Financial volatility, Lévy processes and power variation

Ole E. Barndorff-Nielsen
The Centre for Mathematical Physics and Stochastics (MaPhySto), University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark.
oebn@mi.aau.dk

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

Our papers on this subject are all available at
www.levyprocess.org

This book mauscript is incomplete, comments are very welcome.

June 2002

Contents

Chapter 1

Introduction

Chapter 2

Basics of Lévy processes

Abstract: This Chapter provides a first introduction to the use of Lévy processes as models of log-prices in financial markets, focusing on the probabilistic aspects. Univariate and multivariate models are discussed. A detailed bibliographical review is given at the end of the Chapter. It is important to keep in mind that Lévy processes allow a flexible model for the marginal distribution of returns, but still maintain that returns are iid. Obviously this is a very poor description of reality and later on in our book we extend this framework to allow for stochastic volatility. However, a significant understanding of these processes does help our understanding both of the economics and the development of the later techniques given in the book.

2.1 What is this Chapter about?

In this Chapter we provide a first course on Lévy processes in the context of financial economics. The focus will be on probabilistic and econometric issues; understanding the models and their fit to returns on speculative assets. We leave until our Second book the vital issue of how these models can guide investors in their the allocation of resources between risky and riskless assets and the pricing of derivatives written on Lévy processes. The Chapter will refer to some common datasets discussed in detail in Chapter 1 and will delay the discussion of literature on this topic until the end of this Chapter. Throughout we hope our treatment will be as self-contained as possible. At the end of this book we have given a brief introduction, called "The Primer," to stochastic analysis which may be of help to readers without a strong background in probability theory.

This long Chapter has 8 other sections, whose goals are to:

- Introduce Lévy processes with non-negative increments.
- Extend the analysis to Lévy processes with real increments.
- Introduce time deformation, or time change, where we replace calendar time by a random clock.
- Introduce quadratic variation, a central concept in econometrics and stochastic analysis.
- Brief discuss stochastic analysis in the context of Lévy processes.
- Introduce various methods for building multivariate Lévy processes.
- Draw conclusions to the Chapter.
- Discuss the literature associated with Lévy processes.

This Chapter leads into the next one, which will focus on methods for simulating the paths of Lévy processes and the estimation and testing of these models on financial time series.

2.2 What is a Lévy process?

2.2.1 The random walk

The most basic model of the logarithm of the price of a risky asset is a random walk. It is built by summing independent and identically distributed (i.i.d.) random variables $c_0, c_1, ...$ to deliver

$$z_{n+1} = \sum_{s=0}^{n} c_s, \quad \text{with} \quad z_0 = 0, \quad n = 0, 1, 2, ... \ .$$

The process is written in discrete time and is moved by the i.i.d. increments

$$z_{n+1} - z_n = c_n. \tag{2.1}$$

Hence future changes in a random walk are unpredictable.

Random walks live in discrete time. What is the natural continuous time version of this process? There are at least two strong answers to this question.

2.2.2 Brownian motion

The first approach is based on a central limit type result. Again suppose that $\{c_s\}$ is an i.i.d. sequence whose first two moments exist. Then define the partial sum

$$z_T(t) = \sqrt{T}\frac{1}{T}\sum_{s=1}^{[tT]}\{c_s - \mathrm{E}(c_s)\}, \qquad t \geq 0 \tag{2.2}$$

where t represents time. It means that over any fixed interval for t, that is time, the process is made up of centred and normalised sums of i.i.d. events. We then allow T, the number of these events in any fixed interval of time of unit length, to go off to infinity (this is often labelled "in-fill" asymptotics). As a result $z_T(t)$ obeys a central limit theory and becomes Gaussian. Further, this idea can be extended to show that the whole partial sum, as a random function, converges to a scaled version of Brownian motion, as T goes to infinity. At first sight this suggests the only reasonable continuous time version of a random walk, which will sum up many small events, is Brownian motion. This insight is, however, incorrect.

2.2.3 Infinite divisibility

Our book follows a second approach. Suppose that the goal is to design a continuous time process at time 1, $z(1)$, which has a distribution D. It may be possible to divide the time from zero until one into T pieces, each of which has independent increments from a common distribution $D^{(T)}$ such that the sum

$$z(t) = \sum_{s=1}^{[tT]} c_s^{(T)}, \qquad \text{where} \qquad c_s^{(T)} \overset{i.i.d.}{\sim} D^{(T)},$$

has the distribution D when $t = 1$. Then as T increases we imagine that the division of time between zero and one becomes ever finer. In response, the increments and their distribution $D^{(T)}$ also change, but by construction D, the distribution of the sum, is left unchanged. A simple example of this is where $z(1) \sim Po(1)$, then if

$$z(t) = \sum_{s=1}^{[tT]} c_s^{(T)}, \qquad \text{where} \qquad c_s^{(T)} \overset{i.i.d.}{\sim} Po(1/T),$$

this produces a valid random walk due to the fact that the independent Poisson increments sum to a Poisson. Hence this process makes sense even as T goes to infinity and so this type of construction can be used as a continuous time model — the *Poisson process*. The class of distributions for which this construction is possible is those for which D is *infinitely divisible*. The resulting processes are called Lévy processes. Examples of infinitely divisible distributions include, focusing for the moment on only non-negative random variables, the Poisson, gamma, reciprocal gamma, inverse Gaussian, reciprocal inverse Gaussian, F and positive stable distributions.

2.2.4 The definition of a Lévy process

The natural continuous time version of the discrete time increment given in (2.1) is, for any value of $\Delta > 0$,

$$z(t+\Delta) - z(t), \qquad t \in [0, \infty].$$

Increments play a crucial role in the formal definition of a Lévy process.

Definition 1 *Lévy process. The stochastic process*

$$z(t), \quad t \in [0, \infty], \quad z(0) = 0,$$

is a Lévy process if and only if it has independent and (strictly) stationary increments.

In the definition the first assumption means that the shocks to the process are independent over time and that they are summed, while the second assumption means that the distribution of $z(t+\Delta) - z(t)$ may change with Δ but does not depend upon t. The independence and stationarity of the increments of the Lévy process means that

$$\begin{aligned} \mathrm{C}\left\{\theta \ddagger z(t)\right\} &= \log\left[\mathrm{E}\exp\left\{i\theta z(t)\right\}\right] \\ &= t\log\left[\mathrm{E}\exp\left\{i\theta z(1)\right\}\right] \\ &= t\mathrm{C}\left\{\theta \ddagger z(1)\right\}, \end{aligned}$$

so the distribution of $z(t)$ is completely governed by the cumulants of $z(1)$, the value of the process at time one.

If a Lévy process is used as a model for the log-price of an underlying asset then the increments can be thought of as returns. Consequently Lévy based models provide a potentially flexible framework with which to model the marginal distribution of returns. However, returns will be, measured over a fixed value of Δ, independent and identically distributed. This important observation will imply that Lévy processes can only ever be a rather partially realistic model of asset prices as asset price returns have important serial dependencies such as volatility clustering. This point will be discussed at more length in the fourth chapter of this book. For now we introduce the formal development of Lévy processes.

2.3 Processes with non-negative increments — subordinators

2.3.1 Examples of Lévy processes

Motivation

We start with Lévy processes with non-negative increments. Such processes are often called *subordinators*. This is our focus for two reasons: (i) they are mathematically considerably simpler, (ii) most of models we build in this book will have components which are Lévy processes with non-negative increments and so they are a major concern to us. The discussion of processes on the real line will be given in the next section. In order to reduce the technical demands on the reader we are mostly going to use cumulant functions in this context as this is sufficient for our purposes. As the processes are positive, it is natural to work with the *kumulant function* in the form

$$\overline{\mathrm{K}}\left\{\theta \ddagger z(1)\right\} = \log\left[\mathrm{E}\exp\left\{-\theta z(1)\right\}\right], \qquad \text{where} \qquad \theta \geq 0.$$

Occasionally the more standard cumulant function

$$\mathrm{K}\left\{\theta \ddagger z(1)\right\} = \log\left[\mathrm{E}\exp\left\{\theta z(1)\right\}\right], \qquad \text{where} \qquad \theta \geq 0,$$

will be used.

Poisson process

Introduction Suppose we count the number of events which have occurred from time 0 until $t \geq 0$. The very simplest continuous time model for this type is a Lévy process with independent Poisson increments

$$z(1) \sim Po(\psi), \quad \psi > 0,$$

with density

$$f_{z(1)}(x) = \frac{e^{-\psi}\psi^x}{x!}, \qquad x = 0, 1, 2, ...$$

The process $\{z(t)\}_{t\geq 0}$ is called a (homogeneous) *Poisson process*. Here the mean ψ is often called the intensity of this counting process. A simulated sample path of this process, when $\psi = 1$, is given in Figure 2.1(a). It shows a jumping process, where the jumps are irregularly spaced in time and are of equal size. The times at which events occur are called *arrival times*, and are written as $\tau_1, \tau_2, ..., \tau_{z(t)}$.

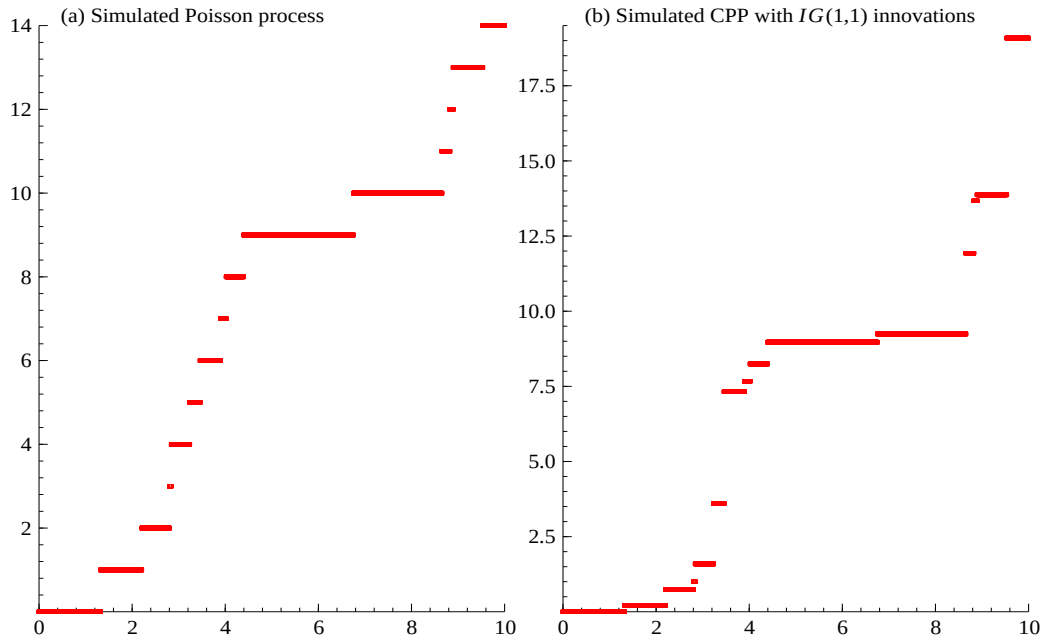

Figure 2.1: *(a) Sample path of a homogeneous Poisson process with intensity $\psi = 1$. Horizontal axis is time t, vertical is $z(t)$. (b) Corresponding compound Poisson process with $c_s \sim IG(1, 1)$. code:* `levy_graphs.ox`.

For the Poisson process

$$\begin{aligned} \overline{\mathrm{K}}\{\theta \ddagger z(1)\} &= \log\left[\mathrm{E}\exp\{-\theta z(1)\}\right] \\ &= \psi(e^{-\theta} - 1) \\ &= -\psi(1 - e^{-\theta}). \end{aligned}$$

Now we know that $-\psi t(1-e^{-\theta})$ corresponds to a cumulant function for a $Po(t\psi)$, which implies that the Poisson distribution is infinitely divisible. Indeed $z(t) \sim Po(t\psi)$.

Poisson process is càdlàg An important feature of some stochastic processes, including Brownian motion, is that they have continuous sample paths with probability one. A weaker assumption is the one exhibited by the Poisson process. This allows jumps but is, with probability

one, right continuous

$$\lim_{s \downarrow t} z(s) = z(t)$$

and has limits from the left

$$z(t-) = \lim_{s \uparrow t} z(s).$$

For such processes the jump just before time t is written as

$$\Delta z(t) = z(t) - z(t-).$$

This notation clashes with our use of Δ to stand for a time interval.

We might expect these types of jumps to appear in financial processes due to dividend payments or news, such as macroeconomic announcements. A process with this mathematical property is called a càdlàg (continue à droit, limite à gauche) or a RCLL (right continuous left limit) process in the literature. It causes no restrictions as regard the finite dimensional distributions and we follow the convention here that all Lévy processes are càdlàg, unless otherwise stated. The similiarly named property càglàd (continue à gauche, limite à droit) plays an important role in our Appendix on stochastic analysis.

Poisson process is a special semimartingale Semimartingales play a central role in modern stochastic analysis, in particular providing a basis for the definition of a stochastic integral. Consequently it is important to note that a Poisson process is a semimartingale. We see this by noting that any semimartingale x can be decomposed into

$$x(t) = a(t) + m(t), \tag{2.3}$$

where a is of local bounded variation and m is a local martingale. Here we have assumed $z(0) = a(0) = m(0) = 0$. This result is discussed at some length in our Primer on stochastic analysis at the end of this Chapter. When $a(t)$ is also a predictable process then the semimartingale is said to be special and the decomposition (2.3) is unique and is called canonical. Informally a process is said to be predictable if its value at time t is known an instant before t, given the filtration. Hence all deterministic functions of time are predictable, as are càglàd processes.

In the case of the Poisson process we can see that z can be decomposed in this way by writing

$$a(t) = \mathrm{E}\left\{z(t)\right\} = \psi t \qquad \text{and} \qquad m(t) = z(t) - \psi t.$$

The m process, which is often called a compensated Poisson process, is a mean-0 martingale, for $\mathrm{E}(m(t)|\mathcal{F}_t) = 0$. It will turn out that all Lévy processes for whom $\mathrm{E}\left\{z(t)\right\}$ exists are special semimartingales. This is an important result.

As z is a Lévy process the theory of semimartingales implies that if h is locally square integrable then the *stochastic integral*

$$y(t) = \int_0^t h(u-)\mathrm{d}z(u),$$

can be constructed. This is often written in the more abstract notation as the process

$$y = h \bullet z.$$

The integral can often be defined as the limit in probability of finite sums of the form

$$\sum_{i=0}^{n-1} h(u_i)(z(u_{i+1}) - z(u_i)),$$

where $0 = u_0 < u_1 < ... < u_n = t$. Any such integral process y is itself a semimartingale. More formal details of stochastic integrals are given in our Primer. When z is a Poisson process y simplifies to the random sum

$$y(t) = \sum_{j=1}^{z(t)} h(\tau_j -)$$

where $\tau_1, \tau_2, ..., \tau_{z(t)}$ are the arrival times of z.

In the case of a special semimartingale we can view $\mathrm{d}a(t)$ informally as the continuously updated expected return conditional on the information just before t, while $\mathrm{d}x(t)$ is the return and $\mathrm{d}m(t)$ is the unanticipated return. Without predictability on $a(t)$ this decomposition is not possible. Thus, from an economic viewpoint special semimartingales seem of fundamental importance.

Ito formula for a Poisson process Suppose we wish to look at a function of some stock price process s, $y(t) = f(s(t))$ where $s(t) = \mu t + z(t)$, which is drift plus a Poisson process z. This is a continuous time version of a binomial tree model frequently used in financial economics[1]. Ito's formula for semimartingales given in (A.11) applies here, for all Lévy processes are semimartingales. In this case

$$y(t) = y(0) + \int_0^t f'\{s(u-)\}\,\mathrm{d}s(u) + \sum_{0<u\leq t} \{f\{s(u)\} - f\{s(u-)\} - f'\{s(u-)\}\,\Delta u\}, \qquad (2.4)$$

which can be written in the form of a stochastic differential equation (SDE) as

$$\mathrm{d}y(t) = f'\{s(t-)\}\,\mathrm{d}s(t) + \left[f\{s(t)\} - f\{s(t-)\} - f'\{s(t-)\}\,\Delta s\right].$$

This can be simplified as $\mathrm{d}s(t) = \mu \mathrm{d}t + \Delta z(t)$ and $\Delta s(t) = \Delta z(t)$ to

$$\mathrm{d}y(t) = f'\{s(t-)\}\,\mu \mathrm{d}t + f\{s(t)\} - f\{s(t-)\}.$$

The above analysis is interesting for it means that the portfolio $V = y - \delta s$, where $\delta = f\{s(t-) + 1\} - f\{s(t-)\}$ has the important property that, if a riskless interest rate r exists, then

$$\mathrm{d}V = \mathrm{d}y - \delta \mathrm{d}s = f'\{s(t-)\}\,\mu \mathrm{d}t$$

is instantly riskless and so the porfolio must grow at the riskless rate $\mathrm{d}V = rV\mathrm{d}t$. As this argument does not need us to specify a utility function we can price y as if all agents were risk neutral (Cox and Ross (1976)) and so price contingent assets not under s but under the risk neutral process $rt + \{z(t) - \mathrm{E}z(t)\}$. This argument is simply a continuous time version of a standard binomial tree, for it relies on knowing that the value of the portfolio will either jump or not jump by one unit. It implies that familiar Black-Scholes analysis, based on Brownian motion and hedging, is not upset by jumps of known size.

Equivalent martingale measure (EMM) An *equivalent measure* is defined as a measure Q_t such that the *Radon-Nikodym derivative* $\mathrm{d}Q_t/\mathrm{d}P_t$ is strictly positive, where P_t is the measure of the original z process from time 0 up to time t. It implies that the support of the process under Q_t must be the same as the original process for the z process. This means Q must correspond to the measure generated by a *counting process* to be an equivalent measure. However, equivalence does not constrain the intensity of the counting process and so there are an infinite number of equivalent measures for this problem. This is standard in financial economics.

[1]This exposition was suggested to us by George Konaris, who we thank for allowing us to use it here.

In order for Q_t to be an *equivalent martingale measure (EMM)* we additionally need

$$\mathrm{E}_{Q_t}\left\{e^{-r(t-s)}z(t)|\mathcal{F}_{s-}\right\} = z(s-), \tag{2.5}$$

that is the discounted price process must be a martingale under Q_t. Thus we have to find the intensity (which can depend upon the filtration) which deliver this martingale. By taking t to be very close to s in (2.5) it is easy to see there is only one possibility when $z(t-) > 0$,

$$\psi(t) = rz(t-),$$

so the intensity will increase with the level of the Poisson process.

Compound Poisson process

Introduction Suppose $\{N(t)\}_{t\geq 0}$ is a Poisson process and $\{c_s\}$ is an i.i.d. sequence. Then define a *compound Poisson process* as

$$z(t) = \sum_{s=1}^{N(t)} c_s, \quad \text{where} \qquad z(0) = 0 \qquad \text{and} \qquad \sum_{s=1}^{0} c_s = 0. \tag{2.6}$$

That is $z(t)$ is made up of the addition of a random number $N(t)$ of i.i.d. random variables. This is a Lévy process for the increments of this process

$$z(t+\Delta) - z(t) = \sum_{s=N(t)+1}^{N(t+\Delta)} c_s$$

are independent and are stationary as the increments of the Poisson process are independent and stationary.

At this point it is important to note that there is no added flexibility if the distribution of the $\{c_s\}$ is allowed to have an atom at zero for this would, in effect, just knock out or *thin* some of the Poisson process arrivals. Hence this is ruled out a priori. This point will recur in our later exposition. It is informative to note that

$$\begin{aligned}
\overline{\mathrm{K}}\left\{\theta \ddagger z(1)\right\} &= \log\left[\mathrm{E}\exp\left\{-\theta z(1)\right\}\right] \\
&= \log\left[\mathrm{E}_{N(1)}\mathrm{E}\exp\left\{-\theta z(1)\right\}|N(1)\right] \\
&= \log\left\langle \mathrm{E}\exp\left[N(1)\overline{\mathrm{K}}\left\{\theta \ddagger c_1\right\}\right]\right\rangle \\
&= \mathrm{K}\left\{\overline{\mathrm{K}}\left(\theta \ddagger c_1\right) \ddagger N(1)\right\} \\
&= -\psi\left\{1 - \exp\overline{\mathrm{K}}\left(\theta \ddagger c_1\right)\right\}.
\end{aligned}$$

Example 1 *Figure 2.1(b) gives a simulation using* $c_s \overset{i.i.d.}{\sim} IG(1,1)$*, taking exactly the same Poisson process draws as used in Figure 2.1(a). The resulting* $\overline{\mathrm{K}}\left\{\theta \ddagger z(1)\right\}$ *is*

$$-\psi\left\langle 1 - \exp\left[\delta\left\{\gamma - \left(\gamma^2 + 2\theta\right)^{1/2}\right\}\right]\right\rangle.$$

Ito formula for a compound Poisson process Suppose $y(t) = f(s(t))$ where $s(t) = \mu t + z(t)$. This case repeats the SDE from the Poisson process example which again simplifies to

$$\mathrm{d}y(t) = f'\left\{s(t-)\right\}\mu\mathrm{d}t + \left[f\left\{s(t)\right\} - f\left\{s(t-)\right\}\right].$$

However, now

$$\begin{aligned} f\{s(t)\} - f\{s(t-)\} &= \begin{cases} 0, & \text{if} \quad \Delta N(t) = 0 \\ f\left\{s(t-) + c_{N(t)}\right\} - f\{s(t-)\}, & \text{if} \quad \Delta N(t) = 1. \end{cases} \\ &= \left[f\left\{s(t-) + c_{N(t)}\right\} - f\{s(t-)\}\right] \Delta N(t). \end{aligned}$$

Here $c_{N(t)}$ is the random innovation of the compound Poisson process.

When we construct a portfolio $V = y - \delta s$ it is not possible to find a δ to make $\mathrm{d}V$ deterministic for $c_{N(t)}$ is not known at time $t-$. This is a crucial observation for it implies any compound Poisson process yields an incomplete financial market and prices of contingent claims cannot be determined simply using hedging. Some introduction of a utility function will be necessary. This will be discussed in detail in our second book.

Examples of infinite activity Lévy processes

Gamma process The Poisson process and the compound process are by far the most well-known non-negative Lévy processes. The jumps happen, typically, rather rarely. Consequently increments to these processes are often exactly zero, even when measured over quite large time intervals. This feature of the process is fundamentally different from the gamma Lévy process. A *gamma Lévy process* z makes $z(1)$ obey a gamma law

$$z(1) \sim \Gamma(\nu, \alpha), \qquad \nu, \alpha > 0,$$

with density

$$f_{z(1)}(x) = \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} \exp(-\alpha x), \qquad x > 0. \tag{2.7}$$

Here 2ν is thought of as a degrees of freedom parameter, controlling the skewness of the distribution. The other parameter, α, is a scale parameter.

The kumulant function of the gamma distribution is

$$\overline{\mathrm{K}}(\theta \ddagger z(1)) = \nu \log\left(1 + \frac{\theta}{\alpha}\right),$$

which implies $z(t) \sim \Gamma(\nu t, \alpha)$. The gamma process has the useful property that it has increments which are strictly positive whatever small time interval has elapsed. Such Lévy processes are said to have *infinite activity*. This feature puts them apart from a compound Poisson process.

A sample path of a gamma process is drawn in Figure 2.2(a). The path is not continuous (anywhere). It was drawn by splitting time into intervals of length $1/2000$ and sampling from the implied random walk with $\Gamma(\nu/2000, \alpha)$ distribution. Very similar paths are produced by using smaller time intervals. The process is a rough upward trend with occasional large shifts.

Inverse Gaussian process An *inverse Gaussian* (IG) process z requires the Lévy process at time one to follow an inverse Gaussian distribution $z(1) \sim IG(\delta, \gamma)$, where $\delta > 0, \gamma \geq 0$, with density

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

This implies

$$\overline{\mathrm{K}}(\theta \ddagger z(1)) = \delta\left\{\gamma - \left(\gamma^2 + 2\theta\right)^{1/2}\right\}.$$

Like the gamma process, the IG process has an infinite number of jumps in any small interval of time. Most of these jumps are tiny, while occasionally there are larger upwards movements. The form of the cumulant function implies $z(t) \sim IG(t\delta, \gamma)$.

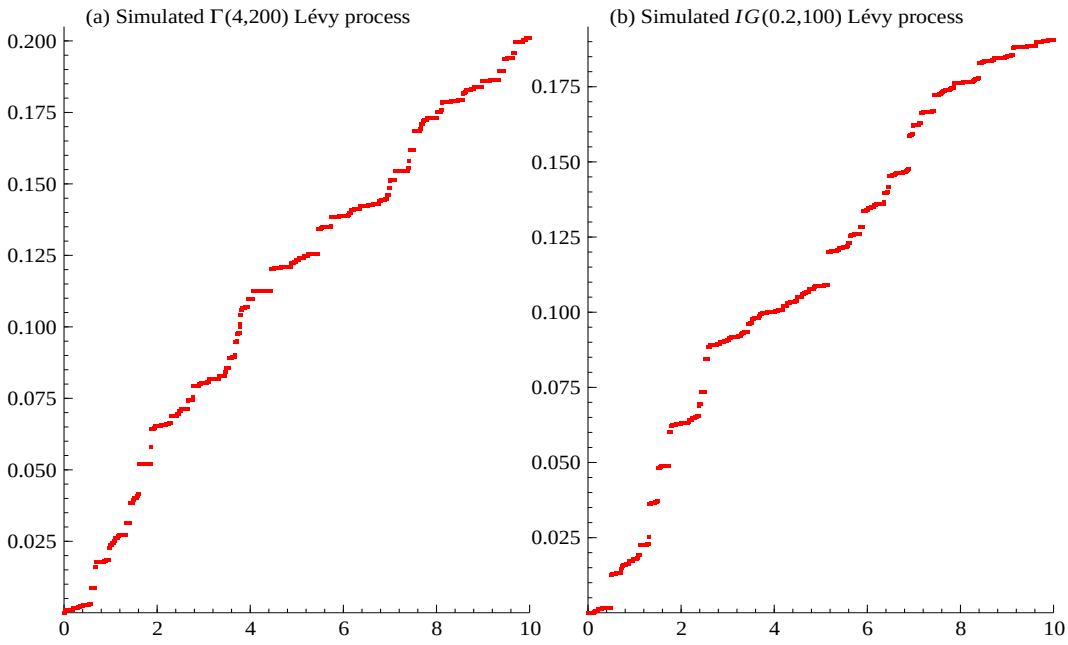

Figure 2.2: *Simulated* Γ *and IG Lévy processes, using intervals of length 1/2000. Code* `levy_graphs.ox`.

A sample path of an IG Lévy process is drawn in Figure 2.2(b). The parameters were selected to have the same mean and variance of $z(1)$ as that used to draw the path of the gamma process given in Figure 2.2. Again the process is a rough upward trend with occasional large shifts.

Some other non-negative processes A *reciprocal* (inverse) *gamma* ($R\Gamma$) process z requires the Lévy process at time one to be a reciprocal gamma variable $z(1) \sim R\Gamma(\nu, \alpha)$, $\alpha, \nu > 0$,with density

$$f_{z(1)}(x) = \frac{\alpha^\nu}{\Gamma(\nu)x^{\nu+1}} \exp\left(-\alpha x^{-1}\right), \qquad x > 0.$$

We should note that, by construction, $z(1)^{-1} \sim \Gamma(\nu, \alpha)$. Only the moments of order less than ν exist for this distribution.

Problematically, sums of independent reciprocal gamma variables are not distributed as reciprocal gamma. However, it has been shown by rather involved methods that the reciprocal gamma is infinitely divisible (and so yields a Lévy process), although we do not know the distribution of $z(t)$ in closed form. This makes simulation of this process more difficult, however we will show later that we can use computationally intensive methods to simulate the process.

A *lognormal* (LN) process z requires the Lévy process at time one to be a lognormal variable $z(1) \sim LN(\mu, \sigma^2)$, $\sigma^2 \geq 0$, with density

$$f_{z(1)}(x) = \frac{1}{x\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma^2}\left(\log x - \mu\right)^2\right\}, \quad x > 0.$$

The proof that the lognormal is infinitely divisible is probabilistically challenging. However it has been established and so the lognormal provides a valid basis for a Lévy process even though sums of independent lognormals are not lognormal.

A *reciprocal* (inverse) Gaussian (RIG) process z requires the Lévy process at time one to be

a reciprocal inverse Gaussian variable $z(1) \sim RIG(\delta, \gamma)$, $\delta > 0$, $\gamma \geq 0$, with density

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

The corresponding kumulant function is

$$\overline{\mathrm{K}}(\theta \ddagger z(1)) = -\frac{1}{2}\log\left(1 + 2\theta/\gamma^2\right) + \delta\gamma\left\{1 - \left(1 + 2\theta/\gamma^2\right)^{1/2}\right\}.$$

We should note that, by construction, $z(1)^{-1} \sim IG(\gamma, \delta)$. Again sums of independent RIG variables are not distributed as RIG, however we show in Part II that the RIG distribution is infinitely divisible and so supports a Lévy process.

Similar remarks hold for the *positive hyperbolic* (PH) process which requires that $z(1) \sim PH(\delta, \gamma)$, $\delta > 0$, $\gamma \geq 0$, with density

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

Here $K_1(\cdot)$ is a modified Bessel function of the third kind. All of the moments of this infinitely divisible distribution do exist if $\gamma > 0$, however, the distribution of $z(t)$ is unknown (except when $t = 1$) and we cannot directly simulate from it without using intensive methods.

In the special case of $\delta \to 0$, then

$$f_{z(1)}(x) = \frac{\gamma^2}{2} \exp\left(-\frac{1}{2}\gamma^2 x\right), \quad x > 0,$$

which is the exponential distribution $\Gamma(1, \gamma^2/2)$. This is also a special case of the gamma process given in (2.7). Indeed at time t the exponential Lévy process has $z(t) \sim \Gamma(t, \gamma^2/2)$.

We will call the inverse of a positive hyperbolic random variable a *reciprocal positive hyperbolic* (RPH). It leads to an RPH process which requires that $z(1) \sim RPH(\delta, \gamma)$, $\delta > 0$, $\gamma \geq 0$, with density

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

This distribution is again infinitely divisible but is again difficult to work with.

Generalised inverse Gaussian process The above infinite activity processes are all special cases of the *generalized inverse Gaussian* (GIG) process. This puts

$$z(1) \sim GIG(\nu, \delta, \gamma),$$

with GIG density

$$f_{z(1)}(x) = \frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} x^{\nu-1} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}, \quad x > 0, \tag{2.8}$$

where again $K_\nu(\cdot)$ is a modified Bessel function of the third kind. This density has been shown to be infinitely divisible and so supports a whole nesting class of Lévy processes. Prominent special cases are achieved in the following ways:

$$\begin{array}{ll} IG(\delta, \gamma) = GIG(-\frac{1}{2}, \delta, \gamma), & PH(\delta, \gamma) = GIG(1, \delta, \gamma), \\ R\Gamma(\nu, \delta^2/2) = GIG(-\nu, \delta, 0), & \Gamma(\nu, \gamma^2/2) = GIG(\nu > 0, 0, \gamma), \\ RIG(\delta, \gamma) = GIG(\frac{1}{2}, \delta, \gamma), & PHA(\delta, \gamma) = GIG(0, \delta, \gamma) \\ RPH(\delta, \gamma) = GIG(-1, \delta, \gamma). & \end{array}$$

Here all these distributions are familiar except for the positive hyperbola distribution, which is denoted PHA. In order to obtain these results we have to allow δ or γ to be zero 0. In these cases the GIG's density has to be interpreted in the limiting sense, using the well-known results that for $x \downarrow 0$ we have

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

In general, of course, the distribution of the increments to this process are unknown and we cannot directly simulate from it without using intensive methods.

2.3.2 Lévy measures for non-negative processes

Cumulant function

It should be clear by now that the cumulant function of $z(1)$ plays an important role in Lévy processes. In this subsection this observation will be further developed in order to build towards the vital Lévy-Khintchine representation which shows us the form characteristic functions of Lévy processes must obey. As this representation is so important, and is also mathematically involved, development will be carried out in stages. At first sight this looks unnecessary from a modelling viewpoint, however we will see that practical modelling will sometimes be carried out directly via some of the terms which make up the Lévy-Khintchine representation. Hence a good understanding of this section is essential for later developments.

Poisson and compound processes

To start off with think of a Poisson process, so that $z(1) \sim Po(\psi)$. Then, writing $\delta_1(x)$ as the Dirac delta centred at $x = 1$, we write

$$\begin{aligned} \overline{\mathrm{K}}\{\theta \ddagger z(1)\} &= -\psi(1 - e^{-\theta}) \\ &= -\psi \int_0^\infty (1 - e^{-\theta x})\delta_1(x)dx, \\ &= -\psi \int_0^\infty (1 - e^{-\theta x})P(\mathrm{d}x), \end{aligned}$$

where P is the Dirac delta probability measure centred at one. The introduction of the probability measure is entirely expository in this context, however expressing kumulant functions in this type of way will become essential later. Before proceeding another level of abstraction has to be introduced. Instead of working with probability measures we will have to use more general measures W concentrated on R_+. An important point is that some of the measures that will be used later will not be integrable (that is $\int_0^\infty W(\mathrm{d}x) = \infty$) and so probability measures are insufficient for a discussion of Lévy processes. In the simple Poisson case measures are introduced by expressing

$$\overline{\mathrm{K}}\{\theta \ddagger z(1)\} = -\int_0^\infty (1 - e^{-\theta x})W(\mathrm{d}x) \tag{2.9}$$

where $W = \psi\delta_1$ is called the *Lévy measure*. Of course this measure is integrable, indeed it integrates to ψ.

Let us now generalise the above setup to the compound Poisson process (2.6), but still requiring $\{c_s\}$ to be strictly positive — ruling out the possibility that c_s can be exactly zero with non-zero probability. Then, writing the distribution function of c_1 as $P(x \ddagger c_1)$,

$$\overline{\mathrm{K}}\{\theta \ddagger z(1)\} = -\psi \left\{1 - \exp \overline{\mathrm{K}}\left(\theta \ddagger c_1\right)\right\}$$

$$
\begin{aligned}
&= -\psi \int_0^\infty (1 - e^{-\theta x}) P(\mathrm{d}x \ddagger c_1) \\
&= -\int_0^\infty (1 - e^{-\theta x}) W(\mathrm{d}x),
\end{aligned}
$$

again, but now with $W(\mathrm{d}x) = \psi P(\mathrm{d}x \ddagger c_1)$. Again this measure is integrable as it is proportional to the probability measure. In the simple case where c_1 has a density we write

$$W(\mathrm{d}x) = w(x)\mathrm{d}x$$

and call $w(x)$ (which is ψ times the density of c_1) the *Lévy density*. In such cases the kumulant function becomes

$$\overline{\mathrm{K}}\{\theta \ddagger z(1)\} = -\int_0^\infty (1 - e^{-\theta x}) w(x)\mathrm{d}x.$$

A simple example of this is where $c_s \overset{i.i.d.}{\sim} \Gamma(\nu, \alpha)$. Then the Lévy density is

$$w(x) = \psi \frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} \exp(-\alpha x).$$

Of course this Lévy density integrates to ψ — not one.

Although Poisson and compound processes have integrable Lévy measures (for W is proportional to a probability measure which integrates to one) theoretically more general Lévy processes can be constructed without abandoning the form (2.9). The non-integrable measures W will not correspond to compound Poisson processes. To ensure that they yield a valid kumulant function we require that $\int_0^\infty (1 - e^{-\theta x}) W(\mathrm{d}x)$ exists, while continuing to rule out the possibility that W has an atom at zero. It is simple to prove that a necessary and sufficient condition for $\int_0^\infty (1 - e^{-\theta x}) W(\mathrm{d}x)$ to exist is that

$$\int_0^\infty \min(1, x)\, W(\mathrm{d}x) < \infty.$$

If the Lévy measure is absolutely continuous then we can define $w(x)$ as the Lévy density where $W(\mathrm{d}x) = w(x)\mathrm{d}x$. However, as some Lévy measures are not integrable, it follows that Lévy densities are not necessarily integrable. This is at first sight confusing. This point comes up in the following two examples.

Example 2 *It turns out that the Lévy density of $z(1) \sim IG(\delta, \gamma)$ is*

$$w(x) = (2\pi)^{-1/2} \delta x^{-3/2} \exp(-\gamma^2 x/2), \quad x \in R_+. \tag{2.10}$$

This Lévy density is not integrable as it goes off to infinity too rapidly as x goes to zero. This is important for it implies an IG process is not a compound Poisson process. Although the Lévy density is not integrable it does satisfy the finiteness condition on $\int_0^\infty \min(1, x)\, W(\mathrm{d}x)$ for the addition of the x factor regularises the density near zero (it behaves proportionally to a $\Gamma(1/2, \gamma^2/2)$ variable for $x \leq 1$).

Example 3 *It can also be shown that the Lévy density of $z(1) \sim \Gamma(\nu, \alpha)$ is*

$$w(x) = \nu x^{-1} \exp(-\alpha x), \quad x \in R_+. \tag{2.11}$$

Again this is not an integrable Lévy density although it is slower to go off to infinity than the inverse Gaussian case. This mean in practice that it will have a smaller number of very small jumps than the IG process.

These two results are special cases of the result for the $GIG(\nu, \delta, \gamma)$ probability density (2.8). The corresponding Lévy density is then

$$w(x) = x^{-1}\left[\frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\delta^{-2}x\xi} g_\nu(\xi)\mathrm{d}\xi + \max\{0,\nu\}\lambda\right]\exp\left(-\gamma^2 x/2\right) \tag{2.12}$$

where

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

and J_ν and N_ν are Bessel functions. This result is derived in Part II of this book. Although this looks forbidding, when ν is a half integer these functions are analytically tractable.

Example 4 *An example outside the GIG class is the positive stable $PS(\alpha, \delta)$ process. Although the probability density of this variable is unknown in general, the cumulant function is*

$$\overline{\mathrm{K}}\left\{\theta \ddagger z(1)\right\} = -\delta\left(2\theta\right)^\alpha, \qquad 0 < \alpha < 1, \quad \delta > 0,$$

which implies it does not possess any moment. The Lévy density for a positive stable Lévy process is given by

$$w(x) = Cx^{-1-\alpha}, \quad \textit{where} \quad C = \delta 2^\alpha \frac{\alpha}{\Gamma\left(1-\alpha\right)}, \tag{2.13}$$

while $z(t) \sim PS(\alpha, t\delta)$.

2.3.3 Lévy-Khintchine representation for non-negative processes

Representation

Having allowed the Lévy measure not to be integrable, a single extra step is required in order to produce a general setup. We allow a drift $a > 0$ to be added to the cumulant function. This is carried out in the following fundamental theorem.

Theorem 2.1 *Lévy-Khintchine representation for non-negative Lévy processes. Suppose z is a Lévy process with non-negative increments. Then the kumulant function can be written as*

$$\overline{\mathrm{K}}\{\theta \ddagger z(1)\} = -a\theta - \int_0^\infty \left(1 - e^{-\theta x}\right) W(\mathrm{d}x) \tag{2.14}$$

where $a \geq 0$ and W is a measure on R_+ such that

$$\int_0^\infty \min\{1, x\} W(\mathrm{d}x) < \infty. \tag{2.15}$$

Conversely, any pair (a, W) with these properties determines a non-negative Lévy process z such that $z(1)$ has kumulant function determined by (2.14).

The importance of this representation is that the kumulant function of all non-negative Lévy processes can be written in this form. In other words, non-negative Lévy processes are completely determined by a and the Lévy measure W (which has to satisfy (2.15)). No other feature is necessary.

In the special case when $\int_0^\infty W(\mathrm{d}x) < \infty$ we say that z is of *finite activity* — indeed all such processes can be written as a compound Poisson process. In cases where this does not hold, z is said to be an *infinite activity* process.

Models via the Lévy density: tempered stable process

An important implication of the Lévy-Khintchine representation is that Lévy processes can be built by specifying a and W directly, implying the probability density of $z(1)$. An important example of this is the tempered stable, $TS(\kappa, \delta, \gamma)$, class which tilts the Lévy density of the positive stable. The result is

$$w(x) = \delta 2^{\kappa} \frac{\kappa}{\Gamma(1-\kappa)} x^{-\kappa-1} \exp\left(-\frac{1}{2}\gamma^{1/\kappa} x\right), \quad x, \delta > 0, \quad 0 < \kappa < 1, \quad \gamma \geq 0, \tag{2.16}$$

which means the process has infinite activity. The density of $z(1)$ is not generally known, however the Lévy density plays such a crucial role that this difficulty is not overly worrying. Special cases of this structure include the IG Lévy density (2.10) and the Γ Lévy density (2.11), which is the limiting case of $\kappa \downarrow 0$. Notice the constraint that $\kappa < 1$ is essential in order to satisfy the condition (2.15) in the Lévy-Khintchine representation. The corresponding kumulant function is

$$\overline{\mathrm{K}}\{\theta \ddagger z(1)\} = \delta\gamma - \delta\left(\gamma^{1/\kappa} + 2\theta\right)^{\kappa},$$

while the corresponding first two cumulants are

$$2\kappa\delta\gamma^{(\kappa-1)/\kappa} \quad \text{and} \quad 4\kappa(1-\kappa)\delta\gamma^{(\kappa-2)/\kappa}.$$

Finally the cumulant function implies the convenient property that $z(t) \sim TS(\kappa, t\delta, \gamma)$.

2.4 Processes with real increments

2.4.1 Examples of Lévy processes

Motivation

In this section the focus will be on Lévy processes with innovations which are on the real line. Many of them play important roles in financial economics as direct models of financial assets.

Brownian motion

In financial economics the most frequently used Lévy process is Brownian motion. It is the basic model of the log-price of a risky asset. In the simplest case of (standard) Brownian motion we write

$$z(1) \sim N(0, 1),$$

with density

$$f_{z(1)}(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \qquad x \in R,$$

while

$$\mathrm{K}\{\theta \ddagger z(1)\} = \log\left[\mathrm{E}\exp\{\theta z(1)\}\right] = \frac{1}{2}\theta^2.$$

The implication of this is that marginally $z(t) \sim N(0, t)$, while increments

$$z(t+\Delta) - z(t) \sim N(0, \Delta).$$

A standard Brownian motion, written $b(t)$, can be generalised to allow for innovations with a non-zero mean and a different scale than one. A *drift* μ and a *volatility* term σ can be introduced to deliver the Lévy process

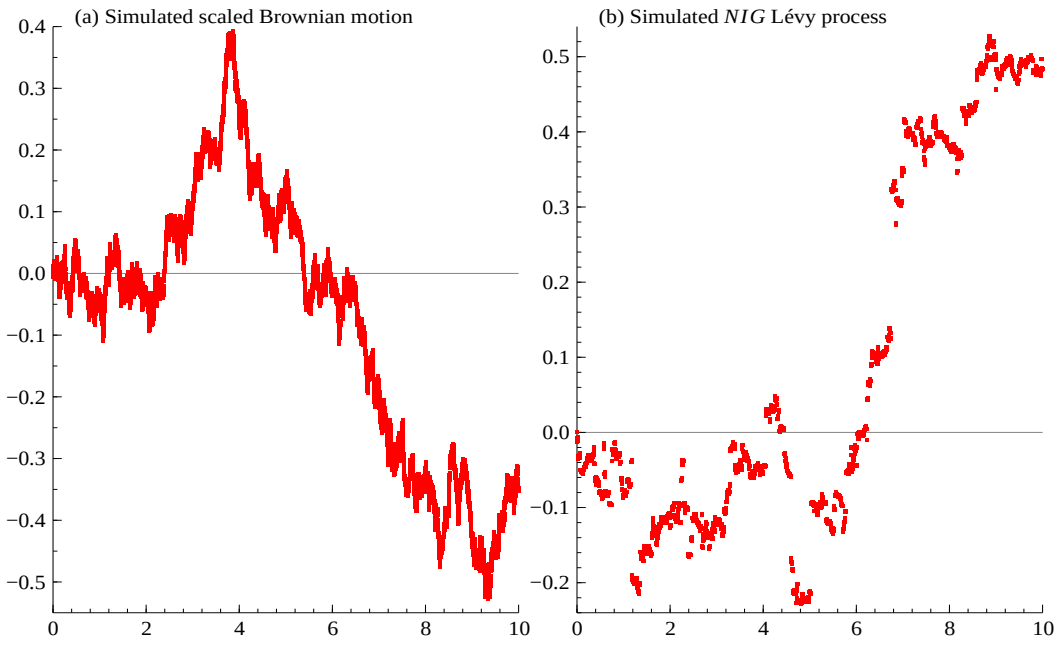

Figure 2.3: *(a) Sample path of* $\sqrt{0.02}$ *times standard Brownian motion with* $z(0) = 0$. *(b) Sample path of a NIG(0.2,0,0,10) Lévy process with* $z(0) = 0$. *Thus the increments of both processes have the same variance. Code:* `levy_graphs.ox`.

$$\begin{aligned} z(t) &= \mu t + \sigma b(t) \\ &\sim N(\mu t, t\sigma^2), \end{aligned}$$

with increments

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

The associated cumulant function for $z(1)$ is $\mu\theta + \frac{1}{2}\theta^2\sigma^2$.

A graph of a sample path from standard Brownian motion is displayed in Figure 2.3(a). It illustrates that the path is continuous. In a moment we will see that Brownian motion is the only Lévy process with this property — all other Lévy processes have jumps.

Compound process

Compound processes were introduced in (2.6), but there we required the shocks $\{c_s\}$ to be strictly positive. Here this condition is relaxed, just ruling out that they have an atom at zero. In this case, again,

$$\begin{aligned} \mathrm{K}\left\{\theta \ddagger z(1)\right\} &= \log\left[\mathrm{E}\exp\left\{\theta z(1)\right\}\right] \\ &= \psi\left\{\exp \mathrm{K}\left(\theta \ddagger c_1\right) - 1\right\}. \end{aligned}$$

Example 5 *Suppose* $c_s \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, *then*

$$\mathrm{K}\left\{\theta \ddagger z(1)\right\} = \psi\left\{\exp\left(\mu\theta + \frac{1}{2}\theta^2\sigma^2\right) - 1\right\}.$$

So here the Lévy process is constant until a new arrival from the Poisson process. The arrival then moves the Lévy process by a Gaussian variable. This variable can have a non-zero mean

and a non-unit variance. It was discussed in some detail in the first Chapter of this book. This is an important model in practice for quite a lot of effort has been expended on working on the derivative pricing theory associated with this simple structure.

Location scale mixture processes

Normal inverse Gaussian process If we assume $\sigma^2 \sim IG(\delta, \gamma)$ and ε is an independent standard normal variable then

$$y = \mu + \beta\sigma^2 + \sigma\varepsilon \sim NIG(\alpha, \beta, \mu, \delta), \qquad \text{where} \qquad \alpha = \sqrt{\beta^2 + \gamma^2}$$

has a normal inverse Gaussian (NIG) distribution. If $\mu = \beta = 0$ then this random variable is centred at zero, while if $\beta < 0$ it is skewed with a longer left hand tail. Further, μ is a free parameter which helps control the mean of the process, with $\mathrm{E}(y) = \mu + \beta\delta\gamma^{-1}$. The NIG Lévy process puts

$$z(1) \sim NIG(\alpha, \beta, \mu, \delta), \qquad \mu \in \mathbf{R}, \delta \in \mathbf{R}_+, 0 \leq \beta < \alpha$$

which has the density

$$f_{z(1)}(x) = a(\alpha, \beta, \mu, \delta) q\left(\frac{x-\mu}{\delta}\right)^{-1} K_1\left\{\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right\} \exp\left\{\beta\left(x-\mu\right)\right\}$$

where $q(x) = \sqrt{1+x^2}$ and

$$a(\alpha, \beta, \mu, \delta) = \pi^{-1}\alpha \exp\left\{\delta\sqrt{\alpha^2 - \beta^2} - \beta\mu\right\}.$$

Sometimes it is convenient to reparameterise this model. A popular local-scale invariant choice is achieved by defining

$$\xi = \left(1 + \delta\sqrt{\alpha^2 - \beta^2}\right)^{-1/2} \quad \text{and} \quad \chi = \frac{\beta}{\alpha}\xi, \tag{2.17}$$

where ξ, the steepness parameter, and χ, the asymmetry parameter, obey a triangular constraint

$$\left\{(\chi, \xi) : -1 < \chi < 1, \quad 0 < \xi < 1\right\}. \tag{2.18}$$

The flexibility of the model is shown in Figure 2.4, which displays the log-density for a variety of values of χ, ξ. Such a plot is called a *shape triangle*. As $\xi \to 0$ so the log-density becomes more quadratic, while for values around 0.5 the tails are approximately linear. For larger values of ξ the tails start decaying at a rate which looks appreciably slower than linear. In the limit as $\xi \to 1$ the density becomes a Cauchy variable.

This model has recently received considerably attention as a tractable alternative to Brownian motion as a model of log asset prices. One of its advantages is that the resulting cumulant function is

$$\mathrm{K}(\theta \ddagger z(1)) = \delta\left\{\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta+\theta)^2}\right\} + \mu\theta,$$

implying $z(t) \sim NIG(\alpha, \beta, t\mu, t\delta)$. In particular this implies its increments are non-zero with probability one. A sample path of a NIG Lévy process is drawn in Figure 2.3(b) with the jumps being of irregular size which implies the very jagged shape of the picture.

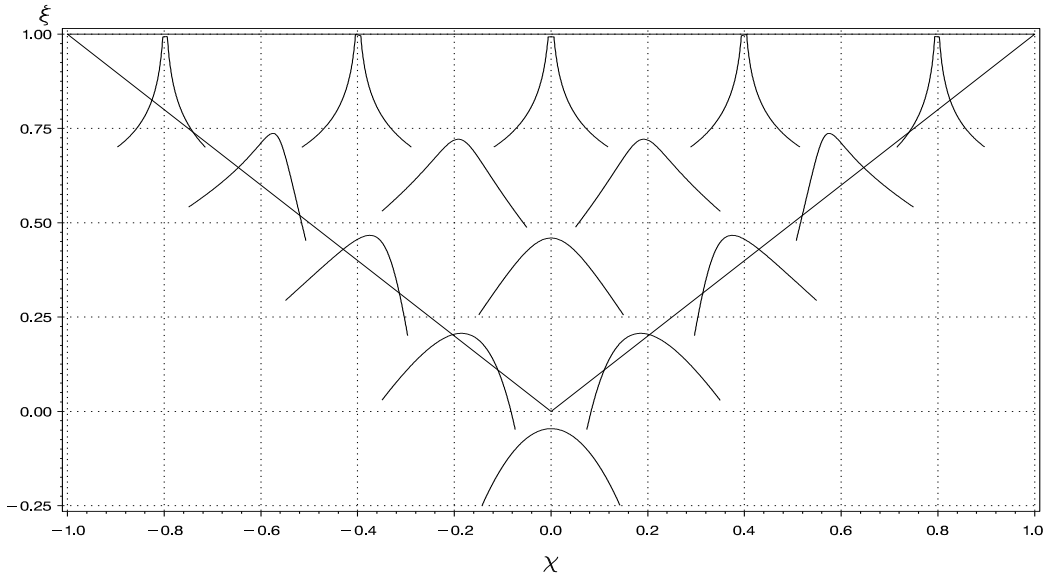

Figure 2.4: *Shape triangle for the NIG model. That is we graph the shape of the log-density for the NIG model for a variety of values of the steepness parameter ξ and the shape parameter χ. This graph was kindly made available to us by Preben Blæsild.*

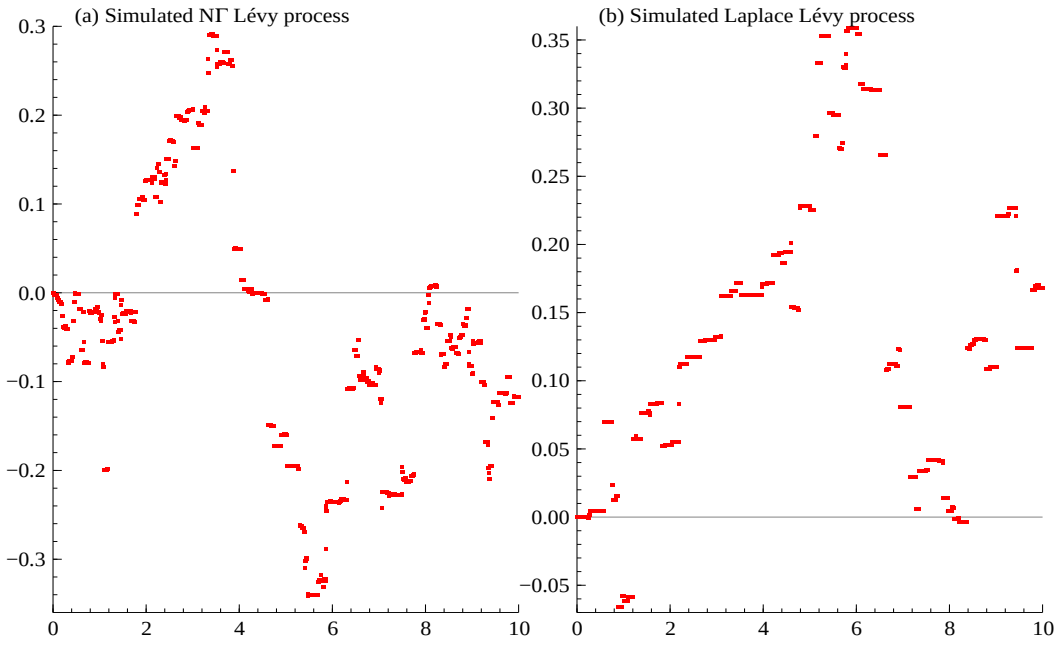

Figure 2.5: (a) *Sample path of a NΓ(4,200,0,0) Lévy process. Such processes are often called variance gamma processes in the literature. (b) Sample path of a La(0.2,0,0) Lévy process. Code:* `levy_graphs.ox`.

Normal gamma process If we assume $\sigma^2 \sim \Gamma(\nu, \gamma^2/2)$ and ε is an independent standard normal variable then

$$y = \mu + \beta\sigma^2 + \sigma\varepsilon \sim N\Gamma(\nu, \gamma, \beta, \mu),$$

which we will call the normal gamma distribution. From the cumulant function

$$\mathrm{K}\left\{\theta \ddagger z(1)\right\} = \mu\theta + \nu \log\left(1 + \frac{\theta\beta + \theta^2/2}{\gamma}\right), \tag{2.19}$$

it follows that $z(t) \sim N\Gamma(t\nu, \gamma, \beta, t\mu)$, which means this process is particularly simple to handle and the increments are non-zero. The density of the process at time one is

$$f_{z(1)}(x) = \frac{\gamma^{2\nu}\left(\frac{\gamma^2}{2}\right)^{1-2\nu}}{\sqrt{2\pi}\Gamma(\nu)2^{\nu-1}}\bar{K}_{\nu-1/2}\left(\frac{\gamma^2}{2}\left|x-\mu\right|\right)\exp\left\{\beta\left(x-\mu\right)\right\}.$$

In the special case of $\beta = 0$ this process has been used extensively in the finance literature where it is often called the *variance gamma* (VG) process.

Figure 2.5(a) graphs a simulated path from a $N\Gamma(4, 200, 0, 0)$ process. As we would expect, the sample path has some of the features of the NIG process we drew in Figure 2.3(b). In particular it is very jagged.

Hyperbolic and Laplace processes If we assume $\sigma^2 \sim PH(\delta, \gamma)$ and ε is an independent standard normal variable then

$$y = \mu + \beta\sigma^2 + \sigma\varepsilon \sim H(\alpha, \beta, \mu, \delta), \qquad \text{where} \qquad \alpha = \sqrt{\beta^2 + \gamma^2},$$

has the hyperbolic distribution. This distribution can be shown to be infinitely divisible, although the proof of this is difficult. The hyperbolic process puts $z(1) \sim H(\alpha, \beta, \mu, \delta)$, where the density is

$$f_{z(1)}(x) = \frac{\gamma}{2\sqrt{\beta^2+\gamma^2}\delta K_1\left(\delta\gamma\right)}\exp\left\{-\alpha\sqrt{\delta^2 + \left(x-\mu\right)^2} + \beta\left(x-\mu\right)\right\}, \qquad x \in R. \tag{2.20}$$

All the moments of this process exist so long as $\gamma > 0$, while the cumulant function is

$$\mathrm{K}(\theta \ddagger x) = \frac{1}{2}\log\left\{\frac{\gamma^2}{\alpha^2 - \left(\beta+\theta\right)^2}\right\} + \log\left\{\frac{K_1\left\{\delta\sqrt{\alpha^2 - \left(\beta+\theta\right)^2}\right\}}{K_1\left(\delta\gamma\right)}\right\} + \theta\mu. \tag{2.21}$$

The hyperbolic model is again rather flexible. We can compare it to the NIG density using the shape triangle. In particular reparameterise into the location-scale invariant parameters given in (2.17), then Figure 2.6 shows the log-densities for this model. We see that again as $\xi \to 0$ we get the normal quadratic log-density. For higher values the log-density gets increasingly linear decay in the tails as $\xi \to 1$. Indeed in the limit we get the Laplace density as the model. This contrasts with the NIG density which has the ability to have thicker tails than this density. Hyperbolic models have the interesting and important feature that the log-density of $z(1)$ is approximately linear in the tails of the distribution (rather than quadratic in the normal case, and the corresponding Brownian motion at time one).

Hyperbolic Lévy processes have the disadvantage that we do not have an exact expression for the density of $z(t)$ for $t \neq 1$, nor can we simulate from the process in a non-intensive manner.

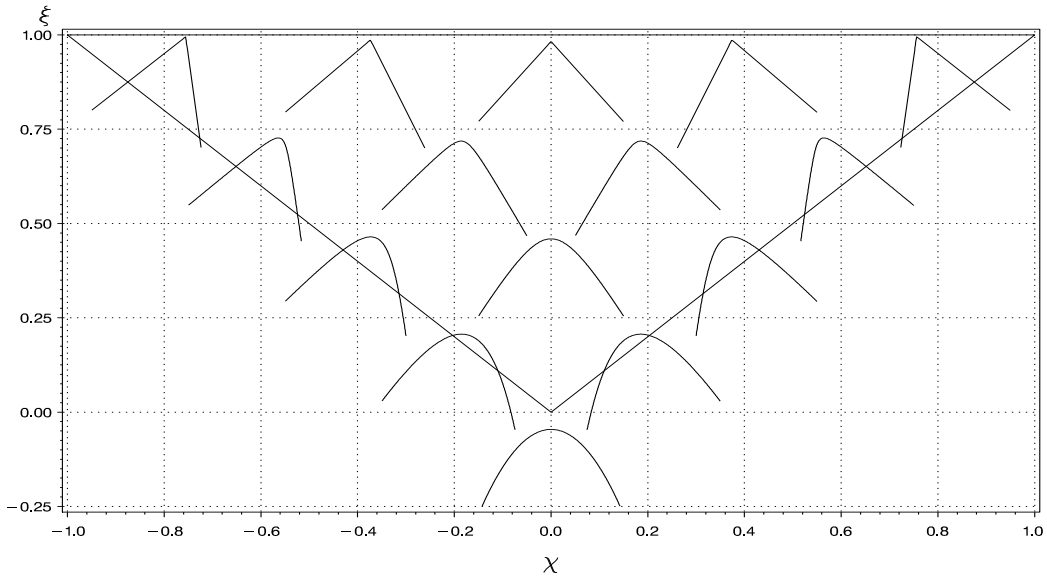

Figure 2.6: *Shape triangle for the hyperbolic model. That is we graph the shape of the log-density for the hyperbolic model for a variety of values of the steepness parameter ξ and the shape parameter χ. This graph was kindly made available to us by Preben Blæsild.*

Both of these properties are inherited from the fact that this is also the case for the positive hyperbolic process we discussed in the previous section.

The Laplace distributions (symmetric and asymmetric) occur as limiting cases of (2.20) for α, β and μ fixed and $\delta \downarrow 0$. We write this as $La(\alpha, \beta, \mu)$. The corresponding density is

$$\frac{\alpha^2 - \beta^2}{2\alpha} \exp\left\{-\alpha\left|x-\mu\right| + \beta\left(x-\mu\right)\right\}, \qquad \text{where} \qquad \alpha = \sqrt{\beta^2+\gamma^2}, \tag{2.22}$$

which is achieved by $\sigma^2 \sim E(\gamma^2/2) = \Gamma(1, \gamma^2/2)$. One of the main features of this model is that $z(t) \sim N\Gamma(t, \gamma, \beta, t\mu)$, which has the advantage of simple tractability — and is a special case of the normal gamma process.

Finally, if $\sigma^2 \sim RPH(\delta, \gamma)$, instead of being positive hyperbolic, then $y \sim RH(\alpha, \beta, \mu, \delta)$, the reciprocal hyperbolic. This distribution is again infinitely divisible. The RH process puts

$$f_{z(1)}(x) = \frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_{-1}\left(\delta\sqrt{\alpha^2-\beta^2}\right)} x^{-2} \exp\left\{-\alpha\sqrt{\delta^2+(x-\mu)^2} + \beta\left(x-\mu\right)\right\}, \qquad x \in R.$$

If we assume $\sigma^2 \sim R\Gamma(\nu, \delta^2/2)$ and ε is an independent standard normal variable then

$$y = \mu + \beta\sigma^2 + \sigma\varepsilon \sim T(\nu, \delta, \beta, \mu),$$

a skewed Student's t distribution, which is infinitely divisible and so can be used as the basis of a Lévy process. The skewed Student's t process puts $z(1) \sim T(\nu, \delta, \beta, \mu)$, where the density is

$$\frac{1}{\sqrt{2\pi}\delta\Gamma(\nu)2^{\nu-1}} q\left(\frac{x-\mu}{\delta}\right)^{-2\nu-1} \bar{K}_{\nu+1/2}\left\{\delta\beta q\left(\frac{x-\mu}{\delta}\right)\right\} \exp\left\{\beta\left(x-\mu\right)\right\},$$

where

$$\overline{K}_\nu(x) = x^\nu K_\nu(x) \qquad \text{and} \qquad q(x) = \sqrt{1+x^2}.$$

The more familiar Student's t distribution is found when we let $\beta \to 0$, then the density becomes

$$\frac{\Gamma(\nu+1/2)}{\sqrt{\delta^2\pi}\Gamma(\nu)}\left\{1+\left(\frac{x-\mu}{\delta}\right)^2\right\}^{-\nu-1/2}.$$

In both cases this process has the interesting feature that only moments of order less than ν will exist — at any time horizon. However, we do not know the distribution of $z(t)$ for this process, while simulation has to be carried out in quite an involved manner. Hence this process is not as easy to handle as the NIG or normal gamma Lévy processes.

A similar type of complexity occurs if we replace $\sigma^2 \sim R\Gamma(\nu, \delta^2/2)$ by a reciprocal inverse Gaussian distribution, $RIG(\delta, \gamma)$. The resulting mixture distribution is called a normal reciprocal inverse Gaussian distribution $NRIG(\alpha, \beta, \mu, \delta)$ which can be shown to be infinitely divisible and so supports a $NRIG$ Lévy process. This process is again hard to work with and so we will not discuss it in detail here.

Generalized hyperbolic process If we assume $\sigma^2 \sim GIG(\nu, \delta, \gamma)$ and ε is an independent standard normal variable then

$$y = \mu + \beta\sigma^2 + \sigma\varepsilon \sim GH(\nu, \alpha, \beta, \mu, \delta), \qquad \text{where} \qquad \alpha = \sqrt{\beta^2+\gamma^2},$$

the *generalised hyperbolic* distribution. This distribution includes as special cases the normal (N), normal inverse Gaussian (NIG), normal reciprocal inverse Gaussian ($NRIG$), hyperbola (HA), hyperbolic (H), skewed Laplace (La), normal gamma ($N\Gamma$) and skewed Student (T) distributions in the following way:

$$\begin{array}{ll}
N(\mu,\sigma^2) = \lim_{\gamma\to\infty} GH(\nu,\gamma,0,\mu,\sigma^2\gamma), & NIG(\alpha,\beta,\mu,\delta) = GH\left(-\tfrac{1}{2},\alpha,\beta,\mu,\delta\right),\\
NRIG(\alpha,\beta,\mu,\delta) = GH\left(\tfrac{1}{2},\alpha,\beta,\mu,\delta\right), & H(\alpha,\beta,\mu,\delta) = GH(1,\alpha,\beta,\mu,\delta),\\
T(\nu,\delta,\beta,\mu) = GH(-\nu,\beta,\beta,\mu,\delta), & La(\alpha,\beta,\mu) = GH(1,\alpha,\beta,\mu,0)\\
N\Gamma(\nu,\delta,\beta,\mu) = GH(\nu,\alpha,\beta,\mu,0), & RH(\alpha,\beta,\mu,\delta) = GH(-1,\alpha,\beta,\mu,\delta),
\end{array}$$

for $\nu > 0$. The generalised hyperbolic distribution is infinitely divisible and so can be used as the basis of a rather general Lévy process whose special cases obviously include Brownian motion with drift and the NIG, hyperbolic, hyperbola, $NRIG$, normal gamma, skewed Laplace and skewed Student Lévy processes. The proof of infinite divisibility of this distribution is involved but will be discussed in Part II of our book. The generalised hyperbolic process puts $z(1) \sim GH(\nu, \alpha, \beta, \mu, \delta)$, where the density is

$$\frac{(\gamma/\delta)^\nu}{\sqrt{2\pi}\alpha^{2\left(\nu-\frac{1}{2}\right)}K_\nu(\delta\gamma)}\left\{\alpha\delta q\left(\frac{x-\mu}{\delta}\right)\right\}^{\left(\nu-\frac{1}{2}\right)}K_{\left(\nu-\frac{1}{2}\right)}\left\{\alpha\delta q\left(\frac{x-\mu}{\delta}\right)\right\}\exp\left\{\beta(x-\mu)\right\}, \quad (2.23)$$

where $q(x) = \sqrt{1+x^2}$ and K_ν is the modified Bessel function of the third kind. It is helpful to recall that $K_\nu(x) = K_\nu(-x)$. The cumulant function is

$$\mathrm{K}(\theta \ddagger x) = \frac{\nu}{2}\log\left\{\frac{\gamma}{\alpha^2-(\beta+\theta)^2}\right\} + \log\left\{\frac{K_\nu\left\{\delta\sqrt{\alpha^2-(\beta+\theta)^2}\right\}}{K_\nu\left\{\delta\sqrt{\alpha^2-\beta^2}\right\}}\right\} + \theta\mu, \quad |\beta+\theta| < \alpha,$$

while the first two moments are

$$\begin{aligned}
\mathrm{E}(X) &= \mu + \beta\frac{\delta}{\gamma}\frac{K_{\nu+1}(\delta\gamma)}{K_\nu(\delta\gamma)} \quad \text{and}\\
\mathrm{Var}(X) &= \delta^2\left\langle\frac{K_{\nu+1}(\delta\gamma)}{\delta\gamma K_\nu(\delta\gamma)} + \frac{\beta^2}{\gamma^2}\left[\frac{K_{\nu+2}(\delta\gamma)}{K_\nu(\delta\gamma)} - \left\{\frac{K_{\nu+1}(\delta\gamma)}{K_\nu(\delta\gamma)}\right\}^2\right]\right\rangle.
\end{aligned}$$

It is helpful to reexpress this density in a new alternative format. We do this by defining $\overline{K}_\nu(x) = x^\nu K_\nu(x)$, noting that $K_\nu(x) = K_{-\nu}(x)$ and setting up some scale invariant parameters

$$\overline{\alpha} = \delta\alpha, \qquad \overline{\beta} = \delta\beta, \qquad \overline{\gamma} = \delta\gamma.$$

Then the density becomes

- for $\nu > 0$
$$\frac{\overline{\gamma}^{2\nu}\alpha^{1-2\nu}}{\delta\sqrt{2\pi}\bar{K}_\nu(\bar{\gamma})}\bar{K}_{\nu-1/2}\left\{\overline{\alpha}q\left(\frac{x-\mu}{\delta}\right)\right\}\exp\left\{\overline{\beta}\left(\frac{x-\mu}{\delta}\right)\right\},$$
- for $\overline{\nu} = -\nu > 0$
$$\frac{1}{\sqrt{2\pi}\delta\bar{K}_{\overline{\nu}}(\bar{\gamma})}q\left(\frac{x-\mu}{\delta}\right)^{-2\overline{\nu}-1}\bar{K}_{\overline{\nu}+1/2}\left\{\overline{\alpha}q\left(\frac{x-\mu}{\delta}\right)\right\}\exp\left\{\overline{\beta}\left(\frac{x-\mu}{\delta}\right)\right\}.$$

Not surprisingly, in general we do not know the GH density of $z(t)$ for $t \neq 1$, nor can we simulate from the process in a non-intensive manner. This model is so general that it is typically difficult to manipulate mathematically and so is not often used empirically. Instead special cases are usually employed.

Symmetric stable processes One of the most studied Lévy processes is the symmetric stable process. This puts

$$z(1) \sim S(\alpha, \delta), \qquad 0 < \alpha \leq 2, \quad \delta > 0,$$

a symmetric stable distribution with index α. Except for the boundary case of $\alpha = 2$, this distribution has the empirically unappealing feature that the variance of $z(1)$ is infinity. The density of this variable is unknown in general, with exceptions being the Gaussian variable ($\alpha = 2$), the Cauchy variable ($\alpha = 1$) and the Lévy variable ($\alpha = 1/2$). Despite the complexity of the density the cumulant function is simply

$$\mathrm{K}\left\{\theta \ddagger z(1)\right\} = \delta\theta^\alpha,$$

which implies $z(t) \sim S(\alpha, t\delta)$. The Lévy density for a symmetric stable process is given by

$$w(x) = \delta\left|x\right|^{-1-\alpha}, \quad 0 < \alpha < 2.$$

Stable processes have the remarkable property that for $\lambda > 0$

$$\left\{z(\lambda t)\right\}_{t\geq 0} \overset{\mathcal{L}}{=} \left\{\lambda^{\alpha/2} z(t)\right\}_{t\geq 0}.$$

Thus, in particular, increments of the Lévy process over time λt are, in distribution, just scaled versions of increments over time t. This fractal like property is called *self-similarity* and the stable Lévy processes (symmetric or not) are the only Lévy processes which possess this feature.

Although stable processes have received considerable attention in financial economics since their introduction into that subject in the early 1960s, it has been known since the late 1960s that they provide a poor fit to the empirical data we usually see in practice. This is because returns over long time intervals, which are sums of returns over finer time intervals, tend to be more Gaussian than ones over short horizons. Hence our interest in this type of process will usually be to provide theoretical illustrations, rather than as practical models.

Normal tempered stable process If we assume $\sigma^2 \sim TS(\kappa, \delta, \gamma)$ and ε is an independent standard normal variable then

$$y = \mu + \beta\sigma^2 + \sigma\varepsilon \sim NTS(\kappa, \delta, \gamma, \beta, \mu),$$

a normal tempered stable distribution, which is infinitely divisible and so can be used as the basis of a Lévy process. This process has as special cases the NIG Lévy process and the normal gamma Lévy process (when $\kappa \downarrow 0$). This process will be discussed in more detail in a later section.

Some other Lévy processes living on the real line

Truncated Lévy flights If the Lévy density of the stable process is truncated, so that

$$w(x) = \begin{cases} c\,|x|^{-1-\alpha} & \text{for} \quad x \in [-l, l] \\ 0 & \text{otherwise,} \end{cases}$$

where $c, l > 0$, then it can be shown that this again supports a Lévy process where $z(1)$ has a finite variance. This process is called a truncated Lévy flights process. This has received some interest in the context of the econophysics literature.

Extended Koponen class Another example of a Lévy process specified through the Lévy density is the KoBoL (after Koponen, Boyarchenko and Levendorskii) or extended Koponen class. This puts

$$w(x) = \begin{cases} C_-|x|^{-1-A}e^{-B_-|x|} & \text{for} \quad x < 0 \\ \\ C_+x^{-1-A}e^{-B_+x} & \text{for} \quad x > 0. \end{cases} \tag{2.24}$$

This process is sometimes referred to as a CGMY Lévy process (after Carr, Geman, Medan and Yor).

Meixner process The density of a Meixner random variable, written Meixner(a, b, d, μ), is

$$f_{z(1)}(x) = \frac{\{2\cos(b/2)\}^{2d}}{2a\pi\Gamma(2d)} \exp\left\{\frac{\beta(x-\mu)}{a}\right\} \left|\Gamma\left(d + \frac{i(x-\mu)}{a}\right)\right|^2, \qquad i = \sqrt{-1},$$

which has the cumulant function

$$\mathrm{C}\{\zeta \ddagger z(1)\} = i\mu\zeta + 2d\log\left\{\frac{\cos(\beta/2)}{\cosh\left(\frac{a\zeta - i\beta}{2}\right)}\right\}.$$

The resulting process is obviously infinitely divisible and so supports a Lévy process with $z(t) \sim Meixner(a, b, dt, \mu t)$. Further, the Lévy density is

$$w(x) = d\frac{e^{bx/a}}{x\sinh(\pi x/a)},$$

and so we can see that this Lévy process has infinite activity. Unfortunately we do not know a simple way of simulating from this distribution.

2.4.2 Lévy-Khintchine representation

The Lévy-Khintchine representation for positive variables given in (2.14) can be generalised to cover Lévy processes with increments on the real line. Three basic developments are needed. First, the Lévy measure must be allowed to have support on the real line, not just the positive half-line, but still excluding the possibility that the measure has an atom at zero. Second, the parameter a needs to be allowed to be a real variable, not just positive. Third, we imagine that an independent Brownian motion component is added to the process. The result is the celebrated Lévy-Khintchine representation for Lévy processes.

Theorem 2.2 *Lévy-Khintchine representation. Suppose z is a Lévy process. Then the log of the characteristic function can be written as*

$$\mathrm{C}\{\zeta \ddagger z(1)\} = ai\zeta - \frac{1}{2}\sigma^2\zeta^2 - \int_{\mathbf{R}} \left\{1 - e^{i\zeta x} + i\zeta x \mathbf{1}_B(x)\right\} W(\mathrm{d}x), \tag{2.25}$$

where $a \in R$, $\sigma \geq 0$, $B = [-1, 1]$ and the Lévy measure W must satisfy

$$\int_{\mathbf{R}} \min\{1, x^2\} W(\mathrm{d}x) < \infty \tag{2.26}$$

and W has no atom at 0.

Lévy processes are completely determined by the *characteristic triplet*: a, the variance σ^2 of the Brownian motion and the Lévy measure W (which has to satisfy (2.26)). No other feature is necessary and every such triplet (a, σ^2, W) specifies a Lévy process. Importantly only processes with $W = 0$ do not have jumps — but in that case z is a scaled Brownian motion.

2.5 Time deformation, chronometers and subordinators

2.5.1 Definitions

Financial markets sometimes seem to move rapidly. One way of starting to model this is to allow the relationship between standard calendar time and the pace of the market to be random. We call a stochastic process which models the random clock a *chronometer*, while the resulting process is said to be *time deformed* or *subordinated*. The use of the nomenclature chronometer in this context is new.

Definition 2 *A chronometer is any non-decreasing random process. The special case where the chronometer has independent and stationary increments is called a subordinator.*

The requirement that the chronometer is non-decreasing rules out the chance that time can go backwards. A special case of a chronometer is a subordinator, while subordinators are special cases of Lévy processes (e.g. Poisson or IG Lévy processes are subordinators). All subordinators are pure upward jumping processes. We should note here that the finance literature typically labels chronometers subordinators, while the probability literature only discusses deformation in the context of Lévy processes.

In this section we will study what happens when a subordinator is used to change the clock, that is deform, a stochastically independent Lévy process. Write $v(t)$ and $\tau(t)$ as independent Lévy processes, the latter being a subordinator used to model the random clock. The result is

$$z(t) = v(\tau(t)).$$

The increments of this process are

$$\begin{aligned} z(t+\Delta) - z(t) &= v(\tau(t+\Delta)) - v(\tau(t)) \\ &= v(\tau(t) + \{\tau(t+\Delta) - \tau(t)\}) - v(\tau(t)), \end{aligned}$$

which are independent and stationary and so z is a Lévy process.

Brownian motion is the only Lévy process with continuous sample paths, however this property does not survive being deformed by a subordinator.

The subordinator must be a pure jump process — jumping upwards at random times. At each instant of a jump $z(t)$ must (with probability one) also jump, while in instants where the subordinator does not jump the level of $z(t)$ is left unchanged.

2.5.2 Examples

Brownian motion with a Poisson subordinator

Assume $v(t) \sim N(\beta t, \sigma^2 t)$ is a scaled Brownian motion with drift and that it is deformed by a Poisson process with intensity ψ. Then

$$z(1)|\tau(1) \sim N(\beta\tau(1), \sigma^2\tau(1)).$$

This is a compound Poisson process

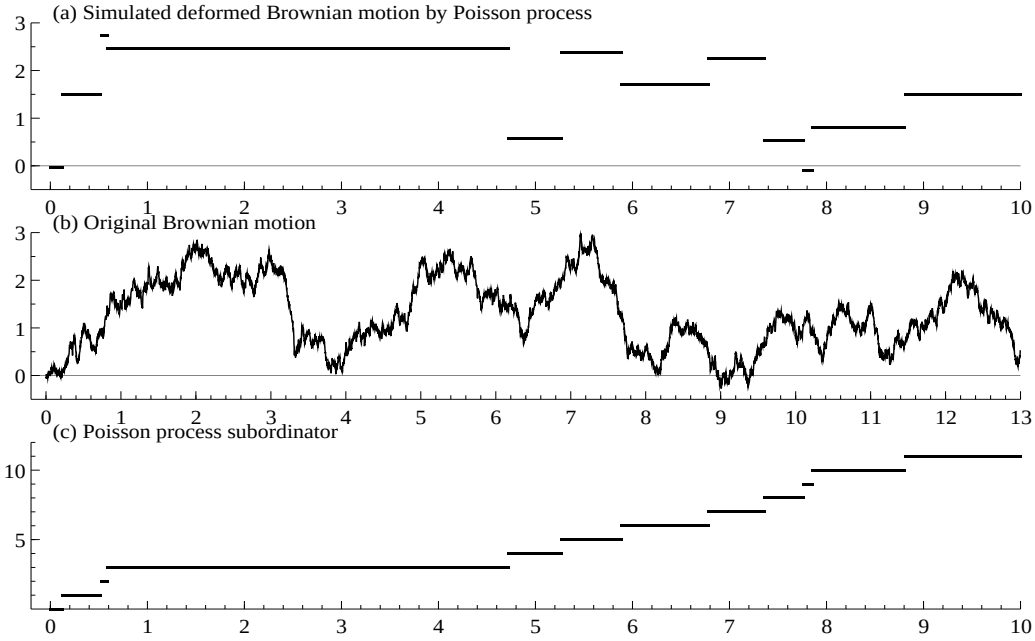

Figure 2.7: *Figure (a) deformed Brownian motion using a Poisson process subordinator. Figure (b) path of the Brownian motion. Figure (c) Poisson process subordinator. Code:* `levy_code.ox`.

$$z(t) = \sum_{j=1}^{\tau(t)} c_j, \quad c_j \overset{i.i.d.}{\sim} N(\mu, \sigma^2), \tag{2.27}$$

with shocks which are Gaussian. The implication is that when the process jumps, the jumps are independent of the time we have waited until the jump. An example of a sample path from this process is given in Figure 2.7. The jumps now go up as well as down, with the jump sizes being random.

Brownian motion subordinated by IG — the NIG Lévy process

Suppose $\tau(t)$ is an inverse Gaussian Lévy process with $\tau(1) \sim IG(\delta, \gamma)$ and $v(t)$ is Brownian motion with drift β. Then $z(t)$ is a Lévy process. In particular

$$z(1)|\tau(1) \sim N(\beta\tau(1), \tau(1))$$

and so unconditionally the increments are independent with

$$z(1) \sim NIG(\alpha, \beta, 0, \delta), \quad \alpha^2 = \beta^2 + \gamma^2.$$

Hence this deformed Brownian motion is a special case of the normal inverse Gaussian Lévy process, which we simulated in Figure 2.2(b).

Normal tempered stable Lévy process

Suppose $\tau(t)$ is a tempered stable $TS(\kappa, \delta, \gamma)$ Lévy process, with Lévy density given in (2.16). It is a subordinator. Then if we assume $b^\beta(\cdot)$ is Brownian motion plus drift and we write

$$z(t) = \mu t + b^\beta(\tau(t)),$$

then z is called a *normal tempered stable* (NTS) Lévy process. We write

$$z(1) \sim NTS(\kappa, \alpha, \beta, \mu, \delta),$$

but the corresponding probability density is generally unknown (except for an infinite series representation, see Feller (1971, p. 583)). The cumulant function, on the other hand, is rather simple

$$\mathrm{K}(\theta \ddagger z(1)) = \mu\theta + \delta\gamma - \delta\left\{\alpha^2 - (\beta+\theta)^2\right\}^\kappa, \quad \text{where} \quad \alpha = \sqrt{\beta^2 + \gamma^{1/\kappa}}.$$

The form of this function implies

$$z(t) \sim NTS(\kappa, \alpha, \beta, \mu t, \delta t).$$

It can be shown (after some considerable work), using the cumulant function of the NTS process, that the Lévy density is

$$w(x) = \frac{\delta}{\sqrt{2\pi}} \frac{\kappa 2^{\kappa+1}}{\Gamma(1-\kappa)} \alpha^{\kappa+\frac{1}{2}} |x-\mu|^{-\left(\kappa+\frac{1}{2}\right)} K_{\kappa+\frac{1}{2}}(\alpha|x-\mu|) \exp\{\beta(x-\mu)\}.$$

The direct use of this Lévy density is obviously going to be difficult due to its complexity. The deformation interpretation will mean that we can usually sidestep this, instead employing the simple Lévy density of the $TS(\kappa, \delta, \gamma)$ Lévy process.

Type G Lévy processes

In the probability literature, Lévy processes which can be written as $z(t) = \mu t + b^\beta(\tau(t))$, for some subordinator τ, which we shall call *type G Lévy processes* — the subset of Lévy processes for which there is a deformation of Brownian motion interpretation. Many well known Lévy processes are not in this class.

2.6 Quadratic variation

2.6.1 Definition and examples

A commonly used measure of continuous time processes in financial economics is the *Quadratic Variation (QV) process*. This has two steps. First, time is split into small intervals

$$t_0^r = 0 < t_1^r < ... < t_{m_r}^r = t.$$

Then the QV process is

$$[z](t) = \text{p}-\lim_{r\to\infty} \sum \{z(t_{i+1}^r) - z(t_i^r)\}^2, \tag{2.28}$$

where

$$\sup_i \{t_{i+1}^r - t_i^r\} \to 0 \quad \text{for} \quad r \to \infty.$$

This series looks at the partial sum of squared increments over tiny intervals of time. In general the QV process of a Lévy process is a (different) Lévy process for the increments are independent and stationary as QV are just sums the squares of independent and stationary increments. Further, it can be regarded as a subordinator for the increments are non-negative. To illustrate these points two examples are given.

Brownian motion

Suppose z is a scaled Brownian motion with drift, such that $z(1) \sim N(\mu, \sigma^2)$. Then

$$z(t_{i+1}^r) - z(t_i^r) \sim N((t_{i+1}^r - t_i^r)\,\mu, (t_{i+1}^r - t_i^r)\,\sigma^2).$$

For small values of $(t_{i+1}^r - t_i^r)$ the variation in the series dominates — the standard deviation and drift are $O\left(\sqrt{t_{i+1}^r - t_i^r}\right)$ and $O\left(t_{i+1}^r - t\right)$, respectively. As a result $[z](t) = t\sigma^2$, whatever the value of μ. The important observation is that the QV is non-stochastic.

This is a particularly interesting result from a statistical viewpoint for it means we can theoretically estimate σ^2 without error using a tiny path of Brownian motion even in the presence of drift. Of course in practice this is highly misleading argument for the continuous time model is unlikely to be perfectly specified at very short time horizons.

Brownian motion deformed by a Poisson process

Suppose z is constructed by time deforming a Brownian motion with drift β with a Poisson process τ, then z is a compound process (2.27) and

$$[z](t) = \sum_{j=1}^{\tau(t)} c_j^2, \quad \text{where} \quad c_j \overset{i.i.d.}{\sim} N(0, \sigma^2).$$

Hence the QV is also a compound process with $[z](t) \sim \chi^2_{\tau(t)}$. Hence $[z](t)$ is a noisy estimator of $\tau(t)$, but they share the same expectation.

Semimartingales

We noted in section 3 that all Lévy processes were semimartingales, and that when the mean of the increments of the process existed then they are special semimartingales. This means we can uniquely write them as

$$z(t) = a(t) + m(t),$$

a predictable component of locally bounded variation and a local martingale, respectively. It is possible to show that for all semimartingales

$$[z](t) = [m](t) + \sum_{0 \le u \le t} \{a(u) - a(u-)\}^2,$$

the quadratic variation of the martingale component plus the sums of squares of the jumps in the predictable component. In Lévy processes the predictable component is always (if it exists) $a(t) = t\mathrm{E}\{z(1)\}$ and hence is continuous. Thus this result specialises to

$$[z](t) = [m](t).$$

Quadratic variation is very important. One of the reasons for this is that for all local semimartingales with continuous predictable components

$$\mathrm{Var}(\mathrm{d}z(t)|\mathcal{F}_t) = \mathrm{E}\left(\mathrm{d}[z](t)|\mathcal{F}_t\right),$$

so long as the moments exist.

2.6.2 Realised variance process

Definition and basics

In applied economics it is often inappropriate to study returns over infinitely small time intervals for our models tend to be highly misspecified at that level due to market microstructure effects. In particular the idea of a unique price is a fiction for the transaction price tends to depend upon, for example, the volume of the deal, the reputation of the buyer and seller, prevailing liquidity (and so time of day) and the initiator (i.e. was it the buyer or the seller). These issues will be discussed at more length in later chapters. To avoid the worst effects of misspecification, a finite version of quadratic variation is often used. This is called the realised volatility or variance process. This splits time into intervals of length δ and computes the corresponding sum of squares.

The *realised variance process* is defined, for $\delta > 0$,

$$[z_\delta](t) = \sum_{j=1}^{M} \left[z(\delta j) - z\{\delta(j-1)\}\right]^2, \quad M = \lfloor t/\delta \rfloor,$$

where $\lfloor t \rfloor$ denotes the largest integer less than or equal to t. We can see that

$$p - \lim_{\delta \downarrow 0} [z_\delta](t) = [z](t),$$

that is the realised variance process is a consistent estimator of quadratic variation. However, the realised variance process is not a Lévy process — rather it jumps upwards at specified points in time and so has a number of features of a discrete time random walk.

A numerical example of the realised variance process is given in Figure 2.8, which computes it for a NIG Lévy process. In this picture we have taken $\delta = 1$ and $\delta = 1/10$, so taking one and 10 squared observations per unit of time, respectively. Also given is the corresponding limit, the quadratic variation. We see that as δ gets small so the realised variance process becomes a good approximation of the QV.

In order to understand the connection between the Lévy process, the realised variance process and the quadratic variation it is helpful to think about the following calculation. Hold t fixed and set the choice of δ so that $M = t/\delta$, then

$$\mathrm{E}\begin{Bmatrix} z(t) \\ [z_\delta](t) \\ [z](t) \end{Bmatrix} = t\begin{pmatrix} \kappa_1 \\ \kappa_2 \\ \kappa_2 \end{pmatrix}, \quad \mathrm{Cov}\begin{Bmatrix} z(t) \\ [z_\delta](t) \\ [z](t) \end{Bmatrix} = t\begin{pmatrix} \kappa_2 & \kappa_3 & \kappa_3 \\ \kappa_3 & \kappa_4 + 3\kappa_2^2 t M^{-1} & \kappa_4 \\ \kappa_3 & \kappa_4 & \kappa_4 \end{pmatrix},$$

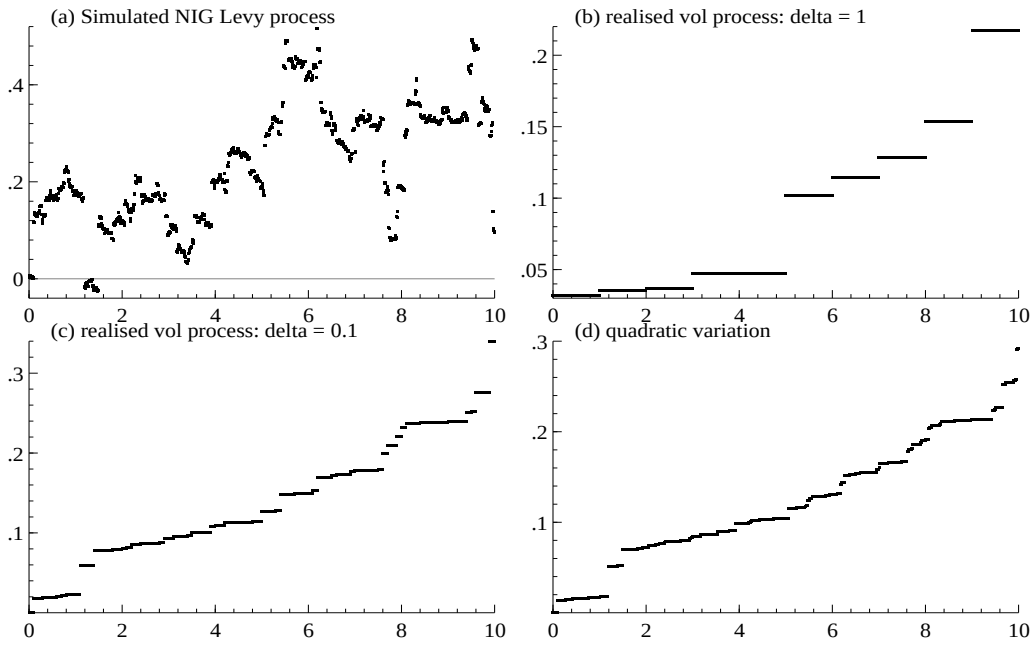

Figure 2.8: *Figure (a) Sample path of NIG(0.2,0,0,10) Lévy process. (b) Sample path of corresponding realised variance process taking $\delta = 1$ — one observation per unit of time. (c) Same but with $\delta = 1/10$ — ten observations per unit of time. (d) Quadratic variation of the process. Code:* `levy_code.ox`.

where κ_r denotes the r-th cumulant of $z(1)$. The only one of these results which is not straightforward is

$$\begin{aligned}\mathrm{Var}([z_\delta](t)) &= M\mu_4\left[z(tM^{-1})\right] \\ &= M\left\{\kappa_4\left[z(tM^{-1})\right] + 3\kappa_2\left[z(tM^{-1})\right]^2\right\} \\ &= M\left(tM^{-1}\kappa_4 + 3t^2M^{-2}\kappa_2^2\right).\end{aligned}$$

Finally we notice the implication that $[z_\delta] - [z]$ has a zero mean, while

$$\mathrm{Var}\left\langle [z](t) - \{z_\delta\}(t)\right\rangle = 3\kappa_2^2 t^2 M^{-1}.$$

Hence we could use $[z_\delta]$ as an estimator of $[z]$.

A simple example of this is where z is standard Brownian motion, then $\kappa_3 = \kappa_4 = 0$, which means that

$$\mathrm{E}\left\{\begin{array}{c} z(t) \\ [z_\delta](t) \\ [z](t) \end{array}\right\} = t\left(\begin{array}{c} \kappa_1 \\ \kappa_2 \\ \kappa_2 \end{array}\right), \quad \mathrm{Cov}\left\{\begin{array}{c} z(t) \\ [z_\delta](t) \\ [z](t) \end{array}\right\} = t\left(\begin{array}{ccc} \kappa_2 & 0 & 0 \\ 0 & 3\kappa_2^2 tM^{-1} & 0 \\ 0 & 0 & 0 \end{array}\right),$$

which makes sense for $[z](t) = t$. If z is the homogeneous Poisson process, then all the cumulants are equal to κ_1. Thus

$$\mathrm{E}\left\{\begin{array}{c} z(t) \\ [z_\delta](t) \\ [z](t) \end{array}\right\} = t\kappa_1\iota, \quad \mathrm{Cov}\left\{\begin{array}{c} z(t) \\ \{z_\delta\}(t) \\ [z](t) \end{array}\right\} = t\kappa_1\left(\begin{array}{ccc} 1 & 1 & 1 \\ 1 & 1+3\kappa_1 tM^{-1} & 1 \\ 1 & 1 & 1 \end{array}\right).$$

Notice the covariance is singular, for $z(t) = [z](t)$.

A more interesting example occurs when z is scaled Brownian motion which is deformed by the subordinator τ. Then

$$\mathrm{K}\left\{\theta \ddagger z(1)\right\} = \mathrm{K}\left\{\frac{1}{2}\theta^2\sigma^2 \ddagger \tau(1)\right\}.$$

This implies

$$\kappa_1 = 0, \quad \kappa_2 = \sigma^2\kappa_1(\tau), \quad \kappa_3 = 0, \quad \kappa_4 = 3\sigma^2\kappa_2(\tau),$$

where $\kappa_s(\tau)$ denotes the s-th cumulant of $\tau(1)$. Hence in this case

$$\mathrm{Var}\left\langle [z](t) - [z_\delta](t)\right\rangle = 3\sigma^4\kappa_1^2(\tau)t^2M^{-1},$$

which only depends upon the mean of the subordinator, not its variance.

2.7 Lévy processes and stochastic analysis

2.7.1 Stochastic integrals

This section will assume a basic knowledge of stochastic analysis — that is the calculus of stochastic integrals based on semimartingales. For those unfamiliar with this background we have provided a very short primer to this material in Section A.

All Lévy processes are semimartingales. So, in particular, we can consider stochastic integrals of the form

$$f(\cdot, A) \bullet z,$$

where z is a Lévy process, f is a real function on $\mathbf{R}_x \times \mathbf{R}$ and A is a càglàd stochastic process, and f satisfies some mild regularity condition ensuring that the process $f(\cdot, A)$ is again càglàd.

2.7.2 Lévy-Ito representation of Lévy processes

Consider first the case of Lévy subordinators. It can be shown that any non-negative Lévy process z is representable in the *Lévy-Ito representation*

$$z(t) = \int_0^t \int_0^\infty xN(\mathrm{d}x, \mathrm{d}s). \tag{2.29}$$

Here $N(\mathrm{d}x, \mathrm{d}t)$ denotes a Poisson field on $\mathbf{R}_+ \times \mathbf{R}_+$ with mean measure $\nu(\mathrm{d}x, \mathrm{d}t)$. While the integral in (2.29) is well defined in wide generality under mild restrictions on the measure $\nu(\mathrm{d}x, \mathrm{d}t)$, to make z a Lévy process we must require the measure to factorise as

$$\nu(\mathrm{d}x, \mathrm{d}t) = W(\mathrm{d}x)\mathrm{d}t,$$

with $W(\mathrm{d}x)$ satisfying

$$\int_0^\infty \min\{1, x\}W(\mathrm{d}x) < \infty,$$

in order for the stochastic integral in (2.29) to exist. It is clear from the expression (2.29) that z is a process with non-negative, independent and stationary increments, i.e. a subordinator.

Suppose $z(t)$ is a finite activity process, then it can be written as a compound Poisson process $z(t) = \sum_j^{N(t)} c_j$, with a mean measure

$$\nu(\mathrm{d}x, \mathrm{d}t) = W(\mathrm{d}x)\mathrm{d}t = \psi P(\mathrm{d}x \ddagger c_1)\mathrm{d}t,$$

where $P(\mathrm{d}x \ddagger c_1)$ is the probability measure and ψ is the intensity of the Poisson process.

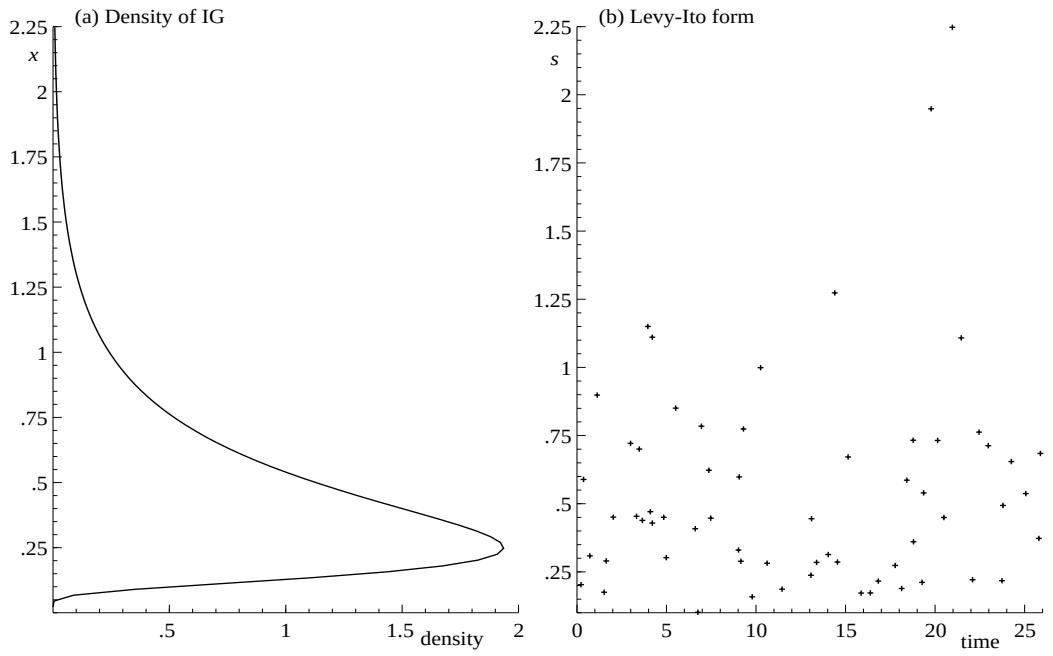

Figure 2.9: *(a) rotates the IG(1,2) density for c_1. (b) is the Lévy-Ito form, displaying the Poisson field $N(x,t)$. File name is* `levy_graphs.ox`.

Example 6 *Figure 2.9 shows a simulation of the resulting Poisson field*

$$N(x,t) = \int_0^x \int_0^t N(\mathrm{d}u, \mathrm{d}s),$$

in the case where $c_1 \sim IG(\delta,\gamma)$, taking $\psi = 3$, $\delta = 1$ and $\gamma = 2$. We drew this by first simulating a homogeneous Poisson process with rate ψ, and then assigning height according to draws from the $IG(\delta,\gamma)$ distribution.

It is not possible to correctly draw an infinite activity process for we would have to draw an infinite number of points in the Poisson field, although most of them would be have very little height.

For a general Lévy process z we have the *Lévy-Ito representation*

$$\begin{aligned} z(t) &= at + bw(t) \\ &\quad + \int_{\{|x|<\varepsilon\}} x\{N(\mathrm{d}x,t) - t\nu(\mathrm{d}x)\} \\ &\quad + \sum_{0<s\le t} \mathbf{1}_{\{|\Delta z_s|\ge\varepsilon\}} \Delta Z(s), \end{aligned} \tag{2.30}$$

where $\varepsilon > 0$, w is a Brownian motion while, for any set Λ, $0 \notin \bar{\Lambda}$ (the closure of Λ), $N_t^\Lambda = \int_\Lambda N_t(\cdot, \mathrm{d}x)$ is a Poisson process with mean $t\nu(\Lambda)$, ν being a Lévy measure on $\mathbf{R}$. Furthermore, N_t^Λ is independent of W and N_t^Λ is independent of N_t^Γ if Λ and Γ are disjoint.

For a proof see, for instance, Jacod and Shiryaev (1987, Section ??).

2.7.3 Quadratic Variation

Section 2.6 showed us that it is obvious that if X is a Lévy process then its quadratic variation, $[X]$, is also a Lévy process, in fact a subordinator. An example of this is where z is a compound

Poisson process

$$z(t) = \sum_{i=1}^{N(t)} y_i \quad \text{and then} \quad [z](t) = \sum_{i=1}^{N(t)} y_i^2.$$

In some cases it is helpful to work with an alternative, and equivalent, definition of QV which is written in terms of a stochastic integral. It is that

$$[X] = X^2 - 2X_- \bullet X. \tag{2.31}$$

This is discussed in some detail in our primer on stochastic analysis. For now we give an example.

Example 7 *Let N be a Poisson process and let us check the consistency of the formulae (2.31) and (2.28). Suppose $N_t = n$. It is immediate from (2.28) that*

$$[N](t) = n$$

while, on the other hand,

$$N^2 - 2\int_0^t N(s-)\mathrm{d}N(s) = n^2 - 2\sum_{i=1}^{n}(i-1) = n^2 - 2\binom{n}{2} = n.$$

Example 8 *More generally still, for z an arbitrary subordinator we find from formula (2.29)*

$$[z](t) = \int_0^t \int_0^\infty x^2 N(\mathrm{d}x, \mathrm{d}s)$$

We can derive this formally by using the Ito algebra rules (described in (A.28, A.29, A.30)). By (2.29),

$$\mathrm{d}z(t) = \int_0^\infty xN(\mathrm{d}x, \mathrm{d}t),$$

so

$$\begin{aligned}(\mathrm{d}z(t))^2 &= \int_0^\infty xN(\mathrm{d}x,\mathrm{d}t)\int_0^\infty yN(\mathrm{d}y,\mathrm{d}t) \\ &= \int_0^\infty\int_0^\infty xy(N(\mathrm{d}x,\mathrm{d}t))^2 \\ &= \int_0^\infty x^2 N(\mathrm{d}x,\mathrm{d}t).\end{aligned}$$

2.7.4 Stochastic exponential of a Lévy process

TO BE ADDED

2.8 Multivariate Lévy processes

2.8.1 Overview

An important question is how to generate multivariate Lévy processes, that is processes with independent and stationary multivariate increments. Here we discuss just two approaches: linear transformation and time deformation.

Suppose $u(t)$ and $v(t)$ are independent Lévy processes and Θ is some non-diagonal matrix. Then the linear combinations of the original Lévy processes

$$z(t) = \Theta \left\{ \begin{array}{c} u(t) \\ v(t) \end{array} \right\},$$

is a bivariate Lévy process. The elements of z are marginally Lévy processes. This type of argument generalises to any dimension.

We saw in Section 2.5 that the use of subordinators can be used to generate compelling Lévy processes. Here we use this idea to put

$$z(t) = \left\{ \begin{array}{c} u(\tau(t)) \\ v(\tau(t)) \end{array} \right\},$$

where τ is an independent, common subordinator. This means that $\{u(\tau(t)), v(\tau(t))\}$ is now a dependent series. A concrete example of this is where $\{u(t), v(t)\}$ are independent standard Brownian motions, then

$$z(t)|\tau(t) \sim N(0, \tau(t)I),$$

which implies the elements of $z(t)$ are uncorrelated but are dependent. In particular

$$\mathrm{Cov}(z_1^2(t), z_2^2(t)) = \mathrm{E}\tau(t).$$

2.8.2 Example: multivariate generalised hyperbolic Lévy process

Suppose we take $v(t)$ as a $d \times 1$ vector of correlated Brownian motions generated by

$$v(t) = t\Sigma\beta + \Sigma^{1/2}u(t),$$

where u is a $d \times 1$ vector of independent, standard Brownian motions. Then we take τ to be an independent subordinator and define the deformed series

$$z(t) = \mu t + v(\tau(t)).$$

Then $z(t)$ is a *multivariate type G* Lévy process with

$$z(t)|\tau(t) \sim N(\mu t + \tau(t)\Sigma\beta, \tau(t)\Sigma).$$

Suppose we choose to make τ a $GIG(\nu, \delta, \gamma)$ Lévy process, then we say that z is a multivariate generalised hyperbolic Lévy process, following our earlier work on the univariate process discussed in Section 2.4.1. In particular the increments of such a process are independent and stationary while the density of $z(1)$ is known to follow a multivariate $GH(\nu, \alpha, \beta, \mu, \delta, \Sigma)$ density

$$f_{z(1)}(x) = \frac{(\gamma/\delta)^{\nu}}{(2\pi)^{d/2}\,\alpha^{2\left(\nu-\frac{d}{2}\right)}K_{\nu}\left(\delta\gamma\right)}\left\{\alpha q\left(x-\mu\right)\right\}^{\left(\nu-\frac{d}{2}\right)}K_{\left(\nu-\frac{d}{2}\right)}\left\{\alpha q\left(x-\mu\right)\right\}\exp\left\{\beta'\left(x-\mu\right)\right\}, \tag{2.32}$$

where

$$q(x) = \sqrt{\delta^2 + (x-\mu)'\,\Sigma^{-1}\,(x-\mu)} \qquad \text{and} \qquad \alpha = \beta'\Sigma\beta.$$

Here Σ allows us to model the correlation between the processes, while ν, δ, and γ controls the tails of the density. We have a whole vector β which freely parameterises the skewness of the returns. In order to enforce identification on this model it is typical to assume that

$$|\Sigma| = 1. \tag{2.33}$$

Of course the multivariate GH density has many interesting special cases such as the multivariate Student t, normal gamma, normal inverse Gaussian, hyperbolic and Laplace. Of course this important distribution can be thought of as a scale location mixture with

$$y = \mu + \Sigma\beta\sigma^2 + \sigma\Sigma^{1/2}\varepsilon, \qquad \varepsilon \sim N(0, I) \perp\!\!\!\perp \sigma^2 \sim GIG(\nu, \delta, \gamma).$$

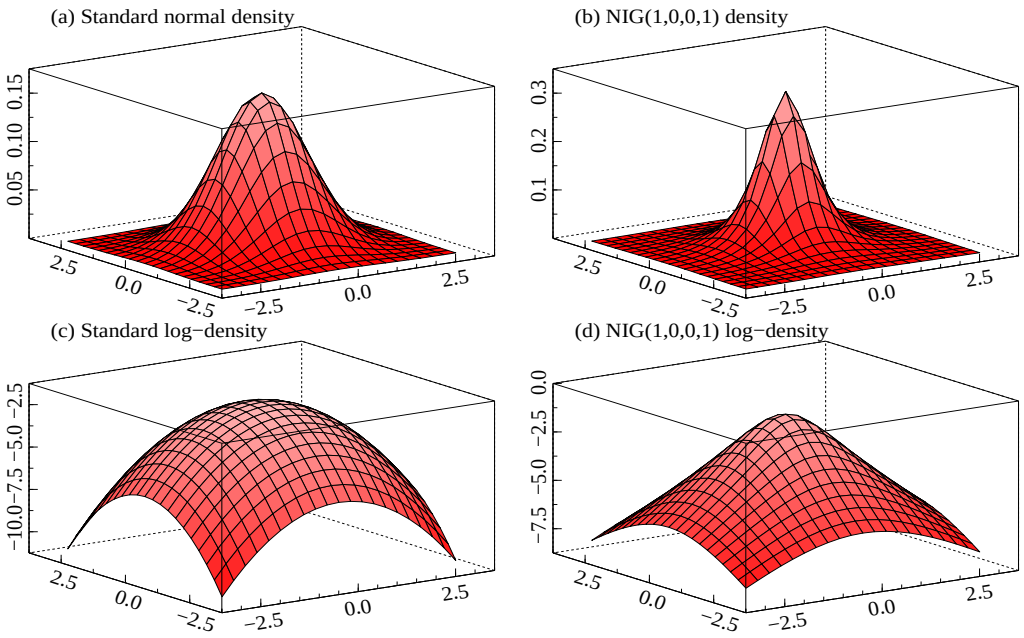

Figure 2.10: *Graph of the densities and log-densities of N(0,I) and NIG(1,0,0,1,I) variables. (a) Density of N(0,I). (b) Density of NIG variables. (c) Log-density of N(0,I). (d) Log-density of bivariate NIG variables. Code:* `levy_graphs.ox`.

This is important, for it immediately follows that linear combinations of y are also GH, while if we write $y = (x', z')'$ then $z|x$ is GH. Hence many of the important attractive features of the multivariate Gaussian distribution carry over to the multivariate generalised hyperbolic distribution.

A simple example of the above multivariate distributions of $z(1)$ is given in Figure 2.10 which draws the density and log-density for the bivariate standard normal and the corresponding $NIG(1, 0, 0, 1)$ (chosen so that the marginal variances of the variables is one) variables. Again the log-densities show that the tails of the NIG variables are much thicker — looking roughly linear in all tails. This has a very big impact on the chance of observing two observations in the tails of the distribution.

2.8.3 Quadratic covariation

Definitions

The idea of quadratic variation extends to the multivariate case. We again split time into small intervals

$$t_0^r = 0 < t_1^r < ... < t_{m_r}^r = t.$$

Then the quadratic covariation series is defined as

$$[z](t) = \text{p}-\lim_{r\to\infty} \sum \{z(t_{i+1}^r) - z(t_i^r)\}\{z(t_{i+1}^r) - z(t_i^r)\}', \tag{2.34}$$

where

$$\sup_i \{t_{i+1}^r - t_i^r\} \to 0 \quad \text{for} \quad r \to \infty.$$

For semimartingales

In the case of multivariate semimartingales we can decompose

$$z(t) = a(t) + m(t),$$

a predictable component of locally bounded variation and a local martingale, respectively. Then for all semimartingales

$$[z](t) = [m](t) + \sum_{0\leq u\leq t} \{a(u) - a(u-)\} \{a(u) - a(u-)\}',$$

the quadratic covariation of the martingale component plus the outer product of the jumps in the predictable component. Quadratic covariation is important. One of the reasons for this is that for all special semimartingales with continuous predictable components

$$\mathrm{Cov}(\mathrm{d}z(t)|\mathcal{F}_t) = \mathrm{E}\left(\mathrm{d}[z](t)|\mathcal{F}_t\right),$$

so long as the moments exist.

2.9 Conclusion

To be added.

2.10 Appendix of derivations and proofs

Dynamics of $\sigma(t)$: Ito's formula and stochastic volatility.

Suppose $\sigma^2(t)$ is the càdlàg version of a positive OU process, satisfying the SDE

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

Then, σ^2 is of bounded variation, implying $[\sigma^2]^c = 0$, and by Ito's formula we obtain for the square root process

$$\begin{aligned}
\sigma(t) &= \frac{1}{2}\int_0^t \frac{1}{\sigma(s)}\mathrm{d}\sigma^2(s) + \sum_{0<s\leq t}\left\{\sigma(s) - \sigma(s_-) - \frac{1}{2}\frac{1}{\sigma(s_-)}\Delta\sigma^2(s)\right\} \\
&= -\frac{1}{2}\lambda\int_0^t \frac{\sigma^2(s)}{\sigma(s_-)}\mathrm{d}s - \frac{1}{2}\lambda\int_0^t \frac{1}{\sigma(s_-)}\mathrm{d}z(\lambda s) \\
&\quad + \sum_{0<s\leq t}\left[\sigma(s) - \sigma(s_-) - \frac{1}{2}\frac{1}{\sigma(s_-)}\left\{\sigma^2(s) - \sigma^2(s-)\right\}\right] \\
&= -\frac{1}{2}\lambda\int_0^t \sigma(s)\mathrm{d}s - \frac{1}{2}\lambda\int_0^t \frac{1}{\sigma(s_-)}\mathrm{d}z(\lambda s) \\
&\quad + \sum_{0<s\leq t}\{\sigma(s) - \sigma(s_-)\}\left\{1 - \frac{1}{2}\frac{\sigma(s) + \sigma(s-)}{\sigma(s_-)}\right\}
\end{aligned}$$

or, equivalently,

$$\begin{aligned}
2\sigma(t) &= -\lambda\sigma^*(t) - \lambda\int_0^t \frac{1}{\sigma(s_-)}\mathrm{d}z(\lambda s) \\
&\quad + \sum_{0<s\leq t}\{\sigma(s) - \sigma(s_-)\}\left\{2 - \frac{\sigma(s)}{\sigma(s_-)}\right\}.
\end{aligned}$$

'OU criterion'

We shall be particularly concerned with integrals of the form

$$X_t = \int_0^t e^{-s} \mathrm{d}Z_s.$$

Let us consider this for the special case where Z is a subordinator. Then we may take the integral as being determined pathwise, as the Stieltjes integral. The process X is clearly nonnegative and increasing and hence a chronometor. We write

$$X_\infty = \int_0^\infty e^{-s} \mathrm{d}Z_s = \lim_{t \uparrow \infty} X_t.$$

Under a mild condition on Z, X_∞ will be finite almost surely, that is it will be a random variable.

To determine the relevant condition we first note that since each X_t is infinitely divisible the same will hold for X_∞ provided it is finite almost surely.

Now, consider the kumulant function of X_t which, by Lévy-Khintchine representation is

$$\bar{\mathrm{K}}\{\theta \ddagger X_t\} = -\int_0^t \int_0^\infty \{1 - \exp(-e^{-s}\theta x)\} W(\mathrm{d}x)\mathrm{d}s$$

Using the substitutions $y = e^{-s}x$ and $r = e^s$ may rewrite this as

$$\begin{aligned} \bar{\mathrm{K}}\{\theta \ddagger X_t\} &= -\int_0^t \int_1^\infty \{1 - \exp(-\theta y)\} W(e^s \mathrm{d}y)\mathrm{d}s \\ &= -\int_1^\infty \int_1^{e^t} \{1 - \exp(-\theta y)\} W(r\mathrm{d}y) r^{-1}\mathrm{d}r \\ &= -\int_1^\infty \{1 - \exp(-\theta y)\} U_t(\mathrm{d}y) \end{aligned}$$

where

$$U_t(\mathrm{d}y) = \int_1^{e^t} W(r\mathrm{d}y) r^{-1}\mathrm{d}r$$

is the Lévy measure of X_t. Now suppose for simplicity that the Lévy measure W is absolutely continuous with a density w. Then, for any $t > 0$, U_t is also absolutely continuous, with density

$$u_t(y) = \int_1^{e^t} w(ry)\mathrm{d}r = y^{-1} \int_y^{e^t y} w(x)\mathrm{d}x$$

and for $t \to \infty$

$$u_t(y) \to u(y) = y^{-1} W^+(y)$$

For X_∞ to be a random variable this limiting function u should be a Lévy density, i.e. it should satisfy the integrability condition (2.15). Noting first that

$$W^+(y) = \int_y^\infty w(x)\mathrm{d}x = y \int_1^\infty w(y\tau)\mathrm{d}\tau$$

we find

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

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

In the latter expression the first and third integrals are finite since W is a Lévy measure. We are thus led to the condition

$$\int_{1}^{\infty}\log y w(y)\mathrm{d}y < \infty$$

for finiteness almost surely of X_∞.

2.11 Exercises

2.12 Bibliographic notes

2.12.1 Lévy processes

Lévy processes were introduced by Lévy (1937) who developed the theory of infinite divisibility. Modern accounts of the probability theory of Lévy processes are given in Bertoin (1996) and Sato (1999). See also Ito (1969), Rogers and Williams (1994, pp. 73–84) and Bertoin (2001). A reasonably accessible overview of the theory and uses of Lévy processes is given in Barndorff-Nielsen, Mikosch, and Resnick (2001). A compact account in the context of finance is given by Shiryaev (1999, pp. 200-206).

The simulation of Lévy processes has to be carried out with some care. There are extensive results available. Some of the most useful are the infinite series representation developed by Rosinski (2001). The special case of gamma process simulation is discussed by Wolpert and Ickstadt (1999), while some more general discussion is given in Walker and Damien (2000). We should also note the important recent contribution of Asmussen and Rosinski (2000).

2.12.2 Flexible distributions

Most of modern financial economics is built out of Brownian motion and the corresponding Itô calculus. In this Chapter we have discussed many familiar alternative distributions like the Poisson, normal gamma, student t and Laplace. The latter two are particularly noteworthy as they have been used as empirical models for log-prices. In early work Praetz (1972) and Blattberg and Gonedes (1974) suggested modelling the increments to log-prices using a student t distribution. This model was not set in continuous time, but we have seen above that it is possible to construct a Lévy process to justify this type of modelling. Further the model can be extended to allow for asymmetry. This is a special case of the generalised hyperbolic distribution, but the fact that it provides a non-symmetric student t distribution has not been discussed explicitly before. More recently Granger and Ding (1995) have advocated the use of Laplace distributions to model discrete time returns, while the non-linear Brownian motion based Cox, Ingersoll, and Ross (1985) processes have gamma marginals and so normal gamma distributions are often implicitly used in econometrics. It turns out that fitted values of the normal gamma distribution are typically thinner tailed than the corresponding student or the Laplace.

In this Chapter we have placed a great deal of emphasis on generalised hyperbolic and generalised inverse Gaussian distributions. We have carried this out for they support Lévy

processes, are empirically flexible, encompass many of the familiar models econometricians are accustomed to and are mathematically tractable. However, their generality and some of the special cases are not so familiar.

The hyperbolic distribution and its extension to the generalised hyperbolic distribution was introduced in Barndorff-Nielsen (1977) in order to describe the size distribution of sand grains in samples of windblown sands. This was motivated by empirical observations due to Brigadier R. A. Bagnold, F.R.S. who noted that in double logarithmic plotting (that is both the horizontal and vertical axes are plotted on the logarithmic scale) the histograms looked strikingly as following hyperbolae, the slopes of the asymptotes being related to the physical conditions under which the sand was deposited; see Bagnold (1941) (note the similarity to the Granger and Ding (1995) empirics). Subsequently, it was discovered that the hyperbolic shape, or shapes very close to that, occur in a very wide range of empirical studies, for instance in other areas of geology, in turbulence, in paleomagnetism, in relativity theory and in biology. For surveys of developments up till the mid-1980ies, see Barndorff-Nielsen, Blæsild, Jensen, and Sørensen (1985). The generalised inverse Gaussian distribution is due to Étienne Halphen in 1946, while it was briefly discussed by Good (1953). A detailed discussion of this distribution was given by Jørgensen (1982).

Following a suggestion by Barndorff-Nielsen, Ernst Eberlein and coworkers began an investigation of the applicability of the hyperbolic laws in finance and this has developed into a major project, see Eberlein and Keller (1995) on their initial work. We will discuss their application of the associated Lévy processes in a moment. Bauer (2000) discusses the use of these models in the context of value at risk.

When deviations from the hyperbolic shape occurred they typically showed somewhat heavier tails than the hyperbolic. This led Barndorff-Nielsen to consider more closely another of the generalised hyperbolic laws, the normal inverse Gaussian, which had until then received scarce attention (although work by Sichel (1973) on the distribution of the size of diamonds is an important exception), but turned out not only to fit a much wider range of data but also to possess various nice mathematical properties not shared by the hyperbolic (Barndorff-Nielsen (1997), Barndorff-Nielsen (1998b), Barndorff-Nielsen (1998a)).

The class of tempered stable distributions was introduced by Tweedie (1984). Hougaard (1986) discussed their applicability in survival analysis. See also Jørgensen (1987) and Brix (1999). The normal variance-mean mixtures with TS mixing was introduced by Barndorff-Nielsen and Shephard (2002a), who also extended this concept to the normal modified stable distribution.

2.12.3 Lévy processes in finance

The use of normal gamma based Lévy processes in finance was pioneered by Madan and Seneta (1990) and Madan, Carr, and Chang (1998) who paid particular attention to their use in option pricing. Recent extensions of this work include Carr, Geman, Madan, and Yor (2002).

The thicker tailed hyperbolic distribution and Lévy process was studied extensively in the context of finance by Eberlein and Keller (1995), who also discussed the corresponding option pricing theory and practice in Eberlein, Keller, and Prause (1998) and Eberlein (2000). This work is possible because the generalised inverse Gaussian distribution were shown to be infinitely divisible by Barndorff-Nielsen and Halgreen (1977). The even thicker tailed normal inverse Gaussian process is studied by Barndorff-Nielsen (1997), while Rydberg (1997b) and Rydberg (1997a) discusses both fitting the process to financial data and simulating from it. Prause (1998) and Raible (1998) have recently written first rate Ph.D. theses on generalised hyperbolic Lévy processes under the supervision of Ernst Eberlein. Both of these theses have a wealth of information on this topic. Bingham and Kiesel (2000) looks at the use of hyperbolic processes

in finance, while Bibby and Sørensen (2001) reviews the area of generalised hyperbolic processes in finance.

The idea of time deformation or subordination is due to Bochner (1949) and Bochner (1955), while it was introduced into economics by Clark (1973) who suggested the use of volume statistics as a subordinator, placing particular weight on studying the implications of using a lognormal subordinator. At that stage we did not know that this was a valid mathematical construction for it was not until Thorin (1977) that the lognormal was shown to be infinitely divisible. See also Bondesson (2000) for up to date treatment of lognormal Lévy processes. Epps and Epps (1976) and Tauchen and Pitts (1983) further studied the relationship between volume and the variance of the increments to prices. Recent discussions of this includes Ané and Geman (2000). Stock (1988) used the concept of subordination in a wider economic context outside finance.

Mandelbrot (1963) and Mandelbrot and Taylor (1967) introduced the concepts of self-similarity and stable Lévy processes into financial economics. Almost immediately the main stream academic profession rejected these models, after some initial support from Fama (1965), due to their lack of empirical fit as most research papers suggested the existence of at least two moments for returns. An elegant discussion of the move away from these models and its importance is given in Campbell, Lo, and MacKinlay (1997, pp. 17-19). However, there still remains a small group of researchers who push in this area. Recent work is discussed by Rachev and Mittnik (2000).

Truncated Lévy flights were introduced by Mantegna and Stanley (1994), while it has been pioneered in finance in Mantegna and Stanley (1996) and Mantegna and Stanley (2000). The extended Koponen class has been considered by Novikov (1994), Koponen (1995), Mantegna and Stanley (2000), Boyarchenko and Levendorskii (1999), Boyarchenko and Levendorskii (2000a), Boyarchenko and Levendorskii (2000b), Boyarchenko and Levendorskii (2000c), Boyarchenko and Levendorskii (2000d), Carr, Geman, Madan, and Yor (2002), and Rosinski (2001). Carr, Geman, Madan, and Yor (2002) called these models CGMY processes after their own initials. We have not followed that nomenclature. Meixner distributions were introduced by Schoutens and Teugels (1998) and have been studied in the context of Lévy based models for finance by Schoutens and Teugels (2001) and Grigelionis (1999). In Geman, Madan, and Yor (2000) what are here called normal tempered stable (NTS) NTS Lévy processes have been studied, in the case of zero drift, from a viewpoint different from the one of the present book. Ben-Hamou (2000) has studied estimating the parameters of the Lévy process from option prices.

Recently Andersen, Bollerslev, Diebold, and Labys (2001a) have discussed the application of the theory of quadratic variation in the context of arbitrage-free financial markets, while Barndorff-Nielsen and Shephard (2001a) emphasised its role in the context of chronometers and stochastic volatility. Both of these papers will be discussed extensively in later chapters of this book, where a detailed treatment of realised variance will be given.

2.12.4 Empirical fit of Lévy processes

There is a large literature on studying the fit of various parametric models to the marginal distribution of returns of speculative assets. Most of these papers are not based on a background of a Lévy process and so risked fitting an incoherent (from a continuous time viewpoint) model. An example of this is Praetz (1972) in his work on the student t distribution. A notable exception is Mandelbrot (1963) where he used stable distributions and related this to stable processes.

The likelihood methods we used to fit the models are entirely standard. We have used profile likelihoods to compute measures of uncertainty as these are known to be more reliable than using the first order Gaussian asymptotic distribution. A discussion of this literature is given in Barndorff-Nielsen and Cox (1994, Section 3.4). The use of profile likelihoods for ν in the generalised hyperbolic is new as was the use of the EM algorithm in this context. Independent and concurrent work on the use of the EM algorithm for this problem was carried out by

Protassov (2001). An elegant discussion of the EM algorithm is given in Tanner (1996). The theory of robust standard errors for maximum likelihood estimation is standard in econometrics and statistics. Leading references are White (1982) and White (1994).

Barndorff-Nielsen and Prause (2001) showed that the Olsen scaling law is explained by the NIG Lévy process.

Chapter 3

Simulation and inference for Lévy processes

Abstract: This Chapter provides an introduction to the simulation of Lévy processes, as well as inference methods for estimating and testing particular parameteric Lévy processes. Univariate and multivariate models are discussed. A detailed bibliographical review is given at the end of the Chapter. It will not come as a surprise to our readers that Lévy processes allow a flexible model for the marginal distribution of returns, but will be rejected as financial returns have important serial dependence structures. This will prompt the development of more general time deformation models, introduced in the next Chapter. Special cases of these will be stochastic volatility models built from Lévy processes.

3.1 What is this Chapter about?

In this Chapter we provide a first course on simulating, estimating and testing Lévy processes in the context of financial economics. The Chapter will refer to some common datasets discussed in detail in Chapter 1 and will delay the discussion of literature on this topic until the end of this Chapter. Throughout we hope our treatment will be as self-contained as possible.

This Chapter has 4 other sections, whose goals are to:

- Describe various methods for simulating from Lévy processes.
- Estimate various Lévy processes from financial data.
- Draw conclusions to the Chapter.
- Discuss the literature associated with Lévy processes.

3.2 Simulating Lévy processes

3.2.1 Simulation

Many aspects of financial economics and statistics require us to be able to simulate from processes in order, for example, to make inference about parameters indexing the model or to price derivatives written on the underlying assets. Here we discuss the simulation of subordinators, which in turn would allow us to simulate processes on the real line by time deforming a simulated Brownian motion.

Suppose we know how to simulate from the marginal distribution of $z(t)$ for any t, then the process z can be exactly simulated at a fixed mesh of points in time $0 = t_0 < t_1 < t_2 < ... < t_n$. In particular we build up the process by iterating

$$z(t_j) = z(t_{j-1}) + u_j, \quad \text{where} \quad u_j \stackrel{\mathcal{L}}{=} z(t_j - t_{j-1}), \quad j = 1, 2, ..., n, \tag{3.1}$$

where the $\{u_j\}$ are independent and $z(0) = 0$. This was the method we used to simulated Poisson, IG, Γ and compound Lévy processes for Figures 2.1 and 2.2. In those pictures we used $t_j - t_{j-1} = 1/2000$.

At a more abstract level we can approximately simulate from the entire process, rather than at a mesh of points, using the construction

$$z_\Delta(t) = \sum_{j=1}^{\lfloor t/\Delta \rfloor} u_j, \quad u_j \stackrel{\mathcal{L}}{=} z(\Delta), \quad \Delta > 0. \tag{3.2}$$

Now as $\Delta \to 0$ the process z_Δ converges to z.

3.2.2 Simulating the paths by rejection in the tempered stable case

Method Many Lévy processes do not have easily computed densities for $z(t)$, which makes it hard to directly simulate from u_j. An important example of this is the TS process, where the probability density is not available but where the Lévy density has the simple form

$$w(x) = Ax^{-\kappa-1}\exp(-Bx), \quad x, A, B > 0 \quad \text{and} \quad \kappa \in (0,1).$$

Recently Rosinski has based an asymptotic representation of the path of a TS Lévy process on its Lévy density.

The *Rosinski rejection method* approximates the path by

$$z_I(t) = \sum_{i=1}^{I} \min\left\{\left(\frac{As}{b_i\kappa}\right)^{1/\kappa}, B^{-1}e_i v_i^{1/\kappa}\right\} I(u_i \le t), \quad \text{for} \quad 0 \le t \le s, \tag{3.3}$$

where $I(\cdot)$ is an indicator function, $\{e_i\}$, $\{v_i\}$, $\{b_i\}$, $\{u_i\}$ are independent of one another and over i except for the $\{b_i\}$ process. Here $u_i \overset{i.i.d.}{\sim} U(0,s)$, $v_i \overset{i.i.d.}{\sim} U(0,1)$, the $\{e_i\}$ are exponential with mean 1. Further the $b_1 < ... < b_i < ...$ are the arrival times of a Poisson process with intensity 1. Then as $I \to \infty$ the process $z_I(t)$ converges uniformly to a sample path of a tempered stable Lévy process. Hence this is a more sophisticated version of the type of result given in (3.2). If we had a more limited goal of simulating the process at a single point $z(\Delta)$, then the summation simplifies to

$$z_I(\Delta) = \sum_{i=1}^{I} \min\left\{\left(\frac{A\Delta}{b_i\kappa}\right)^{1/\kappa}, B^{-1}e_i v_i^{1/\kappa}\right\}.$$

The important features of this method is that

- (4.21) simulates the whole path directly. The random variables $\{e_i\}$, $\{v_i\}$, $\{b_i\}$, $\{u_i\}$ are drawn once and we just plot out (4.21) at whatever resolution is required.
- As $I \to \infty$, z_I converges from below to z. Further $\left(\frac{As}{b_i\kappa}\right)^{1/\kappa}$ monotonically declines with i.
- The underlying random numbers $\{e_i\}$, $\{v_i\}$, $\{b_i\}$, $\{u_i\}$ are parameter free. Suppose we fix these random numbers and computed the z_I function for a variety of values of parameters A, B and κ. The resulting sample paths would change continuously (but not differentiably) with the parameters.
- The rate of convergence of the sum increases as $\kappa \to 0$ for it can be shown that

$$\min\left\{\left(\frac{A\Delta}{b_i\kappa}\right)^{1/\kappa}, B^{-1}e_i v_i^{1/\kappa}\right\} = O_p(i^{-1/\kappa}).$$

Illustration: $IG(\delta,\gamma)$ **process** Figure 3.1 shows the results from simulating an $IG(\delta = 0.2, \gamma = 10)$ Lévy process over the interval $[0,10]$ using the Rosinski approach. This fits in the TS framework by taking $\kappa = 0.5$, $A = (2\pi)^{-1/2}\delta$ and $B = \gamma^2/2$. An important aspect of the Rosinski approach is the choice of I. As I increases we converge to the truth, from below. We can see that the effect of the truncation of the infinite series is small when $I > 1,000$ and tiny for $I > 10,000$.

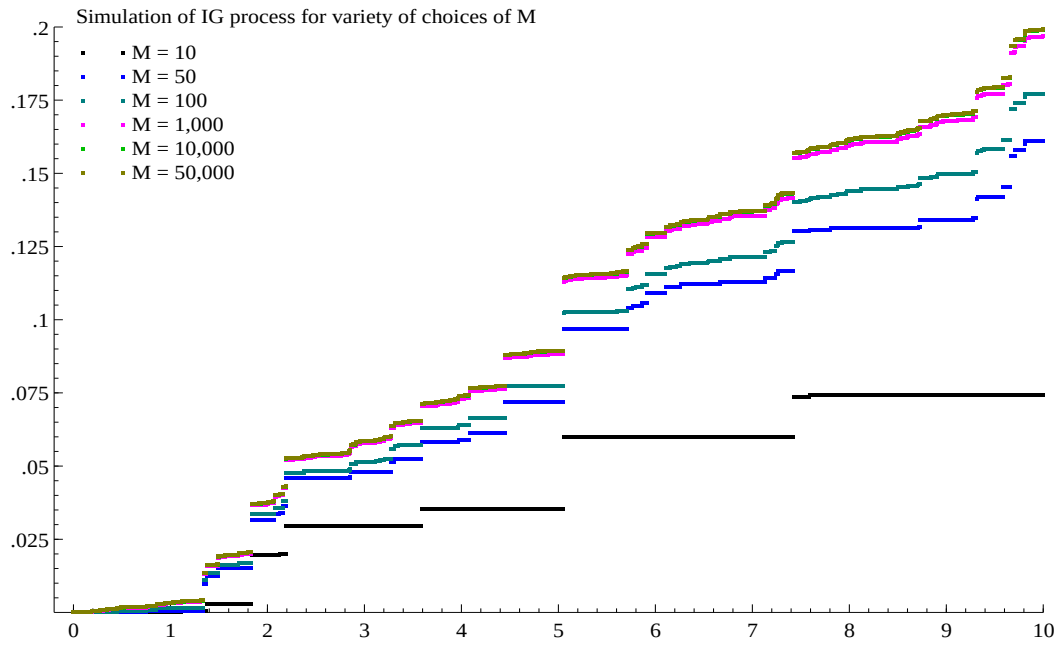

Figure 3.1: *Simulating IG process using Rosinski's method for TS processes. Shows the simulation varying M. SAs M increases the curves move upwards. The lines for $M = 10,000$ and $M = 50,000$ are on top of one another. The line for $M = 1,000$ is slightly below. Code:* `rosin_rej.ox`.

3.2.3 Simulating the paths via the inverse tail integral

The theory

The above rejection based method is rather easy to use, but it only covers the *TS* case. More generally we can work with a related series representation based on the positive Lévy density $w(x)$, $x \in R_+$, for $z(1)$. Then define the tail mass function

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

which is a strictly decreasing function for all $x \in R_+$. Denote the unique inverse function of W^+ by $W^{\leftarrow}$, i.e.

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

Then the desired result, called the *series representation*, is that

$$z_I(t) = \sum_{i=1}^{I} W^{\leftarrow}(b_i/T) I(u_i \leq t), \quad \text{for} \quad 0 \leq t \leq T, \tag{3.4}$$

where $u_i \overset{i.i.d.}{\sim} U(0,T)$, is independent of $b_1 < ... < b_i < ...$ which are the arrival times of a Poisson process with intensity 1. The result is that the process $z_I(t)$ converges uniformly to $z(t)$ as $I \to \infty$. This method also allows us to simulate the increment of the Lévy process. In particular

$$z_I(\Delta) = \sum_{i=1}^{I} W^{\leftarrow}(b_i/\Delta).$$

Clearly the computational speed of these techniques will depend upon the characteristics of the $W^{\leftarrow}$ function — is it expensive to compute $W^{\leftarrow}$ and does $W^{\leftarrow}(x)$ quickly as x increases.

Examples

Compound Poisson process It helps to think about the compound Poisson process example. Let it have intensity ψ and probability density $f(x)$ for the positive jumps, then the Lévy density and tail mass function are $w(x) = \psi f(x)$ and $W^+(x) = \psi\{1 - F(x)\}$, implying

$$W^{\leftarrow}(x) = \begin{cases} F^{-1}\left\{1 - \left(\frac{x}{\psi}\right)\right\}, & x < \psi \\ 0, & x \geq \psi. \end{cases}$$

Hence the inverse only involves computing the quantiles of the jumps. Overall this implies

$$z_I(t) = \sum_{b_i < T\psi}^{I} F^{-1}\left\{1 - \left(\frac{b_i}{\psi T}\right)\right\} I(u_i \leq t).$$

Clearly if $b_I > T\psi$, then there is no approximation by using this series representation. This method has a simple interpretation. If we sample from

$$F^{-1}\left\{1 - \left(\frac{b_i}{\psi T}\right)\right\} \quad \text{until} \quad b_i > T\psi,$$

then an ordered sequence from $f(x)$ of size $Po(\psi T)$ is produced. The effect of the $I(u_i \leq t)$ term is to sample randomly from this ordered sequence a random share of the draws. So the infinite series representation samples compound Poisson processes rather effectively.

A simple important example of this is where $w(x) = \nu\alpha \exp(-\alpha x)$ so that $W^+(x) = \nu e^{-\alpha x}$, which has the convenient property that it can be analytically inverted:

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

Hence as soon as $x > \nu$ then $W^{\leftarrow}(x) = 0$, implying

$$z_I(t) \stackrel{\mathcal{L}}{=} -\frac{1}{\alpha}\sum_{b_i < tv}^{I} \log\left(\frac{b_i}{t\nu}\right).$$

This means that there is a non-zero probability that $z(t)$ is exactly zero.

Gamma process For $z(1) \sim \Gamma(\nu, \alpha)$, the Lévy density is

$$w(x) = vx^{-1}\exp(-\alpha x), \quad x \in R_+, \tag{3.5}$$

which goes to infinity at zero. It follows that

$$W^+(x) = \nu\int_x^\infty s^{-1}\exp(-\alpha s)\mathrm{d}s = \nu E_1(x/\alpha), \tag{3.6}$$

where $E_1(x)$ is the exponential integral $\int_x^\infty y^{-1}e^{-y}\mathrm{d}y$ which can be computed accurately and rapidly using polynomial and rational functions (cf. Abramowitz and Stegun (1970, p. 231) and the FN library of netlib). This has to be numerically inverted in order to simulate using (3.4). For small x,

$$E_1(x) \simeq -\log x - 0.57721 + o(x^{1/2}),$$

implying for large values of x

$$E_1^{-1}(x) \simeq \exp\left(-x - 0.57721\right).$$

The implication is that the infinite summation in (3.4) should converge exponentially quickly and so allows easy truncation. A convenient way of approximating $E_1^{-1}(x)$ is to note that

$$\begin{aligned} E_1(x) &= \lim_{\alpha\to 0}\int_x^\infty y^{a-1}e^{-y}\mathrm{d}y \\ &= \lim_{\alpha\to 0}\Gamma(a)\Pr(X_\alpha > x), \quad \text{where} \quad X_\alpha \sim \Gamma(a), \\ &= \lim_{\alpha\to 0}\frac{1}{a}\Pr(X_\alpha > x), \end{aligned}$$

Hence $E_1^{-1}(x)$ can be found via $E_1^{-1}(x) = \lim_{\alpha\to 0} Q_\alpha(1-ax)$, where $Q_\alpha(p)$ denotes the quantile of the $\Gamma(a)$ variable at probability level p.

3.2.4 Simulation via the characteristic function

Add some stuff from Hubalek.

3.3 Empirical estimation and testing of Lévy processes

3.3.1 A likelihood approach

Estimation of GH Lévy processes

Here we will assess how well Lévy processes fit the marginal distribution of financial returns. Their flexibility allows important improvements over conventional Brownian motion models, however they clearly neglect the dynamics of returns. These results will guide us in building empirically reasonable dynamic models in later chapters of this book.

We assume that we observe the log-price of an asset, written at time t as $y^*(t)$, at equally space intervals of time. Let the interval be Δ, then we write returns as

$$y_n = y^*(\Delta n) - y^*((n-1)\Delta).$$

Further, if we choose to define the continuous time clock in such a way that $\Delta = 1$, then $\{y_n\}$ has the same distribution as $y^*(1)$.

We will estimate GH Lévy processes using likelihood based methods. The assumption of the Lévy process means that the $\{y_n\}$ are assumed to be i.i.d. GH, leading to the likelihood function

$$\log f(y_1, ..., y_T; \theta) = \sum_{n=1}^{T} \log f(y_n; \theta),$$

where θ denotes the unknown parameters which index the density of $y^*(1)$, which is given in equation (2.23). Hence

$$\theta = (\nu, \mu, \beta, \delta, \gamma)'.$$

The key parameter in this model is ν. The maximum likelihood (ML) estimator of θ is given by

$$\widehat{\theta} = \arg\max_{\theta} \log f(y_1, ..., y_T; \theta),$$

which has to be determined by numerical optimisation. This Section will carry this through by using the Broyden, Fletcher, Goldfard and Shanno (BFGS) quasi-Newton algorithm made available in the matrix programming language `Ox` by Jurgen A. Doornik. The GH implementation of it is given in `bfgs_gh.ox`. This algorithm completes its optimisation in around 10 seconds for sample sizes of around $3,000$ on a moderately powerful PC in 2001.

One approach to constructing confidence intervals for the parameters is via the asymptotic distribution of the ML estimator. This is based on the Lévy assumption of *i.i.d.* increments.

We know the independence assumption is unrealistic. In a later subsection we will discuss its impact on confidence intervals. For now we stand by the Lévy assumption.

The asymptotic theory for ML estimators means that

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) \xrightarrow{\mathcal{L}} N(0,\mathcal{I}^{-1}), \qquad \text{as} \qquad T\to\infty, \tag{3.7}$$

where $\mathcal{I}$ is the expected information per observation which, under correct specification of the model, is

$$\mathcal{I} = -\mathrm{E}\left(\frac{\partial^2 \log f(y_n;\theta)}{\partial\theta\partial\theta'}\right) = \mathrm{Cov}\left(\frac{\partial \log f(y_n;\theta)}{\partial\theta}\right). \tag{3.8}$$

The expected information is usually replaced by an averaged observed quantity

$$\mathcal{I}_S = -\frac{1}{T}\sum_{n=1}^{T}\left(\frac{\partial^2 \log f(y_n;\theta)}{\partial\theta\partial\theta'}\right) \qquad \text{or} \tag{3.9}$$

$$\mathcal{I}_O = \frac{1}{T}\sum_{n=1}^{T}\left(\frac{\partial \log f(y_n;\theta)}{\partial\theta}\right)\left(\frac{\partial \log f(y_n;\theta)}{\partial\theta}\right)'. \tag{3.10}$$

Again the score $\partial \log f(y_n;\theta)/\partial\theta$ and observed information $-\partial^2 \log f(y_n;\theta)/\partial\theta\partial\theta'$ are found by numerical differentiation.

The above results allow us to construct asymptotically valid t statistics for elements of $\widehat{\theta}-\theta$. In particular a 95 percent confidence interval for ν can be found as

$$\widehat{\nu} \pm 1.96\sqrt{\frac{1}{T}\left(\mathcal{I}^{-1}\right)_{\nu\nu}}, \tag{3.11}$$

where $\left(\mathcal{I}^{-1}\right)_{\nu\nu}$ denotes the diagonal element of $\mathcal{I}^{-1}$ corresponding to ν. Of course such an interval is only asymptotically valid.

Confidence intervals via likelihood ratio statistics

An alternative way of quantifying uncertainty is based on the likelihood ratio statistic. Again suppose our focus is on ν. Define $\omega = (\mu,\beta,\delta,\gamma)'$, so that $\theta = (\nu,\omega')'$, and

$$\widehat{\omega_\nu} = \underset{\omega}{\arg\max} \log f(y_1,...,y_T;\nu,\omega).$$

$\widehat{\omega_\nu}$ is a *constrained ML estimator* of ω, imposing on θ an a priori fixed value of ν. Likelihood theory tells us that if we constrain ν correctly then the *likelihood ratio statistic*

$$2\left\{\log f(y_1,...,y_T;\widehat{\theta}) - \log f(y_1,...,y_T;\nu,\widehat{\omega_\nu})\right\} \xrightarrow{\mathcal{L}} \chi^2_1, \qquad \text{as} \qquad T\to\infty.$$

This implies we can find a 95 percent confidence interval for ν by plotting the *profile likelihood*

$$\log f(y_1,...,y_T;\nu,\widehat{\omega_\nu}) - \log f(y_1,...,y_T;\widehat{\theta}) \qquad \text{against} \qquad \nu$$

and including all values of ν for which this statistic is greater than -0.5×3.84. Although this interval is only asymptotic valid, higher order likelihood theory strongly suggests that this likelihood ratio based interval has better finite sample coverage than the corresponding t-statistic one given in (3.11).

In order to illustrate the above methods we will look in detail at the case of daily exchange rate movements in the Canadian Dollar rate against the US Dollar. The results are given in Table 3.1. The ML estimate of ν is quite negative, while μ and β are close to one. The asymptotic standard errors for β and ν are quite large and suggest both μ and β are not significantly

	MLE of GH parameters					Likelihoods		
	μ	β	γ	δ	ν	GH	$\beta = 0$	$N(\overline{\mu}, \sigma^2)$
ML estimates	-0.0133	0.179	0.419	1.95	-1.62	-626.74	-627.97	-819.32
Outer product	(.00887)	(.113)	(.0530)	(.438)	(.710)			
Second derivative	(.00897)	(.115)	(.0491)	(.399)	(.651)			

Table 3.1: *ML estimates of GH for the Canadian daily exchange rate. Brackets are the asymptotic standard errors computed using different estimates of the expected information — outer product measure and inverse of the negative of the second derivative matrix. GH column denotes the likelihood for the unrestricted model. $\beta = 0$ imposes symmetry.*

different from zero. Interesting the standard errors based on the outer product measure (3.10) and the second derivative matrix (3.9) are very similar indeed.

The Table also gives the likelihood when β is constrained to be zero. The likelihood drops by around one, which again suggests β can be set to zero. Finally the Table shows the GH model improves upon the Gaussian likelihood fit by around 193, which is a very large improvement.

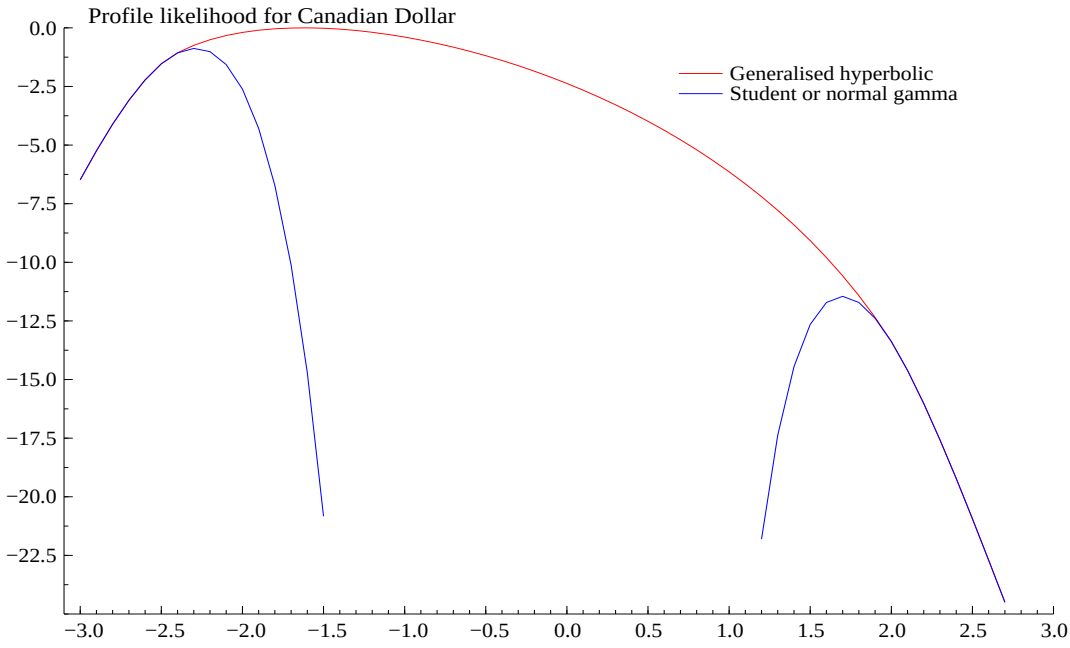

Figure 3.2: *Profile likelihood for ν in the Canadian Dollar example. Three models are considered: GH and the special cases of the student t ($\nu < 0$) and the normal gamma ($\nu > 0$). Code:* `fit_gh.ox`.

Figure 3.2 draws the profile likelihood function for ν for the Canadian Dollar example. This gives a similar result to the t statistics given in Table 3.1 with ranges of approximately -2.5 to -0.2 supported by the data under the assumption of the GH Lévy model. The Figure also shows the profile likelihoods for the normal gamma and Student t special cases of the GH model. Recall in the normal gamma model δ is set to zero, while in the student t case $\gamma = 0$. Of course the likelihoods for these models cannot exceed that of the GH model, but this plot shows how far these models fall behind the GH model. We can see that for very negative ν the likelihood for the GH model is the same as that for the Student t model for the ML of γ turns out to be zero. The same effect can be seen for large values of ν for then the ML of δ is zero. The Figure

shows that the Student t model performs quite well, but the normal gamma process has some very significant difficulties.

Later we will repeat the empirical analysis of many other financial time series using the *GH* model. For now we go on a slight detour which focuses on some numerical issues.

3.3.2 Model misspecification: robust standard errors

Lévy models have i.i.d. increments. Although this allows a flexible framework for modelling the distribution of these increments, we know that financial returns exhibit volatility clustering. In later Chapters of this book we will develop parametric models which will deal with this feature, but for now our Lévy models are misspecified.

Even though our models are incorrect, estimation by ML methods makes sense. We will now be using the *GH* model to model the marginal distribution of the increments and the ML method given above will deliver consistent and asymptotically normally distributed estimators of the parameters. In particular the theory of estimating equations implies

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) \xrightarrow{\mathcal{L}} N\left[0, \mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}\right], \qquad \text{as} \qquad T\to\infty. \tag{3.12}$$

where

$$\mathcal{J} = \lim_{T\to\infty}\frac{1}{T}\mathrm{Cov}\left(\frac{\partial \log f(y;\theta)}{\partial\theta}\right), \qquad \text{and} \qquad \mathcal{I} = -\mathrm{E}\left(\frac{\partial^2 \log f(y_n;\theta)}{\partial\theta\partial\theta'}\right)$$

Under an i.i.d. assumption we know this simplifies to (3.7) due to the information equality (3.8) that $\mathcal{I}=\mathcal{J}$, but this is not true if the data is serially dependent.

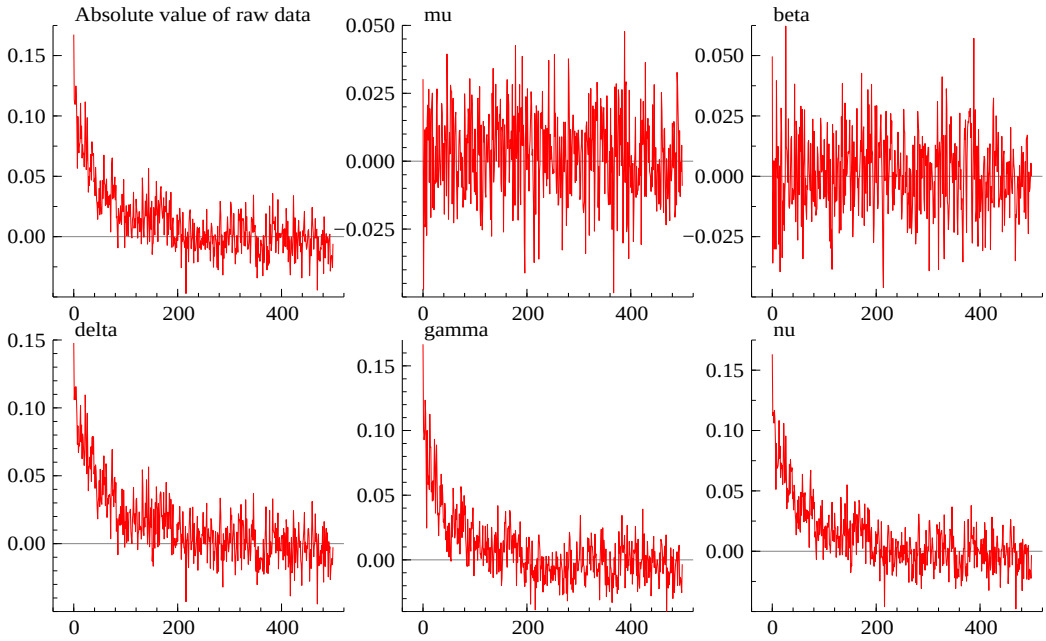

Figure 3.3: *Autocorrelation functions for the Canadian data drawn against lag length. Top left is of* $|y_n|$*, other graphs are for the five elements of* $\partial \log f(y_n;\theta)/\partial\theta$*. Code:* `em_gh.ox`.

In order to make (3.12) practical we need to be able to estimate the element of the *sandwich* $\mathcal{I}^{-1}\mathcal{J}\mathcal{I}^{-1}$ in order to construct robust standard errors for the ML estimator. Now the empirical

average

$$\mathcal{I}_S = -\frac{1}{T}\sum_{n=1}^{T}\left(\frac{\partial^2 \log f(y_n;\theta)}{\partial\theta\partial\theta'}\right),$$

will consistently estimate $\mathcal{I}$ so long as the process is ergodic. However, unless the returns are independent the outer product measure (3.10) $\mathcal{I}_O$ will not correctly estimate $\mathcal{J}$ for the individual scores per observations

$$\frac{\partial \log f(y;\theta)}{\partial\theta} = \sum_{n=1}^{T} s_n, \qquad \text{where} \qquad s_n = \frac{\partial \log f(y_n;\theta)}{\partial\theta},$$

will be dependent through time. To illustrate this we have drawn in Figure 3.3 the autocorrelations corresponding to the Canadian Dollar case (evaluated at $\widehat{\theta}$). The Figure shows the scores for ν, δ and γ have acfs which are close to that of $|y_n|$. The scores for μ and β are much less dependent.

The task of estimating $\mathcal{J}$ is familiar in statistics, for $\mathcal{J}$ is just the zero frequency of the spectral matrix of the vector s_n process

$$\mathcal{J} = \mathrm{Cov}\,(s_n) + \sum_{s=1}^{\infty}\left\{\mathrm{Cov}\,(s_n, s_{n-s}) + \mathrm{Cov}(s_{n-s}, s_n)\right\}.$$

There is a vast literature on this topic both in statistic and econometrics. Here we use the

	MLE of GH parameters				
	μ	β	δ	γ	ν
ML estimates	-0.0133	0.179	0.419	1.95	-1.62
Correct model	(.00887)	(.113)	(.0530)	(.438)	(.710)
Robust: m=250	(.0134)	(.143)	(.0718)	(.394)	(.741)
Robust: m=500	(.0164)	(.172)	(.0768)	(.399)	(.769)

Table 3.2: *Robust s.e.. ML estimates of GH for the Canadian Dollar. Bracketed are the standard outer product information based s.e. together with the robust asymptotic s.e. computed using different values of m. Code:* `fit_gh.ox`.

Newey-West estimator

$$\mathcal{J}_O = \frac{1}{T}\sum_{n=s+1}^{T} s_n s_n' + \sum_{s=1}^{m} K(j;m)\left\{\frac{1}{T}\sum_{n=s+1}^{T} s_n s_{n-s}' + \frac{1}{T}\sum_{n=s+1}^{T} s_{n-s} s_n'\right\},$$

where $K(j;m)$ denotes a non-negative Bartlett smoothing window

$$K(j;m) = \begin{cases} 1 - \left|\frac{j}{m+1}\right|, & \left|\frac{j}{m+1}\right| \le 1, \\ 0, & \left|\frac{j}{m+1}\right| > 1, \end{cases}$$

while m is called the lag truncation parameter. It can be shown that this is enough to guarantee that $\mathcal{J}_O$ is positive semi-definite.

Table 3.2 reports the results from using the above methods to compute the robust standard errors for the Canadian Dollar dataset. The Table reports results for $m = 250$ and 500, although we can see that the standard errors do not vary much with m. Most of the errors have inflated as a result of taking into account the effect of serial dependence in the data, although this is not universally so. Again the normal gamma process looks like a poor fit to this data.

3.3.3 Empirical results

Six daily exchange rate movements

Table 3.3 gives the estimates of the parameters, together with their standard and robust standard errors, for our daily exchange rate return data sets. The corresponding fitted log-density for all

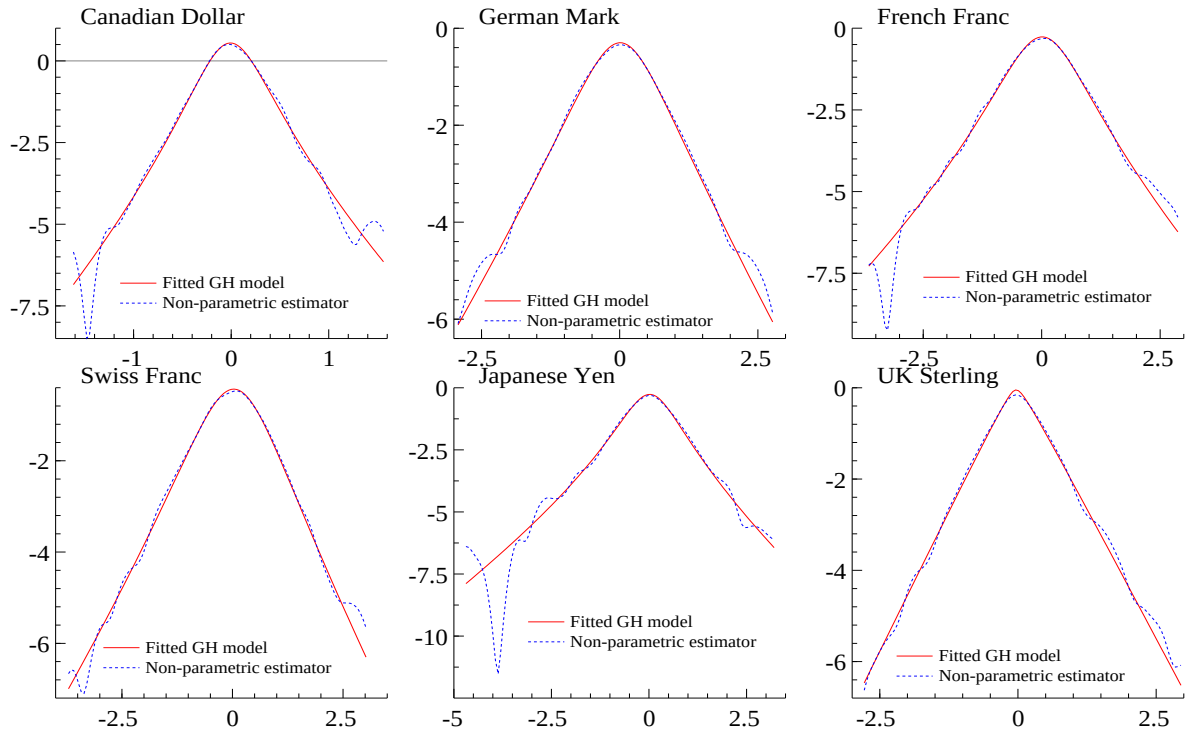

Figure 3.4: *Log of the estimates of the unconditional density of the returns for six exchange rates. Also plotted is the log-density for fitted GH model.*

six series is given in Figure 3.4. This shows pine tree tails in the marginal distributions for all the fitted distributions except for Sterling, which has approximately linear tails. Throughout the fit of the model is very close to the drawn non-parametric estimate of the log-density[1].

There are a number of common features across these results. First all the non-Gaussian models provide dramatic improvements over the fit of the normal likelihood. The ν parameters seems to take values between -2 and 0.5, while neither δ nor γ are close to zero. To reinforce this Figure 3.5 shows the profile likelihood function for each of the datasets. Also drawn are the corresponding profile likelihoods for the skewed Student and normal gamma models (as these are special cases of the *GH* model, naturally these functions are either equal to or below the *GH* curve). The results indicates that the normal gamma model is not really supported by the data. The skewed Student model is typically more supported — primarily as it has fatter tails. Typically ν is around -2, which corresponds to 4 degrees of freedom for the Student distribution. The fit of the distribution is very sensitive to this value. The skewed Student is dominated by *GH* models with $\gamma > 0$. The likelihood function is typically flat for *GH* models with ν between

[1]The non-parametric estimator of the log-density is constructed by using the log of Gaussian kernel estimator coded in Applied Statistics Algorithm AS 176 by Bernard Silverman, which is available at `StatLib` and `NAG` in many statistical software environments such as `Ox`. The bandwidth is chosen to be $1.06\widehat{\sigma}T^{-1/5}$, where T is the sample size and $\widehat{\sigma}$ is the empirical standard deviation of the returns (this is an optimal choice against a mean square error loss for Gaussian data).

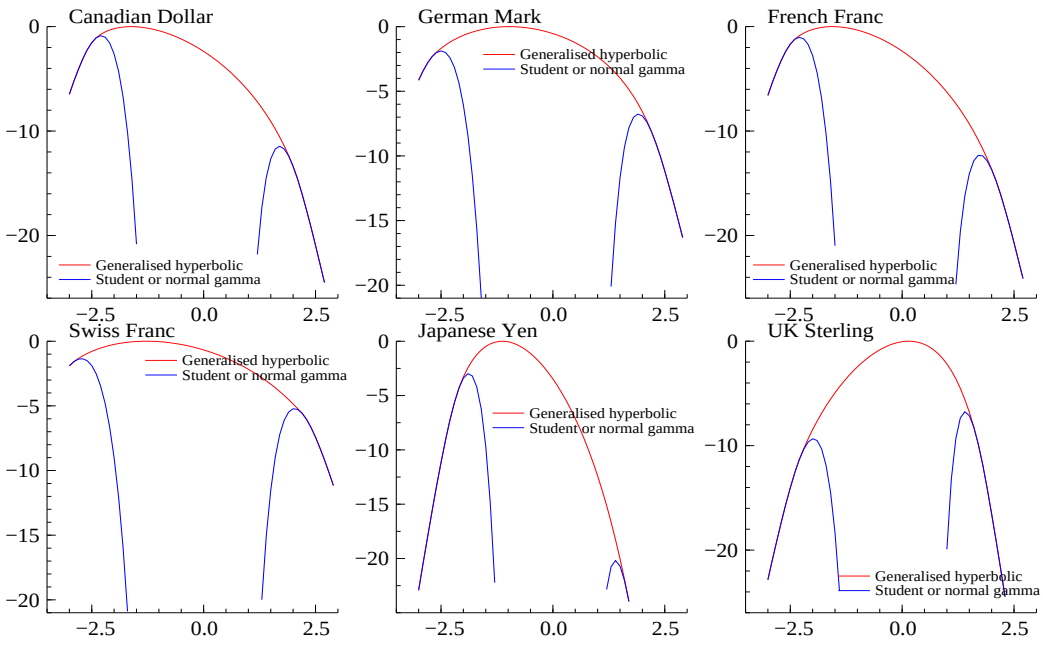

Figure 3.5: *Daily exchange rate data. Profile likelihood (truncated at -25) for the ν parameter of the GH. Also profile likelihood for the skewed Student ($\nu < 0$) and the normal gamma ($\nu > 0$).*

-2 and 2. Overall, however, the values between -2.0 and 0 seem best supported. Finally, the special cases of imposing $\beta = 0$ seems not to harm the fit a great deal for exchange rate data, although there is slight statistical significance in the negative skewness in the UK Sterling, Swiss Franc and Japanese Yen series.

Rate	MLE of GH parameters					Likelihoods				
	μ	β	δ	γ	ν	GH	$\beta = 0$	$\delta = 0$	$\gamma = 0$	N
Canada	-0.013	0.179	0.419	1.95	-1.62	-626.74	-627.97	-638.19	-627.61	-819.32
	(.0164)	(.172)	(.076)	(.399)	(.769)					
DM	0.0243	-0.0647	0.873	1.40	-0.979	-3,903.1	-3,903.8	-3,909.8	-3,905.0	-4,052.2
	(.0330)	(.0631)	(.129)	(.265)	(1.10)					
FF	0.0308	-0.0744	0.930	0.923	-1.55	-3,797.1	-3,798.1	-3,809.3	-3,798.1	-3,988.2
	(.0293)	(.0609)	(.083)	(.208)	(.614)					
SF	0.0731	-0.143	1.08	1.25	-1.27	-4,296.9	-4,300.6	-4,302.1	-4,298.3	-4,428.2
	(.0378)	(.0658)	(.094)	(.214)	(.834)					
JY	0.0447	-0.120	0.800	0.841	-1.12	-4,022.6	-4,027.1	-4,042.8	-4,025.5	-4,310.3
	(.0267)	(.0419)	(.069)	(.144)	(.372)					
Pound	0.0341	-0.0895	0.430	1.89	0.145	-3485.7	-3487.5	-3492.5	-3495.1	-3704.8
	(.0187)	(.0390)	(.090)	(.097)	(.439)					

Table 3.3: *Fit of GH for daily exchange rates. GH denotes unrestricted model. $\beta = 0$ imposes symmetry, $\delta = 0$ normal gamma model, $\gamma = 0$ skewed student. S.E.s ($m = 500$) are in brackets.*

Daily equity indexes

Table 3.4 gives the estimates of our parameters for the daily equity return data. The corresponding profile likelihoods are given in Figure 3.6.

These estimated distributions are more mixed, with values of ν between -1 and 1 being

Index	MLE of GH parameters					Likelihoods				
	μ	β	δ	γ	ν	GH	$\beta = 0$	$\delta = 0$	$\gamma = 0$	N
DAX 30	0.234	-0.113	0.983	0.821	-0.206	-2,735.5	-2,740.6	-2,739.1	-2,740.3	-2,849.4
	(.0915)	(.0377)	(.126)	(.133)	(.836)					
FTSE 100	0.0439	-0.00477	0.734	1.99	1.72	-2,488.7	-2,488.7	-2,488.8	-2,490.4	-2,515.4
	(.0411)	(.0375)	(.328)	(.190)	(1.95)					
S&P 500	0.124	-0.0593	1.05	0.655	-1.03	-2,314.6	-2,315.6	-2,322.3	-2,315.8	-2,444.9
	(.0270)	(.0195)	(.191)	(.236)	(.872)					
Nikkei 500	0.00954	-0.00737	1.08	0.780	-0.654	-2,531.0	-2,531.0	-2,536.1	-2,532.4	-2,638.1
	(.0384)	(.0290)	(.293)	(.322)	(1.75)					

Table 3.4: *Fit of GH for daily equities. GH denotes unrestricted model.* $\beta = 0$ *imposes symmetry,* $\delta = 0$ *normal gamma model,* $\gamma = 0$ *skewed student. S.E.s (*$m = 500$*) are in brackets.*

roughly necessary. Overall again the normal inverse Gaussian usually does pretty well, never fitting really poorly. One conclusion from these fitted models is that there seems very little asymmetry in this data. This is perhaps surprising as this is always an important possibility for equity data. The improvement over the Gaussian fit is picked up very well in the discrepancy between the Gaussian and the *GH* likelihood fits. This holds across all the assets, but is less severe for FTSE — which is not surprising given its normal gamma like behaviour.

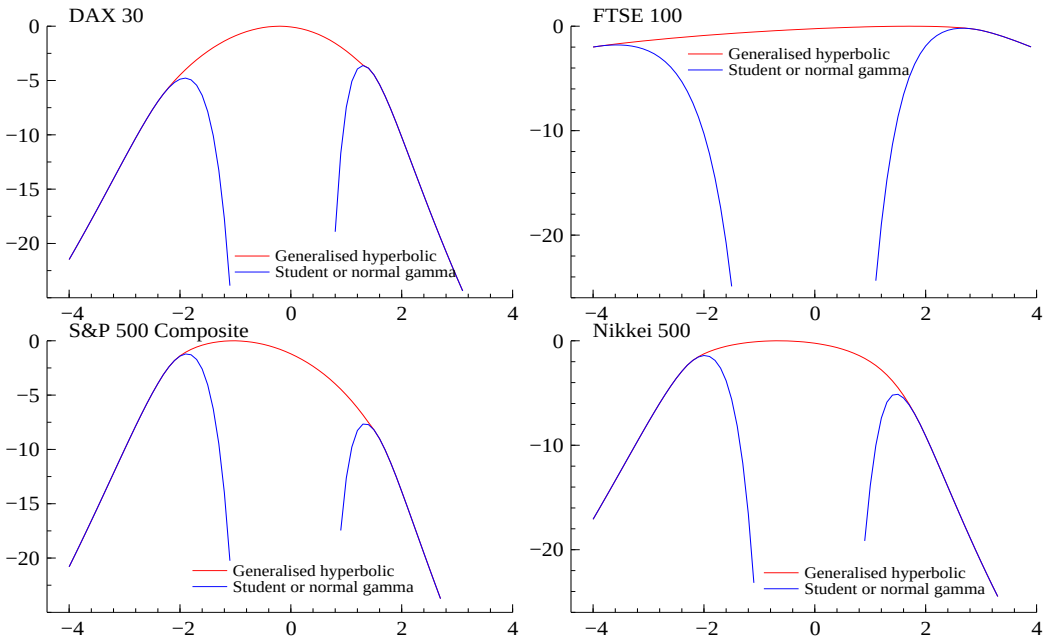

Figure 3.6: *Daily index return data. Profile likelihood for the* ν *in GH model. Also profile for the skewed Student (*$\nu < 0$*) and the normal gamma (*$\nu > 0$*).*

The corresponding fitted log-densities are given in Figure **??** in the first chapter. This showed pine tree tails in the marginal distributions for all the fitted distributions except for Sterling, which has approximately linear tails. Of course this was reflected in the fitted models.

Olsen group's 5 minute data

In Table 3.5 we have reported the results from repeating this experiment but when we have used the 5 minute Olsen exchange rate return data discussed in Chapter 1. Recall there we split the

series into 4 pieces. We follow that here, to try to gain an understanding of the stability of the estimated parameters through time. The results suggest that ν is between -1 and -0.5, which again supports the normal inverse Gaussian distribution. The other parameters in the model are quite stable, although δ falls towards the end of the sample.

	MLE of GH parameters					Likelihoods				
	μ	β	δ	γ	ν	GH	$\beta=0$	$\delta=0$	$\gamma=0$	N
A	.00034	-.203	.0215	15.2	-.485	352,521	352,519	352,137	352,086	317,800
	(.00009)	(.0968)	(.00396)	(.149)	(.0778)					
B	.00038	-.251	.0272	13.7	-.507	322,657	322,652	322,241	322,305	293,460
	(.0001)	(.0907)	(.00526)	(.198)	(.102)					
C	.00020	-.103	.0297	11.4	-.643	319,248	319,247	318,581	318,933	288,941
	(.00009)	(.0781)	(.00333)	(.154)	(.0676)					
D	.00031	-.211	.0295	8.03	-.946	347,149	347,145	346,045	347,011	311,439
	(.00008)	(.0838)	(.00254)	(.176)	(.0559)					

Table 3.5: *ML estimates of GH for the 5 minute exchange rate.* $\beta = 0$ *imposes symmetry,* $\delta = 0$ *implies the normal gamma model,* $\gamma = 0$ *is the skewed student t. Figure in brackets are robust s.e.s computed with* $m = 500$.

The Table suggests that for this exchange rate there is very little non-symmetry in the distribution. Throughout the fit of the normal gamma and Student distributions special cases are significantly worse than the general GH model or indeed the normal inverse Gaussian distribution.

The Olsen dataset allows us to study the empirical effect of time aggregation upon the fit of the these models. This is most conveniently carried out for NIG processes for they are empirically plausible and we know that whatever the value of Δ

$$y_n \sim NIG(\alpha, \beta, \mu\Delta, \delta\Delta).$$

We found that by setting $\gamma = 2.1$ we could produce fits which are broadly consistent with the NIG distribution when the increments were measured over at least 6 hours, with the ML procedure consistently selecting δ (with Δ chosen to represent daily data) to be around 0.96. Throughout β is small, which means $\alpha \simeq \gamma$, while μ is tiny. For finer increments we found that the model was unable to fit the data without changing the NIG's parameters, reflecting the porosity of the independence assumption in the Lévy increments. Figure 3.7 shows the effect of time aggregation on the corresponding shape triangle (2.18) for this data. The points on the triangle are computed based on moving from 6 hour returns up to 36 day returns, plotting a point for every hourly increase. As the interval increases we drift down the shape triangle, starting at $\xi \simeq 0.8$ going down to around $\xi \simeq 0.2$ for increments over a month long.

3.3.4 Olsen scaling rule

The Olsen group in Zurich have reported an apparent scaling result for absolute returns which is of substantial empirical interest. They record the log of the expected absolute returns over an interval of length s,

$$v(t) = \log \mathrm{E}\,|y^*(t)|\,,$$

and find that

$$\frac{\partial v(t)}{\partial \log t} \simeq H,$$

where H is around 0.58 for a freely floating exchange rate, for a wide range of values of t. Of course in the case where $y^*(t)$ is a standard Brownian motion H would be exactly 0.5. Can Lévy

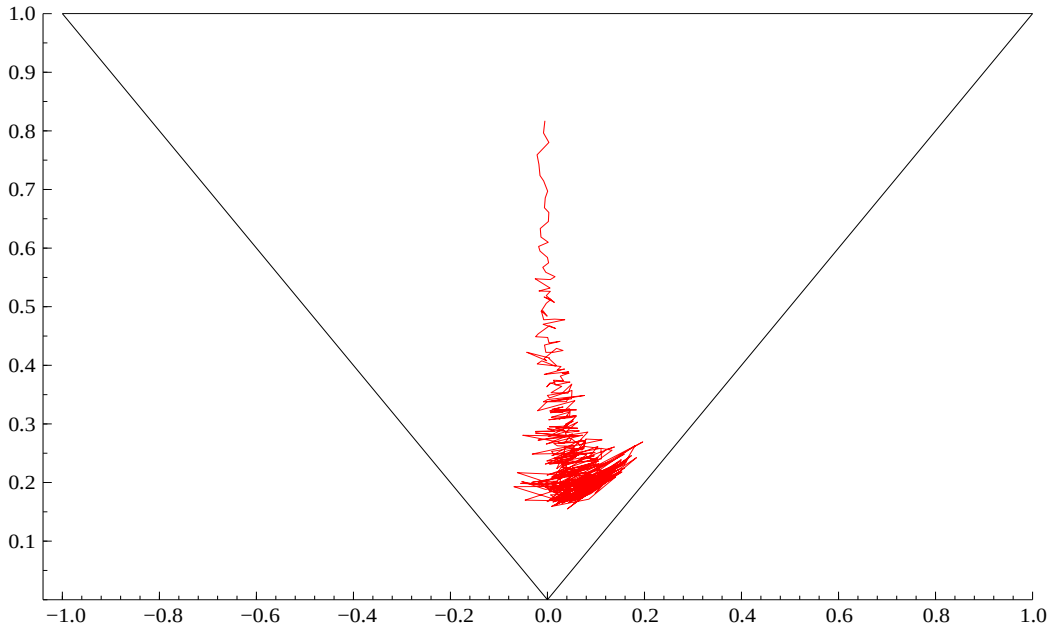

Figure 3.7: *Temporal aggregation of 5 minute returns for exchange rate. Plotted is the estimated NIG distribution, with* γ *fixed at* 2.1, *using a shape triangle.*

processes reproduce this scaling law? This is quite an interesting question for it would mean the apparent scaling law could be explained without the use of dynamics.

For a GH Lévy process with $\mu = \beta = 0$ we have

$$v(t) = \log \sqrt{\frac{2}{\pi}} + \log E\left|z(t)\right|^{1/2},$$

where $z(t)$ is a GIG Lévy process. In the case where $z(t)$ is either a gamma or IG Lévy process we can evaluate this function explicitly.

In the $IG(\delta, \gamma)$ case $z(t) \sim IG(\delta t, \gamma)$ so that

$$\begin{aligned} v(t) &= \log \frac{2}{\pi\gamma} + \log g\left(\delta\gamma t\right), \qquad \text{where} \qquad g(x) = xe^x K_0(x) \\ &= \log \frac{2}{\pi\gamma} + \log g\left(e^{\log t + \log \delta\gamma}\right). \end{aligned}$$

Let us define $\phi(x) = \log g(e^x)$ then, using the asymptotics of $K_0(.)$ we can see that $\phi(x)$ has linear asymptotes both for $x \to -\infty$ and for $x \to \infty$, the slopes of the asymptotes being, respectively, 1 and 1/2. The formula for the slope is

$$\phi'(x) = 1 + e^x - e^x \frac{K_1(e^x)}{K_0(e^x)}, \tag{3.13}$$

which implies

$$\frac{\partial v(t)}{\partial \log t} = \phi'\left\{\log(\delta\gamma) + \log t\right\}.$$

Notice this indicates that the derivative changes very slowly with t when t is around one. Table 3.6 gives some values of $\partial v(t)/\partial \log t$, which suggests this derivative is around 0.55 for the empirically plausible values $\delta = 0.96$ and $\gamma = 2.1$. Importantly the value is always above 0.5,

$\delta\gamma$	$\frac{\partial \log \mathrm{E}\{\lvert z(t)\rvert\}}{\partial \log(t)}$	$\delta\gamma$	$\frac{\partial \log \mathrm{E}\{\lvert z(t)\rvert\}}{\partial \log(t)}$
.0203	.771	1.78	.548
.108	.690	2.31	.539
.235	.646	3.00	.532
.384	.618	3.94	.526
.563	.598	5.25	.520
.778	.582	7.16	.515
1.05	.568	10.1	.511
1.37	.557	15.0	.508

Table 3.6: Slope of $\log \mathrm{E}\{|z(t)|\}$ at t around one.

which implies that the Olsen scaling law can be explained by the non–Gaussian feature of Lévy processes. It is not necessary to have serial dependence in the log-price process to obtain this characteristic.

3.3.5 Fitting multivariate models

General EM framework

The use of BFGS provides a reliable and rather fast way of computing the ML estimator for univariate GH models. However, the direct maximisation of the likelihood in the d dimensional multivariate GH model

$$y_n|\sigma_n^2 \sim N(\mu + \Sigma\beta\sigma_n^2, \sigma_n^2\Sigma), \quad \sigma_n^2 \overset{i.i.d.}{\sim} GIG(\nu, \delta, \gamma), \quad |\Sigma| = 1, \quad n = 1, 2, ..., T,$$

is onerous even though the density of y_n is known to be (2.32). This is due to the presence of the unknown Σ matrix.

Recall the general principle of the EM algorithm. We introduce auxiliary data, $x = (\sigma_1^2, ..., \sigma_T^2)'$ and then work with the augmented likelihood

$$\log f(y, x; \theta) = \log f(y|x; \theta) + \log f(x; \theta).$$

In the EM algorithm we calculate the posterior expectation of the augmented likelihood evaluated at some initial parameter point $\theta^{(0)}$. This is then maximised, giving the algorithm

$$\theta^{(j+1)} = \underset{\theta}{\arg\max} \int \{\log f(y, x; \theta)\} f(x|y; \theta^{(j)}) \mathrm{d}x \qquad j = 1, 2, ...$$

Each step improves $\log f(y; \theta)$ and so iterating it will converge to a maximum.

Two tasks are left in order to implement this method: (i) computing $\sigma_n^2|y_n; \theta^{(j)}$, (ii) maximising the expected likelihood rapidly.

Expectations with respect to $\sigma_n^2|y_n; \theta^{(j)}$

In the GH model we need to work with

$$\begin{aligned} f(\sigma_n^2|y_n) &\propto f(y_n|\sigma_n^2) f(\sigma_n^2) \\ &\propto \frac{1}{(\sigma_n^2)^{d/2}} \exp\left(-\frac{(y_n - \mu)' \Sigma^{-1} (y_n - \mu)}{2\sigma_n^2} + (y_n - \mu)' \beta - \frac{\beta'\Sigma\beta\sigma_n^2}{2}\right) \\ &\quad \times \left(\sigma_n^2\right)^{\nu-1} \exp\left[-\frac{1}{2}\left\{\delta^2 \left(\sigma_n^2\right)^{-1} + \gamma^2 \left(\sigma_n^2\right)\right\}\right]. \end{aligned}$$

This implies

$$\begin{aligned}\sigma_n^2|y_n &\sim GIG\left(\nu-\frac{d}{2},\sqrt{\delta^2+(y_n-\mu)'\Sigma^{-1}(y_n-\mu)},\sqrt{\gamma^2+\beta'\Sigma\beta}\right)\\ &= GIG\left(\nu^p,\delta_n^p,\gamma^p\right).\end{aligned}$$

A useful feature of $X\sim GIG(\nu,\delta,\gamma)$ variable is that

$$\mathrm{E}(X^\alpha)=\frac{(\gamma/\delta)^\nu}{K_\nu(\delta\gamma)}\frac{K_{\alpha+\nu}(\delta\gamma)}{(\gamma/\delta)^{\alpha+\nu}}=\left(\frac{\delta}{\gamma}\right)^\alpha\frac{K_{\alpha+\nu}(\delta\gamma)}{K_\nu(\delta\gamma)},\qquad \delta\gamma>0, \tag{3.14}$$

while

$$\mathrm{E}\log(X)=\left.\frac{\partial\mathrm{E}\left[\exp\left\{\xi\log(X)\right\}\right]}{\partial\xi}\right|_{\xi=0}=\left.\frac{\partial\mathrm{E}(X^\xi)}{\partial\xi}\right|_{\xi=0}, \tag{3.15}$$

which is easy to calculate numerically from (3.14).

Maximisation step

In order to compute $\theta^{(j+1)}$ we have to maximise $\mathrm{E}\log f(y,\sigma^2;\theta)$, where the expectations are fixed and are carried over $\sigma_n^2|y_n;\theta^{(j)}$. This is

$$\mathrm{E}\log f(y,\sigma^2;\theta)=\mathrm{E}\log f(y|\sigma^2;\theta)+\mathrm{E}\log f(\sigma^2;\theta)$$

which equals

$$\begin{aligned}&const+T\nu\log(\gamma/\delta)-T\log K_\nu(\delta\gamma)+T(\nu-1-d/2)\overline{\mathrm{E}(\log\sigma^2)}-\frac{T}{2}\delta^2\overline{\mathrm{E}(\sigma^{-2})}-\frac{T}{2}\gamma^2\overline{\mathrm{E}(\sigma^2)}\\ &-\frac{1}{2}\sum_{n=1}^T(y_n-\mu)'\Sigma^{-1}(y_n-\mu)\,\mathrm{E}\left(\sigma_n^{-2}\right)-\frac{T}{2}\beta'\Sigma\beta\overline{\mathrm{E}(\sigma^2)}+T(\overline{y}-\mu)'\beta,\end{aligned}$$

where

$$\overline{\mathrm{E}(\sigma^{-2})}=T^{-1}\sum_{n=1}^T\mathrm{E}\left(\sigma_n^{-2}\right),\qquad \overline{\mathrm{E}(\sigma^2)}=T^{-1}\sum_{n=1}^T\mathrm{E}\left(\sigma_n^2\right),\qquad \overline{\mathrm{E}(\log\sigma^2)}=T^{-1}\sum_{n=1}^T\mathrm{E}\log\left(\sigma_n^2\right).$$

If we write

$$\overline{y\mathrm{E}(\sigma^{-2})}=T^{-1}\sum_{n=1}^T y_n\mathrm{E}\left(\sigma_n^{-2}\right),$$

then

$$\beta^{(j+1)}=\frac{\Sigma^{-1}\left(\overline{y}-\mu^{(j+1)}\right)}{\overline{\mathrm{E}(\sigma^2)}},\qquad \mu^{(j+1)}=\frac{\overline{y\mathrm{E}(\sigma^2)}-\Sigma\beta^{(j+1)}}{\overline{\mathrm{E}(\sigma^2)}},$$

which implies

$$\mu^{(j+1)}=\left(\overline{\mathrm{E}(\sigma^{-2})}-\frac{1}{\overline{\mathrm{E}(\sigma^2)}}\right)^{-1}\left\{\overline{y\mathrm{E}(\sigma^{-2})}-\frac{\overline{y}}{\overline{\mathrm{E}(\sigma^2)}}\right\}.$$

Concentrating β and μ out of the objective function we just have to maximise with respect to Σ

$$\lambda\left(|\Sigma|-1\right)/2-tr(\Sigma^{-1}S)/2$$

where

$$S=\sum_{n=1}^T\left\{\mathrm{E}\left(\sigma_n^{-2}\right)\left(y_n-\mu^{(j+1)}\right)\left(y_n-\mu^{(j+1)}\right)'\right\}-\frac{T}{\overline{\mathrm{E}(\sigma^2)}}\left(\overline{y}-\mu^{(j+1)}\right)\left(\overline{y}-\mu^{(j+1)}\right)',$$

and λ is an additional Lagrangian parameter to ensure that $|\Sigma| = 1$. This produces[2]

$$\Sigma^{(j+1)} = |S|^{-d} S,$$

which allows us to compute

$$\beta^{(j+1)} = \frac{\left(\Sigma^{(j+1)}\right)^{-1}\left(\overline{y} - \mu^{(j+1)}\right)}{\overline{\mathrm{E}\left(\sigma^2\right)}}.$$

When we plug in the solutions to $\beta^{(j+1)}$, $\Sigma^{(j+1)}$ and $\mu^{(j+1)}$ we can see that all remains is to maximise

$$\nu \log\left(\gamma/\delta\right) - \log K_\nu(\delta\gamma) + (\nu - 1 - d/2)\,\overline{\mathrm{E}\left(\log\sigma^2\right)} - \frac{\delta^2}{2}\overline{\mathrm{E}\left(\sigma^{-2}\right)} - \frac{\gamma^2}{2}\overline{\mathrm{E}\left(\sigma^2\right)},$$

with respect to ν, δ and γ.

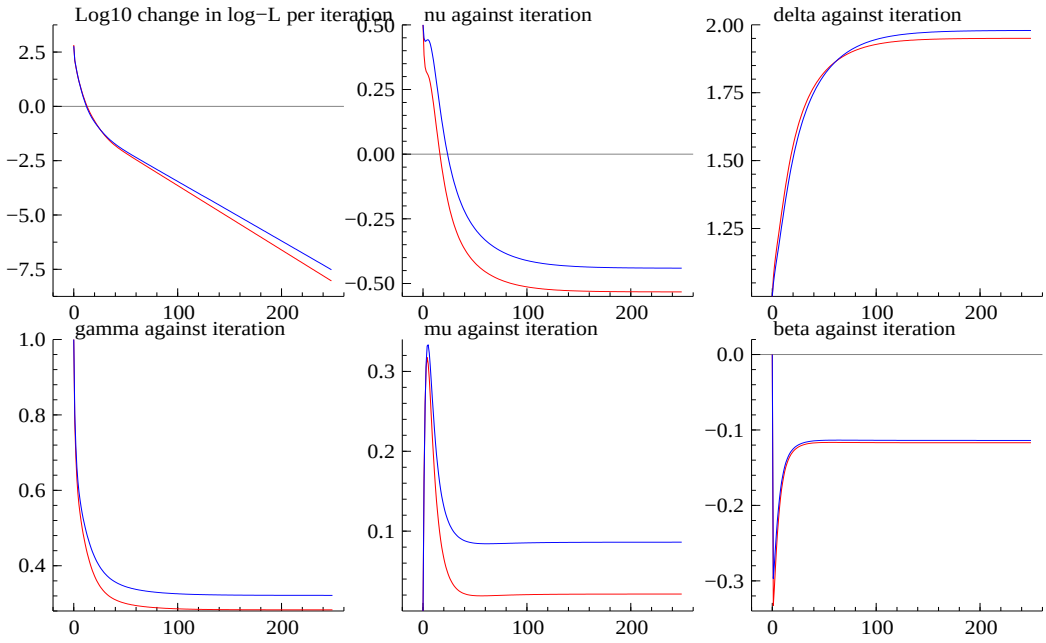

Figure 3.8: *EM algorithm output — based on 2 simulations. Top left:* $\log_{10}$ *transform of change in log-L per iteration. Others:* $\theta^{(j)}$ *against* $j - 1$. *Simulation based* $T = 3,000$ *from a GH model with* $\nu = -0.5$, $\delta = 2.0$, $\gamma = 0.3$, $\mu = 0.0$, $\beta = -0.1$. *Code:* `em_gh.ox`.

Figure 3.8 shows the results of using the EM algorithm on two simulated datasets drawn from a univariate *GH* model with $\nu = -0.5$, $\delta = 2.0$, $\gamma = 0.3$, $\mu = 0.0$, $\beta = -0.1$. In both cases we employed a sample size of $3,000$. The code to carry out the EM for the *GH* problem

[2]Differentiating with respect to Σ and equating to zero gives

$$\lambda\Sigma^{-1} = \Sigma^{-1}S\Sigma^{-1} \quad \text{implying} \quad \Sigma = \lambda^{-1}S.$$

Multiplying the above equation by two and then take determinants of both sides gives

$$|\Sigma| = 1 = \lambda^{-d}|S| \quad \text{implying} \quad \Sigma = |S|^{-d}S.$$

is contained in `em_gh_basic.ox`. It computes around 10 iterations per second on a moderately powerful (2001) notebook. The top left hand graph shows

$$\log_{10}\left\{\log f(y;\theta^{(j+1)}) - \log f(y;\theta^{(j)})\right\} \qquad \text{against} \qquad j.$$

It shows the first few iterations of the algorithm produce enormous improvements in the value of the likelihood, while at larger values of j the improvements tail off. This is a well known feature of the EM algorithm.

Experiments we have conducted suggest that as d increases the speed of the EM algorithm increases, with convergence occurring after a smaller number of iterations. Indeed for 3 or 4 dimensional problems, less than 100 iterations are typically required. This is not surprising for as d increases we have more information about the latent σ_n^2, which is the key to the slowing of EM algorithms. Further, the computational load of each iteration of the EM algorithm increases substantially less quickly than linearly with d.

Empirical results

We start by fitting some bivariate models. The first example of this is a fit of the German DM and French Franc against the US Dollar which is reported in Table 3.7 and Figure 3.9. The Table shows the expected dramatic improvement in fit associated with these multivariate

	MLE of GH parameters							Likelihoods				
	μ	β	δ	γ	ν	Σ		GH	$\beta=0$	$\delta=0$	$\gamma=0$	N
DM	.0221	-.0310	.365	1.03	-1.06	3.83	3.59	-2,865.2	-2,867.6	-2,933.9	-2,869.9	-3,913.7
FF	.0241	-.0143				3.59	3.62	*4,835*	*4,834*	*4,785*	*4,834*	*4,127*

Table 3.7: *ML estimation of bivariate GH model for DM and FF against US Dollar. Figures in italics are improvement in the log-likelihood compared to fit of the two univariate models. Code:* `em_gh_mult.ox`.

models, for the DM and FF are highly related currencies. This is shown up by the estimated Σ matrix. Again ν is estimated to be negative, while the fit of the GH model is very close to the bivariate skewed Student t in this case. The normal gamma model is quite a lot poorer in this multivariate setting. The result in the italics gives the likelihood for the multivariate model minus the sum of the likelihoods for the DM and FF univariate models. So the number for the normal case shows an improvement in the likelihood of $4,127$. Although this is very substantial, the improvements for the other models are much higher. Hence the gains in using GH models is even higher in the multivariate case than one might have expected from the univariate analysis.

Figure 3.9 shows the fitted bivariate normal and GH densities for the DM and FF returns. The graphs have been drawn to show the densities in places where the log-density does not drop 12 from the mode. This gives an impression of the plausible scatter of points from this variable. The bivariate normal density is tightly packed, while the GH model gives a wider range of possible points while the tails of the log-density appear linear or slower in each direction.

Table 3.8 gives the results for all the bivariate relationships which involve the DM. Broadly similar conclusions follow from the above, except the degree of dependence between the currencies is smaller in these other cases. Interestingly the UK Sterling is negatively related to the DM returns. Throughout the table the estimated values of ν ranges between about -0.5 and -1.5. This is an important common theme, again suggesting evidence against the use of normal gamma models.

Table 3.9 gives the GH fit to all six exchange rate return series. This high dimensional model has a low value of ν, which suggests the fit is very close to a skewed multivariate Student t distribution. The skewness parameters are important. The NIG fits worse than the skewed

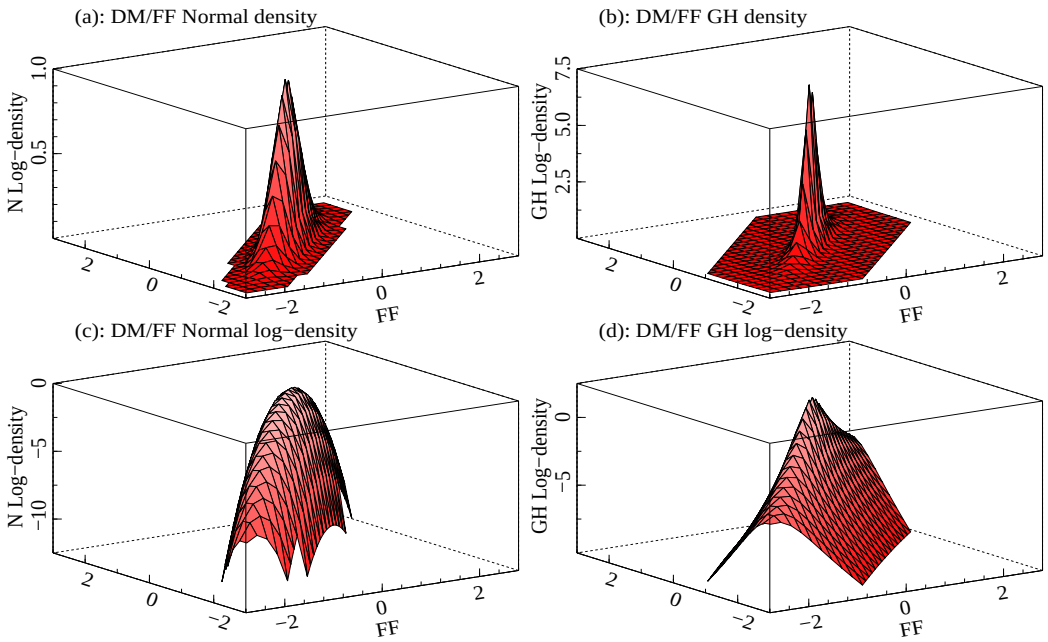

Figure 3.9: *Fit of bivariate Gaussian and GH models for the DM and FF against the Dollar. (a) ML fit of bivariate Gaussian density, (c) gives the log of this density. (b) ML fit of bivariate GH density, (d) gives the log of this density. Code:* `em_gh_mult.ox`.

Student t but is much better than the normal gamma model. All these models are enormous improvements over the multivariate normal fit to the data.

Table 3.10 gives the corresponding result for the four dimensional equity return data. Here the elements of β are all estimated to be negative, indicating common negative skewness. That is the large negative movements have a tendency to occur in all the markets at the same time. In this case the non-symmetries are important, while the normal gamma is again considerably worse than the Student t or the NIG distributions.

3.4 Conclusion

To be added.

3.5 Appendix

3.5.1 Maximum likelihood estimation of GIG models

For known ν and a sample $x_1, ..., x_T$, the likelihood equations for $\widehat{\delta}$, $\widehat{\gamma}$ are (Jørgensen (1982, Section 4.1))

$$\frac{\widehat{\delta}}{\widehat{\gamma}} R_{\nu}\left(\widehat{\delta\gamma}\right) = \frac{1}{T}\sum_{n=1}^{T} x_n \qquad \text{and} \qquad \frac{\widehat{\gamma}}{\widehat{\delta}} R_{-\nu}\left(\widehat{\delta\gamma}\right) = \frac{1}{T}\sum_{n=1}^{T} \frac{1}{x_n}. \tag{3.16}$$

So long as the empirical variance of the $\{x_n\}$ is greater than zero then there always exists a unique solution to these equations. It is quite simple if $|\nu| \leq 1$. Then we can use the following

DM	MLE of GH parameters							Likelihoods				
+	μ	β	δ	γ	ν	Σ		GH	$\beta=0$	$\delta=0$	$\gamma=0$	N
	.0221	-.0310	.365	1.03	-1.06	3.83	3.59	-2,865	-2,867	-2,933	-2,869	-3,913
FF	.0241	-.0143				3.59	3.62	*4,835*	*4,834*	*4,785*	*4,834*	*4,127*
	.0130	-.0434	.655	1.78	-1.46	2.31	.049	-4,518	-4,524	-4,535	-4,520	-4,868
Can	-.0116	.167				.049	.432	*11.8*	*7.4*	*12.6*	*12.2*	*3.2*
	.0429	.0283	.660	1.36	-1.62	2.24	2.27	-4,592	-4,605	-4,615	-4,594	-5,023
SF	.0548	-.129				2.27	2.76	*3,607*	*3,599*	*3,596*	*3,609*	*3,179*
	.0185	.0424	.737	1.37	-.848	1.15	0.60	-7,258	-7,264	-7,281	-7,279	-7,730
JY	.0578	-.175				0.60	1.18	*667*	*666*	*671*	*651*	*632*
	.0315	-.303	.543	1.95	-.391	1.57	-0.98	-5,975	-6,026	-5,993	-5,986	-6,457
BP	.0449	-.361				-0.98	1.24	*1,413*	*1,365*	*1,408*	*1,414*	*1,298*

Table 3.8: *ML estimation of bivariate GH models. Fits DM plus another exchange rate against US Dollar. Figures in brackets are improvement in the log-likelihood compared to fit of the two univariate models. Code:* `em_gh_mult.ox`.

$\delta = .638, \quad \gamma = .586, \quad \nu = -2.22,$

$$\Sigma = \begin{pmatrix} .628 & .0692 & .0754 & .0760 & .0199 & -.115 \\ .0692 & 3.09 & 2.86 & 3.13 & 1.66 & -1.95 \\ .0754 & 2.86 & 2.91 & 2.99 & 1.57 & -1.86 \\ .0760 & 3.13 & 2.99 & 3.82 & 1.85 & -2.08 \\ .0199 & 1.66 & 1.57 & 1.85 & 3.45 & -1.19 \\ -.1157 & -1.95 & -1.86 & -2.08 & -1.19 & 2.54 \end{pmatrix}, \mu = \begin{pmatrix} -.0115 \\ .0307 \\ .0346 \\ .0494 \\ .0605 \\ .0239 \end{pmatrix}, \beta = \begin{pmatrix} .120 \\ .0875 \\ -.0647 \\ -.182 \\ -.148 \\ -.254 \end{pmatrix}.$$

Likelihoods					
GH	$\beta=0$	$\delta=0$	$\gamma=0$	*NIG*	*N*
-9,671	-9,726	-9,791	-9,672	-9,697	-11,635

Table 3.9: *Fit of the multivariate GH model to the six dimensional exchange rate return vector: CD, DM, SF, JY and Sterling.* $\beta = 0$ *imposes symmetry.* $\delta = 0$ *fits the normal gamma model,* $\gamma = 0$ *fits the skewed student t distribution. Code:* `em_gh_mult.ox`.

route. There is a unique $\widehat{\delta\gamma}$ which numerically solves

$$D_\nu\left(\widehat{\delta\gamma}\right) = \left(\frac{1}{T}\sum_{n=1}^{T} x_n\right)\left(\frac{1}{T}\sum_{n=1}^{T}\frac{1}{x_n}\right) \geq 1, \tag{3.17}$$

which implies a value

$$\frac{\widehat{\delta}}{\gamma} = \left(\frac{1}{T}\sum_{n=1}^{T} x_n\right) R_\nu\left(\widehat{\delta\gamma}\right)^{-1}.$$

Jørgensen (1982, Ch. 4) suggests numerically solving (3.17), based on extensive experiments, by finding a value of x such that

$$d(x) = \log\left\{D_\upsilon(e^x) - 1\right\},$$

equals

$$\log\left\{\left(\frac{1}{T}\sum_{n=1}^{T} x_n\right)\left(\frac{1}{T}\sum_{n=1}^{T}\frac{1}{x_n}\right) - 1\right\}.$$

A graph of $d(x)$ against x for various values of ν are given in Jørgensen (1982, Figure 4.1). The expansion

$$D_\upsilon(\omega) = 1 + \frac{1}{\omega} + \frac{-\frac{1}{2}\nu^2 + \frac{1}{8}}{\omega^3} + O(\omega^{-4}),$$

$\delta = 1.47, \quad \gamma = .660, \quad \nu = -1.72,$

$$\Sigma = \begin{pmatrix} 1.54 & .778 & .654 & .145 \\ .778 & 1.18 & .510 & .181 \\ .654 & .510 & .977 & .128 \\ .145 & .181 & .128 & 1.26 \end{pmatrix}, \mu = \begin{pmatrix} .201 \\ .132 \\ .161 \\ .0872 \end{pmatrix}, \beta = \begin{pmatrix} -.0668 \\ -.0105 \\ -.0438 \\ -.0580 \end{pmatrix}.$$

Likelihoods					
GH	$\beta = 0$	$\delta = 0$	$\gamma = 0$	*NIG*	*N*
-9,268	-9,285	-9,293	-9,270	-9,271	-9,694

Table 3.10: *Fit of the GH model to the four dimensional equity return vector: DAX 30, FTSE 100, S&P500, Nikkei 100.* $\beta = 0$ *imposes symmetry,* $\delta = 0$ *implies the normal gamma model,* $\gamma = 0$ *is the skewed student t. Code:* `em_gh_mult.ox`.

suggests the starting value

$$x = -\log\left\{\left(\frac{1}{T}\sum_{n=1}^{T} x_n\right)\left(\frac{1}{T}\sum_{n=1}^{T}\frac{1}{x_n}\right) - 1\right\}.$$

Finally, Jørgensen (1982, Ch. 4) shows that there is always a unique solution to these equations. However, when $|\nu| > 1$ and

$$\mathrm{E}(X)\mathrm{E}(X^{-1}) \geq \frac{|\nu|}{(|\nu| - 1)}$$

we have boundary solutions with

$$\left(\delta^2, \gamma^2\right) = \begin{cases} (0, 2\nu/\mathrm{E}(X)), & \nu > 1 \\ (-2\nu/\mathrm{E}(X^{-1}), 0) & \nu < -1. \end{cases}$$

The first of these two cases corresponds to a gamma distribution, the second to a reciprocal gamma.

3.6 Exercises

3.7 Bibliographic notes

3.7.1 Simulation of Lévy processes

The simulation of Lévy processes has to be carried out with some care. There are extensive results available. Some of the most useful are the infinite series representation developed by Rosinski (2001). The special case of gamma process simulation is discussed by Wolpert and Ickstadt (1999), while some more general discussion is given in Walker and Damien (2000). We should also note the important recent contribution of Asmussen and Rosinski (2000).

3.7.2 Empirical fit of Lévy processes

There is a large literature on studying the fit of various parametric models to the marginal distribution of returns of speculative assets. Most of these papers are not based on a background of a Lévy process and so risked fitting an incoherent (from a continuous time viewpoint) model. An example of this is Praetz (1972) in his work on the student t distribution. A notable exception is Mandelbrot (1963) where he used stable distributions and related this to stable processes.

The likelihood methods we used to fit the models are entirely standard. We have used profile likelihoods to compute measures of uncertainty as these are known to be more reliable

than using the first order Gaussian asymptotic distribution. A discussion of this literature is given in Barndorff-Nielsen and Cox (1994, Section 3.4). The use of profile likelihoods for ν in the generalised hyperbolic is new as was the use of the EM algorithm in this context. Independent and concurrent work on the use of the EM algorithm for this problem was carried out by Protassov (2001). An elegant discussion of the EM algorithm is given in Tanner (1996). The theory of robust standard errors for maximum likelihood estimation is standard in econometrics and statistics. Leading references are White (1982) and White (1994).

Barndorff-Nielsen and Prause (2001) showed that the Olsen scaling law is explained by the NIG Lévy process.

Chapter 4

Time deformation and chronometers

Abstract: This Chapter introduces the general idea of building models as chronometers, which are stochastic processes with non-negative increments. As a generalisation of the notion of a subordinator, a chronometer can be used to change the clock of Brownian motion so that its increments are no just fat tailed but also serially dependent. This is an important feature in financial econometrics and provides a realistic description of the properties of exchange rate and equity market returns. Some of the chronometers are built from Lévy processes, others from non-linear stochastic differential equations. We study some of the properties of the resulting subordinated processes and discuss methods for estimating the models as well as multivariate versions of these models.

4.1 What is this Chapter about?

The sections of this Chapter have the following themes:

- Models built out of general chronometers are analysed. The leading case is where Brownian motion is deformed, postponing a discussion of more general Lévy based models until the end of the Chapter
- Stationary models for τ. The leading cases will be where τ is a
  1. non-Gaussian OU process driven by a Lévy process,
  2. diffusion,
  3. superpositions of stationary processes.
- The probabilistic behaviour of stochastic volatility models.
- Econometrics of stochastic volatility based on low frequency data such as daily returns.
- Multivariate models for asset prices with non-independent increments.
- SV models based on more general Lévy processes than Brownian motion.

Once again we have made the main text as self-contained as possible, leaving a discussion of the corresponding literature to the end of the Chapter. That bibliographical section will focus on historical developments in the subject and papers which develop interesting themes in more depth than we are able to cover.

4.2 General time deformation

4.2.1 Introduction

In financial economics we often build models by time deforming — that is we take a simple stochastic process and construct a more flexible model by replacing its natural time clock with a chronometer (a random process with non-decreasing paths). The most well known example of this is where we change the time clock of a Brownian motion and add a general mean process so that log-prices are modelled as

$$y^*(t) = a^*(t) + w(\tau^*(t)). \tag{4.1}$$

Here w is standard Brownian motion which is assumed independent from a^*, the mean process, and τ^*, the chronometer. A common choice for a^* is to write

$$a^*(t) = \mu t + \beta\tau^*(t), \tag{4.2}$$

which allows the mean process to change with the chronometer. In more general setups w, a^* and τ^* can be jointly dependent processes.

A basic version of the above setup plays a key role in Chapter 2 of this book. There τ^* is taken to be a subordinator, that is a Lévy process, in which case under (4.2) y^* is also a Lévy process and so has independent increments. Empirically we can build rather compelling models by time deformation by a subordinator. An example of this is where τ^* is an generalised inverse Gaussian (GIG) Lévy process (so it is a subordinator), then y^* is a generalised hyperbolic (GH) Lévy process which is rich enough to describe the marginal distribution of returns y_n at a particular time resolution Δ. However, subordinator based models inevitably deliver log-price processes with independent increments implying they can be easily rejected by empirical tests which demonstrate important degrees of serial dependence amongst returns.

Although the construction of Lévy processes via time deformation is stimulating, the general principle of time deformation looks at first sight ad hoc. This turns out not to be true in financial economics. Recall that special semimartingales play a central role in finance and that they can be decomposed as

$$y^*(t) = a^*(t) + m^*(t),$$

where a^* is a predictable process with locally bounded variation and m^* is a local martingale. If we additionally assume that both a^* and m^* have continuous sample paths, then it can be shown that y^* can always be written in the form of (4.1). Hence all financial processes with continuous sample paths fall within the time deformation framework. This Chapter will analyse this fundamental class in some depth.

Mostly our focus will be on chronometers with serially dependent increments — moving us away from Lévy processes. The next section will give results for the general class, while the rest of this Chapter will specialise the framework to chronometers which are defined in the integrated form

$$\tau^*(t) = \int_0^t \tau(u)\mathrm{d}u, \tag{4.3}$$

where τ is a non-negative, càglàd process. This construction implies that these chronometers are integrated processes which are differentiable with

$$\frac{\partial \tau^*(t)}{\partial t} = \tau(t).$$

Hence they do not nest subordinators, for subordinators are pure jump processes. The resulting y^* are called *stochastic volatility models* and have the important property that they have continuous sample paths. Further, the differentiability of τ^* means that y^* can be written in the form of a SDE with

$$\mathrm{d}y^*(t) = a(t)\mathrm{d}t + \sigma(t)\mathrm{d}w(t),$$

where

$$\sigma(t) = \sqrt{\tau(t)}.$$

In this context σ is called the *spot* or *instantaneous volatility*, while τ is the corresponding *spot* (*instantaneous*) *variance*. This setup is by far the most used time deformation model class in all subareas of financial economics.

4.3 Time deformed Brownian motion

4.3.1 Mixture of normals

Recall $\hbar > 0$ is the notation for a fixed interval of time and that the corresponding returns over that time interval are written as

$$y_n = y^*(\hbar n) - y^*(\hbar\,(n-1)-), \qquad n = 1, 2, 3, ... \tag{4.4}$$

The properties of these returns can be derived under a time deformation assumption (4.1). Here will assume τ^* and a^* are jointly independent of w. Then

$$y_n|\tau_n \sim N(a_n, \tau_n),$$

where

$$\tau_n = \tau^*(n\hbar) - \tau^*\left\{(n-1)\,\hbar-\right\} \tag{4.5}$$

and

$$a_n = a^*(n\hbar) - a^*\left\{(n-1)\,\hbar-\right\}.$$

We call τ_n actual variance. By making τ_n time dependent this induces dependence in the returns y_n. When (4.2) holds then

$$a_n = \mu\hbar + \beta\tau_n, \tag{4.6}$$

which is an assumption we often focus on.

In this section we will look at the stochastic properties of the log-price itself $y^*(t)$, recalling that $y^*(0) = 0$.

4.3.2 Cumulant functions of y_1

The conditional cumulant function of y_1 can be calculated just using the properties of the normal distribution. In particular

$$\begin{aligned} \mathrm{K}\{\zeta \ddagger y_1|a_1,\tau_1\} &= \log\left[\mathrm{E}\exp\left\{\zeta \ddagger y_1\right\}|a_1,\tau_1\right] \\ &= \zeta a_1 + \frac{1}{2}\xi^2\tau_1. \end{aligned}$$

implying unconditionally, assuming model (4.6),

$$\mathrm{K}\{\zeta \ddagger y_1\} = \xi\mu t + \mathrm{K}\{\xi\beta + \frac{1}{2}\xi^2 \ddagger \tau_1\} \tag{4.7}$$

which is determined by the cumulant function of τ^*. Important special cases of the above results include that

$$\mathrm{E}\left\{y_1\right\} = \mu t + \beta\mathrm{E}\left\{\tau_1\right\} \qquad \text{and} \qquad \mathrm{Var}\{y_1\} = \beta^2\mathrm{Var}\{\tau_1\} + \mathrm{E}\left\{\tau_1\right\}$$

and

$$\kappa_3\{y_1\} = \beta^3\kappa_3\{\tau_1\} + 3\beta\mathrm{Var}\{\tau_1\},$$

where $\kappa_r\{X\}$ denotes the $r-th$ cumulant of X. Finally,

$$\kappa_4\{y_1\} = 3\mathrm{Var}\{\tau_1\} + 6\beta^2\kappa_3\{\tau_1\} + \beta^4\kappa_4\{\tau_1\}.$$

The same argument delivers the predictive cumulant function

$$\mathrm{K}\{\zeta \ddagger y_1|\mathcal{F}_0\} = \xi\mu t + \mathrm{K}\{\xi\beta + \frac{1}{2}\xi^2 \ddagger \tau_1|\mathcal{F}_0\},$$

and the associated cumulants, e.g.

$$\mathrm{E}\left\{y_1|\mathcal{F}_0\right\} = \mu t + \beta\mathrm{E}\left\{\tau_1|\mathcal{F}_0\right\},$$

and

$$\mathrm{Var}\{y_1|\mathcal{F}_0\} = \beta^2\mathrm{Var}\{\tau_1|\mathcal{F}_0\} + \mathrm{E}\left\{\tau_1|\mathcal{F}_0\right\}.$$

4.4 Non-negative stationary processes

In the introduction we suggested building chronometers as

$$\tau^*(t) = \int_0^t \tau(u)\mathrm{d}u.$$

In this section we will discuss how to build non-negative processes for τ. Broadly this will be based on non-Gaussian Ornstein-Uhlenbeck processes and on discussions. We first focus on the former case, before moving on to the latter class.

4.4.1 OU type processes

First-order autoregression

Independently Slutsky and Yule introduced the linear autoregression into the probability and statistics literature in the early 1920s. Suppose that $\{c_n\}$ is an i.i.d. innovation sequence, then we define a first-order autoregression $\{x_n\}$ as

$$x_{n+1} = \rho x_n + c_n. \tag{4.8}$$

There are two ways of building a parametric model of this type which is strictly stationary when $|\rho| < 1$. We can specify: (i) the distribution of c_n, implying the distribution of x_n (which depends upon ρ), or (ii) the distribution of x_n, implying the distribution of c_n (which also depends upon ρ). Neither route is assumption free, although for the moment we will ignore this, returning to this issue later in this Chapter.

Suppose we have written down a valid autoregression with $|\rho| < 1$ then

$$x_{n+1} = \rho^n x_0 + \sum_{j=0}^{n} \rho^{n-j} c_j = \rho^n x_0 + \sum_{j=0}^{n} \rho^{n-j} \left(z_{j+1} - z_j\right), \tag{4.9}$$

where $z_{n+1} = \sum_{j=0}^{n} c_j$. It is then easy to see that if the first two moments of $\{c_n\}$ exist then $\mathrm{Cor}(x_n, x_{n-s}) = \rho^{|s|}$, which exponentially damps down with lag length s.

Sometimes first-order autoregressions are reparameterised into the *equilibrium correction model* , which writes

$$x_{n+1} - x_n = (\rho - 1)\, x_n + (z_{n+1} - z_n), \qquad \text{where} \qquad z_{n+1} = \sum_{j=0}^{n} c_j. \tag{4.10}$$

Here the increment in the level of the process is regressed on the current level. If the process is stationary the population regression coefficient $\rho - 1$ has to be negative.

OU type processes

Equation (4.9) is insightful as it has a natural continuous time equivalent. The continuous time first-order autoregression, or Ornstein-Uhlenbeck (OU) type process has, for $\lambda, t > 0$,

$$\begin{aligned} x(t) &= e^{-\lambda t} x(0) + \int_0^t e^{-\lambda(t-s)} \mathrm{d}z(s) \\ &= e^{-\lambda t} x(0) + e^{-\lambda t} \int_0^t e^{\lambda s} \mathrm{d}z(s), \end{aligned}$$

where $z(t)$ is a Lévy process. As it drives the OU process we say that $z(t)$ is the *background driving Lévy process* (BDLP). The OU process is often written in the form of an Itô stochastic differential equation (SDE)

$$\mathrm{d}x(t) = -\lambda x(t)\mathrm{d}t + \mathrm{d}z(t),$$

which has many similarities to the ECM briefly discussed in (4.10). Our main concern will be with BDLPs with no negative increment. In such cases a necessary and sufficient condition for the existence of such a process $x(t)$ is that $\mathrm{E}\log(1+z(1)) < \infty$, or equivalently, that $\int_1^\infty \log x W(\mathrm{d}x) < \infty$ where we recall $W(x)$ is the Lévy measure of $z(1)$. The latter case is shown to be sufficient in the Appendix to this Chapter.

The OU process is reasonably familiar when the BDLP is Brownian motion with drift μ and volatility σ. Then the solution to the SDE is

$$x(t)|x(0) \sim N\left\{e^{-\lambda t}x(0) + \frac{\mu}{\lambda}\left(1-e^{-\lambda t}\right), \frac{\tau}{2\lambda}\left(1-e^{-2\lambda t}\right)\right\},$$

implying for regularly spaced data the OU process is exactly a Gaussian first-order autoregression. The unconditional distribution of this process is $x(t) \sim N\left(\mu\lambda^{-1}, \tau\lambda^{-1}2^{-1}\right)$. This shows that λ enters the stationary solution to the OU process.

In our later development the fact that λ enters the stationary solution to the SDE will lead to difficulties and so we prefer to remove this feature now. It is convenient to do this by a simple change of time in the stochastic integral. We will write OU process as

$$x(t) = e^{-\lambda t}x(0) + \int_0^t e^{-\lambda(t-s)}\mathrm{d}z(\lambda s), \tag{4.11}$$

or in the form of a SDE

$$\mathrm{d}x(t) = -\lambda x(t)\mathrm{d}t + \mathrm{d}z(\lambda t). \tag{4.12}$$

This equation will play a crucial role in the rest of this book. In the case where the Lévy process is Brownian motion, $\mathrm{d}z(\lambda t) = \sqrt{\lambda}\mathrm{d}z(t)$, resulting in

$$x(t)|x(0) \sim N\left\{e^{-\lambda t}x(0) + \mu\left(1-e^{-\lambda t}\right), \frac{\tau}{2}\left(1-e^{-2\lambda t}\right)\right\}$$

and that $x(t) \sim N(\mu, \tau/2)$. However for more general Lévy processes the result $\mathrm{d}z(\lambda t) = \lambda \mathrm{d}z(t)$ does not hold (an example of this is where $z(t)$ is an α-stable process then $\mathrm{d}z(\lambda t) = \lambda^{\alpha/2}\mathrm{d}z(t)$). Instead we think of $\mathrm{d}z(\lambda t)$ as speeding up (slowing down) the BDLP for large (small) values of λ compared to $\mathrm{d}z(t)$. In fact, if $x_1(t)$ is an OU process with $\lambda = 1$, then $x(t) = x_1(\lambda t)$ is an OU process with rate λ. Hence λ can be thought of as a time change. In fact if $x_1(t)$ is an OU process with $\lambda = 1$ then

$$x(t) = x_1(\lambda t),$$

is an OU process with rate λ.

Non-negative OU-D processes

Our main concern in this Chapter will be to develop non-negative OU process so that they can be used as the basis of chronometers in (4.3). To emphasise this requirement we will write such SDE based models using the notation for a variance

$$\mathrm{d}\tau(t) = -\lambda\tau(t)\mathrm{d}t + \mathrm{d}z(\lambda t). \tag{4.13}$$

The solution is of the form

$$\begin{aligned} \tau(t) &= e^{-\lambda t}\tau(0) + \int_0^t e^{-\lambda(t-s)}\mathrm{d}z(\lambda s) && (4.14)\\ &= e^{-\lambda t}\tau(0) + e^{-\lambda t}\int_0^{\lambda t} e^{s}\mathrm{d}z(s). && (4.15) \end{aligned}$$

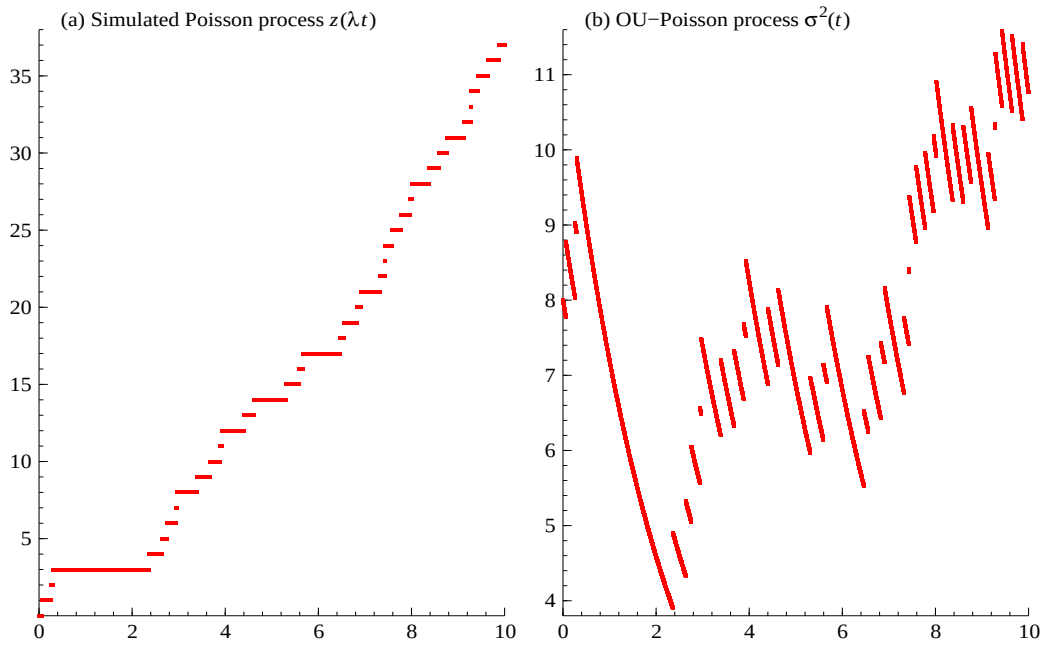

Figure 4.1: *(a) Sample path of Poisson process* $z(\lambda t)$ *with* $\lambda = 0.45$ *and* $\psi = 8$*,(b) Corresponding OU-Poisson process . Code: OU_graphs.ox*

In order to guarantee τ is non-negative we will require that the BDLP has non-negative increments — that is z will be a subordinator. Correspondingly τ moves up entirely by jumps and then tails off exponentially. If $z(t)$ is an infinite activity process then the innovation $e^{-\lambda t}\int_0^{\lambda t} e^s \mathrm{d}z(s)$ must be strictly positive when $t \geq 0$ however small t is. Otherwise there is a positive probability that the innovations from the OU process will be exactly zero.

The very simplest example of a non-negative OU process is where the BDLP is a standard Poisson process with intensity ψ. We call this the *OU-Poisson process*, establishing the notation *OU-D process* which puts the name of the distribution of the BDLP at time one, written generically as D, immediately after the OU title. In the probability literature the OU-Poisson process is often called a *shot noise* and appears commonly in, for example, models of physical storage. In Figure 4.1 we have drawn a sample path from this process taking the intensity as $\psi = 8$ and $\lambda = 0.45$. Here $\tau(t)$ jumps up by one unit, then exponentially decays. This is seen clearly by writing

$$\tau(t) = e^{-\lambda t}\tau(0) + \sum_{j=1}^{N(\lambda t)} e^{-\lambda(t-a_j)},$$

where $N(t)$ is a Poisson process with $\{a_i\}$ being arrivals times so that $a_1 < a_2 < ....$

It is possible to give some rather general properties of OU processes. Here we will focus on two aspects: the autocorrelation function (assuming it exists) and kumulant function of $\tau(t)$.

The autocorrelation function of τ has a simple structure. So long as $\mathrm{Var}(z(1))$ exists,

$$r(u) = \mathrm{Cor}\left\{\tau(t), \tau(t-u)\right\} = \exp\left(-\lambda\left|u\right|\right).$$

It has the important limitation that it can only allow positive serial dependence which damps exponentially.

The kumulant function of $\tau(t)$ can be directly found from the kumulant function of $z(1)$.

Let us define, for $\theta > 0$,

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

These two kumulant functions are related by the fundamental equality

$$\acute{k}(\theta) = \int_0^\infty k(\theta e^{-s})\mathrm{d}s. \tag{4.16}$$

This result is proved in the Appendix. It follows that if we write the cumulants of $\tau(t)$ and $z(1)$ (when they exist) as, respectively, $\acute{\kappa}_m$ and κ_m $(m = 1, 2, ...)$ we have that

$$\kappa_m = m\acute{\kappa}_m, \quad \text{for} \quad m = 1, 2, \ldots .$$

The results on the kumulant function of $\tau(t)$ can be illustrated by working with the OU-Poisson process. Then

$$k(\theta) = -\psi\left(1 - e^{-\theta}\right) \qquad \text{and} \qquad \kappa_m = \psi \qquad m = 1, 2, ...$$

The simple implication of this is that the cumulants of $\tau(t)$ are $\acute{\kappa}_m = m^{-1}\psi$ while the kumulant function is

$$\begin{aligned}
\acute{k}(\theta) &= -\psi\int_0^\infty \left\{1 - \exp\left(-\theta e^{-s}\right)\right\}\mathrm{d}s \\
&= -\psi\int_0^\theta \frac{1 - e^{-t}}{t}\mathrm{d}t \qquad \text{where} \qquad t = \theta e^{-s}, \\
&= -\psi\left\{E_1(\theta) + \log\theta + \gamma\right\},
\end{aligned}$$

where γ is Euler's constant and $E_1(x)$ is the exponential integral.

More OU-D models It is possible to build many OU processes by writing down sensible BDLPs (they just need that $\mathrm{E}\log(1 + |z(1)|) < \infty$, the necessary and sufficient condition for the existence of a valid OU process). Here we build on the BDLPs introduced in Chapter 2: compound processes, gamma processes and inverse Gaussian processes. We could have focused on any member of the *GIG* family for everyone provides a valid basis for an OU process since they satisfy the above condition.

- *OU-compound process.* Suppose the BDLP is a compound process based on a Poisson process $N(t)$ with intensity ψ and i.i.d. positive innovations $\{d_s\}$. Then the process can be represented as

$$\tau(t) = e^{-\lambda t}\tau(0) + \sum_{j=1}^{N(\lambda t)} e^{-\lambda(t - a_j)}d_j.$$

  In the probability literature this type of model is often called *generalised shot noise.*. We saw in Chapter 2 that $k(\theta) = -\psi\left\{1 - \overline{\mathrm{M}}\left(\theta \ddagger d_1\right)\right\}$, where $\overline{\mathrm{M}}\left(\theta \ddagger c_1\right) = \mathrm{E}\exp(-\theta d_1)$. Hence $\kappa_m = \psi\mathrm{E}(d_1^m)$, and so the cumulants of $\tau(t)$ are $\acute{\kappa}_m = m^{-1}\psi\mathrm{E}(d_1^m)$.

- *OU-*Γ. Suppose $z(t)$ is the infinite activity $\Gamma(\nu, \alpha)$ Lévy process. We know that $k(\theta) = \nu\log(1 + \theta/\alpha)$ which implies

$$\begin{aligned}
\acute{k}(\theta) &= \nu\int_0^\infty \log\left(1 + \frac{\theta}{\alpha}e^{-s}\right)\mathrm{d}s \\
&= \nu\int_0^{\theta/\alpha} \frac{1}{t}\log(1 + t)\,\mathrm{d}t \qquad \text{where} \qquad t = \frac{\theta}{\alpha}e^{-s} \\
&= \nu\sum_{j=1}^{\infty}(-1)^j\frac{(\theta/\alpha)^j}{j^2} \qquad \text{for} \qquad 0 \le \theta/\alpha < 1.
\end{aligned}$$

  The cumulants $\kappa_m = m!\nu/\alpha^m$, imply the cumulants of $\tau(t)$ are $\acute{\kappa}_m = \alpha^{-m}(m-1)!\nu$.

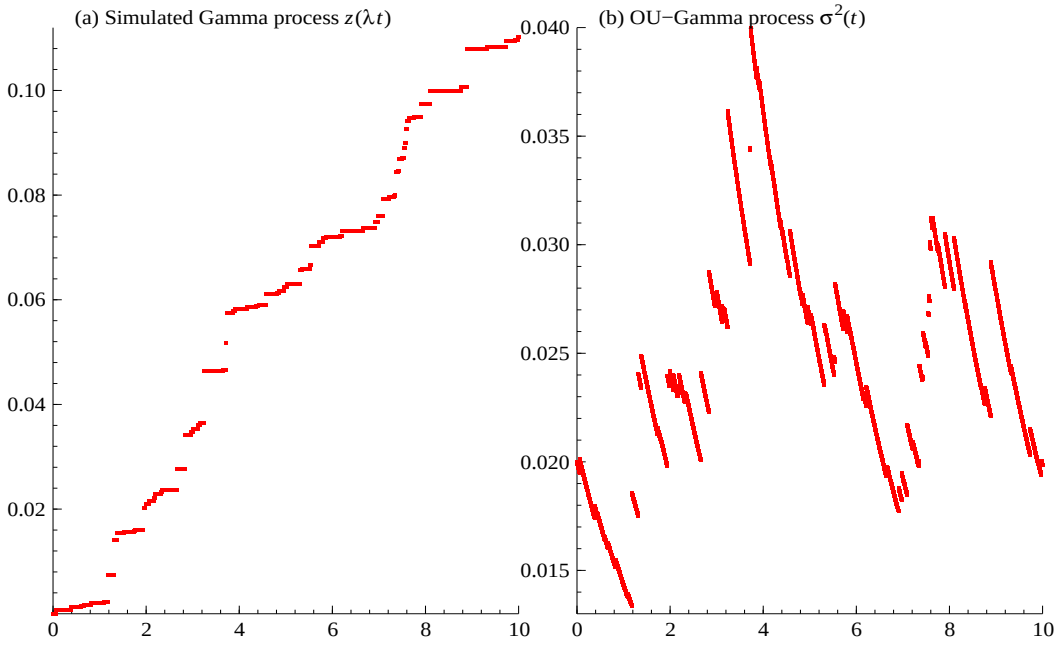

Figure 4.2: *(a) Sample path of* $\Gamma(4, 200)$ *process* $z(\lambda t)$ *with* $\lambda = 0.45$*,(b) Corresponding OU-*Γ *process. Code: OU_graphs.ox*

- *OU-IG process.* The IG based Lévy process can also be used as the BDLP. We know that $k(\theta) = \gamma\delta - \delta\left(\gamma^2 + 2\theta\right)^{1/2}$ which implies

$$\begin{aligned} \acute{k}(\theta) &= \delta\gamma \int_0^\infty \left\{1 - \left(1 + 2\gamma^{-2}\theta e^{-s}\right)^{1/2}\right\} \mathrm{d}s \\ &= \delta\gamma \int_0^{2\theta\gamma^{-2}} \frac{1}{t}\left\{1 - (1+t)^{1/2}\right\} \mathrm{d}t, \qquad t = 2\theta\gamma^{-2}e^{-s}. \end{aligned}$$

  The cumulants

$$\kappa_m = \delta\gamma^{2m-1} 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m-3),$$

  implying the corresponding ones for $\tau(t)$ are

$$\acute{\kappa}_m = \frac{\delta\gamma^{2m-1} 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2m-3)}{m}.$$

Prediction distribution: $\tau(t)|\tau(0)$ Later we will wish to predict future levels of the volatility. Here we will write this as the distribution of $\tau(t)|\tau(0)$. Clearly this has a simple structure for

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

only involves, at time 0, the unknown *innovation*

$$e^{-\lambda t}\int_0^t e^{\lambda s} \mathrm{d}z(\lambda s) = e^{-\lambda t}\int_0^{\lambda t} e^{s} \mathrm{d}z(s). \tag{4.17}$$

Recalling we wrote the first two cumulants of $z(1)$ as κ_1 and κ_2, then

$$\begin{aligned} \mathrm{E}\left\{\tau(t)|\tau(0)\right\} &= e^{-\lambda t}\tau(0) + e^{-\lambda t}\kappa_1 \int_0^{\lambda t} e^s \mathrm{d}s \\ &= e^{-\lambda t}\tau(0) + \kappa_1\left(1 - e^{-\lambda t}\right), \end{aligned}$$

while $\text{Var}\{\tau(t)|\tau(0)\} = 2^{-1}\kappa_2\left(1-e^{-2\lambda t}\right)$. Of course these are reasonably familiar from results on Gaussian OU processes, which are used in financial economics.

More generally we can write the kumulant function of the forecasted volatility as

$$\begin{aligned}\overline{\text{K}}\{\theta \ddagger \tau(t)|\tau(0)\} &= \log \text{E}\left[\exp\{-\theta\tau(t)\}|\tau(0)\right] \\ &= -\theta e^{-\lambda t}\tau(0) + \int_0^{\lambda t} k(e^{-\lambda t}\theta e^s)\text{d}s, \\ &= -\theta e^{-\lambda t}\tau(0) + \int_0^{\lambda t} k(\theta e^{-u})\text{d}u, \qquad e^{-u} = e^s e^{-\lambda t},\end{aligned} \tag{4.18}$$

recalling that $k(\theta) = \log \text{E}\left[\exp\{-\theta z(1)\}\right]$. Of course (4.16) follows from this expression by allowing t to tend to infinity.

Non-negative D-OU processes

An alternative modelling approach So far OU processes have been specified via the model for the BDLP. The stationary distribution of $\tau(t)$ is implied by this model choice, not designed. An alternative is to design directly the marginal distribution of $\tau(t)$ and then work out the implied BDLP. We call these models D-OU processes. From a modelling viewpoint this new setup has advantages for empirical studies often give us a clear impression about good candidates for the marginal distribution of stationary processes, while the innovations driving the stationary process are not observable.

In this section we first ask what class of marginal distributions can yield OU processes? We then relate the unconditional cumulants of $\tau(t)$ to those of the BDLP. Finally we discuss an important class of processes whose marginal law is the GIG distribution.

Self-decomposable distributions Suppose we wish $\tau(t)$ to be an OU process with a marginal distribution D. Is it possible to construct a BDLP to get such a process? The answer is not always. A simple counter example is where $\tau(t)$ is marginally discrete (e.g. binary), then no OU process exists with this marginal law for the exponential damping means the discrete support cannot be satisfied. So what is the required condition?

For D to yield a valid OU process, D has to be *self-decomposable*. This condition is not very familiar in econometrics and finance and so we will spend some time to explain it. It is most easily understood by thinking of our first order autoregression (4.8). If $\{x_n\}$ is stationary with marginal distribution D and associated characteristic function $C(\theta \ddagger x_n) = \text{E}\exp(i\theta x_n)$, then

$$C(\theta \ddagger x_n) = C(\rho\theta \ddagger x_{n-1})C(\theta \ddagger c_n) = C(\rho\theta \ddagger x_n)C(\theta \ddagger c_n).$$

Now as ρ changes so must the distribution of c_n. To reinforce this we rewrite the above equation in the following notation

$$C(\theta) = C(\rho\theta)C_\rho(\theta).$$

It now becomes clear that for D to form a valid OU process a necessary and sufficient condition is that $C(\theta)/C(\rho\theta)$ is a valid characteristic function for all values of $\rho \in (0,1)$. If this is the case D is said to be self-decomposable. It is both necessary and sufficient.

Checking for self-decomposability is often a technically demanding task. Fortunately for us this has been carried out in the probability literature for many of the most familiar models we come across in econometrics and statistics. Here we will only discuss parametric models which are valid, leaving the proof of this assertion to the Appendix. In particular we note that all GIG distributions are self-decomposable and so the GIG-OU model is a valid class of processes.

In the previous section we showed how to compute the cumulants and kumulant function of $\tau(t)$ from the corresponding results for $z(1)$. Here we go the other way around. In particular, differentiating (4.16) we find that

$$\begin{aligned} k(\theta) &= \log \mathrm{E}\left[\exp\left\{-\theta z(1)\right\}\right] \\ &= \theta \acute{k}'(\theta) \end{aligned}$$

(where $\acute{k}'(\theta) = \mathrm{d}\acute{k}(\theta)/\mathrm{d}\theta$), while we recall $\kappa_m = m\acute{\kappa}_m$. A simple example is where $\tau(t) \sim IG(\delta, \gamma)$, then $\acute{k}(\theta) = \delta\gamma - \delta\gamma\left(1 + 2\theta\gamma^{-2}\right)^{1/2}$. As a result $k(\theta) = -\frac{\theta\delta}{\gamma}\left(1 + 2\theta\gamma^{-2}\right)^{-1/2}$. Another important example is where $\acute{k}(\theta) = -\nu\log(1 + \theta\alpha^{-1})$, then $k(\theta) = -\nu\theta\alpha^{-1}\left(1 + \theta\alpha^{-1}\right)^{-1}$.

We have already found in (4.18) that

$$\overline{\mathrm{K}}\left\{\theta \ddagger \tau(t)|\tau(0)\right\} = -\theta e^{-\lambda t}\tau(0) + \int_0^{\lambda t} k(\theta e^{-u})\mathrm{d}u,$$

which, in this context, can be usefully rewritten as

$$\begin{aligned} \overline{\mathrm{K}}\left\{\theta \ddagger \tau(t)|\tau(0)\right\} &= -\theta e^{-\lambda t}\tau(0) + \theta\int_0^{\lambda t} e^{-u}\acute{k}'(\theta e^{-u})\mathrm{d}u \\ &= -\theta e^{-\lambda t}\tau(0) + \int_{\theta e^{-\lambda t}}^{\theta} e^{-u}\acute{k}'(s)\mathrm{d}s, \qquad s = \theta e^{-u}, \\ &= -\theta e^{-\lambda t}\tau(0) + \acute{k}(\theta) - \acute{k}\left(\theta e^{-\lambda t}\right). \end{aligned}$$

This relates the cumulants of the forecast distribution of $\tau(t)$ to the cumulants of the marginal distribution of $\tau(t)$.

Implied Lévy density of the BDLP In D-OU process we specify the Lévy density $w(x)$ of the BDLP $z(1)$ implicitly. These Lévy densities play important roles, both theoretically and in terms of simulation. It is possible to deduce the BDLP's $w(x)$ if we know the Lévy density of D (that is the Lévy density for a Lévy process such that $z(1) \sim D$). We will write the Lévy density corresponding to D as $u(x)$. Then we will prove in the Appendix that

$$w(x) = -u(x) - xu'(x) \qquad \text{and} \qquad W^+(x) = \int_x^\infty w(x)\mathrm{d}x = xu(x).$$

We often write $W^+(x) = \overline{u}(x)$ and note that

$$w(x) = -\overline{u}'(x).$$

This gives us simple expressions for transforming the Lévy densities derived in Chapter 2 for Lévy processes into the Lévy densities of the BDLPs in the D-OU processes. Table 4.1 gives results for the Γ and IG cases, where the results are relatively simple. An interesting feature of the Table is that $w(x)$ is finite when $x = 0$ in the gamma case, which means the corresponding $z(t)$ is a compound Poisson process $\sum_{s=1}^{N(t)} c_s$. Indeed the form of the Lévy density means the Poisson process has rate ν, while the $c_s \overset{i.i.d.}{\sim} \Gamma(1, \alpha)$, implying that the Γ-OU process is particularly easy to simulate.

Simulation of non-negative OU processes

Simulation via the density of the increments We saw in Chapter 2 that we can simulate from Lévy processes by drawing from the increments of the process, which are then summed. This gives the following Euler approximation

	Distributions D	
	$\Gamma(\nu,\alpha)$	$IG(\delta,\gamma)$
$u(x)$	$\nu x^{-1}\exp(-\alpha x)$	$\frac{\delta}{\sqrt{2\pi}}x^{-3/2}\exp\left(-\frac{1}{2}\gamma^2 x\right)$
$w(x)$	$\alpha\nu\exp(-\alpha x)$	$\left(\frac{1}{2}+\frac{1}{2}\gamma^2 x\right)\frac{\delta}{\sqrt{2\pi}}x^{-3/2}\exp\left(-\frac{1}{2}\gamma^2 x\right)$
$W^+(x)$	$\nu\exp(-\alpha x)$	$\frac{\delta}{\sqrt{2\pi}}x^{-1/2}\exp\left(-\frac{1}{2}\gamma^2 x\right)$

Table 4.1: Examples of the Lévy densities for $z(1)$ and $\tau(t)$. They are denoted by $w(x)$ and $u(x)$ respectively.

$$z_\hbar(t)=\sum_{j=1}^{\lfloor t/\hbar\rfloor}u_j,\quad u_j\stackrel{\mathcal{L}}{=}z(\hbar),\quad 0\le\hbar\le t,$$

so that as $\hbar\downarrow 0$ the error in the approximation goes to zero. We can use this approximation to simulate from an OU process (4.15) replacing z by $z_\hbar$, implying the approximate OU process becomes

$$\sigma^2_\hbar(t)=e^{-\lambda t}\sigma^2_\hbar(0)+e^{-\lambda t}\sum_{j=1}^{\lfloor\lambda t/\hbar\rfloor}u_j e^{j\hbar},\quad u_j\stackrel{\mathcal{L}}{=}z(\hbar),\quad 0\le\hbar\le\lambda t. \tag{4.19}$$

Hence for any process where we can simulate from the BDLP we can simulate the corresponding

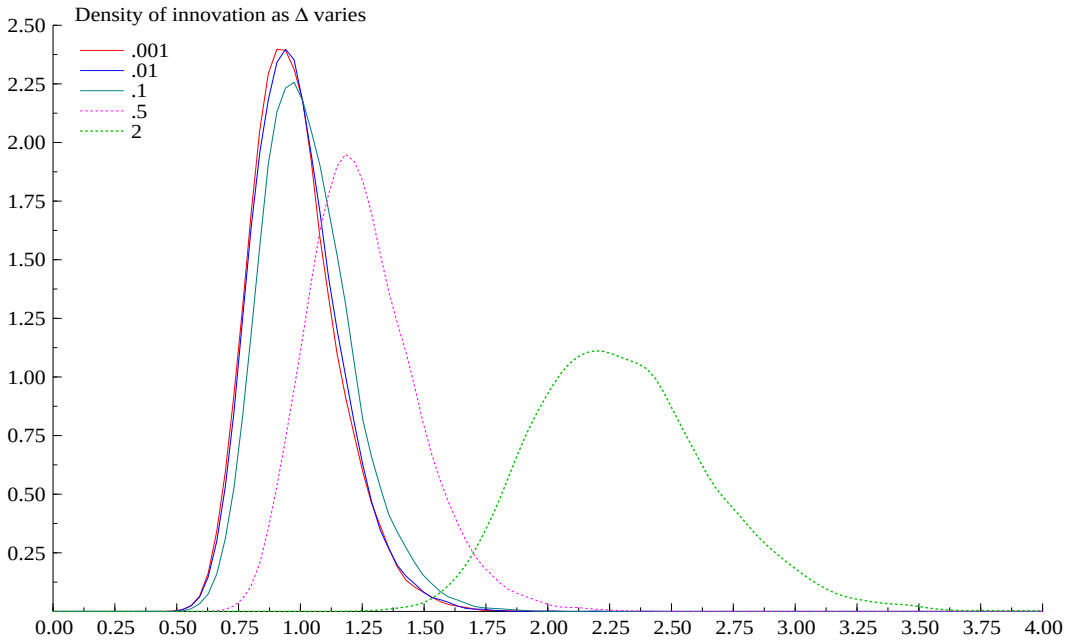

Figure 4.3: *Simulating a OU-IG*$(4,4)$ *process — the approximating the density of* $e^{-\lambda t}\int_0^{\lambda t}e^s\mathrm{d}z(s)$ *with* $t=8$ and $\lambda=0.5$. *Density of* $e^{-\lambda t}\sum_{j=1}^{\lfloor\lambda t/\hbar\rfloor}u_j e^{j\hbar}$ *for various values of* $\hbar$. *Code: OU_graphs.ox*

OU-D process. An example of simulating the approximate innovation $e^{-\lambda t}\sum_{j=1}^{\lfloor\lambda t/\hbar\rfloor}u_j e^{j\hbar}$ is given in Figure 4.3, which displays the resulting density estimated using $160,000$ draws. This is based

on a *OU-IG*$(4,4)$ example with $\lambda = 0.5$, and with $t = 8$. We can see that with $\hbar$ being as low as 0.01 the discretisation error is quite small.

In practice it is perhaps simpler to carry this out recursively, computing

$$\sigma^2_\hbar(\hbar j) = e^{-\lambda\hbar}\sigma^2_\hbar(\delta\,(j-1)) + u_j, \quad \text{where} \quad u_j \overset{\mathcal{L}}{=} z(\lambda\hbar). \tag{4.20}$$

This is the method we employed to simulate the processes reported in Figures 4.1 and 4.2.

Simulation of OU-*TS* processes The tempered stable Lévy process is characterised through its Lévy density which has the simple form

$$w(x) = Ax^{-\kappa-1}\exp(-Bx), \quad x, A, B > 0 \quad \text{and} \quad \kappa \in (0,1).$$

We saw in Chapter 2 that we could use the *Rosinski rejection method* to approximate $z(t)$ by

$$\sum_{i=1}^{I} \min\left\{\left(\frac{At}{b_i\kappa}\right)^{1/\kappa}, B^{-1}e_i v_i^{1/\kappa}\right\}, \tag{4.21}$$

where $I(\cdot)$ is an indicator function, $\{e_i\}$, $\{v_i\}$, $\{b_i\}$, $\{u_i\}$,are independent of one another and over i except for the $\{b_i\}$ process. Here $u_i \overset{i.i.d.}{\sim} U(0,s)$, $v_i \overset{i.i.d.}{\sim} U(0,1)$, the $\{e_i\}$ are exponential

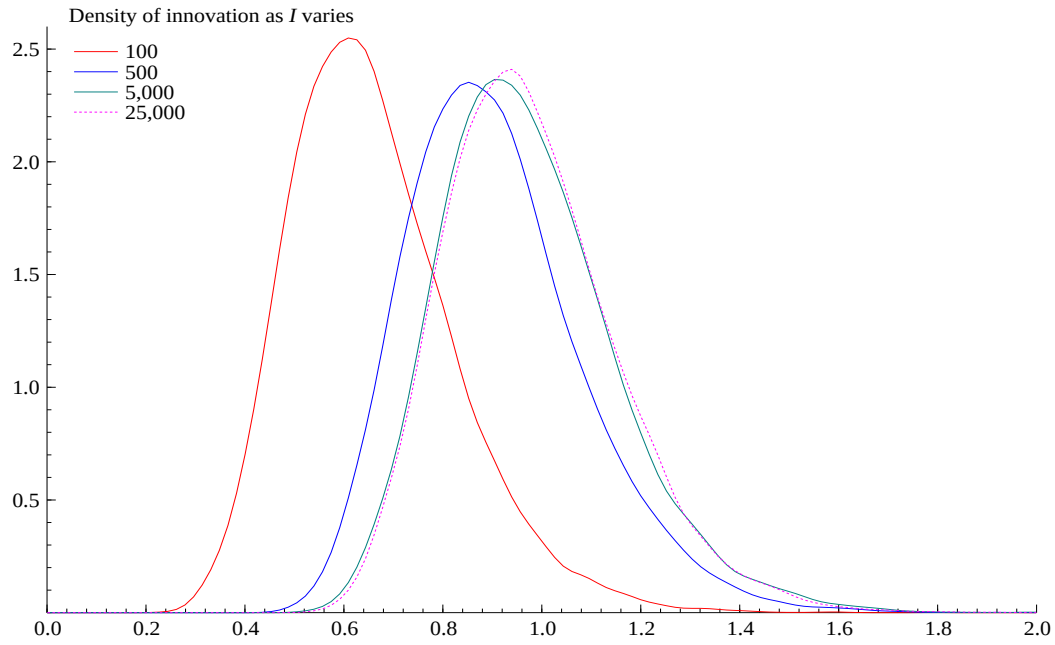

Figure 4.4: *Simulating a OU-IG*$(4,4)$ *process using the TS representation — approximating the density of* $e^{-\lambda t}\int_0^{\lambda t} e^s \mathrm{d}z(s)$. *Density of* $e^{-\lambda t}\sum_{i=1}^{I} \min\left\{\left(\frac{A\lambda t}{b_i\kappa}\right)^{1/\kappa}, B^{-1}e_i v_i^{1/\kappa}\right\}\exp(\lambda t r_i)$*for various values of* I. *Code: OU_graphs.ox*

with mean 1. Further the $b_1 < ... < b_i < ...$ are the arrival times of a Poisson process with intensity 1. These draws from z could be used in (4.19) to simulate the OU process. However, the above approach can be directly extended to allow us to simulate from the $e^{-\lambda t}\int_0^{\lambda t} e^s \mathrm{d}z(s)$ process. In particular Rosinski has showed that process $e^{-\lambda t}\int_0^{\lambda t} e^s \mathrm{d}z(s)$ is approximated in law by

$$e^{-\lambda t}\sum_{i=1}^{I}\min\left\{\left(\frac{A\lambda t}{b_i\kappa}\right)^{1/\kappa},B^{-1}e_iv_i^{1/\kappa}\right\}\exp\left(\lambda t r_i\right), \tag{4.22}$$

where the approximation error goes to zero as $I\to\infty$. Here $r_i \overset{i.i.d.}{\sim} U(0,1)$. Again, each additional term is non-negative and is of $O_p(i^{-1/\kappa})$ which means the rate of convergence of the sum increases as $\kappa\to 0$.

Figure 4.4 shows the simulations in the $IG(\delta,\gamma)$ case where $\kappa=0.5$, $A=(2\pi)^{-1/2}\delta$ and $B=\gamma^2/2$. The picture repeats the experiment reported in Figure 4.3, but now varies I in the sum (4.22). The densities move to the right as I increases, with the result being quite stable by the time I reaches $5,000$. Of course the advantage of this rejection method is that (4.22) potentially represents the whole process as a function of t, not just at a particular value of t.

Tail integral inversion We have just seen that we can directly simulate paths from OU-TS processes. This is a powerful result. However, many interesting Lévy processes cannot be written down as TS. Here we exploit an extension of the infinite series results we used to simulate the increments of Lévy processes in Chapter 2. In Appendix 2 we prove an extension of this result that the stochastic integral

$$\int_0^\lambda f(s)\mathrm{d}z(s)\overset{\mathcal{L}}{=}\sum_{i=1}^{\infty}W^{-1}(a_i/\lambda)f(\lambda r_i), \tag{4.23}$$

where $\{a_i\}$ and $\{r_i\}$ are two independent sequences of random variables with the $r_i's$ independent copies of a uniform random variable r on $[0,1]$ and $a_1<...<a_i<...$ as the arrival times of a Poisson process with intensity 1. Here W^{-1} is the inverse of the tail integral of the BDLP. An application of this result is that the OU's innovation (4.17)

$$e^{-\lambda t}\int_0^{\lambda t}e^s\mathrm{d}z(s)\overset{\mathcal{L}}{=}e^{-\lambda t}\sum_{i=1}^{\infty}W^{-1}(a_i/\lambda t)\exp(\lambda t r_i). \tag{4.24}$$

Hence we are able to produce i.i.d. draws from the innovations or draws from the whole path of the OU process. Two examples of this are given below.

- In the Γ-OU case, Table 4.1 implies

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

  Hence as soon as $x>\nu$ then $W^{-1}(x)=0$, implying that in this case the representation becomes

$$e^{-\lambda t}\int_0^{\lambda t}e^s\mathrm{d}z(s)\overset{\mathcal{L}}{=}-\frac{1}{\alpha}e^{-\lambda t}\sum_{a_i<tv\lambda}\log\left(\frac{a_i}{tv\lambda}\right)\exp(\lambda t r_i),$$

  which is finite as $z(t)$ is a compound Poisson process.

- *OU-*Γ *process.* In Chapter 2 we showed that for $z(1)\sim\Gamma(v,\alpha)$

$$W^+(x)=v\int_x^\infty s^{-1}\exp(-\alpha s)\mathrm{d}s=E_1(x/\alpha),$$

  which is easy to compute and invert numerically. This allows us to simulate from the OU-gamma process without error.

- $IG(\delta, \gamma)$-OU process. Then Table 4.1 records that

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

The focus is often on $W^{-1}(y)$, the value of x which solves $y = W^+(x) > 0$. But, writing $z = \gamma^2 x$, we have to numerically solve

$$\frac{2\pi y^2}{\delta^2 \gamma^2} = r = z^{-1} \exp(-z).$$

For large z we have that $r \simeq z^{-1} - 1$, suggesting z is approximately, for large r, $1/(r+1)$. This implies

$$W^{-1}(y) \simeq \frac{1}{\gamma^2} z \simeq \frac{1}{\gamma^2} \frac{1}{r+1} = \frac{\delta^2}{2\pi y^2 + \gamma^2 \delta^2}.$$

Hence the infinite summation in (4.24) should converge rather slowly and care needs to be taken with the truncation. In practice we can use a numerical method to improve upon this approximation with $z_0 = (r+1)^{-1}$ and then iterating

$$z_{i+1} = z_i + \frac{r - g(z_i)}{g'(z_i)}, \qquad \text{where} \qquad g(z) = z^{-1} \exp(-z),$$

until convergence. In practice we have to truncate the infinite series. We might do this so that $W^{-1}(y) \leq 1 \times 10^{-10}$, which equates to y being less than $\delta 10^5 / \sqrt{2\pi}$.

Small jump approximation

OLE:ADD SOMETHING ABOUT ROSINSKI AND ASMUSSEN HERE.

4.4.2 Non-negative diffusions

Motivation

Non-negative Ornstein-Uhlenbeck type processes are particularly attractive because of their linearity and Markov nature. This allows us to mathematically study and simulate them relatively easily. However, many alternative continuous time, Markov, non-negative processes have been suggested in the literature. Most are based on Brownian motion and have to be non-linear in order to avoid the process being negative. In general these diffusions can be written as

$$\mathrm{d}\tau(t) = \mu\{\tau(t)\}\,\mathrm{d}t + \sigma\{\tau(t)\}\,\mathrm{d}b(t),$$

where $\mu(\cdot)$ and $\sigma(\cdot)$ are general drift and volatility functions. Constraints have to be placed on the relative strengths of the drift and volatility so that τ has a reflecting boundary at zero, otherwise τ can go negative.

WHAT ARE THE CONDITIONS

Here we discuss two cases of these types of process: the Gaussian-OU process for $\log \tau(t)$ and the square root process.

Gaussian-OU process for $\log \tau(t)$

If we model $\tau(t)$ as a Gaussian OU process

$$\mathrm{d}\tau(t) = -\lambda(\tau(t) - \xi)\,\mathrm{d}t + \omega \mathrm{d}b(\lambda t),$$

where $b(t)$ is standard Brownian motion, then there will be a positive probability that $\tau(t)$ will go negative. To avoid this we can model the log of the volatility in this way, producing the process:

$$\mathrm{d}\log\tau(t) = -\lambda\left(\log\tau(t) - \xi_{\log}\right)\mathrm{d}t + \omega_{\log}\mathrm{d}b(\lambda t). \tag{4.25}$$

This has a simple solution

$$\log\tau(t)|\tau(0) \sim N\left\{e^{-\lambda t}\log\tau(0) + \xi_{\log}\left(1-e^{-\lambda t}\right), \frac{\omega_{\log}^2}{2}\left(1-e^{-2\lambda t}\right)\right\},$$

which implies $\tau(t)|\tau(0)$ has a log-normal distribution, while

$$\mathrm{Cor}\left\{\log\tau(t+s), \log\tau(t)\right\} = e^{-\lambda|s|}.$$

The properties of the log-normal distribution imply that

$$\mathrm{E}\left\{\tau(t)\right\} = \exp\left(\xi_{\log} + \frac{\omega_{\log}^2}{2}\right), \qquad \mathrm{Var}\left\{\sigma^2(t)\right\} = \exp\left(2\xi_{\log} + \omega_{\log}^2\right)\left\{\exp\left(\omega_{\log}^2\right) - 1\right\}$$

and

$$\begin{aligned}
\mathrm{Cov}\left\{\tau(t+s), \tau(t)\right\} &= \mathrm{E}\exp\left\{\log\tau(t+s) + \log\tau(t)\right\} - \left[\mathrm{E}\left\{\tau(t)\right\}\right]^2 \\
&= \exp\left(2\xi_{\log}\right)\exp\left[\omega_{\log}^2 + \omega_{\log}^2 e^{-\lambda|s|}\right] - \exp\left(2\xi_{\log} + \omega_{\log}^2\right) \\
&= \exp\left(2\xi_{\log} + \omega_{\log}^2\right)\left[\exp\left\{\omega_{\log}^2 e^{-\lambda|s|}\right\} - 1\right].
\end{aligned}$$

Hence this process is rather tractable. In particular

$$\mathrm{Cor}\left\{\tau(t+s), \tau(t)\right\} = \frac{\exp\left(\omega_{\log}^2\right)\left\{\exp\left(\omega_{\log}^2 e^{-\lambda|s|}\right) - 1\right\}}{\left\{\exp\left(\omega_{\log}^2\right) - 1\right\}}. \tag{4.26}$$

Ito calculus implies that τ follows the solution to the SDE

$$\begin{aligned}
\mathrm{d}\tau(t) &= -\lambda\tau(t)\left(\log\tau(t) - \xi_{\log} - \frac{1}{2}\omega_{\log}^2\right)\mathrm{d}t + \omega_{\log}\tau(t)\mathrm{d}b(\lambda t), \\
&= -\lambda\tau(t)\left(\log\tau(t) - \xi\right)\mathrm{d}t + \omega_{\log}\tau(t)\mathrm{d}b(\lambda t).
\end{aligned}$$

This shows directly the log-normality effect of the model which shifts the mean of the process upwards. The term in front of the noise term forces the noise of the diffusion to be damped when τ becomes small. When τ is large, the process generates large amounts of noise, hence this process has a great deal of heteroskedasticity. This is in marked contrast with the OU processes discussed in the previous section.

Square root process

A popular model in financial economics for a stationary, non-negative process is Feller's *square root process*. It is often called the *Cox-Ingersoll-Ross (CIR) process* after the economists who popularised it in finance. It takes on the form[1]

$$\mathrm{d}\tau(t) = -\lambda\left\{\tau(t) - \xi\right\}\mathrm{d}t + \omega\sqrt{\tau(t)}\mathrm{d}b(\lambda t), \qquad \text{where} \qquad \xi, \lambda, \omega > 0. \tag{4.27}$$

[1]The time change for the Brownian motion is non-standard in the literature, but it is motivated by our discussion of the OU process given above.

Importantly we have to assume that $\xi \geq \omega^2/2$ so there is a reflecting barrier at zero. This model is particularly attractive for the conditional density of $\tau(t)|\tau(0)$ is non-central chi-squared. This follows as it is constructed by squaring a Gaussian OU process. In order to handle this model is is helpful to recall that the kumulant function of the noncentral χ^2-distribution $\chi^2(\phi, \delta)$with ϕ degrees of freedom and noncentrality parameter δ, is given by

$$k(\theta) = -\frac{1}{2}\phi \log(1+2\theta) - \delta\theta(1+2\theta)^{-1}.$$

If we define $\varepsilon(t;\lambda) = \lambda^{-1}(1-e^{-\lambda t})$, then

$$\tau(t)|\tau(0) \stackrel{law}{=} \frac{\omega^2}{4}\lambda\varepsilon(t;\lambda)\acute{\chi}^2,$$

where $\acute{\chi}^2$ is a random variable with law

$$\chi^2\left\{4\xi\omega^{-2}, 4\omega^{-2}\left(e^{\lambda t}-1\right)^{-1}\right\}.$$

Hence, denoting the conditional cumulants of $\tau(t)$ given $\tau(s)$ by $\bar{\tau}_m(t|s)$, $m = 1, 2, ...$, we have that

$$\begin{aligned} \bar{\tau}_1(t|0) &= e^{-\lambda t}\tau(0) + \xi\lambda\varepsilon(t;\lambda) \\ \bar{\tau}_2(t|0) &= \omega^2 e^{-\lambda t}\varepsilon(t;\lambda)\tau(0) + \tfrac{1}{2}\xi\omega^2\lambda^2\varepsilon(t;\lambda)^2 \end{aligned}$$

and for $m = 2, 3, ...$ and $t \downarrow 0$

$$\bar{\tau}_m(t|0) = O(t^{m-1}). \tag{4.28}$$

The fact that $\tau(t)|\tau(0)$ is non-central chi-squared allows us to write down its conditional density explicitly as

$$c\exp\left[-c\left\{\tau(t)+\tau(0)e^{-\lambda t}\right\}\right]\left(\frac{\tau(t)}{\tau(0)e^{-\lambda t}}\right)^{q/2} I_q\left\{2c\sqrt{\tau(t)\tau(0)e^{-\lambda t}}\right\} \tag{4.29}$$

where

$$c = \frac{2}{\omega^2\left(1-e^{-\lambda t}\right)}, \quad \text{and} \quad q = \frac{2\xi}{\omega^2} - 1,$$

and $I_q(\cdot)$ is a modified Bessel function of the first kind. As long as $\lambda > 0$ this process has a stationary distribution which is

$$\tau(t) \sim \Gamma\left(2\xi\omega^{-2}, 2\omega^{-2}\right), \tag{4.30}$$

which has a mean of ξ and variance of $\xi\omega^2/2$. Notice that this distribution does not depend upon λ, a feature which is a result of the use of the time change in (4.27).

The CIR and Γ-OU process have many similiarities. In particular they also share the property that

$$\text{Cor}\left\{\tau(t+s), \tau(t)\right\} = \exp\left(-\lambda\left|s\right|\right),$$

and so the CIR and Γ-OU process are equivalent up to second order, as well as having the same marginal distribution. Both the CIR and OU processes have the interesting property that we can change the persistence of the process by time deformation. In particular suppose τ_1 is a CIR process with law (4.27) but with $\lambda = 1$. We will show in the section on integrated CIR processes that we can reconstitute a general CIR process τ_λ whose law is (4.27) by

$$\tau_\lambda(t) = \tau_1(\lambda t), \tag{4.31}$$

a simple time change of the more limited CIR process. This is a direct verification that the marginal distribution of the CIR process does not depend upon λ.

In terms of differences between the two processes, the Γ-OU process has the advantage that it does not require us to constrain the mean to be bigger than half of the variance in order for the process to be positive, the CIR process has the virtue that the density of $\tau(t)|\tau(0)$ is known analytically rather than by Fourier inversion.

Constant elasticity of variance process

The CIR process allows the volatility of τ to rise with the level of the process, which is sometimes empirically important. A more general class of models which allows this, but still yields some tractability is the constant elasticity of variance (CEV) process. It takes on the form

$$\mathrm{d}\tau(t) = -\lambda\left\{\tau(t) - \xi\right\}\mathrm{d}t + \omega\tau^{\nu}(t)\mathrm{d}b(\lambda t), \qquad \nu \geq 1/2,$$

where ν controls the degrees of responsiveness to the level of the volatility.

Affine models

Simulating non-negative diffusions

One of the virtues of diffusions is that a universal method is available to simulate from them. Suppose

$$\mathrm{d}y = \mu(y)\mathrm{d}t + \sigma(y)\mathrm{d}w,$$

then the Euler discrete approximation to the diffusion is

$$y_\hbar(t+\hbar) - y_\hbar(t) = \mu\left\{y_\hbar(t)\right\}\hbar + \sigma\left\{y_\hbar(t)\right\}\left\{w(t+\hbar) - w(t)\right\}, \tag{4.32}$$

where, of course, $w(t+\hbar) - w(t) \sim N(0,\hbar)$. Then under some weak assumptions,

$$y_\hbar \xrightarrow{L} y, \quad \hbar \downarrow 0,$$

that is the path of the Euler discretisation converges to in law to the path of the desired continuous time process as the time interval becomes small. A detailed discussion of the properties of this Euler scheme is given by Kloeden and Platen (1992).

A major concern for us is that the diffusions we are interested in must obey a non-negativity constraint. This is obtained by using a reflecting barrier at zero. It is clear from (4.32) that $y_\hbar$ does not share this property with probability one. In order to overcome this problem we make an Euler approximation to the $\log y$ transformed process, which is given by Ito calculus as

$$\begin{aligned}
\mathrm{d}\log y &= y^{-1}\mathrm{d}y + \frac{1}{2}y^{-1}\mathrm{d}[y](t) \\
&= y^{-1}\left\{\mu(y) + \frac{1}{2}\sigma^2(y)\right\}\mathrm{d}t + y^{-1}\sigma(y)\mathrm{d}w \\
&= \mu^*(y)\mathrm{d}t + \sigma^*(y)\mathrm{d}w.
\end{aligned}$$

Then when we exponentiate the resulting sample path, this will obey the non-negativity constraint and converge in law to y as $\hbar \downarrow 0$.

4.4.3 Superpositions

Finite superposition

Both OU and diffusion based processes are Markovian. In practice this is often overly restrictive. One tractable approach to expanding this class of processes is by modelling τ as the sum, or *superposition*, of (typically independent) processes each indexed by distinct parameter values. That is

$$\tau(t) = \sum_{j=1}^{m} \tau_j^2(t).$$

In the case of a superposition of OU processes, which we write as an OU_m process, we would take

$$\mathrm{d}\tau_j(t) = -\lambda_j \tau_j(t)\mathrm{d}t + \mathrm{d}z_j(\lambda_j t),$$

where the $\{z_j(t)\}$ are independent (not necessarily identically distributed) BDLPs. By allowing the processes to have different persistence rates, we are able to model processes with quickly decaying short term memory, while allowing for a component of the process to have quite lengthy dependence. This type of flexibility has proved important in empirical volatility modelling. When we work with diffusion based models, similar ideas apply. In particular a CIR_m model is based upon

$$\mathrm{d}\tau_j(t) = -\lambda_j \left\{\tau_j(t) - \xi_j\right\} \mathrm{d}t + \omega_j \sqrt{\tau_j(t)} \mathrm{d}b_j(\lambda_j t),$$

with the Brownian motions being independent processes. Of course, in principle more sophisticated models can be built in this way, allowing dependence across the individual τ_j processes.

In practice it often makes sense to design these superposition models to have particularly attractive features. An example of this is approach is where

$$\tau_j(t) \sim IG(\delta_j, \gamma)\text{-OU}.$$

Then we have that marginally

$$\tau(t) \sim IG(\delta_+, \gamma), \quad \text{where} \quad \delta_+ = \sum_{j=1}^{m} \delta_j. \tag{4.33}$$

However, now the overall process is not an OU process. The second-order dynamics of this process can be expressed as

$$\begin{aligned} \mathrm{Cor}\left\{\tau(t+s), \tau(t)\right\} &= \frac{1}{\mathrm{Var}\left\{\tau(t)\right\}} \sum_{j=1}^{m} \mathrm{Cov}\left\{\tau_j(t+s), \tau_j(t)\right\} \\ &= \sum_{j=1}^{m} \frac{\delta_j}{\delta_+} \exp\left(-\lambda_j |s|\right), \end{aligned} \tag{4.34}$$

which is a weighted sum of exponential damp downs, where the weights δ_j/δ_+ sum to one. For simplicity we will call this type of model an *IG*-OU_m process. Similar types of *D*-OU_m and OU_m-*D* models can be constructed in an obvious manner.

NEIL: REPLACE THIS FIGURE BY THE TRUE ACF OF AN OU$_\{2\}$ PROCESS AND A FIT FROM AN OU$_\{2\}$ AND AN OU$_\{1\}$ PROCESS.

To illustrate the Γ-OU_m process we consider the case where $m = 2$, and $\nu_+ = 3$ and $\alpha = 8.5$. This replicates the marginal distribution for an IG-OU process, a sample path of which was given in Figures ??. However, we take $\nu_1 = 0.8 \times \nu_+$ and $\lambda_1 = 0.2$ and $\lambda_2 = 0.001$. This means the first component has much more variance than the second, but the second component in turn has much more memory than the first. The resulting simulated series of length $8,000$ and corresponding correlogram is given in Figure 4.5. It shows the fast decline in the correlogram for short lags and the very slow decline at larger lags.

Infinite superposition

Theory In the OU case it is possible to produce a model structure which allows m, the number of components to τ, to go off to infinity. This has a number of advantages both from theoretical and modelling viewpoints. It allows for the possibility of long-memory type effects for the τ process as well as a new way of parameterising the model.

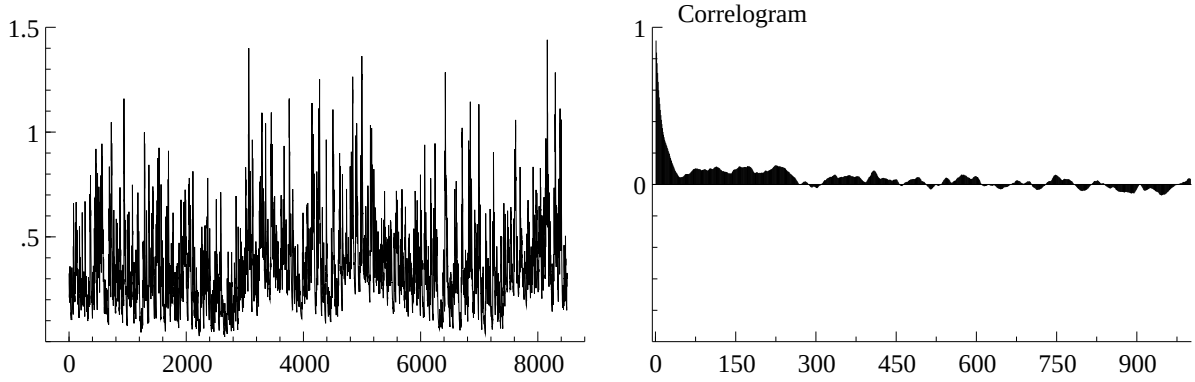

Figure 4.5: $\Gamma-\mathrm{OU}_2$ process with $\nu_+ = 3$, $\alpha = 8.5$. Further, $\nu_1 = 0.8 \times \nu_+$ and $\lambda_1 = 0.2$ and $\lambda_2 = 0.001$. Left: plot of $\tau(n)$ against n. Right: corresponding correlogram.

The rigourous theory of infinite dimensional superpositions is built out of the theory of random fields of OU processes. This work relies on the Lévy fields $z(\mathrm{d}s, \mathrm{d}\xi)$ we introduced in Chapter 2. Recall a field of Lévy processes is obtained as

$$z(t, \mathrm{d}\xi) = \int_0^t \mathrm{d}z(s, \mathrm{d}\xi),$$

where t represents time. Throughout we repeat our discussion of Chapter 2 where we assumed the measure for z is independent over s and $\mathrm{d}\xi$. As a result for a particular choice of ξ this produces a single, univariate Lévy process, and these processes are independent across different ξ.

We can construct a field of OU processes by simply weighting the Lévy increments in a different manner. In particular we can write

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

which is taking the memory parameter of the process as being determined by λ, the index of the field of Lévy processes. Again, a single choice of λ yields an OU process, while pairs of distinct values of λ yield independent OU processes whose marginal distributions (but not its dynamics) will be free of λ.

In the finite superposition model we add together OU processes to produce a richer dynamic. The extension to the infinite dimensional case replaces summation with integration, working with independent OU processes with identically distributed BLDPs. Hence the variability in the processes we integrate will be solely achieved through a random persistence measure λ. The resulting process is

$$\tau(t) = \int_{\mathbf{R}_+} e^{-\lambda t} \int_{-\infty}^{\lambda t} e^s z(\mathrm{d}s, \mathrm{d}\lambda). \tag{4.35}$$

We shall refer to any such process as a *supOU*. Formal calculations with this process implies that we can write it as

$$\mathrm{d}\tau(t) = \int_{\mathbf{R}_+} \left\{ -\lambda \tau(t, \mathrm{d}\lambda)\mathrm{d}t + z(\mathrm{d}t, \mathrm{d}\lambda) \right\}, \tag{4.36}$$

showing that τ is a superposition of OU processes.

Assuming that τ is square integrable, the autocorrelation function r of τ is given by

$$r(u) = \int_0^\infty e^{-u\lambda} \nu(\mathrm{d}\lambda) = \exp\left(\bar{K}\{u \ddagger \lambda\}\right) \tag{4.37}$$

for $u \geq 0$ and where for the last conclusion we interpret λ as a random variable with distribution ν. This is potentially liberating for it means we can model the memory of a linear, non-negative process through the moment generating function of the memory parameter. Note that the choice of this moment generating function has no impact on the marginal distribution of τ. If ν is a distribution made up solely of m atoms then this framework collapses back to the OU_m case discussed in the previous subsection.

Some choices of the moment generating function will deliver models with short term memory, others will deliver long memory models, however all of the models will have correlations which are non-negative at all values of u — an important limitation of this class of model.

Example 9 *Suppose that ν is the gamma law $\Gamma(2\bar{H}, \alpha)$ where $\alpha, \bar{H} > 0$. Then*

$$r(u) = \left(1 + \frac{u}{\alpha}\right)^{-2\bar{H}},$$

which is monotonically declining with u. The process x exhibits second order long range depen-

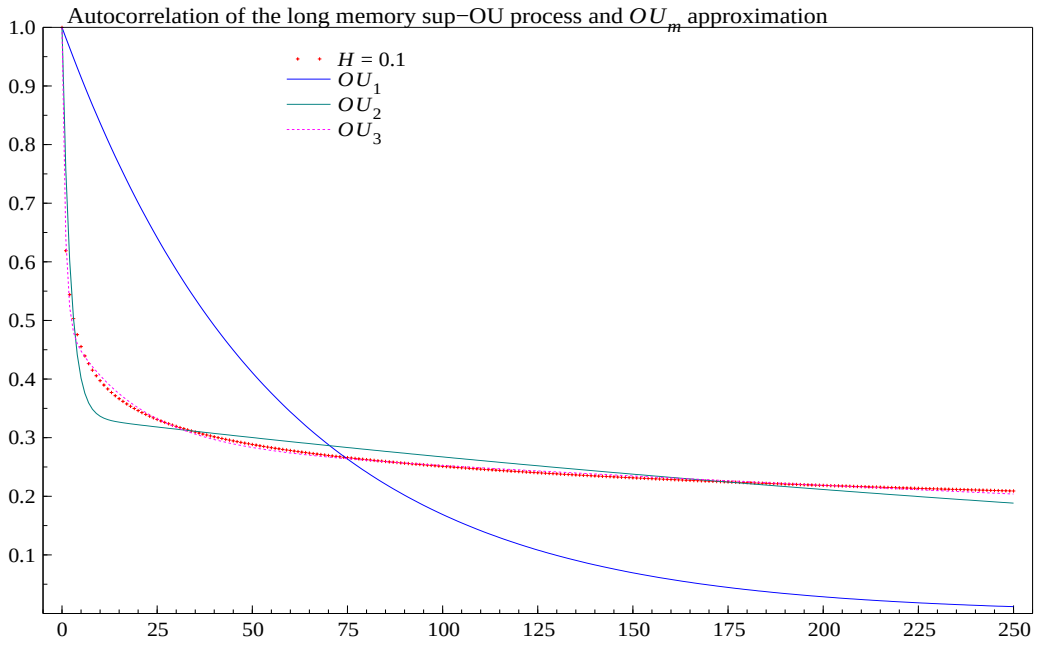

Figure 4.6: Autocorrelation function of sup-OU process with $\overline{H} = 0.1$ and $\alpha = 0.1$. Approximation of this function using the autocorrelation functions of various OU_m processes. Fit is via least squares. Code: `long_memory.ox`.

dence[2] *if $\bar{H} \in \left(0, \frac{1}{2}\right)$, while larger values of $\bar{H}$ give short range dependence. Figure 4.6 shows the autocorrelation function of this type of model in a long memory case with $\bar{H} = 0.1$ and $\alpha = 1$. This has a characteristic initial fast decline followed by a very slow decay in the function.*

Figure 4.6 also displays the non-linear least squares fit of this function by the autocorrelation functions (4.34) of some OU_m processes (which must be short memory processes). In the fitting

[2]Recall a conventional definition of second order long range dependence is that for large u

$$r(u) \propto u^{2d-1}, \quad 0 < d < 0.5.$$

See, for example, the review by Robinson (1994).

exercise we allowed the parameters of the OU_m model to be freely chosen so as to be proxy the autocorrelation of the long-memory process. We can see a simple OU_1 model is a hopeless approximation, while OU_2 gives quite a reasonable fit. By the time we deliver an OU_3 model the fit of this model is roughly on top of the sup-OU process. This shows how well finite superpositions of short memory models can fit long-memory models.

Table 4.2 gives the fitted parameters from the above approximation exercise. The important point is that the OU_m models can fit these autocorrelations by have some components which have small weight (δ_j/δ_+) but very small values of λ. This latter feature allows the autocorrelation function to decline slowly. The components with large weights have large values of λ, which in turn allows the autocorrelation function to initially decay very rapidly.

Model	δ_j/δ_+			λ		
OU_1	1.00			.0177		
OU_2	.337	.662		.00233	.454	
OU_3	.209	.289	.500	.0553	.00140	1.18

Table 4.2: Parameters which approximates the autocorrelation function of a $\Gamma(0.2, 1)$ sup-OU process. The fit is based on a least squares criteria. Code: long_memory.ox

4.4.4 Higher order autoregressive models

By rewriting the formula for an AR(1) process $x_n = xy_{n-1} + z_n$ in equilibrium correction form

$$x_n - x_{n-1} = -(1-\phi)x_{n-1} + z_n$$

one naturally passes to the continuous time analogue of an OU process

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

Such a procedure does not work for higher order AR(p) models with $p > 1$ because of the lack of a proper notion of higher order differentials of a Lévy process. To circumvent this difficulty one may, for instance, in the AR(2) case rewrite

$$x_n = \phi_1 x_{n-1} + \phi_2 x_{n-2} + z_n \tag{4.38}$$

into a bivariate Markov form. This argument is familiar in the statistics literature where it is used to represent higher order models in state space form. We let $c_n = x_n - x_{n-1}$, then

$$\begin{aligned} c_n - c_{n-1} &= x_n - 2x_{n-1} + x_{n-2} \\ &= -(1-\phi_1-\phi_2)x_{n-1} - (1+\phi_2)c_{n-1} + z_n. \end{aligned} \tag{4.39}$$

The bivariate recursive system

$$\begin{aligned} x_n - x_{n-1} &= c_n, \\ c_n - c_{n-1} &= -(1-\phi_1-\phi_2)x_{n-1} - (1+\phi_2)c_{n-1} + z_n, \end{aligned} \tag{4.40}$$

has Markov transition equations but allows x_n to be marginally a second order autoregressive process.

The idea is now, in the light of (4.40), to consider the SDE system

$$\begin{aligned} \mathrm{d}x(t) &= c(t)\mathrm{d}t \\ \mathrm{d}c(t) &= -\{\lambda_1 x(t) + \lambda_2 c(t)\}\,\mathrm{d}t + \mathrm{d}z(t), \end{aligned}$$

with $z(t)$ a Lévy process. According to Brockwell (2001), this bivariate system has a stationary solution if the two zeroes of the characteristic polynomial $z^2+\lambda_1 z+\lambda_2$ both have negative parts and $\mathrm{E}\{|z(1)|^r\}<\infty$ for some $r>0$ [A LESSER RESTRICTION ON $|z(1)|$ WOULD SUFFICE, I SURMISE; WILL THINK ABOUT IT]. The solution for $x(t)$ is then

$$\begin{pmatrix} x(t) \\ c(t) \end{pmatrix} = e^{-\Lambda t}\begin{pmatrix} x(0) \\ c(0) \end{pmatrix} + \int_0^t e^{-\Lambda(t-s)}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \mathrm{d}z(s) \tag{4.41}$$

where

$$\Lambda = \begin{pmatrix} 0 & -1 \\ \lambda_2 & \lambda_1 \end{pmatrix}.$$

Furthermore,

$$\mathrm{C}\{\zeta \ddagger x(t)\} = \int_0^\infty \mathrm{C}\{\zeta(0,1)e^{-\Lambda s}\begin{pmatrix} 0 \\ 1 \end{pmatrix} \ddagger z(1)\}\mathrm{d}s.$$

The above discussion generalises to allow moving average type effects to enter this model structure. Such models are call continuous time autoregressive moving average (CARMA) processes.

4.4.5 General linear models

It is sometimes useful to consider the general case of a linear model, where

$$\tau(t) = \int_{-\infty}^t f(t-s)\mathrm{d}z(s),$$

where f is a deterministic function. Clearly the OU process is recovered from this setup by writing $f(u)=\exp(-\lambda|u|)$. However, the general model structure is itself quite tractable. In particular

$$\mathrm{E}(\tau(t)) = \kappa_1\int_{-\infty}^t f(t-s)\mathrm{d}s \quad \text{and} \quad \mathrm{Var}(\tau(t)) = \kappa_2\int_{-\infty}^t\int_{-\infty}^t f(t-s)f(t-u)\mathrm{d}s\mathrm{d}u.$$

4.5 Integrated non-negative processes

4.5.1 General case under covariance stationarity

In this section we will study the properties of chronometers built by integrating a stationary, non-negative process τ. In particular we will write

$$\tau^*(t) = \int_0^t \tau(u)\mathrm{d}u. \tag{4.42}$$

Leading cases will be where τ is an OU or diffusion process, or superposition of such models. In all such models, by construction of being an integrated process,

$$\frac{\partial\tau^*(t)}{\partial t} = \tau(t). \tag{4.43}$$

It is possible to derive some properties for the above integrated processes. We suppose τ is a covariance stationary process with (when they exist) ξ, ω^2 and r being, respectively, the mean, variance and the autocorrelation function of the process τ. Then

$$\mathrm{E}\{\tau^*(t)\} = \xi t, \quad \text{and} \quad \mathrm{Var}\{\tau^*(t)\} = 2\omega^2 r^{**}(t) \tag{4.44}$$

where

$$r^*(t) = \int_0^t r(u)\mathrm{d}u \quad \text{and} \quad r^{**}(t) = \int_0^t r^*(u)\mathrm{d}u. \tag{4.45}$$

The first result is straightforward, while the second one is to due to the manipulation

$$\begin{aligned}
\mathrm{Var}\{\tau^*(t)\} &= \int_0^t \int_0^t \mathrm{Cov}\{\tau(u), \tau(v)\}\mathrm{d}u\mathrm{d}v \\
&= \omega^2 \int_0^t \int_0^t \mathrm{Cor}\{\tau(u), \tau(v)\}\mathrm{d}u\mathrm{d}v \\
&= \omega^2 \int_0^t \int_0^t r(u-v)\mathrm{d}u\mathrm{d}v \\
&= 2\omega^2 \int_0^t \int_0^v r(u)\mathrm{d}u\mathrm{d}v \\
&= 2\omega^2 \int_0^t r^*(v)\mathrm{d}v \\
&= 2\omega^2 r^{**}(t).
\end{aligned}$$

A plot of $\mathrm{Var}\{\tau^*(t)\}$ against t is called a *variogram*. It will also be helpful to calculate

$$\begin{aligned}
\mathrm{Cov}\{\tau^*(t), \tau^*(t+s)\} &= \omega^2 \int_0^t \int_0^{t+s} r(u-v)\mathrm{d}u\mathrm{d}v \\
&= \omega^2 \int_0^t r^*(t+s-v) - r^*(-v)\mathrm{d}v \\
&= \omega^2 \left[\{r^{**}(t+s) - r^{**}(s)\} - \{-r^{**}(t) - r^{**}(0)\}\right] \\
&= \omega^2 \left[r^{**}(t+s) + r^{**}(t) - r^{**}(s)\right].
\end{aligned}$$

Further we have that if $\tau(u)$ is ergodic then, as $t \to \infty$,

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

Hence a scaled version of the integrated process converges to its expected value ξ as t increases.

Example 10 *Suppose* $r(u) = \exp(-\lambda |u|)$, *implying*

$$r^*(t) = \lambda^{-1}\left(1 - e^{-\lambda t}\right) \tag{4.46}$$

and which means that

$$r^{**}(t) = \lambda^{-2}\left(e^{-\lambda t} - 1 + \lambda t\right). \tag{4.47}$$

Important observations are that $r(s)$ *exponentially damps down to zero, always being less than one. On the other hand* $r^*(s)$ *starts at zero and monotonically rises to* λ^{-1} *as* s *increases. Finally,* $r^{**}(s)$ *again starts at zero but trends upwards, never settling at an asymptote. The implications of these results is that*

$$\mathrm{E}\left\{\tau^*(t)\right\} = \xi t, \quad \textit{and} \quad \mathrm{Var}\{\tau^*(t)\} = \frac{2\omega^2}{\lambda^2}\left(e^{-\lambda t} - 1 + \lambda t\right).$$

For small t *it is sometimes helpful to think of* $\mathrm{Var}\{\tau^*(t)\}$ *as being approximated by* $\omega^2 t^2 + o(t^2)$, *while for large* t

$$\mathrm{Var}\{\tau^*(t)\} \approx \frac{2\omega^2}{\lambda} t.$$

Hence over small intervals the standard deviation of $\tau^*(t)$ is of the same order as the mean, while for large t the mean dominates. Further,

$$\tau^*(t) \underset{\lambda\to\infty}{\overset{a.s.}{\to}} \xi t,$$

that is as the reversion parameter increases the integrated process converges to its expected path.

Example 11 *Suppose we parameterise the autocorrelation function through a distribution. In particular we write*

$$r(u) = \int_0^\infty e^{-u\lambda}\nu(\mathrm{d}\lambda),$$

where v is a measure on λ and

$$r^*(t) = \int_0^\infty \frac{1}{\lambda}\left(1 - e^{-t\lambda}\right)\nu(\mathrm{d}\lambda)$$

and

$$r^{**}(t) = \int_0^\infty \lambda^{-2}\left(e^{-\lambda t} - 1 + \lambda t\right)\nu(\mathrm{d}\lambda).$$

If ν is the form of a GIG distribution then $r^{**}(t)$ can be computed analytically due to the form of the GIG density. This provides a flexible three parameter model for r. In the special case where v corresponds to a $\Gamma(2\bar{H}, \alpha)$ then

$$r(u) = \left(1 + \frac{u}{\alpha}\right)^{-2\bar{H}},$$

while writing $b = 2\bar{H}$ and setting $\alpha = 1$ we find that for $b < 1$ (which deliver long-memory models)

$$r^*(t) = (1-b)^{-1}\left\{(1+t)^{1-b} - 1\right\}$$

and

$$r^{**}(t) = (1-b)^{-1}\left[(2-b)^{-1}\left\{(1+t)^{2-b} - 1\right\} - t\right].$$

Example 12 *Gaussian OU process of $\log\tau(t)$. In the previous Chapter we studied the Gaussian OU process for $\log\tau(t)$ given in (4.25). The unconditional mean and variance of the log process is $\xi_{\log}$ and $\omega^2_{\log}$, respectively. Thus*

$$\mathrm{E}\left\{\tau^*(t)\right\} = \exp\left\{t\xi_{\log} + \frac{1}{2}t\omega^2_{\log}\right\}.$$

We showed in (4.26) that, for $\tau(t)$, writing $r_{\log}(s) = \exp(-\lambda|s|)$ as the autocorrelation function of $\log\tau(t)$,

$$\begin{aligned}\frac{\left\{\exp\left(\omega^2_{\log}\right) - 1\right\}}{\exp\left(\omega^2_{\log}\right)} r(s) &= \exp\left\{\omega^2_{\log} e^{-\lambda s}\right\} - 1 \\ &= \sum_{j=1}^{\infty} \frac{\left(\omega^2_{\log}\right)^j}{j!} r_{\log}(js).\end{aligned}$$

Note that this autocorrelation function is the same as would result from a sup-OU process (4.37) for a particular choice of v. We can note that

$$\begin{aligned}\int_0^t \exp(-j\lambda s)ds &= \frac{1}{j}\int_0^{tj} \exp(-\lambda u)du, \qquad u = js \\ &= \frac{1}{j} r^*_{\log}(jt),\end{aligned}$$

as a result, recalling $r^{**}_{\log}(s) = \lambda^{-2}\left(e^{-\lambda s} - 1 + \lambda s\right)$,

$$\frac{\left\{\exp\left(\omega^2_{\log}\right) - 1\right\}}{\exp\left(\omega^2_{\log}\right)} r^{**}_{\tau}(s) = \sum_{j=1}^{\infty} \frac{\left(\omega^2_{\log}\right)^j}{j! j^2} r^{**}_{\log}(js).$$

4.5.2 Increments of integrated non-negative processes

It is possible to use the above results to study the dynamics of equally spaced increments of the integrated process. Let us write, for some interval $\hbar > 0$,

$$\tau_n = \int_{(n-1)\hbar}^{n\hbar} \tau(u)\mathrm{d}u = \tau^*(n\hbar) - \tau^*((n-1)\hbar).$$

If τ is stationary, then these increments are also stationary. Then (6.94) implies immediately that

$$\mathrm{E}(\tau_n) = \xi\hbar, \quad \text{and} \quad \mathrm{Var}(\tau_n)\} = 2\omega^2 r^{**}(\hbar). \tag{4.48}$$

We can also deduce the autocorrelation function of the discrete time process τ_n, for

$$\begin{aligned}
\mathrm{Cov}(\tau_n, \tau_{n+s}) &= \int_{(n-1)\hbar}^{n\hbar} \int_{(n+s-1)\hbar}^{(n+s)\hbar} \mathrm{Cov}\{\tau(u), \tau(v)\}\mathrm{d}u\mathrm{d}v \\
&= \omega^2 \int_{(n-1)\hbar}^{n\hbar} \int_{(n+s-1)\hbar}^{(n+s)\hbar} r(v-u)\mathrm{d}u\mathrm{d}v \\
&= \omega^2 \int_0^{\hbar} \int_{s\hbar}^{(s+1)\hbar} r(v-u)\mathrm{d}u\mathrm{d}v \\
&= \omega^2 \int_0^{\hbar} \{r^*((s+1)\hbar - u) - r^*(s\hbar - u)\}\mathrm{d}u \\
&= \omega^2 \diamondsuit r^{**}(\hbar s)
\end{aligned} \tag{4.49}$$

where

$$\diamondsuit r^{**}(\hbar s) = r^{**}((s+1)\hbar) - 2r^{**}(\hbar s) + r^{**}((s-1)\hbar). \tag{4.50}$$

Here the notation $\diamondsuit$ is like a second order difference operator (that is $(1-L)^2 r^{**}(s+1)$ where L is the lag operator) in time series analysis. From this result we have that

$$\mathrm{Cor}(\tau_n, \tau_{n+s}) = \frac{\diamondsuit r^{**}(\hbar s)}{2r^{**}(\hbar)}, \quad s = 1, 2, \ldots, \tag{4.51}$$

which only depends upon the autocorrelation function of τ. In particular all models for which we can compute r^{**} allows us to calculate the autocorrelation function of the increments of the processes. Here we discuss in detail just a single example of this to illustrate its structure.

Example 13 *Suppose* $r(u) = \exp(-\lambda|u|)$, *then*

$$\diamondsuit r^{**}(\hbar s) = \lambda^{-2}(1 - e^{-\lambda\hbar})^2 e^{-\lambda\hbar(s-1)},$$

which implies

$$\mathrm{Cor}(\tau_n, \tau_{n+s}) = \frac{(1 - e^{-\lambda\hbar})^2}{2\left(e^{-\lambda\hbar} - 1 + \lambda\hbar\right)} e^{-\lambda\hbar(s-1)} = ce^{-\lambda\hbar(s-1)}, \quad s = 1, 2, \ldots .$$

This has an exponential damp down, so behaves liked the autocorrelation function of an ARMA$(1,1)$ *process with a positive autoregressive root* $(e^{-\lambda\hbar})$ *and a moving average root also determined by* $\lambda\hbar$ *(i.e. there is only a single parameter which determines both the autoregressive and moving average parameter). Of course for* $\hbar$ *small* $c \simeq 1 \simeq e^{-\lambda\hbar}$ *and so the process is close to having the autocorrelation function of an AR*(1) *process.*

4.5.3 intOU processes

Basics

In this subsection we will study the detailed properties of integrated processes based upon the OU processes of the form

$$\mathrm{d}\tau(t) = -\lambda\tau(t)\mathrm{d}t + \mathrm{d}z(\lambda t),$$

for $\lambda > 0$. These processes are labelled intOU processes. If we model $\tau(t)$ as a D-OU process then we say $\tau^*(t)$ is an int$D$-OU process, while an OU-D process delivers an intOU-D process for $\tau^*(t)$.

Although the OU process $\tau(t)$ has jumps, due to the jumps in the Lévy process $z(\lambda t)$, it still holds that $\partial\tau^*(t)/\partial t = \tau(t)$. A major advantage of the OU processes is that

$$\tau^*(t) = \lambda^{-1}\{z(\lambda t) - \tau(t) + \tau(0)\} \tag{4.52}$$

τhas a simple structure — a result which is proved in the Appendix. Hence $\tau^*(t)$ has continuous sample paths because $z(\lambda t)$ and $\tau(t)$ co-break (that is where components of a multivariate series exhibit breaks but a linear combination of that series does not). Figure 4.7 shows this feature for a Γ-OU process, that is a process with a gamma distributed marginal law $\Gamma(\nu, \alpha)$. For the simulated process we plot $\tau^*(t)$ and $z(t\lambda)$ against t. The intOU process has no jumps, although the gradient of the process clearly changes over time. The BDLP has its familiar upward jumps. Further, $\tau^*(t)$ has a lower bound made up of $\lambda^{-1}(1 - e^{-\lambda t})\tau(0)$.

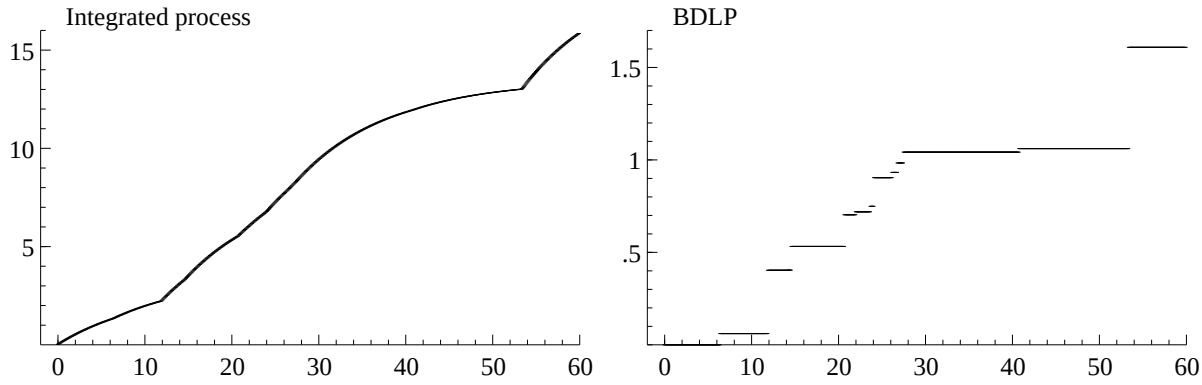

Figure 4.7: $\Gamma-$OU process with $\nu = 3$, $\alpha = 8.5$. Left: plot of $x^*(t)$ against t. Right: plot of $z(\lambda t)$, the BDLP, against t.

The fact that $z(\lambda t)$ and $\tau(t)$ co-break has deep implications for the use of this model. We can then use it as a chronometer of Brownian motion with drift, implying the resulting process $y^*(t)$ has continuous sample paths. This contrasts with the usual case of subordination in the probability literature where the Brownian motion plus drift is subordinated by a Lévy process, $z(t)$. In that case the resulting $y^*(t)$ process must have jumps. Examples of such processes are the normal gamma, normal inverse Gaussian and hyperbolic Lévy processes discussed in Chapter 2.

Although $z(\lambda t)$ and $\tau(t)$ co-break, they do not co-integrate (that is where linear combinations of a non-stationary multivariate system are stationary). Instead, the long-run behaviour of $\tau^*(t)$ is dominated by $z(\lambda t)$. This is clear from rewriting (4.52) as

$$\lambda\tau^*(t) - z(\lambda t) = \tau(0) - \tau(t),$$

which means $\tau^*(t)$ and $z(\lambda t)$ (rather than $\tau(t)$ and $z(\lambda t)$) co-integrate. So roughly $\lambda x^*(t)$ will have the same distribution as $z(\lambda t)$ — the error in this approximation is a stationary process. The distribution of the error, for large t and $\tau(t)$ being a D-OU process, is approximately the difference of two independent random variables drawn from the distribution D.

Prediction distributions for intOU processes

The attractive feature of (4.52) is that the density of the future intOU process $\tau^*(t)|\tau(0)$ is determined by just $z(\lambda t) - \tau(t)|\tau(0)$. As both $z(\lambda t)$ and $\tau(t)$ are linear we can see that this will be mathematically tractable. In particular

$$\tau^*(t) = \lambda^{-1}(1 - e^{-\lambda t})\tau(0) + \lambda^{-1}\int_0^t \left\{1 - e^{-\lambda(t-s)}\right\} \mathrm{d}z(\lambda s)$$

implies we have only to study the stochastic properties of the innovations for the intOU process

$$\lambda^{-1}\int_0^t \left\{1 - e^{-\lambda(t-s)}\right\} \mathrm{d}z(\lambda s) = \int_0^t \varepsilon(t-s;\lambda)\mathrm{d}z(\lambda s) \tag{4.53}$$

where

$$\varepsilon(t;\lambda) = \lambda^{-1}(1 - e^{-\lambda t}). \tag{4.54}$$

and so

$$\lambda^{-1}\int_0^t \left\{1 - e^{-\lambda(t-s)}\right\} \mathrm{d}z(\lambda s) \stackrel{\mathcal{L}}{=} \lambda^{-1}\int_0^{\lambda t} \left\{1 - e^{-s}\right\} \mathrm{d}z(s).$$

In this section we will show how to simulate from this object and also how to compute its cumulants.

The first two moments of $\tau^*(t)|\tau(0)$ can be calculated. We start by working out the mean and variance of this term, before going on to derive the cumulant function. Recalling we wrote the first two cumulants of $z(1)$ as κ_1 (which also equals $\mathrm{E}(\tau(t))$ and κ_2 (which equals $2\mathrm{Var}(\tau(t))$, then

$$\begin{aligned}
\mathrm{E}\left\{\lambda^{-1}\int_0^{\lambda t} (1 - e^{-s})\,\mathrm{d}z(s)\right\} &= \lambda^{-1}\kappa_1\int_0^{\lambda t}(1 - e^{-s})\,\mathrm{d}s \\
&= \lambda^{-1}\kappa_1\left(\lambda t - 1 + e^{-\lambda t}\right).
\end{aligned}$$

The implication is that

$$\mathrm{E}\left\{\tau^*(t)|\tau(0)\right\} = \lambda^{-1}(1 - e^{-\lambda t})\tau(0) + \lambda^{-1}\kappa_1\left(\lambda t - 1 + e^{-\lambda t}\right). \tag{4.55}$$

The corresponding result for the conditional variance is

$$\begin{aligned}
\mathrm{Var}\left\{\tau^*(t)|\tau(0)\right\} &= \lambda^{-2}\kappa_2\int_0^{\lambda t}(1 - e^{-s})^2\mathrm{d}s \\
&= \lambda^{-2}\kappa_2\left(\lambda t - 2 + 2e^{-\lambda t} + \frac{1}{2} - \frac{1}{2}e^{-2\lambda t}\right).
\end{aligned} \tag{4.56}$$

One of the main advantages of basing the model on an OU process is that we are able to derive the conditional cumulant function of the intOU process. In the next chapter we establish the result that

$$\begin{aligned}
\overline{\mathrm{K}}\left\{\theta \ddagger \tau^*(t)|\tau(0)\right\} &= -\theta\varepsilon(t;\lambda)\tau(0) + \lambda\int_0^t k(\theta\varepsilon(s;\lambda))\mathrm{d}s \\
&= -\theta\varepsilon(t;\lambda)\tau(0) + \int_0^{1-e^{-\lambda t}} (1-u)^{-1}k(\lambda^{-1}\theta u)\mathrm{d}u,
\end{aligned} \tag{4.57}$$

recalling that $k(\theta) = \log \mathrm{E}\left[\exp\left\{-\theta z(1)\right\}\right]$. We have derived some examples of this. These results will prove to be very helpful when it comes to option pricing later on in this book. Hence our list of examples is rather extensive.

- *OU-Poisson.* This shot noise process has $k(\theta) = -\psi\left(1 - e^{-\theta}\right)$ and so we have to integrate

$$\int_0^{1-e^{-\lambda t}} (1-u)^{-1} k(\lambda^{-1}\theta u)\mathrm{d}u = -\psi \int_0^{1-e^{-\lambda t}} (1-u)^{-1}\left\{1 - \exp\left(-\chi\theta u\right)\right\}\mathrm{d}u,$$

  where $\chi = \lambda^{-1}\theta$. Now

$$\int (1-u)^{-1}\exp\left(-\chi u\right)\mathrm{d}u = e^{-\chi} E_1\left\{\chi - \chi u\right\},$$

  where $E_1(x)$ is the exponential integral $\int_\chi^\infty y^{-1}e^{-y}\mathrm{d}y$.

- *OU-*Γ. This has $k(\theta) = -\nu\log(1 + \theta\alpha^{-1})$. Then

$$\int_0^{1-e^{-\lambda t}} (1-u)^{-1} k(\lambda^{-1}\theta u)\mathrm{d}u = -\nu \int_0^{1-e^{-\lambda t}} (1-u)^{-1}\log(1+\chi u)\mathrm{d}u,$$

  where $\chi = \lambda^{-1}\theta\alpha^{-1}$. Now

$$\int \frac{\log(1+\chi u)}{(1-u)}\mathrm{d}u = -\log\left(1+\chi u\right)\log\left(1 - \frac{1+\chi u}{1+\chi}\right) - PolyLog\left(2, \frac{1+\chi u}{1+\chi}\right),$$

  where $PolyLog(n,z) = \sum_{k=1}^\infty z^k/k^n$.

- *OU-IG.* This has $k(\theta) = \hbar\gamma - \hbar\gamma(1 + 2\gamma^{-2}\theta)^{1/2}$. Then

$$\begin{aligned}\int_0^{1-e^{-\lambda t}} (1-u)^{-1} k(\lambda^{-1}\theta u)\mathrm{d}u &= \hbar\gamma \int_0^{1-e^{-\lambda t}} \frac{1-(1+\chi u)^{1/2}}{1-u}\mathrm{d}u \\ &= -\hbar\gamma\left\{I(\chi,t) - \lambda t\right\},\end{aligned}$$

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

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

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

- *IG-OU.* This has $\acute{k}(\theta) = \hbar\gamma - \hbar\gamma\left(1 + 2\theta\gamma^{-2}\right)^{1/2}$ and so (**??**) states that

$$k(\theta) = -\frac{\theta\hbar}{\gamma}\left(1 + 2\theta\gamma^{-2}\right)^{-1/2}.$$

  Then

$$\int_0^{1-e^{-\lambda t}} (1-u)^{-1} k(\lambda^{-1}\theta u)\mathrm{d}u = -\frac{\hbar\theta}{\gamma\lambda}\int_0^{1-e^{-\lambda t}} (1-u)^{-1} u\left(1+\chi u\right)^{-1/2}\mathrm{d}u, \tag{4.58}$$

  where $\chi = 2\gamma^{-2}\lambda^{-1}\theta$. Now

$$\int (1-u)^{-1} u\left(1+\chi u\right)^{-1/2}\mathrm{d}u = -\frac{2\sqrt{1+\chi u}}{\chi} + \frac{2\arctan h\left\{\frac{\sqrt{1+\chi u}}{\sqrt{1+\chi}}\right\}}{\sqrt{1+\chi}}.$$

We will return to these types of results in a later chapter.

In a recent paper Lewis (2000) has calculated the moments of $\tau^*(t)|\tau(0)$ for a CIR process. We will report those results here.

Unconditional distributions for intOU processes

Cumulant function of $\tau^*(t)$ can be derived straightforwardly from the conditional cumulant function given in (4.57). The result is

$$\begin{aligned}\overline{\mathrm{K}}\left\{\theta \ddagger \tau^*(t)\right\} &= \acute{k}\left\{\theta\varepsilon(t;\lambda)\right\} + \lambda\int_0^t k\left\{\theta\varepsilon(s;\lambda)\right\}\mathrm{d}s \\ &= \acute{k}\left\{\theta\varepsilon(t;\lambda)\right\} + \int_0^{1-e^{-\lambda t}} (1-u)^{-1}k(\lambda^{-1}\theta u)\mathrm{d}u, \end{aligned} \tag{4.59}$$

recalling that $\varepsilon(t;\lambda) = \lambda^{-1}(1-e^{-\lambda t})$. To be concrete it is helpful to think of an example.

- IG-OU process has $\acute{k}(\theta) = \hbar\gamma - \hbar\gamma\left(1+2\theta\gamma^{-2}\right)^{1/2}$, while the required integral given in (**??**) is solved in (4.58). Hence in principle the law of $\tau^*(t)$ can be found by simply inverting the cumulant function to give the corresponding density or distribution function.

More generally, all of the cases dealt with in the previous section on conditional cumulant functions carry over to the unconditional one.

Conditional simulation

In this subsection we will give methods for simulating from the innovations (4.53). We do this for (i) compound processes, (ii) general processes using infinite series approximation. In the compound process we can write the innovations as equal to

$$\begin{aligned}&\lambda^{-1}\int_0^{\lambda t}\left(1-e^{-s}\right)\mathrm{d}z(s) \\ =\ &\lambda^{-1}\sum_{j=1}^{N(\lambda t)}\left(1-e^{-\lambda t+a_j}\right)c_j,\end{aligned}$$

where $N(t)\sim Po(\psi t)$, $\{a_j\}$ are the arrival times of a Poisson process with intensity ψ and the $\{c_j\}$ are the shocks in the compound Poisson process. In the infinite series approximation we write

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

where $\{r_i\}$ are independent uniforms.

Asymptotic cumulant as $\lambda \downarrow 0$

One way of trying to get an analytic understanding of the behaviour of the intOU process is to think about the process as $\lambda \downarrow 0$, recalling we have already shown that

$$\tau^*(t) \underset{\lambda\to\infty}{\overset{a.s.}{\to}} \xi t.$$

Here we see that

$$\lim_{\lambda\downarrow 0}\overline{\mathrm{K}}\left\{\theta \ddagger \tau^*(t)\right\} = \acute{k}\left(\theta t\right).$$

Of course this simplification is much easier to manipulate.

Example 3.2 ***OU-IG case.*** Then

$$\begin{aligned}k^*(\theta,t,0) &= \int_0^t \left\{\hbar\gamma - \hbar\gamma(1+2\gamma^{-2}\theta)^{1/2}\right\}u^{-1}\mathrm{d}u \\ &= 2\hbar\theta\int_0^t \left\{\hbar\gamma + \hbar\gamma(1+2\gamma^{-2}\theta)^{1/2}\right\}^{-1}\mathrm{d}u.\end{aligned}$$

4.5.4 Integrated diffusion based models

intCIR processes

The square root model

$$\mathrm{d}\tau(t) = -\lambda\left\{\tau(t) - \xi\right\}\mathrm{d}t + \omega\sqrt{\tau(t)}\mathrm{d}b(\lambda t),$$

is analytically quite similiar to the Γ-OU process. Its has the direct solution

$$\tau(t) = \tau(0) + \lambda\xi t - \lambda\tau^*(t) + \omega\int_0^t \sqrt{\tau(u)}\mathrm{d}b(\lambda u).$$

Rearranging we have that

$$\tau^*(t) = \xi t + \lambda^{-1}\left\{\omega\int_0^t \sqrt{\tau(u)}\mathrm{d}b(\lambda u) - \tau(t) + \tau(0)\right\},$$

which has many similiarities to the OU case (4.52). In particular the important predictive role of

$$\omega\int_0^t \sqrt{\tau(u)}\mathrm{d}b(\lambda u) - \tau(t)$$

becomes clear.

Importantly

$$\begin{aligned}\mathrm{E}\left\{\tau^*(t)|\tau(0)\right\} &= \xi t + \lambda^{-1}\left[\tau(0) - E\left\{\tau(t)|\tau(0)\right\}\right] \\ &= \xi t + \varepsilon(t;\lambda)\left[\tau(0) - \xi\right].\end{aligned}$$

A complete characterisation of the conditional distribution is available from the cumulant function, which has a simple affine structure

$$\mathrm{C}\{\zeta \ddagger \tau^*(t)|\tau(0)\} = B(\zeta;t)\tau(0) + A(\zeta;t)$$

where

$$A(\zeta;t) = \omega^{-2}\xi\left\{\lambda t + 2\log(\cosh\frac{\gamma t}{2} + \frac{\lambda}{\gamma}\sinh\frac{\gamma t}{2})\right\}$$

$$B(\zeta;t) = \frac{2i\zeta}{\lambda + \gamma\coth\frac{\gamma t}{2}}$$

with γ given by

$$\gamma = \sqrt{\lambda^2 - 2\omega^2 i\zeta}.$$

OLE: DISCUSS THE DERIVATION OF THIS.

Results on Lewis on moments??. Also add Ole's results.

Affine models

4.5.5 Superposition of integrated non-negative processes

In the previous Section we increased the scope of the dynamics of our models by superposition of non-negative processes. Here we do the same for integrated processes. Suppose that

$$\tau(t) = \sum_{j=1}^{m}\tau_j(t),$$

where the $\tau_j(t)$ are independent non-negative, stationary processes. Then

$$\begin{aligned}\tau^*(t) &= \int_0^t \tau(u)\mathrm{d}u \\ &= \sum_{j=1}^m \tau_j^*(t),\end{aligned}$$

implying the integrated superposition of non-negative processes is equal to a superposition of integrated non-negative processes.

In the superposition case we can calculate the cumulant function of the prediction distribution

$$\tau^*(t)|\tau_1(0),...,\tau_m(0),$$

when we assume the components τ_j are Markov case — as would hold in the OU and diffusion cases. To be specific, in those cases

$$\overline{\mathrm{K}}\left\{\theta \ddagger \tau^*(t)\right\} = \sum_{j=1}^m \overline{\mathrm{K}}\left\{\theta \ddagger \tau_j^*(t)|\tau_j(0)\right\}.$$

An important feature of the above analysis is that they are calculated by conditioning on m initial values, rather than on $\tau(0)$. This is because τ is not Markovian, while the τ_j are assumed to be so.

Corresponding results carry over to the increments of the process. They are again defined as

$$\begin{aligned}\tau_n &= \tau^*(n\hbar) - \tau^*((n-1)\,\hbar) \\ &= \int_{(n-1)\hbar}^{n\hbar} \tau(u)\mathrm{d}u \\ &= \sum_{j=1}^m \tau_{j,n},\end{aligned}$$

where

$$\tau_{j,n} = \tau_j^*(n\hbar) - \tau_j^*((n-1)\,\hbar).$$

Example 14 *Suppose $\tau_j^*(t)$ has mean, variance and autocorrelation function denoted by ξ_j, ω_j^2 and $r_j(u) = \exp(-\lambda_j\,|u|)$, respectively. Then $\mathrm{E}\{\tau_n\} = \xi_+\hbar$ and $\mathrm{Var}\{\tau_n\} = 2\sum_{j=1}^m \omega_j^2 r_j^{**}(\hbar)$, while implies*

$$\mathrm{Var}\{\tau_n\} = 2\sum_{j=1}^n \omega_j^2\lambda_j^{-2}\left(e^{-\lambda_j\hbar} - 1 + \lambda_j\hbar\right).$$

Likewise

$$\begin{aligned}\mathrm{Cov}(\tau_n, \tau_{n+s}) &= \sum_{j=1}^m \omega_j^2 \diamondsuit r_j^{**}(\hbar s) \\ &= \sum_{j=1}^m \omega_j^2\lambda_j^{-2}(1-e^{-\lambda_j\hbar})^2 e^{-\lambda_j\hbar(s-1)}.\end{aligned}$$

This implies

$$\mathrm{Cor}(\tau_n, \tau_{n+s}) = \sum_{j=1}^m c_j e^{-\lambda_j\hbar(s-1)}, \qquad \textit{where} \qquad c_j = \frac{\omega_j^2\lambda_j^{-2}(1-e^{-\lambda_j\hbar})^2}{\mathrm{Var}\{\tau_n\}}.$$

Each of the weights $\{c_j\}$ are non-negative. Of course for small $\hbar$,

$$\lambda_j^{-2}(1-e^{-\lambda_j\hbar})^2 \simeq \hbar^2 \qquad \textit{and} \qquad 2\lambda_i^{-2}\left\{e^{-\lambda_i\hbar} - 1 + \lambda_i\hbar\right\} \simeq \hbar^2,$$

which implies $c_j \simeq \hbar_j/\hbar_+$, which sum to one.

4.6 Conclusion

4.7 Appendix

4.7.1 Conditions for the existence of an OU process

We are concerned with integrals of the form

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

Let us consider this for the special case where z is a subordinator, that is a Lévy process with non-negative increments. Then we may take the integral as being determined pathwise. The process x is clearly nonnegative and increasing and hence a chronometor. We write

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

Under a mild condition on z, $x(\infty)$ will be finite almost surely, that is it will be a random variable, and since each $x(t)$ is infinitely divisible the same will then hold for $x(\infty)$. To determine the relevant conditions under which this holds we consider the kumulant function of $x(t)$ which, by (2.14), is

$$\overline{\mathrm{K}}\{\theta \ddagger x(t)\} = -\int_0^t \int_0^\infty \{1 - \exp(-e^{-s}\theta x)\} W(\mathrm{d}x)\mathrm{d}s.$$

Using the substitutions $y = e^{-s}x$ and $r = e^s$ may rewrite this as

$$\begin{aligned} \overline{\mathrm{K}}\{\theta \ddagger x(t)\} &= -\int_0^t \int_1^\infty \{1 - \exp(-\theta y)\} W(e^s \mathrm{d}y)\mathrm{d}s \\ &= -\int_1^\infty \int_1^{e^t} \{1 - \exp(-\theta y)\} W(r\mathrm{d}y) r^{-1}\mathrm{d}r \\ &= -\int_1^\infty \{1 - \exp(-\theta y)\} U(t)(\mathrm{d}y), \end{aligned}$$

where

$$U(t)(\mathrm{d}y) = \int_1^{e^t} W(r\mathrm{d}y) r^{-1}\mathrm{d}r,$$

is the Lévy measure of $x(t)$. Now suppose for simplicity that the Lévy measure W is absolutely continuous with a density w. Then the $U(t)$ is also absolutely continuous, with density

$$U(t)(y) = \int_1^{e^t} w(ry)\mathrm{d}r = y^{-1} \int_y^{e^t y} w(x)\mathrm{d}x$$

and for $t \to \infty$

$$U(t)(y) \to u(y) = y^{-1} W^+(y).$$

For $x(\infty)$ to be a random variable this limiting function u should be a Lévy density, i.e. it should satisfy the integrability condition (2.15). Noting first that

$$W^+(y) = \int_y^\infty w(x)\mathrm{d}x = y \int_1^\infty w(y\tau)\mathrm{d}\tau$$

we find

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

In the latter expression the first and third integrals are finite since W is a Lévy measure. We are thus led to the condition

$$\int_1^{\infty} \log(y)\ \ w(y)\mathrm{d}y < \infty,$$

for finitenes almost surely of $x(\infty)$.

4.8 Exercises

4.9 Appropriate literature

4.9.1 Time deformation

4.9.2 OU type processes

Autoregressions are one of the most common models used in modern time series, textbook expositions include Brockwell and Davis (1987) and Hamilton (1994). Equilibrium correction models were introduced by Sargen (1964) and were are highlighted by Hendry (1995a), which discusses his pioneering work on the topic as well as discussing in detail the contributions of others. Gaussian OU processes were introduced by Uhlenbeck and Ornstein (1930).

Non-Gaussian OU processes have been discussed in the probability literature for quite some time with precise statements of existence given in cf. Wolfe (1982) and Jurek and Vervaat (1983) (see also Barndorff-Nielsen, Jensen, and Sørensen (1998)). The model has been used in, for example, storage theory by, for example, Cinlar and Pinsky (1972), Harrison and Resnick (1976) and Brockwell, Resnick, and Tweedie (1982). Extensions to the ARMA case are discussed by Brockwell (2001). GIG distributions were shown to be self-decomposable, and so support OU processes, by Halgreen (1979). A wide development of these types of model, in detail, is given in Barndorff-Nielsen and Shephard (2001a). These authors introduced the notation OU-D and D-OU.

The simulation of stochastic integrals based on Lévy processes is not straightforward due to the jump character of the processes. We use infinite series representations of these types of integrals. The required results are, in essence, available from work of Marcus (1987) and Rosinski (1991). A self-contained exposition of this result is given in Barndorff-Nielsen and Shephard (2001b), while recent developments are surveyed in **?**); see also Protter and Talay (1999), Ferguson and Klass (1972), Vervaat (1979) and Walker and Damien (2000).

The idea of incresding the flexibility of the dynamics by using superpositions is very old going back to ?. Granger (1980) and Cox (1991) use this construction, for real valued discrete time processes, to build long-memory processes. A formal theory of an infinite dimensional superposition was developed for non-Gaussian OU processes by Barndorff-Nielsen (2001).

4.9.3 Integrated processes

The idea of integrated processes appears in the work of Box and Jenkins (1970), while cointegration was introduced by Clive Granger in the early 1980s. An elegant framing of these ideas is given in Engle and Granger (1987). Co-breaking was introduced by Hendry (1995b) and eventually published in Clements and Hendry (1999, Ch. 9).

Barndorff-Nielsen and Shephard (2002b) have studied in detail the properties of intOU process. Their analysis covers both non-negative processes as well as processes on the real line. The conditional cumulant function for $\sigma^{2*}(t)|\tau(0)$ has been derived for many different types of D-OU and OU-D processes by **?**) and Tompkins and Hubalek (2000). We have given a selection of results above.

Carr, Geman, Madan, and Yor (2001) gave the expression for the cumulant function for an integrated CIR process. Lewis (2000) has derived all the moments of a CIR process. Variograms play a large role in spatial statistics, see for example Cressie (1993). They have been highlighted recently in transaction level econometrics in some work by Hillman and Salmon (1999) and also appear in Rydberg and Shephard (2000).

Chapter 5

Stochastic volatility

Abstract: In this Chapter we analyse stochastic volatility models, that is time deformation models with differentiable chronometers. Such models are the bedrock of much of modern financial econometrics. Here we discuss the basic properties of these models, leverage effects and estimation and testing strategies for these models.

5.1 What is this Chapter about?

-
-
-
-
-
-
-

5.2 Univariate stochastic volatility

Recall that in Chapter 2 we introduced the idea of subordination, where we took a Brownian motion b and replaced its' clock by a chronometer. This was a positive stochastic process $\tau(t)$ which was monotonically increasing and tending to infinity for t tending to infinity and is independent of the Brownian motion b. The resulting process is $b\{\tau(t)\}$. A simple case of this is where $\tau(t)$ is a Lévy process with non-negative increments, which we called a subordinator in Chapter 2.

Now consider models of the type

$$y^*(t) = \int_0^t \sigma(s)\mathrm{d}w(s), \tag{5.1}$$

where the processes σ and w are independent, w being a Brownian motion and σ being positive and predictable and such that $\sigma^{2*}(t) \to \infty$ for $t \to \infty$. It turns out that, in essence, there is equivalence between the model formulation by (5.1) and the model formulation by subordination with an independent chronometer σ^{2*} when this chronometer has continuous sample paths. This result is proved formally in Chapter ?. This result does not hold when the chronometer is a subordinator such as a compound Poisson process or a inverse Gaussian process, for then it has discontinuous paths. Hence Brownian motions subordinated by Lévy processes are not SV processes.

Our integration based chronometers have the feature that log-prices have continuous sample paths and that increments obey the law

$$y_n|\tau_n \sim N(\mu\hbar + \beta\tau_n, \tau_n), \qquad n = 1, 2, 3, ...$$

where again $\hbar > 0$ and

$$y_n = y^*(\hbar n) - y^*((n-1)\hbar) \qquad \text{and} \qquad \tau_n = \tau^*(\hbar n) - \tau^*((n-1)\hbar).$$

Hence returns y_n will inherit serial dependence via the serial dependence in τ_n, the increments to the chronometer.

In this Chapter we will study some of the generic properties of these chronometers, however our main focus will be on models for which we are able to show that (i) they have empirically desirable properties such that $\{\tau_n\}$ is serial correlated with thick tails, (ii) are mathematically tractable so that we can calculate the cumulant functions of τ_n and its predictive version $\tau_n|\tau\{(n-1)\hbar\}$. We will pay particular attention to non-negative Ornstein-Uhlenbeck (OU) processes which solve the stochastic differential equation (SDE)

$$\mathrm{d}\tau(t) = -\lambda\tau(t)\mathrm{d}t + \mathrm{d}z(\lambda t), \qquad \lambda > 0,$$

where z is a subordinator, whose non-negative increments keep τ from going negative. Hence Lévy processes will again play an important role in our work, however now they appear twice removed, as the process which drives the chronometer which changes the clock of the Brownian motion. We will also discuss the common alternative to this approach which is to use non-linear processes driven by a second Brownian motion b, which is assumed to be independent from w. An example of this is Feller's square root (also called the Cox-Ingersoll-Ross or CIR) process

$$\mathrm{d}\tau(t) = -\lambda\{\tau(t) - \xi\}\,\mathrm{d}t + \omega\tau^{1/2}(t)\mathrm{d}b(\lambda t), \qquad \text{where} \qquad \xi \geq \omega^2/2 > 0, \quad \lambda > 0. \tag{5.2}$$

This has a reflecting boundary at zero, and so can be used as the basis for a chronometer.

This Chapter has ? section. In Section 2 we will introduce non-negative OU processes, while Section 3 will study superposition of such models (the addition of independent copies of OU processes) which allow us to build long-memory models for the increments to the chronometer. Section 4 will look at diffusion based alternatives to these models. Section 5 will be devoted to the properties of chronometers based on integrating the above stationary processes. Section 6 will draw some conclusions from this Chapter, while Section 7 will discuss the corresponding literature on these topics. Finally the Appendix will contain various technical results we need in this Chapter.

5.2.1 Basic model

Continuous time models built out of Brownian motion play a crucial role in modern finance, providing the basis of most option pricing, asset allocation and term structure theory currently being used. An example is the so called Black-Scholes or Samuelson model which models the log of an asset price by the solution to the stochastic differential equation

$$\mathrm{d}y^*(t) = \{\mu + \beta\tau\}\,\mathrm{d}t + \sigma\mathrm{d}w(t), \tag{5.3}$$

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

$$y_n = \int_{(n-1)\hbar}^{n\hbar} \mathrm{d}y^*(t) = y^*(n\hbar) - y^*\{(n-1)\hbar\}$$

implying returns are normal and independently distributed with a mean of $\mu\hbar + \beta\tau\hbar$ and a variance of $\hbar\tau$. Unfortunately we have already seen that for moderate to small values of $\hbar$ (corresponding to returns measured over 5 minute to one day intervals) returns are typically heavy-tailed, exhibit volatility clustering (in particular the $|y_n|$ are correlated) and are skew, although for higher values of $\hbar$ a central limit theorem seems to hold and so Gaussianity becomes a less poor assumption for $\{y_n\}$ in that case. This means that every single assumption underlying the Black-Scholes model is routinely rejected by the type of data usually used in practice.

This common observation, which carries over to the empirical rejection of option pricing models based on this model, has resulted in an enormous effort to develop empirically more reasonable models which can be integrated into finance theory. The most successful of these

are the generalised autoregressive conditional heteroskedastic (GARCH) and the diffusion based stochastic volatility (SV) processes.

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

$$\mathrm{d}y^*(t) = \{\mu + \beta\tau(t)\}\,\mathrm{d}t + \sigma(t)\mathrm{d}w(t), \tag{5.4}$$

where $\tau(t)$, the instantaneous volatility, is going to be assumed to be stationary, latent and stochastically independent of $w(t)$. Typically we will assume $\tau(t)$ is of OU type

$$\mathrm{d}\tau(t) = -\lambda\tau(t)\mathrm{d}t + \mathrm{d}z(\lambda t),$$

where $z(t)$ is a Lévy process with non-negative increments, or a superposition of such processes

$$\tau(t) = \sum_{j=1}^{m} \sigma_j^2(t), \qquad \text{where} \qquad \mathrm{d}\sigma_j^2(t) = -\lambda_j\sigma_j^2(t)\mathrm{d}t + \mathrm{d}z_j(\lambda_j t),$$

where the $\{z_j(t)\}$ are independent (not necessarily identically distributed) BDLPs. However, as we noted in Chapter 5, there are alternative diffusion based alternatives which are commonly used in the literature. In particular, the square root process

$$\mathrm{d}\tau(t) = -\lambda\left\{\tau(t) - \xi\right\}\mathrm{d}t + \omega\sqrt{\tau(t)}\mathrm{d}b(\lambda t),$$

and the Gaussian OU process for $\log\tau(t)$

$$\mathrm{d}\log\tau(t) = -\lambda\left(\log\tau(t) - \xi_{\log}\right)\mathrm{d}t + \omega_{\log}\mathrm{d}b(\lambda t)$$

have both been frequently used in this context. Here $b(t)$ is standard Brownian motion.

Even if $\tau(t)$ exhibits jumps, as it does in the OU case, $y^*(t)$ is a continuous process for all parameter values — a very important result which will be explained more in the next section. This formulation also makes it clear that in the special case where $\mu = \beta = 0$ an SV process can be thought of as a subordinated Brownian motion were the chronometer is *integrated volatility*

$$\int_0^t \tau(u)\mathrm{d}u.$$

We will delay our discussion of this well known connection also until the next section.

We saw in (6.94) that we can write the first two moments of $\sigma^{2*}(t)$ in terms of the mean, variance and the autocorrelation function of the process $\tau(t)$. Recall, we write these as ξ, ω^2 and r, respectively. The implication is that

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

recalling that $r^{**}(t) = \int_0^t\int_0^u r(s)\mathrm{d}s\mathrm{d}u$. Further, when $\beta = 0$,

$$\kappa_3\{y^*(t)\} = 0 \qquad \text{and} \qquad \kappa_4\{y^*(t)\} = 6\omega^2 r^{**}(t).$$

Finally

$$\begin{aligned}\mathrm{Var}(y^*(t)^2) &= \kappa_4\{y^*(t)\} + 2\mathrm{Var}\{y^*(t)\} \\ &= 6\omega^2 r^{**}(t) + 2t\xi.\end{aligned}$$

Statistically $\mathrm{Var}\{y^*(t)\}$ is particularly important, for a plot of it against t is called a variogram. It is worth noting an example of the variogram.

- Suppose $r(s) = \exp(-\lambda u)$, which is the autocorrelation function of $\tau(t)$ for all OU and CIR processes (assuming r exists). Then $r^{**}(t) = \lambda^{-2}\left(e^{-\lambda t} - 1 + \lambda t\right)$ implying

$$\mathrm{Var}\{y^*(t)\} = \frac{2\beta^2\omega^2}{\lambda^2}\left(e^{-\lambda t} - 1 + \lambda t\right) + t\xi,$$

while when $\beta = 0$ the standardised fourth cumulant is

$$\begin{aligned}\frac{\kappa_4\{y^*(t)\}}{\mathrm{Var}\{y^*(t)\}^2} &= \frac{6\omega^2}{\lambda^2 t^2 \xi^2}\left(e^{-\lambda t} - 1 + \lambda t\right) \\ &\simeq \begin{cases} 3\omega^2\xi^{-2}, & \text{for small } t \\ 0 & \text{for large } t \end{cases}.\end{aligned}$$

Hence for small t this is determined by the coefficient of variation of $\tau(t)$, with larger coefficients leading to higher fourth cumulants. For large t the standardised fourth cumulant converges to zero from above, suggesting a standardised version of $y^*(t)$ is converging to Gaussianity. We will formalise this result in a moment in a much wider context.

Second order properties of returns

In this subsection we derive the moments of discrete time returns (4.4) implied by a general continuous time SV model with $\beta = 0$. In particular

$$\begin{aligned}\mathrm{Var}(y_n^2) &= 6\omega^2 r^{**}(\hbar) + 2\hbar\xi, \\ \mathrm{Cov}\left(y_n^2, y_{n+s}^2\right) &= \mathrm{Cov}\left(\sigma_n^2, \sigma_{n+s}^2\right) = \omega^2 r^{**}(\hbar s),\end{aligned}$$

using (**??**). Consequently

$$\mathrm{Cor}\{y_n^2, y_{n+s}^2\} = \frac{\diamondsuit r^{**}(\hbar s)}{6r^{**}(\hbar) + 2\hbar^2(\xi/\omega)^2} = q^{-1}\hbar^{-2}r^{**}(\hbar s), \tag{5.5}$$

where

$$r^{**}(s) = r^{**}(s + \hbar) - 2r^{**}(s) + r^{**}(s - \hbar), \tag{5.6}$$

and

$$q = 6\hbar^{-2}r^{**}(\hbar) + 2(\xi/\omega)^2. \tag{5.7}$$

We give an example of this in practice.

- Suppose $r(s) = \exp(-\lambda s)$, which is the autocorrelation function of $\tau(t)$ for all OU and CIR processes (assuming r exists). Then $r^{**}(t) = \lambda^{-2}\left(e^{-\lambda t} - 1 + \lambda t\right)$ implying

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

This results in

$$\mathrm{Cor}\{\sigma_n^2, \sigma_{n+s}^2\} = de^{-\lambda\hbar(s-1)}, \quad \mathrm{Cor}\{y_n^2, y_{n+s}^2\} = ce^{-\lambda\hbar(s-1)}, \quad s > 0 \tag{5.8}$$

where

$$\begin{aligned}1 &\geq d = \frac{(1 - e^{-\lambda\hbar})^2}{2\left\{e^{-\lambda\hbar} - 1 + \lambda\hbar\right\}} \\ &\geq c = \frac{(1 - e^{-\lambda\hbar})^2}{6\left\{e^{-\lambda\hbar} - 1 + \lambda\hbar\right\} + 2(\lambda\hbar)^2(\xi/\omega)^2} \geq 0.\end{aligned} \tag{5.9}$$

Note that (5.8) implies that σ_n^2 and y_n^2 follow constrained ARMA$(1,1)$ processes with common autoregressive parameters and with the moving average root being stronger for σ_n^2 than for the y_n^2. The ARMA structure implies that y_n is weak GARCH$(1,1)$. Finally, as $\hbar \to 0$ so $d \to 1$ and so σ_n^2 behaves like a first order autoregression.

5.2.2 SV models and stochastic analysis

SV models have continuous sample paths

A key feature of basic SV models is that the solution to (6.12), $y^*(t)$, has continuous sample paths so long as $\tau(t)$ is càdlàg. This follows from $\sigma^{2*}(t)$ having continuous sample paths so long as $\tau(t)$ is càdlàg. Consequently these models are fundamentally different from those generated by subordinating Brownian motion using the subordinators such as the gamma and the inverse Gaussian Lévy processes discussed in Chapter two of this book. This important point is illustrated in Figure ? which shows a gamma Lévy process and the resulting sample paths for the subordinated series, together with an SV model constructed where $\sigma^2(t)$ is a Γ-OU process.

SV models are local martingales

The special case of the SV model given in (5.1)

Quadratic variation of SV models

Ito calculus and SV models

5.2.3 Leverage

Leverage for OU based SV models

It is possible to generalise (6.12) to allow for the feedback of the innovations of the volatility process into the level of the asset price. In particular, we write

$$\mathrm{d}y^*(t) = \{\mu + \beta\tau(t)\}\,\mathrm{d}t + \sigma(t)\mathrm{d}w(t) + \rho\mathrm{d}\bar{z}(\lambda t), \tag{5.10}$$

where $\bar{z}(t) = z(t) - \mathrm{E}z(t)$, the centred version of the BDLP. It is important to understand that ρ does not represent a correlation in this model. Overall this model to deal with the so called leverage type problem which formalises the observation that for equities a fall in the price is associated with an increase in future volatility. In particular we have that

$$y_n|\sigma_n^2, \overline{z}_n \sim N(\mu\hbar + \beta\sigma_n^2 + \rho\overline{z}_n, \sigma_n^2),$$

where

$$\overline{z}_n = \bar{z}(n\lambda\hbar) - \bar{z}\left\{(n-1)\,\lambda\hbar\right\}.$$

We will discuss some aspects of this model in Section 4 of this Chapter.

Leverage for diffusion based OU models

The financial economics literature usually uses an alternative leverage model. This correlates the Brownian motion which appears in the price and volatility equations. Let use write, for a general diffusion based volatility process,

$$\mathrm{d}y^*(t) = \{\mu + \beta\tau(t)\}\,\mathrm{d}t + \sigma(t)\mathrm{d}w(t), \quad \mathrm{d}\tau(t) = g\left\{\tau(t)\right\}\mathrm{d}t + h\left\{\tau(t)\right\}\mathrm{d}b(t),$$

where $w(t)$ and $b(t)$ are correlated standard Brownian motions with

$$\mathrm{Cov}(w(t), b(t)) = \rho t.$$

We can then rewrite the model, in law, as

$$\mathrm{d}y^*(t) = \{\mu + \beta\tau(t)\}\,\mathrm{d}t + \sqrt{1-\rho^2}\sigma(t)\mathrm{d}w_1(t) + \rho\sigma(t)\mathrm{d}b(t),$$

where $w_1(t)$ is stochastically independent from $b(t)$. Then

$$y_n|\sigma_n^2, \overline{\sigma b}_n \sim N(\mu\hbar + \beta\sigma_n^2 + \rho\overline{\sigma b}_n, \left(1-\rho^2\right)\sigma_n^2),$$

with

$$\overline{\sigma b}(t) = \int_0^t \sigma(u)\mathrm{d}b(u).$$

Of course, although $\overline{\sigma b}_n$ is uncorrelated with σ_n^2, these two processes are not independent.

5.2.4 Specific results for OU based SV models

Background

Some of the above results can be made much more explicit when $\sigma^2(t)$ is of OU type. In this subsection we will develop some of these results. In partical we will study the cumulant and conditional cumulant function of $y^*(t)$, as well as some corresponding moments.

We can give an explicit cumulant function for the OU type volatility process. In particular, using (**??**), we have that

$$\mathrm{K}\{\zeta \ddagger y^*(t)\} = \xi\mu t + \mathrm{K}\{\xi\beta + \frac{1}{2}\xi^2 \ddagger \sigma^{2*}(t)\}$$

We have already given the formula for the cumulant generating function of $\sigma^{2*}(t)$ in (**??**). Consequently we can regard this problem as having been solved in general. Concrete analytic soloutions to the required integrals are available in the OU-Poisson, OU-Γ, OU-IG, IG-OU and Γ-OU cases. Hence for all these models we can evaluate the exact density and distribution function for these variables simply by inverting the cumulant function.

Cumulant functions and moments

The cumulant function for $y^*(t)|\tau(0)$ is simply

$$\mathrm{K}\{\zeta \ddagger y^*(t)|\tau(0)\} = \xi\mu t + \mathrm{K}\{\xi\beta + \frac{1}{2}\xi^2 \ddagger \sigma^{2*}(t)|\tau(0)\},$$

which simply involves the cumulant function of $\sigma^{2*}(t)|\sigma^2(0)$. In Chapter 10 we studied this cumulant function for very many different types of OU processes, implying, in particular, that we know the cumulant function in the OU-Poisson, OU-Γ, OU-IG, IG-OU and Γ-OU cases. This will turn out to be an absolutely crucial result for us when we go on to discussing derivative pricing.

It is worthwhile writing down explicitly the special case of the above result which gives the first two moments of $y^*(t)|\tau(0)$. These work off the result given in (4.55) and (4.56). In particular,

$$\begin{aligned}
\mathrm{E}\left\{y^*(t)|\tau(0)\right\} &= \mu t + \beta\mathrm{E}\left\{\sigma^{2*}(t)|\tau(0)\right\} \\
&= \mu t + \beta\left\{\lambda^{-1}(1-e^{-\lambda t})\tau(0) + \lambda^{-1}\xi\left(\lambda t - 1 + e^{-\lambda t}\right)\right\}
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{Var}\left\{y^*(t)|\tau(0)\right\} &= \beta^2\mathrm{Var}\{\sigma^{2*}(t)|\tau(0)\} + \mathrm{E}\left\{\sigma^{2*}(t)|\tau(0)\right\} \\
&= 2\beta^2\lambda^{-2}\omega^2\left(\lambda t - 2 + 2e^{-\lambda t} + \frac{1}{2} - \frac{1}{2}e^{-2\lambda t}\right) \\
&\quad +\lambda^{-1}(1-e^{-\lambda t})\tau(0) + \lambda^{-1}\xi\left(\lambda t - 1 + e^{-\lambda t}\right).
\end{aligned}$$

One of the interesting features of this result is that the conditional variance depending positively on $\tau(0)$ and the conditional mean depends positively on β if β is positive.

Leverage case

In the leverage case (5.10) the calculations are inevitably more specialised. When $\tau(t) \sim OU$ we are able to produce very concrete results. In particular

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

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

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

The effect of the leverage term is to allow $\mathrm{Cov}\left(y_n y_{n+s}^2\right)$ to be negative if $\rho < 0$. However, in addition both $\mathrm{Cov}\left(y_n y_{n+s}^2\right)$ and $\mathrm{Cov}(y_n^2, y_{n+s}^2)$ damp down exponentially with the lag length s. We should note that exactly the same dynamic structure as the discrete time quadratic ARCH model (QARCH). Hence we can think of the QARCH model as a kind of discrete time representation of our continuous time leverage model, generalising the unleveraged result.

5.2.5 Specific results for diffusion based SV models

5.2.6 SV models with added jumps

5.2.7 Lévy processes with SV effects

5.2.8 Stationary SV models

$$\mathrm{d}y(t) = \left\{\mu + \beta\tau(t) - \lambda y(t)\right\}\mathrm{d}t + \sigma(t)\mathrm{d}w(t).$$

Could also look at a reflecting version of this model which puts

$$\mathrm{d}y(t) = \left\{\mu + \beta\tau(t) - \lambda y(t)\right\}\mathrm{d}t + \sigma(t)y^{1/2}(t)\mathrm{d}w(t),$$

5.2.9 Econometrics of SV models on low frequency data

5.2.10 Empirical performance of SV models on low frequency data

5.3 Multivariate stochastic volatility

5.3.1 Introduction

MULTIVARIATE LEVY PROCESSES FROM CHAPTER 2.

A simple multivariate structure for a log-price vector can be generated off a $N \times 1$ multivariate SV model. In particular a simple extension of the univariate setup is to write

$$\mathrm{d}y^*(t) = \left\{\mu + \beta'\Sigma(t)\right\}\mathrm{d}t + \Sigma(t)^{1/2}\mathrm{d}w(t),$$

where $w(t)$ is a vector of independent standard Brownian motions. Then the return vector

$$y_n = y^*\left(n\hbar\right) - y^*\left(\left(n-1\right)\hbar\right),$$

is a multivariate mixture of normals. In particular

$$y_n|\Sigma_n \sim N(\mu\hbar + \beta'\Sigma_n, \Sigma_n),$$

where

$$\Sigma_n = \Sigma^*\left(n\hbar\right) - \Sigma^*\left(\left(n-1\right)\hbar\right) \qquad \text{and} \qquad \Sigma^*(t) = \int_0^t \Sigma(t)\mathrm{d}t.$$

We call $\Sigma^*(t)$ *integrated covolatility* and Σ_n *actual covolatility*. For the above SV model the *quadratic covariation* is $\Sigma^*(t)$, i.e. we have

$$[y^*](t) = \text{p}-\lim_{r\to\infty} \sum \{y^*(t_{i+1}^r) - y^*(t_i^r)\}\{y^*(t_{i+1}^r) - y^*(t_i^r)\}' = \Sigma^*(t) \tag{5.11}$$

for any sequence of partitions $t_0^r = 0 < t_1^r < ... < t_{m_r}^r = t$ with $\sup_i\{t_{i+1}^r - t_i^r\} \to 0$ for $r \to \infty$. Again this is a robust measure as it produces the integrated covolatility even if μ and β are non-zero. However, it is an entirely asymptotic concept and so is not directly applicable in practice.

5.3.2 Factor models

$$\mathrm{d}y^*(t) = \{\mu + \Sigma(t)\beta\}\,\mathrm{d}t + \Sigma(t)^{1/2}\mathrm{d}w(t),$$

How do we drive $\Sigma(t)$?. A traditional finance approach is to use a factor structure. We follow that here.

$$\Sigma(t) = B\Lambda(t)B' + t\Gamma,$$

where Γ is diagonal (or nearly so), while

$$\Lambda(t) = diag(\sigma_1^2(t), ..., \sigma_K^2),$$

where the elements are independent

$$\mathrm{d}\sigma_j^2(t) = -\lambda_j\sigma_j^2(t)\mathrm{d}t + \mathrm{d}z_j(\lambda_j t).$$

In discrete time these models have a long history going back to Diebold and Nerlove (1989).

The implications of this model is that we can write

$$\mathrm{d}y^*(t) \stackrel{L}{=} \{\mu + B\Lambda(t)\beta^{*\prime}\}\,\mathrm{d}t + \sum_{j=1}^{K} b_j\sigma_j(t)\mathrm{d}w_j(t) + \Gamma^{1/2}\mathrm{d}b,$$

where there is independence amoungst the Brownian motion processes.

5.3.3 Quadratic covariation of SV models

5.3.4 Econometrics of multivariate SV models on low frequency data

5.4 Lévy based SV models

5.4.1 Time deformed Lévy processes

General case

Similiar results can be derived for general Lévy processes which are deformed by general time chronometers $\tau^*(t)$. Write

$$y^*(t) = a^*(t) + z(\tau^*(t))$$

Lévy-Khintchine representation

$$\begin{aligned} \mathrm{C}\{\zeta \ddagger y_1|a_1, \tau_1\} &= \log\left[\mathrm{E}\left\{\exp(i\zeta y_1)|a_1, \tau_1\right\}\right] &(5.12)\\ &= \zeta a_1 + \left(ai\zeta - \frac{1}{2}\sigma^2\zeta^2\right)\tau_1 - \tau_1\int_{\mathbf{R}}\left\{1 - e^{i\zeta x} + i\zeta x\mathbf{1}_B(x)\right\}W(\mathrm{d}x). &(5.13)\end{aligned}$$

From now on we will assume $a^*(t) = \mu t + \beta\tau^*(t)$, then unconditionally

$$\mathrm{C}\{\zeta \ddagger y_1\} = \xi\mu\hbar + \mathrm{K}\{ai\zeta - \frac{1}{2}\sigma^2\zeta^2 - \int_{\mathbf{R}}\left\{1 - e^{i\zeta x} + i\zeta x\mathbf{1}_B(x)\right\}W(\mathrm{d}x) \ddagger \tau_1\},$$

yielding similiarly shaped results for $\mathrm{C}\{\zeta \ddagger y_1|\mathcal{F}_0\}$.

NIG based models

Normal gamma based models

5.5 Conclusion

5.6 Appendix of derivations and proofs

5.7 Exercises

5.8 Appropriate literature

5.8.1 Stochastic volatility

The history of stochastic volatility models is particularly confusing. This is because they live on the intersection of many areas in probability theory, finance and econometrics. We will attempt to clarrify issues. Some of the earlier work on this topic is discussed in the review articles by Taylor (1994), Shephard (1996) and Ghysels, Harvey, and Renault (1996).

If $\sigma^{2*}(t)$ were a Lévy process, then it is a subordinator and $y^*(t)$ is itself a Lévy process as it is a subordinated Brownian motion by a classical subordinator. We will choose to say these are not stochastic volatility models for they are entirely satisfactorily called Lévy processes! This means that we will regard the seminal paper by Clark (1973) on subordination as a paper not on SV processes. This goes against quite a lot of the econmetrics literature which regards this paper as the first one on SV paper.

The first real SV paper is due to Taylor (1982), but this was phrased in discrete time. It placed

$$y_n = \varepsilon_n \exp(h_n/2), \quad \varepsilon_n \sim NID(0,1),$$

where h_n is a Gaussian autoregression which is assumed stochastically independent of the ε_n process. We can think of this as a roughly approximate continuous time SV model where the log of the volatility is a Gaussian OU process. This paper has received very little attention in econometrics, although the later exposition in Taylor (1986) is widely cited.

Independently Hull and White (1987) phrased a SV model in continuous time using the above framework, although their volatility process was assumed to be a non-linear diffusion driven by Brownian motion. This paper addressed the pricing of derivatives written on these processes. Although it did not give an analytic solution to this problem, this paper is seminal for it stimulated the development of both finance and econometric work in this area. Immediate finance developments off this initial paper are Chesney and Scott (1989), Scott (1987), Wiggins (1987), Scott (1991), Stein and Stein (1991), Scott (1997), Heston (1993) and Duffie, Pan, and Singleton (2000). The papers by Heston, Duffie, Pan and Singleton are particularly worth emphasising for they work with CIR processes. We have seen that this is basically the only non-linear diffusion for which we are able to carry out analytic calculations. Hence this style of work is closest to our own on non-Gaussian OU processes.

Stochastic volatility models entered the econometrics literature with the paper by Melino and Turnbull (1990) but it was only with Harvey, Ruiz, and Shephard (1994) that the econometrics community became widely aware of this model. Since that time there have been enormous numbers of paper written on the estimation of discrete and continuous time SV models. A very limited list of relevant papers are (i) using Markov chain Monte Carlo by Jacquier, Polson, and Rossi (1994), Kim, Shephard, and Chib (1998) and Elerian, Chib, and Shephard (2001), (ii) using simulated method of moments by Duffie and Singleton (1993), (iii) indirect inference by Gourieroux, Monfort, and Renault (1993), (iv) efficient method of moments by Gallant and Tauchen (1996), (v) simulated likelihood methods by Sandmann and Koopman (1998).

Recently there has been developing a literature on the use of both option and underlying prices to estimate SV models. This is a very interesting area. Earlier work on this topic includes Renault (1997).

Multivariate models ???.

The first paper to deal with SV models built using non-Gaussian OU processes was Barndorff-Nielsen and Shephard (2001a), which developed the basic theory. This paper looked at some preliminary empirics and also derived some simple option pricing models. Some of these issues will be returned to in later chapters. The result on returning being mixing is derived in Genon-Catalot, Jeantheau, and Larédo (2000).

Chapter 6

Realised variation and covariation

Abstract: This Chapter sets out an analysis of realised variance and covariation, that is sums of high frequency squared or outer-producted returns. The probability limit of such statistics is well known using the theory of quadratic variation . We extend this to provide a distribution theory, which looks at the approximate distribution of the difference between realised covariation and quadratic covariation. This theory can be used to provide a distribution theory for regression and correlation between asset returns, as well as economically interesting objects like asset allocation weights and efficiency frontiers for investment decisions. Throughout the theory is illustrated by the application of the theory to financial data.

6.1 What is this Chapter about?

In this Chapter we provide an econometric framework for the recently introduced concepts of realised variance and covariation. These methods, which are based on extensions of quadratic variation and covariation, have become increasingly popular in the context of the advent of high frequency financial data. This framework allows us to have a better understanding of regression and correlation in the context of financial data.

This Chapter has ? other sections. They have the following goals.

- What are realised variances and realised covariation? We set out a framework for understanding their properties.
- A distribution theory for realised variance .
- Empirical examples of the distribution theory.
- Theory and proof of the asymptotic distribution theory for realised variance. This is a starred section.
- A distribution theory for realised covariation and consequently a distribution theory for realised regression and correlation.
- Empirical examples of the distribution theory.
- Theory and proof of the asymptotic distribution theory for realised covariation. This is a starred section.
- Using the theory to perform model based smoothing and forecasting.
- Draw conclusions from the Chapter.
- Discuss the literature on realised quantities.

This Chapter is followed by one of power variation, a generalisation of quadratic variation which looks at limiting properties of sums of powers of absolute high frequency returns. This work will build on the content of this Chapter.

6.2 What is realised variance and covariation?

6.2.1 Introduction

This Chapter analyses econometric strategies for measuring and predicting the variability of returns and the covariability between different asset returns. It is based on the innovative use of high frequency financial data yielding statistics called realised variance and realised covariation. Intimately related to ideas of quadratic variation and covariation for special semimartingales, the

analysis provides an asymptotic distribution theory (not just probability limit) based on a fixed interval of time (e.g. a trading day or a calendar month), allowing the number of high frequency returns during this period to go to infinity. The theory allows us to study how volatilities, covariances, correlations and regression coefficients change through time by computing these quantities over non-overlapping intervals of time and providing confidence limits for them.

The econometrics is motivated by the advent of complete records of quotes or transaction prices for many financial assets. Theoretical and empirical work suggests that the use of such high frequency data is both informative and simplifying for it brings us closer to the theoretical models based on continuous time. However, market microstructure effects (e.g. discreteness of prices, bid/ask bounce, irregular trading etc.) means that there is a mismatch between asset pricing theory based on semimartingales and the data at very fine time intervals. This suggests that we cannot simply rely on empirical computations based on literally infinitesimal returns, instead we need a distribution theory for these estimators. This theory will reflect the fact that we will use a large but finite number of high frequency returns in our empirical work, informing us of the difference between the empirical reality and the theoretical limit of using returns over tiny time intervals.

We suppose there are M intra-$\hbar$ observations during each $\hbar > 0$ time period and that log-price of an asset is written as y^*. Our approach is to think of M as large and increasing. It will drive our limiting theory. Then high frequency observations will be defined as

$$y_{j,i} = y^* \left((i-1)\hbar + \hbar j M^{-1} \right) - y^* \left((i-1)\hbar + \hbar (j-1) M^{-1} \right), \tag{6.1}$$

the j-th intra-$\hbar$ return for the i-th period (e.g. if $\hbar$ is a day, $M = 1440$, then this is the return for the j-th minute on the i-th day). Then our initial focus of attention will be on the *realised variance*

$$[y^*_M]_i = \sum_{j=1}^{M} y_{j,i}^2, \tag{6.2}$$

or the corresponding *realised volatility*

$$\sqrt{\sum_{j=1}^{M} y_{j,i}^2}. \tag{6.3}$$

These statistics measure the variability of the asset during the i-th period. At first sight they look rather odd, for unlike standard variance based measures they do not have their means removed and are not divide by M. The lack of a mean correction is due to it being of negligible importance for moderate values of M, while the lack of a division by M arises as the magnitude of the individual $y_{j,i}$ tends to get smaller with M as they are returns over intervals of length $\hbar M^{-1}$. These issues will become clearer in a moment. Before we proceed we should note that, rather confusingly from our viewpoint, some of the econometric literature on this topic calls $[y^*_M]_i$ the realised volatility. Throughout this book we have followed the finance literature in always calling standard deviation type objects volatility and hence we have called (6.3) the realised volatility and (6.2) the realised variance . This is the nomenclature used in the derivative markets based on these objects. Hopefully econometricians will not be too put off by this change in name.

Later in this Chapter attention will turn to the multivariate version of this, where y^* is a q-dimensional asset price and then we write the k-th element of the vector of returns $y_{j,i}$, defined in (6.1), as $y_{j,i(k)}$. The generalisation of the realised variance is the *realised covariation matrix*

$$[y^*_M]_i = \sum_{j=1}^{M} y_{j,i} y_{j,i}' = \left\{ \sum_{j=1}^{M} y_{j,i(k)} y_{j,i(l)} \right\}_{k,l=1,\ldots,q}, \tag{6.4}$$

which has the realised variances of the individual assets on its leading diagonal and the *realised covariances* on its off-diagonals. This matrix plays a central role in modern financial econometrics for simple functions of it deliver regression and correlation statistics between asset returns. In particular the *realised regression* of returns of asset l on returns of asset k is defined as

$$\widehat{\beta}_{i(lk)} = \frac{\sum y_{j,i(k)} y_{j,i(l)}}{\sum y_{j,i(k)}^2},$$

while the corresponding *realised correlation* is

$$\widehat{\rho}_{i(lk)} = \frac{\sum y_{j,i(k)} y_{j,i(l)}}{\sqrt{\sum y_{j,i(k)}^2 \sum y_{j,i(l)}^2}}.$$

6.2.2 Probability limits and semimartingales

Recall that in financial economics there is little lost in assuming that y^* is a special semimartingale. Under these conditions the probability limit of $[y_M^*]_i$ has been known for many years using the theory of quadratic covariation. Here we remind readers of the substance of that theory, before we go beyond this to develop the asymptotic distribution theory.

Recall a q dimensional special semimartingale y^* can be uniquely decomposed as

$$y^*(t) = \alpha^*(t) + m^*(t), \tag{6.5}$$

where $\alpha^*(t)$, a drift term, is a *predictable process* with *locally bounded variation* paths and $m^*(t)$ is a *local martingale*. One of the most important aspects of semimartingales is the *quadratic covariation* (QV) defined as

$$[y^*](t) = \text{p}-\lim_{M\to\infty} \sum_{j=0}^{M-1} \{y^*(t_{j+1}) - y^*(t_j)\}\{y^*(t_{j+1}) - y^*(t_j)\}', \tag{6.6}$$

for any sequence of partitions $t_0 = 0 < t_1 < ... < t_M = t$ with $\sup_j\{t_{j+1} - t_j\} \to 0$ for $M \to \infty$. Here $\text{p}-\lim$ denotes the probability limit of the sum. Later it will be helpful to label the k, l-th element of the QV $[y^*](t)$ as the *quadratic covariance*

$$\left[y_{(k)}^*, y_{(l)}^*\right](t) = \text{p}-\lim_{M\to\infty} \sum_{j=0}^{M-1} \{y_k^*(t_{j+1}) - y_k^*(t_j)\}\{y_l^*(t_{j+1}) - y_l^*(t_j)\}, \tag{6.7}$$

while the k-th diagonal element of $[y^*](t)$ is the *quadratic variation*

$$[y_{(k)}^*](t) = \text{p}-\lim_{M\to\infty} \sum_{j=0}^{M-1} \{y_k^*(t_{j+1}) - y_k^*(t_j)\}^2.$$

In general

$$[y^*](t) = [y^{*c}](t) + \sum_{0\le s\le t} \Delta y^*(s)\Delta y^*(s)', \tag{6.8}$$

where y^{*c} is the continuous component of y^* and $\Delta y^*(t) = y^*(t) - y^*(t-)$ are the jumps at time t. In the context of the special semimartingale (A.2) the QV becomes

$$\begin{aligned} [y^*](t) &= [m^*](t) + \sum_{0\le s\le t} \Delta\alpha^*(s)\Delta\alpha^*(s)' + \sum_{0\le s\le t} \Delta m^*(s)\Delta\alpha^*(s)' && (6.9)\\ &\quad + \sum_{0\le s\le t} \Delta\alpha^*(s)\Delta m^*(s)', && (6.10) \end{aligned}$$

the QV of m^* plus terms which are influenced by the jumps in α^* and m^*. If α^* is continuous then we obtain the simplification

$$[y^*](t) = [m^*](t), \tag{6.11}$$

irrespective of the presence of jumps in the local martingale component. This holds as the quadratic variation of any continuous, locally bounded variation process is zero. The result (A.10) is powerful for it does not depend upon the model for m^* or α^*, only on the assumption that α^* is continuous.

If we return to the cases of realised covariances and realised volatilities, we have under special semimartingales that as $M \to \infty$

$$[y_M^*]_i = \sum_{j=1}^{M} y_{j,i} y_{j,i}' \xrightarrow{p} [y^*]_i ,$$

where

$$[y^*]_i = [y^*](\hbar i) - [y^*](\hbar (i-1) -).$$

This implies, in particular,

$$\sqrt{\sum_{j=1}^{M} y_{j,i}^2} \xrightarrow{p} \sqrt{[y^*]_i}.$$

Likewise

$$\widehat{\beta}_{i(lk)} = \frac{\sum y_{j,i(k)} y_{j,i(l)}}{\sum y_{j,i(k)}^2} \xrightarrow{p} \frac{\left[y_{(k)}^*, y_{(l)}^*\right]_i}{\left[y_{(k)}^*\right]_i},$$

and

$$\widehat{\rho}_{i(lk)} = \frac{\sum y_{j,i(k)} y_{j,i(l)}}{\sqrt{\sum y_{j,i(k)}^2 \sum y_{j,i(l)}^2}} \xrightarrow{p} \frac{\left[y_{(k)}^*, y_{(l)}^*\right]_i}{\sqrt{\left[y_{(k)}^*\right]_i \left[y_{(l)}^*\right]_i}}.$$

These are entirely general results and cover all the interesting cases in financial economics.

6.2.3 A stochastic volatility model

The above theoretical framework is too general for us to be able to derive a distribution theory for

$$\sum_{j=1}^{M} y_{j,i} y_{j,i}' - [y^*]_i .$$

This is essential in order to be able to construct confidence intervals on realised quantities such as realised regression and correlation. As a result we have had to specialise. In particular we make three assumptions in our treatment of the univariate and multivariate cases

(**A**). That m^* is a multivariate stochastic volatility (SV) process

$$m^*(t) = \int_0^t \Xi(u) \mathrm{d}w(u), \tag{6.12}$$

where $\Xi(t)$, is the *instantaneous* or *spot covolatility matrix*, and w is standard multivariate Brownian motion. We write

$$\Sigma(t) = \Xi(t)\Xi(t)',$$

the *spot covariance matrix* process. By construction $\Sigma(t)$ is positive semi-definite for all values of t while we assume that each element of Σ is a local bounded variation process.

Whatever the model for Σ the implied m^* must be a continuous local martingale. In what follows we will write

$$\Sigma^*(t) = \int_0^t \Sigma(u)\mathrm{d}u$$

the integrated covariance matrix. In the univariate case we will write the volatility in the familiar way as $\sigma(t) = \Xi(t)$, while the variance is

$$\tau(t) = \sigma^2(t) = \Sigma(t),$$

and the integrated variance is again defined as

$$\tau^*(t) = \int_0^t \tau(u)\mathrm{d}u = \int_0^t \sigma^2(u)\mathrm{d}u. \tag{6.13}$$

(**B**). For every $\nu = 1, ..., q$ the mean process α_ν^* is càglàd and satisfies (pathwise)

$$\delta^{-3/4} \max_{1\leq j\leq M} |\alpha_\nu^*(j\delta) - \alpha_\nu^*((j-1)\delta)| = o(1), \tag{6.14}$$

in $\delta \downarrow 0$. This condition implies that the α^* process is continuous.

(**C**). The joint α^*, Σ process is independent of w. This is a strong, undesirable additional assumption for it rules out empirically important dynamic effects such as leverage. We will comment in detail about the import of this assumption later.

Conditions (A) and (B) imply this model structure is a continuous, special semimartingale with

$$y^*(t) = \alpha^*(t) + \int_0^t \Xi(u)\mathrm{d}w(u). \tag{6.15}$$

Example 15 *A general α^* process which satisfies (6.14) is where*

$$\alpha^*(t) = \mu t + \Sigma^*(t)\beta.$$

This links the mean process to the covariance. It has the feature that for all possible models for Σ,

$$\frac{\partial \alpha^*(t)}{\partial t} = \mu + \Sigma(t)\beta.$$

*Differentiability of α^* is a more restrictive than assumption (**B**).*

Importantly, for this model class

$$[y^*](t) = [m^*](t) = \Sigma^*(t),$$

due to condition (**B**) ruling out the possibility of jumps in α^* and (6.12) having continuous sample paths. This implies

$$[y_M^*]_i \xrightarrow{p} \Sigma_i,$$

where

$$\Sigma_i = \Sigma^*(i\hbar) - \Sigma^*\left\{(i-1)\hbar\right\}.$$

We call Σ_i the *actual or notional covariance matrix*. It plays a central role in the probabilistic analysis of SV models. In the univariate case we write

$$\tau_i = \Sigma_i = \tau^*(i\hbar) - \tau^*\left\{(i-1)\hbar\right\},$$

the actual or notional variance.

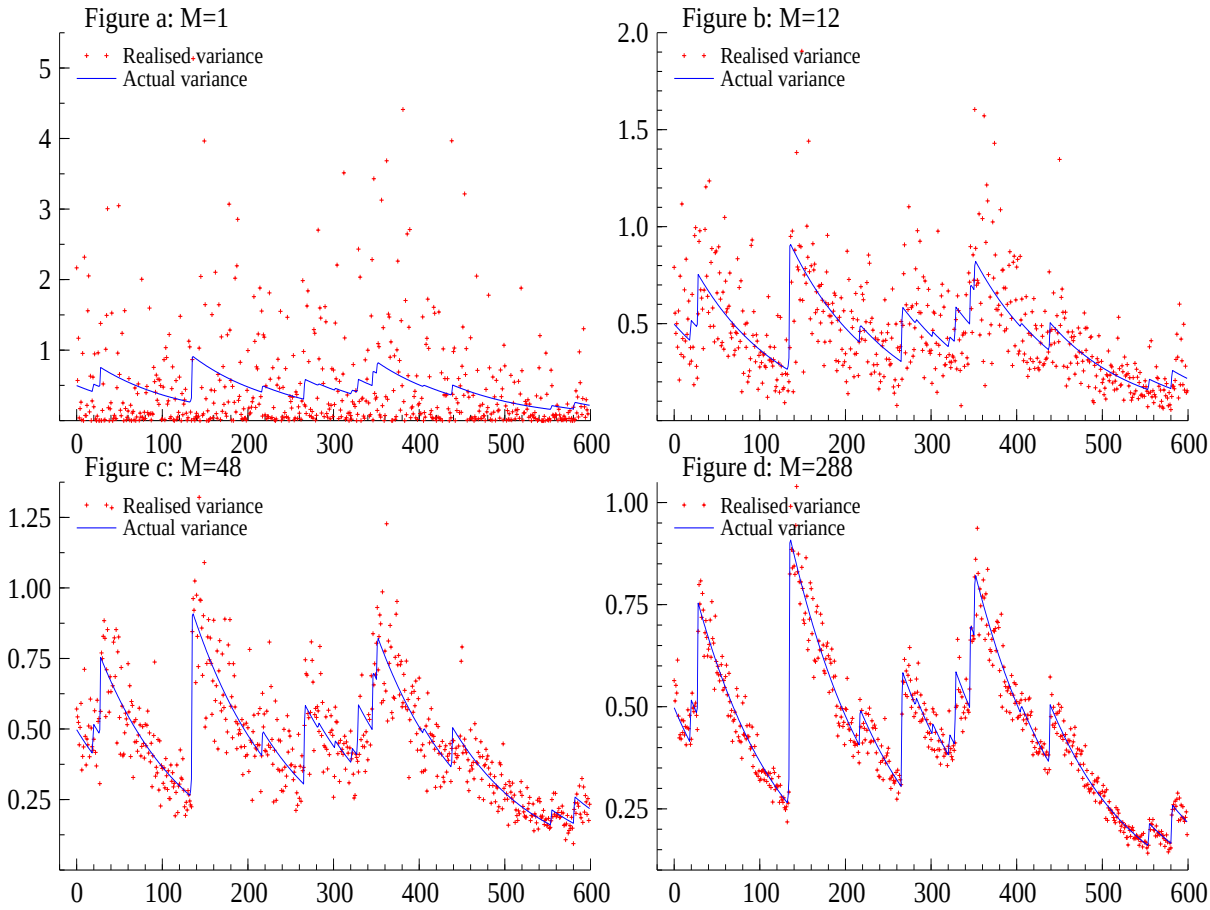

Figure 6.1: *Actual* τ_i *and realised* $\sum_{j=1}^{M} y_{j,i}^2$ *(with* M varying*) volatility based upon a* $\Gamma(4,8)$-OU process with $\lambda = -\log(0.99)$ and $\Delta = 1$. This implies $\xi = 0.5$ and $\xi\omega^{-2} = 8$. Code: `/code/realised/simple.ox`.

Example 16 *Figure 6.1 displays a simulated sample path of integrated variance* τ_i *from an OU process given by the solution to*

$$\mathrm{d}\tau(t) = -\lambda\tau(t)\mathrm{d}t + \mathrm{d}z(\lambda t),$$

where z *is a subordinator. In this example we construct the process so that* $\tau(t)$ *has a* $\Gamma(4,8)$ *stationary distribution,* $\lambda = -\log(0.99)$ *and* $\hbar = 1$. *Also drawn are the sample path of the realised variances* $\sum_{j=1}^{M} y_{j,i}^2$ *(depicted using crosses) where*

$$y^*(t) = \beta\tau^*(t) + \int_0^t \tau^{1/2}(u)\mathrm{d}w(u),$$

and $\beta = 0.5$ *(a very large risk premium). The realised variances are computed using a variety of values of* M. *We see that as* M *increases the size of* $\sum_{j=1}^{M} y_{j,i}^2 - \tau_i$ *falls, illustrating the consistency of* $\sum_{j=1}^{M} y_{j,i}^2$ *for* τ_i *even though* β *is not zero. Further experiments suggest and theory says that the effect of* β *is very small even with quite moderate values of* M*, although its effect is discernible when* $M = 1$.

The quadratic variation result implies that

$$\frac{\partial [y^*](t)}{\partial t} = \Sigma(t),$$

which means in this context we can view the history of the Σ process as observable given the paths of y^*. This implies that $\Sigma(t)\mathrm{d}t$ is the conditional (given the natural filtration) covariance matrix of the infinitesimal return $\mathrm{d}y^*(t)$. As a result $\Sigma^*(t)$ is the integrated conditional covariance matrix, while $\alpha^*(t)$ has the interpretation as the integrated conditional mean process.

6.3 Asymptotic distribution of realised variance

6.3.1 Results and comments

In this section we will specialise the notation to the univariate case, delaying the discussion of the multivariate case for a couple of sections. In particular we will work in terms of volatility, variance and integrated variance processes. We will review results on this topic, giving an intuitive understanding of them and illustrate them on Monte Carlo and real data. The next section will then give a formal proof the result. This is a starred section and so can be skipped at first reading without the loss of the thread of the book for those readers put off by its higher mathematical level.

In the special case of y^* being univariate

$$y^*(t) = \alpha^*(t) + \int_0^t \sigma(u)\mathrm{d}w(u) \tag{6.16}$$

the following three results hold under assumptions (A)-(C). The first result is that as $M \to \infty$ so, recalling $\tau(t) = \sigma^2(t)$,

$$\frac{\sqrt{\frac{M}{\hbar}}\left\{\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u\right\}}{\sqrt{2\int_{(i-1)\hbar}^{i\hbar} \tau^2(u)\mathrm{d}u}} \xrightarrow{L} N(0,1). \tag{6.17}$$

The second result is that

$$\frac{\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u}{\sqrt{\frac{2}{3}\sum_{j=1}^{M} y_{j,i}^4}} \xrightarrow{L} N(0,1). \tag{6.18}$$

These two limit theorems are linked together by the third result which is that

$$\frac{M}{3\hbar}\sum_{j=1}^{M} y_{j,i}^4 \xrightarrow{p} \int_{(i-1)\hbar}^{i\hbar} \tau^2(u)\mathrm{d}u. \tag{6.19}$$

The result (6.18) is statistically feasible, while (6.17) is perhaps more informative from a theoretical viewpoint. In particular the two results imply:

- $\sum_{j=1}^{M} y_{j,i}^2$ converges to $\int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$ at rate $\sqrt{M}$. This considerably strengthens the QV result, for now we know the rate of convergence, not just that it converges.
- The limit theorem is unaffected by the form of the drift process α^*, smoothness condition (6.14) is sufficient that its effect becomes negligible. Again this considerably strengthens the QV result which says the p-lim is unaffected by the drift. Now we know this result extends to the next order term as well.

- Knowledge of the form of the volatility dynamics is not required in order to use this theory.
- The fourth moment of returns need not exist for the asymptotic normality to hold. In such heavy tailed situations, the stochastic denominator $\int_{(i-1)\hbar}^{i\hbar} \tau^2(u)\mathrm{d}u$ loses its unconditional mean. However, this property is irrelevant to the workings of the theory.
- The volatility process $\tau(t)$ can be non-stationary, exhibit long-memory or include intra-day effects.
- $\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$ has a mixed Gaussian limit implying that marginally it will have heavier tails than a normal.
- The magnitude of the error $\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$ is likely to be large in times of high volatility.
- Conditionally on $\int_{\hbar(i-1)}^{\hbar i} \tau^2(u)\mathrm{d}u$ and $\int_{\hbar(k-1)}^{\hbar k} \tau^2(u)\mathrm{d}u$, the errors

$$\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u \qquad \text{and} \qquad \sum_{j=1}^{M} y_{j,k}^2 - \int_{\hbar(k-1)}^{\hbar k} \tau(u)\mathrm{d}u$$

  are asymptotically independent and jointly normal for $i \neq k$.
- Some of the features of (6.17) appear in the usual cross-section asymptotic theory of the estimation of σ^2 when $z_i \sim NID(0, \sigma^2)$. Then

$$\frac{\sqrt{M}\left\{\frac{1}{M}\sum_{j=1}^{M} z_i^2 - \sigma^2\right\}}{\sqrt{2\sigma^4}} \xrightarrow{L} N(0,1),$$

  whose natural feasible version is

$$\frac{\sqrt{M}\left\{\frac{1}{M}\sum_{j=1}^{M} z_i^2 - \sigma^2\right\}}{\sqrt{\frac{2}{3M}\sum_{j=1}^{M} z_i^4}} \xrightarrow{L} N(0,1).$$

  This has quite a few differences from (6.18). In particular the denominator divides by M rather than multiplies by M, while in the numerator $\sum_{j=1}^{M} z_i^2$ is divided by M where as in the theory for realised variance it is left unscaled.
- These results are also quite closely related to some work on the asymptotic distribution theory for an estimator of $\Sigma(t)$, the spot (not integrated) variance. The idea there is to compute a local variance from the lagged data, e.g.,

$$\widehat{\Sigma}(t) = \hbar^{-1}\sum_{j=1}^{M}\left\{y^*\left(t - \hbar j M^{-1}\right) - y^*\left(t - \hbar\left(j-1\right)M^{-1}\right)\right\}^2. \tag{6.20}$$

  They the behaviour of this estimator can then studied as $M \to \infty$ and $\hbar \downarrow 0$ under some assumptions. This "double asymptotics" yields a Gaussian limit theory so long as $\hbar \downarrow 0$ and $M \to \infty$ at the right, related rates. The double asymptotics makes it harder to use in practice than our own simpler analysis, which just needs $M \to \infty$. This is made possible because our goal is to estimate the easier integrated covariation rather than the harder spot covariation.

6.3.2 Intuition about the result

The proof of the result spells out the details of why (6.18) and (6.17) hold. Here we build some intuition for the result in the case where the drift process α^* is set to zero. This maybe helpful to readers before delving into the proof or for readers who do not want to read the proof. As we can see from the result, it holds for all values of i. In order to simplify the notation in the exposition it is helpful to set $i = 1$ and drop reference to that subscript.

We start with

$$y^*(t) = \int_0^t \sigma(u)\mathrm{d}w(u),$$

where σ and w are independent processes. This implies

$$y_j|\tau_j \sim N(\alpha_j, \tau_j),$$

where

$$\tau_j = \tau^*\left(\hbar j M^{-1}\right) - \tau^*\left(\hbar\left(j-1\right)M^{-1}\right), \tag{6.21}$$

the high-frequency increment to integrated variance. Conditional on the path of the variance process τ

$$u = \sum_{j=1}^{M} y_j^2 - \int_0^{\hbar} \tau(u)\mathrm{d}u \overset{\mathcal{L}}{=} \sum_{j=1}^{M} \tau_j\left(\varepsilon_j^2 - 1\right),$$

where $\varepsilon_j \overset{i.i.d.}{\sim} N(0,1)$. Thus u, the realised variance error, is a mixture of weighted centred chi-squared variables. The terms in the sum $\tau_j\left(\varepsilon_j^2 - 1\right)$ are zero means and independent, conditional on the weights. If the weights do not trend upwards or collapse to zero then we might expect $\sqrt{M}u$ to be roughly Gaussian with a mean of zero and variance of

$$2M\sum_{j=1}^{M}\tau_j^2.$$

For large M the locally bounded variation assumption on τ implies, writing

$$\overline{\tau_j} = \frac{M}{\hbar}\int_{\hbar(j-1)M^{-1}}^{\hbar j M^{-1}} \tau(u)\mathrm{d}u,$$

that

$$\begin{array}{rcl} \frac{M}{\hbar}\sum_{j=1}^{M}\tau_j^2 & = & \frac{\hbar}{M}\sum_{j=1}^{M}\overline{\tau_j}^2 \\ & \overset{p}{\to} & \hbar\int_0^{\hbar}\tau^2(u)\mathrm{d}u. \end{array}$$

The theory and proof firms up these approximations, yielding the infeasible limit result given in (6.17). As this result has a limiting distribution which does not depend upon Σ, it holds unconditionally as well as conditionally on τ.

The step to making the result feasible is to prove (6.19). However,

$$\begin{array}{rcl} \frac{M}{3\hbar}\sum_{j=1}^{M} y_j^4 & \overset{\mathcal{L}}{=} & \frac{\hbar}{3M}\sum_{j=1}^{M}\overline{\tau_j}^2\varepsilon_j^4 \\ & \overset{p}{\to} & \hbar\int_0^{\hbar}\tau^2(u)\mathrm{d}u. \end{array}$$

We call

$$\frac{M}{3\hbar}\sum_{j=1}^{M} y_j^4 \tag{6.22}$$

the *quarticity* of the high frequency data. The application of Slutsky's theorem yields the desired feasible limit theory (6.18).

6.3.3 Asymptotically equivalent results

The quarticity is not the only consistent estimator of $\int_0^\hbar \tau^2(u)\mathrm{d}u$, although in practice we have found it to be the most accurate of all the different options we have considered. In particular

$$\begin{array}{rcl} \frac{M}{\hbar}\sum_{j=2}^{M} y_j^2 y_{j+1}^2 & \stackrel{L}{=} & \frac{\hbar}{M}\sum_{j=1}^{M-1}\left(\overline{\tau}_j\right)\left(\overline{\tau}_{j+1}\right)\varepsilon_j^2\varepsilon_{j+1}^2 \\ & \stackrel{p}{\to} & \hbar\int_0^\hbar \tau^2(u)\mathrm{d}u. \end{array}$$

Lagging by a single time unit is not particularly important here, some other small lag could have been used. This implies the non-negative estimator

$$\frac{M}{\hbar}\sum_{j=1}^{M} y_j^4 - \frac{M}{\hbar}\sum_{j=1}^{M-1} y_j^2 y_{j+1}^2 \stackrel{p}{\to} 2\hbar\int_0^\hbar \tau^2(u)\mathrm{d}u,$$

and so provides an alternative denominator in the infeasible limit theory (6.17). In particular this delivers the feasible theories

$$\frac{\sum_{j=1}^{M} y_j^2 - \int_0^\hbar \tau(u)\mathrm{d}u}{\sqrt{\sum_{j=1}^{M} y_j^4 - \sum_{j=1}^{M-1} y_j^2 y_{j+1}^2}} \stackrel{L}{\to} N(0,1),$$

and

$$\frac{\sum_{j=1}^{M} y_j^2 - \int_0^\hbar \tau(u)\mathrm{d}u}{\sqrt{2\sum_{j=1}^{M-1} y_j^2 y_{j+1}^2}} \stackrel{L}{\to} N(0,1).$$

The latter is interesting for it avoids the use of fourth moments of the data in the denominator.

6.3.4 Log transforms and realised volatilities

The basic limit theory results (6.17) and (6.18) can be embellished in a number of ways. One approach, which Monte Carlo experiments suggest improves the finite sample behaviour of asymptotic approximation, is to take a logarithmic transform. A straightforward application of the delta-method[1] yields the infeasible limit theory

$$\frac{\sqrt{\frac{M}{\hbar}}\left\{\log\sum_{j=1}^{M} y_j^2 - \log\int_0^\hbar \tau(u)\mathrm{d}u\right\}}{\sqrt{2\int_0^\hbar \tau^2(u)\mathrm{d}u \Big/ \left(\int_0^\hbar \tau(u)\mathrm{d}u\right)^2}} \stackrel{\mathcal{L}}{\to} N(0,1).$$

The denominator is invariant to scaling the returns so we might expect the denominator not to vary so much through time even when there is volatility clustering. In practice we have to replace the unobserved denominator by an estimator, yielding the feasible approximation

$$\frac{\log\sum_{j=1}^{M} y_j^2 - \log\int_0^\hbar \tau(u)\mathrm{d}u}{\sqrt{2\sum_{j=1}^{M} y_j^4 \Big/ \left(\sum_{j=1}^{M} y_j^2\right)^2}} \stackrel{\mathcal{L}}{\to} N(0,1). \tag{6.23}$$

Confidence limits for $\sum_{j=1}^{M} y_j^2$ based on this theory will be non-symmetric due to the curvature of the log-function.

Example 17 *We continue with the simulation from Example 16 which was based on an SV model with a $\Gamma(4,8)$-OU process for τ. Based on a sample of 12,000 days, we study the performance of the asymptotic theory based on the original feasible version (6.18) and the log-version*

[1] This is based on approximating $\log x$ by

$$\log\mu_x + \frac{x-\mu_x}{\mu_x},$$

where μ_x is the p-lims of x.

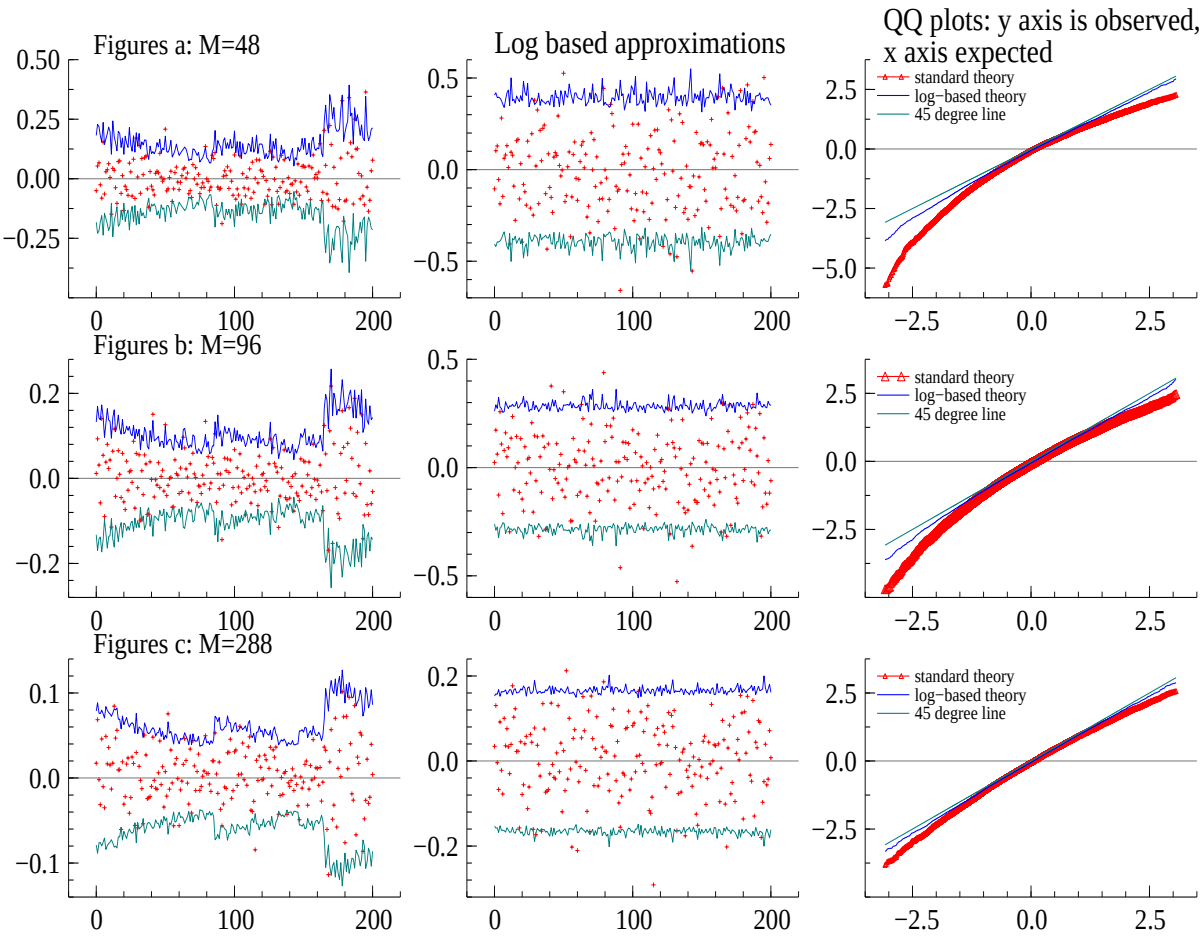

Figure 6.2: *Left graphs: Actual* $\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)du$ *plotted again* i *and twice asymptotic S.E.s. Middle graphs:* $\log \sum_{j=1}^{M} y_{j,i}^2 - \log \int_{\hbar(i-1)}^{\hbar i} \tau(u)du$ *plotted against* i *and twice asymptotic S.E.s. Right graphs: QQ plot of the standardised realised volatility error (X-axis has the expected quantiles, Y-axis the observed). Code:* `simple.ox`.

(6.23). Column one of Figure 6.2 the time series plot of $\sum_{j=1}^{M} y_{j,i}^2 - \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$ *against* i*, given by crosses, together with their 95% confidence intervals based on using two asymptotic standard deviations. These are computed using* $M = 48$*,* 96 *and* 288*. The pictures show a number of features. First, as* M *increases so the confidence intervals shrink. More interestingly, the size of the intervals vary dramatically through time. This was predicted by the theory, but the practical implication is clear that they vary considerably not just in theory. The second column repeats these experiments but based on the log-theory. Now the crosses depict* $\log \sum_{j=1}^{M} y_{j,i}^2 - \log \int_{\hbar(i-1)}^{\hbar i} \tau(u)\mathrm{d}u$*. The confidence limits are now almost constant through time. This holds for any value of* M*. This implies that on the log-scale the realised variance error is approximately Gaussian. Finally, the third column displays the QQ plots of standardised errors (that is it plots the sorted standardised residuals against the expected quantile from the normal distribution). By standardised errors we mean the left hand side of (6.18) and (6.23). This assesses the Gaussianity of the finite sample distribution and so the performance of the asymptotic distribution. The graphs suggest both the original and log-version have a long left hand tail, but the log-version is much more precise.*

The above theory can be used to provide confidence limits for the realised volatilities

$$\sqrt{\sum_{j=1}^{M} y_j^2},$$

by just square rooting the confidence limits for the realised variance . It is of some interest to have a limiting theory directly in terms of the realised volatility however. The resulting theory, again based on the delta method[2] has

$$\frac{\sqrt{\frac{M}{\hbar}}\left\{\sqrt{\sum_{j=1}^{M} y_j^2} - \sqrt{\int_0^\hbar \tau(u)\mathrm{d}u}\right\}}{\sqrt{2\int_0^\hbar \tau^2(u)\mathrm{d}u / \left(\int_0^\hbar \tau(u)\mathrm{d}u\right)}} \xrightarrow{\mathcal{L}} N(0,1).$$

Here the denominator is not scale invariant.

6.4 Empirical examples of realised volatilities

6.4.1 A time series of daily realised volatilities

Figure 6.3(a) displays the first 50 actively traded days after 1st December 1986 of the stochastically interpolated exchange rates for the Dollar/DM and Dollar/Yen Olsen database. The DM series is given by the crosses, while the Yen series is drawn using a line. The fall in both of the series corresponds to a fall in the value of the Dollar during this two month period. Figure 6.3(b) displays crosses to denote the realised volatility $\sqrt{\sum_{j=1}^{M} y_{j,i}^2}$ for the DM series for each of the 50 days using M=144, which corresponds to utilising 10 minute returns. The corresponding bars are the 95 confidence intervals generated from the log-based asymptotic limit theory given in (6.23). These show the important widening and closing of the 95% confidence intervals, with the intervals seemingly being very large when the volatility is high. The picture shows us that the volatility of the exchange changes has move statistically significantly through time with periods of high volatility. Figure 6.3(c) displays the corresponding results for the Yen series, which has a consistently slightly lower level of volatility than the DM series but with common overall effects. Finally, Figure 6.3(d) shows a cross-plot of the realised volatility for the DM series against the realised volatility for the Yen series. This shows that these two quantities are quite closely related, which is not a surprise as they have a common numeraire.

6.4.2 A time series of annual realised volatilities

One use of the asymptotics for realised variances and volatilities is to compute confidence intervals for low frequency data such as annual measures of volatility. Here the high frequency data would be daily observations and our goal in this subsection is to work with realised volatilities

$$\sqrt{\sum_{j=1}^{M} y_{j,i}^2},$$

that is the square root of realised variances . Such historical time series are very common in financial economics.

[2]This is based on approximating $\sqrt{x}$ by

$$\sqrt{\mu_x} + \frac{x - \mu_x}{\sqrt{\mu_x}},$$

where μ_x is the p-lims of x.

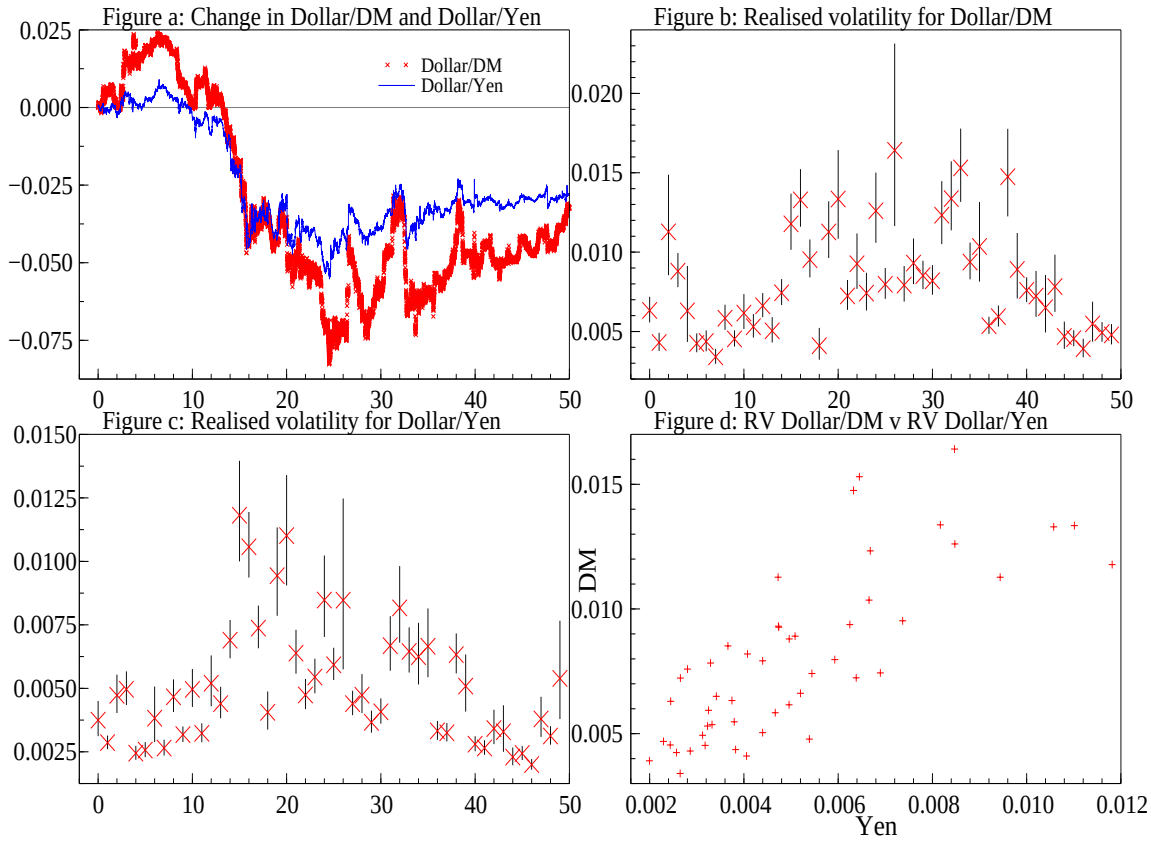

Figure 6.3: *Analysis of two exchange rates Dollar/DM and Dollar/Yen based on the Olsen database which records the rate every five minutes. X-axis is marked in days from 1st December 1986. Figure (a) displays the movement in the log exchange rates since the start of the sample marked off in days. Figure (b) shows the daily realised volatility* $\sqrt{\sum_{j=1}^{M} y_{j,i}^2}$ *for the DM series computed using M=144 as well as the associate 95% confidence interval. Figure (c) gives the corresponding result for the Yen series. Figure (d) plots the daily realised volatilities of the DM series again the corresponding realised volatilities for the Yen series. Code:* `daily_CI.ox`.

In this subsection we take a long series on the closing prices on the Dow Jones Industrial Average, starting on 26th May 1896 and going up to 31st December 2001. This is taken from the Dow Jones website. This is a narrower index than some of the more widely used series discussed in the literature, but it has the advantage of being in the public domain.

This series has a small number of recording breaks, which we have ignored as they make no substantial difference to our analysis. The series has the interesting feature that in the early part of it the markets were open six days a week, while in more recent years this has reduced to five. Of course this makes no difference to the implementation of our theory.

There is a very substantial break from 30th July 1914 until 31st December 1914. This was caused by the start of World War I, with Germany declaring war on Russia on 1st August 1914. This creates some important difficulties for the index was at 71.42 when it closed, while it reopened at 54 after Christmas in 1914. If we ignore this break, it will imply a very high level of volatility for 1914 due to the massive movement in the index.

To reconstruct the missing data we link together the log-prices from the start of August to Christmas using a realisation from a simulated Brownian bridge. The result is shown in Figure 6.4. There is only a single parameter in the linking, the variance of the Brownian motion. This is

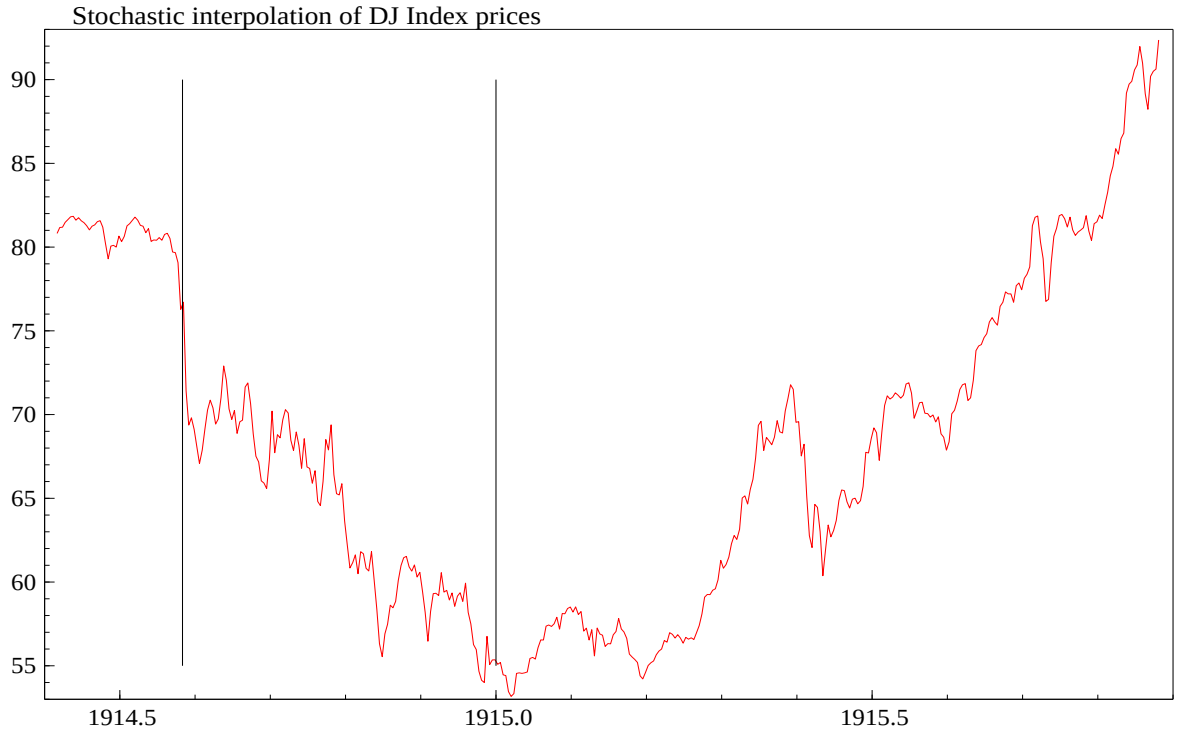

Figure 6.4: *Between the vertical lines prices are interpolated using a Brownian bridge (on the log scale). Code:* `schwert.ox`.

chosen a priori as 0.04/110 per day, which gives a standard deviation of yearly price movements of around 0.33. This is historically moderately high, reflecting the uncertainty of the period. The results we give below are not very sensitive to this choice for we will see 1914 is not a particularly volatile year in this dataset.

The realised volatilities and their 95% confidence intervals are given in Figure 6.5. The confidence intervals use the log-based limit theory given in (6.23). The results again reflect the tendency for the intervals to be wide when the level of volatility is high. However, the results are more varied in this case than in the high frequency analysis we gave for the exchange rate data. In particular the volatility spike in 1987 is poorly measured for it is caused by high levels of price movements over a very short time interval. There is not enough data in the daily observations to pin down precisely the level of volatility in this case. In the 1930s, on the other hand, the high level of movements was sustained over a long time interval and so we produce quite a precise estimate of the level of volatility.

6.5 Theory and proof of asymptotics for realised variance *

6.5.1 A theory and a lemma

In this starred section we give a formal treatment of the asymptotics of the realised variance, including giving proofs. It can be skipped on first reading without losing the thread of the book. Recall that $\delta M = \hbar$ and that for the processes y^* and τ we use the notation

$$y_j = y^*(j\delta\hbar) - y^*((j-1)\delta\hbar)$$

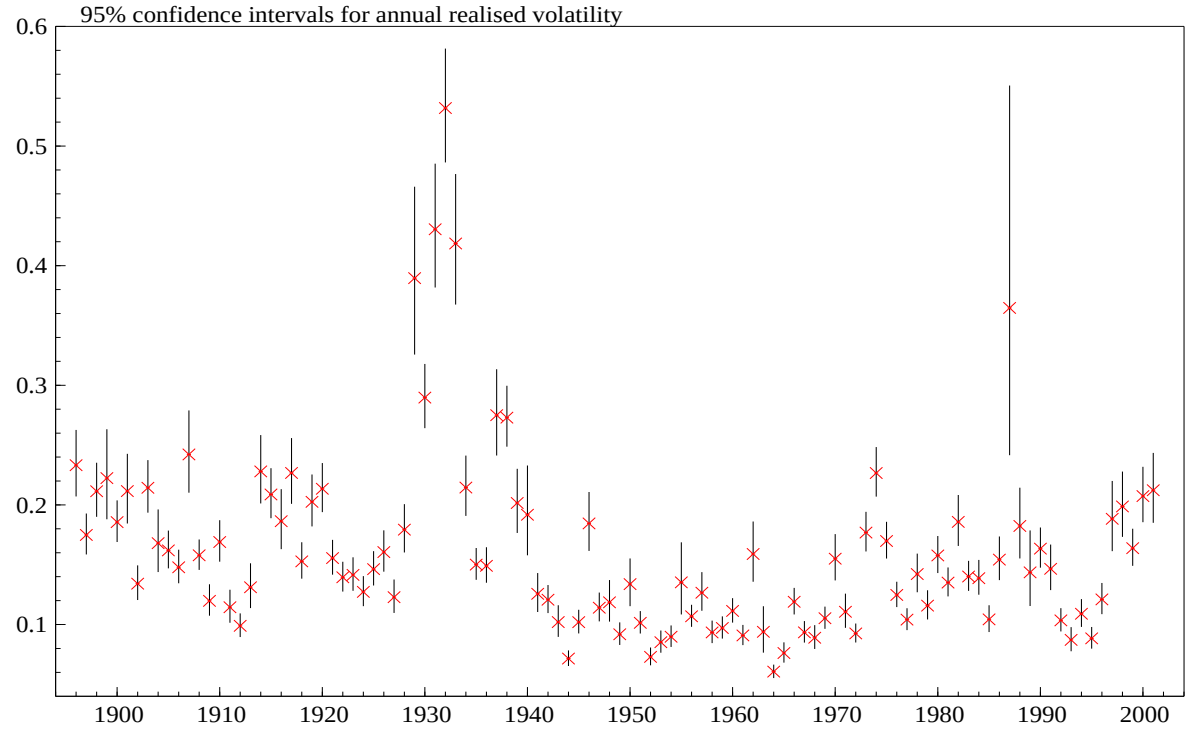

Figure 6.5: *Annual realised volatility* $\sqrt{\sum_{j=1}^{M} y_{j,n}^2}$ *for the Dow Jones Industrial Average (marked with crosses) together with 95% confidence intervals. Code:* `schwert.ox`.

and

$$\tau^{[r]*}(t) = \int_0^t \tau^r(s)\mathrm{d}s \quad \text{and} \quad \tau_j = \tau^*(j\delta\hbar) - \tau^*((j-1)\delta\hbar),$$

where $\tau^*(t) = \tau^{[1]*}(t)$. We shall also use

$$\alpha_j = \alpha^*(j\delta\hbar) - \alpha^*((j-1)\delta\hbar).$$

Lemma 1. *Assume that* $\tau(t)$ *is pathwise locally bounded and Riemann integrable*[34]*. Then, for* $M \to \infty$ *and* $r > 0$

$$\left(\frac{M}{\hbar}\right)^{r-1} \sum_{j=1}^{M} \tau_j^r \xrightarrow{a.s.} \int_0^{\hbar} \tau^r(s)\mathrm{d}s \tag{6.24}$$

Theorem 1. For the SV model in (6.16), i.e.

$$y^*(t) = \alpha^*(t) + \int_0^t \sigma(u)\mathrm{d}w(u)$$

[3] In other words, locally bounded and Riemann integrable with probability 1

[4] A function f on an interval I of R is said to be locally bounded and Riemann integrable if it is bounded and Riemann integrable on every closed subinterval of I. This is the case, in particular, if f is continuous or if it is of local bounded variation. The latter follows from the fact that a bounded function f is Riemann integrable on an interval $[0, t]$ if and only if the set of discontinuity points of f has Lebesgue measure 0 (see Hobson (1927, pp. 465–466), Munroe (1953, p. 174, Theorem 24.4) or Lebesgue (1902)). Any function of bounded variation is the difference between an increasing and a decreasing function and any monotone function has at most countably many discontinuities.

with $(\alpha^*, \tau) \perp\!\!\!\perp w$, suppose the volatility process $\tau = \sigma^2$ is locally bounded and Riemann integrable and that, for any positive $\hbar$, the mean process α^* satisfies condition (B) of Section 2.3, i.e.

$$\delta^{-3/4} \max_{1 \le j \le M} |\alpha^*(j\delta\hbar) - \alpha^*((j-1)\delta\hbar)| = o(\delta^{3/4}) \tag{6.25}$$

pathwise as $\delta \downarrow 0$. Then, for $M = \delta^{-1} \to \infty$,

$$\frac{\sum_{j=1}^M y_j^2 - \int_0^\hbar \tau(u)\mathrm{d}u}{\sqrt{2\sum_{j=1}^M \tau_j^2}} \xrightarrow{\mathcal{L}} N(0,1), \tag{6.26}$$

where

$$\tau_j = \tau^*\left((j\delta\hbar) - \tau^*\left((j-1)\delta\hbar\right)\right. . \tag{6.27}$$

Furthermore,

$$\frac{M}{\hbar} \sum_{j=1}^M \tau_j^2 \xrightarrow{a.s.} \int_0^\hbar \tau^2(s)\mathrm{d}s. \tag{6.28}$$

In particular, then, the limiting law of $\sqrt{M}\left(\sum_{j=1}^M y_j^2 - \int_0^\hbar \tau(u)\mathrm{d}u\right)$ is a normal variance mixture.

Lemma 1. *Assume that $\tau(t)$ locally bounded and Riemann integrable*[5]*. Then, for $M \to \infty$ and r* a positive integer*,*

$$\left(\frac{M}{\hbar}\right)^{r-1} \sum_{j=1}^M \tau_j^r \xrightarrow{a.s.} \int_0^\hbar \tau^r(s)\mathrm{d}s \tag{6.29}$$

6.5.2 Proofs

Proof of Lemma 1. By the definition of τ_j, for every j there exists a c_j such that

$$\inf_{(j-1)M^{-1}\hbar \le s \le jM^{-1}\hbar} \tau(s) \le c_j \le \sup_{(j-1)M^{-1}\hbar \le s \le jM^{-1}\hbar} \tau(s)$$

and

$$\tau_j = c_j \frac{\hbar}{M}. \tag{6.30}$$

Since τ is locally bounded and Riemann integrable the same is true of τ^r for any $r > 0$. Consequently

$$\left(\frac{M}{\hbar}\right)^{r-1} \sum_{j=1}^M \tau_j^r = \sum_{j=1}^M c_j^r \frac{\hbar}{M} \to \int_0^\hbar \tau^r(s)\mathrm{d}s = \tau^{r*}(\hbar).$$

Proof of Theorem 1. Note first that (6.28) follows from Lemma 1.

Next, let

$$u = \sum_{j=1}^M y_j^2 - \tau^*(\hbar)$$

[5] A function f on an interval I of R is said to be locally bounded and Riemann integrable if it is bounded and Riemann integrable on every closed subinterval of I. This is the case, in particular, if f is continuous or if it is of local bounded variation. The latter follows from the fact that a bounded function f is Riemann integrable on an interval $[0,t]$ if and only if the set of discontinuity points of f has Lebesgue measure 0 (see Hobson (1927, pp. 465–466), Munroe (1953, p. 174, Theorem 24.4) or Lebesgue (1902)).Any function of bounded variation is the difference between an increasing and a decreasing function and any monotone function has at most countably many discontinuities.

Conditionally on $\alpha_1, ..., \alpha_M$ and $\tau_1, ..., \tau_M$, the increments $y_1, ..., y_M$ are independent, and $y_j \stackrel{\mathcal{L}}{=} N(\alpha_j, \tau_j)$. Thus, conditionally, y_j^2 is noncentral χ^2 with cumulant function

$$\mathrm{C}\{\zeta \ddagger y_j^2 | \tau_j\} = -\frac{1}{2}\log(1 - 2i\tau_j\zeta) + i\nu_j\zeta(1 - 2i\tau_j\zeta)^{-1}$$

where

$$\nu_j = \alpha_j^2 \tag{6.31}$$

Consequently

$$\mathrm{C}\{\zeta \ddagger u | \tau_1, ..., \tau_M\} = -\sum_{j=1}^{M}\left\{\tfrac{1}{2}\log(1 - 2i\tau_j\zeta) - i\nu_j\zeta(1 - 2i\tau_j\zeta)^{-1} + i\tau_j\zeta\right\}$$

By Taylor's formula with remainder (cf., for instance, Barndorff-Nielsen and Cox (1989, formula 6.122)) we find, provided

$$2|\zeta| \max_{1\le j\le M} \tau_j < 1$$

that

$$\tfrac{1}{2}\log(1 - 2i\tau_j\zeta) - i\nu_j\zeta(1 - 2i\tau_j\zeta)^{-1} + i\zeta\tau_j = \zeta^2\{\tau_j^2 Q_{0j}(\zeta) + 2\nu_j\tau_j Q_{1j}(\zeta)\} - i\nu_j\zeta,$$

where

$$Q_{0j}(\zeta) = 2\int_0^1 \frac{1-s}{(1 - 2i\tau_j\zeta s)^2}\mathrm{d}s$$

and

$$Q_{1j}(\zeta) = 2\int_0^1 \frac{1-s}{(1 - 2i\tau_j\zeta s)^3}\mathrm{d}s.$$

Hence

$$\mathrm{C}\{\zeta \ddagger u | \tau_1, ..., \tau_M\} = i\zeta\sum_{j=1}^{M}\nu_j - \zeta^2\sum_{j=1}^{M}\left\{\tau_j^2 Q_{0j}(\zeta) + 2\nu_j\tau_j Q_{1j}(\zeta)\right\}. \tag{6.32}$$

Now rewrite (6.32) as

$$\begin{aligned}\mathrm{C}\{\zeta \ddagger u | \tau_1, ..., \tau_M\} &= i\zeta\sum_{j=1}^{M}\nu_j - \zeta^2\sum_{j=1}^{M}(\tau_j^2 + 2\nu_j\tau_j) \\ &\quad -\zeta^2\sum_{j=1}^{M}\left[\tau_j^2\{Q_{0j}(\zeta) - 1\} + 2\nu_j\tau_j\{Q_{1j}(\zeta) - 1\}\right] \\ &= \frac{1}{2}\zeta^2 2\sum_{j=1}^{M}\tau_j + R(\zeta),\end{aligned}$$

where

$$R(\zeta) = i\zeta\sum_{j=1}^{M}\nu_j - 2\zeta^2\sum_{j=1}^{M}\nu_j\tau_j - \zeta^2\sum_{j=1}^{M}\left[\tau_j^2\{Q_{0j}(\zeta) - 1\} + 2\nu_j\tau_j\{Q_{1j}(\zeta) - 1\}\right].$$

Thus, to verify (6.26) we must show that

$$\sum_{j=1}^{M}\nu_j \Big/ \sqrt{\textstyle\sum_{j=1}^{M}\tau_j^2} \to 0, \qquad \sum_{j=1}^{M}\nu_j\tau_j \Big/ \textstyle\sum_{j=1}^{M}\tau_j^2 \to 0,$$

$$\sum_{j=1}^{M} \tau_j^2 \left\{ Q_{0j}\left(\zeta/\sqrt{2\sum_{j=1}^{M}\tau_j^2}\right) - 1 \right\} \Big/ \sum_{j=1}^{M}\tau_j^2 \to 0,$$

and

$$\sum_{j=1}^{M} \nu_j\tau_j \left\{ Q_{1j}\left(\zeta/\sqrt{2\sum_{j=1}^{M}\tau_j^2}\right) - 1 \right\} \Big/ \sum_{j=1}^{M}\tau_j^2 \to 0$$

or, equivalently, by (6.29), that

$$\sqrt{M}\sum_{j=1}^{M}\nu_j \to 0, \qquad M\sum_{j=1}^{M}\nu_j\tau_j \to 0, \tag{6.33}$$

$$M\sum_{j=1}^{M} \tau_j^2 \left\{ Q_{0j}\left(\zeta/\sqrt{2\sum_{j=1}^{M}\tau_j^2}\right) - 1 \right\} \to 0, \tag{6.34}$$

and

$$M\sum_{j=1}^{M} \nu_j\tau_j \left\{ Q_{1j}\left(\zeta/\sqrt{2\sum_{j=1}^{M}\tau_j^2}\right) - 1 \right\} \to 0. \tag{6.35}$$

By (6.25),

$$\max_{1\le j\le M}\nu_j = o(M^{-3/2}).$$

Hence $\sqrt{M}\sum_{j=1}^{M}\nu_j = o(1)$ and

$$M\sum_{j=1}^{M}\nu_j\tau_j \le o(M^{-1/2})\sum_{j=1}^{M}\tau_j = \tau^*(\hbar)o(M^{-1/2}) \tag{6.36}$$

implying (6.33).

Finally, to show (6.34)-(6.35) we first note that by (6.30), the local boundedness of τ and (6.29),

$$\tau_j \Big/ \sqrt{\sum_{j=1}^{M}\tau_j^2} = \sqrt{M}\tau_j \Big/ \sqrt{M\sum_{j=1}^{M}\tau_j^2} = M^{-1/2}\hbar c_j \Big/ \sqrt{M\sum_{j=1}^{M}\tau_j^2} = O(M^{-1/2})$$

uniformly in j. Hence

$$Q_{0j}\left(\zeta/\sqrt{2\sum_{j=1}^{M}\tau_j^2}\right) - 1 \to 0 \tag{6.37}$$

and

$$Q_{1j}\left(\zeta/\sqrt{2\sum_{j=1}^{M}\tau_j^2}\right) - 1 \to 0 \tag{6.38}$$

uniformly in j. Moreover, again using (6.29), we have $M\sum_{j=1}^{M}\tau_j^2 = O(1)$ and (6.34)-(6.35) follows from this, (6.36) and (6.37)-(6.38).

6.6 Distribution theory for realised covariation

6.6.1 Results and comments

In order to provide a distribution theory for realised covariation we will work with the multivariate SV model

$$y^*(t) = \alpha^*(t) + \int_0^t \Xi(u)\mathrm{d}w(u), \qquad \text{and} \qquad \Sigma(t) = \Xi(t)\Xi(t)'.$$

Recall that for this model class

$$[y^*](t) = \Sigma^*(t) = \int_0^t \Sigma(u)\mathrm{d}u,$$

that is quadratic covariation equals integrated covariation. This is due to condition (B) ruling out the possibility of jumps in α^* and (6.12) having continuous sample paths. This implies the well known result that

$$[y_M^*]_i = \sum_{j=1}^{M} y_{j,i} y_{j,i}' \xrightarrow{p} \int_{(i-1)\hbar}^{i\hbar} \Sigma(u)\mathrm{d}u, \tag{6.39}$$

as $M \to \infty$. The consistency of realised covariation for increments to actual covariation is of great importance.

Example 18 *Consider the bivariate factor SV model*

$$\mathrm{d}y^*(t) = (\iota \quad I)\,\mathrm{d}s^*(t), \tag{6.40}$$

where s^* is a three dimensional vector of independent, zero mean, SV models with

$$\mathrm{d}s_k^*(t) = \tau_k^{1/2}(t)\mathrm{d}w_k(t), \quad k = 1, 2, 3,$$

where w is a vector of independent, standard Brownian motions. Here ι is a vector of ones, which means that there is a common factor amongst the asset prices. This implies

$$\Sigma(t) = (\iota \quad I)\, diag\left\{\tau_1(t), \tau_2(t), \tau_3(t)\right\} \begin{pmatrix} \iota' \\ I \end{pmatrix}.$$

We will assume that each spot volatility σ follows an independent non-Gaussian Ornstein-Uhlenbeck process

$$\mathrm{d}\tau_k(t) = -\lambda_k \tau_k(t)\mathrm{d}t + \mathrm{d}z_k(\lambda_k t), \quad k = 1, 2, 3, \tag{6.41}$$

where each $z_k(t)$ is an independent Lévy process with non-negative increments. We assume $\tau_k(t) \sim \Gamma(\nu_k, a_k)$, while taking $\hbar = 1$. Then our choices for the parameters will be

$$\begin{pmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{pmatrix} = \begin{pmatrix} 2.0 \\ 1.0 \\ 1.5 \end{pmatrix}, \qquad \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} = \begin{pmatrix} 5.0 \\ 1.0 \\ 1.0 \end{pmatrix}, \qquad \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \end{pmatrix} = \begin{pmatrix} 0.04 \\ 0.13 \\ 3.00 \end{pmatrix}. \tag{6.42}$$

This means that

$$\mathrm{E}\begin{pmatrix} \sigma_1^2(t) \\ \sigma_2^2(t) \\ \sigma_3^2(t) \end{pmatrix} = \begin{pmatrix} 0.4 \\ 1.0 \\ 1.5 \end{pmatrix}, \quad \mathrm{Cov}\begin{pmatrix} \sigma_1^2(t) \\ \sigma_2^2(t) \\ \sigma_3^2(t) \end{pmatrix} = \begin{pmatrix} 0.02 & 0 & 0 \\ 0 & 1.0 & 0 \\ 0 & 0 & 1.5 \end{pmatrix}. \tag{6.43}$$

Thus the common component in the asset prices model (6.40), τ_1, is small in comparison with the individual effects τ_2 and τ_3. However, it is the most persistent of all the three components. Figure 6.6b presents the results from simulating this process. Display (a) shows the simulated daily returns of the two assets. From this picture it is not easy to discern the changing dependence structure in the bivariate process. Display (b) plots the actual covariance $\int_{\hbar(i-1)}^{\hbar i} \Sigma_{12}(u)\mathrm{d}u$ during each day, as well as the associated realised covariance $\sum_{j=1}^{M} y_{j,i(1)} y_{j,i(2)}$ based on $M = 12$. We focus on this as the realised variances were handled in the previous sections. Displays (c) and (d) replicate (b) except we increase M to 48 and then 288. We see that as M increases the estimator $\sum_{j=1}^{M} y_{j,i(1)} y_{j,i(2)}$ becomes closer to $\int_{\hbar(i-1)}^{\hbar i} \Sigma_{12}(u)\mathrm{d}u$, as expected from the convergence in probability result in (A.3).

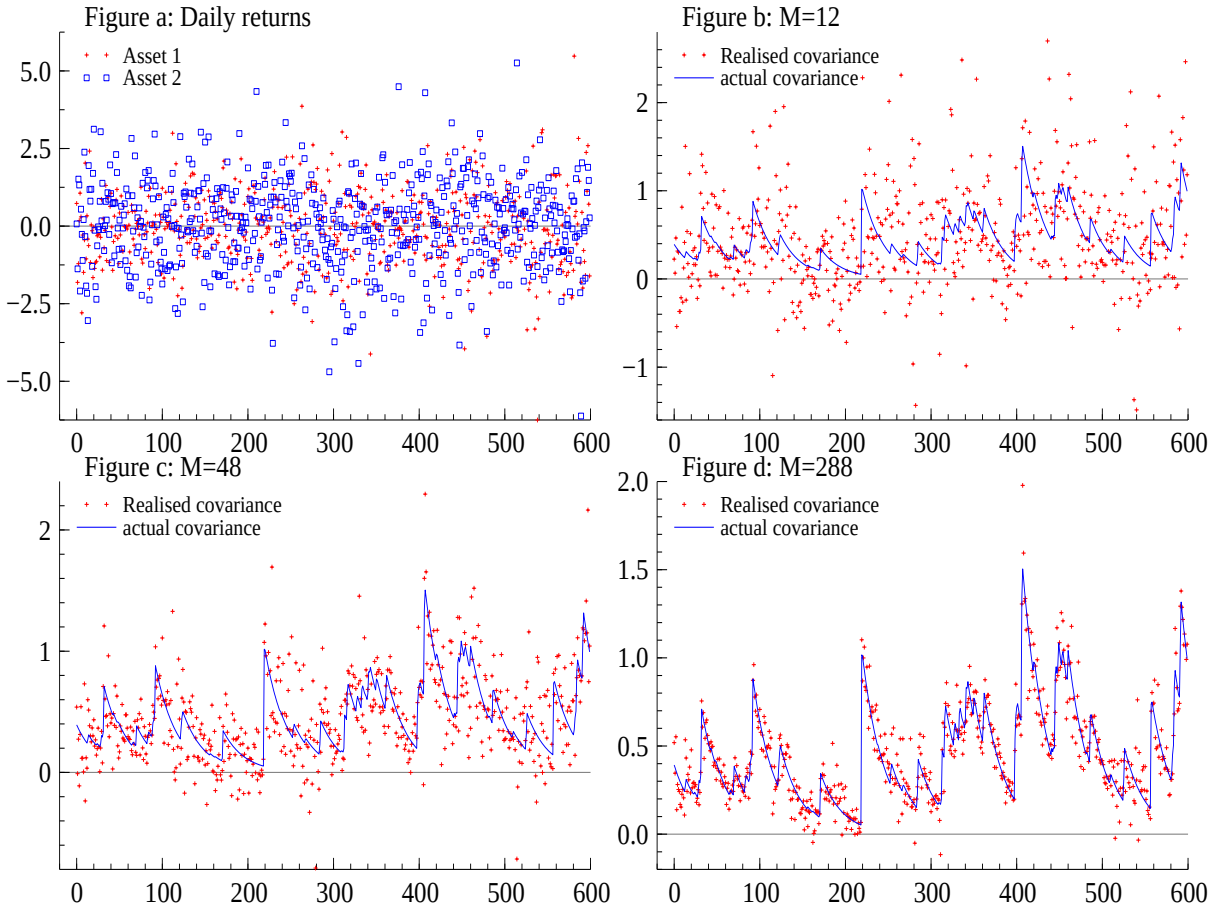

Figure 6.6: *Simulation from a bivariate factor SV model. (a): simulated daily returns. (b): realised covariance based on $M = 12$ together with integrated covariance. (c): realised covariance based on $M = 48$. (d) realised covariance based on $M = 288$. Code is available at:* `simple.ox`

The result that realised covariation converges in probability to the increment of quadratic covariation is a very powerful result for it covers all the interesting cases that arise in financial economics. However, this theory does not provide a guide to the distribution of $[y_M^*]_i - \int_{(i-1)\hbar}^{i\hbar} \Sigma(u)\mathrm{d}u$, the difference between realised covariation and actual covariation. This will be the focus of the next few sections.

The main result discussed here is that the distribution theory for realised variance, discussed in the previous Sections of this Chapter, extends to cover the case of realised covariation. In particular, as $M \to \infty$, conditionally on α^* and Ξ, then under conditions (A)-(C)

$$\sqrt{\frac{M}{\hbar}} \left\{ vech\left(\sum_{j=1}^{M} y_{j,i} y_{j,i}' \right) - vech\left(\int_{\hbar(i-1)}^{\hbar i} \Sigma(u)\mathrm{d}u \right) \right\} \xrightarrow{L} N(0, \Pi_i), \tag{6.44}$$

where the *vech* stacks the (unique) lower triangular elements of the columns of a matrix into a vector. In particular the scaled asymptotic covariance between

$$\sum_{j=1}^{M} y_{j,i(k)} y_{j,i(l)} \qquad \text{and} \qquad \sum_{j=1}^{M} y_{j,i(k')} y_{j,i(l')}$$

is

$$\int_{\hbar(i-1)}^{\hbar i} \{\Sigma_{kk'}(u)\Sigma_{ll'}(u) + \Sigma_{kl'}(u)\Sigma_{lk'}(u)\}\mathrm{d}u. \tag{6.45}$$

Now defining $x_{j,i} = vech(y_{j,i}y'_{j,i})$ and the symmetric positive semi-definite

$$G_i = \sum_{j=1}^{M} x_{j,i}x'_{j,i} - \frac{1}{2}\sum_{j=1}^{M-1}\left(x_{j,i}x'_{j+1,i} + x_{j+1,i}x'_{j,i}\right), \tag{6.46}$$

we have that

$$\frac{M}{\hbar}G_i \xrightarrow{p} \Pi_i. \tag{6.47}$$

The above result provides a general framework for the asymptotics for realised covariation. This is an important result. In particular

- The rate of convergence is $\sqrt{M}$ for all components of the realised covariation.
- No knowledge of the drift process or spot covariance matrix is needed to use this theory.
- The limit theorem is mixed Gaussian, that is Π_i is a stochastic matrix. This means that the difference between realised covariation matrix and actual covariation matrix will be heavier tailed than Gaussian.
- The size of realised covariation matrix errors depends upon the level of volatility of the process. This impacts not just the precision of the realised variance but also the realised covariances.
- The elements of Π_i are explicit, although not observable. They can be consistently estimated by G_i which is a simple function of the high frequency data.
- A convenient feature of the symmetric matrix G_i is that it is positive semi-definite. This follows because for any conformable vector c,

$$c'Hc = \sum_{j=1}^{M}\left(c'x_j\right)^2 - \sum_{j=1}^{M-1}\left(c'x_j\right)\left(c'x_{j+1}\right) \geq 0, \tag{6.48}$$

by the properties of the first serial correlation coefficient.

In order to simplify the notation in the exposition it is helpful to again set $i = 1$ and drop reference to that subscript.

6.6.2 Discussion

The general results are compact. It is helpful to look at special cases in order to gain further understanding. Suppose we are interested in the joint distribution of realised covariation in the bivariate case. Then the Theorem 1 tells us that

$$\sqrt{\frac{M}{\hbar}}\left\{\begin{pmatrix} \sum_{j=1}^{M} y_{j(k)}^2 \\ \sum_{j=1}^{M} y_{j(k)}y_{j(l)} \\ \sum_{j=1}^{M} y_{j(l)}^2 \end{pmatrix} - \int_0^{\hbar}\begin{pmatrix} \Sigma_{kk}(u) \\ \Sigma_{kl}(u) \\ \Sigma_{ll}(u) \end{pmatrix}\mathrm{d}u\right\} \tag{6.49}$$

$$\xrightarrow{L} N\left[0, \int_0^{\hbar}\begin{Bmatrix} 2\Sigma_{kk}^2(u) & 2\Sigma_{kk}(u)\Sigma_{kl}(u) & 2\Sigma_{kl}^2(u) \\ 2\Sigma_{kk}(u)\Sigma_{kl}(u) & \Sigma_{kk}(u)\Sigma_{ll}(u) + \Sigma_{kl}^2(u) & 2\Sigma_{ll}(u)\Sigma_{kl}(u) \\ 2\Sigma_{kl}^2(u) & 2\Sigma_{ll}(u)\Sigma_{kl}(u) & 2\Sigma_{ll}^2(u) \end{Bmatrix}\mathrm{d}u\right]. \tag{6.50}$$

The result on the marginal distribution of realised covariance as $M \to \infty$ is that

$$\frac{\sqrt{\frac{M}{\hbar}}\left\{\sum_{j=1}^{M} y_{j(k)}y_{j(l)} - \int_{\hbar(i-1)}^{\hbar i}\Sigma_{kl}(u)\mathrm{d}u\right\}}{\sqrt{\int_{\hbar(i-1)}^{\hbar i}\{\Sigma_{kk}(u)\Sigma_{ll}(u)+\Sigma_{kl}^2(u)\}\mathrm{d}u}} \xrightarrow{L} N(0,1). \tag{6.51}$$

Notice that when the spot correlation is zero then Σ is diagonal. When $\Sigma_{kk}(t) = \Sigma_{ll}(t)$, then the asymptotic covariance becomes

$$\int_0^{\hbar}\Sigma_{kk}^2(s)\left\{\begin{array}{lll} 2 & 2\rho_{k,l}(u) & 2\rho_{k,l}^2(u) \\ 2\rho_{k,l}(u) & \left(1+\rho_{k,l}^2(u)\right) & 2\rho_{k,l}(u) \\ 2\rho_{k,l}^2(u) & 2\rho_{k,l}(u) & 2 \end{array}\right\}\mathrm{d}u, \tag{6.52}$$

where

$$\rho_{k,l}(u) = \frac{\Sigma_{kl}(u)}{\sqrt{\Sigma_{kk}(u)\Sigma_{ll}(u)}}. \tag{6.53}$$

This last result is a generalisation of the result given in Anderson (1984, p. 121) on the asymptotic joint distribution in the case of i.i.d. Gaussian data.

In order to estimate the Π_i matrix we need some generalisations of realised quarticity. Leading cases are the following. As $M \to \infty$ so

$$\frac{M}{\hbar}\sum_{j=1}^{M} y_{j(k)}^4 \xrightarrow{p} 3\int_0^{\hbar}\Sigma_{kk}^2(u)\mathrm{d}u, \tag{6.54}$$

$$\frac{M}{\hbar}\sum_{j=1}^{M} y_{j(k)}^2 y_{j(l)}^2 \xrightarrow{p} \int_0^{\hbar}\left\{2\Sigma_{kl}^2(u)+\Sigma_{kk}(u)\Sigma_{ll}(u)\right\}\mathrm{d}u. \tag{6.55}$$

Likewise

$$\frac{M}{\hbar}\sum_{j=1}^{M} y_{j(k)}^2 y_{j+1(l)}^2 \xrightarrow{p} \int_0^{\hbar}\Sigma_{kk}(u)\Sigma_{ll}(u)\mathrm{d}u, \tag{6.56}$$

$$\frac{M}{\hbar}\sum_{j=1}^{M}\left(y_{j(k)}y_{j(l)}\right)\left(y_{j+1(k)}y_{j+1(l)}\right) \xrightarrow{p} \int_0^{\hbar}\Sigma_{kl}^2(u)\mathrm{d}u \tag{6.57}$$

Example 19 *A feasible limit theory for the realised covariance is*

$$\frac{\sum_{j=1}^{M} y_{j(k)}y_{j(l)} - \int_0^{\hbar}\Sigma_{kl}(u)\mathrm{d}u}{\sqrt{\sum_{j=1}^{M} y_{j(k)}^2 y_{j(l)}^2 - \sum_{j=1}^{M-1} y_{j(k)}y_{j(l)}y_{j+1(k)}y_{j+1(l)}}} \xrightarrow{L} N(0,1), \tag{6.58}$$

for

$$\sum_{j=1}^{M} y_{j(k)}^2 y_{j(l)}^2 - \sum_{j=1}^{M-1} y_{j(k)}y_{j(l)}y_{j+1(k)}y_{j+1(l)} \xrightarrow{p} \sqrt{\int_0^{\hbar}\{\Sigma_{kk}(u)\Sigma_{ll}(u)+\Sigma_{kl}^2(u)\}\mathrm{d}u},$$

using (6.55) and (6.57). However, an asymptotically equivalent alternative to this is that

$$\frac{\sum_{j=1}^{M} y_{j(k)}y_{j(l)} - \int_0^{\hbar}\Sigma_{kl}(u)\mathrm{d}u}{\sqrt{\frac{1}{2}\left(\sum_{j=1}^{M} y_{j(k)}^2 y_{j(l)}^2 + \sum_{j=1}^{M-1} y_{j(k)}^2 y_{j+1(l)}^2\right)}} \xrightarrow{L} N(0,1). \tag{6.59}$$

6.6.3 Distribution theory for derived quantities

Realised regression

Regression plays a central role both in theoretical and empirical financial economics. For example, the regression of the returns of an individual asset on a wide market index is often called a "beta." In this Section we use our distribution theory for realised covariation to derive a theory for univariate regression. Again this will be based on fixed intervals of time and allowing the number of high frequency observations to go to infinity within that interval. We regress variable l on variable k, then again surpressing subscripts i,

$$\widehat{\beta}_{(lk)} = \frac{\sum_{j=1}^{M} y_{j(k)} y_{j(l)}}{\sum_{j=1}^{M} y_{j(k)}^2}. \tag{6.60}$$

This involves just elements of the realised covariation and so we can use the asymptotic theory of the previous section to derive its asymptotic distribution. The probability limit of regression is known by the theory of QV. In particular

$$\widehat{\beta}_{(lk)} \xrightarrow{p} \frac{\left[y^*_{(k)}, y^*_{(l)}\right]}{\left[y^*_{(k)}\right]} = \beta_{(lk)}. \tag{6.61}$$

Here we extend the theoretical results to derive the asymptotic distribution, under our additional assumptions given above. In this case $\beta_{(lk)}$ has the simpler form of

$$\beta_{(lk)} = \frac{\int_{\hbar(i-1)}^{\hbar i} \Sigma_{kl}(u)\mathrm{d}u}{\int_{\hbar(i-1)}^{\hbar i} \Sigma_{kk}(u)\mathrm{d}u}. \tag{6.62}$$

The asymptotic distribution can be derived using the delta method[6] which yields, as $M \to \infty$, the infeasible limit theory

$$\frac{\sqrt{\frac{M}{\hbar}}\left(\widehat{\beta}_{(lk)} - \beta_{(lk)}\right)}{\sqrt{\left(\int_{\hbar(i-1)}^{\hbar i} \Sigma_{kk}(u)\mathrm{d}u\right)^{-2} g_{(lk)}}} \xrightarrow{L} N(0,1),$$

where

$$g_{(lk)} = d'_{(lk)} \Psi_{(lk)} d_{(lk)}, \qquad d_{(lk)} = \begin{pmatrix} 1 & -\beta_{(lk)} \end{pmatrix} \tag{6.63}$$

and

$$\Psi_{(lk)} = \int_{\hbar(i-1)}^{\hbar i} \begin{Bmatrix} \Sigma_{kk}(u)\Sigma_{ll}(u) + \Sigma_{kl}^2(u) & 2\Sigma_{kk}(u)\Sigma_{kl}(u) \\ 2\Sigma_{kk}(u)\Sigma_{kl}(u) & 2\Sigma_{kk}^2(u) \end{Bmatrix} \mathrm{d}u. \tag{6.64}$$

In practice we have to replace $\Psi_{(lk)}$ and $d_{(lk)}$ by estimators to make the above regression theory feasible. However, the previous section implies this is straightforward. In particular as $M \to \infty$

$$\frac{\widehat{\beta}_{(lk)} - \beta_{(lk)}}{\sqrt{\left(\sum_{j=1}^{M} y_{j(k)}^2\right)^{-2} \widehat{g}_{(lk)}}} \xrightarrow{L} N(0,1). \tag{6.65}$$

[6] This is based on approximating x/y by

$$\frac{\mu_x}{\mu_y} + \frac{x - \mu_x}{\mu_y} - \frac{\left(y - \mu_y\right)\mu_x}{\mu_y^2} = \frac{\mu_x}{\mu_y} + \frac{x}{\mu_y} - \frac{y\mu_x}{\mu_y^2},$$

where μ_x and μ_y are the p-lims of x and y respectively.

where

$$x_j = y_{j(k)}y_{j(l)} - \widehat{\beta}_{(lk)}y_{j(k)}^2 \qquad \text{and} \qquad \widehat{g}_{(lk)} = \sum_{j=1}^{M} x_j^2 - \sum_{j=1}^{M-1} x_j x_{j+1}. \tag{6.66}$$

An attractive feature of this theory is that all of the required terms are straightforward to compute. It is interesting to note that $\sum_{j=1}^{M} x_j = 0$ exactly in this context.

Realised correlation

The same strategy can be used to derive the asymptotic distribution of the realised correlation coefficient. We define

$$\widehat{\rho}_{(lk)} = \frac{\sum_{j=1}^{M} y_{j(k)}y_{j(l)}}{\sqrt{\sum_{j=1}^{M} y_{j(k)}^2 \sum_{j=1}^{M} y_{j(l)}^2}} \xrightarrow{p} \frac{\left[y_{(k)}^*, y_{(l)}^*\right]}{\sqrt{\left[y_{(k)}^*\right]\left[y_{(l)}^*\right]}} = \rho_{(lk)} = \frac{\int_{\hbar(i-1)}^{\hbar i} \Sigma_{kl}(u)\mathrm{d}u}{\sqrt{\int_{\hbar(i-1)}^{\hbar i} \Sigma_{kk}(u)\mathrm{d}u \int_{\hbar(i-1)}^{\hbar i} \Sigma_{ll}(u)\mathrm{d}u}}. \tag{6.67}$$

The infeasible asymptotic distribution can be derived using standard linearisation methods[7]. In particular as $M \to \infty$ so

$$\frac{\sqrt{\frac{M}{\hbar}}\left(\widehat{\rho}_{(lk)} - \rho_{(lk)}\right)}{\sqrt{\left(\int_{\hbar(i-1)}^{\hbar i} \Sigma_{kk}(u)\mathrm{d}u \int_{\hbar(i-1)}^{\hbar i} \Sigma_{ll}(u)\mathrm{d}u\right)^{-1} g_{i(l,k)}}} \xrightarrow{L} N(0,1), \tag{6.68}$$

where

$$g_{(lk)} = d'_{(lk)}\Pi_{(lk)}d_{(lk)}, \quad d_{(lk)} = \begin{pmatrix} -\frac{1}{2}\beta_{(lk)} & 1 & -\frac{1}{2}\beta_{(kl)} \end{pmatrix}' \tag{6.69}$$

and

$$\Pi_{(lk)} = \int_{\hbar(i-1)}^{\hbar i} \begin{Bmatrix} 2\Sigma_{kk}^2(u) & 2\Sigma_{kk}(u)\Sigma_{kl}(u) & 2\Sigma_{kl}^2(u) \\ 2\Sigma_{kk}(u)\Sigma_{kl}(u) & \Sigma_{kk}(u)\Sigma_{ll}(u) + \Sigma_{kl}^2(u) & 2\Sigma_{ll}(u)\Sigma_{kl}(u) \\ 2\Sigma_{k2}^2(u) & 2\Sigma_{ll}(u)\Sigma_{kl}(u) & 2\Sigma_{ll}^2(u) \end{Bmatrix} \mathrm{d}u. \tag{6.70}$$

The feasible limit theory is that as $M \to \infty$ so

$$\frac{\widehat{\rho}_{(lk)} - \rho_{(lk)}}{\sqrt{\left(\sum_{j=1}^{M} y_{j(k)}^2 \sum_{j=1}^{M} y_{j(l)}^2\right)^{-1} \widehat{g}_{(lk)}}} \xrightarrow{L} N(0,1). \tag{6.71}$$

where

$$x_j = y_{j(k)}y_{j(l)} - \frac{1}{2}\widehat{\beta}_{(lk)}y_{j(k)}^2 - \frac{1}{2}\widehat{\beta}_{(kl)}y_{j(l)}^2 \qquad \text{and} \qquad \widehat{g}_{(lk)} = \sum_{j=1}^{M} x_j^2 - \sum_{j=1}^{M-1} x_j x_{j+1}. \tag{6.72}$$

Example 20 *We continue with Example 18. Figures 6.7(a), (b) and (c) show the sample path of $\rho_{(lk)}$ for this model, together with its realised estimator $\widehat{\rho}_{(lk)}$ based on a variety of values of M. We see that for small values of M the estimator is poor, but by the time $M = 288$ it is reasonably precise except when the correlation is low. The bottom row of pictures in Figure 6.7 shows the corresponding errors $\widehat{\rho}_{(lk)} - \rho_{(lk)}$ together with their standard errors computed using (6.71). We see the intervals increasing in size when $\rho_{(lk)}$ is close to zero and reducing otherwise.*

[7] This is based on approximating $x/\sqrt{yz}$ by

$$\begin{aligned} & \frac{\mu_x}{\sqrt{\mu_y\mu_z}} + \frac{(x-\mu_x)}{\sqrt{\mu_y\mu_z}} - \frac{1}{2}\frac{(y-\mu_y)\mu_x}{\mu_y\sqrt{\mu_y\mu_z}} - \frac{1}{2}\frac{(z-\mu_z)\mu_x}{\mu_z\sqrt{\mu_y\mu_z}} \\ = \; & \frac{\mu_x}{\sqrt{\mu_y\mu_z}} + \rho\left(\frac{x}{\mu_x} - \frac{1}{2}\frac{y}{\mu_y} - \frac{1}{2}\frac{z}{\mu_z}\right), \end{aligned}$$

where μ_x, μ_y, μ_z are the p-lims of x, y and z respectively.

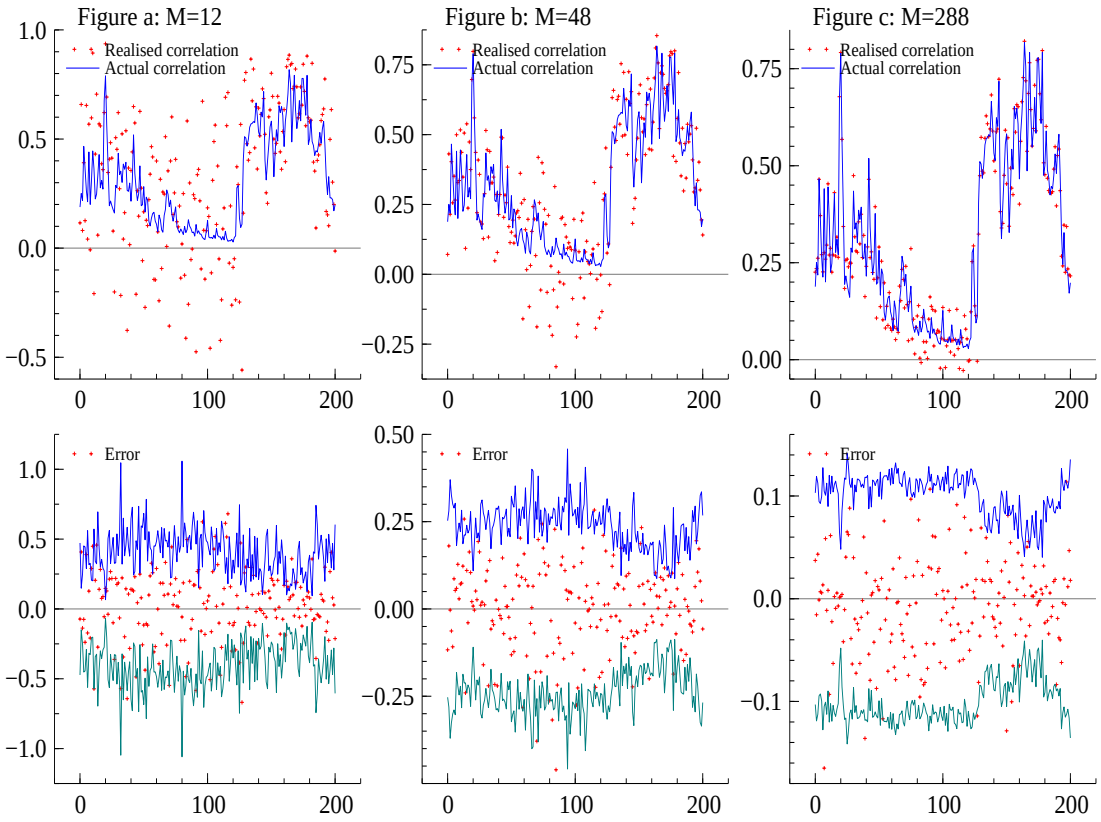

Figure 6.7: *Results from a simulation from a bivariate factor SV model. Plotted on the top graphs are the actual correlation* $\rho_{(lk)}$ *each day. Also drawn are the estimated values* $\widehat{\rho}_{(lk)}$ *based on M=12, 48 and 288 in graphs (a), (b) and (c) respectively. On the bottom row of graphs is* $\rho_{(lk)} - \widehat{\rho}_{(lk)}$ *together with their asymptotic standard errors. Code is available at:* `simple.ox`

One possible way of improving the finite sample behaviour of the asymptotic distribution of $\widehat{\rho}_{(lk)}$ is by using the Fisher-z transformation

$$z_{(lk)} = \frac{1}{2}\log\frac{1+\widehat{\rho}_{(lk)}}{1-\widehat{\rho}_{(lk)}} \quad \text{and} \quad \zeta_{(lk)} = \frac{1}{2}\log\frac{1+\rho_{(lk)}}{1-\rho_{(lk)}}.$$

Recall Fisher's analysis was based on M multivariate, independent and identically distributed Gaussian data, in which case his transformation has the important feature that $\sqrt{M}\left(z_{(lk)} - \zeta_{(lk)}\right)$ has a standard normal limit distribution and it is well known its asymptotic distribution provides an excellent approximation to the exact distribution. In the present case

$$\frac{z_{(lk)} - \zeta_{(lk)}}{\sqrt{\left\{1-\left(\widehat{\rho}_{(lk)}\right)^2\right\}^{-2}\left(\sum_{j=1}^{M} y_{j(l)}^2 \sum_{j=1}^{M} y_{j(k)}^2\right)^{-1}\left(\sum_{j=1}^{M} x_j^2 - \sum_{j=1}^{M-1} x_j x_{j+1}\right)}} \xrightarrow{L} N(0,1). \quad (6.73)$$

Example 21 *We continue with Example 20 but now we focus on the performance of the Fisher based asymptotics (6.73). Using the same simulations as reported in the previous example we now plot* $z_{(lk)} - \zeta_{(lk)}$ *for each day, with their corresponding 95% confidence intervals. This is given in Figure 6.8. The results show that the confidence intervals are now much more stable. In particular they do not seem to move in and out with the value of* $\rho_{(lk)}$*. The finite sample behaviour of the asymptotic theory is studied in the bottom row of Figure 6.8. It shows the QQ*

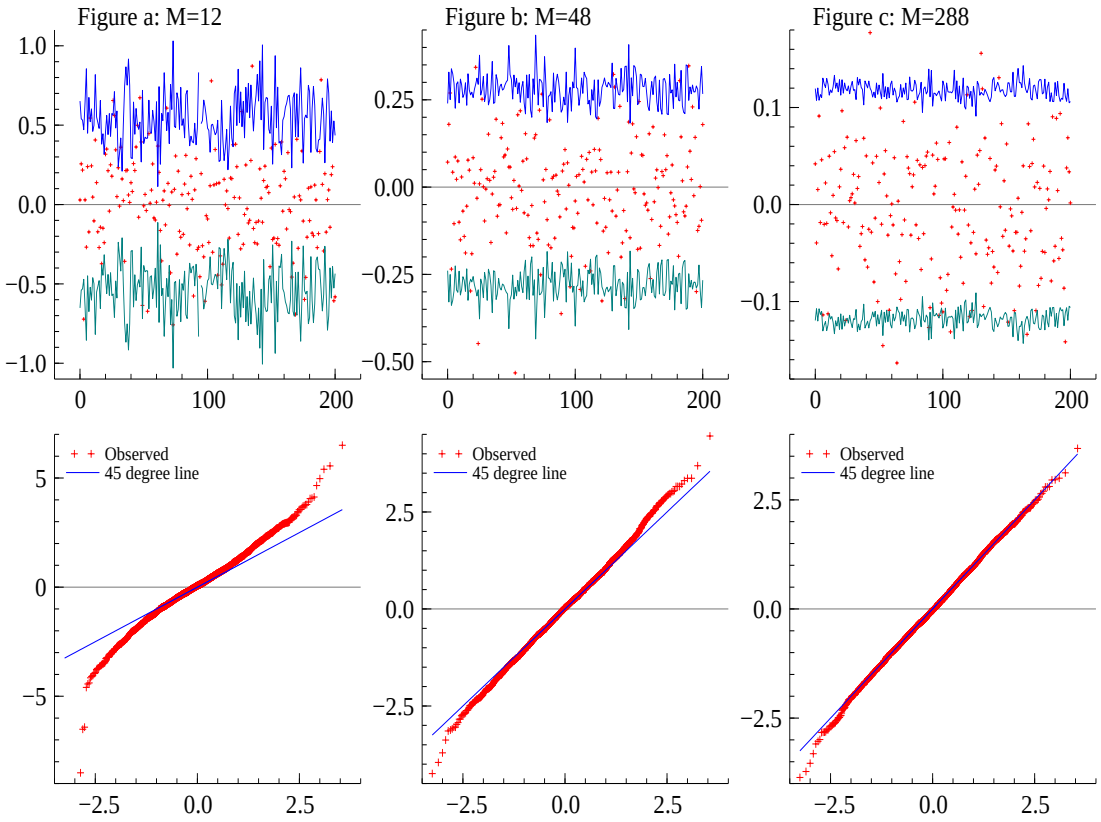

Figure 6.8: *Results from a simulation from a bivariate factor SV model. Plotted on the top graphs are the Fisher transformed* $z_{(lk)} - \zeta_{(lk)}$ *for each of the first 200 days based on M=12, 48 and 288 in graphs (a), (b) and (c) respectively, together with the associate standard errors. On the bottom row of graphs are the corresponding QQ plots to assess normality. The y-axis is the sorted simulations, on the x-axis are the corresponding expected values. Code is available at:* `simple.ox`

plot for (6.73) computed based on 2,400 days. The QQ plot draws the sorted standardised errors (on the y-axis) against that expected under a Gaussian assumption (on the x-axis). Hence if the asymptotics is a perfect description of the finite sample behaviour the QQ plot should be on a 45 degree line. This line is given in the Figure for comparison. We see that for very small values of M *the approximation is poor, but by the time* $M = 48$ *the approximation is quite good and is very accurate by the time* $M = 288$.

An important aspect of the improved finite sample behaviour of the Fisher-based asymptotics for the realised correlation is that it can be used in combination with the theory of the log-realised variances to produce an improved asymptotic theory for the realised covariation matrix. This could be useful in making inference off any function of the covariance matrix. Given the centrality of this measure in financial economics this seems to be of some importance.

Efficiency frontiers

NEIL: TO BE ADDED

Portfolio weights

NEIL: TO BE ADDED

Partial correlation

NEIL: TO BE ADDED

6.7 Empirical example of realised covariation

In this section we extend the work in Section 4 on the analysis of the bivariate return data on the Dollar/DM and Dollar/Yen to consider the dependence structure between the rates for the first 50 days of the sample. Figure 6.9 displays the realised correlation between the two asset returns computed using $M = 144$. This shows the correlation at first falls and then increases to a high level. After maintaining that level it falls again. The movements in the

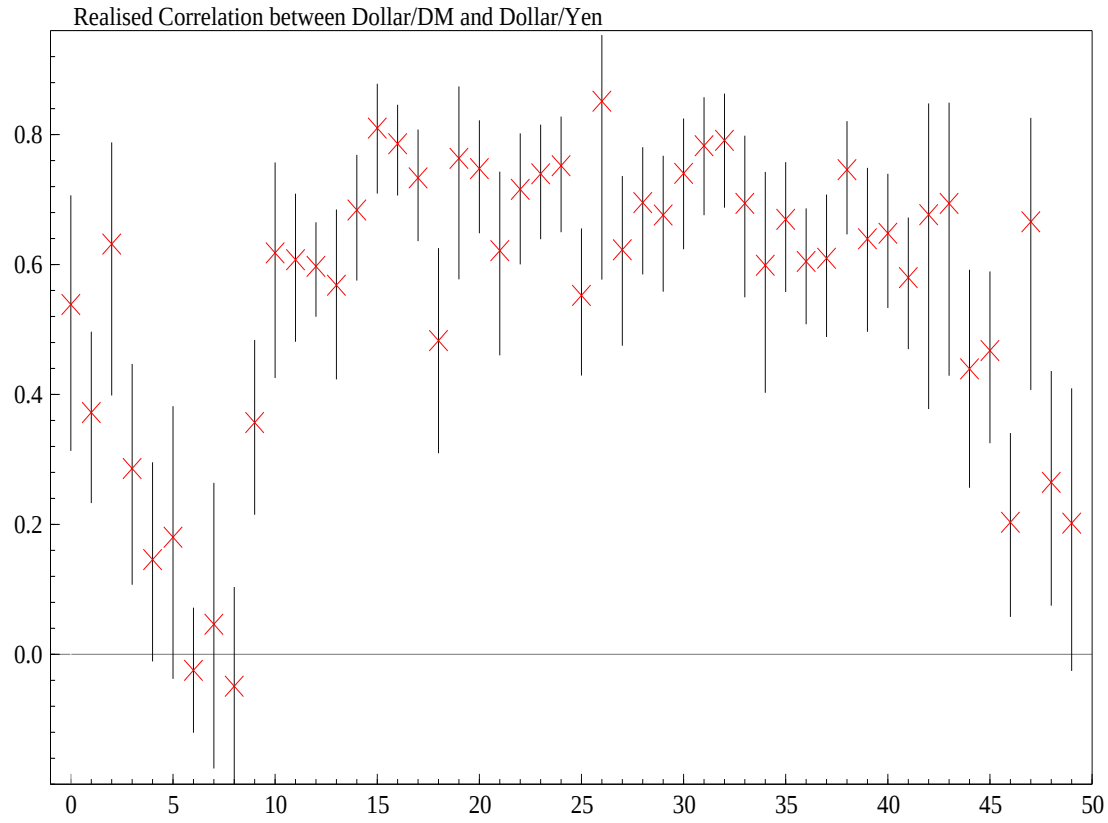

Figure 6.9: *Results from the Dollar/DM and Dollar/Yen return series. Plotted is the realised correlation for the first 50 days of the sample. The 95% confidence intervals are computed using the Fisher-transformed asymptotic theory. Code is available at:* `daily_CI.ox`

realised correlation are significant for the Figure also displays the 95% confidence intervals for these statistics, computed using the Fisher-transform based theory (6.73). The intervals are non-symmetric due to the effect of the transformation. This can be most markedly seen when the correlation is high.

6.8 Theory and proofs of the asymptotics for realised covariance*

6.8.1 Setting

We consider q stochastic processes y^*_ν, determined by a set of stochastic differential equations

$$\mathrm{d}y^*_\nu(t) = \mathrm{d}\alpha^*_\nu(t) + \gamma^a_\nu(t)\mathrm{d}w_a(t) \tag{6.74}$$

with initial condition $y^*_\nu(0) = 0$, $\nu = 1, ..., q$. Here $a = 1, ..., m$ and m may be bigger, equal or smaller than q. Further, $(w_1, ..., w_m)$ is m-dimensional Brownian motion, the processes $\alpha^*_\nu(t)$ and $\gamma^a_\nu(t)$ are assumed to be of locally bounded variation and jointly independent of $(w_1, ..., w_m)$. Throughout we are using the Einstein summation convention[8] to indices a, b, c, d but not to the indices k, l, k', l'. In what is to follow we will write $\gamma^{ab}_{kl} = \gamma^a_k\gamma^b_l$, with similar notation for other index combinations.

Let M denote a positive integer and, for an arbitrary $t \in \mathbf{R}_+$, set[9]

$$\delta = t/M$$

To establish the various limit results on realised covariation it is now convenient to introduce a concept of higher order variation of semimartingales.

6.8.2 Higher order variations of semimartingales

For any q-dimensional semimartingale X with components X_ν, $\nu = 1, ..., q$, we define higher order variations

$$[X_{\nu_1}, ..., X_{\nu_d}](t) = p\text{-}\lim_{\delta\downarrow 0} [X_{\delta\nu_1}, ..., X_{\delta\nu_d}](t),$$

(provided the limit exists). Here d denotes a positive integer and $\nu_1, ..., \nu_d$ is any set of d indices, each index arbitrarily chosen from $\{1, ..., q\}$. Furthermore, $X_{\nu\delta}$ denotes the discrete approximation to X_ν given by

$$X_{\nu\delta}(s) = X_\nu((j-1)\delta) \quad \text{for } (j-1)\delta \le s < j\delta$$

and

$$[X_{\delta\nu_1}, ..., X_{\delta\nu_d}](t) = \delta^{-d/2+1}\sum_{j=1}^{M}\{X_{\nu_1}(\delta j) - X_{\nu_1}((j-1)\delta)\}\cdots\{X_{\nu_d}(j\delta) - X_{\nu_d}((j-1)\delta)\}.$$

Note that for $q = 2$ our notation coincides with the usual notation for the covariation of two semimartingales and we have $[X_{\nu\delta}] = [X_{\nu\delta}, X_{\delta\nu}]$ and $[X_\delta] = [X_\delta, X_\delta]$ where $X_\delta = (X_{1\delta}, ..., X_{q\delta})$.

In studying the higher order variations $[X_{\delta\nu_1}, ..., X_{\delta\nu_d}]$ it is helpful to use the fact, which follows from the multidimensional version of Ito's formula (cf. for instance Protter (1990, p. 74), that for any continuous semimartingales $Y^1_t, ..., Y^m_t$ (with starting value 0) we have

$$Y^1_t\cdots Y^m_t = \sum_{j=1}^{m}\int_0^t \prod_{k\neq j} Y^k_s \mathrm{d}Y^j_s + \sum_{1\le j<k\le m}\int_0^t \prod_{l\neq j,k} Y^l_s \mathrm{d}[Y^j, Y^k]_s. \tag{6.75}$$

For $m = 2$ this reduces to

$$Y^1_t Y^2_t = \int_0^t Y^1_s \mathrm{d}Y^2_s + \int_0^t Y^2_s \mathrm{d}Y^1_s + \int_0^t \mathrm{d}[Y^1, Y^2]_s \tag{6.76}$$

[8] Recall the Einstein summation convention means that if an index is repeated in a single expression then summation over that index is understood.

[9] In reference to the previous Sections, we notice that covering the case $i = 1$ (i.e. one time period) is sufficient, for the results for all other values of i follow immediately. In order to simplify the notation we here drop reference to i and write $\hbar = t$.

while for $m = 4$

$$\begin{aligned} Y_t^1 Y_t^2 Y_t^3 Y_t^4 &= \int_0^t Y_s^1 Y_t^2 Y_t^3 \mathrm{d}Y^4 + \int_0^t Y_s^1 Y_t^2 Y_t^4 \mathrm{d}Y^3 + \int_0^t Y_s^1 Y_t^3 Y_t^4 \mathrm{d}Y^2 + \int_0^t Y_t^2 Y_t^3 Y^4 \mathrm{d}Y_s^1 \\ &+ \int_0^t Y_s^1 Y_t^2 \mathrm{d}[Y^3, Y^4]_s + \int_0^t Y_s^1 Y_t^3 \mathrm{d}[Y^2, Y^4]_s + \int_0^t Y_s^1 Y_t^4 \mathrm{d}[Y^2, Y^3]_s \\ &+ \int_0^t Y_s^2 Y_t^3 \mathrm{d}[Y^1, Y^4]_s + \int_0^t Y_s^2 Y_t^3 \mathrm{d}[Y^1, Y^3]_s + \int_0^t Y_s^3 Y_t^4 \mathrm{d}[Y^1, Y^2]_s. \end{aligned} \tag{6.77}$$

6.8.3 Results

In this subsection we state the limit theorems and corollaries on realised covariation for the processes

$$y_\nu^*(t) = \alpha_\nu^*(t) + \int_0^t \gamma_\nu^a(s) \mathrm{d}w_a(s)$$

defined at the beginning of the Section. We use the notation $y_j = (y_{1j}, ..., y_{qj})'$ where

$$y_{kj} = y_k^*(j) - y_k^*((j-1)\delta).$$

Furthermore,

$$\Sigma_{kl}(t) = \gamma_{kl}^{aa}(t) \tag{6.78}$$

($\Sigma_{kl}(t)$ is the spot covolatility matrix of the SV model) and

$$\Sigma_{kl}^*(t) = \int_0^t \Sigma_{kl}(u) \mathrm{d}u. \tag{6.79}$$

Theorem 2. Conditionally on

$$\{\alpha_\nu, \gamma_\nu^a\}_{\nu=1,...,q;a=1,...,m}$$

the realised covariation matrix

$$[y_M^*] = \left\{ \sum_{j=1}^M y_{kj} y_{lj} \right\}_{k,l=1,2,...,q}, \tag{6.80}$$

follows asymptotically, as $M \to \infty$, the normal law with mean Σ and covariance matrix $\delta\Omega$ where Σ is the $q \times q$ matrix with

$$\Sigma = \Sigma^*(t) = \{\Sigma_{kl}^*(t)\}_{k,l=1,...,q} \tag{6.81}$$

and Ω is the $q^2 \times q^2$ array with elements

$$\Omega = \left\{ \int_0^t \{\Sigma_{kk'}(u)\Sigma_{ll'}(u) + \Sigma_{kl'}(u)\Sigma_{lk'}(u)\} \mathrm{d}u \right\}_{k,k',l,l'=1,...,q}. \tag{6.82}$$

Corollary 1. Unconditionally the asymptotic law of

$$\delta^{-1/2} \left([y_M^*] - \Sigma \right) \tag{6.83}$$

is mixed normal with mean 0 and random covariance matrix Ω.

Theorem 3. Let

$$\widehat{\psi}_{klk'l'} = \delta^{-1} \sum_{j=1}^{M} y_{kj} y_{lj} y_{k'j} y_{l'j}$$

and

$$\widetilde{\psi}_{klk'l'} = \delta^{-1} \sum_{j=1}^{M-1} y_{kj} y_{lj} yk_{k',j+1} y_{l',j+1}.$$

Then

$$\widehat{\psi}_{klk'l'} \xrightarrow{p} \int_0^t \{\Sigma_{kk'}(u)\Sigma_{ll'}(u) + \Sigma_{kl'}(u)\Sigma_{lk'}(u) + \Sigma_{kl}(u)\Sigma_{k'l'}(u)\}\mathrm{d}u$$

while

$$\begin{aligned} \widetilde{\psi}_{klk'l'} &\xrightarrow{p} \int_0^t \Sigma_{kl}(u)\Sigma_{k'l'}(u)\mathrm{d}u \\ &= \operatorname*{p-lim}_{M\to\infty} \widetilde{\psi}_{k'l'kl}. \end{aligned}$$

Corollary 2. Defining

$$\overline{\psi}_{klk'l'} = \widehat{\psi}_{klk'l'} - \frac{1}{2}\left(\widetilde{\psi}_{klk'l'} + \widetilde{\psi}_{k'l'kl}\right)$$

we have

$$\overline{\psi}_{klk'l'} \xrightarrow{p} \int_0^t \{\Sigma_{kk'}(u)\Sigma_{ll'}(u) + \Sigma_{kl'}(u)\Sigma_{lk'}(u)\}\mathrm{d}u,$$

and so there exists a random $q^2 \times q^2$ matrix H

$$H = \sum_{j=1}^{M} x_j x_j' - \frac{1}{2} \sum_{j=1}^{M-1} \left(x_j x_{j+1}' + x_{j+1} x_j'\right), \tag{6.84}$$

where[10] $x_j = vec(y_j y_j')$, such that

$$\delta^{-1} H \xrightarrow{p} \Omega$$

as $M \to \infty$.

Note that H is explicitly calculable in terms of y_M^*.

It is sometimes convenient to avoid the symmetric replication in the realised covariation matrix by employing a *vech* transformation. Then the limit theory can be easily written as

Corollary 3. As $M \to \infty$, conditionally

$$\delta^{-1/2} \{vech([y_M^*]) - vech(\Sigma)\} \xrightarrow{L} N(0, \Pi), \tag{6.85}$$

while, defining[11] $x_j = vech(y_j y_j')$ and

$$G = \sum_{j=1}^{M} x_j x_j' - \frac{1}{2} \sum_{j=1}^{M-1} \left(x_j x_{j+1}' + x_{j+1} x_j'\right), \tag{6.86}$$

we have that

$$\delta^{-1} G \xrightarrow{p} \Pi. \tag{6.87}$$

[10] Recall the *vec* notation stacks the columns of a matrix into a vector.

[11] Recall the *vech* notation stacks the lower triangular elements of a matrix into a vector. See, for example, Lutkepol (1996; Ch. 7).

The matrix G is still only guaranteed to be positive semi-definite, but should be positive definite in practice.

Finally we note the following result.

Corollary 4. The asymptotic unconditional covariance of the error term is, if it exists,

$$\mathrm{E}(\Omega) = \left\{ \int_0^t \left[\mathrm{E}\left\{ \Sigma_{kk'}(u)\Sigma_{ll'}(u) \right\} + \mathrm{E}\left\{ \Sigma_{kl'}(u)\Sigma_{lk'}(u) \right\} \right] \mathrm{d}u \right\}_{k,k',l,l'=1,\ldots,q}. \tag{6.88}$$

6.8.4 Proofs of theorems

We are interested in the limiting behaviour of

$$[y^*_{k\delta}, y^*_{l\delta}](t) = \sum_{j=1}^{M} y_{kj} y_{lj} = \sum_{j=1}^{M} \{y^*_k(\delta j) - y^*_k((j-1)\delta)\}\{y^*_l(j\delta) - y^*_l((j-1)\delta)\}, \tag{6.89}$$

when the processes α^*_k, γ^a_k are considered given by conditioning.

To the asymptotic order considered the limit behaviour of $[y^*_\delta]$ is dominated by the infinitesimal variation of the Brownian motion w, so that - as we shall show - the behaviour of the processes α^*_k does not influence the limit laws.

Case $\alpha^* = 0$. Suppose initially that $\alpha^* = (\alpha^*_1, ..., \alpha^*_q)$ is identically 0, in which case

$$y_{j(k)} = \int_{(j-1)\delta}^{j\delta} \gamma^a_k(s)\mathrm{d}w_a(s). \tag{6.90}$$

The summands in (6.89) are independent and since the processes γ^a_k are locally of bounded variation the matrix $[y^*_\delta]$ must asymptotically be normally distributed. (Detailed verification of this follows standard reasoning and is therefore omitted here.) The task is thus to determine the asymptotic mean and variance of $[y^*_\delta]$.

It is convenient to introduce

$$\Gamma_{kl}(t) = \int_0^t \gamma^{aa}_{kl}(s)\mathrm{d}s$$

as an alternative notation for and $\Sigma_{kl}(t)$ and to write

$$\Gamma_{klj} = \int_{(j-1)\delta}^{j\delta} \gamma^{aa}_{kl}(s)\mathrm{d}s.$$

By (6.90) and (A.25) we find $\mathrm{E}\{y_{kj}y_{lj}\} = \Gamma_{klj}$ and hence we have

$$\mathrm{E}\{[y^*_{k\delta}, y^*_{l\delta}](t)\} = \Gamma_{kl}(t) = \Sigma_{kl}(t).$$

Furthermore, for any indices k, l, k', l' in $\{1, ..., q\}$,

$$\begin{aligned} \mathrm{Cov}\left\{[y^*_{k\delta}, y^*_{l\delta}](t), [y^*_{k'\delta}, y^*_{l'\delta}](t)\right\} &= \mathrm{E}\left\{ \sum_{j'=1}^{M} (y_{kj'}y_{lj'} - \Gamma_{klj'}) \sum_{j=1}^{M} (y_{k'j}y_{l'j} - \Gamma_{k'l'j}) \right\} \\ &= \sum_{j=1}^{M} \mathrm{E}\{y_{kj}y_{lj}y_{k'j}y_{l'j}\} - \sum_{j=1}^{M} \Gamma_{klj}\Gamma_{k'l'j}. \end{aligned} \tag{6.91}$$

Consider now the case $j = 1$. Using (A.27) and similarly for other index combinations, we find

$$\begin{aligned}\mathrm{E}\{y_{k1}y_{l1}y_{k'1}y_{l'1}\} &= \int_0^\delta \mathrm{E}\{\int_0^u \int_0^u \gamma_k^a(s)\gamma_l^b(s)\mathrm{d}b_a(s)\mathrm{d}b_b(s)\}\gamma_{k'}^c(u)\gamma_{l'}^c(u)\mathrm{d}u \quad [6] \\ &= \int_0^\delta \gamma_{k'l'}^{cc}(u)\int_0^u \gamma_{kl}^{aa}(s)\mathrm{d}s\mathrm{d}u \quad [6] \end{aligned} \tag{6.92}$$

the symbol [6] indicating that $\mathrm{E}\{y_{kj}y_{lj}y_{k'j}y_{l'j}\}$ equals the sum of the term given plus 5 similar terms obtained via permutation of the indices k, l, k', l'. Continuing the calculation we have

$$\begin{aligned}\mathrm{E}\{y_{k1}y_{l1}y_{k'1}y_{l'1}\} &= \int_0^\delta \gamma_{k'l'}^{cc}(u)\Gamma_{kl}(u)\mathrm{d}u \quad [6] \\ &= \int_0^\delta \gamma_{k'l'}^{cc}(u)\Gamma_{kl}(u)\mathrm{d}u + \int_0^\delta \gamma_{kl}^{cc}(u)\Gamma_{k'l'}(u)\mathrm{d}u \\ &\quad + \int_0^\delta \gamma_{ll'}^{cc}(u)\Gamma_{kk'}(u)\mathrm{d}u + \int_0^\delta \gamma_{kk'}^{cc}(u)\Gamma_{ll'}(u)\mathrm{d}u \\ &\quad + \int_0^\delta \gamma_{k'l}^{cc}(u)\Gamma_{kl'}(u)\mathrm{d}u + \int_0^\delta \gamma_{kl'}^{cc}(u)\Gamma_{k'l}(u)\mathrm{d}u.\end{aligned}$$

Next we note that

$$\frac{\mathrm{d}}{\mathrm{d}s}\{\Gamma_{kl}(s)\Gamma_{k'l'}(s)\} = \gamma_{kl}^{cc}(s)\Gamma_{k'l'}(s) + \gamma_{k'l'}^{cc}(s)\Gamma_{kl}(s),$$

or, in other words,

$$\int_0^\delta \gamma_{k'l'}^{cc}(u)\Gamma_{kl}(u)\mathrm{d}u + \int_0^\delta \gamma_{kl}^{cc}(u)\Gamma_{k'l'}(u)\mathrm{d}u = \Gamma_{kl}(\delta)\Gamma_{k'l'}(\delta).$$

The terms for other values of the indices behave similarly and all in all we obtain

$$\mathrm{Cov}\,\{[y_{k\delta}^*, y_{l\delta}^*](t), [y_{k'\delta}^*, y_{l'\delta}^*](t)\} = \sum_{j=1}^M (\Gamma_{kk'j}\Gamma_{ll'j} + \Gamma_{kl'j}\Gamma_{lk'j}). \tag{6.93}$$

Now, when $\delta \to 0$ the sum in (6.93) turns into an integral and behaves as $\delta\Omega_{kl,k'l'}(t)$ where

$$\Omega_{kl,k'l'}(t) = \int_0^t \{\gamma_{kk'}^{aa}(s)\gamma_{ll'}^{cc}(s) + \gamma_{kl'}^{aa}(s)\gamma_{lk'}^{cc}(s)\}\mathrm{d}s,$$

as stated in Theorem 2.

Turning then to Theorem 3, recall from (6.91) that

$$\mathrm{Cov}\,\{[y_{k\delta}^*, y_{l\delta}^*](t), [y_{k'\delta}^*, y_{l'\delta}^*](t)\} = \sum_{j=1}^M \mathrm{E}\{y_{kj}y_{lj}y_{k'j}y_{l'j}\} - \sum_{j=1}^M \Gamma_{klj}\Gamma_{k'l'j}$$

From the previous discussion we have

$$\delta^{-1}\mathrm{Cov}\,\{[y_{k\delta}^*, y_{l\delta}^*](t), [y_{k'\delta}^*, y_{l'\delta}^*](t)\} \to \Omega_{kl,k'l'}(t).$$

On the other hand, arguing as above it is seen that

$$\delta^{-1}\sum_{j=1}^M \mathrm{E}\{y_{kj}y_{lj}y_{k'j}y_{l'j}\} \to \int_0^t \{\gamma_{kl}^{aa}(s)\gamma_{k'l'}^{cc}(s) + \gamma_{kk'}^{aa}(s)\gamma_{ll'}^{cc}(s) + \gamma_{kl'}^{aa}(s)\gamma_{lk'}^{cc}(s)\}\mathrm{d}s$$

and

$$\delta^{-1}\sum_{j=1}^{M}\Gamma_{klj}\Gamma_{k'l'j}\to\int_0^t\gamma_{kl}^{aa}(s)\gamma_{k'l'}^{cc}(s)\mathrm{d}s,$$

and, moreover, that

$$[y_{\delta k}^*,y_{\delta l}^*,y_{\delta k'}^*,y_{\delta l'}^*]=\delta^{-1}\sum_{j=1}^{M}y_{kj}y_{lj}y_{k'j}y_{l'j},$$

must converge in probability to the same limit as $\delta^{-1}\sum_{j=1}^{M}\mathrm{E}\{y_{kj}y_{lj}y_{k'j}y_{l'j}\}$. Thus to obtain a consistent estimator of $\Omega_{kl,k'l'}(t)$ it suffices to find a consistent estimator of $\int_0^t\gamma_{kl}^{aa}(s)\gamma_{k'l'}^{cc}(s)\mathrm{d}s$. The quantity

$$\sum_{j=1}^{M-1}y_{kj}y_{lj}y_{k',j+1}y_{l',j+1},$$

solves this problem, as is seen by again applying formula (A.27).

General case To prove that the same limiting laws hold when the mean processes α_ν^* are not 0 (but of locally bounded variation) we first note that

$$y_{kj}y_{lj}=\alpha_{kj}\alpha_{lj}+\alpha_{kj}y_{0lj}+\alpha_{lj}y_{0kj}+y_{0kj}y_{0lj},$$

where

$$y_{0kj}=\int_{(j-1)\delta}^{j\delta}\gamma_k^a(s)\mathrm{d}w_a(s)$$

and

$$\alpha_{kj}=\alpha_k^*(j\delta)-\alpha_k^*((j-1)\delta).$$

Since α_k^* is of locally bounded variation, α_{kj} is $o(\sqrt{\delta})$ uniformly in k and j. Furthermore, conditionally on the α_k^* and γ_k^a processes, we have

$$\alpha_{kj}y_{0lj}+\alpha_{lj}y_{0kj}\sim N(0,\alpha_{kj}^2\Gamma_{llj}+\alpha_{lj}^2\Gamma_{kkj}).$$

Consequently,

$$\begin{aligned}[y_{k\delta}^*,y_{l\delta}^*](t) &= \sum_{j=1}^{M}\alpha_{kj}\alpha_{lj}+\sum_{j=1}^{M}(\alpha_{kj}y_{0lj}+\alpha_{lj}y_{0kj})+\sum_{j=1}^{M}y_{0kj}y_{0lj}\\ &= o(\sqrt{\delta})+o_p(\sqrt{\delta})+[y_{0k\delta}^*,y_{0l\delta}^*](t),\end{aligned}$$

where

$$[y_{0k\delta}^*,y_{0l\delta}^*](t)=\sum_{j=1}^{M}y_{0kj}y_{0lj}.$$

It follows that, conditionally, $\delta^{-1/2}\{[y_\delta^*,y_\delta^*]-\Gamma\}$ has the same limit law as $\delta^{-1/2}\{[y_{0\delta}^*,y_{0\delta}^*]-\Gamma\}$ and the latter is as given in Theorem 2.

An analogous argument applies to Theorem 3.

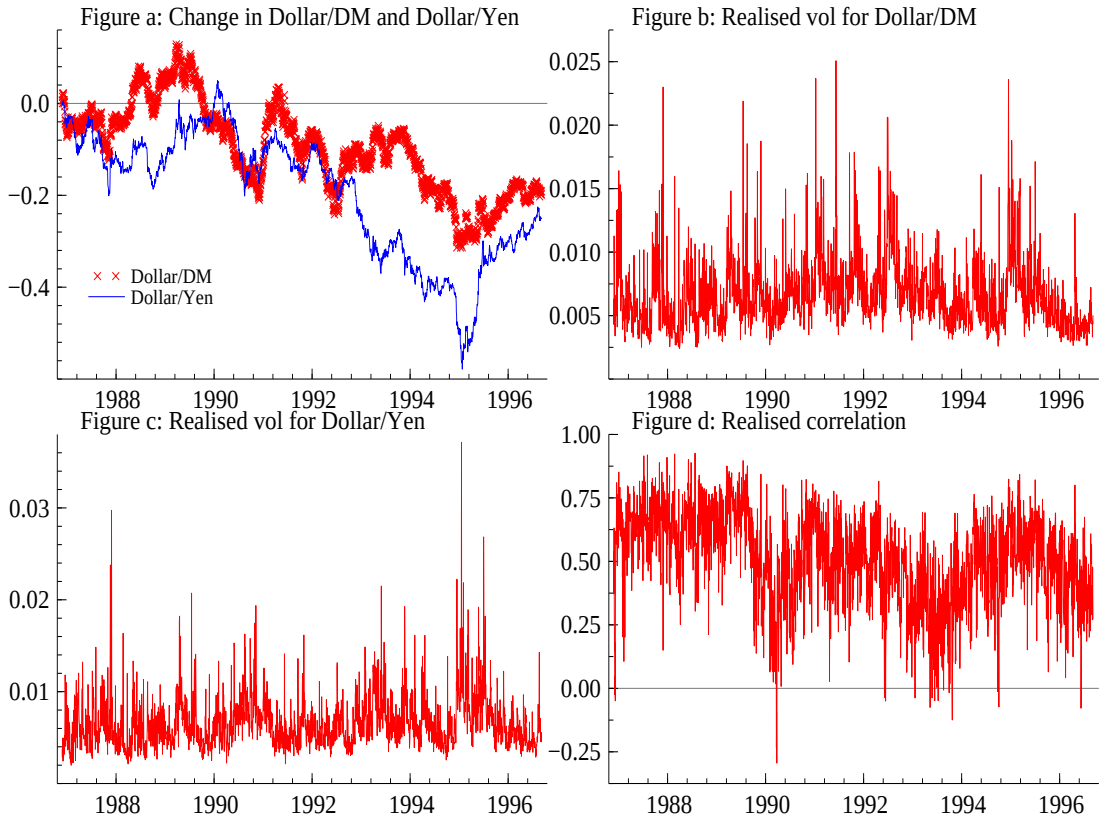

Figure 6.10: *Long time series of the daily movements in the Dollar against the DM and Yen. Figure (a) change in the rates since 1st December 1986. Figure (b) realised volatility each day computed using $M = 144$ for the DM series. Figure (c) realised volatility each day computed using $M = 144$ for the Yen series. Figure (d) realised correlation between DM and Yen series each day computed using $M = 144$. File:* `daily_realised.ox`.

6.9 Time series of realised variances

6.9.1 Framework

So far we have analysed the asymptotics of $\sum_{j=1}^{M} y_{j,i}^2$ as $M \to \infty$ for a single i. In this section we will explicitly analyse a long time series of realised variances, trying to use the time series structure to construct more efficient estimators and forecasts of τ_i. To start out we have drawn Figure 6.10 which displays information on the Olsen data on the DM and Yen against the US Dollar. Figure 6.10(a) shows the movement of the log prices since 1st December 1986 and demonstrates the strengthening of the Dollar against both currencies. It is more marked against the Yen which loses half of its value in this period. Recall this decade corresponds to the start of the long Japanese recession which saw a massive sell off in the value of its stock market while the US markets were booming. Some of these booms and contractions were funded by outflows and inflows of capital from foreign investors. This had the effect of moving the exchange rates, sometimes very rapidly. Figure 6.10(b) shows the daily realised volatility $\sqrt{\sum_{j=1}^{M} y_{j,i}^2}$ drawn against i, the day, for the DM series. It is computed using $M = 144$, corresponding to 10 minute returns. It is quite a ragged series but with periods of increased volatility. A similar picture emerges from the corresponding realised volatility for the Yen given in Figure 6.10(c). The most

stable of the time series is Figure 6.10(d) which displays the realised correlation between the DM and Yen. This is quite a predictable series with seemingly significant movements in the correlation.

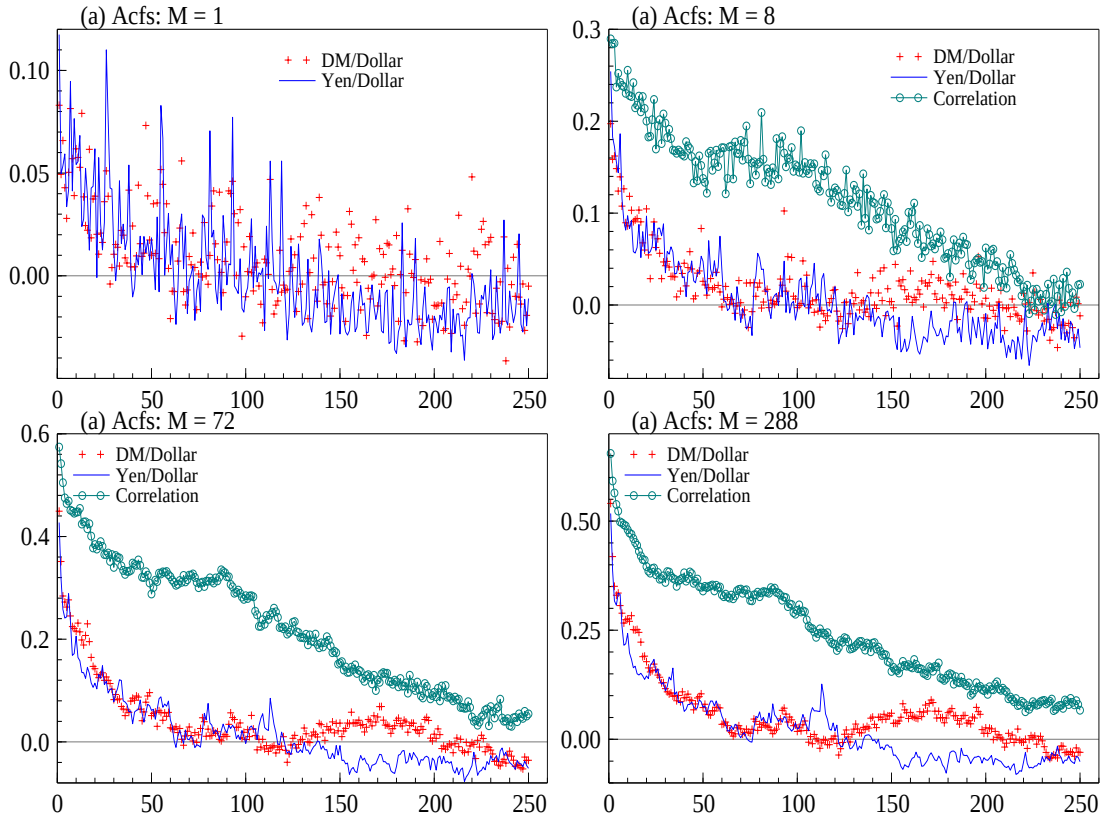

Figure 6.11: *Autocorrelations of realised quantities using a long time series of the movements in the Dollar against the DM and Yen. Plots are for the realised volatilities and the realised correlation. Figure (a) M=1 case, which corresponds to daily returns. No realised correlation is computed in this case. Figure (b) $M = 8$ case. Figure (c) $M = 72$ for the Yen series. Figure (d) $M = 288$. File:* `daily_realised.ox`.

Having computed realised variances , realised volatilities and correlations each day for a long bivariate time series, one approach is now to regard that derived series as a daily time series. An example is the time series of realised variances

$$\sum_{j=1}^{M} y_{j,1}^2, \sum_{j=1}^{M} y_{j,2}^2, ..., \sum_{j=1}^{M} y_{j,T}^2.$$

This new series is of length T, the number of days in the sample. This is an important intellectual step for we can now use time series methods on this derived series.

The correlograms for the daily time series of realised volatilities of these quantities are displayed in Figure 6.11 for a variety of values of M. 250 lags are used in these figures which correspond to measuring correlations over a one year period. Figure 6.11(a) shows the results for $M = 1$. In this case the realised variances are simply squared daily returns while the realised correlations do not make sense in this case and so are not plotted here. The correlogram has the well know slow decay but starting at quite a low level. Figure 6.11(b) shows the

effect of increasing M slightly to 8, now we are computing the realised quantities using 150 minute returns. The level of the correlation in the realised variances have jumped by a factor of around 2, however the two autocorrelations are quite similar with moderately fast decays. Indeed there does not seem to be much dependence in the series after around half a year. The realised correlation is dramatically different. It has a much higher level of autocorrelation whose decay seems to be roughly linear. The fact that the realised correlation has a higher level of autocorrelation than the realised variances is interesting and we will name this the ABDL effect after the authors, Torben Andersen, Tim Bollerslev, Frank Diebold and Paul Labys, who first observed it.

Figure 6.11(c) shows the corresponding results for $M = 72$, which uses 20 minute returns. All the autocorrelations correlations are boosted as M increases from 8, however the broad story is the same. A clear observation is that the autocorrelations are becoming less jagged with the increase in M. The same points appear with Figure 6.11(d) which uses $M = 288$. Using these five minute return based statistics the correlations are now quite high and smooth. The autocorrelation of the realised correlations, in particular, is very high.

Having observed some of the empirical features of the realised variances and correlations we will now set out a theoretical framework for the study of the time series of realised quantities. For the moment we focus on the realised variances.

We define sequences of realised and actual variances

$$[y_M^*]_{s:p} = \left(\sum_{j=1}^{M} y_{j,s}^2, \sum_{j=1}^{M} y_{j,s+1}^2, ..., \sum_{j=1}^{M} y_{j,p}^2 \right)' \qquad \text{and} \qquad \tau_{s:p} = (\tau_s, \tau_{s+1}, ..., \tau_p)',$$

where we recall that $\tau_i = \int_{\hbar(i-1)}^{\hbar i} \tau(u) \mathrm{d}u$. The theory of realised variance implies that

$$\sqrt{\frac{M}{\hbar}} \left([y_M^*]_{s:p} - \tau_{s:p} \right) | \tau_{s:p}^{[2]} \xrightarrow{d} N \left\{ 0, 2 diag \left(\tau_{s:p}^{[2]} \right) \right\},$$

where $\tau_i^{[2]} = \int_{\hbar(i-1)}^{\hbar i} \tau^2(u) \mathrm{d}u$.

Although estimating $\tau_{s:p}$ by $[y_M^*]_{s:p}$ has attractions, the variance of the error is typically quite large even when M is high. More precise estimators could be obtained by pooling neighbouring time series observations for realised variances tend to be highly correlated through time. This pooling will typically reduce the variance of the estimator, but will induce a bias.

To set up a formal framework for this discussion, abstractly write A as a matrix of non-stochastic weights. Then

$$\sqrt{\frac{M}{\hbar}} \left(A[y_M^*]_{s:p} - A\tau_{s:p} \right) | \tau_{[\nu]}^{[2]} \xrightarrow{d} N \left\{ 0, 2A diag \left(\tau_{s:p}^{[2]} \right) A' \right\}.$$

Then consider the estimator

$$\widehat{\tau_{s:p}} = c\mathrm{E} \left(\tau_{s:p} \right) + A[y_M^*]_{s:p}.$$

If we assume that the realised variances constitute a covariance stationary process then the weighted least squares estimator of $\tau_{s:p}$ sets

$$c = (I - A) \iota$$

and

$$\begin{aligned} A &= \left[\mathrm{Cov} \left(\tau_{s:p} \right) + \frac{2\hbar \mathrm{E} \left(\tau_i^{[2]} \right)}{M} I \right]^{-1} \mathrm{Cov} \left(\tau_{s:p} \right) \\ &= \left[\mathrm{Cov} \left([y_M^*]_{s:p} \right) \right]^{-1} \mathrm{Cov} \left(\tau_{s:p} \right). \end{aligned}$$

Of course, as $M \to \infty$ so $A \to I$ and so

$$\widehat{\tau_{s:p}} \xrightarrow{p} \tau_{s:p}.$$

Proposition (due to Bent Nielsen). A necessary and sufficient condition for A to be a non-negative matrix (a matrix with non-negative elements) is that all the partial correlations amongst the elements of $\tau_{s:p}$ are non-negative.

Proof. Given in the Appendix.

In practice A has to be estimated from the data. Broadly this can be carried out in two ways (i) by using empirical averages from the data, (ii) implying them from an estimated model. Here we follow the former approach, delaying until the next subsection a discussion of a model based method.

Thus if we have a large sample of a stationary process of realised variances and the process is ergodic then we have that

$$\mathrm{E}\left(\widehat{\tau_i^{[2]}}\right) = \left(\frac{1}{T}\sum_{i=1}^{T}\frac{M}{3\hbar}\sum_{j=1}^{M} y_{j,i}^4\right) \xrightarrow{p} \mathrm{E}\left(\tau_i^{[2]}\right),$$

as T and M go to infinity. Likewise $\mathrm{Cov}\left([y_M^*]_{s:p}\right)$ can be estimated by averages of the time series of realised variances. Hence A can be replaced by

$$\widehat{A} = \left[\mathrm{Cov}\left(\widehat{[y_M^*]_{s:p}}\right)\right]^{-1}\left\{\mathrm{Cov}\left(\widehat{[y_M^*]_{s:p}}\right) - \mathrm{E}\left(\widehat{\tau_i^{[2]}}\right)\frac{2\hbar}{M}I\right\},$$

which is a feasible weighting matrix. This will imply $\widehat{c} = \left(I - \widehat{A}\right)\iota$ and

$$\widehat{\tau_{s:p}} = \widehat{c}\mathrm{E}\left(\widehat{\tau_{s:p}}\right) + \widehat{A}[y_M^*]_{s:p}.$$

In the case where we estimate a single actual variance using a single realised variance sequence, so $s = p$, we have given the weights for the DM and Yen series in Table 6.1. This is based on the entire time series sample of nearly 2500 days.

	DM		Yen	
M	$\widehat{A}$	$\widehat{c}$	$\widehat{A}$	$\widehat{c}$
1	.182	.817	.229	.770
8	.449	.550	.513	.486
72	.778	.221	.789	.210
288	.877	.122	.906	.093

Table 6.1: *Estimated weights for $\widehat{\tau}_i$, the regression estimator of τ_i which uses only $[y_M^*]_i$ and an intercept. Results for the DM and Yen series against the Dollar. File:* `daily_realised.ox`.

We can see the results do not vary very much with the series being used. In particular, for $M = 8$ then the estimator of τ_i for the DM series would be

$$\widehat{\tau}_i = .449[y_M^*]_i + .550\frac{1}{T}\sum_{j=1}^{T}[y_M^*]_j.$$

Thus for small values of M the regression estimator puts a moderate weight on the realised variance and more on the unconditional mean of the variances. As M increases this situation reverses, but even for large values of M the unconditional mean is still quite highly weighted. From now on we will solely focus on the DM series to make the exposition more compact.

In the dynamic case the results are more complicated to present. Here we start by considering estimating three actual variances using three contiguous realised variances — one lag, one lead and the contemporaneous realised variance. Thus $\widehat{A}$ will be a 3×3 matrix and $\widehat{c}$ a 3×1 vector. In the case of $M = 8$ we have that

$$\left\{\text{Cov}\left([y_M^*]_{1:3}\right)\right\}^{-1} = \frac{1}{100^2}\begin{pmatrix} 2.07 & -.358 & -.258 \\ -.358 & 2.10 & -.358 \\ -.258 & -.358 & 2.07 \end{pmatrix},$$

while

$$\widehat{A} = \begin{pmatrix} .418 & .100 & .072 \\ .100 & .409 & .100 \\ .072 & .100 & .418 \end{pmatrix}, \quad \widehat{c} = \begin{pmatrix} .408 \\ .388 \\ .408 \end{pmatrix}.$$

Thus the second row of $\widehat{A}$ implies the estimator of τ_i is

$$\widehat{\tau_i} = .100[y_M^*]_{i-1} + .409[y_M^*]_i + .100[y_M^*]_{i+1} + .388\frac{1}{T}\sum_{j=1}^{T}[y_M^*]_j.$$

Hence quite a lot of weight is put on neighbouring values of the realised variance and on the intercept, although the weight on $[y_M^*]_i$ is not very much smaller than in the univariate case. The corresponding result for $M = 72$ is

$$\widehat{A} = \begin{pmatrix} .712 & .105 & .053 \\ .105 & .684 & .105 \\ .053 & .105 & .712 \end{pmatrix}, \quad \widehat{c} = \begin{pmatrix} .128 \\ .105 \\ .128 \end{pmatrix}.$$

This shows that the weighting on the diagonal elements of $\widehat{A}$ are much higher, while the size of $\widehat{c}$ has fallen by a factor of around 4.

Figure 6.12 shows selected elements of $\widehat{A}$ for the case of estimating τ_i using 9 realised variances, four lags and four leads together with $[y_M^*]_i$. It displays the weights as a function of M indicating how quickly the weights focus on $[y_M^*]_i$ as M increases. The legend of the Figure also gives the value of the weight put on the unconditional mean of the realised variance. For $M = 72$ it is .0553, which is much lower than in the trivariate case of .105 and univariate case of .221.

Figure 6.13 shows a time series of realised variances for a number of values of M together with the corresponding estimator $\widehat{\tau_i}$ based on nine observations, four leads, the current value and four lags. The smoothed estimator seems to deliver sensible answers, with the results being less sensitive to large values of the realised variances.

Table 6.2 reports, using the DM data,

$$\frac{1}{T}\sum_{i=1}^{T}\left([y_M^*]_i - [y_{288}^*]_i\right)^2,$$

which is an empirical approximation to the mean square error of the realised variance estimator, using $[y_{288}^*]_i$ as a good proxy for τ_i (the model based estimators would turn out to deliver even more accurate estimators, but this could be interpreted as biasing the results towards the model based approach and so here we use the raw realised variance). The Table shows a rapid decline in the mean square error with M. It also shows the corresponding results for the estimators $\widehat{\tau_i}$, based on just a regression on a constant and $[y_M^*]_i$, and $\widetilde{\tau_i}$, which uses $[y_M^*]_{i-4}$, $[y_M^*]_{i-3}$, ...,$[y_M^*]_{i+3}$,$[y_M^*]_{i+4}$. The results reflect the fact that these adjusted estimators are much more efficient than the realised variance, although the difference between using the time series dynamics and the simple regression estimator is modest.

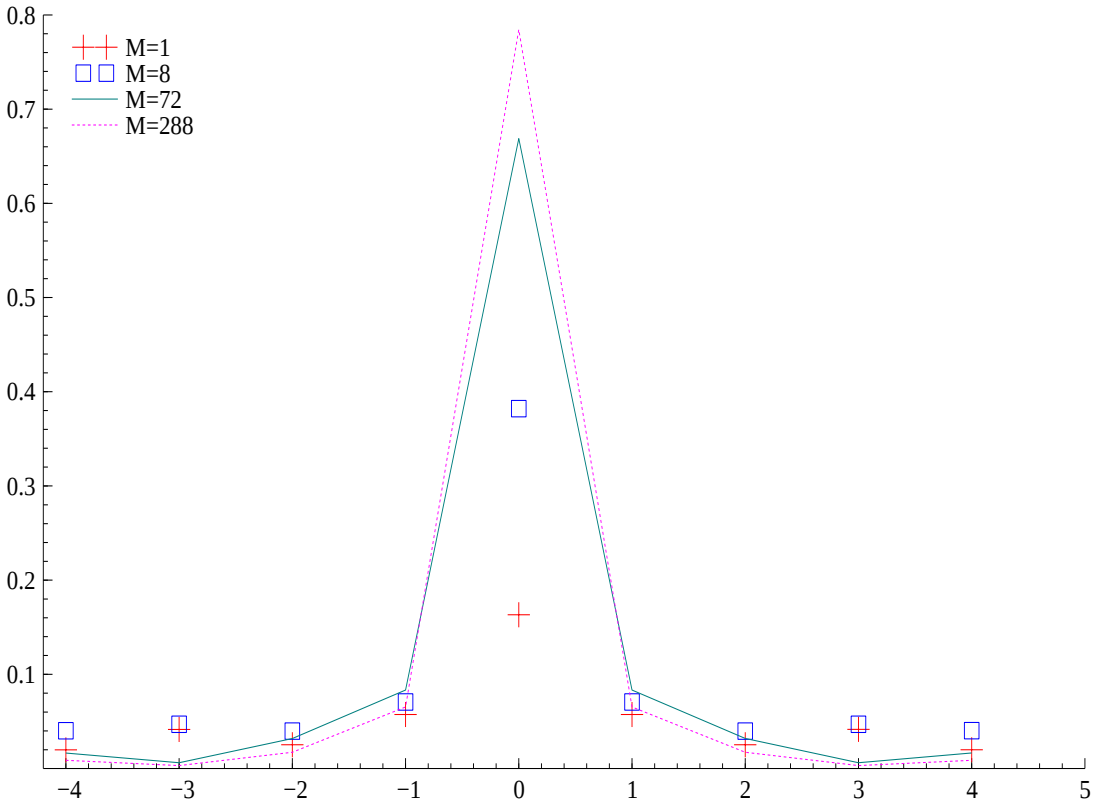

Figure 6.12: *Estimated weight vector for estimating* τ_i *using* $[y_M^*]_{i-4}$,$[y_M^*]_{i-3}$,...,$[y_M^*]_{i+4}$ *drawn against lag length. Computed using the Dollar against the DM. Shows that as* M *increases the weight on* $[y_M^*]_i$ *increases. Corresponding to these results is* $\widehat{c}$*, which moves from* .548, .222, .0553, .026 *as* M *increases through* 1, 8, 72 *to* 288. *File:* `daily_realised.ox`.

A similar style of argument could have been used based on the log-realised variances, where the pooled estimator has the asymptotic distribution

$$\sqrt{\frac{M}{\hbar}}\left(A\log[y_M^*]_{s:p} - A\log\tau_{s:p}\right)|\tau_{s:p}^{[2]},\tau_{s:p} \xrightarrow{d} N\left\{0, 2A\mathrm{E}\begin{pmatrix}\tau_s^{[2]}/\left(\tau_s^2\right) & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \tau_p^{[2]}/\left(\tau_p^2\right)\end{pmatrix}A'\right\},$$

which would allow us to choose A as a least squares estimator of $\log\tau_{s:p}$ repeating the above argument. Weighting based on the log-realised variances has the advantage that the Monte Carlo evidence suggests that the asymptotics for the log-realised variance is accurate with the errors being approximately homoskedastic which suggests the weighting will be more effective.

The important result that we need to use this result is that

$$\left\{\frac{1}{T}\sum_{i=1}^{T}\frac{\frac{M}{3\hbar}\sum_{j=1}^{M}y_{j,i}^4}{\left(\sum_{j=1}^{M}y_{j,i}^2\right)^2}\right\} \xrightarrow{p} \mathrm{E}\left(\frac{\tau_i^{[2]}}{\tau_i^2}\right),$$

and hence

$$\widehat{\log\tau_{s:p}} = \widehat{c}\mathrm{E}\left(\log\tau_{s:p}\right) + \widehat{A}\log[y_M^*]_{s:p}.$$

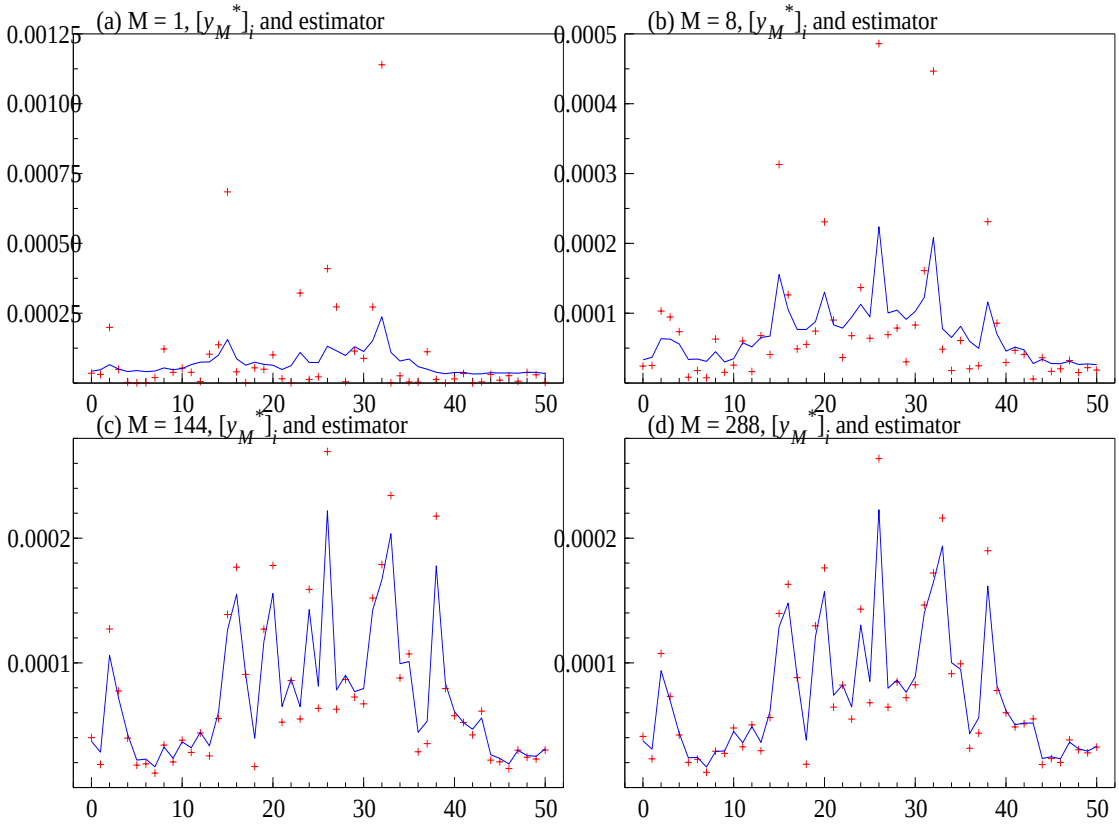

Figure 6.13: *The estimated τ_i using realised variance and weighted version of $[y_M^*]_{i-4}$,$[y_M^*]_{i-3}$,...,$[y_M^*]_{i+4}$. Computed using the Dollar against the DM. (a) M=1, (b) M=8, (c) M=144 and (d) M=288. File:* `daily_realised.ox`.

Of course

$$\frac{1}{T}\sum_{i=1}^{T}\log[y_M^*]_i \xrightarrow{p} \mathrm{E}\left(\log\tau_i\right),$$

hence we are left with just determining $\widehat{c}$ and $\widehat{A}$. If we assume that the realised variances are a covariance stationary process then the weighted least squares estimator of $\log\tau_{s:p}$ sets

$$\widehat{c} = \left(I - \widehat{A}\right)\iota$$

and

$$\begin{aligned}\widehat{A} &= \left[\mathrm{Cov}\left(\log\tau_{s:p}\right) + \frac{2\hbar\mathrm{E}\left(\tau_i^{[2]}/\tau_i^2\right)}{M}I\right]^{-1}\mathrm{Cov}\left(\log\tau_{s:p}\right)\\ &= \left[\mathrm{Cov}\left(\log[y_M^*]_{s:p}\right)\right]^{-1}\mathrm{Cov}\left(\log\tau_{s:p}\right).\end{aligned}$$

Of course for this estimator

$$\widehat{\log\tau_{s:p}} \to \tau_{s:p}$$

as $M\to\infty$, as expected.

This style of approach extends to the multivariate case where the focus is on estimating the actual covariance matrix. Then it makes sense to use these regression approaches based

	DM			Yen		
	$[y_M^*]_i$	$(1-\widehat{A})E(\tau_i)+\widehat{A}[y_M^*]_i$	$\widehat{\tau}_i$	$[y_M^*]_i$	$(1-\widehat{A})E(\tau_i)+\widehat{A}[y_M^*]_i$	$\widehat{\tau}_i$
$M=1$	.822	.175	.145	1.16	.198	.168
$M=8$	.207	.0989	.0769	.186	.117	.0985
$M=72$	.0377	.0345	.0317	.0424	.0406	.0378

Table 6.2: *Mean square error of the realised variance and the regression estimator and the time series estimators $\widehat{\tau}_i$, which is based on $[y_M^*]_{i-4}$, $[y_M^*]_{i-3}$, ...,$[y_M^*]_{i+3}$,$[y_M^*]_{i+4}$. These are computed using $M = 1$, 8 and 72. The true value is taken as $[y_M^*]_i$ for 288. File:* `daily_realised.ox`.

on the logs of the realised variances and the Fisher transformation of the realised correlation. The asymptotic theory of the realised covariation allows this approach to be feasible without specifying a parametric model for the spot covariance matrix.

6.9.2 Model based approach

General discussion and example

Suppose we write (when they exist) ξ, ω^2 and r, respectively, as the mean, variance and the autocorrelation function of the continuous time stationary variance process τ. Here we will discuss estimating and forecasting τ_i based upon a parametric models for τ and the time series of realised variances . Let us write $u_i = [y_M^*]_i - \tau_i$, then the asymptotic theory tells us that for large M the u_i are approximately uncorrelated with

$$\text{Var}\left(\sqrt{M}u_i\right) \to 2\hbar^2\left(\omega^2+\xi^2\right)$$

as $M \to \infty$. Thus the second order properties of $[y_M^*]_i$ can be approximated. In particular $\text{E}([y_M^*]_i) = \hbar\xi + o(1)$ and

$$\begin{aligned}\text{Var}([y_M^*]_i) &= 2M^{-1}\hbar^2\left(\omega^2+\xi^2\right)+\text{Var}(\tau_i)+o(1),\\ \text{Cov}([y_M^*]_i,[y_M^*]_{i+s}) &= \text{Cov}(\tau_i,\tau_{i+s})+o(1).\end{aligned}$$

$\text{Var}(\tau_i)$ and $\text{Cov}(\tau_i,\tau_{i+s})$ were given for all covariance stationary processes in the previous Chapter on chronometers. In particular

$$\text{Var}(\tau_i) = 2\omega^2 r^{**}(\hbar) \qquad \text{and} \qquad \text{Cov}\{\tau_i,\tau_{i+s}\} = \omega^2 \diamondsuit r^{**}(\hbar s), \tag{6.94}$$

where

$$\diamondsuit r^{**}(s) = r^{**}(s+\hbar) - 2r^{**}(s) + r^{**}(s-\hbar), \tag{6.95}$$

and

$$r^*(t) = \int_0^t r(u)\mathrm{d}u \quad \text{and} \quad r^{**}(t) = \int_0^t r^*(u)\mathrm{d}u. \tag{6.96}$$

Thus, for a given model for the covariance stationary process τ we can compute the second order properties of the time series of $[y_M^*]_i$ and τ_i.

The above theory implies we can calculate asymptotically approximate best linear filtered, smoothed and forecast values of τ_i using standard regression theory. In particular suppose we wish to estimate $\tau_{s:p}$ using $[y_M^*]_{s:p}$. Then the best linear estimator is

$$\widehat{\tau}_{s:p} = \widehat{A}\left\{[y_M^*]_{s:p} - \hbar\xi\iota\right\} + \hbar\xi\iota,$$

where

$$\begin{aligned}\widehat{A} &= \{\mathrm{Cov}([y_M^*]_{s:p})\}^{-1}\mathrm{Cov}\left(\tau_{s:p},[y_M^*]_{s:p}\right)\\ &= \left\{\mathrm{Cov}\left(\tau_{s:p}\right)+2M^{-1}\hbar^2\left(\omega^2+\xi^2\right)I\right\}^{-1}\mathrm{Cov}\left(\tau_{s:p}\right).\end{aligned}$$

This has a variance of

$$2M^{-1}\hbar^2\left(\omega^2+\xi^2\right)\widehat{A}\widehat{A}'+\left(I-\widehat{A}\right)\mathrm{Cov}\left(\tau_{s:p}\right)\left(I-\widehat{A}\right)'.$$

Notice that as $M\to\infty$ so $\widehat{A}\to I$ and $\widehat{\tau}_{s:p}\xrightarrow{p}\tau_{s:p}$. The simplest special case of this is where $s=p=i$, that is we use a single realised variance to estimate actual variance. Then the theory above suggests the efficient linear estimator is constructed using

$$\widehat{X}=\left\{r^{**}(\hbar)+M^{-1}\hbar^2\left(1+\xi^2/\omega^2\right)\right\}^{-1}r^{**}(\hbar)\in[0,1], \tag{6.97}$$

which implies $\widehat{\tau}_i\geq 0$. In econometrics the corresponding $\widehat{\tau}_i$ is called the Meddahi regression estimator of actual variance. It is always a consistent estimator of τ_i, but is more efficient than realised variance under the covariance stationarity assumptions.

In practice it is helpful to use the special structure of the $\mathrm{Cov}\left(\tau_{s:p}\right)$ in order to carry out the required matrix inverse of $\mathrm{Cov}([y_M^*]_{s:p})$.

Special case

Suppose τ has the autocorrelation function $r(t)=\exp(-\lambda|t|)$. This implies that

$$\mathrm{E}\left(\tau_i\right)=\hbar\xi,\qquad \mathrm{Var}\left(\tau_i\right)=2\omega^2\lambda^{-2}\left(e^{-\lambda\hbar}-1+\lambda\hbar\right),$$

and

$$\mathrm{Cor}\{\tau_i,\tau_{i+s}\}=de^{-\lambda\hbar(s-1)},\qquad s=1,2,..., \tag{6.98}$$

where

$$d=\frac{(1-e^{-\lambda\hbar})^2}{2\left(e^{-\lambda\hbar}-1+\lambda\hbar\right)}\leq 1.$$

This implies τ_i has the autocorrelation function of an ARMA$(1,1)$ model. Its autoregressive root is $e^{-\lambda\hbar}$, which will be typically close to one unless $\hbar$ is very large, while the moving average root θ is also determined by $e^{-\lambda\hbar}$ but has to be found numerically. So for this case, in particular, the Meddahi regression has

$$\widehat{A}=\left\{\lambda^{-2}\left(e^{-\lambda\hbar}-1+\lambda\hbar\right)+M^{-1}\hbar^2\left(1+\xi^2/\omega^2\right)\right\}^{-1}\lambda^{-2}\left(e^{-\lambda\hbar}-1+\lambda\hbar\right)$$

This argument extends to the case of a superposition where $r(t)=\sum_{j=1}^J w_j\exp(-\lambda_j|t|)$, then τ_i can be represented as the sum of J ARMA$(1,1)$ processes.

In calculating $\widehat{\tau}_{s:p}$ it is computationally convenient to place $[y_M^*]_i$ into a linear state space representation for the filtering, smoothing and forecasting can be carried out using the Kalman filter. In particular write $\alpha_{1i}=(\tau_i-\hbar\xi)$ and $u_i=\sqrt{2M^{-1}\hbar^2\left(\omega^2+\xi^2\right)}v_{1i}$, then

$$\begin{aligned}[y_M^*]_i &= \hbar\xi+\begin{pmatrix}1 & 0\end{pmatrix}\alpha_i+\sqrt{2M^{-1}\hbar^2\left(\omega^2+\xi^2\right)}v_{1i},\\ \alpha_{i+1} &= \begin{pmatrix}\phi & 1\\ 0 & 0\end{pmatrix}\alpha_i+\begin{pmatrix}\sigma_\sigma\\ \sigma_\sigma\theta\end{pmatrix}v_i,\end{aligned} \tag{6.99}$$

M	$\xi = 0.5$, $\xi\omega^{-2} = 8$			$\xi = 0.5$, $\xi\omega^{-2} = 4$			$\xi = 0.5$, $\xi\omega^{-2} = 2$		
$e^{-\Delta\lambda} = 0.99$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$
1	.0134	.0226	.624	.0209	.0369	.749	.0342	.0625	.998
12	.00383	.00792	.0520	.00586	.0126	.0624	.00945	.0211	.0833
48	.00183	.00430	.0130	.00276	.00692	.0156	.00440	.0116	.0208
288	.000660	.00206	.00217	.000967	.00343	.00260	.00149	.00600	.00347
$e^{-\Delta\lambda} = 0.9$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$	Smooth	Predict	$[y_M^*]_i$
1	.0345	.0456	.620	.0569	.0820	.741	.0954	.148	.982
12	.0109	.0233	.0520	.0164	.0396	.0624	.0259	.0697	.0832
48	.00488	.0150	.0130	.00707	.0260	.0156	.0108	.0467	.0208
288	.00144	.00966	.00217	.00195	.0178	.00260	.00280	.0338	.00347

Table 6.3: *Exact mean square error (steady state) of the estimators of actual volatility. The first two estimators are model based (smoother and predictor) and the third is* $[y_M^*]_i$. *These measures are calculated for different values of* $\omega^2 = \mathrm{Var}(\tau(t))$ *and* λ, *keeping* $\xi = \mathrm{E}(\tau(t))$ *fixed at 0.5. File:* `ssf_mse.ox`.

where v_i is a zero mean, unit variance, white noise sequence. The parameters ϕ, θ and σ_σ^2 represent the autoregressive root, the moving average root and the variance of the innovation to this process. The extension to the superposition case is straightforward.

Table 6.3 reports the mean square error of the model based predictor and smoother of actual variance, as well as the corresponding result for $[y_M^*]_i$. The results in the left hand block of the Table corresponds to the model which was simulated in Figure 6.1, while the other blocks vary the ratio of ξ to ω^2. The exercise is repeated for two values of λ.

The main conclusion from the results in Table 6.3 is that model based approaches can potentially lead to very significant reductions in mean square error, with the reductions being highest for persistent (low value of λ) variance processes with high values of $\xi\omega^{-2}$. Even for moderately large values of M the model based predictor can be more accurate than realised variance , sometimes by a considerable amount. This is an important result from a forecasting viewpoint. However, when there is not much persistence and M is very large, this result is reversed and realised variance can be moderately more accurate. The smoother is always substantially more accurate than realised variance , even when M is very large and there is not much memory in variance. This suggests that model based methods may be particularly helpful in estimating historical records of actual variance. Finally, we should place a number of caveats on these conclusions. The above results represent a somewhat favourable setup for the model based approach. In the above calculations we have assumed knowledge of the second order properties of variance while in practice we will have to build such a model and then estimate it, inducing additional biases that we have not reported on. Of course, these criticisms do not apply to the estimators in the previous subsection which were model free.

Estimating parameters: a numerical illustration

Estimating the parameters of continuous time stochastic volatility models is known to be difficult due to our inability to compute the appropriate likelihood function. This has prompted the development of a sizable collection of methods to deal with this problem. Here we advocate the use of quasi-likelihood estimation methods based on the time series of realised variance . The quasi-likelihood is constructed using the output of the Kalman filter. It is suboptimal for it does not exploit the non-Gaussian nature of the variance dynamics, but it provides a consistent and asymptotically normal set of estimators. This follows from the fact that the Kalman filter builds the Gaussian quasi-likelihood function for the ARMA representation of the process, where the noise in the representation is both white and strong mixing. This means we can immediately apply the asymptotic theory results of Francq and Zakoïan (2000) in this context so long as $\tau(t)$

is strong mixing. Monte Carlo results reported in research papers indicate that the finite sample behaviour of this approach is quite good. Further the estimation takes only a few seconds on a modern notebook computer.

Empirical illustration

To illustrate some of these results we have fitted a set of superposition based models to the realised variance time series constructed from the five minute exchange rate return data discussed above. Here we use the quasi-likelihood method to estimate the parameters of the model — ξ, ω^2, $\lambda_1, ..., \lambda_J$ and $w_1, ..., w_J$. We do this for a variety of values of M, starting with $M = 6$, which corresponds to working with four hour returns. The resulting parameter estimates are given in Table 6.4. For the moment we will focus on this case.

M	J	ξ	ω^2	λ_1	λ_2	λ_3	w_1	w_2	Quasi-L	BP
6	3	0.4783	0.376	0.0370	1.61	246	0.212	0.180	-113,258	11.2
6	2	0.4785	0.310	0.0383	3.76	—	0.262	—	-113,261	11.3
6	1	0.4907	0.358	1.37	—	—	—	—	-117,397	302
18	3	0.460	0.373	0.0145	0.0587	3.27	0.0560	0.190	-101,864	26.4
18	2	0.460	0.533	0.0448	4.17	—	0.170	—	-101,876	26.5
18	1	0.465	0.497	1.83	—	—	—	—	-107,076	443
144	3	0.508	4.79	0.0331	0.973	268	0.0183	0.0180	-68,377	15.3
144	2	0.509	0.461	0.0429	3.74	—	0.212	—	-68,586	23.3
144	1	0.513	0.374	1.44	—	—	—	—	-76,953	765

Table 6.4: *Fit of the superposition of J volatility processes for a SV model based on realised variance computed using $M = 6$, $M = 18$ and $M = 144$. We do not record w_J as this is 1 minus the sum of the other weights. Estimation method: quasi-likelihood using output from a Kalman filter. BP denotes Box–Pierce statistic, based on 20 lags, which is a test of serial dependence in the scaled residuals. File:* `ssf_empirical.ox`.

The fitted parameters suggests a dramatic shift in the fitted model as we go from $J = 1$ to $J = 2$ or 3. The more flexible models allow for a factor which has quite a large degree of memory, as well as a more rapidly decaying component or two. A simple measure of fit of the model is the Box–Pierce statistic, which shows a large jump from a massive 302 when $J = 1$, down to an acceptable number for a superposition model.

To provide a more detailed assessment of the fit of the model we have drawn a series of graphs in Figure 6.14 based on $M = 8$ and $M = 144$. Figure 6.14(a) draws the computed realised variance $[y_M^*]$, together with the corresponding smoothed estimate (based on $J = 3$) of actual variance using the model. These are based on the $M = 8$ case. We can see that realised variance is much more jagged than the smoothed quantity. These are quite close to the semi-parametric estimator given in Figure 6.13. Figure 6.14(b) shows the corresponding autocorrelation function for the realised variance series together with the corresponding empirical correlogram. We see from this figure that when $J = 1$ we are entirely unable to fit the data, as its autocorrelation function starts at around 0.6 and then decays to zero in a couple of days. A superposition of two processes is much better, picking up the longer-range dependence in the data. The superposition of two and three processes give very similar fits, indeed in the graph they are indistinguishable.

We next ask how these results vary as M increases. We reanalyse the situation when $M = 144$, which corresponds to working with ten minute returns. Figure 6.14(c) and (d) gives the corresponding results. Broadly the smoother has not produced very different results, while the $J = 3$ case now gives a slightly different fit to the Acf than the $J = 2$. The latter result is

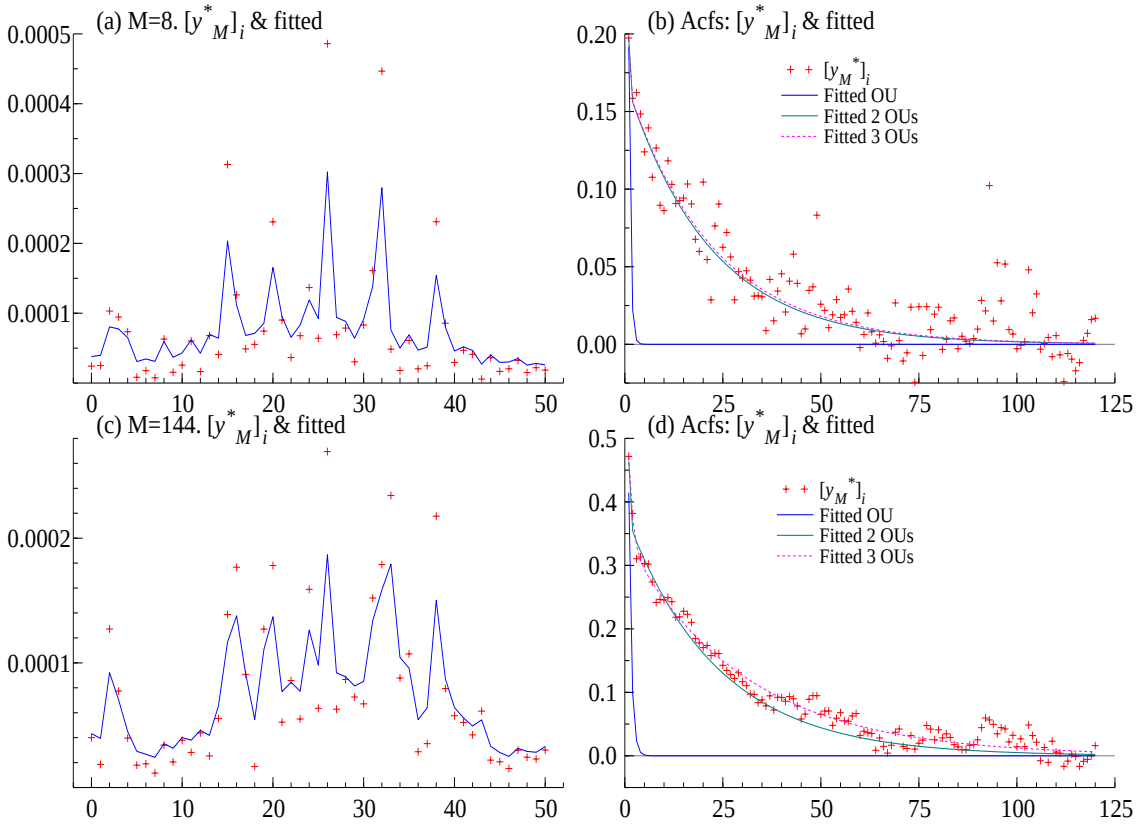

Figure 6.14: *Results from $M = 8$ and $M = 144$. (a) Using M=8, first 50 observations of $[y^*_M]_i$ & smoother. (b) Using M=8, Acf of $[y^*_M]_i$ and the fitted version for various values of J. (c) Using M=144, first 50 observations of $[y^*_M]_i$ & smoother. (d) Using M=144, Acf of $[y^*_M]_i$ and the fitted version for various values of J.* File: ***daily_timeseries.ox***.

of importance, for as M increases the correlogram becomes more informative, allowing us to discriminate between different models more easily.

Table 6.4 contains the estimated parameters for this problem. They suggest that moving to a superposition of three processes has an important impact on the fit of the model. Again the fitted models indicate variance has elements which have a substantial memory, while other components are much more transitory. An important feature of this table is the jump in the value of the estimated ω^2 when we move to having $J = 3$. This is caused by the third component which has a very high value of λ, which does not overly change the variance of actual variance.

6.10 Conclusion

In this Chapter we have reviewed some recent work on realised variance and covariation. Intimately related to probabilistic ideas of quadratic variation and covariation, the quantities have associated distribution theories allowing us to perform econometric inference on derived quantities such as realised regressions, correlations and asset allocation.

The asymptotic theory for these realised quantities looks, at first sight, rather familiar from variance and covariance estimation of Gaussian i.i.d. models, however a deeper understanding of the results shows they are subtly different in interesting ways. Time series of realised quantities can be pooled to allow us to reduce the amount of noise associated with estimated actual

variances and covariances, however this typically biases the resulting estimator. We discuss how to optimally trade the bias against the variance in a number of practical situations. We show that time series of realised variances can be used to estimate parametric stochastic volatility models in computationally simple and precise ways.

6.11 Bibliographical information

6.11.1 Realised variance and empirical finance

Realised variances and volatilities have been used in financial econometrics for many years. Examples include Poterba and Summers (1986), Schwert (1989), Hsieh (1991), Taylor and Xu (1997) and Christensen and Prabhala (1998). Sums of squared returns are often called realised volatility in econometrics, while we use the name realised variance for that term and realised volatility for the corresponding square root. The use of volatility to denote standard deviations rather than variances is standard in financial economics. See, for example, the literature on volatility and variance swaps, which are derivatives written on realised volatility or variance, which includes Demeterfi, Derman, Kamal, and Zou (1999), Howison, Rafailidis, and Rasmussen (2000) and Chriss and Morokoff (1999). We have chosen to follow this nomenclature rather than the one more familiar in econometrics.

6.11.2 Quadratic variation, realised variance and econometrics

Under very weak assumptions the probability limit of $[y_M^*]_i$ has been known for many years using the theory of quadratic covariation. Textbook expositions include Jacod and Shiryaev (1987, p. 55) and Protter (1990). A discussion of some of the economic implications of quadratic variation is given in Back (1991).

Quadratic variation for semimartingales in the univariate case is discussed in the econometric literature by independent and concurrent work by Comte and Renault (1998), Barndorff-Nielsen and Shephard (2001a) and Andersen and Bollerslev (1998a). It was later developed and applied systematically in some empirical work by Andersen, Bollerslev, Diebold, and Labys (2001a). See also Barndorff-Nielsen and Shephard (2001a) and Andersen, Bollerslev, Diebold, and Labys (2001b) for a discussion of the multivariate case and Andersen, Bollerslev, and Diebold (2002) for an incisive survey of this area. Andersen, Bollerslev, Diebold, and Ebens (2001) discusses the use of the multivariate theory in the context of equity prices.

A significant contribution made by the Andersen, Bollerslev, Diebold, et al research group was their emphasis on looking at the conditional (on past data) mean of realised variance as a good approximation to the conditional variance of returns. This line of thinking has not been the focus of our attention, but is certainly stimulating from a modelling viewpoint.

The asymptotic distribution of realised variance was first given in Barndorff-Nielsen and Shephard (2002c). A discussion of the use of this work on empirical data is given in Barndorff-Nielsen and Shephard (2002d) while the small sample performance of the distribution theory is discussed at length in Barndorff-Nielsen and Shephard (2003). This latter paper also studied the asymptotics and finite sample behaviour of the log of realised variance.

Related to this work is that of Foster and Nelson (1996) (note also the work of Genon-Catalot, Laredo, and Picard (1992), Florens-Zmirou (1993) and Hansen (1995)). In the univariate SV case, where the volatility follows a scalar diffusion, they provided an asymptotic distribution theory for an estimator of $\Sigma(t)$, the spot (not integrated) variance (6.20).

Meddahi (2002) followed Barndorff-Nielsen and Shephard (2002c) in studying the moments of the realised variance error. For a rather general model structure he showed that leverage made no difference to the asymptotic moments of the error. In this paper he also suggested the model based regression estimator of integrated variance that we call the Meddahi regression estimator

(6.97). More sophisticated model based time series estimators were introduced by Barndorff-Nielsen and Shephard (2002c) and further studied by Ysusi (2001). The use of this approach to estimate the parameters of continuous time SV models appears in Barndorff-Nielsen and Shephard (2002c) and Bollerslev and Zhou (2001). The development of a corresponding model free approach seems to be new.

Other papers on various econometric aspects of realised variance includes Maheu and McCurdy (2001), Areal and Taylor (2002), Bollerslev and Forsberg (2002) and Fleming, Kirby, and Ostdiek (2002).

6.11.3 Quadratic covariation

The theory of quadratic covariation is discussed in Jacod and Shiryaev (1987, p. 55) and Protter (1990), while the economics is again emphasised in Back (1991). Andersen, Bollerslev, Diebold, and Labys (2001a) informally discussed some of the econometric implications of this theory for realised covariation, but this is carried out more precisely in Andersen, Bollerslev, and Diebold (2002) and Barndorff-Nielsen and Shephard (2002e).

Some literature on the topic of SV based factor models includes Diebold and Nerlove (1989), Meddahi and Renault (1996), Meddahi and Renault (2002), Pitt and Shephard (1999), Chib, Nardari, and Shephard (1999), Barndorff-Nielsen and Shephard (2001a, Section 6.5) and Hubalek and Nicolato (2001).

Barndorff-Nielsen and Shephard (2002e) derived the asymptotic distribution of realised covariation and implied the distribution theory for realised regression, correlation, etc.

6.11.4 Model based estimation of integrated variance

An important, but more intricate area, is where we use the covariance properties of the time series of realised variances to improve the estimation of integrated variance. This was discussed around the same time by Barndorff-Nielsen and Shephard (2002c) and Andreou and Ghysels (2001), from rather different standpoints. Both sets of authors argued there were potentially very significant efficiency gains to be made by using this approach. A important subsequent contribution was Meddahi (2002) who extended some of the results of Barndorff-Nielsen and Shephard (2002c) to cover the leverage case.

Chapter 7

Power variation

Abstract: This Chapter looks at a generalisation of quadratic variation called power variation. This is the probability limit of a scaled sum of powers of absolute returns. In financial econometrics we use realised variance to approximate increments to quadratic variation, while in the more general case we employ realised power variation. We are able to establish conditions under which derive the asymptotic distribution for realised power variation. When we use low powers, such as absolute values themselves, there is some hope that this measure of variability may be more robust to jumps than realised variance. We provide a theory for why this should be the case.

7.1 What is this Chapter about?

7.2 Introduction

Stochastic volatility processes play an important role in financial economics, generalising Brownian motion to allow the scale of the increments (or returns in economics) to change through time in a stochastic manner. We show such intermittency can be coherently measured using sums of absolute powers of increments, which we name realised power variation. This paper derives limit theorems for these measures, over a fixed interval of time, as the number of high frequency increments goes off to infinity.

A referee has drawn our attention to an unpublished thesis by Etienne Becker (1998) that develops a number of results that are closely related to those of the present paper. We outline the relation to Becker's work in the conclusion of this paper.

This paper has six other sections. In Section 2 we establish our idea of realised power variation as well as define the regularity assumptions we need to derive our limit theorems. Section 3 contains our main results, while the proofs of them are given in Section 4. Section 5 gives some examples of the processes covered by our theory, while a Monte Carlo experiment to assess the accuracy of our limit theory approximations is conducted in Section 6. Finally, Section 7 gives some concluding remarks including a discussion of the use of these ideas in other areas of study, for instance turbulence and image analysis.

7.3 Models, notation and regularity conditions

We first introduce some notation for realised power variation quantities of an arbitrary semimartingale x. Let δ be positive real and, for any $t \geq 0$, define

$$x_\delta(t) = x(\lfloor t/\delta \rfloor \, \delta),$$

where $\lfloor a \rfloor$ for any real number a denotes the largest integer less than or equal to a. The process $x_\delta(t)$ is a discrete approximation to $x(t)$. Further, for r positive real we define the *realised power variation of order* r[1] or *realised* r*-tic variation* of $x_\delta(t)$ as

$$[x_\delta]^{[r]}(t) \quad = \quad \sum_{j=1}^{M} |x_\delta\,(j\delta) - x_\delta((j-1)\delta)|^r$$

[1]The similarly named p-variation, $0 < p < \infty$, of a real-valued function f on $[a, b]$ is defined as

$$\sup_\kappa \sum |f(x_i) - f(x_{i-1})|^p \,,$$

where the supremum is taken over all subdivisions κ of $[a, b]$. If this function is finite then f is said to have bounded p-variation on $[a, b]$. The case of $p = 1$ gives the usual definition of bounded variation.

This concept has been studied recently in the probability literature. See the work of, for example, Lyons (1994) and Mikosch and Norvaisa (2000).

$$= \sum_{j=1}^{M} |x(j\delta) - x((j-1)\delta)|^r \tag{7.1}$$

where $M = M(t) = \lfloor t/\delta \rfloor$. Then, in particular, for $M \to \infty$, *realised quadratic variation*

$$[x_\delta]^{[2]}(t) \xrightarrow{p} [x](t),$$

where $[x]$ is the *quadratic variation* process of the semimartingale x. Note also that,

$$[x_\delta]^{[2]} = [x_\delta].$$

Henceforth, for simplicity of exposition, we fix t and take δ so that $M = \lfloor t/\delta \rfloor$ is an integer (and then $\delta M = t$).

Our detailed results will be established for the stochastic volatility (SV) model where basic Brownian motion is generalised to allow the volatility term to vary over time and there to be a rather general drift. Then the y^* follows

$$y^*(t) = \alpha(t) + \int_0^t \sigma(s)\mathrm{d}w(s), \qquad t \geq 0, \tag{7.2}$$

where $\sigma > 0$ and α are assumed to be stochastically independent of the standard Brownian motion w. Throughout this paper we will assume that the processes $\tau = \sigma^2$ and α are of locally bounded variation. This implies that τ and α are locally bounded Riemann integrable functions and that y^* is a semimartingale with a continuous local martingale component. We call σ the *spot volatility* process and α the *mean* or *risk premium* process. (For some general information on processes y^* of this type, see for example Ghysels, Harvey, and Renault (1996) and Barndorff-Nielsen and Shephard (2001a)). By allowing the spot volatility to be random and serially dependent, this model will imply its increments will exhibit volatility clustering and have unconditional distributions which are fat tailed. This allows it to be used in finance and econometrics as a model for log-prices. In turn, this provides the basis for option pricing models which overcome some of the major failings in the Black-Scholes option pricing approach. Leading references in this regard include Hull and White (1987), Heston (1993) and Renault (1997). See also the recent work of Nicolato and Venardos (2001).

For the price process (7.2) the *realised power variation of order* r of y^* is, at time t and discretisation δ, $[y^*_\delta]^{[r]}(t)$. Letting

$$y_j(t) = y^*(j\delta) - y^*((j-1)\delta)$$

we have that

$$[y^*_\delta]^{[r]}(t) = \sum_{j=1}^{M} |y_j(t)|^r .$$

We use the notation $\tau(t) = \sigma^2(t)$ and

$$\tau^*(t) = \int_0^t \tau(s)\mathrm{d}s$$

and, more generally, we consider the *integrated power volatility of order* r

$$\tau^{r*}(t) = \int_0^t \tau^r(s)\mathrm{d}s.$$

That τ^r is Riemann integrable for every $r > 0$ follows from the assumed locally bounded variation of τ and the fact, due to Lebesgue, that a bounded function f on a finite interval I is Riemann

integrable on I if and only if the Lebesgue measure of the set of discontinuity points of f is equal to 0 (see Hobson (1927, pp. 465–466), Munroe (1953, p. 174, Theorem 24.4) or Lebesgue (1902)). In our case the latter property follows immediately from the bounded variation of τ (any function of bounded variation is the difference between an increasing and a decreasing function and any monotone function has at most countably many discontinuities).

Throughout the following, r denotes a positive number. Moreover we shall refer to the following conditions on the volatility and mean processes:

(V) The volatility process $\tau = \sigma^2$ is (pathwise) locally bounded away from 0 and has, moreover, the property

$$p\text{-}\lim_{\delta \downarrow 0} \delta^{1/2} \sum_{j=1}^{M} |\tau^r(\eta_j) - \tau^r(\xi_j)| = 0 \tag{7.3}$$

for some $r > 0$ (equivalently for all $r > 0$)[2] and for any ξ_j and η_j such that

$$0 \leq \xi_1 \leq \eta_1 \leq \delta \leq \xi_2 \leq \eta_2 \leq 2\delta \leq \cdots \leq \xi_j \leq \eta_j \leq M\delta = t.$$

(M) The mean process α satisfies (pathwise)[3]

$$\overline{\lim_{\delta \downarrow 0}} \max_{1 \leq j \leq M} \delta^{-1} |\alpha(j\delta) - \alpha((j-1)\delta)| < \infty. \tag{7.4}$$

These regularity conditions are quite mild.[4] Of some special interest are cases where α is of the form

$$\alpha(t) = \int_0^t g(\sigma(s))\mathrm{d}s,$$

for g a smooth function. Then regularity of τ will imply regularity of α.

Note that the assumptions allow the spot volatility to have, for example, deterministic diurnal effects, jumps, long memory, no unconditional mean or to be non-stationary.

7.4 Results

Our main theoretical result is

Theorem 1 For $\delta \downarrow 0$ and $r \geq 1/2$, under conditions (**V**) and **(M)**,

$$\mu_r^{-1} \delta^{1-r/2} [y_\delta^*]^{[r]}(t) \xrightarrow{p} \tau^{r/2*}(t) \tag{7.5}$$

and

$$\frac{\mu_r^{-1} \delta^{1-r/2} [y_\delta^*]^{[r]}(t) - \tau^{r/2*}(t)}{\mu_r^{-1} \delta^{1-r/2} \sqrt{\mu_{2r}^{-1} v_r [y_\delta^*]^{[2r]}(t)}} \xrightarrow{\mathcal{L}} N(0,1), \tag{7.6}$$

[2]The equivalence follows on noting that for each j there exists an ω_j with

$$\inf_{(j-1)\delta \leq s \leq j\delta} \tau(s) \leq \omega_j \leq \sup_{(j-1)\delta \leq s \leq j\delta} \tau(s)$$

such that

$$\left|\tau^r(\eta_j) - \tau^r(\xi_j)\right| = r\omega_j^{r-1} \left|\tau(\eta_j) - \tau(\xi_j)\right|$$

and then using that τ is pathwise bounded away from 0 and ∞.

[3]This condition is implied by Lipschitz continuity and itself implies continuity of α.

[4]Condition **(V)** is satisfied in particular if τ is of OU type and condition **(M)** is valid if, for instance, α is the intOU process plus drift.

where $\mu_r = \mathrm{E}\{|u|^r\}$ and $v_r = \mathrm{Var}\{|u|^r\}$, with $u \sim N(0,1)$.

This theorem tells us that, for $\delta \downarrow 0$, scaled realised power variation converges in probability to integrated power volatility and follows asymptotically a normal variance mixture distribution with variance distributed as

$$\delta \mu_r^{-2} v_r \tau^{r*}(t),$$

which is consistently estimated by the square of the denominator in (7.6). Hence the limit theory is statistically feasible and does not depend upon knowledge of α or σ^2.

Leading cases are *realised quadratic variation*, which is usually called *realised volatility* in the finance and econometrics literature,

$$[y_\delta^*]^{[2]}(t) = \sum_{j=1}^{M} y_j^2(t),$$

in which case

$$\frac{\sum_{j=1}^{M} y_j^2(t) - \tau^*(t)}{\sqrt{\frac{2}{3}\sum_{j=1}^{M} y_j^4(t)}} \xrightarrow{\mathcal{L}} N(0,1), \tag{7.7}$$

and *realised absolute variation*

$$[y_\delta^*]^{[1]}(t) = \sum_{j=1}^{M} |y_j(t)|,$$

when

$$\frac{\sqrt{\pi/2}\sqrt{\delta}\sum_{j=1}^{M} |y_j(t)| - \sigma^*(t)}{\sqrt{(\pi/2-1)\,\delta\sum_{j=1}^{M} y_j^2(t)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{7.8}$$

In the case of $r = 2$ the result considerably strengthens the well known quadratic variation result that realised quadratic variation converges in probability to integrated volatility $\int_0^t \sigma^2(s)\mathrm{d}s$ — which was highlighted in concurrent and independent work by Andersen and Bollerslev (1998a) and Barndorff-Nielsen and Shephard (2001a). The asymptotic distribution of realised quadratic variation was discussed by Barndorff-Nielsen and Shephard (2002c) in the special case where $\alpha(t) = \mu t + \beta \int_0^t \sigma^2(s)\mathrm{d}s$. To our knowledge the probability limit of (normalised) realised absolute variation has not been previously derived, let alone its asymptotic distribution.

Taking sums of squares of increments of log-prices has a very long tradition in financial economics — see, for example, Poterba and Summers (1986), Schwert (1989), Taylor and Xu (1997), Christensen and Prabhala (1998), Dacorogna, Muller, Olsen, and Pictet (1998), Andersen, Bollerslev, Diebold, and Labys (2001a) and Andersen, Bollerslev, Diebold, and Ebens (2001). However, for a long time no theory was known for the behaviour of such sums outside the Brownian motion case. Since the link to quadratic variation has been made there has been a remarkably fast development in this field. Contributions include Corsi, Zumbach, Muller, and Dacorogna (2001), Andersen, Bollerslev, Diebold, and Labys (2001a), Andersen, Bollerslev, Diebold, and Ebens (2001), Barndorff-Nielsen and Shephard (2002c), Andreou and Ghysels (2001), Bai, Russell, and Tiao (2000), Maheu and McCurdy (2001), Areal and Taylor (2002), Galbraith and Zinde-Walsh (2000), Bollerslev and Zhou (2001) and Bollerslev and Forsberg (2002).

Andersen and Bollerslev (1998b) and Andersen and Bollerslev (1997) empirically studied the properties of $\sum_{j=1}^{M} |y_j(t)|$ computed using sums of absolute values of intra-day returns on speculative assets (many authors in finance have based their empirical analysis on absolute values of returns — see, for example, Taylor (1986, Ch. 2), Cao and Tsay (1992), Ding, Granger, and Engle (1993), West and Cho (1995), Granger and Ding (1995), Jorion (1995), Shiryaev (1999,

Ch. IV) and Granger and Sin (1999)). This was empirically attractive, for using absolute values is less sensitive to possible large movements in high frequency data. There is evidence that if returns do not possess fourth moments then using absolute values rather than squares would be more reliable (see, for example, the work on the distributional behaviour of the correlogram of squared returns by Davis and Mikosch (1998) and Mikosch and Starica (2000)). However, the approach was abandoned in their subsequent work reported in Andersen and Bollerslev (1998a), Andersen, Bollerslev, Diebold, and Ebens (2001) and Andersen, Bollerslev, Diebold, and Labys (2001a) due to the lack of appropriate theory for the sum of absolute returns as $\delta \downarrow 0$, although recently Andreou and Ghysels (2001) have performed some interesting Monte Carlo studies in this context, while Shiryaev (1999, pp. 349–350) and Maheswaran and Sims (1993) mention interests in the limit of sums of absolute returns. Our work provides a theory for the use of sums of absolute returns.

7.5 Proofs

Since the mean and volatility processes α and τ are jointly independent of the Brownian motion w we need only argue conditionally on (α, τ). Define α_j, τ_j and σ_j by

$$\alpha_j = \alpha(j\delta) - \alpha((j-1)\delta),$$

$$\tau_j = \tau^*(j\delta) - \tau^*((j-1)\delta)$$

and

$$\sigma_j = \sqrt{\tau_j}.$$

As a preliminary step we show

Lemma 1 For $\delta \to 0$,

$$\delta^{1-r}[\tau^*_\delta]^{[r]}(t) \to \tau^{r*}(t), \tag{7.9}$$

pathwise.

PROOF By the boundedness of $\tau(t)$ we have that for every $j = 1, ..., M$ there exists a constant θ_j such that

$$\inf_{(j-1)\delta \le s \le j\delta} \tau(s) \le \theta_j \le \sup_{(j-1)\delta \le s \le j\delta} \tau(s)$$

and

$$\tau_j = \theta_j \delta, \tag{7.10}$$

and using this and the Riemann integrability of $\tau^r(t)$ we obtain

$$\begin{aligned}
\delta^{1-r}[\tau^*_\delta]^{[r]} &= \delta^{1-r}\sum_{j=1}^{M} \tau_j^r \\
&= \sum_{j=1}^{M} (\tau_j/\delta)^r \delta \\
&= \sum_{j=1}^{M} \theta_j^r \delta \\
&\to \int_0^t \tau^r(s)\mathrm{d}s \\
&= \tau^{r*}(t).
\end{aligned}$$

It is now convenient to write $y^*(t)$ as

$$y^*(t) = \alpha(t) + y_0^*(t),$$

where

$$y_0^*(t) = \int_0^t \sigma(s)\mathrm{d}w(s)$$

and to introduce

$$y_{0j} = y_0^*(j\delta) - y_0^*((j-1)\delta).$$

The joint law of $y_{01}, ..., y_{0M}$ is equal to that of $v_1, ..., v_M$ where

$$v_j = \sigma_j u_j$$

and $u_1, ..., u_M$ are i.i.d. standard normal and independent of the process σ. It follows that

$$[y_{0\delta}^*]^{[r]}(t) \stackrel{\mathcal{L}}{=} \sum_{j=1}^{M} \tau_j^{r/2} |u_j|^r . \tag{7.11}$$

The conditional mean and variance of $[y_{0\delta}^*]^{[r]}(t)$ are then

$$\mathrm{E}\{[y_{0\delta}^*]^{[r]}(t)|\tau\} = \mu_r \sum_{j=1}^{M} \tau_j^{r/2} = \mu_r [\tau_\delta^*]^{[r/2]}(t) \tag{7.12}$$

and

$$\mathrm{Var}\{[y_{0\delta}^*]^{[r]}(t)|\tau\} = v_r \sum_{j=1}^{M} \tau_j^{r} = v_r [\tau_\delta^*]^{[r]}(t). \tag{7.13}$$

Hence, letting

$$D_0 = \mu_r^{-1}[y_{0\delta}^*]^{[r]}(t) - [\tau_\delta^*]^{[r/2]}(t)$$

we have

$$\mathrm{E}\{D_0|\tau\} = 0$$

and

$$\mathrm{Var}\{D_0|\tau\} = \mu_r^{-2} v_r [\tau_\delta^*]^{[r]}(t).$$

By Lemma 1 as $\delta \to 0$,

$$\delta^{1-r}\mathrm{Var}\{D_0|\tau\} \to \mu_r^{-2} v_r \tau^{r*}(t)$$

indicating the validity of

Proposition 1 For $\delta \to 0$,

$$\delta^{(1-r)/2} \frac{\mu_r^{-1}[y_{0\delta}^*]^{[r]}(t) - [\tau_\delta^*]^{[r/2]}(t)}{\sqrt{\mu_r^{-2} v_r \tau^{r*}(t)}} \stackrel{\mathcal{L}}{\to} N(0,1). \tag{7.14}$$

PROOF To establish this proposition we recall Taylor's formula with remainder term:

$$f(x) = f(0) + f'(0)x + x^2 \int_0^1 (1-s) f''(sx)\mathrm{d}s. \tag{7.15}$$

By (7.11) and (7.12) we find, using (7.15), that the conditional cumulant transform of D_0 is of the form

$$\begin{aligned}\log \mathrm{E}\{\exp(i\zeta D_0)|\tau\} &= \zeta^2 \sum_{j=1}^{M} \tau_j^r \int_0^1 (1-s)\kappa_r''(\tau_j^{r/2}\zeta s)\mathrm{d}s \\ &= \zeta^2 \delta^r \sum_{j=1}^{M} \theta_j^r \int_0^1 (1-s)\kappa_r''(\delta^{r/2}\theta_j^{r/2}\zeta s)\mathrm{d}s,\end{aligned}$$

where θ_j is given by (7.10) and κ_r denotes the cumulant transform of $\mu_r^{-1}|u|^r$ for u a standard normal random variable. Consequently,

$$\log \mathrm{E}\{\exp(i\zeta\delta^{(1-r)/2}D_0)|\tau\} = \frac{1}{2}\zeta^2\delta R$$

where

$$R = 2\sum_{j=1}^{M} \theta_j^r \int_0^1 (1-s)\kappa_r''(\delta^{1/2}\theta_j^{r/2}\zeta s)\mathrm{d}s.$$

The boundedness of τ on $[0,t]$ implies

$$\lim_{\delta\downarrow 0} \delta^{1/2} \max_j \theta_j^{r/2} = 0$$

and hence, for $\delta \downarrow 0$,

$$\begin{aligned}\delta R &\to 2\int_0^1 (1-s)\mathrm{d}s\kappa_r''(0) \lim_{\delta\downarrow 0} \sum_{j=1}^{M} \theta_j^r \delta \\ &= \kappa_r''(0)\tau^{r*}(t) = -\mu_r^{-2} v_r \tau^{r*}(t).\end{aligned}$$

Therefore

$$\log \mathrm{E}\{\exp(i\zeta\delta^{(1-r)/2}D_0)|\tau\} = -\frac{1}{2}\zeta^2\mu_r^{-2}v_r\tau^{r*}(t) + o(1) \tag{7.16}$$

and Proposition 1 follows.

Lemma 1 uses only the local boundedness and Riemann integrability of τ. Invoking condition **(V)** we may strengthen the result (7.9) as follows.

Lemma 2 Under condition **(V)** we have

$$\delta^{1-r}[\tau_\delta^*]^{[r]}(t) - \tau^{r*}(t) = o_p(\delta^{1/2}).$$

PROOF For each j there exists a number ψ_j such that

$$\inf_{(j-1)\delta\le s\le j\delta} \tau(s) \le \psi_j \le \sup_{(j-1)\delta\le s\le j\delta} \tau(s)$$

and

$$\int_{(j-1\delta)}^{j\delta} \tau^r(s)\mathrm{d}s = \psi_j^r\delta. \tag{7.17}$$

Using this and (7.10) we find

$$\begin{aligned}\delta^{1-r}[\tau_\delta^*]^{[r]}(t) - \tau^{r*}(t) &= \delta^{1-r}\sum_{j=1}^{M}\tau_j^r - \int_0^t \tau^r(s)\mathrm{d}s \\ &= \delta\sum_{j=1}^{M}(\theta_j^r - \psi_j^r), \end{aligned} \tag{7.18}$$

and the conclusion now follows from assumption **(V)**.

Proposition 2 Under condition **(V)**, for $\delta \to 0$,

$$\frac{\delta^{1-r/2}\mu_r^{-1}[y_{0\delta}^*]^{[r]}(t) - \tau^{r/2*}(t)}{\delta^{1/2}\sqrt{\mu_r^{-2}v_r\tau^{r*}(t)}} \xrightarrow{\mathcal{L}} N(0,1). \tag{7.19}$$

Proof We may rewrite the left hand side of (7.14) as

$$\begin{aligned}\frac{\delta^{1-r/2}\mu_r^{-1}[y_{0\delta}^*]^{[r]}(t) - \delta^{1-r/2}[\tau_\delta^*]^{[r/2]}(t)}{\delta^{1/2}\sqrt{\mu_r^{-2}v_r\tau^{r*}(t)}} &= \frac{\delta^{1-r/2}\mu_r^{-1}[y_{0\delta}^*]^{[r]}(t) - \tau^{r/2*}(t)}{\delta^{1/2}\sqrt{\mu_r^{-2}v_r\tau^{r*}(t)}} \\ &\quad + \frac{\delta^{1-r/2}[\tau_\delta^*]^{[r/2]}(t) - \tau^{r/2*}(t)}{\delta^{1/2}\sqrt{\mu_r^{-2}v_r\tau^{r*}(t)}}\end{aligned}$$

and Lemma 2 then implies the result.

As an immediate consequence of Proposition 2 we have

Corollary 1 Under condition **(V)**, for $\delta \to 0$,

$$\delta^{1-r}\mu_{2r}^{-1}[y_{0\delta}^*]^{[2r]}(t) \xrightarrow{p} \tau^{r*}(t).$$

In other words, when normalised, $[y_{0\delta}^*]^{[2r]}(t)$ provides a consistent estimate of $\tau^{r*}(t)$. Combining this with (7.19) yields the conclusion of Theorem 1 for the special case where the mean process α is identically 0.

The remaining task is to show that, to the order concerned, α does not affect the asymptotic limit behaviour, provided conditions **(V)** and **(M)** hold. For this it suffices to show that, under **(V)** and **(M)**,

$$\delta^{(1-r)/2}\left\{[y_\delta^*]^{[r]}(t) - [y_{0\delta}^*]^{[r]}(t)\right\} = o_p(1),$$

cf. Proposition 1.

We shall in fact prove the following stronger result.

Proposition 3 Under conditions **(V)** and **(M)**,

$$\delta^{-r/2}\left\{[y_\delta^*]^{[r]}(t) - [y_{0\delta}^*]^{[r]}(t)\right\} = O_p(1).$$

PROOF Let

$$\underline{\tau} = \inf_{0\le s\le t} \tau(s) \quad \text{and} \quad \bar{\tau} = \sup_{0\le s\le t} \tau(s)$$

and

$$\gamma_j = \delta^{-1}\alpha_j$$

and note that (pathwise for (α, τ)), by assumption,

$$0 < \underline{\tau} \le \bar{\tau} < \infty,$$

implying

$$0 < \min_j \theta_j \le \max_j \theta_j < \infty,$$

while, due to assumption **(M)**, there exists a constant c for which

$$\max_j |\gamma_j| \le c,$$

whatever the value of M.

We have

$$\begin{aligned}
[y^*_\delta]^{[r]}(t) - [y^*_{0\delta}]^{[r]}(t) &= \sum_{j=1}^{M} (|\alpha_j + y_{0j}|^r - |y_{0j}|^r) \\
&= \sum_{j=1}^{M} \left(|\delta\gamma_j + \delta^{1/2}\theta_j^{1/2} u_j|^r - |\delta^{1/2}\theta_j^{1/2} u_j|^r \right) \\
&= \delta^{r/2} \sum_{j=1}^{M} \theta_j^{r/2} \left\{ \left| \left(\gamma_j/\theta_j^{1/2} \right) \delta^{1/2} + u_j \right|^r - |u_j|^r \right\}
\end{aligned}$$

and hence

$$\delta^{-r/2} \left\{ [y^*_\delta]^{[r]}(t) - [y^*_{0\delta}]^{[r]}(t) \right\} \stackrel{\mathcal{L}}{=} \sum_{j=1}^{M} \theta_j^{r/2} h_r \left(u_{0j}; \gamma_j/\theta_j^{1/2} \right)$$

where

$$h_r(u;\rho) = \left| \rho\delta^{1/2} + u \right|^r - |u|^r.$$

The conclusion of Proposition 3 now follows from Lemma 3 below.

Lemma 3 For $r \ge 1/2$, u a standard normal random variable and ρ constant,

$$\mathrm{E}\{h_r(u;\rho)\} = O(\delta)$$

and

$$\mathrm{Var}\{h_r(u;\rho)\} = O(\delta).$$

PROOF With φ denoting the standard normal density we obtain

$$\begin{aligned}
\mathrm{E}\{|\rho\delta^{1/2} + u|^r\} &= \int_{-\infty}^{\infty} \left| \rho\delta^{1/2} + x \right|^r \varphi(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty} |x|^r \varphi(x) e^{\rho\delta^{1/2}x} \mathrm{d}x e^{-\rho^2\delta/2} \\
&= \int_{-\infty}^{\infty} |x|^r \varphi(x) e^{\rho\delta^{1/2}x} \mathrm{d}x + O(\delta) \\
&= \mathrm{E}\{|u|^r\} + \int_{-\infty}^{\infty} |x|^r \varphi(x) \left(e^{\rho\delta^{1/2}x} - 1 \right) \mathrm{d}x + O(\delta),
\end{aligned}$$

i.e.

$$\begin{aligned}
\mathrm{E}\{h_r(u;\rho)\} &= \delta^{1/2}\int_{-\infty}^{\infty} x|x|^r\varphi(x)\frac{e^{\rho\delta^{1/2}x}-1}{\delta^{1/2}x}\mathrm{d}x + O(\delta) \\
&= \delta^{1/2}\int_{-\infty}^{\infty} x|x|^r\varphi(x)\frac{e^{\rho\delta^{1/2}x}-1-\rho\delta^{1/2}x}{\delta^{1/2}x}\mathrm{d}x + \rho\delta^{1/2}\int_{-\infty}^{\infty} x\,|x|^r\,\varphi(x)\mathrm{d}x + O(\delta) \\
&= O(\delta).
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\mathrm{E}\{h_r(u;\rho)^2\} &= \mathrm{E}\{h_{2r}(u;\rho)\} + 2\mathrm{E}\left\{|u|^r\left(|u|^r - |\rho\delta^{1/2}+u|^r\right)\right\} \\
&= O(\delta) + 2\mathrm{E}\left\{|u|^r\left(|u|^r - |\rho\delta^{1/2}+u|^r\right)\right\} \qquad (7.20)
\end{aligned}$$

by the previous result. Here

$$\begin{aligned}
\mathrm{E}\{|u|^r|\rho\delta^{1/2}+u|^r\} &= \int_{-\infty}^{\infty}|x|^r|\rho\delta^{1/2}+x|^r\varphi(x)\mathrm{d}x \\
&= \int_{-\infty}^{\infty}|x^2-\frac{1}{4}\rho^2\delta|^r\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x e^{-\frac{1}{8}\rho^2\delta} \\
&= \int_{-\infty}^{\infty}|x^2-\frac{1}{4}\rho^2\delta|^r\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x + O(\delta) \\
&= \int_{-\infty}^{\infty}|x|^{2r}\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x + O(\delta) \\
&\quad + \int_{-\infty}^{\infty}\left(\left|x^2-\frac{1}{4}\rho^2\delta\right|^r - |x^2|^r\right)\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x. \qquad (7.21)
\end{aligned}$$

Now, for a and b nonnegative numbers we have the inequality

$$|b^r - |b-a|^r| \le \begin{cases} a^r & \text{for} \quad 0 \le b \le a \\ rb^{r-1}a & \text{for} \quad b > a. \end{cases} \qquad (7.22)$$

Using this and $r \ge 1/2$ we find that

$$\begin{aligned}
&\left|\int_{-\infty}^{\infty}\left(|x^2-\frac{1}{4}\rho^2\delta|^r - |x^2|^r\right)\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x\right| \\
\le\ &\left(\frac{r}{4}\rho^2\delta\right)^r\int_{|x|\le\rho\delta^{1/2}/2}\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x + \frac{r}{4}\rho^2\delta\int_{-\infty}^{\infty}|x|^{2(r-1)}\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x \qquad (7.23) \\
=\ &O(\delta). \qquad (7.24)
\end{aligned}$$

Thus, combining (7.20), (7.21) and (7.24),we have

$$\begin{aligned}
\mathrm{E}\{h_r(u;\rho)^2\} &= 2\left\{\int_{-\infty}^{\infty}|x|^{2r}\varphi(x)\mathrm{d}x - \int_{-\infty}^{\infty}|x|^{2r}\varphi(x)e^{\frac{1}{2}\rho\delta^{1/2}x}\mathrm{d}x\right\} + O(\delta) \\
&= 2\int_{-\infty}^{\infty}|x|^{2r}\varphi(x)\left(1-e^{\frac{1}{2}\rho\delta^{1/2}x}\right)\mathrm{d}x + O(\delta) \\
&= 2\delta^{1/2}\int_{-\infty}^{\infty}x|x|^{2r}\varphi(x)\frac{1-e^{\frac{1}{2}\rho\delta^{1/2}x}}{\delta^{1/2}x}\mathrm{d}x + O(\delta) \\
&= O(\delta),
\end{aligned}$$

as was to be shown.

7.6 Examples

The following two examples show that conditions **(V)** and **(M)** are satisfied for OU models used by Barndorff-Nielsen and Shephard (2001a) in the context of SV models.

Example 1 *Volatility process τ of OU type.* Without loss of generality we suppose that τ satisfies the equation

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

where z is a subordinator, i.e. a positive Lévy process, which is referred to as the BDLP (background driving Lévy process).

The number of jumps of the BDPL z on the interval $[0,t]$ is at most countable. Let $z_1 \geq z_2 \geq ...$ denote the jump sizes given in decreasing order and let $u_1, u_2, ...$ be the corresponding jump times. Then, for any $s \in [0,t]$,

$$\tau(s) = \tau(0)e^{-\lambda s} + \sum_{n=1}^{\infty} z_n e(s;u_n),$$

where

$$e(s;u) = e^{-\lambda(s-u)}\mathbf{1}_{[u,t]}(s)$$

and we have

$$\sum_{j=0}^{M} |e(j\delta;u) - e((j-1)\delta;u)| \leq 2.$$

Hence

$$\sum_{j=0}^{M} |\tau(j\delta) - \tau((j-1)\delta)| \leq 2\left\{\tau(0) + \sum_{n=1}^{\infty} z_n\right\} = 2\left\{\tau(0) + z(\lambda t)\right\},$$

showing that condition **(V)** is amply satisfied.

Example 2 *OU volatility and intOU risk premium* In this particular case the volatility process τ is as in the previous example and the mean process is of the form

$$\alpha(t) = \mu\delta + \beta\tau^*(t),$$

where μ and β are arbitrary real parameters.

We then have

$$\alpha_j = \delta\{\mu + \beta\theta_j\},$$

implying

$$\max_{1\leq j\leq M} \delta^{-1}|\alpha(j\delta) - \alpha((j-1)\delta)| \leq |\mu| + |\beta|\bar{\tau} < \infty,$$

so that condition **(M)** is indeed satisfied.

7.7 A Monte Carlo experiment

7.7.1 Multiple realised power variations

We have stated the definition and results for realised power variation for a single fixed t. It is clear that the theory can also be applied repeatedly on non-overlapping increments to the

process. Let us write $\Delta > 0$ as a time interval and focus on the n-th such interval. Then define the intra-Δ increments as

$$y_{j,n} = y^*\left((n-1)\Delta + \frac{\Delta j}{M}\right) - y^*\left((n-1)\Delta + \frac{\Delta(j-1)}{M}\right), \qquad \delta = \Delta M^{-1},$$

which allows us to construct the n-th realised power variation

$$[y^*_\delta]^{[r]}_n = \sum_{j=1}^{M} |y_{j,n}|^r .$$

The implication is that

$$\frac{\left[\mu_r^{-1}\left(\Delta M^{-1}\right)^{1-r/2}[y^*_\delta]^{[r]}_n\right] - \int_{(n-1)\Delta}^{n\Delta}\sigma^r(s)\mathrm{d}s}{\mu_r^{-1}\left(\Delta M^{-1}\right)^{1-r/2}\sqrt{\frac{v_r}{\mu_{2r}}[y^*_\delta]^{[2r]}_n}} \xrightarrow{\mathcal{L}} N(0,1).$$

An important observation is that the *realised power variation errors*

$$\left\{\mu_r^{-1}\left(\Delta M^{-1}\right)^{1-r/2}[y^*_\delta]^{[r]}_n\right\} - \sigma^{[r]}_n, \tag{7.25}$$

where $\sigma^{[r]}_n$, the *actual power volatility*, is defined as

$$\sigma^{[r]}_n = \sigma^{r*}(n\Delta) - \sigma^{r*}((n-1)\Delta), \qquad \text{where} \qquad \sigma^{r*}(t) = \int_0^t \sigma^r(s)\mathrm{d}s.$$

will be asymptotically uncorrelated through n, although they will not be independent.

7.7.2 Simulated example

Realised power variation and actual power volatility

The above distribution theory says in particular that realised power variation error will converge in probability to zero as $\delta \downarrow 0$. To see the magnitude of this error we have carried out a simulation. This will allow us to see how accurate our asymptotic analysis is in practice. Throughout we have set the mean process $\alpha(t)$ to zero. Our experiments could have been based on the familiar constant elasticity of variance (CEV) process which is the solution to the SDE

$$\mathrm{d}\sigma^2(t) = -\lambda\left\{\sigma^2(t) - \xi\right\}\mathrm{d}t + \omega\sigma(t)^\eta \mathrm{d}b(\lambda t), \qquad \eta \in [1,2],$$

where b is standard Brownian motion uncorrelated with w. Of course the special cases of $\eta = 1$ delivers the square root process, while when $\eta = 2$ we have Nelson's GARCH diffusion. These models have been heavily favoured by Meddahi and Renault (2002) in the context of SV models. Instead of this we will work with the non-Gaussian Ornstein-Uhlenbeck process, or OU process for short, which is the solution to the stochastic differential equation

$$\mathrm{d}\sigma^2(t) = -\lambda\sigma^2(t)\mathrm{d}t + \mathrm{d}z(\lambda t), \tag{7.26}$$

where z is a subordinator (that is a Lévy process with non-negative increments). These models have been developed in this context by Barndorff-Nielsen and Shephard (2001a). In Figure 7.1(a), (c), (e) we have drawn a curve to represent a simulated sample path of $\sigma^{[2]}_n$ from an OU process where $z(t)$ has a $\Gamma(t4, 8)$ marginal distribution, $\lambda = -\log(0.99)$ and $\Delta = 1$, along with

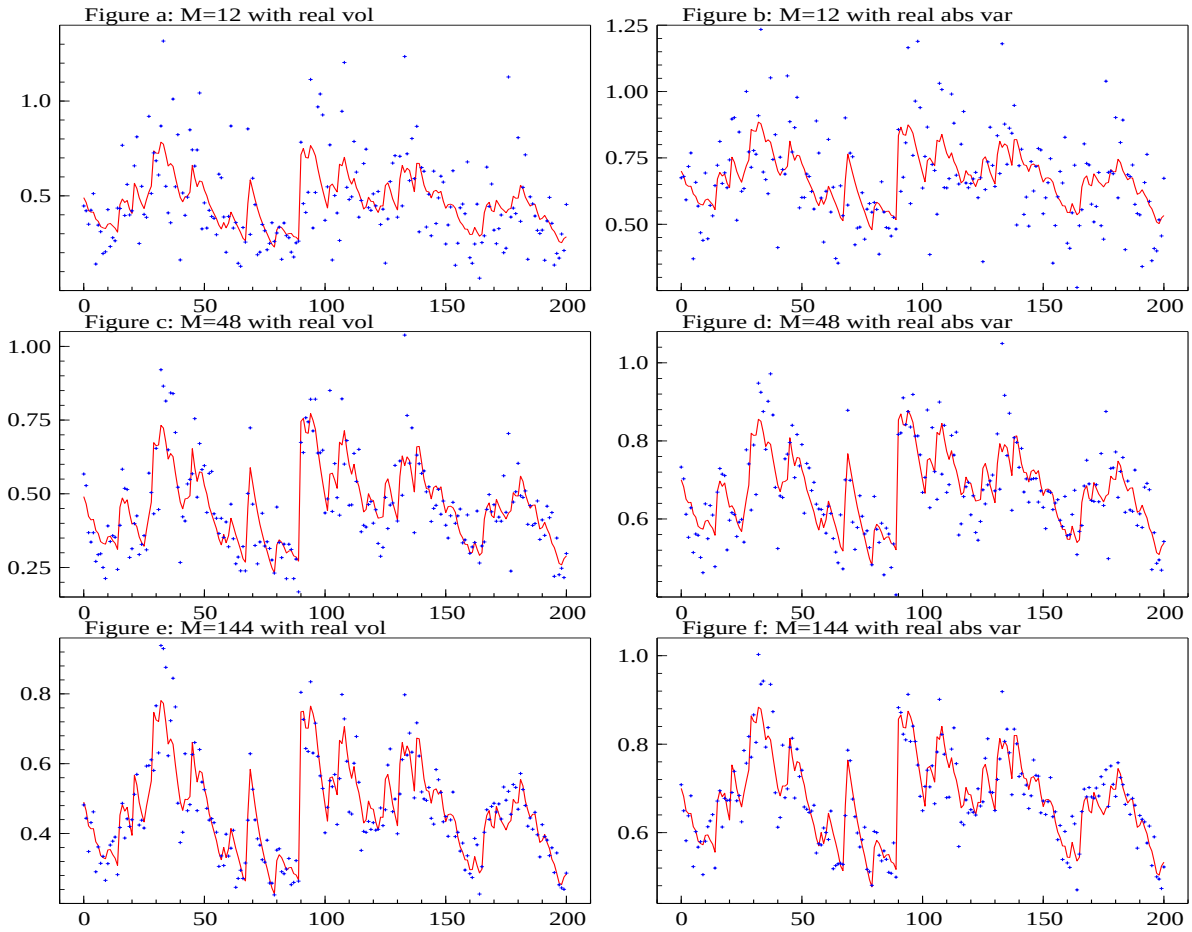

Figure 7.1: *Simulation from an OU-*$\Gamma(4,8)$ *process. Drawn is* $\sigma_n^{[2]}$ *and* $\sigma_n^{[1]}$ *against time, together with their associated realised power variation estimators. Graph is computed for* $M = 12$, 48 *and* 144.

the associated realised quadratic variation (depicted using crosses) computed using a variety of values of M. It is helpful to keep in mind that

$$\mathrm{E}\left(\sigma^2(t)\right) = \mathrm{E}\left(z(1)\right) = \frac{1}{2} \quad \text{and} \quad \mathrm{Var}\left(\sigma^2(t)\right) = \frac{1}{2}\mathrm{Var}\left(z(1)\right) = \frac{1}{32}.$$

The corresponding results for $\sigma_n^{[1]}$ and realised absolute variation is given in Figure 7.1(b), (d), (f). We see that as M increases the precision of realised power variation increases, while Figure 7.1 shows that the variance of the realised power variation increases with the level of volatility. This is line with the prediction from the asymptotic theory, for the denominator increases with the level of volatility.

QQ plots

To assess the finite sample performance of the asymptotic distributions of the realised quadratic and absolute variation we have constructed some QQ plots based upon the standardised errors (7.7) and (7.8). Our main focus will be on the absolute variation case, leaving the quadratic case to be covered in detail in a follow up paper by Barndorff-Nielsen and Shephard (2003). However, to start off with we give both cases, in order to allow an easy comparison.

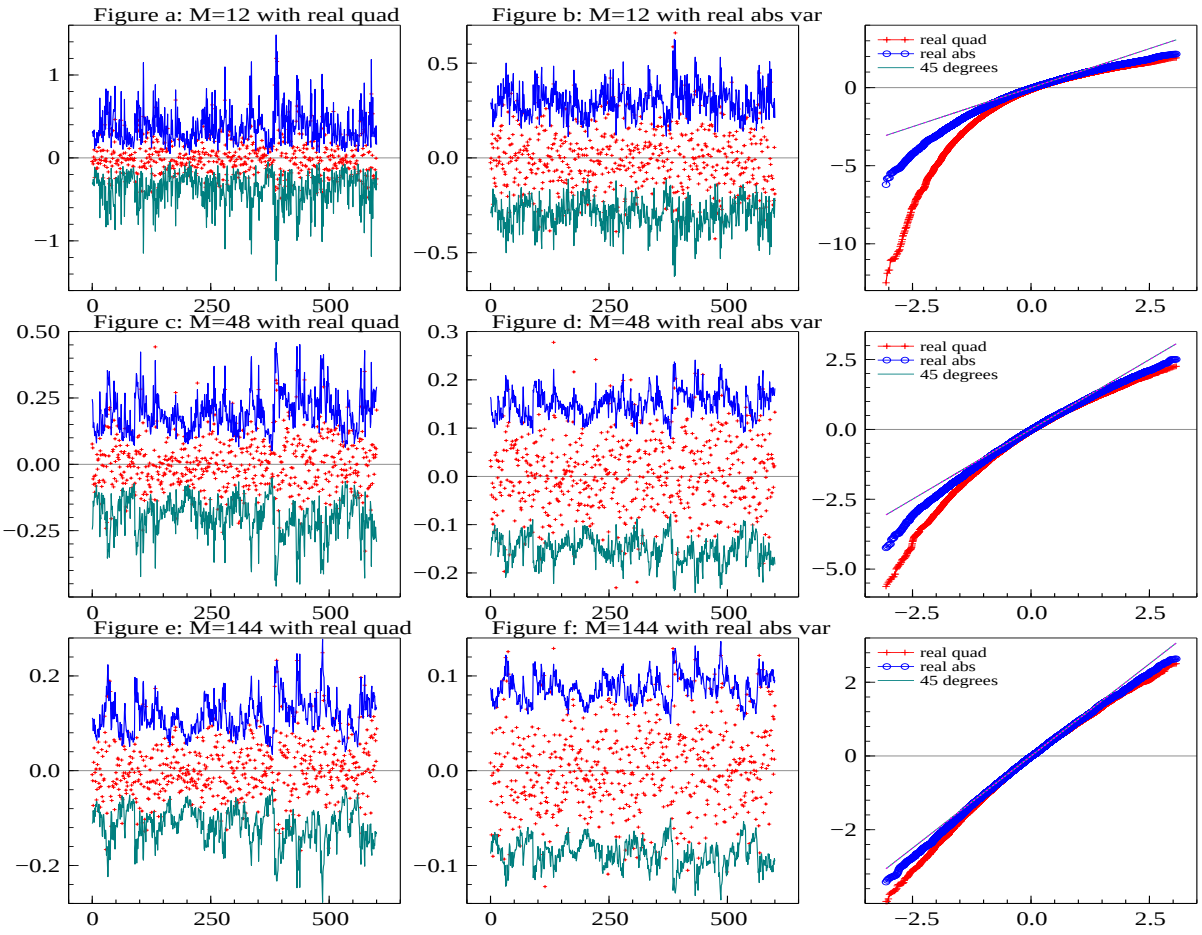

Figure 7.2: *Plots for the realised quadratic variation error and the realised absolute variation error plus twice their asymptotic standard errors. (a) and (b) have* $M = 12$*, while (c) and (d) has* $M = 48$ *and while (e) and (f) has* $M = 144$*. Corresponding QQ plots are on the right hand side, based on the standardised realised quadratic variations and the realised absolute variations.*

Figure 7.2 gives QQ plots based on the simulation experiment reported in the previous subsection, with a sample size of $10,000$. Again we vary M over 12, 48 and 144. The left hand side graphs give the realised quadratic variation errors $[y^*_\delta]^{[2]}_n - \sigma^{[2]}_n$ as well as plus and minus two times the asymptotic standard errors. The plot is based on the first 600 simulations. This graph shows how much the standard errors change through time. This continues to happen with large values of M and reflects the stochastic denominator in the limit theory.

The middle graphs of Figure 7.2 give the corresponding realised absolute variation errors

$$\sqrt{\frac{\pi\delta}{2}}\sum_{j=1}^{M}|y_{j,n}| - \sigma^{[1]}_n,$$

together with twice standard error bounds. The conditional standard errors are more stable through time, especially when M is small. The unconditional variance of the errors is approximately

$$\delta\left(\mu_1^{-2}-1\right)\Delta\mathrm{E}\left\{\sigma^2(t)\right\}.$$

The right hand side of Figure 7.2 gives the associated QQ plots for the standardised residuals from the realised quadratic and absolute variation measures. These use all $10,000$ observations.

The results are clear, in comparison with the asymptotic limit laws both random variables are too fat tailed in small samples, with this problem reducing as M increases. The realised absolute variation version of the statistic has much better finite sample behaviour, while the realised quadratic variation is quite poorly behaved.

Logarithmic transformation

The realised power variation $[y_\delta^*]^{[r]}(t)$ is the sum of non-negative items and so is non-negative. It would seem sensible to transform this variable to the real line in order to improve its finite sample performance. Hence we use the standard logarithmic transformation (that is for a consistent estimator $\widehat{\theta}$ of θ we approximate $\log(\widehat{\theta})$ by $\log(\theta)+\left(\widehat{\theta}-\theta\right)/\theta$, hence the asymptotic distribution of $\widehat{\theta}-\theta$ can be used to deduce the asymptotic distribution of $\log(\widehat{\theta})-\log(\theta)$). For the general realised power variation this implies

$$\frac{\log\left[\mu_r^{-1}\delta^{1-r/2}[y_\delta^*]^{[r]}(t)\right]-\log\int_0^t\sigma^r(s)\mathrm{d}s}{\mu_r^{-1}\delta^{1-r/2}\sqrt{\frac{v_r}{\mu_{2r}}\frac{[y_\delta^*]^{[2r]}(t)}{\left[\mu_r^{-1}\delta^{1-r/2}[y_\delta^*]^{[r]}(t)\right]^2}}}\xrightarrow{\mathcal{L}}N(0,1). \tag{7.27}$$

Referring to the denominator we should note that

$$\frac{\mu_{2r}^{-1}\delta^{1-r}[y_\delta^*]^{[2r]}(t)}{\left[\mu_r^{-1}\delta^{1-r/2}[y_\delta^*]^{[r]}(t)\right]^2}\xrightarrow{\mathcal{L}}\frac{\int_0^t\sigma^{2r}(s)\mathrm{d}s}{\left\{\int_0^t\sigma^r(s)\mathrm{d}s\right\}^2}\geq 1,$$

by Jensen's inequality. In the realised absolute variation case (7.27) simplifies to

$$\frac{\log\left\{\sqrt{\frac{\pi\delta}{2}}\sum_{j=1}^M|y_j(t)|\right\}-\log\int_0^t\sigma(s)\mathrm{d}s}{\sqrt{\delta\left(\frac{\pi}{2}-1\right)\frac{\sum_{j=1}^M y_j^2(t)}{\left\{\sqrt{\frac{\pi\delta}{2}}\sum_{j=1}^M|y_j(t)|\right\}^2}}}\xrightarrow{\mathcal{L}}N(0,1). \tag{7.28}$$

Using the same simulation setup as that employed in the previous subsection, we plot in Figure 7.3 the log version of the realised absolute variation error

$$\log\left\{\sqrt{\frac{\pi\delta}{2}}\sum_{j=1}^M|y_{j,n}|\right\}-\log\sigma_n^{[1]},$$

plus and minus twice the corresponding standard errors using (7.28). The standard errors have now stabilised, almost not moving with n. This is not surprising for t times

$$\frac{t^{-1}\int_0^t\sigma^2(s)\mathrm{d}s}{\left\{t^{-1}\int_0^t\sigma(s)\mathrm{d}s\right\}^2}$$

is empirically very close to one for small values of t while for large t it converges to

$$\frac{\mathrm{E}\left\{\sigma^2(t)\right\}}{\left[\mathrm{E}\left\{\sigma(t)\right\}\right]^2},$$

so long as the volatility process is ergodic and the moments exist.

The corresponding QQ plots in Figure 7.3 have also improved, with the normality approximation being accurate even for moderate values of M. This result carries over to wider simulations we have conducted.

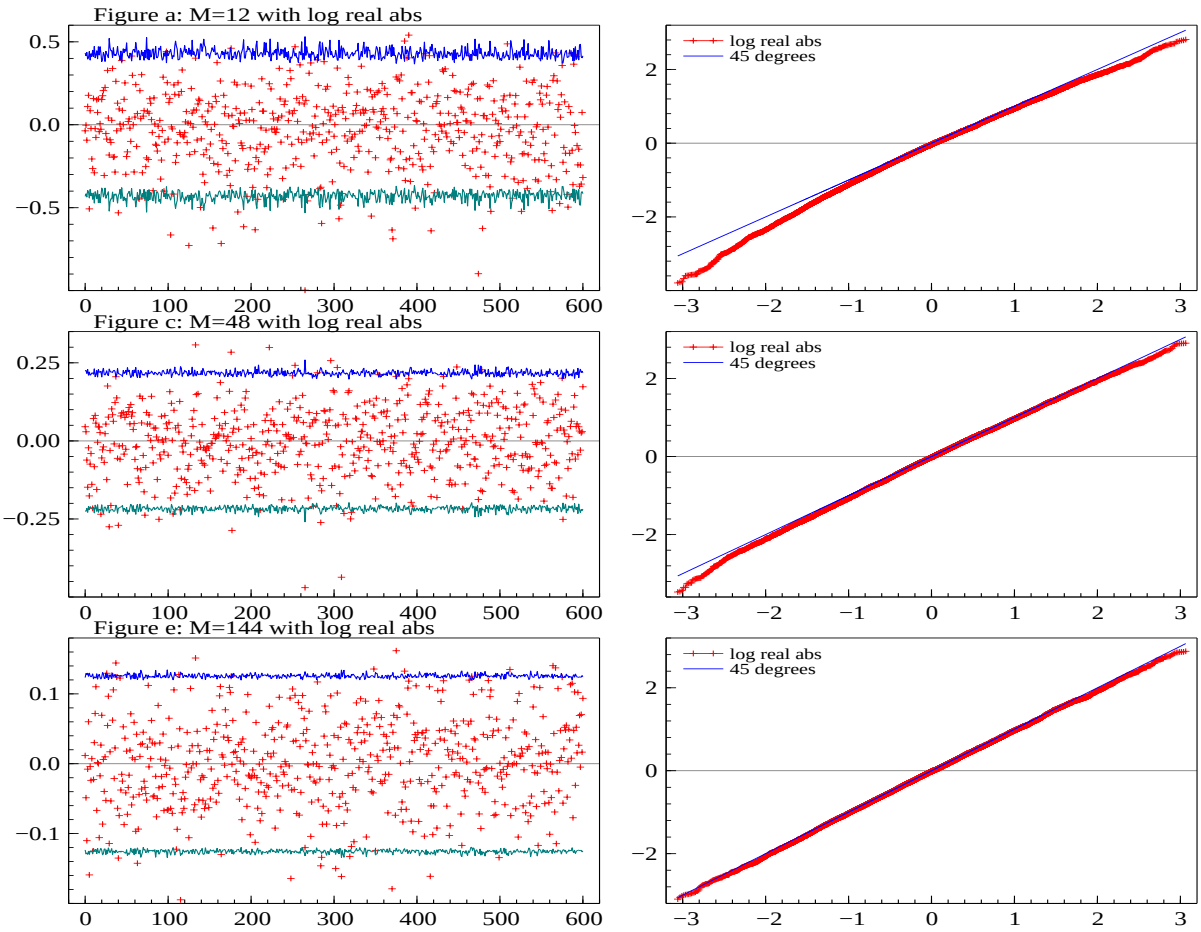

Figure 7.3: *Plots for the log transform of the realised absolute variation. Left hand plots are the errors plus and minus twice their asymptotic standard errors. Corresponding QQ plots are on the right hand side, based on the standardised log realised absolute variations.*

7.8 Conclusions

This paper has introduced the idea of realised power variation, which generalises the concept of realised volatility. The asymptotic analysis we provide, for $\delta \downarrow 0$, represents a significant extension of the usual quadratic variation result. Further, we provide a limiting distribution theory which considerably strengthens the consistency result and allows us to understand the variability of the difference between the realised power variation and the actual power volatility.

We have seen that when we take a log transformation of the realised power variation then the finite sample performance of the asymptotic approximation to the distribution of this estimator improves and seems to be accurate even for moderate values of M.

Our motivation for the study reported in this paper came originally from mathematical finance and financial econometrics where volatility is a key object of study. However, stochastic models in the form of a 'signal' α plus a noise term e where e is (conditionally) Gaussian with a variance that varies from 'site' to 'site' are ubiquitous in the natural and technical sciences, and we believe that results similar to those discussed here will be of interest for applications in a variety of other fields, for instance in turbulence and in spatial statistics.

Finally, the thesis of Becker (1998), briefly referred to in the Introduction, consists in a

comprehensive study of the limit behaviour as $M \to \infty$ of processes of the type

$$y_M(t) = \sum_{j=1}^{[Mt]} f\left[\frac{j-1}{M}, \frac{1}{\sqrt{M}}\left\{x\left(\frac{j}{M}\right) - x\left(\frac{j-1}{M}\right)\right\}\right]$$

where f is function of two variables and x denotes a Brownian semimartingale (of a certain kind, see below). The diffusion case is especially important. Extensions to general continuous or purely discontinuous semimartingales and even combination of the two are presented. The thesis is partly based on an earlier report by Jacod (1992), see also Delattre and Jacod (1997) and Florens-Zmirou (1993). Both x and f may be multidimensional, and generalisations to cases where not only the increment of x over the j-th interval but the whole trajectory over that interval occur in the second argument of f are also considered.

We shall not here attempt to indicate the precise results and the accompanying regularity conditions of Becker's thesis in any detail, but we wish to underline that the setting of his study is extremely general. Of immediate interest in connection with the present paper are his results when x is a Brownian semimartingale. More specifically, Becker considers the case where x is of the form

$$x(t) = \int_0^t c(s)\mathrm{d}s + \int_0^t \sigma(s)\mathrm{d}w(s)$$

where w is Brownian motion and c and σ are predictible and subject to restrictions on their variational behaviour. He shows, in particular, that $y_M(t)$ after a suitable centering converges to a stochastic process which is representable as a certain type of stochastic integral where the integration is with respect to a 'limit martingale-measure tangential to x'. A key point of our present work is that for the kind of functions f we consider, i.e. absolute powers, we are able to identify the limit behaviour as mixed Gaussian and, crucially for the statistical applicability, from this to establish a standard normal limit statement using random rescaling by observable scale factors.

7.9 Bibliographical information

7.10 Generalising results on realised power variation

For continuous sample path SV models, that is time deformed Brownian motion by a general chronometer, we can find the probability limit and asymptotic distribution of realised power variation. An important question is whether we can extend these ideas outside the continuous sample path case? One approach to dealing with this question is to add to the SV model an independent process which has jumps, another is to replace the Brownian motion assumption by a more general process which allows for the possibility of jumps.

In this section we analyse the latter problem, replacing Brownian motion with a stable process.

TO BE ADDED: *remarks about what happens if instead we replace Brownian motion by a NIG process.*

7.10.1 Stable innovations

Setup

Suppose that z is a symmetric α-stable process with $0 < \alpha < 2$. This means it has the cumulant function

$$\mathrm{C}\{\zeta \ddagger z(t)\} = \log \mathrm{E} e^{i\zeta z(t)} = -t|\zeta|^{\alpha}.$$

Except for the boundary case of $\alpha = 2$, this distribution has the empirically unappealing feature that the variance of $z(1)$ is infinity. The density of this variable is unknown in general, with exceptions being the zero mean, variance of 2 Gaussian variable ($\alpha = 2$), the Cauchy variable ($\alpha = 1$) and the Lévy variable ($\alpha = 1/2$). Importantly, all stable processes are semimartingales.

This process is representable by subordination as

$$z(t) \overset{law}{=} b(q(t)),$$

where q is the positive $\alpha/2$-*stable subordinator* with kumulant function

$$\bar{\mathrm{K}}\{\theta \ddagger q(t)\} = \log \mathrm{E} e^{-\theta q(t)} = -t(2\theta)^{\alpha/2}.$$

When $\alpha = 2$, $q(t)$ is simply a drift $2t$, otherwise $q(t)$ is a jump process. The $z(t)$ process inherits these properties, with only the $\alpha = 2$ case having a continuous sample path.

Assume that τ is locally bounded and that $\tau \perp\!\!\!\perp z$. Then a stable-SV model

$$y^*(t) = \int_0^t \sigma(s) \mathrm{d}z(s),$$

is well-defined for all $t \geq 0$. This process has discontinuous sample paths except in the case where $\alpha = 2$. This structure has the disadvantage that the unconditional variance of the price process does not exist, whatever the volatility process we use.

We define the increment to the price process over an interval of length $\delta > 0$ as

$$y_j = y^*(j\delta) - y^*((j-1)\,\delta),$$

then the following result holds:

Remark 7.1

$$y_j \overset{law}{=} \left(\int_{(j-1)\delta}^{j\delta} \sigma^{\alpha}(s) \mathrm{d}s \right)^{1/\alpha} z(1). \tag{7.29}$$

Proof. Let f be some deterministic function, then for any Lévy processes z we have that

$$\text{Cum}\left\{\zeta \ddagger \int_0^\infty f(s)\text{d}z(s)\right\} = \int_0^\infty \text{Cum}\{f(s)\zeta \ddagger z(1)\}\text{d}s,$$

by using the Lévy-Khintchine representation (e.g. Barndorff-Nielsen and Shephard (2001b, p. 287)). This gives that, conditionally on σ,

$$\text{Cum}\left\{\zeta \ddagger \int_0^t \sigma(s)\text{d}z(s)\right\} = \int_0^t \text{Cum}\{\sigma(s)\zeta \ddagger z(1)\}\text{d}s.$$

Then, since $\text{Cum}\{\zeta \ddagger z(1)\} = -|\zeta|^\alpha$ we obtain

$$\text{Cum}\{\zeta\sigma(s) \ddagger z(1)\} = -|\sigma(s)\zeta|^\alpha = -\sigma^\alpha(s)|\zeta|^\alpha,$$

i.e.

$$\begin{aligned}
\text{Cum}\left\{\zeta \ddagger \int_0^t \sigma(s)\text{d}z(s)\right\} &= -\int_0^t \sigma^\alpha(s)\text{d}s|\zeta|^\alpha \\
&= -\left|\left(\int_0^t \sigma^\alpha(s)\text{d}s\right)^{1/\alpha}\zeta\right|^\alpha \\
&= \text{Cum}\left\{\left(\int_0^t \sigma^\alpha(s)\text{d}s\right)^{1/\alpha}\zeta \ddagger z(1)\right\} \\
&= \text{Cum}\left\{\zeta \ddagger \left(\int_0^t \sigma^\alpha(s)\text{d}s\right)^{1/\alpha} z(1)\right\}.
\end{aligned}$$

■

For realised power variation we look at sums of powers of absolute returns. This means we focus on, from (7.29),

$$|y_j|^r \overset{law}{=} \left(\int_{(j-1)\delta}^{j\delta} \sigma^\alpha(s)\text{d}s\right)^{r/\alpha} |z(1)|^r.$$

We may reexpress this as

$$|y_j|^r \overset{law}{=} \left(\int_{(j-1)\delta}^{j\delta} \sigma^\alpha(s)\text{d}s\right)^{r/\alpha} |z_j|^r,$$

where $z_1, ..., z_M$ are i.i.d. with the same distribution as $z(1)$, or equivalently, using the subordination property, as

$$|y_j|^r \overset{law}{=} \left(\int_{(j-1)\delta}^{j\delta} \sigma^\alpha(s)\text{d}s\right)^{r/\alpha} q_j^{r/2}|u_j|^r,$$

where the $q_1, ..., q_M$ are i.i.d., with the same law as $q(1)$, and independent of $u_1, ..., u_M$ which are i.i.d. standard normal.

In view of these representations of $|y_j|^r$ it would be rather simple to give a complete description of the various possible limiting behaviours of $[y_\delta^*]^{[r]}$ as $\delta \to 0$. Here we shall only discuss some particular cases.

Quadratic variation

For $r = 2$ we have that realised variance is

$$[y_\delta^*](t) \overset{law}{=} \left\{\sum_{j=1}^M |u_j|^\alpha \int_{(j-1)\delta}^{j\delta} \sigma^\alpha(s)\text{d}s\right\}^{2/\alpha} q(1).$$

Under the usual type of regularity condition on τ the term in braces satisfies, conditionally on τ, as $\delta \downarrow 0$

$$\sum_{j=1}^{M} |u_j|^\alpha \int_{(j-1)\delta}^{j\delta} \sigma^\alpha(s)\mathrm{d}s \xrightarrow{p} \mu_\alpha \tau^{\alpha/2*}(t)$$

with $\mu_c = \mathrm{E}\{|u_j|^c\}$. Consequently the quadratic variation converges to

$$[y^*](t) \overset{law}{=} \{\mu_\alpha \tau^{\alpha/2*}(t)\}^{2/\alpha} q(1),$$

where

$$\tau^{\alpha/2*}(t) = \int_0^t \sigma^\alpha(s)\mathrm{d}s,$$

is the integrated power volatility.

Remark 7.2 *When $\alpha = 2$, then $q(t) = 2t$ and so $\mu_2 = \mathrm{E}\{|u_j|^2\} = 1$*

$$[y^*](t) \overset{law}{=} 2\tau^*(t), \qquad \textit{where} \qquad \tau^*(t) = \int_0^t \sigma^2(s)\mathrm{d}s$$

The 2 appears as the stable variable has a variance of 2 over a unit interval of time when $\alpha = 2$.

Power variation

The probability limit of realised variance, that is quadratic variation, is quite complicated in the stable-SV case. It turns out that much simpler and more powerful results are available if we use realised power variation instead of realised variance.

The moments of $|z(1)|^r$ exist up to, but not including, order $(1+\alpha-r)/r$. Hence, still given τ, if $r < (1+\alpha)/2$ then

$$\delta^{1-r/\alpha}[y_\delta^*]^{[r]}(t) \xrightarrow{p} \mu_{\alpha r}\tau^{r/2*}(t), \tag{7.30}$$

where $\mu_{\alpha,r} = \mathrm{E}\{|z(1)|^r\}$. This provides a simple generalisation of the use of quadratic variation for Brownian motion based SV models for then $r = 2$ and

$$[y_\delta^*]^{[2]}(t) \xrightarrow{p} \tau^*(t),$$

exactly.

As $\alpha \leq 2$ it implies we must use $r \leq 2$ for the limit (7.30) to hold. In particular as α falls, so must r in order for us to be able to analyse the problem. Overall, this is a powerful result for it means that by differentiating realised power variation we reveal the spot volatility even in the stable-SV case. However, the observation has an important limitation for it implies that we need to know the value of α to use it, for the normalisations involve $\delta^{1-r/\alpha}$ and $\mu_{\alpha,r}$. We will see why this result remains important in a moment, even though the above is a limitation.

And in case $r < \min\{\alpha, (1+\alpha)/3\}$ we have the stronger result that

$$\frac{\delta^{1-r/\alpha}\mu_{\alpha,r}^{-1}[y_{0\delta}^*]^{[r]}(t) - \tau^{r/2*}(t)}{\delta^{1/2}\sqrt{\mu_{\alpha,r}^{-2}v_{\alpha,r}\tau^{r*}(t)}} \overset{law}{\to} N(0,1) \tag{7.31}$$

both conditionally and unconditionally; here $v_{\alpha,r} = \mathrm{Var}\{|z(1)|^r\}$. Of course in practice the above limit theory is infeasible as we do not know the value of $\tau^{r*}(t)$, however so long as $r < (1+\alpha)/4$ we can use formula (7.30) to replace $\tau^{r*}(t)$ by a feasible quantity.

Asymptotic negligibility

A limitation with the above limit theorems is that we need to know the value of α. However, in practice that would be typically unavailable to use. In this part of this paper we show that in some sense the realised absolute variation is, for $r < 2$, in any case small for stable-SV models, compared to the same quantity for Browian-SV models.

In the Brownian-SV case we have proved elsewhere that

$$\frac{\delta^{1-r/2}\mu_{2,r}^{-1}[y_{0\delta}^*]^{[r]}(t) - \tau^{r/2*}(t)}{\delta^{1/2}\sqrt{\mu_{2,r}^{-2}v_{2,r}\tau^{r*}(t)}} \xrightarrow{law} N(0,1).$$

Thus, in particular that

$$\delta^{1-r/2}\mu_{2,r}^{-1}[y_{0\delta}^*]^{[r]}(t) \xrightarrow{p} \tau^{r/2*}(t).$$

In the stable-SV case $r = 1$

$$[y_\delta^*]^{[1]}(t) \overset{law}{=} \sum_{j=1}^{M}\left(\int_{(j-1)\delta}^{j\delta}\sigma^\alpha(s)\mathrm{d}s\right)^{1/\alpha}|z_j|$$

For $1 < \alpha < 2$ we have, by (7.30),

$$\delta^{1-1/\alpha}[y_\delta^*]^{[1]}(t) \xrightarrow{p} \mu_{\alpha,1}\tau^{1/2*}(t)$$

This implies

$$\delta^{1/2}[y_\delta^*]^{[1]}(t) \xrightarrow{p} 0$$

in contrast to the case where z is Brownian motion where the probability limit is non-negative. This is a feasible result, for it implies we do not need to know α to use it. More generally,

$$\delta^{1-r/2}[y_{0\delta}^*]^{[r]}(t) \xrightarrow{p} 0 \qquad \text{if} \qquad r < (1+\alpha)/2.$$

Other observations

Suppose now, for simplicity, that τ is constant and equal to 1. Then $|y_j| \overset{law}{=} \delta^{1/\alpha}|z(1)|$ and

$$[y_\delta^*]^{[1]}(t) \overset{law}{=} \delta^{1/\alpha}\sum_{j=1}^{M}|z_j|$$

The random variables $|z_j|$ belong to the 'domain of normal attraction' of a stable law with index α. Hence, on account of **?**, pp. 580–581), we may conclude:

- If $1 < \alpha < 2$ then, for a certain α-stable law S_α,

$$[y_\delta^*]^{[1]}(t) - \delta^{-1+1/\alpha}\mu_{\alpha,1} \xrightarrow{law} S_\alpha.$$

- If $0 < \alpha < 1$ then, for a certain positive α-stable law $S_{+\alpha}$,

$$[y_\delta^*]^{[1]}(t) \xrightarrow{law} S_{+\alpha}.$$

- If $\alpha = 1$ then, for a certain 1-stable law S_1,

$$[y_\delta^*]^{[1]}(t) - b_\delta \xrightarrow{law} S_1$$

  where

$$\int_{-\infty}^{\infty}\sin(\delta x)\mathrm{d}P\{|z(1)| \le x\}.$$

In all three cases, $\delta^{1/2}[y_\delta^*]^{[1]}(t) \xrightarrow{p} 0$.

Chapter 8

Conclusions

Appendix A

Primer on stochastic analysis

Abstract: This book assumes some knowledge of the properties of semimartingales. This may be unfamiliar to some readers, particularly those with backgrounds in econometrics and statistics. Thus we have prepared this Appendix which provides some of the basic information about stochastic analysis. No proofs are given, instead we focus on concepts and definitions.

A.1 Introduction

This book will assume a basic knowledge of semimartingales. This material, which is familiar in probability theory and finance, is less well known in econometrics and statistics. In order to bridge this gap we have provided this section as a primer to this material. We state some definitions and concepts, as well as some results. No proofs are given. For further reading on this topic we highly recommend the excellent Protter (1990) for a self-contained treatment. We have deliberately used the notation from that book here in order to allow it to be more accessible. (The books by Karatzas and Shreve (1991) and Øksendal (1995) are also highly recommendable and provide additional material). For a very comprehensive treatment, at a high level of abstraction, see Jacod and Shiryaev (1987).

We first recall some basic concepts.

It is sometimes essential to specify the *filter* in the background of the stochastic processes considered. A *filtered probability space* is a quadruplet $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, P)$ where $(\Omega, \mathcal{F}, P)$ constitutes an ordinary probability space (i.e. Ω is the sample space, $\mathcal{F}$ is the σ-algebra of events, and P is the probability measure) and the $\mathcal{F}_t$, $0 \leq t < \infty$, form an increasing sequence of sub-σ-algebras of $\mathcal{F}$ such that for any stochastic process X under consideration events determined by $X^t = \{X_s : 0 \leq s \leq t\}$ belong to $\mathcal{F}_t$ (the process X is then said to be *adapted* to the filtration).

A stochastic process X is said to be *càdlàg* if, with probability 1, its sample paths are continuous from the right with limits from the left, i.e. except for a subset of Ω of probability 0 we have for all $t \geq 0$ that $\lim_{s \downarrow t} X_s(\omega) = X_t(\omega)$ and that $\lim_{s \uparrow t} X_s(\omega)$ exists. Analogously, a *càglàd* process is defined as having limits from the right and being continuous from the left.

Predictability (some authors use instead the term *previsibility*) of a process means essentially (see Protter (1990, p. 89) for a precise definition) that it is càglàd. Any càglàd process is predictable, and any deterministic process is predictable (whether càglàd or not).

A stochastic process X on a filtered probability space $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, P)$ is a *martingale* if $\mathrm{E}\{X_t\} < \infty$ and

$$\mathrm{E}\{X_t | \mathcal{F}_s\} = X_s \quad \text{almost surely,} \tag{A.1}$$

for all $s \leq t$. (Note that (A.1) implies that $\mathrm{E}\{X_t\}$ is constant independently of t. In particular, one speaks of mean-0 and mean-1 martingales.) We write $M \in \mathcal{M}_{loc}$ to indicate that a process M is a local martingale, and $M \in \mathcal{M}_{loc}^c$ if it is also continuous[1]. Processes that belong to $\mathcal{M}_{loc}$ without, in fact, being martingales are rather exceptional and we shall not indicate the somewhat technical definition of a local martingale here (see Protter (1990, p. 33)).

Example 22 *The following has been proposed as model for the log price process of a typical stock (Bibby and Sørensen (1997)). The process is defined as the stationary solution to the SDE*

$$\mathrm{d}x_t = v(x_t)\sigma \mathrm{d}w_t$$

where w is Brownian motion, $\sigma > 0$, and v is specified by

$$v^2(x) = \exp\{\alpha\sqrt{\delta^2 + (x-\mu)^2} - \beta(x-\mu)\}$$

[1] When we speak of continuous processes we mean processes whose sample paths are continuous (almost surely).

and $0 \leq \beta < \alpha$. *With this choice of* v *the one-dimensional marginals of the process follow the hyperbolic distribution whose density is proportional to*

$$\exp\{-\alpha\sqrt{\delta^2 + (x-\mu)^2} + \beta(x-\mu)\}.$$

The stationary solution to the SDE is a uniformly integrable local martingale which is not a martingale (for details, see Bibby and Sørensen (1997)). Speaking qualitatively, the reason x_t *is not a martingale is that the volatility factor* $v(x_t)$ *is very large when the process is far away from the origin so that the process has too limited a tendency to return to* 0.

A.2 Bounded variation

A real function g on a finite interval $[a, b]$ is said to be of *bounded variation* if there exists an $M > 0$ such that for any partition $(t_0, t_1, ..., t_n)$ of $[a, b]$ with

$$a = t_0 < t_1 < \cdots < t_n = b$$

we have

$$\sum_{j=1}^{n} |g(t_j) - g(t_{j-1})| < M.$$

Any function g of bounded variation can be written as a difference $g = h - k$ of two nondecreasing bounded functions h and k. Recall that Stieltjes integration deals with integrals of the type

$$\int f(s)\mathrm{d}g(s)$$

for f and g of bounded variation. Standard references for the basic theory of Stieltjes integrals are Apostol (1957) and Widder (1946).

A.3 Semimartingales and stochastic integrals

Let $X = \{X_t\}_{t\geq 0}$ be a stochastic process. Then X_t is said to be càdlàg if it has sample paths that, at each point of time t, are continuous from the right with limits from the left. For any such process we let $X_{t-} = \lim_{s\uparrow t} X_s$ and then the process X_{t-} is càglàd, i.e. having sample paths that are continuous from the left with limits from the right. A càdlàg process is a *semimartingale* if it is decomposable as

$$X_t = X_0 + A_t + M_t, \tag{A.2}$$

where A_t, with $A_0 = 0$, is a stochastic process whose paths are of *locally bounded variation* (i.e. of bounded variation on any finite subinterval of $[0, \infty)$) and M_t, with $M_0 = 0$, is a *local martingale*. In general, the decomposition is not unique. However, if there is a decomposition for which A_t is *predictable* then that decomposition is unique and the process is called a *special semimartingale*. We denote the classes of martingales, semimartingales and special semimartingales by $\mathcal{M}$, $\mathcal{SM}$, and $\mathcal{SSM}$, respectively, and we write $\mathcal{M}^c$, $\mathcal{SM}^c$, and $\mathcal{SSM}^c$ for the respective subclasses of path continuous processes. Similiarly, $\mathcal{M}_{loc}$ stands for the class of local martingales, etc. The class of locally bounded variation semimartingales is denoted $\mathcal{BV}$.

Remark A.1 *If* $X \in \mathcal{M}^c$ *then* $X \notin \mathcal{BV}$ *unless* $X = 0$.

Both M_t and A_t can be decomposed into continuous parts, written M_t^c and A_t^c, and discontinuous parts, M_t^d and A_t^d, respectively. For all semimartingales the decomposition into M_t^c and A_t^c

is unique, while this is not so for M_t^d and A_t^d. When we additionally assume the semimartingale is special, then the decomposition of the discontinuous part also becomes unique.

In the case where $X_0 = 0$ so that $X = A + M$ with $A \in BV$ and $M \in \mathcal{M}_{loc}$ we write $X^c = A^c + M^c$ and $X^d = X - X^c$ for the continuous and discontinuous paths of X respectively.

Example 23 *Let $X = B + C$ where B is Brownian motion, C is a compound Poisson process with $\mathrm{E}\{C\} = 0$ and $B \perp\!\!\!\perp C$. Then two of possible decompositions $X = A + M$ (cf. (A.2)) are $A \equiv 0$, $M = B + C$ and $A = C$, $M = B$. In this case $X^c = B$.*

Predictable stochastic processes can, in wide generality, be integrated with respect to an arbitrary semimartingale. In particular if H is predictable and locally square integrable and X is a semimartingale then the *stochastic integral*

$$Y = H \bullet X$$

is the stochastic process

$$Y_t = \int_0^t H_{s-} \mathrm{d}X_s,$$

where the integral can often be defined pathwise (Mikosch and Norvaisa (2000)), and more generally as the limit in probability of finite sums of the form

$$\sum_{i=0}^{n-1} H_{s_i}(X_{s_{i+1}} - X_{s_i}),$$

where $0 = s_0 < s_1 < ... < s_n = t$. The latter is the case, in particular, if H is càglàd. In full generality, a slightly weaker form of convergence holds. Any such integral process Y is itself a semimartingale (for details, see Protter (1990, Sections II.4 and IV.2)).

Given a semimartingale X, the class of processes H for which the integral $Y = H \bullet X$ can be naturally defined depends to some extent on the properties of X. In particular, if X equals Brownian motion B and H is a process such that

$$\int_0^t H_s^2 \mathrm{d}s < \infty$$

then the integral exists and Y is a continuous local martingale. A detailed account of integration of deterministic functions with respect to a Lévy process is given in Rocha-Arteaga and Sato (2001, Section 2.1), see further in Section 11.

In case Y_t is of the form

$$Y_t = \int_0^t H_{s-} \mathrm{d}M_s$$

with $M \in \mathcal{M}_{loc}$ and H predictable then Y is itself a local martingale. If, moreover, $\mathrm{E}\{[Y]_t^{1/2}\} < \infty$ for all $t > 0$ then Y is, in fact, a martingale (here $[Y]$ denotes the quadratic variation of $[Y]$, see section 12 below). On the other hand, if

$$Y_t = \int_0^t H_{s-} \mathrm{d}A_s$$

with A being of bounded variation then Y is also a bounded variation process.

A.4 Quadratic variation

Let X_t be a semimartingale with $X_0 = 0$. The *quadratic variation* (process) $[X]$ of X is defined by

$$[X] = X^2 - 2X_- \bullet X, \tag{A.3}$$

or equivalently by

$$[X]_t = p\text{-}\lim \sum_{i=0}^{n-1} (X_{s_{i+1}} - X_{s_i})^2, \tag{A.4}$$

where again $0 = s_0 < s_1 < ... < s_n = t$ and the limit is for the mesh size

$$\max_{1 \le j \le n} |s_j - s_{j-1}| \to 0$$

as $n \to \infty$. It is this latter construction that we have emphasised in our book.

Example 24 *Suppose X_t is of locally bounded variation then*

$$[X]_t = \sum_{0 \le s \le t} (\Delta X_s)^2 .$$

So all continuous processes with bounded variation have $[X]_t = 0$.

If X and Y are semimartingales then one defines the quadratic covariation $[X, Y]$ of X and Y as

$$[X, Y] = \frac{1}{2}[X + Y] - [X] - [Y]. \tag{A.5}$$

(Note $[X, X] = [X]$.) The covariation can also be calculated directly by

$$[X, Y]_t = p\text{-}\lim \sum_{i=0}^{n-1} (X_{s_{i+1}} - X_{s_i})(Y_{s_{i+1}} - Y_{s_i}), \tag{A.6}$$

provided $X_0 = Y_0 = 0$.

The process $[X, Y]$ is always of local bounded variation (Protter (1990, Corollary 1, Section II.6)), and

$$[X, Y]_t = [X^c, Y^c]_t + \sum_{0 \le s \le t} \Delta X_s \Delta Y_s \tag{A.7}$$

where X^c and Y^c denote the continuous local martingale components of X and Y, respectively.

For arbitrary càglàd processes H and K we have the important formula

$$[H \bullet X, K \bullet Y]_t = \int_0^t H_s K_s \mathrm{d}[X, Y]_s. \tag{A.8}$$

Suppose Y_t is an m-dimensional special semimartingale, that is an m-dimensional semimartingale each of whose components $Y_{1t}, ..., Y_{mt}$ belongs to $\mathcal{SSM}$, and let $Y_t = A_t + M_t$ be the (component by component) unique decomposition into a predictable semimartingale A_t and a local martingale M_t. The matrix QV, which contains the quadratic variations on the diagonal and the corresponding covariations off the diagonal, becomes

$$[Y]_t = [M]_t + \sum_{0 \le s \le t} \Delta A_s \Delta A_s{}' + \sum_{0 \le s \le t} \Delta M_s \Delta A_s{}' + \sum_{0 \le s \le t} \Delta A_s \Delta M_s{}', \tag{A.9}$$

the QV of M plus terms which are influenced by the jumps in A and M. If A is continuous then we obtain the simplification

$$[Y]_t = [M]_t, \tag{A.10}$$

which holds irrespective of the presence of jumps in the local martingale component.

A.5 Ito's formula

Suppose X_t is a semimartingale and let f be a twice continuously differentiable function. Then *Ito's formula* for semimartingales states that

$$\begin{aligned} f(X_t) &= f(X_0) + \int_0^t f'(X_{s-})\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_{s-})\mathrm{d}[X]_s^c \\ &\quad + \sum_{0<s\leq t}\{f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\}, \end{aligned} \tag{A.11}$$

where $[X]^c$ denotes the path by path continuous part of the quadratic variation process $[X]$. (We have that $[X]^c = [X^c]$, as follows from the definition by (A.2) above). Note that $f(X_t)$ is again a semimartingale. In the special case where X_t is continuous then

$$f(X_t) = f(X_0) + \int_0^t f'(X_s)\mathrm{d}X_s + \frac{1}{2}\int_0^t f''(X_s)\mathrm{d}[X]_s.$$

Example 25 *For a general semimartingale X_t, suppose $Y_t = \exp(X_t)$, then Ito's formula implies*

$$Y_t = Y_0 + \int_0^t Y_{s-}\mathrm{d}X_s + \frac{1}{2}\int_0^t Y_{s-}\mathrm{d}[X]_s^c + \sum_{0<s\leq t}\left(\Delta Y_s - Y_{s-}\Delta X_s\right).$$

When X_t is continuous this simplifies to

$$Y_t = Y_0 + \int_0^t Y_s\mathrm{d}X_s + \frac{1}{2}\int_0^t Y_s\mathrm{d}[X]_s.$$

This result extends to cover the case where $Y_t = \exp(i\zeta X_t)$. OLE: ADD MATERIAL.

Example 26 *For a general semimartingale X_t, suppose $Y_t = X_t^2$, then Ito's formula implies*

$$Y_t = Y_0 + 2\int_0^t X_{s-}\mathrm{d}X_s + [X]_t^c + \sum_{0<s\leq t}\left(\Delta Y_s - Y_{s-}\Delta X_s\right).$$

When X_t is continuous this simplifies to

$$Y_t = Y_0 + 2\int_0^t X_s\mathrm{d}X_s + [X]_t.$$

When $Y_t = X_t^4$, the corresponding results are

$$Y_t = Y_0 + 4\int_0^t X_{s-}^3\mathrm{d}X_s + 6\int_0^t X_{s-}^2\mathrm{d}[X]_s^c + \sum_{0<s\leq t}\left(\Delta Y_s - Y_{s-}\Delta X_s\right),$$

and in the continuous case

$$Y_t = Y_0 + 4\int_0^t X_s^3\mathrm{d}X_s + 6\int_0^t X_s^2\mathrm{d}[X]_s^c.$$

These results will be used in, for example, our analysis of realised variances.

There is an important extension of Ito's formula for the case where f is only assumed to be a convex function while on the other hand the semimartingale X_t belongs to $\mathcal{SM}^c$. It says, in particular, that convex functions of continuous semimartingales are themselves continuous semimartingales.

Theorem A.2 *Let $Y \in \mathcal{SM}^c$ and let f be a convex function on $\mathbf{R}$. Then $f(Y_t) \in \mathcal{SM}^c$ and*

$$f(Y_t) = f(Y_0) + \int_0^t D^- f(Y_s)\mathrm{d}Y_s + \int_{\mathbf{R}} \Lambda_t(a)\mu(\mathrm{d}a)$$

where

$$\mu([a,b)) = D^- f(b) - D^- f(a)$$

and $\Lambda_t(a)$ is the local time of Y at level a.

For a detailed discussion of this result see Karatzas and Shreve (1991, p.218-219).

Example 27 *Suppose $f(x) = |x|^r$ which is convex for $r \geq 1$. If $r = 1$ then the measure μ of the above Theorem is $2 \times \delta_0$ (δ_0 delta measure at 0) while for $r > 1$ we have that μ is absolutely continuous with density $2\binom{r}{2}|x|^{r-2}$.*

Remark A.3 *Let M be a continuous local martingale. Then $|M|^r$ is a semimartingale if and only if $r \geq 1$. (Cf. Revuz and Yor (1999, p.231)).*

A.6 Stochastic differential equations

One is interested in defining stochastic processes Y such that these satisfy *stochastic differential equations* (*SDE*s) of the form

$$\mathrm{d}Y_t = F(t; Y_{t-})\mathrm{d}X_t, \tag{A.12}$$

with an initial condition specifying the value of Y at $t = 0$ and where X is a given semimartingale. This question is not well posed until it is specified what is meant by a solution to the SDE. One says that Y solves (A.12) with initial condition Y_0 if

$$Y_t = Y_0 + \int_0^t F(s, Y_{s-})\mathrm{d}X_s. \tag{A.13}$$

Under relatively mild conditions on the function F the solution exists uniquely and is itself a semimartingale, see Protter (1990, Section V.3).

We can write Ito's formula in terms of an SDE. The expression is

$$\mathrm{d}f(X_t) = f'(X_{t-})\mathrm{d}X_t + \frac{1}{2}f''(X_{t-})\mathrm{d}[X^c]_t + \left\{f(X_t) - f(X_{t-}) - f'(X_{t-})\Delta X_t\right\}. \tag{A.14}$$

When the process is continuous this simplifies to

$$\mathrm{d}f(X_t) = f'(X_t)\mathrm{d}X_t + \frac{1}{2}f''(X_t)\mathrm{d}[X]_t. \tag{A.15}$$

Example 28 *Suppose $Y_t = \exp(X_t)$, then*

$$\mathrm{d}Y_t = Y_t\mathrm{d}X_t + \frac{1}{2}Y_t\mathrm{d}[X]_t.$$

Example 29 *If X_t follows a general diffusion*

$$\mathrm{d}X_t = \mu(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}w_t,$$

then

$$\mathrm{d}[X]_s = \sigma^2(X_s)\mathrm{d}t,$$

which implies

$$\mathrm{d}f(X_t) = \left\{f'(X_t)\mu(X_t) + \frac{1}{2}f''(X_t)\sigma^2(X_t)\right\}\mathrm{d}t + \sigma(X_t)\mathrm{d}w_t,$$

from (A.14).

A.7 Stochastic exponentials

For a continuous semimartingale X the *stochastic (or Doléans-Dade) exponential* is given by

$$\mathcal{E}(X) = e^{X-\frac{1}{2}[X]} \tag{A.16}$$

and we have that

$$X \in \mathcal{M}^c_{loc} \Longleftrightarrow \mathcal{E}(X) \in \mathcal{M}^c_{loc}. \tag{A.17}$$

Moreover,

$$e^X = \mathcal{E}(\hat{X}) \tag{A.18}$$

where

$$\hat{X} = X + \frac{1}{2}[X], \tag{A.19}$$

(as follows as the quadratic variation of quadratic variation is zero).

It follows that

$$e^X \in \mathcal{M}^c_{loc} \Longleftrightarrow \mathcal{E}(\hat{X}) \in \mathcal{M}^c_{loc} \Longleftrightarrow H + \frac{1}{2}[H] \in \mathcal{M}^c_{loc}. \tag{A.20}$$

More generally, suppose that X is a semimartingale. Then the stochastic exponential $\mathcal{E}(X)$ of X is defined as

$$\mathcal{E}(X)_t = e^{X_t-\frac{1}{2}[X]_t} \prod_{0<s\leq t} (1+\Delta X_s) e^{-\Delta X_s + \frac{1}{2}(\Delta X_s)^2} \tag{A.21}$$

and $Y = \mathcal{E}(X)$ is the unique solution to the SDE

$$\mathrm{d}Y_t = Y_{t-}\mathrm{d}X_t$$

with initial condition $Y_0 = 1$. The process $\mathcal{E}(X)$ is a semimartingale.

Formula (A.21) may equivalently be expressed as

$$\mathcal{E}(X)_t = e^{X^c_t - \frac{1}{2}[X^c]_t} \cdot e^{X^d_t} \prod_{0<s\leq t} (1+\Delta X_s) e^{-\Delta X_s} \tag{A.22}$$

where X^c denotes the (unique) continuous local martingale part of X and $X^d = X - X^c$. Although it may seem puzzling at first sight, the formulae (A.21) and (A.22) are indeed even if $1 + \Delta x_s$ is 0 or negative for some values of s.

For these results, see Protter (1990, Section II.8).

A.8 The likelihood ratio process

To introduce ideas, suppose we have a discrete time stochastic process y with natural filtration $\mathcal{F}_{t-}$, which (in discrete time) holds all the observations up to time $t-1$. We compare two different probability measures P and Q which live on the filtered space $(\Omega, \{\mathcal{F}_t\}, \mathcal{F})$. We write f and g for densities under P and Q, respectively, (assuming these densities exist). Then we can compare their fit via the likelihood ratio

$$L_t = \frac{g(y_1, ..., y_t)}{f(y_1, ..., y_t)}.$$

Of course, using the prediction decomposition,

$$L_t = \frac{g(y_t|\mathcal{F}_{t-})}{f(y_t|\mathcal{F}_{t-})} L_{t-1} = \frac{g(y_t, y_{t-1}|\mathcal{F}_{t-1-})}{f(y_t, y_{t-1}|\mathcal{F}_{t-1-})} L_{t-2}.$$

If we calculate this for each (discrete) value of t, taking $L_0 = 1$, then L_t is called the *likelihood ratio process* and under P, L_t is a mean-1 martingale with respect to $\mathcal{F}_{t-}$. Thus

$$\mathrm{E}_P\left(L_t|\mathcal{F}_{s+1-}\right) = L_s.$$

The same arguments can be used in continuous time, although now we have to explicitly use measures instead of densities. Let P_t and Q_t denote the restrictions of P and Q to the σ-algebra $\mathcal{F}_t$ and assume P and Q are locally equivalent, i.e.

$$P(A) = 0 \quad \Leftrightarrow \quad Q(A) = 0$$

whenever $A \in \mathcal{F}_t$ for some t. Then we use the following notation

$$L_t = \frac{\mathrm{d}Q_t}{\mathrm{d}P_t}$$

The quantity $\frac{\mathrm{d}Q_t}{\mathrm{d}P_t}$ is called the *Radon-Nikodym derivative* of Q_t with respect to P_t, and L_t is the likelihood ratio function based on observation of the process up to and including time t. The Radon-Nikodym derivative decomposes, for $0 < s < t$, as[2]

$$\frac{\mathrm{d}Q_t}{\mathrm{d}P_t} = \frac{\mathrm{d}Q_s}{\mathrm{d}P_s}\frac{\mathrm{d}Q_t^{\mathcal{F}_s}}{\mathrm{d}P_t^{\mathcal{F}_s}}$$

into the product of the restriction to $\mathcal{F}_s$ times the conditional law given the restriction $\mathcal{F}_s$. This implies the likelihood ratio process satisfies

$$L_t = L_s\frac{\mathrm{d}Q_t^{\mathcal{F}s}}{\mathrm{d}P_t^{\mathcal{F}s}}.$$

Consequently, by the calculus of conditional expectations

$$\mathrm{E}_P\{L_t|\mathcal{F}_s\} = L_s\mathrm{E}_P\left\{\frac{\mathrm{d}Q_t^{\mathcal{F}s}}{\mathrm{d}P_t^{\mathcal{F}s}}|\mathcal{F}_s\right\} = L_s$$

(almost surely) showing that the likelihood process L_t is a mean-1 martingale under the probability measure P. See, for example, Barndorff-Nielsen and Sørensen (1995) for a more formal view of this material.

A.9 Girsanov-Meyer Theorem

Let P and Q be *locally equivalent* probability measures on the filtered measure space $(\Omega, \{\mathcal{F}_t\}, \mathcal{F})$ and, as in the previous section, let $L_t = \frac{\mathrm{d}Q_t}{\mathrm{d}P_t}$.

[2]We are using here the general fact that if P and Q are two equivalent probability measures on a measure space $(\Omega, \mathcal{F})$ then for any sub-σ-algebra $\mathcal{B}$ of $\mathcal{F}$ the Radon-Nikodym derivative of Q with respect to P factorises as

$$\frac{\mathrm{d}Q}{\mathrm{d}P} = \frac{\mathrm{d}Q_{\mathcal{B}}}{\mathrm{d}P_{\mathcal{B}}}\frac{\mathrm{d}Q^{\mathcal{B}}}{\mathrm{d}P^{\mathcal{B}}}$$

into the Radon-Nikodym derivative of the marginal laws corresponding to $\mathcal{B}$ times the Radon-Nikodym derivative of the conditional laws given $\mathcal{B}$.

Theorem A.4 (Girsanov-Meyer) *Suppose that X is a semimartingale under P, with decomposition $X = A + M$ where A is a càdlàg process of locally bounded variation and M is a local martingale. Then X is also a semimartingale under Q. Furthermore*

$$N = M - L^{-1} \bullet [M, L] \tag{A.23}$$

is a Q-local martingale and $X - N$ is of locally bounded variation under Q.

Proof. See Protter (1990, Section III.6).

Example 30 *Let X be the solution of the SDE*

$$\mathrm{d}X_t = \left\{\mu + \beta\sigma^2(t)\right\}\mathrm{d}t + \sigma(t)\mathrm{d}B_t$$

i.e.

$$X_t = \mu t + \beta\sigma^{2*}(t) + \int_0^t \sigma(s)\mathrm{d}B_s$$

where B is Brownian motion and

$$\sigma^{2*}(t) = \int_0^t \sigma^2(s)\mathrm{d}t$$

and we assume that $\sigma(t)$ is a positive bounded càglàd process independent of B, all under a filtered probability space $(\Omega, \{\mathcal{F}_t\}, \mathcal{F}, P)$. We seek an equivalent martingale measure (EMM)*, that is a probability measure locally equivalent to P and such that X is a local martingale under the new measure. Let*

$$H_t = \mu + \beta\sigma^2(t)$$

and note that

$$Y_t = -\int_0^t \frac{H_s}{\sigma(s)}\mathrm{d}B_s$$

is a local P-martingale with

$$\mathrm{d}[Y]_t = \sigma^{-2}(t)H_t^2\mathrm{d}t$$

By the section on exponential martingales we therefore have that

$$L_t = \mathcal{E}(Y)_t = \exp\left\{-\int_0^t \frac{H_s}{\sigma(s)}\mathrm{d}B_s - \frac{1}{2}\int_0^t \frac{H_s^2}{\sigma^2(s)}\mathrm{d}s\right\}$$

is a local martingale under P. If L is, in fact, a martingale then it has mean 1. Consequently, in that case, we may for each $t > 0$ define a probability measure Q_t by

$$\frac{\mathrm{d}Q_t}{\mathrm{d}P_t} = L_t.$$

We are now in position to show that for any $T > 0$ the process $\{X_t : 0 \le t \le T\}$ is a Q_T martingale. Letting

$$A_t = \mu t + \beta\sigma^{2*}(t) \quad \textit{and} \quad M_t = \int_0^t \sigma(s)\mathrm{d}B_s$$

we have that $X = A + M$ is a decomposition of X under P as required in the Girsanov-Meyer Theorem, and with

$$N_t = M_t - \int_0^t L_s^{-1}\mathrm{d}[L, M]_s$$

the theorem gives that for any $T > 0$ the process $\{N_t : 0 \leq t \leq T\}$ is a Q_T-local martingale. What remains to verify is thus that $X = N$. By construction, L is the solution of the SDE

$$\begin{aligned} \mathrm{d}L_t &= L_t \mathrm{d}Y_t \\ &= L_t \frac{H_t}{\sigma(t)} \mathrm{d}B_t \end{aligned}$$

cf. the section on exponential martingales and the fact that L is continuous. Therefore, by Ito calculus,

$$\begin{aligned} \mathrm{d}[L, M]_t &= -L_t \frac{H_t}{\sigma(t)} \mathrm{d}B_t \cdot \sigma(t) \mathrm{d}B_t \\ &= -L_t H_t \mathrm{d}t. \end{aligned}$$

It follows that

$$\begin{aligned} N_t &= M_t - \int_0^t L_s^{-1} \mathrm{d}[L, M]_s \\ &= M_t + \int_0^t H_s \mathrm{d}s \\ &= M_t + A_t \\ &= X_t \end{aligned}$$

as was to be shown.

A.10 Multivariate versions

There are multivariate versions of all the concepts discussed above. In particular we will discuss the multivariate version of Ito's formula. Let $X = (X_1, ..., X_m)$ be an m-tuple of semimartingales and let f be a function from $\mathbf{R}^m$ into $\mathbf{R}$ such that f has continuous second order derivatives. Here we use the notation that $f_{/i}$ is the partial derivative of f with respect to its i-th coordinate, $f_{/ij}$ is the second order partial derivative, etc.

Then $f(X)$ is a semimartingale and

$$\begin{aligned} f(X_t) &= f(X_0) + \sum_{i=1}^m \int_0^t f_{/i}(X_{s-}) \mathrm{d}X_{is} \\ &\quad + \frac{1}{2} \sum_{1 \leq i,j \leq m} \int_0^t f_{/ij}(X_{s-}) \mathrm{d}[X_i, X_j]_s^c \\ &\quad + \sum_{0 < s \leq t} \left\{ f(X_s) - f(X_{s-}) - \sum_{i=1}^m f_{/i}(X_{s-}) \Delta X_{is} \right\}. \end{aligned} \tag{A.24}$$

Example 31 *For any continuous semimartingales $Y_t^1, ..., Y_t^m$ (with starting value 0) we have*

$$Y_t^1 \cdots Y_t^m = \sum_{i=1}^m \int_0^t \prod_{j \neq i} Y_s^j \mathrm{d}Y_s^i + \sum_{1 \leq i < j \leq m} \int_0^t \prod_{k \neq i,j} Y_s^k \mathrm{d}[Y^i, Y^j]_s \tag{A.25}$$

For $m = 2$ this reduces to

$$Y_t^1 Y_t^2 = \int_0^t Y_s^1 \mathrm{d}Y_s^2 + \int_0^t Y_s^2 \mathrm{d}Y_s^1 + \int_0^t \mathrm{d}[Y^1, Y^2]_s \tag{A.26}$$

while for $m = 4$

$$\begin{aligned} Y_t^1 Y_t^2 Y_t^3 Y_t^4 &= \int_0^t Y_s^1 Y_t^2 Y_t^3 \mathrm{d}Y^4 + \int_0^t Y_s^1 Y_t^2 Y_t^4 \mathrm{d}Y^3 + \int_0^t Y_s^1 Y_t^3 Y_t^4 \mathrm{d}Y^2 + \int_0^t Y_t^2 Y_t^3 Y^4 \mathrm{d}Y_s^1 \\ &\quad + \int_0^t Y_s^1 Y_t^2 \mathrm{d}[Y^3, Y^4]_s + \int_0^t Y_s^1 Y_t^3 \mathrm{d}[Y^2, Y^4]_s + \int_0^t Y_s^1 Y_t^4 \mathrm{d}[Y^2, Y^3]_s \\ &\quad + \int_0^t Y_s^2 Y_t^3 \mathrm{d}[Y^1, Y^4]_s + \int_0^t Y_s^2 Y_t^3 \mathrm{d}[Y^1, Y^3]_s + \int_0^t Y_s^3 Y_t^4 \mathrm{d}[Y^1, Y^2]_s. \end{aligned} \tag{A.27}$$

A.11 Ito algebra

With A_t a process of locally bounded variation and B_t the Brownian motion we have

$$\mathrm{d}B_t \cdot \mathrm{d}B_t = \mathrm{d}t \tag{A.28}$$

and

$$\mathrm{d}B_t \cdot \mathrm{d}t = 0 \quad \text{and} \quad \mathrm{d}A_t \cdot \mathrm{d}t = 0. \tag{A.29}$$

If N is a Poison process then

$$(\mathrm{d}N_t)^2 = \mathrm{d}N_t \tag{A.30}$$

while for $s \neq t$

$$\mathrm{d}N_s \mathrm{d}N_t = 0. \tag{A.31}$$

The formulae (A.28) and (A.30) are, in fact, special cases of

$$\mathrm{d}X_t \cdot \mathrm{d}Y_t = \mathrm{d}[X, Y]_t$$

which holds for arbitrary semimartingales X and Y.

These simple *Ito algebra* results are helpful in carrying out calculations.

A.12 Results for Lévy processes

A.12.1 Types of Lévy processes

Recall that the law of a Lévy process z is determined by the Lévy-Khintchine formula for the cumulant function of $z(t)$, which is of the form

$$\mathrm{C}\{z(t) \ddagger y\} = -\frac{1}{2} t b \zeta^2 + i t a \zeta + t \int_{\mathbf{R}} \{e^{i\zeta x} - 1 - i\mathbf{1}_{[-1,1]}(x)\zeta x\} U(\mathrm{d}x) \tag{A.32}$$

where U is a Lévy measure. It is useful to introduce the following classification of Lévy processes.

Definition 3 *A Lévy process having characteristic triplet* (a, b, U) *is said to be of*
(i) class A if $\int_{\mathbf{R}} U(\mathrm{d}x) < \infty$*;*
(ii) class B if $\int_{\mathbf{R}} U(\mathrm{d}x) = \infty$ *and* $\int_{\mathbf{R}} \min\{1, |x|\} U(\mathrm{d}x) < \infty$*;*
(iii) class C if $\int_{\mathbf{R}} \min\{1, |x|\} U(\mathrm{d}x) = \infty$*.*
If, moreover, $b = 0$ *we write* A_0*,* B_0*,* C_0*.*

A Lévy process without Gaussian component (i.e. $b = 0$) is of locally bounded variation if and only if it is of type A_0 or type B_0 (see Sato (1999, Theorem 21.9)). Such a process is sometimes said to be of *finite variation*, while processes of class B_0 and C_0 are said to exhibit *infinite activity*. Note that the class A_0 is precisely the class of compound Poisson processes. If a Lévy process has a Gaussian component then it is of unbounded variation on any finite interval and hence not of locally bounded variation.

A.12.2 Stochastic integration

All Lévy processes are semimartingales. So, in particular, we can consider stochastic integrals of the form

$$f(\cdot, A) \bullet Z,$$

where Z is a Lévy process, f is a real function on $\mathbf{R}_+ \times \mathbf{R}$ and A is a càglàd stochastic process, and f satisfies some mild regularity condition ensuring that the process $f(\cdot, A)$ is again càglàd.

Integration (over bounded intervals) of deterministic functions with respect to a Lévy process is discussed in detail by Rocha-Arteaga and Sato (2001, Section 2.1).

A.12.3 Lévy-Ito formula for Lévy processes

For subordinators Z (without a drift term) the Lévy-Ito representation has the simple form

$$Z_t = \int_0^\infty x N_t(\mathrm{d}x). \tag{A.33}$$

For general Lévy processes, the expression is considerably more complicated. Then Z has the *Lévy-Ito representation*

$$\begin{aligned} Z_t &= t\mathrm{E}\left\{Z_1 - \int_{\{|x|\geq 1\}} x N_t(\mathrm{d}x)\right\} + bB_t \\ &\quad + \int_{\{|x|<1\}} x\{N_t(\mathrm{d}x) - t\nu(\mathrm{d}x)\} \\ &\quad + \int_{\{|x|\geq 1\}} x N_t(\mathrm{d}x), \end{aligned} \tag{A.34}$$

where B is Brownian motion while, for any set Λ, $0 \notin \bar{\Lambda}$ (the closure of Λ), $N_t(\Lambda) = \int_\Lambda N_t(\mathrm{d}x)$ is a Poisson process with mean $t\nu(\Lambda)$, ν being a Lévy measure on $\mathbf{R}$. Furthermore, the Poisson process $N(\Lambda)$ is independent of B and $N(\Lambda)$ is independent of $N(\Gamma)$ if Λ and Γ are disjoint. The representation may be given the alternative form

$$\begin{aligned} Z_t &= at + bB_t \\ &\quad + \int_{\{|x|<\varepsilon\}} x\{N_t(\mathrm{d}x) - t\nu(\mathrm{d}x)\} \\ &\quad + \sum_{0<s\leq t} \mathbf{1}_{\{|\Delta Z_s|\geq\varepsilon\}} \Delta Z_s, \end{aligned} \tag{A.35}$$

where ε is an arbitrarily chosen positive number. For details, see Protter (1990, Section I.4).

A.12.4 Quadratic variation of Lévy processes

The quadratic variation $[Z]$ of an arbitrary Lévy process Z is again a Lévy process, in fact a subordinator. This is an immediate consequence of the definition of quadratic variation. Furthermore, by the Lévy-Ito representation (A.34) we have, writing $N(\mathrm{d}t, \mathrm{d}x)$ for $\mathrm{d}N_t(\mathrm{d}x)$, that

$$\begin{aligned} \mathrm{d}Z_t &= a\mathrm{d}t + b\mathrm{d}B(t) \\ &\quad + \int_{\{|x|<1\}} x\{N(\mathrm{d}t, \mathrm{d}x) - \mathrm{d}t\nu(\mathrm{d}x)\} \\ &\quad + \int_{\{|x|<1\}} x N(\mathrm{d}t, \mathrm{d}x), \end{aligned}$$

from which, using the Ito algebra formulae, we obtain

$$(\mathrm{d}Z_t)^2 = b^2\mathrm{d}t + \int_{\mathbf{R}\setminus\{0\}} x^2 N(\mathrm{d}t, \mathrm{d}x).$$

Hence the Lévy-Ito representation of $[Z]$ is

$$[Z]_t = b^2 t + \int_{\mathbf{R}\setminus\{0\}} x^2 N(t, \mathrm{d}x).$$

A.12.5 Density transformations

Sato (2000).

Theorem A.5 *Let X be a Lévy process with characteristic triplet (a, b, U) under some probability measure P, and let Q be a probability measure on the same filtered space as P. Then (i) and (ii) below are equivalent.*

(i) $Q \overset{loc}{\sim} P$ *and X is a Lévy process under Q, with characteristic triplet $(\tilde{a}, \tilde{b}, \tilde{U})$.*

(ii) The following conditions hold:

- $\tilde{U}(\mathrm{d}x) = k(x)U(\mathrm{d}x)$ *for some Borel function k.*
- $\int\{1 - k(x)\}^2 U(\mathrm{d}x) < \infty$
- $\tilde{a} = a + \int \frac{x}{1+x^2}\{k(x) - 1\}U(\mathrm{d}x) + c\sqrt{b}$ *for some $c \in \mathbf{R}$*
- $\tilde{b} = b$.

Appendix B

Collections of definitions and notation

B.1 Motivation

This book covers a great deal of material drawn from a number of fields. As a result it is easy to lose track of the notation or not recall a term. Hence we have provided an extensive list of definitions and notation. We have split this into two sections: notation and some distributions.

B.2 Notation

- $[y](t)$. The quadratic variation (QV) of the semimartingale process $y(t)$. Split time into small intervals
$$t_0^r = 0 < t_1^r < ... < t_{M_r}^r = t,$$
then
$$[z](t) = \text{p}\!-\!\lim_{r\to\infty} \sum_{j=1}^{M_r} \{y(t_j^r) - y(t_{j-1}^r)\}^2, \quad \text{where} \quad \sup_j \{t_j^r - t_{j-1}^r\} \to 0 \quad \text{for} \quad r \to \infty.$$

- $[y]_i$. Increment to quadratic variation
$$[y]_i = [y](i\hbar) - [y]\{(i-1)\hbar\}.$$

- $[y_M]_i$. Realised variance of y over an interval $\hbar > 0$ using M high frequency observations. It is defined as
$$[y_M]_i = \sum_{j=1}^{M} \left\{ y\left((i-1)\hbar + \frac{j}{M}\hbar\right) - y\left((i-1)\hbar + \frac{(j-1)}{M}\hbar\right)\right\}^2.$$

- $c(s)$. The autocovariance function
$$c(s) = \text{Cov}(x(t), x(t+s)).$$

- $\text{C}\{\zeta \ddagger x\}$. The cumulant function of the random variable x using complex arguments. It is defined as
$$\text{C}\{\zeta \ddagger x\} = \log \text{E}\left(e^{i\zeta x}\right).$$
Thus it is the log of the characteristic function.

- CIR. In economics Feller's square root process is often called the CIR process after Cox, Ingersoll, and Ross (1985) who were the first economists to use this process in economics. It is defined as the solution to the SDE
$$\text{d}\tau(t) = -\lambda\left\{\tau(t) - \xi\right\}\text{d}t + \omega\sqrt{\tau(t)}\text{d}b(\lambda t), \qquad \text{where} \qquad \xi, \lambda, \omega > 0,$$
where b is Brownian motion.

- $D_\upsilon(\omega)$. Functions of Bessel functions
$$D_\nu(\omega) = R_\nu(\omega) R_{-\nu}(\omega) = \frac{K_{\nu+1}(\omega)K_{\nu-1}(\omega)}{K_\nu^2(\omega)},$$
noting that $D_\nu(\omega) = D_{-\nu}(\omega)$ and that $D_\upsilon(\omega)$ is strictly decreasing. Further,
$$D_\nu'(\omega) = D_\nu(\omega)\left\{R_\nu(\omega) - R_\nu(\omega)^{-1} + R_{-\nu}(\omega) - R_{-\nu}(\omega)^{-1} - \frac{2}{\omega}\right\}.$$

We note that

$$\lim_{\omega \downarrow 0} D_\nu(\omega) = \begin{cases} \infty & |\nu| \leq 1 \\ \frac{|\nu|}{|\nu|-1} & |\nu| > 1 \end{cases} \qquad \text{and} \qquad \lim_{\omega \to \infty} D_\nu(\omega) = 1.$$

Hence if $|\nu| \leq 1$ there is a unique ω which solves $r = D_\nu(\omega)$ for any $r \geq 1$..

- D-OU process. A parametric OU process specified through the one-dimensional marginal distribution D of the stationary solution, $\tau(t)$.
- D-OU$_m$. Sum of m independent OU processes whose sum follows a stationary process having one-dimensional marginal law D.
- GH$(\nu, \alpha, \beta, \mu, \delta)$. Generalised hyperbolic distribution. See Table B.3.
- GIG(ν, δ, γ). Generalised inverse Gaussian distribution**.** See Table B.1 and Table B.2.
- $\Gamma(\nu)$. Gamma function

$$\Gamma(\nu) = \int_0^\infty x^{\nu-1} \exp(-x) \mathrm{d}x, \qquad \nu > 0.$$

- $\Gamma(\nu, \alpha)$. Gamma distribution. See Table B.1 and Table B.2.
- HA$(\alpha, \beta, \mu, \delta)$. Hyperbola distribution. See Table B.3.
- H$(\alpha, \beta, \mu, \delta)$. Hyperbolic distribution. See Table B.3.
- $\hbar$. Time space. Could represent, for example, a day or a month. Pronounced hbar.
- i. Represents the index of a period of time, e.g. i-th day or month.
- IG(δ, γ). Inverse Gaussian distribution.
- $J_\nu(x)$. Bessel function of the first kind. Defined as

$$J_\nu(x) = \frac{x^\nu}{2^\nu} \sum_{k=0}^{\infty} (-1)^k \frac{x^{2k}}{2^{2k} k! \Gamma(\nu + k + 1)}. \tag{B.1}$$

For ν of the form $\nu = n + \frac{1}{2}$, where n is an integer, explicit expressions for J_ν exist. Specifically we have

$$J_{n+\frac{1}{2}}(z) = (-1)^n N_{-n-\frac{1}{2}}(z) = (-1)^n z^{n+\frac{1}{2}} \sqrt{(2/\pi)} \frac{\mathrm{d}^n}{(z \mathrm{d}z)^n} \frac{\sin z}{z} \tag{B.2}$$

$$J_{-n-\frac{1}{2}}(z) = (-1)^{n-1} N_{n+\frac{1}{2}}(z) = z^{n+\frac{1}{2}} \sqrt{(2/\pi)} \frac{\mathrm{d}^n}{(z \mathrm{d}z)^n} \frac{\cos z}{z} \tag{B.3}$$

In particular,

$$J_{\frac{1}{2}}(z) = \sqrt{(2/\pi)} z^{-1/2} \sin z \tag{B.4}$$

$$J_{-\frac{1}{2}}(z) = \sqrt{(2/\pi)} z^{-1/2} \cos z \tag{B.5}$$

$$J_{\frac{3}{2}}(z) = \sqrt{(2/\pi)} z^{-1/2} (\frac{\sin z}{z} - \cos z) \tag{B.6}$$

$$J_{-\frac{3}{2}}(z) = -\sqrt{(2/\pi)} z^{-1/2} (\sin z + \frac{\cos z}{z}) \tag{B.7}$$

- $\mathrm{K}(\theta \ddagger x)$. The kumulant function of the random variable x. It is defined as

$$\mathrm{K}(\theta \ddagger x) = \log \mathrm{E}\left(e^{\theta x}\right).$$

  This is the log of the moment generating function.

- $\overline{\mathrm{K}}(\theta \ddagger x)$. The kumulant function of the random variable x using a negative argument θ is written as

$$\overline{\mathrm{K}}\{\theta \ddagger x\} = \log \mathrm{E}\left(e^{-\theta x}\right).$$

- κ_r. The r-th cumulant is given by

$$\kappa_r = \left.\frac{\partial^r \mathrm{K}(\theta \ddagger x)}{\partial \theta^r}\right|_{\theta=0}.$$

  These are related to the centred moments in the following way:

$$\kappa_1 = \mu_1, \qquad \kappa_2 = \mu_2, \qquad \kappa_3 = \mu_3$$

  and

$$\kappa_4 = \mu_4 - 3\mu_2^2, \qquad \kappa_5 = \mu_5 - 10\mu_2\mu_3.$$

- $\acute{\kappa}_r$. Special notation for the cumulants of an OU process. We then reserve κ_r for the $r-th$ cumulant of $z(1)$, the BDLP at time 1. It is then possible to show that

$$\kappa_r = r\acute{\kappa}_r, \quad \text{for} \quad r = 1, 2, ...$$

- $k(\theta)$. Cumulant function of $z(1)$, the BDLP at time one. In particular

$$k(\theta) = \log \mathrm{E}\left[\exp\left\{-\theta z(1)\right\}\right].$$

  An important result is that

$$k(\theta) = \theta \acute{k}'(\theta),$$

  where $\acute{k}'(\theta) = \mathrm{d}\acute{k}(\theta)/\mathrm{d}\theta$.

- $\acute{k}(\theta)$. Kumulant function of an OU process τ

$$\acute{k}(\theta) = \log \mathrm{E}\left[\exp\left\{-\theta \tau(t)\right\}\right].$$

  An important feature is that

$$\acute{k}(\theta) = \int_0^\infty k(\theta e^{-s})\mathrm{d}s.$$

- $K_\nu(x)$. Modified Bessel functions of the third kind are written as $K_\nu(x)$, where

$$K_\nu(x) = \frac{1}{2}\int_0^\infty z^{\nu-1}\exp\left\{-\frac{1}{2}x\left(z + z^{-1}\right)\right\}\mathrm{d}z, \qquad x > 0.$$

  Mathematically tractable special cases include

$$K_{1/2}(x) = \sqrt{\frac{\pi}{2x}}e^{-x} = K_{-1/2}(x),$$

  while

$$K_{n+\frac{1}{2}}(x) = K_{\frac{1}{2}}(x)\left(1 + \sum_{i=1}^{n}\frac{(n+i)!}{i!(n-i)!}2^{-i}x^{-i}\right).$$

It is helpful to note that

$$K'_\nu(x) = -\frac{1}{2}\left\{K_{\nu-1}(x) + K_{\nu+1}(x)\right\} \quad \text{and} \quad K_\nu(x) = K_{-\nu}(x),$$

while

$$K_{\nu+1}(x) = 2\nu x^{-1} K_\nu(x) + K_{\nu-1}(x).$$

The functions $K_0(x)$ and $K_1(x)$ appear in most mathematical and statistical libraries, while general code for computing $K_\nu(x)$ is available at the FN library of NETLIB. It is ported to most scientific software such as `Splus`, `MatLab`, `R` and `Ox`. Generally $K_\nu(x)$ has the important property that for $x \downarrow 0$ we have

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

For the definitions and properties Bessel functions see, for example, Gradstheyn and Ryzhik (1965, pp. 958-71). Finally as $x \to \infty$ so

$$K_\nu(x) \sim \sqrt{(\pi/2)}x^{-1/2}e^{-x}\left(1 + \frac{4\nu^2 - 1}{8}x^{-1} + O(x^{-2})\right),$$

thus, in particular

$$K_\nu(x) \sim \sqrt{(\pi/2)}x^{-1/2}e^{-x} \tag{B.8}$$

for $x \to \infty$.

- $\overline{K}_\nu$. Normalised Bessel functions of the third kind $\overline{K}_\nu(x) = x^\nu K_\nu(x)$. As a result for $\nu \neq 0$ we have that for $x \downarrow 0$

$$\overline{K}_\nu(x) \sim \Gamma(|\nu|)2^{|\nu|-1}.$$

- $\mathrm{La}(\alpha, \beta, \mu)$. Laplace distribution. See Table B.3.

- $\mathrm{LN}(\mu, \sigma^2)$. Log-normal distribution. Parametric distribution on the positive half-line such that $\log x \sim N(\mu, \sigma^2)$. Then density is

$$\frac{1}{\sqrt{2\pi\sigma^2}}\frac{1}{x}\exp\left\{-\frac{1}{2\sigma^2}\left(\log x - \mu\right)^2\right\}, \quad x > 0.$$

The implication is that

$$\mathrm{E}\left(x^r\right) = \exp\left(r\mu + \frac{1}{2}r^2\sigma^2\right).$$

Finally the log-normal does not possess a moment generating function.

- $\mathrm{M}(\theta \ddagger x)$. Moment generating function for x. It is defined as

$$\mathrm{M}(\theta \ddagger x) = \mathrm{E}\left(e^{\theta x}\right).$$

An important feature of the moment generating function is that it can produce the uncentred moments via

$$\mu'_r = E(x^r) = \left.\frac{\partial^r \mathrm{M}(\theta \ddagger x)}{\partial\theta^r}\right|_{\theta=0}.$$

- μ_r. The r-th centred moment. Assuming they exist, let $\mu_1 = \mathrm{E}\left(x\right)$ and then

$$\mu_r = \mathrm{E}\left(x - \mu_1\right)^r, \qquad r = 2, 3, \ldots .$$

- μ'_r. The r-th uncentred moment. Assuming they exist

$$\mu'_r = \mathrm{E}x^r, \qquad r = 2, 3, \ldots .$$

- Meixner(a, b, d, μ). Meixner density is

$$f(x) = \frac{\{2\cos(b/2)\}^{2d}}{2a\pi\Gamma(2d)} \exp\left\{\frac{\beta(x-\mu)}{a}\right\} \left|\Gamma\left(d + \frac{i(x-\mu)}{a}\right)\right|^2, \qquad i = \sqrt{-1},$$

  which has the cumulant function

$$\mathrm{C}\{\zeta \ddagger x\} = i\mu\zeta + 2d \log\left\{\frac{\cos(\beta/2)}{\cosh\left(\frac{a\zeta - i\beta}{2}\right)}\right\}.$$

- N_ν. Bessel function of the second kind. Defined as

$$N_\nu(x) = \frac{1}{\sin\nu\pi}\{\cos\nu\pi J_\nu(x) - J_{-\nu}(x)\}. \tag{B.9}$$

- N(μ, σ^2). Normal distribution**.** See Table B.3.
- N$\Gamma(\nu, \delta, \beta, \mu)$. Normal gamma distribution. See Table B.3.
- NIG$(\alpha, \beta, \mu, \delta)$. Normal inverse Gaussian distribution. See Table B.3.
- NRIG$(\alpha, \beta, \mu, \delta)$. Normal reciprocal inverse Gaussian distribution. See Table B.3.
- OU process. Ornstein-Uhlenbeck process follows the solution to

$$\mathrm{d}\tau(t) = -\lambda\tau(t)\mathrm{d}t + \mathrm{d}z(\lambda t), \qquad \lambda > 0,$$

  where z is a Lévy process. Often we are interested in the special case of non-negative processes, in which case z must be a subordinator.

- OU-D process. A parametric OU process specified through the law of the BDLP at time one, $z(1)$.
- OU$_m$. Sum of independent m independent OU processes.
- PHA(δ, γ). Positive hyperbola distribution. See Table B.1 and Table B.2.
- PH(δ, γ). Positive hyperbolic distribution. See Table B.1 and Table B.2.
- Po(ψ). Poisson distribution. Parametric distribution for the non-negative integers. The density function is, for the parameter ψ,

$$\frac{e^{-\psi}\psi^x}{x!}, \qquad x = 0, 1, 2, \ldots .$$

  We usually write the distribution as $X \sim Po(\psi)$. Finally

$$\overline{\mathrm{K}}(\theta \ddagger x) = \psi\left(1 - e^{-\theta}\right).$$

- PolyLog(n, z). The polylog function.

$$PolyLog(n, z) = \sum_{k=1}^{\infty} z^k / k^n.$$

- $\phi\left(\zeta \ddagger x\right)$. The characteristic function of the random variable x. It is

$$\phi\left(\zeta \ddagger x\right) = \mathrm{E}\left(e^{i\zeta x}\right).$$

- $\Phi\left(.\right)$. Standard normal distribution function.

- $r(s)$. Autocorrelation function. That is

$$r(s) = \mathrm{Cor}(x(t), x(t+s)).$$

  Sometimes we use $r(s \ddagger x)$ to indicate ownership, e.g. the autocorrelation of x is $r(s \ddagger x)$, the autocorrelation of y is $r(s \ddagger y)$.

- $R_\nu\left(\omega\right)$. Ratios of Bessel functions

$$R_\nu\left(\omega\right) = \frac{K_{\nu+1}(\omega)}{K_\nu(\omega)}.$$

  Important properties of these ratios are that

$$R_{-\nu}\left(\omega\right) = R_{\nu-1}\left(\omega\right)^{-1}$$

  The first derivative of this ratio is

$$R'_\nu\left(\omega\right) = R^2_\nu\left(\omega\right) - \frac{2\nu+1}{\omega} R_\nu\left(\omega\right) - 1.$$

- R$\Gamma(\nu, \alpha)$. Reciprocal gamma distribution. See Table B.1 and Table B.2.

- RIG(δ, γ). Reciprocal inverse Gaussian distribution. See Table B.1 and Table B.2.

- $r^*(s)$. Integrated autocorrelation function. That is

$$r^*(s) = \int_0^s r(u)\mathrm{d}u.$$

- $r^{**}(s)$. Doubly integrated autocorrelation function. That is

$$r^{**}(s) = \int_0^s r^*(u)\mathrm{d}u.$$

- $\sigma(t)$. A non-negative stochastic process, usually employed to model the instantaneous or spot volatility in an stochastic volatility model.

- $\tau(t) = \sigma^2(t)$. A non-negative process, usually employed to model the spot variance.

- τ_i. Integrated variance over the period $\hbar\left(i-1\right)$ to $\hbar i$, that is

$$\tau_i = \int_{\hbar(i-1)}^{\hbar i} \tau(t)\mathrm{d}t = \tau^*(\hbar i) - \tau^*(\hbar\left(i-1\right)).$$

- Semimartingale. A process $y(t)$ which can be decomposed as
$$y(t) = a(t) + m(t),$$
where $m(t)$ is a local martingale and $a(t)$ is of locally bounded variation is called a semimartingale. The decomposition is not, in general unique.
- Special semimartingale. A semimartingale whose local bounded variation process is additionally predictable is called a special semimartingale. In such cases the decomposition into a local martingale and a locally bounded variation process is unique and is called canonical.
- Subordinator. Any Lévy process with non-negative increments.
- Superposition. Sum of independent processes.
- supOU. Superposition of OU processes.
- $\tau^*(t)$. Chronometer or integrated variance. Let $\tau(t)$ denote the instantaneous variance of a SV process. Then
$$\tau^*(t) = \int_0^t \tau(t)\mathrm{d}t,$$
is called the integrated variance of the process.
- t. Represents the continuous time clock.
- $\mathrm{T}(\nu, \delta, \beta, \mu)$. Skewed Student's t distribution. In the case of $\beta = 0$ see the symmetric Student's t distribution. See Table B.3.
- y_i. Return over integral of length $\hbar$. That is
$$y_i = y^*(i\hbar) - y^*\left\{(i-1)\,\hbar\right\}.$$
- $u(x)$. Density of the Lévy measure of the BDLP of an OU process.
- $\overline{u}(x)$. Shorthand for $xu(x)$.
- $w(x)$. Density of the Lévy measure $W(x)$. For an OU process with measure $u(x)$, the density of the BDLP measure is
$$\begin{aligned} w(x) &= -u(x) - xu'(x) \\ &= -\overline{u}'(x). \end{aligned}$$
Lévy densities are non-negative (like probability densities), but do not necessarily integrate to one or indeed do not necessarily even integrate.
- $W(x)$. Lévy measure. Lévy measures appear in the Lévy-Khintchine representation and play an important role in manipulating Lévy processes.
- $W^+(x)$. Upper tail integral. It is defined as
$$\begin{aligned} W^+(x) &= \int_x^\infty w(y)\mathrm{d}y \\ &= \overline{u}(x). \end{aligned}$$
- $W^{-1}(x)$. Inverse of the upper tail integral. That is
$$W^{-1}(x) = \inf\left\{y > 0 : W^+(y) \leq x\right\}.$$
- $y^*(t)$. Log-price of an asset at time t.
- $z(t)$. A Lévy process.

B.3 Distributions

B.3.1 Generalised inverse Gaussian (GIG) distributions

Table B.1 and Table B.2 gives the density function of the GIG distribution as well as the corresponding result for many of its most well known special cases. The GIG distribution is important for it provides a rather general parametric form for many positive random variables. The following results are important and hold generally for GIG variables so long as $\min(\delta, \gamma) > 0$.

1. It is easy to see that generically if $X \sim GIG(\nu, \delta, \gamma)$ then

$$X^{-1} \sim GIG(-\nu, \gamma, \delta) \qquad \text{and} \qquad aX \sim GIG\left(\nu, a^{1/2}\gamma, a^{-1/2}\delta\right).$$

2. The density is unimodal with the mode at

$$\sqrt{\left(\frac{\nu-1}{\gamma^2}\right)^2 + \frac{\delta^2}{\gamma^2}} - \frac{\nu-1}{\gamma^2}.$$

   The density is also log-concave if $\nu \geq 1$.

3. If $\delta\gamma > 0$ then

$$\mathrm{E}(X) = \left(\frac{\delta}{\gamma}\right) R_\nu(\delta\gamma) \qquad \text{and} \qquad \mathrm{E}(X^{-1}) = \frac{\gamma}{\delta} R_{-\nu}(\delta\gamma).$$

4. A simple generic method has been derived by Dagpunar (1988, pp. 133-5) to sample from the GIG distribution (see also Atkinson (1982)).

In addition the following special results are useful. We can simulate from a gamma distribution using rejection (e.g. Ripley (1987, pp. 88-90)) or, more slowly, via inverting the distribution function. The former is coded in most statistical packages. Sums of n independent $\Gamma(\nu, \alpha)$ variables are distributed as a $\Gamma(n\nu, \alpha)$ variable. Finally, we recall $\Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$ is the same as a χ^2_ν distribution, while $\Gamma(1, \alpha) = Exp(\alpha)$. We recall the $R\Gamma\left(\frac{\nu}{2}, \frac{1}{2}\right)$ is the same as a χ_ν^{-2} distribution.

The inverse Gaussian distribution function is

$$\Pr(X \leq x) = \Phi\left\{\delta x^{-1/2}\left(\frac{x\gamma}{\delta} - 1\right)\right\} + \exp(2\delta\gamma)\Phi\left\{-\delta x^{-1/2}\left(\frac{x\gamma}{\delta} + 1\right)\right\}.$$

This result is due to Schrödinger (1915). The first three cumulants are

$$\kappa_1 = \frac{\delta}{\gamma}, \quad \kappa_2 = \frac{\delta}{\gamma^3}, \quad \kappa_3 = 3\frac{\delta}{\gamma^5}.$$

We can simulate from the IG distribution without rejection using a method discussed by Devroye (1986, p. 149) (notice Ripley (1987, p. 94) has a small typo which is corrected on his homepage). Sums of n independent $IG(\delta, \gamma)$ variables are distributed as a $IG(n\delta, \gamma)$ variable. Finally, we should note that some textbooks use a different parameterisation than above, writing the density as

$$\frac{\lambda^{1/2}}{\sqrt{2\pi}} x^{-3/2} \exp\left\{-\frac{\lambda}{2\mu^2 x}(x-\mu)^2\right\}, \qquad x > 0.$$

The parameters in the model are related via $\lambda = \delta^2$ and $\mu^2 = \gamma^{-2}\lambda$.

Distribution	Density function	Lévy density $w(x)$
$GIG(\nu, \delta \geq 0, \gamma \geq 0)$	$\frac{(\gamma/\delta)^\nu}{2K_\nu(\delta\gamma)} x^{\nu-1} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}$	$\left\{\frac{1}{2}\int_0^\infty e^{-\frac{1}{2}\delta^{-2}x\xi} g_\nu(\xi) d\xi + \max(0,\nu)\,\lambda\right\}$ $\times \exp\left(-\gamma^2 x/2\right) x^{-1}$
$\Gamma(\nu > 0, \alpha = \gamma^2/2)$ $= GIG(\nu > 0, 0, \gamma)$	$\frac{\alpha^\nu}{\Gamma(\nu)} x^{\nu-1} \exp(-\alpha x)$	$\nu x^{-1} \exp(-\alpha x)$
$IG(\delta > 0, \gamma \geq 0)$ $= GIG\left(-\frac{1}{2}, \delta, \gamma\right)$	$\frac{\delta e^{\delta\gamma}}{\sqrt{2\pi}} x^{-3/2} \exp\left\{-\frac{1}{2}\left(\delta^2 x^{-1} + \gamma^2 x\right)\right\}$	$\frac{\delta}{\sqrt{2\pi}} x^{-3/2} \exp\left(-\frac{1}{2}\gamma^2 x\right)$
$PH(\delta > 0, \gamma \geq 0)$ $= GIG(1, \delta, \gamma)$	$\frac{(\gamma/\delta)}{2K_1(\delta\gamma)} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}$	
$\mathrm{R}\Gamma(\nu > 0, \alpha = \delta^2/2)$ $= GIG(-\nu, \delta, 0)$	$\frac{\alpha^\nu}{\Gamma(\nu) x^{\nu+1}} \exp\left(-\alpha x^{-1}\right)$	
$RIG(\delta > 0, \gamma \geq 0)$ $= GIG\left(\frac{1}{2}, \delta, \gamma\right)$	$\frac{\gamma e^{\delta\gamma}}{\sqrt{2\pi}} x^{-1/2} \exp\left\{-\frac{1}{2}\left(\delta^2 x^{-1} + \gamma^2 x\right)\right\}$	
$RPH(\delta, \gamma)$ $= GIG(-1, \delta, \gamma)$	$\frac{(\gamma/\delta)^{-1}}{2K_1(\delta\gamma)} x^{-2} \exp\left\{-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right\}$	

Table B.1: Summary of the GIG distribution and its special cases. Recorded are the densities and Lévy densities. Here K_ν is a modified Bessel function of the third kind. Also $g_\nu(x) = \frac{2}{x\pi^2}\left\{J^2_{|\nu|}(\sqrt{x}) + N^2_{|\nu|}(\sqrt{x})\right\}^{-1}$.

Distribution	Cumulant function $\bar{K}(\theta)$	Moments
$GIG(\nu, \delta \geq 0, \gamma \geq 0)$	$\log K_\nu \left\{\delta\gamma \left(1 + 2\theta/\gamma^2\right)^{1/2}\right\} - \log K_\nu(\delta\gamma)$ $+\nu \log \gamma - \frac{\nu}{2} \log\left(\gamma^2 + 2\theta\right)$	
$\Gamma(\nu > 0, \alpha = \gamma^2/2)$ $= GIG(\nu > 0, 0, \gamma)$	$-\nu \log\left(1 + \frac{\theta}{\alpha}\right)$	$\kappa_r = \frac{(r-1)!\nu}{\alpha^r}$
$IG(\delta > 0, \gamma \geq 0)$ $= GIG\left(-\frac{1}{2}, \delta, \gamma\right)$	$\delta\gamma - \delta\left(\gamma^2 + 2\theta\right)^{1/2}$	$\kappa_r = 1 \cdot 3 \cdots (2r-3)$ $\times \delta\gamma^{-2r+1}$
$PH(\delta > 0, \gamma \geq 0)$ $= GIG(1, \delta, \gamma)$	$\log\{1 + 2\theta/\gamma^2\}^{1/2} - \log K_1(\delta\gamma)$ $+ \log K_1 \left\{\delta\gamma \left(1 + 2\theta/\gamma^2\right)^{1/2}\right\}$	
$\mathrm{R}\Gamma(\nu > 0, \alpha = \delta^2/2)$ $= GIG(-\nu, \delta, 0)$		$\mu_r = \frac{1}{\alpha^r} \frac{\Gamma(\nu - r)}{\Gamma(\nu)}$
$RIG(\delta > 0, \gamma \geq 0)$ $= GIG\left(\frac{1}{2}, \delta, \gamma\right)$	$\log \gamma^2 - \frac{1}{2} \log\left(\gamma^2 + 2\theta\right)$ $+\delta\gamma - \delta\left(\gamma^2 + 2\theta\right)^{1/2}$	
$RPH(\delta, \gamma)$ $= GIG(-1, \delta, \gamma)$	$\log K_1 \left\{\delta\gamma \left(1 + 2\theta/\gamma^2\right)^{1/2}\right\} - \log K_1(\delta\gamma)$ $- \log \gamma + \frac{1}{2} \log\{\gamma^2 + 2\theta\}$	

Table B.2: Summary of the cumulant structure of the GIG distribution and its special cases. Recorded are the cumulant function and either the moments or cumulants (which ever has a simpler form). Here K_ν is a modified Bessel function of the third kind.

B.3.2 Generalised hyperbolic (GH) distributions

The GH distribution covers many of the special cases of distributions with support on the real line used in this book**.** It is constructed by

$$x = \mu + \beta\sigma^2 + \sigma\varepsilon,$$

where $\sigma^2 \sim GIG(\nu, \delta, \gamma)$ and is independent of $\varepsilon \sim N(0,1)$. Throughout we write $\alpha = \sqrt{\beta^2 + \gamma^2}$. The normal variance-mean mixture representation allows easy simulation, while its cumulant function also follows immediately as

$$\mathrm{K}(\theta \ddagger x) = \mu\theta + \mathrm{K}\left\{\left(\beta + \frac{1}{2}\sigma^2\right) \ddagger \theta\right\},$$

that is the cumulant function is

$$\mathrm{K}(\theta \ddagger x) = \frac{\nu}{2}\log\left\{\frac{\gamma}{\alpha^2 - (\beta+\theta)^2}\right\} + \log\left\{\frac{K_\nu\left\{\delta\sqrt{\alpha^2 - (\beta+\theta)^2}\right\}}{K_\nu\left\{\delta\sqrt{\alpha^2-\beta^2}\right\}}\right\} + \theta\mu, \quad |\beta+\theta| < \alpha.$$

The first two moments are

$$\begin{aligned}
\mathrm{E}(X) &= \mu + \beta\frac{\delta}{\gamma}\frac{K_{\nu+1}(\delta\gamma)}{K_\nu(\delta\gamma)} \quad \text{and} \\
\mathrm{Var}(X) &= \delta^2\left\langle\frac{K_{\nu+1}(\delta\gamma)}{\delta\gamma K_\nu(\delta\gamma)} + \frac{\beta^2}{\gamma^2}\left[\frac{K_{\nu+2}(\delta\gamma)}{K_\nu(\delta\gamma)} - \left\{\frac{K_{\nu+1}(\delta\gamma)}{K_\nu(\delta\gamma)}\right\}^2\right]\right\rangle.
\end{aligned}$$

K_ν is the modified Bessel function of the third kind. Special cases of the class of distributions are given in Table B.3.

Some results for the special cases given in Table B.3 turn out to be helpful.

- **Hyperbolic distribution.** Parametric distribution on the real line (special case of generalised hyperbolic with $\nu = 1$). Cumulant generating function

$$\mathrm{K}(\xi \ddagger x) = \log\xi - \log K_1(\xi) + \log K_1\left(\sqrt{\xi^2 + \delta^2 u^2}\right) - \frac{1}{2}\log\left(\xi^2 + \delta^2 u^2\right).$$

- **Normal gamma** distribution has the cumulant function

$$\mu\theta + \nu\log\left(1 + \frac{\theta\beta + \theta^2/2}{\gamma}\right).$$

- **Normal inverse Gaussian distribution.** NIG's cumulant generating function

$$\mathrm{K}(\xi \ddagger x) = \delta\left\{\sqrt{\alpha^2 - \beta^2} - \sqrt{\alpha^2 - (\beta+\xi)^2}\right\} + \mu\xi.$$

  The first four cumulants are, writing $\rho = \beta/\alpha$,

$$\kappa_1 = \mu + \frac{\delta\rho}{\sqrt{1-\rho^2}}, \quad \kappa_2 = \frac{\delta^2}{(\delta\alpha)(1-\rho^2)^{3/2}}$$

  and

$$\kappa_3 = \frac{3\delta^3\rho}{(\delta\alpha)^2(1-\rho^2)^{5/2}}, \quad \kappa_4 = \frac{3\delta^4(1+4\rho^2)}{(\delta\alpha)^3(1-\rho^2)^{7/2}}.$$

- **Student's t distribution.** Parametric model on the real line which is symmetric about μ with scale parameter σ. The density function is, for d degrees of freedom,

$$\frac{\Gamma\left(\frac{d+1}{2}\right)}{(\pi d\sigma^2)^{1/2}\,\Gamma\left(\frac{d}{2}\right)}\left[1+\left\{\frac{(x-\mu)^2}{\sigma^2 d}\right\}\right]^{-(d+1)/2}, \quad d>0.$$

This distribution does not have a cumulant generating function. $\mathrm{E}(x^r)$ exists as long as $r<d$. In particular

$$\kappa_1=\mu, \quad \text{and} \quad \kappa_2=\sigma^2\frac{d}{d-2}.$$

Special case of the generalised hyperbolic distribution with $\gamma=0$ and $\nu<0$. Importantly the Student t distribution is obtained as a mixture of normals with

$$x=\mu+\sigma\frac{g}{\sqrt{c/d}}, \quad g\sim N(0,1), \quad c\sim\Gamma\left(\frac{d}{2},\frac{1}{2}\right),$$

where g and c are independent. This structure is just a reparameterisation of the symmetric $T(\nu,\delta,\mu)$ given in Table B.3, where the density is

$$\frac{\Gamma(\nu+1/2)}{\sqrt{\pi}\delta\Gamma(\nu)}\left[1+\left\{\frac{(x-\mu)^2}{\delta^2}\right\}\right]^{-\nu-1/2}.$$

In particular we have that

$$\nu=\frac{d}{2}, \qquad \delta^2=\sigma^2 d.$$

This demonstrates that the more general $T(\nu,\delta,\beta,\mu)$ can be thought about as a skewed student-t distribution.

Distribution	σ^2	Density function
$GH(\nu,\alpha,\beta,\mu,\delta)$	$GIG(\nu,\delta,\gamma)$	$\frac{(\gamma/\delta)^{\nu}}{\sqrt{2\pi}\alpha^{\left(\nu-\frac{1}{2}\right)}K_{\nu}(\delta\gamma)}\left\{\delta^2+(x-\mu)^2\right\}^{\frac{1}{2}\left(\nu-\frac{1}{2}\right)}K_{\left(\nu-\frac{1}{2}\right)}\left(\alpha\sqrt{\delta^2+(x-\mu)^2}\right)e^{\beta(x-\mu)}$
$N(\mu,\sigma^2)=\lim_{\gamma\to\infty}$ $GH(\nu,\gamma,0,\mu,\sigma^2\gamma)$	σ^2	$\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$
$RH(\alpha,\beta,\mu,\delta)$ $=GH(-1,\alpha,\beta,\mu,\delta)$	$RPH(\delta,\gamma)$	
$NIG(\alpha,\beta,\mu,\delta)$ $=GH\left(-\frac{1}{2},\alpha,\beta,\mu,\delta\right)$	$IG(\delta,\gamma)$	$\pi^{-1}\alpha\exp\left\{\delta\sqrt{\alpha^2-\beta^2}-\beta\mu\right\}q\left(\frac{x-\mu}{\delta}\right)^{-1}K_1\left\{\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right\}e^{\beta(x-\mu)}$
HA $=GH(0,\alpha,\beta,\mu,\delta)$	$GIG(0,\delta,\gamma)$	$\frac{1}{2\alpha^{-1}K_0(\delta\gamma)}\left\{\delta^2+(x-\mu)^2\right\}^{-\frac{1}{2}}\exp\left\{-\alpha\sqrt{\delta^2+(x-\mu)^2}\right\}\exp\{\beta(x-\mu)\}$
$NRIG(\alpha,\beta,\mu,\delta)$ $=GH\left(\frac{1}{2},\alpha,\beta,\mu,\delta\right)$	$RIG(\delta,\gamma)$	
$H(\alpha,\beta,\mu,\delta)$ $=GH(1,\alpha,\beta,\mu,\delta)$	$PH(\delta,\gamma)$	$\frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1\left(\delta\sqrt{\alpha^2-\beta^2}\right)}\exp\left\{-\alpha\sqrt{\delta^2+(x-\mu)^2}+\beta(x-\mu)\right\}$
$La(\alpha,\beta,\mu)$ $=GH(1,\alpha,\beta,\mu,0)$	$\Gamma\left(1,\frac{\gamma^2}{2}\right)$	$\frac{\alpha^2-\beta^2}{2\alpha}\exp\{-\alpha\lvert x-\mu\rvert+\beta(x-\mu)\}.$
$N\Gamma(\nu,\gamma,\beta,\mu)$ $=GH(\nu,\alpha,\beta,\mu,0)$	$\Gamma\left(\nu,\frac{\gamma^2}{2}\right)$	$\frac{\gamma^{2\nu}\left(\frac{\gamma^2}{2}\right)^{1-2\nu}}{\sqrt{2\pi}\Gamma(\nu)2^{\nu-1}}\bar{K}_{\nu-1/2}\left(\frac{\gamma^2}{2}\lvert x-\mu\rvert\right)\exp\{\beta(x-\mu)\}$
$T(\nu,\delta,\beta,\mu)$ $=GH(-\nu,\beta,\beta,\mu,\delta)$	$R\Gamma\left(\nu,\frac{\delta^2}{2}\right)$	$\frac{1}{\sqrt{2\pi}\delta\Gamma(\nu)2^{\nu-1}}q\left(\frac{x-\mu}{\delta}\right)^{-2\nu-1}\bar{K}_{\nu+1/2}\left\{\delta\beta q\left(\frac{x-\mu}{\delta}\right)\right\}\exp\{\beta(x-\mu)\}$
$T(\nu,\delta,\mu)$ $=GH(-\nu,0,0,\mu,\delta)$	$R\Gamma\left(\nu,\frac{\delta^2}{2}\right)$	$\frac{\Gamma(\nu+1/2)}{\sqrt{\pi}\delta\Gamma(\nu)}\left[1+\left\{\frac{(x-\mu)^2}{\delta^2}\right\}\right]^{-\nu-1/2}$

Table B.3: Summary of the GH distribution and its special cases. Recorded are the densities. where $\overline{K}_\nu(x)=x^\nu K_\nu(x)$, $q(x)=\sqrt{1+x^2}$ and $\alpha=\sqrt{\beta^2+\gamma^2}$. K_ν is the modified Bessel function of the third kind. Here K_ν is a modified Bessel function of the third kind.

B.3.3 Stable based distributions

This class of models is based on aspects of the stable distribution. Let $p(x;\kappa,\delta)$ denote the probability density function of the positive κ-stable law $S(\kappa,\delta)$ with cumulant transform

$$-\delta(2\theta)^{\kappa}, \qquad 0<\kappa<1.$$

Let $p(x;\kappa,\delta,\gamma)$ denote the tempered (exponentially tilted) version of $p(x;\kappa,\delta)$ defined by

$$p(x;\kappa,\delta,\gamma)=e^{\delta\gamma}p(x;\kappa,\delta)e^{-\frac{1}{2}\gamma^{1/\kappa}x}, \qquad \kappa\in(0,1),\ \delta>0,\ \gamma\geq 0. \tag{B.10}$$

The distribution with density (B.10) will be referred to as a *tempered stable law* and we denote it by $TS(\kappa,\delta,\gamma)$.

Next, consider for any $\nu\in\mathbf{R}$ and $\gamma\vee(-\nu)>0$ the derived probability density

$$p(x;\kappa,\nu,\delta,\gamma)=c(\kappa,\nu,\delta,\gamma)x^{\nu+\kappa}p(x;\kappa,\delta,\gamma), \tag{B.11}$$

where $c(\kappa,\nu,\delta,\gamma)$ is a norming constant. We denote by MS (*modified stable*) the class of distributions on the positive halfline whose densities are of the form $p(x;\kappa,\nu,\delta,\gamma)$. Correspondingly, the law determined by $p(x;\kappa,\nu,\delta,\gamma)$ is denoted $MS(\kappa,\nu,\gamma,\delta)$. The subclass of the family of MS laws obtained for $\kappa=\frac{1}{2}$ is the class of GIG (generalised inverse Gaussian) distributions.

Correspondingly, and in analogy with the construction of the generalised hyperbolic distributions, a random variable x is said to be distributed according to the *normal modified stable* law $NMS(\kappa,\nu,\gamma,\beta,\mu,\delta)$ if it is of the normal variance-mean mixture form

$$x=\mu+\beta\tau+\sqrt{\tau}\varepsilon,$$

with $\varepsilon\sim N(0,1)$ and $\tau\sim MS(\kappa,\nu,\gamma,\delta)$ and τ and ε independent. The special case of $\kappa=-\nu$ yields the simpler *normal tempered stable* $NTS(\kappa,\gamma,\beta,\mu,\delta)$ law which is based on a normal variance-mean mixture using the tilted stable distribution.

Bibliography

Abramowitz, M. and I. A. Stegun (1970). *Handbook of Mathematical Functions*. New York: Dover Publications Inc.

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

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

Andersen, T. G. and T. Bollerslev (1998b). Deutsche mark-dollar volatility: intraday activity patterns, macroeconomic announcements, and longer run dependencies. *Journal of Finance 53*, 219–265.

Andersen, T. G., T. Bollerslev, and F. X. Diebold (2002). Parametric and nonparametric measurement of volatility. In Y. Ait-Sahalia and L. P. Hansen (Eds.), *Handbook of Financial Econometrics*. Amsterdam: North Holland. Forthcoming.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens (2001). The distribution of realized stock return volatility. *Journal of Financial Economics 61*, 43–76.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001a). The distribution of exchange rate volatility. *Journal of the American Statistical Association 96*, 42–55.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and P. Labys (2001b). Modeling and forecasting realized volatility. Unpublished paper: Department of Economics, Duke University.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis* (2 ed.). New York: John Wiley and Sons.

Andreou, E. and E. Ghysels (2001). Rolling-sampling volatility estimators: some new theoretical, simulation and empirical results. *Journal of Business and Economic Statistics 19*. Forthcoming.

Ané, T. and H. Geman (2000). Order flow, transaction clock and normality of asset returns. *Journal of Finance 55*, 2259–2284.

Apostol, T. M. (1957). *Mathematical Analysis*. London: Addison-Wesley.

Areal, N. M. P. C. and S. J. Taylor (2002). The realized volatility of FTSE-100 futures prices. *Journal of Futures Markets 22*. Forthcoming.

Asmussen, S. and J. Rosinski (2000). Approximation of small jumps of Lévy processes with a view towards simulation. Research Report no. 26. MaPhySto, Aarhus University.

Atkinson, A. C. (1982). The simulation of generalised inverse Gaussian and hyperbolic random variables. *SIAM Journal of Scientific and Statistical Computing 3*, 502–17.

Bachelier, L. (1900). *Theorie del la speculation*. Paris Doctoral Dissertation in Mathematics.

Back, K. (1991). Asset pricing for general processes. *Journal of Mathematical Economics 20*, 371–395.

Bagnold, R. A. (1941). *The Physics of Blown Sand and Desert Dunes*. London: Methuen.

Bai, X., J. R. Russell, and G. C. Tiao (2000). Beyond Merton's utopia: effects of non-normality and dependence on the precision of variance estimates using high-frequency financial data. Unpublished paper: Graduate School of Business, University of Chicago.

Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences 353*, 401–419.

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

Barndorff-Nielsen, O. E. (1998a). Probability and statistics: selfdecomposability, finance and turbulence. In L. Accardi and C. C. Heyde (Eds.), *Probability Towards 2000. Proceedings of a Symposium held 2-5 October 1995 at Columbia University*, pp. 47–57. New York: Springer-Verlag.

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

Barndorff-Nielsen, O. E. (2001). Superposition of Ornstein-Uhlenbeck type processes. *Theory of Probability and its Applications 45*, 175–194.

Barndorff-Nielsen, O. E., P. Blæsild, J. L. Jensen, and M. Sørensen (1985). The fascination of sand. In A. C. Atkinson and S. E. Fienberg (Eds.), *A Celebration of Statistics*, pp. 57–87. New York: Springer-Verlag.

Barndorff-Nielsen, O. E. and D. R. Cox (1989). *Asymptotic Techniques for use in Statistics*. London: Chapman & Hall.

Barndorff-Nielsen, O. E. and D. R. Cox (1994). *Inference and Asymptotics*. London: Chapman & Hall.

Barndorff-Nielsen, O. E. and C. Halgreen (1977). Infinite divisibility of the hyperbolic and generalised inverse Gaussian distributions. *Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 38*, 309–311.

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

Barndorff-Nielsen, O. E., T. Mikosch, and S. Resnick (2001). *Lévy Processes – Theory and Applications*. Boston: Birkhäuser.

Barndorff-Nielsen, O. E. and K. Prause (2001). Apparent scaling. *Finance and Stochastics 5*, 103–113.

Barndorff-Nielsen, O. E. and N. Shephard (2001b). Modelling by Lévy processes for financial econometrics. In O. E. Barndorff-Nielsen, T. Mikosch, and S. Resnick (Eds.), *Lévy Processes – Theory and Applications*, pp. 283–318. Boston: Birkhäuser.

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

Barndorff-Nielsen, O. E. and N. Shephard (2002e). Econometric analysis of realised covariation: high frequency covariance, regression and correlation in financial economics. Unpublished paper: Nuffield College, Oxford.

Barndorff-Nielsen, O. E. and N. Shephard (2002c). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society, Series B 64*, 253–280.

Barndorff-Nielsen, O. E. and N. Shephard (2002d). Estimating quadratic variation using realised variance. *Journal of Applied Econometrics*. Forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2002b). Integrated OU processes and non-Gaussian OU-based stochastic volatility. *Scandinavian Journal of Statistics*. Forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2002a). Normal modified stable processes. *Theory of Probability and Mathematical Statistics*. Forthcoming.

Barndorff-Nielsen, O. E. and N. Shephard (2003). How accurate is the asymptotic approximation to the distribution of realised volatility? In D. Andrews, J. Powell, P. A. Ruud, and J. H. Stock (Eds.), *Identification and Inference for Econometric Models. A Festschrift in Honour of Thomas J. Rothenberg*, Econometric Society Monograph Series. Cambridge: Cambridge University Press. Forthcoming.

Barndorff-Nielsen, O. E. and M. Sørensen (1995). A review of some aspects of asymptotic likelihood theory for stochastic processes. *International Statistical Review 62*, 133–165.

Bauer, C. (2000). Value at risk using hyperbolic distributions. *Journal of Economics and Business 52*, 455–467.

Becker, E. (1998). Theorems limite pour des processus discretises. These de Doctorat de l'Universite Paris 6.

Ben-Hamou, E. (2000). Option pricing with levy process.

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

Bertoin, J. (2001). *Subordinators: Examples and Applications*. Berlin: Springer. Presented at Ecole d'Ete de St-Flour, 1997. Forthcoming: Lecture Notes in Mathematics.

Bibby, B. M. and M. Sørensen (1997). A hyperbolic diffusion model for stock prices. *Finance and Stochastics 1*, 25–41.

Bibby, B. M. and M. Sørensen (2001). Hyperbolic processes in finance. Unpublished paper: Department of Statistics, University of Copenhagen.

Bingham, N. and R. Kiesel (2000). Modelling asset returns with hyperbolic distributions. In J. Knight and S. E. Satchell (Eds.), *Asset Return Distributions*. Butterworth-Heinemann.

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

Blattberg, R. C. and N. Gonedes (1974). A comparison of the stable and student distributions as models for stock prices. *Journal of Business 47*, 244–280.

Bochner, S. (1949). Diffusion equation and stochastic processes. *Proc. National Academy of Science, USA 85*, 369–370.

Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*. Berkeley: University of California Press.

Bollerslev, T. and L. Forsberg (2002). The distribution of realised volatility and the normal inverse gaussian GARCH model: An application to the ECU and EURO exchange rates. *Journal of Applied Econometrics*. Forthcoming.

Bollerslev, T. and H. Zhou (2001). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*. Forthcoming.

Bondesson, L. (2000). On the Lévy measure of the lognormal and the logcauchy distributions. Research Report No. 10, Dept. Mathematical Statistics, Umeå University.

Box, G. E. P. and G. M. Jenkins (1970). *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day.

Boyarchenko, S. I. and S. Z. Levendorskii (1999). Generalizations of the Black-Scholes equation for truncated Lévy processes. Working paper.

Boyarchenko, S. I. and S. Z. Levendorskii (2000d). Barrier options and touch-and-out options under regular Lévy processes. Working paper.

Boyarchenko, S. I. and S. Z. Levendorskii (2000c). European barrier options and American digitals under Lévy processes. Working paper.

Boyarchenko, S. I. and S. Z. Levendorskii (2000a). Option pricing for truncated Lévy processes. *International Journal of Theoretical and Applied Finance 3*, 549–552.

Boyarchenko, S. I. and S. Z. Levendorskii (2000b). Perpetual American options under Lévy processes. Working paper.

Brix, A. (1999). Generalized gamma measures and shot-noise Cox processes. *Advances in Applied Probability 31*, 929–953.

Brockwell, P. J. (2001). Lévy driven CARMA processes. *Annals of the Institute for Statistical Mathematics 52*, 113–124.

Brockwell, P. J. and R. A. Davis (1987). *Time Series: Theory and Methods*. New York: Springer-Verlag.

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

Cai, J., Y.-L. Cheung, R. S. K. Lee, and M. Melvin (1999). 'once-in-a-generation' yen volatility in 1998: fundamentals, intervention and order flow.

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

Cao, C. Q. and R. S. Tsay (1992). Nonlinear time-series analysis of stock volatilities. *Journal of Applied Econometrics 7*, S165–S185.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2001). Stochastic volatility for Lévy processes. Unpublished paper: Robert H Smith School of Business, Maryland University.

Carr, P., H. Geman, D. B. Madan, and M. Yor (2002). The fine structure of asset returns: an empirical investigation. *Journal of Business 76*. Forthcoming.

Chesney, M. and L. O. Scott (1989). Pricing European options: a comparison of the modified Black-Scholes model and a random variance model. *J. Financial and Qualitative Analysis 24*, 267–84.

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

Chriss, N. and W. Morokoff (1999). Volatility and variance swaps. *Risk 12*, 55–59.

Christensen, B. J. and N. R. Prabhala (1998). The relation between implied and realized volatility. *Journal of Financial Economics 37*, 125–150.

Cinlar, E. and M. Pinsky (1972). On dams with additive inputs and a general release rule. *Journal of Applied Probability 9*, 422–429.

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

Clements, M. P. and D. F. Hendry (1999). *Forecasting Non-stationary Economic Time Series: The Zeuthen Lectures on Economic Forecasting*. Cambridge, Mass.: MIT Press.

Comte, F. and E. Renault (1998). Long memory in continuous-time stochastic volatility models. *Mathematical Finance 8*, 291–323.

Corsi, F., G. Zumbach, U. Muller, and M. Dacorogna (2001). Consistent high-precision volatility from high-frequency data. Unpublished paper: Olsen and Associates, Zurich.

Cox, D. R. (1991). Long-range dependence, non-linearity and time irreversibility. *Journal of Time Series Analysis 12*, 329–335.

Cox, J. C., J. E. Ingersoll, and S. A. Ross (1985). A theory of the term structure of interest rates. *Econometrica 53*, 385–407.

Cox, J. C. and S. Ross (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics 3*, 145–166.

Cressie, N. (1993). *Statistics for Spatial Data*. New York: Wiley.

Dacorogna, M. M., R. Gencay, U. A. Muller, R. B. Olsen, and O. V. Pictet (2001). *An Introduction to High-Frequency Finance*. San Diego: Academic Press.

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

Dagpunar, J. (1988). *Principles of Random Number Generation*. Oxford: Oxford University Press.

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

Davis, R. A. and T. Mikosch (1998). The limit theory for the sample ACF of stationary process with heavy tails with applications to ARCH. *Annals of Statistics 26*, 2049–2080.

Delattre, S. and J. Jacod (1997). A central limit theorem for normalized functions of the increments of a diffusion process in the presence of round off errors. *Bernoulli 3*, 1–28.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou (1999). A guide to volatility and variance swaps. *Journal of Derivatives 6*, 9–32.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York: Springer-Verlag.

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

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

Duffie, D. (1996). *Dynamic Asset Pricing Theory* (2 ed.). New Jersey: Princeton University Press.

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

Duffie, D. and K. J. Singleton (1993). Simulated moments estimation of markov models of asset proces. *Econometrica 61*, 929–952.

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

Eberlein, E. and U. Keller (1995). Hyperbolic distributions in finance. *Bernoulli 1*, 281–299.

Eberlein, E., U. Keller, and K. Prause (1998). New insights into smile, mispricing and value at risk: the hyperbolic model. *Journal of Business 71*, 371–405.

Elerian, O., S. Chib, and N. Shephard (2001). Likelihood inference for discretely observed non-linear diffusions. *Econometrica 69*, 959–993.

Engle, R. F. and C. W. J. Granger (1987). Co-integration and error correction: representation, estimation and testing. *Econometrica 55*, 251–276.

Epps, T. W. and M. L. Epps (1976). The stochastic dependence of security price changes and transaction volumes: implications for the mixture-of-distributions hypothesis. *Econometrica 44*, 305–321.

Fama, E. (1965). The behaviour of stock market prices. *J Business 38*, 34–105.

Feller, W. (1971). *An Introduction to Probability Theory and Its Applications* (2 ed.), Volume 2. New York: John Wiley.

Ferguson, T. S. and M. J. Klass (1972). A representation of independent increment processes without Gaussian components. *Annals of Mathematical Statistics 43*, 1634–1643.

Fleming, J., C. Kirby, and B. Ostdiek (2002). The economic values of volatility timing using 'realized' volatility. *Journal of Financial Economics*. Forthcoming.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability 30*, 790–804.

Foster, D. P. and D. B. Nelson (1996). Continuous record asymptotics for rolling sample variance estimators. *Econometrica 64*, 139–174.

Francq, C. and J.-M. Zakoïan (2000). Covariance matrix estimation for estimators of mixing weak ARMA models. *Journal of Statistical Planning and Inference 83*, 369–94.

Galbraith, J. W. and V. Zinde-Walsh (2000). Properties of estimates of daily GARCH parameters based on intra-day observations. Unpublished paper: Economics Department, McGill University.

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

Geman, H., D. Madan, and M. Yor (2000). Time changes hidden in Brownian subordination.

Genon-Catalot, V., T. Jeantheau, and C. Larédo (2000). Stochastic volatility as hidden Markov models and statistical applications. *Bernoulli 6*, 1051–1079.

Genon-Catalot, V., C. Laredo, and D. Picard (1992). Non-parametric estimation of the diffusion coefficient by wavelet methods. *Scandinavian Journal of Statistics 19*, 317–335.

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

Good, I. J. (1953). The population frequencies of species and the estimation of population parameters. *Biometrika 40*, 237–264.

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

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

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

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics 14*, 227–238.

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

Granger, C. W. J. and C.-Y. Sin (1999). Modelling the absolute returns of different stock indices: exploring the forecastability of an alternative measure of risk. Working paper: Department of Economics, University of California at San Diego.

Grigelionis, B. (1999). Processes of Meixner type. *Lithuanian Mathematics Journal 39*, 33–41.

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

Hamilton, J. (1994). *Time Series Analysis*. Princeton: Princeton University Press.

Hansen, B. E. (1995). Regression with non-stationary volatility. *Econometrica 63*, 1113–1132.

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

Harvey, A. C., E. Ruiz, and N. Shephard (1994). Multivariate stochastic variance models. *Review of Economic Studies 61*, 247–264.

Hasbrouck, J. (1996). Modeling Market Microstructure Time Series. In C. R. Rao and G. S. Maddala (Eds.), *Statistical Methods in Finance*, pp. 647–692. Amsterdam: North-Holland.

Hasbrouck, J. (1999). The dynamics of discrete bid and ask quotes. *Journal of Finance 54*, 2109–2142.

Hendry, D. F. (1995a). *Dynamic Econometrics*. Oxford: Oxford University Press.

Hendry, D. F. (1995b). A theory of co-breaking. Unpublished paper: Nuffield College, Oxford.

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

Hillman, R. and M. Salmon (1999). From market micro-structure to macro fundamentals: is there predictability in the Dollar Deutsche Mark exchange rate. City University Business School working paper.

Hobson, E. W. (1927). *The Theory of Functions of a Real Variable and the Theory of Fourier's Series* (3 ed.). Cambridge: Cambridge University Press.

Hougaard, P. (1986). Survival models for heterogeneous populations derived from stable distributions. *Biometrika 73*, 387–396.

Howison, S. D., A. Rafailidis, and H. O. Rasmussen (2000). A note on the pricing and hedging of volatility derivatives. Unpublished paper: Mathematical Institute, University of Oxford.

Hsieh, D. (1991). Chaos and nonlinear dynamics: Application to financial markets. *Journal of Finance 46*, 1839–1877.

Hubalek, F. and E. Nicolato (2001). On multivariate extensions of Lévy driven Ornstein-Uhlenbeck type stochastic volatility models and multi-asset options. Working Paper, University of Aarhus.

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

Ito, K. (1969). *Stochastic Processes*. Aarhus: Aarhus University. Lecture Notes Series 16. To be reprinted by Springer 2002.

Jacod, J. (1992). Limit of random measures associated with the increments of a brownian semimartingale. Publ. Lab. Probabilite N. 120, Paris 6.

Jacod, J. and A. N. Shiryaev (1987). *Limit Theorems for Stochastic Processes*. Springer-Verlag: Berlin.

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

Jørgensen, B. (1982). *Statistical Properties of the Generalised Inverse Gaussian Distribution*. New York: Springer-Verlag.

Jørgensen, B. (1987). Exponential dispersion models (with discussion). *Journal of the Royal Statistical Society, Series B 49*, 127–162.

Jorion, P. (1995). Predicting volatility in the foreign exchange market. *Journal of Finance 50*, 507–528.

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

Karatzas, I. and S. E. Shreve (1991). *Brownian Motion and Stochastic Calculus* (2 ed.), Volume 113 of *Graduate Texts in Mathematics*. Berlin Heidelberg New York: Springer–Verlag.

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

Kloeden, P. E. and E. Platen (1992). *Numerical Solutions to Stochastic Differential Equations*. New York: Springer.

Koponen, I. (1995). Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physics Review E 52*, 1197–1199.

Lebesgue, H. (1902). Integrale, longuer, aire. *Annali di Mathematica pura ed applicata 7*, 231–359.

Lévy, P. (1937). *Théories de L'Addition Aléatories*. Paris: Gauthier-Villars.

Lewis, M. A. (2000). Analytical expressions for the moments of the integrated volatility in affine stochastic volatility models. Working paper, CIRANO, Montreal.

Lyons, T. (1994). Differential equations driven by rough signals. I. An extension of an inequality by L.C.Young. *Mathematical Research Letters 1*, 451–464.

Madan, D. B., P. Carr, and E. Chang (1998). The variance gamma process and option pricing. *European Finance Review 2*, 79–105.

Madan, D. B. and E. Seneta (1990). The VG model for share market returns. *Journal of Business 63*, 511–524.

Maheswaran, S. and C. A. Sims (1993). Empirical implications of arbitrage-free asset markets. In P. C. B. Phillips (Ed.), *Models, Methods and Applications of Econometrics*. Basil Blackwell.

Maheu, J. M. and T. H. McCurdy (2001). Nonlinear features of realised FX volatility. *Review of Economics and Statistics 83*. Forthcoming.

Mandelbrot, B. (1963). The variation of certain speculative prices. *Journal of Business 36*, 394–419.

Mandelbrot, B. and H. Taylor (1967). On the distribution of stock price differences. *Operations Research 15*, 1057–1062.

Mantegna, R. and H. E. Stanley (2000). *Introduction to Econophysics: Correlation and Complexity in Finance*. Cambridge: Cambridge University Press.

Mantegna, R. N. and H. E. Stanley (1994). Stochastic process with ultraslow convergence to a Gaussian: the truncated lévy flight. *Physics Reviews Letters 73*, 2946–2949.

Mantegna, R. N. and H. E. Stanley (1996). Turbulence and financial markets. *Nature 383*, 587–588.

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

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatilities. *Journal of Applied Econometrics*. Forthcoming.

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

Meddahi, N. and E. Renault (2002). Temporal aggregation of volatility models. *Journal of Econometrics*. Forthcoming.

Melino, A. and S. M. Turnbull (1990). Pricing foreign currency options with stochastic volatility. *Journal of Econometrics 45*, 239–265.

Merton, R. (1973). Rational theory of option pricing. *Bell Journal of Economics and Management Science 4*, 141–83.

Mikosch, T. and R. Norvaisa (2000). Stochastic integral equations without probability. *Bernoulli 6*, 401–434.

Mikosch, T. and C. Starica (2000). Limit theory for the sample autocorrelations and extremes of a GARCH(1,1) process. *Annals of Statistics 28*, 1427–1451.

Munroe, M. E. (1953). *Introduction to Measure and Integration*. Cambridge, MA: Addison-Wesley Publishing Company, Inc.

Nicolato, E. and E. Venardos (2001). Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type. *Mathematical Finance*. Forthcoming.

Novikov, E. A. (1994). Infinitely divisible distributions in turbulence. *Physics Reviews E 50*, R3303–R3305.

O'Hara, M. (1995). *Market Microstructure Theory*. Oxford: Blackwell Publishers.

Øksendal, B. (1995). *Stochastic Differential Equations. An Introduction with Financial Applications* (5 ed.). Berlin: Springer.

Pagan, A. R. and A. Ullah (1999). *Nonparametric Econometrics*. Cambridge: Cambridge University Press.

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

Poterba, J. and L. Summers (1986). The persistence of volatility and stock market fluctuations. *American Economic Review 76*, 1124–1141.

Praetz, P. D. (1972). The distribution of share prices. *Journal of Business 45*, 45–55.

Prause, K. (1998). The generalised hyperbolic models: estimation, financial derivatives and risk measurement. PhD Thesis. Mathematics Faculty, Freiburg University.

Protassov, R. S. (2001). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ. Department of Statistics, Harvard University.

Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*. New York: Springer-Verlag.

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

Rachev, S. and S. Mittnik (2000). *Stable Paretian Models in Finance*. New York: Wiley.

Raible, S. (1998). Lévy processes in finance: Theory, numerics and empirical facts. PhD Thesis. Mathematics Faculty, Freiburg University.

Renault, E. (1997). Econometric models of option pricing errors. In D. M. Kreps and K. F. Wallis (Eds.), *Advances in Economics and Econometrics: Theory and Applications*, pp. 223–78. Cambridge: Cambridge University Press.

Revuz, D. and M. Yor (1999). *Continuous Martingales and Brownian motion* (3 ed.). Heidelberg: Springer-Verlag.

Ripley, B. D. (1987). *Stochastic Simulation*. New York: Wiley.

Robinson, P. M. (1994). Time series with strong dependence. In C. A. Sims (Ed.), *Advances in Econometrics, Sixth World Congress*, Volume 1, pp. 47–95.

Rocha-Arteaga, A. and K. Sato (2001). Topics in infinitely divisible distributions and Lévy processes. Research Report. CIMAT, Guanajuato, Mexico.

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

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

Rosinski, J. (2001). Contribution to the discussion of a paper by Barndorff-Nielsen and Shephard. *Journal of the Royal Statistical Society, Series B 63*, 230–231.

Rydberg, T. H. (1997a). Existence of unique equivalent martingale measures in a Markovian setting. *Finance and Stochastics 1*, 251–257.

Rydberg, T. H. (1997b). The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in Statistics: Stochastic Models 13*, 887–910.

Rydberg, T. H. and N. Shephard (2000). A modelling framework for the prices and times of trades made on the NYSE. In W. J. Fitzgerald, R. L. Smith, A. T. Walden, and P. C. Young (Eds.), *Nonlinear and Nonstationary Signal Processing*, pp. 217–246. Cambridge: Isaac Newton Institute and Cambridge University Press.

Sandmann, G. and S. J. Koopman (1998). Estimation of stochastic volatility models via Monte Carlo maximum likelihood. *Journal of Econometrics 87*, 271–301.

Sargen, J. D. (1964). Wages and prices in the Uunited Kingdom: a study in econometric methodology (with discussion). In P. E. Hart, G. Mills, and J. K. Whitaker (Eds.), *Econometric Analysis of National Economic Planning, Volume 16 of the Colston Papers*, pp. 23–63. London: Butterworth and Company.

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

Sato, K. (2000). *Density Transformations for Lévy Processes*. MaPhySto Lecture Notes 7. University of Aarhus.

Schoutens, W. and J. L. Teugels (1998). Levy processes, polynomials and martingales. *Statistics and Stochastic Models 14*, 335–349.

Schoutens, W. and J. L. Teugels (2001). Meixner processes in finance. Unpublished paper: Eurandom, K.U. Leuven.

Schrödinger, E. (1915). Zur theorie der fall - und steigversuche an teilschen mit brownscher bewegung. *Physikalische Zeitschrift 16*, 289–295.

Schwert, G. W. (1989). Why does stock market volatility change over time? *Journal of Finance 44*, 1115–1153.

Scott, L. (1987). Options pricing when the variance changes randomly: theory, estimation and an application. *J. Financial and Quantitative Analysis 22*, 419–438.

Scott, L. (1991). Random-variance option pricing. *Advances in Future and Options Research 5*, 113–135.

Scott, L. (1997). Pricing stock options in a jump-diffusion model with stochastic volatility and interest rates: Applications of Fourier inversion methods. *Mathematical Finance 7*, 413–26.

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

Shiryaev, A. N. (1999). *Essentials of Stochastic Finance: Facts, Models and Theory*. Singapore: World Scientific.

Sichel, H. S. (1973). Statistical valuation of diamondiferous deposits. *Journal of South African Institute of Mining and Metallurgy 73*, 235–243.

Stein, E. M. and J. Stein (1991). Stock price distributions with stochastic volatility: an analytic approach. *Review of Financial Studies 4*, 727–752.

Stock, J. H. (1988). Estimating continuous-time processes subject to time deformation: an application to postwar U.S. GNP. *Journal of the American Statistical Association 83*, 77–85.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for Exploration of Posterior Distributions and Likelihood Functions* (3 ed.). New York: Springer-Verlag.

Tauchen, G. and M. Pitts (1983). The price variability-volume relationship on speculative markets. *Econometrica 51*, 485–505.

Taylor, S. J. (1982). Financial returns modelled by the product of two stochastic processes — a study of daily sugar prices 1961-79. In O. D. Anderson (Ed.), *Time Series Analysis: Theory and Practice, 1*, pp. 203–226. Amsterdam: North-Holland.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley.

Taylor, S. J. (1994). Modelling stochastic volatility. *Mathematical Finance 4*, 183–204.

Taylor, S. J. and X. Xu (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance 4*, 317–340.

Thorin, O. (1977). On the infinite divisibility of the lognormal distribution. *Scandinavian Actuarial Journal 47*, 121–148.

Tompkins, R. and F. Hubalek (2000). On closed form solutions for pricing options with jumping volatility. Unpublished paper: Technical University, Vienna.

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

Uhlenbeck, G. E. and L. S. Ornstein (1930). On the theory of Brownian motion. *Physical Review 36*, 823–841.

Vervaat, W. (1979). On a stochastic differential equation and a representation of nonnegative infinitely divisible random variables. *Advances in Applied Probability 11*, 750–783.

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

West, K. D. and D. Cho (1995). The predictive ability of several models for exchange rate volatility. *Journal of Econometrics 69*, 367–391.

White, H. (1982). Maximum likelihood estimation of misspecified models. *Econometrica 50*, 1–25.

White, H. (1994). *Estimation, Inference and Specification Analysis*. Cambridge: Cambridge University Press.

Widder, D. V. (1946). *The Laplace Transform*. Princeton: Princeton University Press.

Wiggins, J. B. (1987). Option values under stochastic volatilities. *Journal of Financial Economics 19*, 351–372.

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

Wolpert, R. L. and K. Ickstadt (1999). Simulation of Lévy random fields. In D. D. Dey, M. Muller, and D. Sinha (Eds.), *Practical Nonparametric and Semiparametric Bayesian Statistics*, pp. 227–242. New York: Springer-Verlag.

Ysusi, C. (2001). M.Sc. thesis. Unpublished, Department of Statistics, University of Oxford.

Index