The Variance Gamma (VG) Model with Long Range Dependence

Richard Finlay

A thesis submitted in fulfilment of
the requirements for the degree of
Doctor of Philosophy

School of Mathematics and Statistics
University of Sydney

February 2009

CONTENTS

List of Figures

List of Tables

Glossary

$\{G(t;a,b)\}$**:** a gamma process with independent increments where $G(t;a,b) \sim \Gamma(ta,b)$ for each t. We also define $\{G(t;b)\} \stackrel{def}{=} \{G(t;1,b)\}$ and $\{G(t)\} \stackrel{def}{=} \{G(t;1)\}$, 95

$\{P_t\}$**:** the price process of some financial asset, 11

$\{T_t\}$**:** the time over which market prices evolve – see **Activity time**, 11

$\{X_t\}$**:** the process of log increments (continuously compounded returns) of $\{P_t\}$, 11

$\{\tau_t\}$**:** the process of unit increments of $\{T_t\}$, 11

Activity time: the increasing stochastic process $\{T_t\}$ which can be interpreted as the time over which market prices evolve, as opposed to standard deterministic clock-time t, 11

Average relative measure (ARM): a measure of model fit taken as the average over the parameters of the model of each measure of fit (MAD score, bias, standard deviation) divided by the true parameter value, 73

Bridge sampling: sampling from a stochastic process by 'filling in the gaps' using the conditional distribution given the next and last points, 94

Difference of Gammas (DG) model: a model of log stock prices taken as the difference to two independent gamma processes, 102

Efficient markets hypothesis (EMH): the assertion that current market prices reflect all available information, so that past information cannot be used to predict future stock price returns, 13

Empirical characteristic function (ECF) estimation: estimation of a model from data via the minimisation the distance between the theoretical and empirical characteristic functions, 77

Kolmogorov distance: for two random variables X and Y, the quantity $\sup_z |P(X \leq z) - P(Y \leq z)|$, 54

Long range dependence (LRD): when the sum of the ultimately non-negative autocovariances of a process diverges, 8

Mean absolute deviation (MAD): a measure of model fit taken as the average deviation of estimated parameter values from their true value, 66

Mean-correcting martingale: when a martingale is obtained from a stock price model for historical data by constraint of the mean parameter, 108

Product-density maximum likelihood estimation (PMLE): estimation of a model from data, irrespective of its correlation structure, via the maximisation of the product of marginal densities as performed in classical maximum likelihood estimation, 60

Root mean square error (RMSE): The square root of the average squared error between actual prices and model determined prices, 104

Self similarity: for a process Y_t, when $Y_{ct} \stackrel{\mathcal{D}}{=} c^H Y_t$, 8

Short range dependence (SRD): when the sum of the autocovariances of a process converges, 8

Skew-correcting martingale: when a martingale is obtained from a stock price model for historical data by constraint of the skew parameter, 109

Subordinator model: a model driven by time-changed Brownian motion, 10

Variance Gamma (VG) distribution: The unconditional distribution that results from a normal with conditional variance given by a gamma random variable, 16

Abstract

This thesis mainly builds on the Variance Gamma (VG) model for financial assets over time of Madan & Seneta (1990) and Madan, Carr & Chang (1998), although the model based on the t distribution championed in Heyde & Leonenko (2005) is also given attention.

The primary contribution of the thesis is the development of VG models, and the extension of t models, which accommodate a dependence structure in asset price returns. In particular it has become increasingly clear that while returns (log price increments) of historical financial asset time series appear as a reasonable approximation of independent and identically distributed data, squared and absolute returns do not. In fact squared and absolute returns show evidence of being long range dependent through time, with autocorrelation functions that are still significant after 50 to 100 lags. Given this evidence against the assumption of independent returns, it is important that models for financial assets be able to accommodate a dependence structure.

The VG and t are both subordinator models, whereby log stock prices P_t are driven by Brownian motion $B(t)$ evaluated at a random time-change, the so-called 'activity time' T_t, for $t \geq 0$. That is, we assume

$$\log(P_t) = \log(P_0) + \mu t + \theta T_t + \sigma B(T_t)$$

for $\mu, \theta \in \mathbb{R}$ and $\sigma > 0$, where μ and σ correspond to the drift and diffusion coefficients of Brownian motion respectively. Here $\{T_t\}$ has the attractive interpretation of information

flow or trading activity – the more frenzied trading becomes, or the more information released to the market on a given day, the faster 'time' flows. The VG, which is a pure jump process, is obtained when the unit increments of $\{T_t\}$, $\tau_t = T_t - T_{t-1}$, have the gamma distribution, while the t model, based on a generalisation of the classical Student's t distribution, is obtained for τ_t with the inverse gamma distribution. The process $\{T_t\}$ is assumed to be *strictly stationary*, but with dependent as opposed to independent increments. It is the dependence structure of the $\{T_t\}$ process that determines the dependence structure of the log price increments process, so it is via $\{T_t\}$ that we construct models with dependence structure.

This thesis is set out as follows: in Chapter 1 we provide evidence in support of the general subordinator model and against the classical assumption of independent and identically normally distributed asset price returns. In particular we argue that the geometrical Brownian motion model of asset prices is an oversimplification, and that a realistic model for returns should allow for skewness, excess kurtosis, heteroscedasticity, and little autocorrelation in returns but a dependence structure in squared returns. We go on to introduce the VG and t models, as well as the Generalised Hyperbolic model, of which the VG and t are special cases and in a sense dual to each other, and discuss the relative merits of these models as well as their historical development.

Chapter 2 deals with the activity time process $\{T_t\}$. In Section 2.1 we construct a discrete time process $\{T_t\}$ whereby the discrete increments process $\{\tau_t\}$ has marginal gamma distribution of arbitrary parameter; is allowed an arbitrary convex autocorrelation function; and whereby, for a certain choice of autocorrelation function, the log price increments of $\{P_t\}$ are VG distributed and long range dependent, and an appropriately normed $\{T_t\}$ converges weakly to the self similar continuous time Rosenblatt process. These results are one of the main contributions of this thesis, and have appeared as Finlay & Seneta (2006) and Finlay & Seneta (2007). Section 2.2 deals with the t distribution case and extends a result

of Heyde & Leonenko (2005). In particular, a discrete time process $\{T_t\}$ is constructed whereby the discrete increments process $\{\tau_t\}$ has marginal inverse gamma distribution with integer parameter; is allowed an arbitrary convex autocorrelation function; and whereby, for a certain choice of autocorrelation function, the log price increments of $\{P_t\}$ over unit time are t distributed and long range dependent, and an appropriately normed $\{T_t\}$ converges weakly to a process which is the negative of the above Rosenblatt process. Heyde & Leonenko (2005) constructed the case for the increments process $\{\tau_t\}$ with inverse gamma distribution, integer parameter and a specific autocorrelation function, and we extend their result to allow for arbitrary convex autocorrelation function. In the later sections of Chapter 2 we also numerically investigate some properties of our constructed asymptotically self similar $\{T_t\}$ processes, and review some alternative methods for constructing dependent activity time processes.

Chapter 3 studies the simulation of long range dependent VG and t models for financial assets, and tests various estimation techniques on the simulated data. In Section 3.1 we detail a method for simulating long range dependent VG and t data and propose four methods for recovering the (in our case known) model parameter values of the simulated data: the method of moments, a minimum χ^2 method, product-density maximum likelihood estimation (sometimes called 'pseudolikelihood'), and empirical characteristic function estimation. Section 3.2 and Section 3.3 detail the results of our simulation study. We find that product-density maximum likelihood is the most successful estimation method of the four considered. In Section 3.3 we also review early work by Madan & Seneta (1987b) and Madan & Seneta (1989), based on a Chebyshev polynomial expansion of the likelihood of a transformed VG variable, and compare the fit thus obtained to those considered earlier. Section 3.4 considers whether it is possible to choose between the VG and t models based on a χ^2 goodness of fit test on data simulated from each, and concludes that it is not. Section 3.5 fits the VG and t to some actual financial data sets, finding in some cases that the VG provides a superior fit and in other cases that the two models more or less fit equally. For three

of the data sets considered we find evidence of negative skewness, while for a fourth positive skewness is found. Skewness, a topic of some current interest for historical financial data, is allowed for in our model through presence of the parameter θ. Finally, Section 3.6 reviews other approaches to simulation, estimation and model fit that have appeared in the literature. The bulk of the contents of Chapter 3 have appeared as Finlay & Seneta (2008a).

While Chapters 1 to 3 concern models for and the analysis of historical data, in Chapter 4 we leave the *real-world* setting of historical data and move to the *risk-neutral world* of option pricing. In this case we consider a VG model which allows for the long range dependence of squared returns, as well as a simpler model based on the work of Madan, Carr & Chang (1998) which does not allow for dependence, but which introduces one extra degree of freedom over the classical independent increments VG. Section 4.2 details current methods for pricing options in the VG framework, while Sections 4.3 and 4.4 detail option pricing under our two new models. Here we use data from and base our approach to model fit on Schoutens (2003), while to actually compute option prices in accordance with our models we rely on the results of Carr & Madan (1999). We find that the current approach to VG option pricing has a number of drawbacks, and that in terms of fit to market data, our two suggestions improve on other comparable models. This work has appeared as Finlay & Seneta (2008b).

Finally, Chapter 5 summarises our results.

Acknowledgements

A number of people have kindly provided me with assistance over the last few years and I would like to thank and acknowledge them here.

Professor Nikolai Leonenko was kind enough to send hand-written notes on the theory of orthogonal polynomials, which were a tremendous help in my understanding of the paper 'Student Processes' (Heyde & Leonenko (2005)), and thus in the writing of the paper Finlay & Seneta (2006).

C.C. Heyde, N.N. Leonenko and an anonymous referee provided comments which much improved the paper that is the basis for Chapter 2, Finlay & Seneta (2007). Regarding Finlay & Seneta (2007), I would also like to thank A. Sly for access to his Master's thesis.

Thomas Fung provided considerable assistance with model fitting via WinBUGS, as detailed in Chapter 3 and Finlay & Seneta (2008a). Regarding Chapter 4 and the paper Finlay & Seneta (2008b) on which it is based, W. Schoutens was kind enough to send me the option pricing data to which my analysis is applied, while an anonymous referee very helpfully pointed out the link between the C3 model of this paper and the absence of static arbitrage.

The most thanks and acknowledgement however is owed to my supervisor Professor Eugene Seneta. This thesis is built round the contents of four joint papers with Professor Seneta. Their form would not have been possible without his guidance, insights, and encouragement. In addition, he has been a great help in my synthesizing the four papers into

one (hopefully!) cohesive and well-presented whole. So I wish to thank Professor Seneta for the last few years – he has made my postgraduate experience extremely rewarding, and the writing of this thesis a pleasure to undertake.

On a personal front I would also like to thank my partner Clare for her support, encouragement and understanding over the last few years, as well as my parents Roger and Ulla, and my Associate Supervisor, Professor Neville Weber, for their support, encouragement and advice.

CHAPTER 1

Motivation, definitions and the model

The paradigm model for asset price movements, geometric Brownian motion, has afforded the financial world numerous insights into how markets function, as well as spawning a multi-billion dollar global derivatives industry. The model is relatively simple, giving the price of an asset at time $t \geq 0$ as

$$P_t = P_0 e^{\mu t + \sigma B(t)}$$

for $\mu \in \mathbb{R}$ and $\sigma > 0$ with $\{B(t)\}$ standard Brownian motion. Log price increments (continuously compounded returns) are then given as

$$X_t = \log(P_t) - \log(P_{t-1}) = \mu + \sigma(B(t) - B(t-1))$$

which in particular implies that returns are independent and identically distributed (iid) normal random variables. This is opposed to typical asset price data which display the following characteristics (see for example Heyde & Liu (2001) and the references therein):

(1) a leptokurtic distribution (kurtosis greater than three) – higher peaks above the mean, and thicker tails, than a normal distribution;

(2) a heteroscedastic time series (time-dependent conditional variance), unlike the geometric Brownian motion model;

(3) little or no autocorrelation present in returns, at least past one or two lags, but a long range dependence structure in squared and absolute returns, violating the independence assumption;

(4) occasionally skewed distributions, as opposed to the symmetry of a normal distribution.

We outline some of the evidence against geometric Brownian motion below, and introduce the Variance Gamma (VG) and t models, which seek to circumvent these problems, in the sections that follow. Our primary focus is on the VG; we consider the t due to its dualism with the VG, and as it can be viewed as a direct competitor to the VG for the modelling of financial assets.

1.1. Evidence against geometric Brownian motion

1.1.1. Kurtosis. The log increments of geometric Brownian motion are normally distributed, and as such have a kurtosis of 3, where kurtosis is given by $\mathbb{E}(X_t - \mathbb{E}X_t)^4/(\mathbb{E}(X_t - \mathbb{E}X_t)^2)^2$, the fourth central moment divided by the square of the second central moment. Actual log price increments however display much greater kurtosis, due both to heavier tails and greater peakedness over the mean, which respectively increase the fourth moment and reduce the second moment relative to those of the normal. In effect, actual log price increments are more prone to either small, or alternatively very large, movements, and less prone to moderate movements than they should be if prices followed geometric Brownian motion.

For example the estimated kurtosis of daily returns (log price increments) of the Standard & Poor's 500 Index (S&P 500) and Microsoft's share price (Microsoft) between 1 January 1996 and 31 December 2005 are respectively 5.95 and 8.32, roughly twice the theoretical value given the normality assumption. This excess tail weight and higher peakedness about the mean can be seen in Figures 1.1 and 1.2, which display histograms of the two series mentioned, as well as theoretical normal densities scaled to have the same means and variances as the series. Note that due to the weak law of large numbers, the proportion of observations

falling into a given interval converges to the probability of landing in that interval, so that construction of histograms is valid even if the data has a dependence structure.

Alternatively, if the log returns of these two financial time series followed the normal distribution, we would expect to find 7 out of 2500 observations outside of three standard deviations about the mean, whereas we have 32 and 30 for the S&P 500 and Microsoft respectively; and we would expect to find roughly 1700 out of 2500 observations within one standard deviation of the mean, whereas we have roughly 1900 for each of the two series.

1.1.2. Volatility and intermittency. The geometric Brownian motion asset price model implies that log increments of the asset price are independently normally distributed with constant variance, so that the standard deviation (alternatively volatility) of a sample of returns should show no pattern, and tend to fluctuate about some mean value. In contrast, actual data often displays varying volatility, that is, heteroscedasticity, with periods of low variation in asset returns followed by periods of higher variation. In Figure 1.3 we plot the annualised rolling standard deviation of returns over the previous 30 days for the S&P 500 Index and Microsoft data sets.

It can be seen that there does appear to be some pattern to the volatility, with intermittent peaks and troughs appearing in a way not predicted by geometric Brownian motion, but allowed for in the models that we shall consider. Note that the models we shall consider are strictly stationary, with constant variance of asset returns but fluctuating *conditional* variance of returns.

1.1.3. Dependence. If an asset price follows geometric Brownian motion, log increments should be iid normal. In contrast we find that asset returns display a long range dependence structure, with significant autocorrelation occurring in the squared and absolute returns series. This would tend to indicate that geometric Brownian motion oversimplifies

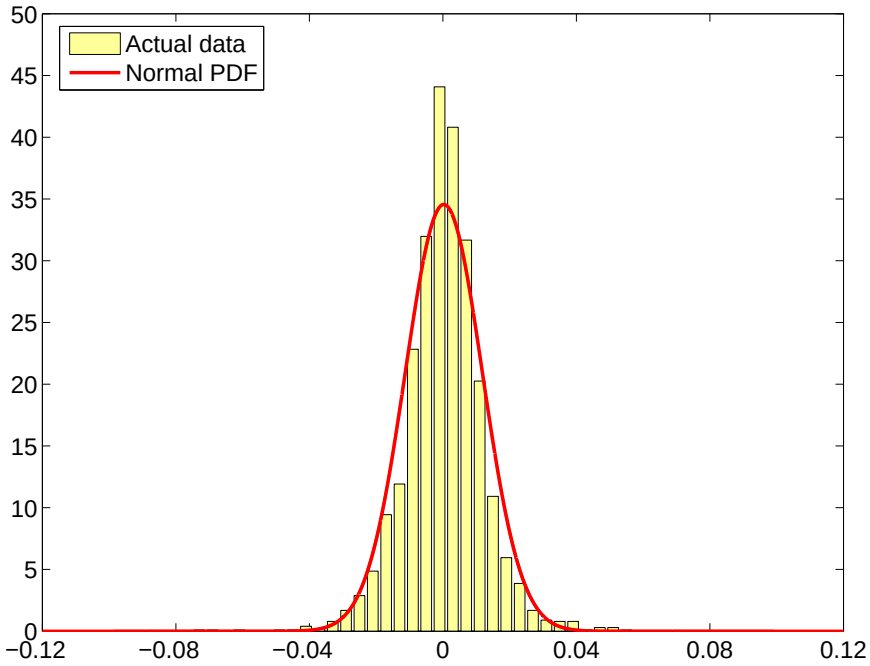

FIGURE 1.1. Histogram of S&P 500 index returns with fitted normal model.

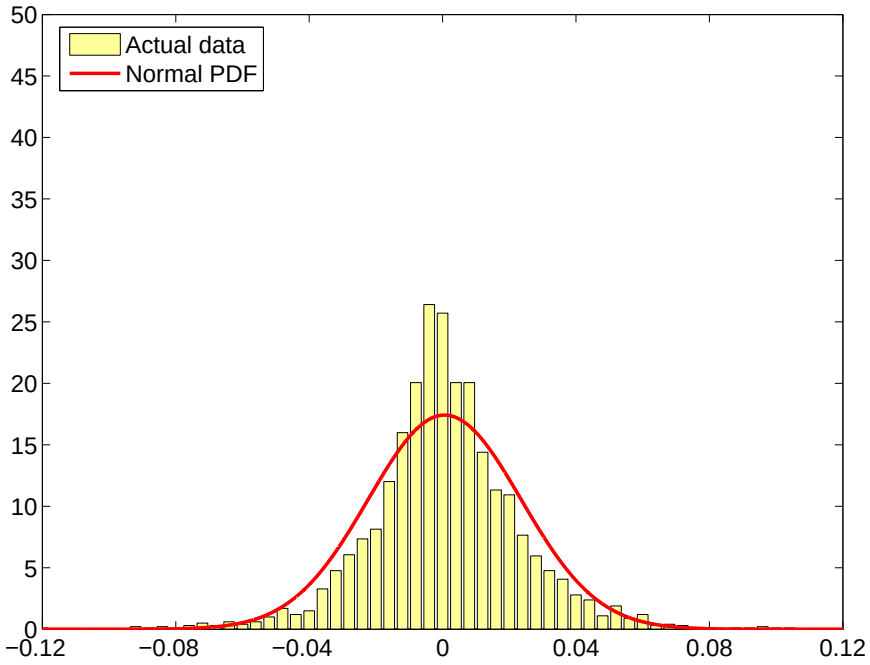

FIGURE 1.2. Histogram of Microsoft Corporation's share price returns with fitted normal model.

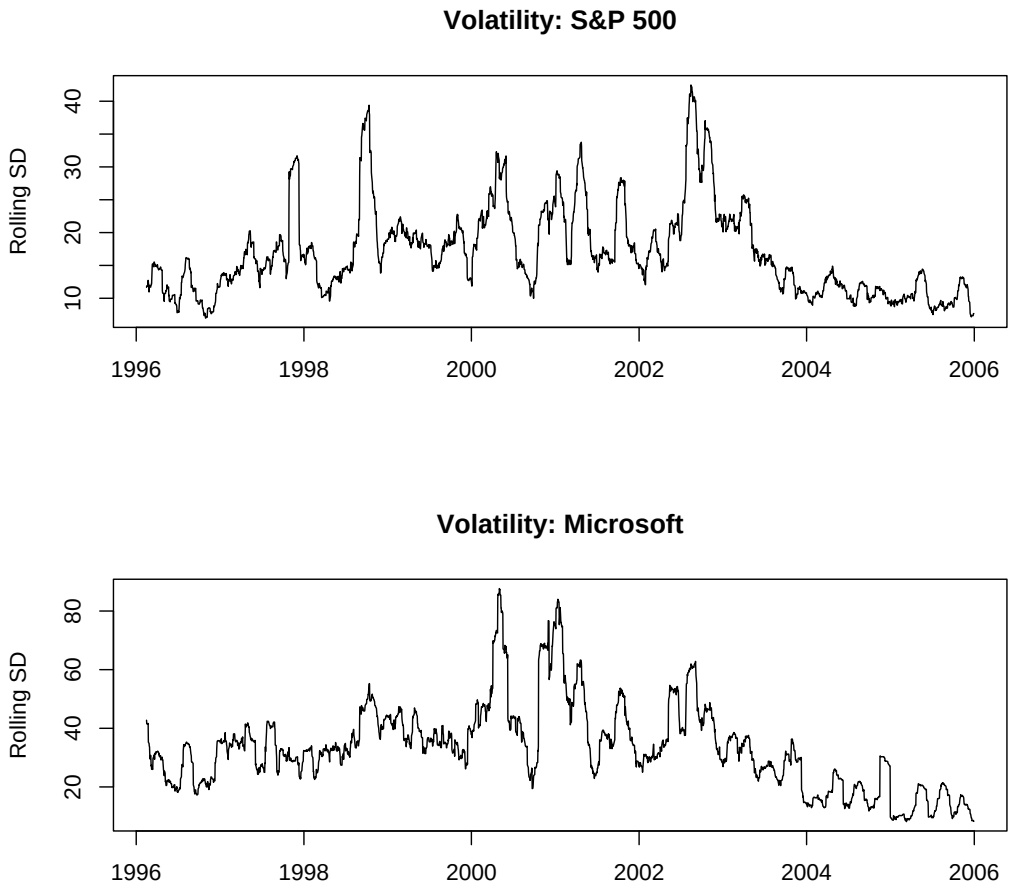

FIGURE 1.3. 30 day rolling annualised volatility (in per cent).

the nature of asset price dynamics with respect to dependence through time, in that the independent increments assumption is incorrect.

Figure 1.4 shows the autocorrelation function (acf) values of log increments of the S&P 500 index, Microsoft, and a simulated sample of 2500 standard normal random variables. As can be seen, while acfs of the log increments themselves die out relatively quickly for the two financial series, the acfs of their squares and absolute values are significant out to a number of lags, indicating a strong, long range dependence structure. This is in contrast to the standard normal sample where all acf series die out quickly.

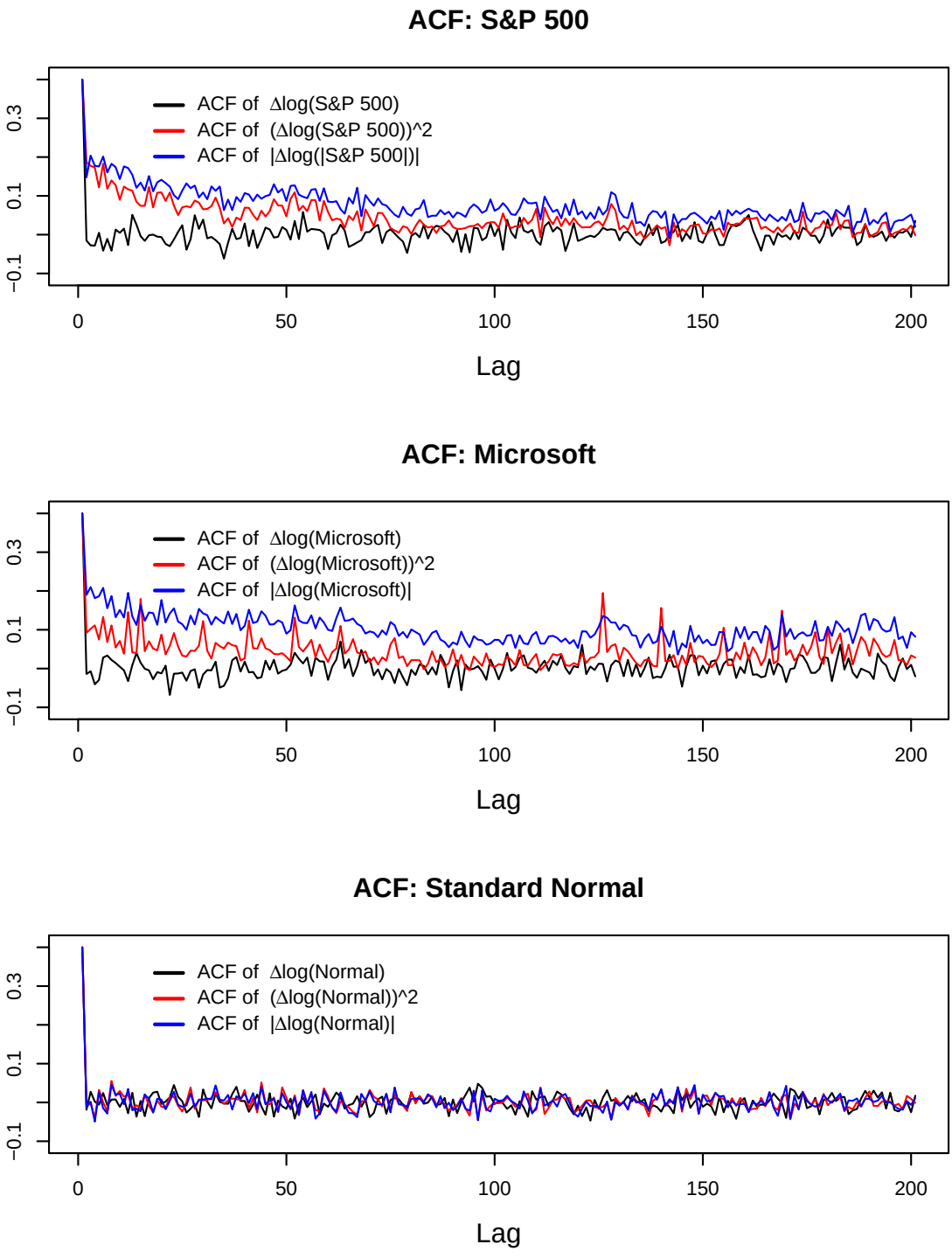

FIGURE 1.4. ACF of returns, absolute returns and squared returns for the S&P 500 index, Microsoft, and a simulated standard normal series.

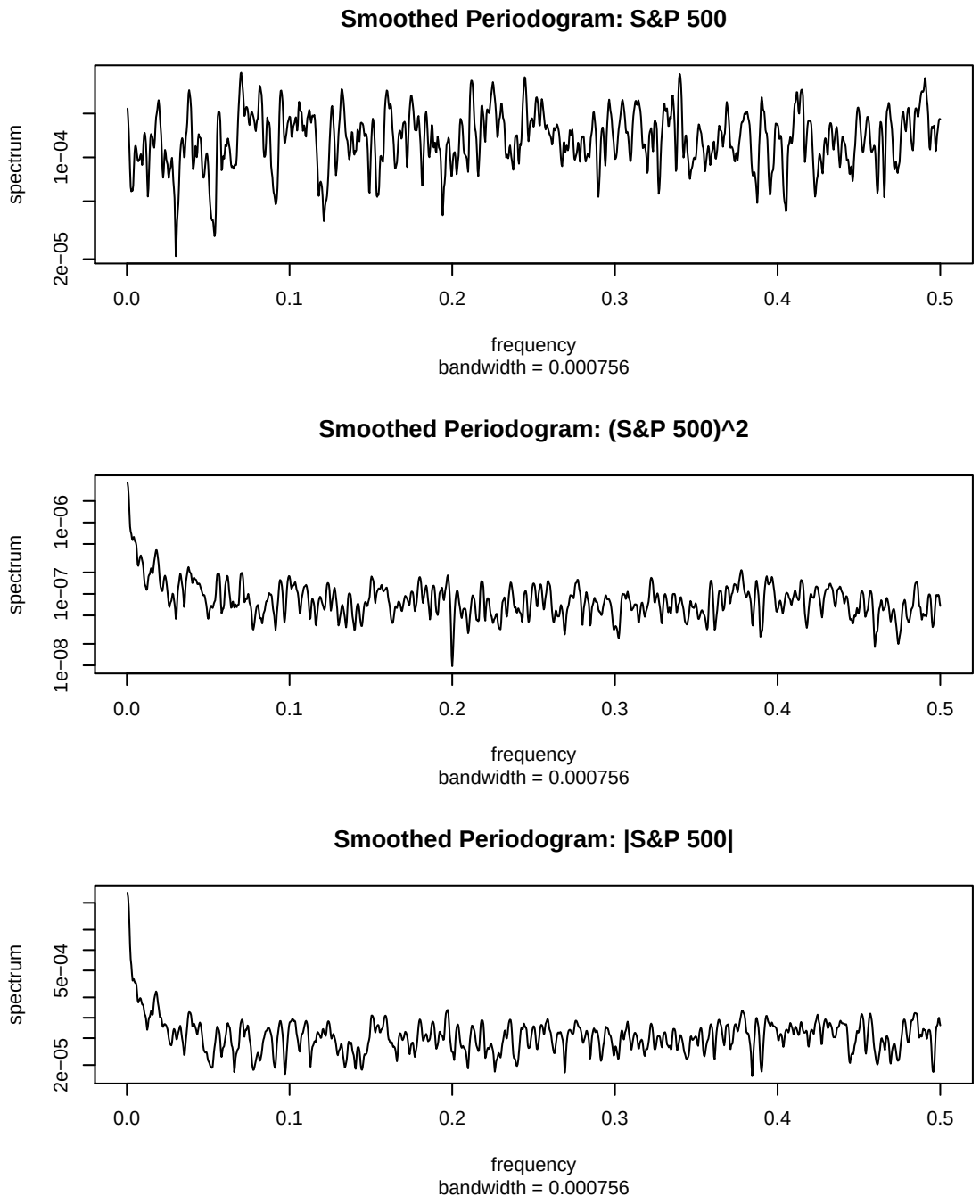

FIGURE 1.5. Smoothed Periodogram for S&P 500, (S&P 500)^2 and |S&P 500| returns.

This claim of long range dependence in the squared and absolute returns series is further supported by spectral analysis. Figure 1.5 shows the periodogram of the S&P 500 returns series and the squared and absolute returns series. As can be seen, the estimated spectral density of the original series resembles that of a white noise process, with many random fluctuations but no pattern. Conversely, the periodograms of the squared and absolute series show a very different picture. Here the estimated spectral densities have clear spikes around $\omega = 0$ and decay as ω gets large, which constitute criteria for long range dependent time series.

1.2. Self similarity and long range dependence

In the preceding section we talked about long range dependence (LRD) without defining it, which we now remedy. LRD is said to hold, for a stationary process on the integers with ultimately non-negative autocovariances $\{\gamma_k\},\ k \geq 1$, if

$$\sum_{k=1}^{\infty} \gamma_k = \infty,$$

that is, if the sum of the autocovariances of the process diverges. As the name suggests, if a process is long range dependent then events that occurred long ago still have an impact on future values of the process. Analogously, we define short range dependence (SRD) as holding if the acf decays sufficiently fast so as to be summable.

This definition of LRD only makes sense if the process under consideration has finite variance. This is sufficient for our purposes, although we note that Heyde & Yang (1997) provide an alternative definition, based on the concept of Allen Variance, which accommodates processes of infinite variance and which coincides with the above definition for processes of finite variance.

A related concept to that of LRD is self similarity. Notions of self similarity can be found in many fields, but for our purposes we shall say that a process $\{Y_t\}$ in continuous time with $Y_0 = 0$ is H self similar if there is some H, $\frac{1}{2} < H < 1$, such that

$$Y_{ct} \stackrel{\mathcal{D}}{=} c^H Y_t,$$

where '$\stackrel{\mathcal{D}}{=}$' denotes equality in distribution. Then putting $\sigma^2 = \mathbb{V}\text{ar}(Y_1)$ assuming finite variance, and taking $t > s$, we have

$$\begin{aligned}
\sigma^2 t^{2H} &= \mathbb{V}\text{ar}(Y_t) \\
&= \mathbb{V}\text{ar}(Y_t - Y_s + Y_s) \\
&= \mathbb{V}\text{ar}(Y_{t-s}) + \mathbb{V}\text{ar}(Y_s) + 2\mathbb{C}\text{ov}(Y_s, Y_t - Y_s) \\
&= \sigma^2(t-s)^{2H} + \sigma^2 s^{2H} + 2\mathbb{C}\text{ov}(Y_s, Y_t - Y_s)
\end{aligned}$$

which follows since the increments of $\{Y_t\}$ are assumed stationary and $Y_0 = 0$. Then for $t > s$,

$$\begin{aligned}
\mathbb{C}\text{ov}(Y_s, Y_t) &= \mathbb{C}\text{ov}(Y_s, Y_s + Y_t - Y_s) \\
&= \sigma^2 s^{2H} + \mathbb{C}\text{ov}(Y_s, Y_t - Y_s) \\
&= \frac{1}{2}\sigma^2(t^{2H} + s^{2H} - (t-s)^{2H}).
\end{aligned}$$

Now, for $y_t = Y_t - Y_{t-1}$ we have

$$\begin{aligned}
\mathbb{C}\text{ov}(y_t, y_{t+k}) &= \mathbb{C}\text{ov}(Y_t - Y_{t-1}, Y_{t+k} - Y_{t+k-1}) \\
&= \mathbb{C}\text{ov}(Y_t, Y_{t+k}) - \mathbb{C}\text{ov}(Y_t, Y_{t+k-1}) - \mathbb{C}\text{ov}(Y_{t-1}, Y_{t+k}) + \mathbb{C}\text{ov}(Y_{t-1}, Y_{t+k-1}) \\
&= \frac{1}{2}\sigma^2((k+1)^{2H} + (k-1)^{2H} - 2k^{2H}).
\end{aligned}$$

Hence we obtain that ρ_k, the autocorrelation function of the unit increments process $\{y_t\}$, is given by,

$$\rho_k = \frac{1}{2}((k+1)^{2H} + (k-1)^{2H} - 2k^{2H}) \tag{1.1}$$

$$= H(2H-1)k^{2H-2} + \sum_{i=2}^{\infty} \binom{2H}{2i} k^{2H-2i}$$

$$\geq H(2H-1)k^{2H-2}. \tag{1.2}$$

From (1.1) it follows that ρ_k increases with H, with H close to $\frac{1}{2}$ giving correlation close to zero, and H close to 1 giving correlation close to unity. Hence the level of H determines the strength of dependence between increments. Equation (1.2) further implies that for $H > \frac{1}{2}$,

$$\sum_{k=1}^{\infty} \rho_k \geq H(2H-1) \sum_{k=1}^{\infty} k^{2H-2} = \infty \tag{1.3}$$

indicating long range dependence. That is, if a process $\{Y_t\}$ is self similar with $\frac{1}{2} < H < 1$, then the process of unit increments $\{y_t\}$ is LRD (see Beran (1994) for an overview of self similarity and LRD).

1.3. The subordinator model

We now introduce the subordinator model, our alternative to geometric Brownian motion. Let P_t be the price of a risky asset at time t and assume $\{P_t\}$ in continuous time follows subordinated geometric Brownian motion. Specifically,

$$P_t = P_0 e^{\mu t + \theta T_t + \sigma B(T_t)} \tag{1.4}$$

where μ, θ and $\sigma > 0$ are constants, and $\{B(t)\}$ is standard Brownian motion independent of $\{T_t\}$, which is a positive non-decreasing random process with stationary but not necessarily independent increments, denoted over unit time by $\tau_t = T_t - T_{t-1}$. Without loss of generality we take $\mathbb{E}\tau_t = 1$, since any scaling can be absorbed into θ and σ as required, assuming $\mathbb{E}\tau_t < \infty$. In the context of this model, the 'activity time' process $\{T_t\}$ can be interpreted as the time over which market prices evolve, and is often associated with trading volume

or the flow of new price-sensitive information. That is, the more share trades that occur, or the more information released to the market on a given day, the faster 'time' progresses. In fact $\{T_t\}$ plays a crucial role in our model, determining both the distribution of the log price increments and their correlation structure.

The log increments of $\{P_t\}$ (continuously compounded returns) are then given by

$$\begin{aligned} X_t &= \log(P_t) - \log(P_{t-1}) \\ &= \mu + \theta\tau_t + \sigma(B(T_t) - B(T_{t-1})) \qquad (1.5) \\ &\stackrel{\mathcal{D}}{=} \mu + \theta\tau_t + \sigma\tau_t^{1/2}B(1). \qquad (1.6) \end{aligned}$$

From (1.6) it follows that $\mathbb{E}X_t = \mu + \theta$ and (Seneta (2004)),

$$\mathbb{E}(X_t - \mathbb{E}X_t)^2 = \sigma^2 + \theta^2 M_2 \qquad (1.7)$$

$$\mathbb{E}(X_t - \mathbb{E}X_t)^3 = 3\theta\sigma^2 M_2 + \theta^3 M_3 \qquad (1.8)$$

$$\mathbb{E}(X_t - \mathbb{E}X_t)^4 = 3\sigma^4(1 + M_2) + 6\sigma^2\theta^2(M_2 + M_3) + \theta^4 M_4 \qquad (1.9)$$

where $M_i = \mathbb{E}(\tau_t - 1)^i,\ i = 2, 3, 4$ and $\mathbb{E}\tau_t = 1$, assuming these moments exist. Therefore, the coefficients of skewness and kurtosis are

$$\beta = \frac{3\theta\sigma^2 M_2 + \theta^3 M_3}{(\sigma^2 + \theta^2 M_2)^{3/2}} \qquad (1.10)$$

and

$$\kappa = \frac{3\sigma^4(M_2 + 1) + 6\sigma^2\theta^2(M_2 + M_3) + \theta^4 M_4}{(\sigma^2 + \theta^2 M_2)^2} \qquad (1.11)$$

respectively. From (1.6) one can immediately see that in the case $\theta = 0$ a symmetric model results, whereas from (1.10) for example, for $\theta \neq 0$ the model implies skewed returns.

The two specific models that we shall be concerned with (the VG and t) were first considered as symmetric models, taking P_t as

$$P_t = P_0 e^{\mu t + \sigma B(T_t)}. \qquad (1.12)$$

Later these models were extended to allow for skewness by introducing the θ parameter. Although we shall in general be concerned with models that allow for skewness via θ, we occasionally restrict discussion to the symmetric $\theta = 0$ case given by (1.12); we consider the $\theta = 0$ case either to discuss the historical development of the models, or in some instances to simplify the mathematics.

The idea of subordination was developed by Bochner (1955), and later taken up by many authors as a flexible method for modelling financial time series data. See for example Hurst, Platen & Rachev (1997) for an overview of a number of popular subordinator models. The flexibility of the subordinator model owes to the fact that most important properties of the $\{X_t\}$ process, including its moments and degree of autocorrelation, are determined by the $\{\tau_t\}$ process. Hence, via the specification of the distribution of T_t, one can accommodate a wide variety of models within the same subordinator model framework.

In fact

$$\begin{aligned}\mathbb{C}\text{ov}(X_t, X_{t+k}) &= \mathbb{C}\text{ov}(\theta\tau_t + \sigma\tau_t^{1/2}B_1(1), \theta\tau_{t+k} + \sigma\tau_{t+k}^{1/2}B_2(1)) \\ &= \theta^2(\mathbb{E}(\tau_t\tau_{t+k}) - \mathbb{E}\tau_t\mathbb{E}\tau_{t+k}) \\ &= \theta^2\mathbb{C}\text{ov}(\tau_t, \tau_{t+k}) \qquad (1.13)\end{aligned}$$

for B_1 and B_2 independent Brownian motions, so that at $\theta = 0$

$$\mathbb{C}\text{ov}(X_t, X_{t+k}) = 0. \qquad (1.14)$$

For $\mu = \theta = 0$ we also have

$$\begin{aligned}\mathbb{C}\text{ov}(|X_t|, |X_{t+k}|) &= \mathbb{C}\text{ov}(|\sigma\tau_t^{1/2}B_1(1)|, |\sigma\tau_{t+k}^{1/2}B_2(1)|) \\ &= \sigma^2\mathbb{E}(|B_1(1)|)\mathbb{E}(|B_2(1)|)\mathbb{C}\text{ov}(\tau_t^{1/2}, \tau_{t+k}^{1/2}) \\ &= \frac{2}{\pi}\sigma^2\mathbb{C}\text{ov}(\tau_t^{1/2}, \tau_{t+k}^{1/2})\end{aligned}$$

while, in regards to $\{X_t^2\}$ in the general $\mu, \theta \neq 0$ case, we have that

$$\begin{aligned}\mathbb{C}\text{ov}(X_t^2,\ X_{t+k}^2) &= \mathbb{C}\text{ov}((\mu + \theta\tau_t + \sigma\tau_t^{1/2}B_1(1))^2,\ (\mu + \theta\tau_{t+k} + \sigma\tau_{t+k}^{1/2}B_2(1))^2) \\ &= (\sigma^4 + 4\theta^2\mu^2 + 4\theta\mu\sigma^2)\mathbb{C}\text{ov}(\tau_t,\ \tau_{t+k}) + \theta^4\mathbb{C}\text{ov}(\tau_t^2,\ \tau_{t+k}^2) \\ &\quad + (\theta^2\sigma^2 + 2\theta^3\mu)(\mathbb{C}\text{ov}(\tau_t^2,\ \tau_{t+k}) + \mathbb{C}\text{ov}(\tau_t,\ \tau_{t+k}^2)) \qquad (1.15)\end{aligned}$$

which for $\theta = 0$ reduces to

$$\mathbb{C}\text{ov}(X_t^2,\ X_{t+k}^2) = \sigma^4\mathbb{C}\text{ov}(\tau_t,\ \tau_{t+k}). \qquad (1.16)$$

For $\{\tau_t\}$ with dependence structure, $\{X_t\}$ also displays conditional heteroscedasticity, that is, time dependent conditional variance. Let $\mathcal{F}_t = \sigma(\{B(u), u \leq T_t\}, \{T_u, u \leq t\})$ which can be thought of as information available up to time t. Then

$$\begin{aligned}\mathbb{V}\text{ar}(X_t|\mathcal{F}_{t-1}) &= \mathbb{E}(X_t^2|\mathcal{F}_{t-1}) - (\mathbb{E}(X_t|\mathcal{F}_{t-1}))^2 \\ &= \theta^2\ \mathbb{E}(\tau_t^2|\mathcal{F}_{t-1}) + (\sigma^2 + 2\mu\theta)\mathbb{E}(\tau_t|\mathcal{F}_{t-1}) \\ &\quad - (2\mu\theta\ \mathbb{E}(\tau_t|\mathcal{F}_{t-1}) + \theta^2\ (\mathbb{E}(\tau_t|\mathcal{F}_{t-1}))^2) \\ &= \theta^2\ \mathbb{V}\text{ar}(\tau_t|\mathcal{F}_{t-1}) + \sigma^2\ \mathbb{E}(\tau_t|\mathcal{F}_{t-1}).\end{aligned}$$

Under the restricted model of $\theta = 0$, the above expression reduces to $\mathbb{V}\text{ar}(X_t|\mathcal{F}_{t-1}) = \sigma^2\ \mathbb{E}(\tau_t|\mathcal{F}_{t-1})$ (Heyde & Liu (2001)).

The remarkable results (1.14) and (1.16) in the symmetric case $\theta = 0$ can already be found in Heyde & Liu (2001), and the main point to note is that for small θ, $\{X_t\}$ will appear to have an autocorrelation function of zero, while the acf of $\{X_t^2\}$ will mimic that of $\{\tau_t\}$. In particular if $\{\tau_t\}$ has a dependence structure then so has $\{X_t^2\}$.

1.3.1. The efficient markets hypothesis and the subordinator model. Returning to (1.5), it is readily seen that $\mathbb{E}(B(T_t)|\{B(T_u), u \leq t-1\}, \{T_u, u \leq t-1\}) = B(T_{t-1})$, which may be interpreted (c.f., Campbell, Lo & MacKinlay (1997, pp. 23–24)) as the efficient markets hypothesis holding for the symmetric version of the model, since $\{B(T_u), u = 1, 2, \ldots\}$ is a martingale and, hence, $\{X_t\}$ is a martingale difference sequence (Heyde & Liu (2001)

expressed this in a related, continuous time way). Here we take the efficient markets hypothesis as amounting to the assertion that current market prices reflect all available information, so that, in particular, past information about the stock price process cannot be used to predict future stock price returns. The striking property that $\mathbb{C}\text{ov}(X_t, X_{t+k}) = 0$ in the case $\theta = 0$ may consequently be regarded as a manifestation of the efficient markets hypothesis, and as such is supportive of our model.

For the general $\theta \neq 0$ model we do not have $\mathbb{C}\text{ov}(X_t, X_{t+k}) = 0$, but we see from (1.13) and (1.7) that, for $k \geq 1$,

$$|\mathbb{C}\text{orr}(X_t, X_{t+k})| = \frac{\theta^2 |\mathbb{C}\text{ov}(\tau_t, \tau_{t+k})|}{\theta^2 \mathbb{V}\text{ar}(\tau_t) + \sigma^2} \leq \theta^2 \frac{|\mathbb{C}\text{orr}(\tau_t, \tau_{t+k})|}{\sigma^2} \mathbb{V}\text{ar}(\tau_t) \leq \theta^2 \frac{\mathbb{V}\text{ar}(\tau_t)}{\sigma^2}. \tag{1.17}$$

From estimation procedures applied to the asymmetric VG and t distributional models for returns to be discussed shortly, for four sets of observations $\{X_t\}$, each of approximate length $N = 2500$, evaluation of the right-hand side of this equation resulted in a bound of about 0.02. Inasmuch as white noise would produce about 1 in 20 sample autocorrelations outside the range $\pm 2/\sqrt{N} = \pm 0.04$, it seems unlikely that data actually described by our general model would have a sample acf distinguishable from that of white noise. Hence, for all practical applications, the efficient markets hypothesis criteria $\mathbb{C}\text{ov}(X_t, X_{t+k}) = 0$ as given above would also appear to hold, and could not be rejected, in the asymmetric case also.

It could be argued that the efficient markets hypothesis in any case would imply the statistical independence of returns, which would imply that $\mathbb{C}\text{ov}(X_t^r, X_{t+k}^r) = 0$, $r = 1, 2, \ldots$. However, there is evidence that for $r = 2$ this is not the case, with actual data showing $\mathbb{C}\text{ov}(X_t^2, X_{t+k}^2) > 0$ even at relatively long lag lengths k (see for example Ding, Granger & Engle (1993) or Taylor (1986)). It may also be argued that real-world markets are not entirely efficient, in support of our general model.

For completeness we note that there is a body of work, based on data analysis for symmetric models continuing from that of Heyde & Liu (2001), which suggests that $\mathbb{E}(X_t^4)$ may be infinite. This corresponds to $\mathbb{V}\text{ar}(\tau_t)$ being infinite, in which case the above argument, based on (1.17), would not be applicable.

1.3.2. Volatility and the interpretation of $\{\tau_t\}$. In the probabilistic guise of (1.6), when $\theta = 0$ and $\{\tau_t\}, t = 1, 2, \ldots$ are independent and identically distributed, we note that our model reduces to the familiar random volatility model (see for example Taylor (1994)). Thus in our model it is natural to interpret $\sigma\tau_t^{1/2}$ as the volatility at time t and, hence, the stationary process $\{\sigma\tau_t^{1/2}\}$ as describing a stationary random process of stochastically dependent volatilities. This can be seen from an equivalent standpoint by noting that the conditional distribution of the right-hand side of (1.6), given $V \equiv \tau_t$ and $\theta = 0$ is $N(\mu, \sigma^2 V)$. In the general model (1.5) for $\theta \neq 0$ then, where the conditional distribution of the right-hand side is

$$N(\mu + \theta V, \sigma^2 V),$$

we may think of the size and direction of the fluctuation in random mean as being driven by volatility. We recognise the (marginal) distribution described by the above expression as being that of the normal mixed distributions (terminology due to Barndorff-Nielsen, Kent & Sørensen (1982)).

In the case $\theta = 0$, as we have seen, the autocorrelations $\mathbb{C}\text{orr}(X_t, X_{t+k})$ are all zero. One of the motivations for investigating this model was to reflect the fact that asset returns, while displaying a sample acf plot characteristic of white noise, no longer do so in sample acf plots of squared returns and absolute values of returns. This, however, should be the case if the process $\{X_t\}$ is indeed a sequence of independent, identically distributed random variables. Hence a plausible model for $\{X_t\}$ should have the capability of reflecting this long term dependence in $\{X_t^2\}$ as well as the weak autocorrelation structure of the values

$\{X_t\}$ themselves. Indeed, inasmuch as the values X_t^2 are themselves measures of volatility, such a model would impart a desirable dependence structure on the evolution of volatilities.

1.4. The VG and the t models

We now specialise our general model for $\{X_t\}$ by focusing on two distributions for τ_t, the gamma (Γ) and the inverse or reciprocal gamma ($R\Gamma$). These respectively are the distributions which result in X_t increments with the VG and t distribution. Note that taking $T_t \equiv t$ ($\tau_t \equiv 1$) of course results in the classical log normal distribution for P_t and so a normal distribution for X_t.

1.4.1. The VG distribution. We say that $\tau_t \sim \Gamma(\alpha, \lambda)$ for $\alpha, \lambda > 0$ has a marginal gamma distribution if its probability density function (pdf) is of the form

$$f_\Gamma(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x},\ x > 0. \tag{1.18}$$

We choose $\lambda = \alpha$ so that $\mathbb{E}(\tau_t) = 1$ and

$$\mathbb{V}\text{ar}(\tau_t) = \frac{1}{\alpha}. \tag{1.19}$$

This choice of a gamma distribution for τ_t, coupled with the model (1.6), results in X_t having the marginal (skew) Variance Gamma distribution (Madan, Carr & Chang (1998)) with pdf

$$f_{\text{VG}}(x) = \sqrt{\frac{2}{\pi}} \frac{\lambda^\alpha e^{\frac{(x-\mu)\theta}{\sigma^2}}}{\sigma\Gamma(\alpha)} \left(\frac{|x-\mu|}{\sqrt{\theta^2 + 2\lambda\sigma^2}}\right)^{\alpha-\frac{1}{2}} K_{\alpha-\frac{1}{2}}\left(\frac{|x-\mu|\sqrt{\theta^2+2\lambda\sigma^2}}{\sigma^2}\right) \tag{1.20}$$

for $x \in \mathbb{R}$ and characteristic function (cf)

$$\phi_{\text{VG}}(u) = e^{\mathrm{i}\mu u}(1 - \mathrm{i}\theta u/\lambda + \frac{1}{2}\sigma^2 u^2/\lambda)^{-\alpha} \tag{1.21}$$

for $\lambda = \alpha$. Denote the distribution defined by (1.20) and (1.21) by VG$(\mu, \theta, \sigma^2, \alpha, \lambda)$. The symmetric VG distribution and model, which results from taking $\theta = 0$, was introduced in the financial context by Madan & Seneta (1990). See Section 1.7 for a discussion of

the historical development of the VG and t models. The VG, a pure jump process, was called VG since the log increments X_t have conditional variance given by a gamma random variable – from (1.6) for $\theta = 0$, $X_t|V \stackrel{\mathcal{D}}{=} N(\mu, \sigma^2 V)$ where V has the gamma distribution. Through out this text we use the term 'VG' interchangeably for the VG distribution and the VG model or stochastic process, and trust that the meaning is made clear by the context.

In the above, $K_\eta(\omega)$ for $\eta \in \mathbb{R}$ and $\omega > 0$, given by

$$K_\eta(\omega) = \frac{1}{2}\int_0^\infty z^{\eta-1} e^{-\frac{\omega}{2}(z+\frac{1}{z})} \mathrm{d}z, \tag{1.22}$$

is a modified Bessel function of the third kind (Erdélyi, Magnus, Oberhettinger, and Tricomi, 1954) with index η, and the VG is sometimes also known as the Bessel K-function distribution (see Johnson, Kotz & Balakrishnan (1994, pp. 50–51)). Note that $K_\eta(\omega)$ is referred to as a modified Bessel function of the *second* kind in some texts.

Specialising (1.7), (1.8) and (1.9), the central moments of X_t are then given by

$$\begin{aligned}
\mathbb{E}(X_t - \mathbb{E}X_t)^2 &= \sigma^2 + \frac{\theta^2}{\alpha} \\
\mathbb{E}(X_t - \mathbb{E}X_t)^3 &= \frac{3\theta\sigma^2}{\alpha} + \frac{2\theta^3}{\alpha^2} \\
\mathbb{E}(X_t - \mathbb{E}X_t)^4 &= \frac{3\sigma^4(\alpha+1)}{\alpha} + \frac{6\sigma^2\theta^2(2+\alpha)}{\alpha^2} + \frac{3\theta^4(\alpha+2)}{\alpha^3}.
\end{aligned}$$

Before turning to the t, we lastly note that while Brownian motion is of infinite variation, the VG, which is Brownian motion time-changed by a gamma process, is of bounded variation (see for example Madan, Carr & Chang (1998)). In fact from Yor (2007) this result applies more generally, with the gamma process time-changing many 'erratic' processes into processes of bounded variation. In particular, for any process $\{Y(t)\}$, $t \geq 0$ such that there exists constants K and c for which

$$\mathbb{E}(|Y(u) - Y(v)|) \leq K|u - v|^c \tag{1.23}$$

for all $u, v \geq 0$, the process $\{Y(G(t;a,b))\}$, where $\{G(t;a,b)\}$ is an independent increments gamma process with $G(t;a,b) \sim \Gamma(ta,b)$ for any given t, and where $\{G(t;a,b)\}$ is independent of $\{Y(t)\}$, is of bounded variation. This follows since for any partition $\mathcal{P} = \{x_0, \cdots, x_{n_\mathcal{P}}\}$ of $[0,T]$,

$$\mathbb{E}\left(\sum_{i=0}^{n_\mathcal{P}-1} |Y(G(x_{i+1};a,b)) - Y(G(x_i;a,b))|\right)$$

$$\leq K \sum_{i=0}^{n_\mathcal{P}-1} \mathbb{E}(G(x_{i+1};a,b) - G(x_i;a,b))^c \tag{1.24}$$

$$= \frac{K}{b^c} \sum_{i=0}^{n_\mathcal{P}-1} \frac{\Gamma((x_{i+1}-x_i)a + c)}{\Gamma((x_{i+1}-x_i)a)} \tag{1.25}$$

$$= \frac{aK}{b^c} \sum_{i=0}^{n_\mathcal{P}-1} \frac{\Gamma((x_{i+1}-x_i)a + c)(x_{i+1}-x_i)}{\Gamma((x_{i+1}-x_i)a + 1)}$$

where (1.24) follows from (1.23) by conditioning on $\{G(t;a,b)\}$, (1.25) follows since $\{G(t;a,b)\}$ has independent gamma increments and $\mathbb{E}X^c = \beta^{-c}\Gamma(\alpha + c)/\Gamma(\alpha)$ for $X \sim \Gamma(\alpha, \beta)$. Now as the partition $\mathcal{P}$ gets finer and $x_{i+1} - x_i \to 0$ for each i,

$$\frac{\Gamma((x_{i+1}-x_i)a + c)(x_{i+1}-x_i)}{\Gamma((x_{i+1}-x_i)a + 1)} \sim \Gamma(c)(x_{i+1}-x_i)$$

so that

$$\mathbb{E}\left(\sum_{i=0}^{n_\mathcal{P}-1} |Y(G(x_{i+1};a,b)) - Y(G(x_i;a,b))|\right) \leq \frac{aK}{b^c}\Gamma(c)T$$

which establishes the result.

1.4.2. The t distribution. We say that $\tau_t \sim R\Gamma(\delta, \epsilon)$ for $\delta, \epsilon > 0$ has a marginal inverse gamma distribution if its pdf is of the form

$$f_{R\Gamma}(x;\delta,\epsilon) = \frac{\epsilon^\delta}{\Gamma(\delta)} x^{-\delta-1} e^{\frac{-\epsilon}{x}}, \; x > 0. \tag{1.26}$$

We choose $\epsilon = \delta - 1$ so that $\mathbb{E}(\tau_t) = 1$, which gives, for $\delta > 2$,

$$\mathbb{V}\text{ar}(\tau_t) = \frac{1}{\delta - 2}. \tag{1.27}$$

This choice of an inverse gamma distribution for τ_t, coupled with the model (1.6), results in X_t having the marginal (skew) t distribution with pdf for $x \in \mathbb{R}$,

$$f_{\text{t-dist}}(x) = \begin{cases} \frac{\Gamma(\delta+\frac{1}{2})}{\sqrt{2\sigma^2\epsilon}\sqrt{\pi}\Gamma(\delta)} \frac{1}{(1+(\frac{x-\mu}{\sqrt{2\sigma^2\epsilon}})^2)^{\delta+\frac{1}{2}}} & \text{for } \theta = 0 \\ \sqrt{\frac{2}{\pi}} \frac{(\delta-1)^\delta e^{\frac{(x-\mu)\theta}{\sigma^2}}}{\sigma\Gamma(\delta)} (\frac{\theta^2}{2\epsilon\sigma^2+(x-\mu)^2})^{\frac{\delta+1/2}{2}} K_{\delta+\frac{1}{2}}(\frac{|\theta|\sqrt{2\epsilon\sigma^2+(x-\mu)^2}}{\sigma^2}) & \text{for } \theta \neq 0 \end{cases}$$

(see Sørensen & Bibby (2003)), and cf

$$\phi_{\text{t-dist}}(u) = \frac{2^{1-\frac{\delta}{2}}\, e^{i\mu u}}{\Gamma(\delta)} (\epsilon(\sigma^2u^2 - 2i\theta u))^{\frac{\delta}{2}} K_\delta(\sqrt{2\epsilon(\sigma^2u^2 - 2i\theta u)})$$

for $\epsilon = \delta - 1$. The number of degrees of freedom, ν, is defined by $\nu = 2\delta$, this being a direct generalisation of the concept for the classical t-distribution. In the symmetric case in which $\theta = 0$, the distribution is the scaled t-distribution and is well known. It is a slight generalisation of the classical Student's t-distribution, and is at the heart of the paper of Heyde & Leonenko (2005). Seneta (2004) Section 6 discussed its parallelism and duality with the (symmetric) VG distribution. Here $X_t|V \overset{\mathcal{D}}{=} N(\mu + \theta V, \sigma^2 V)$, where V has an inverse gamma distribution as above (and thus $1/V$ has a gamma distribution, but not with unit expectation).

From (1.7), (1.8) and (1.9) and assuming $\delta > 4$, the central moments of X_t in the t distribution case are given by,

$$\mathbb{E}(X_t - \mathbb{E}X_t)^2 = \sigma^2 + \frac{\theta^2}{\delta - 2}$$

$$\mathbb{E}(X_t - \mathbb{E}X_t)^3 = \frac{3\theta\sigma^2}{\delta - 2} + \frac{4\theta^3}{(\delta-2)(\delta-3)} \tag{1.28}$$

$$\mathbb{E}(X_t - \mathbb{E}X_t)^4 = 3\sigma^4 + \frac{3\sigma^2(2\theta^2 + \sigma^2)}{\delta - 2} + \frac{24\sigma^2\theta^2}{(\delta-2)(\delta-3)} + \frac{3\theta^4(\delta+5)}{(\delta-2)(\delta-3)(\delta-4)}. \tag{1.29}$$

Note that $\delta \leq 2$ in the t distribution case implies the infinite variance of τ_t, and indeed of X_t in the case $\theta \neq 0$.

1.4.3. Comparison between the VG and t. Both the VG and t distribution have the normal distribution as a limiting case, but are perhaps illuminated in general by noting that the symmetric t distribution is a generalisation of the Cauchy distribution (the case $\theta = 0$, $\delta = \frac{1}{2}$ in the above), while the symmetric VG distribution is a generalisation of the Laplace distribution, also sometimes called the double exponential distribution, which results from putting $\theta = 0$, $\alpha = 1$ in $f_{\text{VG}}(x)$, since $K_{\frac{1}{2}}(\omega) = e^{-\omega}(\pi/(2\omega))^{1/2}$ for $\omega > 0$ (Barndorff-Nielsen & Blæsild (1981)).

Notice that putting $\alpha = 1$ in $f_{\Gamma}(x)$ makes it the pdf of an exponential distribution, which in the symmetric case $\theta = 0$ results in a parallelism with the double exponential distribution of X_t. In the case when the τ_t's can be taken as independent as well as exponentially distributed, the $\{T_t\},\ t = 1, 2, \cdots$ process may be understood as the sequence of time points at which events of a Poisson process occur.

Regarding the question of tailweight, we have that for $\theta = \mu = 0$,

$$P_{\text{VG}}(|X_t| > x) \sim Const.(\alpha, \sigma))x^{\alpha-1}e^{-x\sqrt{2\alpha/\sigma^2}} \tag{1.30}$$
$$P_t(|X_t| > x) \sim Const.(\delta, \sigma)x^{-2\delta},\ x \to \infty.$$

Thus, statistical techniques which indicate non-existence of higher moments (as in Heyde & Liu (2001)) may incline the potential user towards the t in preference to the VG for which all moments exist.

However, Heyde & Kou (2004) demonstrate that in practice as many as 100,000 readings may be necessary to distinguish between exponential (Laplace distribution type) and power law (t distribution type) tails. Now, as we have seen, the Laplace distribution is a special case of the symmetric VG distribution. So allowing for a power law modification of exponential decay as in the symmetric VG makes distinction even more difficult. This result is supported by Fung & Seneta (2007), which compared the symmetric VG and t distributions, and found the two models almost impossible to distinguish between on the basis of

tailweight, by choice of compatible parameters. Note that these results just mentioned are based in particular on iid simulations.

For the VG at $\alpha = 1$ where X_t has Laplace distribution the relation (1.30) is exact and not just asymptotic. For $\alpha > 1$ the pdf of the VG is smooth and increasingly bell-shaped for large α. For $\frac{1}{2} < \alpha \leq 1$ the pdf is cusped at the origin, while for $0 < \alpha \leq \frac{1}{2}$ it is unbounded at the origin. Decreasing the value of α from a value > 1 has the effect of increasing probability near the origin and in the tails, at the expense of probability in the middle range (Madan & Seneta (1990)). This may be seen clearly by numerical investigation of quantiles after standardisation as done in Fung & Seneta (2007), which also displays graphics of the symmetric VG and t distribution for various parameter values. Note that standardisation there is by standard deviation; standardisation by interquartile range is not appropriate for this purpose since that controls the amount of probability near the origin.

1.4.4. The GH as a generalisation of the VG and t. Although we shall concentrate on the VG and t, it is interesting to note that both distributions are in fact special cases of the Generalised Hyperbolic (GH) distribution (Barndorff-Nielsen & Halgreen (1977)), owing to the fact that the gamma and inverse gamma are special cases of the Generalised Inverse Gaussian (GIG) distribution. Here we say that $\tau_t \sim GIG(\alpha, \beta, \gamma)$ has marginal GIG distribution when its pdf is of the form

$$f_{\text{GIG}}(x) = \frac{(\gamma/\beta)^{\alpha/2}}{2K_\alpha(\sqrt{\beta\gamma})} x^{\alpha-1} e^{-\frac{1}{2}(\frac{\beta}{x}+\gamma x)}, \; x > 0.$$

Due to the complexity of the GIG distribution we cannot impose the unit expectation constraint for τ_t explicitly to reduce the number of parameters. In the above the admissible parameter values are $\beta > 0, \gamma \geq 0$ for $\alpha < 0$; $\beta > 0, \gamma > 0$ for $\alpha = 0$; and $\beta \geq 0, \gamma > 0$ for $\alpha > 0$.

For τ_t with the GIG distribution, X_t from (1.6) will have marginal (skew) Generalised Hyperbolic distribution with pdf

$$f_{\text{GH}}(x) = \frac{(\frac{\gamma}{\beta})^{\frac{\alpha}{2}}(\frac{\beta\sigma^2+(x-\mu)^2}{\gamma\sigma^2+\theta^2})^{\frac{\alpha}{2}-\frac{1}{4}}K_{\alpha-\frac{1}{2}}(\sqrt{(\gamma+\frac{\theta^2}{\sigma^2})(\beta+\frac{(x-\mu)^2}{\sigma^2})})e^{\frac{(x-\mu)\theta}{\sigma^2}}}{\sqrt{2\pi\sigma^2}K_\alpha(\sqrt{\beta\gamma})}$$

and cf

$$\phi_{\text{GH}}(u) = \frac{K_\alpha(\sqrt{\beta(\gamma-2i\theta u+\sigma^2u^2)})}{K_\alpha(\sqrt{\beta\gamma})}(\frac{\gamma}{\gamma-2i\theta u+\sigma^2u^2})^{\alpha/2}e^{i\mu u}.$$

But from Barndorff-Nielsen & Blæsild (1981) for example, as $\omega \downarrow 0$ the Bessel function (1.22) has the asymptotic properties:

$$K_\eta(\omega) \sim \begin{cases} -\log(\omega) & \text{for } \eta = 0 \\ 2^{|\eta|-1}\Gamma(|\eta|)\omega^{-|\eta|} & \text{for } \eta \neq 0 \end{cases}$$

from which we obtain as $\beta \to 0$, $\alpha > 0$ and $\gamma \to 0$, $\alpha < 0$ the gamma and inverse gamma respectively for the distribution of τ_t, and correspondingly the VG and t from the Generalised Hyperbolic.

In the symmetric case $\theta = 0$ when $\gamma \neq 0$, the hyperbolic family, which includes the VG, clearly have distribution tails decreasing asymptotically at power-law modified exponential rate, while when $\gamma = 0$ the rate of decay of the tails is slower: power-law (Pareto tails). It is this particular property of the symmetric scaled t distribution which is thought by some to make it preferable in practical modelling, and is used to justify the focus on 'Student' processes in the partly review paper of Heyde & Leonenko (2005), although the gamma distribution and symmetric VG distribution are accorded some attention in that paper also.

1.5. Related Models

While researching the VG model we came across a number of other models which are related or can be considered as generalisations of the VG. We detail two of these below.

1.5.1. Heston and the VG. A popular model in finance is that of Heston (1993), which we very briefly touch on in relation to the VG. The Heston model is couched in terms of stochastic differential equations, and defines P_t, the stock price at time t, by

$$\mathrm{d}P_t = \mu P_t \mathrm{d}t + \sqrt{v_t} P_t \mathrm{d}B_t^{(1)} \tag{1.31}$$

$$\mathrm{d}v_t = -\gamma(v_t - \vartheta)\mathrm{d}t + \kappa\sqrt{v_t}\mathrm{d}B_t^{(2)} \tag{1.32}$$

where $\mathrm{d}B_t^{(2)} = \rho \mathrm{d}B_t^{(1)} + \sqrt{1-\rho^2}\mathrm{d}B_t^{(3)}$ for $B_t^{(3)}$ and $B_t^{(1)}$ independent Brownian motions, and $\rho \in [-1, 1]$. In this model v_t, the variance of P_t, is a mean-reverting stochastic process with long-term mean ϑ, relaxation rate to this mean γ, and 'volatility of volatility' κ.

Drăgulescu & Yakovenko (2002) show that for $t \ll \frac{1}{\gamma}$, that is, t much smaller that the relaxation time of variance, the unconditional distribution of $\log(P_t/P_0)$ under the Heston model reduces to the VG (their Equations (47) and (49)). The result comes about by noting that for small t, variance does not have time to change from its initial value v_0. Conditional on v_0 however, (1.31) describes geometric Brownian motion, so that

$$\log(P_t/P_0)|v_i \sim N((\mu - \frac{v_i}{2})t, \sqrt{tv_i}).$$

That is, the distribution of $\log(P_t/P_0)$ is conditionally normal with conditional variance of tv_i. The stationary distribution of v_0 from (1.32) is the gamma distribution however (see Feller (1951)), so integrating over v_0 to reach the unconditional distribution of $\log(P_t/P_0)$ results in the VG.

1.5.2. A correlated difference of gammas (DG) model. A possible variant on the VG would be to model log stock prices as the difference of two mutually independent but internally correlated gamma processes, since as well shall see, the standard symmetric VG distribution results from the difference of two independent and identically distributed gamma random variables. That is, we model stock prices by

$$P_t = P_0 e^{\mu t + G_1^*(t;a,b) - G_2^*(t;c,d)}.$$

This is similar to the process considered in Madan, Carr & Chang (1998), as well as the DG model we will consider in Chapter 4 in relation to option pricing, but in this case we take $\{G_1^*(t;a,b)\}$ and $\{G_2^*(t;c,d)\}$ such that $(G^*(t;\alpha,\beta) - G^*(t-1;\alpha,\beta)) \sim \Gamma(\alpha,\beta)$ with the increments process having a dependence structure, as opposed no dependence. As such we have

$$\begin{aligned} X_t &= \log(P_t) - \log(P_{t-1}) \\ &= \mu + (G_1^*(t) - G_1^*(t-1)) - (G_2^*(t) - G_2^*(t-1)) \\ &= \mu + A_t - B_t \end{aligned}$$

say, for $A_t \sim \Gamma(a,b)$ and $B_t \sim \Gamma(c,d)$, with

$$\begin{aligned} \mathbb{E}X_t &= \frac{a}{b} - \frac{c}{d} \\ \mathbb{V}\text{ar}(X_t) &= \frac{a}{b^2} + \frac{c}{d^2}. \end{aligned}$$

and more interestingly,

$$\begin{aligned} \mathbb{C}\text{ov}(X_t, X_{t+k}) &= \mathbb{C}\text{ov}(A_t - B_t, A_{t+k} - B_{t+k}) \\ &= \mathbb{C}\text{ov}(A_t, A_{t+k}) + \mathbb{C}\text{ov}(B_t, B_{t+k}) \end{aligned}$$

and

$$\begin{aligned} \mathbb{C}\text{ov}(X_t^2, X_{t+k}^2) &= \mathbb{C}\text{ov}(A_t^2, A_{t_k}^2) + \mathbb{C}\text{ov}(B_t^2, B_{t_k}^2) + \mathbb{C}\text{ov}(A_t B_t, A_{t_k} B_{t_k}) \\ &+ (4\mu^2 + 8\mu\frac{c}{d})\mathbb{C}\text{ov}(A_t, A_{t_k}) + (2\mu + 2\frac{c}{d})(\mathbb{C}\text{ov}(A_t^2, A_{t_k}) + \mathbb{C}\text{ov}(A_t, A_{t_k}^2)) \\ &+ (4\mu^2 + 8\mu\frac{a}{b})\mathbb{C}\text{ov}(B_t, B_{t_k}) + (2\mu + 2\frac{a}{b})(\mathbb{C}\text{ov}(B_t^2, B_{t_k}) + \mathbb{C}\text{ov}(B_t, B_{t_k}^2)). \end{aligned}$$

The $\{A_t\}$ and $\{B_t\}$ here are merely assumed to be increments of some gamma process, similar to the marginally gamma distributed $\{\tau_t\}$ we will be concerned with in Chapter 2, and so we have many options in choosing the desired correlation structure. Note however that this correlated DG model implies $\mathbb{C}\text{ov}(X_t, X_{t+k}) \neq 0$ as well as $\mathbb{C}\text{ov}(X_t^2, X_{t+k}^2) \neq 0$, which as we have seen is not supported by financial data. Such a dependence structure may however be appropriate in other, non-financial applications.

1.6. Lévy processes

Continuous time models for the evolution of log prices are generally taken to be processes with stationary independent increments. These properties are fundamental in describing Lévy processes (Schoutens (2003), p. 44).

The classical instance of a Lévy process is Brownian motion. The original VG process as described in Madan & Seneta (1990) is another Lévy process. In this sense Lévy processes form a foundation for our development, although our interest is focussed on processes with dependent returns (log-price increments over unit time), which are therefore *not* Lévy processes.

The distribution of an increment of a Lévy process is infinitely divisible, and for each infinitely divisible distribution one can construct a Lévy process. Schoutens (2003) details a number of Lévy process applicable in finance, including the independent increments VG.

While the returns for the processes with which we shall be concerned form a strictly stationary sequence, and in the VG and t-cases the distribution of returns is in fact infinitely divisible, as continuous-time processes the log-price processes will not be Lévy processes, due to the dependence structure introduced.

From Grosswald (1976) (see also Barndorff-Nielsen & Halgreen (1977)) we recall that the t distribution is infinitely divisible and so consistent with a Lévy process. The Lévy process in continuous time associated with a t distribution for returns is quite complex and difficult to work with however. In particular, the probability distributions of increments over intervals of arbitrary length do not belong to the same simple t family as t-distributed returns. This is in contrast to the classical VG process, where the distributions of increments over intervals of arbitrary length all belong to the same VG family.

1.7. Historical discussion

Before leaving this chapter we briefly review the historical development and predecessor papers to the VG and t, borrowing quite heavily from Seneta (2007) and Seneta (2009) for reference material.

The t model for financial assets essentially started with Praetz (1972), who argued for independent log price increments given by variance-mixing on the normal according to $X|V \stackrel{\mathcal{D}}{=} N(\mu, \sigma^2 V)$ for V with the inverse gamma distribution, which is to say our (symmetric) t subordinator model described above. As discussed, this distribution is a generalisation of the classical Student's t-distribution, allowing for fractional degrees of freedom. The classical distribution itself was popularised by Student (1908), although in fact was known even before this: Lüroth (1876) and Edgeworth (1883) both derived the t-distribution, in a Bayesian context, as a posterior distribution for a population mean using a flat prior. See Zabell (2008) and Seneta (2008) for a discussion of the historical context and predecessor papers to Student's result.

For our own direction, the papers of Heyde (1999), Heyde & Liu (2001) and Heyde & Leonenko (2005) are key motivators – in increasing specialisation, they deal with 'fractal activity time' models, where the $\{T_t\}$ process from above is chosen such that τ_t has marginal inverse gamma distribution so that X_t has the t distribution; $\{T_t\}$ appropriately normed is asymptotically self similar (hence a 'fractal' model); and $\{\tau_t\}$ and therefore $\{X_t^2\}$ is LRD. The motivation for such models was the increasing evidence, already discussed, that while financial asset returns appear as a reasonable approximation to white noise, squared and absolute returns do not. As such, it was argued that any model for financial assets should be able to accommodate little dependence structure in returns, while at the same time allowing for LRD in squared returns. The use of a self similar activity time process coupled with

the subordinator model achieves this, is supported in the literature, and imparts desirable properties of its own (see Chapter 2).

Turning to the history of the VG, similar distributions also appeared very early in the literature. What we now call the VG was studied as far back as Pearson, Jeffery & Elderton (1929) and Pearson, Stouffer & David (1932), who defined the '$T_m(x)$' density function as (1.20) with $\mu = \theta = 0$, $\sigma^2 = 2$ and $\lambda = 1$ instead of $\lambda = \alpha$. Here the distribution was studied in relation to testing for differences between χ^2 values in contingency table data. A special case of the VG distribution was also studied by Kullback (1934), who considered the distribution of the difference and quotient of two independent $\Gamma(\alpha, 1)$ random variables; while from the variance-mixing structure of the distribution of X_t, Teichroew (1957) obtained the VG pdf (1.20) in the $\mu = \theta = 0$ case in terms of a Hankel function. More recently, McLeish (1982) studied the VG distribution and model, and we review his work below.

McLeish (1982) described what we now call the (symmetric) VG distribution as a normal multiplied by the square-root of a gamma random variable, which he was interested in as a robust alternative to the normal distribution. The mean-corrected pdf and more general moment generating function (mgf) given by McLeish in his Equations (3) and (5) are equivalent to our (1.20) with $\mu = \theta = 0$, $\lambda = 1$, $\sigma = \sqrt{2}$, and our (1.21) with $\theta = 0$, $\lambda = 1$ respectively. In addition, below his Equation (5) McLeish also points out that the symmetric VG arises from the difference of two independent and identically distributed gamma random variables, a point made also in Madan & Seneta (1990).

McLeish also examines the problem of constructing a VG process with dependence structure, suggesting two alternatives. In the first case for X_t as in (1.20) with distribution VG$(0, 0, \sigma^2, \alpha, \frac{1}{2})$, the model

$$X_{t+1} = \sqrt{B_t} X_t + e_t \tag{1.33}$$

is suggested. Here $B_t \sim B(p\alpha,(1-p)\alpha)$ has the Beta distribution while the e_t are again VG but with distribution VG$(0,0,\sigma^2,(1-p)\alpha,\frac{1}{2})$ for $0<p<1$, all independent. In this case, by independence we have

$$\begin{aligned}\mathbb{C}\text{orr}(X_t,X_{t+k}) &= \mathbb{C}\text{orr}(X_t,\sqrt{B_{t+k-1}}X_{t+k-1}+e_{t+k-1})\\ &= \mathbb{E}\sqrt{B_{t+k-1}}\,\mathbb{C}\text{orr}(X_t,X_{t+k-1})\\ &= (\mathbb{E}\sqrt{B_{t+k-1}})^k\,\mathbb{C}\text{orr}(X_t,X_t)\\ &= \rho^k\end{aligned}$$

where $\rho=\mathbb{E}\sqrt{B_t}=\frac{\Gamma(p\alpha+\frac{1}{2})\Gamma(\alpha)}{\Gamma(p\alpha)\Gamma(\alpha+\frac{1}{2})}$. Similarly, for $\{X_t^2\}$ we have

$$\begin{aligned}\mathbb{C}\text{orr}(X_t^2,X_{t+k}^2) &= \mathbb{C}\text{orr}(X_t^2,B_{t+k-1}X_{t+k-1}^2+2\sqrt{B_{t+k-1}}X_{t+k-1}e_{t+k-1}+e_{t+k-1}^2)\\ &= \mathbb{C}\text{orr}(X_t^2,B_{t+k-1}X_{t+k-1}^2)\\ &= \mathbb{E}B_{t+k-1}\,\mathbb{C}\text{orr}(X_t^2,X_{t+k-1}^2)\\ &= (\mathbb{E}B_{t+k-1})^k\,\mathbb{C}\text{orr}(X_t^2,X_t^2)\\ &= p^k.\end{aligned}$$

Note that here both $\{X_t\}$ and $\{X_t^2\}$ display non-zero short range dependence, since the acf decays sufficiently fast so as to be summable, as opposed to actual data which typically show $\mathbb{C}\text{orr}(X_t^2,X_{t+k}^2)>0$ with $\{X_t^2\}$ long range dependent, but $\mathbb{C}\text{orr}(X_t,X_{t+k})=0$.

Before moving on we should check that (1.33) does actually define a stationary sequence of correlated VG increments. If we denote cf of X_t with VG$(0,0,\sigma^2,\alpha,\frac{1}{2})$ distribution by $\phi_{\text{VG}}(u;\alpha)$, then the result will follow if $\sqrt{B_t}X_t$ has the VG$(0,0,\sigma^2,p\alpha,\frac{1}{2})$ distribution (and therefore has cf $\phi_{\text{VG}}(u;p\alpha)$), since

$$\phi_{\text{VG}}(u;p\alpha)\phi_{\text{VG}}(u;(1-p)\alpha)=\phi_{\text{VG}}(u;\alpha),$$

a property the VG inherits from the gamma distribution. Now

$$\sqrt{B_t}X_t \stackrel{\mathcal{D}}{=} \sigma\sqrt{B_tG}Z$$

for B_t, G and Z mutually independent with $G \sim \Gamma(\alpha, \frac{1}{2})$ and $Z \sim N(0,1)$. Hence (1.33) defines a stationary sequence if

$$`B(p\alpha,(1-p)\alpha) \times \Gamma(\alpha,\frac{1}{2}) \stackrel{\mathcal{D}}{=} \Gamma(p\alpha,\frac{1}{2})',$$

which is a distribution theory result proved for example in Yeo & Milne (1991).

The second suggestion from McLeish (1982) for incorporating a dependence structure is much closer to our own. In this case X_t is defined as

$$X_t = \sqrt{\tau_t} Z_t \tag{1.34}$$

for the $\{Z_t\}$ standard normal and the $\{\tau_t\}$ gamma distributed, both sequences stationary and independent of each other, but possibly internally dependent. McLeish gives in his Equations (21) and (22) the autocovariance functions of $\{X_t\}$ and $\{X_t^2\}$ as

$$\mathbb{C}\text{ov}(X_t, X_{t+k}) = \psi(k)\mathbb{E}\sqrt{\tau_t \tau_{t+k}} \tag{1.35}$$

$$\mathbb{C}\text{ov}(X_t^2, X_{t+k}^2) = (1+2\psi^2(k))\mathbb{E}(\tau_t\tau_{t+k}) - \mathbb{E}\tau_t \mathbb{E}\tau_{t+k} \tag{1.36}$$

where $\psi(k)$ is the autocorrelation function of $\{Z_t\}$. Once again for $\psi(k) \neq 0$ this model implies a non-zero correlation for $\{X_t\}$, but if one takes $\psi(k) = 0$ for $k \neq 0$ then (1.35) and (1.36) reduce to our (1.14) and (1.16).

We can extend the model (1.34) by continuing to allow the $\{Z_t\}$ to be autocorrelated and introducing skewness via the model

$$X_t = \theta\tau_t + \sqrt{\tau_t} Z_t.$$

In this case $\{X_t\}$ will have the properties

$$\mathbb{C}\text{ov}(X_t, X_{t+k}) = \theta^2 \mathbb{C}\text{ov}(\tau_t, \tau_{t+k}) + \psi(k)\mathbb{E}\sqrt{\tau_t\tau_{t+k}}$$

$$\begin{aligned}\mathbb{C}\text{ov}(X_t^2, X_{t+k}^2) = \theta^2 &(\mathbb{C}\text{ov}(\tau_t^2, \tau_{t+k}) + \psi(k)\mathbb{E}(\tau_t^{1\frac{1}{2}}\tau_{t+k}^{1\frac{1}{2}}) + \mathbb{C}\text{ov}(\tau_t, \tau_{t+k}^2)) \\ &+ 2\theta(\mathbb{E}(\tau_t^{1\frac{1}{2}}\tau_{t+k})\mathbb{E}(Z_t Z_{t+k}^2) + \mathbb{E}(\tau_t \tau_{t+k}^{1\frac{1}{2}})\mathbb{E}(Z_t^2 Z_{t+k})) \\ &+ \theta^4 \mathbb{C}\text{ov}(\tau_t^2, \tau_{t+k}^2) + (1+2\psi^2(k))\mathbb{E}(\tau_t\tau_{t+k}) - \mathbb{E}\tau_t\mathbb{E}\tau_{t+k}.\end{aligned}$$

Regarding the introduction of dependence into the $\{\tau_t\}$ sequence, which our Chapter 2 focuses on, McLeish suggests a fairly simple autoregression

$$\tau_{t+1} = B_t\tau_t + \delta_t \tag{1.37}$$

where B_t, τ_t and δ_t are independent variables having respectively the $B(p\alpha, (1-p)\alpha)$, $\Gamma(\alpha, 1/(2\sigma^2))$ and $\Gamma((1-p)\alpha, 1/(2\sigma^2))$ distributions. Again this results in $\mathbb{C}\text{orr}(\tau_t, \tau_{t+k}) = (\mathbb{E}B_t)^k = p^k$ and $\mathbb{C}\text{orr}(\tau_t^2, \tau_{t+k}^2) = (\mathbb{E}B_t^2)^k = (p(p\alpha+1)/(\alpha+1))^k$ which leads to an $\{X_t\}$ process for log price increments for which $\{X_t^2\}$ is short range dependent. Again, as with (1.33), (1.37) also defines a stationary sequence.

CHAPTER 2

Activity time

As we have seen, given the general model (1.6), the properties and distribution of $\{X_t\}$ are essentially determined by those of $\{\tau_t\}$ and therefore $\{T_t\}$. In particular, the dependence structure of $\{\tau_t\}$ determines that of $\{X_t^2\}$. While most interest in modelling financial asset returns has focussed on models which imply independence of returns, and returns from real data do appear as a reasonable approximation to being independently and identically distributed by showing no significant autocorrelation, at least past one or two lags, squared returns often exhibit long range dependence. We presented evidence of this LRD in squared and absolute returns in Section 1.1.3, and there is growing recognition of this fact in the literature: see for example Greene & Fielitz (1979), Taylor (1986), Ding, Granger & Engle (1993), Ding & Granger (1996) and Willinger, Taqqu & Teverovsky (1999).

Now although the non-decreasing process $\{T_t\}$ cannot be exactly self-similar, as Heyde & Leonenko (2005) note between their Equations (5.4) and (5.5), there is in fact growing evidence – see for example Heyde (1999) and Heyde & Liu (2001) – that at least asymptotic self similarity of $\{T_t\}$, to a good degree of approximation, is supported by financial data. From Section 1.2 a self similar $\{T_t\}$ process with self similarity parameter $H \in (\frac{1}{2}, 1)$ implies the LRD of the $\{\tau_t\}$ increments process, which via (1.15) or (1.16) renders the stochastic process $\{X_t^2\}$ LRD as desired.

As such, in order to accommodate in our model these features of self similarity and LRD, this chapter concerns the construction of activity time processes $\{T_t\}$ with *identically but not independently distributed* gamma or inverse gamma unit increments, which display

LRD, and which converge to a self similar limit. The results pertaining to the gamma case are one of the main contributions of this thesis.

The main guide for our work comes from Heyde (1999) and Heyde & Leonenko (2005): as mentioned, Heyde (1999) introduces the symmetric subordinator model (1.12) driven by a fractal (self similar) activity time process, which Heyde & Leonenko (2005) extend to allow for t distributed asset price returns. That is to say, Heyde & Leonenko (2005) construct a process $\{T_t\}$ with inverse gamma distributed increments, which displays LRD, and which converges to a self similar limit. We follow their method in constructing an analogous gamma process, and extend their result to allow the gamma and inverse gamma distributed processes to have a wider range of autocorrelation function and the gamma process to have distributional parameter which is allowed to be non-integer (that is, the α parameter from Chapter 1 is allowed to be non-integer).

2.1. The gamma case

First we construct a discrete activity time process $\{T_t\}$ which has gamma unit increments, displays LRD, and has asymptotically a (continuous time) self similar limit. Results pertaining to the integer ν construction for $\nu = 2\alpha$ have appeared in Finlay & Seneta (2006), while the extension to allow for non-integer ν was made in Finlay & Seneta (2007). To accord with Finlay & Seneta (2007) and to stress dependence on ν, we write $\tau_\nu(t)$ instead of τ_t as earlier for the increments of the $\{T_t\}$ process.

2.1.1. Construction of the gamma activity time process. For $\mathbb{N} = \{1, 2, 3, \cdots\}$ let $\{\eta_i(t),\ t \in \mathbb{N}\},\ i = 1, \ldots, [\nu],\ \nu \geq 1$ and $[\cdot]$ denoting integer part, be independent and identically distributed stationary Gaussian processes with zero mean, unit variance, and

autocorrelation function $\rho(s),\ s \in \mathbb{N}$. Define the stationary process $\{\tau_{[\nu]}(t)\}$, $t \in \mathbb{N}$ by

$$\tau_{[\nu]}(t) = (\eta_1^2(t) + \cdots + \eta_{[\nu]}^2(t))/[\nu]. \tag{2.1}$$

Then we can set $T_t = \sum_{i=1}^t \tau_{[\nu]}(i)$ so that for each integer $t \geq 1,\ T_t - T_{t-1} = \tau_{[\nu]}(t) \stackrel{\mathcal{D}}{=} \Gamma(\frac{[\nu]}{2}, \frac{[\nu]}{2})$ with $\mathbb{E}\tau_{[\nu]}(t) = 1$, $\mathbb{V}\text{ar}(\tau_{[\nu]}(t)) = \frac{2}{[\nu]}$ and

$$\mathbb{C}\text{ov}(\tau_{[\nu]}(t),\ \tau_{[\nu]}(t+s)) = \frac{1}{[\nu]}\mathbb{C}\text{ov}(\eta_1^2(t),\ \eta_1^2(t+s)) \tag{2.2}$$

$$= \frac{1}{[\nu]}(\mathbb{E}(\eta_1^2(t)\eta_1^2(t+s)) - 1) \tag{2.3}$$

$$= \frac{2}{[\nu]}\rho^2(s). \tag{2.4}$$

Here we have set $T_0 = \tau_{[\nu]}(0) = \eta_1(0) = \cdots = \eta_{[\nu]}(0) = 0$. Equation (2.2) follows from the independence of $\eta_i(t),\ \eta_j(t),\ i \neq j$, (2.3) follows since $\eta_1(t)$ has marginal $N(0,1)$ distribution, and (2.4) follows since, for (X, Y) bivariate normal with zero mean, unit variance and correlation coefficient ρ, and $Z_1,\ Z_2 \sim NID(0,1)$, we have that $(X, Y) \stackrel{\mathcal{D}}{=} (Z_1, \rho Z_1 + \sqrt{1-\rho^2}Z_2)$ so that

$$\mathbb{E}(X^2Y^2) = \mathbb{E}(\rho^2 Z_1^4 + \rho\sqrt{1-\rho^2}Z_1^2 Z_2 + (1-\rho^2)Z_1^2 Z_2^2) = 2\rho^2 + 1. \tag{2.5}$$

Assumption 2.1. *Put $Z(s) = \rho^2(s) - \rho^2(s+1)$ and assume that $Z(s-1) - Z(s) \geq 0$ for $s \in \mathbb{N}$, which is equivalent to $\rho^2(s)$ being convex on the integers. We also require $Z(s) \geq 0$.*

Theorem 2.1. *Under Assumption 2.1 there exists a process $\{\tau_\nu(t)\}$, $t \in \mathbb{N}$, with $\nu > 0$ (not necessarily integer) and marginal $\Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ distribution such that $\mathbb{C}ov(\tau_\nu(t),\ \tau_\nu(t+s)) = \frac{2}{\nu}\rho^2(s)$ for $s \in \mathbb{N}$, parallel to (2.4).*

Putting $T_t = \sum_{i=1}^t \tau_\nu(i)$ and choosing $\rho(s)$ such that $\{\tau_\nu(t)\}$ is LRD results in a discrete LRD VG process with non-integer ν parameter. We now prove Theorem 2.1, with the main steps set out in Lemmas 2.1, 2.2 and 2.3.

First we construct two τ_ν's such that they have covariance of the form given by (2.4). Fix $n \in \mathbb{N}$ and set $\iota = \frac{\nu - [\nu]}{2}$ and $Y_{i,*}^n \stackrel{\mathcal{D}}{=} Y_{i,\circ}^n \stackrel{\mathcal{D}}{=} \Gamma(\iota\frac{1}{n}, \frac{1}{2}),\ i = 1, \cdots, n$, all independent and

independent of the η's. Then set

$$V_*^n := \sum_{i=1}^{n} Y_{i,*}^n \stackrel{\mathcal{D}}{=} \Gamma(\iota, \frac{1}{2}) \quad \text{and} \quad V_\circ^n := \sum_{i=1}^{k} Y_{i,*}^n + \sum_{i=1}^{n-k} Y_{i,\circ}^n \stackrel{\mathcal{D}}{=} \Gamma(\iota, \frac{1}{2}) \tag{2.6}$$

so there is an overlap of k of the $Y_{i,*}^n$'s between V_*^n and $V_\circ^n$. Now set $\tau_\nu^n(t) = (\eta_1^2(t) + \cdots + \eta_{[\nu]}^2(t) + V_*^n)/\nu$ and $\tau_\nu^n(t+s) = (\eta_1^2(t+s) + \cdots + \eta_{[\nu]}^2(t+s) + V_\circ^n)/\nu$, both $\Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ random variables.

Lemma 2.1. *For any fixed $t \in \mathbb{N}$ and fixed single temporal lag $s \in \mathbb{N}$, $\tau_\nu^n(t)$ and $\tau_\nu^n(t+s)$ as defined above with $k = [n\rho^2(s)]$ result in $\mathbb{C}ov(\tau_\nu^n(t), \tau_\nu^n(t+s)) \to \frac{2}{\nu}\rho^2(s)$ as $n \to \infty$, with error bounded by $\frac{4\iota}{n\nu^2}$ independent of t and s. (In this case we do not need Assumption 2.1.)*

Proof. From (2.6), $\mathbb{C}ov(V_*^n, V_\circ^n) = \mathbb{V}ar(Y_{1,*}^n + \cdots + Y_{k,*}^n) = 4\iota\frac{k}{n}$ so that

$$\begin{aligned}\mathbb{C}ov(\tau_\nu^n(t), \tau_\nu^n(t+s)) &= \frac{1}{\nu^2}\mathbb{C}ov(\sum_{i=1}^{[\nu]} \eta_i^2(t),\ \sum_{i=1}^{[\nu]} \eta_i^2(t+s)) + \frac{1}{\nu^2}\mathbb{C}ov(V_*^n, V_\circ^n) \\ &= \frac{2}{\nu}\rho^2(s) + \frac{4\iota}{\nu^2}(\frac{[n\rho^2(s)] - n\rho^2(s)}{n}).\end{aligned}$$

□

The above shows how we construct a process $\{\tau_\nu\}$ with the desired correlation structure at lag s. Constructing a stationary process $\{\tau_\nu\}$ that has the correct correlation at all lags is more involved. We now give a procedure to this end.

Again fix $n \in \mathbb{N}$, set $\iota = \frac{\nu - [\nu]}{2}$, $Y_{i,j}^n \stackrel{\mathcal{D}}{=} \Gamma(\iota\frac{1}{n}, \frac{1}{2})$, $i = 1, \cdots, [n\rho^2(1)]$ for $j = 0$ and $i = 1, \cdots, n - [n\rho^2(1)]$ for $j = 1, 2, \cdots$ with all $Y_{i,j}^n$'s mutually independent. Then set

$$V_t^n = \sum_{i=1}^{[n\rho^2(t)]} Y_{i,0}^n + \sum_{j=1}^{t}(\sum_{i=1}^{[n\rho^2(t-j)] - [n\rho^2(t-j+1)]} Y_{i,j}^n) \stackrel{\mathcal{D}}{=} \Gamma(\iota, \frac{1}{2}) \tag{2.7}$$

for $t = 1, 2, \cdots$ (assuming $Z(s) \geq 0$, setting $V_0^n = 0$ and noting that $\rho^2(0) = 1$), and

$$\tau_\nu^n(t) = (\eta_1^2(t) + \cdots + \eta_{[\nu]}^2(t) + V_t^n)/\nu \stackrel{\mathcal{D}}{=} \Gamma(\frac{\nu}{2}, \frac{\nu}{2}),$$

$$T_t^n = \sum_{i=1}^{t} \tau_\nu^n(i). \tag{2.8}$$

Lemma 2.2. *Under Assumption 2.1, for any time $t \in \mathbb{N}$ and temporal lag $s \in \mathbb{N}$, and $\tau_\nu^n(t)$ and $\tau_\nu^n(t+s)$ as defined above, $\mathbb{C}ov(\tau_\nu^n(t), \tau_\nu^n(t+s)) \to \frac{2}{\nu}\rho^2(s)$ as $n \to \infty$.*

Proof. Consider any V_t^n and V_{t+s}^n for $t, s \in \mathbb{N}$. Then for any j such that $1 \leq j \leq t$, V_t^n contains the first $[n\rho^2(t-j)] - [n\rho^2(t-j+1)]$ of the $Y_{i,j}^n$'s, while V_{t+s}^n contains the first $[n\rho^2(t+s-j)] - [n\rho^2(t+s-j+1)]$ of the same $Y_{i,j}^n$'s. But $s > 0$ so by Assumption 2.1, $[n\rho^2(t+s-j)] - [n\rho^2(t+s-j+1)] \leq [n\rho^2(t-j)] - [n\rho^2(t-j+1)]$ for large n, so the overlap of $Y_{i,j}^n$'s between V_t^n and V_{t+s}^n is simply $[n\rho^2(t+s-j)] - [n\rho^2(t+s-j+1)]$. For $j > t$, V_t^n contains none of the $Y_{i,j}^n$'s, while for $j = 0$, V_t^n contains the first $[n\rho^2(t)]$ of the $Y_{i,0}^n$'s, while V_{t+s}^n contains the first $[n\rho^2(t+s)]$ of the $Y_{i,0}^n$'s. Hence the total number of overlapping $Y_{i,j}^n$'s between V_t^n and V_{t+s}^n is

$$\sum_{j=1}^{t}([n\rho^2(t+s-j)] - [n\rho^2(t+s-j+1)]) + [n\rho^2(t+s)] = [n\rho^2(s)].$$

But from Lemma 2.1 this delivers the correct correlation. □

Lemma 2.3. *Under Assumption 2.1, $\{V_t^n\}$ for $t \in \mathbb{N}$ as defined by (2.7) converges weakly to a well defined stochastic process $\{V_t\}$ as $n \to \infty$.*

Proof. Fix $p \in \mathbb{N}$ and let $a_1, \cdots, a_p \in \mathbb{R}$. To ease notation set $f(t) = [n\rho^2(t)]$ and $g(t) = [n\rho^2(t-1)] - [n\rho^2(t)]$. Then starting from (2.7), one can show that $\sum_{t=1}^{p} a_t V_t^n$ is given by

$$\begin{aligned}(2.9)\qquad &\sum_{i=1}^{f(p)}(\sum_{t=1}^{p} a_t)Y_{i,0}^n + \sum_{j=1}^{p-1}\sum_{i=f(j+1)+1}^{f(j)}(\sum_{t=1}^{j} a_t)Y_{i,0}^n \\ &+ \sum_{k=1}^{p}\sum_{i=1}^{g(k)}(\sum_{t=p-k+1}^{p} a_t)Y_{i,p-k+1}^n \\ &+ \sum_{k=1}^{p-1}\sum_{j=1}^{p-k}\sum_{i=g(j+1)+1}^{g(j)}(\sum_{t=k}^{k+j-1} a_t)Y_{i,k}^n.\end{aligned}$$

Now each $Y_{i,j}^n$ is iid $\Gamma(\iota\frac{1}{n}, \frac{1}{2})$ distributed so the characteristic function of $(V_1^n, \cdots, V_p^n)$ is given by

$$\begin{aligned}\phi_p^n(a_1, \cdots, a_p) =&(1 - 2i(\sum_{t=1}^{p} a_t))^{-\iota\frac{f(p)}{n}} \prod_{j=1}^{p-1}(1 - 2i(\sum_{t=1}^{j} a_t))^{-\iota\frac{g(j+1)}{n}} \\ &\times \prod_{k=1}^{p}(1 - 2i(\sum_{t=p-k+1}^{p} a_t))^{-\iota\frac{g(k)}{n}} \\ &\times \prod_{k=1}^{p-1}\prod_{j=1}^{p-k}(1 - 2i(\sum_{t=k}^{k+j-1} a_t))^{-\iota\frac{g(j)-g(j+1)}{n}}.\end{aligned}$$

As $n \to \infty$ we have $\frac{f(t)}{n} \to \rho^2(t)$ and $\frac{g(t)}{n} \to \rho^2(t-1) - \rho^2(t) = Z(t-1)$ so that $\phi_p^n(a_1, \cdots, a_p)$ converges to a function $\phi_p(a_1, \cdots, a_p)$ given by

$$\begin{aligned}(2.10)\qquad \phi_p(a_1, \cdots, a_p) =&(1 - 2i(\sum_{t=1}^{p} a_t))^{-\iota\rho^2(p)} \prod_{j=1}^{p-1}(1 - 2i(\sum_{t=1}^{j} a_t))^{-\iota Z(j)} \\ &\times \prod_{k=1}^{p}(1 - 2i(\sum_{t=p-k+1}^{p} a_t))^{-\iota Z(k-1)} \\ &\times \prod_{k=1}^{p-1}\prod_{j=1}^{p-k}(1 - 2i(\sum_{t=k}^{k+j-1} a_t))^{-\iota(Z(j-1)-Z(j))}.\end{aligned}$$

Hence $\phi_p^n(a_1, \cdots, a_p)$ converges pointwise to $\phi_p(a_1, \cdots, a_p)$ and $\phi_p(a_1, \cdots, a_p)$ is continuous about the origin, which from Billingsley (1968) Theorem 7.6 (see also the second paragraph on p. 30) implies weak convergence. One can also verify that V_t has the

gamma distribution $\Gamma(\iota, \frac{1}{2})$ for $t = 1, 2, \cdots$ by considering $\phi_t(a_1, \cdots, a_t)$ and choosing $a_1 = \cdots = a_{t-1} = 0$. □

If we choose $\rho(s) = (1 + \omega|s|^\alpha)^{(H-1)/\alpha}$ for $\omega > 0,\ 0 < \alpha \leq 2$ and $\frac{1}{2} < H < 1$, i.e. an autocorrelation function from the so-called Cauchy family detailed in Gneiting (2000), our construction will lead to a LRD VG model.

Now working with the limit processes $\{T_t\} = \{T_t^\infty\}$, $\{\tau_t\} = \{\tau_t^\infty\}$ and $\{V_t\} = \{V_t^\infty\}$, from (2.8) we can take our activity time process $\{T_t\}$ as the sum of two independent parts:

$$T_t = \frac{1}{\nu}\sum_{j=1}^{[\nu]}\sum_{i=1}^{t}\eta_j^2(i) + \frac{1}{\nu}\sum_{i=1}^{t}V_i = A_t + B_t \text{ say.} \tag{2.11}$$

We now show that $\mathbb{V}\text{ar}(A_k)$ and $\mathbb{V}\text{ar}(B_k)$ are both $\mathcal{O}(k^{2H})$, and that $\zeta_k(t) = \frac{1}{k^H}(A_{[kt]} - \mathbb{E}A_{[kt]})$ for $t \in [0, 1]$ converges weakly as $k \to \infty$ to a self similar process with parameter H. We also give a proof in Theorem 2.4 below that $\frac{1}{k^H}(B_k - \mathbb{E}B_k)$ converges in probability to 0, which is enough to demonstrate that our discrete time $\{T_t\}$ process (2.11) has asymptotically a self similar limit.

2.1.2. Convergence to a self similar limit. First from Taqqu (1975), Theorem and Proposition 6.1, we take Theorem 2.2:

Theorem 2.2. *For Y_i a stationary Gaussian sequence with $\mathbb{E}Y_i = 0$ and $\mathbb{V}ar(Y_i) = 1$, such that $\rho_Y(\tau) \sim \tau^{H-1}L(\tau)$ as $\tau \to \infty$ with $\frac{1}{2} < H < 1$ and L slowly varying,*

$$Z_k(t) = \frac{1}{k^H}\sum_{i=1}^{[kt]}(Y_i^2 - 1)$$

converges weakly as $k \to \infty$ to a process $R(t)$ which has properties,

(1) *$R(t)$ has strictly stationary increments.*

(2) *$R(t)$ is H self similar.*

(3) *$\mathbb{E}R(t) = 0$ and $\mathbb{E}|R(t)|^\gamma < \infty$ for $\gamma \leq \frac{1}{H}$.*

(4) $R(t)$ *is separable and almost surely continuous.*

Further, the characteristic function of $R(t)$ *admits the representation*

$$\phi(u) = \exp\{\frac{1}{2}\sum_{k=2}^{\infty}\frac{(2iu)^k}{(k+1)!}S_k\} \tag{2.12}$$

for $S_k = \int_0^1\cdots\int_0^1 \frac{\mathrm{d}x_1\cdots\mathrm{d}x_k}{(|x_1-x_2||x_2-x_3|\cdots|x_{k-1}-x_k||x_k-x_1|)^{1-H}}$, *which is valid for small values of* $|u|$.

Theorem 2.3. *When* $\rho(s)$ *is given by a member of the Cauchy family,* $\mathbb{V}ar(A_k)$ *and* $\mathbb{V}ar(B_k)$ *are both* $\mathcal{O}(k^{2H})$ *and the standardised process* $\zeta_k(t) = \frac{1}{k^H}(A_{[kt]} - \mathbb{E}A_{[kt]})$ *for* $t \in [0,1]$ *converges weakly as* $k \to \infty$ *to a self similar process with parameter* H.

Proof. We have that $\mathbb{V}\mathrm{ar}(A_k) = \frac{[\nu]}{\nu^2}\mathbb{V}\mathrm{ar}(\sum_{i=1}^k \eta_1^2(i))$ and from (2.5),

$$\mathbb{V}\mathrm{ar}(\sum_{i=1}^k \eta_1^2(i)) = \sum_{s=1}^k\sum_{s^*=1}^k \mathbb{C}\mathrm{ov}(\eta_1^2(s),\eta_1^2(s^*)) = 2\sum_{s=1}^k\sum_{s^*=1}^k \rho^2(s-s^*).$$

Similarly, $\mathbb{V}\mathrm{ar}(B_k) = \frac{1}{\nu^2}\mathbb{V}\mathrm{ar}(\sum_{i=1}^k V_i)$ with

$$\mathbb{V}\mathrm{ar}(\sum_{i=1}^k V_i) = \sum_{s=1}^k\sum_{s^*=1}^k \mathbb{C}\mathrm{ov}(V_s,V_{s^*}) = 4\iota\sum_{s=1}^k\sum_{s^*=1}^k \rho^2(s-s^*).$$

Now

$$\begin{aligned}\sum_{s=1}^k\sum_{s^*=1}^k \rho^2(s-s^*) &= \sum_{s=1}^k\sum_{s^*=1}^k (1+\omega|s-s^*|^\alpha)^{2(H-1)/\alpha}\\ &\sim \int_0^k\int_0^k \frac{\mathrm{d}s\mathrm{d}s^*}{(1+\omega|s-s^*|^\alpha)^{(2-2H)/\alpha}}\\ &= k^{2H}\int_0^1\int_0^1 \frac{\mathrm{d}u\mathrm{d}v}{(\frac{1}{k^\alpha}+\omega|u-v|^\alpha)^{(2-2H)/\alpha}} \qquad (2.13)\\ &\to k^{2H}\, c(H) \quad \text{as} \quad k\to\infty \qquad (2.14)\end{aligned}$$

where (2.13) follows by taking $u = \frac{s}{k},\ v = \frac{s^*}{k}$, and

$$\begin{aligned} c(H) &= \omega^{(2H-2)/\alpha} \int_0^1 \int_0^1 \frac{\mathrm{d}y\mathrm{d}x}{|x-y|^{2-2H}} \\ &= 2\omega^{(2H-2)/\alpha} \int_0^1 \int_0^x \frac{\mathrm{d}y\mathrm{d}x}{(x-y)^{2-2H}} \\ &= \frac{2\omega^{(2H-2)/\alpha}}{2H-1} \int_0^1 x^{2H-1}\mathrm{d}x \\ &= \frac{\omega^{(2H-2)/\alpha}}{(2H-1)(H)} \ < \infty \ \text{ for } \frac{1}{2} < H < 1. \end{aligned}$$

Hence $\mathbb{V}\mathrm{ar}(A_k)$ and $\mathbb{V}\mathrm{ar}(B_k)$ are both $\mathcal{O}(k^{2H})$.

A direct result of Theorem 2.2 is then that

$$\frac{1}{k^H} \sum_{s=1}^{[kt]} (\eta_1^2(s) - 1) \Rightarrow R(t) \ \text{ as } k \to \infty$$

for $R(t)$ the so-called (H self similar with strictly stationary increments) Rosenblatt process, where $H \in (\frac{1}{2},\ 1)$. Here '$\Rightarrow$' denotes the weak convergence of one stochastic process to another, which implies the convergence of finite dimensional distributions. Consequently we have that as $k \to \infty$,

$$\begin{aligned} \zeta_k(t) &= \frac{1}{k^H}(A_{[kt]} - \mathbb{E}A_{[kt]}) \\ &= \frac{1}{\nu} \sum_{j=1}^{[\nu]} (\frac{1}{k^H} \sum_{i=1}^{[kt]} (\eta_j^2(i) - 1)) \Rightarrow \frac{1}{\nu} \sum_{j=1}^{[\nu]} R_j(t) \end{aligned} \tag{2.15}$$

for each $R_j(t)$ an independent copy of the Rosenblatt process. □

Theorem 2.4. *When $\rho(s)$ is given by a member of the Cauchy family, the sequence $\zeta_k = \frac{1}{k^H} \sum_{i=1}^k (V_i - \mathbb{E}V_i) = \frac{\nu}{k^H}(B_k - \mathbb{E}B_k)$ for $k = 1, 2, \cdots$ and $\frac{1}{2} < H < 1$ converges in distribution, and therefore probability, to 0 as $k \to \infty$.*

Proof. We give the proof taking $\rho(s)$ as any member of the Cauchy family which satisfies Assumption 2.1, in order to use (2.10) from Lemma 2.3.

First consider the cf $\phi_p^*(a_1, \cdots, a_p)$ of $(V_1 - \mathbb{E}V_1, \cdots, V_p - \mathbb{E}V_p)$. By replacing each $Y_{i,j}^n$ in (2.9) with $Y_{i,j}^n - \mathbb{E}Y_{i,j}^n = Y_{i,j}^n - \frac{2\iota}{n}$, one can show that $\phi_p^*(a_1, \cdots, a_p)$ is given by (2.10), but with each expression of the form $(1-2ix)^{-\iota y}$ replaced by $(1-2ix)^{-\iota y}e^{-2ix\iota y}$. Now for $a \in \mathbb{R}$, the cf of ζ_k is given by

$$\varphi_k(a) = \mathbb{E}\left\{\exp\left(\frac{ia}{k^H}\sum_{j=1}^{k}(V_j - 2\iota)\right)\right\} = \phi_k^*(\frac{a}{k^H}, \cdots, \frac{a}{k^H}).$$

From (2.10), $\varphi_k(a)$ is a product comprising the four factors

$$(1-2iak^{1-H})^{-\iota\rho^2(k)} \times e^{-2iak^{1-H}\iota\rho^2(k)} \tag{2.16}$$

$$\prod_{j=1}^{k-1}((1-2ijak^{-H})^{-\iota Z(j)} \times e^{-2ijak^{-H}\iota Z(j)}) \tag{2.17}$$

$$\prod_{j=1}^{k}((1-2ijak^{-H})^{-\iota Z(j-1)} \times e^{-2ijak^{-H}\iota Z(j-1)}) \tag{2.18}$$

$$\prod_{m=1}^{k-1}\prod_{j=1}^{k-m}((1-2ijak^{-H})^{-\iota(Z(j-1)-Z(j))} \times e^{-2ijak^{-H}\iota(Z(j-1)-Z(j))}). \tag{2.19}$$

We shall use Markov's inequality to show that the random variables whose cfs are given by (2.16), (2.17) and (2.18) converge in probability to 0 as $k \to \infty$, and show directly that the moment generating function of the random variable with cf given by (2.19) converges to 1 as $k \to \infty$, thus establishing the result.

First note that for a non-negative random variable Y_k say, Markov's inequality states that for any fixed $\epsilon > 0$,

$$P(Y_k > \epsilon) \leq \frac{\mathbb{E}Y_k}{\epsilon}$$

so that if $\mathbb{E}Y_k \to 0$ then $Y_k \xrightarrow{\mathcal{P}} 0$ and therefore $(Y_k - \mathbb{E}Y_k) \xrightarrow{\mathcal{P}} 0$, where '$\xrightarrow{\mathcal{P}}$' denotes convergence in probability. Note also that each of (2.16), (2.17) and (2.18) represent the cf of a sum (mean-corrected) of independent and non-negative gamma random variables, so that if we show that the mean of each such sum (before mean correction) converges to 0 we are done.

Now (2.16) before mean-correction is the cf of a $\Gamma(\iota\rho^2(k), \frac{1}{2k^{1-H}})$ random variable with mean of $2k^{1-H}\iota\rho^2(k)$. For $\rho(s) = (1+\omega|s|^\alpha)^{(H-1)/\alpha}$, $\rho^2(k) = \mathcal{O}(k^{2H-2})$ so that $2k^{1-H}\iota\rho^2(k) = \mathcal{O}(k^{H-1}) \to 0$ as $k \to \infty$. Similarly the mean of the sum of gamma random variables with cf (2.17) is given by $2k^{-H}\iota\sum_{j=1}^{k-1} jZ(j)$. Here we have

$$0 \le k^{-H}\sum_{j=1}^{k-1} jZ(j) = k^{-H}((\sum_{j=1}^{k-1}\rho^2(j)) - (k-1)\rho^2(k))$$

but both $k^{-H}(k-1)\rho^2(k)$ and $k^{-H}\sum_{j=1}^{k-1}\rho^2(j)$ are $\mathcal{O}(k^{H-1}) \to 0$ as $k \to \infty$ so that $2k^{-H}\iota\sum_{j=1}^{k-1} jZ(j) \to 0$. A similar result holds for (2.18).

Finally consider (2.19). In this case the mean is $\mathcal{O}(k^{1-H}) \to \infty$ and so we cannot use Markov's inequality. Instead change the order of multiplication to write the *mgf* of the *negative* of the random variable with cf (2.19) as

$$M_k(a) = \exp\left(\iota\sum_{j=1}^{k-1}(Z(j-1)-Z(j))(k-j)(2jak^{-H} - \log(1+2jak^{-H}))\right). \tag{2.20}$$

Working with the mgf instead of the cf simplifies matters, since $M_k(a)$ is well defined for all $a \ge 0$, and from Mukherjea, Rao & Suen (2006), Theorem 2, pointwise convergence of $M_k(a)$ in some fixed interval (b,d), $0 < b < d < \infty$, as $k \to \infty$, to the mgf $M(a)$ of some random variable implies weak convergence to the associated limit distribution. Thus if $M_k(a)$ converges to 1, the mgf of 0, we are done. Now $x \ge x - \log(1+x) \ge 0$ and $x^2 \ge x - \log(1+x) \ge 0$ for $x \ge 0$, so that

$$\begin{aligned}
0 &\le \sum_{j=1}^{k-1}(Z(j-1)-Z(j))(k-j)(2jak^{-H} - \log(1+2jak^{-H})) \\
&\le \sum_{j=1}^{[k^H]-1}(Z(j-1)-Z(j))(k-j)(2jak^{-H})^2 \\
&\quad + \sum_{j=[k^H]}^{k-1}(Z(j-1)-Z(j))(k-j)2jak^{-H} \\
&\le c_1 k^{-2H}\sum_{j=1}^{[k^H]-1} j^{2H-2}(k-j) + c_2 k^{-H}\sum_{j=[k^H]}^{k-1} j^{2H-3}(k-j) \qquad (2.21)
\end{aligned}$$

for c_1, c_2 constants, since $Z(j-1) - Z(j) = \mathcal{O}(j^{2H-4})$ by repeated application of the mean value theorem, using Assumption 1 and the explicit form of $\rho^2(s)$. But

$$k^{-2H} \int_1^{k^H} j^{2H-2}(k-j)\mathrm{d}j = \frac{k^{1-3H+2H^2} - k^{1-2H}}{2H-1} - \frac{k^{2H(H-1)} - k^{-2H}}{2H}$$

converges to 0 as $k \to \infty$ since each exponent of k is negative for $\frac{1}{2} < H < 1$, and

$$k^{-H} \int_{k^H}^{k} j^{2H-3}(k-j)\mathrm{d}j = \frac{k^{H-1} - k^{1-3H+2H^2}}{2H-2} - \frac{k^{H-1} - k^{2H(H-1)}}{2H-1}$$

converges to 0 as $k \to \infty$ so that (2.21) converges to 0 and (2.20) converges to 1. □

2.1.3. Relaxing the assumptions on $Z(s)$. Recall that for $Z(s) = \rho^2(s) - \rho^2(s+1)$ we require $Z(s) \geq 0$ and $Z(s)$ decreasing with $s \in \mathbb{N}$ according to Assumption 2.1. It is clear that all members of the Cauchy family satisfy the first property, but the same is not true of the second. For example $\alpha = \omega = 2$ satisfies the second property for any $\frac{1}{2} < H < 1$, whereas $\alpha = 2,\ \omega = 1$ for $0.648 < H < 1$ does not. In the latter case the acf value for $\{V_t\}$ at lag 1 will be $1 - \rho^2(1) + \rho^2(2)$ instead of the larger $\rho^2(1)$, but acf values at larger lags will be unaffected (at lags greater than 1, the requirement on $Z(s)$ is satisfied if $1 - \alpha + \omega|s|^\alpha(3-2H) \geq 0$, which for any given $\omega > 0,\ 0 < \alpha \leq 2$ and $\frac{1}{2} < H < 1$ will be the case for sufficiently large values of s). Therefore we briefly discuss how our results are affected when the second property (monotonicity) fails to hold for the first few lags.

As touched on above, if $Z(s)$ does not decrease with s for the first m lags say, then Lemma 2.2 will fail for the first m lags, and the first m acf values will be lower than those given by $\rho^2(\cdot)$ (acf values at lags greater than m will be unaffected). The main result of Section 2.1 was to show the convergence of our add-on processes $\{V_t^n\}$ to $\{V_t\}$, and to further show convergence of $\{V_t\}$ (appropriately normed) to zero, and both these convergence results are unaffected by relaxing the assumption that $Z(s)$ decreases with s.

In Lemma 2.3 we undertook to partition $\sum_{t=1}^{p} a_t V_t^n$ into groups of iid $Y_{i,j}^n$'s with the same coefficients (some sum of a_t's) in order to compute the characteristic function of $\{V_t^n\}$,

and therefore prove weak convergence to $\{V_t\}$. Now for each t, and ignoring rounding issues associated with taking the integer part, V_t^n from (2.7) is given by

$$V_t^n = \sum_{i=1}^{[n\rho^2(t)]} Y_{i,0}^n + \sum_{j=1}^{t}(\sum_{i=1}^{nZ(t-j)} Y_{i,j}^n) \tag{2.22}$$

so that the first $nZ(t-j)$ of the $Y_{i,j}^n$'s for $j = 1, \cdots, t$ are included in V_t^n. That is, the relative sizes of $Z(0), Z(1), \cdots$ determine which of the $Y_{i,j}^n$'s are included in V_t^n for each t. This is important in collecting the $Y_{i,j}^n$'s into groups with the same coefficients, and therefore computing the characteristic function of $\{V_t^n\}$.

To aid understanding we write expressions for $a_t V_t^{*,n}$ for $t = 1, 2, 3, 4, p$ below, where we set $V_t^{*,n} = V_t^n - \sum_{i=1}^{[n\rho^2(t)]} Y_{i,0}^n$. We subtract the $Y_{i,0}^n$ terms to simplify the expressions – we do not need to consider these terms since they are unaffected by relaxing the assumption that $Z(s)$ decreases with s (for the Cauchy family we always have that $\rho^2(s) > \rho^2(s+1)$).

$$\begin{aligned}
a_1 V_1^{*,n} &= a_1 \sum_{i=1}^{nZ(0)} Y_{i,1}^n \\
a_2 V_2^{*,n} &= a_2 \sum_{i=1}^{nZ(1)} Y_{i,1}^n + a_2 \sum_{i=1}^{nZ(0)} Y_{i,2}^n \\
a_3 V_3^{*,n} &= a_3 \sum_{i=1}^{nZ(2)} Y_{i,1}^n + a_3 \sum_{i=1}^{nZ(1)} Y_{i,2}^n + a_3 \sum_{i=1}^{nZ(0)} Y_{i,3}^n \\
a_4 V_4^{*,n} &= a_4 \sum_{i=1}^{nZ(3)} Y_{i,1}^n + a_4 \sum_{i=1}^{nZ(2)} Y_{i,2}^n + a_4 \sum_{i=1}^{nZ(1)} Y_{i,3}^n + a_4 \sum_{i=1}^{nZ(0)} Y_{i,4}^n \\
&\vdots \qquad\qquad\qquad \vdots \\
a_p V_p^{*,n} &= a_p \sum_{i=1}^{nZ(p-1)} Y_{i,1}^n + a_p \sum_{i=1}^{nZ(p-2)} Y_{i,2}^n + a_p \sum_{i=1}^{nZ(p-3)} Y_{i,3}^n + \cdots + a_p \sum_{i=1}^{nZ(0)} Y_{i,p}^n.
\end{aligned} \tag{2.23}$$

Hence for $Z_0^{(s)} = \max(Z(0), Z(1), \cdots, Z(s))$ with $Z_1^{(s)}$ the next biggest through to $Z_s^{(s)} = \min(Z(0), Z(1), \cdots, Z(s))$, $Z_{p-1}^{(p-1)}$ of the $Y_{i,1}^n$'s will have coefficient $\sum_{t=1}^{p} a_t$, $Z_{p-2}^{(p-1)} - Z_{p-1}^{(p-1)}$ of the $Y_{i,1}^n$'s will have as coefficient all bar one of the a_t's, and $Z_0^{(p-1)} - Z_1^{(p-1)}$ of the $Y_{i,1}^n$'s will have as coefficient one of the a_t's. Similar results hold for the $Y_{i,2}^n$'s, $Y_{i,3}^n$'s, $\cdots$.

Next we re-arrange (2.23) in terms of the $Y_{i,j}^n$'s:

(2.24)

$$\begin{aligned}
\sum_{t=1}^{p} a_t V_t^{*,n} = & \sum_{i=1}^{nZ_{p-1}^{(p-1)}} (\sum_{t=1}^{p} a_t^{(1)}) Y_{i,1}^n + \sum_{i=nZ_{p-1}^{(p-1)}+1}^{nZ_{p-2}^{(p-1)}} (\sum_{t=1}^{p-1} a_t^{(1)}) Y_{i,1}^n + \cdots + \sum_{i=nZ_1^{(p-1)}+1}^{nZ_0^{(p-1)}} a_1^{(1)} Y_{i,1}^n \\
& + \sum_{i=1}^{nZ_{p-2}^{(p-2)}} (\sum_{t=2}^{p} a_t^{(2)}) Y_{i,2}^n + \sum_{i=nZ_{p-2}^{(p-2)}+1}^{nZ_{p-3}^{(p-2)}} (\sum_{t=2}^{p-1} a_t^{(2)}) Y_{i,2}^n + \cdots + \sum_{i=nZ_1^{(p-2)}+1}^{nZ_0^{(p-2)}} a_2^{(2)} Y_{i,2}^n \\
& + \sum_{i=1}^{nZ_{p-3}^{(p-3)}} (\sum_{t=3}^{p} a_t^{(3)}) Y_{i,3}^n + \sum_{i=nZ_{p-3}^{(p-3)}+1}^{nZ_{p-4}^{(p-3)}} (\sum_{t=3}^{p-1} a_t^{(3)}) Y_{i,3}^n + \cdots + \sum_{i=nZ_1^{(p-3)}+1}^{nZ_0^{(p-3)}} a_3^{(3)} Y_{i,3}^n \\
& \vdots \qquad\qquad\qquad \vdots \\
& + \sum_{i=1}^{nZ_1^{(1)}} (\sum_{t=p-1}^{p} a_t^{(p-1)}) Y_{i,p-1}^n + \sum_{i=nZ_1^{(1)}+1}^{nZ_0^{(1)}} a_{p-1}^{(p-1)} Y_{i,p-1}^n \\
& + \sum_{i=1}^{nZ_0^{(0)}} a_p^{(p)} Y_{i,p}^n.
\end{aligned}$$

Here we define $a_s^{(s)}$ to be the first member of the re-ordered set $\{a_s, \cdots, a_p\}$, where the $\{a_s, \cdots, a_p\}$ are ordered according to the decreasing size of the $\{Z(0), \cdots, Z(p-s)\}$, with $a_{s+1}^{(s)}$ the second member of the re-ordered set and so on. So for example if $Z(2)$ is the largest of $\{Z(0), \cdots, Z(p-3)\}$, followed by $Z(4)$, then the decreasing size ordering of $\{Z(0), \cdots, Z(p-3)\}$ begins $\{Z(2), Z(4), \cdots\}$ and the ordering of $\{a_3, \cdots, a_p\}$ begins $\{a_5, a_7, \cdots\}$, so that $a_2^{(2)} = a_5$ and $a_3^{(2)} = a_7$.

If Assumption 2.1 holds, then $Z_j^{(s)} = Z(j)$ for each s, but if it does not hold then the order of the largest few $Z_j^{(s)}$'s may change. With this in mind, using (2.24) we can modify

(2.9) as follows, where again $f(t) = [n\rho^2(t)]$:

$$\sum_{t=1}^{p} a_t V_t^n = \sum_{i=1}^{f(p)} (\sum_{t=1}^{p} a_t) Y_{i,0}^n + \sum_{j=1}^{p-1} \sum_{i=f(j+1)+1}^{f(j)} (\sum_{t=1}^{j} a_t) Y_{i,0}^n$$
$$+ \sum_{k=1}^{p} \sum_{i=1}^{nZ_{k-1}^{(k-1)}} (\sum_{t=p-k+1}^{p} a_t^{(p-k+1)}) Y_{i,p-k+1}^n$$
$$+ \sum_{k=1}^{p-1} \sum_{j=1}^{p-k} \sum_{i=nZ_j^{(p-k)}+1}^{nZ_{j-1}^{(p-k)}} (\sum_{t=k}^{k+j-1} a_t^{(k)}) Y_{i,k}^n.$$

From here it is a simple matter to derive a new expression for the cf of $(V_1^n, \cdots, V_p^n)$ as $n \to \infty$ analogous to (2.10),

$$\phi_p^*(a_1, \cdots, a_p) = (1 - 2i(\sum_{t=1}^{p} a_t))^{-\iota\rho^2(p)} \prod_{j=1}^{p-1} (1 - 2i(\sum_{t=1}^{j} a_t))^{-\iota Z(j)}$$
$$\times \prod_{k=1}^{p} (1 - 2i(\sum_{t=p-k+1}^{p} a_t^{(p-k+1)}))^{-\iota Z_{k-1}^{(k-1)}}$$
$$\times \prod_{k=1}^{p-1} \prod_{j=1}^{p-k} (1 - 2i(\sum_{t=k}^{k+j-1} a_t^{(k)}))^{-\iota(Z_{j-1}^{(p-k)} - Z_j^{(p-k)})}. \tag{2.25}$$

Weak convergence of $\{V_t^n\}$ to a process $\{V_t\}$ with cf (2.25) follows from Billingsley (1968) Theorem 7.6. Again if Assumption 2.1 holds then $Z_j^{(s)} = Z(j)$ and $a_j^{(s)} = a_j$ for each s, and (2.25) reduces to (2.10).

Next consider Theorem 2.4 in light of our new cf. In Theorem 2.4 we take $a_t = \frac{a}{k^H}$ for each t, so that the distinction between $a_j^{(s)}$ and a_j is no longer relevant – only results that may be effected by the distinction between $Z(j)$ and $Z_j^{(s)}$ need checking. Theorem 2.4 showed that the random variables with cfs (2.16), (2.17), (2.18) and (2.19) all converge to zero. Working from our new cf (2.25) instead of (2.10), note that the analogues of (2.16) and (2.17) do not change so we need not consider them.

Next, the proof that the random variable with cf (2.18) converged to zero relied on asymptotic properties of $Z(j)$, that is, the proof relied on the behavior of $Z(j)$ as j gets

large. But as stated above, the requirement that $Z(s)$ decrease with s is satisfied if $1 - \alpha + \omega|s|^{\alpha}(3-2H) \geq 0$, which for any given $\omega > 0,\ 0 < \alpha \leq 2$ and $\frac{1}{2} < H < 1$ will be the case for sufficiently large values of s. Hence our arguments still hold, since for any $\omega > 0,\ 0 < \alpha \leq 2$ and $\frac{1}{2} < H < 1$ only a finite number, m say, of $Z(j)$ will be effected by relaxing the assumption that $Z(s)$ decrease with s, and $Z_j^{(s)}$ will equal $Z(j)$ for $j > m$.

Finally, consider the proof that the random variable with cf (2.19) converged to zero. Analogous to (2.20), we can write the mgf of the negative of the corresponding random variable when we relax the assumption on $Z(s)$ as

$$\exp\left(\iota \sum_{j=1}^{k-1}(2jak^{-H} - \log(1+2jak^{-H}))\sum_{l=1}^{k-j}(Z_{j-1}^{(k-l)} - Z_j^{(k-l)})\right)$$

Therefore, similarly to (2.21), using the fact that there exists some fixed m such that $Z_j^{(s)} = Z(j)$ for $j > m$,

$$\begin{aligned}
0 &\leq \sum_{j=1}^{k-1}(2jak^{-H} - \log(1+2jak^{-H}))\sum_{l=1}^{k-j}(Z_{j-1}^{(k-l)} - Z_j^{(k-l)}) \\
&\leq \sum_{j=1}^{m}(2jak^{-H})^2\sum_{l=1}^{k-j}(Z_{j-1}^{(k-l)} - Z_j^{(k-l)}) \\
&\qquad + \sum_{j=m+1}^{[k^H]-1}(Z(j-1) - Z(j))(k-j)(2jak^{-H})^2 \\
&\qquad\qquad + \sum_{j=[k^H]}^{k-1}(Z(j-1) - Z(j))(k-j)2jak^{-H} \\
&\leq c_0\sum_{j=1}^{m} j^2k^{-2H}(k-j) + c_1k^{-2H}\sum_{j=m+1}^{[k^H]-1} j^{2H-2}(k-j) + c_2k^{-H}\sum_{j=[k^H]}^{k-1} j^{2H-3}(k-j).
\end{aligned}$$

Now $\sum_{j=1}^{m} j^2k^{-2H}(k-j) < m^3k^{1-2H} \to 0$ as $k \to \infty$ so that once again the random variable with cf corresponding to (2.19) converges to zero.

Hence the activity time process generated from a member of the Cauchy family that does not satisfy Assumption 2.1 will still be LRD and have self similar limit, but the first few acf values will be lower than in the integer ν case.

2.1.4. Conclusion. We have constructed correlated $\{\tau_\nu(t)\}$ increments processes with marginal gamma distribution, such that $\{T_t\}$, appropriately normed, is asymptotically H self similar, being weakly convergent to the sum of $[\nu]$ independent Rosenblatt processes. This in turn leads to correlated $\{X_t\}$ log price increments processes with marginal VG distribution, in which both $\{\tau_\nu(t)\}$ and $\{X_t^2\}$ are long range dependent. This self similarity is a desirable property to endow $\{T_t\}$ with, since as previously discussed, there is growing evidence that historical financial data is LRD and driven by a self similar activity time process.

2.2. The inverse gamma case

In this section we construct a discrete activity time process $\{T_t\}$ which has inverse gamma increments, displays LRD, and has asymptotically a (continuous time) self similar limit. The results extend those of Heyde & Leonenko (2005), which dealt only with the integer ν case for $\nu = 2\delta$ and $\rho(s) = (1 + s^2)^{(H-1)/2}$.

2.2.1. Construction of the inverse gamma activity time process. To accord with Heyde & Leonenko (2005) and enable the direct application of results contained in Leonenko (1999), we make a re-parameterisation: consider again the $\{\tau_\nu(t)\}$ process constructed in Section 2.1, restrict ν to integer values, and set $\chi^2_\nu(t) = \frac{\nu}{2}\tau_\nu(t)$. Then the density of $\chi^2_\nu(t)$, which for each t is a scaled χ^2_ν random variable, is of the form

$$p_{\frac{\nu}{2}}(x) = \frac{x^{\frac{\nu}{2}-1}e^{-x}}{\Gamma(\frac{\nu}{2})} \tag{2.26}$$

and from the Hille-Hardy formula (Erdélyi, Magnus, Oberhettinger and Tricomi (1953)), the bivariate density of $(\chi^2_\nu(t), \chi^2_\nu(s))$ is given by (Leonenko (1999))

$$p_{\frac{\nu}{2}}(x,y;\gamma) = p_{\frac{\nu}{2}}(x)p_{\frac{\nu}{2}}(y)[1+\sum_{j=1}^{\infty}\gamma^j e_j(x)e_j(y)] \tag{2.27}$$

where

$$e_k(x) = L_k^{\frac{\nu}{2}-1}(x)\{\frac{k!\Gamma(\frac{\nu}{2})}{\Gamma(\frac{\nu}{2}+k)}\}^{1/2} \tag{2.28}$$

and

$$L_k^{\beta}(x) = \frac{1}{k!}x^{-\beta}e^x\frac{d^k}{dx^k}\{x^{\beta+k}e^{-x}\}$$

are generalised Laguerre polynomials of index β (Leonenko (1999)). If we define $e_0(x) \equiv 1$ we have (from Leonenko (1999) Section 2.1.10 for example) that $\{e_k(x)\}_{k=0}^{\infty}$ form a complete orthogonal system of functions in the Hilbert space $\mathbf{L_2}((0,\infty), p_{\frac{\nu}{2}}(x)dx)$ where $p_{\frac{\nu}{2}}(x)$ is as in (2.26). That is, for $G \in \mathbf{L_2}((0,\infty), p_{\frac{\nu}{2}}(x)dx)$, we have that

$$G(x) = \sum_{k=0}^{\infty} C_k e_k(x), \quad C_k = \int_0^{\infty} G(x)e_k(x)p_{\frac{\nu}{2}}(x)\mathrm{d}x, \tag{2.29}$$

$$\int_0^{\infty} e_k(x)e_m(x)p_{\frac{\nu}{2}}(x)\mathrm{d}x = \delta_{km}, \tag{2.30}$$

$$\sum_{k=0}^{\infty} C_k^2 = \int_0^{\infty} G^2(x)p_{\frac{\nu}{2}}(x)\mathrm{d}x < \infty \tag{2.31}$$

where δ_{km} is the Kronecker-delta function.

If we now choose $G(x) = (\frac{\nu}{2}-1)\frac{1}{x}$ then we have, for $\nu > 4$ so that (2.31) holds, that $G(\chi^2_\nu(t))$ is stationary with $R\Gamma(\frac{\nu}{2}, \frac{\nu}{2}-1)$ marginals. Further, using (2.29) we can expand $G(\chi^2_\nu(t))$ as

$$G(\chi^2_\nu(t)) = \sum_{k=0}^{\infty} C_k e_k(\chi^2_\nu(t)), \quad C_k = \int_0^{\infty} \frac{\nu/2-1}{x} e_k(x)p_{\frac{\nu}{2}}(x)\mathrm{d}x$$

where for example $e_0(x) = 1,\ C_0 = 1$ and

$$e_1(x) = \sqrt{\frac{2}{\nu}}(\frac{\nu}{2} - x), \quad C_1 = \sqrt{\frac{2}{\nu}}. \tag{2.32}$$

And since, for $k \geq 1$ or $m \geq 1$,

$$\begin{aligned}
\mathbb{C}\text{ov}(e_k(\chi^2_\nu(t)),\ e_m(\chi^2_\nu(t+\tau))) &= \int_0^\infty \int_0^\infty e_k(u)\ e_m(v) p_{\frac{\nu}{2}}(u,v;\rho^2(\tau))\mathrm{d}u\mathrm{d}v \\
&\quad - \mathbb{E}(e_k(\chi^2_\nu(t)))\ \mathbb{E}(e_m(\chi^2_\nu(t+\tau))) \\
&= \int_0^\infty \int_0^\infty e_k(u)\ e_m(v) p_{\frac{\nu}{2}}(u) p_{\frac{\nu}{2}}(v)[1 + \sum_{j=1}^\infty \rho^{2j}(\tau) e_j(u) e_j(v)]\mathrm{d}u\mathrm{d}v \\
&\quad - \mathbb{E}(e_k(\chi^2_\nu(t)))\ \mathbb{E}(e_m(\chi^2_\nu(t+\tau))) \\
&= \sum_{j=1}^\infty (\rho^{2j}(\tau) \int_0^\infty e_k(u) e_j(u) p_{\frac{\nu}{2}}(u))\mathrm{d}u \int_0^\infty e_m(v) e_j(v) p_{\frac{\nu}{2}}(v))\mathrm{d}v) \\
&= \delta_{mk}\ \rho^{2k}(\tau), \qquad (2.33)
\end{aligned}$$

while for $m = k = 0,\ e_k(x) = e_m(x) = 1$ so that $\mathbb{C}\text{ov}(e_k(\chi^2_\nu(t)),\ e_m(\chi^2_\nu(t+\tau))) = 0$, we have that

$$\mathbb{C}\text{ov}(G(\chi^2_\nu(t)),\ G(\chi^2_\nu(t+\tau))) = \sum_{k=1}^\infty C_k^2 \rho^{2k}(\tau). \qquad (2.34)$$

Also, since $\mathbb{V}\text{ar}(G(\chi^2_\nu(t))) = \frac{2}{\nu-4}$ for $\nu > 4$, we have that the autocorrelation function of $G(\chi^2_\nu(t))$ is given by

$$\rho_G(\tau) = \frac{\nu-4}{2} \sum_{k=1}^\infty C_k^2 \rho^{2k}(\tau).$$

Of course we may choose as $\rho(t)$ any consistent Gaussian autocorrelation function, and so we have a fair degree of flexibility in choosing the autocorrelation function of $G(\chi^2_\nu(t))$. Analogous with the gamma case, we take $T_t = \sum_{s=1}^t G(\chi^2_\nu(s))$ so that $\tau^*_\nu(t) = T_t - T_{t-1} = G(\chi^2_\nu(t)) \sim R\Gamma(\frac{\nu}{2}, \frac{\nu}{2} - 1)$.

Further, expanding $T_t - t$ as $T_t - t = \sum_{k=0}^\infty \sum_{s=1}^t C_k e_k(\chi^2_\nu(s)) - t = \sum_{k=1}^\infty \sum_{s=1}^t C_k e_k(\chi^2_\nu(s))$ leads to

$$T_t - t = \sum_{s=1}^t C_1 e_1(\chi^2_\nu(s)) + \sum_{k=2}^\infty \sum_{s=1}^t C_k e_k(\chi^2_\nu(s)) = \sum_{s=1}^t C_1 e_1(\chi^2_\nu(s)) + R_t$$

for $R_t = \sum_{k=2}^{\infty} \sum_{s=1}^{t} C_k e_k(\chi_\nu^2(s))$. Hence, for $\rho(s) = (1+\omega|s|^\alpha)^{(H-1)/\alpha}$, for $\omega > 0,\ 0 < \alpha \leq 2$ and $\frac{1}{2} < H < 1$, we have

$$\mathbb{V}\text{ar}(T_t - t) = \mathbb{V}\text{ar}(\sum_{s=1}^{t} C_1 e_1(\chi_\nu^2(s))) + \mathbb{V}\text{ar}(R_t) \tag{2.35}$$

$$= \sum_{s=1}^{t}\sum_{s^*=1}^{t} C_1^2\ \rho^2(|s-s^*|) + \mathbb{V}\text{ar}(R_t) \tag{2.36}$$

$$\sim C_1^2 c(H) t^{2H} + \mathbb{V}\text{ar}(R_t). \tag{2.37}$$

Here (2.35) and (2.36) follow from the orthogonality property given by Equation (2.33), and (2.37) follows from (2.14) for $c(H) = \frac{\omega^{(2H-2)/\alpha}}{(2H-1)(H)}$.

Now consider $R_t = \sum_{k=2}^{\infty} \sum_{s=1}^{t} C_k e_k(\chi_\nu^2(s))$,

$$\begin{aligned}
\mathbb{V}\text{ar}(R_t) &= \sum_{k=2}^{\infty}\sum_{s=1}^{t}\sum_{s^*=1}^{t} C_k^2\ \rho^{2k}(|s-s^*|) \\
&= \sum_{k=2}^{\infty} C_k^2 \sum_{s=1}^{t}\sum_{s^*=1}^{t} (1+\omega|s-s^*|^\alpha)^{2k(H-1)/\alpha} \\
&= \sum_{k=2}^{\infty} C_k^2 (t + 2\sum_{s=1}^{t-1}(t-s)(1+\omega|s|^\alpha)^{2k(H-1)/\alpha})
\end{aligned}$$

so that

$$\frac{1}{t^{2H}}\mathbb{V}\text{ar}(R_t) = \sum_{k=2}^{\infty} C_k^2 (t^{1-2H} + 2t^{-2H}\sum_{s=1}^{t-1}(t-s)(1+\omega|s|^\alpha)^{2k(H-1)/\alpha}).$$

But $\sum_{k=0}^{\infty} C_k^2 < \infty$ from (2.31) so if $t^{1-2H} + 2t^{-2H}\sum_{s=1}^{t-1}(t-s)(1+\omega|s|^\alpha)^{2k(H-1)/\alpha} \to 0$ as $t \to \infty$ then $\frac{1}{t^{2H}}\mathbb{V}\text{ar}(R_t) \to 0$. In fact we need only consider the $k = 2$ case, since $(1+\omega|s|^\alpha)^{2k(H-1)/\alpha} < (1+\omega|s|^\alpha)^{4(H-1)/\alpha}$ for $k > 2$. Now $t^{1-2H} \to 0$ for $\frac{1}{2} < H < 1$, and for $\frac{1}{2} < H < 1,\ H \neq \frac{3}{4}$,

$$t^{-2H}\sum_{s=1}^{t-1}(t-s)(1+\omega|s|^\alpha)^{4(H-1)/\alpha} < \omega^{4(H-1)/\alpha} t^{1-2H}\sum_{s=1}^{t} s^{4(H-1)} \tag{2.38}$$

$$\sim t^{1-2H}\int_1^t s^{4(H-1)}\mathrm{d}s \tag{2.39}$$

$$= \frac{t^{2H-2} - t^{1-2H}}{4H-3}$$

which $\to 0$ as $t \to \infty$. Here (2.38) follows since $(1+\omega|s|^{\alpha})^{4(H-1)/\alpha} < \omega^{4(H-1)/\alpha} s^{4(H-1)}$ for $\frac{1}{2} < H < 1$, and $t - s < t$. For $H = \frac{3}{4}$, line (2.39) becomes

$$t^{-\frac{1}{2}} \int_1^t \frac{\mathrm{d}s}{s} = t^{-\frac{1}{2}} \log(t)$$

which also goes to zero as t gets large. Hence $\frac{1}{t^{2H}} \mathbb{V}\mathrm{ar}(R_t) \to 0$ as $t \to \infty$ for $\frac{1}{2} < H < 1$ and so, in terms of asymptotic behavior, the $k = 1$ term is dominant in the expansion $T_{[nt]} - [nt] = \sum_{s=1}^{[nt]} G(\chi^2_\nu(s)) - [nt] = \sum_{s=1}^{[nt]} \sum_{k=1}^{\infty} C_k e_k(\chi^2_\nu(s))$. This follows from the results above since

$$\begin{aligned} &\mathbb{V}\mathrm{ar}\Big(\frac{1}{n^H}\Big(\sum_{s=1}^{[nt]} G(\chi^2_\nu(s)) - [nt]\Big) - \frac{C_1}{n^H}\sum_{s=1}^{[nt]} e_1(\chi^2_\nu(s))\Big) \\ &\qquad = \frac{t^{2-2\alpha}}{(nt)^{2-2\alpha}} \mathbb{V}\mathrm{ar}(R_{[nt]}) \to 0 \ \text{ as } \ n \to \infty \ \ \forall\, t > 0. \end{aligned} \tag{2.40}$$

Now $\mathbb{E}(\frac{1}{n^H}(\sum_{s=1}^{[nt]} G(\chi^2_\nu(s)) - [nt])) = \mathbb{E}(\frac{C_1}{n^H}\sum_{s=1}^{[nt]} e_1(\chi^2_\nu(s))) = 0$, the second equality following from the orthogonality property of $\{e_k(x)\}_{k=0}^{\infty}$ given by (2.30), since $e_0(x) = 1$. Hence Chebyshev's inequality implies that for each t, $\frac{1}{n^H}(\sum_{s=1}^{[nt]} G(\chi^2_\nu(s)) - [nt])$ and $\frac{C_1}{n^H}\sum_{s=1}^{[nt]} e_1(\chi^2_\nu(s))$ converge in probability, and therefore in distribution, to the same random variable, if such a limiting random variable exists. Now from (2.32) we have that $e_1(x) = \sqrt{\frac{2}{\nu}}(\frac{\nu}{2} - x)$ and $C_1 = \sqrt{\frac{2}{\nu}}$, and so

$$\begin{aligned} \frac{C_1}{n^H}\sum_{s=1}^{[nt]} e_1(\chi^2_\nu(s)) &= -\frac{1}{\nu}\frac{2}{n^H}\sum_{s=1}^{[nt]}(\chi^2_\nu(s) - \frac{\nu}{2}) \\ &= -\frac{1}{n^H}\sum_{s=1}^{[nt]}(\tau_\nu(s) - 1) \Rightarrow -\frac{1}{\nu}\sum_{j=1}^{[\nu]} R_j(t) \end{aligned} \tag{2.41}$$

for $R_j(t)$ independent copies of the H self similar Rosenblatt process. Here (2.41) follows from Equation (2.15). Equations (2.40) and (2.41) now imply that as $n \to \infty$,

$$\frac{1}{n^H}\Big(\sum_{s=1}^{[nt]} G(\chi^2_\nu(s)) - [nt]\Big) \xrightarrow{\mathcal{D}} -\frac{1}{\nu}\sum_{i=1}^{[\nu]} R_i(t)$$

where convergence is in the sense of finite dimensional distributions.

2.2.2. Conclusion. As with the gamma case, we have constructed $\{\tau_\nu^*(t)\}$ so as to make $\{T_t\}$, appropriately normed, asymptotically H self similar.

More generally, we have succeeded in constructing some correlated $\{\tau_\nu^*(t)\}$ increments processes with marginal inverse gamma distribution, which in turn lead to correlated $\{X_t\}$ log price increments processes with marginal t distribution. Indeed, as earlier we have constructed as a special case a process $\{T_t\}$ whereby an appropriately normed $\{T_t - t\}$ is asymptotically self similar. Thus we have a model that incorporates t-distributed log price marginal increments, and in which both $\{\tau_\nu^*(t)\}$ and $\{X_t^2\}$ are long range dependent.

2.3. Asymmetry and sign reversal in the asymptotic $\{T_t\}$ process

We now draw some parallels between the development of the inverse gamma process described above and the gamma process from Section 2.1. In both cases the ultimate process of interest in determining the limit distribution was

$$\frac{1}{n^H}(T_{[nt]} - [nt]) = \frac{1}{n^H}\sum_{s=1}^{[nt]}(G(\psi_\nu(s)) - 1)$$

as $n \to \infty$, where in the inverse gamma case $G(x) = G^{\mathrm{R}\Gamma}(x) = (\frac{\nu}{2} - 1)/x$, and in the gamma case $G(x) = G^{\Gamma}(x) = \frac{2}{\nu}x$, where $2\psi_\nu(t) \sim \chi_\nu^2$ for each t. As we have seen, the modified Laguerre expansion of $G^{\mathrm{R}\Gamma}(x)$ has first two terms 1 and $1 - \frac{2}{\nu}x$, followed by additional terms, whereas the modified Laguerre expansion of $G^{\Gamma}(x)$ can be thought of as consisting of two summands, 1 and $\frac{2}{\nu}x - 1$ only. Thus

$$G^{\mathrm{R}\Gamma}(\psi_\nu(t)) - 1 = 1 - \frac{2}{\nu}\psi_\nu(t) + R_t$$

where R_t consists of higher order terms, which as shown become asymptotically negligible, while

$$G^{\Gamma}(\psi_\nu(t)) - 1 = \frac{2}{\nu}\psi_\nu(t) - 1.$$

Hence we see once again that the weak limit as $n \to \infty$ in the inverse gamma case is the negative of the limit in the gamma case, which can now be understood to arise from the form of the modified Laguerre expansions of $G^{\mathrm{R}\Gamma}(x)$ and $G^{\Gamma}(x)$.

So for $\xi_k^{\Gamma}(t) \stackrel{\text{def}}{=} \frac{1}{k^H}(T_{[kt]} - [kt])$ for $\tau_\nu(t)$ with the gamma distribution and $-\xi_k^{R\Gamma}(t) \stackrel{\text{def}}{=} \frac{-1}{k^H}(T_{[kt]} - [kt])$ for $\tau_\nu^*(t)$ with the inverse gamma distribution, both $\xi_k^{\Gamma}(t)$ and $-\xi_k^{R\Gamma}(t)$ will have the same self similar asymptotic distribution, with self similarity parameter H, since both $\xi_k^{\Gamma}(t)$ and $-\xi_k^{R\Gamma}(t)$ converge weakly as $k \to \infty$ to the average of $[\nu]$ independent Rosenblatt processes. There are two somewhat surprising aspects to this result: the asymmetry of the limiting distribution of the difference in activity time T_t and clock time t, which can be seen from the cf of the Rosenblatt process (2.12) by noting that $\phi(-u) \neq \phi(u)$; and the sign reversal of the limiting distribution in the inverse gamma and gamma cases discussed above. We briefly illustrate empirically these features. This is done by taking $\rho(s) = (1+s^2)^{(H-1)/2}$ and simulating for various H values, 5000 chains of $\{\tau_5(t)\}$ and $\{\tau_5^*(t)\}$ of length 1000 and 4000, summing these chains to form T_{1000} and T_{4000}, then comparing the $\frac{1}{k^H}(T_k - k)$ from gamma distributed $\tau_5(t)$ with the $\frac{-1}{k^H}(T_k - k)$ from inverse gamma distributed $\tau_5^*(t)$ via a boxplot, which is shown in Figure 2.1.

Here the first box corresponds to $H = 0.65$, the second to $H = 0.75$, and the third to $H = 0.85$. In each box, the first boxplot corresponds to the negative of the inverse gamma case of length 1000, the second to the gamma case of length 1000, while the third and fourth are similar but with length 4000.

The main point to note is that the asymptotic distributional equivalence can be observed empirically, with the length 4000 pairs more alike than the length 1000 pairs, but that the rate of convergence seems to slow as H gets larger. This empirical observation is consistent with the theoretical result given in Theorem 3.1 of Leonenko & Anh (2001), which gives an upper bound for the Kolmogorov distance between a process similar to our gamma and

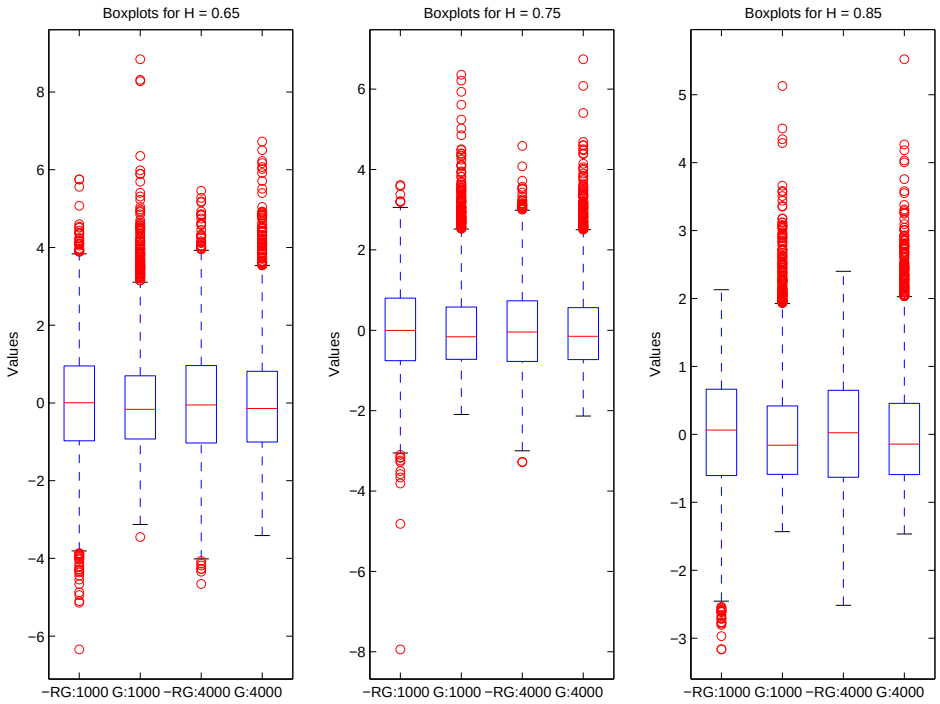

FIGURE 2.1. Boxplots of $\frac{1}{k^H}(T_k - k)$ and $\frac{-1}{k^H}(T_k - k)$ for gamma and inverse gamma distributed increments.

inverse gamma distributed $\{T_t\}$, and the Rosenblatt process as $t \to \infty$. Leonenko & Anh's result shows that the upper bound on the rate of convergence to the Rosenblatt process falls as H falls (or in the notation of their paper, falls as α gets larger), in agreement with our empirical result.

In addition to the rate of convergence, the skewness of the common asymptotic distribution of the difference between market time and clock time (T_t and t) is also evident empirically through the boxplots.

2.4. Other possible activity time processes

In Sections 2.1 and 2.2 above we described processes with an acf given essentially by a generalisation of $\rho(s) = (1 + s^2)^{(H-1)/2}$. Related work on the theory of LRD processes is

contained in Anh, Knopova & Leonenko (2004), in which a (possibly negative and therefore non-activity time) process $\{\xi(t), t \in \mathbb{R}\}$ is considered which displays *cyclical* LRD, with acf of the form

$$\mathbb{C}\text{ov}(\xi(t), \xi(t+s)) = \frac{\cos(\kappa s)}{(1+s^2)^{\alpha/2}}. \tag{2.42}$$

From the same paper we see that the spectral density associated with the acf above for $\kappa = 0$ is given by

$$f(\lambda) = \frac{2^{(1-\alpha)/2}}{\sqrt{\pi}\Gamma(\alpha/2)} |\lambda|^{(\alpha-1)/2} K_{(\alpha-1)/2}(|\lambda|). \tag{2.43}$$

Equation (2.42) for $\kappa = 0$ is essentially the cf of the symmetric VG and the pdf of the symmetric t distribution; while (2.43) is essentially the pdf of the symmetric VG and the cf of the symmetric t. This is a reflection of the duality between the symmetric VG and t distributions mentioned by Seneta (2004).

Before moving on we briefly describe two other processes that *do* result in LRD activity time constructions. The aim is only to introduce other possible approaches to constructing an activity time process, with proofs and greater detail available in the papers mentioned.

From Sly (2006) (see also Taqqu (1979)): for $\eta_1(t)$ as in Section 2.1, set $\Phi(\cdot)$ the distribution function of a standard normal, and F the distribution function of a gamma or inverse gamma random variable for example. Then for each t, $\Phi(\eta_1(t))$ has the uniform distribution and

$$\tau_t = F^{-1}(\Phi(\eta_1(t)))$$

has the distribution of F. Assume that τ_1 has finite variance, which holds for example for any gamma random variable and any inverse gamma random variable with $\delta > 2$ from (1.26), and set $T_t = \sum_{j=1}^{t} \tau_j$ as before. Now $F^{-1}(\Phi(x))$ is monotone increasing so that $C = \mathbb{E}(F^{-1}(\Phi(\eta_1(t)))\eta_1(t)) > 0$, which implies, from Taqqu (1979) for example, that

$$\frac{1}{n^H}(T_{nt} - nt) \xrightarrow{\mathcal{D}} CB_H(t) \tag{2.44}$$

as $n \to \infty$ where $B_H(t)$ is fractional Brownian motion.

For inverse gamma distributed τ_t with $1 < \delta \leq 2$, that is, τ_t with finite expectation but infinite variance, if $\frac{1}{2} < H < 1$ with $\delta H > 1$ then (2.44) again holds, while if $\frac{1}{2} < H < 1$ with $\delta H < 1$ then

$$\frac{1}{n^{1/\delta}}(T_{nt} - nt) \xrightarrow{\mathcal{D}} C^* R^*(t)$$

where C^* is a constant and $R^*(t)$ is a Lévy-stable process with parameter α (see Sly (2006), Section 4.3).

The second example of a LRD model comes from Taqqu & Levy (1986) (see also Liu (2000)): set $W(t) = \sum_{k=0}^{\infty} W_k I(S_{k-1} < t \leq S_k) = W_{N(t)}$ where $S_k = S_0 + \sum_{j=1}^{k} U_j$ is a renewal sequence with positive integer-valued inter-arrival times U_j, and $N(t)$ is the associated counting process. $W(t)$ therefore takes the random value W_k for the duration of the k^{th} inter-arrival time. Also assume that the $\{U_k\}$ are iid with $P(U_1 \geq u) \sim u^{-a}h(u)$ for $1 < a < 2$ and $h(\cdot)$ slowly varying with $\mathbb{E}U_1 = \mu$; that the $\{W_k\}$ are iid with $\mathbb{E}W_1 = 0$ and $\mathbb{E}W_1^2 < \infty$; and that the $\{U_k\}$ and $\{W_k\}$ are independent. So that $\{S_k\}$ is stationary choose $P(S_0 = u) = \mu^{-1}P(U_i \geq u + 1),\ u = 0, 1, \ldots$ so that by Karamata's theorem $P(S_0 \geq u) = \sum_{k=u}^{\infty} P(S_0 = k) \sim \mu^{-1}(a-1)^{-1}u^{-(a-1)}h(u)$, which implies $\mathbb{E}S_0 = \sum_{u=1}^{\infty} P(S_0 \geq u) = \infty$. Then

$$\mathbb{C}\text{ov}(W(t), W(t+s)) = \mathbb{E}W_k^2 \sum_{k=0}^{\infty} P(S_{k-1} < t < t + s \leq S_k) = \mathbb{E}W_k^2 P(S_0 \geq s)$$

so that $\sum_{s=0}^{\infty} \mathbb{C}\text{ov}(W(t), W(t+s)) = \infty$, giving LRD.

In fact, similar to above,

$$\zeta_k^{\text{TL}}(t) = \frac{\sum_{i=1}^{[kt]} W(i)}{k^{1/a}L(k)}$$

for $L(\cdot)$ slowly varying and $t \in [0, 1]$, converges in finite dimensional distribution as $k \to \infty$ to a self similar Lévy-stable process with parameter a. Using the notation from Section 2.1, one could set $\tau_t = W(t) + 1$ and $T_t = \sum_{j=1}^{t} \tau_j$, taking $W_k,\ k = 1, 2, \cdots$ to be mean-corrected iid gamma or inverse gamma random variables for example.

Interestingly, a restriction parallel to that imposed on $Z(s)$ from Section 2.1 must be imposed on the Taqqu & Levy construction. In this case we require $P(S_0 = u) = P(S_0 \geq u) - P(S_0 \geq u + 1)$ to be positive and decreasing with u. If the latter does not hold then the distribution of the U_i will admit negative probabilities. Hence as earlier, there is no distribution of U_i which gives $P(S_0 \geq s) = (1 + \omega|s|^a)^{(H-1)/a}$ where $a = 2$, $\omega = 1$ for $0.648 < H < 1$.

CHAPTER 3

Simulation, estimation and fit to data

This chapter explicitly details simulation of LRD VG and t models, and assesses a number of estimation techniques (the method of moments, product-density maximum likelihood, non-standard minimum χ^2, and characteristic function based estimation) to recover the (in our case known) model parameter values of the simulated data. The motivation for comparing estimation techniques is the apparent poor performance of the method of moments on real data (Tjetjep & Seneta (2006)), and the apparent inappropriateness of product-density maximum likelihood estimation given non-independence.

In the event we find that product-density maximum likelihood estimation performs best, followed by minimum χ^2 estimation. Thus although it assumes independence of returns, product-density maximum likelihood outperforms procedures such as the method of moments which do not assume independence. We also find that classical χ^2 goodness of fit tests can be carried out as if the data were independent.

In addition to simulated data, we also fit VG and t models to historical financial data and assess goodness of fit. In this case we find that for three data sets, the VG and t give reasonably similar fits, while for one, Microsoft, the VG fit is clearly superior. The bulk of the contents of this chapter has been published as Finlay & Seneta (2008a).

3.1. Simulation and estimation techniques

To simulate LRD VG and t log price increments at integer points in time, we follow the procedure laid out in Chapter 2 above. In both cases the acf of the underlying Gaussian processes in continuous time is $\rho(s) = (1+s^2)^{-\frac{\gamma}{2}},\ 0 < \gamma < \frac{1}{2}$, and we restrict ourselves to the integer ν case.

The VG procedure is as follows, where we take $\nu = 2\alpha$ integer valued:

(1) Generate an $n \times n$ symmetric positive definite matrix Σ where the $(i,j)^{\text{th}}$ entry is $\Sigma_{ij} = (1+(i-j)^2)^{-\frac{\gamma}{2}}$ for some γ with $0 < \gamma < \frac{1}{2}$.

(2) Calculate $\Sigma^{\frac{1}{2}}$, the 'square-root' of Σ ($\Sigma^{\frac{1}{2}} = ED^{\frac{1}{2}}E^{-1}$ where D is the diagonal matrix of the eigenvalues of Σ and the columns of E consist of the corresponding orthonormal eigenvectors of Σ).

(3) Generate ν column n-vectors of $N(\mathbf{0}, I)$ independent normal random variables, and left multiply each of these by $\Sigma^{\frac{1}{2}}$. Call the ν vectors $\boldsymbol{\eta}_1, \cdots, \boldsymbol{\eta}_\nu$. Each $\boldsymbol{\eta}_i$ is now an n-vector of correlated normals, with correlation matrix given by Σ.

(4) Calculate the column vector $\boldsymbol{\tau}_{\text{VG}}$ where $\tau_{\text{VG},j} = (\eta_{1,j}^2 + \cdots + \eta_{\nu,j}^2)/\nu$ for $j = 1, \ldots, n$. Here $\tau_{\text{VG},j}$ is the j^{th} element of $\boldsymbol{\tau}_{\text{VG}}$ and $\eta_{i,j}$ is the j^{th} element of $\boldsymbol{\eta}_i$. Then each element of $\boldsymbol{\tau}_{\text{VG}}$ has a marginal $\Gamma(\alpha, \alpha)$ distribution, for $\alpha = \frac{\nu}{2}$, and the correlation matrix of $\boldsymbol{\tau}_{\text{VG}}$ is given by $\rho_{ij} = (1+(i-j)^2)^{-\gamma}$.

(5) Calculate the column vector $\mathbf{X}_{\text{VG}}$ where $X_{\text{VG},j} = \mu + \theta\tau_{\text{VG},j} + \sigma\sqrt{\tau_{\text{VG},j}}Z_j$. Here $X_{\text{VG},j}$ is the j^{th} element of $\mathbf{X}_{\text{VG}}$, and $Z_j \sim N(0,1)$ is independent of $\tau_{\text{VG},j}$ and $Z_i,\ i \neq j$. $\mathbf{X}_{\text{VG}}$ is then an n-vector of long range dependent VG increments, where the LRD comes from the asymptotically self similar $\{T_t - t\}$ process, where $T_k = \sum_{j=1}^{k} \tau_j$, and which has self similarity parameter given by $H = 1 - \gamma$.

To simulate t increments, replace $\boldsymbol{\tau}_{\text{VG}}$ in point (4) above by $\boldsymbol{\tau}_{\text{t-dist}}$, where

$$\tau_{\text{t-dist},j} = \frac{\nu - 2}{\eta_{1,j}^2 + \cdots + \eta_{\nu,j}^2}.$$

Then each element of $\boldsymbol{\tau}_{\text{t-dist}}$ has the $R\Gamma(\delta, \delta - 1)$ distribution for $\delta = \frac{\nu}{2}$. The autocorrelation function is considerably more complicated, but again leads to a LRD $\boldsymbol{X}_{\text{t-dist}}$ with marginal t distribution and asymptotic self similar $\{T_t - t\}$. Note however that unlike the VG, the t construction requires $\nu > 4$ ($\delta > 2$). This bound ensures $\mathbb{V}\text{ar}(\tau_t) < \infty$, a condition needed to prove asymptotic self similarity.

Next to create some actual data series we choose $n = 2500$ along with various γ and ν to simulate $\boldsymbol{\tau}_{\text{VG}}$ and $\boldsymbol{\tau}_{\text{t-dist}}$, and $\mu,\ \theta$ and σ to simulate $\boldsymbol{X}_{\text{VG}}$ and $\boldsymbol{X}_{\text{t-dist}}$.

To then re-estimate the known parameter values $\sigma,\ \alpha$ or $\delta,\ \theta$ and μ of the marginal distribution we employ four techniques: the method of moments (MOM), a minimum χ^2 method, product-density maximum likelihood estimation (PMLE), as well as empirical characteristic function estimation (ECF) which shall be dealt with separately in Section 3.3.

Here and through out this chapter *we take PMLE to mean the maximisation of the product of marginal densities*, as opposed to the maximisation of the full joint density of all 2500 readings (of course if the data are independent the two coincide). That is, we maximise $\sum_{t=1}^{n} \log f_{\text{VG}}(X_t)$ or $\sum_{t=1}^{n} \log f_{\text{t-dist}}(X_t)$ as a function of the parameters. Estimation via the product of marginal (or bivariate) densities has also been considered by other authors, and is sometimes referred to as *pseudolikelihood* in the literature (see for example Cox & Reid (2004) who detail certain cases where pseudolikelihood results in consistent estimators, considering in particular some normal models where the full likelihood is difficult to work with). In our case maximising the full joint likelihood would likely produce better results than PMLE, but does not appear feasible. However, in an attempt to incorporate dependence structure into our estimation technique in a somewhat parallel fashion, we use the ECF technique described in Section 3.3.

Estimation via MOM is performed as in Tjetjep & Seneta (2006). Specifically, note that for μ or $\theta = 0$ we can solve the MOM equations directly, whereas for $\mu,\ \theta \neq 0$ we cannot. To circumvent this we minimise, for each simulation run,

$$f_{\text{MOM}}(\sigma, \{\alpha \text{ or } \delta\}, \theta, \mu) = \sum_{i=1}^{4} (\frac{O_i - E_i}{O_i})^2 \tag{3.1}$$

where O_i is the i^{th} sample central moment as calculated from the data, and E_i is the i^{th} theoretical central moment given the parameter set, as detailed in Section 1.3. Note that we divide by O_i and not E_i as this aids numerical stability. If the moments of the data are consistent with the model, then there will exist a set of parameters such that $f_{\text{MOM}}(\hat{\sigma}, \{\hat{\alpha} \text{ or } \hat{\delta}\}, \hat{\theta}, \hat{\mu}) = 0$. If the moments of the data are not consistent with the model, then we choose the set of parameters that minimises (3.1). The method is motivated by the Generalised Method of Moments (GMM) of Hansen (1982), and is intended as a simple way to solve the moment equations exactly if possible, or choose a set of compromise parameters which deliver moments close to those of the data if not. In fact of the models fitted via MOM in Section 3.2.2 for $\mu,\ \theta \neq 0$, 99.3% of fitted VG models returned $f_{\text{MOM}}(\hat{\sigma}, \hat{\alpha}, \hat{\theta}, \hat{\mu}) = 0$, whereas only 79.2% of fitted t models did the same. As touched on earlier, MOM fitting for the t requires us to impose $\hat{\delta} > 4$ when we numerically minimise (3.1) – this is to ensure that we have four finite moments to identify the four parameters with – and as such is likely to produce a bad fit when the true δ value is less than 4.

The minimum χ^2 approach we use is non-standard. Our minimum χ^2 procedure, for a single realisation of $n = 2500$ increments, involves numerically minimising a statistic of the form

$$\chi^2_{\text{data}}(\sigma, \{\alpha \text{ or } \delta\}, \theta, \mu) = \sum_{i=1}^{100} \frac{(O_i^* - E_i^*)^2}{E_i^*} \tag{3.2}$$

where for $i = 1, \ldots, 100,\ O_i^*$ is the total number of observations in the data set divided by 100 (so in our case $O_i^* = 25$), and E_i^* is the expected number of observations falling within the i^{th} 1% sample quantile band (that is, the area under the marginal pdf between order

statistic $2500 * (i-1)/100$ and order statistic $2500 * i/100$, given the chosen parameters). The minimum χ^2 method is again motivated by the GMM, and results in a good fit between the estimated marginal pdf and the empirical pdf, at least as measured by the value of the χ^2 goodness of fit statistic. In contrast, method of moments fitting merely ensures that the first four moments of the fitted model agree with the data, and as such provides no guarantee of a holistic fit. Further, our minimum χ^2 procedure only requires that the proportion of readings falling into an (albeit data-determined) interval converges to the probability of landing in that interval, and so can sensibly be used even when data are not independent. In contrast, the PMLE procedure we employ maximises the product of marginal likelihoods, and so may not be appropriate given dependent data. Hence minimum χ^2 estimation was expected to have advantages over both the MOM and PMLE.

3.1.1. Estimation of the correlation structure. To estimate H and hence γ we use least squares and restrict ourselves to symmetric $\mu = \theta = 0$ models and mean–adjusted data. From (1.1), and assuming that $\{T_t - t\}$ is exactly self similar, we can estimate H by minimising

$$f(H) = \sum_{k=1}^{300} (\hat{\rho}_k - \frac{1}{2}((k+1)^{2H} + (k-1)^{2H} - 2k^{2H}))^2. \tag{3.3}$$

Here we use (1.16) in the form

$$\rho_k \stackrel{def}{=} \mathbb{C}\text{orr}(\tau_t, \tau_{t+k}) = \frac{\mathbb{C}\text{ov}(X_t^2, X_{t+k}^2)}{\sigma^4 \mathbb{V}\text{ar}(\tau_t)} \tag{3.4}$$

to estimate ρ_k, since all components occurring on the right-hand side of (3.4) may be estimated from observations on $\{X_t\}$ over unit time points. For our specific VG and t models, we do this via the sample covariance of the squared return series for $\mathbb{C}\text{ov}(X_t^2, X_{t+k}^2)$, and via the fitted marginal model parameters using Equations (1.19) and (1.27) for σ^4 and $\mathbb{V}\text{ar}(\tau_t)$, where respectively $\alpha = \frac{\nu}{2}$ and $\delta = \frac{\nu}{2}$.

In fact one can also estimate σ^4 and $\mathbb{V}\text{ar}(\tau_t)$ without appealing to a marginal model fit. We can estimate σ^2 by $\mathbb{E}(X_t - \mathbb{E}X_t)^2$ from (1.7) and $\mathbb{V}\text{ar}(\tau_t)$, which we referred to as M_2

in Chapter 1, by $\mathbb{E}(X_t - \mathbb{E}X_t)^4/(3\sigma^4) - 1$ from (1.9). Expressed differently, the existence of such a 'data only' estimation method for ρ_k implies that for any subordinator model whose parameters are fitted via the method of moments, the marginal distribution of τ_t and therefore X_t chosen (VG, t, GH, or some other distribution) has no impact on the estimate of ρ_k, and therefore H, produced – that is, for method of moments fitted models, estimates of the correlation structure are independent of any marginal distribution assumptions.

We in fact estimate ρ_k using both the marginal model dependent and marginal model independent methods described above, and find that both methods give broadly similar results. It is important to note however that we only have self similarity of an asymptotic normed $\{T_t - t\}$, and so the autocorrelation function of $\{\tau_t\}$ given by the right-hand side of (1.1) as $\rho_k = \frac{1}{2}((k+1)^{2H} + (k-1)^{2H} - 2k^{2H})$ will not be exact.

Although we do not pursue the technique here, we note that one can also use minimum contrast methods for estimating H. In this case H is estimated by minimising a certain distance between the spectral density and periodogram (see Leonenko & Sakhno (2006) and Anh, Leonenko & Sakhno (2007)). Under certain assumptions, such an estimator is consistent and asymptotically normal.

3.2. Estimation results for the simulated data

Recall from (1.14) that $\theta = 0$ implies the increments $\{X_t\}$ are uncorrelated, irrespective of the autocorrelation structure of the increments $\{\tau_t\}$. If $\theta = 0$ and the increments $\{\tau_t\}$ are also uncorrelated, then from (1.16) the increments $\{X_t^2\}$ are uncorrelated. If the $\{\tau_t\}$ are iid, then the $\{X_t\}$ are iid. For our autocorrelated $\{\tau_t\}$ we require $\frac{1}{2} < H < 1$. As $H \downarrow \frac{1}{2}$, from (1.1) the autocorrelation at lag k, ρ_k, approaches zero, so the $\{\tau_t\}$ and $\{X_t\}$ might closely resemble iid behavior, irrespective of the value of θ.

3.2.1. Estimation results for $\mu = \theta = 0$. Beginning with the $\mu = \theta = 0$ case, we simulate 1000 runs of $n = 2500$ VG increments for all combinations of $\sigma = 0.01$, $\alpha = \{2, 2.5, 3\}$ and $H = \{\text{iid}, 0.75, 0.85, 0.95\}$. By $H = \text{iid}$ we mean that we simulate independent increments $\{\tau_t\}$, that is, we choose Σ to be the identity matrix in Section 3.1, so that each $\boldsymbol{\eta}_i$ is an n-vector of uncorrelated normals. Although H plays no role in this case we can still 'estimate' it via (3.4) and (3.3), and expect the estimate to be close to $\frac{1}{2}$. Similarly, we simulate 1000 runs of $n = 2500$ t distributed increments for all combinations of $\sigma = 0.01$, $\delta = \{2.5, 3, 3.5\}$ and $H = \{\text{iid}, 0.75, 0.85, 0.95\}$. We then jointly estimate the model parameters using the three methods described in Section 3.1. Here the sets of α and δ were chosen to resemble those found in estimation of financial data (see Section 3.5), to ensure that 2α and $2\delta \in \mathbb{N}$, and to accommodate the fact that we need $\delta > 2$ in order to construct our increments.

For expository purposes we first look at 4 individual simulation runs (each of length $n = 2500$) in some detail; they are the first simulated VG increments of each of the $\sigma = 0.01,\ \alpha = 2.5$ and $H = \{\text{iid}, 0.75, 0.85, 0.95\}$ runs. Table 3.1 provides some descriptive statistics of each of these.

H	Sample Mean	Sample Std. Deviation	Sample Skewness	Sample Kurtosis
iid	8.1e-5	1.0e-2	0.14	4.50
0.75	-2.2e-5	9.9e-3	-0.01	4.12
0.85	7.9e-5	9.7e-3	-0.13	4.44
0.95	1.5e-4	8.0e-3	0.12	4.40

TABLE 3.1. Descriptive statistics of VG simulation runs.

Here the theoretical mean and standard deviation of each run are given by $\mu + \theta$ and $\sqrt{\sigma^2 + \frac{\theta^2}{\alpha}}$ which evaluate to 0 and 0.01 respectively, from (1.10) the theoretical skewness is given by

$$\beta = \frac{2\theta^3 + 3\theta\sigma^2\alpha}{\alpha^{1/2}(\theta^2 + \alpha\sigma^2)^{3/2}} = 0,$$

while from (1.11) the theoretical kurtosis is given by

$$\kappa = \frac{3\theta^4(\alpha+2) + 6\sigma^2\theta^2\alpha(2+\alpha) + 3\sigma^4\alpha^2(\alpha+1)}{\alpha(\theta^2+\alpha\sigma^2)^2} = 4.2.$$

Table 3.2 shows the results of parameter estimation.

	MOM estimates			PMLE estimates			Min χ^2 estimates		
H	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{H}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{H}$	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{H}$
iid	1.0e-2	1.99	0.50	1.0e-2	2.56	0.50	1.0e-2	2.62	0.50
0.75	9.9e-3	2.67	0.78	1.0e-2	2.34	0.78	1.0e-2	2.37	0.78
0.85	9.7e-3	2.09	0.87	9.7e-3	2.27	0.88	9.7e-3	2.21	0.88
0.95	8.0e-3	2.12	0.86	8.0e-3	2.42	0.87	7.9e-3	2.41	0.87

TABLE 3.2. Parameter estimation results of VG simulation runs.

Even from this limited sample one can see that the method of moments generally gives less precise estimates of the true α parameter value in the presence of substantial dependence than the other two methods, and that the 'Brownian motion driven parameter' σ appears easier to estimate than the other parameters. The fact that $\hat{H}$ is estimated to be larger when $H = 0.85$ than when $H = 0.95$ looks a little strange, but could be due to sampling error from just one simulation run.

Analogously, some descriptive statistics of the first simulated t increments for each of the $\sigma = 0.01,\ \delta = 2.5$ and $H = \{\text{iid}, 0.75, 0.85, 0.95\}$ runs are given in Table 3.3 below.

H	Sample Mean	Sample Std. Deviation	Sample Skewness	Sample Kurtosis
iid	-1.6e-4	1.0e-2	0.05	6.40
0.75	1.3e-4	9.8e-3	-0.04	4.83
0.85	-6.8e-5	6.3e-3	-0.12	4.73
0.95	1.2e-4	1.1e-2	0.34	5.94

TABLE 3.3. Descriptive statistics of t simulation runs.

Here the theoretical mean and standard deviation of each run are given by $\mu + \theta$ and $\sqrt{\sigma^2 + \frac{\theta^2}{\delta-2}}$ which evaluate to 0 and 0.01 respectively, since $\theta = 0$ the model is symmetric, although from (1.10) the theoretical skewness for $\delta > 3$ is given in general by

$$\beta = \frac{4\theta^3 + 3\theta\sigma^2(\delta-3)}{(\delta-2)(\delta-3)(\sigma^2+\theta^2/(\delta-2))^{3/2}}$$

and since $\delta = 2.5$ the theoretical kurtosis is infinite, although in general from (1.11), the kurtosis is for $\delta > 4$ given by

$$\frac{3\theta^4(\delta+5)+24\sigma^2\theta^2(\delta-4)+3\sigma^2(2\theta^2+\sigma^2)(\delta-3)(\delta-4)}{(\delta-2)(\delta-3)(\delta-4)(\sigma^2+\theta^2/(\delta-2))^2} + \frac{3\sigma^4(\delta-2)(\delta-3)(\delta-4)}{(\delta-2)(\delta-3)(\delta-4)(\sigma^2+\theta^2/(\delta-2))^2}.$$

Table 3.4 shows the results of parameter estimation.

	MOM estimates			PMLE estimates			Min χ^2 estimates		
H	$\hat{\sigma}$	$\hat{\delta}$	$\hat{H}$	$\hat{\sigma}$	$\hat{\delta}$	$\hat{H}$	$\hat{\sigma}$	$\hat{\delta}$	$\hat{H}$
iid	1.0e-2	2.88	0.50	1.0e-2	2.41	0.50	1.0e-2	2.25	0.50
0.75	9.8e-3	3.63	0.80	9.8e-3	3.18	0.72	9.8e-3	2.96	0.71
0.85	6.3e-3	3.72	0.77	6.3e-3	5.14	0.76	6.3e-3	5.84	0.78
0.95	1.1e-2	3.01	0.91	1.1e-2	3.01	0.83	1.1e-2	3.12	0.84

TABLE 3.4. Parameter estimation results of t simulation runs.

Over the entire 1000 simulation runs, comparison between estimation methods is made on the basis of mean absolute deviation (MAD) of estimated parameter values from the true values, a standard measure in financial mathematics. So for example the MAD score for the σ parameter would be given by $\text{MAD}(\hat{\sigma}) = \frac{1}{1000}\sum_{i=1}^{1000}|\hat{\sigma}_i - \sigma|$, where σ is the true value of the parameter, and $\hat{\sigma}_i$ is the estimated parameter value from the i^{th} simulation run.

To separately examine the accuracy (how close our estimators are to the true parameter value) and precision (the spread of estimates produced) of our estimators, we also report the bias of each estimation method as well as the standard deviation of the estimates produced. Again for σ the true parameter and $\hat{\sigma}_i$ the estimated parameter value from the i^{th} simulation run, the bias is given by $\text{Bias}(\hat{\sigma}) = \bar{\sigma} - \sigma$ for $\bar{\sigma} = \frac{1}{1000}\sum_{i=1}^{1000}\hat{\sigma}_i$ the average of the estimates; and the standard deviation is given by $\text{SD}(\hat{\sigma}) = \sqrt{\sum_{i=1}^{1000}(\hat{\sigma}_i - \bar{\sigma})^2/1000}$. Note also that the root mean square error (RMSE), another measure of overall accuracy analogous to the MAD and defined as $\text{RMSE}(\hat{\sigma}) = \sqrt{\frac{1}{1000}\sum_{i=1}^{1000}(\hat{\sigma}_i - \sigma)^2}$, is given by $\sqrt{(\text{SD}(\hat{\sigma}))^2 + (\text{Bias}(\hat{\sigma}))^2}$.

Tables 3.5, 3.6 and 3.7 gives the MAD scores, bias and standard deviation of estimates produced for the VG model, while Tables 3.8, 3.9 and 3.10 gives similar information for the t model. In each case, subscripts in the top row give the true α or δ value.

	H	σ_2	$\sigma_{2.5}$	σ_3	α_2	$\alpha_{2.5}$	α_3
MOM	iid	1.6e-4	1.4e-4	1.4e-4	0.37	0.52	0.66
	0.75	6.7e-4	5.4e-4	5.5e-4	0.47	0.64	0.78
	0.85	1.0e-3	1.0e-3	8.8e-4	0.72	0.97	1.24
	0.95	2.0e-3	1.8e-3	1.6e-3	3.29	4.34	4.71
PMLE	iid	1.6e-4	1.4e-4	1.4e-4	0.21	0.31	0.42
	0.75	6.7e-4	5.4e-4	5.5e-4	0.27	0.40	0.55
	0.85	1.0e-3	1.0e-3	8.8e-4	0.49	0.70	0.94
	0.95	2.0e-3	1.8e-3	1.7e-3	2.82	3.86	4.16
Min χ^2	iid	1.6e-4	1.5e-4	1.4e-4	0.23	0.33	0.47
	0.75	6.7e-4	5.5e-4	5.5e-4	0.27	0.42	0.59
	0.85	1.0e-3	1.0e-3	8.9e-4	0.49	0.70	0.95
	0.95	2.0e-3	1.8e-3	1.6e-3	2.79	4.03	4.23

TABLE 3.5. MAD statistics for VG models with $\mu = \theta = 0$.

	H	σ_2	$\sigma_{2.5}$	σ_3	α_2	$\alpha_{2.5}$	α_3
MOM	iid	-4.4e-6	-3.2e-6	-1.1e-5	0.12	0.18	0.24
	0.75	-2.6e-5	-3.5e-5	1.7e-5	0.28	0.40	0.43
	0.85	-2.6e-5	-8.7e-5	-3.6e-5	0.62	0.83	1.03
	0.95	-2.6e-4	-2.2e-4	-3.8e-4	3.26	4.30	4.66
PMLE	iid	-3.8e-6	-2.1e-6	-9.8e-6	0.04	0.05	0.09
	0.75	-2.3e-5	-3.3e-5	1.7e-5	0.13	0.21	0.29
	0.85	-2.1e-5	-8.4e-5	-3.5e-5	0.42	0.60	0.78
	0.95	-2.4e-4	-2.0e-4	-3.4e-4	2.80	3.84	4.14
Min χ^2	iid	-5.7e-6	-3.3e-6	-1.2e-5	0.04	0.06	0.10
	0.75	-2.1e-5	-3.0e-5	1.5e-5	0.13	0.20	0.31
	0.85	-1.9e-5	-8.2e-5	-3.5e-5	0.41	0.59	0.78
	0.95	-2.3e-4	-2.0e-4	-3.4e-4	2.77	4.00	4.21

TABLE 3.6. Bias for VG models with $\mu = \theta = 0$.

A few things are immediately obvious: estimation generally becomes more inaccurate as α or δ get larger, and as H and therefore correlation gets larger. In addition, regarding the MAD of the α and δ coefficient estimates, the MOM method always performs worse than PMLE or minimum χ^2 estimation. The PMLE and minimum χ^2 methods are closer,

	H	σ_2	$\sigma_{2.5}$	σ_3	α_2	$\alpha_{2.5}$	α_3
MOM	iid	2.0e-4	1.8e-4	1.7e-4	0.47	0.67	0.83
	0.75	8.5e-4	6.8e-4	6.9e-4	0.58	0.81	1.01
	0.85	1.3e-3	1.2e-3	1.1e-3	0.81	1.14	1.51
	0.95	2.5e-3	2.2e-3	2.0e-3	4.98	5.82	6.39
PMLE	iid	2.0e-4	1.8e-4	1.7e-4	0.27	0.40	0.55
	0.75	8.5e-4	6.8e-4	6.9e-4	0.35	0.53	0.72
	0.85	1.3e-3	1.2e-3	1.1e-3	0.58	0.89	1.15
	0.95	2.5e-3	2.2e-3	2.0e-3	3.67	4.47	4.57
Min χ^2	iid	2.1e-4	1.9e-4	1.8e-4	0.30	0.45	0.62
	0.75	8.5e-4	6.9e-4	6.9e-4	0.35	0.56	0.78
	0.85	1.3e-3	1.3e-3	1.1e-3	0.63	0.96	1.26
	0.95	2.5e-3	2.2e-3	2.0e-3	3.70	5.02	4.91

TABLE 3.7. Standard deviation for VG models with $\mu = \theta = 0$.

	H	$\sigma_{2.5}$	σ_3	$\sigma_{3.5}$	$\delta_{2.5}$	δ_3	$\delta_{3.5}$
MOM	iid	2.1e-4	1.8e-4	1.6e-4	0.47	0.49	0.61
	0.75	6.4e-4	5.5e-4	5.2e-4	0.52	0.56	0.72
	0.85	1.2e-3	1.0e-3	9.0e-4	0.72	0.87	0.96
	0.95	2.2e-3	1.9e-3	1.8e-3	3.43	3.95	5.19
PMLE	iid	2.0e-4	1.7e-4	1.6e-4	0.20	0.30	0.40
	0.75	6.5e-4	5.6e-4	5.2e-4	0.27	0.36	0.48
	0.85	1.2e-3	1.0e-3	9.0e-4	0.48	0.64	0.75
	0.95	2.2e-3	1.9e-3	1.8e-3	3.03	3.59	4.63
Min χ^2	iid	2.2e-4	1.9e-4	1.7e-4	0.24	0.36	0.48
	0.75	6.6e-4	5.6e-4	5.3e-4	0.31	0.41	0.55
	0.85	1.2e-3	1.0e-3	9.1e-4	0.54	0.70	0.88
	0.95	2.2e-3	1.9e-3	1.8e-3	3.23	3.85	4.85

TABLE 3.8. MAD statistics for t models with $\mu = \theta = 0$.

but PMLE does generally outperform. As remarked on earlier, the accuracy and precision of the σ estimates are similar for all methods.

In Tables 3.11, 3.12 and 3.13 are the MAD scores, bias and standard deviation of the H estimates for the $\mu = \theta = 0$ models achieved by minimising (3.3). The minimum χ^2 method is used to estimate the marginal distribution parameters α or δ; using maximum likelihood to estimate the marginal distribution parameters results in virtually the same values as those shown.

	H	$\sigma_{2.5}$	σ_3	$\sigma_{3.5}$	$\delta_{2.5}$	δ_3	$\delta_{3.5}$
MOM	iid	1.3e-6	3.1e-6	-5.4e-6	0.41	0.34	0.36
	0.75	-5.8e-5	4.0e-5	-4.8e-5	0.47	0.43	0.53
	0.85	-8.3e-5	-1.2e-4	-4.4e-5	0.69	0.78	0.79
	0.95	-4.8e-4	-2.7e-4	-2.2e-4	3.42	3.93	5.17
PMLE	iid	7.2e-6	1.0e-6	-2.4e-6	0.04	0.07	0.09
	0.75	-5.4e-5	4.4e-5	-4.6e-5	0.14	0.14	0.26
	0.85	-8.1e-5	-1.3e-4	-4.5e-5	0.38	0.52	0.60
	0.95	-5.0e-4	-2.8e-4	-2.3e-4	3.01	3.58	4.61
Min χ^2	iid	6.8e-6	-2.1e-6	-3.9e-6	0.05	0.09	0.11
	0.75	-5.4e-5	4.6e-5	-4.4e-5	0.15	0.16	0.28
	0.85	-9.9e-5	-1.3e-4	-5.0e-5	0.43	0.57	0.69
	0.95	-5.1e-4	-2.9e-4	-2.4e-4	3.21	3.84	4.82

TABLE 3.9. Bias for t models with $\mu = \theta = 0$.

	H	$\sigma_{2.5}$	σ_3	$\sigma_{3.5}$	$\delta_{2.5}$	δ_3	$\delta_{3.5}$
MOM	iid	2.8e-4	2.5e-4	2.0e-4	0.40	0.53	0.70
	0.75	8.1e-4	6.9e-4	6.5e-4	0.45	0.59	0.80
	0.85	1.4e-3	1.3e-3	1.1e-3	0.74	0.88	1.13
	0.95	2.7e-3	2.3e-3	2.2e-3	5.79	6.01	6.79
PMLE	iid	2.5e-4	2.2e-4	2.0e-4	0.26	0.38	0.53
	0.75	8.2e-4	6.9e-4	6.5e-4	0.34	0.46	0.62
	0.85	1.4e-3	1.3e-3	1.1e-3	0.64	0.76	0.98
	0.95	2.7e-3	2.3e-3	2.2e-3	3.40	3.72	4.65
Min χ^2	iid	2.7e-4	2.4e-4	2.2e-4	0.31	0.45	0.63
	0.75	8.3e-4	7.0e-4	6.5e-4	0.39	0.55	0.76
	0.85	1.4e-3	1.3e-3	1.1e-3	0.68	0.91	1.20
	0.95	2.7e-3	2.3e-3	2.2e-3	3.68	4.11	5.00

TABLE 3.10. Standard deviation for t models with $\mu = \theta = 0$.

	VG models			t models		
True $H \backslash \alpha$ or δ	2	2.5	3	2.5	3	3.5
iid	0.03	0.03	0.03	0.01	0.02	0.02
0.75	0.04	0.04	0.04	0.07	0.05	0.04
0.85	0.05	0.05	0.05	0.12	0.09	0.07
0.95	0.11	0.11	0.10	0.13	0.12	0.12

TABLE 3.11. MAD of $\hat{H}$ estimates using the minimum χ^2 method.

If we use rather the marginal distribution free method of estimating H as detailed in the second paragraph of Section 3.1.1 (i.e., estimation via moments – the 'MOM' method), we

	VG models			t models		
True $H\backslash\alpha$ or δ	2	2.5	3	2.5	3	3.5
iid	0.03	0.03	0.03	0.01	0.02	0.02
0.75	0.02	0.02	0.03	-0.07	-0.04	-0.01
0.85	-0.04	-0.04	-0.04	-0.12	-0.08	-0.07
0.95	-0.11	-0.11	-0.10	-0.13	-0.12	-0.11

TABLE 3.12. Bias of $\hat{H}$ estimates using the minimum χ^2 method.

	VG models			t models		
True $H\backslash\alpha$ or δ	2	2.5	3	2.5	3	3.5
iid	0.04	0.04	0.05	0.02	0.03	0.04
0.75	0.04	0.04	0.04	0.06	0.05	0.04
0.85	0.04	0.05	0.05	0.07	0.05	0.05
0.95	0.05	0.05	0.05	0.06	0.06	0.05

TABLE 3.13. Standard deviation of $\hat{H}$ estimates using the minimum χ^2 method.

arrive at Tables 3.14, 3.15 and 3.16. Here the $\hat{H}$ estimates produced are quite similar, if generally a little better, than those from Tables 3.11, 3.12 and 3.13.

	VG models			t models		
True $H\backslash\alpha$ or δ	2	2.5	3	2.5	3	3.5
iid	0.03	0.03	0.03	0.02	0.02	0.02
0.75	0.04	0.04	0.04	0.05	0.04	0.03
0.85	0.05	0.05	0.05	0.09	0.07	0.07
0.95	0.10	0.10	0.10	0.13	0.12	0.11

TABLE 3.14. MAD of $\hat{H}$ estimates using the moment method.

	VG models			t models		
True $H\backslash\alpha$ or δ	2	2.5	3	2.5	3	3.5
iid	0.03	0.03	0.03	0.02	0.02	0.02
0.75	0.02	0.03	0.03	-0.03	-0.02	0.00
0.85	-0.04	-0.04	-0.04	-0.09	-0.07	-0.06
0.95	-0.10	-0.10	-0.10	-0.13	-0.12	-0.11

TABLE 3.15. Bias of $\hat{H}$ estimates using the moment method.

As mentioned earlier, our data are only asymptotically self similar. As such, we would not necessarily expect the estimated H values to be very close to the 'true' values. They do however generally conform to expectations, with the bias generally quite small, although

	VG models			t models		
True $H\backslash\alpha$ or δ	2	2.5	3	2.5	3	3.5
iid	0.04	0.04	0.05	0.03	0.03	0.04
0.75	0.04	0.04	0.05	0.05	0.04	0.04
0.85	0.04	0.05	0.05	0.05	0.05	0.05
0.95	0.05	0.06	0.06	0.05	0.05	0.06

TABLE 3.16. Standard deviation of $\hat{H}$ estimates using the moment method.

increasing in magnitude with H, and the standard deviation of estimates also small. It is apparent that the bias of the t estimates is generally more negative than for the VG estimates, but this may be due to the simulation procedure: to show that the t increments are self similar requires two asymptotic arguments, as opposed to one for the VG.

3.2.2. Estimation results for $\mu, \theta \neq 0$. One explanation for the superior performance of PMLE in Section 3.2.1 may be that although our simulated increments were dependent, they were not strongly so. Indeed from (1.14), for $\theta = 0$ the increments are uncorrelated. As such, the product of the univariate densities that we are maximising in the PMLE procedure may be a good approximation to the true multivariate density, at least in regards to parameter estimation. To test this conjecture we simulate some further, more strongly correlated, increments.

First we simulate increments where $\sigma = 0.01$, $\theta = -5 \times 10^{-4}$, $\mu = 10^{-3}$ and $H = \{\text{iid}, 0.75, 0.85, 0.95\}$, with $\alpha = 2.5$ for the VG and $\delta = 2.5$ for the t, where the θ and μ parameters were again chosen to resemble financial data.

Once again we look at the first simulated set of VG increments from each of the 1000 runs in some detail. Table 3.17 gives the descriptive statistics (here the theoretical mean, standard deviation, skewness and kurtosis are given by 5×10^{-4}, 0.01, -0.06 and 4.2 respectively), while Table 3.18 gives the results of parameter estimation.

H	Sample Mean	Sample Std. Deviation	Sample Skewness	Sample Kurtosis
iid	5.8e-4	1.0e-2	0.08	4.45
0.75	5.9e-4	9.5e-3	0.14	4.38
0.85	4.5e-4	9.0e-3	-0.06	4.36
0.95	5.5e-4	9.4e-3	-0.04	3.94

TABLE 3.17. Descriptive statistics of VG simulation runs.

	H	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\mu}$
MOM	iid	1.0e-2	2.07	5.3e-4	5.2e-5
	0.75	9.5e-3	2.19	1.0e-3	-4.1e-4
	0.85	9.0e-3	2.20	-3.9e-4	8.4e-4
	0.95	9.4e-3	3.19	-3.8e-4	9.3e-4
PMLE	iid	1.0e-2	2.50	-7.0e-4	1.3e-3
	0.75	9.5e-3	1.98	1.3e-5	5.8e-4
	0.85	9.0e-3	2.03	-5.7e-4	1.0e-3
	0.95	9.4e-3	2.79	-6.9e-4	1.2e-3
Min χ^2	iid	1.0e-2	2.60	-9.4e-4	1.5e-3
	0.75	9.5e-3	2.06	-8.5e-5	6.5e-4
	0.85	9.0e-3	2.02	-5.5e-4	9.9e-4
	0.95	9.4e-3	2.74	-7.5e-4	1.3e-3

TABLE 3.18. Parameter estimation results of VG simulation runs.

As before, the method of moments generally gives less accurate estimates than the other two estimation methods.

Analogously, descriptive statistics of the first simulated set of t increments from each of the 1000 runs are given in Table 3.19 (here the theoretical mean and standard deviation are given by 5×10^{-4} and 0.01, while the skewness and kurtosis are both infinite). Table 3.20 gives the results of parameter estimation.

H	Sample Mean	Sample Std. Deviation	Sample Skewness	Sample Kurtosis
iid	4.2e-4	9.6e-3	-0.13	4.70
0.75	6.1e-4	1.0e-2	-0.06	6.49
0.85	6.3e-4	1.1e-2	0.25	5.37
0.95	5.2e-4	8.1e-3	0.04	3.23

TABLE 3.19. Descriptive statistics of t simulation runs.

	H	$\hat{\sigma}$	$\hat{\alpha}$	$\hat{\theta}$	$\hat{\mu}$
MOM	iid	9.5e-3	4.01	-8.4e-4	1.3e-3
	0.75	1.1e-2	4.00	-3.4e-4	1.0e-3
	0.85	1.2e-2	4.01	9.2e-4	-2.3e-4
	0.95	8.1e-3	15.19	1.5e-3	-9.5e-4
PMLE	iid	9.6e-3	3.35	-4.3e-4	8.5e-4
	0.75	1.0e-2	2.66	4.4e-4	1.7e-4
	0.85	1.2e-2	2.55	8.2e-4	-1.8e-4
	0.95	8.1e-3	17.12	1.3e-3	-7.8e-4
Min χ^2	iid	9.5e-3	3.35	-3.0e-4	7.6e-4
	0.75	1.0e-2	2.67	6.5e-4	-1.1e-5
	0.85	1.2e-2	2.42	9.1e-4	-2.6e-4
	0.95	8.1e-3	20.99	5.0e-4	5.1e-5

TABLE 3.20. Parameter estimation results of t simulation runs.

For ease of comparison between models over the entire 1000 simulation runs, and given the extra parameters introduced, we include one extra statistic – the average relative measure (ARM). For each model this is given by the average of the absolute value of the MAD score, bias or standard deviation of each parameter divided by the absolute value of the true value of the parameter. So for example, the ARM statistic for MAD scores for VG fitted models is given by

$$\frac{1}{4}\left(\left|\frac{\text{MAD}(\hat{\sigma})}{\sigma}\right| + \left|\frac{\text{MAD}(\hat{\alpha})}{\alpha}\right| + \left|\frac{\text{MAD}(\hat{\theta})}{\theta}\right| + \left|\frac{\text{MAD}(\hat{\mu})}{\mu}\right|\right)$$

(replacing α with δ for t models, and MAD$(\cdot)$ with Bias$(\cdot)$ or SD$(\cdot)$ for the ARM applied to the bias or standard deviation). For the MAD, bias and standard deviation, the ARM number is intended as a simple measure of overall distance from zero, and therefore goodness of fit.

The results of the simulation runs are presented in Tables 3.21 to 3.26 below. Note that for the t, $\delta = 2.5$ implies infinite skewness and kurtosis (equations (1.28) and (1.29)), whereas to use the MOM we must assume that these are finite, as 4 moments are needed in order to identify the 4 model parameters. As such, we must impose $\hat{\delta} > 4$ as a constraint when we numerically minimise (3.1) and determine our MOM parameter estimates. Estimating a model that assumes $\hat{\delta} > 4$ when in fact $\delta = 2.5$ of course results in a bad fit.

	H	σ	α	θ	μ	ARM
MOM	iid	1.5e-4	0.53	6.2e-4	6.1e-4	0.52
	0.75	5.7e-4	0.69	7.0e-4	6.7e-4	0.60
	0.85	9.7e-4	0.97	7.0e-4	6.6e-4	0.64
	0.95	1.8e-3	4.66	1.3e-3	1.3e-3	1.48
PMLE	iid	2.5e-4	0.31	5.2e-4	4.9e-4	0.42
	0.75	6.8e-4	0.43	6.4e-4	6.0e-4	0.53
	0.85	1.1e-3	0.67	6.6e-4	6.2e-4	0.58
	0.95	1.8e-3	3.54	1.1e-3	1.1e-3	1.24
Min χ^2	iid	1.7e-4	0.41	6.5e-4	6.3e-4	0.53
	0.75	5.7e-4	0.45	5.5e-4	5.1e-4	0.46
	0.85	9.7e-4	0.68	6.0e-4	5.5e-4	0.53
	0.95	1.8e-3	3.61	1.1e-3	1.1e-3	1.23

TABLE 3.21. MAD statistics for VG models with θ small.

	H	σ	α	θ	μ	ARM
MOM	iid	-1.7e-5	0.20	-3.2e-5	3.6e-5	0.05
	0.75	-4.1e-5	0.44	-3.0e-5	3.1e-5	0.07
	0.85	-1.3e-4	0.80	1.3e-6	1.2e-5	0.09
	0.95	-2.4e-4	4.63	-7.1e-5	6.9e-5	0.52
PMLE	iid	9.5e-5	0.05	-2.8e-5	-5.2e-5	0.03
	0.75	8.0e-5	0.25	-7.2e-5	-1.1e-4	0.09
	0.85	-3.4e-5	0.55	-6.1e-5	-1.1e-4	0.11
	0.95	-2.0e-4	3.52	-1.2e-4	-7.2e-5	0.44
Min χ^2	iid	-3.4e-5	0.14	-1.6e-4	1.7e-4	0.14
	0.75	-4.0e-5	0.25	1.8e-5	-1.5e-5	0.04
	0.85	-1.3e-4	0.54	1.8e-5	-5.4e-6	0.07
	0.95	-2.5e-4	3.58	-6.4e-5	6.3e-5	0.41

TABLE 3.22. Bias of estimates from VG models with θ small.

In fact the results for the small θ simulations are substantially similar to those for the $\theta = 0$ case. That is, the MOM is almost always the worst method, and in most cases PMLE does better than minimum χ^2 fitting. Note however that the MAD scores and standard deviation for the θ parameter are generally of the same order of magnitude as the parameter itself, as opposed to all other parameters for which the MAD scores and standard deviation are generally 1 to 2 orders of magnitude smaller. This indicates that estimation of the θ parameter in practice, especially if it is quite small, is likely to be very imprecise.

	H	σ	α	θ	μ	ARM
MOM	iid	1.8e-4	0.67	7.8e-4	7.6e-4	0.65
	0.75	7.2e-4	0.88	8.7e-4	8.4e-4	0.75
	0.85	1.2e-3	1.66	8.9e-4	8.5e-4	0.85
	0.95	2.2e-3	10.51	2.7e-3	2.7e-3	3.11
PMLE	iid	1.0e-3	0.40	1.2e-3	1.3e-3	0.99
	0.75	1.2e-3	0.58	1.6e-3	1.7e-3	1.34
	0.85	1.5e-3	1.02	1.7e-3	1.7e-3	1.40
	0.95	2.3e-3	4.07	2.2e-3	2.3e-3	2.13
Min χ^2	iid	3.7e-4	1.38	2.8e-3	2.8e-3	2.22
	0.75	7.2e-4	0.63	6.9e-4	6.5e-4	0.59
	0.85	1.2e-3	0.88	7.7e-4	7.2e-4	0.69
	0.95	2.2e-3	4.29	2.2e-3	2.2e-3	2.10

TABLE 3.23. Standard deviation of estimates from VG models with θ small.

	H	σ	δ	θ	μ	ARM
MOM	iid	4.7e-4	1.53	1.4e-3	1.4e-3	1.24
	0.75	7.7e-4	1.54	1.4e-3	1.4e-3	1.22
	0.85	1.2e-3	1.62	1.3e-3	1.2e-3	1.16
	0.95	2.3e-3	3.51	1.2e-3	1.1e-3	1.27
PMLE	iid	1.9e-4	0.20	3.7e-4	3.3e-4	0.29
	0.75	6.6e-4	0.26	3.7e-4	3.3e-4	0.31
	0.85	1.1e-3	0.47	4.2e-4	3.7e-4	0.38
	0.95	2.2e-3	2.95	7.4e-4	6.7e-4	0.89
Min χ^2	iid	2.1e-4	0.23	4.0e-4	3.5e-4	0.32
	0.75	6.6e-4	0.32	4.2e-4	3.6e-4	0.35
	0.85	1.2e-3	0.54	4.6e-4	4.0e-4	0.41
	0.95	2.2e-3	3.26	8.3e-4	7.6e-4	0.99

TABLE 3.24. MAD statistics for t models with θ small.

In regards to the PMLE beating minimum χ^2 fitting, this may be due to the θ and μ parameters chosen being quite small, and as such not increasing the degree of autocorrelation between increments significantly. Accordingly, we simulate one further realisation with substantially stronger autocorrelation, choosing parameters $\sigma = 0.01$, $\theta = -5 \times 10^{-2}$, $\mu = 10^{-3}$ and $H = 0.95$, with $\alpha = 2.5$ for the VG and $\delta = 2.5$ for the t. Results of the simulation are given in Tables 3.27 to 3.29.

In this case the minimum χ^2 method does give a more accurate and precise estimate of α or δ (but not the other parameters) than the PMLE, and indeed is more accurate and precise

	H	σ	δ	θ	μ	ARM
MOM	iid	9.5e-5	1.53	-8.2e-4	8.2e-4	0.77
	0.75	1.0e-5	1.54	-7.3e-4	7.3e-4	0.70
	0.85	-5.3e-6	1.62	-6.5e-4	6.5e-4	0.65
	0.95	-3.7e-4	3.51	-3.8e-4	3.8e-4	0.64
PMLE	iid	-1.3e-5	0.04	-1.1e-5	1.4e-5	0.01
	0.75	-8.5e-5	0.14	-9.7e-8	7.8e-6	0.02
	0.85	-1.0e-4	0.40	-2.9e-5	3.1e-5	0.06
	0.95	-3.7e-4	2.93	-4.8e-5	4.8e-5	0.34
Min χ^2	iid	-2.6e-5	0.05	-1.2e-5	1.6e-5	0.02
	0.75	-1.1e-4	0.19	-1.2e-5	1.9e-5	0.03
	0.85	-1.2e-4	0.46	-4.4e-5	4.5e-5	0.08
	0.95	-3.9e-4	3.24	-5.4e-5	5.2e-5	0.37

TABLE 3.25. Bias of estimates from t models with θ small.

	H	σ	δ	θ	μ	ARM
MOM	iid	8.8e-4	0.13	2.2e-3	2.1e-3	1.66
	0.75	1.0e-3	0.16	2.0e-3	1.9e-3	1.51
	0.85	1.6e-3	0.65	1.9e-3	1.9e-3	1.52
	0.95	2.9e-3	3.55	1.7e-3	1.6e-3	1.69
PMLE	iid	2.4e-4	0.26	4.7e-4	4.2e-4	0.37
	0.75	8.2e-4	0.32	4.8e-4	4.3e-4	0.40
	0.85	1.4e-3	0.53	5.2e-4	4.6e-4	0.46
	0.95	2.7e-3	3.30	9.7e-4	9.0e-4	1.11
Min χ^2	iid	2.6e-4	0.30	5.1e-4	4.5e-4	0.41
	0.75	8.3e-4	0.40	5.3e-4	4.7e-4	0.44
	0.85	1.4e-3	0.64	5.7e-4	5.1e-4	0.51
	0.95	2.7e-3	4.03	1.1e-3	1.1e-3	1.31

TABLE 3.26. Standard deviation of estimates from t models with θ small.

	Model	σ	α or δ	θ	μ	ARM
MOM	VG	9.4e-3	6.22	2.8e-2	2.0e-2	5.89
PMLE	VG	5.3e-3	4.78	2.3e-2	1.2e-2	3.73
Min χ^2	VG	1.6e-2	2.03	4.4e-2	4.3e-2	11.59
MOM	t	2.9e-2	4.10	3.7e-2	2.4e-2	7.39
PMLE	t	3.7e-3	3.28	2.1e-2	7.4e-3	2.38
Min χ^2	t	2.9e-2	2.13	3.5e-2	2.9e-2	8.32

TABLE 3.27. MAD statistics for models with θ large.

than when it was applied to less correlated increments. This latter increase in accuracy is

	Model	σ	α or δ	θ	μ	ARM
MOM	VG	-6.8e-3	6.17	-1.3e-2	1.3e-2	4.02
PMLE	VG	-3.8e-3	4.77	-9.1e-3	8.9e-3	2.84
Min χ^2	VG	1.4e-2	1.56	4.0e-2	-3.6e-2	9.77
MOM	t	2.0e-2	4.10	5.1e-3	2.6e-3	1.57
PMLE	t	3.8e-4	3.24	9.5e-4	-3.5e-4	0.43
Min χ^2	t	2.8e-2	0.80	1.9e-2	-2.1e-2	6.24

TABLE 3.28. Bias of estimates from models with θ large.

	Model	σ	α or δ	θ	μ	ARM
MOM	VG	7.0e-3	7.11	3.9e-2	2.5e-2	7.38
PMLE	VG	4.5e-3	4.69	3.1e-2	1.5e-2	4.46
Min χ^2	VG	1.7e-2	3.54	3.4e-2	4.5e-2	12.17
MOM	t	4.7e-2	4.62	5.6e-2	4.0e-2	11.84
PMLE	t	4.7e-3	4.46	2.7e-2	1.0e-2	3.26
Min χ^2	t	2.9e-2	3.96	4.2e-2	3.0e-2	8.96

TABLE 3.29. Standard deviation of estimates from models with θ large.

likely due to the larger θ value, which gives τ_t more prominence in the realised X_t readings (c.f., Equation (1.6)). Once again the MOM performed worst in estimating the α or δ parameter.

The parameters chosen in this last case are probably not particularly realistic values of actual stock price processes however. This leads us to the conclusion that, unless one has particularly skewed and/or correlated data, and is particularly interested in the α or δ parameter, PMLE estimation is probably the best of these three methods to use. This is somewhat surprising given that in the non-independence case PMLE has theoretical drawbacks compared with the other two methods so far considered. Given that MLE is the superior estimation method for iid data, it would appear that, at least for the cases considered, the dependence structure introduced did not render the true multivariate density sufficiently different from the product of marginal densities to cause either of the other methods to perform better than PMLE. Note however that one *would* expect full multivariate maximum likelihood estimation to outperform PMLE, as such estimation would take into account the dependence of squared returns.

3.3. The ECF method

In addition to the three methods described above, we also estimate using the empirical characteristic function technique. The ECF technique has the advantage of utilising the joint distribution of $\{X_t\}$, which depends on the H parameter and should aid in estimation, and is based on the joint characteristic function of readings, which unlike the joint pdf, we know in the case of the VG. Estimation via the ECF is performed similarly to Yu (2004) for the VG only. Note that the asymptotic results regarding ECF estimation in Yu (2004) – that ECF estimates are consistent and asymptotically normal – apply only to iid processes and processes following a form of weak dependence (SRD). In particular, the results do not necessarily carry over to LRD case, which is the case we are concerned with (see Yu (2004), Section 3.1 for the list of assumptions needed to ensure a consistent and asymptotically normal estimator). To perform ECF estimation we numerically minimise

$$\int_{\mathbb{R}^p} |\phi_{\text{VG}}(\boldsymbol{\omega}) - \phi_{\text{ECF}}(\boldsymbol{\omega})|^2 e^{-\sum_{j=1}^p |\omega_j|} \mathrm{d}^p\boldsymbol{\omega} \tag{3.5}$$

where $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_p)'$, $e^{-\sum_{j=1}^p |\omega_j|}$ is an exponential weighting function,

$$\phi_{\text{VG}}(\boldsymbol{\omega}) = e^{i\boldsymbol{\omega}'\boldsymbol{\mu}} |I - \frac{1}{\alpha}(\Sigma^{\frac{1}{2}} \text{diag}(i\theta\boldsymbol{\omega} - \frac{\sigma^2}{2}\boldsymbol{\omega^2})\Sigma^{\frac{1}{2}})|^{-\alpha} \tag{3.6}$$

is the joint cf of $\{X_1, \cdots, X_p\}$ with $\boldsymbol{\mu}$ a p-vector of μ's and $\Sigma^{\frac{1}{2}}$ a $p \times p$ matrix, and

$$\phi_{\text{ECF}}(\boldsymbol{\omega}) = \frac{1}{n-p+1} \sum_{k=0}^{n-p} e^{i\sum_{j=1}^p \omega_j X_{k+j}}$$

is the joint p-dimensional ECF of the readings $X_1, \cdots, X_n$. In fact ECF estimation belongs to the class of GMM estimators as it minimises a distance between sample and population moments, as given respectively by the ECF and cf evaluated at different $\boldsymbol{\omega}$ values. To actually perform ECF estimation we choose $p = 6$ and numerically minimise (3.5) as a function of $\sigma, \alpha, \theta, \mu$ and H. To calculate (3.5) for a given set of parameters we integrate numerically using a 15,625 point quadrature for the $\mu = \theta = 0$ case and a 46,656 point quadrature for the $\mu,\ \theta \neq 0$ case. Larger p values capture more of the joint structure of $\{X_t\}$, at the expense of slower computation, and the choice $p = 6$ was made with this trade-off in mind.

3.3.1. Estimation results for the ECF method. Estimation results for the symmetric VG model using the same 1000 runs of $\sigma = 0.01$, $H = \{\text{iid}, 0.75, 0.85, 0.95\}$ and $\alpha = 2.5$ data as in Section 3.2.1, fitted via the ECF method, are given in Tables 3.30 to 3.32 below. In this case, in addition to detailing results for the $\hat{H}$ estimates produced via the ECF procedure, we give for comparison results for $\hat{H}$ as estimated via the least squares procedure detailed in Section 3.1.1 (these are given in the H_{LS} column of the tables). Tables 3.30 to 3.32 can be compared with Tables 3.5 to 3.7 in Section 3.2.1.

Results for the skew VG model using the same 1000 runs of $\sigma = 0.01$, $\theta = -5 \times 10^{-4}$, $\mu = 10^{-3}$, $H = \{\text{iid}, 0.75, 0.85, 0.95\}$ and $\alpha = 2.5$ data as in Section 3.2.2 are given in Tables 3.33 to 3.35, which can be compared with Tables 3.21 to 3.23 in Section 3.2.2. We provide no H_{LS} column as estimation of H via a procedure similar to that detailed in Section 3.1.1, although possible for the skew case, is quite involved.

H	σ	α	H	H_{LS}
iid	1.4e-4	2.48	0.10	0.03
0.75	5.4e-4	2.68	0.18	0.06
0.85	1.0e-3	2.21	0.24	0.10
0.95	1.8e-3	2.39	0.32	0.21

TABLE 3.30. MAD statistics for symmetric VG models fitted via ECF.

H	σ	α	H	H_{LS}
iid	-2.2e-6	0.85	0.10	0.03
0.75	-3.4e-5	0.75	-0.11	-0.01
0.85	-8.6e-5	0.32	-0.21	-0.09
0.95	-2.0e-4	0.42	-0.31	-0.21

TABLE 3.31. Bias for symmetric VG models fitted via ECF.

H	σ	α	H	H_{LS}
iid	1.8e-4	4.65	0.15	0.05
0.75	6.8e-4	4.96	0.16	0.08
0.85	1.3e-3	4.25	0.16	0.08
0.95	2.2e-3	4.10	0.18	0.09

TABLE 3.32. Standard deviation for symmetric VG models fitted via ECF.

H	σ	α	θ	μ	H	ARM
iid	3.9e-4	5.55	1.3e-3	1.3e-3	0.21	1.65
0.75	8.3e-4	3.97	1.2e-3	1.2e-3	0.18	1.36
0.85	1.1e-3	3.42	1.1e-3	1.1e-3	0.20	1.25
0.95	1.9e-3	5.24	1.1e-3	1.1e-3	0.24	1.43

TABLE 3.33. MAD statistics for skew VG models fitted via ECF.

H	σ	α	θ	μ	H	ARM
iid	-2.6e-4	3.86	5.4e-4	-5.4e-4	0.21	0.90
0.75	-3.1e-4	2.23	4.8e-4	-4.8e-4	-0.01	0.59
0.85	-2.9e-4	1.85	5.4e-4	-5.2e-4	-0.12	0.62
0.95	-3.9e-4	3.72	5.1e-4	-5.1e-4	-0.21	0.81

TABLE 3.34. Bias for skew VG models fitted via ECF.

H	σ	α	θ	μ	H	ARM
iid	1.5e-3	12.35	2.3e-3	2.3e-3	0.20	3.09
0.75	1.7e-3	9.24	2.0e-3	2.0e-3	0.20	2.52
0.85	1.7e-3	7.76	1.8e-3	1.8e-3	0.20	2.22
0.95	2.5e-3	9.42	1.7e-3	1.6e-3	0.20	2.23

TABLE 3.35. Standard deviation for skew VG models fitted via ECF.

Contrary to expectations, the results indicate that in these two cases ECF estimation performs worse than the simpler methods considered previously, particularly with regards to the α and H parameters. Focusing first on the symmetric models, while the MSE, bias and standard deviation of the σ estimates produced via ECF estimation are of roughly equal magnitude to those from estimates produced by the three methods considered earlier, results regarding estimation of α are generally worse (except for the $H = 0.95$ case), and results regarding estimation of H are uniformly worse. Estimation results for the skew models are similar: the MAD, bias and standard deviation of σ estimates produced by the ECF method are of roughly equal magnitude to those from estimates produced by the three methods considered earlier; results regarding the α parameter are much worse than earlier, except for the $H = 0.95$ case where the MAD, bias and standard deviation are of a similar magnitude; and results regarding estimation of θ and μ are also typically worse than earlier.

As noted above, ECF estimation does perform relatively better in the $H = 0.95$ case, which corresponds to the strongest correlation case and therefore would benefit most from a procedure like the ECF which incorporates the correlation structure, but even here the estimates produced are quite inaccurate and imprecise. As mentioned earlier, in order to keep calculations tractable we use only a 6-dimensional cf in estimation, which may limit the degree of the joint structure captured. The extra complexity, data manipulation and numerical integration involved in working with the joint cf also introduces more margin for rounding error, and is more removed from the data than the previous methods.

Given the poor performance of ECF estimation recorded here we shall not proceed with it past this section, although we will derive the joint cf $\phi_{\text{VG}}(\boldsymbol{\omega})$ given by (3.6).

Similar to Section 3.1, let $\boldsymbol{\tau}_{\text{VG}}$ be a p-vector of correlated random variables with marginal $\Gamma(\alpha, \alpha)$ distribution. If $\mathbf{Z}$ is a vector of uncorrelated standard normals, then $\mathbf{X}_{\text{VG}} = \boldsymbol{\mu} + \theta\boldsymbol{\tau}_{\text{VG}} + \sigma\sqrt{\boldsymbol{\tau}_{\text{VG}}} \otimes \mathbf{Z}$ describes our observed log price increments. Here $\sqrt{\boldsymbol{\tau}_{\text{VG}}}$ is the element-wise square-root of $\boldsymbol{\tau}_{\text{VG}}$, and '$\otimes$' gives element-wise multiplication. The characteristic function of $\mathbf{X}_{\text{VG}}$ is then given by

$$\phi_{\text{VG}}(\boldsymbol{\omega}) = \mathbb{E}(e^{i\boldsymbol{\omega}'\mathbf{X}_{\text{VG}}}) = e^{i\boldsymbol{\omega}'\boldsymbol{\mu}}\mathbb{E}(e^{(i\theta\boldsymbol{\omega} - \frac{\sigma^2}{2}\boldsymbol{\omega}^{\mathbf{2}})'\boldsymbol{\tau}_{\text{VG}}}) \tag{3.7}$$

where $\boldsymbol{\omega}^{\mathbf{2}}$ is the element-wise square of $\boldsymbol{\omega}$, and the right-hand side of (3.7) follows by conditioning on $\boldsymbol{\tau}_{\text{VG}}$, since $\mathbf{Z} \sim N(\mathbf{0}, I)$. Now $\boldsymbol{\tau}_{\text{VG}} = (\boldsymbol{\eta}_1^2 + \cdots + \boldsymbol{\eta}_\nu^2)/\nu$ where for $i = 1, \ldots, \nu$, $\boldsymbol{\eta}_i = \Sigma^{\frac{1}{2}}\mathbf{Z}_i$, each $\mathbf{Z}_i$ is an independent vector of independent standard normals, and $\boldsymbol{\eta}_i^2$ is the element-wise square of $\boldsymbol{\eta}_i$. As such, for $\mathbf{w}' = (w_1, \cdots, w_p)$,

$$\mathbf{w}'\boldsymbol{\tau}_{\text{VG}} = \sum_{i=1}^{\nu}(w_1\eta_{i,1}^2 + \cdots + w_p\eta_{i,p}^2)/\nu = \sum_{i=1}^{\nu}(\mathbf{Z}_i'\Sigma^{\frac{1}{2}}W\Sigma^{\frac{1}{2}}\mathbf{Z}_i)/\nu$$

where $W = \text{diag}(\mathbf{w})$. Therefore (see for example Chapter 12 of Hogg & Craig (1978))

$$\mathbb{E}(e^{\mathbf{w}'\boldsymbol{\tau}_{\text{VG}}}) = (\mathbb{E}(e^{\mathbf{Z}_1'(\Sigma^{\frac{1}{2}}W\Sigma^{\frac{1}{2}})\mathbf{Z}_1/\nu}))^\nu = |I - \frac{1}{\alpha}(\Sigma^{\frac{1}{2}}W\Sigma^{\frac{1}{2}})|^{-\alpha}$$

for $\alpha = \frac{\nu}{2}$, so that

$$\phi_{\text{VG}}(\boldsymbol{\omega}) = e^{i\boldsymbol{\omega}'\boldsymbol{\mu}}|I - \frac{1}{\alpha}(\Sigma^{\frac{1}{2}}\text{diag}(i\theta\boldsymbol{\omega} - \frac{\sigma^2}{2}\boldsymbol{\omega}^{\mathbf{2}})\Sigma^{\frac{1}{2}})|^{-\alpha}.$$

Finally, recall that $\Sigma_{ij} = (1 + (i-j)^2)^{-\frac{\gamma}{2}}$ for $H = 1 - \gamma$ from Section 3.1, so it is via Σ that H enters into $\phi_{\text{VG}}(\boldsymbol{\omega})$.

3.3.2. Other characteristic function-based approaches. Before moving on we note that Madan & Seneta (1987a) also consider the problem of estimation of parameters on simulated data via an ECF procedure, although only in the symmetric, iid increments case. They report mixed results, with two of the five parametric classes considered being essentially inestimable (the normal compound Poisson and Brownian compound Poisson distributions), two being well estimated (the normal and stable distributions), and one, the VG, being in one case estimable and in one case not.

Related work is contained in Madan & Seneta (1987b) and Madan & Seneta (1989), where instead of ECF estimation on the original $\{X_t\}$ increments, PMLE is suggested on a transformed variable Y_t, given by $Y_t = \cos(uX_t)$ or $Y_t = uX_t(\text{mod } 2\pi)$ for u a parameter to be chosen. Madan & Seneta (1987b) show that for X_t symmetric, the density of Y_t in the $Y_t = \cos(uX_t)$ transform case is given by

$$g(y) = \pi^{-1}(1-y^2)^{-1/2}\sum_{n=0}^{\infty} 2^n\phi(nu)q_n(y),\ y \in (-1,1), \tag{3.8}$$

where $\phi(\cdot)$ is the (real) characteristic function of X_t, $q_0(y) = 1$ and $q_n(y) = T_n(y)/2^{n-1}$, $n \geq 1$, for $T_n(y)$ the Chebyshev polynomial of the first kind of degree n (see Szegö (1959)). Thus one can express the density of the transformed variable $Y_t = \cos(uX_t)$ as a series involving evaluations of the characteristic function of X_t multiplied by Chebyshev polynomials. This method therefore allows for near-PMLE to be performed on symmetric distributions for which the cf is available but the pdf is not, which is carried out by truncating (3.8) for some n, $n = 30$ for example. Regarding the transform $Y_t = uX_t(\text{mod } 2\pi)$, Madan &

Seneta (1989) show that the density of Y_t in this case is given by

$$g(y) = \frac{1}{2\pi} + \pi^{-1} \sum_{n=1}^{\infty} (\Re\{\phi(nu)\} \cos(nux) + \Im\{\phi(nu)\} \sin(nux)) \tag{3.9}$$

which is valid for X_t with a not necessarily symmetric distribution. Here $\Re\{z\}$ gives the real part of the complex number z, while $\Im\{z\}$ gives the imaginary part.

We perform this Madan & Seneta ('MS') estimation using the $Y_t = \cos(uX_t)$ transform on the 1000 runs of symmetric VG simulated data (that is, simulated data with $\sigma = 0.01$, $H = \{\text{iid}, 0.75, 0.85, 0.95\}$ and $\alpha = 2.5$) as in Section 3.2.1 and 3.3.1, using a truncation of $n = 30$ in (3.8). We also perform MS estimation using the $Y_t = uX_t (\text{mod } 2\pi)$ transform on the 1000 runs of skew VG simulated data (with $\sigma = 0.01$, $\theta = -5 \times 10^{-4}$, $\mu = 10^{-3}$, $H = \{\text{iid}, 0.75, 0.85, 0.95\}$ and $\alpha = 2.5$) as in Section 3.2.2 and 3.3.1, this time using a truncation of $n = 100$ in (3.9). In both cases we choose $u = 50$, which corresponds to the original papers' recommendation to take $u = 0.5$ for standardised data with unit variance. Both sets of results are given in Tables 3.36 to 3.38 below.

	$\theta = 0$ case		$\theta \neq 0$ case				
H	$\sigma = 0.01$	$\alpha = 2.5$	$\sigma = 0.01$	$\alpha = 2.5$	$\theta = -5 \times 10^{-4}$	$\mu = 10^{-3}$	ARM
iid	2.3e-4	0.52	2.3e-4	0.53	8.0e-4	7.7e-4	0.65
0.75	6.9e-4	0.69	7.2e-4	0.77	9.0e-4	8.7e-4	0.76
0.85	1.2e-3	1.27	1.1e-3	1.24	1.1e-3	1.0e-3	0.96
0.95	1.9e-3	5.48	1.9e-3	5.46	2.1e-3	2.1e-3	2.16

TABLE 3.36. MAD statistics for symmetric and skew VG models fitted via the Madan & Seneta ECF procedure.

	$\theta = 0$ case		$\theta \neq 0$ case				
H	$\sigma = 0.01$	$\alpha = 2.5$	$\sigma = 0.01$	$\alpha = 2.5$	$\theta = -5 \times 10^{-4}$	$\mu = 10^{-3}$	ARM
iid	-1.0e-6	0.16	-2.3e-5	0.15	-2.0e-5	1.9e-5	0.03
0.75	-4.4e-5	0.44	-9.6e-5	0.52	1.1e-5	-2.1e-5	0.07
0.85	-1.1e-4	1.11	-1.9e-4	1.05	-3.1e-5	3.8e-5	0.14
0.95	-2.5e-4	5.45	-3.3e-4	5.43	-1.5e-4	1.5e-4	0.67

TABLE 3.37. Bias for symmetric and skew VG models fitted via the Madan & Seneta ECF procedure.

	$\theta = 0$ case		$\theta \neq 0$ case				
H	$\sigma = 0.01$	$\alpha = 2.5$	$\sigma = 0.01$	$\alpha = 2.5$	$\theta = -5 \times 10^{-4}$	$\mu = 10^{-3}$	ARM
iid	2.9e-4	0.73	2.9e-4	0.72	1.0e-3	9.8e-4	0.83
0.75	8.7e-4	1.03	8.9e-4	1.25	1.2e-3	1.2e-3	1.07
0.85	1.4e-3	2.12	1.4e-3	2.14	2.4e-3	2.4e-3	2.07
0.95	2.3e-3	6.17	2.4e-3	6.48	3.9e-3	3.9e-3	3.61

TABLE 3.38. Standard deviation for symmetric and skew VG models fitted via the Madan & Seneta ECF procedure.

Tables 3.36 to 3.38 can be compared with Tables 3.30 to 3.32 and 3.33 to 3.35, which give statistics for the symmetric and skew VG models fitted via the ECF procedure, as well as Tables 3.5 to 3.7 and 3.21 to 3.23, which respectively display statistics for the symmetric and skew VG models, fitted via the MOM, PMLE and minimum χ^2 procedures. For both the symmetric and skew fits using the MS procedure, the MAD scores achieved are comparable, although slightly worse than, the MAD scores achieved by the MOM in Sections 3.2.1 and 3.2.2, and in all but the $H = 0.95$ case the MS procedure does better than the ECF method from Section 3.3.1. The MS procedure is however beaten by the PMLE and minimum χ^2 estimation methods, as may be expected – the procedure gives a way to perform near-PMLE estimation when one does not know the likelihood function, as was the case with the VG at the time of publication of the two papers mentioned. For the VG, the MS procedure has now been somewhat superseded since the density function in now available and actual PMLE can be performed. The procedure may still be of value in cases where the characteristic function is known but the density is not, such as with the stable laws for example.

3.4. Rejection of models

In practice we would like to be able to choose, based on a χ^2 goodness of fit statistic, between competing models for a given set of data. To assess the feasibility of this, for each group of 1000 simulation runs, each of length 2500, generated from the $\sigma = 0.01,\ \theta = -5 \times 10^{-4},\ \mu = 10^{-3},\ H = \{\text{iid}, 0.75, 0.85, 0.95\}$ and $\alpha = 2.5$ for the VG or $\delta = 2.5$ for

the t models, we fit both VG and t models to the same data and compare results. The χ^2 statistic we base our comparison on has the same functional form as (3.2), but here we use a single set of fixed bins for all data sets (we take the first bin as $(-\infty, -0.025]$, the next 10 as 1.55×10^{-3} units wide and lying between -0.025 and -0.0095, the next 39 as 5.13×10^{-4} units wide and lying between -0.0095 and 0.0105, the next 10 as 1.55×10^{-3} units wide and lying between 0.0105 and 0.0260, and the last bin as $(0.0260, \infty)$).

For example, for the 1000 simulation runs generated using a VG model with σ, θ and μ as above, $\alpha = 2.5$ and $H = 0.75$ say, we fit the VG model to each simulation, calculate and record the 1000 values of the χ^2 goodness of fit statistic, and from these empirically determine the average and the 95% quantile value of the distribution of this χ^2 statistic. We then fit a t model to the same 1000 simulated runs, calculate and record the 1000 values of the χ^2 goodness of fit statistic, and determine what proportion of these χ^2 values (fitting an inappropriate t model to VG simulated data) exceed the recorded 'correct model' 95% quantile. This gives the column with entries (55.7, 73.1, 9.4%) in Table 3.39. Then we obtain 1000 simulation runs generated using a t model with σ, θ and μ as above, $\delta = 2.5$ and $H = 0.75$ say, obtain the empirical 95% quantile, use the same simulated 1000 t-runs to fit the VG, and determine what proportion of inappropriately fitted VG models exceed the empirical 't generated' 95% quantile. This gives the column with entries (56.1, 74.9, 11.5%) in Table 3.39. The procedure for PMLE fitted models is similar, with the χ^2 statistic associated with each fit obtained using the PMLE estimated parameters in (3.2).

Under a null hypothesis that the observations are iid, the χ^2 statistic would asymptotically have a χ^2_n distribution, where n is the degrees of freedom. Here we expect the degrees of freedom of the χ^2 statistic to be $61-5 = 56$ since in each case 4 parameters are estimated, and the χ^2 calculation is based on 61 bins. Since the expected value of a χ^2_n distributed random variable is n, the average of our 1000 calculated values under the null hypothesis would reflect the degrees of freedom.

	H	iid	iid	0.75	0.75	0.85	0.85	0.95	0.95
	Correct Model	VG	t	VG	t	VG	t	VG	t
Min χ^2	Average χ^2 (CM)	56.4	56.2	55.7	56.1	56.2	56.1	56.0	57.1
	95^{th} percentile (CM)	75.9	74.1	73.1	74.9	74.5	74.4	73.6	75.1
	Percent rejected (IM)	7.9%	12.7%	9.4%	11.5%	9.5%	11.8%	7.0%	8.2%
PMLE	Average χ^2 (CM)	55.9	56.5	55.8	56.4	57.2	56.2	55.9	56.3
	95^{th} percentile (CM)	76.7	74.5	75.4	75.6	75.6	73.9	74.2	74.5
	Percent rejected (IM)	8.1%	15.7%	7.9%	13.9%	8.8%	12.8%	6.9%	8.1%

TABLE 3.39. Average χ^2 values and 95^{th} percentiles for correct models (CM), and per cent of incorrect models (IM) rejected, fitted via minimum χ^2 and PMLE.

Although in fact the underlying data are not independent for each case of H considered, nor are the parameters estimated in the conventional way, we see from the averages in Table 3.39 that it is nevertheless reasonable to use in every case the χ^2_{56}, or perhaps more conservatively the χ^2_{57}, 95% quantile value as cut-off to measure the proportion of incorrect models rejected. This is done in Table 3.40, since $\chi^2_{56}(0.95) = 74.5$, and $\chi^2_{57}(0.95) = 75.6$. The percentages rejected are similar to those in Table 3.39.

In fact the use of conventional χ^2 quantile values is a conservative choice – Gleser & Moore (1983) show that the null distribution of the χ^2 statistic is stochastically larger in the positively correlated case as compared to the iid case. In particular, Gleser & Moore show that under quite general conditions, the vector with i^{th} entry $(O^*_i - E^*_i)/\sqrt{E^*_i}$, $i = 1, \cdots, 61$, for O^* and E^*_i defined as in (3.2), has a limiting $N(\mathbf{0}, S)$ distribution in the positively correlated case and limiting $N(\mathbf{0}, S_{\text{IID}})$ distribution in the iid case, with $S - S_{\text{IID}}$ positive semi-definite. Hence positive dependence is confounded with lack of fit, and the χ^2 goodness of fit test will too often reject the null hypothesis if the underlying data are dependent.

From both Table 3.39 and 3.40, one of the things that we can say is that it appears very hard to distinguish between VG and t models, as both are flexible enough to accommodate

	H	iid	iid	0.75	0.75	0.85	0.85	0.95	0.95
	Correct Model	VG	t	VG	t	VG	t	VG	t
Min χ^2	Cutoff: 74.5	9.5%	11.3%	7.7%	12.2%	9.5%	11.7%	6.2%	9.0%
	Cutoff: 75.6	8.2%	9.6%	6.4%	10.4%	7.9%	9.7%	4.4%	7.0%
PMLE	Cutoff: 74.5	10.4%	15.7%	9.0%	16.6%	10.5%	11.7%	6.9%	8.1%
	Cutoff: 75.6	9.5%	13.1%	7.4%	13.9%	8.8%	9.7%	4.5%	6.1%

TABLE 3.40. Per cent of incorrect models rejected, using χ^2_{56} and χ^2_{57} cutoffs.

data from the other distribution. This result accords with Fung & Seneta (2007), which compared the symmetric VG and t distributions, and found the two models almost impossible to distinguish on the basis of tailweight, by choice of compatible parameters. Beyond this, it appears slightly easier to reject a VG fitted to t data than a t fitted to VG data.

Secondly, it emerges from the above that whatever the value of H in our postulated model, either a VG or a t-process, and whichever of our estimation methods for parameters is used, the conventional $\chi^2_{56}(0.95)$ cut off value can be used for rejection of the postulated model when testing at 5% significance level.

3.5. Fitting to financial data

We now fit the VG and t models to some financial data sets and assess the fit. The aim is not to decide between models by exhaustively fitting every stock in the S&P 500 for example, but to put the prior discussion into action on a select few series.

The data are end of day prices, between 1 January 1996 and 31 December 2005, on the S&P 500 index, Microsoft Corporation's share price, the Australian dollar/US dollar exchange rate, and the spot price of West Texas Intermediate crude oil. Each data set has length of around 2500 readings, with the exact figure varying with the number of non-trading days in the 10-year period, and determined by the particular asset and the exchange on which

it is traded. Some descriptive statistics are given in Table 3.41, the results for minimum χ^2 fitting are given in Table 3.42, while those for the PMLE are in Table 3.43.

In both, 'χ^2' is the χ^2 goodness of fit statistic computed according to bins as in Section 3.4.

	Sample Size	Sample Mean	Sample Std. Deviation	Sample Skewness	Sample Kurtosis
S&P 500	2518	2.8e-4	1.2e-2	-0.09	5.95
Microsoft	2518	6.1e-4	2.3e-2	-0.12	8.32
AUD/US	2609	-5.3e-6	6.7e-3	0.06	5.24
WTI	2503	4.5e-4	2.5e-2	-0.28	6.35

TABLE 3.41. Descriptive statistics of data.

Model	Security	$\hat{\sigma}$	$\hat{\alpha}$ or $\hat{\delta}$	$\hat{\theta}$	$\hat{\mu}$	χ^2
VG	S&P 500	0.011	1.71	-4.6e-4	7.7e-4	69.0
	Microsoft	0.022	1.29	1.7e-3	-1.1e-3	58.1
	AUD/US	0.007	2.11	-5.5e-4	5.2e-4	52.2
	WTI	0.024	2.27	-3.5e-3	3.8e-3	72.6
t	S&P 500	0.012	2.13	-3.4e-4	5.8e-4	69.2
	Microsoft	0.024	1.78	1.3e-3	-4.8e-4	84.5
	AUD/US	0.007	2.83	-4.7e-4	4.3e-4	55.3
	WTI	0.024	2.97	-3.5e-3	3.8e-3	71.9

TABLE 3.42. Estimated parameters for data fitted via minimum χ^2.

Model	Security	$\hat{\sigma}$	$\hat{\alpha}$ or $\hat{\delta}$	$\hat{\theta}$	$\hat{\mu}$	χ^2
VG	S&P 500	0.011	1.63	-6.4e-4	9.2e-4	69.5
	Microsoft	0.022	1.21	1.6e-3	-9.5e-4	58.2
	AUD/US	0.007	1.99	-4.5e-4	4.5e-4	51.2
	WTI	0.025	1.93	-2.7e-3	3.2e-3	73.7
t	S&P 500	0.012	2.29	-4.1e-4	6.8e-4	69.3
	Microsoft	0.024	1.90	7.9e-4	-1.6e-4	85.8
	AUD/US	0.007	2.99	-3.3e-4	3.2e-4	55.2
	WTI	0.025	2.70	-2.6e-3	3.1e-3	72.3

TABLE 3.43. Estimated parameters for data fitted via PMLE.

Firstly, on the basis of Section 3.4 we note that the t-model fitted to the Microsoft series can be rejected at the 5% significance level, since the values of the χ^2 goodness of fit statistic 84.5 and 85.8 far exceed $\chi^2_{56}(0.95) = 74.5$. Indeed, the optimal δ estimate for fitting the t

distribution to Microsoft is less than 2, implying an infinite variance of returns, which would seem implausible for a stock price process.

The VG model, on the other hand, fits the Microsoft data well.

For the other 3 data sets, neither the VG nor t can be rejected, with the AUD/US data being fitted very well by both models.

Hurst, Platen & Rachev (1997) used Dow Jones Industrial Average and Nikkei 225 index data, over the period 4 January 1973 to 30 July 1993, to compare a number of (symmetric) subordinator models, and on the basis of the maximised log likelihood value concluded that the t model was best. They also transformed their data to take account of the number of calendar days between successive readings. This adjustment is typically not made, as the number of trading days, as opposed to calendar days, has been found to be the more relevant measure of time for financial asset returns – see for example Fama (1965). Our limited analysis indicates that the Hurst, Platen & Rachev (1997) conclusion does not always hold. Even leaving aside the Microsoft case, and using the smallness of the χ^2 goodness of fit statistic as a measure of fit, the VG model for the AUD/US data fits better than the t, and there is relatively little to choose between the VG and t fits for the S&P 500 and WTI.

Madan, Carr & Chang (1998) and Seneta (2004) also examine the fit of a VG model using PMLE, this time to the S&P 500 index; our parameter estimates are reasonably similar to those reported there, given that data spanning different time periods was used.

The S&P 500 line of Table 3.41, and for the VG model in Tables 3.42 and 3.43 can be compared with Tables 1 and 2 in Tjetjep & Seneta (2006, p. 121), although there the period of observation was earlier (January 1977 to December 1981, 1261 readings) and half as long. The shorter period showed smaller sample mean and smaller standard deviation, and numerically smaller negative skew, although larger kurtosis. These likely reflect a period

of less volatility. The comparable estimate of α for the shorter period was 1/0.4242 = 2.36, compared with our present 1.63 (PMLE) and 1.71 (minimum χ^2).

We now focus on visual comparison of model fits to the four data sets. Figures 3.1 to 3.4 below are drawn to the same scale for ease of comparison, and show histograms and minimum χ^2 fitted model densities. In each case, the uppermost curve at the origin is the fitted VG model, while the lower one is the fitted t model. The PMLE fitted curves, which we omit, are almost identical to the minimum χ^2 fitted curves shown.

According to the almost identical values 69.0 and 69.2 of the χ^2 goodness of fit statistic, there is little to choose between the VG and t fits for the S&P 500 data. Figure 3.1 suggests that this is due to the fact that the VG overestimates peakedness at the mean at the expense of weight in the mid-range, while the t underestimates peakedness at the mean. The data are close to symmetric. It is the peakedness at the mean of the data, rather than heavy-tails, that appears to be responsible for the sample kurtosis of 5.95, compared to the theoretical value of 3 for the normal distribution.

For Microsoft Corporation on the other hand, which has the largest kurtosis of the four series considered, the VG fit is clearly superior from a visual examination. In this case the t places too much weight in the tails at the expense of weight in the center of the distribution, thereby markedly understating peakedness at the mean, whereas the VG manages to fit the center of the distribution well. As discussed in Fung & Seneta (2007), the kurtosis coefficient is strongly affected by the concentration of probability near the origin, as well as in the tails, and it would appear that given the large kurtosis seen in the Microsoft Corporation data, the VG achieves a better trade-off between heavy tails and peakedness at the mean than does the t.

According to the minimised χ^2 statistics, the VG gives a slightly better goodness of fit to the AUD/US exchange rate data and a slightly worse goodness of fit to the WTI series than

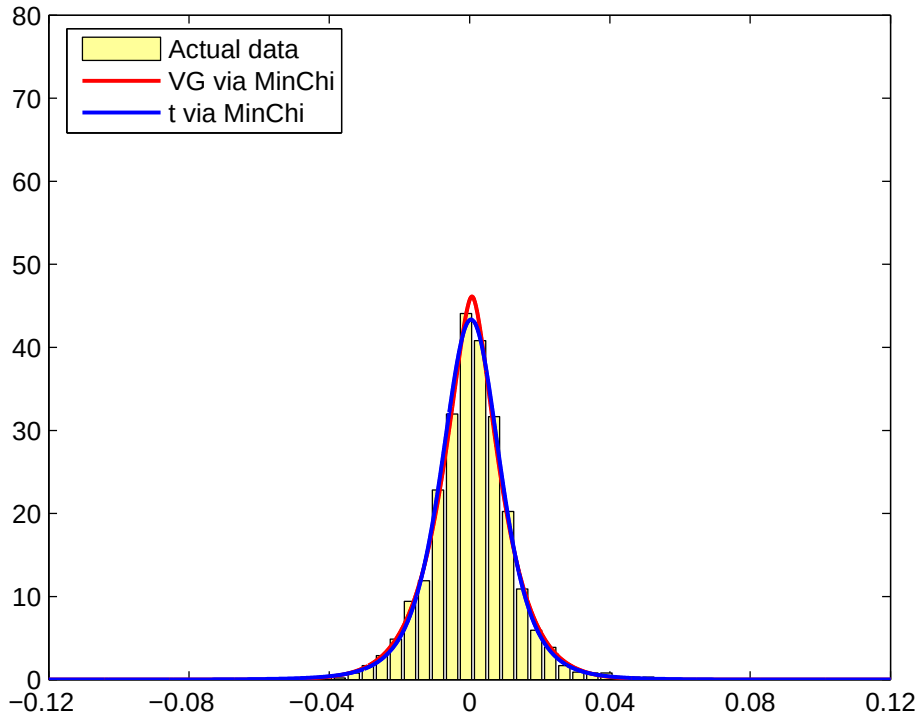

FIGURE 3.1. Histogram of S&P 500 index returns with fitted models.

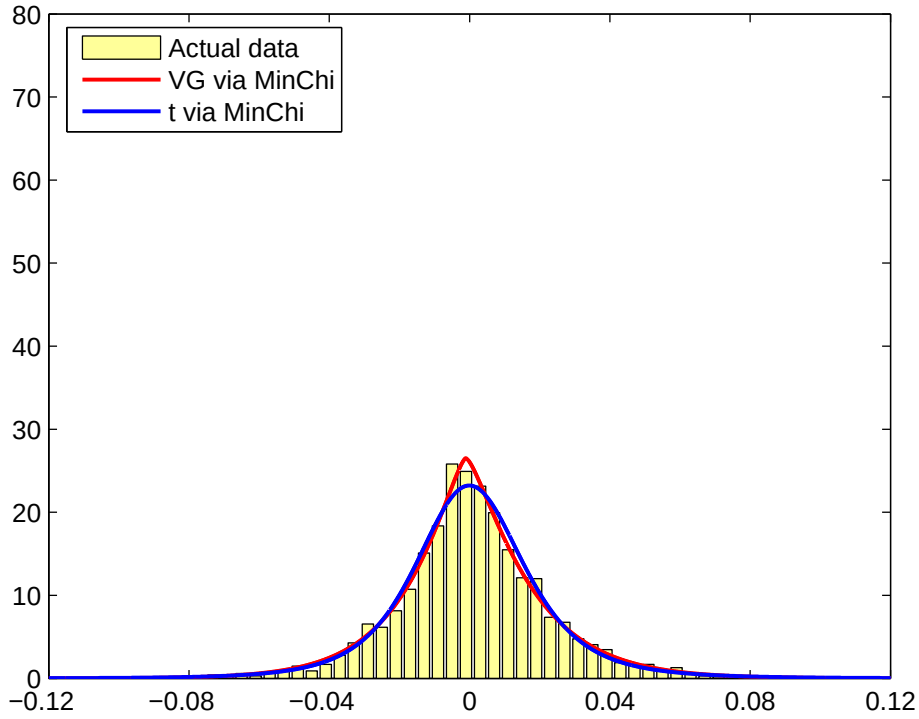

FIGURE 3.2. Histogram of Microsoft Corporation's share price returns with fitted models.

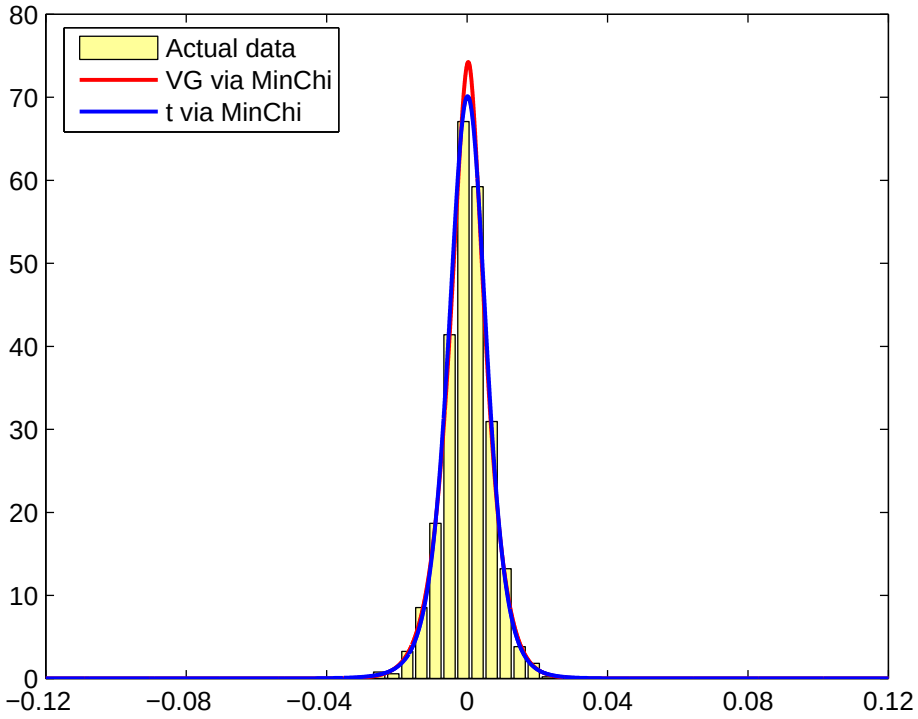

FIGURE 3.3. Histogram of the AUD/US exchange rate returns with fitted models.

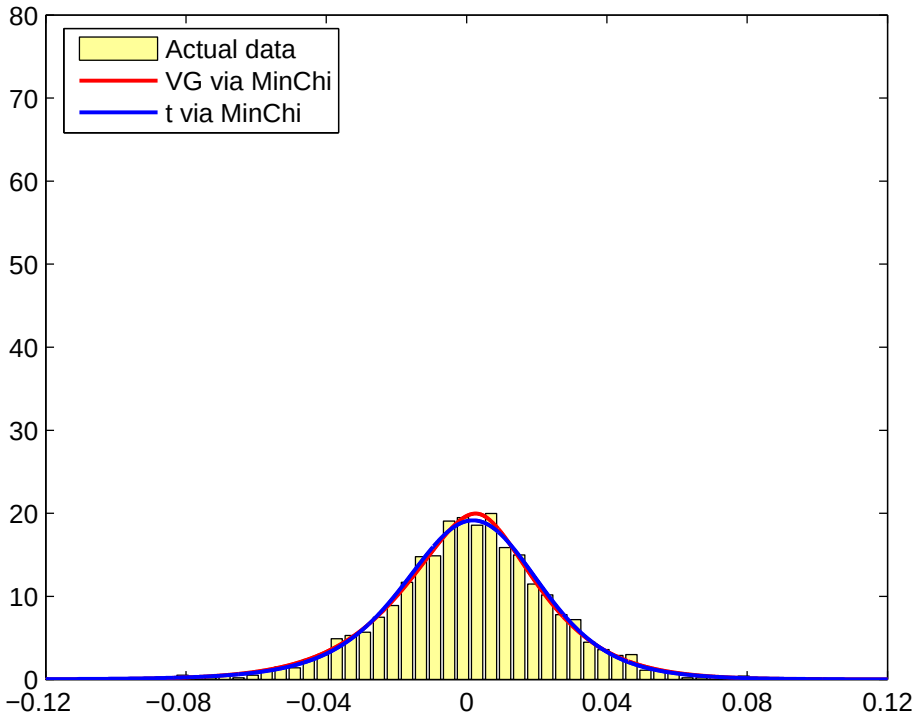

FIGURE 3.4. Histogram of WTI returns with fitted models.

does the t. The AUD/US exchange rate has the smallest kurtosis and standard deviation of the four series considered, and this is clearly evident in Figure 3.3. Here the t fit about the origin appears superior from a visual inspection, although as mentioned the VG recorded the smaller χ^2 value overall. The WTI series on the other hand has the largest standard deviation of the four series considered, although not the largest kurtosis, indicating a less peaked profile than Microsoft for example. In this case both fits appear equally adequate, which accords with the almost identical minimum χ^2 values recorded by the VG and t in Tables 3.42 and 3.43.

We also estimate H values, using both the marginal model independent 'MOM' method, and via the marginal model dependent method using univariate parameter estimates from a $\mu = \theta = 0$ model with mean–adjusted data, fitted via minimum χ^2 (taking $\mu = \theta = 0$ greatly simplifies the calculations involved, and while it does imply that the log price increments $\{X_t\}$ are uncorrelated, data analysis typically bears this out, with $\{X_t^2\}$ showing strong autocorrelation but $\{X_t\}$ showing very little). The results are given in Table 3.44, which can be compared with Tables 3.11 to 3.16 which show H estimates from simulated symmetric VG and t data.

Securities	S&P 500	Microsoft	AUS/US	WTI
VG	0.90	0.92	0.85	0.75
t	0.69	—	0.77	0.70
'MOM'	0.86	0.85	0.81	0.70

TABLE 3.44. $\hat{H}$ estimates for financial data.

In this case the VG gives larger estimates than the t, with estimates from the model independent 'MOM' method usually falling somewhere in between. From Equation (3.4), estimation of H is mainly impacted by the marginal distribution parameters via $\mathbb{V}\text{ar}(\tau_t)$, and indeed estimates of $\mathbb{V}\text{ar}(\tau_t)$ from the t model are substantially larger than from the VG model. In the case of Microsoft, the t model implies an infinite $\mathbb{V}\text{ar}(\tau_t)$ since $\hat{\delta} < 2$, while for the S&P 500, the t estimate is over 10 times as large as the VG estimate (for the other

two models the t variance is about 2.5 times larger). These larger $\mathbb{V}\text{ar}(\tau_t)$ estimates from the t shrink the estimated $\mathbb{C}\text{orr}(\tau_t, \tau_{t+k})$ number towards zero, reducing the fitted H value.

The MOM estimates of $\mathbb{V}\text{ar}(\tau_t)$ on the other hand fall between those of the VG and t. For the S&P 500, the MOM estimate of $\mathbb{V}\text{ar}(\tau_t)$ is twice the VG estimate, but 8 times smaller than the t estimate, and the MOM H estimate is correspondingly closer to the VG than the t; for Microsoft the MOM estimate of $\mathbb{V}\text{ar}(\tau_t)$ is roughly 2.5 times larger than the VG estimate; for the AUD/US exchange rate, the MOM estimate is roughly 1.5 times larger than the VG estimate and 1.5 times smaller than the t estimate, with the H estimate of the MOM also falling between the VG and the t estimates; while for the WTI series the MOM estimate of $\mathbb{V}\text{ar}(\tau_t)$ is 2.5 times larger than the VG estimate and roughly equal to the t estimate, and here the H estimate from the MOM correspondingly matches the H estimate from the t model.

3.6. Other approaches to simulation, estimation and model fit

Finally, we review some other approaches to simulation, estimation, and the fitting of VG models which have appeared in the literature.

3.6.1. Simulation. Avramidis, L'Ecuyer & Tremblay (2003) (see also Avramidis & L'Ecuyer (2006)) also consider the problem of simulating VG increments, although from a different viewpoint and for the iid case only. In particular, they propose an algorithm to sample gamma and VG processes not sequentially, but by continually 'halving the interval'. That is, they sample the given process at the last time point of interest, then at the midpoint, then at the first and third quarter points, and so on, so that for a process of length T, the time points sampled are in order: $T, T/2, T/4, 3T/4, T/8, 3T/8, 5T/8, 7T/8, \cdots$. This is done by making use of the conditional distribution given knowledge of the last and next point, and is analogous to Brownian bridge sampling (Brownian bridge sampling involves sampling a Brownian motion path in such a 'halving the interval' fashion. The name 'Brownian bridge'

is used since the conditional distribution of a point on the Brownian motion path, given the previous and next point, is that of the Brownian bridge).

In particular, for $\{G(t;a,b)\}$ a gamma process with independent increments and distribution for each t given by $G(t;a,b) \sim \Gamma(ta,b)$, we have that for any time $t \geq 0$ and nonnegative increments Δt_1, Δt_2, the conditional distribution of $G(t+\Delta t_1;a,b)$ given $G(t;a,b)=\gamma_0$ and $G(t+\Delta t_1+\Delta t_2;a,b)=\gamma_2$ is (Avramidis & L'Ecuyer (2006))

$$\gamma_0 + (\gamma_2 - \gamma_0)Y$$

for $Y \sim B(\Delta t_1 \alpha, \Delta t_2 \alpha)$, so that one can sample a gamma process recursively from $t=0$ to T by generating one $\Gamma(Ta,b)$ random variable and as many beta random variables as needed to 'fill in' the time points required between 0 and T.

As an aside, Yor (2007) also considers the gamma bridge although from a different perspective, noting that the only known Lévy processes for which one can present an explicit construction of their bridges in terms of the original Lévy process are Brownian motion and the gamma process. From Yor (2007), for $\{G(t;b)\} \stackrel{def}{=} \{G(t;1,b)\}$ a gamma process, $\{G(u;b)/G(t;b)\}$, $u \leq t$ is independent of $G(t;b)$. Moreover the distribution of $\{G(u;b)/G(t;b)\}$, which for fixed t and a given u is a beta $B(u,t-u)$ random variable, does not depend on b, so that a gamma bridge starting at 0 and ending at $a > 0$ may be obtained from $\{aG(u)/G(t)\}$, $0 \leq u \leq t$, where $\{G(t)\} \stackrel{def}{=} \{G(t;1)\}$. The process $\{D_u^{(t)}\} = \{G(u)/G(t)\}$ is often called the Dirichlet process with parameter t – see for example Cifarelli & Melilli (2000) and Cifarelli & Regazzini (1990).

Returning to Avramidis, L'Ecuyer & Tremblay (2003), to sample the VG process one can either combine gamma bridge sampling as detailed above with Brownian bridge sampling, or simply use gamma bridge sampling on its own. The Brownian bridge option first: for $B(t)$ standard Brownian motion and for any increments Δt_1, Δt_2, the conditional distribution of

$B(t + \Delta t_1)$ given $B(t) = b_0$ and $B(t + \Delta t_1 + \Delta t_2) = b_2$ is given by

$$N(b_0 \Delta t_2/(\Delta t_1 + \Delta t_2) + b_2 \Delta t_1/(\Delta t_1 + \Delta t_2),\ \Delta t_1 \Delta t_2/(\Delta t_1 + \Delta t_2)).$$

Combining this with gamma bridge sampling one can recursively sample the VG, which would be given by Brownian bridge samples conditional on a gamma bridge-sampled variance. The alternative method is to sample two appropriately chosen independent gamma processes via gamma bridge sampling, and take the difference, since the VG process is also given by the difference of two gamma processes (see the discussion around the DG model in Chapter 4).

These techniques, combined with quasi-Monte Carlo (QMC) methods, are used to improve the speed and efficiency of estimating stochastic integrals involving the gamma or VG process, which is to say estimating integrals defined against the paths of gamma and VG processes. The idea is to concentrate the variance of the given integrand on a few coordinates, then combine QMC methods with bridge sampling to efficiently estimate the integral. The authors find that this combination outperforms ordinary Monte Carlo simulations, sometimes by orders of magnitude. Efficient estimation of stochastic integrals involving the gamma and VG processes is the focus of Avramidis, L'Ecuyer & Tremblay (2003), and underpins the pricing of some exotic options and structured credit securities which are modeled using gamma and VG processes.

3.6.2. Estimation with WinBUGS. WinBUGS is a programme for Bayesian estimation of parameters. Observed iid data is from a specified distribution (in our study the t or VG) regarded as a conditional distribution in which the parameters are regarded as fixed values of random variables. The statistician specifies a prior distribution (usually the uniform) of the parameters, and uses the observed sample values $x_1, x_2, \cdots x_N$ to modify the prior distribution to a posterior distribution of parameters. WinBUGS uses the Gibbs sampler to simulate realisations of a Markov chain whose states are parameters, and whose limiting

distribution is the posterior distribution of parameters. Thus once a realisation of a chain has been allowed to settle down (the 'burn-in' time), the Bayes estimators of parameters are obtained from averaging the Markov-dependent values of the Markov chain.

For a large size N of the iid sample $x_1, x_2, \cdots x_N$, the Bayes estimators so obtained are equivalent to PMLE estimators. As we have N close to 2500, out parameter estimates using WinBUGS are effectively PMLE estimates. Given we have already dealt extensively with PMLE estimation there is not much more to be gained from a lengthy discussion of estimation via WinBUGS, but we nonetheless briefly detail the procedure and empirically demonstrate the closeness of the PMLE estimated parameters and the WinBUGS estimated parameters.

The normal mean-variance-mixing representation of the (skew) VG and t distributions (see for example Tjetjep & Seneta (2006))

$$X \sim N(\mu + \theta W, \sigma^2 W)$$

where W is, respectively, gamma and inverse gamma distributed, makes the input code straightforward. Thus to fit the t to the $N = 2518$ returns for the S&P 500 data, we used, after entering the data, the input code:

```
for (i in 1:N) {
   x[i] ~ dnorm(mu[i], tau[i])
   mu[i] <- mu.c + theta/s[i]
   tau[i] <- 1/sigma2[i]
   sigma2[i] <- sigma*sigma/s[i]
   s[i] ~ dgamma(u,d) }
```

before specifying the prior distributions for the parameters mu.c, theta, sigma, and u, and making the specification d $\leftarrow$ u-1.

We asked WinBUGS to initialise the parameters and generate 3 separate Markov chains. After 10,000 iterations of which the first 5,000 are treated as burn-in, we had 5,000 stable-state values from each chain, making 15,000 in all, from which to obtain the Bayes estimators. The table below compares the results obtained by PMLE (from Table 3.43), and by WinBUGS, from which it is clear that the two are almost identical.

Model	Method	$\hat{\sigma}$	$\hat{\alpha}$ or $\hat{\delta}$	$\hat{\theta}$	$\hat{\mu}$
VG	PMLE	0.011	1.63	-6.4e-4	9.2e-4
	WinBUGS	0.011	1.63	-6.6e-4	9.4e-4
t	PMLE	0.012	2.29	-4.1e-4	6.8e-4
	WinBUGS	0.012	2.32	-4.4e-4	7.0e-4

TABLE 3.45. Estimated parameters for S&P 500 data fitted via WinBUGS and PMLE.

3.6.3. Model fit. Regarding goodness of fit criteria used by other authors, Tjetjep & Seneta (2006) employ not a log likelihood or χ^2 statistic as we have done, but a statistic of the form

$$\text{AD} = \max_{y\in\mathbb{R}} \frac{F_E(y) - F_M(y)}{\sqrt{F_M(y)(1-F_M(y))}}$$

where $F_M(\cdot)$ is the model distribution function and $F_E(\cdot)$ is the empirical distribution function. This statistic is a normed distance between the model and empirical distribution functions, specifically an 'Anderson-Darling' modification, to random norming, of the Kolmogorov-Smirnov statistic, designed to give greater test sensitivity in the tails (see Anderson & Darling (1936) or Press, Teukolsky, Vetterling & Flannery (1992)). In the event, of two large ($1000+$ readings) and three small (132 readings) data sets considered in Tjetjep & Seneta (2006), the VG was preferred over the the skewed normal, the 'NVM Exponential' and the 'NM Exponential' models for both of the large and one of the small data sets, and came a close second to the NVM Exponential model for the other two small data sets. The t distribution was not considered.

Regarding our own financial data, the AD statistic for each fit is given in Table 3.46 below. (The statistic was calculated via numerical integration of the appropriate pdf, and

based on 301 equally spaced bins between -0.15 and 0.15, after discarding any data points for which $F_M(y)$, due to the imprecision of the numerical integration, was outputted as 0 or 1, and after correcting the outputted $F_M(y)$ to ensure that it was non-decreasing. Note that Tjetjep & Seneta (2006) use a smaller band of ± 0.04 as opposed to our ± 0.15 for computation of the AD statistic. They are able to do this as their data sets are smaller and contain fewer outliers, whereas we must widen the band to capture the entire range of data.)

Securities	S&P 500	Microsoft	AUS/US	WTI
PMLE t	0.04	0.04	0.04	0.05
PMLE VG	0.13	0.19	0.09	0.15
Min χ^2 t	0.04	0.05	0.04	0.08
Min χ^2 VG	0.14	0.23	0.49	0.29

TABLE 3.46. AD statistics for data.

Although the t appears to uniformly outperform the VG, this is not in fact the case. Table 3.47 below gives empirically determined 95% quantile values for the AD statistic fitted to the 1000 simulated runs of $\sigma = 0.01,\ \theta = -5 \times 10^{-4},\ \mu = 10^{-3}$ and $H = \{\text{iid}, 0.75, 0.85, 0.95\}$ data, with $\alpha = 2.5$ for the VG or $\delta = 2.5$ for the t, as in Section 3.4. So for example, regarding the VG data set with $H = 0.75$, the 95% quantile value of the AD statistic for the fit of a (correct) VG model via PMLE was 0.11, whereas the 95% quantile value of the AD statistic for the fit of an (incorrect) t model via PMLE was 0.05. This pattern is repeated throughout, with the t models fitted to VG data producing lower AD statistics than the VG models fitted to VG data, and indeed than the t models fitted to t data, and Table 3.46 must be interpreted in light of this.

	H	iid	iid	0.75	0.75	0.85	0.85	0.95	0.95
	Correct Model	VG	t	VG	t	VG	t	VG	t
Min χ^2	95^{th} percentile (CM)	0.11	0.16	0.11	0.17	0.11	0.22	0.11	0.21
	95^{th} percentile (IM)	0.06	28.14	0.06	14.23	0.06	15.49	0.08	1.48
PMLE	95^{th} percentile (CM)	0.11	0.12	0.11	0.12	0.11	0.12	0.10	0.11
	95^{th} percentile (IM)	0.05	11.55	0.05	5.91	0.05	5.75	0.06	0.98

TABLE 3.47. 95^{th} percentiles of the AD statistic calculated based on correct models (CM) and incorrect models (IM), fitted via minimum χ^2 and PMLE.

The somewhat unexpected pattern seen in Table 3.47 is likely due to a few overlapping factors. The AD statistic's extreme sensitivity to tail behavior, whereby even for VG data, inappropriate t models which put more weight in the tails score lower AD scores than appropriate VG models, means that even one outlying data point, which our minimum χ^2 and PMLE fitting techniques would not put excessive weight on, can have a disproportionate impact on the computed AD score. In addition to this, computation of the AD statistic involves numerically integrating some quite complex pdfs involving in particular bessel functions, and as such is likely to be relatively imprecise, which was demonstrated for instance in our outputted model distribution functions which, before correction, were not always non-decreasing, especially in the VG case. We note that the apparent superior performance of the VG models fitted to t data with $H = 0.95$, as compared to the VG models fitted to t data with H less than 0.95, is somewhat illusory – in this case the outputted $F_M(y)$ function was especially badly behaved, and so our 'data cleaning' procedures outlined above had an especially large impact in moderating the computed AD statistics.

In conclusion, the fact that inappropriate t models fitted to VG data scored lower AD statistics that appropriate VG models fitted to the same VG data would tend to indicate that in this case the AD statistic is not a good one for choosing between competing fits, as too much weight is put on outlying data points, resulting in the distorted outcome of models which are known to be inappropriate being favored over models which are known to be correct.

CHAPTER 4

Option pricing

In this chapter we consider some adaptations of the VG to a risk-neutral process for option pricing, as opposed to considering the VG as a model for historical data as in the preceding chapters. The bulk of the work presented here has been published as Finlay & Seneta (2008b).

Recall that the VG model for asset prices

$$P_t = P_0 e^{\mu t + \theta T_t + \sigma B(T_t)} \tag{4.1}$$

gives a log price increment X as having, conditional on a gamma random variable V say, a normal distribution, i.e.,

$$X|V \sim N(\mu + \theta V, \sigma^2 V) \tag{4.2}$$

where $\mu, \theta \in \mathbb{R}$ and $\sigma > 0$ are constants. As such, the realisations of V drive both volatility, via $\sigma^2 V$, and skewness, via θV. Recall also that the characteristic function of X has the simple representation, for $V \sim \Gamma(\alpha, \alpha)$,

$$\phi_{\text{VG}}(u; \mu, \theta, \sigma, \alpha) = e^{\mathrm{i}\mu u}(1 - \mathrm{i}\theta u/\alpha + \frac{1}{2}\sigma^2 u^2/\alpha)^{-\alpha}. \tag{4.3}$$

There are two stock price process representations that result in VG distributed returns; the subordinator model representation considered in the preceding chapters, and the difference of two gamma processes representation.

What we shall call the Difference of Gammas (DG) process results from the exponentiated difference of two independent gamma processes which have independent gamma distributed increments, one representing log price gains, the other losses. That is, we choose

$$P_t = P_0 e^{\mu t + G_1(t;a,b) - G_2(t;c,d)} \tag{4.4}$$

where for example $\{G(t;\alpha,\beta)\}$ is a gamma process with pdf given by $f_\Gamma(x;t\alpha,\beta)$ from (1.18) for any given t. For each t the returns or log price increments for P_t as in (4.4) have characteristic function

$$\phi_{\text{DG}}(u;\mu,a,b,c,d) = e^{\mathrm{i}\mu u}(1-\frac{\mathrm{i}u}{b})^{-a}(1+\frac{\mathrm{i}u}{d})^{-c}. \tag{4.5}$$

Comparing (4.3) and (4.5) it is clear that choosing $c = a$ results in a VG distribution with parameters $\alpha = a$, $\theta = a(\frac{1}{b} - \frac{1}{d})$ and $\sigma^2 = \frac{2a}{bd}$. Madan, Carr & Chang (1998) consider this restricted version of (4.4) to develop the theory of the subordinator model (4.1).

We propose two extensions to VG option pricing. The first drops the $c = a$ restriction from the difference of two gammas representation and simply works with the process described by (4.4) and (4.5). This introduces one extra degree of freedom into the model while still retaining most of the appealing properties of the VG process, such as its representation as the difference of two gamma processes and its simple characteristic function. The relaxation of this constraint does mean that we can no longer write down a closed-form expression for the pdf of returns, but this is not a problem for option pricing since we only need the characteristic function to compute prices.

For our second extension we consider option pricing when one drops the traditional assumption in the subordinator model (4.1) that the increments of $\{T_t\}$ are independent, and instead look at a few models to price options with a strictly stationary returns process. Although this chapter is only concerned with the *risk-neutral world*, our motivation here comes from the *real-world* setting of historical data, for which it is becoming increasingly

clear that squared returns may not be serially independent, and may actually be long range dependent, as discussed in the preceding chapters.

Although we do not pursue the point here, further motivation for considering long range dependent option pricing models comes from observed prices for long-dated options on realised variance. The 'asset' underlying these options is average realised variance, and in standard models this average realised variance converges to its expected value, which implies that out-of-the-money options on realised variance should have negligible value. In actual fact, such options have considerable value, which implies that the average realised variance does not converge quickly to a constant. Models with long range dependence accommodate this empirical observation, as opposed to standard models without long range dependence which do not (see for example Madan & Eberlein (2007)).

We use the data set of Schoutens (2003) which consists of 77 call option prices written on the S&P 500 index as at 18 April 2002, and compare the fit of our models to the fits obtained in Schoutens (2003), which examines and fits a large number of popular option pricing models. (In fact only 75 of the 77 option prices used in Schoutens (2003) are listed (they are given on pages 155–166) – the missing two are (1100, 45.9, June 2002) and (1200, 88.2, December 2003), where we have written (Strike, Option Price, Expiry date).) This data set is freely available and allows for easy comparison with a number of popular pricing models already in use in Schoutens (2003).

To fit our models, that is, to estimate model parameters, we follow Schoutens (2003) and minimise the root mean square error (RMSE) of our computed option prices. Here the error on option price j is the actual (market) price minus the model determined price. We also adjust for dividends by replacing P_0 by $e^{-q\Upsilon}P_0$, where Υ specifies the remaining life of the option and q is the dividend yield – see for example Schoutens (2003), Section 2.6. If the model perfectly described the asset price process then the RMSE value should be zero, with

all model prices matching market prices, given the single true set of parameters. As such, the size of the RMSE value gives an indication of how well each model matches the market, and provides a basis for comparison between models.

There is somewhat of a dichotomy in option pricing research into (i) specific models of the stock price process and, (ii) general methods for actually computing option prices given any one of a number of specific models. We focus on the first of these two branches and make use of established and very generally applying results to actually compute option prices in accordance with our various models. We choose to price via the characteristic function and make use of the formulae given in Carr & Madan (1999) and Lee (2004) as detailed below.

For $C(\Upsilon, k)$ the price of a European call option with time to maturity Υ and strike price K, where $k = \log(K)$, following Carr & Madan (1999) we define the modified call price as

$$c(\Upsilon, k) = e^{\gamma k} C(\Upsilon, k)$$

for some γ such that $\mathbb{E}(P_\Upsilon^{\gamma+1}) < \infty$. The Fourier transform of $c(\Upsilon, k)$ is then given by

$$\begin{aligned}
\Psi_\Upsilon(x) &= \int_{-\infty}^{\infty} e^{\mathrm{i}xk} c(\Upsilon, k) \mathrm{d}k \\
&= \int_{-\infty}^{\infty} e^{\mathrm{i}xk} \int_k^{\infty} e^{\gamma k} e^{-r\Upsilon} (e^p - e^k) q_\Upsilon(p) \mathrm{d}p \, \mathrm{d}k \\
&= \int_{-\infty}^{\infty} e^{-r\Upsilon} q_\Upsilon(p) \int_{-\infty}^{p} (e^{p+\gamma k} - e^{(1+\gamma)k}) e^{\mathrm{i}xk} \mathrm{d}k \, \mathrm{d}p \\
&= \int_{-\infty}^{\infty} e^{-r\Upsilon} q_\Upsilon(p) \left(\frac{e^{(\gamma+1+\mathrm{i}x)p}}{\gamma + \mathrm{i}x} - \frac{e^{(\gamma+1+\mathrm{i}x)p}}{\gamma + 1 + \mathrm{i}x} \right) \mathrm{d}p \\
&= \frac{e^{-r\Upsilon} \phi_\Upsilon(x - (\gamma+1)\mathrm{i})}{\gamma^2 + \gamma - x^2 + \mathrm{i}x(2\gamma + 1)} \qquad (4.6)
\end{aligned}$$

where for the specific model under consideration, $q_\Upsilon(p)$ is the risk-neutral density of $\log(P_\Upsilon)$, the log stock price at time Υ, $\int_k^\infty e^{-r\Upsilon}(e^p - e^k) q_\Upsilon(p) \mathrm{d}p = e^{-r\Upsilon} \mathbb{E}(P_\Upsilon - K)^+ = C(\Upsilon, k)$, and $\phi_\Upsilon(x)$ is the characteristic function of the log stock price at time Υ. Since $C(\Upsilon, k)$ is real, the real part of $\Psi_\Upsilon(x)$ is even while the imaginary part is odd, so that taking the inverse

transform of $\Psi_\Upsilon(x)$ gives

$$C(\Upsilon,k)=\frac{e^{-\gamma k}}{2\pi}\int_{-\infty}^{\infty}e^{-ixk}\Psi_\Upsilon(x)\mathrm{d}x=\frac{e^{-\gamma k}}{\pi}\int_0^{\infty}\Re\{e^{-ixk}\Psi_\Upsilon(x)\}\mathrm{d}x.$$

In fact we use a modified version of the above suggested in Lee (2004) and given by

$$C(\Upsilon,k)=R_\gamma+\frac{e^{-\gamma k}}{\pi}\int_0^{\infty}\Re\{e^{-ixk}\Psi_\Upsilon(x)\}\mathrm{d}x. \tag{4.7}$$

Here the R_γ term results from shifting an integral through or across a pole in the complex plane, and is given by $R_\gamma=0$ for $\gamma>0$, $R_\gamma=\phi_\Upsilon(-i)/2$ for $\gamma=0$, $R_\gamma=\phi_\Upsilon(-i)$ for $-1<\gamma<0$, $R_\gamma=\phi_\Upsilon(-i)-e^k\phi_\Upsilon(0)/2$ for $\gamma=-1$ and $R_\gamma=\phi_\Upsilon(-i)-e^k\phi_\Upsilon(0)$ for $\gamma<-1$ (the choice of γ impacts on the error generated by the numerical approximation of (4.7), and is discussed extensively in the two papers mentioned). Finally, the option price (4.7) is computed via numerical integration.

We note that although the characteristic function method of option pricing just described is more efficient and convenient, when dealing with the subordinator model under the restriction $\mu=r$, $\theta=-\frac{1}{2}\sigma^2$ one can also proceed with a Merton-Black-Scholes type method for option pricing. Here the original Merton-Black-Scholes option pricing formula, valid when the stock price process is taken as geometric Brownian motion and derived by both Black & Scholes (1973) and Merton (1973), is given by $P_0\Phi(d)-Ke^{-r\Upsilon}\Phi(d')$ for $d=\frac{\log(\frac{P_0}{K})+r\Upsilon+\frac{1}{2}\sigma^2\Upsilon}{\sigma\sqrt{\Upsilon}}$, $d'=\frac{\log(\frac{P_0}{K})+r\Upsilon-\frac{1}{2}\sigma^2\Upsilon}{\sigma\sqrt{\Upsilon}}$, and where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal. If one replaces d and d' by $d=\frac{\log(\frac{P_0}{K})+r\Upsilon+\frac{1}{2}\sigma^2T_\Upsilon}{\sigma\sqrt{T_\Upsilon}}$ and $d'=\frac{\log(\frac{P_0}{K})+r\Upsilon-\frac{1}{2}\sigma^2T_\Upsilon}{\sigma\sqrt{T_\Upsilon}}$, both functions of the random variable T_Υ, then one can arrive at the VG option price by integrating over the density of T_Υ.

4.1. Market completeness

Market completeness is important in option pricing and is related to uniqueness of option prices. A market, or more specifically the market for securities implied by the chosen stock price process, is typically said to be complete if for every contingent claim or option

there exists an admissible self-financing predictable strategy that can replicate the claim. A predictable strategy replicates a claim if it specifies a dynamic portfolio consisting of a cash account and stock holdings which matches the value of the claim at every time point. It is self-financing if no money is put in or pulled out along the way. And it is admissible if the portfolio's value is bounded below by a constant. A market is incomplete if it is not complete. Essentially, a market is complete if one can perfectly hedge an option by trading in the stock, and is incomplete otherwise.

For a given model, the uniqueness of an equivalent martingale measure (that is, the uniqueness of option prices implied by the model) implies market completeness. The converse is by and large true – when there is not one but a number of equivalent martingale measures with which one could price options, the market is typically incomplete – but not always true, with complete markets existing where the equivalent martingale measure is not unique (see Artzner & Heath (1995) or Jarrow, Jin & Madan (1999)).

Most markets are in fact incomplete. Of all Lévy processes, only Brownian motion and the Poisson process imply complete markets (Schoutens (2003), p. 46). The classical VG as well as the processes we consider imply incomplete markets, due essentially to jumps in the sample paths which make constructing a perfect hedge as discussed above impossible. Brownian motion is continuous and hence has no jumps, while the Poisson process has jumps of size one only, which as it turns out also leads to a complete market. See for example Schoutens (2003) p. 18 and Section 6.2.1 for a discussion of market completeness in the context of Lévy processes.

The fact that the VG implies incomplete markets means that there is not a unique way of going from the real to the risk-neutral world. That is, there is not a single unique equivalent martingale measure with which to price options. We do not attempt a comprehensive study of how one should choose between martingales but rather calibrate our models to market

data and compare the performance of different models on the basis of in-sample fit as discussed below (see for example Fujiwara & Miyahara (2003) for a more general discussion on the choice of martingale measure in the Lévy process framework).

As an aside, the distinction between complete and incomplete markets is also an important one in macroeconomics. Similar to the discussion above, in the macroeconomic nomclementure a complete market is said to exist if consumers can insure for every eventuality, or equivalently, if for every possible state of the world consumers can purchase a security which pays off in that state. Conversely, a market is said to be incomplete if consumers cannot insure for every eventuality and must therefore bear risk. See Cochrane (2001) Chapter 3 or Ljungqvist & Sargent (2004) Chapters 8 and 17 for a discussion of market completeness in the macroeconomic context.

4.2. Established VG option pricing models

When constructing a model for option pricing one typically ensures that the discounted stock price process $\{e^{-rt}P_t\}$ is a martingale. This is because martingale stock price processes represent a 'fair game' and exclude arbitrage, with the conditional expectation of future values of the process being equal to the current value. Now for either the difference of gammas or subordinator model and for r the interest rate, consider the decomposition of the discounted stock price $e^{-rt}P_t$ given by

$$e^{-rt}P_t = e^{-rs}P_s \times A_{s,t}$$

for $s \leq t$. To move from a real-world model of a stock price process $\{P_t\}$ and the associated historical returns, to a risk-neutral model of $\{e^{-rt}P_t\}$ for option pricing, one typically imposes parameter restrictions to ensure that $\{e^{-rt}P_t\}$ is a martingale, that is, to ensure that $\mathbb{E}(A_{s,t}|\mathcal{F}_s) = 1$ where $\mathcal{F}_s$ represents information available up until time s.

From (4.1), for $\mathcal{F}_s = \sigma(\{B(u), u \leq T_s\}, \{T_u, u \leq s\})$ and $\mathcal{F}_* = \sigma(\{B(u), u \leq T_s\}, \{T_u, u \leq s\}, T_t)$ for $s < t$ and $\{B(t)\}$ independent of $\{T_t\}$, we have that irrespective of the distribution of T_t,

$$\mathbb{E}(e^{-rt}P_t|\mathcal{F}_s) = P_0\, \mathbb{E}(e^{(\mu-r)t+\theta(T_t-T_s+T_s)+\sigma(B(T_t)-B(T_s)+B(T_s))}|\mathcal{F}_s)$$

$$= P_s\, e^{(\mu-r)t-\mu s}\, \mathbb{E}(\mathbb{E}(e^{\theta(T_t-T_s)+\sigma(B(T_t)-B(T_s))}|\mathcal{F}_*)|\mathcal{F}_s) \tag{4.8}$$

$$= e^{-rs}P_s \times e^{(\mu-r)(t-s)}\mathbb{E}(e^{(\theta+\frac{1}{2}\sigma^2)(T_t-T_s)}|\mathcal{F}_s). \tag{4.9}$$

Here (4.8) follows from the 'tower' (repeated expectation) property of conditional expectation, since $\mathcal{F}_s \subset \mathcal{F}_*$, and (4.9) follows from the moment generating function of a normal random variable. Now for $T_t \sim \Gamma(\alpha t, \alpha)$, that is, T_t with iid gamma increments as in Madan, Carr & Chang (1998), we have that

$$\mathbb{E}(e^{(\theta+\frac{1}{2}\sigma^2)(T_t-T_s)}|\mathcal{F}_s) = \mathbb{E}(e^{(\theta+\frac{1}{2}\sigma^2)T_{t-s}}) = (1 - \frac{\theta+\frac{1}{2}\sigma^2}{\alpha})^{-\alpha(t-s)} \tag{4.10}$$

for $\alpha > \theta + \frac{1}{2}\sigma^2$ which ensures that (4.10) is finite. Hence taking

$$\mu = r + \alpha \log(1 - \frac{\theta+\frac{1}{2}\sigma^2}{\alpha}) \tag{4.11}$$

results in $e^{(\mu-r)(t-s)}\mathbb{E}(e^{(\theta+\frac{1}{2}\sigma^2)(T_t-T_s)}|\mathcal{F}_s) = 1$ so that $\mathbb{E}(e^{-rt}P_t|\mathcal{F}_s) = e^{-st}P_s$ as desired. This is an example of a 'mean-correcting martingale' model, whereby a risk-neutral model is obtained from a real-world model by restricting the mean parameter (see for example Schoutens (2003), Section 6.2.2 and Madan, Carr & Chang (1998)).

The above approach is a standard method for obtaining a martingale, and results in only one parameter constraint being imposed. It relies on the computation of the left hand side of (4.10) however, which in effect is the moment generating function of T_t. Thus if the mgf fails to exist, as is the case for the t-distribution for example where the increments of T_t are inverse gamma distributed, or is not tractable, then the method will fail – see Seneta (2004) for a more general discussion of the mgf and martingales.

An alternative advocated in Heyde & Leonenko (2005) is to take $\mu = r$ and $\theta = -\frac{1}{2}\sigma^2$ in (4.9) so that the right hand side of (4.9) becomes $e^{-st}P_s \times 1 = e^{-st}P_s$ (note that Heyde & Leonenko (2005) start with a real-world model with no θ parameter, i.e. (4.1) with $\theta = 0$, and introduce θ, in the form of $-\frac{1}{2}\sigma^2$, only when they come to option pricing). This construction is simple and quite general, as it does not in fact depend on the distribution of T_t (the distribution of T_t is of course needed however when one comes to actually compute the price of the option). It is somewhat restrictive however, in that two parameters are constrained, μ and θ. We shall refer to this construction as a 'skew-correcting martingale' as θ, the parameter that determines skewness, is constrained.

4.3. Option pricing under the DG model

Consider the process described by (4.4) and (4.5). The simple representation means that it is quite easy to calculate various model properties, so for example the n^{th} moment of the stock price process at time t in the future is given by, for $b > n$,

$$P_0^n e^{n\mu t}\left(\frac{b}{b-n}\right)^{at}\left(\frac{d}{d+n}\right)^{ct}. \tag{4.12}$$

Further, one can show that

$$\mathbb{E}(e^{-rt}P_t|\mathcal{F}_s) = e^{-rs}P_s \times e^{(\mu-r)(t-s)}\left(\frac{b}{b-1}\right)^{a(t-s)}\left(\frac{d}{d+1}\right)^{c(t-s)}$$

so that imposing the restriction

$$\mu = r - a\log\left(\frac{b}{b-1}\right) - c\log\left(\frac{d}{d+1}\right)$$

for $b > 1$ renders $\{e^{-rt}P_t\}$ a martingale with four free parameters: a, b, c and d. Given this, it is an in principle simple matter to price options via (4.7) where in this case, for an option with time to maturity Υ, $\phi_\Upsilon(u)$ from (4.6) is given by $\phi_{\text{DG}}(u; \log(P_0) + \mu\Upsilon, a\Upsilon, b, c\Upsilon, d)$ for $\phi_{\text{DG}}(u; \mu, a, b, c, d)$ as in (4.5).

Model	RMSE	a	b	c	d
DG	2.24	4.35	240.86	9.79×10^{-6}	2.65×10^{-7}

TABLE 4.1. Fit of DG model to Schoutens (2003) option data.

Table 4.1 gives the minimised RMSE of the fit as well as the estimated parameter values, where the parameters are based on the time scale of a day, assuming 250 trading days per year (to convert to the time scale of a year one simply multiplies a and c by 250). The optimal parameters imply a stable log price gains process (mean and variance of 0.02 and 7.49×10^{-5}) and a very volatile log price losses process (mean and variance of 36.98 and 1.40×10^{8}), although from (4.12) the implied moment properties of the stock price process itself are relatively moderate – over a one year horizon, for P_0 normalised to unity, P_t has a standard deviation of 24 per cent, skewness of -2.3 and kurtosis of 11.4 (the mean is constrained to ensure we have a martingale).

Turning to goodness of fit, our four parameter model (not counting the constrained mean parameter) does better as expected than the standard three parameter VG model given by (4.1) with mean-correction as per (4.11), considered in Schoutens (2003) on page 83, which recorded an RMSE of 3.56. It also does better than the four parameter CGMY and GH models considered (RMSEs of 2.76 and 2.88 respectively), although as expected, models considered in Schoutens (2003) with five or more parameters gave a lower RMSE value.

We note that the RMSE surface in this case was found to be quite flat, with a number of different parameter values giving essentially the same RMSE of 2.24, to 2 decimal places at least. The parameters reported above gave an RMSE 0.001 units smaller than the next-smallest RMSE found. All such parameter values implied a stable gains process and very volatile losses process however, as well as moment properties of P_t almost identical to those reported here. One of many checks on our optimisation procedures which we performed for our pricing model was to calculate the RMSE of the standard VG as reported in Schoutens (2003) p. 82, where the parameters of $C = 1.3574$, $G = 5.8704$ and $M = 14.2699$

correspond for our DG model to $a = c = C/250$, $b = M$ and $d = G$. Our calculated RMSE using these parameters was 3.57, as compared to the 3.56 reported. We also optimised in this setting, starting from $a = c = 0.01$, $b = d = 10$ and got back to Schoutens' parameters and RMSE.

4.4. Option pricing under a LRD VG model

Return to the subordinator process as described by (4.1). We again take $\{T_t\}$ to be a positive non-decreasing random process with stationary, identically distributed gamma unit increments, but now drop the independence assumption and allow the $\{\tau_t\}$ increments to be a strictly stationary process through time.

Specifically, as in Chapter 2 we choose $T_t = \sum_{i=1}^{t} \tau_\alpha(i)$ with

$$\tau_\alpha(t) = (\eta_1^2(t) + \cdots + \eta_{2\alpha}^2(t))/2\alpha \sim \Gamma(\alpha, \alpha)$$

for 2α integer valued. Here each $\{\eta_i(t), t \geq 0\}$ is an independent stationary Gaussian process with zero mean, unit variance, and autocorrelation function given by

$$\rho(s) = (1 + s^2)^{(H-1)/2} \tag{4.13}$$

but this time with $H < 1$ instead of $\frac{1}{2} < H < 1$. It follows that $\mathbb{C}\text{ov}(\tau_t, \tau_{t+s}) = \rho^2(s)/\alpha$, where we write τ_t for $\tau_\alpha(t)$ to simplify notation. Then

$$T_t \stackrel{\mathcal{D}}{=} \frac{1}{2\alpha} \sum_{i=1}^{2\alpha} \mathbf{Z}_i' \Sigma(t) \mathbf{Z}_i$$

for each $\mathbf{Z}_i$ a length t vector of uncorrelated $N(0, I)$ random variables, $\mathbf{Z}_i'$ the transpose of $\mathbf{Z}_i$, and $\Sigma(t)$ the $t \times t$ correlation matrix of $\{\eta(i), i = 1, \cdots, t\}$. Note that Σ from Chapter 3 is equivalent to $\Sigma(2500)$; we change notation here to stress dependence on t. For $\frac{1}{2} \leq H < 1$ the correlation structure chosen implies the long range dependence of $\{\tau_t\}$ and therefore of $\{X_t^2\}$. From Chapter 2 we know that for $\frac{1}{2} < H < 1$, $\{T_t\}$ appropriately normed converges

to the H self similar Rosenblatt process. We shall refer to the model that results from using this autocorrelated $\{T_t\}$ as the LRD VG.

From Hogg & Craig (1978) Section 12.2 for example, the characteristic function of T_t is then given by

$$\mathbb{E}(e^{\mathrm{i}uT_t}) = |I - \mathrm{i}u\Sigma(t)/\alpha|^{-\alpha} = (\prod_{j=1}^{t}(1 - \mathrm{i}ur_j(t)/\alpha))^{-\alpha} \tag{4.14}$$

where $r_j(t)$ is the j^{th} largest eigenvalue of $\Sigma(t)$. As such, the characteristic function of $\log(P_t)$ for P_t as in (4.1) and T_t as in (4.14) is given by

$$\phi_{\text{LRD}}(u) = P_0^{\mathrm{i}u} e^{\mathrm{i}\mu tu}(\prod_{j=1}^{t}(1 + (\frac{1}{2}u^2\sigma^2 - \mathrm{i}u\theta)\frac{r_j(t)}{\alpha}))^{-\alpha}, \tag{4.15}$$

which can be compared with the joint cf of $\{X_t\}$ for $X_t = \log(P_t) - \log(P_{t-1})$ given in Section 3.3. Note that the model described by (4.15) has five parameters: μ, θ, σ, α, as well as H, which via (4.13) determines the strength of correlation between increments, thereby determining $\Sigma(t)$ and so each $r_j(t)$.

As $H \uparrow 1$ the characteristic function (4.15) reduces to

$$P_0^{\mathrm{i}u} e^{\mathrm{i}\mu tu}(1 + (\frac{1}{2}u^2\sigma^2 - \mathrm{i}u\theta)\frac{t}{\alpha})^{-\alpha}$$

since in this case $\Sigma(t)$ is a matrix of ones so that $r_1(t) = t$ and $r_j(t) = 0$, $j = 2, \cdots, t$. This is the model (4.1) with $T_t \sim \Gamma(\alpha, \alpha/t)$ instead of $\Gamma(t\alpha, \alpha)$. While both the $\Gamma(t\alpha, \alpha)$ and $\Gamma(\alpha, \alpha/t)$ distributions have expectation t, the first has variance t/α while the second has variance t^2/α. Hence as H increases 'time' becomes more variable and the log stock price becomes more fat-tailed ($\kappa = 3(1 + \mathbb{V}\text{ar}(T_t)/t^2)$ where κ is the kurtosis of $\log(P_t)$ for the model (4.1) with $\theta = 0$).

Conversely as $H \to -\infty$ we return to the independent increments case with (4.15) reducing to

$$P_0^{\mathrm{i}u} e^{\mathrm{i}\mu tu}(1 + (\frac{1}{2}u^2\sigma^2 - \mathrm{i}u\theta)/\alpha)^{-t\alpha}$$

since in this case $\rho(s) \to 0$ for $s > 0$ from (4.13), so that $r_j(t) \to 1$ for $j = 1, \cdots, t$.

In the real-world context there is growing recognition that historical squared returns data are not independent, and indeed show evidence of long range dependence through time. Although we are only concerned with the risk-neutral setting here, we take this real-world evidence as motivation for new risk-neutral models, and wish to determine, via a goodness of fit criterion, whether the idea of LRD returns in a VG model is applicable in the risk-neutral world also.

The purpose of the next few sections is then to detail and discuss methods for pricing options under the LRD VG, and compare the fit to that obtained from other comparable models.

4.4.1. Some martingale constructions. Recall that to construct a martingale we must first compute $\mathbb{E}(e^{(\theta+\frac{1}{2}\sigma^2)(T_t-T_s)}|\mathcal{F}_s)$. For T_t as in (4.14), $\{\tau_t\}$ is LRD and so definitely not Markovian. Hence the distribution of $(T_t - T_s)|\mathcal{F}_s$ will depend on $\{T_u,\ 0 \leq u \leq s\}$, and this complicates matters. We shall consider three constructions which seek to overcome this problem in different ways.

4.4.1.1. *A skew-correcting martingale: 'M1'.* The simplest solution is to employ a skew-correcting martingale as described in Section 4.2. In this case we take $\mu = r$ and $\theta = -\frac{1}{2}\sigma^2$ which renders $\{e^{-rt}P_t\}$ a martingale with three free parameters: σ, α and H.

4.4.1.2. *A second skew-correcting martingale: 'M2'.* Another approach is to take the skew-correcting martingale M1 and include an extra term to drive skewness. The idea here is to allow for flexibility in determining the skewness of the stock price process, unlike M1 where θ is constrained. In particular, let the stock price P_t be given by

$$P_t = P_0 e^{\mu t + \theta T_t + \sigma B(T_t) + \theta^* T_t^*} \tag{4.16}$$

where $\{T_t\}$ is as in (4.14), and $\{T_t^*\}$ is an independent gamma process which has independent $\Gamma(\beta, \beta)$ distributed unit increments. In this case the characteristic function of $\log(P_t)$ is given by

$$\phi_{\text{M2}}(u) = P_0^{\mathrm{i}u} e^{\mathrm{i}\mu t u} (\prod_{j=1}^{t} (1 + (\frac{1}{2}u^2\sigma^2 - \mathrm{i}u\theta)\frac{r_j(t)}{\alpha}))^{-\alpha} (1 - \frac{\mathrm{i}\theta^* u}{\beta})^{-\beta t}.$$

Taking $\theta = -\frac{1}{2}\sigma^2$ and $\mu = r + \beta \log(1 - \frac{\theta^*}{\beta})$ for $\theta^* < \beta$ will render the discounted stock price process a martingale. In this case the discounted stock price process $\{e^{-rt}P_t\}$ has five free parameters: σ, α, H, θ^* and β.

4.4.1.3. *A mean-correcting construction: 'C3'.* An attractive alternative to the martingales above would be a mean-correcting martingale, where only μ was constrained. The construction of such a martingale would require the computation of $\mathbb{E}(e^{(\theta+\frac{1}{2}\sigma^2)(T_t - T_s)}|\mathcal{F}_s)$ however, which is determined by and dependent on the unobservable process $\{T_u,\ 0 \leq u \leq s\}$, and so is not possible in practice. Were a mean-correcting martingale construction possible however it could have advantages over M1 and M2, due to its greater flexibility (regarding M1) and its more natural accommodation of skewness (regarding M2). As such, to stimulate a critical comparison of M1 and M2, we construct a mean-correcting process which, while not a martingale, is close to one, in the sense that our process $\{P_t\}$ will have the property, for fixed s the current time and t some time in the future, $e^{-rt}P_t = e^{-rs}P_s \times A_{s,t}$ with $\mathbb{E}(A_{s,t}|\mathcal{F}_s)$ possibly not equal to one but $\mathbb{E}(\mathbb{E}(A_{s,t}|\mathcal{F}_s)) = \mathbb{E}(A_{s,t}) = 1$. Here we follow the approach of Drăgulescu & Yakovenko (2002) who, in the context of the Heston (1993) model for historical returns, 'integrated out' the unobservable volatility parameter to focus on returns. In our case we 'integrate out' the dependence on past unobservable realisations of $\{T_u,\ 0 \leq u \leq s\}$ by taking the expectation of the conditional expectation.

Now $e^{-rt}P_t = e^{-rs}P_s \times e^{(\mu-r)(t-s)+\theta(T_t-T_s)+\sigma(B(T_t)-B(T_s))}$ and since the $\{\tau_t\}$ process is strictly stationary, we have by conditioning on $\{T_s, T_t\}$ that

$$\begin{aligned}\mathbb{E}(e^{\theta(T_t-T_s)+\sigma(B(T_t)-B(T_s))}) &= \mathbb{E}(e^{(\theta+\frac{1}{2}\sigma^2)(T_t-T_s)}) \\ &= (\prod_{j=1}^{t-s}(1-(\theta+\frac{1}{2}\sigma^2)r_j(t-s)/\alpha))^{-\alpha}\end{aligned}$$

for $\alpha > (\theta+\frac{1}{2}\sigma^2)r_1(t-s)$. Therefore setting the (time-dependent) μ as

$$\mu = r + (\alpha/(t-s))\sum_{j=1}^{t-s}\log\left(1-(\theta+\frac{1}{2}\sigma^2)r_j(t-s)/\alpha\right)$$

results in

$$e^{-rt}P_t = e^{-rs}P_s \times A_{s,t}$$

with $E(A_{s,t}) = 1$ as desired. In this case the discounted stock price $e^{-rt}P_t$ has four free parameters: θ, σ, α and H.

In fact the construction of our C3 model is analogous to the construction advocated in Carr, Geman, Madan & Yor (2003), Section 4.2 for mean-corrected stochastic volatility Lévy models, and the property $\mathbb{E}(A_{s,t}) = 1$ enjoyed by our C3 model is equivalent to that described in Carr, Geman, Madan & Yor (2003), Section 5 as ensuring a lack of *static* arbitrage. (Here static arbitrage is said to hold if there exists a costless trading strategy, dependent only on the time and contemporaneous stock price, which provides for a positive profit with positive probability and has no possibility of a loss. This is opposed to *dynamic* arbitrage which is defined analogously but for the fact that the trading strategy is allowed to employ knowledge of all past values of all variables, observable or not.) Carr, Geman, Madan & Yor (2003) note the practical impossibility of accessing dynamic arbitrage opportunities which depend on unobserved processes such as our $\{T_t\}$ process, and rather consider a number of models which, similar to our C3 model, are *not* martingales, but which like martingales possess a similar 'fair game' property and which exclude static arbitrage. (For further results on static arbitrage see for example Carr & Madan (2005) or

Davis & Hobson (2007), which consider the conditions necessary for static arbitrage to be excluded from a set of observed option prices.)

4.4.2. Fit to data. To fit the model we again minimise the RMSE of our computed option prices, with the results given in Table 4.2. Here the θ parameters for the M1 and M2 models, given in bold-face, are not free but are constrained so that $\theta = -\frac{1}{2}\sigma^2$. Again we work on the time scale of a day, assuming 250 trading days per year.

Model	RMSE	σ	θ	α	θ^*	β	H
M1	6.35	0.012	$\mathbf{-7.0 \times 10^{-5}}$	0.02	–	–	-1.58
M2	2.27	0.009	$\mathbf{-3.8 \times 10^{-5}}$	6.39	-5.1×10^{-3}	3.0×10^{-5}	-23.50
C3	0.76	0.010	-7.5×10^{-4}	1.66	–	–	0.99

TABLE 4.2. Fit of LRD VG models to Schoutens (2003) option data.

Regarding Table 4.2, it seems clear that the three parameter M1 model imposes too binding a constraint to produce a good fit to actual data – the relatively large σ and small $|\theta|$ fitted by the M1 model relative to the parameter values fitted by the more flexible C3 model indicates that $\theta = -\frac{1}{2}\sigma^2$ is not a good assumption, with σ pushed up to accommodate θ and $|\theta|$ pushed down to accommodate σ. Indeed the skew-correcting M1 model does worse than the standard independent increments VG model considered in Schoutens (2003), which recorded an RMSE of 3.56, indicating that having a flexible skewness parameter is more important than allowing for dependence of squared returns.

The five parameter M2 model does achieve a lower RMSE, but is still beaten by the four parameter DG model considered in Section 4.3. The M2 model allows for dependence of squared returns as well as flexible skewness, but does so in a somewhat artificial manner. In particular the M2 model does not incorporate in a flexible manner the so called 'leverage effect', an empirical phenomenon documented in the literature whereby large negative returns on the stock market and high volatility usually accompany one another. This can be most easily seen by considering the distribution of returns conditional on T_t and T_t^*. In this case

we have, for $X = \log(P_t) - \log(P_{t-1})$ a typical log price increment, with $V = T_t - T_{t-1}$ and $W = T_t^* - T_{t-1}^*$ similar to (4.2), that

$$X|\{V, W\} \sim N(r + \beta \log\,(1 - \frac{\theta^*}{\beta}) - \frac{1}{2}\sigma^2 V + \theta^* W,\ \sigma^2 V).$$

Therefore high volatility in the stock price, which is to say large $\sigma^2 V$ values resulting from large realisations of V, result in a negative contribution to returns of V multiplied by the *fixed* coefficient $-\frac{1}{2}\sigma^2$. Similarly, large negative returns caused by a negative *non-constrained* θ^* parameter and large realisations of W, do not increase the conditional variance. The H value estimated for the M2 model in fact implies only a very weak dependence structure, but it has relatively little effect on computed option prices; increasing H to 0.60 increases the RMSE by only 0.006, while increasing H to 0.80 increases the RMSE by only 0.031. Similarly, decreasing the α parameter to 1.00 say increases the RMSE by only 0.001 while decreasing α to 0.10 increases the RMSE by only 0.039. It therefore seems clear that the M2 model has a quite flat RMSE surface, perhaps due to too many model parameters (θ, θ^*; α, β) playing similar roles.

The four parameter C3 model by comparison does accommodate the leverage effect in a flexible way, as well as allowing for flexible skewness and the dependence of squared returns. From (4.2) for example, $X|V \sim N(\mu + \theta V, \sigma^2 V)$ so that highly volatile stock price movements – large $\sigma^2 V$ values resulting form large realisations of V – result in large realised negative returns $\mu + \theta V$, where the θ parameter, assumed here to be negative, is not constrained. In fact our four parameter C3 model seems to fit the option data quite well, doing better than the five parameter BNS models fitted in Schoutens (2003) on page 95, the best of which achieves a RMSE of 1.33 (the BNS IG-OU model). Our model is however beaten by the seven and eight parameter Lévy SV models given on page 98, which have RMSE values ranging from 0.36 to 0.50 (for the CGMY-Gamma-OU and VG-CIR respectively). In the C3 case H is estimated close to but below the upper bound of unity, implying a strong dependence structure in $\{T_t\}$ and returns.

Table 4.2 can also be compared with Tables 3.42 and 3.43 in Section 3.5, where a LRD VG model was fitted to historical data from the S&P 500 index. There the estimated real-world parameters σ, θ, α and H were given as 0.011, -6.4×10^{-4}, 1.63 and 0.90 respectively, which are in fact quite close to those recorded for our C3 model. The estimated parameters for the M2 model in Table 4.2 are not directly comparable, due to the inclusion of the $\theta^* T_t^*$ term in (4.16). The M1 estimated parameters are comparable, and are reasonably different from the C3 parameters and the real-world parameters from Section 3.5; the high value of the M1 model's RMSE however indicates that M1 is not a good model for fitting option price data.

The fact that the C3 model does better than other models with higher degrees of freedom would seem to indicate that allowing for dependence of returns, skewness and the leverage effect may be important aspects of option pricing.

CHAPTER 5

Conclusion

There is growing recognition that returns data from various financial assets are dependent through time, and not independent as has previously been assumed. It is therefore important that models for financial assets be able to accommodate this dependence structure, which typically manifests itself in seemingly uncorrelated returns series coupled with strongly dependent squared and absolute return series. This thesis is primarily concerned with extending the VG model to allow for such a dependence structure.

In Chapter 2 we constructed a process $\{T_t\}$ whereby the increments over unit time $\{\tau_t\}$ are long range dependent, which from Chapter 1 results in a VG model for financial assets in which returns are uncorrelated or lightly correlated, but squared returns display long range dependence. In particular, we constructed increments $\{\tau_t\}$ which can have arbitrary marginal gamma distribution and an arbitrary convex correlation function, and consider as a special case $\{\tau_t\}$ with a Cauchy correlation function and for which $\{T_t\}$, appropriately normed, converges to the self similar Rosenblatt process. This result was then extended to the inverse gamma case and the t model for financial assets where we extended a result of Heyde & Leonenko (2005) to allow for $\{\tau_t\}$ with arbitrary Cauchy correlation.

In Chapter 3 we detailed a technique to simulate such dependent VG and t returns data, and compared four estimation methods to recover the (in our case known) parameter values which described the distribution of the returns. In this case we found that in most instances product-density maximum likelihood estimation was the superior method, even when the assumption of serial independence did not hold. Minimum χ^2 estimation was on average

the next best method, followed by the method of moments. ECF estimation was found to perform poorly.

In fitting to actual financial data, we found that for one heavy-tailed data set (Microsoft), the t fit was clearly inappropriate, and for another (AUD/US), while the t fit was acceptable, the VG fit was better. For the other two data sets there was little to choose between the models.

Inasmuch as all 4 data sets considered were in essence symmetric about the origin, the conclusions in Fung & Seneta (2007), that the essential difference between the symmetric VG and t distributions is the differential concentration of probability around the point of symmetry and in the middle range, was borne out by the S&P 500 and Microsoft data fits.

In Chapter 4 we detailed two new models for option pricing based on the VG process. The first relaxed a parameter constraint to describe log prices as the difference of two gamma processes. It retained many of the appealing features of the standard VG model such as its simplicity and workable characteristic function, and resulted in an improved fit to option price data.

The second model was that of a long range dependent VG process. In this case we found that the current approach to pricing options (what we have called the skew-correcting method) has shortcomings. A second martingale considered, which allowed for the flexible determination of skewness but did not accommodate the leverage effect in a flexible manner, provided a better although still disappointing fit to data. That a third construction incorporating both flexible skewness driven by volatility and dependence of squared returns produced a better fit to data than other comparable models, and even models with higher degrees of freedom, seems to indicate the importance of accommodating these phenomena in option pricing models.

References

[1] ANDERSON, T.W., & DARLING, D.A. (1936). Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *The Annals of Mathematical Statistics*, **23**, 193–212.

[2] ANH, V.V., KNOPOVA, V.P. & LEONENKO, N.N. (2004). Continuous-time stochastic processes with cyclical long-range dependence. *Australian & New Zealand Journal of Statistics*, **46**, 275–296.

[3] ANH, V.V., LEONENKO, N.N. & SAKHNO, L.M. (2007). Statistical inference using higher-order information. *Journal of Multivariate Analysis*, **98**, 706–742.

[4] ARTZNER, P. & HEATH, D. (1995). Approximate completeness with multiple martingale measures. *Mathematical Finance*, **5**, 203–228.

[5] AVRAMIDIS, A.N., L'ECUYER, P. & TREMBLAY P.A. (2003). *Efficient simulation of gamma and Variance-Gamma processes.* In: Chick, S., Sánchez, P.J., Ferrin, D. & Morrice, D.J. (Eds.), Proceedings of the 2003 Winter Simulation Conference. Piscataway, New Jersey, 319–326.

[6] AVRAMIDIS, A.N. & L'ECUYER, P. (2006). Efficient Monte Carlo and quasi-Monte Carlo option pricing under the Variance Gamma model. *Management Science*, **52**, 1930–1944.

[7] BARNDORFF-NIELSEN, O.E. & BLÆSILD P. (1981). *Hyperbolic distributions and ramifications: contributions to theory and application.* In: Taillie, C., Patil, G. & Baldessari, B. (Eds.), Statistical Distributions in Scientific Work. Riedel, Dordrecht, 19–44.

[8] BARNDORFF-NIELSEN, O.E. & HALGREEN, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **38**, 309–311.

[9] BARNDORFF-NIELSEN, O.E., JENSEN, J.L. & SØRENSEN, M. (1998). Some stationary processes in discrete and continuos time. *Advances in Applied Probability*, **30**, 989–1007.

[10] BARNDORFF-NIELSEN, O.E., KENT, J. & SØRENSEN, M. (1982). Normal variance-mean mixtures and z distributions. *International Statistical Review*, **50**, 145–159.

[11] BARNDORFF-NIELSEN, O.E. & SHEPHARD, N. (2001). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial econometrics. *Journal of the Royal Statistical Society: Series B*, **63**, 167–241.

[12] BERAN, J. (1994). *Statistics for Long-Memory Processes.* Chapman and Hall, New York.

[13] BILLINGSLEY, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons, New York.

[14] BLACK, F. & SCHOLES, M. (1973). The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, **81**, 637–654.

[15] BOCHNER, S. (1955). *Harmonic Analysis and the Theory of Probability.* University of California Press, Berkeley.

[16] CAMPBELL, J.Y., LO, A.W. & MACKINLAY, A.C. (1997). *The Econometrics of Financial Markets.* Princeton University Press, Princeton.

[17] CARR, P., GEMAN, H., MADAN, D. & YOR, M. (2003). Stochastic volatility for Lévy processes. *Mathematical Finance*, **13**, 345–382.

[18] CARR, P. & MADAN, D. (1999). Option valuation using the fast Fourier transform. *Journal of Computational Finance*, **2**, 61–73.

[19] CARR, P. & MADAN, D. (2005). A note on sufficient conditions for no arbitrage. *Finance Research Letters*, **2**, 125–130.

[20] CIFARELLI, D.M. & MELILLI, E. (2000). Some new results for Dirichlet priors. *The Annals of Statistics*, **28**, 1390–1413.

[21] CIFARELLI, D.M. & REGAZZINI, E. (1990) Distribution functions of means of a Dirichlet process. *The Annals of Statistics*, **18**, 429–442.

[22] COCHRANE, J.H. (2001). *Asset Pricing.* Princeton University Press, Princeton.

[23] COX, D.R. & REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737.

[24] DAVIS, M. & HOBSON, D. (2007). The range of traded option prices. *Mathematical Finance*, **17**, 1–14.

[25] DING, Z., GRANGER, C.W.J. & ENGLE, R. (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance*, **1**, 83–106.

[26] DING, Z. & GRANGER, C.W.J. (1996). Modeling volatility persistence of speculative returns: a new approach. *Journal of Econometrics*, **73**, 185–215.

[27] DRĂGULESCU, A. & YAKOVENKO, V. (2002). Probability distribution of returns in the Heston model with stochastic volatility. *Quantitative Finance*, **2**, 443–453.

[28] EDGEWORTH, F.Y. (1883). The method of least squares. *Philosophical Magazine*, **16**, 360–375.

[29] ERDÉLYI, A., MAGNUS, W., OBERHETTINGER, F. & TRICOMI, F.G. (1953). *Bateman Manuscript Project: Higher Transcendental Functions (Vol 2).* McGraw-Hill, New York.

[30] ERDÉLYI, A., MAGNUS, W., OBERHETTINGER, F. & TRICOMI, F.G. (1954). *Bateman Manuscript Project: Tables of Integral Transforms (Vol 2).* McGraw-Hill, New York.

[31] FAMA, E.F. (1965). The behavior of stock-market prices. *Journal of Business*, **38**, 34–105.

[32] FELLER, W. (1951). Two singular diffusion problems. *The Annals of Mathematics*, **54**, 173–182.

[33] FINLAY, R. & SENETA, E. (2006). Stationary-increment Student and Variance-Gamma Processes. *Journal of Applied Probability*, **43**, 441–453. (Correction: 1207.)

[34] FINLAY, R. & SENETA, E. (2007). A gamma activity time process with noninteger parameter and self-similar limit. *Journal of Applied Probability*, **44**, 950–959.

[35] FINLAY, R. & SENETA, E. (2008A). Stationary-increment Variance-Gamma and t models: simulation and parameter estimation. *International Statistical Review*, **76**, 167–186.

[36] FINLAY, R. & SENETA, E. (2008B). Option pricing with VG–like models. *International Journal of Theoretical and Applied Finance*, **11**, 943–955.

[37] FUJIWARA, T. & MIYAHARA, Y. (2003). The minimal entropy martingale measures for geometric Lévy processes. *Finance and Stochastics*, **7**, 509–531.

[38] FUNG, T. & SENETA, E. (2007). Tailweight, quantiles and kurtosis: a study of competing distributions. *Operations Research Letters*, **35**, 448–454.

[39] GLESER, L.J. & MOORE, D.S. (1983). The effect of dependence on chisquared and empirical distribution test of fit. *The Annals of Statistics*, **11**, 1100–1108.

[40] GNEITING, T. (2000). Power-law correlations, related models for long-range dependence and their simulation. *Journal of Applied Probability*, **37**, 1104–1109.

[41] GREENE, M.T. & FIELITZ, B.D. (1979). The effect of long term dependence on risk-return models of common stocks. *Operations Research*, **27**, 944–951.

[42] GROSSWALD, E. (1976). The Student t-distribution of any degree of freedom is infinitely divisible. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **36**, 103–109.

[43] HALGREEN, C. (1979). Self-decomposability of the generalized inverse Gaussian and hyperbolic distributions. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **47**, 13–17.

[44] HANSEN, L.P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, **50**, 1029–1054.

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

[46] HEYDE, C.C. (1999). A risky asset model with strong dependence through fractal activity time. *Journal of Applied Probability*, **36**, 1234–1239.

[47] HEYDE, C.C. & KOU, S.G. (2004). On the controversy over tailweight of distributions. *Operations Research Letters*, **32**, 399–408

[48] HEYDE, C.C. & LEONENKO, N.N. (2005). Student processes. *Advances in Applied Probability*, **37**, 342–365.

[49] HEYDE, C.C. & LIU, S. (2001). Empirical realities for a minimal description risky

asset model. The need for fractal features. *Journal of the Korean Mathematical Society (5)*, **38**, 1047–1059.

[50] HEYDE, C.C. & YANG, Y. (1997). On defining long-range dependence. *Journal of Applied Probability*, **34**, 939–944.

[51] HOGG, R. & CRAIG, A. (1978). *Introduction to Mathematical Statistics, Fourth Edition.* Macmillan, New York.

[52] HURST, S.R., PLATEN, E. & RACHEV, S.T. (1997). Subordinated market index models: a comparison. *Financial Engineering and the Japanese Markets*, **4**, 97–124.

[53] JARROW, R., JIN, X. & MADAN, D. (1999). The second fundamental theorem of asset pricing. *Mathematical Finance*, **9**, 255–273.

[54] JOHNSON, N.L., KOTZ, S. & BALAKRISHNAN, N. (1994). *Continuous Univariate Distributions, Vol. 1, 2nd edn.* John Wiley, New York.

[55] KULLBACK, S. (1936). The distribution laws of the difference and quotient of variables independently distributed in Pearson type III laws. *The Annals of Mathematical Statistics*, **7**, 51–53.

[56] LEE, R.W. (2004). Option pricing by transform methods: extensions, unification and error control. *Journal of Computational Finance*, **7**, 51–86.

[57] LEONENKO, N.N. (1999). *Limit Theorems for Random Fields with Singular Spectrum.* Kluwer Academic Publishers, Dordrecht.

[58] LEONENKO, N.N. & ANH, V.V. (2001). Rate of convergence to the Rosenblatt distribution for additive functionals of stochastic processes with long-range dependence. *Journal of Applied Mathematics and Stochastic Analysis*, **14**, 27–46.

[59] LEONENKO, N.N. & SAKHNO, L.M. (2006). On the Whittle estimators for some classes of continuous-parameter random processes and fields. *Statistics & Probability Letters*, **76**, 781–795.

[60] LIU, M. (2000). Modeling long memory in stock market volatility. *Journal of Econometrics*, **99**, 139–171.

[61] LJUNGQVIST, L. & SARGENT, T. (2004). *Recursive Macroeconomic Theory, 2nd edn.* MIT Press, Cambridge

[62] LÜROTH, J. (1876). Vergleichung von zwei Werten des wahrscheinlichen Fehlers.

Astronomische Nachrichten, **87**, 209–220.

[63] MADAN, D.B., CARR, P.P. & CHANG, E.C. (1998). The Variance-Gamma process and option pricing. *European Finance Review*, **2**, 79–105.

[64] MADAN, D.B., & EBERLEIN, E. (2007). Sato Processes and the Valuation of Structured Products. SSRN Working Paper. Available at: `http://ssrn.com/abstract=957167`.

[65] MADAN, D.B. & SENETA, E. (1987A). Simulation of estimates using the empirical characteristic function. *International Statistical Review*, **55**, 153–161.

[66] MADAN, D.B. & SENETA, E. (1987B). Chebyshev polynomial approximations and characteristic function estimation. *Journal of the Royal Statistical Society: Series B*, **49**, 163–169.

[67] MADAN, D.B. & SENETA, E. (1989). Chebyshev polynomial approximation for characteristic function estimation: some theoretical supplements. *Journal of the Royal Statistical Society: Series B*, **51**, 281–285.

[68] MADAN, D.B. & SENETA, E. (1990). The Variance Gamma (V.G.) model for share market returns. *Journal of Business*, **63**, 511–524.

[69] MCLEISH, D.L. (1982). A robust alternative to the normal distribution. *The Canadian Journal of Statistics*, **10**, 89–102.

[70] MERTON, R. (1973). Theory of Rational Option Pricing. *Bell Journal of Economics and Management Science*, **4**, 141–183.

[71] MUKHERJEA, A., RAO, M. & SUEN, S. (2006). A note on moment generating functions. *Statistics & Probability Letters*, **76**, 1185–1189.

[72] PEARSON, K., JEFFERY, G.B. & ELDERTON, E.M. (1929). On the distribution of the first product-moment coefficient in samples drawn from an indefinitely large normal population. *Biometrika*, **21**, 164–201.

[73] PEARSON, K., STOUFFER, S. & DAVID, F. (1932). Further applications in statistics of the $T_m(x)$ Bessel function. *Biometrika*, **24**, 293–350.

[74] PRAETZ, P.D. (1972). The distribution of share price changes. *Journal of Business*, **45**, 49–55.

[75] PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T., & FLANNERY, B.P. (1992).

Numerical Recipes in C, Second Edition. Cambridge University Press, Cambridge.

[76] SCHOUTENS, W. (2003). *Lévy Processes in Finance. Pricing Financial Derivatives.* John Wiley & Sons, England.

[77] SENETA, E. (2004). Fitting the Variance-Gamma model to financial data. *Stochastic Methods and their Applications*, Journal of Applied Probability Special Volume 41(A), 177–187.

[78] SENETA, E. (2007). *The early years of the Variance-Gamma process.* In: Fu, M.C., Jarrow, R.A., Yen, J.J. & Elliott, R.J. (Eds.), Advances in Mathematical Finance. Birkhäuser, Boston, 3–19.

[79] SENETA, E. (2008). Comment: On Student's 1908 paper "The probable error of the mean". *Journal of the American Statistical Association*, **103**, 13–15.

[80] SENETA, E. (2009). *Variance Gamma model.* In: Cont, R. & Gatheral, J. (Eds.), Encyclopedia of Quantitative Finance. John Wiley & Sons, Chichester, to appear.

[81] SLY, A. (2006). *Self-similarity, Multifractionality and Multifractality.* Master of Philosophy thesis, Australian National University.

[82] STUDENT (1908). The probable error of a mean. *Biometrika*, **6**, 1–25.

[83] SZEGÖ, G. (1959). *Orthogonal Polynomials (Vol 23).* American Mathematical Society, New York.

[84] SØRENSEN, M. & BIBBY, M. (2003). *Hyperbolic processes in finance.* In: Rachev, S.T. (Ed.), Handbook of Heavy Tailed Distributions in Finance. Elsevier, Amsterdam, 211–248.

[85] TAQQU, M.S. (1975). Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **31**, 287–302.

[86] TAQQU, M.S. (1979). Convergence of integrated processes of arbitrary Hermite rank. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **50**, 53–83.

[87] TAQQU, M.S. & LEVY, J.B. (1986). *Using renewal processes to generate long range dependence and high variability.* In: Eberlein, E. & Taqqu, M.S. (Eds.), Dependence in Probability and Statistics: A Survey of Recent Results. Birkhauser, Boston, 73–89.

[88] TAYLOR, S.J. (1986). *Modelling Financial Time Series.* John Wiley & Sons, New

York.

[89] TAYLOR, S.J. (1994). Modeling stochastic volatility: a review and comparative study. *Mathematical Finance*, **4**, 183–204.

[90] TEICHROEW, D. (1957). The mixture of normal distributions with different variances. *The Annals of Mathematical Statistics*, **28**, 510–512.

[91] TJETJEP, A. & SENETA, E. (2006). Skewed normal variance-mean models for asset pricing and the method of moments. *International Statistical Review*, **74**, 109–126.

[92] WILLINGER, W., TAQQU, M. & TEVEROVSKY V. (1999). Stock price return indices and long-range dependence. *Finance and Stochastics*, **3**, 1–13.

[93] YEO, G.F. & MILNE, R.K. (1991). On characterizations of beta and gamma distributions. *Statistics & Probability Letters*, **11**, 239–242.

[94] YOR, M. (2007). *Some remarkable properties of gamma processes.* In: Fu, M.C., Jarrow, R.A., Yen, J.J. & Elliott, R.J. (Eds.), Advances in Mathematical Finance. Birkhäuser, Boston, 37–47.

[95] YU, J. (2004). Empirical characteristic function estimation and its applications. *Econometric Reviews*, **23**, 93–123.

[96] ZABELL, S.L. (2008). On Student's 1908 paper "The probable error of the mean". *Journal of the American Statistical Association*, **103**, 1–7.