).

For tests which are not testing the different marginal distributions, the *Gamma* marginal was used as it is the fastest to run.

This section demonstrates that the MCMC has been implemented correctly and the empirical performance of the models, for real data sets, will be examined in Chapter 5.

4.4.1 Test1

Test1 was constructed to test the correct implementation of proposals (1), (2), (3), (4), (11), (12) and (14). These are proposals for the parameters controlling the exact form of the marginal distribution, $\sigma^2(0\Delta)$, μ and A and R. Test1 does not involve a leverage parameter for reasons which were briefly mentioned in Section 3.4 and these will be discussed in Test2.

Parameters of the marginal distribution were chosen to give the marginal a unit mean and variance and $\sigma^2(0\Delta)$ was set to 0.5. This allows us to see how dominant the prior for $\sigma^2(0\Delta)$ is, as the prior should have a mean of approximately 1. The training data were generated with $\lambda = 0.05$ and $\mu = 0.0001$ (3.7% interest rate per year). This value of λ is typical for observed financial data (see for example Griffin and Steel (2003)). For Test1, we set $\rho = 0$ and this was held constant (no MCMC was performed on ρ). For the *Generalised Inverse Gaussian* distribution, the data were generated with $\gamma = -0.5$ and for the *Tempered Stable* distribution, the data were generated with $\kappa = 0.5$ (these both correspond to $IG(1,1)$ distributions). For the *Positive Hyperbolic* marginal, it is not obvious how to pick ν and α to give unit mean and variance. Instead, training data were generated with $\nu = 0.15$ and $\alpha = 1.4$ (which approximately give unit mean and variance). For the *Inverse Gamma* marginal, picking ν and α to give unit mean and variance gives $\nu = 2$, so $E\left[\sigma^8(t)\right]$ is infinite. Instead, the training data were generated with $\nu = 5$ and $\alpha = 2$ (corresponding to a mean and variance of $\frac{1}{2}$ and $\frac{1}{12}$).

Histograms are posteriors of 30,000 samples, taken after a burn-in period of 10,000. Trace plots for the unnormalised log-likelihood (ll) and λ parameter are also given to demonstrate the convergence of the chain.

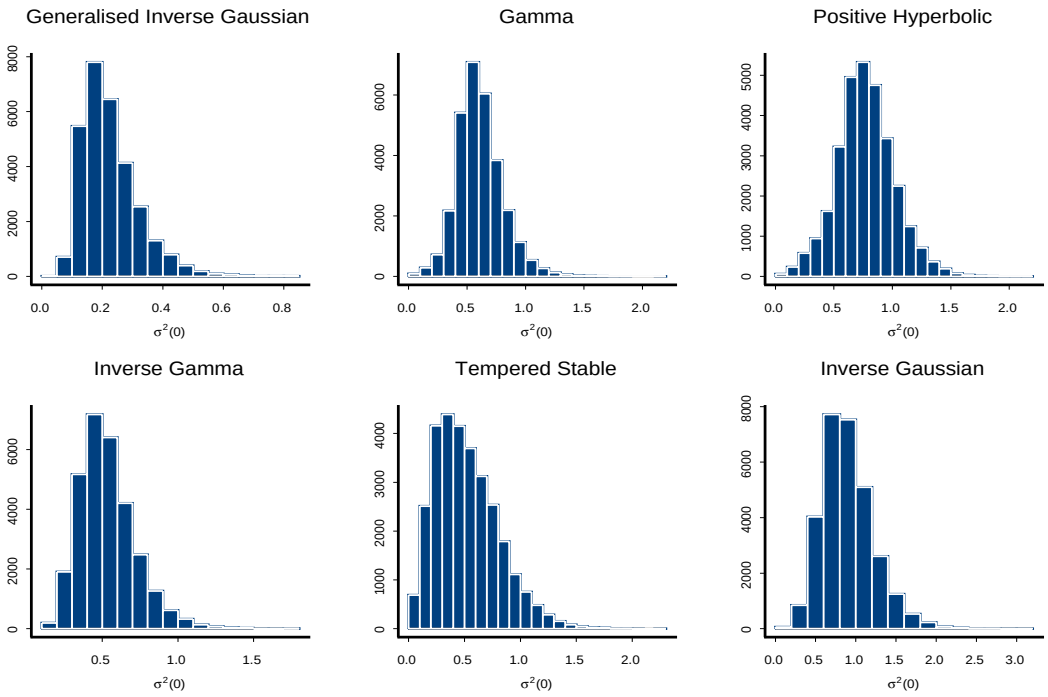


Figure 4.2: Histograms of the posterior distribution of $\sigma^2(0\Delta)$ for Test1. The true value is $\sigma^2(0\Delta) = 0.5$.

For the six marginals, the posterior supports $\sigma^2(0\Delta) = 0.5$. Estimating this parameter is more difficult than some of the other parameters, as only the first few data points are important in determining the distribution $\sigma^2(0)$.

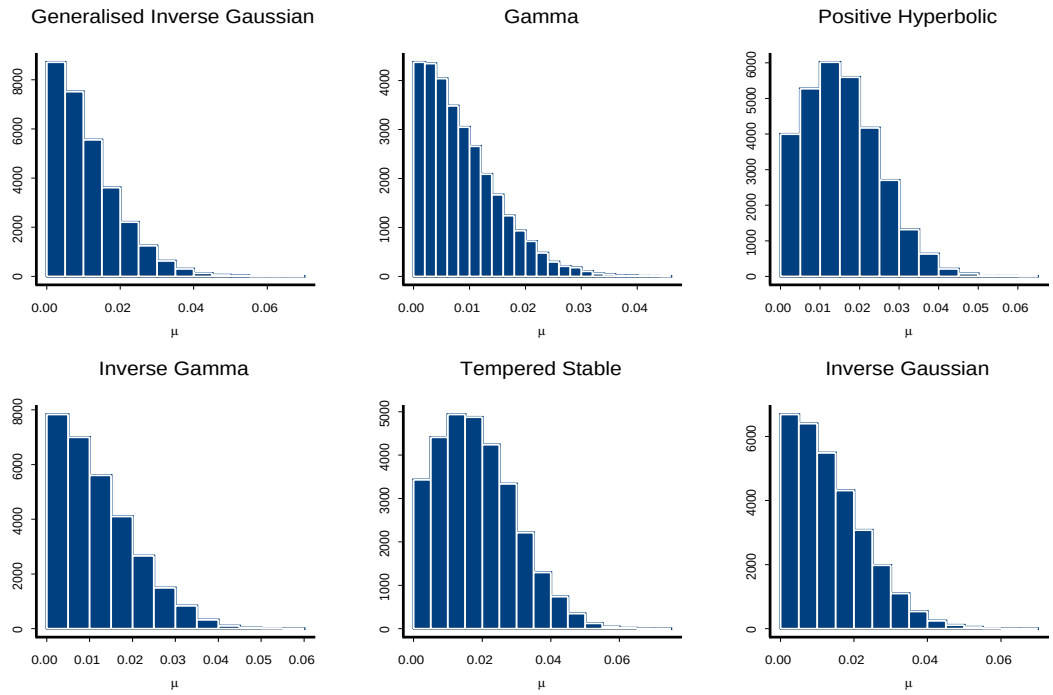


Figure 4.3: Histograms of the posterior distribution of μ for Test1. The true value is $\mu = 0.0001$.

All posteriors are able to identify that μ is small but are unable to pick out $\mu = 0.0001$ accurately. Histograms of the posterior of μ are not given in Griffin and Steel (2003), so all that can be concluded is that, for small μ, the model is not significantly altered as μ varies, provided it remains small. Although a prior which is more concentrated around $\mu = 0.0001$ could be used to improve identifiability, for a generic financial time series, it was felt that this would be unknown and could not justify using a much more informative prior than the $Ga(1, 250)$ used here.

	$\mu \times 10000$
GIG	84.3 (0.00, 403)
Ga	72.6 (0.01, 381)
RPH	148 (0.04, 491)
IGa	101 (0.00, 437)
TS	173 (0.03, 573)
IG	116 (0.02, 514)

Table 4.3: Posterior medians and 95% credible intervals (in brackets) for μ for Test1.

For all marginal distributions, Table 4.3 supports the true value of $\mu \times 10000 = 1$, though each marginal does not have a posterior for μ concentrated near the true value from which the data were generated.

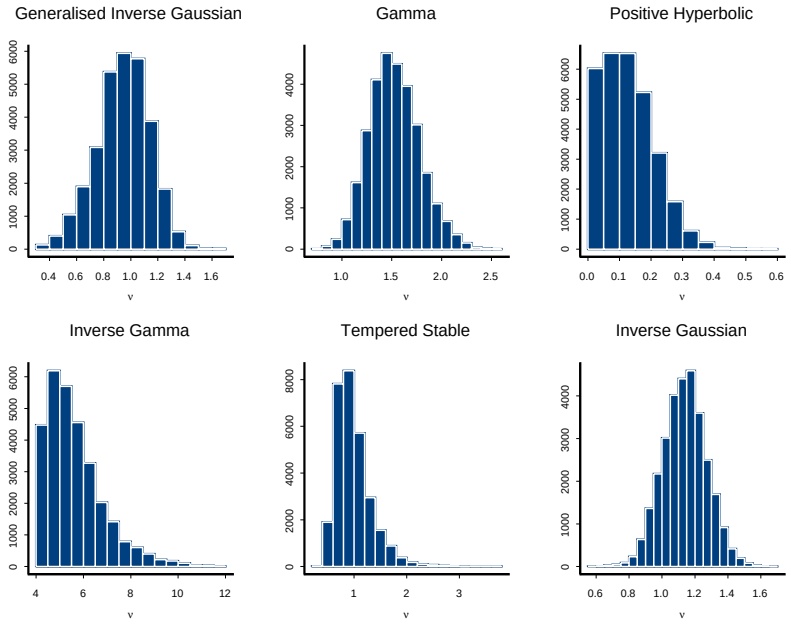


Figure 4.4: Histograms of the posterior distribution of ν for Test1.

The inference of the MCMC is much better for ν than for μ. These posteriors have a mean and mode at approximately the same ν value from which the training data were generated (see Table 4.4) and are similar to those given in Griffin and Steel (2003). The posterior for the *Gamma* marginal is not as concentrated near its true value as the other distributions for this simulation. For other seeds, the posterior of ν for the *Gamma* marginal is often closer to $\nu = 1$ than in Figure 4.4. This property can also be seen in Griffin and Steel (2003) and occurs because of the high correlation in the volatility and the size of the data set.

	ν
GIG	1
Ga	1
RPH	0.15
IGa	5
TS	1
IG	1

Table 4.4: True values of ν for Test1.

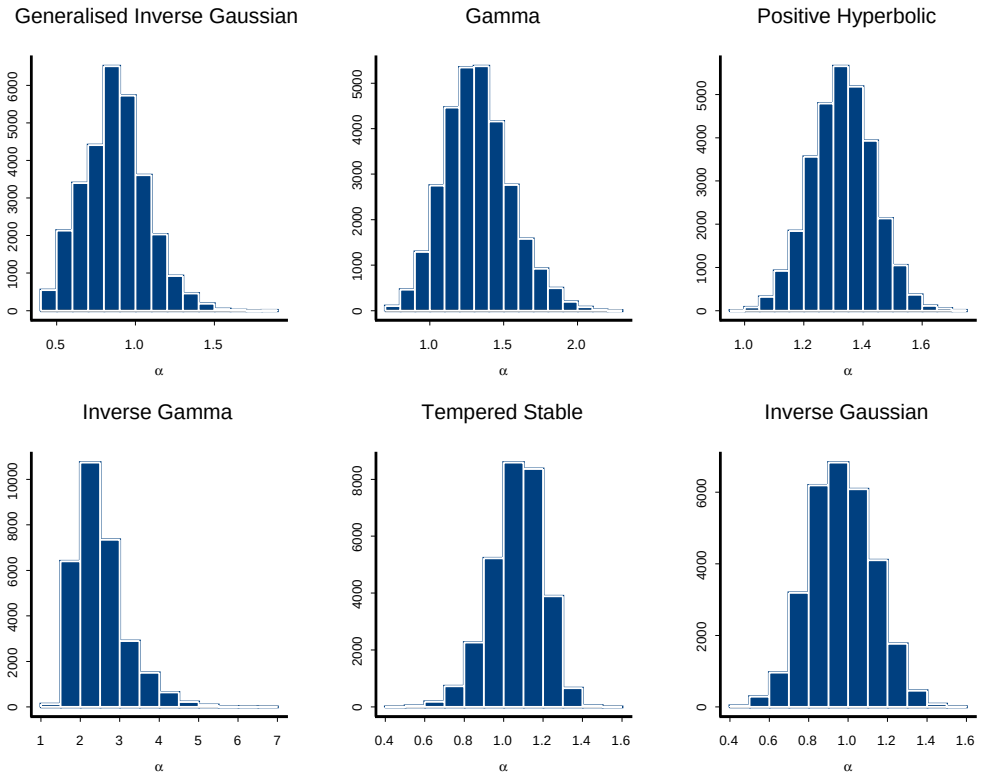


Figure 4.5: Histograms of the posterior distribution of α for Test1.

These posteriors are similar to those given in Griffin and Steel (2003) and all have a mean and mode at approximately the same α value from which the training data were generated (see Table 4.5). Again, for the *Gamma* marginal, the posterior for α is not as concentrated near its true value as the other marginals, which was justified for the ν parameter previously. For the *Gamma* marginal, the joint posterior of ν, α gives a volatility process with approximately unit mean (like the training data).

	α
GIG	1
Ga	1
RPH	1.4
IGa	2
TS	1
IG	1

Table 4.5: True values of α for Test1.

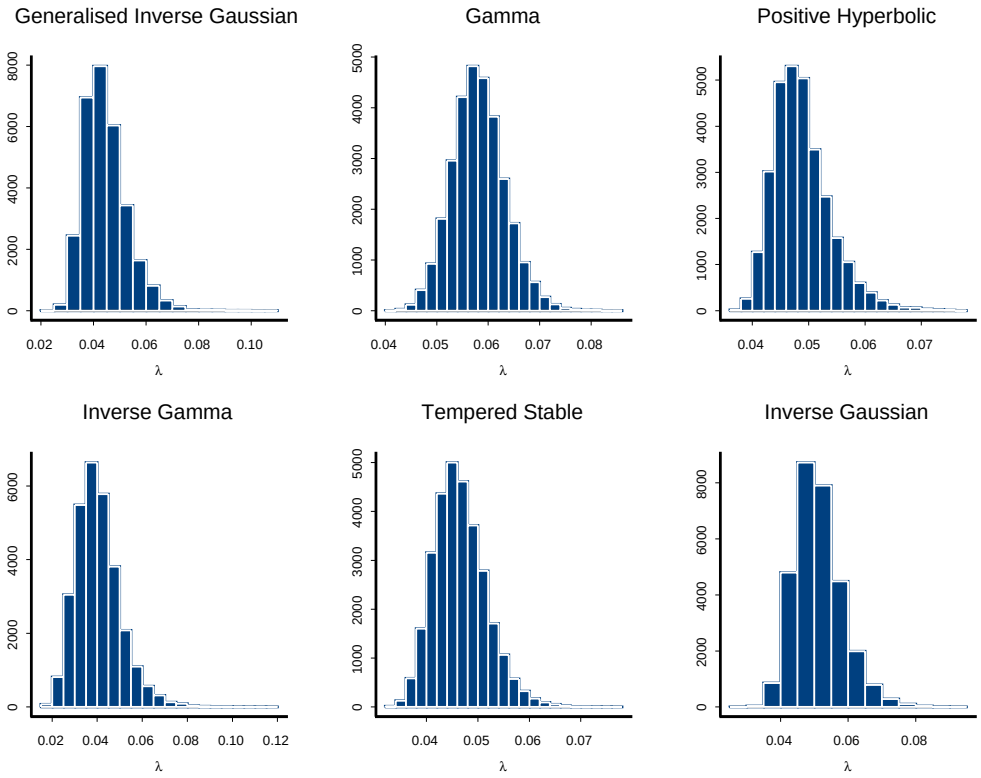


Figure 4.6: Histograms of the posterior distribution of λ for Test1. The true value is $\lambda = 0.05$.

These posteriors have a mean and mode at approximately the same λ value from which the training data were generated ($\lambda = 0.05$) and are similar to those given in Griffin and Steel (2003). The trace plot for this parameter is now given, as it is easy to compare this parameter across the different marginal distributions.

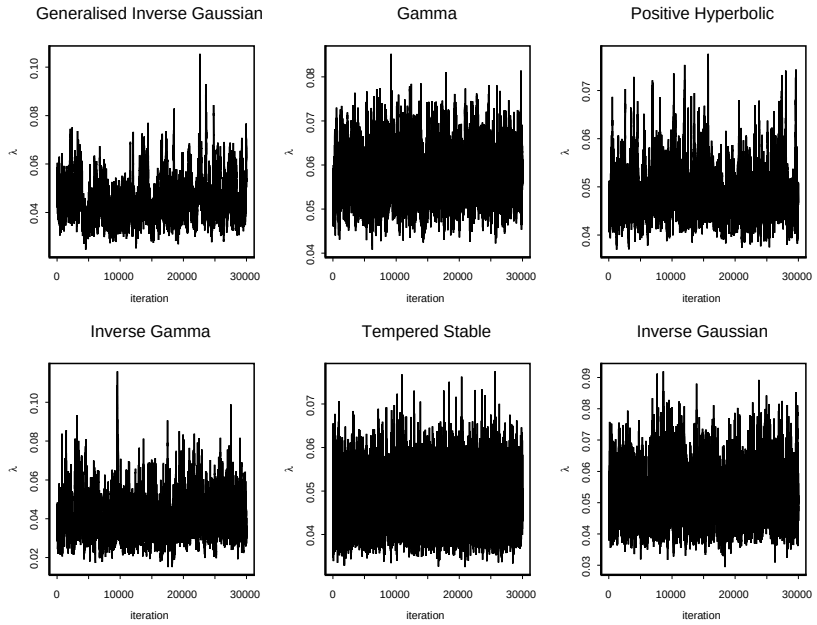


Figure 4.7: Trace plots of λ for Test1.

All simulations have converged and the posterior contains the value from which the training data were simulated, $\lambda = 0.05$.

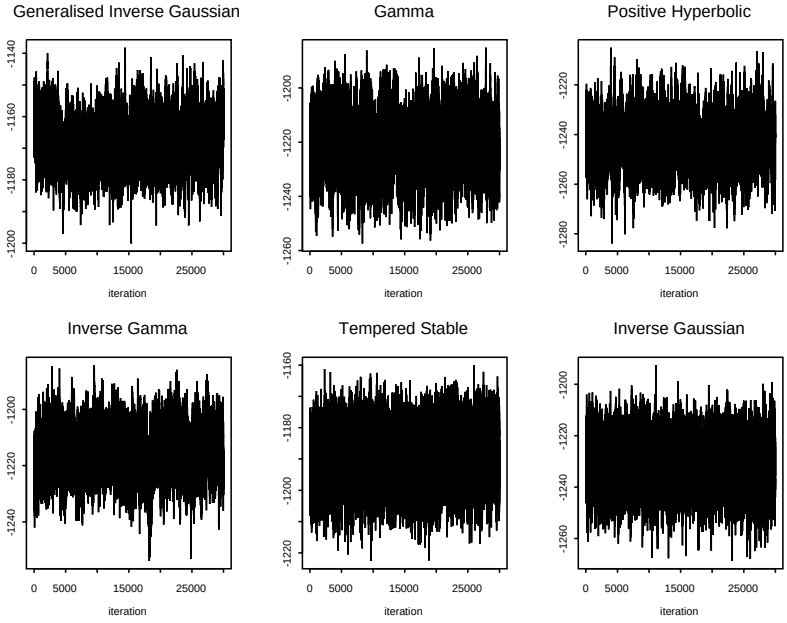


Figure 4.8: Plots of the unnormalised log-likelihood for Test1.

All simulations have converged and the likelihood is stable.

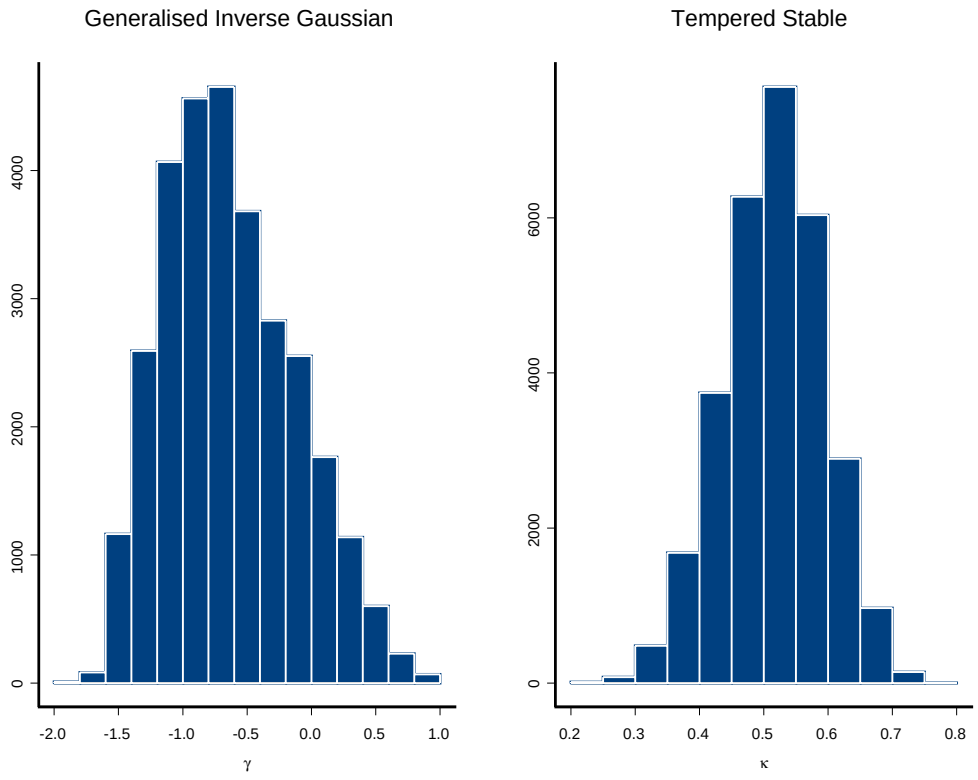


Figure 4.9: Histograms of the posterior of γ and κ for the *Generalised Inverse Gaussian* and *Tempered Stable* marginals for Test1.

For both marginal distributions, the posterior supports the values from which the training data were generated. Given that γ determines the thickness of the tails of the log returns and $\kappa \in (0, 1)$, it might be hoped that the variance in the posterior was less. For example, for the GIG, the posterior correctly identifies that the *Positive Hyperbolic* ($\gamma = 1$) distribution is highly unlikely, though it is not clear if the training data were generated from an *Inverse Gaussian* ($\gamma = -0.5$) or *Inverse Gamma* ($\gamma < 0$) marginal. The *Inverse Gamma* marginal looks unlikely, but this is not as strongly rejected as the *Positive Hyperbolic*. It is difficult to estimate all three parameters of the $GIG(\gamma, \nu, \alpha)$ (or $TS(\kappa, \nu, \alpha)$) distribution unless the data set is very large. It is particularly difficult to estimate the third parameter of the GIG when the data are from an *Inverse Gamma* distribution (so $\alpha = 0$), even when the data are independent IGa variates and MCMC is used to estimate the parameters of the GIG directly. This can be verified by generating variables from the $GIG(\gamma, \nu, \alpha = 0)$ distribution and estimating the three parameters using MCMC (this is not MCMC for the volatility process, improving α estimation). $10,000$ random variables

from the $GIG\left(-5, \sqrt{2}, 0\right)$ $(IGa\left(5,1\right))$ distribution were generated and $50,000$ iterations were taken after a burn-in of $10,000$. Histograms of the posterior of the γ and α parameters are given in Figure 4.10. The priors used were the same as for the volatility process MCMC.

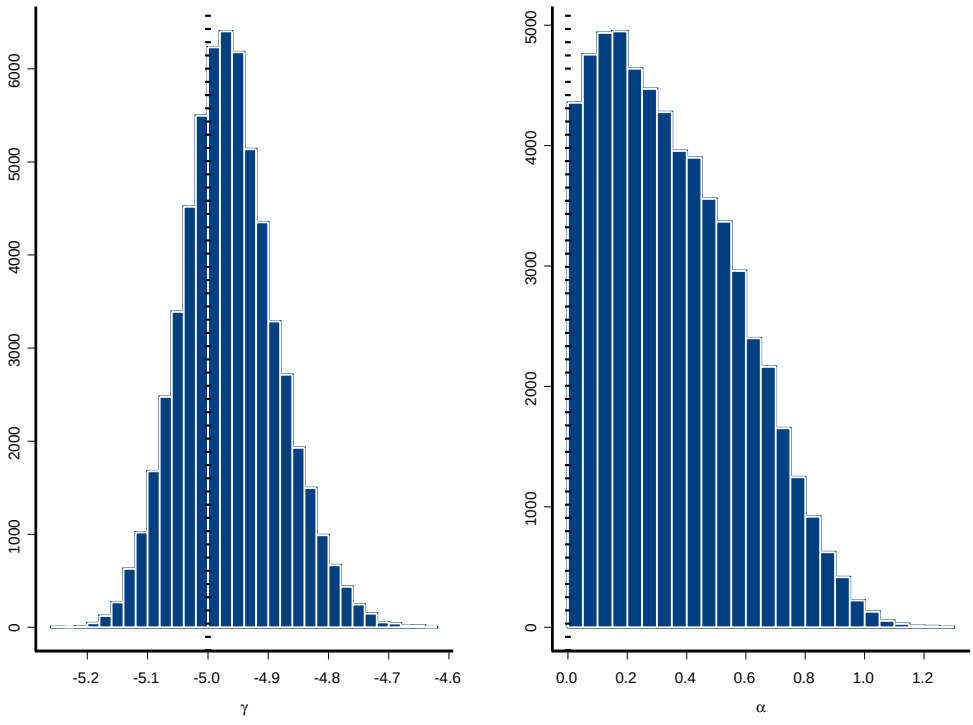


Figure 4.10: Histograms of the posterior of the first and third parameters of the *GIG* distribution on independent $IGa\left(5,1\right)$ variables.

Figure 4.10 shows that estimating the third parameter of the *GIG* distribution is difficult if it is small, even when the data set is $10,000$ independent $IGa\left(5,1\right)$ variates. The financial data sets we investigate are typically of size 1000 and are assumed to follow the BNS SV model, so observations are correlated. This makes it difficult to assess if the posterior distribution of α supports the *IGa* distribution. Instead we shall use the posterior distribution of γ to assess if the posterior of the *GIG* parameters supports the *IGa* marginal, by checking if the posterior only supports $\gamma < 0$.

Additionally, for inference using the BNS SV models, only the first two moments of the marginal distribution are used in the likelihood and estimating three parameters is difficult when only examining two moments of the data. This is not a major concern

here, as later the performance of the different models is tested empirically, where any over parameterised models (such as perhaps the $GIG(\gamma, \nu, \alpha)$ and $TS(\kappa, \nu, \alpha)$ marginals) will not predict "unseen" asset movement well.

To summarise the results of Test1, there are identifiability problems for the interest rate parameter, μ, for all marginal distributions. The MCMC is able to infer that μ is small but is not able to accurately pick out its value. Griffin and Steel (2003) do not give posterior histograms for μ, but their posterior median and 95% credible intervals for μ are influenced by the model that they are imposing. i.e. with/without leverage alters their credible intervals.

All the data contribute to estimating the most important parameters ν, α and λ (and additionally γ and κ for the *Generalised Inverse Gaussian* and *Tempered Stable* marginals) and these are well estimated by the MCMC (though the variance of the posterior of γ and κ are higher than those of ν and α).

The main focus is the performance of the MCMC on real data and so scatterplots of posterior samples are not given here. In Section 4.4.5, S&P 500 data are used and a discussion of the MCMC algorithm and possible reparameterisations are discussed. MCMC algorithms for the *Gamma* marginal can be found in Roberts et al. (2004) and Griffin and Steel (2003). The efficiency of our algorithm is compared with the preferred algorithm of Roberts et al. (2004) in Section 4.4.6.

4.4.2 Test2

Test2 is similar to Test1 but with leverage (see Section 3.4). The reason this was not included in Test1 is because the strength of the leverage induced is not only determined by the magnitude of the leverage parameter, ρ, but also by the variance of the marginal distribution of $\sigma^2(t)$. Here, histograms of the posterior of ρ are given for training data generated with the same parameters as Test1 (but with the additional leverage parameter). The training data were generated with $\rho = -3.0$, which is typical for financial data (see Griffin and Steel (2003) and Section 4.4.5). Only histograms of ρ are given. The other histograms were similar to Test1. 50,000 iterations were taken after a burn-in of 10,000.

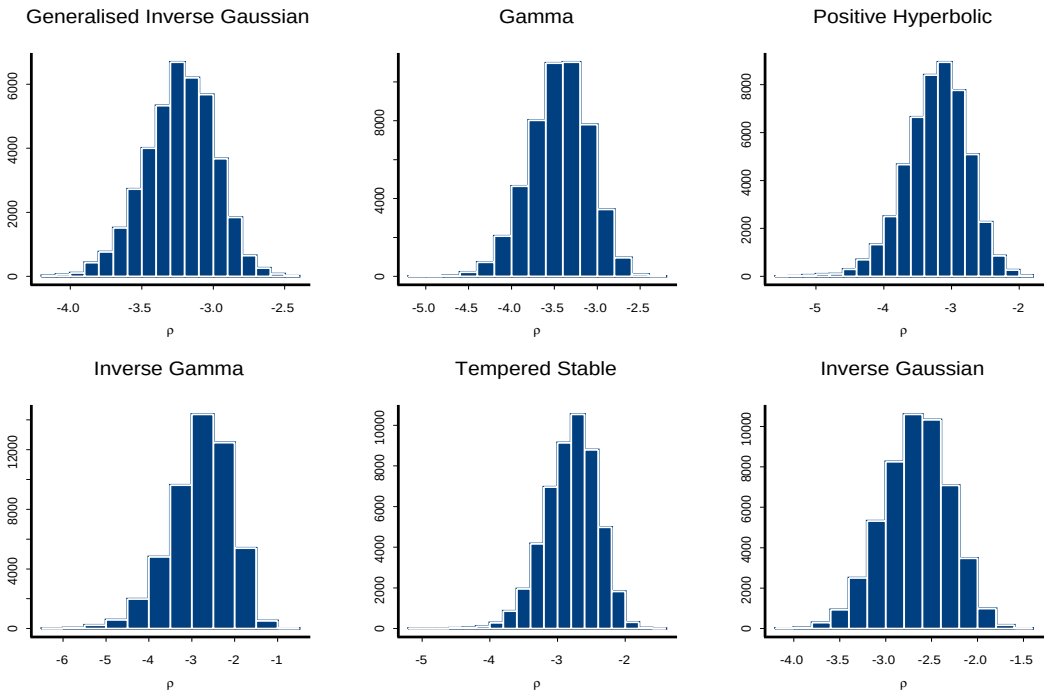


Figure 4.11: Histograms of the posterior distribution of ρ for Test2. The true value is $\rho = -3.0$.

The training data were generated with $\rho = -3.0$ and the posterior has mean and mode at approximately this value for each marginal.

4.4.3 Test3

Test3 was designed to test the correct implementation of the MCMC for the Hurst parameter introduced in Section 3.5. The same parameters were used as in Test2 to generate the training data, with the Hurst parameter set to $H = 0.6$. Histograms are given for the posterior of H (other histograms were similar to Test1 and Test2). 10,000 iterations were taken after a burn-in of 10,000.

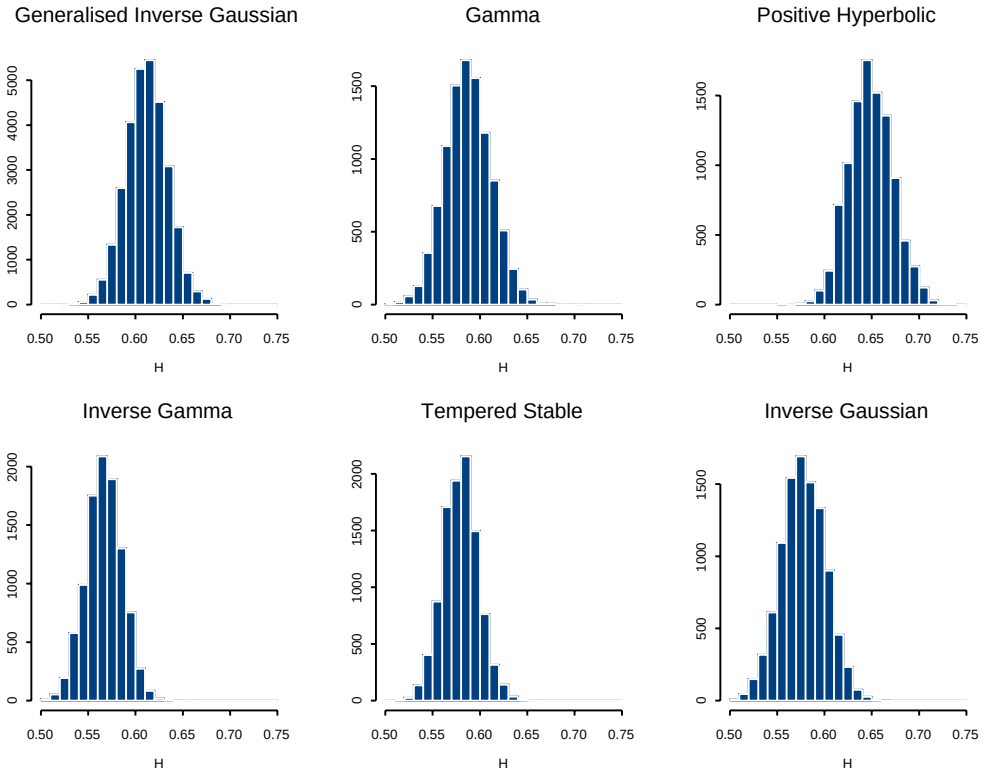


Figure 4.12: Histograms of the posterior distribution of H for Test3. The true value is $H = 0.6$.

The posterior contains $H = 0.6$ and only supports H close to 0.6, suggesting the algorithm is working correctly. The posterior does not support $H = 0.5$ or $H = 1.0$. Figure 4.13 is a histogram of the posterior of H for a *Gamma* marginal on training data with the same parameters as Figure 4.12 apart from $H = 0.5$. This is included so we can compare the posterior histograms of the Hurst parameter for real data with data for which we know $H = 0.5$, to try to assess if real data supports $H = 0.5$.

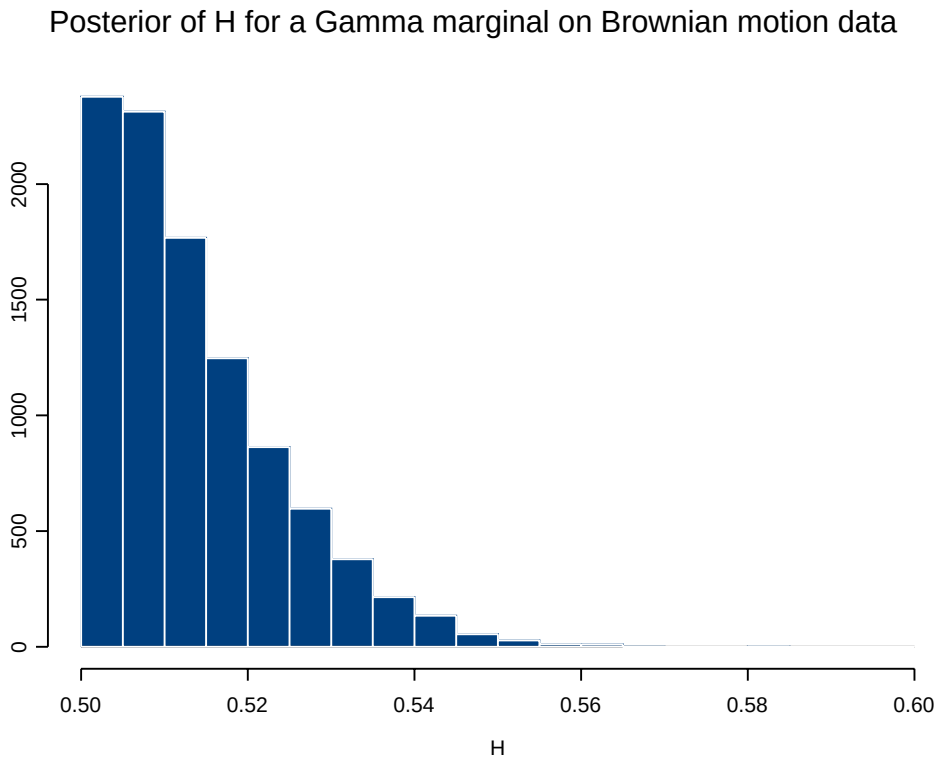


Figure 4.13: Histogram of the posterior distribution of H for Brownian motion data. The true value is $H = 0.5$.

Figure 4.13 shows that, for Brownian motion data, the posterior has a mode at $H = 0.5$ and is monotonic decreasing. There is very little posterior support for $H > 0.56$.

4.4.4 Test4

Test4 was designed to test the correct implementation of the MCMC for the stochastic λ process introduced in Section 3.3. The same parameters were used as Test3 to generate

the training data. There were three different λ_i values used for this test, namely

$$\begin{aligned}
\text{TestA} \quad \lambda_i &= 0.05 \qquad i \leq 1000 \\
\text{TestB} \quad \lambda_i &= \begin{cases} 0.01 & i \leq 500 \\ 0.5 & 500 < i \leq 1000 \end{cases} \\
\text{TestC} \quad \lambda_i &= \begin{cases} 0.05 & i \leq 1000 \\ 0.3 & 1000 < i \leq 2000 \\ 0.05 & 2000 < i \leq 3000 \\ 0.3 & 3000 < i \leq 4000 \end{cases}
\end{aligned}$$

40,000 iterations were taken after a burn-in of 10,000. The prior for r was $Ga\,(1, 0.5)$. This has higher variance than the $Ga\,(1, 1)$ prior suggested in Section 3.3. In Section 5.3.3, where MCMC is performed on real data, simulations are run for both priors for r. TestA and TestB are data sets of size 1000 with no jumps and one jump respectively. TestC has three jumps in the stochastic λ process and is of size 4000, as it is difficult to make inference about the jump times and sizes when the jumps are close together.

For TestA, histograms are given for the posterior of N_2 (number of jumps in the stochastic λ process) and λ_0 when $N_2 = 0$ (other histograms were similar to Test1, Test2 and Test3).

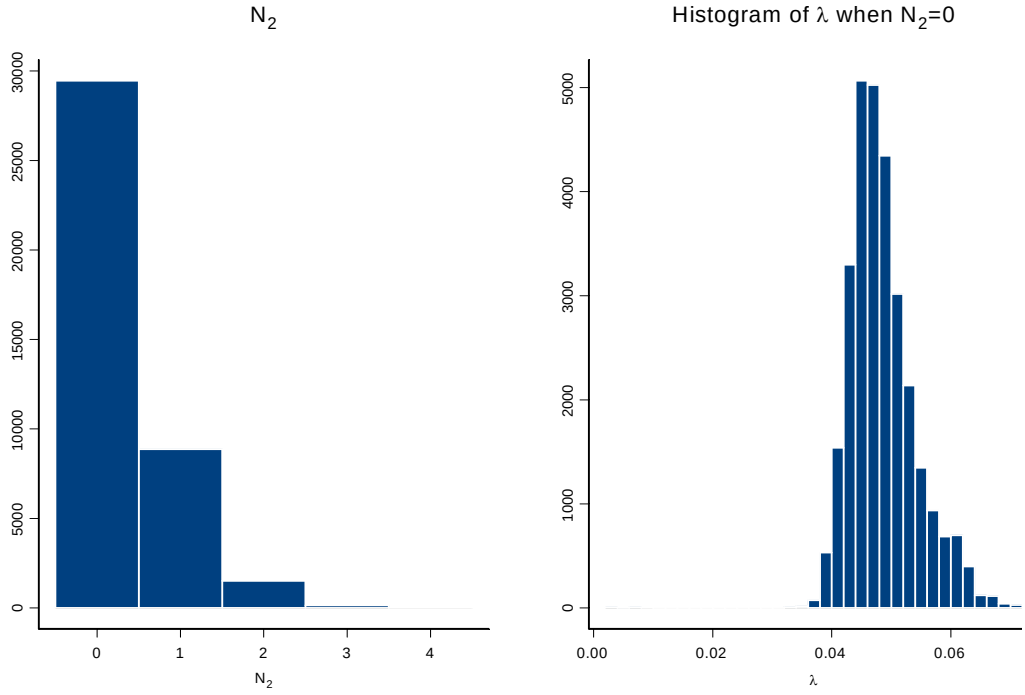


Figure 4.14: Histograms of N_2 and λ, when $N_2 = 0$, for Test4 TestA.

The model identifies that it is most likely there are no jumps and when $N_2 = 0$, estimates λ to be approximately 0.05.

n_2	Prior $P(N_2 = n_2)$	Posterior $P(N_2 = n_2)$
0	0.333	0.736
1	0.222	0.222
2	0.148	0.038
3	0.099	0.004

Table 4.6: Prior and posterior probabilities for N_2 for Test4 TestA.

Table 4.6 shows the posterior for N_2 has more support at $N_2 = 0$ than the prior.

For TestB, histograms are given for the posterior of N_2 (number of jumps in the λ process), the positioning of the jumps when $N_2 = 1$, as well as histograms of the posterior of the values of λ in the first and second segment when $N_2 = 1$ (other histograms were similar to Test1, Test2 and Test3).

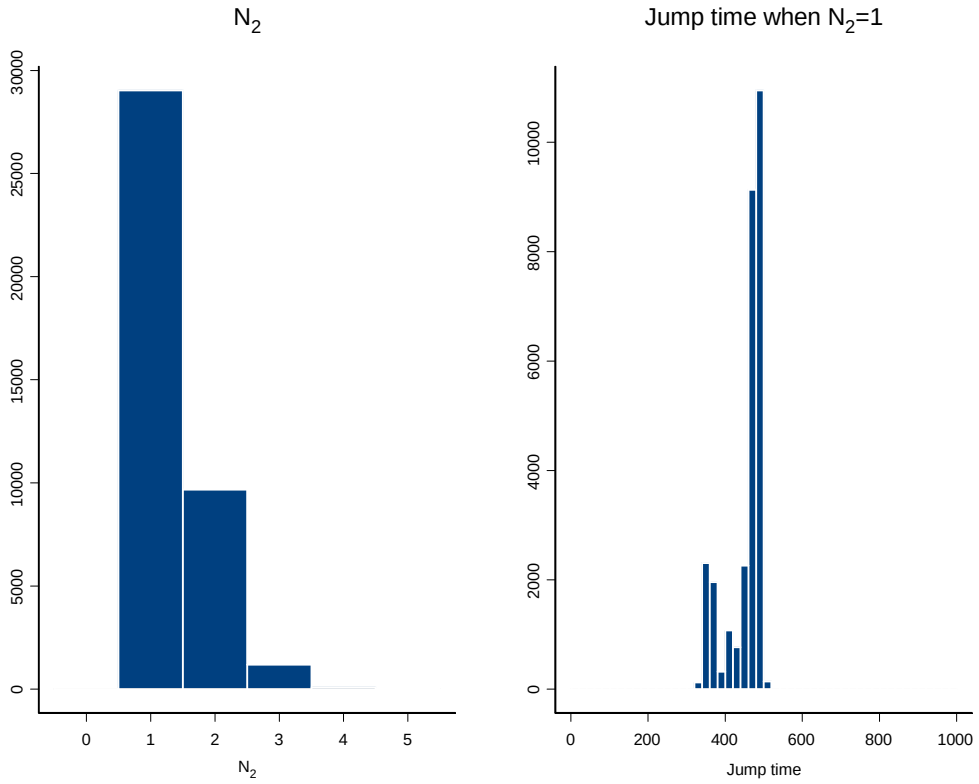


Figure 4.15: Histograms of N_2 and jump time, when $N_2 = 1$, for Test4 TestB.

The model identifies that one jump is most plausible and locates this jump in the correct place, $T = 500$.

n_2	Prior $P(N_2 = n_2)$	Posterior $P(N_2 = n_2)$
0	0.333	0.000
1	0.222	0.675
2	0.148	0.272
3	0.099	0.007

Table 4.7: Prior and posterior probabilities for N_2 for Test4 TestB.

Table 4.7 shows that the posterior has more support at $N_2 = 1$ and $N_2 = 2$ than the prior. The posterior probabilities for 0 jumps is 0.000 to three decimal places.

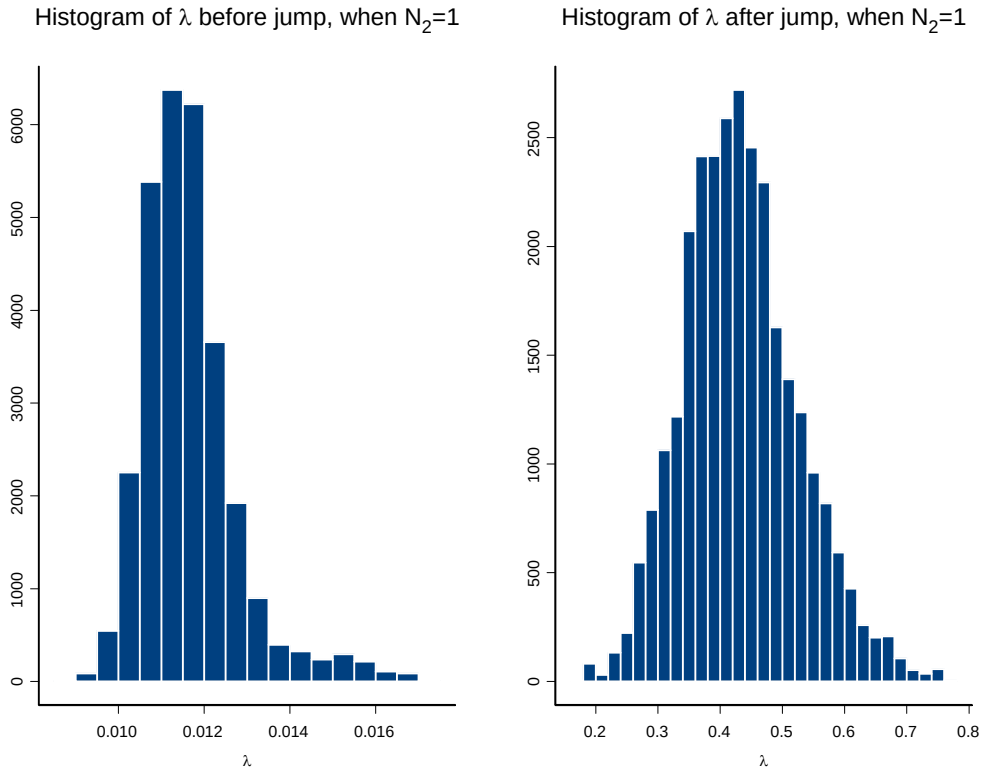


Figure 4.16: Histograms of λ before and after the jump, when $N_2 = 1$, for Test4 TestB.

The MCMC picks out the correct values of λ to the left and right of the jump.

For TestC, histograms are given for the posterior of N_2 (number of jumps in the λ process), the positioning of the jumps when $N_2 = 3$, as well as histograms of the posterior

of λ values when $N_2 = 3$ (other histograms were similar to Test1, Test2 and Test3).

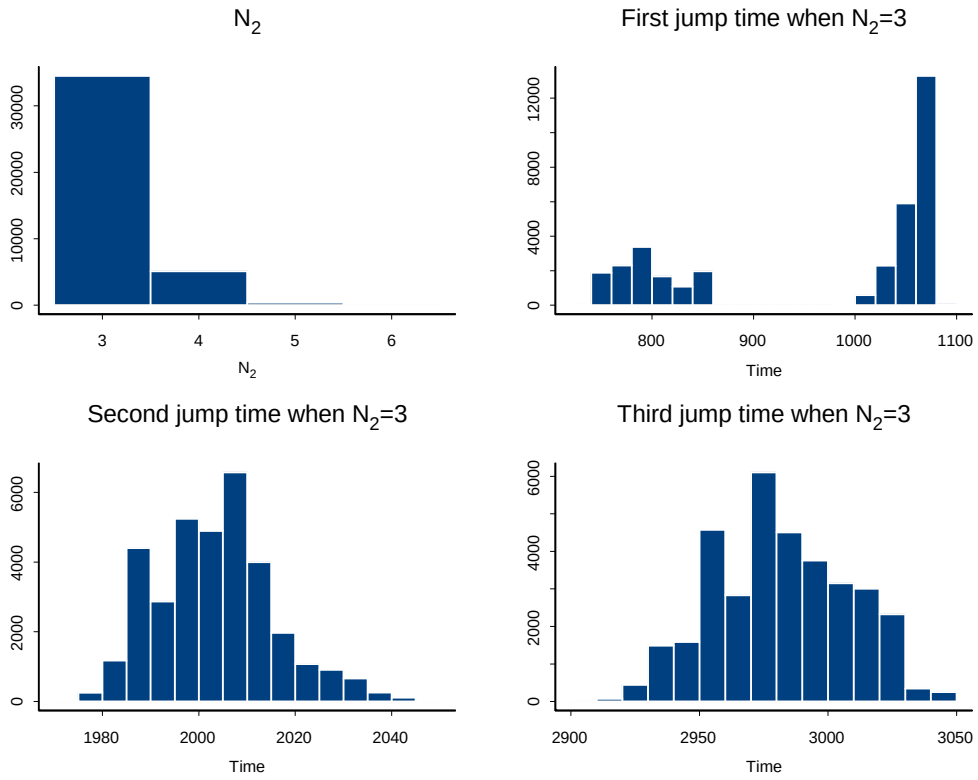


Figure 4.17: Histograms of N_2 and jump time, when $N_2 = 3$, for Test4 TestC.

The model identifies that three jumps is most plausible and locates these jumps accurately.

n_2	Prior $P(N_2 = n_2)$	Posterior $P(N_2 = n_2)$
0	0.333	0.000
1	0.222	0.000
2	0.148	0.000
3	0.099	0.862
4	0.066	0.128
5	0.044	0.010

Table 4.8: Prior and posterior probabilities for N_2 for Test4 TestC.

Table 4.8 illustrates the strong posterior support for $N_2 = 3$. The following graphs are also conditional on $N_2 = 3$.

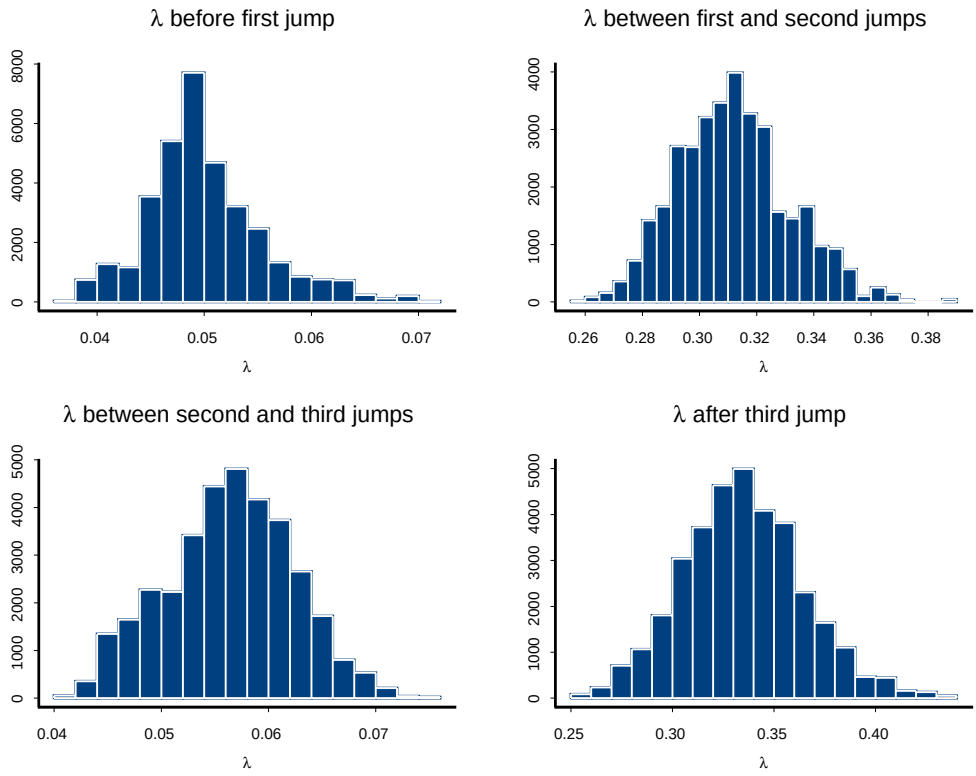


Figure 4.18: Histograms of λ at the beginning, end and in between jumps, when $N_2 = 3$, for the stochastic λ process, for Test4 TestC.

The MCMC picks out the correct values of λ in each segment.

4.4.5 Test5

The previous tests were on training data. Test5 is a test on S&P 500 data from 2^{nd} January 1980 to 30^{th} December 1987. All marginals are run on this data and the mean/variance of the different marginals are compared. Results for the *Gamma* marginal can also be compared with those of Griffin and Steel (2003) and Nicolato and Venardos (2003) (which fits the BNS SV models to S&P 500 data over a different time period).

Before running any simulations, it was noted that the results from these two sources are quite different. Nicolato and Venardos (2003) report a λ value of approximately 2.4958 and 1.6787 for the *Gamma* and *Inverse Gaussian* marginals respectively (this is a concern, as the marginal distribution of $\sigma^2(t)$ and λ value should be unrelated (see Appendix D)), while Griffin and Steel (2003) reports $\lambda \in (0.004, 0.05)$ for the *Gamma* marginal. The

Griffin and Steel (2003) estimate for λ is more in keeping with the general opinion that the correlation in volatility decays slowly for financial time series. Nicolato and Venardos (2003) also do not give any credible intervals for their estimates and this makes it difficult to check if their results are similar to ours.

Initially, the MCMC was run for all the model generalisations that have been tested so far but without the stochastic λ process, to make it easier to compare results with other sources (i.e. the model was run with leverage and the MVN approximation to fBm). Histograms are given for the posterior of H in Figure 4.19. 10,000 iterations were taken after a burn-in of 10,000. The data were rescaled by multiplying y_i by 88.8 so that $E\left[y_i^2\right] \approx 1$. The prior for κ was $Beta\,(1, 15)$.

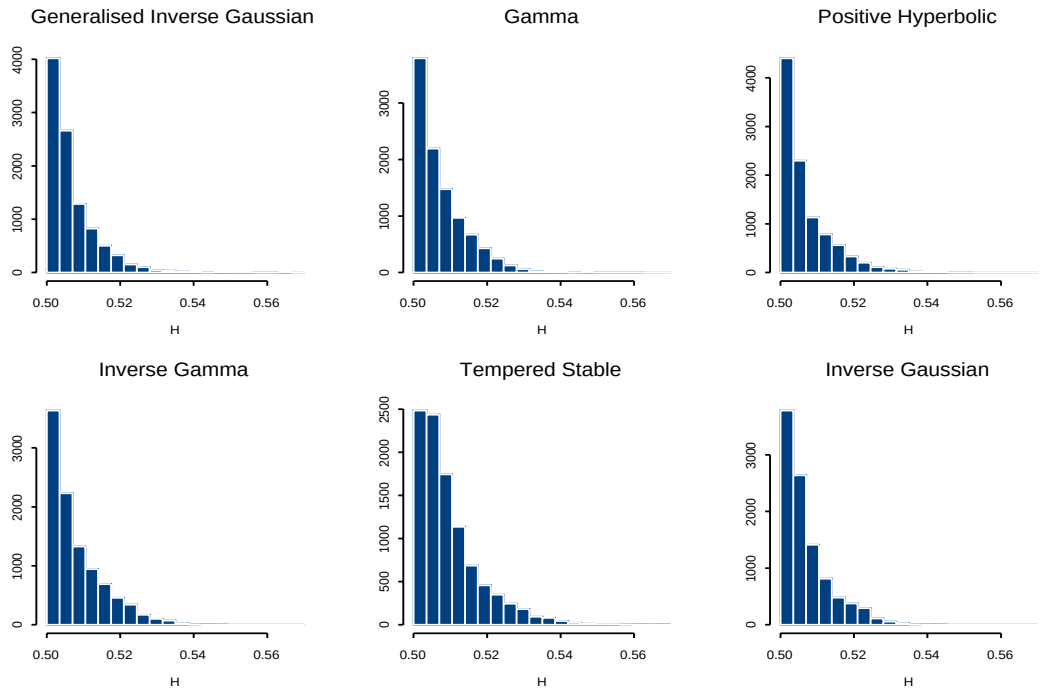


Figure 4.19: Histograms of the posterior distribution of H for Test5.

	H
GIG	0.505 (0.500, 0.524)
Ga	0.505 (0.500, 0.524)
RPH	0.505 (0.500, 0.525)
IGa	0.505 (0.500, 0.527)
TS	0.507 (0.500, 0.532)
IG	0.505 (0.500, 0.524)

Table 4.9: Posterior medians and 95% credible intervals (in brackets) for the Hurst parameter for Test5.

For each marginal, the posterior has most support at $H = 0.5$, and is similar to Figure 4.13, suggesting $H \approx 0.50$ and fBm is not required to model the S&P 500 data. The MCMC was rerun without the Hurst parameter, so our results can be directly compared with those of Nicolato and Venardos (2003) and Griffin and Steel (2003). A sample of 50,000 iterations were taken after a burn-in of 10,000. Results are summarised in Tables 4.10 and 4.11.

	ρ	λ	ll
GIG	-3.88 (-6.99, -1.81)	0.047 (0.027, 0.082)	-2349 (-2376, -2322)
Ga	-5.31 (-11.17, -1.95)	0.040 (0.023, 0.080)	-2377 (-2403, -2349)
RPH	-2.37 (-5.34, -1.58)	0.039 (0.022, 0.079)	-2377 (-2403, -2339)
IGa	-1.89 (-3.91, -0.62)	0.052 (0.030, 0.091)	-2352 (-2383, -2323)
TS	-3.49 (-7.53, -2.41)	0.051 (0.027, 0.095)	-2361 (-2390, -2330)
IG	-3.11 (-4.69, -0.68)	0.045 (0.027, 0.079)	-2365 (-2395, -2336)

Table 4.10: Posterior medians and 95% credible intervals (in brackets) for parameters which do not directly influence the marginal for Test5.

For each marginal, the posterior credible interval for ρ does not contain 0, suggesting that, for this data set, a model with leverage is required. The summaries for the different marginals for the leverage parameter are less similar to each other than the posterior for other parameters; this is discussed in Sections 3.4 and Appendix A.9. For the *Gamma* and *Inverse Gaussian* marginals, the credible intervals for the leverage parameters are similar to those estimated in Nicolato and Venardos (2003) and Griffin and Steel (2003).

For λ, the summaries suggest the parameter is small and hence there is a slow decay in correlation of the volatility. The posterior of λ is similar for each marginal and these are similar to the values reported in Griffin and Steel (2003). Our posteriors for λ do not agree with the large values estimated by Nicolato and Venardos (2003) and, even though they fit the model to observed market option prices (and the data set is over different time periods),

the results should be similar, as they both characterise the S&P 500 movement (assuming it follows a stationary process). The value of λ should be similar for each marginal and this is the case for all of our marginals. Even though the marginal process does not have the same mean and variance in the prior, the posterior mean and variance are similar for each marginal and this is not the case for the results of Nicolato and Venardos (2003). Our λ values agree with those of Griffin and Steel (2003) and suggests the results of Nicolato and Venardos (2003) need further examination for verification.

Finally, the log-likelihood (ll) suggests the *Inverse Gamma* and *Generalised Inverse Gaussian* marginals fit the data best. The log-likelihood for the *Inverse Gamma* distribution is larger than that of the three parameter *Tempered Stable* marginal. The worst performing is the *Gamma* marginal. The aim of Test5 is not to compare the performance of different models; this is left until Chapter 5, where a more thorough comparison of the different models' empirical performance is given.

	$E\left[\sigma^2(t)\,\middle\vert\,\nu,\alpha,\kappa\right]$	$V\left[\sigma^2(t)\,\middle\vert\,\nu,\alpha,\kappa\right]$	γ or κ
$GIG(\gamma,\nu,\alpha)$	0.94 (0.71, 1.28)	0.34 (0.15, 1.16)	-0.23 (-3.27, 2.19)
$Ga(\nu,\alpha)$	1.07 (0.74, 1.79)	0.37 (0.19, 1.18)	
$RPH(\nu,\alpha)$	1.20 (0.89, 1.58)	0.46 (0.21, 0.93)	
$IGa(\nu,\alpha)$	0.83 (0.71, 0.99)	0.27 (0.17, 0.42)	
$TS(\kappa,\nu,\alpha)$	0.96 (0.75, 1.27)	0.36 (0.18, 0.76)	0.50 (0.33, 0.67)
$IG(\nu,\alpha)$	1.04 (0.81, 1.36)	0.31 (0.13, 0.88)	

Table 4.11: Posterior medians and 95% credible intervals (in brackets) for parameters which do directly influence the marginal for Test5.

For the *Generalised Inverse Gaussian* distribution, the posterior 95% credible intervals for γ are unable to reject any of the other marginal distributions, as the credible interval supports γ positive (*Gamma*), γ negative (*Inverse Gamma*), $\gamma = -0.5$ (*Inverse Gaussian*) and $\gamma = 1$ (*Positive Hyperbolic*). We shall not overly concern ourselves with this here and leave model selection until Chapter 5. For the *Tempered Stable* marginal, the credible interval for κ contains 0.5 and suggests the *Inverse Gaussian* distribution is a reasonable marginal to use for volatility for the S&P 500 data set. This is reflected in the large log-likelihood for the *Inverse Gaussian* distribution in Table 4.10. The informative $Beta(1, 15)$ prior keeps κ small, suggesting the log returns require a fat-tailed distribution; it is not surprising that the *Inverse Gamma* distribution has a large log-likelihood for this reason.

The data were rescaled so $E\left[y_i^2\right] = 1$ and this can be seen in the median and credible interval for $E\left[\sigma^2(t)\,\middle\vert\,\nu,\alpha,\kappa\right]$, which all contain 1, except for the $IGa(\nu,\alpha)$ marginal which contains 0.99. Posterior summaries for $V\left[\sigma^2(t)\,\middle\vert\,\nu,\alpha,\kappa\right]$ are similar for each marginal and are in keeping with the results of Griffin and Steel (2003). This is not

the case in Nicolato and Venardos (2003), where for the *Gamma* and *Inverse Gaussian* marginals, $V\left[\sigma^2(t)\,|\nu,\alpha,\kappa\right]$ is 114 and 18,000 respectively. It might be expected that $V\left[\sigma^2(t)\,|\nu,\alpha,\kappa\right] \approx V\left[y_i^2\right]$ (which is 88.5 for this data set) and this is not the case. This is because the model fits a smoothed process to the observed data and the data set has a small number of extreme observations. Removing the four largest y_i^2 from the rescaled data alters $E\left[y_i^2\right]$ from 1 to 0.74 and $V\left[y_i^2\right]$ from 88.5 to 2.34. Such extreme observations are often explained through the leverage term, rather than the volatility process (i.e. $V\left[\sigma^2(t)\,|\nu,\alpha,\kappa\right]$ need not necessarily be similar to $V\left[y_i^2\right]$). Griffin and Steel (2003) also find the leverage term explains the large (and negative) log returns, particularly when they use a "quadratic leverage term".

Note that the mean and variance of each marginal distribution is similar, unlike those of Nicolato and Venardos (2003).

4.4.6 Test6

The correlation structure between parameters in the posterior can be complex (particularly for observed data) and this makes efficient MCMC implementation difficult. A discussion of the correlation of the parameters for the *Gamma* marginal is given in this test and the performances of different MCMC algorithms are compared. We only consider the simple forms of the model (no fBm approximation or stochastic λ process) so the algorithm can be compared with that of Roberts et al. (2004).

To increase the acceptance rate and/or size of the MCMC moves, it is preferable that the posterior of the parameters have low correlation (see for example Robert and Casella (2002) for discussion). When a *Gamma* distribution is used as the marginal for the volatility, there are five non-latent parameters (μ, ρ, ν, α and λ) and many latent parameters ($\sigma^2(0)$, the Poisson point process, A, and the uniforms, R). Griffin and Steel (2003) consider proposals which are based on the knowledge of the high correlation between λ and n, where

$$n = \sum_{i=1}^{T} n_i.$$

This correlation occurs because as λ increases, the value at which the Poisson points is truncated increases (see equation (4.22)). Although it is intuitive that λ and n are likely to be positively correlated, the correlation between other parameters (such as ν and α) and n is less clear. A table of correlations between the different parameters is now given for the *Gamma* marginal run on S&P 500 data, where sum_a is

$$sum\_a = \sum_{i=1}^{T} \sum_{j=1}^{n_i} a_{i,j}.$$

	ρ	ν	α	λ	n	sum_a
μ	0.171	-0.091	-0.020	-0.087	-0.092	-0.096
ρ		-0.518	-0.580	-0.522	-0.713	-0.665
ν			0.901	0.114	0.632	0.595
α				0.072	0.596	0.540
λ					0.802	0.758
n						0.964

Table 4.12: Correlation in the posterior of the parameters of a Gamma marginal on S&P 500 data.

The two most notable values from Table 4.12 are that of $corr\,(\nu, \alpha)$ and $corr\,(\lambda, n)$, which have a high correlation in the posterior. It was expected that $corr\,(\lambda, n)$ would be large and the proposal described in Appendix C.5 reflects this. Additionally, as it is known that ν and α specify the value, a_c, at which the Poisson points are truncated, the proposals for an update in either ν or α consider adding or removing Poisson points and uniforms, to keep constant accuracy of the approximation of the infinite sum in equation (4.21). As the likelihood is *Normal*, two moments of the volatility process are fitted and, for the *Gamma* distribution, the first moment is ν/α and this suggests why there may be high correlation between these two parameters in the posterior.

There is strong correlation in other parameters which might not be expected, for example (ρ, α), (ρ, n) and (ν, n). To try to combat the high correlation between ν and α, Griffin and Steel (2003) reparameterise, using the parameters $(\nu/\alpha, \alpha)$. Figure 4.20 is a scatter plot, demonstrating the correlation in the posterior of the parameters (ν, α) and $(\nu/\alpha, \alpha)$.

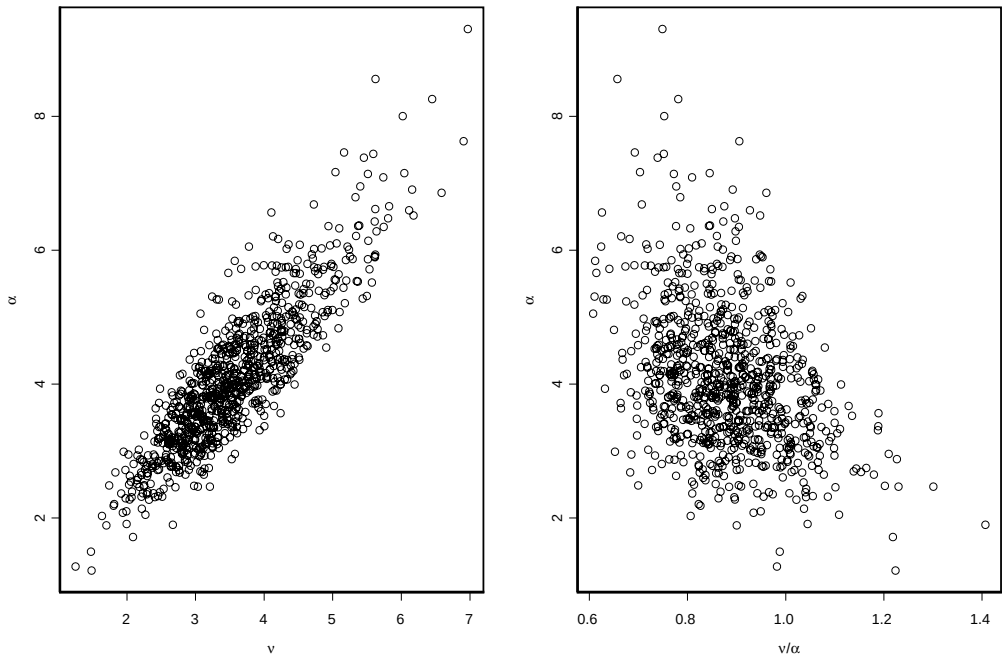


Figure 4.20: Scatterplots demonstrating the correlation in the posterior of (ν, α) and $(\nu/\alpha, \alpha)$ for a $Ga(\nu, \alpha)$ marginal distribution on S&P 500 data.

The correlation is higher for the original parameterisation and without considering joint moves on (ν, α), the MCMC algorithm might be inefficient. For the $(\nu/\alpha, \alpha)$ parameterisation, there is less correlation, so the reparameterisation might improve the convergence rate and/or correlation in the chain. A table of correlations between the new parameters is now given (with the old parameterisation values in **bold**).

	ρ	ν/α	α	λ	n	sum_a
μ	0.171	-0.157 **-0.091**	-0.020	-0.087	-0.092	-0.096
ρ		0.180 **-0.518**	-0.580	-0.522	-0.713	-0.665
ν/α			-0.304 **0.901**	0.089 **0.114**	0.018 **0.632**	0.057 **0.595**
α				0.072	0.596	0.540
λ					0.802	0.758
n						0.964

Table 4.13: Correlation in the posterior of the reparameterisation for a $Ga(\nu, \alpha)$ marginal distribution on S&P 500 data.

Table 4.13 suggests that the reparameterisation of Griffin and Steel (2003) could improve the efficiency of the chain, assuming that there is not much correlation structure between the parameters in the table and the latent parameters. ACF plots of α (which is common to both parameterisations) are now given for the two parameterisations.

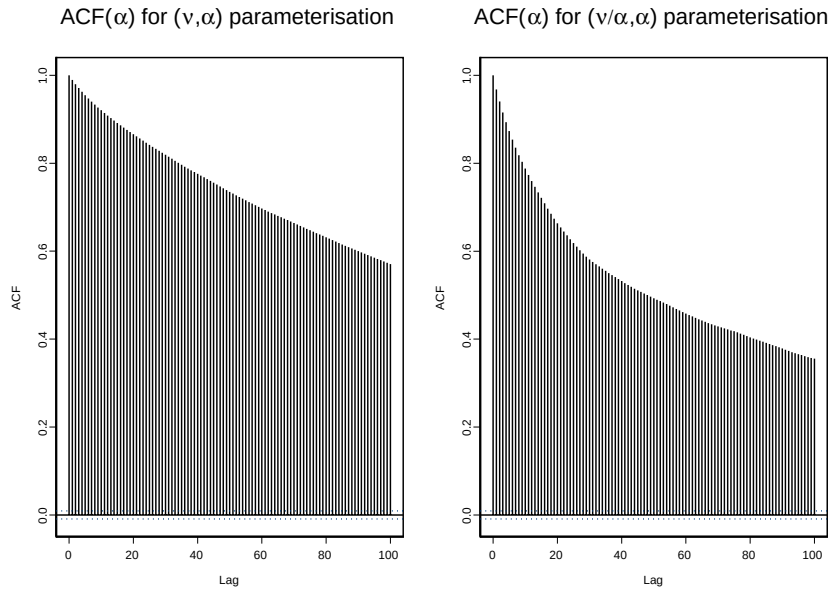


Figure 4.21: ACF(α) plots for the (ν, α) and $(\nu/\alpha, \alpha)$ parameterisations for a $Ga(\nu, \alpha)$ marginal distribution on S&P 500 data.

Figure 4.21 shows that the reparameterisation has improved the efficiency of the chain, as the ACF of the common parameter decays more rapidly for the reparameterisation. However, there could be another parameter which converges/mixes more slowly than the ν and α parameters and so plots of the log-likelihood of the two different parameterisations are now examined. This compares the efficiency of the two MCMC algorithms, without focusing on one individual parameter. A parameterisation which decreases the correlation in the log-likelihood will provide a less correlated sample from the predictive density, which we sample from in Chapter 5.

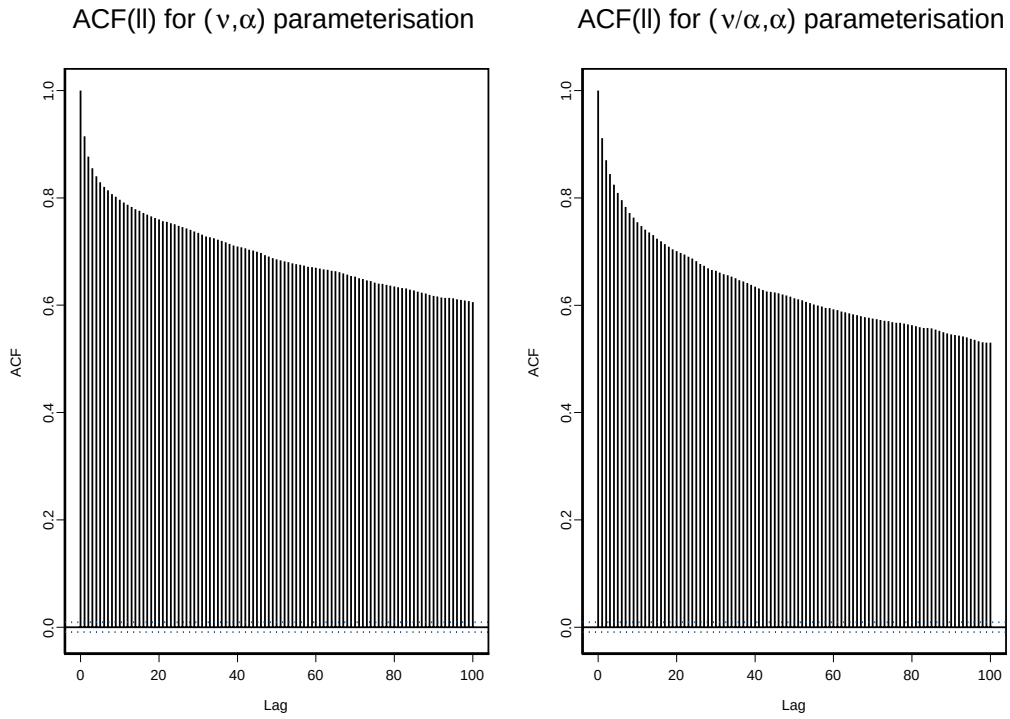


Figure 4.22: ACF(log-likelihood) plots for the (ν, α) and $(\nu/\alpha, \alpha)$ parameterisations for a $Ga(\nu, \alpha)$ marginal distribution on S&P 500 data.

The ACF plots are similar for both parameterisations. At lag 100 the (ν, α) parameterisation has correlation 0.61 and the $(\nu/\alpha, \alpha)$ parameterisation has correlation 0.53. The improvement is not as noticeable as in Figure 4.21 because there are other parameters converging/mixing more slowly (than ν and α), limiting the effectiveness of the reparameterisation. For example, $ACF(\lambda)$ tends to decay slowly, even though our algorithm already uses knowledge of the correlation between λ and the number of Poisson points and uniforms. Similar reparameterisations (where one parameter is the mean of the marginal distribution) are not trivial for all of the marginal distributions and as the performance

increase is slight, these reparameterisations are not implemented in later sections. As our focus is prediction (using predictive densities and option pricing results), these MCMC algorithms perform similarly.

Roberts et al. (2004) focus on the performance of different MCMC algorithms and their preferred algorithm and ours is now compared. They implement two algorithms, which they denote the CA (centred algorithm) and NCA (non-centred algorithm). They conclude that a hybrid of the two algorithms performs best (our algorithm is a non-centred algorithm). This is referred to as the Hybrid algorithm. As our interest is for the MCMC acting on real data, timings of our algorithm and the Hybrid algorithm[2] for $50,000$ iterations (after $10,000$ burn-in) on six of the real data sets are now given. Each data set is size $T = 1000$ and the same computer was used for the timings. Timing results are summarised in Table 4.14. These data sets are used in Chapter 5 and are not described fully here, as we are only focusing on the MCMC algorithm performance.

Date Set	Hybrid algorithm time/s	Our Algorithm time/s	Hybrid thinning	Our thinning
British Airways PLC	1267	1652	3	2
Coca-Cola Co	1281	1678	3	2
HJ Heinz Co	1365	1681	4	3
JP Morgan Chase & Co	1281	1580	4	3
McDonald's Corp	1270	4782	4	0
Procter & Gamble Co	1272	610	0	1

Table 4.14: Timings for the Hybrid algorithm of Roberts et al. (2004) and our algorithm to perform 50,000 iterations on six share data sets.

The Hybrid algorithm of Roberts et al. (2004) takes a similar length of time to run for every data set, whilst our algorithm speed changes across data sets. The most obvious difference between data sets which will alter the speed of our algorithm is the posterior for λ, as this alters the point at which the Poisson point process is truncated, a_c, linearly (for the simple truncation scheme).

[2] We are grateful to Dr O. Papaspiliopoulos for generously allowing us to use his code.

Date Set	Posterior 95% credible interval for λ
British Airways PLC	0.077 (0.045, 0.155)
Coca-Cola Co	0.056 (0.010, 0.120)
HJ Heinz Co	0.117 (0.079, 0.197)
JP Morgan Chase & Co	0.114 (0.070, 0.173)
McDonald's Corp	0.215 (0.038, 0.668)
Procter & Gamble Co	0.029 (0.018, 0.038)

Table 4.15: Posterior 95% credible intervals for λ for six share data sets.

Table 4.15 shows the first four data sets have similar posteriors for λ and therefore take a similar length of time to run in Table 4.14. For the McDonald's data set the posterior supports larger λ values and Procter & Gamble Co supports only smaller λ values. These posteriors partly explain the time differences for our algorithm on these data sets. However, ν also alters a_c in a similar way to λ, so the posterior for λ is not able to entirely explain speed differences of our algorithm across the data sets.

The thinning of Table 4.14 ensures that each MCMC algorithm uses almost the same cpu time for a given data set. ACF plots of $\nu\alpha$ (which Roberts et al. (2004) examined to compare MCMC algorithm performance) for these thinnings are given in Figure 4.23, where $500,000$ samples were taken after a burn-in of $50,000$ iterations. As the thinning of the MCMC samples is not constant for each data set, Figure 4.23 cannot be used to compare the efficiency of the algorithms across data sets and should only be used to compare the performance of the two MCMC algorithms.

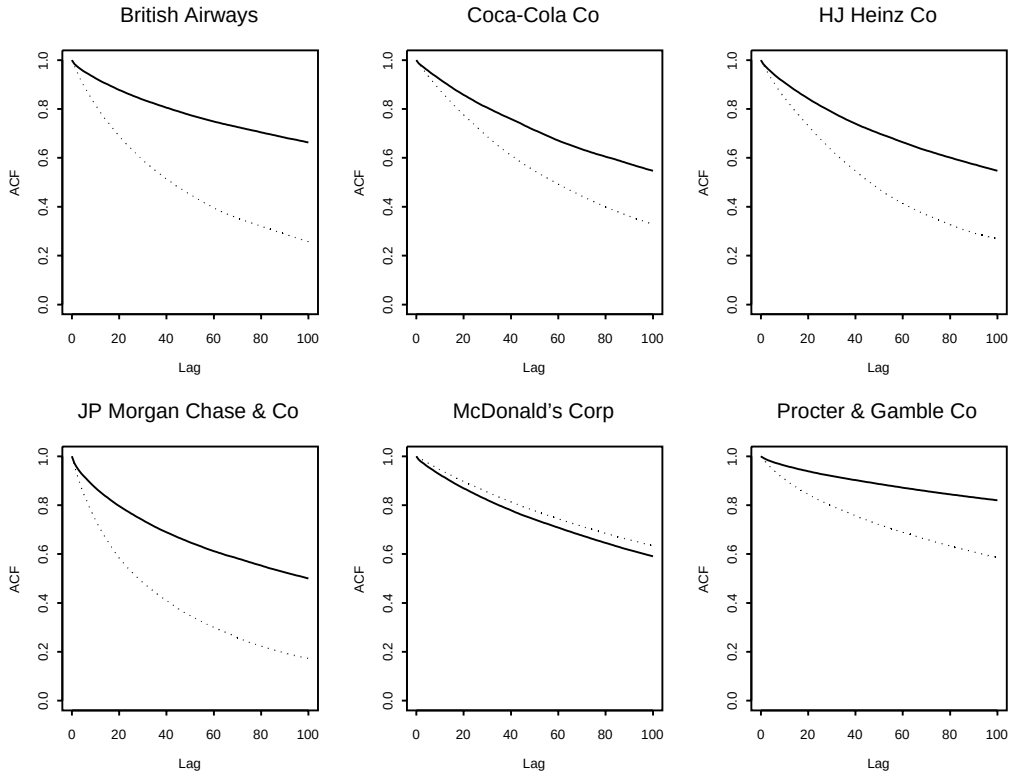


Figure 4.23: ACF($\nu\alpha$) plots of the Hybrid MCMC algorithm of Roberts et al. (2004) (solid line) and our algorithm (dashed line).

Our algorithm outperforms the Hybrid algorithm for five of the six data sets. For the remaining McDonald's Corp data set, the Hybrid algorithm performs very slightly better. The aim of Test6 was to show it is not easy to significantly improve our MCMC algorithm. Now this has been achieved, we assume the MCMC algorithm is working correctly (from previous tests) and that provided the chain is thinned sufficiently, we can sample (approximately) independently from the posterior. The empirical performances of the models on real data are investigated in the next chapter.

Chapter 5

Model selection: Empirical performance of the models

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

5.1 Predictive densities

Predictive densities assess how likely "unseen" data (on which the model was not fitted) is for different models. The method used is similar to the one in Vrontos et al. (2003), where the performance of some popular discrete time financial models is investigated.

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

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

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 in Section 4.3 and Appendix C.5) and denoted by $\underline{\theta}_1, \ldots, \underline{\theta}_B$. The "Rao-Blackwellized"

MC estimate for the predictive density is

$$\widehat{p}\left(\underline{Y}^{'}|\underline{y}\right)=\frac{1}{B}\sum_{i=1}^{B}p\left(\underline{Y}^{'}|\underline{\theta}_{i},\underline{y}\right),$$

where $p\left(\underline{Y}^{'}|\underline{\theta}_{i},\underline{y}\right)$ is estimated by averaging the likelihood function given in equation (3.11) over many volatilities generated from the non-latent parameters $\underline{\theta}_{i}$. This method is similar to those described in Pitt and Shephard (1999) and Vrontos et al. (2003).

The predictive density given $\underline{\theta}_{i}$, $p\left(\underline{Y}^{'}|\underline{\theta}_{i},\underline{y}\right)$, is estimated using

$$p\left(\underline{Y}^{'}|\underline{\theta}_{i},\underline{y}\right)=\frac{1}{B_{2}}\sum_{j=1}^{B_{2}}p\left(\underline{Y}^{'}|\underline{\sigma_{j}},\underline{y}\right), \tag{5.1}$$

where $\underline{\sigma_{j}}$ is a volatility generated using $\underline{\theta_{i}}$ and B_2 is large enough so that the approximation is sufficiently accurate. This will increase rapidly as the number of "unseen" data points and dimension of the Monte Carlo integration increases.

Models, which are fitted to $\underline{y}$ and have large $\widehat{p}\left(\underline{Y}^{'}|\underline{y}\right)$, explain the "unseen" data, $\underline{Y}^{'}$, well. This gives a way of comparing the out of sample fit of the different models.

5.2 Option pricing

All our option pricing assumes that investors are completely indifferent to the risk involved in an investment and is only concerned about expected return, and so are risk-neutral.

For the models described in Chapter 3, in general, option prices must be evaluated numerically, computing the expected discounted payoff of the option (with Monte Carlo integration say). For certain European options, techniques have been developed to try to improve the numerical evaluation of these integrals (see Nicolato and Venardos (2003) and Hubalek and Tompkins (2001)), though these techniques perform best when the payoff is only dependent on the expiry (or strike) value of the asset, such as for a European call or put. For more complicated options, such as an Asian option (which is dependent on the average value of the asset over some time period), the methods are less efficient. For this reason, all options are priced using forward simulation of the share, from equation (3.11), as the forward simulation of the asset must be performed for some of the more complicated options. Evaluating the simpler payoffs at the same time does not alter the efficiency significantly.

To price a European option, the expected discounted payoff is calculated (using the estimated constant interest rate). The prices of the algorithm are indifferent to risk and

therefore risk neutral. To test the performance of the different models, the models were fitted to 1000 data points and the expected discounted payoff of a variety of different options was calculated. The rescaled sum of the squared error loss between predicted and actual discounted option payoffs is then reported. The rescaling is such that the *Inverse Gamma* marginal has a rescaled error of 1. The actual discounted payoffs are evaluated by discounting using the estimated drift, as although the actual drift for shares at a given time can be looked up (as it is the interest rate), the actual drift of FX data cannot.

If many different option payoffs are considered, models which predict paths close to the actual share path will perform well. In this case, although it is the path of the share which is important (as is also the case for predictive densities), conclusions based on predictive densities and option pricing results need not be the same. Predictive densities have a loss function controlled by the likelihood, whilst for option pricing results, the loss function is chosen by the user, which we choose to be the rescaled squared error loss. For each data set and marginal distribution we price 27 options and report

$$\frac{\sum_{i=1}^{27}\left(P_{M,i}-O_i\right)^2}{\sum_{i=1}^{27}\left(P_{IGa,i}-O_i\right)^2},$$

where O_i is the observed discounted payoff of option number i and $P_{M,i}$ is the predicted discounted payoff of option number i for marginal distribution M.

5.2.1 Simulation from the asset process and rescaling

From equation (3.11), the log asset, $x\left(t\right)$, follows the equation

$$dx\left(t\right)=\mu dt-\frac{\sigma^2\left(t\right)}{2}\left(dt\right)^{2H}+\sigma\left(t\right)dW_{MVN}\left(t\right)+\rho d\overline{z}\left(\lambda t\right).$$

The discrete solution to this is

$$x_{t+\Delta}-x_t=\mu\Delta-\frac{\sigma_t^2\Delta^{2H-1}}{2}+\sigma_t A_i+\rho z_i, \tag{5.2}$$

where $\underline{A}\sim MVN\left(0,\Sigma\right)$ and σ_t are discrete simulations from the volatility process of the MCMC applied to the original observed log returns (see Hu and Øksendal (2003)).

It is sometimes favourable to rescale the data, so that parameters in the MCMC are of reasonable size. Let $x_2=\beta x$ (this is the equivalent to running the model on the rescaled log returns $\underline{y_2}=\beta\underline{y}$), then equation (3.11) becomes

$$dx_2\left(t\right)=\beta\left\{\mu dt-\frac{\sigma^2\left(t\right)}{2}\left(dt\right)^{2H}\right\}+\beta\sigma\left(t\right)dW_{MVN}\left(t\right)+\beta\rho d\overline{z}\left(\lambda t\right),$$

where $\sigma\left(t\right)$ is the volatility for the original log returns.

If $\sigma(t)$ and $z(t)$ are the volatility and BDLP which drive the Ornstein-Uhlenbeck equation (1.8) for the original data and $\xi(t)$ and $z_2(t)$ are the volatility and BDLP for the rescaled log returns respectively, then $\xi(t) = \beta\sigma(t)$, $z_2(t) = \beta^2 z(t)$ (from equation (4.4)) and so

$$dx_2(t) = \left\{\beta\mu dt - \frac{\xi^2(t)}{2\beta}(dt)^{2H}\right\} + \xi(t)\, dW_{MVN}(t) + \frac{\rho}{\beta} d\overline{z_2}(\lambda t) \tag{5.3}$$

and the likelihood for $\underline{Y_2}$ is specified by the likelihood for $\underline{Y}$ given in equation (3.12). This is proportional to

$$f_{\underline{Y_2}}(\underline{y_2}) = \left(\prod_{i=1}^{T} \frac{1}{\xi_i}\right) f_{\underline{G}}(\underline{g}),$$

where

$$g_i = \frac{(\underline{y_2})_i - \beta\mu\Delta + \xi_i^2\Delta^{2H-1}/2\beta - \rho z_i/\beta}{\xi_i}$$

and $\underline{G} \sim MVN(0, \Sigma)$. Comparing equation (3.11) with equation (5.3) and using the solution given in equation (5.2), gives

$$x_{2,t+\Delta} - x_{2,t} = \left(\beta\mu\Delta - \frac{\xi_i^2}{2\beta}\Delta^{2H-1}\right) + \xi_i A_i + \frac{\rho}{\beta}(z_2)_i,$$

where $\underline{A} \sim MVN(0, \Sigma)$.

As $x_2 = \beta x$, the discrete solution for the original share is

$$x_{t+\Delta} - x_t = \left(\mu\Delta - \frac{\xi_i^2}{2\beta^2}\Delta^{2H-1}\right) + \frac{\xi_i}{\beta} A_i + \frac{\rho}{\beta^2}(z_2)_i.$$

Although this is easier to evaluate by setting $\xi(t) = \beta\sigma(t)$ in equation (5.2), it is still useful as it forces the correct derivation of the likelihood. Simulation of the share requires generation of $MVN(0, \Sigma)$ random variates and this was discussed in Section 3.5.

5.2.2 General option pricing algorithm

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

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

(2) Simulate $\sigma_1^2, \ldots, \sigma_t^2 | \lambda, \gamma, \kappa, \nu, \alpha$ from equation (4.4) 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 forward (using ρ, 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 $\rho, \mu, \lambda, \gamma, \kappa, \nu, \alpha$ is sufficiently accurate.

(5) go to (1) and take another sample from the posterior of $\rho, \mu, \lambda, \gamma, \kappa, \nu, \alpha$ 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 payoffs of the options are then examined to compare the performance of the different models. Histograms of the expected discounted payoff from (4), for two popular options on real data sets, are given in Figures 5.1 and 5.2 and demonstrate the BNS SV model accurately predicting the payoff of two exotic options.

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.

Nicolato and Venardos (2003) derive the set of equivalent martingale measures (that is 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.

The BNS SV model was constructed to try to improve on shortcomings in the Black-Scholes model for option pricing over small time periods. For this reason, typically t will be small and it will be feasible to perform the Monte Carlo integration in step (3) of the option pricing algorithm.

Our focus is to compare the performance of the different marginal distributions and so all options considered are European (meaning that the option expires at a fixed time in the future and the holder cannot exercise before this time). This simplifies the Monte Carlo integration as it does not require knowledge of how to exercise the option.

For options with payoffs which are a function of the continuous movement of the underlying (i.e. the payoff is not determined by the underlying at a finite number of discrete time points), the option pricing method described above can give biased estimates for the fair price (see for example Ribeiro and Webber (2003)), though the algorithm can be adjusted to give unbiased estimates. Here, all option prices are determined by the open prices of the asset over 20 days, so are a function of the underlying at discrete time points

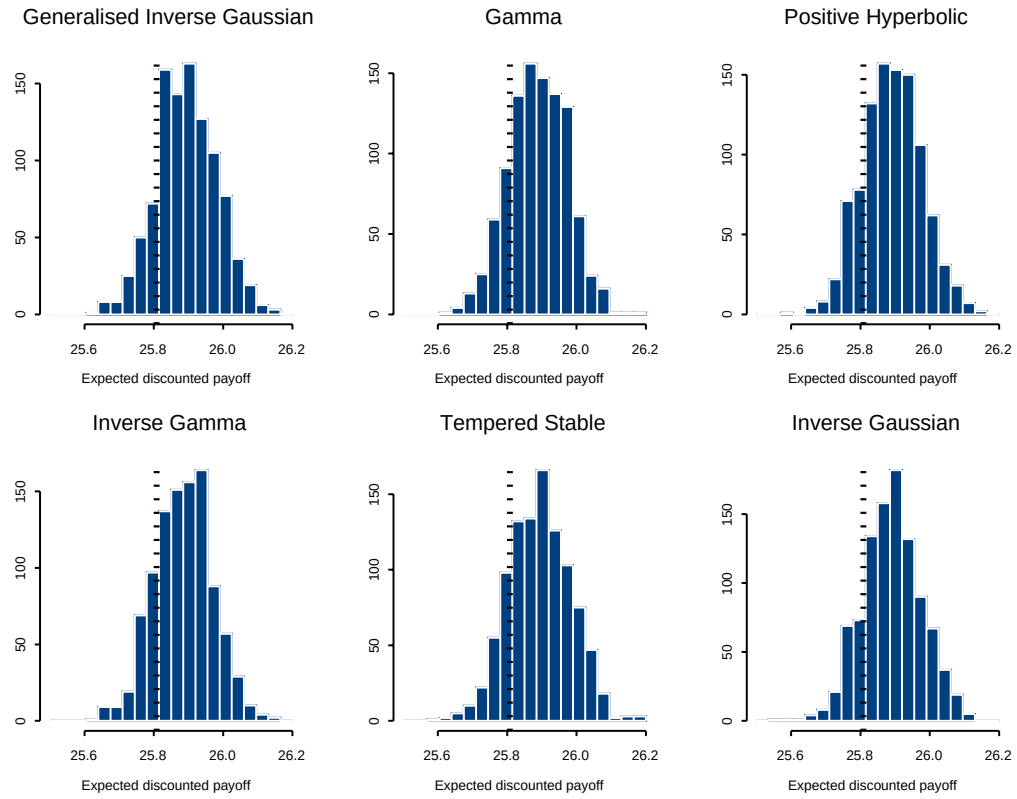


Figure 5.1: Histograms of the expected discounted payoff for different posterior samples from the MCMC, for an arithmetic Asian option on Microsoft shares.

and the unbiasing technique need not be implemented. For example, for a continuous Asian option, the payoff is an average of the underlying over some time period (and so the payoff is a function of the continuously moving asset). The Asian options we consider are averages of the underlying open prices on each day (and so the payoff is a function of the asset at discrete time points).

5.2.3 Test example: constant volatility

For the constant volatility Black-Scholes model, the fair price of an option can be calculated by solving equation (1.4), for appropriate boundary conditions. For the Test example, consider this model, for two of the simplest standard options: the European call and put (sometimes referred to as the 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

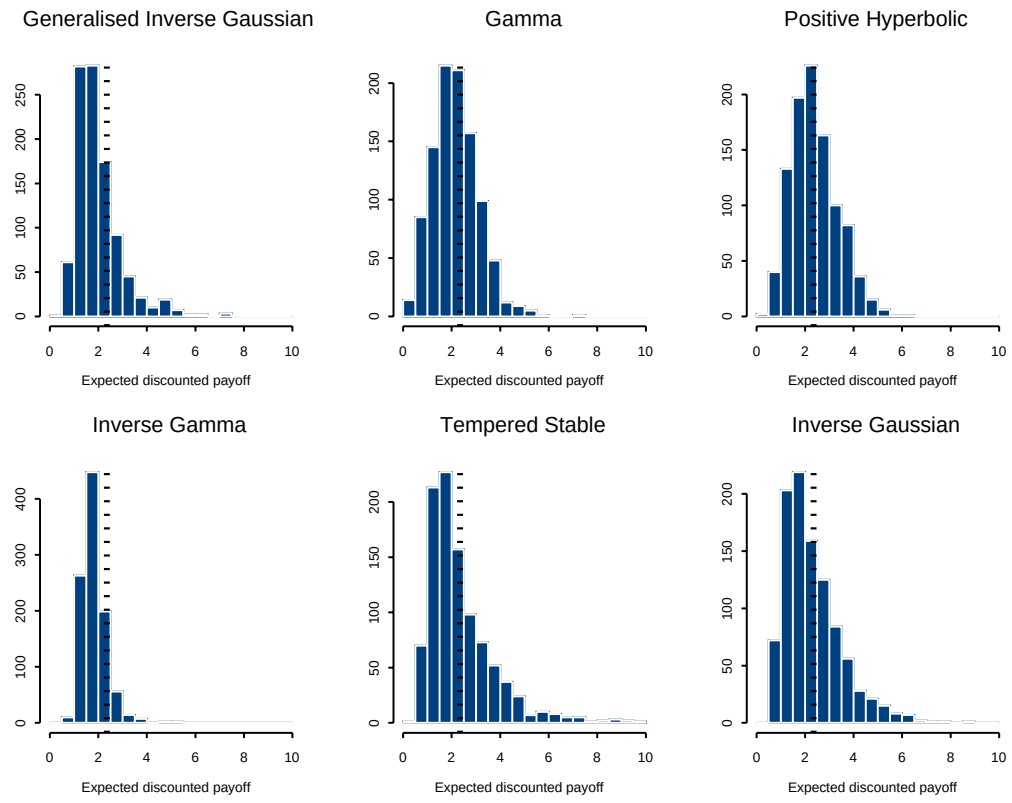


Figure 5.2: Histograms of the expected discounted payoff for different posterior samples from the MCMC, for a knock in option with vanilla call payoff on Procter & Gamble Co shares.

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 expiring at the same time t and 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 $t = 0$, $T = 20$, $S(0) = 100$, $E = 97$ 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 was made on the non-latent parameters of the BNS SV model using MCMC. Volatilities were then generated using these non-latent parameters and the expected discounted payoff, given these simulated volatilities, was 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. For each marginal distribution, the convergence of the method to the correct price can be seen in Figures C.5, C.6 and C.7. The thick line is the expected discounted payoff given $\sigma = 0.03$.

Note that the Black-Scholes equation with constant volatility will give uncorrelated $\sigma^2(t)$, so the MCMC should give large λ (corresponding to a rapid decay in the correlation of the volatility). Our experience is that this tends to be the case, though sometimes the chain can initially move to a state where λ is very small and σ is almost constant (corresponding to very slow exponential decay in the volatility). If the chain is run for long enough, it eventually leaves this state and converges to the "expected" large λ.

Another implementation problem, for constant volatility data, is that it is not obvious when the chain has converged. It is to be expected that the simulated σ_i^2 be approximately σ^2 and, in practice, this can mean that ν, α and λ all increase in size giving the correct mean, almost zero variance and low correlation in σ^2. This makes A and R (see equation (4.10)) grow in size, as more points are needed before the summation is truncated. This slows the algorithm down and can cause the program to crash due to lack of memory. However, the difference in the simulated σ_i^2 does not vary significantly as soon as ν and α get large, almost regardless of the parameters λ, A and R. The λ prior suggested by Griffin and Steel (2003), $Ga(1,1)$, gives smaller λ than the flatter $Ga(1,0.001)$ prior (used for testing purposes in Chapter 4) and this can stop the system expanding too much. This predominantly occurs on training data where the correlation in volatility is zero.

5.2.4 Battery of tests

Having verified the correct option pricing of some of the models for a vanilla call and put on data generated from the constant volatility Black-Scholes model (in Section 5.2.3), options are now priced on real data sets with more complicated payoffs, to test the empirical performance of the models.

The 27 (European) option payoffs are combinations of the following standard options:-

(1) European call/put.

(2) Binary call/put.

(3) Asian (arithmetic and geometric averages).

(4) Barrier (knock in/out).

(5) Parisian (price based on how often the share is within some specified range).

(6) Lookback (assumes the option expired at the time which would have given the maximum payoff).

More details on different option payoffs can be found in Hull (2000). The exact payoffs used can be found in Appendix C.7.

The FX rates are the interbank exchange rate between the US Dollar and various currencies and the shares data sets are open prices on the NYSE for the shares used in Section 3.1.3. The data sets are listed in Table 5.1.

Exchange rates (US Dollar vs)	Shares
Australian Dollar	British Airways PLC
Brazilian Real	Citigroup Inc
British Pound	Coca-Cola Co
Canadian Dollar	General Motors Corp
Colombian Peso	HJ Heinz Co
Danish Krone	Host Marriott Corp
Euro	JP Morgan Chase & Co
Iceland Krona	Kellogg Co
Japanese Yen	McDonald's Corp
Moroccan Dirham	Microsoft
Singapore Dollar	Procter & Gamble Co
Thai Baht	S&P 500
	Textron Inc
	Time Warner Inc

Table 5.1: FX rates and shares used for model selection.

For FX options, the data are daily from 6^{th} March 2001 to 1^{st} December 2003, whilst for share options, the data are daily (excluding days when the NYSE market was closed) from 29^{th} November 1999 to 1^{st} December 2003.

The chains were run for 10,000 iterations (burn-in) and then run for a further 10,000 iterations, calculating the predictive densities at each of the second batch of 10,000 iterations. After these 20,000 iterations, the options were priced, thinning the MCMC by taking every 250^{th} generated volatility. A minimum of 250 and maximum of 1000 volatilities were used to price the options. More volatilities were used until the maximum variance of each of these estimates for the option prices was less than 0.02. If the maximum variance of the prices was still greater than 0.02 after 1000 volatilities, the simulation was stopped and the estimated prices from 1000 volatilities were reported.

For a given volatility, 20 estimates for the fair prices were obtained, each using 1000 paths in the Monte Carlo integration. More estimates were then generated for the fair price, each using a further 1000 points. Once the variance of the estimates of the fair prices were less than 0.02, the fair prices of the options for that specific stochastic volatility were taken to be the averages of these estimates.

5.3 Model selection results

The results for the predictive densities and option pricing for real data are now summarised. For predictive densities, the entries in the tables are posterior median and 95% credible intervals on the log scale. Results for posterior summaries of individual parameters are recorded similarly. For predictive densities over 20 "unseen" data points, 1000 volatilities were used to estimate $\widehat{p}\left(\underline{Y}'|\underline{\theta}_i, \underline{y}\right)$ in equation (5.1) (i.e. B_2 was set equal to 1000) and for 80 "unseen" data points 10,000 volatilities were used. For option pricing, the rescaled sum of the squared error between expected discounted and actual discounted payoffs are reported.

Comments on the performance of the different models are given for the FX rates and shares individually, in the section where the results are displayed. After the results for both sectors are presented, the results are discussed.

5.3.1 Different marginal distributions: Predictive densities over 20 "unseen" data points and option pricing results

a) FX rates

Predictive Density	$GIG(\gamma, \nu, \alpha)$	$TS(\kappa, \nu, \alpha)$	$IG(\nu, \alpha)$
Australian Dollar	-38.4 (-39.4, -36.7)	-38.6 (-39.6, -37.1)	-38.7 (-39.2, -38.0)
Brazilian Real	-17.1 (-20.7, -11.5)	-18.1 (-21.6, -12.8)	-17.1 (-18.5, -15.5)
British Pound	-41.2 (-42.7, -38.7)	-41.2 (-42.6, -38.9)	-42.6 (-43.1, -42.1)
Canadian Dollar	-54.5 (-56.7, -52.6)	-54.9 (-56.7, -53.8)	-54.7 (-55.6, -53.9)
Colombian Peso	-38.1 (-40.2, -35.1)	-38.2 (-40.8, -35.3)	-39.9 (-42.7, -37.1)
Danish Krone	-43.7 (-44.8, -42.2)	-43.5 (-44.6, -42.2)	-43.4 (-44.2, -42.7)
Euro	-43.6 (-44.5, -42.6)	-43.4 (-44.2, -42.4)	-43.2 (-43.9, -42.6)
Iceland Krona	-31.4 (-35.0, -26.8)	-31.6 (-34.8, -27.1)	-37.8 (-38.9, -35.9)
Japanese Yen	-39.7 (-40.8, -37.8)	-39.4 (-40.5, -37.7)	-39.4 (-40.1, -38.5)
Moroccan Dirham	-63.1 (-70.1, -57.5)	-61.7 (-67.8, -56.8)	-61.3 (-64.6, -58.4)
Singapore Dollar	-45.5 (-46.7, -44.2)	-45.7 (-47.0, -44.4)	-44.5 (-45.6, -43.7)
Thai Baht	-29.7 (-30.8, -28.6)	-30.6 (-31.9, -28.8)	-29.2 (-30.3, -28.0)

Table 5.2: Median and 95% credible intervals for predictive densities of GIG, TS and IG marginal distributions for FX data.

Predictive Density	$Ga(\nu, \alpha)$	$RPH(\nu, \alpha)$	$IGa(\nu, \alpha)$
Australian Dollar	-38.6 (-39.6, -37.8)	-38.6 (-39.4, -37.9)	-38.7 (-39.3, -38.3)
Brazilian Real	-17.4 (-21.1, -11.9)	-25.2 (-27.3, -21.4)	-24.8 (-26.8, -23.0)
British Pound	-40.8 (-41.6, -39.4)	-42.2 (-43.0, -40.7)	-43.4 (-43.8, -43.0)
Canadian Dollar	-54.7 (-56.0, -53.7)	-54.5 (-55.6, -53.8)	-54.7 (-55.7, -54.0)
Colombian Peso	-38.0 (-40.2, -35.0)	-37.8 (-38.4, -36.9)	-38.2 (-38.5, -37.9)
Danish Krone	-43.7 (-44.8, -42.2)	-42.6 (-43.4, -41.6)	-42.0 (-42.5, -41.5)
Euro	-43.5 (-44.3, -42.5)	-42.7 (-43.1, -42.2)	-42.1 (-42.5, -41.7)
Iceland Krona	-31.3 (-34.8, -26.7)	-38.2 (-39.3, -35.3)	-40.2 (-40.7, -39.7)
Japanese Yen	-39.5 (-40.7, -37.5)	-39.9 (-40.6, -39.0)	-41.0 (-41.5, -40.6)
Moroccan Dirham	-59.0 (-60.0, -58.2)	-60.8 (-63.3, -58.6)	-60.6 (-62.0, -59.6)
Singapore Dollar	-45.6 (-46.9, -44.3)	-43.4 (-43.8, -43.1)	-42.9 (-43.2, -42.6)
Thai Baht	-31.0 (-32.2, -29.3)	-30.1 (-31.4, -28.1)	-31.3 (-32.5, -29.9)

Table 5.3: Median and 95% credible intervals for predictive densities of Ga, RPH and IGa marginal distributions for FX data.

There is no clear trend to suggest any of the marginal distributions have a noticeably larger predictive density than the others. Although this is not helpful is assessing which marginal distribution should be used for FX data, there are still the option pricing results to try to answer this. Predictive densities on their own were unable to direct us to which marginal is most appropriate.

Option Pricing	$GIG(\gamma, \nu, \alpha)$	$TS(\kappa, \nu, \alpha)$	$IG(\nu, \alpha)$
Australian Dollar	0.96	1.01	1.03
Brazilian Real	0.97	1.00	1.12
British Pound	1.36	1.37	1.07
Canadian Dollar	0.85	1.05	1.01
Colombian Peso	0.80	0.80	1.23
Danish Krone	1.70	1.72	1.53
Euro	1.33	1.34	1.20
Iceland Krona	1.46	1.45	1.52
Japanese Yen	1.27	1.23	1.17
Moroccan Dirham	1.42	1.37	1.26
Singapore Dollar	1.05	1.01	1.05
Thai Baht	0.97	1.40	1.05
Sum	14.14	14.76	14.25

Table 5.4: Summaries of option pricing performance of GIG, TS and IG marginal distributions for FX data; rescaled squared errors.

Option Pricing	$Ga(\nu, \alpha)$	$RPH(\nu, \alpha)$	$IGa(\nu, \alpha)$
Australian Dollar	1.01	1.00	1.00
Brazilian Real	0.97	0.90	1.00
British Pound	1.43	1.18	1.00
Canadian Dollar	1.01	1.01	1.00
Colombian Peso	0.74	0.66	1.00
Danish Krone	1.69	1.47	1.00
Euro	1.33	1.24	1.00
Iceland Krona	1.46	1.20	1.00
Japanese Yen	1.23	1.14	1.00
Moroccan Dirham	1.39	1.13	1.00
Singapore Dollar	1.05	1.03	1.00
Thai Baht	1.50	0.97	1.00
Sum	14.81	12.91	12.00

Table 5.5: Summaries of option pricing performance of Ga, RPH and IGa marginal distributions for FX data; rescaled squared errors.

For the three parameter distributions, we find the GIG to predict the payoff more accurately than 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 the rescaled 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. This shows that, when financial models are tested, they should be tested for the purpose for which the models were developed, as tests need not have identical conclusions.

b) Shares

Predictive Density	$GIG(\gamma, \nu, \alpha)$	$TS(\kappa, \nu, \alpha)$	$IG(\nu, \alpha)$
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)

Table 5.6: Median and 95% credible intervals for predictive densities of GIG, TS and IG marginal distributions for share data.

Predictive Density	$Ga(\nu, \alpha)$	$RPH(\nu, \alpha)$	$IGa(\nu, \alpha)$
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 5.7: Median and 95% credible intervals for predictive densities of Ga, RPH and IGa marginal distributions for share data.

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 5.3.

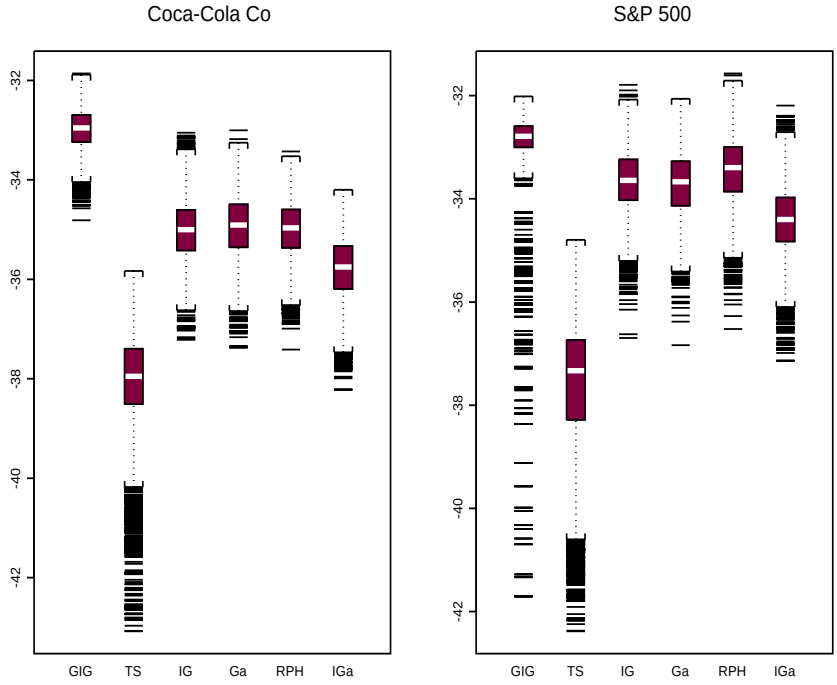


Figure 5.3: Boxplots of samples from the predictive densities of the six different marginals on Coca-Cola Co and S&P 500 data.

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

Table 5.8: Summaries of option pricing performance of GIG, TS and IG marginal distributions for share data; rescaled squared errors.

Option Pricing	$Ga(\nu, \alpha)$	$RPH(\nu, \alpha)$	$IGa(\nu, \alpha)$
British Airways PLC	1.02	1.01	1.00
Citigroup Inc	1.00	1.00	1.00
Coca-Cola Co	1.05	1.05	1.00
General Motors Corp	1.01	1.00	1.00
HJ Heinz Co	1.03	1.05	1.00
Host Marriott Corp	1.00	1.00	1.00
JP Morgan Chase & Co	1.00	1.00	1.00
Kellogg Co	1.05	1.04	1.00
McDonald's Corp	0.99	0.99	1.00
Microsoft	1.03	1.02	1.00
Procter & Gamble Co	0.99	1.03	1.00
S&P 500	0.99	0.99	1.00
Textron Inc	0.99	1.00	1.00
Time Warner Inc	0.99	1.00	1.00
Sum	14.15	14.19	14.00

Table 5.9: Summaries of option pricing performance of Ga, RPH and IGa distributions for share data; rescaled squared errors.

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.

The important conclusion from this section is that the *Inverse Gamma* marginal distribution should be used when the stochastic volatility follows the BNS SV models and the focus is option pricing. The models have similar predictive densities for FX rates and often the *Generalised Inverse Gaussian* distribution has the largest predictive density for share data.

For the remaining results, only predictive densities are summarised. In principle, option pricing results could be calculated, though are computationally intensive. We continue using the *Gamma* marginal distribution, as this is the fastest to implement.

5.3.2 Posterior distributions of parameters of the *GIG* distribution

We now look at 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.

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 5.10: Special cases of the GIG distribution.

The posterior summaries for γ, ν, α (for the GIG distribution) are now given. Posterior summaries for the TS distribution are not given as the only standard special case of this is the IG distribution, which is also a special case of the GIG.

FX rate	γ	ν	α
Australian Dollar	0.39 (0.25, 0.50)	0.02 (0.01, 0.08)	0.91 (0.79, 1.04)
Brazilian Real	0.20 (0.17, 0.24)	0.00 (0.00, 0.00)	0.69 (0.62, 0.76)
British Pound	0.42 (0.29, 0.51)	0.02 (0.00, 0.07)	0.91 (0.80, 1.04)
Canadian Dollar	0.40 (0.32, 0.54)	0.01 (0.00, 0.03)	0.88 (0.78, 0.99)
Colombian Peso	0.22 (0.19, 0.27)	0.00 (0.00, 0.00)	0.75 (0.67, 0.85)
Danish Krone	0.40 (0.32, 0.44)	0.00 (0.00, 0.10)	0.85 (0.76, 0.94)
Euro	0.44 (0.35, 0.54)	0.02 (0.00, 0.05)	0.93 (0.82, 1.05)
Iceland Krona	0.25 (0.22, 0.29)	0.00 (0.00, 0.00)	0.72 (0.65, 0.80)
Japanese Yen	0.37 (0.26, 0.48)	0.04 (0.00, 0.14)	0.89 (0.73, 1.03)
Moroccan Dirham	0.27 (0.23, 0.31)	0.00 (0.00, 0.00)	0.79 (0.72, 0.87)
Singapore Dollar	0.32 (0.27, 0.38)	0.00 (0.00, 0.00)	0.79 (0.72, 0.90)
Thai Baht	0.46 (0.10, 0.55)	0.03 (0.00, 0.22)	1.01 (0.79, 1.14)

Table 5.11: 95% credible intervals for the parameters of the GIG distribution for FX data.

For FX data, the posterior distributions of the three parameters of the *Generalised Inverse Gaussian* support the *Gamma* distribution. The three other special cases of the *GIG* distribution are not supported. A graph of the posterior of the (γ, ν) plane for the British Pound data set is given in Figure 5.4 to illustrate this.

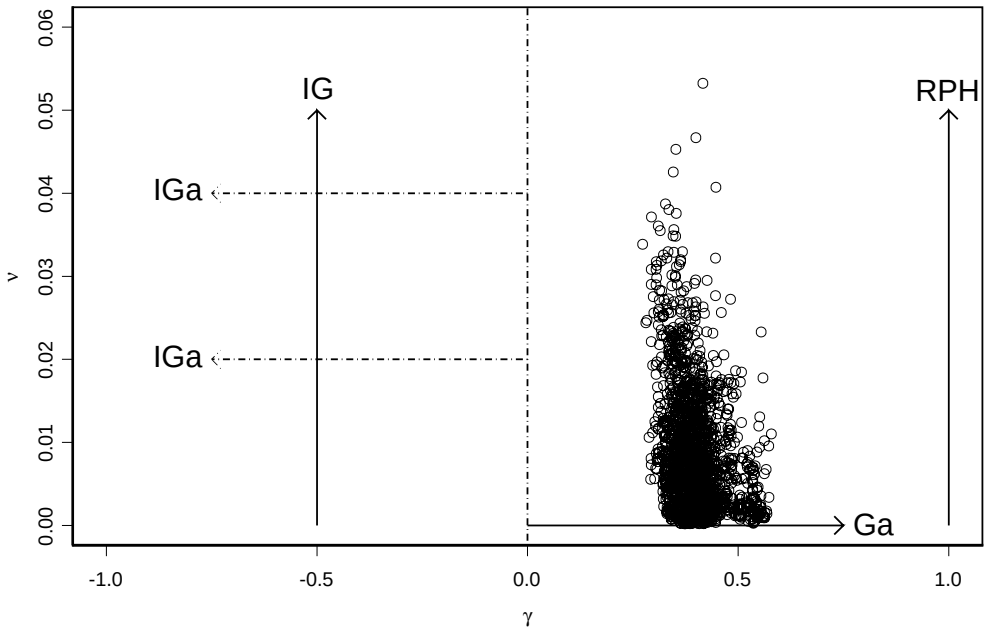


Figure 5.4: Posterior samples in the (γ, ν) plane for the British Pound data set.

Figure 5.4 shows that the *Ga* distribution is supported by the posterior of γ (note the scale of the ν axis) and the three other special cases of the *GIG* are not. Tables 5.2 and 5.3 reflect this, where the *Ga* marginal has the largest predictive density of the two parameter distributions for the British Pound data set. Although this posterior for ν is typical for the FX data sets investigated, the *Gamma* distribution does not always have a large predictive density or perform well for option pricing for FX data set. Models should therefore be accepted or rejected based on tests which concentrate on what the model will be used for.

Share	γ	ν	α
British Airways PLC	-1.44 (-2.11, -0.49)	1.21 (0.92, 1.48)	0.37 (0.06, 1.06)
Citigroup Inc	-1.59 (-2.75, -0.77)	1.40 (1.11, 1.77)	0.59 (0.06, 1.25)
Coca-Cola Co	-0.63 (-2.44, 1.41)	1.40 (0.69, 2.02)	1.41 (0.45, 2.14)
General Motors Corp	0.22 (-2.34, 1.88)	1.37 (0.43, 2.28)	1.68 (0.42, 2.35)
HJ Heinz Co	0.78 (-1.08, 2.73)	0.86 (0.26, 1.66)	1.69 (0.91, 2.49)
Host Marriott Corp	-1.67 (-2.08, -1.26)	1.14 (0.98, 1.34)	0.10 (0.00, 0.40)
JP Morgan Chase & Co	-1.44 (-1.81, -0.92)	1.14 (0.94, 1.33)	0.40 (0.04, 0.87)
Kellogg Co	-1.12 (-2.70, -0.12)	1.37 (1.00, 1.99)	0.95 (0.18, 1.55)
McDonald's Corp	-0.83 (-1.50, 0.65)	1.27 (0.68, 1.61)	1.06 (0.59, 1.65)
Microsoft	-1.80 (-2.31, 1.18)	1.18 (0.96, 1.39)	0.22 (0.03, 0.58)
Procter & Gamble Co	-1.39 (-1.81, -1.03)	0.84 (0.70, 0.98)	0.16 (0.00, 0.41)
S&P 500	-0.21 (-3.29, 2.20)	1.38 (0.72, 2.34)	1.57 (0.67, 2.47)
Textron Inc	-0.87 (-1.83, 1.27)	1.18 (0.45, 1.56)	0.96 (0.38, 1.88)
Time Warner Inc	-0.26 (-1.54, 2.05)	1.04 (0.18, 1.55)	1.09 (0.28, 2.20)

Table 5.12: 95% credible intervals for the parameters of the GIG distribution for share data.

For share data, the 95% credible interval for ν is not concentrated near $\nu = 0$ (unlike the results of Table 5.11), so the *Gamma* distribution is not supported by the *GIG*. The 95% credible interval for the posterior of γ supports $\gamma = 1$ (*RPH*) for six data sets and $\gamma = -0.5$ (*IG)* is supported for nine data sets. As discussed in Section 4.4.1, even for data from the $GIG\left(-\nu, \sqrt{2\alpha}, 0\right)$ $(IGa\left(\nu, \alpha\right))$ distribution, the third parameter need not necessarily be estimated to be very close to zero. For this reason, conclusions relating to the *IGa* distribution are not based on the third parameter of the *GIG* but are based the first parameter only. For six data sets, the posterior of γ is strictly negative, so the *IGa* distribution is only partly supported by the *GIG* (and this is illustrated in the very negative log-likelihood of the *Inverse Gamma* marginal of Table 5.7). As the posteriors of γ, ν, α for share data are less consistent than FX data, two data sets with typical posterior distributions are displayed in Figure 5.5.

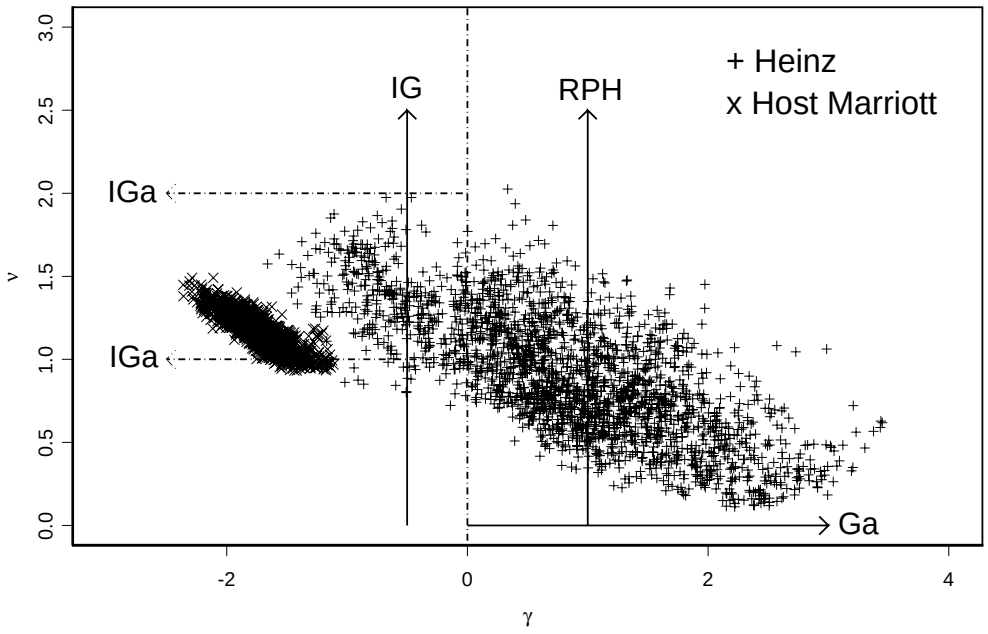


Figure 5.5: Posterior samples in the (γ, ν) plane for the Heinz and Host Marriott data sets.

Figure 5.5 illustrates that the *Gamma* marginal is not supported by the *GIG* distribution (note the difference in the $\nu - axis$ scale for Figures 5.4 and 5.5) and this is the case for all share data sets. For Heinz share data, the *RPH* marginal looks like the most suitable two parameter marginal, though as the posterior is not concentrated near $\gamma = 1$, the generalisation from the *RPH* to the *GIG* marginal looks necessary. The *IG* marginal is supported, though less than the *RPH*. The *IGa* is unsupported as almost all the posterior is in the positive γ plane. For the Host Marriott data set, the only two parameter marginal which is supported is the *IGa*.

5.3.3 Stochastic λ process: Predictive densities over 20 and 80 "unseen" data points

a) FX rates

Predictive Density	Constant λ	Stochastic λ
Australian Dollar	-38.6 (-39.6, -37.8)	-38.5 (-39.7, -37.0)
Brazilian Real	-10.9 (-14.6, -5.38)	-15.1 (-15.5, -14.9)
British Pound	-40.8 (-41.6, -39.4)	-41.3 (-45.2, -39.0)
Canadian Dollar	-54.7 (-56.0, -53.7)	-55.4 (-58.2, -55.0)
Colombian Peso	-38.0 (-40.2, -35.0)	-38.2 (-40.5, -35.2)
Danish Krone	-43.7 (-44.8, -42.2)	-43.4 (-43.6, -43.2)
Euro	-43.5 (-44.3, -42.5)	-43.5 (-44.5, -42.6)
Iceland Krona	-31.3 (-34.8, -26.7)	-31.6 (-35.1, -27.2)
Japanese Yen	-39.5 (-40.7, -37.5)	-39.6 (-40.7, -37.7)
Moroccan Dirham	-59.0 (-60.0, -58.2)	-60.8 (-62.3, -60.0)
Singapore Dollar	-45.6 (-46.9, -44.3)	-45.2 (-46.4, -43.9)
Thai Baht	-31.0 (-32.2, -29.3)	-27.9 (-30.8, -27.4)

Table 5.13: Summaries of predictive densities of models with constant and stochastic λ processes for FX data.

Results are mixed and it is not clear that there is any significant performance difference between using a constant or stochastic λ. Although for eight of the data sets the constant λ model has a larger predictive density (compared to only four for the stochastic λ), the differences in the median of the predictive densities are often small and it is difficult to draw a firm conclusion.

Predictive Density	Posterior for r	Posterior for ε^2
Australian Dollar	2.98 (0.247, 10.2)	0.017 (0.000, 1.20)
Brazilian Real	0.422 (0.094, 1.30)	4.99 (2.32, 11.9)
British Pound	3.23 (0.121, 13.5)	0.014 (0.000, 4.75)
Canadian Dollar	0.327 (0.034, 1.35)	4.06 (1.60, 10.3)
Colombian Peso	1.39 (0.289, 15.3)	0.851 (0.000, 4.12)
Danish Krone	0.126 (0.009, 0.731)	8.36 (2.75, 19.6)
Euro	3.80 (0.229, 18.5)	0.008 (0.000, 1.40)
Iceland Krona	0.404 (0.069, 1.33)	4.14 (2.01, 8.60)
Japanese Yen	2.52 (0.060, 13.6)	0.030 (0.000, 2.54)
Moroccan Dirham	0.400 (0.054, 1.42)	3.89 (1.66, 9.66)
Singapore Dollar	0.550 (0.072, 2.07)	2.23 (0.940, 6.01)
Thai Baht	0.368 (0.085, 2.58)	6.06 (0.113, 13.1)

Table 5.14: Summaries of the posterior of the stochastic λ process parameters for FX data.

When the posterior median of ε^2 is small (such as the Australian Dollar, British Pound, Euro and Japanese Yen) any jumps in λ will be small, so the generalisation to the stochastic λ process is probably not necessary and this is reflected in the similar predictive densities for these data sets for the two models. The posterior summaries for r and ε^2 are noticeably different for different data sets.

b) Shares

Predictive Density	Constant λ	Stochastic λ
British Airways PLC	-34.0 (-34.8, -33.3)	-33.7 (-35.3, -32.1)
Citigroup Inc	-28.8 (-30.0, -28.0)	-28.6 (-29.9, -27.9)
Coca-Cola Co	-34.9 (-36.3, -33.9)	-35.2 (-37.3, -33.9)
General Motors Corp	-45.9 (-46.3, -45.5)	-46.0 (-46.5, -45.6)
HJ Heinz Co	-26.1 (-28.6, -24.5)	-26.2 (-28.6, -24.9)
Host Marriott Corp	-36.1 (-36.9, -35.4)	-36.4 (-39.1, -35.6)
JP Morgan Chase & Co	-24.4 (-26.1, -23.0)	-23.8 (-24.5, -23.2)
Kellogg Co	-36.3 (-37.0, -35.7)	-35.9 (-36.7, -35.2)
McDonald's Corp	-48.8 (-49.7, -48.3)	-48.7 (-50.9, -48.2)
Microsoft	-22.4 (-24.4, -21.2)	-22.0 (-25.2, -21.2)
Procter & Gamble Co	-28.6 (-29.2, -28.1)	-28.2 (-28.8, -27.5)
S&P 500	-33.6 (-34.9, -32.5)	-33.3 (-34.9, -32.3)
Textron Inc	-42.6 (-43.0, -42.1)	-42.7 (-43.1, -42.2)
Time Warner Inc	-32.1 (-33.2, -31.5)	-32.6 (-33.8, -31.7)

Table 5.15: Summaries of predictive densities of models with constant and stochastic λ processes for share data.

Again the results are not conclusive. For eight data sets, the stochastic λ model appears to perform best while a constant λ performs best in the remaining six.

Predictive Density	Posterior for r	Posterior for ε^2
British Airways PLC	0.627 (0.098, 2.19)	1.63 (0.209, 4.96)
Citigroup Inc	0.537 (0.071, 2.14)	2.01 (0.406, 8.43)
Coca-Cola Co	0.504 (0.031, 2.22)	1.50 (0.051, 5.93)
General Motors Corp	0.526 (0.055, 2.21)	1.56 (0.128, 7.12)
HJ Heinz Co	0.333 (0.030, 1.87)	2.87 (0.293, 12.9)
Host Marriott Corp	0.601 (0.028, 2.57)	0.813 (0.032, 5.09)
JP Morgan Chase & Co	0.648 (0.110, 2.20)	2.05 (0.675, 7.51)
Kellogg Co	0.622 (0.089, 2.14)	1.86 (0.550, 5.16)
McDonald's Corp	0.617 (0.042, 2.54)	0.829 (0.036, 5.53)
Microsoft	0.451 (0.028, 2.12)	1.77 (0.135, 12.5)
Procter & Gamble Co	0.654 (0.110, 2.34)	2.72 (0.795, 7.18)
S&P 500	0.623 (0.034, 2.45)	1.14 (0.064, 5.30)
Textron Inc	0.637 (0.081, 2.43)	1.10 (0.075, 4.74)
Time Warner Inc	0.528 (0.051, 2.12)	1.89 (0.446, 8.11)

Table 5.16: Summaries of the posterior of the stochastic λ processes parameters for share data.

For the share data sets, the posteriors of r and ε^2 are similar between data sets (unlike for FX data sets). For example, the median of the posterior of ε^2 for each share data set is not near 0, whilst for FX rates sometimes it is.

It was hoped that the stochastic λ process would be able to accurately pick out a sensible value of λ at the end of each data set (which is then used as λ_0 for the predictive density calculations). The constant λ series cannot do this as it must pick a λ which fits the entire data series on which the MCMC is run. This is why it was felt that there may be an advantage using the stochastic λ process, even for predictive densities over small time periods, though it turned out not to be the case.

The stochastic λ process predicts jumps in λ. When the predictive densities are calculated using a small number of "unseen" data points, very few jumps will be predicted and this might make the predictive densities of the constant and stochastic λ process similar. To try to distinguish between the two models, predictive densities over 80 "unseen" days are now given using two different priors for the number of jumps in the "seen" 1000 data points. For many "unseen" data points, the stochastic λ process will predict more jumps (and so hopefully have a different predictive density to the constant λ process) but it becomes unfeasible to perform the Monte Carlo integration (of equation (5.1)) accurately for many "unseen" data points.

a) FX rates

Predictive Density	Constant λ	Stochastic λ	
		prior $r_p = 1.0$	prior $r_p = 0.5$
Australian Dollar	-196.8 (-207.9, -187.5)	-196.8 (-207.5, -187.6)	-192.3 (-199.3, -185.4)
Brazilian Real	-117.9 (-126.3, -105.5)	-94.5 (-103.8, -91.9)	-95.5 (-108.6, -89.4)
British Pound	-225.3 (-242.1, -211.5)	-224.8 (-242.8, -209.4)	-217.6 (-234.4, -209.0)
Canadian Dollar	-223.9 (-226.4, -222.3)	-229.6 (-271.9, -223.0)	-238.8 (-284.1, -229.3)
Colombian Peso	-194.8 (-250.3, -166.2)	-205.9 (-263.6, -168.1)	-173.6 (-173.9, -173.6)
Danish Krone	-215.3 (-238.9, -199.5)	-197.4 (-197.6, -195.4)	-204.3 (-214.2, -196.1)
Euro	-210.3 (-221.5, -200.1)	-209.2 (-220.0, -199.4)	-207.9 (-221.3, -198.7)
Iceland Krona	-208.7 (-247.7, -182.2)	-188.1 (-207.6, -175.5)	-181.1 (-181.6, -180.8)
Japanese Yen	-154.5 (-159.0, -148.1)	-154.7 (-159.3, -148.4)	-154.6 (-159.5, -148.6)
Moroccan Dirham	-232.1 (-237.9, -229.6)	-229.9 (-231.1, -229.2)	-229.8 (-230.5, -229.3)
Singapore Dollar	-183.8 (-205.9, -171.0)	-181.5 (-198.2, -170.7)	-180.9 (-196.9, -169.7)
Thai Baht	-178.3 (-185.7, -170.9)	-177.0 (-182.5, -170.7)	-182.5 (-189.0, -174.8)

Table 5.17: Summaries of predictive densities over 80 "unseen" data points of models with constant and stochastic λ processes for FX data.

For both priors, the stochastic λ process has a larger predictive density than the constant λ model for ten of the twelve data sets. For eight of the twelve data sets, the prior $r_p = 0.5$ has a larger predictive density than that of $r_p = 1.0$. The stochastic λ process was a methodological improvement of the BNS SV models and we have verified that this can predict "unseen" data more accurately than the constant λ model for FX data. We will not focus on how to pick r_p. The posterior distribution of $N_2 = 0$ for FX rates are summarised in Table 5.18 for prior $r_p = 0.5$.

Predictive Density	Posterior $P(N_2 = 0)$	Posterior mode
Australian Dollar	0.000	$P(N_2 = 2) = 0.82$
Brazilian Real	0.000	$P(N_2 = 1) = 0.57$
British Pound	0.000	$P(N_2 = 2) = 0.81$
Canadian Dollar	0.000	$P(N_2 = 2) = 0.71$
Colombian Peso	0.000	$P(N_2 = 1) = 0.93$
Danish Krone	0.000	$P(N_2 = 2) = 0.86$
Euro	0.003	$P(N_2 = 2) = 0.49$
Iceland Krona	0.000	$P(N_2 = 2) = 0.56$
Japanese Yen	0.000	$P(N_2 = 1) = 0.48$
Moroccan Dirham	0.000	$P(N_2 = 2) = 0.78$
Singapore Dollar	0.022	$P(N_2 = 2) = 0.28$
Thai Baht	0.026	$P(N_2 = 1) = 0.69$

Table 5.18: Posterior probability $P(N_2 = 0)$ for FX data.

Table 5.18 provides further evidence that the stochastic λ process is favourable to the constant λ model for FX data as there is very little support in the posterior for $N_2 = 0$. For many of the data sets, the posterior is concentrated almost entirely at one N_2 value.

b) Shares

Predictive Density	Constant λ	Stochastic λ	
		prior $r_p = 1.0$	prior $r_p = 0.5$
British Airways PLC	-168.5 (-170.0, -166.1)	-167.4 (-170.3, -164.6)	-169.6 (-172.8, -165.6)
Citigroup Inc	-112.1 (-115.9, -109.6)	-111.5 (-114.8, -109.8)	-111.6 (-115.7, -109.9)
Coca-Cola Co	-137.6 (-141.7, -135.3)	-137.1 (-141.2, -135.5)	-137.3 (-141.5, -135.6)
General Motors Corp	-160.4 (-161.9, -158.9)	-161.3 (-162.4, -159.7)	-161.2 (-163.2, -159.5)
HJ Heinz Co	-131.6 (-135.5, -129.1)	-131.7 (-135.8, -128.5)	-131.6 (-137.5, -130.2)
Host Marriott Corp	-143.0 (-145.6, -140.7)	-144.5 (-147.8, -141.4)	-144.6 (-147.5, -141.3)
JPM Chase & Co	-105.1 (-112.3, -99.2)	-94.5 (-98.1, -93.0)	-95.1 (-97.5, -93.5)
Kellogg Co	-124.8 (-127.9, -122.9)	-125.9 (-131.2, -123.5)	-126.4 (-130.3, -123.3)
McDonald's Corp	-164.0 (-167.6, -161.3)	-170.8 (-171.5, -165.7)	-170.3 (-171.9, -162.8)
Microsoft	-93.9 (-101.9, -86.7)	-78.2 (-99.4, -76.7)	-77.0 (-78.1, -76.6)
Procter&Gamble Co	-113.9 (-115.1, -112.8)	-112.1 (-113.0, -111.7)	-112.1 (-112.9, -111.5)
S&P 500	-138.7 (-141.7, -137.1)	-138.2 (-142.7, -136.9)	-137.6 (-139.6, -136.6)
Textron Inc	-150.9 (-152.8, -149.6)	-150.6 (-152.0, -149.7)	-150.8 (-152.7, -149.6)
Time Warner Inc	-115.5 (-119.0, -114.0)	-115.7 (-119.5, -114.3)	-115.9 (-123.2, -114.5)

Table 5.19: Summaries of predictive densities over 80 "unseen" data points of models with constant and stochastic λ processes for share data.

For eight of the fourteen data sets, the stochastic λ model has a larger predictive density for $r_p = 1.0$, whilst for $r_p = 0.5$ the predictive density is larger than the constant λ model in seven of the fourteen data sets. It is again not clear if the generalised model is warranted by the share data. The posterior distribution of $N_2 = 0$ for shares are summarised in Table 5.20 for prior $r_p = 0.5$.

Predictive Density	Posterior $P(N_2 = 0)$	Posterior mode
British Airways PLC	0.010	$P(N_2 = 1) = 0.65$
Citigroup Inc	0.000	$P(N_2 = 2) = 0.51$
Coca-Cola Co	0.000	$P(N_2 = 1) = 0.72$
General Motors Corp	0.032	$P(N_2 = 1) = 0.47$
HJ Heinz Co	0.000	$P(N_2 = 1) = 0.62$
Host Marriott Corp	0.236	$P(N_2 = 1) = 0.44$
JPM Chase & Co	0.000	$P(N_2 = 1) = 0.53$
Kellogg Co	0.000	$P(N_2 = 2) = 0.39$
McDonald's Corp	0.039	$P(N_2 = 1) = 0.59$
Microsoft	0.000	$P(N_2 = 3) = 0.88$
Procter&Gamble Co	0.000	$P(N_2 = 2) = 0.47$
S&P 500	0.093	$P(N_2 = 1) = 0.39$
Textron Inc	0.214	$P(N_2 = 1) = 0.53$
Time Warner Inc	0.050	$P(N_2 = 1) = 0.50$

Table 5.20: Posterior probability $P(N_2 = 0)$ for share data.

Table 5.20 suggests that the stochastic λ process is favourable to the constant λ model for shares as there is very little support in the posterior for $N_2 = 0$, apart from the Host Marriott Corp and Textron Inc data sets. For most data sets, the posterior is not as concentrated at one individual N_2 value as it is for FX rates.

Comments on results for both sectors

For both FX and share data, it is not possible to distinguish between the predictive densities of the stochastic and constant λ models for 20 "unseen" data points. For predictive densities over 80 "unseen" data points, the stochastic λ model has a larger predictive density than the constant λ model for FX data. For share data, the results are mixed and it is unclear if the generalised model or original model is preferable.

5.3.4 Long-memory: Predictive densities over 20 and 80 "unseen" data points

a) FX rates

Predictive Density	Brownian Motion	Approximate fBm	Posterior for H
Australian Dollar	-31.6 (-32.6, -30.8)	-38.4 (-39.4, -36.8)	0.510 (0.501, 0.538)
Brazilian Real	-17.4 (-21.1, -11.9)	-18.1 (-21.8, -12.4)	0.508 (0.500, 0.535)
British Pound	-40.8 (-41.6, -39.4)	-41.4 (-42.7, -38.9)	0.510 (0.501, 0.540)
Canadian Dollar	-54.7 (-56.0, -53.7)	-54.7 (-56.1, -53.9)	0.510 (0.500, 0.538)
Colombian Peso	-38.0 (-40.2, -35.0)	-38.2 (-40.5, -35.2)	0.515 (0.503, 0.542)
Danish Krone	-43.7 (-44.8, -42.2)	-43.6 (-44.7, -42.2)	0.509 (0.501, 0.534)
Euro	-43.5 (-44.3, -42.5)	-43.5 (-44.4, -42.5)	0.507 (0.500, 0.527)
Iceland Krona	-31.3 (-34.8, -26.7)	-31.2 (-34.9, -26.5)	0.531 (0.504, 0.571)
Japanese Yen	-39.5 (-40.7, -37.5)	-39.5 (-40.7, -37.5)	0.511 (0.501, 0.540)
Moroccan Dirham	-59.0 (-60.0, -58.2)	-60.9 (-68.3, -57.5)	0.506 (0.501, 0.524)
Singapore Dollar	-45.6 (-46.9, -44.3)	-45.5 (-46.7, -44.2)	0.505 (0.500, 0.531)
Thai Baht	-31.0 (-32.2, -29.3)	-31.1 (-32.2, -29.4)	0.506 (0.501, 0.527)

Table 5.21: Summaries of predictive densities over 20 "unseen" data points of models with Brownian and approximate fractional Brownian motion for FX data.

For each data set, the posterior distribution of H does not deviate far from $H = 0.5$ and this suggests that if long-memory is required in the share equation, it is not very strong. The performance of the two models is very similar with and without long-memory and it is not possible to decide which model should be used to predict "unseen" data. As the posterior for H is concentrated near $H = 0.5$, histograms of the posterior are examined. Figure 5.6 shows posterior histograms of the Hurst parameter for the first three FX data sets.

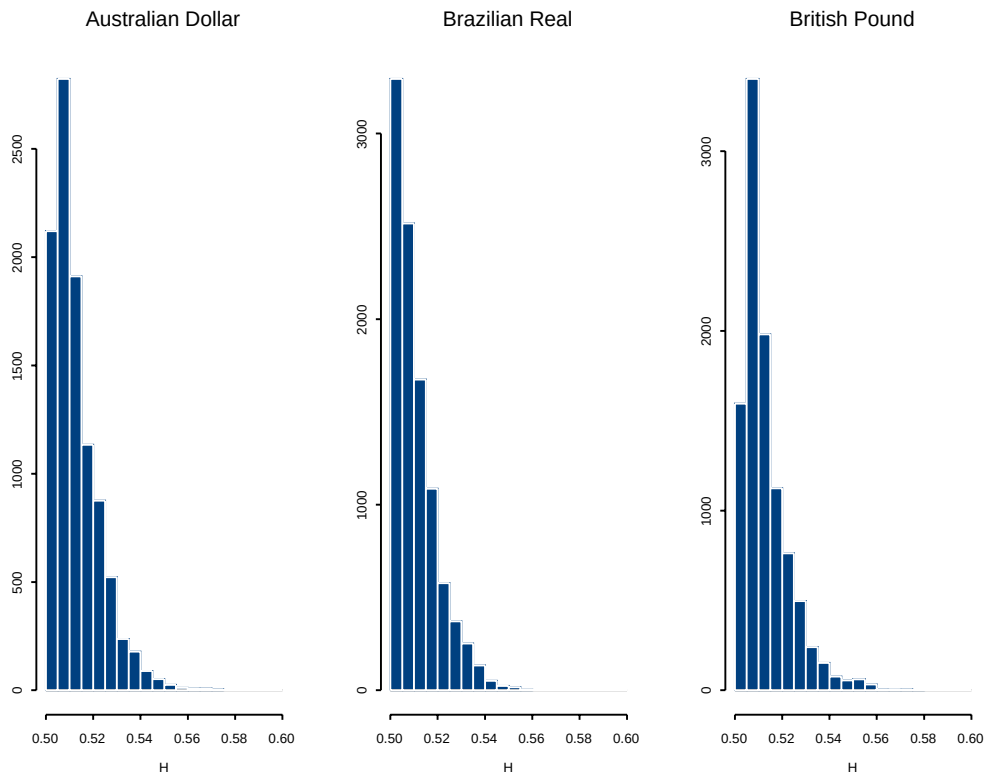


Figure 5.6: Posterior histograms of the Hurst parameter for Australian Dollar, Brazilian Real and British Pound vs US Dollar data sets.

Figure 5.6 suggests that the approximate fBm model is not required for the Brazilian Real data set as the support is maximum at $H = 0.5$ (compare this histogram with Figure 4.13). For the Australian Dollar and British Pound data sets, although the posterior suggests H is small, the mode of the posterior is not at $H = 0.5$ and this suggests that the long-memory model may be required, even though any long-memory should be weak. Posteriors for the remaining data sets which do not have a mode at $H = 0.5$ are given in Figure 5.7.

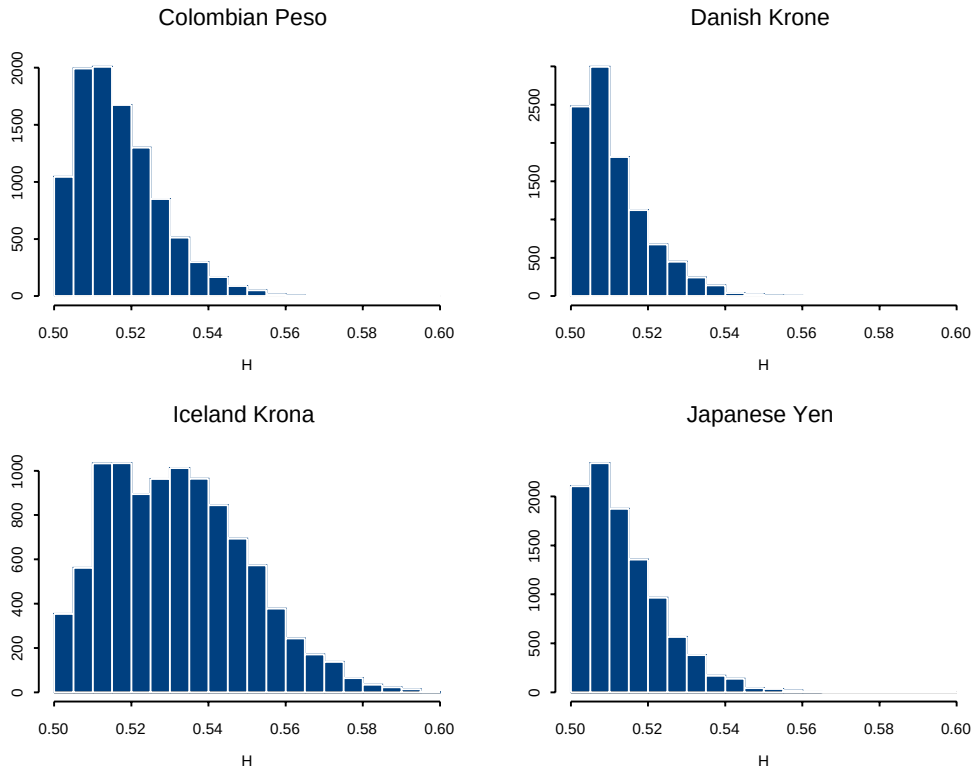


Figure 5.7: Posterior histograms of the Hurst parameter for Colombian Peso, Danish Krone, Iceland Krona and Japanese Yen vs US Dollar data sets.

For these data sets, there is evidence that the fBm model is preferable to the standard Brownian motion model. Posteriors for the remaining six FX data sets were similar to that of the Brazilian Real and do not require the fBm approximation. Boxplots are given of the correlation of the MVN approximation to fBm for the Brazilian Real and Iceland Krona data sets.

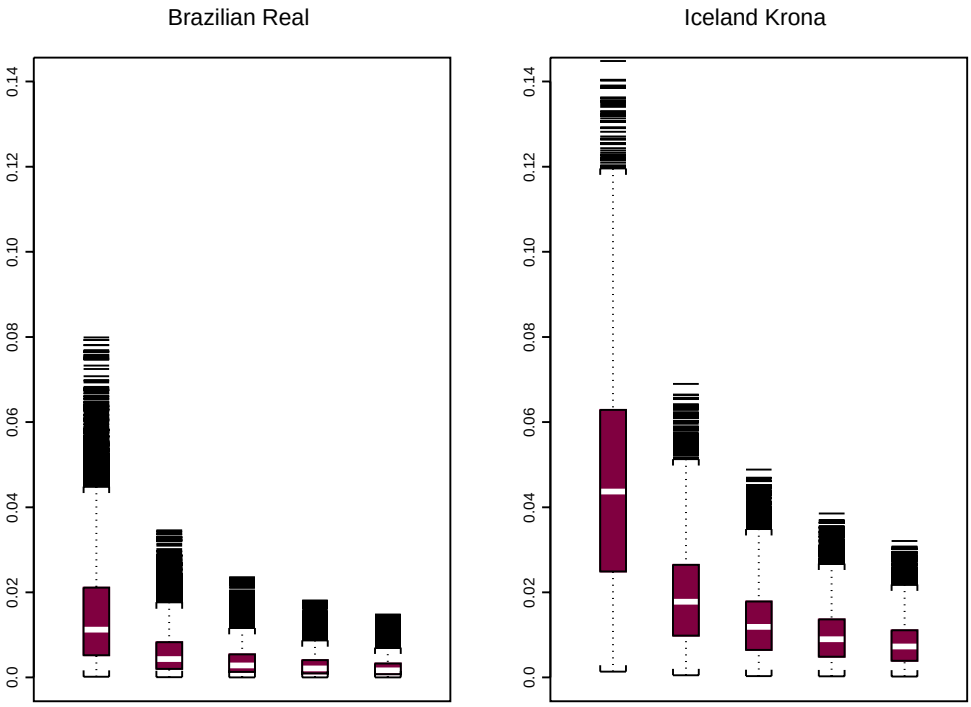


Figure 5.8: Boxplots of the correlation for lags one to five of the MVN approximation to fBm for the Brazilian Real and Iceland Krona vs US Dollar data sets.

Figure 5.8 shows the decay in the correlation of the fBm approximation for the Iceland Krona is slower than that of the Brazilian Real.

b) Shares

Predictive Density	Brownian Motion	Approximate fBm	Posterior for H
British Airways PLC	-34.0 (-34.8, -33.3)	-33.9 (-34.6, -33.0)	0.512 (0.501, 0.548)
Citigroup Inc	-28.8 (-30.0, -28.0)	-29.1 (-30.4, -28.2)	0.505 (0.501, 0.523)
Coca-Cola Co	-34.9 (-36.3, -33.9)	-34.7 (-36.0, -33.6)	0.523 (0.502, 0.563)
General Motors Corp	-45.9 (-46.3, -45.5)	-45.8 (-46.2, -45.3)	0.512 (0.500, 0.542)
HJ Heinz Co	-26.1 (-28.6, -24.5)	-26.2 (-30.2, -24.3)	0.503 (0.500, 0.517)
Host Marriott Corp	-36.1 (-36.9, -35.4)	-36.1 (-36.8, -35.4)	0.512 (0.503, 0.551)
JP Morgan Chase & Co	-24.4 (-26.1, -23.0)	-24.3 (-25.9, -23.1)	0.510 (0.501, 0.540)
Kellogg Co	-36.3 (-37.0, -35.7)	-36.4 (-37.3, -35.8)	0.504 (0.501, 0.514)
McDonald's Corp	-48.8 (-49.7, -48.3)	-49.0 (-50.0, -48.4)	0.510 (0.502, 0.535)
Microsoft	-22.4 (-24.4, -21.2)	-22.6 (-24.1, -21.4)	0.506 (0.501, 0.529)
Procter & Gamble Co	-28.6 (-29.2, -28.1)	-28.8 (-29.8, -28.1)	0.508 (0.500, 0.528)
S&P 500	-33.6 (-34.9, -32.5)	-33.8 (-35.0, -32.8)	0.507 (0.500, 0.530)
Textron Inc	-42.6 (-43.0, -42.1)	-42.6 (-43.1, -42.2)	0.518 (0.501, 0.554)
Time Warner Inc	-32.1 (-33.2, -31.5)	-32.1 (-33.4, -31.4)	0.510 (0.502, 0.535)

Table 5.22: Summaries of predictive densities over 20 "unseen" data points of models with Brownian and approximate fractional Brownian motion for share data.

The results are similar to the results for the FX data. If the data requires long-memory in the share equation, it is only weak long memory. As the posterior for H is concentrated near $H = 0.5$, histograms of the posterior are examined. Figure 5.9 shows posterior histograms of the Hurst parameter for Coca-Cola Co and British Airways PLC.

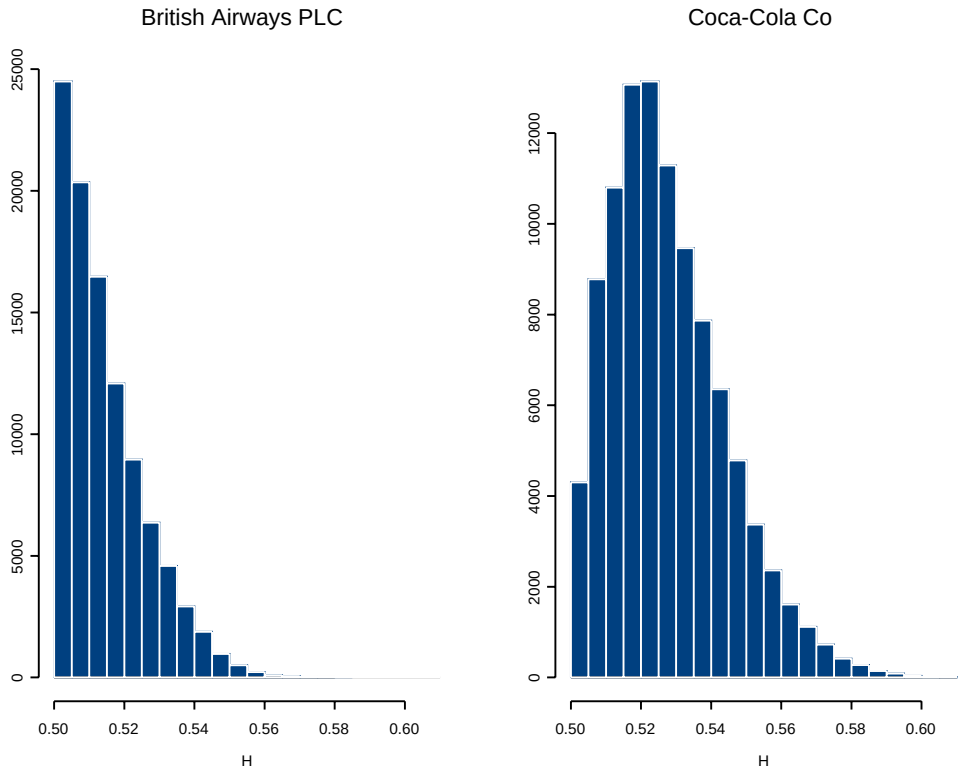


Figure 5.9: Posterior histograms of the Hurst parameter for British Airways PLC and Coca-Cola Co.

Figure 5.9 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 a lot of support for $H = 0.5$, suggesting that the long-memory model is not necessary for this data set. For the twelve remaining share 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 to the long-memory model.

Comments on results for both sectors

Results between the two sectors are similar; posteriors for H are close to $H = 0.5$, suggesting standard Brownian motion may be suitable (ignoring posterior histograms of H). However, predictive density estimates do not suggest a performance advantage using either model. Before the tests were carried out, it was not known how strong the long-memory parameter would be, so predictive densities were taken over only 20 "unseen" data points (as the volatility model was constructed to improve the Black-Scholes model particularly

over small time periods). As the Hurst parameter appears to be quite small, the predictive densities are again calculated but over 80 "unseen" data points and $10,000$ volatilities were used in equation (5.1) to estimate $\widehat{p}\left(\underline{Y}^{'}|\underline{y}\right)$ (i.e. B_2 was set equal to $10,000$). More points were needed than for 20 "unseen" points as it is a higher dimensional integral. The predictive densities over 80 "unseen" data points are summarised in Tables 5.23 and 5.24.

Predictive Density	Brownian Motion	Approximate fBm
Australian Dollar	-196.8 (-207.9, -187.5)	-194.7 (-203.5, -186.4)
Brazilian Real	-117.9 (-126.3, -105.5)	-117.1 (-126.1, -104.6)
British Pound	-225.3 (-242.1, -211.5)	-225.4 (-241.8, -210.8)
Canadian Dollar	-223.9 (-226.4, -222.3)	-224.3 (-227.0, -222.4)
Colombian Peso	-194.8 (-250.3, -166.2)	-197.9 (-241.4, -167.3)
Danish Krone	-215.3 (-238.9, -199.5)	-212.3 (-230.5, -198.1)
Euro	-210.3 (-221.5, -200.1)	-210.3 (-223.2, -199.7)
Iceland Krona	-208.7 (-247.7, -182.2)	-199.3 (-231.9, -178.3)
Japanese Yen	-154.5 (-159.0, -148.1)	-154.8 (-159.7, -148.5)
Moroccan Dirham	-232.1 (-237.9, -229.6)	-282.4 (-312.8, -254.3)
Singapore Dollar	-183.8 (-205.9, -171.0)	-182.1 (-195.8, -170.8)
Thai Baht	-178.3 (-185.7, -170.9)	-178.8 (-186.0, -171.4)

Table 5.23: Summaries of predictive densities over 80 "unseen" data points of models with Brownian and approximate fractional Brownian motion for FX data.

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

Table 5.24: Summaries of predictive densities over 80 "unseen" data points of models with Brownian and approximate fractional Brownian motion for share data.

Again, there is no clear performance advantage for the long-memory model. We therefore conclude from the posteriors of H (see Figures 5.6, 5.7 and 5.9) that the generalisation to fractional Brownian motion may be warranted for some of the FX data sets (six out of twelve data sets) and very rarely for the asset data sets (one in fourteen of the data sets). For all data sets, the long-memory is weak (even if the posterior does not support $H = 0.5$) and this makes it difficult to distinguish between the models based on their empirical performance. It is currently not feasible (computationally) to consider much larger data sets and predictive densities over a longer period of "unseen" data, though as computers become more powerful, it would be interesting to test if the fBm model has a larger predictive density for the data sets where $H = 0.5$ is not supported.

5.3.5 Leverage: Predictive densities over 20 "unseen" data points

a) FX rates

Predictive Density	Without Leverage	With Leverage	Posterior for ρ
Australian Dollar	-31.6 (-32.6, -30.8)	-38.3 (-39.4, -36.7)	-5.49 (-16.9, 1.74)
Brazilian Real	-10.9 (-14.6, -5.38)	-18.09 (-21.8, -12.6)	0.704 (-1.17, 2.21)
British Pound	-40.8 (-41.6, -39.4)	-41.2 (-42.7, -38.7)	-5.98 (-23.5, 3.70)
Canadian Dollar	-54.7 (-56.0, -53.7)	-54.8 (-55.9 -54.0)	-9.61 (-24.8, 2.55)
Colombian Peso	-38.0 (-40.2, -35.0)	-57.5 (-66.1, -39.4)	5.60 (4.81, 21.5)
Danish Krone	-43.7 (-44.8, -42.2)	-43.2 (-44.3, -41.8)	-9.97 (-16.6, -3.48)
Euro	-43.5 (-44.3, -42.5)	-43.4 (-44.3, -42.4)	-1.95 (-11.4, 7.16)
Iceland Krona	-31.3 (-34.8, -26.7)	-31.2 (-34.7, -26.5)	-8.18 (-15.1, -3.38)
Japanese Yen	-39.5 (-40.7, -37.5)	-39.5 (-40.6, -37.5)	-1.32 (-12.7, 10.1)
Moroccan Dirham	-59.0 (-60.0, -58.2)	-60.4 (-62.6, -58.8)	15.2 (10.2, 23.3)
Singapore Dollar	-45.6 (-46.9, -44.3)	-45.7 (-46.9, -44.4)	11.4 (1.14, 24.9)
Thai Baht	-31.0 (-32.2, -29.3)	-31.2 (-32.3, -29.6)	4.72 (-9.14, 19.4)

Table 5.25: Summaries of predictive densities of models with and without leverage for FX data.

First note that only two of the data sets have a 95% credible interval for the posterior of the leverage parameter which is strictly negative and here the predictive density is larger than the model without leverage. Three of the data sets have posteriors for ρ which are strictly positive (contrary to what might be expected) and here the model with leverage has a smaller predictive density than the model without leverage.

For the remaining data sets (where the credible interval for ρ contains zero), six out of eight data sets have a larger predictive density for the model without leverage. When the credible interval is strictly negative, the model with leverage predicts "unseen" data better than without the leverage parameter. When the credible interval is strictly positive, the leverage model predicts "unseen" data worse than the model without the leverage parameter. For the FX data sets considered here, most of the time the credible interval contains zero and there is little evidence to suggest the requirement of a leverage parameter. This is in agreement with accepted wisdom.

b) Shares

Predictive Density	Without Leverage	With Leverage	Posterior for ρ
British Airways PLC	-34.0 (-34.8, -33.3)	-33.9 (-34.7, -33.4)	-1.64 (-2.54, -0.901)
Citigroup Inc	-28.8 (-30.0, -28.0)	-28.6 (-30.1, -27.4)	-4.33 (-5.73, -3.11)
Coca-Cola Co	-34.9 (-36.3, -33.9)	-34.9 (-36.3, -33.8)	-4.77 (-8.53, -1.81)
General Motors Corp	-45.9 (-46.3, -45.5)	-46.3 (-46.8, -45.7)	-3.40 (-6.04, -0.244)
HJ Heinz Co	-26.1 (-28.6, -24.5)	-27.7 (-31.1, -25.5)	-5.08 (-8.98, -3.12)
Host Marriott Corp	-36.1 (-36.9, -35.4)	-37.9 (-38.6, -36.3)	-5.35 (-7.73, -3.39)
JP Morgan Chase & Co	-24.4 (-26.1, -23.0)	-23.9 (-26.0, -22.4)	-4.59 (-6.57, -3.08)
Kellogg Co	-36.3 (-37.0, -35.7)	-36.1 (-37.0, -35.3)	-1.01 (-3.58, 1.86)
McDonald's Corp	-48.8 (-49.7, -48.3)	-50.1 (-51.3, -48.9)	-3.82 (-9.02, -1.38)
Microsoft	-22.4 (-24.4, -21.2)	-21.3 (-24.0, -20.2)	-3.30 (-5.16, -0.313)
Procter & Gamble Co	-28.6 (-29.2, -28.1)	-28.2 (-29.3, -27.7)	-4.20 (-5.59, -3.40)
S&P 500	-33.6 (-34.9, -32.5)	-33.3 (-34.6, -32.4)	-6.23 (-14.1, -2.34)
Textron Inc	-42.6 (-43.0, -42.1)	-42.6 (-42.9, -42.3)	-3.90 (-7.62, -0.507)
Time Warner Inc	-32.1 (-33.2, -31.5)	-31.7 (-33.0, -30.9)	-1.88 (-3.26, -0.961)

Table 5.26: Summaries of predictive densities of models with and without leverage for share data.

Here thirteen of the fourteen data sets have a posterior for ρ which is strictly negative and suggests that leverage is present in the share data sets. The predictive density is larger for the model with leverage for ten of the fourteen data sets and there is a performance advantage of using the leverage model for share data.

Comments on results for both sectors

For both sectors, if the posterior for ρ was strictly negative, generally, the leverage model had a large predictive density. This was almost always the case for share data and rarely for FX data. A useful pricing strategy could be to assume leverage is present (without assuming $\rho < 0$) and examine the posterior for ρ. If this is strictly negative, continue pricing using leverage, otherwise price without leverage.

5.3.6 Results summary

The results for all the model generalisations are summarised:-

(1) Predictive density results for FX data are mixed and it is not clear which marginal distribution is most suitable.

(2) For share data the *Generalised Inverse Gaussian* and *Inverse Gamma* distributions have the largest predictive density.

(3) For both FX and share data the *Inverse Gamma* distribution predicts the actual discounted payoff most accurately and is the marginal of choice for option pricing.

(4) The stochastic λ process has a larger predictive density than the constant λ model for FX rates, whilst for share data the performance of the models is similar.

(5) For the fBm model, all data sets either do not require the generalisation to the long-memory model, or require weak long-memory with a Hurst parameter near to $H = 0.5$. There is evidence to suggest that half of the FX data sets and one of the share data sets require the long-memory model. Further testing, possibly using predictive densities over large "unseen" data sets, is required to verify this, though this is not feasible with current computing power.

(6) There is strong evidence that the leverage model of Barndorff-Nielsen and Shephard (2001a) is required for share data sets and not required for FX data sets.

Chapter 6

Non Ornstein-Uhlenbeck Lévy processes

This chapter describes some recent developments of potential continuous time stochastic volatility models. Some of these models are generalisations of the BNS SV models and have a more flexible correlation structure than the exponential decay of the models investigated in Chapters 4 and 5. The models of this chapter are new and there is little in the literature on them, particularly when they are used as stochastic volatility models. These models were introduced in Wolpert and Taqqu (2004) to model the Telecom process and important results from this paper are included here for completeness. To the best of our knowledge the simulation (and therefore inference) for such models has not been performed and the models have yet to be used in a stochastic volatility setting. The purpose of this chapter is to introduce a new class of stochastic volatility models and some of its properties, along with methods to simulate from them. These SV models do not need a superposition of processes to give long-memory. Related work can be found in Brockwell (2001), where models similar to those in Wolpert and Taqqu (2004) are investigated.

The main original contribution of this chapter is showing how to simulate from such models, using the series representation of Barndorff-Nielsen and Shephard (2000) and we introduce a volatility process whose correlation structure decays asymptotically like $t^{-\lambda}$, where $\lambda > 1$ is a parameter of the model.

6.1 Ornstein-Uhlenbeck processes: Alternative series representation

The OU process defined previously is

$$d\sigma^2(t) = -\lambda\sigma^2(t)\,dt + dz(\lambda t), \tag{6.1}$$

which has solution

$$\begin{aligned}
\sigma^2(t) &= \int_{-\infty}^{t} f_2(\lambda, t, s)\,dz(\lambda s) && (6.2)\\
&= \int_0^{\infty} f_1(\lambda, t, s)\,dz(\lambda s) + \int_0^{t} f_2(\lambda, t, s)\,dz(\lambda s) && (6.3)\\
&= e^{-\lambda t}\sigma^2(0) + e^{-\lambda t}\int_0^{t} e^{\lambda s} dz(\lambda s), && (6.4)
\end{aligned}$$

where the two Lévy processes of equation (6.3) are independent copies of each other (i.e. series representations for the stochastic integrals use independent realisations from the same Lévy process) and

$$f_1(\lambda, t, s) = e^{-\lambda(t+s)}$$

and

$$f_2(\lambda, t, s) = e^{-\lambda(t-s)}.$$

It is already known that for these f_1 and f_2, the process has correlation structure

$$corr\left(\sigma^2(t), \sigma^2(t+j)\right) = \exp(-\lambda j)$$

and that $\sigma^2(t)$ is stationary and positive for a wide range of functions f_1 and f_2 (see Barndorff-Nielsen and Shephard (2001b)). Barndorff-Nielsen and Shephard (2001b) mention using models with more general functions f_1 and f_2 and decide to concentrate on OU models, where f_1 and f_2 are as described above.

The timing of the BDLP, $dz(\lambda s)$, in equation (6.1) was chosen so that λ does not influence the marginal distribution of $\sigma^2(t)$. Instead of considering equation (6.3), consider

$$\begin{aligned}
\sigma^2(t) &= \int_0^{\infty} f_1(\lambda, t, s)\,dz(s) + \int_0^{t} f_2(\lambda, t, s)\,dz(s) && (6.5)\\
&= I_{1,t} + I_{2,t}.
\end{aligned}$$

This is the same representation as used in Wolpert and Taqqu (2004). Unlike the BNS SV models, the rate of jumps of the Lévy process is not controlled by λ.

Simulation from the OU process is relatively straightforward because the time dependent term of f_1 and f_2 can be removed from the stochastic integrals of equation

(6.3) and this allows $I_{1,t}$ to be written in terms of the volatility at time zero, $\sigma^2(0)$, in equation (6.4). If the OU equation is generalised, so integrands are not of the form $f_1(t,s) = g_{1,1}(t)\, g_{1,2}(s)$, then $\sigma^2(t)$ can no longer be expressed in terms of $\sigma^2(0)$. For general f_1 and f_2, it is also not possible to separate the t and s terms in $I_{2,t}$. This makes simulating from such models more complicated than the original BNS SV OU models.

We now describe how to sample from such models. Consider the approximation for $I_{1,t}$,

$$I_{1,t} \approx \int_0^d f_1(\lambda, t, s)\, dz(s),$$

where d is large enough so the approximation is sufficiently accurate. The Barndorff-Nielsen and Shephard (2000) series representation is then

$$\int_0^d f_1(\lambda, t, s)\, dz(s) \overset{\mathfrak{L}}{=} \sum_{j=0}^{n_{1,j}} W^{-1}(a_j/d)\, f_1(\lambda, t, r_j), \tag{6.6}$$

where $da_{1,c}$ is the value at which the Poisson point process (order statistics of uniform random variables) is truncated, $n_{1,j} \sim Po(da_{1,c})$, a_j are the order statistics of $n_{1,j}$ $U(0, da_{1,c})$ random variables, $r_j \overset{iid}{\sim} U(0,d)$, all variables are independent and W^{-1} is the Inverse Tail Mass function as defined previously. For every t, the same Poisson points, a_j, and uniforms, r_j, are used and this induces the correlation in $I_{1,t}$, so $I_{1,t} = e^{-\lambda t}\sigma^2(0)$ for the OU case. This allows us to sample from $I_{1,t}$.

For the finite integral, the situation is more complex. Previously a series representation was used, based on independent Poisson point processes and uniforms and this was possible because the volatility could be written in terms of the previous volatility and a stochastic integral (independent of previous stochastic integrals). Further, the stochastic integrals were unaltered by t. For more general functions than $f_2(\lambda, t, s) = e^{-\lambda(t-s)}$, it is not possible to write the volatility in terms of previous volatilities, though we are able to express the integral as a summation of integrals on disjoint domains and then use independent series representations for these integrals. The second integral at time $t-1$ is

$$I_{2,t-1} = \int_0^{t-1} f_2(\lambda, t-1, s)\, dz(s)$$

and now consider $I_{2,t}|I_{2,t-1}$

$$I_{2,t} = \int_0^{t-1} f_2(\lambda, t, s)\, dz(s) + \int_{t-1}^{t} f_2(\lambda, t, s)\, dz(s). \tag{6.7}$$

The domains of these two integrals are disjoint and so any realisations from these integrals use independent series representations. Equation (6.7) can be rewritten as

$$I_{2,t} = \sum_{j=0}^{t-2} \int_j^{j+1} f_2(\lambda, t, s)\, dz(s) + \int_{t-1}^{t} f_2(\lambda, t, s)\, dz(s)$$

but the integrals of the summation are also disjoint, so by the independent increments assumption,

$$\begin{aligned} I_{2,t} &\overset{\mathfrak{L}}{=} \sum_{j=0}^{t-2} \int_0^1 f_2(\lambda, t, s+j)\, dz(s) + \int_0^1 f_2(\lambda, t, s+t-1)\, dz(s) \\ &= \sum_{j=0}^{t-1} \int_0^1 f_2(\lambda, t, s+j)\, dz(s), \end{aligned}$$

where integral terms in the sum are all with respect to independent realisations of the BDLP (as they represent partitions of the integrals in equation (6.7)). This gives t disjoint independent integrals and these can be simulated using the series representation derived in Barndorff-Nielsen and Shephard (2000) and given in equation (4.8). If the series are again truncated by discarding all Poisson points which are greater than $a_{2,c}$, then the series representation is

$$I_{2,t} \overset{\mathfrak{L}}{=} \sum_{j=0}^{t-1} \sum_{i=0}^{n_{2,j}} W^{-1}(a_{2,j,i})\, f_2(\lambda, t, r_{2,j,i}+j), \tag{6.8}$$

where $n_{2,j} \sim Po(a_{2,c})$, $a_{2,j}$ are the order statistics of $n_{2,j}$ $U(0, a_{2,c})$ random variables, $r_{2,j,i} \overset{iid}{\sim} U(0,1)$, all variables are independent of each other and $W^{-1}()$ is the Inverse Tail Mass function as defined previously. For the OU process, simulating using these series representations gives the properties of $\sigma^2(t)$ that were discussed in Chapters 3 and 4 (this is illustrated in Figure 6.1). We are now able to simulate from processes of the form

$$\sigma^2(t) = \int_0^\infty f_1(\lambda, t, s)\, dz(s) + \int_0^t f_2(\lambda, t, s)\, dz(s)$$

for general f_1 and f_2. Note that simulating from the instantaneous volatility using the series representation of equation (6.8) is an order t^2 algorithm, unlike the series representation that was used for the OU process in Chapter 4, which was order t. We now consider what forms of these functions should be examined.

6.2 Continuous time SV models driven by Lévy processes

Continue assuming that we wish to sample from stochastic integrals with respect to the homogeneous BDLP, $z(t)$, where $z(1)$ has Lévy measure

$$w(x) = -u(x) - xu(x),$$

and $u(x)$ is the Lévy measure of the marginal distribution of the BNS SV model with the same BDLP. We will focus on marginal distributions on the positive real line, so z is a subordinator. The Lévy-Khintchine formula for $z(1)$ is

$$\log\left[E\left[e^{i\theta z(1)}\right]\right] = \int_{-\infty}^{\infty} \left(e^{i\theta x} - 1\right) w(x)\, dx,$$

and $w(x)$ is zero for $x \leq 0$.

Definition 19 *A stochastic process, X_t, is a continuous* ***moving average*** *process if it can be expressed as*

$$X_t = \int_0^t f(t,s)\, dz(s),$$

where $z(s)$ is a Lévy process.

Definition 20 *The function $G(s)$ is* ***non-anticipating*** *with respect to $dz(s)$ if $G(s)$ cannot be used to predict future movement in $dz(s)$. The process*

$$X_t = \int_{t_0}^{t_n} G(s)\, dz(s)$$

is then also non-anticipating.

Consider non-anticipating moving average processes of the form

$$\sigma^2(t) = \int_{-\infty}^t h_1(t-s)\, dz(s), \tag{6.9}$$

where the Lévy measure of $z(1)$ is $w(x)$ and $h_1(t-s) \geq 0$ for $s < t$ (so $\sigma^2(t)$ has only positive jumps). Assume that

$$\int_{-\infty}^t h_1(t-s)\, ds < \infty, \tag{6.10}$$

so that the stochastic integral in equation (6.9) exists. Ignoring the timing of the BDLP, this is a generalisation of the solution given in equation (6.2). Therefore

$$\sigma^2(t) = \int_{-\infty}^{\infty} h(t-s)\, dz(s), \tag{6.11}$$

where

$$h(x) = \begin{cases} 0 & x < 0 \\ h_1(x) & x \geq 0 \end{cases}.$$

For models of the form of equation (6.11), the negative of the characteristic exponent is

$$\begin{aligned} \log\left\{E\left[e^{i\theta\sigma^2(t)}\right]\right\} &= \log\left\{E\left[\exp\left(i\theta\int_{-\infty}^{\infty} h(t-s)\, dz(s)\right)\right]\right\} \\ &= \log\left\{E\left[\exp\left(i\theta\sum_{j=-\infty}^{\infty}\int_{j\Delta}^{(j+1)\Delta} h(t-s)\, dz(s)\right)\right]\right\} \end{aligned}$$

and as $\sigma^2(t)$ is non-anticipative,

$$\begin{aligned}\log\left\{E\left[e^{i\theta\sigma^2(t)}\right]\right\} &= \log\left\{E\left[\exp\left(i\theta\sum_{j=-\infty}^{\infty}h(t-j\Delta)\left(z\left((j+1)\Delta\right)-z(j\Delta)\right)\right)\right]\right\}\\ &= \log\left\{E\left[\prod_{j=-\infty}^{\infty}\exp\left(i\theta h(t-j\Delta)z_j(\Delta)\right)\right]\right\},\end{aligned}$$

where $z_j(t)$ are independent and identical homogeneous Lévy processes with Lévy measure $w(x)$. Then

$$\begin{aligned}\log\left\{E\left[e^{i\theta\sigma^2(t)}\right]\right\} &= \sum_{j=-\infty}^{\infty}E\left[\exp\left(i\theta h(t-j\Delta)z_j(\Delta)\right)\right]\\ &= \sum_{j=-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\exp\left(i\theta h(t-j\Delta)x\right)-1\right\}w(x)\,dx\end{aligned}$$

and letting $\Delta\to 0$ this gives

$$\log\left\{E\left[e^{i\theta\sigma^2(t)}\right]\right\}=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\exp\left(i\theta h(t-s)x\right)-1\right\}w(x)\,dxds \tag{6.12}$$

and

$$\log\left[E\left[e^{i\theta\sigma^2(t)}\right]\right]=\int_{-\infty}^{\infty}\int_{0}^{\infty}\left(e^{i\theta xh(s)}-1\right)w(x)\,dsdx, \tag{6.13}$$

which can be found in Wolpert and Taqqu (2004). The variance and covariance of the process can be calculated by considering the joint characteristic function[1]. Using equation (6.12), the joint characteristic function of $\sigma^2(t)$ and $\sigma^2(0)$ is

$$E\left[e^{i\theta_1\sigma^2(t)+i\theta_2\sigma^2(0)}\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{e^{i\theta_1h(t-s)x+i\theta_2h(-s)x}-1\right\}w(x)\,dxds$$

and then the covariance is

$$\begin{aligned}Cov\left[\sigma^2(t),\sigma^2(0)\right] &= -\left.\frac{\partial^2}{\partial\theta_1\partial\theta_2}\right|_{\theta_1=\theta_2=0}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{e^{i\theta_1h(t-s)x+i\theta_2h(-s)x}-1\right\}w(x)\,dxds\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^2h(t-s)\,h(-s)\,w(x)\,dxds\\ &= \sigma^2\int_{-\infty}^{\infty}h(t+s)\,h(s)\,ds\\ &= \sigma^2\int_{0}^{\infty}h_1(t+s)\,h_1(s)\,ds\end{aligned}$$

where

$$\sigma^2=\int_{-\infty}^{\infty}x^2w(x)\,dx<\infty.$$

[1] I am grateful to Prof R. Wolpert for his enlightening comments.

The variance of the process is therefore

$$\sigma^2 \int_0^\infty h_1^2(s)\, ds, \tag{6.14}$$

which we require to be finite. The correlation at lag t is

$$\rho(t) = \frac{\int_0^\infty h_1(|t|+s)\, h_1(s)\, ds}{\int_0^\infty h_1^2(s)\, ds}. \tag{6.15}$$

By picking suitable functions for $h_1(x)$, we are able to generate from a wide range of distributions and correlation structures which have Lévy measure and correlation structure specified by equations (6.13) and (6.15) respectively. In general, the discretely observed volatility (see equation (1.6)) is not readily available and this makes it difficult to fit SV models of this form using the discretely observed volatility.

Two examples of the flexibility of models of this form are now given, before fractional Ornstein-Uhlenbeck processes are introduced.

6.2.1 Ornstein-Uhlenbeck process

The marginal distribution of the BNS SV OU volatility models is unaltered by the λ parameter. Instead of using this OU process, consider

$$h_1(t-s) = \sqrt{2\lambda} e^{-\lambda(t-s)}$$

in equation (6.9). This is the OU process used in Wolpert and Taqqu (2004) (and given in equation (6.18)), where λ influences the marginal distribution of $\sigma^2(t)$. The correlation is specified by equation (6.15) and is

$$\rho(t) = e^{-\lambda t},$$

as for the BNS SV OU models. The relationship between the marginal and λ is now given, when the BDLPs of Chapter 4 drive the OU process.

Substituting $r = xh(s)$ in equation (6.13) implies the negative of the characteristic exponent is

$$\frac{1}{\lambda} \int_0^\infty \int_0^{x\sqrt{2\lambda}} \left\{e^{itr} - 1\right\} r^{-1} w(x)\, dr dx$$

and swapping the order of integration (taking care with domains of the integrals), this is

$$\frac{1}{\lambda} \int_0^\infty \int_{r/\sqrt{2\lambda}}^\infty \left\{e^{itr} - 1\right\} r^{-1} w(u)\, dx dr$$

and so $\sigma^2(t)$ has Lévy measure

$$\begin{aligned}\frac{1}{\lambda}r^{-1}\int_{r/\sqrt{2\lambda}}^{\infty} w(x)\,dx &= \frac{1}{\lambda}r^{-1}\left[-xu(x)\right]_{r/\sqrt{2\lambda}}^{\infty} \\ &= \frac{1}{\lambda\sqrt{2\lambda}}u\left(r/\sqrt{2\lambda}\right). \end{aligned} \tag{6.16}$$

When the BDLP, which gives a $GIG(\gamma,\nu,\alpha)$ marginal for the BNS SV OU model, is used to drive equation (6.18), using equation (6.16) and equation (2.1), the Lévy measure of $\sigma^2(t)$ is

$$\frac{1}{\lambda}x^{-1}\left[\left\{\frac{1}{2}\int_0^{\infty}\exp\left(-\frac{x\xi}{2\nu^2\sqrt{2\lambda}}\right)g_\gamma(\xi)\,d\xi+\max(0,\gamma)\right\}\exp\left(-\frac{\alpha^2 x}{2\sqrt{2\lambda}}\right)\right].$$

In general, it is not possible to write the distribution of $\sigma^2(t)$ in terms of a GIG distribution because of the complex nature of the integrand. However, to demonstrate these models, we use the BDLP which gives a $Ga(\nu,\alpha)$ ($GIG\left(\nu,0,\sqrt{2\alpha}\right)$) distribution for the BNS SV model, so the integral is zero. Then

$$\sigma^2(t)\sim Ga\left(\frac{\nu}{\lambda},\frac{\alpha}{\sqrt{2\lambda}}\right).$$

This marginal distribution is verified in Figure 6.1, which also demonstrates that the correlation structure is $e^{-\lambda t}$, as given by equation (6.15).

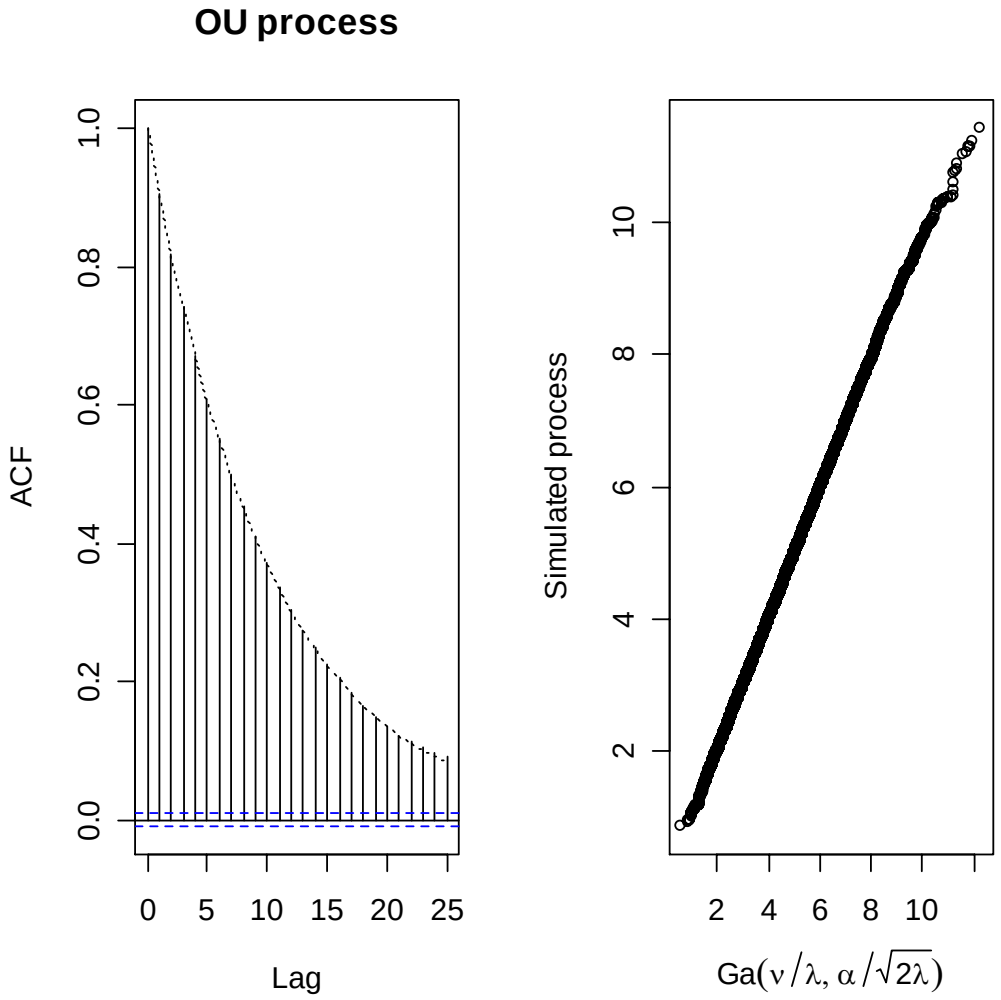


Figure 6.1: ACF of the OU process of Wolpert and Taqqu (2004) for $\lambda = 0.1$ for a $Ga(1,1) - OU$ BDLP using the series representation of Section 6.1.

The simulation results of Figure 6.1 are as the theory suggests. This demonstrates the correct implementation of the series representation of Section 6.1 for the OU process.

When the BDLP, which gives a $TS(\kappa, \nu, \alpha)$ marginal for the BNS SV OU model, is used to drive equation (6.18), using equation (6.16) and equation (4.12), the Lévy measure of $\sigma^2(t)$ is

$$A' r^{-B'-1} e^{-C' x},$$

where

$$A' = \frac{A}{\lambda (2\lambda)^{\kappa/2}}, \quad B' = B \quad \text{and} \quad C' = \frac{C}{\sqrt{2\lambda}}$$

and A, B and C are as defined under equation (4.12). From this it can be shown that

$$\sigma^2(t) \sim TS\left(\kappa, \nu\lambda^{-1}(2\lambda)^{-\kappa/2}, \alpha(2\lambda)^{-\kappa/2}\right)$$

and so the *IG-OU* BDLP generates $\sigma^2(t) \sim IG$ (with different parameters).

6.2.2 Power Decay process

As we have seen in Chapter 3, for observed financial data, it is sometimes argued that the square of the log returns have long-memory or that the correlation of them decays more slowly than exponentially. We now consider a SV model whose correlation decays asymptotically like a power. Let

$$h_1(t-s) = \frac{1}{(\alpha + \beta|t-s|)^{\lambda}} \tag{6.17}$$

in equation (6.9), where $\lambda > 1$ (so that the integral in equation (6.15) exists). We will focus on the case $\alpha = 1$, as other α values do not offer a richer correlation structure, as their effect only rescales the β parameter. Substituting $r = xh(s)$ in equation (6.13) implies the negative of the characteristic exponent is

$$\frac{1}{\beta\lambda}\int_{-\infty}^{\infty}\int_{0}^{1}\left(e^{itr}-1\right)x^{1/\lambda}r^{-(1+1/\lambda)}drw(x)\,dx.$$

Therefore the Lévy measure of $\sigma^2(t)$ is

$$\frac{1}{\beta\lambda}r^{-(1+1/\lambda)}\int_{0}^{1}x^{1/\lambda}w(x)\,dx.$$

Both the Lévy measure and correlation structure of $\sigma^2(t)$ can be expressed in terms of standard numerical functions, though these expressions are complex. For this reason we focus on the cases $\lambda = 1.5$ and $\lambda = 2$, which will be used for simulation purposes later. In the case of the *Ga(ν, α_2)-OU* BDLP, the Lévy measure is

$$\frac{3\nu}{5\alpha_2^{1/3}}e^{-\alpha_2/2}WhittakerM\left(\frac{1}{3}, \frac{5}{6}, \alpha_2\right)r^{-5/3} \qquad \lambda = 1.5$$

$$\frac{\nu}{2\sqrt{\alpha_2}}\left\{\sqrt{\pi}\,\mathrm{erf}\left(\sqrt{\alpha_2}\right) - 2\sqrt{\alpha_2}e^{-\alpha_2}\right\}r^{-3/2} \qquad \lambda = 2,$$

where $WhittakerM(\mu, \nu, z)$ is a standard numerical function. These are limiting cases of the $TS(1/\lambda, \nu_2, \alpha_3)$ distribution as $\alpha_3 \to 0$ (where ν_2 is determined by ν, α_2 and λ). The correlation is specified by equation (6.15) and is

$$\frac{4}{\beta^2}\frac{\left(2+\beta t - 2\sqrt{1+\beta t}\right)}{t^2\sqrt{1+\beta t}} \qquad \lambda = 1.5$$

$$\frac{3}{\beta^3 t^3(1+\beta t)}\left\{2(1+\beta t)\log\left(\frac{1}{1+\beta t}\right) + \beta t(2+\beta t)\right\} \qquad \lambda = 2.$$

The asymptotic decay of the correlation is proportional to

$$\lim_{t\to\infty}\int_0^\infty \frac{1}{\{(1+\beta s)\,(1+\beta\,(t-s))\}^\lambda}ds,$$

truncating the integral at some large K ($\ll t$) gives the correlation proportional to

$$\lim_{t\to\infty}\int_0^K \frac{1}{\{(1+\beta s)\,(\beta t)\}^\lambda}ds$$

and so the asymptotic decay in the correlation is $t^{-\lambda}$. Recall we require $\lambda > 1$ so equation (6.10) is satisfied. As $\lambda \to 1$, the asymptotic decay in the correlation tends to t^{-1} so the model mimics a long-memory model. For $\lambda = 1.5$ the correlation decays asymptotically like $t^{-3/2}$ and for $\lambda = 2$ is decays like t^{-2}. This gives a slower decay than the BNS SV OU models. Figure 6.2 are ACF plots for simulations of size $50,000$ from this process, when $\beta = 0.1$ and for $\lambda = 1.5$ and $\lambda = 2$, using the $Ga\,(1,1) - OU$ BDLP and demonstrate the correct decay of the correlation of the volatility process.

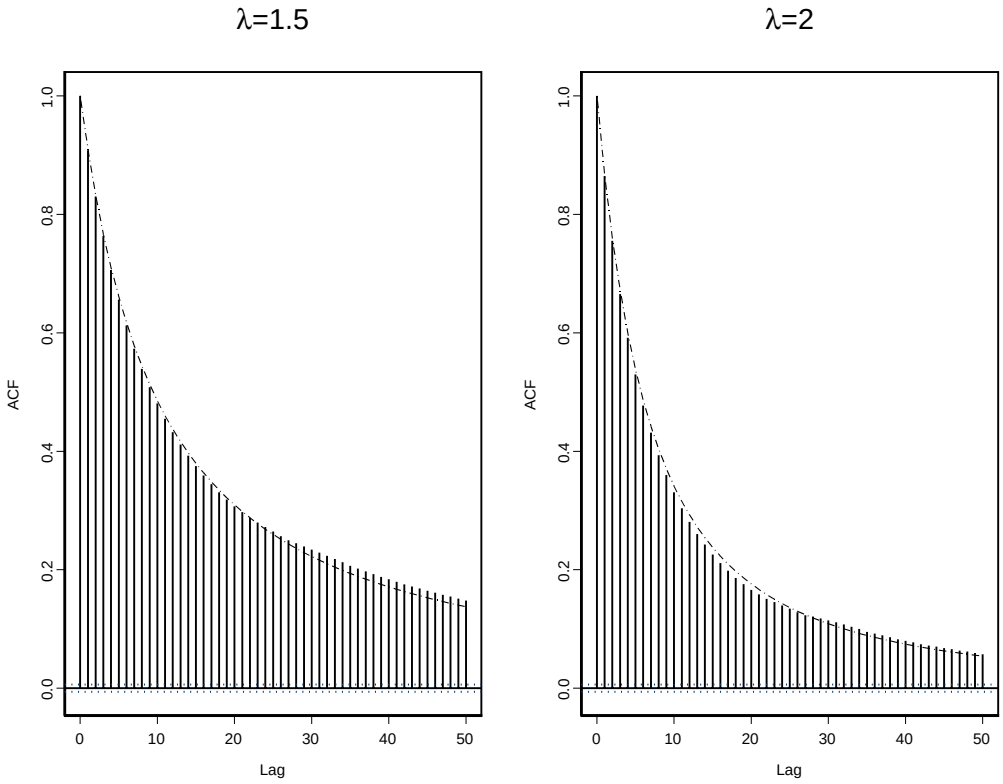


Figure 6.2: ACF of the Power Decay volatility process for $\beta = 0.1$, $\lambda = 1.5$ and $\lambda = 2$.

As λ increases the asymptotic decay of the process increases. The dashed line shows the theoretical correlation and suggests that the series representation of Section 6.1 has been

implemented correctly. The β parameter can be used to further control the correlation structure. Figure 6.2 are the same ACF plots as Figure 6.3 but for $\beta = 1$.

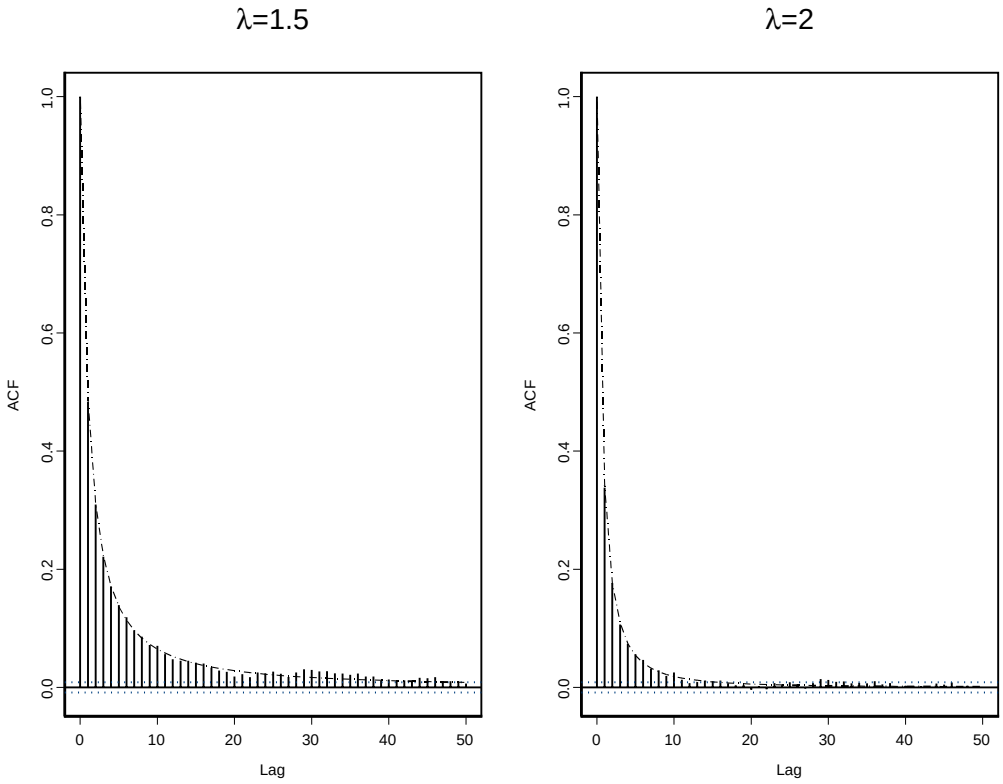


Figure 6.3: ACF of the Power Decay volatility process for $\beta = 1$, $\lambda = 1.5$ and $\lambda = 2$.

For $\beta = 1$, the initial decay in the correlation is faster than when $\beta = 0.1$. Models with $h_1(t-s)$ given by equation (6.17) can control the initial decay of the volatility through the β parameter and the asymptotic decay by the λ parameter.

6.2.3 Fractional Ornstein-Uhlenbeck process

Instead of using the Ornstein-Uhlenbeck process of previous chapters we consider the OU process with solution

$$\sigma^2(1,t) = \sqrt{2\lambda}\int_{-\infty}^{t} e^{-\lambda(t-s)} dz(s), \tag{6.18}$$

as was used in Section 6.2.1.

Definition 21 *The **Riemann-Liouville operator of fractional integration** of a func-*

tion, $f(s)$*, is defined by*

$$D^{-n} f(s) = \frac{1}{\Gamma(n)} \int_a^t (t-s)^{n-1} f(s)\, ds, \tag{6.19}$$

where D^{-n} *is the n-fold integral (see Anh and McVinish (2003)).*

Define $\sigma^2(\kappa, t)$ as

$$\sigma^2(\kappa, t) = \int_{-\infty}^t \lambda e^{-\lambda(t-s)} \sigma^2(\kappa - 1, t)(s)\, ds, \tag{6.20}$$

for $\kappa \neq 1$ (the $\kappa = 1$ case has been covered in Section 6.2.1). It can be shown that

$$\sigma^2(\kappa, t) = \sqrt{2\lambda} \int_{-\infty}^t \frac{\lambda^{\kappa-1}}{\Gamma(\kappa)} (t-s)^{\kappa-1} e^{-\lambda(t-s)} dz(s)$$

(see Wolpert and Taqqu (2004) and Appendix A.11). The process $\sigma^2(\kappa, t)$ is therefore called the fractional Ornstein-Uhlenbeck Lévy (fOUL) process as it is of the form of equation (6.19) with $n = \kappa$ and

$$f(s) = \sqrt{2\lambda} \lambda^{\kappa-1} e^{-\lambda(t-s)}.$$

Equations (6.18) and (6.20) are equivalent to those used in Wolpert and Taqqu (2004) to define fOUL processes. Unlike the BNS SV OU models, the marginal distribution of the volatility is influenced by λ for these OU processes. We shall not change the timing of the BDLP to avoid this, as although this is possible with the OU solution, for the fOUL solution, κ also alters the marginal and it is difficult to manipulate equation (6.20) to ensure this is not the case (and therefore we will be unable to make the marginal independent of both λ and κ). Wolpert and Taqqu (2004) use fOUL processes to model the Telecom process and are interested in the covariance function of these processes, whilst we are usually concerned with the correlation function. The fOUL process is a special case of equation (6.11), when

$$h(x) = \begin{cases} 0 & x < 0 \\ \sqrt{2\lambda} \dfrac{\lambda^{\kappa-1}}{\Gamma(\kappa)} x^{\kappa-1} e^{-\lambda x} & x \geq 0 \end{cases}.$$

The variance of the process is given by equation (6.14) and so we restrict our attention to the finite variance processes, where $\kappa > 1/2$. Equation (6.15) gives the correlation function

$$\rho^{\kappa}(t) = \frac{2}{\Gamma(\kappa - 1/2)} \left(\frac{\lambda t}{2}\right)^{k-1/2} K_{\kappa-1/2}(\lambda |t|).$$

For small lags, the correlation decays like a power for the fOUL process (unlike the OU process). Although the fOUL process has a more flexible correlation structure than the OU, both processes decay exponentially for large lags and do not have long-memory. The

decay of the correlation for small lags is controlled by κ and the decay for large lags is determined by λ. Figure 6.4 demonstrates how the correlation structure varies with κ for constant $\lambda = 0.1$.

Unfortunately, using a similar method to that used in Section 6.2.1, to calculate the marginal distribution of the volatility, it is difficult to derive the Lévy measure or distribution of $\sigma^2(\kappa, t)$ for fOUL processes in general for the BDLPs of Chapter 4. As before, consider the homogeneous BDLP, $z(t)$, with Lévy measure

$$w(x) = -u(x) - xu'(x),$$

where $u(x)$ is the Lévy measure of the marginal distribution of the BNS SV model with the same BDLP.

From equation (6.18), the negative of the characteristic exponent of $\sigma^2(\kappa, t)$ is

$$\log\left\{E\left[e^{it\sigma^2(\kappa,t)}\right]\right\} = \int_0^\infty \int_0^\infty \left\{e^{ituh(s)} - 1\right\} w(x)\, ds dx,$$

where

$$h(s) = \sqrt{2\lambda}\frac{\lambda^{\kappa-1}}{\Gamma(\kappa)} x^{\kappa-1} e^{-\lambda x},$$

from equation (6.20). Previously, we performed the substitution $r = uh(s)$ and more care must be taken to perform this substitution for general $h(s)$ as, for some κ values, the substitution is not one to one on the domain of integration. The difficulties for the case $1/2 < \kappa < 1$ are discussed below.

The substitution $r = uh(s)$ is one-to-one on $\mathbb{R}^+$ for $1/2 < \kappa < 1$ (unlike when $\kappa > 1$). However, this is still more complex than for the OU process, as

$$dr = uh'(s)\, ds$$

and

$$\frac{1}{uh'(s)} = \frac{1}{uh(s)} \frac{s}{(\kappa - 1 - \lambda s)} = r^{-1} \frac{h^{-1}(r/u)}{(\kappa - 1 - \lambda h^{-1}(r/u))},$$

where

$$h^{-1}(s) = x^{1/(1-\kappa)} \exp\left[-LambertW\left(\frac{\lambda x^{1/(1-\kappa)}}{1-\kappa}\right)\right].$$

Then

$$\log\left\{E\left[e^{it\sigma^2(\kappa,t)}\right]\right\} = \int_0^\infty \int_0^\infty \left\{e^{itr} - 1\right\} r^{-1} \frac{h^{-1}(r/u)}{(\lambda h^{-1}(r/u) + 1 - \kappa)} dr w(u)\, du$$

and $\sigma^2(\kappa, t)$ has Lévy measure

$$r^{-1} \int_0^\infty \frac{h^{-1}(r/u)}{(\lambda h^{-1}(r/u) + 1 - \kappa)} w(u)\, du.$$

Due to the complex nature of h^{-1}, it is not possible to simplify this further, even when we have a $Ga - OU$ BDLP. This demonstrates that both λ and κ specify the exact form of the marginal distribution of $\sigma^2(\kappa, t)$. Even for specific κ values (such as $\kappa = 3/4$) it is not possible to evaluate this integral.

We will concentrate on simulating from such models with the $Ga - OU$ BDLP, as this gives finite summations in equations (6.6) and (6.8) and the Inverse Tail Mass function is available directly. Using the numerical methods described in Chapter 4 and Appendix B.1, along with the series representation of Section (6.1), allows us to simulate from integrals with respect to any of the BDLPs used in Chapter 4.

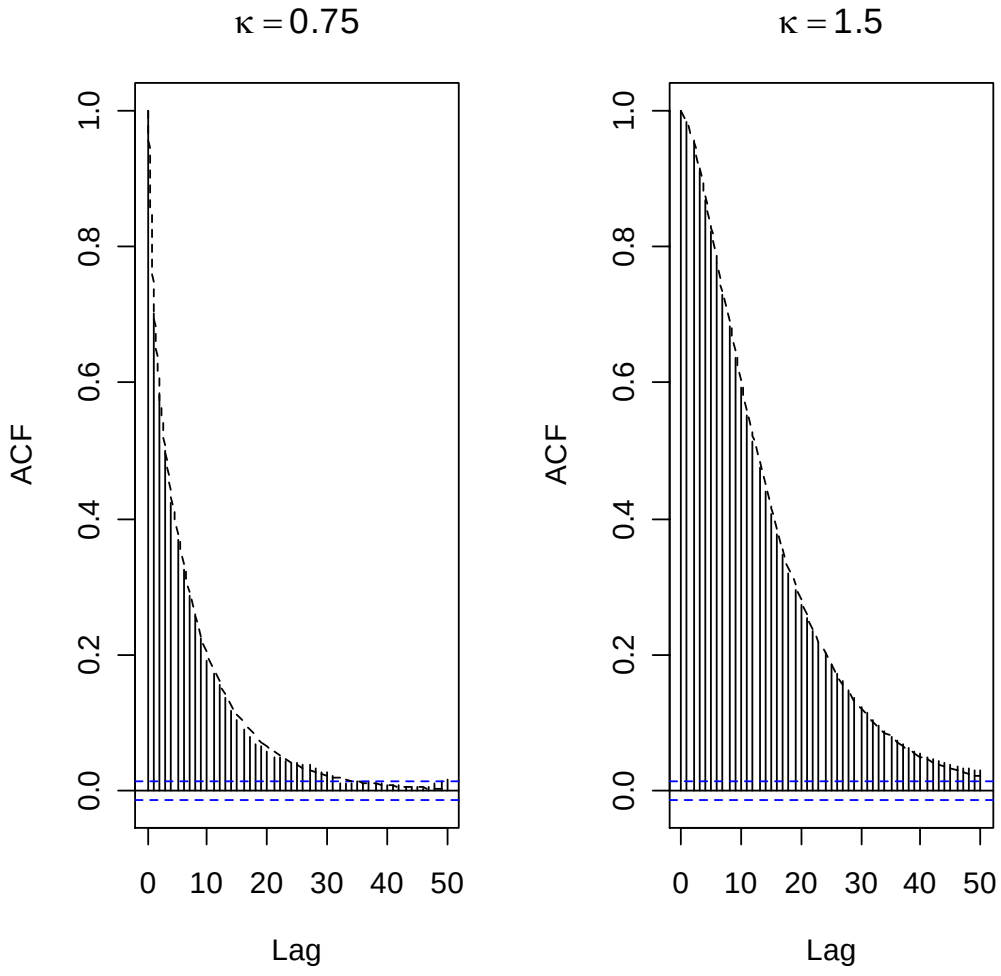


Figure 6.4: ACF of the fOUL process for $\lambda = 0.1$, $\kappa = 0.75$ and $\kappa = 1.5$.

Even though we were unable to derive the marginal distribution of the volatility for this BDLP, empirical results suggest the volatility might be distributed *Gamma* when the $Ga - OU$ BDLP is used to drive the fOUL process. Figure 6.5 are QQplots of the simulated volatility processes against two *Gamma* distributions, with parameters of the *Gamma* distributions chosen from moment matching.

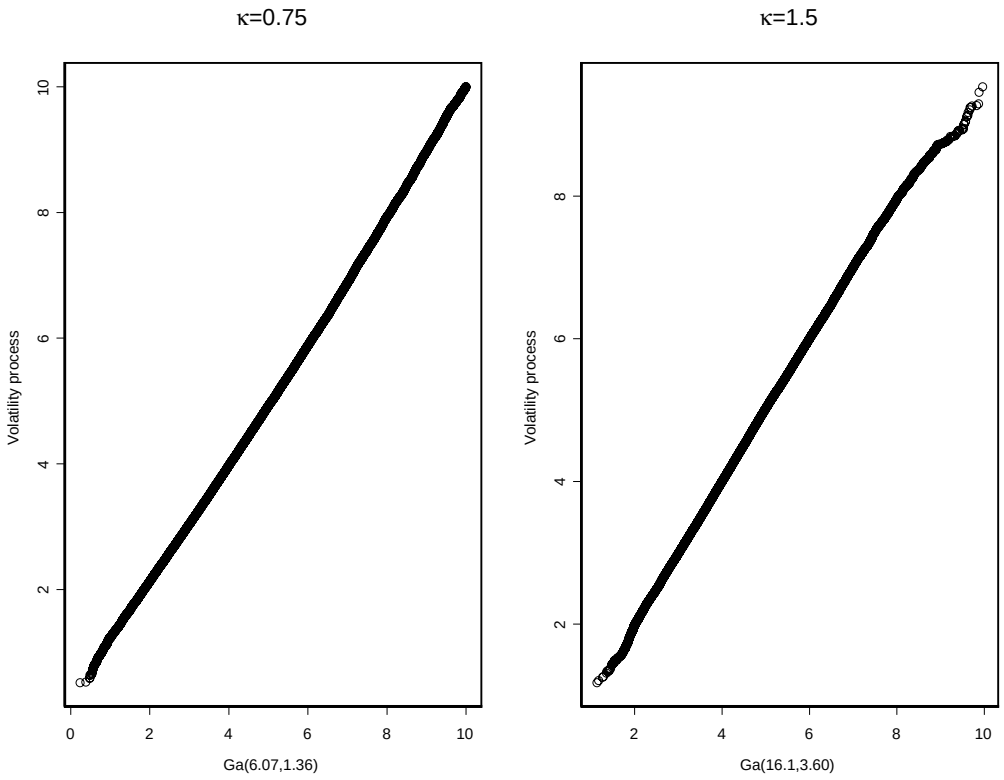


Figure 6.5: QQplots of the fOUL process for $\lambda = 0.1$, $\kappa = 0.75$ and $\kappa = 1.5$ against *Gamma* distributions with parameters chosen from moment matching.

6.3 Inference using MCMC

An MCMC algorithm to estimate the parameters of the BNS OU SV models was described in Chapter 4. This algorithm simulates from the stochastic integral of equation (4.5) using the series representation of equation (4.9). For a data set of size T this is an $O(T)$ algorithm. For the models described in this chapter, in general, we need to sample from stochastic integrals of the form of equation (6.5) and can use a similar series representation, given in equations (6.6) and (6.8), which gives an $O\left(T^2\right)$ algorithm to sample from the instantaneous volatilities, $\sigma^2(0\Delta), \ldots, \sigma^2(T\Delta)$.

The likelihood for BNS OU SV models in Chapter 4 is specified by the discretely observed volatility, σ_i^2, defined in equation (1.6) as

$$\sigma_i^2 = \int_{(i-1)\Delta}^{i\Delta} \sigma^2(u)\, du,$$

which has a simple form for the BNS OU SV models, given in equation (4.2). In general, for the models of this chapter, this discretely observed volatility is

$$\int_{(i-1)\Delta}^{i\Delta} \int_0^{\infty} f_1(\lambda, u, s)\, dz(s)\, du + \int_{(i-1)\Delta}^{i\Delta} \int_0^{t} f_2(\lambda, u, s)\, dz(s)\, du.$$

As the time dependent term of $\sigma^2(t)$ cannot be separated from the stochastic integral term, this cannot be simplified to a single integral, as was the case for the BNS OU SV models. The series representations for the models of this chapter are slower to implement because of the $O\left(t^2\right)$ series representation of equation (6.8). The double integrals for the discretely observed volatility are very intensive to compute and not currently feasible to implement for an MCMC algorithm. However, the instantaneous volatility and discretely observed volatilities have similar properties and so we can fit the models of this chapter using the same likelihood as before but with the approximation

$$\sigma_i^2 = \int_{(i-1)\Delta}^{i\Delta} \sigma^2(u)\, du \approx \sigma^2((i-1)\Delta)\,\Delta.$$

Alternatively, we could use the approximation

$$\int_{(i-1)\Delta}^{i\Delta} \sigma^2(u)\, du \approx \sigma^2(i\Delta)\,\Delta,$$

or use the trapezium rule to make the approximation

$$\int_{(i-1)\Delta}^{i\Delta} \sigma^2(u)\, du \approx \left\{\frac{\sigma^2(i\Delta) - \sigma^2((i-1)\Delta)}{2}\right\}\Delta.$$

For simulation purposes, each approximation gives similar results. For this reason we use the approximation $\sigma_i^2 \approx \sigma^2(i\Delta)\,\Delta$, as the correlation structure of $\sigma^2(i\Delta)$ is already known and gives a simple correlation structure for σ_i^2.

To test the correct fit of the models to observed data, we can look at the observed and theoretical correlation structure of the square of the log returns given the estimated model parameters. To estimate

$$corr\left[y_i^2, y_{i+s}^2\right], \qquad \text{for } s > 0,$$

make the approximation

$$y_i \sim N\left(0, \sigma^2(i\Delta)\right).$$

Then $E\left[y_i^2\right] = E\left[\sigma^2\left(i\Delta\right)\right]$ and so

$$\begin{aligned} Cov\left[y_i^2, y_{i+s}^2\right] &= E\left[\left(\sigma^2\left(i\Delta\right)X_i^2 - \overline{\sigma^2\left(i\Delta\right)}\right)\left(\sigma^2\left(\left(i+s\right)\Delta\right)X_{i+1}^2 - \overline{\sigma^2\left(\left(i+s\right)\Delta\right)}\right)\right] \\ &= \rho\left(s\right)V\left[\sigma^2\left(i\Delta\right)\right], \end{aligned}$$

where $X_i \overset{iid}{\sim} N\left(0,1\right)$ is independent of the volatility process and $\rho\left(t\right)$ is as given in equation (6.15). We also have

$$V\left[y_i^2\right] = V\left[y_{i+s}^2\right] = V\left[\sigma^2\left(i\Delta\right)\right]V\left[X_i^2\right] = 2V\left[\sigma^2\left(i\Delta\right)\right]$$

and so

$$Corr\left[y_i^2, y_{i+s}^2\right] = \frac{\rho\left(s\right)}{2}.$$

The MCMC algorithm is the same as previously, using the new stochastic integrals and series representations. Due to the intensive series representation, efficient coding is very important, so that the models run in sensible time. Note that when the b^{th} row of Poisson points and uniforms are updated, as

$$I_{2,t} \overset{\mathfrak{L}}{=} \sum_{j=0}^{t-1}\sum_{i=0}^{n_{2,j}} W^{-1}\left(a_{2,j,i}\right) f_2\left(\lambda, t, r_{2,j,i} + j\right),$$

we have that

$$I_{2,t}^{'}|I_{2,t} \overset{\mathfrak{L}}{=} I_{2,t} - \sum_{i=0}^{n_{2,j}} W^{-1}\left(a_{2,b,i}\right) f_2\left(\lambda, t, r_{2,b,i} + b\right) + \sum_{i=0}^{n_{2,j}} W^{-1}\left(a_{2,b,i}^{'}\right) f_2\left(\lambda, t, r_{2,b,i}^{'} + b\right),$$

which does not require $O\left(t^2\right)$ operations. To further improve the speed, the values $W^{-1}\left(a_{2,b,i}\right)$ and $f_2\left(\lambda, t, r_{2,b,i} + b\right)$ can be stored, to avoid repeat calculations. We will focus on the *Ga-OU* BDLP to facilitate algorithm run time, though any of the previous BDLPs could be used. The MCMC algorithm for the *Ga-OU* BDLP for the new models runs at a similar speed to the *GIG-OU* BDLP for BNS OU SV models for data sets of size $T = 1000$. Our focus is to show the MCMC inference is possible and demonstrate this on training with the Power Decay process and real data, using both processes. For the Power Decay process, we use a $Ga\left(1, 0.1\right)$ prior for $\lambda + 1$ and a $Ga\left(1, 0.5\right)$ prior for β. For the fOUL process, we use a $Ga\left(1, 0.5\right)$ prior for λ and a $Ga\left(1, 2/3\right)$ prior for $\kappa + \frac{1}{2}$. Posteriors of λ and β for the Power Decay process on training data are shown in Figure 6.6, where $100,000$ iterations were taken (and thinned by recording every 10^{th} value) after a burn-in of $10,000$, for the simplest model (no leverage, stochastic λ process etc.). As the models of this chapter give flexible correlation structures, we concentrate on the parameters which determine this structure.

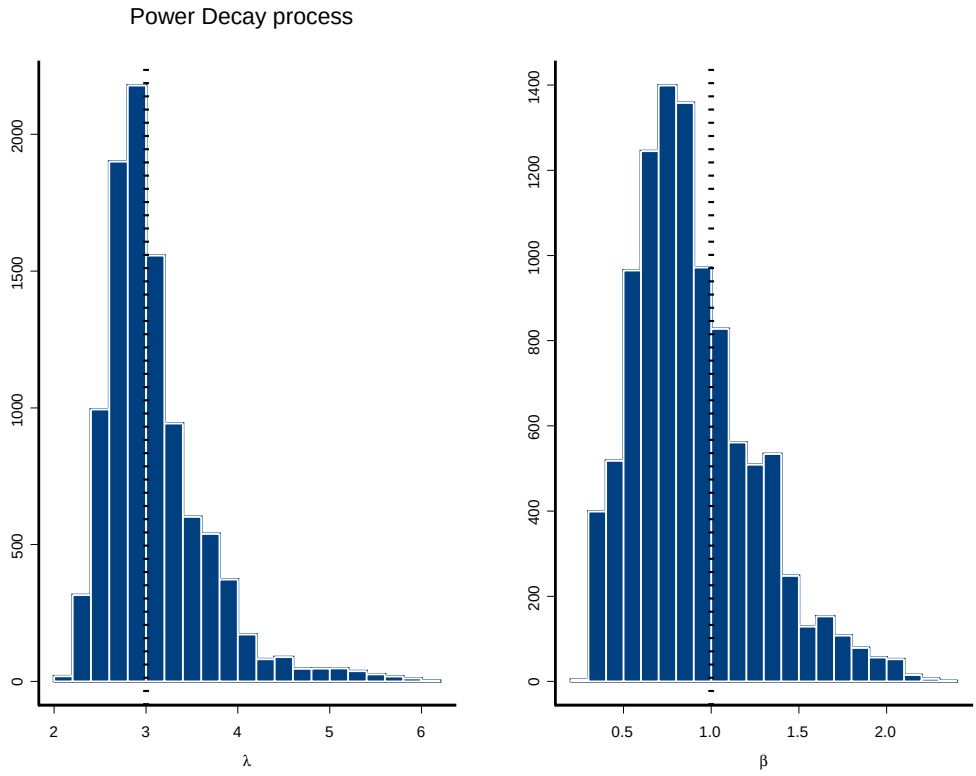


Figure 6.6: Histograms of the posterior distribution of λ and β for training data.

The posterior supports the true values from which the data were generated for the Power Decay process.

The inference for the fOUL process is the same as for the Power Decay process (apart from the form of the functions f_1 and f_2) and so we now assume the MCMC algorithm is working correctly and fit the models to the S&P 500 data set. Posterior histograms of the parameters determining the correlation of the square of the log returns are given in Figures 6.7 and 6.8. $10,000$ iterations were taken, thinning by recording every 10^{th} value, after a burn-in of $10,000$.

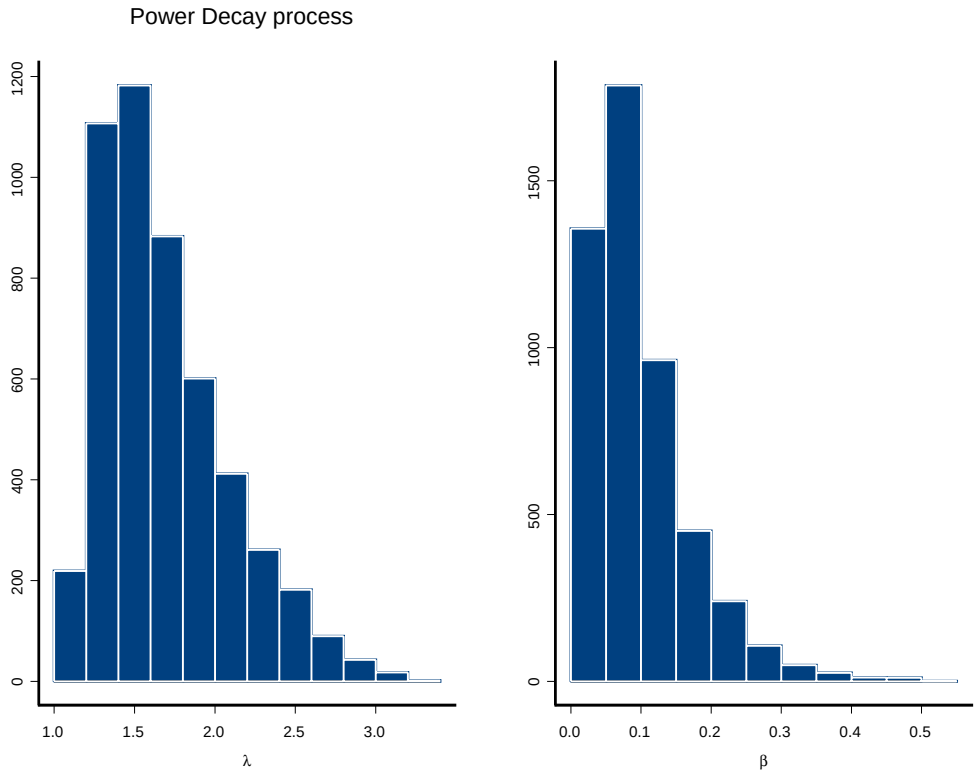


Figure 6.7: Power Decay process: Histograms of λ and β for S&P 500 data.

The posterior for λ supports small λ values and fits a volatility process which decays asymptotically at a rate between t^{-1} and t^{-3}. The posterior is not concentrated at $\lambda = 1$, so the volatility process does not have long-memory.

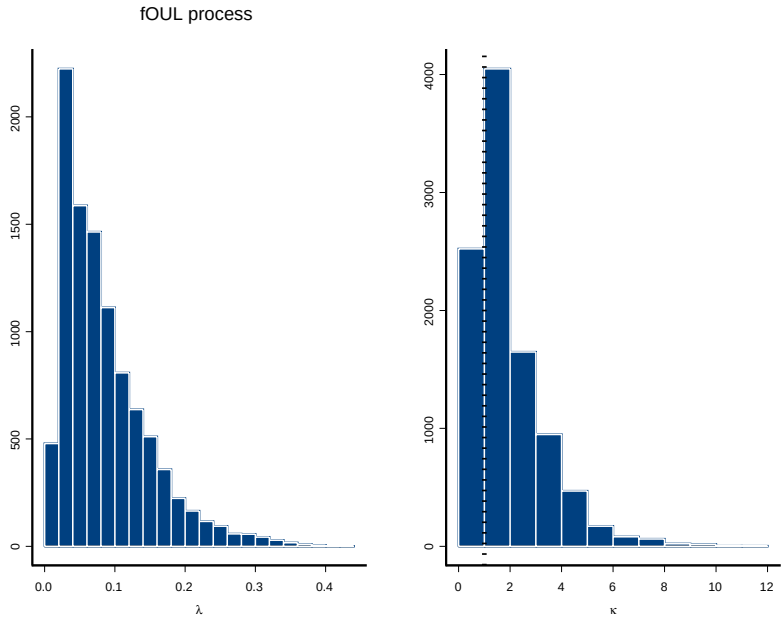


Figure 6.8: fOUL process: Histograms of λ and κ for S&P 500 data.

The posterior supports $\kappa = 1$ and so supports the BNS OU model.

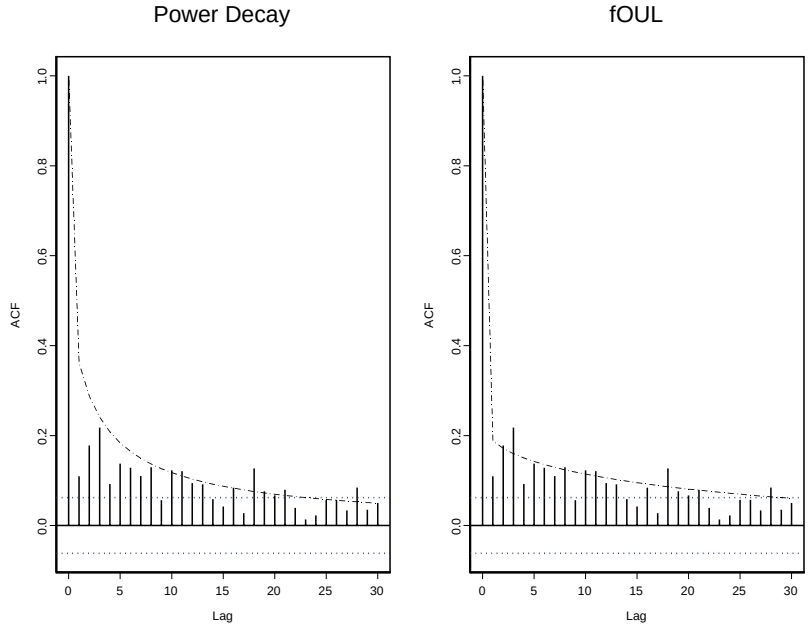


Figure 6.9: ACF of the square of the log returns of S&P 500 data and theoretical ACF of the fitted Power Decay and fOUL processes.

Figure 6.9 demonstrates the Power Decay and fOUL models accurately fitting the correlation structure of the S&P 500 data set. These graphs are ACF plots of the square of the log returns and the theoretical distribution of the square of the log returns of the fitted processes for one set of parameters, taken after the MCMC had converged.

6.4 Chapter summary

Wolpert and Taqqu (2004) consider a class of stochastic processes driven by Lévy processes. We recall these models and suggest they could be used for stochastic volatility models because of their rich correlation structure. We describe how to simulate from such models and some of the properties of them. Although the models of this chapter can have a more flexible correlation structure than the BNS SV models, they are less tractable because the stochastic integrands have a more complex form. For example, for the BNS SV model, the relationship between the BDLP and marginal distribution of the volatility is simple, whilst for the models of this chapter it is often not available analytically. The models of this chapter require a more complex series representation and this makes simulation slower. Therefore MCMC inference for these models, using the series representation of this chapter and a similar algorithm to that of Chapter 4, is slower, though still feasible for the *Ga-OU* BDLP.

Chapter 7

Further Work

Some ideas for potential further work are suggested. These have not been implemented as they are not feasible on modern computers with current mathematical theory and/or were considered beyond the scope of this thesis, which largely focuses on the implementation and performance of the stochastic volatility models proposed in Barndorff-Nielsen and Shephard (2001b).

7.1 Multivariate volatility models

Although the thesis focuses on Lévy measures of univariate distributions, the Lévy measure definition (see Theorem 2) is for multivariate distributions. Barndorff-Nielsen and Shephard (2001b) concentrate on univariate SV models but observe that the models extend to multivariate Lévy processes. In principle, there is no reason why an infinitely divisible distribution cannot be used as a marginal distribution for the volatility between several underlying assets. The *multivariate Inverse Gaussian* distribution (see Minami (2003)) is an example of an infinitely divisible multivariate distribution. The multivariate model allows correlation between the volatilities of different assets, rather than modelling the volatility for each asset independently. Given the difficulty of the MCMC implementation in the univariate case and the time taken to run it, MCMC inference for multivariate models is not feasible with current computing power and the techniques used in this thesis (at least not with multivariate generalisations of all the marginal distributions that are used in Chapters 4 and 5). It might be possible to perform the inference for a small number of assets with carefully chosen marginal distributions which have simple Inverse Tail Mass functions (such as a multivariate generalisation of the *Gamma* distribution). It would be interesting to compare the empirical performance of multivariate BNS SV models for a

small number of assets and univariate BNS SV models (with the same marginal distribution) to see if the extra complexity of the multivariate models improves the prediction of future assets movements.

7.2 Comparison with discrete time models

All the models used in this thesis were continuous time models (apart from the stochastic λ process in Section 3.3). Although the asset was only simulated at discrete time points, the differential equations driving the underlying are continuous (such as the Black-Scholes and Ornstein-Uhlenbeck equations). There are many stochastic volatility models in discrete time, such as those given in Bollerslev (1987) and Shephard and Pitt (1997). The MCMC inference for such models is generally more straightforward than for the BNS SV models. It has been argued that continuous time models are more appropriate than discrete ones as observations occur in continuous time. Certain analytical results are available for the continuous BNS SV model which are less tractable for the discrete time models (such as a simple expression for the integrated volatility). However, there is no concrete mathematical justification as to why a continuous or discrete model should have any advantage over the other for option pricing or forecasting.

It would be interesting to compare some of the popular discrete time models with the Ornstein-Uhlenbeck model to see if any of the models are able to outperform the others empirically.

7.3 Long-memory models

Although significant evidence of long-memory was found in real financial data in Section 3.1.3, results in the literature are less conclusive (and we were unable to identify any empirical advantage of the long-memory generalisation in Chapter 5). To try to cater for long-memory, an approximation to fractional Brownian motion was introduced in Section 3.5. Quasi long-memory can also be created by considering a superposition of Ornstein-Uhlenbeck processes though this introduces many more parameters. Various long-memory models in discrete time exist, such as the Long-Memory Stochastic Volatility (LMSV), introduced by Breidt et al. (1998).

It would be interesting to try to generalise an Ornstein-Uhlenbeck model for the volatility so it could generate long-memory in the absolute and square log returns in the original Black-Scholes equation. Sadly such generalisations, which keep the analytic tractability of the model proposed in Barndorff-Nielsen and Shephard (2001b), are not obvious.

Theorem 5 (proved by Jurek and Vervaat (1983)) guarantees that all self-decomposable distributions can be used as a marginal distribution for volatility and written in terms of a BDLP. Differentiating this Lévy representation of the infinitely divisible distribution generates the Ornstein-Uhlenbeck equation. Similar theorems to that of Jurek and Vervaat (1983), which can be manipulated into an attractive stochastic differential equation for the stochastic volatility, are not readily available. If there were other infinitely divisible distributions that could be written in terms of a different stochastic integral (with respect to a different type of Lévy process) then other models would be easier to discover. SV models with polynomial decay in the correlation of the volatility process were considered in Chapter 6 and these could be used to try to better explain the correlation structure of observed financial data.

An example of a long-memory model in continuous time is where the log of the volatility follows the Ornstein-Uhlenbeck equation (see Comte and Renault (1998), though this paper uses Brownian motion to drive the OU equation). For this model (not necessarily driven by Brownian motion) the integrated volatility is not available analytically (unlike the BNS SV models). Additionally, for the BNS SV models, it is easy to specify the marginal distribution of the volatility and this is a very appealing feature of the model. The model of Comte and Renault (1998) could be compared with the BNS SV model to try to see if there are any data sets which can take advantage of the long-memory model (though this is currently difficult to test empirically as observed in Section 5.3.4).

7.4 Lévy processes in the Black-Scholes equation

The standard Black-Scholes equation uses Brownian motion to drive the asset equation. As shown in Chapter 2, Brownian motion is a special example of a Lévy process. Lévy processes other than Brownian motion could be used to drive the Black-Scholes equation, which would generate jumps in the share process. It is not obvious how to pick this Lévy process from observed financial data, though this would be a logical way to further generalise the model. In the finance literature, parameters for such models are often estimated by minimizing some function of the difference between model and market option prices (see for example Carr et al. (2003)). It would be interesting to investigate which Lévy processes can be used to modify the standard Black-Scholes equation and most accurately predict "unseen" data, in a similar way to Section 5.1, again, possibly estimating model parameters using MCMC.

7.5 Improved truncation of the infinite sum of the random shock vector

For marginals other than the *Gamma* distribution, the infinite sum given in equation (4.20) must be truncated at some point. The truncation algorithms suggested in Section 4.3.5 and Appendix B.2 are sufficiently accurate, as samples from the model generate marginal distributions with the correct mean and variance for a wide range of parameter values of the marginal distributions, as well as the correct correlation structure for the volatility process. However, if it was possible to accurately estimate the error of the truncation (such as was performed in Appendix B.2 for the *Tempered Stable* marginal), for the *Generalised Inverse Gaussian*, *Positive Hyperbolic* and *Inverse Gamma* marginals, the infinite sums could be truncated earlier, without significantly altering the accuracy, and the speed of the MCMC would increase. Alternatively, a higher accuracy could be obtained with a similar computer run time. Working out the asymptotic behaviour of the Inverse Tail Mass function for the *Generalised Inverse Gaussian* distribution (see equation (4.11)) is not obvious and makes it difficult to find improved truncation schemes. The speed (and/or accuracy) of the MCMC could be increased if we had an adaptive z_{tol}, similar to the one used for the *Tempered Stable*.

7.6 Comparing leverage parameters for different marginal distributions

One of the disadvantages of the volatility models introduced by Barndorff-Nielsen and Shephard (2001b) is that the leverage parameter is influenced by the marginal distribution of the volatility (see for example Table 4.10). This makes it hard to interpret, quantitatively, the posterior distribution of leverage parameters between different marginals and different data sets. The ingenious choice and timing of λ in equation (1.8) means that it is unaltered by the marginal distribution used for $\sigma^2(t)$. When the MCMC was run on data from the Standard and Poor's 500 stock price (see Section 4.4.5) it can be seen that the posterior distributions for λ are very similar for each different marginal, whilst the posteriors for ρ differ slightly (though are still of the same order of magnitude). It would be useful to investigate why these posteriors are different and how to compare the leverage parameters of different marginals and data sets.

7.7 Different jump distributions in the stochastic λ process

Section 3.3 allowed λ to vary over time, with jumps from a *Normal* distribution (on the log scale) occurring as a Poisson process. This choice of jump distribution is somewhat arbitrary, as there is no compelling reason to believe jumps should be distributed in this way. It would be interesting to allow for jumps from other distributions, though the distributions must be able to cause λ to jump both up and down and keep λ positive. The most obvious way to do this is by considering jumps on the log scale and allowing positive and negative jumps. Other distributions which could be investigated include the *Generalised Asymmetric Laplace, Generalised Cauchy* and *Cosh/Sinh* distributions (see Chapter 2 for example). Empirical tests similar to those described in Chapter 5 could be used to assess which jump distributions fit observed financial data well. For a carefully chosen jump distribution, the stochastic λ process may perform better in Section 3.3 and clearly outperform the constant λ model.

7.8 Empirical performance of the models of Chapter 6

Chapter 6 motivated and explained how to simulate from a new class of SV models. The empirical performances of these models were not investigated. It would be interesting to see if these models can be fitted using MCMC, with current computing power and similar algorithms of Chapters 4 and 6, for more complex BDLPs than the *Ga-OU* BDLP. If it is feasible to fit the models using MCMC, it would be useful to test the empirical performance of the different models. This might give an insight as to what the correlation structure of SV models should be and whether capturing the asymptotic decay of the correlation in the volatility is as important as it is often made out to be.

Appendix A

Analytical Results

A.1 Derivation of the log asset equation

From equation (1.4), the asset equation is

$$dS(t) = \mu S dt + \sigma S dW(t).$$

Let $x(t) = \log(S(t))$, so

$$dx = \frac{\partial x}{\partial t}dt + \frac{\partial x}{\partial S}dS + \frac{1}{2}\frac{\partial^2 x}{\partial S^2}(dS)^2$$

and, as $W(t)$ is Brownian motion, $dW \sim N(0, dt)$, so $(dS)^2 = \sigma^2 S^2 dt + h.o.t.$ in dt. Then

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

as in equation (1.5). If $W(t)$ is fractional Brownian motion (see Definition 12), with Hurst parameter H, $(dS)^2 = \sigma^2 S^2 (dt)^{2H} + h.o.t.$ (see Hu and Øksendal (2003)).

A.2 Kurtosis of the log returns

From equation (1.7), the log asset, $x(t)$, obeys the stochastic differential equation

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

Assuming the dt term can be integrated out/ignored, the likelihood for the unobserved log returns, Y_i, is

$$Y_i \overset{\mathfrak{L}}{=} \sigma_i \varepsilon_i,$$

where $\varepsilon_i \overset{iid}{\sim} N(0,1)$ and σ_i is the root of the discretely observed or actual volatility as defined in Barndorff-Nielsen and Shephard (2001b) and equation (4.3). Then

$$E[Y_i] = 0$$

and

$$E\left[(Y_i - E[Y_i])^2\right] = E\left[Y_i^2\right] = \Delta E\left[\sigma_i^2\right]$$

and finally

$$E\left[(Y_i - E[Y_i])^4\right] = E\left[Y_i^4\right] = 3\Delta^2 E\left[\sigma_i^4\right].$$

The kurtosis of random variable Z is defined as

$$K_Z = \frac{m_4}{(m_2)^2},$$

where m_j is the j^{th} centralised moment of Z. The kurtosis of the log returns is therefore

$$K_Y = \frac{E\left[(Y_i - E[Y_i])^4\right]}{\left\{E\left[(Y_i - E[Y_i])^2\right]\right\}^2} = 3\frac{E\left[\sigma_i^4\right]}{\left(E\left[\sigma_i^2\right]\right)^2}.$$

This is not equal to $3K_{Z^2}$, where Z^2 is a random variable with the same distribution as the marginal of $\sigma^2(t)$, as the terms $E\left[\sigma_i^4\right]$ and $E\left[\sigma_i^2\right]$ are not centralised. The log returns have heavier tails than a *normal* distribution because $K_Y > 3$.

Non-centralised moments m_2, m_4, for three of the six different marginal distributions, are now given, along with the corresponding kurtosis of the log returns, K_y (note K_y is kurtosis of the log returns and is not a Bessel function). The table has the second and fourth moments of the different marginal distributions, though only the first and second moments contribute to the kurtosis.

	$IG(\nu,\alpha)$	$Ga(\nu,\alpha)$	$IGa(\nu,\alpha)$
m_2	$\frac{\nu(1+\nu\alpha)}{\alpha^3}$	$\frac{\nu(1+\nu)}{\alpha^2}$	$\frac{\alpha^2}{(\nu-1)(\nu-2)}$
m_4	$\frac{\nu\left(\nu^3\alpha^3+6\nu^2\alpha^2+15\nu\alpha+15\right)}{\alpha^7}$	$\frac{\nu(1+\nu)(2+\nu)(3+\nu)}{\alpha^4}$	$\frac{\alpha^4}{(\nu-1)(\nu-2)(\nu-3)(\nu-4)}$
K_y	$3\frac{1+\nu\alpha}{\nu\alpha}$	$3\frac{1+\nu}{\nu}$	$3\frac{\nu-1}{\nu-2}$
Kurtosis parameters	$\nu\alpha$	ν	ν

Table A.1: Non-centralised moments for three of the six different marginal distributions and the kurtosis of the log returns.

The $GIG(\gamma,\nu,\alpha)$, $TS(\kappa,\nu,\alpha)$ and $RPH(\nu,\alpha)$ distributions are not included in Table A.1 as m_2, m_4 and K_Y are complicated. These missing values can be found in Section 4.2.3.

A.3 Representations of the Stable distribution

The representation used in Feller (1971), pg 549, for the density of the $Stable(\alpha, \gamma)$ density (for $0 < \alpha < 1$) is

$$\begin{aligned} f_X(x) &= \frac{1}{\pi x}\sum_{j=1}^{\infty}\frac{\Gamma(j\alpha+1)}{j!}\left(-x^{-\alpha}\right)^j \sin\left\{\frac{j\pi}{2}(\gamma-\alpha)\right\}, \quad \text{for} \quad x>0 \\ &= \frac{1}{\pi x}\sum_{j=1}^{\infty}\frac{\Gamma(j\alpha+1)}{j!}(-1)^{j-1}x^{-\alpha j}\sin\left\{\frac{j\pi}{2}(\alpha-\gamma)\right\}. \end{aligned}$$

Picking $\alpha = \kappa$ and $\gamma = -\kappa$ gives

$$f_X(x) = \frac{1}{\pi}\sum_{j=1}^{\infty}\frac{\Gamma(j\kappa+1)}{j!}(-1)^{j-1}x^{-\kappa j-1}\sin(j\pi\kappa), \quad \text{for} \quad x>0.$$

Define the random variable $Y = cX$ (for $c > 0$), then

$$\begin{aligned} f_Y(y) &= \frac{1}{c}f_X\left(\frac{y}{c}\right), \quad \text{for} \quad y>0 \\ &= \frac{1}{\pi c}\sum_{j=1}^{\infty}\frac{\Gamma(j\kappa+1)}{j!}(-1)^{j-1}\left(\frac{y}{c}\right)^{-\kappa j-1}\sin(j\pi\kappa). \end{aligned}$$

Picking $c = 2\delta^{1/\kappa}$ gives

$$f_Y(y) = \frac{1}{\pi c}\sum_{j=1}^{\infty}(-1)^{j-1}\sin(j\pi\kappa)\frac{\Gamma(j\kappa+1)}{j!}2^{\kappa j+1}\left(\frac{y}{\delta^{1/\kappa}}\right)^{-\kappa j-1}, \quad \text{for} \quad y>0.$$

This is the representation used in Barndorff-Nielsen and Shephard (2001c) and Section 4.2.2 for the positive $\kappa - Stable$ distribution.

A.4 Evaluation of $\int_{-\infty}^{\infty}\frac{1-\cos(tx)}{x^2}dx$

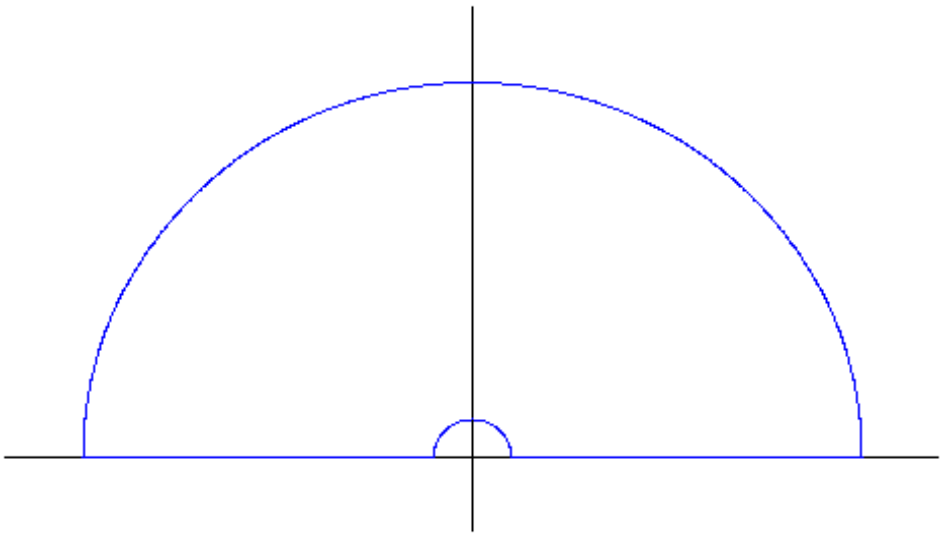
Consider the function

$$\begin{aligned} f(z) &= \frac{1-e^{itz}}{z^2} \\ &= -\frac{it}{z} - \frac{i^2t^2}{2!} - \frac{i^3t^3}{3!}z - \ldots, \end{aligned}$$

where $t \in \mathbb{R}^+$ and $z \in \mathbb{C}$. Therefore $f(z)$ has a simple zero at $z = 0$ with residue $-it$.

Use the contour which is a semi-circle, of radius R, in the positive imaginary plane,

indented at the origin with a semi-circle of radius ε, to avoid the pole ($0 < \varepsilon < R$).



Use the star domain $\mathbb{C} \setminus \{z = x + iy : y \leq 0\}$. There are no singularities inside the contour above so

$$\int_{\gamma_1+\gamma_2+\gamma_3+\gamma_4} f(z)\, dz = 0.$$

Considering real parts,

$$\int_{\gamma_1+\gamma_3} f(z)\, dz = \int_{-R}^{-\varepsilon} \frac{1-\cos(tx)}{x^2} dx + \int_{\varepsilon}^{R} \frac{1-\cos(tx)}{x^2} dx$$

and (indentation lemma)

$$\begin{aligned} \int_{\gamma_2} f(z)\, dz &= -\pi i \left[residue \left\{ f(z), z = 0 \right\} \right] \\ &= -\pi t. \end{aligned}$$

The extra minus occurs as the small semi-circle contour is clockwise.

Let $g(z) = e^{-itz} f(z)$ and write $z = x + iy$ $(x, y \in \mathbb{R})$, then on γ_4

$$\begin{aligned} |g(z)| &= \frac{\left|e^{-itz} - 1\right|}{|z^2|} \\ &\leq \frac{\left|e^{-itx-ty}\right| + 1}{R^2} \\ &= \frac{e^{-ty} + 1}{R^2} \end{aligned}$$

and $t \geq 0$ so

$$|g(z)| \leq \frac{2}{R^2},$$

which tends to zero as $R \to \infty$. Applying Jordan's lemma gives

$$\int_{\gamma_4} e^{itz} g(z)\, dz \to 0 \quad \text{as} \quad R \to \infty$$

and therefore letting $R \to \infty$ and $\varepsilon \to 0$ we obtain

$$\int_{-\infty}^{\infty} \frac{1-\cos(tx)}{x^2} dx = \pi t,$$

for $t \geq 0$. Note that for $t \geq 0$

$$\int_{-\infty}^{\infty} \frac{1-\cos(-tx)}{x^2} dx = \pi t,$$

as cos is even. Hence

$$\int_{-\infty}^{\infty} \frac{1-\cos(tx)}{x^2} dx = \pi |t|.$$

A.5 Derivation of the Lévy measure of the Cosh distribution

The density function is

$$f_X(x) = \frac{1}{\pi \cosh(x)}, \quad \text{for} \quad -\infty < x < \infty. \tag{A.1}$$

It has been shown that

$$\phi_1(t) = \int_{-\infty}^{\infty} \frac{e^{itx}}{\pi \cosh(x)} dx = \frac{1}{\cosh(\pi t/2)}.$$

Though not proved here, a technique to prove this is described in Feller (1971), pg 503. Letting $Y = X/a$, for $a \in \mathbb{R}$, gives

$$f_Y(y) = \frac{1}{\pi |a| \cosh(y/a)}, \quad \text{for} \quad -\infty < y < \infty$$

and

$$\phi_2(t) = \int_{-\infty}^{\infty} \frac{e^{ity}}{|a| \pi \cosh(y/a)} dy = \frac{1}{\cosh(\pi t a/2)}.$$

Then

$$\Psi(t) = -\log\{\phi_2(t)\} = \log\{\cosh(\pi t a/2)\} \tag{A.2}$$

and

$$\frac{d^2}{dt^2}\{\Psi(t)\} = \frac{\pi^2 a^2}{4}\{\phi_2(t)\}^2,$$

so $f_Y(y)$ is infinitely divisible.

Also, $(\phi_2(t))^2$ is the characteristic function of the random variable $Z = X + Y$, where X and Y are independent and identically distributed with density function $f_X(.)$ from equation (A.1). This characteristic function is derived in Feller (1971). This gives

$$f_Z(z) = \frac{4sign(a)}{\pi^2 a^2} \frac{z}{e^{z/a} - e^{-z/a}}.$$

The derivation of $f_Z(z)$ is given in Appendix A.6.

So,

$$\frac{d^2}{dt^2}[\Psi(t)] = sign(a) \int_{-\infty}^{\infty} e^{itx} \frac{x}{e^{x/a} - e^{-x/a}} dx$$

and integrating with respect to t gives

$$\frac{d}{dt}[\Psi(t)] = sign(a) \int_{-\infty}^{\infty} \frac{e^{itx} + f_1(x) + if_2(x)}{ix} \frac{x}{e^{x/a} - e^{-x/a}} dx, \tag{A.3}$$

where $f_1(x), f_2(x) \in \mathbb{R}$. From equation (A.2), we then have

$$\frac{d}{dt}[\Psi(t)] = \frac{\pi a}{2} \tanh(\pi t a/2),$$

so

$$i\frac{\pi a}{2} \tanh(\pi t a/2) = sign(a) \int_{-\infty}^{\infty} \frac{e^{itx} + f_1(x) + if_2(x)}{x} \frac{x}{e^{x/a} - e^{-x/a}} dx, \quad \forall t \in \mathbb{R}.$$

Taking real and imaginary parts, we have $f_1(x)$ and $f_2(x)$ satisfy

$$\int_{-\infty}^{\infty} \frac{\cos(tx) + f_1(x)}{e^{x/a} - e^{-x/a}} dx = 0 \tag{A.4}$$

and

$$sign(a) \int_{-\infty}^{\infty} \frac{\sin(tx) + f_2(x)}{e^{x/a} - e^{-x/a}} x dx = \frac{\pi a}{2} \tanh(\pi t a/2). \tag{A.5}$$

To prevent a pole occurring in equation (A.4), we require

$$\lim_{x \to 0} \frac{\cos(tx) + f_1(x)}{e^{x/a} - e^{-x/a}} = c.$$

Therefore the power series for $f_1(x)$ has only positive powers. Letting $t = 0$, for $f_1(x)$, we require

$$\int_{-\infty}^{\infty} \frac{1 + f_1(x)}{e^{x/a} - e^{-x/a}} dx = 0.$$

An obvious $f_1(x)$ which satisfies this is $f_1(x) = -1$. However, this is not the only such function which would work. For example $f_1(x) = -1 + x^2$ also satisfies equation (A.4).

For $f_2(x)$, letting $t = 0$,

$$\int_{-\infty}^{\infty} \frac{f_2(x)}{e^{x/a} - e^{-x/a}} x dx = 0. \tag{A.6}$$

The most obvious such $f_2(x)$ is $f_2(x) = 0$. Again, this is not unique as $f_2(x) = x^2$ satisfies equations (A.5) and (A.6). Continuing with the "obvious" $f_1(x)$ and $f_2(x)$, equation (A.3) becomes

$$\frac{d}{dt}\left[\Psi(t)\right] = sign(a)\int_{-\infty}^{\infty}\frac{e^{itx}-1}{ix}\frac{x}{e^{x/a}-e^{-x/a}}dx, \quad \forall t \in \mathbb{R}.$$

Integrating again, we obtain

$$\Psi(t) = -sign(a)\int_{-\infty}^{\infty}\frac{e^{itx}-itx+f_3(x)+if_4(x)}{x^2}\frac{x}{e^{x/a}-e^{-x/a}}dx,$$

where $f_3(x), f_4(x) \in \mathbb{R}$. Using equation (A.2) we have

$$\log\left\{\cosh(\pi ta/2)\right\} = -sign(a)\int_{-\infty}^{\infty}\frac{e^{itx}-itx+f_3(x)+if_4(x)}{x^2}\frac{x}{e^{x/a}-e^{-x/a}}dx$$

and using an argument similar to that used previously, we try $f_3(x) = -1$ and $f_4(x) = 0$, then

$$\Psi(t) = sign(a)\int_{-\infty}^{\infty}\frac{1+itx-e^{itx}}{x^2}\frac{x}{e^{x/a}-e^{-x/a}}dx. \tag{A.7}$$

This is not quite in the canonical form that appears in Theorem 4. However, note that

$$\int_{-\infty}^{\infty}\frac{itx}{x^2}\frac{x}{e^{x/a}-e^{-x/a}}dx = 0$$

and

$$\int_{-\infty}^{\infty}\frac{it\sin(x)}{x^2}\frac{x}{e^{x/a}-e^{-x/a}}dx = 0,$$

so equation (A.7) can be rewritten as

$$\Psi(t) = sign(a)\int_{-\infty}^{\infty}\frac{1+it\sin(x)-e^{itx}}{x^2}\frac{x}{e^{x/a}-e^{-x/a}}dx$$

and the canonical measure, $M(x)$, is given by

$$M(x) = sign(a)\frac{x}{e^{x/a}-e^{-x/a}}.$$

Setting $u(x) = M(x)x^{-2}$ gives the Lévy measure (i.e. satisfies Theorem 1.1, the Lévy-Khintchine formula). Setting $W = Y - \mu$, where $\mu \in \mathbb{R}$, gives

$$f_W(w) = \frac{1}{\pi|a|\cosh\left(\frac{w-\mu}{a}\right)}, \quad \text{for} \quad -\infty < w < \infty.$$

It is already known from Section 2.2.1 that the Lévy measure will be unaltered by the shift.

A.6 Derivation of the density of the convolution of two iid Cosh variates

Let X and Y be two independent and identically distributed random variables with density function

$$f_X(x) = \frac{1}{\pi |a| \cosh(x/a)}, \quad \text{for} \quad a \in \mathbb{R} \quad \text{and} \quad -\infty < x < \infty.$$

Let $Z = X + Y$, so

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{\infty} f_X(z-y) f_x(y)\, dy \\ &= \frac{4}{\pi^2 a^2} \int_{-\infty}^{\infty} \frac{dy}{\left(e^{z/a}e^{-y/a} + e^{-z/a}e^{y/a}\right)\left(e^{y/a} + e^{-y/a}\right)}. \end{aligned}$$

Substitute $t = e^{y/a}$

$$\begin{aligned} f_Z(z) &= \frac{4}{\pi^2 a} \int \frac{dt}{t\left(\frac{e^{z/a}}{t} + e^{-z/a}t\right)\left(t + \frac{1}{t}\right)} \\ &= \frac{4}{\pi^2 a} \int \frac{tdt}{\left(e^{z/a} + e^{-z/a}t^2\right)(t^2+1)}, \end{aligned}$$

for suitable limits in the integration. Using partial fractions

$$\begin{aligned} f_Z(z) &= \frac{4}{\pi^2 a} \int \frac{At}{e^{z/a} + e^{-z/a}t^2} + \frac{Bt}{t^2+1} dt \\ &= \frac{2}{\pi^2 a} \left\{ Ae^{z/a} \log\left(t^2 e^{-z/a} + e^{z/a}\right) + B \log\left(t^2+1\right) \right\}, \end{aligned}$$

where

$$A = \frac{1}{1 - e^{2z/a}} \qquad \text{and} \qquad B = \frac{1}{e^{z/a} - e^{-z/a}}.$$

So

$$Ae^{z/a} = -B$$

and

$$f_Z(z) = \frac{2B}{\pi^2 a} \left[\log\left(\frac{t^2+1}{t^2 e^{-z/a} + e^{z/a}} \right) \right].$$

For $a > 0$ this gives

$$\begin{aligned} f_Z(z) &= \frac{2B}{\pi^2 a} \left[\log\left(\frac{t^2+1}{t^2 e^{-z/a} + e^{z/a}} \right) \right]_0^{\infty} \\ &= \frac{2B}{\pi^2 a} \left\{ \log\left(e^{z/a}\right) - \log\left(e^{-z/a}\right) \right\} \\ &= \frac{4B}{\pi^2 a^2} z \end{aligned}$$

and if $a<0$

$$\begin{aligned} f_Z(z) &= \frac{2B}{\pi^2 a}\left[\log\left(\frac{t^2+1}{t^2e^{-z/a}+e^{z/a}}\right)\right]_{\infty}^{0} \\ &= -\frac{4B}{\pi^2a^2}z. \end{aligned}$$

Therefore

$$f_Z(z)=\frac{4sign(a)}{\pi^2a^2}\frac{z}{e^{z/a}-e^{-z/a}},$$

which is the density of the *Sinh* distribution (see Section 2.2.8).

A.7 Evaluation of $\int_{-\infty}^{\infty}\left\{1+it\sin(x)-e^{itx}\right\}x^{-2}c\delta(x)\,dx$

$$\begin{aligned} \int_{-\infty}^{\infty}\left\{\frac{1+it\sin(x)-e^{itx}}{x^2}\right\}c\delta(x)\,dx &= c\lim_{x\to0}\frac{1+it\sin(x)-e^{itx}}{x^2} \\ &= c\lim_{x\to0}\frac{1+it\sin(x)-\cos(tx)-i\sin(tx)}{x^2} \\ &= c\lim_{x\to0}\frac{-\left(-\frac{t^2x^2}{2!}+\ldots\right)-i\left(-\frac{t^3x^3}{3!}+\ldots\right)+\ldots}{x^2} \\ &= c\frac{t^2}{2} \end{aligned}$$

A.8 Relationship between the leverage parameters of two popular SV models

A popular discrete time model for the log asset, $x(t)$, is

$$dx(t)=\left\{\mu-\frac{\sigma^2(t)}{2}\right\}dt+\sigma(t)\,dW_1(t),$$

where the stochastic volatility follows the equation

$$d\log\left[\sigma^2(t)\right]=\alpha+\beta\log\left[\sigma^2(t)\right]dt+\sigma_v dW_2(t)$$

and $W_1(t)$ and $W_2(t)$ are two Brownian motions with $corr\left[dW_1(t),dW_2(t)\right]=\rho_1$ (so ρ_1 is a correlation between two Brownian motions and is the leverage parameter). Further details on this model can be found in Yu (2002).

For the BNS OU model with leverage, the log underlying, $x(t)$, follows the equation

$$dx(t)=\left\{\mu-\frac{\sigma^2(t)}{2}\right\}dt+\sigma(t)\,dW(t)+\rho_2 d\overline{z}(\lambda t), \tag{A.8}$$

where ρ_2 is the leverage parameter and $\overline{z}$ is a "centred" BDLP defined as

$$\overline{z}(t) = z(t) - E[z(t)] \tag{A.9}$$

and the stochastic volatility follows the Ornstein-Uhlenbeck equation

$$d\sigma^2(t) = -\lambda\sigma^2(t)\,dt + dz(\lambda t).$$

To compare ρ_1 and ρ_2, calculate the correlation between the two random processes in the share and volatility equation for the continuous model. First rewrite equation (A.8) as

$$dx(t) = \left\{\mu - \frac{\sigma^2(t)}{2}\right\}dt + \sigma(t)\{dW(t) + kd\overline{z}(\lambda t)\},$$

where $k = \rho_2/\sigma(t)$. We are interested in the correlation

$$C(\rho_2) = corr[dz(\lambda t), dW(t) + kd\overline{z}(\lambda t)]$$

and from equation (A.9), $d\overline{z}(t) = dz(\lambda t)$, so

$$C(\rho_2) = corr[d\overline{z}(t), dW(t) + kd\overline{z}(\lambda t)].$$

As $d\overline{z}(t)$ and $dW(t)$ are independent

$$\begin{aligned} cov[d\overline{z}(\lambda t), dW(t) + kd\overline{z}(\lambda t)] &= E[\{d\overline{z}(\lambda t) - E[d\overline{z}(\lambda t)]\}\,k\,\{d\overline{z}(\lambda t) - E[d\overline{z}(\lambda t)]\}] \\ &= kVar[d\overline{z}(\lambda t)]. \end{aligned}$$

Then,

$$\begin{aligned} C(\rho_2) &= \frac{kVar[d\overline{z}(\lambda t)]}{\sqrt{Var[d\overline{z}(\lambda t)]}\sqrt{Var[dW(t)] + k^2 Var[d\overline{z}(\lambda t)]}} \\ &= \frac{k}{\sqrt{k^2 + \dfrac{Var[dW(t)]}{Var[d\overline{z}(\lambda t)]}}}. \end{aligned}$$

As $W(t)$ is Brownian motion, $dW(t) = W(t+\delta t) - W(t) \sim N(0, \delta t)$, so

$$Var[dW(t)] = \delta t.$$

For $Var[d\overline{z}(\lambda t)]$, we note that

$$Var[d\overline{z}(\lambda t)] = Var[dz(\lambda t)]$$

and use a similar technique to the one used for Brownian motion. That is

$$\begin{aligned} dz(\lambda t) &= z(\lambda(t+\delta t)) - z(\lambda t) \\ &\stackrel{\mathfrak{L}}{=} z(\lambda\delta t) \end{aligned}$$

and then

$$\begin{aligned} Var\left[dz\left(\lambda t\right)\right] &= Var\left[z\left(\lambda\delta t\right)\right] \\ &= \lambda\delta t Var\left[z\left(1\right)\right] \\ &= 2\lambda\delta t Var\left[\sigma^2\left(t\right)\right]. \end{aligned}$$

Finally,

$$\begin{aligned} C\left(\rho_2\right) &= \frac{k}{\sqrt{k^2 + \frac{1}{2\lambda Var\left[\sigma^2\left(t\right)\right]}}} \\ &= \frac{sign\left(\rho_2\right)}{\sqrt{1 + \frac{1}{\rho_2^2}\frac{\sigma^2\left(t\right)}{2\lambda Var\left[\sigma^2\left(t\right)\right]}}}. \end{aligned}$$

This is a monotonic increasing function of ρ_2, like the leverage parameter in the discrete model. When ρ_1 and ρ_2 are negative, there is negative correlation in both models and the strength of this correlation becomes stronger as the parameters become more negative.

Unlike the discrete model, the correlation between the two driving processes (for the asset and volatility) is not completely specified by the leverage parameter; it is also determined by $\lambda, \sigma^2\left(t\right)$ and $Var\left[\sigma^2\left(t\right)\right]$.

A.9 Alternative leverage parameter

From Section 3.4, the leverage parameter enters the likelihood expression in the form ρz_i, where

$$\begin{aligned} z_i &= \eta_{i,2}/\lambda - \Delta E\left[\sigma^2\left(t\right)\right] \\ &\stackrel{\mathfrak{L}}{=} \Delta\left\{z\left(1\right) - E\left[\sigma^2\left(t\right)\right]\right\} \end{aligned}$$

and satisfies $E\left[z_i\right] = 0$.

Consider $V\left[z_i\right] = \Delta^2 V\left[z\left(1\right)\right]$ and let k_m be the m^{th} cumulant of $z\left(1\right)$. That is

$$k_m = E\left[\left\{z\left(1\right) - \overline{z\left(1\right)}\right\}^m\right]$$

and also let d_m be the m^{th} cumulant for the marginal distribution which is used for $\sigma^2\left(t\right)$. Then Barndorff-Nielsen and Shephard (2001a) have shown

$$k_m = m d_m.$$

If $m = 2$

$$k_2 = V\left[z\left(1\right)\right] = 2V\left[\sigma^2\left(t\right)\right],$$

and

$$V[z_i] = 2\Delta^2 V\left[\sigma^2(t)\right],$$

so define

$$\begin{aligned} z_i^{'} &= \frac{z_i}{\Delta\sqrt{2V\left[\sigma^2(t)\right]}} \\ &= \frac{\eta_{i,2}/\lambda - \Delta E\left[\sigma^2(t)\right]}{\Delta\sqrt{2V\left[\sigma^2(t)\right]}}. \end{aligned}$$

Then $E\left[z_i^{'}\right] = 0$ and $V\left[z_i^{'}\right] = 1$ (so the mean and variance of the leverage are independent of the marginal distribution used for $\sigma^2(t)$ as well as being independent of λ).

A.10 $GIG(\nu, 0, \gamma)$ is the only $GIG(\nu, \delta, \gamma)$ distribution with finite RSV sum

It is already known that the $GIG(\nu, 0, \gamma)$ (*Gamma*) distribution has a finite sum in equation (4.9) and does not require truncation. Continue with $\delta \neq 0$ and

$$\eta_i = \left\{ \begin{array}{c} e^{-\lambda\Delta} \sum\limits_{j=1}^{\infty} W^{-1}\left(\frac{a_j}{\lambda\Delta}\right) e^{\lambda\Delta r_j} \\ \sum\limits_{j=1}^{\infty} W^{-1}\left(\frac{a_j}{\lambda\Delta}\right) \end{array} \right\}.$$

In order that the sum be finite, it must be possible to truncate exactly at some n_i. As a_j's are the arrival times of a Poisson process with intensity 1, there must be some $K \in \mathbb{R}^+$, such that $W^{-1}\left(\frac{a_j}{\lambda\Delta}\right) = 0$ for all $a_j > K$. From equation (4.7)

$$W^{-1}\left(\frac{a_j}{\lambda\Delta}\right) = \inf\left\{y > 0 : W^+(y) \leq \frac{a_j}{\lambda\Delta}\right\}$$

so

$$\inf\left\{y \geq 0 : W^+(y) \leq l\right\} = 0 \quad \forall \quad l \geq \frac{K}{\lambda\Delta}.$$

The sum is finite if and only if $\lim_{y\to 0} W^+(y)$ is finite. Using equation (4.6),

$$\begin{aligned} \lim_{y\to 0} W^+(y) &= \lim_{y\to 0} yu(y) \\ &= \lim_{y\to 0}\left\{\frac{1}{2}\int_0^{\infty} \exp\left(-\frac{y\xi}{2\delta^2}\right) g_\nu(\xi)\, d\xi + \max(0, \nu)\right\} \exp\left(-\frac{\gamma^2 y}{2}\right). \end{aligned}$$

This is finite when

$$\lim_{y\to 0}\int_0^{\infty} \exp\left(-\frac{y\xi}{2\delta^2}\right) g_\nu(\xi)\, d\xi$$

is finite. Recall equation (2.2),

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

and as

$$\lim_{\xi \to \infty} g_\nu(\xi) = 0 \quad \forall \quad \nu \in \mathbb{R},$$

we are interested in when

$$\int_0^\infty g_\nu(y)\, dy$$

is finite. Consider asymptotic expansions for $J^2_{|\nu|}(\sqrt{x})$ and $N^2_{|\nu|}(\sqrt{x})$ (details of these expansions can be found in Gradshteyn and Ryzhik (1965)):

$$J_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}} \left\{ \cos\left(x \mp \frac{\pi}{2}\nu - \frac{\pi}{4}\right)(S1_{n,\nu,x} + R_1) - \sin\left(x \mp \frac{\pi}{2}\nu - \frac{\pi}{4}\right)(S2_{n,\nu,x} + R_1) \right\}$$

and

$$N_{\pm\nu}(x) = \sqrt{\frac{2}{\pi x}} \left\{ \sin\left(x \mp \frac{\pi}{2}\nu - \frac{\pi}{4}\right)(S1_{n,\nu,x} + R_2) + \cos\left(x \mp \frac{\pi}{2}\nu - \frac{\pi}{4}\right)(S2_{n,\nu,x} + R_2) \right\},$$

where

$$S1_{n,\nu,x} = \sum_{k=0}^{n-1} \frac{(-1)^k \Gamma\left(\nu + 2k + \frac{1}{2}\right)}{(2x)^{2k}(2k)!\Gamma\left(\nu - 2k + \frac{1}{2}\right)}$$

and

$$S2_{n,\nu,x} = \sum_{k=0}^{n-1} \frac{(-1)^k \Gamma\left(\nu + 2k + \frac{3}{2}\right)}{(2x)^{2k+1}(2k+1)!\Gamma\left(\nu - 2k - \frac{1}{2}\right)}$$

and R_1 and R_2 are the remainders

$$|R_1| < \left| \frac{\Gamma\left(\nu + 2n + \frac{1}{2}\right)}{(2x)^{2n}(2n)!\Gamma\left(\nu - 2n + \frac{1}{2}\right)} \right|, \quad \text{for} \quad n > \frac{\nu}{2} - \frac{1}{2}$$

and

$$|R_2| \leq \left| \frac{\Gamma\left(\nu + 2n + \frac{3}{2}\right)}{(2x)^{2n+1}(2n+1)!\Gamma\left(\nu - 2n - \frac{1}{2}\right)} \right|, \quad \text{for} \quad n \geq \frac{\nu}{2} - \frac{3}{4}.$$

These remainders tend to zero as $x \to \infty$ and are smaller than the other terms in the summations. Consider the behaviour of $J_{|\nu|}(\sqrt{x})$ and $N_{|\nu|}(\sqrt{x})$ as $x \to \infty$

$$\begin{aligned} J_{|\nu|}(\sqrt{x}) &\sim \frac{1}{\sqrt[4]{x}} \left\{ \cos\left(x \mp \frac{\pi}{2}|\nu| - \frac{\pi}{4}\right) - \sin\left(x \mp \frac{\pi}{2}|\nu| - \frac{\pi}{4}\right) \frac{1}{2x} \frac{\Gamma\left(|\nu| + \frac{3}{2}\right)}{\Gamma\left(|\nu| - \frac{1}{2}\right)} \right\} \\ &\sim \frac{1}{\sqrt[4]{x}} \cos\left(x \mp \frac{\pi}{2}|\nu| - \frac{\pi}{4}\right) \end{aligned}$$

and

$$\begin{aligned} N_{|\nu|}(\sqrt{x}) &\sim \frac{1}{\sqrt[4]{x}} \left\{ \sin\left(x \mp \frac{\pi}{2}|\nu| - \frac{\pi}{4}\right) + \cos\left(x \mp \frac{\pi}{2}|\nu| - \frac{\pi}{4}\right) \frac{1}{2x} \frac{\Gamma\left(|\nu| + \frac{3}{2}\right)}{\Gamma\left(|\nu| - \frac{1}{2}\right)} \right\} \\ &\sim \frac{1}{\sqrt[4]{x}} \sin\left(x \mp \frac{\pi}{2}|\nu| - \frac{\pi}{4}\right), \end{aligned}$$

so

$$\begin{aligned} J^2_{|\nu|}\left(\sqrt{x}\right)+N^2_{|\nu|}\left(\sqrt{x}\right) &\sim \frac{1}{\sqrt{x}}\left\{\cos^2\left(x\mp\frac{\pi}{2}\left|\nu\right|-\frac{\pi}{4}\right)+\sin^2\left(x\mp\frac{\pi}{2}\left|\nu\right|-\frac{\pi}{4}\right)\right\} \\ &= \frac{1}{\sqrt{x}}. \end{aligned}$$

Therefore, as $x\to\infty$, $g_\nu(x)=O\left(x^{-1/2}\right)$ and $\int_0^\infty g_\nu(y)\,dy$ is never finite for $\delta\neq 0$. Therefore the only $GIG(\nu,\delta,\gamma)$ distribution, which has a finite sum for η_i, occurs when $\delta=0$ and this is the $Ga\left(\nu,\gamma^2/2\right)$ distribution.

It is easy to see that there are no cases of the $TS(\kappa,\nu,\alpha)$ distribution (for $0<\kappa<1$) which give a finite summation by inspection of equation (4.13).

A.11 Proof of the solution of the fOUL process

We have

$$\sigma^2(1,t)=\sqrt{2\lambda}\int_{-\infty}^{t}e^{-\lambda(t-s)}dz(\lambda s)$$

and the recursive definition

$$\sigma^2(\kappa,t)=\int_{-\infty}^{t}\lambda e^{-\lambda(t-s)}\sigma^2(\kappa-1,s)\,ds \tag{A.10}$$

and want to prove the statement $p(k)$:

$$\sigma^2(\kappa,t)=\sqrt{2\lambda}\int_{-\infty}^{t}\frac{\lambda^{\kappa-1}}{\Gamma(\kappa)}(t-s)^{\kappa-1}e^{-\lambda(t-s)}dz(\lambda s).$$

Observe that $p(1)$ is true and assume $p(k)$ is true. Equation (A.10) implies that

$$\sigma^2(\kappa+1,t)=\int_{-\infty}^{t}\lambda e^{-\lambda(t-s)}\sigma^2(\kappa,s)\,ds$$

and, as we have assumed $p(k)$ is true, that is

$$\begin{aligned} \sigma^2(\kappa+1,t) &= \sqrt{2\lambda}\int_{-\infty}^{t}\lambda e^{-\lambda(t-s)}\int_{-\infty}^{s}\frac{\lambda^{\kappa-1}}{\Gamma(\kappa)}\left(s-s'\right)^{\kappa-1}e^{-\lambda\left(s-s'\right)}dz\left(\lambda s'\right)ds \\ &= \sqrt{2\lambda}\frac{\lambda^{\kappa}}{\Gamma(\kappa)}\int_{-\infty}^{t}\int_{s'}^{t}e^{-\lambda\left(t-s'\right)}\left(s-s'\right)^{\kappa-1}dsdz\left(\lambda s'\right) \\ &= \sqrt{2\lambda}\frac{\lambda^{\kappa}}{\Gamma(\kappa+1)}\int_{-\infty}^{t}\left(t-s'\right)^{\kappa-1}e^{-\lambda\left(t-s'\right)}dz\left(\lambda s'\right), \end{aligned}$$

so $p(k+1)$ is true by induction.

Appendix B

Numerical Algorithms

B.1 Evaluating the Inverse Tail Mass function for the IGa marginal distribution

When using an $IGa(\nu, \alpha)$ marginal for $\sigma^2(t)$, we must solve for z (for a given ν, α, x)

$$x\pi^2 = \int_0^\infty \frac{\exp\left(-\frac{zy}{4\alpha}\right)}{y\left\{J_{|\nu|}^2\left(\sqrt{y}\right) + N_{|\nu|}^2\left(\sqrt{y}\right)\right\}} dy. \tag{B.1}$$

For the *GIG* and *RPH* marginals, similar integrals must be evaluated.

Bounds for z can be found by using the monotonic property of $W_{\nu,\alpha}^{-1}(y)$, provided this integral can be evaluated numerically. Once bounds for z are known, a binary search method (and look up table) is used to solve for z. The difficult part of this is evaluating the integral itself for a given z, ν, α.

It proves numerically favourable to rewrite the integral as

$$\int_0^\infty g_{\nu,\alpha,z}(y)\, dy,$$

where

$$g_{\nu,\alpha,z}(y) = \frac{\exp(-y)}{y\left\{J_{|\nu|}^2\left(2\sqrt{\frac{\alpha y}{z}}\right) + N_{|\nu|}^2\left(2\sqrt{\frac{\alpha y}{z}}\right)\right\}}. \tag{B.2}$$

Consider the integral on $(0, 1)$. As $y \to 0$,

$$\begin{aligned} \exp(-y) &\sim 1 - y \\ J_\nu(y) &\sim \frac{1}{2^\nu \Gamma(\nu+1)} y^\nu \\ N_\nu(y) &\sim -\frac{2^\nu \Gamma(\nu)}{\pi} y^{-\nu} \end{aligned}$$

and

$$g_{\nu,\alpha,z}(y) \sim \frac{\pi^2}{\{\Gamma(\nu)\}^2}\left(\frac{\alpha}{z}\right)^{\nu} y^{\nu-1}.$$

This has very important implications as far as the numerical evaluation of the integral is concerned.

(1) If $\nu > 1$, as $y \to 0$ the integrand converges to 0.

(2) If $\nu \geq 2$, the integrand has a finite derivative at the origin, whilst if $1 < \nu < 2$ it is not differentiable as $y \to 0$.

(3) If $\nu = 1$, the integrand is some non zero constant at $y = 0$ and is smooth at the origin.

(4) If $\nu < 1$, the integrand explodes to infinity as $y \to 0$.

It is often easier to perform numerical integration on finite ranges, as it does not require approximation of the tail behaviour of the integrand. However, in this case, the integrand is not always well behaved near the origin and special care must be taken.

Due to possible problems in performing the integration near the origin, the integral was split:

$$\begin{aligned}\int_0^{\infty} g_{\nu,\alpha,z}(y)\,dy &= \int_0^{1} g_{\nu,\alpha,z}(y)\,dy + \int_1^{\infty} g_{\nu,\alpha,z}(y)\,dy \\ &= I_1 + I_2.\end{aligned}$$

Numerical methods can perform poorly on integrands that have infinite derivatives or that become infinite. To cope with these problems, an exponential mapping can be used to "crush" the integrand so that it (and the derivative) remain finite whenever $\nu < 1$ or $1 < \nu < 2$.

For I_1, use the change of variable

$$\phi(t) = \frac{\int_0^t e^{-c/u}du}{\int_0^1 e^{-c/u}du},$$

where c is some positive constant. Then

$$I_1 = \int_0^1 g_{\nu,\alpha,z}(\phi(t))\,\phi'(t)\,dt \tag{B.3}$$

and manipulation of $\phi(t)$ and $\phi'(t)$ gives

$$\begin{aligned}\phi(t) &= \frac{c\,\mathrm{Ei}(1,c/t) - te^{-c/t}}{c\,\mathrm{Ei}(1,c) - e^{-c}} \\ \phi'(t) &= -\frac{e^{-c/t}}{c\,\mathrm{Ei}(1,c) - e^{-c}}.\end{aligned}$$

The integrand of equation (B.3) is smoother than $g_{\nu,\alpha,z}(y)$, in the sense that it does not have an infinite derivative anywhere. This improves the numerical integration and the constant c can be chosen to further improve efficiency (though the choice of c is not obvious).

A linear map was then used to change the range of integration to $(-1,1)$ and Gaussian Quadrature (see Atkinson (1988)) was used, with $w(x)=1$, to evaluate the integral. One disadvantage of Gaussian Quadrature is that, although it can give a very high accuracy for a given number of points, estimating the error is difficult/impossible. In practice, the best way to ensure the integration is accurate is to keep on evaluating the integral with more and more points until the difference in the estimate is sufficiently small. Even using this, it is not obvious how to increase the number of points. e.g. having used 16 points, the number of points to use for the next integral estimate is not clear. The approach we take is to double the number of points, until the estimate is sufficiently accurate.

Now the integrand on the infinite range is examined. As $y \to \infty$,

$$\begin{aligned} J_\nu(y) &\sim \sqrt{\frac{2}{\pi}}\cos\left(y-\frac{\pi\nu}{2}-\frac{\pi}{4}\right)\frac{1}{\sqrt{y}} \\ N_\nu(y) &\sim \sqrt{\frac{2}{\pi}}\sin\left(y-\frac{\pi\nu}{2}-\frac{\pi}{4}\right)\frac{1}{\sqrt{y}} \end{aligned}$$

and

$$g_{\nu,\alpha,z}(y) \sim \pi\sqrt{\frac{\alpha}{z}}y^{-\frac{1}{2}}e^{-y},$$

so the integrand decays at least exponentially and Gauss-Laguerre integration can be used. Gauss-Laguerre integration assumes that the tail behaviour of the integrand dies at least as fast as e^{-y}. The change in integrand from equation (B.1) to (B.2) satisfies this assumption and this is why it proves more numerically efficient to use the second form of the integral.

For I_2, use the change of variable $x=\frac{y-1}{p}$, for p some positive constant, then

$$I_2 = p\int_0^\infty g_{\nu,\alpha,z}(xp+1)\,dx.$$

Gauss-Laguerre integration was then used to compute this as it performs very well when the integrand decays exponentially (as we have here). The constant p can again be chosen to improve the performance of the numerical integration.

B.2 Improved truncation for the Tempered Stable marginal

Assume an initial truncation has been made as suggested in Section 4.3.5 and consider the error term, E_i, for $\eta_{i,2}$ (note that the terms for $\eta_{i,1}$ are less than the terms for $\eta_{i,2}$).

Recall equation (4.9),

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

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

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

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

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

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

The central idea is then to approximate E_i analytically and pick a new truncation point (if the accuracy is not sufficient) making E_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 (4.12) and using equations (4.6) and (4.7), 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, E_i, will be approximately

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

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

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

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

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

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

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

so consider using the new truncation point

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

where $E\left[E_i\right]$ is our desired expected error in the summation (for our simulations we chose $E\left[E_i\right] = 0.001$). 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

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

B.3 Evaluating the Tempered Stable density function

If $\sigma^2\left(0\right)$ is assigned a $TS\left(\kappa, \nu, \alpha\right)$ prior, if $\sigma^2\left(0\right) = X$, simulating from the model requires simulation from

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

where

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

As j gets sufficiently large, $\Gamma\left(j\kappa+1\right)/j!$ ensures that the individual terms tend to zero (recall $0 < \kappa < 1$).

It has been shown that the *Stable* density ($f_{Y|\kappa,\nu}\left(x\right)$) can be represented as a one dimensional integral on a finite domain in Nolan (1997). Although this makes the numerical evaluation of the *Stable* density easier, it can still run into difficulties unless an advanced numerical integration algorithm is used. For certain values of κ, ν and x, the integrand becomes spikey and this can lead to inaccuracies in the density evaluation. The integral representation used in Nolan (1997) was implemented, but with Gaussian Quadrature (see Atkinson (1988) Appendix B.1), the inaccuracies sometimes proved significant in the MCMC implementation and so could not be reliably used.

Similar problems occur when evaluating the density by using the infinite sum and these can cause MCMC convergence problems. Numerical techniques are available to improve the convergence of the sum (see Higham (1996)) though these tend only to postpone the accuracy problems until the parameters become more extreme.

Graphs are given demonstrating the numerical problems which can occur when evaluating the infinite sum. All graphs are for a $TS\left(0.5, 1, 1\right)$ density. This *Tempered Stable*

distribution is the *Inverse Gaussian* distribution with unit parameters. The log of the absolute value of the terms is plotted for different values of x. Only odd terms are plotted, as even terms are zero for $\kappa = 0.5$.

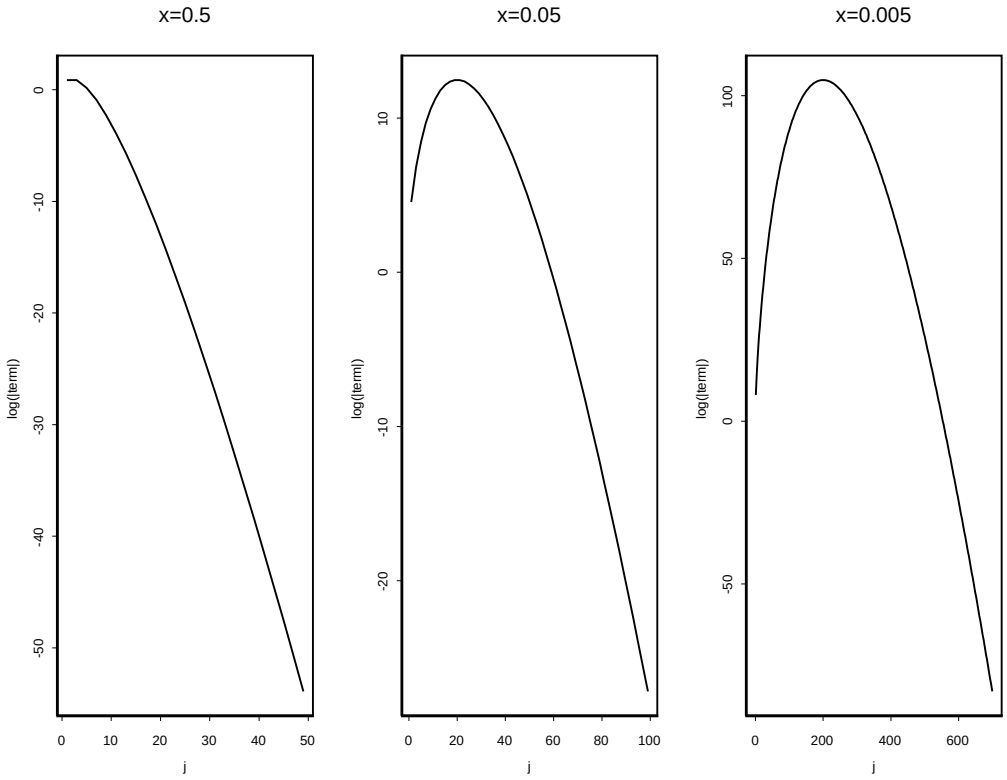


Figure B.1: Plots of the log of the absolute values of terms of the $TS\left(0.5, 1, 1\right)$ density sum for different arguments.

The terms often alternate in sign and, for small x, this can lead to large numerical errors because of the subtraction of large numbers of similar magnitude.

Instead, a series representation for $\sigma^2\left(0\right)$ is used. Details on possible series representations are given in Rosiński (2000) and Section 4.3.2. These are easier to implement accurately than using a Metropolis-Hastings update for $\sigma^2\left(0\right)$ with a $TS\left(\kappa, \nu, \alpha\right)$ prior.

Appendix C

Simulation Results

C.1 Sampling from the Cosh distribution

We wish to simulate a random variable, X, with density

$$f_X(x) = \frac{1}{\pi |a| \cosh\left(\frac{x-\mu}{a}\right)} f_X(x), \quad \text{for} \quad -\infty < x < \infty,$$

where $\mu, a \in \mathbb{R}$.

This has the same distribution as $X = \mu + aY$, where Y has density given by

$$f_Y(y) = \frac{1}{\pi \cosh(y)}, \quad \text{for} \quad -\infty < y < \infty.$$

Then $F_Y(y) = 2 \arctan(e^y)/\pi$ and to generate from Y, set $Y = \log\{\tan(\pi u/2)\}$, where $u \sim U(0,1)$.

C.2 Constant $\sigma^2(0\Delta)$

The following graphs are boxplots of σ_i^2 when a *Gamma* marginal is used on training data from the Black-Scholes model, with constant volatility $\sigma^2 = 0.5$. They demonstrate that, if $\sigma^2(0\Delta)$ is not set at the correct value, even on "simple" training data, the MCMC may not generate the σ_i^2 which might be expected.

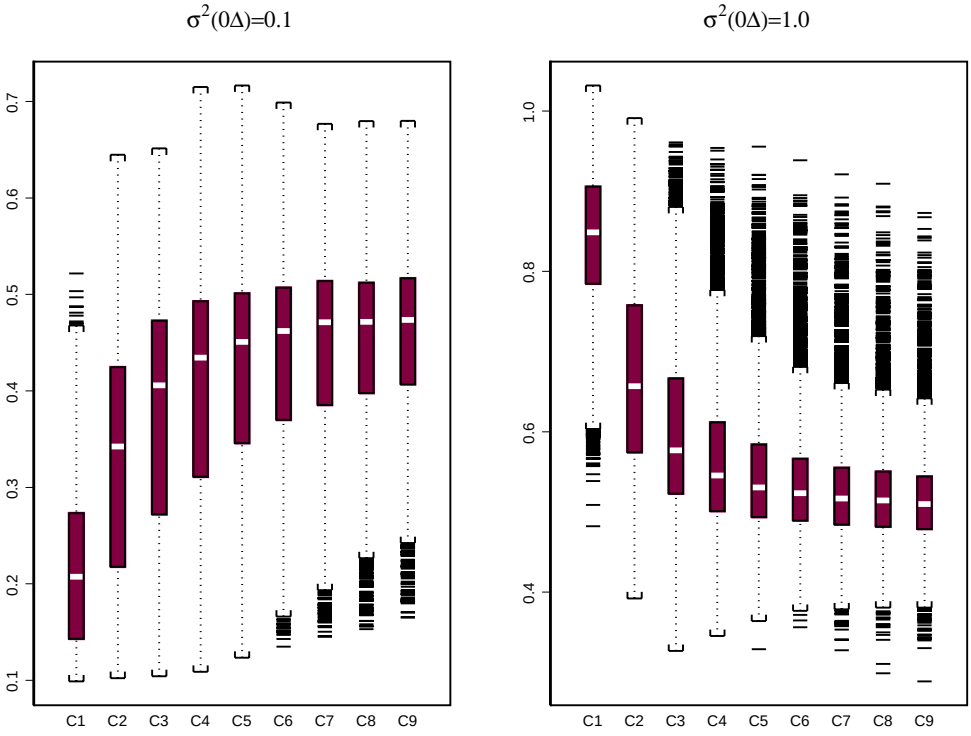


Figure C.1: Boxplots of $\sigma^2(t)$ for constant $\sigma^2(0\Delta) = 0.1$ and $\sigma^2(0\Delta) = 1.0$ for training data generated with $\sigma^2(t) = 0.5$.

For both $\sigma^2(0\Delta) = 0.1$ and $\sigma^2(0\Delta) = 1.0$, the $\sigma^2(i\Delta)$ converge to the expected constant 0.5 but for small i the choice of $\sigma^2(0\Delta)$ is very important. This becomes even more important when λ is small (as there is higher correlation in the volatility - see equation (4.1)). For constant volatility, the MCMC converges to larger values of λ than would be expected for real data and so this problem would be worse in reality than illustrated above.

In this test case, it would be feasible to estimate $\sigma^2(0\Delta)$ using some likelihood based method. This becomes more difficult if it is not known that the training data has constant volatility.

If $\sigma^2(0\Delta)$ was not treated as another parameter in the MCMC, it would not be obvious how many of the first σ_i^2 should be discarded before the solution becomes independent of the $\sigma^2(0\Delta)$ from which the chain was primed. Provided enough $\sigma^2(i\Delta)$ are discarded, the initial choice of $\sigma^2(0\Delta)$ will be unimportant. Typically there is a lot of available financial data, so the missing data approach used by Roberts et al. (2004) seems more favourable

than fixing $\sigma^2(0\Delta)$ as some constant as Griffin and Steel (2003) have done. However, the number of data points required before the process $\sigma^2(i\Delta)$ is stationary is highly influenced by λ and this can make the implementation of the method suggested by Roberts et al. (2004) difficult.

Even if the system is primed with the correct $\sigma^2(0\Delta)$, in practice, the mean of $\sigma^2(i\Delta)$ becomes stationary but the variance does not. Figure C.2 is a graph of $\sigma^2(i\Delta)$ for the same data used in the two boxplots in Figure C.1 but using $\sigma^2(0\Delta) = 0.5$.

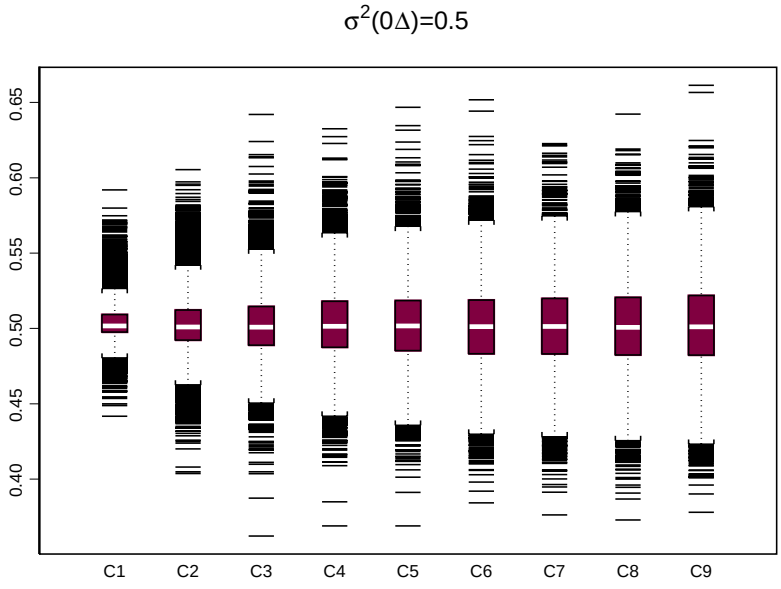


Figure C.2: Boxplot of $\sigma^2(t)$ for training data generated with $\sigma^2(t) = 0.5$.

The mean is constant for each $\sigma^2(i\Delta)$ but the variance is smaller for small i. This is because each volatility plays a part in generating the next volatility. If the variance of $\sigma^2(i\Delta)$ is small then the variance of $\sigma^2((i+1)\Delta)$ will also be small.

There are several possible ways to try to fix this problem. It could be assumed that $\sigma^2(0\Delta)$ was known sufficiently accurately and then a shift/scale transformation on the instantaneous volatility could be performed to try to make the mean and variance stationary. In practice, although some success can be gained doing this, there are problems knowing exactly how to pick the shift and scale change and $\sigma^2(0\Delta)$ must still be estimated accurately.

The second fix is to consider $\sigma^2(0\Delta)$ as another latent variable in the model and

perform an additional MCMC Metropolis-Hastings update for this point. Naïvely, a flat prior for $\sigma^2(0\Delta)$ could be used. This works to some extent - the correct mean for $\sigma^2(1\Delta)$ is obtained but the variance of the initial volatilities tends to be large. On closer inspection, a flat prior does not seem totally logical, as the marginal distribution ($Ga(\nu, \alpha)$ say) for the volatility is already known and a prior more in agreement with this would seem more sensible. When a $Ga(\nu, \alpha)$ prior is used for $\sigma^2(0\Delta)$ on the above training data, the boxplot of $\sigma^2(i\Delta)$ is given in Figure C.3.

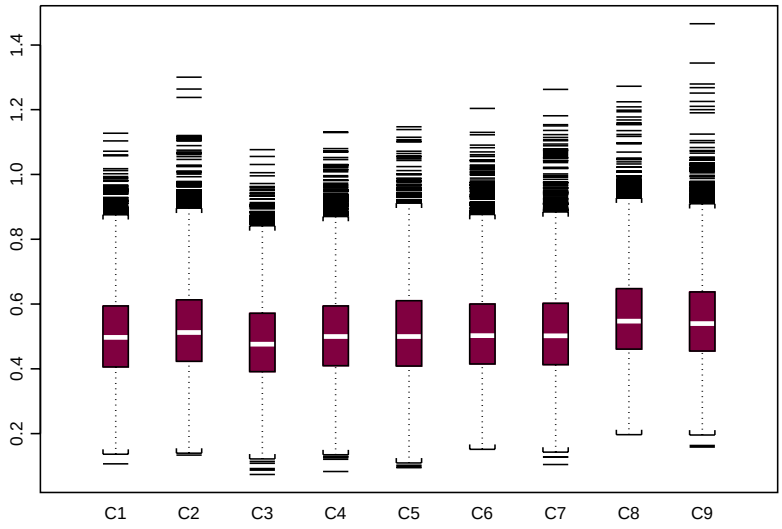


Figure C.3: Boxplot of $\sigma^2(i\Delta)$ on Black-Scholes data when a $Ga(\nu, \alpha)$ prior is used for $\sigma^2(0\Delta)$.

Here the mean and variance are stationary and the correct constant volatility is obtained. The instantaneous volatility at time 0, $\sigma^2(0\Delta)$, was therefore treated as a latent parameter in the model, with a prior the same as the marginal distribution of $\sigma^2(t)$.

C.3 Implied prior for x_1 and x_2

Recall x_1 and x_2 are defined as

$$\begin{aligned} x_1 &= \alpha^{c_1}\nu^{c_2} \\ x_2 &= \alpha^{c_3}\nu^{c_4}. \end{aligned}$$

Let $c_1c_4 - c_2c_3 = K$, so

$$\nu^{-K} = \frac{x_1^{c_3}}{x_2^{c_1}}$$

and

$$\alpha^K = \frac{x_1^{c_4}}{x_2^{c_2}}.$$

It is required that $c_1c_4 - c_2c_3 \neq 0$, so the transformation can be inverted. Then

$$\begin{aligned}\nu &= \left(\frac{x_1^{c_3}}{x_2^{c_1}}\right)^{-1/K} \\ \alpha &= \left(\frac{x_1^{c_4}}{x_2^{c_2}}\right)^{1/K}\end{aligned}$$

and

$$p(x_1,x_2) = p(\nu(x_1,x_2),\alpha(x_1,x_2))\,\|J\|,$$

where

$$\begin{aligned}|J| &= \begin{vmatrix} \frac{\partial\nu}{\partial x_1} & \frac{\partial\nu}{\partial x_2} \\ \frac{\partial\alpha}{\partial x_1} & \frac{\partial\alpha}{\partial x_2} \end{vmatrix} \\ &= \frac{1}{K^2}\begin{vmatrix} c_3x_1^{-c_3/K-1}x_2^{c_1/K} & c_1x_1^{-c_3/K}x_2^{c_1/K-1} \\ c_4x_1^{c_4/K-1}x_2^{-c_2/K} & c_2x_1^{c_4/K}x_2^{-c_2/K-1} \end{vmatrix} \\ &= \frac{1}{K}\left[x_1^{(c_4-c_3)/K-1}x_2^{(c_1-c_2)/K-1}\right] \\ &= \frac{1}{K}\alpha^{1-c_1-c_3}\nu^{1-c_2-c_4}\end{aligned}$$

and $\alpha = \alpha(x_1,x_2)$ and $\nu = \nu(x_1,x_2)$. Then

$$\frac{p\left(x_1',x_2'\right)}{p(x_1,x_2)} = \frac{p\left(\nu'\left(x_1',x_2'\right),\alpha'\left(x_1',x_2'\right)\right)}{p(\nu(x_1,x_2),\alpha(x_1,x_2))}\left\{\frac{\alpha'\left(x_1',x_2'\right)}{\alpha(x_1,x_2)}\right\}^{1-c_1-c_3}\left\{\frac{\nu'\left(x_1',x_2'\right)}{\nu(x_1,x_2)}\right\}^{1-c_2-c_4}.$$

C.4 Prior for κ for the $TS(\kappa,\nu,\alpha)$ marginal distribution

The $TS(\kappa,\nu,\alpha)$ distribution has mean

$$M = 2\nu\kappa\alpha^{1-1/\kappa}$$

and variance

$$V = 4\kappa(1-\kappa)\nu\alpha^{1-2/\kappa},$$

where $0 < \kappa < 1$ and $\nu,\alpha > 0$. The density of the $TS(\kappa,\nu,\alpha)$ distribution is given in equation (2.3) and is not easy to interpret. For this reason, the prior for κ is based on the mean and variance, as well as simulation arguments, which are now discussed.

The mean and variance of the $TS(\kappa, \nu, \alpha)$ distribution are rescaled by ν and so we will not concern ourselves with this parameter or multiplicative constants contributing to the mean or variance. For all valid parameter values we have

$$\lim_{\kappa \to 1} M = 2\nu.$$

Consider the two cases:-

a) $\alpha < 1$

Then

$$\lim_{\kappa \to 0} M = \infty$$

and

$$\lim_{\kappa \to 0,1} V = \infty.$$

b) $\alpha \geq 1$

Then

$$\lim_{\kappa \to 0} M = 0$$

and

$$\lim_{\kappa \to 0,1} V = 0.$$

In practice, the MCMC is able to reject moves to very small κ values as the mean of the marginal distribution becomes 0 or ∞ and the likelihood rejects these moves. For moves proposed to large κ values, when $\alpha < 1$, the variance of the marginal becomes infinite and so the MCMC is also able to reject moves to these states. However, when $\alpha \geq 1$, moves to large κ values generate a volatility process which is concentrated at $\sigma^2(t) = 2\nu$ and the MCMC is not able to guarantee rejection of these moves from the likelihood alone. This constant volatility model can be generated from the $TS\left(\frac{1}{2}, \nu, \alpha\right)$ distribution for suitably large ν and α. If κ gets too close to 1, then too many terms are required in the summation to feasibly store them (see Figure 4.1). This is why, for observed data, an informative prior for κ is used, keeping it away from 1. For example in Section 4.4.5, where the S&P 500 data set is fitted, the prior is $Beta(1, \kappa 1 = 15)$. A prior which does not support both $\kappa = 0$ and $\kappa = 1$ could also be used, such as $Beta(\kappa 2, \kappa 1)$, where $\kappa 2 > 1$, though this does not alter the simulation results noticeably as the likelihood is already able to reject moves to small κ values.

C.5 Additional details of the MCMC algorithm

From Section 4.3, the MCMC algorithm is of the form:-

(1) γ, κ move (reverse jump) (if a *Generalised Inverse Gaussian* or *Tempered Stable* marginal is used).

(2) x_1 move (reverse jump).

(3) x_2 move (reverse jump).

(4) λ move (reverse jump) or λ_0 move if we have a stochastic λ process as in Section 3.3.

(5) ε^2 move (fixed dimension).

(6) r move (fixed dimension).

(7) N_2 move (reverse jump).

(8) Update the jump times of the stochastic λ process (reverse jump).

(9) Update the jump sizes of the stochastic λ process (reverse jump).

(10) H move (fixed dimension).

(11) $\sigma^2(0\Delta)$ move (fixed dimension) (if we are not using the *Tempered Stable* marginal).

(12) μ move (fixed dimension).

(13) ρ move (fixed dimension).

(14) Joint A and R move (fixed dimension). If we are using the *Tempered Stable* marginal this will update $\sigma^2(0\Delta)$ using equation (4.19).

Let the prior for state s be $p(s)$ and propose a move up in dimension from state $s = (s_1, \ldots, s_{d_1})$ to $s' = \left(s'_1, \ldots, s'_{d_1+d_2}\right)$, with probability $q\left(s \to s'\right)$. Let $l\left(y_i|\sigma_i^2\right)$ be the likelihood of the data given the discretely observed volatility process σ_i^2.

For (5), (6), (10), (11), (12), (13) and (14) (fixed dimension moves), $d_2 = 0$, and the acceptance probability is

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

For updates which alter the accuracy of the finite truncation of the infinite sum of the random shock vector given in equation (4.9) (i.e. moves which alter the critical value, a_c, at which the infinite sum is truncated (see Section 4.3.5 and Appendix B.2)), reverse jump MCMC is used to add or remove points from the Poisson point processes, $a_1, \ldots, a_T$, and the uniform random variables, $r_1, \ldots, r_T$ at the same time as the parameter update.

Consider the case when a move up from dimension d_1 to $d_1 + d_2$ is proposed, with probability $p(d_1 \to d_1 + d_2)$. Generate random variables $u \sim q(u)$ and let $g(s, u)$ be the function which generates the new state $s^{'}$ (i.e. $s^{'} = g(s, u)$). The move is accepted with probability

$$\min\left[1, \frac{l\left(y_i|\left(\sigma_i^2\right)^{'}\right)}{l\left(y_i|\sigma_i^2\right)} \frac{p\left(s^{'}\right)}{p(s)} \frac{q\left(s^{'} \to s\right)}{q\left(s \to s^{'}\right)} \frac{1}{q(u)} \left|\left(\frac{\partial s^{'}}{\partial(s,u)}\right)\right|\right], \tag{C.1}$$

where elements of the Jacobian matrix are given by

$$\left(\frac{\partial s^{'}}{\partial(s,u)}\right)_{i,j} = \frac{\partial s_i^{'}}{\partial s_j}, \qquad \text{for} \qquad 1 \leq j \leq d_1$$

$$\left(\frac{\partial s^{'}}{\partial(s,u)}\right)_{i,j} = \frac{\partial s_i^{'}}{\partial u_{j-d_1}}, \qquad \text{for} \qquad d_1 < j \leq d_2.$$

The acceptance probability for the jump down in dimension from state $s^{'}$ to s is

$$\min\left[1, \frac{l\left(y_i|\sigma_i^2\right)}{l\left(y_i|\left(\sigma_i^2\right)^{'}\right)} \frac{p(s)}{p\left(s^{'}\right)} \frac{q\left(s \to s^{'}\right)}{q\left(s^{'} \to s\right)} q(u) \frac{1}{\left|\left(\frac{\partial s^{'}}{\partial(s,u)}\right)\right|}\right]$$

because of the nature of reverse jump MCMC. Individual acceptance probabilities are now given for each of the updates listed above.

C.5.1 (2) x_1 update

Let $x_1^{'}|x_1$ be generated with probability $q\left(x_1 \to x_1^{'}\right)$. The new critical value, $a_c^{'}$, at which the Poisson points are truncated (see Section 4.3.5 and Appendix B.2), is calculated given $x_1^{'}$. For $a_c^{'} > a_c$, $N \sim Po\left(a_c^{'} - a_c\right)$ new Poisson points are generated, for each Poisson process, as the order statistics of N uniform random variables in $\left(a_c, a_c^{'}\right)$. The random uniforms are generated direct from their prior. As the new Poisson points and uniforms are direct from their priors, many of the terms in equation (C.1) cancel. The acceptance probability is

$$\min\left[1, \frac{l\left(y_i|\left(\sigma_i^2\right)^{'}\right)}{l\left(y_i|\sigma_i^2\right)} \frac{p\left(x_1^{'}\right)}{p(x_1)} \frac{q\left(x_1^{'} \to x_1\right)}{q\left(x_1 \to x_1^{'}\right)}\right].$$

The removal of Poisson points and uniforms is deterministic (when $a_c^{'} < a_c$) because of the nature of reverse jump MCMC, with the acceptance probability as described previously. The acceptance probabilities for (1), (3) and (4) are similar.

For (14), as the Poisson points and uniforms are generated direct from their prior, the acceptance probability is

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

When λ is allowed to vary over time, the acceptance probabilities are more complicated. Extra details for these moves are now given.

C.5.2 (5) ε^2 update

The acceptance probability is

$$\min\left[1, \frac{p\left(\varepsilon^{2'}|r\right)}{p\left(\varepsilon^2|r\right)} \frac{p\left(\lambda_t|\lambda_0, \varepsilon^{2'}, N_2\right)}{p\left(\lambda_t|\lambda_0, \varepsilon^2, N_2\right)} \frac{q\left(\varepsilon^{2'} \to \varepsilon^2\right)}{q\left(\varepsilon^2 \to \varepsilon^{2'}\right)}\right],$$

where $p\left(\lambda_t|\lambda_0, \sigma, N\right)$ is given in equation (3.7). As a move in ε^2 does not explicitly alter the stochastic λ_t process, the volatility process is unaltered.

C.5.3 (6) r update

The acceptance probability is

$$\min\left[1, \frac{p\left(r^{'}\right)}{p\left(r\right)} \frac{p\left(\varepsilon^2|r^{'}\right)}{p\left(\varepsilon^2|r\right)} \frac{p\left(N_2|r^{'}\right)}{p\left(N_2|r\right)} \frac{q\left(r^{'} \to r\right)}{q\left(r \to r^{'}\right)}\right],$$

where $N_2|r \sim Po\left(r\right)$. As a move in r does not explicitly alter the stochastic λ_t process, the volatility process is unaltered.

C.5.4 (7) N_2 update

If $N_2 = 0$, a move up in dimension is proposed, whilst if $N_2 = T - 1$ a move down in dimension is proposed. When adding in an extra jump, the new jump is proposed uniformly at times where there are no current jumps (removal of jumps is also proposed uniformly). The inclusion of a new jump at time i randomly alters λ to the left or right of the new jump time, only altering λ up until the point where it reaches the previous/next jump.

Originally, the jump removal algorithm uniformly proposed to remove a jump and altered λ from where the jump was removed until the next jump (so the λ process was unaltered apart from in one segment to the right of where the jump was removed). This leads to mixing problems in the chain because it only removes a jump and alters the future λ values after this jump time. A proposal which is able to alter λ before or after the removed jump time was found to be more successful. Again, a jump is randomly selected to be removed uniformly and λ values to the left or right of this jump are altered with equal probability.

$\lambda_i^{'}$ is generated from a $Ga\left(c, c/\lambda_{i-1}\right)$ distribution and again the Poisson points and uniforms are proposed direct from their priors, given the change in λ_t (c was tuned to give a suitable acceptance rate). The acceptance probability for a move up in dimension is

$$\min\left[1, \frac{l\left(y_i|\left(\sigma_i^2\right)^{'}\right)}{l\left(y_i|\sigma_i^2\right)} \frac{p\left(N_2^{'}|r\right)}{p\left(N_2|r\right)} \frac{p\left(\lambda_t^{'}|\lambda_0, \varepsilon^2, N_2^{'}\right)}{p\left(\lambda_t|\lambda_0, \varepsilon^2, N_2\right)} \frac{q\left(s^{'} \rightarrow s\right)}{q\left(s \rightarrow s^{'}\right)}\right]$$

and the move down in dimension is specified by this as described in Section 1.4.1.

C.6 Proposals for parameters on finite domains

If the volatility has a *Tempered Stable* marginal distribution, the parameter controlling the tail behaviour of the distribution must satisfy $0 < \kappa < 1$. For the MVN approximation to fBm, the Hurst parameter, H, must satisfy $0.5 < H < 1$ (assuming we wish to induce positive correlation in the fBm motion).

Consider a random variable, X, restricted to $a < X < b$, then

$$0 < Y = \frac{b - X}{X - a} < \infty.$$

Generate $Y^{'} \sim Ga\left(c, c/Y\right)$ so $E\left[Y^{'}|Y\right] = Y$ and $Var\left[Y^{'}|Y\right] = Y^2/c$. The constant c controls the variance of the proposal and how local the move in X is. Then

$$\begin{aligned} f_{X^{'}|X}\left(x^{'}|x\right) &= \frac{\left(1+y^{'}\right)^2}{|a-b|} f_{Y^{'}|Y}\left(y^{'}|y\right) \\ &\propto \frac{\left(1+y^{'}\right)^2}{(y)^c}\left(y^{'}\right)^{c-1} \exp\left\{-\frac{c}{y}y^{'}\right\}, \end{aligned}$$

so

$$\frac{q\left(X^{'} \rightarrow X\right)}{q\left(X \rightarrow X^{'}\right)} = \left(\frac{1+y}{1+y^{'}}\right)^2 \left(\frac{y}{y^{'}}\right)^{2c-1} \exp\left\{c\left(\frac{y^{'}}{y} - \frac{y}{y^{'}}\right)\right\}.$$

C.7 Different option payoffs for the battery of tests

The different payoffs used for Section 5.2.4 are all based on the daily open price of the asset and are:-

Very standard options:-

(1) European vanilla call, with exercise price, $E = 0.99 * S_0$.

(2) European binary call, with exercise price, $E = S_0$.

Asian options:-

(3) Arithmetic Asian.

(4) Geometric Asian.

Knock in options:-

(5) Knock in at $1.005 * S_0$ then European call, with exercise price, $E = S_0$.

(6) Knock in at $0.995 * S_0$ then European put, with exercise price, $E = S_0$.

(7) Knock in at $0.995 * S_0$ then binary call, with exercise price, $E = S_0$.

(8) Knock in at $0.995 * S_0$ then binary put, with exercise price, $E = S_0$.

(9) Knock in at $0.995 * S_0$ then arithmetic Asian.

(10) Knock in at $0.995 * S_0$ then geometric Asian.

(11) Knock in with outstrike. Option is knocked out if the asset hits $1.05 * S_0$ otherwise it must be knocked in by hitting $1.005 * S_0$, then payoff is a European call with exercise price, $E = S_0$.

(12) Knock in with outstrike. Option is knocked out if the asset hits $1.05 * S_0$ otherwise it must be knocked in by hitting $1.005 * S_0$, then payoff is a European put with exercise price, $E = S_0$.

Knock out options:-

(13) European call with exercise price, $E = S_0$ and with knock out at $0.99 * S_0$.

(14) European put with exercise price, $E = S_0$ and with knock out at $0.97 * S_0$.

(15) Binary call with exercise price, $E = S_0$ and with knock out at $0.995 * S_0$.

(16) Binary put with exercise price, $E = S_0$ and with knock out at $0.99 * S_0$.

(17) Arithmetic Asian with knock out at $0.995 * S_0$.

(18) Geometric Asian with knock out at $0.995 * S_0$.

(19) Double knock out option. Knock out if S_0 hits either of the barriers $0.97 * S_0$ or $1.05 * S_0$ otherwise European call with exercise price, $E = S_0$.

(20) Double knock out option. Knock out if S_0 hits either of the barriers $0.95 * S_0$ or $1.03 * S_0$ otherwise European put with exercise price, $E = S_0$.

Parisian options:-

(21) Proportion of days the asset is greater than $1.01 * S_0$.

(22) Proportion of days the asset is less than $1.01 * S_0$.

(23) Proportion of days the asset is between $0.98 * S_0$ and $1.02 * S_0$.

Lookback options:-

(24) Lookback for vanilla call, with exercise price, $E = S_0$.

(25) Lookback for vanilla put, with exercise price, $E = S_0$.

(26) Lookback for binary call, with exercise price, $E = S_0$.

(27) Lookback for binary put, with exercise price, $E = S_0$.

C.8 Sampling from the Tempered Stable distribution

Option pricing when the volatility has a *Tempered Stable* marginal distribution, requires samples from the distribution with density

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

where $f_{Y|\kappa,\nu}$ is the positive $\kappa - Stable$ density.

It is relatively unimportant that the generation is very efficient and, as the density for $f_{Y|\kappa,\nu}$ is not straightforward or easy to interpret, consider rejection using the positive κ-*Stable* density (to generate X from the positive κ-*Stable* density see Section C.9).

The algorithm is:-

(1) Set $M = \max_{x>0} \frac{f_X(x)}{g_X(x)} = e^{\nu\alpha}$.

(2) Generate a positive $\kappa - Stable$ random variable, Y, using the method given in Appendix C.9 and a $U(0,1)$ variate, U.

(3) if $\frac{f_X(y)}{M g_Y(y)} = \exp\left(-\alpha^{1/\kappa} y/2\right) > U$, set $X = Y$, else go to (2).

As both $f_X(x)$ and $g_Y(y)$ are normalised pdfs, the acceptance probability is $\frac{1}{M} = e^{-\nu\alpha}$ and is most efficient for "small" $(\nu\alpha)$.

C.9 Sampling from the positive κ-Stable distribution

Appendix C.8 requires samples from the *Tempered Stable* distribution. This can be done using rejection sampling, provided it is known how to sample from the positive $\kappa - Stable$ distribution. To generate from the *positive Stable* distribution, the method of Chambers et al. (1971) was used.

For the representation used in Feller (1971), the characteristic exponent is

$$\Psi(\lambda) = -|\lambda|^{\alpha} e^{i\pi\gamma}/2, \quad \text{for} \quad \lambda > 0$$

and in the Chambers et al. (1971) representation for the $Stable(\alpha,\beta)$ distribution the characteristic exponent is

$$\Psi(\lambda) = -|\lambda|^{\alpha} e^{i\pi/2 - \alpha\beta}, \quad \text{for} \quad \lambda > 0, \quad 0 < \kappa < 1.$$

We need $\gamma = -\alpha\beta$ for these to be the same. Recalling $\alpha = \kappa$ and $\gamma = -\kappa$ (see Appendix A.3), it can be seen that $\beta = 1$.

For the Chambers et al. (1971) representation, sampling from the $Stable(\kappa,1)$ distribution, for $0 < \kappa < 1$, is straightforward. This will have the same distribution as a $Stable(\kappa,-\kappa)$ distribution in the representation used by Feller (1971). So, to generate X from the positive κ-*Stable* distribution with density

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

generate $Y \sim Stable(\kappa, 1)$, using the notation and algorithm given in Chambers et al. (1971), then set $X = 2\nu^{1/\kappa}Y$ (see Appendix A.3).

C.10 95% credible intervals for $\sigma^2(t)$ for four marginals on S&P 500 data

The MCMC was run on the S&P 500 data set for the *Gamma*, *Inverse Gaussian*, *Positive Hyperbolic* and *Inverse Gamma* marginal distributions. 95% credible intervals for each marginal were then calculated using the MCMC output. The x-axis is σ^2 and y-axis is $f_{\Sigma^2}(\sigma^2)$.

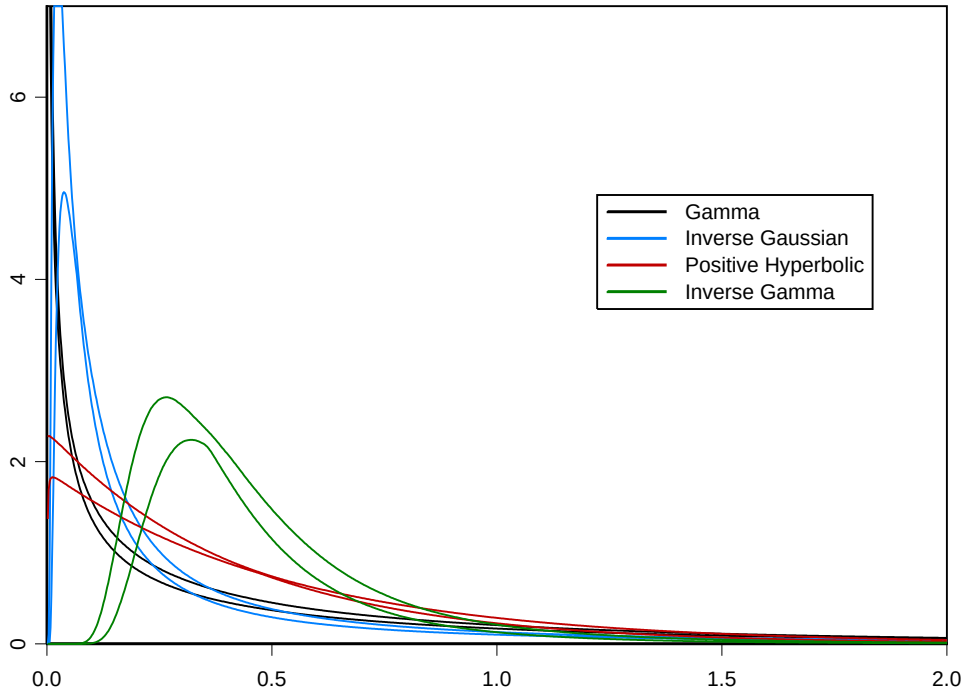


Figure C.4: 95% credible intervals for the volatility process of four different marginal distributions applied to the S&P 500 data set.

Although each MCMC run gives almost the same mean for $\sigma^2(t)$, the marginals are noticeably different. The *Generalised Inverse Gaussian* and *Tempered Stable* distributions are not included to preserve interpretability of the graph and because of numerical difficulties

evaluating the *Tempered Stable* distribution near the origin.

C.11 Results for Black-Scholes option pricing on constant volatility data

Graphs are displayed of the numerical evaluation of the fair price for the Test example options (see Section 5.2.3) on training data generated with constant volatility.

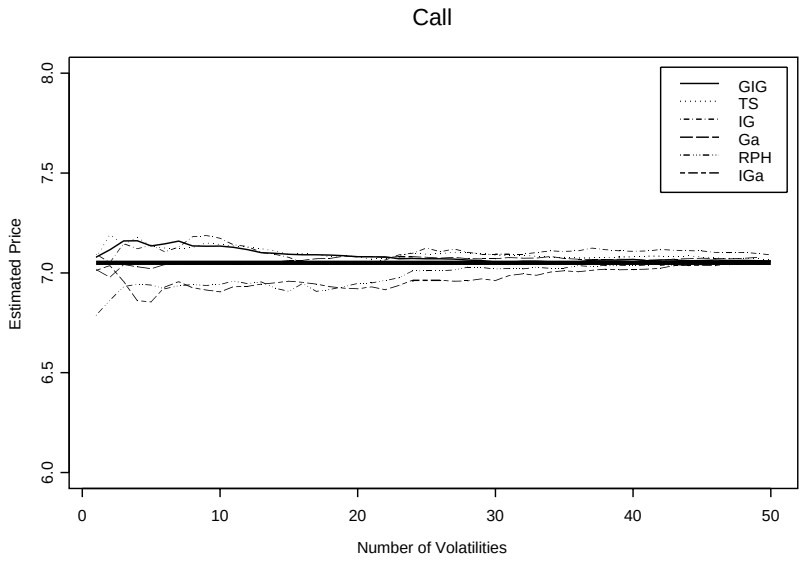


Figure C.5: Graphs of the estimated fair price of a vanilla call for constant volatility, $\sigma = 0.03$.

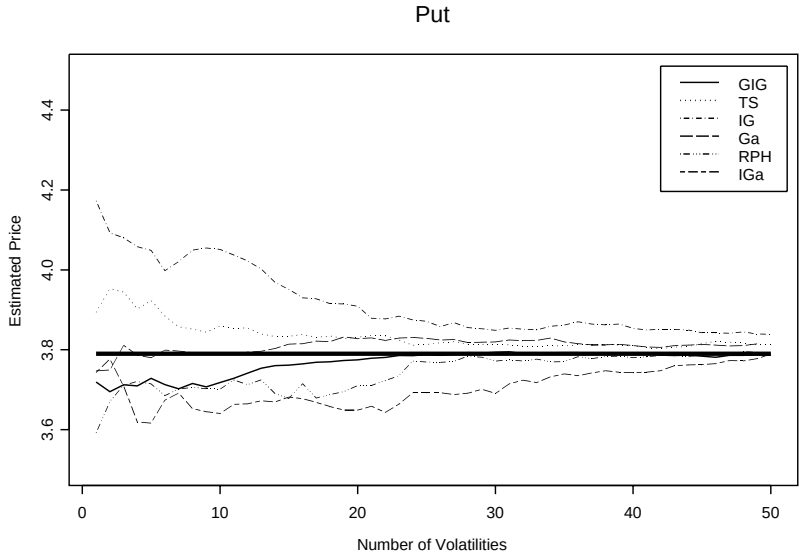


Figure C.6: Graphs of the estimated fair price of a vanilla put for constant volatility, $\sigma = 0.03$.

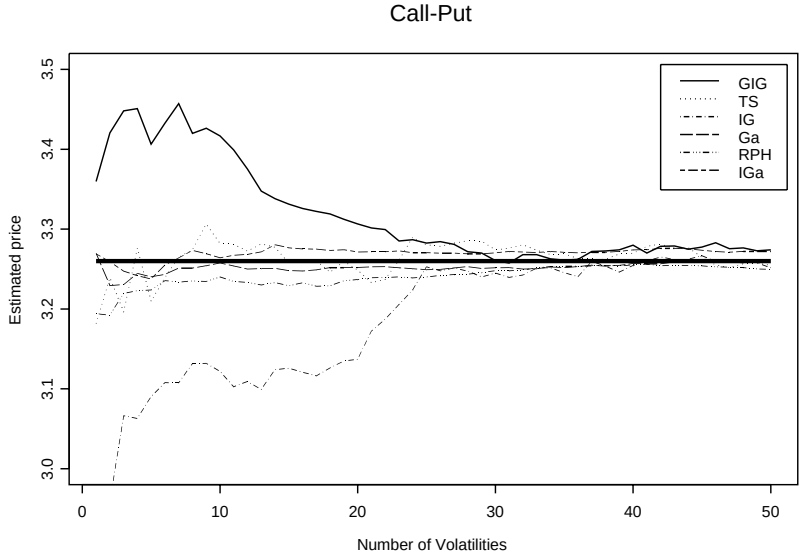


Figure C.7: Graphs of the estimated fair price of a vanilla call-put for constant volatility, $\sigma = 0.03$.

The thick line is the expected result knowing the correct constant value for σ^2 and demonstrates the correct implementation of the option pricing algorithm.

Appendix D

Theory behind solutions of the Ornstein-Uhlenbeck equation

This appendix describes some of the more technical details and theorems behind the solution to the Ornstein-Uhlenbeck equation given in Sections 1.3.2 and 3.2. Most of the proofs and theorems of this appendix can be found in Barndorff-Nielsen (1998), Barndorff-Nielsen and Shephard (2000) and Barndorff-Nielsen and Shephard (2001b) and are included here for completeness. If the reader does not wish to get involved in the specific details of the solution to the Ornstein-Uhlenbeck equation, this Appendix can be skipped; it is here to provide more details on some of the results which are quoted in earlier chapters.

D.1 Existence of solutions to the Ornstein-Uhlenbeck equation

Using Theorem 1, the distribution x is self-decomposable if and only if there is a stochastic process $\sigma^2(t)$ that has the same distribution as x and, for all $\lambda > 0$, can be written as

$$\sigma^2(t) = \int_{-\infty}^{0} \exp(s)\, dz(\lambda t + s), \tag{D.1}$$

where $z\left(t'\right)$ is a homogeneous Lévy process (see Definition 14). Note that solutions of this type are stationary, as the Lévy process, $z\left(t'\right)$, is homogeneous and the stochastic integral is only determined by the length of the range over which $\exp(s)$ is integrated with respect to $z\left(t'\right)$ and this is constant for each value of t.

The returns of financial series are often rescaled so they are a reasonable size and so it is attractive for volatility to have a self-decomposable distribution, as the marginal distri-

bution is altered in a predictable way by rescaling. Any such self-decomposable marginal distribution for the volatility process, $\sigma^2(t)$, can be written in the form of equation (D.1). Instead of considering this type of volatility process directly, consider

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

This process was suggested (but not examined in detail, for reasons which were discussed in Chapter 6) in Barndorff-Nielsen and Shephard (2001b). Manipulating this equation gives

$$\begin{aligned}\sigma^2(t) &= \int_{-\infty}^{0} f(\lambda s)\, dz(\lambda(t+s)) \\ &= \int_{-\infty}^{0} f\{\lambda(s-t)\}\, dz(\lambda s) + \int_{0}^{t} f\{\lambda(s-t)\}\, dz(\lambda s),\end{aligned}$$

where $f(s)$ is a function such that the integrals exists. When $f(s) = e^s$, the distribution of $\sigma^2(t)$ is self-decomposable by Theorem 1. If there is no such Lévy process with $f(s) = e^s$, then the marginal distribution of $\sigma^2(t)$ is not self-decomposable. Assume that

$$f(s_1 + s_2) = f(s_1) f(s_2),$$

which implies

$$f(s) = k^s, \qquad \text{for some } k \in \mathbb{R}^+$$

and therefore

$$\begin{aligned}\sigma^2(t) &= k^{-\lambda t} \int_{-\infty}^{0} k^{\lambda s} dz(\lambda s) + \int_{0}^{t} k^{\lambda(s-t)} dz(\lambda s) \\ &= k^{-\lambda t}\sigma^2(0) + k^{-\lambda t} \int_{0}^{t} k^{\lambda s} dz(\lambda s). \end{aligned} \tag{D.2}$$

In the absence of any jumps in the Lévy process, we would like the contribution of $\sigma^2(0)$ to $\sigma^2(t)$ to decrease as t increases, so require $k > 1$. Differentiating this gives

$$\begin{aligned}d\sigma^2(t) &= -\lambda \log(k) k^{-\lambda t}\sigma^2(0)\, dt + dz(\lambda t) - \lambda \log(k) k^{-\lambda t} dt \int_{0}^{t} k^{\lambda s} dz(\lambda s) \\ &= -\lambda \log(k)\sigma^2(t)\, dt + dz(\lambda t), \end{aligned} \tag{D.3}$$

Therefore $\sigma^2(t)$ is an Ornstein-Uhlenbeck process (see Definition 16). If $z()$ is restricted to be a subordinator (i.e. a Lévy process with non-negative jumps), then jumps in $z()$ cause positive jumps in the volatility. When there are no jumps in $(t, t+s)$

$$\sigma^2(t+s) = \sigma^2(t) k^{-\lambda s}$$

and $\sigma^2(t)$ is a positive process (as required for volatility models).

To sum up so far: Equation (D.3) has solution given by equation (D.2) and this solution is stationary. Further, if $k = e$, then the marginal distribution of $\sigma^2(t)$ is self-decomposable and for any self-decomposable distribution, there exists a corresponding homogeneous Lévy process which will generate this marginal distribution. The relationship between $\sigma^2(t)$ and $z(t)$ is now given.

Barndorff-Nielsen (1998) has generalised some of the results in this Chapter so they hold on the real line (i.e. when the OU equation is not driven by a subordinator). Our focus is on volatility modelling, which requires $\sigma^2(t) > 0$, so these generalisations are not included.

D.2 Relationship between the volatility and the BDLP

Jurek and Vervaat (1983) proved the following theorem.

Theorem 5 *A random variable x is self-decomposable if and only if*

$$x \overset{\mathfrak{L}}{=} \int_0^\infty e^{-t} dz(t),$$

for some homogeneous Lévy process $z(t)$.

Further, if $u(x)$ and $w(x)$ are the Lévy measures of x and $z(1)$ respectively, then

$$U(a,b) = \int_0^\infty W\left\{e^t(a,b)\right\} dt,$$

where $b \geq a$ and

$$U(a,b) = \int_a^b u(x)\, dx$$

and

$$W\left\{f(t)(a,b)\right\} = \int_a^b f(t)\, w(x)\, dx.$$

So, for $x > 0$,

$$U([x,\infty)) = \int_0^\infty W\left(e^t[x,\infty)\right) dt$$

and letting $s = e^t$ gives

$$U([x,\infty)) = \int_1^\infty s^{-1} W([sx,\infty))\, ds,$$

so

$$\int_x^\infty u(s)\, ds = \int_x^\infty s^{-1} W([s,\infty))\, ds.$$

Differentiating this with respect to x gives

$$u(x) = x^{-1} W([x, \infty)),$$

so

$$W([x, \infty)) = W^{+}(x) = \int_{x}^{\infty} w(s)\, ds = x u(x)$$

and a further differentiation yields

$$w(x) = -u(x) - x \frac{du(x)}{dx}. \tag{D.4}$$

As the solution to equation (D.1) is stationary, the Lévy measure of $\sigma^2(t)$ can be chosen and this specifies Lévy measure, $w(x)$, of $z(1)$. Consider $\sigma^2(0)$ in equation (D.1) and, comparing it with Theorem 5, it can be seen that the Lévy measure of $\sigma^2(0)$ and $z(1)$ specify each other through equation (D.4). Further, as the Lévy process is homogeneous, the entire Lévy process is specified by the Lévy measure of $z(1)$, which is specified by the marginal distribution of $\sigma^2(t)$.

D.3 Equations for the integrated and discretely observed volatility

From equation (D.2), it follows that

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

which satisfies equation (D.3), that is

$$d\sigma^2(t) = -\lambda \log(k)\, \sigma^2(t)\, dt + dz(\lambda t).$$

The integrated volatility is

$$\begin{aligned} \sigma^{2*}(t) &= \int_0^t \sigma^2(u)\, du \\ &= \frac{\sigma^2(0)}{\lambda \log(k)} \left(1 - k^{-\lambda t}\right) + \int_0^t k^{-\lambda u} \int_0^u k^{\lambda s} dz(\lambda s)\, du. \end{aligned}$$

Switching the order of integration (taking care with the ranges) gives

$$\begin{aligned} \sigma^{2*}(t) &= \frac{\sigma^2(0)}{\lambda \log(k)} \left(1 - k^{-\lambda t}\right) + \int_0^t \int_0^s k^{-\lambda u} du k^{\lambda s} dz(\lambda s) \\ &= \frac{1}{\lambda \log(k)} \left\{ \sigma^2(0) \left(1 - k^{-\lambda t}\right) + \int_0^t \left(k^{-\lambda s} - k^{-\lambda t}\right) k^{\lambda s} dz(\lambda s) \right\} \\ &= \frac{1}{\lambda \log(k)} \left\{ \sigma^2(0) \left(1 - k^{-\lambda t}\right) + \int_0^t dz^{(\lambda)}(s) - \int_0^t k^{\lambda(s-t)} dz(\lambda s) \right\}. \end{aligned}$$

Using equation (D.2),

$$\begin{aligned}\sigma^{2*}(t) &= \frac{1}{\lambda \log(k)}\left\{\sigma^2(0)\left(1-k^{-\lambda t}\right)+z(\lambda t)-\sigma^2(t)+k^{-\lambda t}\sigma^2(0)\right\} \\ &= \frac{1}{\lambda \log(k)}\left\{z(\lambda t)-\sigma^2(t)+\sigma^2(0)\right\}.\end{aligned}$$

This quantity is important in pricing European options (see Hull and White (1987)). This relatively simple form for the integrated volatility is an attractive feature of the BNS SV models.

If observations are separated by Δ days, the discretely observed or actual volatility is

$$\begin{aligned}\sigma_i^2 &= \sigma^{2*}(i\Delta)-\sigma^{2*}((i-1)\Delta) \\ &= \frac{1}{\lambda \log(k)}\left\{z(\lambda i\Delta)-\sigma^2(i\Delta)-z(\lambda(i-1)\Delta)+\sigma^2((i-1)\Delta)\right\},\end{aligned}$$

where the Lévy process, $z(t)$, satisfies

$$\begin{aligned}z(\lambda i\Delta) &= z(\lambda(i-1)\Delta)+\int_{\lambda(i-1)\Delta}^{\lambda i\Delta} dz(t) \\ &\overset{\mathfrak{L}}{=} z(\lambda(i-1)\Delta)+\int_0^{\lambda\Delta} dz(t)\end{aligned}$$

and, for the instantaneous volatility,

$$\begin{aligned}\sigma^2(i\Delta) &= k^{-\lambda\Delta}\sigma^2((i-1)\Delta)+k^{-\lambda\Delta}\int_{\lambda(i-1)\Delta}^{\lambda i\Delta} k^t dz(t) \\ &\overset{\mathfrak{L}}{=} k^{-\lambda\Delta}\sigma^2((i-1)\Delta)+k^{-\lambda\Delta}\int_0^{\lambda\Delta} k^t dz(t).\end{aligned}$$

D.4 Infinite series representation for stochastic integrals

The following proof is derived in Barndorff-Nielsen and Shephard (2000) and allows the stochastic integrals given in Chapter 4 to be written as an infinite sum of random variables, which facilitates MCMC inference.

Assume z is a subordinator, so the Lévy-Khintchine formula (see Theorem 2) can be written as

$$C(f \ddagger z) = \log\left\{E\left[e^{ifz}\right]\right\} = \frac{1}{\Delta}\int_0^{\infty}\int_0^{\Delta}\left(e^{if(\omega)x}-1\right)u(x,\omega)\,d\omega dx, \tag{D.5}$$

where $f()$ is some non-negative integrable function. Define

$$\overline{K}(\theta \ddagger x) = \log\left\{E\left[e^{-\theta x}\right]\right\} \qquad \text{and} \qquad \overline{L}(\theta \ddagger x) = \exp\left[\overline{K}(\theta \ddagger x)\right], \tag{D.6}$$

so

$$\overline{K}\left(\theta \ddagger x\right) = C\left(i\theta \ddagger x\right). \tag{D.7}$$

Substituting this into equation (D.5) gives

$$\overline{K}\left(\theta \ddagger z\right) = \frac{1}{\Delta}\int_0^\infty \int_0^\Delta \left(e^{-\theta x} - 1\right) u\left(x,\omega\right) d\omega dx.$$

Let $a_1, a_{2,\ldots}$ be the arrival times of a Poisson point process of unit intensity and $a_c \in \mathbb{R}^+$ and let n_{a_c} be the number of Poisson points less than a_c. Let

$$\Omega_i \overset{iid}{\sim} U\left(0,\Delta\right)$$

and define

$$\sigma = \sum_{i=1}^{\infty} f\left(\omega_i\right) W^{-1}\left(a_i;\omega_i\right)$$

and

$$\sigma_{a_c} = \sum_{i=1}^{n_{a_c}} f\left(\omega_i\right) W^{-1}\left(a_i;\omega_i\right),$$

where

$$W^+\left(x,\omega\right) = \int_x^\infty u\left(s,\omega\right) ds$$

and

$$W^{-1}\left(a;\omega\right) = \inf\left[x > 0 : W^+\left(x,\omega\right) \leq a\right].$$

Then

$$\begin{aligned}
\overline{L}\left(\theta \ddagger \sigma_{a_c}\right) &= E_{N_\tau}\left[E_{\Sigma_{a_c}|N_\tau=n_\tau}\left[e^{-\theta\sigma_{a_c}}\right]\right] \\
&= E_{N_\tau}\left[\prod_{i=1}^{n_\tau} E\left[\exp\left(-\theta f\left(\omega_i\right) W^{-1}\left(a_i;\omega_i\right)\right)\right]\right] \\
&= E_{N_\tau}\left[\prod_{i=1}^{n_\tau} \exp\left\{\overline{K}\left(\theta f\left(\omega_i\right) \ddagger W^{-1}\left(a_i;\omega_i\right)\right)\right\}\right].
\end{aligned}$$

Given n_τ, the arrival times of a Poisson point process have the same distribution as the order statistics of n_τ uniform variables on $(0, a_c)$, so dropping the ordering of the a_i gives

$$\overline{L}\left(\theta \ddagger \sigma_{a_c}\right) = E_{N_\tau}\left[\exp\left\{n_\tau \overline{K}\left(\theta f\left(\omega\right) \ddagger W^{-1}\left(ra_c;\omega\right)\right)\right\}\right],$$

where r and ω are independent random variables with distributions $r \sim U\left(0,1\right)$ and $\omega \sim U\left(0,\Delta\right)$. Further, as $N_\tau \sim Po\left(a_c\right)$,

$$\begin{aligned}
\overline{L}\left(\theta \ddagger \sigma_{a_c}\right) &= e^{-\tau}\sum_{n=0}^{\infty}\frac{a_c^n}{n!}\exp\left\{n\overline{K}\left(\theta f\left(\omega\right) \ddagger W^{-1}\left(ra_c;\omega\right)\right)\right\} \\
&= \exp\left[a_c\left(\exp\left\{\overline{K}\left(\theta f\left(\omega\right) \ddagger W^{-1}\left(ra_c;\omega\right)\right)\right\} - 1\right)\right]
\end{aligned}$$

and comparing this with equation (D.6), gives

$$\overline{L}\left(\theta \ddagger \sigma_{a_c}\right) = \exp\left[\overline{K}\left(\theta \ddagger \sigma_{a_c}\right)\right] = \exp\left[a_c\left\{\overline{L}\left(\theta f\left(\omega\right) \ddagger W^{-1}\left(ra_c;\omega\right)\right) - 1\right\}\right]. \tag{D.8}$$

Note that

$$\begin{aligned}\overline{L}\left(\theta \ddagger W^{-1}\left(ra_c;\omega\right)\right) &= E_{f_\Omega}\left[\int_0^1 \exp\left\{-\theta W^{-1}\left(ra_c;\omega\right)\right\} dr\right] \\ &= E_{f_\Omega}\left[\int_0^1 \exp\left\{-\theta W^{-1}\left(ra_c;\omega\right)\right\} - 1 dr\right] + 1,\end{aligned}$$

so letting $x = W^{-1}\left(ra_c;\omega\right)$, gives

$$\begin{aligned}\overline{L}\left(\theta \ddagger W^{-1}\left(ra_c;\omega\right)\right) = - &\quad \frac{1}{a_c} E_{f_\Omega}\left[\int_\infty^{W^{-1}(a_c;\omega)} \left\{e^{-\theta x} - 1\right\} u\left(x,\omega\right) dx\right] + 1 \\ = &\quad \frac{1}{a_c} E_{f_\Omega}\left[\int_{W^{-1}(a_c;\omega)}^{\infty} \left\{e^{-\theta x} - 1\right\} u\left(x,\omega\right) dx\right] + 1\end{aligned}$$

and

$$\lim_{a_c \to \infty} a_c\left\{\overline{L}\left(\theta \ddagger W^{-1}\left(ra_c;\omega\right)\right) - 1\right\} = E_{f_\Omega}\left[\int_0^\infty \left\{e^{-\theta x} - 1\right\} u\left(x,\omega\right) dx\right]. \tag{D.9}$$

From equations (D.5) and (D.6) it is already known that

$$\begin{aligned}\overline{K}\left(\theta \ddagger \sigma\right) &= C\left(i\theta \ddagger \sigma\right) \\ &= \frac{1}{\Delta}\int_0^\infty \int_0^\Delta \left(e^{-\theta f(\omega)x} - 1\right) u\left(x,\omega\right) d\omega dx.\end{aligned}$$

Using equations (D.8) and (D.9) gives

$$\begin{aligned}\overline{K}\left(\theta \ddagger \sigma\right) &= \lim_{a_c \to \infty} \overline{K}\left(\theta \ddagger \sigma_{a_c}\right) \\ &= \frac{1}{\Delta}\int_0^\Delta \int_0^\infty \left(e^{-\theta f(\omega)x} - 1\right) u\left(x,\omega\right) dx d\omega\end{aligned}$$

and from equations (D.5) and (D.7),

$$\overline{K}\left(\theta \ddagger \sigma\right) = \overline{K}\left(\theta f \ddagger z\right). \tag{D.10}$$

As in Barndorff-Nielsen and Shephard (2000), define

$$f\left(\omega\right) \bullet z\left(t\right) = \int_0^\Delta f\left(s\right) dz\left(s\right).$$

Jacod and Shiryaev (1987) have proved that

$$\overline{K}\left(\theta f \ddagger z\right) = \overline{K}\left(\theta \ddagger f \bullet z\right),$$

so equation (D.10) can be rewritten as

$$\overline{K}\left(\theta \ddagger \sigma\right) = \overline{K}\left(\theta \ddagger f \bullet z\right).$$

Therefore

$$\lim_{\tau\to\infty} \sigma_{a_c} = \sum_{i=1}^{\infty} f\left(\omega_i\right) W^{-1}\left(a_i;\omega_i\right) \overset{\mathfrak{L}}{=} \int_0^{\Delta} f\left(s\right) dz\left(s\right)$$

and the integral with respect to the Lévy process can be written in terms of an infinite sum of random variables.

D.5 Correlation structure of the OU volatility process

Let $X\left(t\right) = \sigma^2\left(t\right)$ be a solution to the OU equation (D.3), so for $s > 0$,

$$X\left(t+s\right)|X\left(t\right) \overset{\mathfrak{L}}{=} k^{-\lambda s} X\left(t\right) + k^{-\lambda s} \int_0^{\lambda s} k^u dz\left(u\right)$$

and continue assuming $V\left[X\left(t\right)\right] = V$ is finite. Then $cov\left[X\left(t\right), X\left(t+s\right)\right]$ is

$$\int_0^{\infty} \left(X\left(t\right) - M\right)\left(k^{-\lambda s} X\left(t\right) + k^{-\lambda s} \int_0^{\lambda s} e^u dz\left(u\right) - M\right) f_{X(t)}\left(x\right) dx,$$

where $M = E\left[X\left(t\right)\right]$. Noting that

$$\int_0^{\infty} \left(X\left(t\right) - M\right) f_{X(t)}\left(x\right) dx = 0,$$

it can be seen that

$$\begin{aligned} cov\left[X\left(t\right), X\left(t+s\right)\right] &= k^{-\lambda s} \int_0^{\infty} \left(X\left(t\right) - M\right) X\left(t\right) f_{X(t)}\left(x\right) dx \\ &= k^{-\lambda s}\left[V + M^2 - M^2\right] \\ &= V k^{-\lambda s}, \end{aligned}$$

so the correlation is

$$corr\left[X\left(t\right), X\left(t+s\right)\right] = k^{-\lambda s}.$$

At first, this may look like a more general correlation structure than the BNS SV models, where $k = e$. To ensure that the correlation between $\sigma^2\left(t\right)$ and $\sigma^2\left(0\right)$ decreases as t increases, we require $k > 1$, so

$$corr\left[X\left(t\right), X\left(t+s\right)\right] = d^s,$$

where $0 < d = k^{-\lambda} < 1$ and the extra parameter, k, only alters the scale of λ, without providing a different correlation structure to that of the BNS SV models. This is frustrating because models of the form

$$\sigma^2\left(t\right) = \int_{-\infty}^{t} f\left\{\lambda\left(s-t\right)\right\} dz\left(\lambda s\right)$$

are difficult to simulate from for $f(s) \neq k^s$, as it is not easy to rewrite $\sigma^2(t)$ in terms of $\sigma^2(0)$ (and some integral with respect to the BDLP on a finite range). To simulate from such models requires the simulation from an integral with respect to the BDLP on the negative real line and this is more involved than evaluating $k^{-\lambda t}\sigma^2(0)$ and makes parameter estimation more difficult. Models where $f(s) \neq k^s$ are considered in Chapter 6.

Bibliography

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

Ammar, G. S. and Gragg, W. B. (1988). Superfast solution of real positive definite Toeplitz systems, *SIAM Journal on Matrix Analysis and Applications* **9**: 61–76.

Anh, V. V. and McVinish, R. (2003). Fractional diffrerential equations driven by Lévy noise, *Journal of Applied Mathematics and Stochastic Analysis* **16**: 97–119.

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

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

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

Barndorff-Nielsen, O. E. (2001). Superposition of Ornstein-Uhlenbeck type processes, *Theory of Probability and Its Applications* **45**: 175–194.

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

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

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

Barndorff-Nielsen, O. E. and Thorbjørnsen, J. (2002). Selfdecomposability and Lévy processes in free probability, *Bernoulli* **8**: 323–366.

Baten, W. D. (1934). The probability law for the sum of n independent variables, each subject to the law (1/(2h))sech(pi*x.(2h)), *Bulletin of the American Mathematical Society* **40**: 284–290.

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

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

Birge, J. R. (1995). Quasi-Monte Carlo approaches to option pricing, *Technical Report* . Department of Industrial and Operations Engineering, University of Michigan.

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

Bollerslev, T. (1987). A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics* **69**: 542–547.

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

Bondesson, L. (1992). *Generalized Gamma Convolutions and related classes of distributions and densities*, Springer-Verlag.

Breidt, F. J., Crato, N. and Lima, P. (1998). The detection and estimation of long memory in stochastic volatility, *Journal of Econometrics* **83**: 325–348.

Brockwell, P. J. (2001). Lévy-driven CARMA processes, *Annals of the Instititute of Statistical Mathematics* **53**: 113–124.

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

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

Carr, P., Geman, H., Medan, D. and Yor, M. (2002). The Fine Structure of Asset Returns: An Empirical Investigation, *Journal of Business* **75**: 305–332.

Chambers, J. M., Mallows, C. L. and Stuck, W. (1971). A method for simulating stable random variables, *Journal of the American Statistical Association* **71**: 340–344.

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

Comte, F. and Renault, E. (1998). Long memory in continuous-time stochastic volatility models, *Mathematical Finance* **8**: 291–323.

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

Fama, E. (1965). The behavior of stock market prices, *Journal of Business* **38**: 84–105.

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

Ferguson, T. S. and Klass, M. J. (1972). A representation of independent increment processes without Gaussian components, *Annals of Mathematical Statistics* **43**: 1634–1643.

Frühwirth-Schnatter, S. and Sögner, L. (2001). Bayesian Estimation of Stochastic Volatility Models Based on OU Processes with Marginal Gamma Law, *Technical Report* . Department of Statistics, Vienna University of Economics and Business Administration.

Gallant, A. R., Rossi, P. E. and Tauchen, G. (1992). Stock prices and volume, *Review of Financial Studies* **5**: 199–242.

Geman, S. and Geman, D. (1994). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *Transactions on Pattern Analysis and Machine Intelligence* **6**: 721–741.

Geweke, J. and Porter-Hudak, S. (1983). The estimation and application of long memory time series models, *Journal of Time Series Analysis* **4**: 221–237.

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

Gradshteyn, I. S. and Ryzhik, I. M. (1965). *Table of Integrals, Series and Products*, Academic Press.

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

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

Grimmett, G. and Stirzaker, D. (2001). *Probability and Random Processes*, Oxford University Press.

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

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications, *Biometrika* **57**: 97–109.

Higham, N. J. (1996). *Accuracy and Stability of Numerical Algorithms*, SIAM.

Hu, Y. and Øksendal, B. (2003). Fractional white noise calculus and applications to finance, *Infinite Dimensional Analysis, Quantum Probability and Related Topics* **6**: 1–32.

Hubalek, F. and Tompkins, R. (2001). On explicit option pricing for stochastic volatility models driven by a superposition of Ornstein-Uhlenbeck type jump processes, *Unpublished paper* .

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

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

Hurst, H. E. (1951). Long term storage capacity of resevoirs, *Transactions of the American Society of Civil Engineers* **116**: 770–799.

Jacod, J. and Shiryaev, A. N. (1987). *Limit Theorems for Stochastic Processes*, Springer-Verlag: Berlin.

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

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

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

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

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

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

Leipnik, R. B. (1991). On lognormal random variables: I-th characteristic function, *Journal of the Australian Mathematical Society, Series B* **32**: 327–347.

Lo, A. W. (1991). Long term memory stock market prices, *Econometrica* **59**: 1279–1313.

Lukacs, E. (1970). *Characteristic Functions*, Griffin.

Mandelbrot, B. B. (1975). Limit theorems on the self-normalized range for weakly and strongly dependent processes, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **31**: 271–285.

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

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. and Teller, E. (1953). Equation of state calculations by fast computing machines, *Journal of Chemical Physics* **21**: 1087–1092.

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

Minami, M. (2003). A multivariate extension of inverse Gaussian distribution derived from inverse relationship, *Communications in Statistics: Theory and Methods* **32**: 2285–2304.

Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* **61**: 631–653.

Nicolato, E. and Venardos, E. (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 Shephard, N. (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*, Oxford: Oxford University Press, pp. 547–570.

Ribeiro, C. and Webber, N. (2003). Correcting for simulation bias in Monte Carlo methods to value exotic options in models driven by Lévy processes, *Unpublished paper* .

Robert, C. P. and Casella, G. (2002). *Monte Carlo Statistical Methods*, Springer.

Roberts, G., Papaspiliopoulos, O. and Dellaportas, P. (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.

Rogers, L. C. G. and Williams, D. (2000). *Diffusions, Markov Processes and Martingales*, Cambridge University Press.

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 Taqqu, M. S. (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.

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

Tezuka, S. (1995). *Uniform Random Numbers: Theory and Practice*, Kluwer Academic Publishers.

Tolvi, J. (2003). Long memory in a small stock market, *Economics Bulletin* **7**: 1–13.

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

Walker, S. G. and Damien, P. (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.

Wolpert, R. L. and Taqqu, M. S. (2004). Fractional Ornstein-Uhlenbeck Lévy processes and the telecom process: Upstairs and downstairs, *Unpublished paper* .

Yu, J. (2002). MCMC methods for estimating stochastic volatility models with leverage effects: Comments on Jacquier, Polson and Rossi (2002), *Unpublished paper* .