Price feedback and hybrid diffusions in finance

A thesis submitted for the degree of

Doctor of Philosophy

by

Marcus Schofield

March 3, 2015

Department of Mathematics

University College London

Declaration

I, Marcus Schofield, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

It is well-known that the probabilistic behaviour of financial asset returns is not captured well by the classical Black-Scholes model. The true behaviour will never be perfectly captured in any model, but insight is continually being obtained into our understanding of more sophisticated and realistic models. Much research has been published recently exploring the use of Lévy process models, which maintain the original *independent stationary increments* assumption present in the Black-Scholes model, but incorporate jumps in the modelling. This investigation seeks to motivate a new class of models, throwing out the stationary increments hypothesis. We argue that certain techniques of trading decision-making are not independent of previous price movements, and the returns, being driven by the trade order flow, will reflect that. From here, we develop two particular such models, which are both diffusion models, and study them for their probabilistic behaviour. The first of these models is a hybrid of the arithmetic and geometric Brownian motions, which has transition probabilities expressible in terms of a spectral expansion involving Legendre functions. The second is a hybrid of the arithmetic Brownian motion and the Cox-Ingersoll-Ross process, and its spectral expansions involve the confluent hypergeometric functions. Having developed these expressions in sufficient detail to do so, we consider the calculation of value-at-risk and expected shortfall in these two models.

Acknowledgements

I would first like to thank my PhD supervisor, Professor William Shaw. I am grateful to him for inspiring this research, for teaching me and passing on his enthusiasm for the subject, for guiding my research and writing over these four years and for many exciting and lively discussions throughout. I am deeply grateful to the EPSRC for financial support under my DTA grant. This was initially awarded by the Department of Mathematics at Kings College London, and I am grateful to Prof Jurgen Berndt for facilitating the transfer to University College London.

I would like to thank the others from whom I have received mathematical guidance, Professor Valery Smyshlyaev, Dr Cristin Buescu and Dr Reimer Kühn, as well as Suraj Gosai for his insight and motivation on computational matters. I am also extremely grateful to other members of the mathematics department at UCL, including Helen Higgins, Soheni Francis and Professor Robb McDonald, who have helped me make my time here rewarding and enjoyable.

Finally, I would like to express my gratitude to my family. Thank you to Diana, Nick and June for your continued moral and emotional support. I wish to devote a particularly special thank you to Karina, whose love and patience has been momentous in getting me through this work. Most importantly, I wish to thank my father, Sandy, and late mother, Sarah, for their devotion that has provided this opportunity for me, and for the extra-ordinary sacrifice of time and effort that my father has made throughout this time in order to give me the best help possible in preparing this work.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Gaussian hypothesis for financial asset returns

The purpose of this investigation is to motivate and study a class of stochastic processes as dynamic models for a single financial asset price. The processes are in continuous time and we assume that they possess the Markov property so as to obtain analytic tractability. Our aim is really to begin from fairly simple, natural modelling assumptions and find that they lead to a wider, and more statistically plausible, class of return probability distributions than the most basic models do. This is not only of theoretical interest, but also of practical importance if it can be shown that, actually, fat-tailed return distributions should arise just as naturally as does the Gaussian.

The first continuous-time stochastic model in finance, developed by Louis Bachelier in 1900 in his PhD thesis *Théorie de la spéculation*, used a scaled Brownian motion with drift to model a security price directly. It was later realised[1] that such a process might more appropriately describe the price's logarithm, or the returns process. In 1973, Black & Scholes [18] wrote down, in closed form, the price of European put and call options on this asset under this latter model.

In modern notation, we write the stochastic dynamics of the underlying asset price S_t, under the Black-Scholes model, as $dS_t = \mu S_t dt + \sigma S_t dW_t$, where $\mu \in \mathbf{R}$ and $\sigma > 0$ are constants (termed respectively *drift* and *volatility*) and $(W_t)_{t\geq 0}$ is a standard one-dimensional Wiener process. Using Itô's lemma, one finds that the model may equally well have been specified in terms of the returns process $X_t = \log(S_t) - \log(S_0)$, whose dynamics are governed by

$$dX_t = \mu_1 dt + \sigma dW_t, \qquad t > 0, \quad X_0 = x, \tag{1.1}$$

[1] This was an observation made independently around the same time (in the 1950s) by M. F. M. Osborne, P. A. Samuelson and S. S. Alexander. See the introduction in part 1 of Cootner [32].

in which the drift coefficient is the constant $\mu_1 = \mu - \sigma^2/2$. According to this model then, the returns on an investment made for a period of length t have the $N(\mu_1 t, \sigma^2 t)$ distribution. In this study, we shall refer to any of the distributions $N(\mu, \sigma^2)$ as *normal* or *Gaussian*. We furthermore refer to any model using (1.1) for the underlying returns process as the Black-Scholes model, or the arithmetic Brownian motion (aBm) model, since (1.1) is the SDE for an aBm.

Unfortunately, empirical investigations strongly reject the hypothesis of Gaussian distributed returns. Whole books have been written about non-Gaussian financial modelling, see for instance Boyarchenko & Levendorskiĭ [23], Campbell et al. [25], Embrechts et al. [41] and Jondeau et al. [71]. Non-normality of financial asset returns had been observed long before Black & Scholes' fundamental paper, notably already in 1963 by Fama [47] and Mandelbrot [86], but it had in fact been observed and discussed before that. Mandelbrot [86] refers to an article[2] as far back as 1915.

For an easy explanation of what is wrong with the Gaussian hypothesis, we shall consider a test based on moments. If X denotes a real-valued random variable, for instance the return on some investment over a known time-period, with distribution F, we define its mean and variance

$$m = EX = \int_{-\infty}^{\infty} x\, F(dx),$$

$$\sigma^2 = \mathrm{var}\,(X) = \int_{-\infty}^{\infty} (x-m)^2 F(dx) = E[X^2] - (EX)^2.$$

We then define its skewness and kurtosis as

$$\mathrm{sk}(X) = E\left[\left(\frac{X-m}{\sigma}\right)^3\right] \quad \text{and} \quad \mathrm{ku}(X) = E\left[\left(\frac{X-m}{\sigma}\right)^4\right]$$

respectively. Skewness is a measure of the asymmetry of a distribution. Kurtosis is a measure of the spread of the distribution, and it tells you how heavy the tails are, without distinguishing the left and right ones. If X belongs to the normal family of distributions, then it has zero skewness and a kurtosis of 3. For this reason, one often defines the excess kurtosis of X as

$$\text{excess kurtosis}(X) = \mathrm{ku}(X) - 3.$$

Empirical investigations tend to show that returns have slightly negative skewness and large positive excess kurtosis. Indeed let us now conduct a test

[2] 'The Making and Using of Index Numbers', by Wesley C. Mitchell

for normality on some share price indices, based on skewness and excess kurtosis. For simplicity, we shall suppose that the data are modelled as i.i.d. random variables, with their observed mean and variance, and test the assumption that they are from a normal sample. We consider daily log-returns of seven world indices, each sampled between 2nd January 1996 and 30th December 2005 (31st was a Saturday). These can be sampled using the FinancialData[] function built in to Mathematica 8, or using the **get.hist.quote**() function in R [95] in the tseries package, see Daníelsson [34]. Skewness and kurtosis are estimated from i.i.d. data $(r_n)_{n=1}^N$ by

$$\widehat{\mathrm{sk}} = N^{-1}\sum_{n=1}^{N}\left(\frac{r_n-\overline{r}}{\hat{\sigma}}\right)^3, \quad \widehat{\mathrm{ku}} = N^{-1}\sum_{n=1}^{N}\left(\frac{r_n-\overline{r}}{\hat{\sigma}}\right)^4,$$

where mean and variance were estimated as usual:

$$\overline{r} = N^{-1}\sum_{n=1}^{N} r_n, \quad \hat{\sigma}^2 = \frac{1}{N-1}\sum_{n=1}^{N}(r_n-r)^2.$$

We use a Jarque-Bera (JB) test[3], which uses the fact that as $N\to\infty$,

$$\frac{\widehat{\mathrm{sk}}}{\sqrt{6/N}} \xrightarrow{d} N(0,1) \quad \text{and} \quad \frac{\widehat{\mathrm{ku}}-3}{\sqrt{24/N}} \xrightarrow{d} N(0,1).$$

The Jarque-Bera test statistic JB is defined as and satisfies

$$JB = N \times \left[\frac{\widehat{\mathrm{sk}}^2}{6} + \frac{(\widehat{\mathrm{ku}}-3)^2}{24}\right] \xrightarrow{d} \chi^2_2$$

where the distribution on the right is the chi-square distribution with 2 degrees of freedom. JB cannot be negative; large values of JB are likely to reject the Gaussian hypothesis. The χ^2_2-distribution attributes[4] 99% of its mass to the interval $(0, 9.210)$, and 99.9% of its mass to the interval $(0, 13.816)$. So values of JB greater than 13.816 reject the Gaussian hypothesis even at the 99.9% confidence level.

The results are summarised in table 1.1. We can see from them that excess kurtosis is very positive, and the JB test statistic gives an idea of how non-normal the data are (normality is rejected at the 99% level when $JB > 10$). JB allows us, for every one of the indices in table 1.1, to reject the Gaussian

[3]This method is outlined in Jondeau et al. [71], section 2.2.3.
[4]Rohlf & Sokal [98]

Index	mean (%, 3 s.f.)	std. dev. (%, 3 s.f.)	skewness	excess kurtosis	JB
FTSE 100	0.0167	1.29	-0.140084	2.52965	681.499
S & P 500	0.0277	1.33	-0.0907735	2.95578	920.075
DAX	0.0341	2.54	-0.146489	2.35538	593.41
Cac 40	0.0357	2.04	-0.113291	2.61001	724.672
Hang Seng Index	0.0153	2.93	0.123453	11.0659	12608.9
Nikkei 225	-0.0100	2.09	-0.0271506	1.94153	386.208
Ibovespa Sao Paulo	0.0821	5.38	0.441001	14.5207	21824.3

Table 1.1: JB test for normal distribution of seven financial time series sampled between 2/01/1996 and 30/12/2005

hypothesis based on both the skewness and excess kurtosis of the data, but in fact for all these indices the Gaussian hypothesis is also rejected by a normal test on only the excess kurtosis.

The tails of the Gaussian model can be fattened to some extent by allowing for non-constant volatility. For example, we might consider a simple Bernoulli mixture distribution $F = \theta N(\mu, \sigma_1^2) + (1-\theta)N(\mu, \sigma_2^2)$, where $\theta \in (0,1)$ is called the mixing probability. This distribution has four parameters. In general, taking mixtures of k normal distributions with the same mean μ but different volatilities gives us a distribution with $2k$ parameters, because the use of another volatility also introduces an extra mixing parameter. In Chapter 7, once we have developed our models, we shall consider a likelihood test on using data from these same world indices and compare the fit of our models with the classical Gaussian and a Gaussian with a random volatility.

1.2 A review of the Gaussian Black-Scholes theory

Since the theory is elementary, we shall now summarise the Black-Scholes model. When we later generalize its underlying assumptions (in particular, the underlying equation of motion (1.1)), this should serve as a template for investigation of analogous properties of our new model. First, consider measures of risk associated with the aBm model. The $100\alpha\%$ value-at-risk (VaR) is defined as the minimum relative loss made on an investment in the worst $100(1-\alpha)\%$ of cases. Assuming that under our model, X_t has an invertible distribution function, this definition states that if $\mathrm{VaR}_\alpha(X_t) = \xi_\alpha$, then $\alpha = P\{X_t > -\xi_\alpha\}$. Expected shortfall at the $100\alpha\%$ confidence level is then defined as the conditional expected loss, given that the loss exceeds VaR_α. Defining the functions

$$u_k(x) = |x|^k 1_{(-\infty, -\xi_\alpha)}(x),$$

it is easily seen that we can use the formulae

$$1-\alpha = E[u_0(X_t)], \qquad \mathrm{ES}_\alpha(X_t) = \frac{1}{1-\alpha} E[u_1(X_t)] \tag{1.2}$$

to define or compute VaR and ES. Some authors use higher values of k for more general risk measures, extracting information deeper in the tails[5]. These risk measures are a function of the probability level and the random variable or, more properly, its distribution. For a specific financial asset returns process $X = (X_t)$, they are functions of α and t (Hull [63]).

For the Gaussian distribution $N(\mu_1 t, \sigma^2 t)$, distribution function and its inverse, the quantile function, are known and can be easily implemented in certain software packages[6]. For example, under these assumptions the $100\alpha\%$ value-at-risk is

$$\mathrm{VaR}_\alpha(X_t) = \sigma\sqrt{t}\,\Phi^{-1}(\alpha) - \mu_1 t,$$

where Φ^{-1} denotes the normal quantile function. The expected shortfall is

$$\mathrm{ES}_\alpha(X_t) = \sigma\sqrt{t}\,\frac{\varphi(\Phi^{-1}(1-\alpha))}{1-\alpha} - \mu_1 t.$$

where we have denoted the standard normal pdf

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Secondly, since the transition function of the Markov process $(X_t)_{t\geq 0}$ is also known explicitly, maximum likelihood estimation of the parameters μ and σ is easy to perform. The transition density is

$$p(t;x,y) = p(t;x,y,\theta) = (2\pi\sigma^2 t)^{-1/2} \exp\left(-\frac{(y-(x+\mu_1 t))^2}{2\sigma^2 t}\right).$$

where $\theta = (\mu, \sigma^2)$ is the vector containing the parameters. Suppose that we are given a time series $(X_{n\Delta})_{n=0}^N$, in which the time-lag $\Delta > 0$ is known. Under the aBm assumption, the likelihood function is then

$$\begin{aligned} L(\theta;X) &= \prod_{n=1}^{N} p(\Delta; X_{(n-1)\Delta}, X_{n\Delta}, \theta) \\ &= (2\pi\sigma^2\Delta)^{-N/2} \exp\left(-\frac{1}{2\sigma^2\Delta}\sum_{n=1}^{N}(X_{n\Delta} - X_{(n-1)\Delta} - \mu_1\Delta)^2\right), \end{aligned}$$

[5] See Jondeau et al. [71], Section 9.2, the definition is from Berkelaar & Kouwenberg [12]. For example, $k = 2$ can give us the 'downside variance'.

[6] R, MATLAB, *Mathematica* and *Maple* all have built-in functions for computing these.

The usual choice for estimating the parameter θ in situations where this function is known explicitly is to choose it so that the function L is maximized. It is equivalent to maximize the log-likelihood

$$\ell(\theta; X) = \log L(\theta; X) = -\frac{N}{2}\log(2\pi\sigma^2\Delta) - \frac{1}{2\sigma^2\Delta}\sum_{n=1}^{N}(X_{n\Delta} - X_{(n-1)\Delta} - \mu_1\Delta)^2.$$

This is performed by differentiating with respect to each of the parameters in turn, so defining the score function

$$S_N(\theta) = \partial_\theta \ell(\theta; X),$$

then setting this equal to 0 and solving for θ. In the aBm model this gives explicit formulae

$$\widehat{\mu_1} = \frac{1}{N\Delta}\sum_{n=1}^{N}(X_{n\Delta} - X_{(n-1)\Delta}), \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{N\Delta}\sum_{n=1}^{N}(X_{n\Delta} - X_{(n-1)\Delta} - \hat{\mu}_1\Delta)^2.$$

Finally, it may be shown that explicit formulae may be found for European call and put options on an asset with returns process (1.1). In the full Black-Scholes set-up of their paper [18], one considers a market in which two assets are traded, each with perfect liquidity: a risky asset whose returns process is given by (1.1) and a riskless asset (β_t) which accumulates at a constant rate r; i.e. $\beta_t = \beta_0 e^{rt}$. This market is free from arbitrage and is complete[7], so that there is a unique pricing measure Q under which contingent claims are priced. For a European option, with payoff function g and time t remaining until expiry, it is

$$u(t, X_0) = E[e^{-rt} g(X_t) \mid X_0] \tag{1.3}$$

where the expectation is taken under the pricing measure Q. This formula is actually a special case of that given on page 6 of Dynkin [38] volume I. Write $S_0 = e^{X_0}$. Under the pricing measure Q, the returns process X has stochastic dynamics

$$dX_t = (r - \frac{1}{2}\sigma^2)dt + \sigma dW_t^Q$$

where W^Q is a Q-Brownian motion, so $X_t - X_0 \sim N((r - \sigma^2/2)t, \sigma^2 t)$ under Q. Put and call options have respectively payoffs $g_{\mathrm{P}}(x) = [K - e^x]^+$ and $g_{\mathrm{C}}(x) = [e^x - K]^+$ and, for these functions, (1.3) may be written down in

[7]For definitions of arbitrage and market completeness, see any standard introductory text on mathematical finance. Good introductions include Björk [15], Etheridge [45] and Shreve [103]; I recommend Klebaner [77] Chapter 11. Absence of arbitrage corresponds to the existence of an equivalent martingale measure Q, while completeness is equivalent to its uniqueness.

closed form. One easily finds by direct calculation that

$$E[e^{-rt}g_{\mathrm{C}}(X_t) \mid X_0] = S_0\Phi(d_1) - e^{-rt}K\Phi(d_2)$$

and

$$E[e^{-rt}g_{\mathrm{P}}(X_t) \mid X_0] = e^{-rt}K\Phi(-d_2) - S_0\Phi(-d_1).$$

where we have used the usual notational abbreviations for these cases

$$d_1 = \frac{X_0 - \log(K) + (r + \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}}, \quad d_2 = \frac{X_0 - \log(K) + (r - \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}}$$

and Φ is the standard normal distribution function.

We note that formula (1.3) is the Feynman-Kač formula[8], telling us the transition operator of the process X killed at the constant rate r. Thus the option pricing formula, like the formula defining risk measures or the likelihood function for parametric estimation, may be found directly from the transition function of the aBm.

1.3 Fundamental and technical analysis

Before discussing which of the properties of the aBm model it is desirable for us to retain in our search for more realistic financial models, we shall briefly initiate our attempts to provide some explanation for this observed deviation of the returns distribution from normality. Consider once again the model in question, namely (1.1), thinking this time about where it comes from. One might envisage the security S being traded on an extremely liquid market, in which are executed so many transactions, each one at a slightly different price from the previous one, that a plot of the returns process against time looks *from sufficiently far out* like a continuous diffusion process. From this point of view, the market participants, in making these trades, are driving the asset price up and down continuously.

It makes sense that the buyer-initiated trades result in upward price movement, while seller-initiated trades result in downward price movement. The net effect on the market state (i.e. X_t) of trading over small time periods is probably partly determined by the imbalance of inflow of buy and sell trade orders[9]. The equation of motion (1.1) says that, while we do not know the change in price over a small time-period, we do know, or at least we can calculate, its mean and variance. The model then supposes further that the price

[8] Çinlar [30] Chapter 9, formula 2.49, and Rogers & Williams [97] III(18.13).

[9] We discuss this further in section **??**, where we define limit orders and market orders

is driven in the same way over a succession of these trading periods.

If it is these market participants who are driving the returns process X, then in order to discover why it is that the returns process moves about in a certain way, we shall probably wish to study these traders' trading motivations. One dichotomy of trading motivations is particularly popular in the literature, into trading based on fundamental analysis and trading based on technical analysis. Traders employing fundamental analysis are termed *fundamentalists*, those employing technical analysis, *technicians*. Both fundamentalist and technician are seeking to gain knowledge about the general direction (upward or downward) of an asset's price and trade based upon that knowledge, but their methods of obtaining this information differ.

The fundamentalists look for external factors that lie behind price changes. For example, when trading the stock of a particular company, the fundamentalist will look at the general health of that company and its profit prospects. On the other hand, the technicians look at the history of all sorts of market statistic: a chart of X_t against t over a previous time period will be useful to a technician, and he or she also trades based on information such as the history of trading volumes. Since, quite often, the work of a market technician involves looking at and analysing charts, technicians are often called *chartists*. Key textbooks aimed at the market technician are Kirkpatrick & Dalquist [76] and Murphy [90]. In our rôle of market modelling, we need not comment upon the relative merits of either style of market participation, which have been discussed elsewhere; we need only accept the presence of both in the market. We should also note that the actual traders in the market are probably employing analysis of both types in some form, but this does not affect our modelling: we simply need to acknowledge that both types of motivation are present in the orders that arrive at the market.

1.4 Non-Gaussian returns and the Markovian assumption

We know from the above discussions that the assumption of Gaussian-distributed returns fails to capture observed financial market behaviour. The observed negative skewness in the data means that market returns are asymmetric in the sense that market crashes occur more frequently than booms. Excess kurtosis, on the other hand, is observed because extreme events occur much more frequently than a Gaussian model is capable of predicting. The true distribution of financial asset returns should have much heavier tails than the normal

distribution. This will be our main concern in developing models to replace the aBm model developed above, though we would also like them to be capable of exhibiting asymmetry when the need should arise.

Despite the observed drawbacks of the aBm model, it has one notable advantage over its competitors: tractability. We have already noted that risk measures, parameter estimation and derivatives pricing are all elementary in this model. The reason for this is that they have closed-form expressions based on the transition function of the underlying Markov process, which is in turn known in terms of its density with respect to the Lebesgue measure.

In stating that the process X describing the market state possesses a transition function of the form $P_t(x, A) = P^x\{X_t \in A\}$, we are assuming that it is Markovian. The equality $P^x\{X_t \in A\} = P[X_{s+t} \in A \mid X_s = x]$ then further assumes that it is time-homogeneous. Time-homogeneity can probably be justified over short to medium trading periods, but the Markov assumption is fairly specific and restrictive. In fact, the Markov assumption essentially says that all the relevant information about the asset's future price fluctuations contained in the asset price's history is encoded in its current price. This implies that any attempts, by chartism for instance, to determine the price's future movements are futile.

The aBm model (1.1) actually contains all of the following assumptions:

(i) X is a Markov process, possessing a transition function;

(ii) X is temporally homogeneous;

(iii) X has continuous paths, i.e. the function $t \mapsto X_t$ is continuous;

(iv) X is spatially homogeneous.

Conversely, if X is assumed to satisfy these conditions (i)-(iv), then it is possible to show that X satisfies equation (1.1) for suitably chosen constants μ_1 and σ. If we are to discover more general models, one or more of these conditions must be relaxed, or dropped completely.

We now discuss these assumptions in a little more detail. Starting from the top, a Markov process X on a time-parameter set T with topological state space S may (and is usually assumed to) possess a transition function $P_{s,t}(x, dy)$, intuitively[10] given by

$$P_{s,t}(x, B) = P[X_t \in B \mid X_s = x], \quad s \leq t \text{ in } T,\ x \in S,\ B \in \mathcal{S},$$

[10]To define rigorously, one must take care to remember that conditional probabilities are defined uniquely only up to a set of measure 0, see Blumenthal & Getoor [19].

where $\mathcal{S}$ denotes the class of Borel subsets of S. The assumption that X is Markovian does not imply that a transition function should exist: this property must be assumed. Transition functions, however, make sense only for processes possessing at least the simple Markov property that, at each fixed time t (called the 'present'), past and future are conditionally independent given the present state X_t. By conditioning on intermediate times $u \in [s,t]$, one sees that the transition function $P_{s,t}(x,dy)$ must satisfy the so-called Chapman-Kolmogorov relation

$$P_{s,t}(x,B) = \int_S P_{s,u}(x,dy)P_{u,t}(y,B), \qquad \text{for any } u \in [s,t] \subset T,\ x \in S,\ B \in \mathcal{S}.$$

Roughly speaking, we say that a Markov process is *strong Markov* if the (simple) Markov property is satisfied when the 'present' fixed time t is replaced with *stopping* times τ. See Blumenthal & Getoor [19] for the precise definitions.

It is worth noting at this point what is possible if all four conditions, including the Markovian assumption, are dropped. That is, we consider a market, still described by a one-dimensional stochastic process (X_t), which is now not even Markovian. It is usually assumed in this situation that X is a semimartingle. That is, X is a rcll[11] process, decomposable as

$$X_t = X_0 + M_t + A_t, \quad t \geq 0$$

where M is a local martingale and A is a process of finite variation, both starting at zero. These processes are discussed in section 2.4. The problems of risk measurement and option valuation then require the computation of expectations, without using transition semigroups. Specifically, the value at time t of a European option with payoff function u is given by the expectation

$$V(t,x) = E^{(t,x)}[e^{-rt}u(X_T)]$$

under some measure $P^{(t,x)}$ (under which $P^{(t,x)}\{X_t = x\} = 1$). Similarly, the α-level value-at-risk may be defined as the unique value of $\xi > 0$ such that

$$P^{(t,X_t)}\{X_T < -\xi\} = E^{(t,X_t)}[1\{X_T < -\xi\}] = 1 - \alpha.$$

If it is possible to simulate the process X, then such expectations may at least be computed by Monte Carlo methods. For the most part, since the analytic tools on which we later rely are not available in the general non-Markov set-up,

[11] Right-continuous with left limits, also often described as càdlàg or corlol.

the solution of these problems is necessarily computationally intensive.

It is noted in Çinlar [30], Chapter 9 remark 1.12, that, if tractability is permitted to be completely disregarded, any stochastic process may be made Markovian by a sufficient expansion of its state space[12]. Indeed, if a stochastic process X is given on a probability space, then one may construct a new process Y on the same space, such that the state Y_t describes the whole history of X up till time t. By this construction then, Y is Markovian, but there is nothing further we can do with it since the present state Y_t is in general too complicated for us to condition on.

There do, however, exist situations in which non-Markovian processes may be made Markovian without the loss of too much tractability. We should like to call *Markov reducible* any process X with the property that, with the introduction of an $\mathbf{R}^d$-valued process Y, the stochastic process (X, Y) is then Markovian with respect to its natural filtration. So long as d is a sufficiently small number, the system should remain tractable. An example is furnished by the evolution of the value A_t of an Asian option, see for instance Wilmott et al. [111]. Let X again represent the return on the underlying asset, and introduce the price averages

$$\begin{aligned} I_1(t) &= \frac{1}{t}\int_0^t e^{X_u} du \quad \text{(arithmetic average)}, \\ I_2(t) &= \exp\left(\frac{1}{t}\int_0^t X_u du\right) \quad \text{(geometric average)}. \end{aligned}$$

The payoff of the option is then given by some expression of the form $A_T = u(X_T, I(T))$ (where $I \in \{I_1, I_2\}$) for the suitable payoff function u. The process A is not Markovian, but if X is itself Markov with respect to its natural filtration, then so is the process (A, X, I). It is then possible to study this as a 3-dimensional Markov process.

In order to preserve analytic tractability, we have decided to consider only the situation in which the one-dimensional process X satisfies the Markov property. We hope that, at some stage, the ideas sketched above may be used to conduct a more realistic investigation into the return process under pressure from fundamental and technical trading, for instance by the introduction of a finite-dimensional market statistic process, say Y, such that (X, Y) satisfies a stochastic differential equation. Y_t would represent a vector of statistics considered by technicians. In fact, Brody et al. [24] (referred to as BHM) already went even further than this, by developing a general framework in

[12] Çinlar [30] attributes this observation to Doob.

which the information available to market participants is modelled directly. In the BHM framework, investors gain their information from a separate process (ξ_t). For example, in the simplest case, the value at time $t \leq T$ of an asset entitling the holder to only a single cashflow D_T at time T, would be equal to $E[P(t,T)D_T|\xi_t]$, where $P(t,T)$ is the time t value of a unit cashflow at time T. Thus the information in the marketplace is generated by the process (ξ_t), rather than by the Brownian motion W which drives the asset price S_t in models such as that of Black & Scholes [18]. For an equity share, D_T would represent a dividend paid by the company at time T; the authors went on to consider shares paying multiple dividends and dividend growth. Hughston & Macrina developed the framework further in [61], by pricing fixed-income securities in it.

We return to the discussion of assumptions (i)-(iv) of the elementary aBm model. A Markov process X, with state space S, is said to be temporally homogeneous if $P_{s,t}(x,A)$ depends on the time variables s and t only through $t-s$, in which case we write

$$P_{s,t}(x,A) = P_{t-s}(x,A).$$

Çinlar [30] (exercise 1.40) notes that, if X is a temporally inhomogeneous Markov process in S, the process (t, X_t), with state space $\mathbf{R}_+ \times S$, is temporally homogeneous. For a time-homogeneous Markov process X with transition function (P_t), each P_t is a kernel from S to itself, but can of course also be viewed as an operator defined on the space $\mathcal{S}_b$ of bounded Borel-measurable functions $u : S \to \mathbf{R}$. From this perspective, the Chapman-Kolmogorov relation is equivalent to the semigroup property

$$P_{s+t} = P_s P_t.$$

Since the analytic theory of semigroups arising in this way is quite extensively developed, this will provide us with a good deal of information about the probabilistic behaviour of the process X.

It is assumptions (iii) and (iv) that must be relaxed to furnish more realistic yet usable return models. Before looking at any examples, let us classify the possibilities. Roughly speaking, we call a time-homogeneous continuous strong Markov process a *diffusion*, and a space- and time- homogeneous strong Markov process a *Lévy process*. Diffusion models are expressed by means of a stochastic integral equation against a driving Brownian motion (though these integral equations are then usually written in differential form). To write

down the equation of motion of a Lévy process requires an extra integral term, against a Poisson random measure, which represents the jumps of the motion. The general (i.e. space-inhomogeneous) process whose equation of motion is an integral equation involving integrals against both a Brownian motion and a standard Poisson random measure on $\mathbf{R}_+ \times \mathbf{R}_+$ is called an *Itô process*; such a process is a diffusion between its jumps. If the process has only finitely many jumps in each finite time-interval, then we call it a *jump-diffusion*.[13]

1.5 Some specific non-Gaussian Markov models

1.5.1 Lévy processes

Lévy processes have been proposed to generalize the aBm model in finance by a large number of proponents. Indeed, when Mandelbrot [86] noted the deficiencies in the Gaussian models, he was proposing the immediate replacement of the aBm with stable processes, a subclass of Lévy processes. We would like to give some examples of Lévy process models which have been proposed, but first recall some basic definitions needed to specify these models.

We have described Lévy processes as the continuous-time Markov processes which are time and space homogeneous. A regularity assumption is also required: usually one assumes that a Lévy process is *stochastically continuous*[14], see Sato [99] definition 1.5. Non-overlapping increments are independent of each other, and have a distribution depending only on the length of the time-increment. The distribution of the whole process is then determined by its starting point (often taken to be zero) and the distribution of one of the X_t, usually X_1. This distribution is infinitely divisible and there is a bijection between the class of Lévy processes which start at zero and the class of infinitely divisible distributions.

From these observations, every Lévy process X is characterized by the distribution of X_1, which is in turn characterized by a triple (μ, σ, ν), where $\mu \in \mathbf{R}$, $\sigma \geq 0$ and ν is a *Lévy measure*[15] on $\mathbf{R} \setminus \{0\}$. If, for example, $\nu = 0$, then

$$X_t = \mu t + \sigma W_t;$$

a continuous Lévy process is therefore an arithmetic Brownian motion. The

[13] This terminology agrees with that used in Çinlar [30], Chapter IX.

[14] Note that 'stochastically continuous' does not even imply that any of the paths are continuous. See Kallenberg [72] Theorem 15.1, which states that, together with the assumption of independent increments, it does imply that the process has an rcll version with no fixed (non-random) jumps, which is what we work with.

[15] i.e. a measure satisfying the integrability condition $\int (1 \wedge |x|^2)\nu(dx) < \infty$.

full characterization of the general Lévy process in one dimension is now standard, and we quote the formula from Kallenberg [72], Corollary 15.8, see also Boyarchenko & Levendorskiĭ [23] formulae (2.14) and (2.15). The characteristic function of X_t is

$$E[e^{i\xi X_t}] = e^{t\psi(\xi)}$$

where the *characteristic exponent* ψ possesses a representation, the so-called *Lévy-Khintchine formula*,

$$\psi(\xi) = i\mu\xi - \frac{1}{2}\sigma^2\xi^2 + \int_{\mathbf{R}\setminus\{0\}} (e^{i\xi x} - 1 - i\xi x 1_{[-1,1]}(x))\nu(dx). \tag{1.4}$$

The jump measure ν is called the *Lévy measure*: if it possesses a Lebesgue density $g(x) = \nu(dx)/dx$, we call this the *Lévy density*. The Lévy measure ν is concentrated on $\mathbf{R}\setminus\{0\}$, and must satisfy

$$\int_{\mathbf{R}\setminus\{0\}} (1 \wedge |x|^2)\nu(dx) < \infty.$$

If it satisfies the stronger condition

$$\int_{\mathbf{R}\setminus\{0\}} (1 \wedge |x|)\nu(dx) < \infty,$$

then the Lévy process with characteristics $(\mu, 0, \nu)$ (i.e. X with its Brownian component removed) has finite variation, and if $\nu(\mathbf{R}\setminus\{0\}) < \infty$, then X has finitely many jumps in any bounded time-interval $[t_1, t_2]$.

An increasing Lévy process A is called a *subordinator*. In this case, the distribution of A_1 is concentrated on $[0, \infty)$. Instead of using the characteristic function and characteristic exponent, such a process is characterized instead by its *Laplace exponent* ψ, defined via

$$E[e^{-\lambda A_t}] = e^{t\psi(\lambda)}.$$

In these cases, the Lévy-Khintchine representation takes on the form

$$\psi(\lambda) = \mu\lambda + \int_{(0,\infty)} (e^{-\lambda x} - 1)\nu(dx).$$

Since a subordinator is increasing, it has finite variation and the Lévy measure must satisfy

$$\int_{\mathbf{R}\setminus\{0\}} (1 \wedge x)\nu(dx) < \infty.$$

Subordinators are unsuitable for modelling financial asset returns, but they

are often used as random time-change processes. This technique, called subordination, is largely due to S. Bochner[16]. Here we note that if an arbitrary Lévy process $Y(t)$) has characteristic exponent χ and a subordinator (A_t) has Laplace exponent ψ, then the subordinated process X given by $X_t = Y(A_t)$ has characteristic exponent $\psi(\chi)$, i.e.

$$E[e^{i\xi X_t}] = e^{t\psi(\chi(\xi))}.$$

We may now introduce the class of stable processes. First consider an arbitrary Lévy process X, with $X_0 = 0$. The distribution F_1 of X_1 is infinitely divisible, and the distribution F_t of X_t is then uniquely determined by F_1. It may happen that, for every $t > 0$, F_t is a change of scale and a change of location from F_1; in this case F_1 is called a stable distribution, and X is called a stable process. (It follows that F_t is a stable distribution for every $t > 0$.) A stable process satisfies the scaling relation[17]

$$(c^{1/\alpha}X_t + \mu t)_{t\geq 0} \sim (X_{ct})_{t\geq 0}, \quad c > 0$$

for some $\alpha \in (0, 2]$. The arithmetic Brownian motion satisfies this relation with $\alpha = 2$, and it is fairly intuitive that any stable process with variance should satisfy the same, and is therefore an aBm. Thus, any stable process which is not an aBm has infinite variance. Mandelbrot [86] conducted his own empirical investigation on cotton price fluctuations using the class of stable distributions, which he termed *stable Paretian*, and found an approximate value $\alpha \simeq 1.7$. This indicates that, for the data with which he worked, infinite variance stable distributions provide a better fit than the Gaussian distribution.

However, the use of stable distributions in modelling financial asset returns is also usually found to be somewhat unrealistic. Authors including Boyarchenko & Levendorskiĭ [23] and Jondeau et al. [71], who attribute the observation to DuMouchel, argue that the tails of the stable distributions appear *too heavy*: return distributions in reality appear to have finite second, third and probably higher moments. Boyarchenko & Levendorskiĭ go further by observing that, if (X_t) is α-stable with $\alpha < 2$, then $E[e^{X_t}] = \infty$, so that options pricing by usual methods runs into problems at the outset. So the heavy tails in the stable models lead to complications that even financial time series do not warrant.

By using the subordination technique described above, we can obtain the variance gamma (VG) model, advocated by Madan & Seneta [84] in 1990. In

[16] See Jacob [67], which provides technical details as well as historical notes.
[17] See Sato [99], DEFINITION 13.4. If $\mu = 0$, then X is called *strictly stable*.

fact either of the following yield a VG process:

(i) Let C and D be independent gamma processes with identical distributions (i.e. Lévy processes such that $C_1 \sim D_1 \sim \text{Gamma}(c, \sqrt{2\rho})$ for some pair (c, ρ)), and let $X = C - D$.

(ii) Let B be a Brownian motion, and let γ be an independent gamma process with parameters c and ρ. Define $X_t = B_{\gamma(t)}$.

The processes as constructed by these two methods have the same distribution. Since for a Brownian motion B, $\text{var}\,(B_t) = t$, we have in (ii), $\text{var}\,(X_t \mid \gamma(t)) = \gamma(t)$, which explains the name *variance gamma*. More generally, we permit the two gamma subordinators in (i) to have their own parameters, $C_1 \sim \text{Gamma}(c_+, \rho_+)$ and $D_1 \sim \text{Gamma}(c_-, \rho_-)$. This process then has Lévy-Khintchine triple $(0, 0, \nu)$ where ν has Lebesgue density

$$g(x) = \begin{cases} c_+ x^{-1} e^{-\rho_+ x}, & x > 0, \\ c_- |x|^{-1} e^{-\rho_- |x|}, & x < 0. \end{cases}$$

Here, c_+, c_-, ρ_+ and ρ_- are positive parameters. These processes are called bilateral gamma processes in Küchler & Tappe [79]. The VG process was later (2002) generalized by Carr, Geman, Madan and Seneta [26], by simply defining the Lévy measure of a new infinitely divisible distribution, the CGMY distribution, to have the more general density

$$g(x) = \begin{cases} C_+ x^{-Y-1} e^{-Gx}, & x > 0, \\ C_- |x|^{-Y-1} e^{-M|x|}, & x < 0 \end{cases}$$

for some $Y \in [0, 2)$. The CGMY model contains the VG model with $Y = 0$, so that it is a proper generalization of that model. When applying the model to financial data, Carr et al. [26] found significant evidence that $Y > 0$, meaning that the CGMY model performed better than the VG model. This family of models is also called KoBoL in Boyarchenko & Levendorskiĭ [23].

Another possibility of subordination was explored by Barndorff-Nielsen & Levendorskiĭ [11]. They use a *tempered stable* subordinator: let $p(x; \alpha, \delta)$ denote the probability density function of an α-stable distribution concentrated on $(0, \infty)$, with Laplace transform

$$\int_0^\infty e^{-\lambda x} p(x; \alpha, \delta) dx = \exp\left(-\delta(2\lambda)^\alpha\right).$$

The tempered stable distribution is then defined by taking another parameter

$\gamma > 0$, and observing that with $\lambda = \gamma^2/2$, the function of x given by

$$p(x;\alpha,\gamma,\delta) = e^{\delta\gamma^{2\alpha}} p(x;\alpha,\delta) e^{-\frac{1}{2}\gamma^2 x}$$

is the density of an infinitely divisible distribution on $(0,\infty)$. The associated subordinator is called a *tempered stable subordinator*. Barndorff-Nielsen and Levendorskiĭ take a Brownian motion with drift, and subordinate by the tempered stable subordinator. The resulting process is called a *Normal Tempered Stable* (NTS) Lévy process. The special case $\alpha = 1/2$ gives us the *Normal Inverse Gaussian Process*, proposed for finance by Barndorff-Nielsen [10].

In 1977, O. Barndorff-Nielsen [9] proposed the generalized hyperbolic (GH) distributions to model the size of sand particles. A subclass of the class of GH distributions is that of the hyperbolic distributions. These are most conveniently defined through their probability density functions because, while the logarithm of the Gaussian pdfs is a parabola, the logarithm of a hyperbolic distribution is a hyperbola: thus, in the present terminology, we would call the Gaussian distribution a parabolic distribution. Hyperbolic distributions are considered for financial applications by Eberlein & Keller [39]. There they define the pdf of a hyperbolic distribution as

$$\mathrm{hyp}(x) = \frac{\sqrt{\alpha^2-\beta^2}}{2\alpha\delta K_1(\delta\sqrt{\alpha^2-\beta^2})} \exp\left(-\alpha\sqrt{\delta^2+(x-\mu)^2} + \beta(x-\mu)\right).$$

The introduction of a fifth parameter gives us the pdf of a GH distribution. For the sake of completeness, we record this as

$$\mathrm{gh}(x) = a(\lambda,\alpha,\beta,\delta)(\delta^2+(x-\mu)^2)^{(\lambda-1/2)/2} K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2}) e^{\beta(x-\mu)}.$$

Eberlein & Prause [40] say that these distributions are well suited to describing asset returns.

Other more general types of Lévy processes have been considered. A Lévy process is said to be *completely asymmetric* if its Lévy measure ν is concentrated on a half-axis $(-\infty,0)$ or (o,∞). If it is just concentrated on the negative half-axis, it could also be called a *spectrally negative* Lévy process. These processes have also been the subject of attention in finance and insurance applications. The simplest non-trivial example is the well-known Cramér-Lundberg model from collective risk theory $X_t = u + ct - C_t$ where (C_t) is a compound-Poisson process, see for example Embrechts et al. [41]. Thus, the setting of a spectrally negative Lévy process, used by Avram et al. [8], is a generalisation of this model. Spectrally negative Lévy processes were also used by Avram et

al. [7] to model the underlying asset returns, after which they proceed to solve the problem of pricing Russian options and Canadized Russian options on the underlying asset. The authors comment that the assumption of only negative jumps was largely for the tractability gained, since it then permits the use of results on spectrally negative Lévy processes, but that Carr & Wu [27] also found some empirical evidence that a spectrally negative Lévy model, specifically an α-stable process with $\alpha \in (1, 2)$ may be appropriate for the modelling of stock returns.

1.5.2 Diffusion models

Diffusion processes may in some generality be thought of as continuous, strong Markov processes, with a locally compact topological state space S, but in financial applications the model is likely to be specified by a stochastic differential equation (SDE). We shall consider only SDEs in one dimension, with state space $\mathbf{R}$, or a subset of $\mathbf{R}$. Such a diffusion model requires specification of three infinitesimal characteristics, namely the infinitesimal mean (*drift coefficient*), infinitesimal variance (*diffusion coefficient*) and the *infinitesimal killing rate*. Assuming a constant zero killing rate, a diffusion model is of the form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t, \qquad t > 0, \quad X_0 = x \tag{1.5}$$

for some $x \in I$, where the state space I is an interval in $\mathbf{R}$, W is a standard Wiener process, $b : \mathbf{R}_+ \times \mathbf{R} \to \mathbf{R}$ is the drift coefficient and $\sigma^2 : \mathbf{R}_+ \times \mathbf{R} \to \mathbf{R}_+$ is the diffusion coefficient, b and σ satisfying for example the conditions of Klebaner [77] **Theorem 5.2** so that a (strong) solution of (1.5) exists and is unique[18]. If the coefficients do not depend on time t, then the model is time-homogeneous: our diffusion models are always time-homogeneous. The aBm model corresponds to the case in which the coefficients b and σ depend on neither the state space variable nor the time parameter - they are simply constants. The only time-homogeneous diffusion model which is also space-homogeneous is therefore the aBm model, so that generalisations of this process as a diffusion involve removing the spatial homogeneity. Because of this, diffusion models, unlike Lévy processes, are capable of exhibiting mean-reverting features or trends (developing momentum).

Mean reversion makes diffusions suitable for modelling spot rate dynamics.

[18]The conditions are that $b(t, x)$ and $\sigma(t, x)$ are locally Lipschitz in x uniformly in t, and that they satisfy a linear growth condition in x. They ensure existence and uniqueness of a solution of (1.5), and are given also in Kallenberg [72] **Theorem 21.3** and Rogers & Williams [97] . In our investigation, we shall devote attention to the one-dimensional time-homogeneous case, in which the slightly sharper Yamada-Watanabe conditions existence and uniqueness of a strong solution, see Theorem 5.

Examples of diffusion models developed for spot rates are

- the *Merton* model

$$dX_t = \mu dt + \sigma dW_t, \qquad t > 0, \quad X_0 = x,$$

  Merton [88];

- the *Vasicek* model

$$dX_t = \kappa(\mu - X_t)dt + \sigma dW_t, \qquad t > 0, \quad X_0 = x,$$

  Vasicek [110];

- the *Cox-Ingersoll-Ross* (CIR) model

$$dX_t = \kappa(\mu - X_t)dt + \sigma\sqrt{X_t}\, dW_t, \qquad t > 0, \quad X_0 = x,$$

  Cox et al. [33];

- The *Ho-Lee* model

$$dX_t = \mu_t dt + \sigma dW_t, \qquad t > 0, \quad X_0 = x,$$

  Ho & Lee [60]

- The *Black-Derman-Toy* model

$$dX_t = \mu_t X_t dt + \sigma_t dW_t, \qquad t > 0, \quad X_0 = x,$$

  Black et al. [16]

Hull & White [62] generalize the Vasicek model to

$$dX_t = \kappa_t(\mu_t - X_t)dt + \sigma_t dW_t, \qquad t > 0, \quad X_0 = x,$$

while Black & Karasinski [17] specify

$$dX_t = X_t(\theta_t - \varphi_t \log X_t)dt + \sigma_t dW_t, \qquad t > 0, \quad X_0 = x.$$

Merton's model is simply an aBm. The Vasicek model uses an Ornstein-Uhlenbeck (OU) process to model the short rate. A process satisfying the SDE of the CIR model will be called a *CIR process*. For all these models the transition density is known: it is elementary to derive it for the OU process, while for the CIR model it had previously been found by Feller [48]. The Vasicek models (including the Merton model) have Gaussian transition probabilities,

concentrated on $\mathbf{R}$. The CIR model has state-space $(0, \infty)$ or $[0, \infty)$, and its transition probability distributions are non-central chi-square distribution. So long as $\kappa > 0$, these (OU and CIR) processes are mean-reverting. They have a long-term mean of μ, and an invariant distribution π, so that as $t \to \infty$, $X_t \xrightarrow{d} \pi$. In the OU process, π is Gaussian, while in the CIR process, π is a gamma distribution.

These models exhibit some elements of current state sensitivity (i.e. dependence of dX_t on X_t), but fall short of a full feedback of the price in the returns' dynamics, especially in the volatility. Our efforts are to be concentrated then in expanding the class of known diffusion models available to the financial modeller.

Before concluding our remarks on diffusion models, we just mention that, by Itô's lemma, it is possible to form new diffusions from existing ones. If X is a diffusion process satisfying (1.5) and f is a C^2 function, then the process $f(X) = (f(X_t))_{t \geq 0}$ is a diffusion process satisfying the equation of motion

$$df(X_t) = Af(X_t)dt + \sigma(X_t)f'(X_t)dW_t, \qquad t > 0, \quad X_0 = x,$$

where

$$Af(x) = \frac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x).$$

In a formal sense, A here is the infinitesimal generator of the process X. We shall have more to say on infinitesimal generators and their rôles in the study of diffusion processes in section 2.3, but we just mention here that, as shown in Shaw & Schofield [102], an instance of this arises in the models that we develop in this investigation, specifically the process $\sinh(W_t)$. This process can be understood through Bougerol's identity [22]. Furthermore, Alili et al. [3] showed a beautiful generalisation of Bougerol's identity which relates to the process $\sinh(Y_t)$ when Y_t is specified via the dynamics

$$dY_t = \left(c_0 \tanh(Y_t) + \frac{c_1}{\cosh(Y_t)} \right) dt + dW_t, \qquad t > 0, Y_0 = y.$$

Shaw & Schofield [102] noted further that this process, $\sinh(Y_t)$ with $c_0 = 1$ and $c_1 = 0$, also arises in the context of the hybrid diffusions we are about to develop, and noted a symmetry between the two processes $\sinh(W_t)$ and $\sinh(Y_t)$, which can be interpreted in terms of the Legendre symmetry $P_\nu \leftrightarrow P_{-\nu-1}$. For a collection of results and survey of literature relating to Bougerol's identity, see Vakeroudis [109].

1.5.3 Jump-diffusion models

The theoretical generalization of diffusion processes and Lévy processes, which includes both those classes, is the class of Itô processes. These are described by an integral equation. (The proper interpretation of an SDE of the form (1.5) is as an integral equation). In addition to a standard Wiener process W, we introduce an independent standard Poisson random measure N on $\mathbf{R}_+ \times \mathbf{R}_+$, i.e. such that

$$\text{for all } B \in \mathcal{B}(\mathbf{R}_+ \times \mathbf{R}_+), \quad N(B) \sim \text{Poi}(|B|)$$

where $|B|$ denotes the Lebesgue measure of the set B. We further denote the *compensated* Poisson random (signed) measure $\tilde{N} = N - EN$, i.e. $\tilde{N}(B) = N(B) - |B|$; then the equation of motion for an Itô process is[19]

$$X_t = X_0 + \int_0^t \Big\{ b(X_s)ds + \quad \sigma(X_s)dW_s + \int_{\mathbf{R}} j(X_{s-}, y) 1_D(X_{s-}, y) N(ds, dy)$$

$$+ \int_{\mathbf{R}} j(X_{s-}, y) 1_{D^c}(X_{s-}, y) \tilde{N}(ds, dy) \Big\}, \quad t \geq 0,$$

where

$$D = \{(x, y) \in \mathbf{R} \times \mathbf{R}_+ : |j(x, y)| > 1\},$$

and where $b : \mathbf{R} \to \mathbf{R}$, $\sigma : \mathbf{R} \to \mathbf{R}_+$ and $j : \mathbf{R} \times \mathbf{R}_+ \to \mathbf{R}$ are deterministic functions. The function j gives the size of the jumps.

This level of generality is too great, for our purposes of investigating the distributions of the underlying asset's returns, and we do not expect such models to be tractable. Instead, we shall consider a special case, in which only finitely many jumps occur in each bounded time-interval. We shall call the processes we are about to describe *jump-diffusions*. They are assumed to satisfy the equation of motion

$$X_t = X_0 + \int_0^t \left\{ b(X_s)ds + \sigma(X_s)dW_s + \int_{\mathbf{R}} j(X_{s-}, y) N(ds, dy) \right\}, \qquad t \geq 0, \tag{1.6}$$

where it is assumed that there is a $\lambda > 0$ such that $j(x, y) = 0$ for all $y \geq \lambda$. Given that a jump occurs from the point $X_{t-} = x$, the jump has size $\Delta X_t = j(x, U)$ where $U \sim \text{Unif}(0, \lambda)$. The conditional distribution of this jump is denoted[20] $J(x, \cdot)$: J is a probability kernel from $\mathbf{R}$ to $\mathbf{R}$. Between its jumps

[19] Çinlar [30], 3.95.

[20] This is non-standard notation, adopted only here for convenience. $y \mapsto j(x, \lambda y)$ is the quantile function of $J(x, \cdot)$.

the process acts as a diffusion

$$d\bar{X}_t = b(\bar{X}_t)dt + \sigma(\bar{X}_t)dW_t, \qquad \tau_{n-1} < t < \tau_n,$$

(where τ_n denotes the time of the nth jump) killed at a rate given by

$$k(x) = |\{y \in \mathbf{R}_+ : j(x,y) \neq 0\}|.$$

At a 'killing time' τ, instead of being sent to the cemetery state, X simply chooses a new starting point, according to the law $J(X_{\tau-}, \cdot)$.

It is clear how this generalizes the notion of a diffusion. It does not generalize the notion of a Lévy process in the generality with which we have defined them. The motion (1.6) is a Lévy process if b and σ do not depend on X_s and j does not depend on X_{s-}. The Lévy processes this motion is capable of producing are therefore those of the form

$$X_t = bt + \sigma W_t + C_t, \qquad t \geq 0,$$

where the *compound Poisson process* (C_t) is given by

$$C_t = \sum_{n=1}^{N_t} Y_n,$$

for some independent i.i.d. sequence (Y_n) and an independent Poisson process (N_t) with rate λ. The characteristic exponent for this particularly simple Lévy process is

$$\frac{1}{t}\log Ee^{i\xi X_t} = ib\xi - \frac{1}{2}\sigma^2\xi^2 + \int_{-\infty}^{\infty}(e^{i\xi y} - 1)\lambda F_Y(dy)$$

where F_Y denotes the distribution of each Y_n. Thus our jump-diffusions cannot be Lévy processes of infinite activity.

Some jump-diffusion processes which have been considered in the literature are Lévy processes of finite activity, but they arose differently to those mentioned above in the class of Lévy models. The models described as an aBm with a compound Poisson process of jumps superimposed are sometimes themselves referred to as jump-diffusions. Examples of jump-diffusion models of this type are those of Merton [88] and Press [92], in which the jump distribution F_Y was assumed to belong to the Gaussian family.

In the full level of generality, an approach to pricing double barrier options on an underlying Itô process (which they call a jump-diffusion) was presented in Eriksson & Pistorius [44]. The authors also mention that commodity prices

exhibit features of jumps and mean-reversion, see the references therein. The equation of motion is actually given as

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dW_t + \lambda(t, X_t)dJ_t, \qquad t > 0, \quad X_0 = x,$$

where (J_t) is a pure-jump Lévy process (i.e. with Lévy-Khintchine triple $(\mu, \sigma, \nu) = (0, 0, \nu)$). Since J may have infinite activity, this process is an Itô process, rather than a jump-diffusion process as we have defined them. A barrier option becomes worthless when the underlying X leaves some set B. The authors show how to use a method of moments approach, based on the discounted exit location measure

$$\xi(A) = E[e^{-\alpha_\tau} 1_{\{(\tau, X_\tau) \in A\}}]$$

and the discounted occupation measure

$$\omega(A) = E\left[\int_0^\tau e^{-\alpha_\tau} 1_{\{(s, X_x) \in A\}} ds\right],$$

where

- τ is the first time at which X leaves the set B;
- α is given by

$$\alpha_t = \int_0^t r(s, X_s)ds$$

  for the rate r of discounting.

1.6 Plan of investigation

We have now reviewed the basic Gaussian model employed in finance and highlighted some major flaws, which are well-known. We have also discussed some more sophisticated models which have been developed to address to some extent some of these flaws. Our aim is now to motivate, study and illustrate a further class of such models, that we have termed *hybrid diffusion models*. The structure of the remainder of this investigation is then as follows.

In Chapter 2, we review the theory and associated literature needed for each of the subsequent chapters in turn. In Chapter 3, we develop a market microstructure model in which the impact of each individual trade order on the market state is modelled. The latter model results in a process in which the market state makes successive and finitely many discrete jumps between successive states. Chapter 4 describes the procedure of approximating this

discrete jump process with a continuous one. An equation of motion is derived, which allows us to discuss the possible additional assumptions one can make in the feedback model and the resulting specific model.

At this point, we choose to specify two models in particular, the arithmetic-geometric hybrid Brownian motion of Shaw & Schofield [102], studied in Chapter 5, and the arithmetic-Cox-Ingersoll-Ross process which is new, and which we study in Chapter 6. In Chapter 7 we conduct a maximum likelihood test to show that our models do indeed fit return distributions better than classical models. We then give some numerical illustrations of these two processes in Chapter 8 before concluding.

Chapter 2

A theory and literature review

Here we look forward to each of the subsequent sections in turn and point to the relevant mathematical headings under which they fall. The minimal theory for our investigation is developed and with each topic summary the reader is pointed to some of the associated literature. Before we begin, here is a summary of the following discussions.

In section 2.1, we give an exposition of the theory of renewal processes and renewal-reward processes, whose importance is that we can use them to describe the trade order flow. Given certain other assumptions regarding the price formation process, we obtain a discrete renewal model of the market. The main result from renewal theory that we are interested in is the renewal theorem, Theorem 1, which tells us that under relatively weak conditions the underlying renewal process converges in law to a stationary renewal process. As a consequence, we get a simple description of the large-time mean and variance of the renewal-reward process.

Section 2.2 is a summary of the concepts and results we require from the theory of weak convergence of probability measures, which is relevant to our investigation through our aim of approximating the discrete renewal-type model mentioned above by a continuous diffusion process. In this section, we give theorems 2-4. Theorem 2 tells us that a sequence of processes which converge in distribution in the space $C[0,T]$ of continuous functions also converge in distribution to the same limit in the Skorohod space $D[0,T]$. Theorem 3 is Donsker's invariance principle, which is a functional Central Limit Theorem, stating essentially that a suitable successive scalings of a random walk (based on a distribution with zero mean and finite variance) converge in distribution to a standard Brownian motion. A variation of this is Theorem 4, which replaces the random walk (S_n) in Donsker's Theorem with the renewal-reward process $(S_{N(t)})$.

The longest section of this chapter is section 2.3, which goes through the concepts that we shall require from the theory of Markov processes, including the theory of one-dimensional diffusions. This is necessarily in some detail, since then a direct application of this thoery allows us to present a characterisation of the probability distributions associated with our models (in most cases, this is through the associated Green function). The general theory of Markov processes and the more specific theory of one-dimensional diffusions are quite well-known, so we simply explain and summarise these concepts without formal statements of the results. We formally present three theorems in this section, theorems 5, 6 and 7. Theorem 5 gives conditions under which a one-dimensional SDE makes sense in the sense that a strong solution exists and is unique. Theorem 6 is Linetsky's spectral classification of one-dimensional diffusions, based on the Sturm-Liouville theory associated with its generator, while Theorem 7 characterises the oscillatory/non-oscillatory nature of the boundary points via the Liouville transformation.

In section 2.4 we give some definitions from the general theory of stochastic processes. This material follows Jacod & Shiryaev [68]. The reason for introducing this material is only to state two convergence theorems for semimartingales when we come to section 4.6. We hope these theorems will be useful in a further investigation to prove that our convergence results hold when the Markov assumption is removed. The situation is then one in which the approximating processes are semimartingales. We introduce the set-up on which semimartingales are defined, and go on to define the characteristics of semimartingales, in terms of which these convergence theorems will be stated. This material is relevant only to the open question of convergence of the microstructure model and may therefore be omitted if the reader is interested only in the development and implementation of hybrid diffusion models.

Once we have applied the analysis of section 2.3 to our models, our solutions will be in the form of special functions, whose behaviour we shall need to understand. In sections 2.5 and 2.6, we summarise the basic properties of the Legendre functions and the confluent hypergeometric functions. We do not present any major results in these sections; rather these sections are intended as a reference for properties of these functions that we intend to apply later.

In section 2.7, we discuss the definition of a risk measure, and desirable properties of risk measures. We then go on to define value-at-risk (VaR) and expected shortfall (ES), two risk measures which are used in finance. The reason for doing this is that we would like to be able to show how these risk measures may be computed for our new models, and use them to compare our

model with the Gaussian model.

The final section of this chapter is section 2.8, in which we show some methods in which parameters may be fitted to our models. We begin by showing that, via a rather crude approximation, we could use a standard likelihood technique, based on the invariant distribution of our model (assuming it exists). This also allows us to perform an investigation quantifying the goodness of fit. For a dynamic parameter estimation, this technique is not sufficiently accurate and we suggest the use of martingale estimating functions. We give the basic definitions and the usual method of constructing them, but there is much further work to do in this area.

2.1 Renewal and semi-Markov processes

Our feedback models are based on a random arrival of trade orders to a market. We suppose that the trades each affect the security price directly and proportionally to the size of the trade order. These natural assumptions are the simplest within this type of model and it makes sense to investigate the consequences of these simplest assumptions before generalizing. The two sources of randomness in the model are then the randomness of the time instant at which a trade arrives at the market and the random size of the trade order. The total quantity which has been ordered since a certain fixed time in the past then forms a renewal-reward process; the number of trades gives a renewal process. The theory of such processes is called renewal theory and is used for example in modelling the flow of insurance claims arrivals.

A modern systematic treatment of renewal theory begins with a transient random walk (T_n) on $\mathbf{R}$. Thus, we do not need the increments $\tau_n = T_n - T_{n-1}$ to be positive, just to have a distribution such that $|T_n| \to \infty$ as $n \to \infty$ with probability one. However, we shall develop the theory assuming that the τ_n are positive, since they shall represent waiting times in our framework. Then (T_n) is a transient random walk based on a distribution F_τ concentrated on $(0, \infty)$. The initial position T_0 is permitted to have an arbitrary distribution F_0, so the renewal sequence (T_n) is actually defined as

$$T_0 \sim F_0, \qquad T_n = T_{n-1} + \tau_n, \quad n \geq 1,$$

where (τ_n) is an i.i.d. sequence of F_τ-distributed random variables, the so-called *holding times*. When $T_0 = 0$, the renewal process is called *pure*, otherwise we call it *delayed*.

The *renewal process* is the random measure

$$N = \#\{T_n \in \cdot\} = \sum_{n\geq 0} \delta_{T_n}, \quad N(B) = \sum_{n\geq 0} 1\{T_n \in B\} \tag{2.1}$$

where δ_x denotes unit mass at x. By (N_t) we denote its (random) distribution function, also called the renewal process; thus,

$$N_t = N[0,t] = \sum_{n\geq 0} 1\{T_n \leq t\}, \quad t \geq 0.$$

By the *intensity* (or *intensity measure*) of N, we mean the (nonrandom) measure

$$EN = F_0 * \sum_{n\geq 0} F_\tau^{n*} \tag{2.2}$$

where F^{n*} denotes the nth convolution power[1] of the measure F. Sometimes $EN(dt) = \lambda dt$ for a constant $\lambda > 0$ (dt denoting integration with respect to Lebesgue measure) in which case λ is also referred to as the intensity, or the *rate*, of (N_t).

Introduce the shift operators ϑ_t on the set of all measures μ on $\mathbf{R}_+$ by setting

$$\vartheta_t \mu(B) = \mu(B + t), \quad B \in \mathcal{B}(\mathbf{R}_+),\, t \geq 0.$$

The renewal process N is said to be stationary if $\vartheta_t N \sim N$ for all $t \geq 0$. One of the basic results of renewal theory[2] is that for each F_τ having a finite mean $m \in (0,\infty)$, there is a unique choice of initial distribution F_0, such that the renewal process N is stationary, namely

$$F_0(t) = m^{-1} \int_0^t F_\tau(s,\infty) ds, \quad t \geq 0.$$

Conversely also, if F_τ does not have a mean belonging to $(0,\infty)$ then N cannot be stationary. A characterisation of the Lebesgue measure[3] as the only shift invariant measure (on any of the Euclidean spaces $\mathbf{R}^d$) then implies that $EN(dt) = \lambda\, dt$ for some $\lambda > 0$. In fact, it is easily found that $\lambda = m^{-1}$. In developing renewal models, we shall hereafter begin from a random walk based on a distribution F_τ with mean λ^{-1}. Now, the renewal theorem, usually attributed to D. Blackwell and W. Smith, states[4] the following:

[1] If independent random variables X_1 and X_2 have distributions F_1 and F_2, then the distribution of $X_1 + X_2$ is $F_1 * F_2$. If $F_1 = F_2 = F$, this is written $F * F = F^{2*}$.

[2] See Kallenberg [72] Proposition 9.18; Feller [53] Chapter XI, equation (4.6) and the discussion thereafter; Resnick [96] Theorem 3.9.1.

[3] Kallenberg [72] **Theorem 2.6**

[4] This is the one-sided version of Kallenberg [72] **Theorem 9.20**

Theorem 1 (renewal theorem)**.** *Let N be a renewal process based on a nonarithmetic*[5] *distribution F_τ with mean λ^{-1} and an initial distribution F_0. Let $\tilde{N}$ be a stationary renewal process based on the same transition distribution F_τ. Then as $t \to \infty$, $\vartheta_t N \xrightarrow{d} \tilde{N}$ and $\vartheta_t EN(dx) \xrightarrow{v} \lambda dx$ where $\mu_n \xrightarrow{v} \mu$ denotes convergence in the vague topology.*

Here, the vague topology is defined by $\mu_n \xrightarrow{v} \mu$ if $\mu_n u \to \mu u$ for all non-negative continuous functions u having compact support. The vague topology makes sense on the space $\mathcal{M}(S, \mathcal{S})$ of locally finite measures on a locally compact second countable Hausdorff (lcscH) space endowed with its Borel σ-field $\mathcal{S}$. Vague convergence is defined, for example, in Billingsley [14] section 28, and Kallenberg [72] Chapter 16 (p316). For convergence in distribution ($\xrightarrow{d}$), see section 2.2.

From Theorem 1, the mean number of renewals occurring in $[0, t]$ is asymptotically equal to λt,

$$EN_t \sim \lambda t, \quad t \to \infty, \tag{2.3}$$

with equality for all t in the case of a stationary renewal process. A corresponding result concerning the asymptotic behaviour of the variance is also valid when the holding times have a finite variance, and may be deduced as follows (Feller [53] Chapter XI, Problem 13; see also Resnick [96] Exercise 3.21. Taylor & Karlin also quote the result in [107] Chapter VII, equation (4.4).). First, assume that the renewal process is pure. By considering the representation of the renewal process

$$N_t = \sum_{n \geq 0} 1\{T_n \leq t\},$$

we find that

$$E[N_t^2] = EN_t + 2 \sum_{m \geq 1} m F_\tau^{m*}(t) = 2\, U * U(t) - U(t),$$

where U is the *renewal measure*

$$U(t) = \sum_{n \geq 0} F_\tau^{n*}(t),$$

(equal to EN_t, because the renewal process is pure). Integration by parts leads to

$$E[N_t^2] = 2\lambda \int_0^t U(s) ds + \sigma_\tau^2 \lambda^3 t,$$

[5] A distribution is said to be nonarithmetic if it is not concentrated on $h\,\mathbf{Z}$ for some $h > 0$

from which we obtain

$$\text{var}\,(N_t) \sim \sigma_\tau^2 \lambda^3 t, \quad t \to \infty. \tag{2.4}$$

For delayed renewal processes N, (2.4) still holds, whatever the delay distribution F_0, but with equality if and only if N is stationary.

Introducing next an independent i.i.d. sequence of rewards R_n, we may define a new process

$$D_t = \sum_{n=1}^{N_t} R_n = \sum_{n \geq 1} R_n 1_{[0,t]}(T_n).$$

This *renewal-reward process* generalizes the compound Poisson process, which is the instance of it when N is a Poisson process or, equivalently, the i.i.d. holding times τ_n are exponentially distributed. It takes only straightforward conditioning to observe that the mean satisfies

$$ED_t = ER_1 \cdot EN_t \sim \mu\lambda t, \quad t \to \infty$$

where it is supposed that $\mu = ER_1$ exists, and that the variance satisfies

$$\text{var}\,(D_t) = EN_t \text{var}\,(R_1) + \text{var}\,(N_t)(ER_1)^2 \sim [\sigma_R + (\mu\sigma_\tau\lambda)^2]\lambda t, \quad t \to \infty$$

when $\text{var}\,(R_1) = \sigma_R^2 < \infty$.

Each R_n may be described as a step or increment in a random walk. The random walk is then the discrete-parameter process (S_n) where

$$S_0 = 0, \qquad S_n = R_1 + \cdots + R_n = S_{n-1} + R_n, \quad n \geq 1.$$

The reward-renewal process may then be written

$$D_t = S_{N(t)}, \quad \text{or} \quad D = S \circ N$$

for the underlying renewal process N. Evidently D is Markovian if and only if N is a Poisson process. In this construction, the processes (S_n) and N were assumed independent.

If N is a Poisson process, then the process $D = S \circ N$ is, of course, a compound Poisson process. As time passes, the jumps arrive according to a Poisson process and the increments are independent of N and i.i.d. If we replace the random walk with a discrete-parameter Markov process S, then the resulting process $D = S \circ N$ is called a pseudo-Poisson process, see Feller [53]

X.1. When the process is at x, waiting for its next jump, the time until the next jump has an exponential distribution with some rate, denoted ρ. If now this rate $\rho(x)$ is permitted to depend on the present state of the process, we have the most general pure-jump Markov process. If the rate function $\rho \geq 0$ is bounded above on $\mathbf{R}$, then it is possible[6] to write it as a pseudo-Poisson process (with constant rate function) in which virtual jumps (i.e. jump from x to x) are permitted.

To generalize these constructions further, let N be any renewal process. If S is a random walk, we see that $S \circ N$ is then a reward-renewal process as defined above. If instead, the process S is a discrete-time Markov process and at the time of jumping to x, the waiting time then has some distribution (not necessarily exponential) depending on x, we call X a semi-Markov process. It is not a Markov process, because at fixed times t the time that has since the last jump is relevant in determining the distribution of waiting time until the next jump, but it is Markov at its jump times. Furthermore, the bivariate process (D_t, A_t) where A_t denotes the time which has elapsed since the last jump, is Markov, see Gihman & Skorohod [55].

The point of introducing semi-Markov processes is that they allow us to incorporate a dependence of the holding time and jump distributions on the current state x. This will be relevant in developing feedback models in which traders, who are driving the asset price movements, react to the current state of the market.

For further discussions of renewal theory, the reader is referred to the relevant sections in Embrechts et al. [41], Feller [53], Kallenberg [72], and Resnick [96]. Semi-Markov processes are discussed in Çinlar [28] Chapter 10 and [29] and in Gihman & Skorohod [55] Chapter III §3.

2.2 Theory of weak convergence

A dynamic model of returns produced using a renewal-reward model of trade order arrival will involve discrete jumps. To take advantage of the stochastic calculus of Itô, we intend to use a continuous approximation to it. If we imagine the returns process as being represented by a graph of X_t against time t, then a diffusion approximation is obtained by 'zooming out' from the details of the frequent small price fluctuations. This is illustrated in figure 4.1. It may be the case that some changes in the return process are so large in such a small time period that they cannot be described by continuous diffusion and

[6] Kallenberg [72] Proposition 12.20, Ethier & Kurtz section 4.2.

so we should also like to consider how models of *jump-diffusion* type can arise.

Let (S, d) be a separable metric space, and denote by $\mathcal{S}$ the class of Borel sets in S. It is important that we are able to develop this theory without assuming even that the space S is locally compact, because typical applications of this theory are to stochastic processes, which are random elements living in one of the non-locally compact spaces $C[0,\infty)$ or $D[0,\infty)$. The set of all probability measures on $(S, \mathcal{S})$ is denoted $\mathcal{P}(S) = \mathcal{P}(S, \mathcal{S})$. If (μ_n) is a sequence and μ a point in $\mathcal{P}(S)$, we say that (μ_n) converges weakly to μ, and write $\mu_n \xrightarrow{w} \mu$, if

$$\mu_n(u) \to \mu(u) \quad \text{for all } u \in C_b(S),$$

$C_b(S)$ denoting the real Banach space of all bounded continuous functions $u : S \to \mathbf{R}$, with the norm of uniform convergence,

$$\|u\|_\infty = \sup_{x \in S} |u(x)|.$$

Thus, by the *weak* topology on $\mathcal{P}(S)$, we mean what functional analysts would mean by the weak* topology[7] on $\mathcal{P}(S)$ when it is regarded as a subset of $C_b(S)^*$.

The weak topology on $\mathcal{P}(S)$ is metrizable. In fact, it is induced by the metric defined by

$$d(P, Q) = \inf\{\varepsilon > 0 : QB \le PB^\varepsilon + \varepsilon,\ PB \le QB^\varepsilon + \varepsilon,\ B \in \mathcal{S}\}$$

where we have written

$$B^\varepsilon = \{x \in S : d(x, B) < \varepsilon\}.$$

This metric is often known as the Lévy-Prohorov metric, or as the Lévy metric or the Prohorov metric. See Merkle [87] for further details; also useful are Billingsley [13], Ethier & Kurtz [46] and Stroock [106].

Let X_n, X be random elements of S with distributions μ_n, μ. We say that X_n converges in distribution to X, and write $X_n \xrightarrow{d} X$, if $\mu_n \xrightarrow{w} \mu$. If the X_n and X are all defined on the same probability space, this is the same as

$$Eu(X_n) \to Eu(X) \quad \text{for all } u \in C_b(S).$$

We must choose, for approximations by our feedback model, a suitable space S. The random processes of the approximating sequence and the limit process

[7] Though our chosen terminology appears ambiguous, it is in fact standard in probability theory, and can cause no confusion. The 'real' weak convergence of functional analysis is never used in probability because it is too strong a form of convergence to be of any practical value. This is pointed out in Bobrowski [20] section 5.6 and Stroock [106] section 3.1. See also the discussion in Rogers & Williams [97] preceding II.80.

are all elements of the space $D(\mathbf{R}_+, \mathbf{R})$, which consists of the functions $x : \mathbf{R}_+ \to \mathbf{R}$ which are right continuous with left limits (rcll). For the definitions of the Skorohod topology and notions of weak convergence in $D(\mathbf{R}_+, \mathbf{R})$, see Billingsley [13] Chapter 4, Ethier & Kurtz [46] Chapter 3, Iglehart [65] Section 2, Jacod & Shiryaev [68] Chapter 6 and Kallenberg [72] Chapter 16. We consider first the geometry of $D([0,T], \mathbf{R})$. The uniform norm does actually make sense on this space, but it is not separable.

The index set $[0,T]$ is always thought of as time. A time-deformation of $[0,T]$ is a bijective continuous increasing map $\lambda : [0,T] \to [0,T]$, and we denote by Λ_T the set of all time-deformations of $[0,T]$. For $\lambda \in \Lambda_T$, we define

$$\|\lambda\| = \sup_{s \neq t} \left| \log \left(\frac{\lambda(t) - \lambda(s)}{t - s} \right) \right|$$

Since $(\lambda(t) - \lambda(s))/(t-s)$ approximates the derivative of λ if it exists when $t-s$ is small, $\|\lambda\|$ can be thought of as telling us how much λ deforms time over its worst interval (s,t). The Skorohod metric d is defined on $D_T = D([0,T], \mathbf{R})$ as

$$d(x,y) = \inf\{\varepsilon > 0 : \|x - y \circ \lambda\|_\infty < \varepsilon \text{ for some } \lambda \text{ with } \|\lambda\| \leq \varepsilon\}. \qquad (2.5)$$

Under the Skorohod metric, elements x and y of D_T are within ε of each other if and only if a time-deformation can be found, itself not larger than ε, to make y within ε of x in the uniform norm. The metric d is complete and induces a separable topology.

Our next step is to consider some concrete weak convergence results in D_T and in $D := D(\mathbf{R}_+, \mathbf{R})$. The now classical functional central limit theorem of Donsker [36] (Billingsley [13] Theorem 8.2) is a weak convergence result in $C = C(\mathbf{R}_+)$ but may be adapted to D, by a result of Liggett & Rosén to obtain Theorem 3 below (quoted from Billingsley [13] Theorem 14.1 and Kallenberg [72] Theorem 14.9). Liggett & Rosén's result is Theorem 3 of Iglehart's paper [65] and says

Theorem 2 (weak convergence for function spaces C and D)**.** *Let (X_n) be a sequence of random elements of $C[0,T]$, (Y_n) a sequence of random elements of $D[0,T]$ and X a random element in $C[0,T]$. Denote the Skorohod metric on $D[0,T]$ by ρ. If $\rho(X_n, Y_n) \xrightarrow{d} 0$, then $X_n \xrightarrow{d} X$ in $C[0,T]$ if and only if $Y_n \xrightarrow{d} X$ in $D[0,T]$. In particular, $X_n \xrightarrow{d} X$ in $C[0,T]$ if and only if $X_n \xrightarrow{d} X$ in $D[0,T]$.*

In the space D_T, Donsker's invariance principle (Theorem 14.1 of Billingsley [13], Kallenberg [72] Theorem 14.9) states

Theorem 3 (Donsker's invariance principle)**.** *Let $X_1, X_2, \ldots$ be i.i.d. random variables with mean 0 and variance σ^2. Define*

$$X_t^n = \frac{1}{\sigma\sqrt{n}} \sum_{k \le nt} X_k, \quad t \le T;$$

then $X^n \xrightarrow{d} W$ where W is a standard Wiener process on $[0, T]$.

Billingsley, [13] Chapter 14, then shows how to place this in the context of renewal theory. Let (T_n) be a renewal sequence, so that $N_t = \sum_k 1\{T_k \le t\}$ is a renewal process. His **Theorem 14.6** there asserts the following. Assuming that constants λ^{-1} and σ_τ exist so that, for

$$X_t^n = n^{-1/2} \sum_{k=1}^{[nt]} (\tau_k - \mu), \quad \tau_k = T_k - T_{k-1},\, k \ge 1$$

we have $X^n \xrightarrow{d} \sigma W$; then with

$$Z_t^n = \frac{N_{nt} - \lambda n t}{\lambda^{3/2}\sqrt{n}}$$

we have $Z^n \xrightarrow{d} \sigma W$.

The central result of Iglehart's paper [65] (i.e. **Theorem 6**) extends this result to random sums, though his notation is designed for approximating risk reserve processes $u + ct - S_{N(t)}$. We quote essentially the same result from Embrechts et al. [41], **Theorem 2.5.17**, which states:

Theorem 4 (functional central limit theorem for random sum processes)**.** *Let (N_t) be a renewal process based on a distribution with mean λ^{-1} and variance $\sigma_\tau^2 > 0$. Let (R_n) be an independent i.i.d. sequence of random variables with mean μ and variance $\sigma_R^2 > 0$. Denote $S_n = X_1 + \cdots + X_n$, $n \ge 1$ and*

$$X_t^n = \frac{S_{N(nt)} - \lambda\mu n t}{\sqrt{(\sigma_R^2 + (\mu\lambda\sigma_\tau)^2)\lambda n}} \quad ; \tag{2.6}$$

then $X^n \xrightarrow{d} W$ in $D[0, \infty)$.

The standard reference for the theory of weak convergence of probability is the textbook by Billingsley [13]. For weak convergence of the distributions of processes, the reader is referred to Ethier & Kurtz [46], Jacod & Shiryaev [68] and Kallenberg [72] as general references. We shall extend the theory of this section in section 4.6 by stating two convergence theorems for semimartingales from Jacod & Shiryaev's text that we believe will be useful in extending the

diffusion approximation results obtained in this thesis to the case of non-Markov approximating processes. As far as I know, the idea of approximating renewal-reward processes using diffusion processes is due to Iglehart [65], who was working with an inflow of insurance claims. It is based on the invariance principle obtained by Donsker [36]. Also Embrechts et al. [41] mentions and summarises these ideas in the context of approximating random sum processes.

2.3 Markov processes and transition functions

The correct mathematical framework in which to study Markov processes is set out in Blumenthal & Getoor [19]. For general Markov process background theory, the reader is referred also to Ethier & Kurtz [46], Kallenberg [72] and Rogers & Williams [97]. Also, of historical significance and containing a vast quantity of relevant theory, we have the books of Dynkin [38] and of Itô & McKean [66].

For diffusion processes in one dimension, the theory is very complete. The process is characterized by a monotone increasing bijective *scale function*, a *speed measure* and a *killing measure* all defined on the state-space. The state space of the diffusion is an interval $I \subset \mathbf{R}$, and these characteristics determine the nature of the two boundary points. Borodin & Salminen [21] is a useful reference, since it points to how to determine scale and speed from the infinitesimal characteristics, and shows how to determine the boundary classification from them. This theory is quite standard now, and can be found for example in Karlin & Taylor [73] and Rogers & Williams [97]. It was born in the 1950s with the work of Feller [49, 50, 51], McKean [57] and others.

From analysis, we need the theory of strongly continuous operator semigroups. The transition function (i.e. the set of transition probabilities) of a Markov process, when considered as a family of operators on an appropriate function space, forms a strongly continuous contraction semigroup of positive operators. For one-dimensional diffusion processes, with a certain choice of measure m on the state space I, this semigroup is also self-adjoint on $L^2(I, m)$ (Linetsky [83]). The general theory of strongly continuous operator semigroups is in Engel & Nagel [42] and I have also used extensively the monograph of Jacob [67], since he restricts attention to results and topics relevant to the probabilist concerned with Markov processes. Classic monographs are Yosida [113] and Hille & Phillips [59]. Virtually any text on Markov processes in continuous time will include a preliminary discussion on strongly continuous semigroups. I mention again Ethier & Kurtz [46], as well as Feller [53] Chapter

X (also IX).

The other analysis topic we must consider is spectral theory, needed to study the generators of such semigroups. The generator uniquely determines the semigroup and is actually a more natural starting point for our investigations, because the returns process is specified by its infinitesimal characteristics rather than its transition probabilities. Ultimately our aim would be to obtain closed form expressions for the transition probabilities and their associated operators but, since these are often out of reach, we shall then settle for obtaining approximations, via a spectral analysis of the generator. Indeed in most situations we shall encounter, the transition semigroup my be recovered by means of a generalization of Dunford's functional calculus for bounded operators. The resulting expressions may be viewed as spectral decompositions of the transition semigroup and density, though they need not be calculated this way. In addition to the references given above, I mention the treatises of Dunford & Schwartz [37] and of Kato [74], and the textbook by Abramovich & Aliprantis [1]. Based on the well-known classification of Sturm-Liouville problems (Fulton et al. [54]), Linetsky [82, 83] has developed a spectral classification of all one-dimensional diffusion processes, and we shall base our analysis on it.

Let X be a strong Markov process with state space $S \subset \mathbf{R}$. Associated with each $x \in S$ we have a probability measure P^x, with the property $P^x\{X_0 = x\} = 1$. The transition function is defined as

$$P_t(x, A) = P^x\{X_t \in A\}.$$

Each P_t is a kernel from the state space to itself. We recall that a kernel $(x, A) \mapsto K(x, A)$ from a measurable space S to another measurable space T (here $x \in S$, $A \subset T$) induces a linear operator $Ku(x) = \int_T K(x, dy)u(y)$ formally mapping functions u defined on T to functions Ku defined on S. Thus P_t induces an operator

$$P_t u(x) = \int P_t(x, dy)u(y) = E^x u(X_t)$$

on, for example, $C_b(S)$. With abuse of terminology, we sometimes say that kernels are operators, or that operators are kernels. We shall have more to say on the choice of function space later, but for now we assume that the operators P_t are simply defined on a sufficiently rich Banach space $\mathbf{B}$. The fact that each $P_t(x, dy)$ is a (sub-) probability measure makes it a contraction (i.e. $\|P_t u\| \leq \|u\|$).

The kernels P_t satisfy the Chapman-Kolmogorov equations

$$P_{s+t}(x, B) = \int_I P_s(x, dy) P_t(y, A), \quad s, t \geq 0, x \in S, B \in \mathcal{S};$$

these are always assumed to hold for any Markov process possessing a transition function ($\mathcal{S}$ denotes the σ-field on S turning S into a measurable space). In terms of the operators P_t, this is equivalent to the semigroup property

$$P_{s+t} = P_t P_s, \quad s, t \geq 0.$$

For tractability reasons, it is usually advantageous to develop models in which the transition kernels P_t possess a density with respect to the Lebesgue measure. This is called the transition density and is denoted $p(t; x, y)$, that is,

$$P_t(x, dy) = p(t; x, y) dy.$$

We denote a transition density with respect to another measure m by $p_m(t; x, y)$, i.e.

$$P_t(x, dy) = p_m(t; x, y) m(dy).$$

This is particularly useful when the process X is symmetric with respect to a measure m. By this we mean that the transition density $p_m(t; x, y)$ is symmetric with respect to (x, y) (that is, $p_m(t; x, y) = p_m(t; y, x)$), in which case the transition operators are symmetric (hence self-adjoint, since they are bounded) in the Hilbert space $L^2(S, m)$. This case provides a dramatic analytic simplification which, encouragingly, is always available when X is a one-dimensional diffusion in the state space $S = I \subset \mathbf{R}$, an interval on the line.

2.3.1 Transition semigroups

Since it is quite standard, instead of developing the theory of these strongly continuous semigroups systematically, we give a list of the basic objects (numbers, functions, operators etc.) and their definitions. These concepts are mostly found in Engel & Nagel [42] and Jacob [67]. At the beginning here, the semigroup is defined only on an arbitrary Banach space $\mathbf{B}$. However, these operators will be defined by formulae which make sense on more than one Banach space, so that all these notions introduced here are dependent on $\mathbf{B}$.

- The semigroup itself is denoted (P_t). It is a family of operators in the space $\mathcal{L}(\mathbf{B})$ of bounded linear operators on $\mathbf{B}$, such that the map $t \mapsto P_t$ is continuous with respect to the strong topology on $\mathcal{L}(\mathbf{B})$.

- The *type* of the semigroup is defined as

$$\omega_0 = \inf\{w \in \mathbf{R} : \text{ there exists } M_w > 0 \text{ with } \|P_t\| \leq M_w e^{wt}\}$$

Note that $\|P_t\| \leq M_w e^{wt}$ for every $w > \omega_0$, but not necessarily for $w = \omega_0$. Strong continuity of the semigroup implies that $\omega_0 < \infty$, but it is possible that $\omega_0 = -\infty$.

- The generator

$$Au = \lim_{t \to 0} t^{-1}(P_t u - u),$$

is defined on

$$D(A) = \{u \in \mathbf{B} : \lim_{t \to 0} t^{-1}(P_t u - u) \text{ exists in } \mathbf{B}\}.$$

The generator is a closed densely defined operator in $\mathbf{B}$. If (P_t) is a contraction semigroup ($\forall t$, $\|P_t\| = 1$) then A is dissipative, i.e.

$$\|(\lambda - A)u\| \geq \lambda \|u\|, \quad \lambda > 0, u \in \mathbf{B}.$$

A is bounded (hence $D(A) = \mathbf{B}$) if and only if (P_t) is continuous in the uniform operator topology, in which case $P_t = e^{tA}$ in the usual sense for bounded operators. Of the processes mentioned in this thesis, the Markov processes with a bounded generator are precisely the pseudo-Poisson processes with a bounded rate function.

- The semigroup has a finite **spectral bound**

$$s(A) = \sup\{\operatorname{Re}\lambda : \lambda \in \sigma(A)\}.$$

From standard semigroup theory,

$$s(A) \leq \omega_0. \tag{2.7}$$

The spectral mapping theorem $\sigma(P_t) \setminus \{0\} = e^{t\sigma(A)}$ holds if and only if equality holds in (2.7).

- The **resolvent family** is $\{R(\lambda, A) = (\lambda - A)^{-1} : \lambda \in \rho(A)\}$. For $\operatorname{Re}\lambda > \omega_0$, one has

$$R_\lambda = R(\lambda, A) = \int_0^\infty e^{-\lambda t} P_t dt,$$

exhibiting the fact that the resolvent may be obtained as the Laplace transform of the semigroup. Indeed this definition may be used to prove (2.7). When (P_t) is a contraction semigroup, the fact that the generator

is dissipative implies that $(R_\lambda : \lambda > 0)$ is a *contraction resolvent*, i.e.

$$\|\lambda R_\lambda\| \leq 1, \quad \lambda > 0.$$

Note that this terminology means that the operators λR_λ are contractions, rather than the resolvent operators themselves.

- **Sectorial operators and Laplace inversion.** A densely defined operator $(A, D(A))$ in a Banach space $\mathbf{B}$ is called sectorial (of angle $\theta \in (0, \pi/2)$) if its resolvent set contains a sector

$$S_{\theta,w} = \{\lambda \in \mathbf{C} \setminus \{w\} : |\arg(\lambda - w)| < \pi/2 + \theta\} \subset \rho(A),$$

and for $\lambda \in S_{\theta,w}$, the estimate

$$\|R(\lambda, A)\| \leq \frac{M}{|\lambda - w|}$$

holds for a constant M independent of λ. Such an operator is automatically closed. The generator of a self-adjoint contraction semigroup on $L^2(I, m)$ is sectorial of angle $\pi/2$ with $w = 0$ and $M = 1$. For sectorial operators, the semigroup may be obtained from the resolvent by the Laplace inversion formula

$$P_t = \frac{1}{2\pi i} \int_\Gamma e^{\lambda t} R(\lambda, A) d\lambda, \tag{2.8}$$

where one picks a $\theta' \in (\pi/2, \pi/2+\theta)$ and defines the contour of integration by

$$\Gamma = \{re^{-i\theta'} : \infty > r \geq 1\} \cup \{e^{i\varphi} : -\theta' < \varphi < \theta'\} \cup \{re^{i\theta'} : 1 \leq r < \infty\} \tag{2.9}$$

In fact this definition allows us to extend the semigroup to $\{t \in \mathbf{C} : |\arg t| < \theta\}$, and one often says that the semigroup (P_t) is analytic (of angle θ). For analytic semigroups, (2.7) holds with equality.

2.3.2 One-dimensional diffusion processes

By a one-dimensional diffusion, we mean a path-continuous strong-Markov process whose state space is an interval $I \subset \mathbf{R}$. We denote its left endpoint l and its right endpoint r, where $-\infty \leq l < r \leq \infty$.

Scale, speed and killing.

In this introduction, we follow Borodin & Salminen [21] very closely. One-dimensional diffusions have three characteristics, namely speed measure, scale function and killing measure. Following Blumenthal & Getoor [19], it is standard in Markov process theory to adjoin to the state space a *cemetery state* $\Delta \notin I$, and allow the process X to be killed at some time $\zeta \in [0, \infty]$, called its lifetime, $\zeta = \inf\{t : X_t \notin I\}$, at which point the process is transported instantaneously to the cemetery state. The killing measure k of our diffusion is associated with the distribution of the location of X at its lifetime:

$$P^x[X_{\zeta-} \in A; \zeta < t] = \int_0^t ds \int_A k(dy) p_m(s; x, y)$$

where p_m is the density of the transition kernel with respect to the speed measure.

Let us suppose first for the ease of exposition that the killing measure is zero. The scale function S is chosen so that the scaled diffusion $S(X)$ is a continuous local martingale on the scaled state-space $S(I)$. S is a continuous strictly increasing function, a homeomorphism $S : I \to S(I)$. If $X_0 = x$, the probability that X hits $b > x$ before $a < x$ is

$$P^x\{H_b < H_a\} = \frac{S(x) - S(a)}{S(b) - S(a)}.$$

The scale function being strictly increasing induces a measure, the scale measure also denoted S, on I. For $x \in [a, b]$ the probability above is $S[a, x]/S[a, b]$.

It is a well-known result[8] of Dambis, Dubins and Schwarz that every continuous local martingale is a time-change of Brownian motion. Write $\tilde{X}_t = S(X_t)$, $t \geq 0$. Then $\tilde{X}_t = B(\gamma_t)$ for a random time-change γ. To get the time-change, we put

$$A_t = \int_{S(I)} l_t^x \tilde{m}(dx)$$

for the speed measure $\tilde{m}$ of $\tilde{X}$ and the Brownian local time (l_t^x), and then let

$$\gamma_t = \inf\{s : A_s \geq t\}$$

be the right-continuous inverse of A. $\tilde{m}$ is a measure on $S(I)$; the speed measure of X is then

$$m(a, b) = \tilde{m}(S(a), S(b)), \quad a, b \in I.$$

[8] Klebaner [77] **Theorem 7.37**, Rogers & Williams [97] (I.(2.3)), IV.34, Kallenberg [72] **Theorem 18.4**.

The speed measure m is invariant with respect to the transition function (P_t). Thus

$$mp_t := \int_I m(dx)p_t(x,\cdot) = m(\cdot).$$

For this reason the transition operators are self-adjoint in $L^2(I,m)$ and m may also be referred to as the symmetry measure.

Boundary classification

We allow non-zero killing again from here. Each of the boundary points l and r is classified according to the behaviour of the process when it is near that boundary. The terminology is motivated by the notion of a diffusing particle in I, its position at time t being denoted X_t. We use notation from Karlin & Taylor [73].

Let X be a regular diffusion on I with scale measure S, speed measure m and killing measure k. Let $\xi \in (l,r) = I^\circ$ be an arbitrary but fixed point of the interior of the state space. The classification is independent of the choice of ξ.

Let

$$\Sigma(l) = \int_l^\xi S(l,x]\,(m(dx)+k(dx)), \quad \Sigma(r) = \int_\xi^r S[x,r)\,(m(dx)+k(dx)). \tag{2.10}$$

We say that the boundary e is *exit* if $\Sigma(e) < \infty$. Obviously a sufficient condition that e not be exit is that S be infinite in neighbourhoods of e. If S is finite in neighbourhoods of an endpoint e, then we call e *attracting*. Thus a non-attracting endpoint is never exit, and attracting boundaries may or may not be exit.

Likewise, define

$$N(l) = \int_l^\xi S[x,\xi]\,(m(dx)+k(dx)), \quad N(r) = \int_\xi^r S[\xi,x]\,(m(dx)+k(dx)). \tag{2.11}$$

We say that the boundary e is *entrance* if $N(e) < \infty$. Using Fubini's theorem, one can deduce that if the speed measure m is infinite in neighbourhoods of an endpoint e, then that endpoint is not entrance.

The boundary classification is as follows. The boundary $e \in \{l,r\}$ is said to be:

- **regular** if e is both entrance and exit, i.e.

$$\Sigma(e) < \infty, \quad N(e) < \infty;$$

- **exit-not-entrance** if

$$\Sigma(e) < \infty, \quad N(e) = \infty;$$

- **entrance-not-exit** if

$$\Sigma(e) = \infty, \quad N(e) < \infty;$$

- **natural** if it is neither entrance nor exit, i.e.

$$\Sigma(e) = N(e) = \infty.$$

SDEs

The calculus becomes much easier to perform when we are given an SDE. In what follows, we show how to proceed from knowing the infinitesimal parameters of our diffusion process (i.e. b, σ and c below) to the scale, speed and killing measures.

The solution X of the general time-homogeneous stochastic differential equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad t > 0, \quad X_0 = x, \tag{2.12}$$

is a diffusion. A sufficient condition that (2.12) makes sense is given in Klebaner [77], Theorem 5.3 (note that this is for one-dimensional SDEs only), that b satisfies a Lipschitz condition and σ satisfies a Hölder condition of order α, with $\alpha \geq 1/2$. That is, there exists constants K_b and K_σ such that

$$|b(x) - b(y)| \leq K_b|x - y|, \quad \text{for all } x, y \in I \tag{2.13}$$

and

$$|\sigma(x) - \sigma(y)| \leq K_\sigma|x - y|^\alpha, \quad \text{for all } x, y \in I, \text{ where } \alpha \geq 1/2. \tag{2.14}$$

The existence of a strong solution of (2.12) is guaranteed by the following theorem[9].

Theorem 5 (Yamada-Watanabe)**.** *Let $b, \sigma : \mathbf{R} \to \mathbf{R}$ be functions satisfying (2.13) and (2.14) respectively. Then equation (2.12) has a strong solution and it is unique.*

[9] Klebaner [77], **Theorem 5.3**.

In addition to the infinitesimal mean b and standard deviation (or volatility) σ, the diffusion should have specified an infinitesimal killing rate $c : I \to \mathbf{R}_+$. The speed, scale and killing measures are assumed to be absolutely continuous with respect to Lebesgue measure, i.e.

$$m(dx) = m(x)dx,\ k(dx) = k(x)dx,\ S(x) = \int_\xi^x s(y)dy \text{ for some } \xi \in I.$$

(The choice of $\xi \in I$ may be made arbitrarily.)

First, the generator A may be defined on the subset $C_c^\infty(I)$ of its domain by

$$Au(x) = \frac{1}{2}\sigma^2(x)u''(x) + b(x)u'(x) - c(x)u(x), \quad u \in C_c^\infty(I), x \in I \tag{2.15}$$

To determine scale, speed and killing densities, one takes

$$B(x) = \int_\xi^x \frac{2b(y)}{\sigma^2(y)}dy.$$

Then the scale density s is

$$s(x) = e^{-B(x)}, \tag{2.16}$$

the speed density is

$$m(x) = \frac{2}{\sigma^2(x)s(x)} \tag{2.17}$$

and the killing density is

$$k(x) = c(x)m(x). \tag{2.18}$$

Note that S satisfies

$$\frac{1}{2}\sigma^2(x)\frac{d^2S(x)}{dx^2} + b(x)\frac{dS(x)}{dx} = 0$$

while m enjoys the adjoint property

$$\frac{1}{2}\frac{d^2}{dx^2}(\sigma^2(x)m(x)) - \frac{d}{dx}(b(x)m(x)) = 0.$$

The latter is another statement that m is an invariant measure when $c = 0$. In fact every invariant measure is a multiple of m, so that the diffusion has a unique invariant distribution precisely if $m(I) < \infty$. In this case, the measure

$$\pi = \frac{m}{m(I)}, \quad \pi(dx) = \pi(x)dx = \frac{m(x)dx}{\int_I m(y)dy} \tag{2.19}$$

is its unique invariant distribution and, for any initial distribution μ (of X_0)

on I, as $t \to \infty$,

$$X_t \xrightarrow{d} \pi$$

A diffusion with an invariant probability distribution is positive recurrent as a Markov process.

Sturm-Liouville problems and Green functions

The Sturm-Liouville (SL) equation is

$$Au(x) = \lambda u(x) \tag{2.20}$$

Observe at this point that the generator may be written in formally self-adjoint form, so that the SL equation is

$$\frac{1}{m(x)}\left[\frac{d}{dx}\left\{\frac{1}{s(x)}\frac{du(x)}{dx}\right\} - k(x)u(x)\right] = \lambda u(x)$$

It has two independent fundamental solutions, determined uniquely up to a multiplicative constant. The first is increasing and the second decreasing, denoted respectively ψ_λ and χ_λ. If a boundary point is regular, then the definition is not unique, and requires specification of a boundary condition, of ψ_λ at l and of χ_λ at r. By taking our derivatives with respect to the *scale measure*, the Wronskian here is defined as

$$w_\lambda = \frac{d\psi_\lambda(x)}{s(x)dx}\chi_\lambda(x) - \psi_\lambda(x)\frac{d\chi_\lambda(x)}{s(x)dx} \tag{2.21}$$

and is independent of x.

The Green function is the Laplace transform of the transition density:

$$G_\lambda(x,y) = \int_0^\infty e^{-\lambda t} p_m(t;x,y)dt.$$

As a consequence of this definition, it is the integral kernel of the resolvent operator:

$$R(\lambda, A)u(x) = (\lambda - A)^{-1}u(x) = \int_I u(y)G_\lambda(x,y)m(y)dy.$$

According to classical theory of differential equations, the Green function is

$$G_\lambda(x,y) = \begin{cases} w_\lambda^{-1}\psi_\lambda(x)\chi_\lambda(y), & x \le y \\ w_\lambda^{-1}\psi_\lambda(y)\chi_\lambda(x), & y \le x. \end{cases} \tag{2.22}$$

The transition operators are then given by Laplace transform inversion

$$P_t u(x) = \frac{1}{2\pi i}\int_\Gamma e^{\lambda t} R_\lambda u(x)\, d\lambda = \frac{1}{2\pi i}\int_\Gamma e^{\lambda t}\int_{-\infty}^{\infty} u(y) G_\lambda(x,y) m(y) dy\, d\lambda, \tag{2.23}$$

and the transition density by

$$p(t;x,y) = \frac{1}{2\pi i}\int_\Gamma e^{\lambda t} G_\lambda(x,y) m(y) dy, \tag{2.24}$$

where Γ is a contour to the right of the spectrum of A. From our development of the theory, it is natural to use the contour given by (2.9), but this is not the only possibility: since $\sigma(A)$ is contained in the real negative half-axis $(-\infty, 0]$, this contour may be deformed. The classical Laplace transform theory will suggest the Bromwich contour

$$\mathcal{B} = \{\gamma + is : -\infty < s < \infty\}, \tag{2.25}$$

but computational considerations will usually lead us to use Talbot's method, which involves the contour

$$\mathcal{C} = \{\lambda(\theta) = r\theta(\cot(\theta) + i) : -\pi < \theta < \pi\} \tag{2.26}$$

where $r > 0$ is a parameter. These contours and methods are defined and examined in Cohen [31] and Davies [35]. See also LePage [80] for the Bromwich inversion integral.

2.3.3 Linetsky's spectral classification

Linetsky [82, 83] has used the classification of Sturm Liouville problems of Fulton et al. [54] to provide a spectral classification of one-dimensional SDEs. Because of the constraints on the generator of our diffusion, there are only three categories that may arise, and they depend on the nature of the solutions of the SL equation at the boundary points. Recall that $\sigma(A) \subset (-\infty, 0]$, when A is considered as an operator in $L^2(I, m)$, because (a) A being self-adjoint, the spectrum is real, and (b) we also have $s(A) \leq \omega_0 \leq 0$ from the analytic theory of strongly continuous contraction semigroups.

We say that the SL equation $(A-\lambda)u = 0$ is *oscillatory* at the boundary $e \in \{l, r\}$ if every solution u has infinitely many zeros in every neighbourhood of e, otherwise we call it *non-oscillatory* at e. We similarly describe the endpoint e as being non-oscillatory or oscillatory. This boundary classification then depends on λ. The Liouville transformation we introduce later will classify

the non-oscillatory/oscillatory nature of each of the boundary points for all values of $\lambda \in \mathbf{R}$. There are two possibilities[10]:

(i) The SL equation is non-oscillatory at e for all $\lambda \in \mathbf{R}$;

(ii) There exists a *cut-off point* $-\Lambda \leq 0$ such that the SL equation is non-oscillatory at e for $\lambda > -\Lambda$ and oscillatory at e for $\lambda < -\Lambda$.

In the latter case, the SL equation may be either oscillatory or non-oscillatory for $\lambda = -\Lambda$, except in the case $\Lambda = 0$, because the SL equation is always non-oscillatory for $\lambda = 0$. We refer in case (ii) to the boundary e as *oscillatory with cut-off* $-\Lambda$. Any combination of the classifications for the two endpoints is possible, so there are three possibilities, namely that both boundaries may be non-oscillatory, that one may be non-oscillatory for all λ and the other oscillatory below a cut-off $-\Lambda$ or that both boundaries are oscillatory below their own cut-off points $-\Lambda_l$ and $-\Lambda_r$. In Feller's boundary classification given above, it is only the natural boundaries which may be oscillatory with a finite cut-off $-\Lambda > -\infty$.

Linetsky's spectral classification[11] is as follows:

Theorem 6. *Let X be a one-dimensional diffusion on the interval I with boundaries $l < r$, with infinitesimal generator $(A, D(A))$. Then (X_t) belongs to precisely one of the following:*

- **Spectral Category I.** *Both endpoints are non-oscillatory for all real λ and $\sigma(A) = \sigma_d(A)$ consists[12] only of an infinite sequence of simple eigenvalues, accumulating at $-\infty$.*

- **Spectral Category II.** *One endpoint is non-oscillatory for all real λ, and the other is oscillatory with cut-off $-\Lambda$. Then $\sigma_e(A) \subset (-\infty, -\Lambda]$ and $\sigma_d(A) \subset [-\Lambda, 0]$. The discrete part of the spectrum is finite if the oscillatory endpoint is non-oscillatory when $\lambda = -\Lambda$, and otherwise consists of countably infinitely many simple eigenvalues accumulating at $-\Lambda$.*

- **Spectral Category III.** *The two endpoints are oscillatory, with cut-offs $-\Lambda_l$ and $-\Lambda_r$. We take $\underline{\Lambda} = \Lambda_l \wedge \Lambda_r$ and $\overline{\Lambda} = \Lambda_l \vee \Lambda_r$. The spectrum is simple in $(-\overline{\Lambda}, 0]$ and has multiplicity two in $(-\infty, -\overline{\Lambda})$. Furthermore $\sigma_d(A) \subset [-\underline{\Lambda}, 0]$. There are finitely many eigenvalues in this interval if and only if the equation is non-oscillatory for $\lambda = -\underline{\Lambda}$. If there are*

[10] Linetsky [82] **Theorem 1**, [83] **Theorem 3.1**

[11] Linetsky [82] **Theorem 2**, [83] **Theorem 3.2**

[12] We adopt the convention of partitioning the spectrum $\sigma(A)$ into its *discrete* $\sigma_d(A)$ and *essential* $\sigma_e(A)$ components. The discrete spectrum consists of eigenvalues of finite multiplicity, and the essential spectrum consists of every other spectral value. For differential operators, all eigenvalues have finite multiplicity.

infinitely many eigenvalues (i.e. the equation is oscillatory for one of the endpoints when $\lambda = -\underline{\Lambda}$) then they cluster at $-\underline{\Lambda}$.

Spectral Category I is obviously the simplest. Suppose that we have a diffusion which does fall into Spectral Category I. Recall that the Green function is given by (2.22); as a function of λ it may be analytically continued to a complex-valued function on $\rho(A) \subset \mathbf{C}$. In light of the straightforward structure of the spectrum, we see that this function (of λ) possesses simple poles at each of the simple eigenvalues of A, which form a decreasing sequence $(-\lambda_n)$, where $0 \leq \lambda_n < \lambda_{n+1} \uparrow \infty$. These poles occur at the zeros of the function $\lambda \mapsto w_\lambda$. The eigenvalue of smallest absolute value may or may not be zero. Now we may observe that (2.23) reduces via the Cauchy residue theorem to

$$P_t u(x) = \sum_n e^{-\lambda_n t} (\varphi_n, u)_m \varphi_n(x), \quad u \in L^2(I, m),\, x \in I,\, t \geq 0, \tag{2.27}$$

where φ_n is the unit eigenfunction corresponding to $-\lambda_n$, and $(\cdot, \cdot)_m$ denotes the inner product in $L^2(I, m)$. The transition density (with respect to Lebesgue measure) may then be written in the form

$$p(t; x, y) = \sum_n e^{-\lambda_n t} \varphi_n(x) \varphi_n(y) m(y), \quad x, y \in I,\, t \geq 0.$$

This does not provide us with a representation of the transition operator and density in closed form because it is, in general, difficult to find the eigenvalues $-\lambda_n$ exactly and difficult to normalise the eigenfunctions φ_n.

For spectral categories II and III, the eigenfunction expansion (2.27) must be replaced by a more general *spectral expansion*, in which an integral replaces the sum below the first cut-off point $-\underline{\Lambda}$.

The Liouville transformation

The following form of the transformation of the SL equation is taken from Linetsky [83]. Assume that the coefficients σ and b are twice continuously differentiable. By a change of the independent variable then a change of the dependent variable, one may transform the SL equation (2.20) to the so-called *Liouville normal form* (LNF). For historical reasons, one usually works with the positive operator $-A$, and we shall also make the simplifying assumption that there is no killing. Thus we start from the SL equation written in the form

$$-Au(x) = -\frac{1}{m(x)} \frac{d}{dx} \left\{ \frac{u'(x)}{s(x)} \right\} = \lambda u(x)$$

The Liouville normal form for our SL equation is

$$-\frac{1}{2}v''(z) + Q(z)v(z) = \lambda v(z), \tag{2.28}$$

with no first order term, and the second order term always has constant coefficient[13] $-1/2$. The operator

$$\frac{1}{2}\frac{d^2}{dz^2} - Q(z)$$

may, by comparison with (2.15), be thought of as the generator of a standard Brownian motion on $z(I)$ killed at the rate Q. The change of independent variable is the *Lamperti transform* z (see Iacus [64]), chosen to satisfy

$$z'(x) = \frac{1}{\sigma(x)},$$

and the change of the dependent variable is

$$u(x) = v(z(x))\sqrt{(\sigma(x)s(x))}$$

In (2.28) the potential function is given by $Q(z) = U(x(z))$, where $x(z)$ is the inverse of the Lamperti transform, and

$$U(x) = \frac{1}{8}(\sigma'(x))^2 - \frac{1}{4}\sigma''(x)\sigma(x) + \frac{1}{2}\frac{b^2(x)}{\sigma^2(x)} + \frac{1}{2}b'(x) - \frac{b(x)\sigma'(x)}{\sigma(x)}$$

We shall refer to the function Q as the *Schrödinger potential.*

Recall that only a natural boundary may be oscillatory. Linetsky [82] **Theorem 3** and [83] **Theorem 3.3** tell us the following:

Theorem 7. *Suppose that the boundary $e \in \{l, r\}$ is a natural boundary. Then the oscillatory/non-oscillatory nature of e is unchanged under the Liouville transformation, and:*

(i) If the Lamperti transform takes e to a finite point $z(e) \in \mathbf{R}$, then that endpoint is non-oscillatory for all real λ;

(ii) If $z(e)$ is infinite, and $U(e) = +\infty$, then e is non-oscillatory for all real λ. If instead $U(e) = \Lambda$, then e is oscillatory with cut-off $-\Lambda$. If $\Lambda = 0$, then e is non-oscillatory for $\lambda = -\Lambda = 0$. If $\Lambda > 0$ and $\lim_{x\to e} z^2(x)(U(x) - \Lambda) > -1/4$ then e is non-oscillatory for $\lambda = -\Lambda$, while if $\Lambda > 0$ and $\lim_{x\to e} z^2(x)(U(x) - \Lambda) < -1/4$, then e is oscillatory for $\lambda = -\Lambda$.

[13] In physics, one often chooses this coefficient equal to one. Then the operator on the left is then the generator of a Brownian motion, running twice as fast as a standard Brownian motion, in the potential Q.

2.4 Some definitions from the general theory of stochastic processes

Here we introduce some definitions for the general theory of stochastic processes. This material is used only to give the definitions that allow the statement of the theorems in section 4.6 and is not used elsewhere in this thesis.

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ be a filtered probability space[14]. The time-parameter set is the set $\mathbf{R}_+ = [0, \infty)$. A stochastic process is a measurable map $X : \Omega \to S^{\mathbf{R}_+}$ into the set of S-valued functions $x : \mathbf{R}_+ \to S$, but we may also be interpret X as a mapping of the set $\mathbf{R}_+ \times \Omega$ into the state space S. We take $S = \mathbf{R}$ in this investigation. However, we additionally assume that our stochastic processes are regular, that is, that $X(\omega)$ is right-continuous with left-hand limits or left-continuous with right-hand limits. The *predictable* σ-field $\mathcal{P}$ is defined as that generated by the left-continuous adapted processes. The *optional* σ-field $\mathcal{O}$ is defined as that generated by the right-continuous adapted processes. One has $\mathcal{P} \subset \mathcal{O}$, meaning that every predictable process is optional. The standard Poisson process is an example of an optional process that is not predictable.

A process X is said to be integrable if $\sup_t EX_t < \infty$. A property of a stochastic process is said to hold locally if there exists a sequence of stopping times τ_n such that $\tau_n \to \infty$ a.s. and that property holds for the stochastic process $(X_{t \wedge \tau_n})$ stopped at τ_n for every n. The Doob-Meyer decomposition[15], states the following: a local submartingale X has an almost surely unique decomposition $X_t = M_t + A_t$, where M is a local martingale and A is a locally integrable increasing, predictable process. The process A is called the compensator of X. Notice that, if a process A has locally finite variation, and its variation is locally integrable, then it may be decomposed into two locally integrable increasing processes, to which the Doob-Meyer decomposition may be applied. Thus a locally finite variation process A of locally integrable variation also possesses a compensator. The compensator of a process A is a predictable process $\tilde{A}$ such that $A - \tilde{A}$ is a local martingale.

A semimartingale is a process X of the form

$$X_t = X_0 + M_t + A_t$$

where M is a local martingale and A has locally finite variation, both starting

[14] It is always assumed that the filtration $(\mathcal{F}_t)$ is right-continuous, but not necessarily complete.
[15] Kallenbergy [72] **Theorem 25.5**.

at 0. If the variation of A is locally integrable, then we may write

$$X_t = X_0 + M_t + (A_t - \tilde{A}_t) + \tilde{A}_t = X_0 + M'_t + \tilde{A}_t$$

which expresses the decomposition into a local martingale M' and a predictable locally finite variation process $\tilde{A}$. In other words the process A in the decomposition of the semimartingale X is predictable. In this case we call X a *special* semimartingale. Thus X is a special semimartingale if A may be chosen predictable, and this occurs if the variation of A is locally integrable. A sufficient condition for X to be a special semimartingale is that the jumps of A are bounded, $|\Delta A_t| \leq a$ for some fixed $a > 0$. Note that, by the Doob-Meyer decomposition, every local submartingale is a special semimartingale.

Now let us consider the local martingale part M of X. Two local martingales M and N are said to be orthogonal if their product MN is a local martingale. A local martingale is said to be *purely discontinuous* if it is orthogonal to every continuous local martingale. This does not mean that it equals the sum of its jumps, but if M is a purely discontinuous local martingale of locally finite variation, then it equals the *compensated* sum of its jumps. A general result of the theory states[16] that any local martingale M has an almost surely unique decomposition $M_t = M_0 + M_t^c + M_t^d$ where M^c is a continuous local martingale and M^d is a purely discontinuous local martingale. Since the decomposition is essentially unique, we call the process M^c the *continuous martingale part* of M. For a semimartingale X with the decomposition given above, this decomposition applies to the local martingale part M, so that X has a continuous martingale part, written X^c.

If M is a local martingale, the process M^2 is a local submartingale and its compensator is the process denoted $\langle M \rangle$. $\langle M \rangle$ is called the predictable quadratic variation of M. The quadratic variation of M is the process $[M]$, where

$$[M]_t = \lim_{n \to \infty} \sum_{k=1}^{n} (M(t_k^n) - M(t_{k-1}^n))^2$$

where the limit is in probability and is taken over finite partitions $0 = t_0^n < t_1^n < \cdots < t_n^n = t$ of $[0, t]$ which satisfy $\max(t_k^n - t_{k-1}^n) \to 0$ as $n \to \infty$. This process does exist (indeed even so if M is only a semimartingale) and is increasing. The compensator of $[M]$ is $\langle M \rangle$.

In order to define the characteristics of a semimartingale, there is one final concept to explain, compensators of random measures. For our purposes, it suffices to consider random measures on $\mathbf{R}_+ \times \mathbf{R}$, where $\mathbf{R}_+$ is the time-

[16] See Kallenberg [72] **Theorem 26.14**.

parameter set and $\mathbf{R}$ is the state space. If N is a random measure on this space, then a measure-valued process is formed by taking $N_t = N([0,t] \times \cdot)$ and considering the process (N_t). (N_t) takes values in the measurable space $\mathcal{M}(\mathbf{R})$ of σ-finite measues on $\mathbf{R}$, which is endowed with the σ-field generated by the projections $\pi_B : \mu \mapsto \mu(B)$ for $B \in \mathcal{B}(\mathbf{R})$. The random measure N is predictable if (N_t) is predictable, and adapted if (N_t) is adapted. Finally, we call N integrable if $N([0,\infty) \times \mathbf{R})$ is an integrable random variable. A general result[17] of the theory is that if N is a locally integrable, adapted random measure then there is a predictable random measure $\tilde{N}$ on $\mathbf{R}_+ \times \mathbf{R}$, called the compensator of N, such that $E[\int W dN] = E[\int W d\tilde{N}]$ for every predictable process $W \geq 0$ on $\mathbf{R}_+ \times \mathbf{R}$.

Now it is time to motivate the concept of characteristics of a semimartingale. These generalise the Lévy-Khintchine characteristics of a Lévy process. For technical reasons, we introduce a truncation function h, which is a bounded function which agrees with the identity (i.e. $h(x) = x$) in a neighbourhood of zero. The classical choice is $h(x) = x \cdot 1_{[-1,1]}(x)$, but for limit theorems h is often chosen continuous instead. However, in the discussion below, we shall assume this form for h. Now the distribution of a Lévy process (X_t) starting at zero is characterised by constants $\mu \in \mathbf{R}$, $\sigma > 0$ and a measure ν concentrated on $\mathbf{R} \setminus \{0\}$, subject to the condition

$$\int_{\mathbf{R}\setminus\{0\}} (1 \wedge x^2)\nu(dx) < \infty,$$

via

$$E\left[e^{i\xi X_t}\right] = e^{\psi_t(\xi)}, \qquad \text{for all } \xi \in \mathbf{R},$$

where

$$\psi_t(\xi) = ib_t\xi - \frac{1}{2}c_t\xi^2 + \int_{\mathbf{R}\setminus\{0\}} (e^{i\xi x} - 1 - i\xi h(x))\nu_t(dx).$$

with $b_t = \mu t$, $c_t = \sigma^2 t$ and $\nu_t(dx) = \nu(dx)t$, and where h is a truncation function. It follows that $e^{i\xi X_t - \psi_t(\xi)}$, is a martingale. The idea behind characteristics of a semimartingale is this: replace the functions b_t, c_t and the measure-valued function ν_t above with processes (B_t), (C_t) and a random measure N on $\mathbf{R}_+ \times \mathbf{R}$, so that $e^{i\xi X_t - \psi_t(\xi)}$ is a martingale. Of course, it will only be possible to choose deterministic functions

$$B_t = \mu t, \qquad C_t = \sigma^2 t \qquad \text{and } N([0,t] \times dx) = \nu(dx)t$$

if the process X is a Lévy process. For a general semimartingale X, one

[17] Kallenberg [72] **Theorem 25.22**.

requires B, C and N to be predictable processes.

For an example, let X be the diffusion process which arises as the solution of a stochastic differential equation, say

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \quad t \geq 0, \qquad X_0 = x,$$

where the coefficients b and σ satisfy the conditions given in Theorem 5. Define a new process Y by

$$dY_t = dX_t - b(X_t)dt = \sigma(X_t)dW_t,$$

which, as a stochastic integral with respect to the Brownian motion W, is a local martingale. By considering the complex stochastic exponential[18] of the process $i\xi Y$, we therefore see that candidates for the characteristics of X are

$$\begin{aligned} B_t &= \int_0^t b(X_s)ds, \\ C_t &= \int_0^t \sigma^2(X_s)ds, \\ N &= 0. \end{aligned}$$

The details of defining the characteristics of a semimartingale X are too involved to go through here: they may be found in Jacod & Shiryaev [68] section II.2a, to which the reader is referred if he or she is looking to make rigorous the following discussion. Roughly speaking, B is the finite variation part of the process $X(h)$ that is formed from X by subtracting its large jumps. The large jumps are defined by

$$\check{X}(h)_t = \sum_{s \leq t} [\Delta X_s - h(\Delta X_s)]$$

and we take

$$X(h)_t = X_t - \check{X}(h)_t,$$

so that $X(h)$ is a special semimartingale: this means that $X(h)_t = X_0 + M(h)_t + B(h)_t$ for a local martingale $M(h)$ and a finite variation predictable process $B(h)$, and the first characteristic is given by $B = (B(h)_t)$. The second characteristic C is the predictable quadratic variation of the continuous martingale part of (X_t), usually written $C_t = \langle X^c \rangle_t$ (in the example above, for our diffusion X the quadratic variation was continuous, implying it is predictable). The predictable random measure N is the compensator of the random measure

[18] See for example Kallenberg [72] **Lemma 18.1**.

μ^X associated with the jumps[19] of X. N and C do not depend on the choice of h, while B does.

Furthermore, the modified characteristics may be defined by taking

$$X(h)_t = X_0 + M(h)_t + B(h)_t$$

where M is a local martingale and $B(h)$ is the first of the characteristics of X. Then instead of $C_t = \langle X^c \rangle_t$, put

$$\tilde{C}_t = \langle M \rangle_t.$$

The modified characteristics of a semimartingale form the triple $(B, \tilde{C}, N)$. Note that $\tilde{C}$, like B, now depends on the choice of truncation function.

2.5 Legendre functions

The Legendre functions arise as the solution of certain second-order ordinary differential equations related to the hypergeometric equation. We require them as solutions of the Sturm-Liouville equation associated with the arithmetic-geometric hybrid semigroup studied in Chapter 5. The associated Schrödinger potential is a case of *modified Pöschl-Teller potential*, see Grosche & Steiner [56] (6.2.3). The properties of these functions that we need, we take mostly from Erdélyi et al. [43] Chapter 3, but many facts and formulae are also in Abramowitz & Stegun [2] and the Wolfram functions site [112].

Legendre's differential equation is

$$(1-z^2)w''(z) - 2zw'(z) + \left[\nu(\nu+1) - \frac{\delta^2}{1-z^2}\right] = 0.$$

Solutions are the Legendre functions

$$P_\nu^\delta(z) = \frac{1}{\Gamma(1-\delta)} \left(\frac{z+1}{z-1}\right)^{\delta/2} {}_2F_1\left(-\nu, \nu+1; 1-\delta; \frac{1-z}{2}\right)$$

and

$$\begin{aligned} Q_\nu^\delta(z) = \; & e^{\delta i\pi} 2^{-\nu-1} \sqrt{\pi} \tfrac{\Gamma(1+\nu+\delta)}{\Gamma(\nu+3/2)} z^{-\nu-\delta-1} (z^2-1)^{\delta/2} \\ & {}_2F_1\left(\tfrac{\nu+\delta}{2}+1, \tfrac{\nu+\delta+1}{2}; \nu+3/2; \tfrac{1}{z^2}\right). \end{aligned}$$

In fact, with these definitions, the functions $P_\nu^{\pm\delta}(\pm z)$, $P_{-\nu-1}^{\pm\delta}(\pm z)$, $Q_\nu^{\pm\delta}(\pm z)$ and $Q_{-\nu-1}^{\pm\delta}(\pm z)$ are all solutions of Legendre's differential equation.

[19] This is from Jacod & Shiryaev [68] II.1.16: $\mu^X(\omega; dt, dx) = \sum_s 1_{\{\Delta X_s(\omega)\neq 0\}} \delta_{(s, \Delta X_s(\omega))}(dt, dx)$.

For this investigation, we need only the values $P_\nu^\delta(z)$, for $z = x \in (-1, 1)$. The definition above works for the whole complex plane, with branch cuts $(-\infty, -1)$ and $(-1, 1)$, and the Legendre function on the cut $(-1, 1)$ is defined by the averaging formula

$$P_\nu^\delta(x) = \frac{1}{2}[e^{\delta i\pi/2} P_\nu^\delta(x + i0) + e^{-\delta i\pi/2} P_\nu^\delta(x - i0)],$$

so as to obtain real values. Thus the definition we use is

$$P_\nu^\delta(x) = \frac{1}{\Gamma(1-\delta)} \left(\frac{1+x}{1-x}\right)^{\delta/2} {}_2F_1\left(-\nu, \nu+1; 1-\delta; \frac{1-x}{2}\right),$$

in agreement with Erdélyi et al. [43] 3.4(6). These are sometimes referred to as Legendre polynomials, although that term is at other times reserved for the case $\delta = 0$ and $\nu \in \mathbf{Z}_+$.

2.5.1 Differentiation

We have

$$\frac{dP_\nu^\delta}{dx}(x) = \frac{(\nu+1)xP_\nu^\delta(x) - (\nu-\delta+1)P_{\nu+1}^\delta(x)}{1-x^2} = \frac{-\nu x P_\nu^\delta(x) + (\nu+\delta)P_{\nu-1}^\delta(x)}{1-x^2},$$

from Erdélyi et al. [43] 3.8(19).

2.5.2 Asymptotic behaviour

Since we are working on the cut only, we need to consider the behaviour of $P_\nu^\delta(x)$ as $x \to \pm 1$. We also shall find that we are only interested in $\delta < 0$. Erdélyi et al. [43] 3.9(8) and 3.9(14) tell us that

$$P_\nu^\delta(x) \sim \tfrac{2^{\delta/2}(1-x)^{-\delta/2}}{\Gamma(1-\delta)}, \qquad x \to 1,\ \delta \neq 1, 2, \ldots$$

$$P_\nu^\delta(x) \sim \tfrac{2^{-\delta/2}\Gamma(-\delta)(1+x)^{\delta/2}}{\Gamma(1+\nu-\delta)\Gamma(-\nu-\delta)} \quad x \to -1,\ \mathrm{Re}\delta < 0.$$

2.6 Confluent hypergeometric functions

Confluent hypergeometric functions are special functions which arise as solution of ordinary differential equations. We require them for solutions of the Sturm-Liouville equation associated with the generator of the hybrid arithmetic-CIR semigroup.

The confluent hypergeometric equation, also called Kummer's equation, is

$$zw'' + (c - z)w' - aw = 0$$

where a and c are constants. Two independent solutions are given in the case $c \notin \mathbf{Z}$ by

$$M(a, c, z) = \Phi(a, c; z) = {}_1F_1(a; c; z)$$

and

$$\begin{aligned} U(a, c, z) &= \Psi(a, c; z) \\ &= \frac{\pi}{\sin(\pi c)} \left\{ \frac{M(a, c, z)}{\Gamma(1 + a - c)\Gamma(c)} - z^{1-c} \frac{M(1 + a - c, 2 - c, z)}{\Gamma(a)\Gamma(2 - c)} \right\} \end{aligned}$$

Properties of these functions are studied in Erdélyi [43] and Slater [104], and listed in Abramowitz & Stegun [2]; see also the wolfram functions site [112]. We shall refer to the function $M(a, c, z)$ as Kummer's function, and to $U(a, c, z)$ as Tricomi's function. They have the integral representations

$$M(a, c, z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c - a)} \int_0^1 e^{zt} t^{a-1} (1 - t)^{c-a-1} \, dt, \; \operatorname{Re} b > \operatorname{Re} a > 0,$$

$$U(a, c, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1 + t)^{c-a-1} \, dt, \; \operatorname{Re} a > 0, \operatorname{Re} z > 0$$

In discussing these functions (and only in this context), the following notation is standard.

$$(a)_0 = 1$$

$$(a)_1 = a$$

$$(a)_n = \frac{\Gamma(a + n)}{\Gamma(a)} = a(a + 1) \dots (a + n - 1), \qquad a \in \mathbf{C}, n \in \mathbf{Z}_+$$

2.6.1 Analyticity

The confluent hypergeometric function of the first kind $M(a, c, z)$, is an entire function of both a and z for $c \notin -\mathbf{Z}_+$. For fixed values of a and z, the function $c \mapsto M(a, c, z)$ has simple poles at each of the points $c = -n$ $(n = 0, 1, 2, \cdots)$. The regularised version

$${}_1\tilde{F}_1(a, c, z) := \frac{M(a, c, z)}{\Gamma(c)}$$

is an entire function of a, c and z. $U(a,c,z)$ is the confluent hypergeometric function of the second kind. It is an entire function of a, c, and z, even defined and finite as $c \to n$, $n \in \mathbf{Z}$.

2.6.2 Differentiation

The following differential calculus holds for Kummer's and Tricomi's functions.

$$\frac{d}{dz}M(a,c,z) = M'(a,c,z) = \frac{a}{c}M(a+1,c+1,z).$$

$$\frac{d}{dz}U(a,c,z) = U'(a,c,z) = -aU(a+1,c+1,z).$$

For indefinite integration, the following relations (see Slater [104], (3.2.7)-(3.2.12)) are also helpful:

$$e^{-z}z^{c-1}U(a,c,z) = -\frac{d}{dz}\left[e^{-z}z^{c}U(a+1,c+1,z)\right], \tag{2.29}$$

$$z^{c-1}U(a,c,z) = \frac{1}{c-a}\frac{d}{dz}\left[z^{c}U(a,c+1,z)\right], \quad a \neq c, \tag{2.30}$$

$$e^{-z}U(a,c,z) = -\frac{d}{dz}\left[e^{-z}U(a,c-1,z)\right]. \tag{2.31}$$

2.6.3 Wronskian

For functions u and v, denote $W\{u,v\} = -W\{v,u\} = uv' - u'v$, called the Wronskian. Then[20] with $U(z) = U(a,c,z)$ and $M(z) = M(a,c,z)$, we have

$$W\{U,M\} = \frac{\Gamma(c)}{\Gamma(a)}e^{z}z^{-c}.$$

2.6.4 Behaviour for large $|z|$

From Abramowitz & Stegun [2], as $|z| \to \infty$

$$M(a,c,z) = \frac{\Gamma(c)}{\Gamma(a)}e^{z}z^{a-c}\left(1+O(|z|^{-1})\right)$$

$$U(a,c,z) = z^{-a}\left(1+O(|z|^{-1})\right) \tag{2.32}$$

We see that, when $a > 0$, the linear space spanned by Tricomi's function contains every solution of Kummer's equation which is zero at $z = +\infty$.

[20] Abramowitz & Stegun **13.1.22**

2.6.5 Laguerre polynomials

Let m denote a (not necessarily normalised) Gamma(c, ρ) distribution on $(0, \infty)$, with pdf

$$m(x) = m_0 e^{-\rho x} x^{c-1}.$$

Then a sequence of (unnormalised) orthogonal polynomials in the Hilbert space $H = L^2([0, \infty), m)$ is given by

$$\varphi_n(x) = L_n^{(c-1)}(x) = \sum_{k \le n} \frac{(c+k)_{n-k}(-x)^k}{(n-k)!k!}.$$

The polynomial $L_n^{(c-1)}$ is referred to as a generalized Laguerre polynomial. They are special cases of the confluent hypergeometric functions, by the following relations: (Abramowitz & Stegun [2] formulae **13.6.9** and **13.6.27**)

$$M(-n, c, x) = \frac{n!}{(c)_n} L_n^{(c-1)}(x), \tag{2.33}$$

$$U(-n, c, x) = (-1)^n n! L_n^{(c-1)}(x). \tag{2.34}$$

2.7 Risk measures

Intuitively speaking, a risk measure is a statistic (number) associated with a distribution, which gives an idea of how risky an investment with that return (or relative loss) distribution is. For elementary examples, (minus) expectation and variance are both risk measures. Other examples are value-at-risk (VaR) and expected shortfall (ES). Our aim here is only to provide the means for calculating VaR and ES within the models we develop. We do not intend to enter the debate as to what constitutes a good risk measure. The term *coherence* is applied to risk measures satisfying certain properties as set out in Artzner et al. [5, 6] (and discussed below), but we are interested in this only so that we know certain elementary properties of VaR and ES. Risk measures are well introduced and studied in Jondeau et al. [71] and Klugman et al. [78].

Consider a real linear space $\mathbf{L}$ consisting of random variables which could represent the return obtained on an investment; the space $\mathbf{L}$ should include all degenerate random variables $X = \mu \in \mathbf{R}$. A risk measure is a functional $\rho : \mathbf{L} \to \mathbf{R}$ such that $\rho(X)$ is determined by the distribution of X. Such functionals exist: elementary examples are the expectation and the variance functionals[21]. The quantity $\rho(X)$ is then meant to describe in some way the

[21] Comparison with the expectation functional E is the reason we have chosen to define ρ on a space of random variables rather than a space of probability distributions on $\mathbf{R}$.

risk associated with taking a position which gives a return of X. In order to fix the terminology, we list four desirable properties for risk measures to have. Here, X and Y denote arbitrary elements of $\mathbf{L}$.

- *monotonicity* $X \leq Y$ implies $\rho(X) \geq \rho(Y)$.
- *subadditivity* $\rho(X+Y) \leq \rho(X) + \rho(Y)$.
- *positive homogeneity* If $\sigma > 0$ then $\rho(\sigma X) = \sigma\rho(X)$.
- *translation invariance*[22] If $\mu \in \mathbf{R}$, then $\rho(X+\mu) = \rho(X) - \mu$.

A risk measure satisfying all four is said to be *coherent*, see Artzner et al. [5]. Subadditivity means that diversification cannot result in increased risk.

We now define the two risk measures most used in finance. The questions of how (or even whether) to use these measures in finance have been debated elsewhere, but mathematically speaking these are useful summary statistics of any probability measure on $\mathbf{R}$.

Value-at-risk (VaR)

Let X be a random variable with a distribution F. If $\alpha \in (0,1)$ is a fixed probability (confidence level), we may define the quantile function

$$F^{-1}(\alpha) = \inf\{x \in \mathbf{R} : F(x) \geq \alpha\}.$$

The value-at-risk at the confidence level α of X (or of F) is the quantile

$$\mathrm{VaR}_\alpha(X) = \xi_\alpha = -F^{-1}(\alpha). \tag{2.35}$$

It is easily checked that VaR is monotone, positive homogeneous and translation invariant. It is not subadditive, so VaR is not a coherent risk measure in the sense of Artzner et al. [5].

Expected shortfall (ES)

Expected shortfall has been variously termed tail value-at-risk, conditional value-at-risk and tail conditional expectation. It has been shown (Artzner *et al.* [5]) that it is a coherent risk measure. At the confidence level $\alpha \in (0,1)$, ES is the conditional expectation of X given that the relative loss $(-X)$exceeds

[22] Of course, the risk measure is not invariant under translations! Unfortunately this terminology is standard.

the α-level value-at-risk ξ_α. That is,

$$\mathrm{ES}_\alpha(X) = E[-X \mid X < -\xi_\alpha] = \frac{E[-X; X < -\xi_\alpha]}{P\{-X < \xi_\alpha\}}. \tag{2.36}$$

One may compute this quantity with any of the formulae

$$\mathrm{ES}_\alpha(X) = \begin{cases} -\frac{1}{1-\alpha}\int_{-\infty}^{-\xi_\alpha} xF(dx) \\ \\ \xi_\alpha + \int_{-\infty}^{-\xi_\alpha} \frac{F(y)}{1-\alpha} dy \\ \\ \frac{1}{1-\alpha}\int_0^{1-\alpha} \xi_u du, \end{cases} \tag{2.37}$$

where F is the distribution (function) of X. The first and third equations are obtained by direct integration, whilst the second is obtained from the first by a simple application of Fubini's theorem. The second relation expresses the difference between VaR and ES as the *mean excess loss* above the VaR (see Klugman *et al.* [78]).

Given a return model (X_t), VaR and ES are functions of the probability level α and the time horizon t, as stated in Hull [63]. This is because the distribution of X_t is a (measure-valued) function of the time t.

2.8 Model fitting and parameter estimation

When attempting to use these diffusion models, the following problem is likely to arise. Suppose that a time-series $(Y_n)_{0\le n\le N} = (X_{n\Delta})_{0\le n\le N}$, where $\Delta > 0$ is a fixed known time-lag, of $N+1$ observations of the process X taken from the proposed model, is given. We would like to be able to be able to estimate the parameters. The problem of testing the model's fit to the given data is related and also of interest.

For the parametric estimation, we assume a given model

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t. \tag{2.38}$$

Note that throughout, the functions depend now also on a parameter $\theta \in \Theta \subset \mathbf{R}^K$. If the transition densities are $p(t; x, y, \theta)$, then the likelihood function is

$$L_N(\theta, (Y_n)) = \prod_{n=1}^{N} p(\Delta; Y_{n-1}, Y_n, \theta),$$

where $\Delta > 0$ is assumed known and $Y_0 = 0$. The log-likelihood is

$$\ell_N(\theta, (Y_n)) = \sum_{n=1}^{N} \log(p(\Delta; Y_{n-1}, Y_n, \theta)).$$

and the score function is defined as

$$S_N(\theta, (Y_n)) = \partial_\theta \ell_N(\theta, (Y_n)) = \sum_{n=1}^{N} \frac{1}{p(\Delta; Y_{n-1}, Y_n, \theta)} \partial_\theta p(\Delta; Y_{n-1}, Y_n, \theta)$$

where ∂_θ denotes the gradient operator $(\partial_{\theta_1}, \ldots, \partial_{\theta_K})$. This is a vector-valued function, taking values in $\mathbf{R}^K$ where K is the number of dimensions in the parameter space. If the transition densities are known in closed form, then so is the score function, and the optimal estimate of θ is obtained by solving

$$S_N(\theta) = 0.$$

In all but the most elementary models (2.38), the transition densities are not known in closed form. Approaches which have been developed to circumvent this problem are based on numerical approximation of the transition density, and constructing *estimating functions* to mimic the properties of the score function.

2.8.1 Static estimation

The crudest approximation of the transition density is simply the invariant distribution (2.19). To use this, we suppose that

$$p(\Delta; x, y, \theta) = \pi(y - x, \theta), \qquad x, y \in \mathbf{R},$$

with π from (2.19), write down the likelihood function

$$L(\theta, (Y_n)) = \prod_{n=1}^{N} \pi(Y_n - Y_{n-1}, \theta) = \prod_{n=1}^{N} \pi(\Delta Y_n, \theta),$$

and maximise over θ. An equivalent formulation supposes that the increments

$$\Delta Y_n = Y_n - Y_{n-1}, \qquad n \geq 1.$$

(for example daily returns) are i.i.d. π, and uses a maximum likelihood estimation.

The technique of maximum likelihood tends to favour a model with more free parameters. This can result in an *overfitting* of the model to the data, so

that we fit the model to noise rather than the underlying trend. Intuitively we understand that this is likely to happen in cases where the number of observations is not large enough compared with the dimension of the parameter space. Fortunately for us, there are a few *information criteria* which have been introduced in statistics to quantify the goodness of fit of the model, taking into account its number of free parameters. We mention two of them: the Akaike information criterion (AIC) and the Schwarz (or Bayesian) information criterion. These are defined respectively as

$$\begin{array}{ll} \mathrm{AIC} = -2\log(L)/N + 2k/N & \text{(Akaike)} \\ \mathrm{BIC} = -2\log(L)/N + 2k\log(N)/N & \text{(Schwarz)} \end{array} \tag{2.39}$$

where L denotes the likelihood, N denotes the number of observations and k denotes the number of free parameters in the model. Of these two, BIC penalises models for having more parameters more than AIC.

2.8.2 Estimating functions

The simplest construction of an estimating function is to define functions $h_1, \ldots, h_M$ which have conditional mean zero, i.e. for each $m = 1, \ldots, M$,

$$E_\theta^x h_m(Y_0, Y_1, \theta) = \int_{\mathbf{R}} h_m(x, y, \theta) p(\Delta; x, y, \theta) dy = 0, \qquad x \in \mathbf{R},$$

and weight functions $a_1, \ldots, a_M : \mathbf{R} \times \Theta \to \mathbf{R}^K$. Thus $a(Y_n, \theta)$ is a $K \times M$ matrix and

$$f(x, y, \theta) = a(x, \theta)h(x, y, \theta) = \sum_{m=1}^{M} a_m(x, \theta)h_m(x, y, \theta)$$

defines an $\mathbf{R}^K$-valued function. The estimating function is defined by

$$F_N(\theta, (Y_n)) = \sum_{n=1}^{N} f(Y_{n-1}, Y_n, \theta),$$

which is a martingale. Then we have

$$E_{\theta_0} F_N(\theta, (Y_n)) = 0, \text{ if and only if } \theta = \theta_0.$$

where θ_0 is the true value of θ. Thus we estimate θ by solving

$$F_N(\theta, (Y_n)) = 0.$$

Most progress on the problem of parameter estimation in diffusion models from discrete observations has been made relatively recently. To gain an overview of the subject, the survey by H. Sørensen [105] is extremely useful, summarising the techniques which have been explored. For implementation of these techniques for one-dimensional diffusion processes, see Iacus [64].

Chapter 3

Feedback models

We now begin the procedure of specifying a probabilistic model of the financial asset price. At this first stage, the intention is to look at certain aspects of the market microstructure which drive the market state as it changes over time. We will write down the various components of the change in the market state, drawing attention to their dependencies so as to obtain a mathematical expression of the market dynamics.

For simplicity, we assume that a single number X_t is enough to tell us the current state of the market. Fixing a time $t \geq 0$, we define the (log-) return

$$X_t = \log\left(\frac{S_t}{S_0}\right), \tag{3.1}$$

where S_t denotes the asset price at time t. $X = (X_t)$ is a real-valued stochastic process, and the state of the market at time t is X_t. X_t is the return on an investment made at time 0. Note that, as long as the price change $S_t - S_0$ is sufficiently small relatively to S_0, the log-return (3.1) approximates the simple return

$$R_t = \frac{S_t - S_0}{S_0},$$

This is the Taylor development of (3.1) up to order 1.

3.1 A renewal-reward trade order model

In our feedback models, the market state X_t is driven by the arrival of buy and sell orders to the market. Let us fix the present time $t = 0$ and consider the next trade which is to arrive at the market. There are obviously two random quantities we are interested in knowing, namely the duration of the waiting period between now and the next trade order and the size of the trade order.

By modelling the size of the order as a $(-\infty, \infty)$-valued random variable, this quantity then also tells us whether it is a buy order or a sell order (sell orders are negative). Of course, an order may not be of size 0.

Consider then a time interval $[0, T]$ or $[0, \infty)$ of market activity. For simplicity we shall suppose that the waiting durations between successive trades forms an i.i.d. sequence. Then let (τ_k) be an i.i.d. sequence of holding (waiting) times, so that trade orders arrive at the successive time instants

$$T_1 = \tau_1, \qquad T_k = T_{k-1} + \tau_k, \quad k \geq 2.$$

The sequence (T_n) is an increasing random walk so that only finitely many T_n can appear in any bounded trading period $[0, t]$. Its occupation measure

$$N(B) = \#\{T_n \in B : n \in \mathbf{N}\}$$

is a *renewal process*, and its (random) cumulative distribution function is defined as

$$N_t = N[0, t] = \sum_{n \geq 1} 1_{[0,t]}(T_n). \tag{3.2}$$

(also often referred to as a renewal process). The random walk (T_n) is referred to, in this context, as a renewal sequence.

Attached to the nth trade, which occurs at time T_n, is the trade size. We shall suppose again that the successive sizes form an i.i.d. sequence, this time denoted (R_n), and shall further suppose that the two sequences (τ_n) and (R_n) are independent of one another. The aggregate order in any time interval $I \subset [0, \infty)$ is

$$D(I) = \sum_{T_n \in I} R_n = \sum_{n \geq 1} R_n 1_I(T_n).$$

D is not a measure, since it must be permitted to take negative values (in neighbourhoods of sell orders). D might be viewed mathematically as an atomic signed random measure. We let

$$D_t = D[0, t] = \sum_{n=1}^{N_t} R_n = \sum_{n \geq 1} R_n 1_{[0,t]}(T_n). \tag{3.3}$$

be the *renewal-reward process*, telling us the aggregate trade order in $[0, t]$.

3.2 Choice of inter-trade waiting time and trade order size distributions

Thanks to the technique of using an invariance principle for the diffusion approximation below, actual specification of the *reward* (trade size) and *transition* (waiting duration) distributions is unnecessary, further than assuming the finiteness of variance, for the models we are to study. However, for modelling asset returns by other processes which might include jumps, such as Lévy processes or jump diffusions, it will be necessary to have some sort of understanding of these two distributions. The invariance principle will not apply and the final mathematical expression for the market state dynamics will still depend upon the chosen distributions.

The most natural choice of inter-trade waiting time distribution F_τ is the exponential distribution. Indeed, our renewal and renewal-reward processes are not Markovian unless the holding times are exponential. If T_n are assumed exponentially distributed, with mean λ^{-1} say, then (N_t) is a Poisson process with intensity λ. The demand process (D_t) is then a compound Poisson process, the most tractable of renewal-reward processes. Unfortunately, it is found in empirical studies that the exponential distribution does not fit inter-trade data well. We mention the investigations of Mainardi et al. [85], Politi & Scalas [91], Sazuka et al. [100] and Scalas et al. [101]. In those papers, the authors conclude that a much better fit is provided by Weibull and Mittag-Lefler distributions. Jiang et al. [70] also suggest a Weibull distribution provides a good fit for the inter-trade distribution in their analysis of trading in 23 liquid Chinese stocks listed on the Shenzhen Stock Exchange in 2003. The study of Cartea & Meyer-Brandis [4] notes investigates the impact of the duration between trades on the security price process and on option prices. They find that the volatility smile arises as a consequence of incorporating duration between trades in the model.

For the distribution F_R of trade sizes, empirical investigations largely ignore the buy-sell property of the trade orders and concentrate only on absolute value. Therefore, an appropriate choice may well be a Bernoulli mixture of a buy-order distribution concentrated on $(0, \infty)$ and a sell-order distribution concentrated on $(-\infty, 0)$. It appears from the studies of Mu et al. [89] and Queirós [94] that the distribution of best fit for trading volumes is an F- or beta prime distribution. These are Type VI in the Pearson classification.

3.3 Feedback model

Having summarised the basic ingredients of the market microstructure model, we turn our attention to the price formation process. Let us suppose that the state of the market X_t is known at some time $t > 0$ and that it will change at time $t + \Delta t$, by a quantity $\Delta X_t = X_{t+\Delta t} - X_t$. We previously asserted that a price change occurs because of an imbalance between buy and sell orders. Mathematically, this means that we may express the change in the market state over the trading-period $[t, t + \Delta t)$ as a function of the aggregate demand over this period ΔD_t, i.e.

$$\Delta X_t = \mathcal{I}(\Delta D_t),$$

where the *impact function* $\mathcal{I}$ is to be specified. The only of its properties we know for certain are that it should be non-decreasing, so that when demand exceeds supply the price is pushed upwards, and $\mathcal{I}(0) = 0$, so that when demand and supply match the price does not move.

Let us explain why the quantities of the asset being bought or sold in this market is important to the price formation. The following is based on the discussion in Jondeau et al. [71], section 3.1. In an order-driven market, traders enter their orders into an electronic trading system. Orders may be of various types, and we consider here only *limit orders* and *market orders*. A trader placing a limit order would indicate his or her willingness to trade (buy or sell) a certain quantity at a certain price. In some markets, the trader may also provide additional information such as a time-limit after which he or she would no longer be willing to trade that quantity at that price. The collection of all the limit orders forms the order book. A market order would execute the transaction immediately at the best available prices. For example, if the highest limit bid orders are for 10, 25 and 15 units at prices 35.0, 34.5 and 34.0 respectively, and a market sell order is made for 20 units, then the trader placing the market order will sell 10 units at a price of 35.0, and a further 10 units at a price of 34.5. If instead a market sell order is made for 40 units, they will sell 10 units at a price of 35.0, 25 units at a price of 34.5 and the remaining 5 units at 34.0. We see that a greater number or volume of sell orders results in a more noticeable fall in the transaction price, while large numbers or volumes of buy orders result in large increases in the transaction price.

We opt for mathematically the simplest model, that of a linear price impact function, i.e. $\mathcal{I}(x) = kx$ for some $k > 0$. In terms of the order book, we attempt a partial justification as follows. The actual price adjustments that occur do so

at the execution of market orders. When a market order is executed, it might leave a 'hole' in the order book (that is, there are fewer limit orders remaining in the system); however, we assume that this market is sufficiently liquid that more limit orders are placed as soon as the market order was executed and the structure of the order book remains intact. The impact function $\mathcal{I}$ is then just a description of the order book, and we suppose that it is approximately linear.

The actual structure of the order book is probably some sort of staircase function, but investigations into its structure consider possible smooth approximations to it. Kempf & Korn [75] observe from intraday data on German index futures that the impact function is nonlinear, stating 'Large orders lead to relatively small price changes whereas small orders lead to relatively large price changes'. This means a sublinear impact function, examples of which are $(kx)^\alpha$ with $0 < \alpha < 1$. Zhang [114] suggests a square root function, i.e. $\mathcal{I}(x) = (kx)^{1/2}$, while Lillo et al. [81] find that $\mathcal{I}(x) \propto x^\alpha$, where the exponent α changes - it is close to 0.5 for small volumes and decreases to about 0.2 for larger volumes.

In any case, for our model we have chosen

$$\Delta X_t = k \Delta D_t$$

for some fixed $k > 0$. It remains to incorporate the effects of fundamental analysis and technical analysis in ΔD_t. The excess demand due to fundamentally motivated trades during the trading period of interest is $\Delta D_F(t)$ and that due to the technically motivated trades is $D_T(t)$. The overall excess demand is then $\Delta D_t = \Delta D_F(t) + \Delta D_T(t)$. Writing R_k for the fundamental trade order sizes and W_k for the technical trade order sizes,

$$\begin{aligned} \Delta D_F(t) &= \sum_{k=N_t+1}^{N_{t+\Delta t}} R_k; \\ \Delta D_T(t) &= \sum_{k=N_t+1}^{N_{t+\Delta t}} W_k. \end{aligned}$$

The difference between the above expressions is that the distribution of the technical trade order size W_k is assumed to depend directly on X_t, the current market state. When the renewal process (N_t) is stationary, we have

$$E[\Delta D_F(t)] = \mu_F \Delta t, \qquad \text{var}\left(\Delta D_F(t)\right) = \sigma_F^2 \Delta t$$

for suitably chosen constants $\mu_F \in \mathbf{R}$ and $\sigma_F > 0$, and

$$E[\Delta D_T(t)] = -\mu_T(X_t)\Delta t, \qquad \text{var}\left(\Delta D_F(t)\right) = \sigma_T^2(X_t)\Delta t$$

likewise. The minus sign is included without any loss in generality, since then an increasing choice of function μ_T indicates the traders' employment of a mean-reverting strategy. It is assumed that the technical traders do not trade when there is no movement of the price, so that $\mu_T(0) = \sigma_T(0) = 0$. Using our linear price impact model, we obtain

$$E[\Delta X_t] = k\mu_F\Delta t - k\mu_T(X_t)\Delta t$$

and

$$\text{var}\left(\Delta X_t\right) = k^2\left(\sigma_F^2 + 2r\sigma_F\sigma_T(X_t) + \sigma_T^2(X_t)\right)\Delta t,$$

where $r \in [-1, 1]$ is the correlation between $\Delta D_F(t)$ and $\Delta_T(t)$, the effects over the time-period of interest of fundamental trading and technical trading respectively. Finally, we write

$$\Delta X_t = (\mu_1 - f(X_t))\Delta t + \sigma_1\Delta W_t^1 + g(X_t)\Delta W_t^2, \tag{3.4}$$

where

- $\mu_1 = k\mu_F$,
- $\sigma_1 = k\sigma_F$,
- $f(x) = k\mu_T(x)$,
- $g(x) = k\sigma_T(x)$,
- $f(0) = g(0) = 0$,
- ΔW_t^1 and ΔW_t^2 are random variables with mean 0 and variance Δt.

3.4 Immediate price changes

To summarise the model above, we have allowed trades to accumulate over a short time period $[t, t+\Delta t)$ and then adjusted the price according to excess buy or sell orders. This description is useful in that it leads very transparently to the discrete equation (3.4), which in turn suggests an obvious continuous-time analogue. However, it is also useful mathematically to consider the possibility in which an arriving trade order causes an immediate price fluctuation proportional to the order size. The resulting process is then of pure-jump type and is

right-continuous with respect to the discrete topology on $\mathbf{R}$. Such a process is expressed more easily than the one considered earlier, in terms of the random variables from which it is constructed.

Suppose then that the return X_t is adjusted at the moment a new trade order arrives. Assume for a first illustration that the market is composed entirely of fundamentalists. Then the first trade order arrives at some time $T_1 = \tau_1$ and has size R_1. Since these quantities never depend on the current state of the market, the process (X_t) is just a renewal-reward process

$$X_t = \sum_{n=1}^{N_t} kR_n,$$

where (N_t) is the renewal process associated with the i.i.d. holding times sequence (τ_n) (and k is the price impact constant).

Now assume that some of the trades are technically motivated. Suppose that the process has had its $(n-1)$th jump at time $T_{n-1} = \tau_1 + \cdots + \tau_{n-1}$, and that now we have

$$T_{n-1} \leq t < T_n, \quad X_t = x, \quad A_t = t - T_{n-1}.$$

Then the time T_n of the next jump has some conditional distribution governed by

$$P[\tau_n \in \cdot | \tau_n > A_t] = \frac{F_\tau(x, \cdot)}{1 - F_\tau(x, A_t)}.$$

The jump size also has a distribution dependent on x, which means the sequence (Y_n) of states that X visits, $Y_n = X_{\tau_n}$, is a discrete-time Markov process. The process (X_t) is then a semi-Markov process. It is Markov if the jump times are exponentially distributed. In that case, the time until the next jump never depends on A_t.

The whole process can be constructed from scratch as follows. For each possible state x of the market, there is a waiting-time distribution $F_\tau(x, \cdot)$ and a trade order distribution $F_R(x, \cdot)$. For the given start point, $Y_0 = X_0$, the variables of interest are the first arrival time $\tau_1 \sim F_\tau(Y_0, \cdot)$ and the first order size $R_1 \sim F_R(Y_0, \cdot)$. We define $T_1 = \tau_1$ and $Y_1 = kR_1 + Y_0$. Now given Y_1, we are interested in $\tau_2 \sim F_\tau(Y_1, \cdot)$ and $R_2 \sim F_R(Y_1, \cdot)$. We define $Y_2 = kR_2 + Y_1$ and $T_2 = \tau_2 + T_1$.

Once we know the pair (T_{n-1}, Y_{n-1}), we may obtain the next pair (T_n, Y_n) by letting

$$\tau_n \sim F_\tau(Y_{n-1}, \cdot) \quad \text{and} \quad R_n \sim F_R(Y_{n-1}, \cdot)$$

and then letting $Y_n = kR_n + Y_{n-1}$ and $T_n = \tau_n + T_{n-1}$. The sequence of points (T_n, Y_n) gives the arrivals of the market state process (X_t) at each of its successive states. The whole process is given by

$$X_t = Y_N = X_0 + \sum_{n=1}^{N_t} kR_n, \qquad T_N \leq T_{N+1}. \tag{3.5}$$

where $N_t = \sup\{n \geq 1 : t \geq T_n\}$. (N_t) generalizes the notion of a renewal process by introducing some dependency structure to the sequence (τ_n) of holding times.

We note finally that, if the τ_n are exponentially distributed, then the processes (X_t) and (N_t) are Markov. We use $\rho(x)$ to denote the rate parameter (inverse of the mean) of the holding time τ_n given that X is in the state x. Then the generator of this process is

$$Au(x) = \rho(x) \int (u(y) - u(x)) K(x, dy)$$

where the transition kernel K is related to the trade size kernel via

$$K(x, B) = F_R\left(x, \frac{B - x}{k}\right), \quad B \in \mathcal{B}(\mathbf{R}).$$

Chapter 4

Process approximation

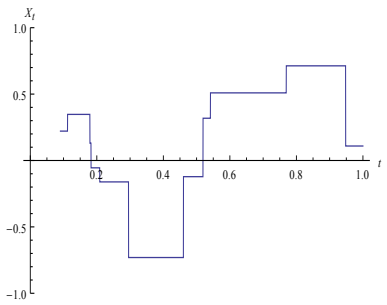
Now that we have described a discrete price formation process, we would like to consider the validity of using a continuous approximation to it. We consider first the most direct approach, which is a continuous approximation to the discrete equation of motion (3.4). Continuous movement may be observed in the limit if the number of trades during any fixed trading period becomes very large while their respective price impacts become very small, which suggests that it would be appropriate to appeal to a functional version of the Central Limit Theorem. We then look at other possible approximating sequences to the same continuous model.

The motivation for a diffusion approximation is that a graph of the step process $(X_t^{(\rho)})$ against time looks from a distance like continuous diffusion. We give a simple illustration in figure 4.1, which shows simulated compound Poisson processes with increasing rates. The renewal-reward process starts to run very quickly and a diffusion (X_t) is obtained by considering a time-and-space scale on which the time between individual trades and their impacts are small. The asymptotic results with $t \to \infty$ are available from renewal theory.

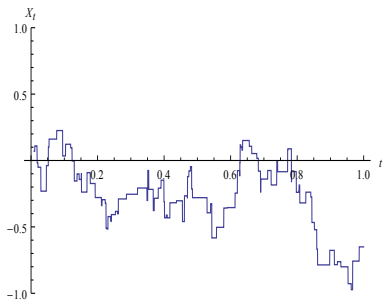
4.1 Direct approximation by the Central Limit Theorem

Equation (3.4) is obviously suggestive of a possible continuous approximation, so we now argue that the continuum limit may be taken as $\Delta t \to 0$, by appealing to a functional version of the Central Limit Theorem. The following is not the only possible direct method of diffusion approximation, but it serves as an example to take us through the issues the approximation presents and leads us to the financial return models we are interested in investigating. We

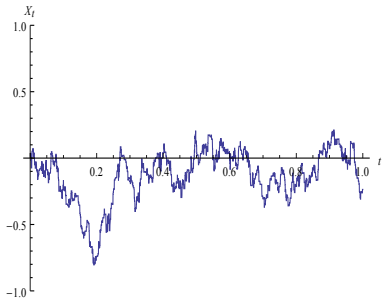
Figure 4.1: Simulated scaled compound Poisson processes with rate ρ and normal increments.



(a) $\rho = 10$.



(b) $\rho = 100$.



(c) $\rho = 1000$.

may then study in more detail the consequences of each modelling assumption on the resulting model, before turning back to rigorize the approximation argument.

First, let $\delta > 0$ denote the slope of the order book, i.e. $\mathcal{I}(x) = \delta x$. Consider first only the fundamental trades and their impact $\Delta_F X_t$ of trading in $[t, t + \Delta t)$. We shall now allow the trade order rate $((E\tau_1)^{-1})$ to depend on δ. This already assumes that $E\tau_1 \in (0, \infty)$, but let us also assume that $\text{var}\,(\tau_1) < \infty$. Small values of δ would be expected with greater liquidity in the market, so we might reasonably expect that the trade order rate increases as $\delta \downarrow 0$. In fact, where we previously had an underlying renewal process (N_t) with arbitrary rate $\lambda > 0$, let us have a renewal process $N_t^{(\rho)} = N_{\rho t}$, where ρ is another parameter which measures liquidity, and is not independent of δ. Our renewal process has a rate $\lambda\rho$ and, by Theorem 1, satisfies the asymptotic expressions[1]

$$EN_t^{(\rho)} \sim \lambda\rho t, \quad \text{var}\,(N_t^{(\rho)}) \sim \sigma_\tau^2 \lambda^3 \rho t, \quad \rho \to \infty.$$

We also allow the trade order size distribution to depend on ρ and δ. The trade sizes need to become small to make the diffusion approximation work,

[1]These asymptotic expressions are from (2.3) and (2.4) which were consequences of the renewal theorem, Theorem 1.

but we also need to be careful that the variance does not become too small, otherwise the limit process will be deterministic. Let us explicitly take

$$ER_1^{(\rho)} = \mu\delta, \quad \text{var}\,(R^{(\rho)}) = \sigma_R^2.$$

To make the diffusion approximation work, suppose that $\rho \to \infty$ and $\delta \to 0$ in such a way[2] that $\delta^2\rho \to 1$ (this needs to be a positive constant; there is no loss in generality in taking this constant to be 1). Then

$$E[\Delta_F X_t^{(\rho)}] = \mu\lambda(\delta^2\rho)\Delta t, \quad \text{var}\,(\Delta_F X_t^{(\rho)}) = [\sigma_R^2 + (\mu\delta\sigma_\tau\lambda)^2]\lambda(\delta^2\rho)\Delta t$$

As $\rho \to \infty$, the number of trades grows large: $N^{(\rho)}(t) \stackrel{d}{=} N(\rho t) \xrightarrow{a.s.} \infty$, and, dropping the (ρ) from our notation, the distribution of the sum

$$\Delta_F X_t = \sum_{n-N_t=1}^{n-N_t=\Delta N_t} \delta R_n$$

is approximately normally distributed according to the central limit theorem. We can therefore approximate the increment $\Delta_F X_t$ by

$$\Delta_F X_t = \mu\lambda\Delta t + \sigma_R\sqrt{\lambda}\,\Delta W_t^1, \tag{4.1}$$

where

$$\Delta W_t^1 = \Delta W_{t+\Delta t}^1 - \Delta W_t^1 \sim N(0, \Delta t)$$

is an increment of a standard Brownian motion in $\mathbf{R}$ over the time interval $[t, t+\Delta t)$. If we treat disjoint intervals as being independent of one another, equation (4.1) may be supposed to hold for all values t for the one Brownian motion $\{W_t^1\}$.

For the technical traders, we again suppose that X has a known value at time t and that it is to be adjusted at time $t + \Delta t$. For the traders using technical analysis, the distributions of the quantities $\Delta N(t)$ and R_k depend upon the value X_t. The most straightforward way to model the change $\Delta_T X_t$ in X due to technical analysis is by taking the corresponding parameters λ_T, μ_T and σ_T to be functions of X_t. Therefore,

$$\Delta_T X_t = \mu_T\lambda_T\Delta t + \sigma_T\sqrt{\lambda_T}\,\Delta W_t^2, \tag{4.2}$$

where W^2 is another Brownian motion, and where we have suppressed the X_t-dependence of the parameters λ_T, μ_T and σ_T in the notation, for readability.

[2]The reader may like to compare this technique with that outlined in Feller [52] XIV.6, particularly the limiting procedure (6.4) there.

The increment of the process X is found by putting together the two components due to fundamental and technical trading,

$$\Delta X_t = \Delta_F X_t + \Delta_T X_t \,.$$

We end up with the discrete hybrid process

$$\Delta X_t = (\mu\lambda + \mu_T\lambda_T)\Delta t + \sigma\sqrt{\lambda}\Delta W_t^1 + \sigma_T\sqrt{\lambda_T}\Delta W_t^2 \tag{4.3}$$

where subscript T (for *technical*) reminds us of dependence on the current level X_t. It is natural to take $\lambda_T = \mu_T = \sigma_T = 0$ when $X_t = 0$.

Next, we let the time length Δt over which the price remains unchanged approach zero. The limit in probability of the sequence of processes satisfying the discrete equation (4.3) is a process X satisfying the continuous equation

$$dX_t = (\mu_1 - f(X_t))dt + \sigma_1 dW_t^1 + g(X_t)dW_t^2, \qquad t > 0, \quad X_0 = x, \tag{4.4}$$

with fundamental parameters $\mu_1 = \lambda\mu$ and $\sigma_1 = \sigma\sqrt{\lambda}$ and technical functions $-f = \mu_T\lambda_T$ and $g = \sigma_T\sqrt{\lambda_T}$, with $f(0) = 0$ and $g(0) = 0$ (since technical excess demand is supposed not to be triggered by zero price-change). Here, f must satisfy condition (2.13), while g should satisfy (2.14), so that (4.4) has a unique strong solution.

4.2 Specific models

The next step in specifying the return model is choosing the functions λ_T, μ_T and σ_T (hence f and g in (4.4)). The simplest models, save the arithmetic model (in which they are all constant in X_t), are the following.

- **Arithmetic-geometric hybrid**

  If the rate of trading does not vary with X_t, but the trade size distribution does with both μ_T and σ_T varying linearly with X_t, say

  $$f(X_t) = \mu_2 X_t, \quad g(X_t) = \sigma_2 X_t \,,$$

  for certain constants μ_2 and σ_2, we reach the model

  $$dX_t = (\mu_1 - \mu_2 X_t)dt + \sigma_1 dW_t^1 + \sigma_2 X_t dW_t^2, \qquad t > 0, \quad X_0 = x, \tag{4.5}$$

  which is a hybrid of arithmetic and geometric Brownian motions.

- **Arithmetic-CIR hybrid**

Supposing the trade size distribution does not depend on X_t, but the rate λ_T of trading depends linearly on X_t, we obtain technical functions of the form

$$f(X_t) = \mu_2 X_t, \quad g(X_t) = \sigma_2 \sqrt{|X_t|}\,,$$

and hence

$$dX_t = (\mu_1 - \mu_2 X_t)dt + \sigma_1 dW_t^1 + \sigma_2 \sqrt{|X_t|} dW_t^2, \qquad t > 0, \quad X_0 = x, \quad (4.6)$$

where μ_1 and σ_2 are constants.

The arithmetic-geometric process was the subject of study in Shaw & Schofield [102].

4.3 Distributional impact of fundamental and technical trading

Now we remark briefly upon the respective rôles in these models of each of its two essential components, the separate pressures exerted on the market state by the fundamentally motivated trades and the technical trades. If we consider the general diffusion model (4.4) that we arrive at, the effects are made explicit in the notation, in that μ_1 and σ_1 are effects of the fundamental trading, while f and g reflect the technical trading. If only the fundamentalists trade, we should make the simplifications $f = g = 0$, and if only the technicians trade, we would take $\mu_1 = \sigma_1 = 0$.

If there is no technical pressure, the process we arrive at via this diffusion approximation is a Lévy process. This would be true by design even if we did not approximate the whole process by this diffusion and left in some of the jumps - the assumptions of this microstructure model still lead us to a spatially homogeneous process. As remarked in the introduction, the process is then characterized by a Lévy-Khintchine triple (μ, σ, ν).

By introducing the technical pressure then, we have sought to remove from that sort of model the assumption of spatial homogeneity. The technical trading brings, at each time $t \geq 0$, a dependence of the next price movements on the present state X_t of the market. Jump-diffusion models (without killing) are described by three components, namely

- a drift coefficient $b : \mathbf{R} \to \mathbf{R}$;
- a diffusion coefficient $a = \sigma^2 : \mathbf{R} \to \mathbf{R}+$;
- a jump kernel $L(x, dy)$.

The triple (b, a, L) generalizes[3] the Lévy-Khintchine triple (μ, σ, ν). At least, this is the most natural generalization of the Lévy processes discussed above to spatially inhomogeneous Markov processes obtainable from our fundamental-technical price feedback model. Our investigation of *hybrid diffusion* models only, leads us, after this Chapter, to consider only models with $L = 0$, but we believe it is important that it is possible to generalize this analysis to situations in which the returns process exhibits jumps. We note from Çinlar[30], section 9.3, that the restriction of the generator of such a process (with only finitely many jumps in any time-period) to, for example $C_c^\infty(\mathbf{R})$, is given by

$$Au(x) = \frac{1}{2}a(x)u''(x) + b(x)u'(x) + \int_{-\infty}^{\infty} L(x, dy)(u(y) - u(x)).$$

This is the sort of process we are aiming at. Consideration of the market microstructure, and of the motivations of fundamentalists and technicians, serves to suggest what sort of form the coefficients a, b and L should take. Our aim in the present Chapter is to bridge that gap between the microstructure concepts and the model specification in terms of these infinitesimal coefficients.

4.4 Use of the theory of weak convergence

Let X be a process satisfying (4.4) and, for each $\rho > 0$, let $X^{(\rho)}$ be the step process constructed to satisfy (3.4) at each of its jump times $t = n\Delta t$. X and $X^{(\rho)}$ are random elements in the space $D = D[0, \infty)$. The validity of using X to approximate $X^{(\rho)}$ is summarised by the statement

$$X^{(\rho)} \xrightarrow{d} X \text{ in } D, \quad \text{as } \rho \to \infty. \tag{4.7}$$

Denoting by F the distribution of (X_t) and by F_ρ the distribution of $(X_t^{(\rho)})$ (4.7) means

$$F_\rho \xrightarrow{w} F \quad \text{as } t \to \infty,$$

where $\xrightarrow{w}$ means convergence in the 'weak' sense of Billingsley [13]. In other words, (4.7) is really a statement not about the random elements $X^{(\rho)}$ and X themselves but rather about their distributions.

Consider the approximation problem when only fundamentalists trade. We aim to manipulate either side of (2.6), so that we can obtain a convergence result of the form

$$(\delta S_{N(nt)})_{t\geq 0} \xrightarrow{d} (\mu t + \sigma W_t)_{t\geq 0}.$$

[3]More properly, the diffusion coefficient a replaces σ^2 in a Lévy process model.

In this expression, we recall that $\delta > 0$ is the price impact constant and that (S_n) gives the successive price levels. Let us recall how we made our previous attempt at this process approximation. The $\mathbf{N}$-valued parameter n is replaced with a real one ρ. We shall allow both the mean 'μ' of $R^{(\rho)}$ and the rate λ of $N^{(\rho)}$ to depend on ρ in a certain way. Finally, we recall that, with the rate of trading increasing to infinity, the market liquidity also increases, so that the *impact* upon the price, which is what we are interested in, is given by $\rho^{-1/2} S_{N(\rho t)}$, rather than the random sum $S_{N(\rho t)}$ itself.

Assume then that we have a sequence of holding times τ_n with positive mean λ^{-1} and finite variance σ_τ^2. Let $(N(t))$ be the renewal process associated with the τ_n. We take $ER_1^{(\rho)} = \mu\rho^{-1/2}$ for a suitable constant μ; then we see that

$$E\left[\rho^{-1/2} S_{N(\rho t)}\right] = E[N(\rho t)]\, E[R_1^{(\rho)}]\rho^{-1/2} \sim \lambda\mu t \quad \text{as } t \to \infty$$

and

$$\begin{aligned} \text{var}\left(\rho^{-1/2} S_{N(\rho t)}\right) &= \rho^{-1}\left\{E\left[\text{var}\left(S_{N(\rho t)} | N(\rho t)\right)\right] + \text{var}\left(E\left[S_{N(\rho t)} | N(\rho t)\right]\right)\right\} \\ &\sim \rho^{-1}\left\{\sigma_R^2 \rho\lambda t + \mu^2 \rho^{-1} \sigma_\tau^2 \lambda^3 \rho t\right\} \\ &\to \sigma_R^2 \lambda t \qquad \text{as } \rho \to \infty. \end{aligned}$$

So by Theorem 4,

$$(\rho^{-1/2} S_{N(\rho t)})_{t \geq 0} \xrightarrow{d} (\lambda\mu t + \sigma_R \sqrt{\lambda} W_t),$$

where the convergence is weak convergence of the distributions on the Skorohod space $D[0, \infty)$.

The final stage of our approximation procedure is to use the outline above for the fundamentally motivated trades, and a spatially inhomogeneous generalization of the same thing for the technicians. We would like to make these parameters λ, μ and σ_R^2 depend on the present state X_t of the market, and make the price adjustments at the points of an increasingly fine mesh $\{n\Delta t : n \in \mathbf{N}\}$. However, it is difficult to argue rigorously that the discrete processes should converge in distribution to a stochastic integral $\int (b(X_s)ds + \sigma(X_s)dW_s)$. We see that extending these arguments to spatial inhomogeneity is not easy, and we shall actually approach it via a different method.

4.5 Approximation by pseudo-Poisson processes

Given the difficulties encountered in the direct approach above, we seek an alternative approach to obtaining the desired weak convergence results. One possible strategy is to use the following result, taken from Billingsley [13], Theorem 3.1.

Theorem 8 (weak convergence of two approximating sequences)**.** *Suppose that* (X_n) *and* (Y_n) *are two sequences of random elements of a metric space* (S, d)*, and that* X *is another random element in* S*. If* $X_n \xrightarrow{d} X$ *and* $d(X_n, Y_n) \xrightarrow{d} 0$*, then* $Y_n \xrightarrow{d} X$*.*

This result may potentially be usefully employed in our context by using the random elements X_n as a bridge between the discrete processes Y_n from our feedback model, satisfying (3.4), and the process X of (4.4).

The candidates for X_n are *pseudo-Poisson processes*. We recall that a pseudo-Poisson process is a process of the form $(S_{N(t)})$ where (S_n) is a discrete-time Markov process and $(N(t))$ is a Poisson process. Unfortunately, these processes are not permitted to arise directly from the feedback model since they have a Poisson process which counts the trade order arrivals, a model empirically refuted in financial markets by the investigations of E. Scalas and others, as previously mentioned. On the other hand, like the process X of (4.4), they are time-homogeneous Markov processes, which means that they possess a generator that may be used to characterise the process. The generator provides a more straightforward description of the infinitesimal characteristics of the process, and our intention is to choose the approximating sequence (X_n) of pseudo-Poisson processes in such a way that we can see their generators converging, in the sense discussed below, to the generator of X.

Here is a plan of our discussion:

(i) We intend to use a result linking strong convergence of the generators with convergence in distribution of the associated Markov processes.

(ii) We give a characterization of pseudo-Poisson processes by their generators.

(iii) Generators of diffusion processes are obtained, for instance in Feller [53], by Taylor development of the transition operator near $t = 0$. We would like to use a similar argument, replacing the transition operator of the diffusion itself with the jump kernel of a pseudo-Poisson process.

For the convergence result at the end, we intend to employ Theorem 9 below[4], but in order to state it, some preparation is needed. It concerns a set-up in which Markov processes are defined, taking values in some locally compact, second countable Hausdorff space S. A *Feller semigroup* is a transition semigroup (of a Markov process) (P_t) which maps $C_0(S)$ into itself, and a Feller process is a process whose transition semigroup is Feller. All the Markov processes that we consider are Feller processes[5]. The Alexandroff compactification S_Δ of S is constructed by taking a point $\Delta \notin S$ and adjoining it to the state space to form $S_\Delta = S \cup \{\Delta\}$. Δ is thought of as the cemetery state, so that a process X which dies at some point $\zeta = \inf\{t : X_t \notin S\}$ is then transported immediately to Δ and stays there for ever. This construction always makes the extended semigroup $(\hat{P}_t)$ conservative in the sense that $\hat{P}_t(x, S_\Delta) = 1$ for all $x \in S_\Delta$ and for all $t \geq 0$. Recall that a core D for a closed operator $(A, D(A))$ is a subset of the domain $D(A)$ which is dense in $D(A)$ with respect to the graph norm.

Theorem 9 (convergence of Feller processes)**.** *Let X^n and X be Feller processes in some locally compact, separable metric state space S with transition semigroups (P^n_t) and (P_t), and generators $(A^{(n)}, D(A^{(n)}))$ and $(A, D(A))$, defined on $C_0(S)$, respectively. Suppose that there is a core D for $(A, D(A))$. Then the following conditions are equivalent:*

(i) For every $u \in D$ there are $u_n \in D(A^{(n)})$ with $u_n \to u$ and $A_n u_n \to Au$.

(ii) $P^n_t \to P_t$ operator strongly for each $t > 0$.

(iii) If $X^n_0 \xrightarrow{d} X_0$ in S then $X^n \xrightarrow{d} X$ in $D(\mathbf{R}_+, S_\Delta)$.

This theorem is stronger than we need. In fact, there are useful pseudo-Poisson processes with bounded generators so that, in the first condition of this theorem, we require only that, for all $u \in D$, $A_n u \to Au$ (since $D(A_n) = C_0(S)$).

The classic text in which pseudo-Poisson processes are defined is Feller [53] Chapter X. Another very clear definition is given in Kallenberg [72], p241. Let us start however, by considering the space-homogeneous case, with the intention of generalizing afterwards. Consider then a space- and time- homogeneous pure-jump Markov process, taking its values in $\mathbf{R}$. It is described by two characteristics:

(i) the rate $\rho > 0$ of jumps;

[4]Kallenberg [72] **Theorem 19.25**

[5]See the discussion at the end of Çinlar [30], Section 9.5.

(ii) the distribution K of each jump size.

By the Markov property, inter-arrival durations are exponentially distributed random variables, and we define ρ as the reciprocal of their means. The underlying renewal process is then a Poisson process (N_t), and the pure-jump process has i.i.d. jumps $R_1, R_2, \ldots$. The overall process is the *compound Poisson process*

$$C_t = \sum_{n=1}^{N_t} R_n.$$

By conditioning on the number of jumps in $[0, t]$, the transition probability of transition from x to $B \subset \mathbf{R}$ can easily be calculated as

$$P_t(x, B) = \sum_{n=1}^{\infty} e^{-\rho t} \frac{(\rho t)^n K^{n*}(B - x)}{n!} \tag{4.8}$$

which exhibits the process as having generator

$$A = \rho(K * -1), \qquad \text{i.e. } Au(x) = \rho \int (u(x - y) - u(x))K(dy). \tag{4.9}$$

A is a bounded operator on $C_0(\mathbf{R})$ and $P_t = e^{tA} = e^{\rho t(K*-1)}$ holds for (4.8) and (4.9) in the usual sense for bounded operators.

To generalize these notions to allow state-space dependency, we allow the rate ρ to depend on x, and replace the jump distribution with a probability kernel K from $\mathbf{R}$ to itself. For our purposes, it suffices that the rate function $\rho : \mathbf{R} \to \mathbf{R}_+$ is assumed bounded. The generator of the resulting process is then

$$Au(x) = \rho(x) \int (u(y) - u(x))K(x, dy) = \int (u(y) - u(x))L(x, dy),$$

where $L(x, dy) = \rho(x)K(x, dy)$ is the *rate kernel* (see Kallenberg [72] p238, or Çinlar [30] Chapter 9, formulas (3.11) and (3.52), where L is called the Lévy kernel for its analogy with the Lévy measure for the jumps of Lévy processes). According to Proposition 12.20 of Kallenberg [72], or the discussion on pp162-164 of Ethier & Kurtz [46], such a process is equivalent to a pseudo-Poisson process, i.e. a process of the form $X_t = Y_{N(t)}$ where (Y_n) is a Markov chain and N is a Poisson process with constant rate $\bar{\rho} = \sup_{x \in \mathbf{R}} \rho(x) > 0$. The homogeneous case discussed above is the special case when Y is a random walk.

Recall that the process X satisfying (4.4) has formal generator

$$Au(x) = \frac{1}{2}a(x)u''(x) + b(x)u'(x).$$

where $a(x) = \sigma_1^2 + 2r\sigma_1 g(x) + g^2(x)$, r being the correlation coefficient between the two Brownian motions, and $b(x) = \mu_1 - f(x)$. We look for generators $\rho(K^{(\rho)} - 1)$ of pseudo-Poisson processes which converge, as $\rho \to \infty$, to A pointwise on a core D of A. The most obvious example is yielded by the definition

$$Au = \lim{}_{\rho\to\infty}\, \rho(P_{1/\rho}u - u)$$

where (P_t) is the transition semigroup of the diffusion X itself; that this is the definition of the generator of (P_t) is seen by taking $\rho = t^{-1}$. Feller [53], in section X.4, defines a diffusion process by the following postulates[6] on the transition function (P_t): for all $\delta > 0$, as $t \to \infty$,

$$\begin{aligned} t^{-1}\textstyle\int_{|y-x|\geq\delta} P_t(x, dy) &\to 0; \\ t^{-1}\textstyle\int_{|y-x|<\delta}(y-x)P_t(x, dy) &\to b(x); \\ t^{-1}\textstyle\int_{|y-x|<\delta}(y-x)^2 P_t(x, dy) &\to a(x). \end{aligned} \tag{4.10}$$

For $u \in C_c^\infty(\mathbf{R})$, he uses the Taylor development of u around x, and the fact that $u^{(3)}$ is bounded:

$$\begin{aligned} \frac{P_t u(x) - u(x)}{t} &= \frac{1}{t}\int (u(y) - u(x))P_t(x, dy) \\ &= \frac{1}{t}\int \left((y-x)u'(x) + \frac{1}{2}(y-x)^2 u''(x) + \frac{1}{6}(y-x)^3 u^{(3)}(\xi)\right) P_t(x, dy) \\ &\to b(x)u'(x) + \frac{1}{2}a(x)u''(x), \quad t \to 0, \end{aligned}$$

where we have denoted by ξ some point lying between x and y. The strong convergence

$$t^{-1}(P_t - 1) \to \frac{1}{2}a(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx} \quad \text{on } D(A),$$

depends on the postulates (4.10), rather than the fact that P_t is the transition operator associated with X. Thus our approach is to look for pseudo-Poisson processes with generator $\rho(K^{(\rho)} - 1)$ satisfying the following three analogous postulates for arbitrary $\delta > 0$:

[6] Feller [53] Chapter X, formulae (4.2)-(4.4); see also Prohorov & Rozanov [93]

(i)

$$\lim_{\rho\to\infty} \rho \int_{(x-\delta,x+\delta)^c} K^{(\rho)}(x, dy) = 0;$$

(ii)

$$\lim_{\rho\to\infty} \rho \int_{x-\delta}^{x+\delta} (y-x) K^{(\rho)}(x, dy) = b(x);$$

(iii)

$$\lim_{\rho\to\infty} \rho \int_{x-\delta}^{x+\delta} (y-x)^2 K^{(\rho)}(x, dy) = a(x).$$

Example: Two-point distributions. Consider a pseudo-Poisson process sitting at x which, at the next arrival of its underlying Poisson process, will jump upwards to a point $u(x)$ with probability $1/2$ or downwards to $d(x)$ with probability $1/2$. By the choice

$$\begin{aligned} u(x) &= x + \rho^{-1} b(x) + \rho^{-1/2}\sigma(x), \\ d(x) &= x + \rho^{-1} b(x) - \rho^{-1/2}\sigma(x), \end{aligned}$$

where $\sigma^2(x) = a(x)$ we find that the distribution

$$K^{(\rho)}(x, \cdot) = \frac{1}{2}(\delta_{u(x)} + \delta_{d(x)})$$

satisfies the conditions (i)-(iii). Note that this distribution has mean $x + \rho^{-1} b(x)$ and standard deviation $\rho^{-1/2}\sigma(x)$ (hence variance $\rho^{-1} a(x)$).

Arbitrary jump distributions. $K^{(\rho)}(x, \cdot)$ satisfies conditions (ii) and (iii) if it has mean $x + \rho^{-1} b(x)$ and variance $\rho^{-1} a(x)$.

We have therefore derived the following.

Convergence Result 1. *Assume that $b : \mathbf{R} \to \mathbf{R}$ and $\sigma = \sqrt{a} : \mathbf{R} \to \mathbf{R}_+$ are functions satisfying the conditions (2.13) and (2.14) of Theorem 5. Let $\rho_0 > 0$ and for each $\rho > \rho_0$, let $(X_t^{(\rho)})$ be a pseudo-Poisson process taking values in $\mathbf{R}$, with $X_0^{(\rho)} = x$, with rate ρ, and with jump probability kernel $K^{(\rho)}(x, dy)$, so that the family $\{K^{(\rho)} : \rho > \rho_0\}$ satisfies (i), (ii) and (iii) above. Then $X^{(\rho)} \xrightarrow{d} X$ in the Skorohod space $D[0, \infty)$, where X is the unique strong solution of the SDE*

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad t > 0, \quad X_0 = x.$$

Example: Convergence of the feedback model.

Let us proceed to show a possible link between the feedback models of section 3.4 and the limit process X in Convergence Result 1 under the elementary (and unrealistic) assumption of exponentially distributed holding times. For the sake of simplicity, we shall also assume that the fundamentally and technically motivated market participants act independently of one another.

Assume then that fundamental trades arrive according to a Poisson process with rate $\theta\rho$, and that the technical trades arrive according to an independent Poisson process with rate $(1-\theta)\rho$, where $\rho > 0$, and where $\theta \in (0,1)$ is fixed. Assume that at some time instant t, the market state is $X_t = x$. Then the Poisson process of trades has rate ρ and the size of the next trade order to arrive has a distribution which is a Bernoulli mixture of F_R, the distribution of the fundamental trade size, and $F_W(x,\cdot)$ the distribution of the technical trade size, with mixing parameter θ (i.e. it is a fundamental trade with probability θ). Let $\mu_1 \in \mathbf{R}$, $\sigma_1 > 0$ and assume that f and g are functions which satisfy the conditions (2.13) and (2.14) respectively. Suppose that the mean and variance of the fundamental and technically motivated trades are as follows:

- The mean fundamental trade size (mean of F_R) is

$$\mu_F = \frac{\mu_1}{\theta\rho^{1/2}}. \tag{4.11}$$

- The mean technical trade size is

$$\mu_T = -\frac{f(x)}{(1-\theta)\rho^{1/2}}. \tag{4.12}$$

- The variance of the fundamental trade size distribution F_R is

$$\sigma_F^2 = \frac{\sigma_1^2}{\theta}. \tag{4.13}$$

- The variance of the technical trade size distribution $F_W(x,\cdot)$ is

$$\sigma_T^2 = \frac{g(x)^2}{(1-\theta)}. \tag{4.14}$$

Then with price impact function $\mathcal{I}(x) = \rho^{-1/2}x$, it can be seen that the distribution $K^{(\rho)}(x,\cdot)$ of the next jump (which we shall denote Y) of $X^{(\rho)}$ satisfies conditions (ii) and (iii) above. In order to show that (i) also holds, it is natural to try applying Markov's inequality. Recall that Markov's inequality states[7]

[7] Kallenberg [72] **Lemma 4.1**.

that, for a nonnegative random variable Z and a constant $\delta > 0$, one has

$$P\{Z \geq \delta\} \leq \frac{E[Z]}{\delta}.$$

Applying this with $Z = (Y - E[Y])^2$ gives

$$\begin{aligned} \rho\, P\{|Y - E[Y]| \geq \delta\} &\leq \rho \cdot \frac{E\left[\left(Y - x - \frac{b(x)}{\rho}\right)^2\right]}{\delta^2} \\ &\to \frac{a(x)}{\delta^2} \quad \text{as } \rho \to \infty \end{aligned}$$

by (ii) and (iii) above (note that the distribution of Y depends on ρ, but this is hidden in the notation) and we proceed no further. The bound on the probability on the left yielded by Markov's inequality is not sharp enough, because it is valid for any probability distribution.

In order for Convergence Result 1 to apply then, we need further assumptions. One possibility is to require that F_R and $F_T(x, \cdot)$ both have a finite standardised kurtosis which is independent of ρ, or else does not increase with ρ. In this case, applying Markov's inequality with $Z = (Y - E[Y])^4$ yields

$$\begin{aligned} \rho\, P\{|Y - E[Y]| \geq \delta\} &\leq \rho \frac{E[(Y - E[Y])^4]}{\delta^4} \\ &= \rho \frac{(\sigma_1 + g(x)^2)^2 \,\mathrm{ku}(Y)}{\rho^2 \delta^4} \\ &\to 0, \quad \text{as } \rho \to \infty. \end{aligned}$$

This condition on the kurtosis is satisfied, for instance, by a normal distribution.

Hence if conditions (4.11) to (4.14) are assumed for every ρ, and for each ρ and x the kurtosis of the distribution $K^{(\rho)}(x, \cdot)$ is non-increasing with ρ, we have convergence to the diffusion which solves the SDE

$$dX_t = (\mu_1 - f(X_t))dt + \sqrt{\sigma_1^2 + g(X_t)^2} dW_t.$$

Jumps in the limiting process.

It is the small jump condition (i) that ensures that the process obtained in the limit $\rho \to \infty$ has continuous paths. Let us therefore see what happens when we no longer assume it. Suppose instead that the jump distribution is a mixture of small and large jumps

$$K^{(\rho)}(x, dy) = \theta S^{(\rho)}(x, dy) + (1 - \theta) L(x, dy)$$

where the large jump component $L(x, dy)$ does not depend on ρ, while the small jump component $S^{(\rho)}$ satisfies conditions (i)-(iii) above. For readability, we wrote θ for the mixing probability, but it actually is permitted to depend upon ρ and x. where the large jump component $L(x, dy)$ does not depend on ρ, while the small jump component $S^{(\rho)}$ satisfies conditions (i)-(iii) above. For readability, we wrote θ for the mixing probability, but it actually is permitted to depend upon ρ and x.

The generator of the resulting pseudo-Poisson process is then

$$\rho(K^{(\rho)}u(x) - u(x)) = \rho\theta(S^{(\rho)}u(x) - u(x)) + \rho(1-\theta)(Lu(x) - u(x)).$$

By looking at the second term on the right here, dependence of θ on ρ is evidently needed; more precisely we shall take

$$\rho(1 - \theta(\rho, x)) = k(x)$$

where $k(x) \geq 0$ is independent of ρ. These definitions make sense when $\rho > k(x)$, and we shall henceforth assume that k is bounded, so that we may take $\rho > k$ uniformly. The jump kernel of the approximating process of rate ρ is then

$$K^{(\rho)}u(x) = \left(1 - \frac{k(x)}{\rho}\right) S^{(\rho)}u(x) + \frac{k(x)}{\rho} Lu(x), \tag{4.15}$$

hence the process itself has generator

$$\rho(K^{(\rho)}u(x) - u(x)) = (\rho - k(x))S^{(\rho)}u(x) + kLu(x) - \rho u(x).$$

Rearranging then letting $\rho \to \infty$, in view of

$$\rho(S^{(\rho)}u - u) \to \frac{1}{2}au'' + bu', \quad u \in C_c^\infty(\mathbf{R}),$$

we obtain

$$\lim_{\rho\to\infty} \rho(K^{(\rho)}u - u) = \frac{1}{2}au'' + bu' + k(Lu - u), \quad u \in C_c^\infty(\mathbf{R})$$

the limit on the left here being taken in the uniform topology on $C_0(\mathbf{R})$. Note that the coefficients a, b and k are functions defined on the state space, and L is a probability kernel from $\mathbf{R}$ to itself.

The process $X^{(\rho)}$ with generator $\rho(K^{(\rho)} - 1)$, where $K^{(\rho)}$ is given by (4.15), has rate of jump arrivals ρ and jump kernel which is a mixture of the small jump kernel $S^{(\rho)}$ and the large jump kernel L. If $X_t^{(\rho)} = x$ at some time t, then the position that the process reaches at its next jump is chosen from

the large jump distribution $L(x,\cdot)$ with probability $k(x)/\rho$ and from the small jump distribution $S^{(\rho)}(x,\cdot)$ with probability $1-k(x)/\rho$. From Theorem 9, we therefore have the following result.

Convergence Result 2. *Assume that $b:\mathbf{R}\to\mathbf{R}$ and $\sigma=\sqrt{a}:\mathbf{R}\to\mathbf{R}_+$ are functions satisfying conditions (2.13) and (2.14). Assume further that $L(x,dy)$ is a probability kernel, that there exists a maximum jump size J so that $L(x,[x-J,x+J])=1$ for all x, and that $k:\mathbf{R}\to\mathbf{R}_+$ is a bounded measurable function. For each $\rho>\sup_x k(x)$, let $X^{(\rho)}$ be a pseudo-Poisson process taking values in $\mathbf{R}$, with $X_0^{(\rho)}=x$, with rate ρ, and with jump kernel $K^{(\rho)}(x,dy)$ of the form (4.15)where $\{S^{(\rho)}\}$ is a family of probability kernels satisfying the conditions (i), (ii) and (iii) above. Then there exists a process X with paths in $D[0,\infty)$, satisfying*

$$X_t = x+\int_0^t b(X_s)ds+\int_0^t \sigma(X_s)dW_s+\int_{s=0}^{s=t}\int_{y\in\mathbf{R}} j(X_{s-},y)M(ds,dy), \qquad t>0$$

where $M(\cdot,\cdot)$ is a standard Poisson random measure on $[0,\infty)\times[0,\infty)$ which is independent of the Brownian motion (W_t), and where, for each $x\in\mathbf{R}$, $\alpha\mapsto j(x,\alpha J)$ is the quantile function of the probability measure $L(x,\cdot)$. Moreover,

$$X^{(\rho)}\xrightarrow{d}X \text{ in } D[0,\infty), \qquad \text{as } \rho\to\infty \text{ in the Skorohod space } D[0,\infty).$$

4.6 Approximation by semi-Markov processes

In this section, we sketch the approach that will hopefully be developed in subsequent investigations and used to prove that (4.4) is the limit of processes of the form (3.5). Given a process of the form (4.4), we are now able to find an approximating sequence $(X^{(\rho)})$ of Markov jump processes with rates ρ. What we would like to be able to do next is use an approximating sequence $(X^{(\rho)})$ of jump processes, which are only semi-Markov because the underlying jump counting process is not Poisson. The semi-Markov process is a time-change of the Markov process that we have already used to approximate the diffusion X. Unfortunately, the approximation by semi-Markov processes has proved just out of reach for the present investigation. For this discussion then, we intend simply to state the two theorems, from Jacod & Shiryaev [68], that are candidates to be applied to the processes $X^{(\rho)}$ and limit X to conclude that $X^{(\rho)}\xrightarrow{d}X$.

In order even to state these two theorems a good deal of preparation is needed. We have included a short exposition of the necessary definitions from

the general theory of stochastic processes in section 2.4. Jacod & Shiryaev's presentation is concerned with semimartingales, as both the approximating processes $X^{(\rho)}$ and the limit processes X. This choice is widely applicable and we note that all the processes that we are considering here are semimartingales. Our preparation for stating these limit theorems consists of the following:

(i) Define the characteristics of a semimartingale.

(ii) Explain what is meant by a solution of the martingale problem, in the setting of a given semimartingale.

(iii) Introduce a class of functions $C_1(\mathbf{R})$.

In developing this theory, we take much material from Jacod & Shiryaev [68]. We then use it subsequently solely to state these convergence results, and it is not needed in the rest of the thesis.

Characteristics of semimartingales

Recall that a semimartingale is an adapted process X that has a decomposition

$$X_t = X_0 + M_t + A_t$$

where M is a local[8] martingale starting at zero and A is a process with locally finite variation also starting at zero. The process $X^{(\rho)}$ defined by (3.5) has, by construction, locally finite variation, hence is a semimartingale. The limit process in (4.4) is also a semimartingale.

Let h be a truncation function (i.e. a bounded function $h : \mathbf{R} \to \mathbf{R}$ which agrees with the identity in a neighbourhood of zero). Define

$$\check{X}(h)_t = \sum_{s \le t} [\Delta X_s - h(\Delta X_s)]$$

and

$$X(h)_t = X_t - \check{X}(h)_t,$$

so that $X(h)$ is a special semimartingale: this means that $X(h)_t = X_0 + M(h)_t + B(h)_t$ for a local martingale $M(h)$ and a finite variation predictable process $B(h)$. The first characteristic of X is $B = (B(h)_t)$, the compensator of $X(h)$. The second characteristic C is the predictable quadratic variation of the continuous martingale part of X, usually written $C_t = \langle X^c \rangle_t$ (in the example above, for our diffusion (X_t) the quadratic variation was continuous, implying

[8] In this context, a property of a process X holds *locally* if there exists a sequence of finite stopping times τ_n such that $\tau_n \to \infty$ a.s and the process $X^{\tau_n} = (X_{t \wedge \tau_n})$ stopped at τ_n has that property.

it is predictable). The predictable random measure N is the compensator of the random measure μ^X associated with the jumps[9] of X. N and C do not depend on the choice of h, while B does.

Furthermore, the modified characteristics may be defined by taking

$$X(h)_t = X_0 + M(h)_t + B(h)_t$$

where $M(h)$ is a local martingale and B is the first of the characteristics of X. Then instead of $C_t = \langle X^c \rangle_t$, put

$$\tilde{C}_t = \langle M(h) \rangle_t.$$

In terms of the characteristics (B, C, N), we have

$$\tilde{C}_t = C_t + \int h(x) N([0,t] \times dx) - \sum_{s \le t} (\Delta B_s)^2.$$

The modified characteristics of a semimartingale form the triple $(B, \tilde{C}, N)$. Note that $\tilde{C}$, like B, now depends on the choice of truncation function.

The Martingale Problem

Consider a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t))$ without a probability measure on it. We insist that the filtration is right-continuous, but there is no notion of completeness for these σ-fields. Let $\mathcal{H}$ be a sub-σ-field of the initial σ-field $\mathcal{F}_0$ and let P_H be a probability measure, called the initial condition, on $(\Omega, \mathcal{H})$. Let $\mathcal{C}$ be a class of $\mathbf{R}$-valued processes that are optional[10] (e.g. all rcll processes are optional). A solution of the martingale problem associated with $(P_H, \mathcal{C})$ is a probability measure P on $(\Omega, \mathcal{F})$ such that $P|_{\mathcal{H}} = P_H$ (i.e. $P(A) = P_H(A)$ for all $A \in \mathcal{H}$) and every $X \in \mathcal{C}$ is a martingale on the filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$.

For a simple example to illustrate the context, take the canonical space $\Omega = C[0, \infty)$ of continuous functions, let $W_t(\omega) = \omega(t)$ for $\omega \in \Omega$, $Y_t = W_t^2 - t$, $\mathcal{C} = \{W, Y\}$ and assume that W is adapted to a right-continuous filtration $(\mathcal{F}_t)$. Then Lévy's characterisation of Brownian motion says that (W_t) is a standard Brownian motion with respect to P if and only if P is a solution of the martingale problem associated with $(P_H, \mathcal{C})$, where $P_H\{X_0 = 0\} = 1$.

For a semimartingale X, we say that a probability measure on $(\Omega, \mathcal{F})$ solves

[9] This is from Jacod & Shiryaev [68] II.1.16: $\mu^X(\omega; dt, dx) = \sum_s 1_{\{\Delta X_s(\omega) \neq 0\}} \delta_{(s, \Delta X_s(\omega))}(dt, dx)$.

[10] The optional σ-field is the σ-field on $\Omega \times \mathbf{R}_+$ generated by the rcll processes when they are considered as mappings of $\Omega \times \mathbf{R}_+$.

the martingale problem associated with $(\mathcal{H}, X)$ and $(P_H; B, C, N)$ if $P|_{\mathcal{H}} = P_H$ and X is a semimartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ with characteristics (B, C, N). The martingale problem associated with $(\mathcal{H}, X)$ and $(P_H; B, C, N)$ is denoted $\mathcal{S}(\mathcal{H}, X|P_H; B, C, N)$. The processes B, C and N are not martingales with respect to P in this set-up, but it turns out that a solution P of this martingale problem is characterised as a solution of the martingale problem associated with $(P_H, \mathcal{C})$ where $\mathcal{C}$ contains the following processes:

(i) $M(h)$, the martingale part in the semimartingale decomposition $X(h)_t = X_0 + M(h)_t + B(h)_t$. (Recall that $X(h)$ is the result of removing the large jumps from X.)

(ii) The process
$$M(h)^2 - \tilde{C}$$
where
$$\tilde{C}_t = C_t + \int_{x \in \mathbf{R}} h(x)^2 N([0,t] \times dx) - \sum_{s \leq t} (\Delta B_s)^2$$

(iii) $g \cdot \mu^X - g \cdot N$, defined by
$$g \cdot \mu^X_t - g \cdot N_t = \int_{\mathbf{R}} g(x) \mu^X([0,t] \times dx) - \int_{\mathbf{R}} g(x) N([0,t] \times dx)$$
where g is bounded, continuous and vanishes inside a neighbourhood of 0. (Recall that μ^X was the random measure associated with the jumps of X.)

Let X be an rcll adapted process on a fixed filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t))$, h a truncation function, B and C predictable processes and N a predictable random measure on $\mathbf{R}_+ \times \mathbf{R}$. The *uniqueness-measurability hypothesis* is that

(i) for each $x \in \mathbf{R}$ the martingale problem[11] $\mathcal{S}(\sigma(X_0), X|\delta_x; B, C, N)$ has a unique solution P^x;

(ii) $x \mapsto P^x(A)$ is a Borel function for all measurable sets A.

Spaces of functions

There is one final technical concept that we must introduce from Jacod & Shiryaev [68], two classes[12] of functions $C_1(\mathbf{R})$ and $C_2(\mathbf{R})$. $C_2(\mathbf{R})$ is the set of all continuous bounded functions which are 0 around 0, and $C_1(\mathbf{R})$ is defined as a subclass of $C_2(\mathbf{R})$ consisting of nonnegative functions, containing all the

[11] Here $\sigma(X_0)$ means 'the σ-field generated by X_0'.
[12] See Jacod & Shiryaev [68] VII.2.7.

functions g_a where $g_a(x) = (a|x| - 1)^+ \wedge 1$ for all positive rationals a, and with the following property: let η_n, η be positive measures on $\mathbf{R}$ with no mass at 0 and finite on the complement of every neighbourhood of zero; then $\eta_n(f) \to \eta(f)$ for all $f \in C_1(\mathbf{R})$ implies $\eta_n(f) \to \eta(f)$ for all $f \in C_2(\mathbf{R})$. These properties do not uniquely determine $C_1(\mathbf{R})$, but the point is that $C_1(\mathbf{R})$ is a convergence-determining class for the weak convergence induced by $C_2(\mathbf{R})$ (that is, convergence only needs to be checked on $C_1(\mathbf{R})$ before one knows that it holds for $C_2(\mathbf{R})$.

Statement of theorems

We now state the two promised theorems on the convergence of semimartingales to jump diffusions. These are given in Jacod & Shiryaev [68], Chapter IX, theorems 4.8 and 4.15.

Theorem 10. *Let X and $X^{(\rho)}$ for each $\rho \in \{R, R+1, R+2, \ldots\}$, where R is some sufficiently large number, be semimartingales with $X_0 = x$ and with characteristics and modified characteristics all of the following form:*

$$\begin{aligned} B_t &= \int_0^t b(X_s)ds, \\ C_t &= \int_0^t \tilde{c}(X_s)ds, \\ N(dt, dx) &= dt\, K(X_t, dx) \\ \tilde{C}_t &= \int_0^t = \tilde{c}(X_s)ds \end{aligned}$$

where $b : \mathbf{R} \to \mathbf{R}$, $c : \mathbf{R} \to \mathbf{R}$ are Borel, K is a Borel kernel from $\mathbf{R}$ to itself with $\int K(x, dy)(|y|^2 \wedge 1) < \infty$, and where

$$\tilde{c}(x) = c(x) + \int K(x, dy)h^2(x),$$

where h is a truncation function. Assume the uniqueness measurability hypothesis for X (but not for $X^{(\rho)}$). Assume further that the characteristics of X are such that

$$\limsup_{\xi \uparrow \infty} \sup_{|x| \le a} K(x, \mathbf{R} \setminus [-\xi, \xi]) = 0 \quad \text{for all } a > 0,$$

that b, $\tilde{c}$ and $Kg = K(\cdot, dy)g(y)$ are continuous functions on $\mathbf{R}$ for all $g \in C_1(\mathbf{R})$, and that

$$b^{(\rho)} \to b, \quad \tilde{c}^{(\rho)} \to \tilde{c},$$

and $K^{(\rho)}g \to Kg$ *locally uniformly on* $\mathbf{R}$ *for all* $g \in C_1(\mathbf{R})$. *Denote the distribution of* $X^{(\rho)}$ *by* $P^{(\rho)}$*; then* $P^{(\rho)} \xrightarrow{w} P^x$.

Theorem 11. *Let* $X^{(\rho)}$ *and* X *be as in theorem 10. Assume the uniqueness-measurability hypothesis for* X *and that*

$$\limsup_{b\uparrow\infty} \sup_{|x|\le a} \int_{|y|>b} K(x,dy)|y|^2 = 0 \quad \text{for all } a > 0.$$

Set

$$b'(x) = b(x) + \int K(x,dy)(y - h(y)) \quad \text{and} \quad \tilde{c}'(x) = c(x) + \int K(x,dy)y^2,$$

and similarly for $b'^{(\rho)}$ *and* $c'^{(\rho)}$, *and suppose that* b', c' *and* Kg *are continuous for all* $g \in C_1(\mathbf{R})$. *Also assume that* $K^{(\rho)}$ *integrates* $|y|^2$ *and that* $b'^{(\rho)} \to b'$, $\tilde{c}'^{(\rho)} \to \tilde{c}'$ *and* $K^{(\rho)}g \to Kg$ *locally uniformly on* $\mathbf{R}$ *for all* $g \in C_1(\mathbf{R})$. *Write* $P^{(\rho)}$ *for the distribution of* $X^{(\rho)}$ *(recall that it starts at* x*); then* $P^{(\rho)} \xrightarrow{w} P$.

Chapter 5

Arithmetic-geometric hybrid Brownian motion

The hybrid arithmetic-geometric hybrid Brownian motion process arises naturally from the feedback model in the form

$$dX_t = (\mu_1 - \mu_2 X_t)dt + \sigma_1 dW_t^1 + \sigma_2 X_t dW_t^2, \qquad t > 0, \quad X_0 = x, \qquad (5.1)$$

driven by the two Brownian motions (W_t^1) and (W_t^2), having a correlation coefficient $r \in (-1, 1)$. It is studied in Shaw & Schofield [102]. Let us define a process W by setting

$$dW_t = \frac{\sigma_1}{\sqrt{\sigma_1^2 + 2r\sigma_1\sigma_2 X_t + \sigma_2^2 X_t^2}} dW_t^1 + \frac{\sigma_2 X_t}{\sqrt{\sigma_1^2 + 2r\sigma_1\sigma_2 X_t + \sigma_2^2 X_t^2}} dW_t^2.$$

By Lévy's characterization[1] of Brownian motion, the process W above is a Brownian motion. The hybrid SDE (5.1) can thus be written in the equivalent form

$$dX_t = (\mu_1 - \mu_2 X_t)dt + \sqrt{(\sigma_1^2 + 2r\sigma_1\sigma_2 X_t + \sigma_2^2 X_t^2)}\, dW_t, \qquad t > 0, \quad X_0 = x. \qquad (5.2)$$

Theorem 5 guarantees existence and uniqueness of a strong solution (X_t). This process is a regular diffusion on the whole line $I = (-\infty, \infty)$.

The parameters μ_1 and μ_2 are real, σ_1 and σ_2 are positive (nonzero to avoid trivialities) and $-1 \le r \le 1$. With the usual notation

$$B(x) = \int_0^x \frac{2b(y)}{\sigma^2(y)} dy, \quad x \in I,$$

[1] Lévy's characteriztion of Brownian motion says the following. Let M be a continuous local martingale with $M_0 = 0$ and quadratic variation process $[M]_t = t$; then M is Brownian motion. See, for example, Kallenberg [72] **Theorem 18.3** or Klebaner [77] **Theorem 7.32**.

we have in this case

$$B(x) = \nu_2 \left[-\arcsin(r) + \arctan\left(\frac{x-\eta}{x_0}\right)\right] - \frac{\mu_2}{\sigma_2^2}\log\left(\frac{\sigma^2(x)}{\sigma_1^2}\right),$$

where

$$\begin{aligned} \nu &= \tfrac{2\mu_2}{\sigma_2^2} + 1 \\ \nu_2 &= \tfrac{2(r\mu_2\sigma_1 + \mu_1\sigma_2)}{\sqrt{(1-r^2)}\,\sigma_1\sigma_2^2} \\ x_0 &= \tfrac{\sigma_1}{\sigma_2}\sqrt{1-r^2} \\ \eta &= -\tfrac{r}{\sqrt{(1-r^2)}} x_0 = -r\tfrac{\sigma_1}{\sigma_2}. \end{aligned} \tag{5.3}$$

Observe that for the expression inside the log function,

$$\frac{\sigma^2(x)}{\sigma_1^2} = 1 + \left(\frac{x-\eta}{x_0}\right)^2.$$

5.1 Scale

The diffusion (5.2) has scale density

$$s(x) = e^{-B(x)} = e^{\nu_2 \arcsin(r)} \left(1 + \frac{(x-\eta)^2}{x_0^2}\right)^{\frac{\nu-1}{2}} e^{-\nu_2 \arctan\left(\frac{x-\eta}{x_0}\right)}. \tag{5.4}$$

The arctan function is bounded (taking values in $(-\pi/2, \pi/2)$) and it is easily seen that the scale measure has finite mass if and only if $2(\nu-1)/2 < -1$. Thus, the scale measure is finite if and only if $\nu < 0$.

5.2 Speed

The speed measure associated with (5.2) has Lebesgue density

$$m(x) = \frac{2}{\sigma_1^2} e^{-\nu_2 \arcsin(r)} \left(1 + \frac{(x-\eta)^2}{x_0^2}\right)^{-\frac{\nu+1}{2}} e^{\nu_2 \arctan\left(\frac{x-\eta}{x_0}\right)}. \tag{5.5}$$

The speed measure is finite precisely if $\nu > 0$, in which case an invariant distribution exists.

5.3 Boundary classification

The scale measure is finite at both endpoints if $\nu < 0$ and infinite at both endpoints if $\nu \geq 0$. Both boundaries are always natural, that is, the diffusing particle can neither enter the state space from $\pm\infty$ nor reach $\pm\infty$ from the interior of the state space.

5.4 Equilibria

The simplest way of calculating the density of the invariant distribution, when it exists, is simply to normalise the speed density. We denote invariant distributions, and their densities with respect to the Lebesgue measure, by π, i.e. $\pi(dx) = \pi(x)dx$. These invariant distributions exist precisely when $\nu > 0$.

- **Symmetric case $\mu_1 = r = 0$.** In this case, normalising the speed density leads us to

$$\pi(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)x_0}\left(1+\frac{x^2}{x_0^2}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbf{R}. \tag{5.6}$$

  This is easily recognised as a scaled Student's t-distribution, with ν degrees of freedom and scale parameter $x_0/\sqrt{\nu}$. If T is a Student random variable with ν degrees of freedom, then the invariant distribution is the distribution of $x_0T/\sqrt{\nu}$. It is well-known that this distribution has significantly fatter tails than the normal distribution. Indeed it has only an nth moment for $n < \nu$.

  Denoting by X a stochastic variable having this invariant distribution, we have

$$\begin{aligned} EX &= 0 \\ E[X^2] &= \operatorname{var}(X) = \frac{x_0^2}{\nu-2} \end{aligned}$$

  Of course, for $\nu > 3$ it has zero skewness and, for $\nu > 4$, an excess kurtosis of

$$\text{ex. kurtosis } (X) = \frac{6}{\nu-4}.$$

- **General case.** For general values of the parameters $\mu_1, \mu_2 \in \mathbf{R}$, $\sigma_1, \sigma_2 > 0$, $-1 < r < 1$, we find the invariant probability density

$$\pi(x) = C\left(1+\left(\frac{x-\eta}{x_0}\right)^2\right)^{-\frac{\nu+1}{2}} e^{\nu_2 \arctan\left(\frac{x-\eta}{x_0}\right)}, \tag{5.7}$$

where the normalising constant is[2] given by

$$C = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{x_0\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left|\frac{\Gamma\left(\frac{\nu+1-i\nu_2}{2}\right)}{\Gamma\left(\frac{\nu+1}{2}\right)}\right|^2$$

The invariant distribution belongs to the Pearson Type IV class of distributions. It is a kind of skew-Student distribution. A useful guide to the properties of the Pearson Type IV class of distributions is Heinrich [58]. We note here the first four moments:

1. For $\nu > 1$, the mean exists and equals

$$m = EX = \eta + \frac{\nu_2 x_0}{\nu - 1}.$$

2. For $\nu > 2$, we can calculate the variance

$$\sigma^2 = \mathrm{var}\,(X) = \eta + \frac{x_0^2[(\nu-1)^2 + \nu_2^2]}{(\nu-1)^2(\nu-2)}.$$

3. The standardized skewness is, for $\nu > 3$,

$$s = E\left[\left(\frac{X-m}{\sigma}\right)^3\right] = \frac{4\nu_2}{(\nu-3)}\sqrt{\frac{\nu-2}{(\nu-1)^2+\nu_2^2}}.$$

4. The excess kurtosis is, for $\nu > 4$,

$$\kappa - 3 = E\left[\left(\frac{X-m}{\sigma}\right)^4\right] - 3 = \frac{6[(\nu-1)^2(\nu-3) + (5\nu-11)\nu_2^2]}{(\nu-3)(\nu-4)((\nu-1)^2+\nu_2^2)}.$$

5.5 Spectral classification

We consider only the simplest case with $\mu_1 = r = 0$, the model for which we have exhibited a Student t equilibrium. In this case, the Lamperti transform is

$$z(x) = \frac{1}{\sigma_2}\mathrm{arsinh}\left(\frac{x}{x_0}\right), \quad x(z) = x_0\sinh(\sigma_2 z), \tag{5.8}$$

and direct application of Itô's lemma yields the SDE

$$dZ_t = b_Z(Z_t)dt + dW_t$$

[2] Heinrich [58] and Jeffreys [69].

for $Z_t = z(X_t)$, where

$$b_Z(z) = -\frac{1}{2}\nu\sigma_2 \tanh(\sigma_2 z).$$

The Schrödinger potential is $Q(z) = U(x(z))$ where

$$U(x) = \frac{-2\sigma_1^2\sigma_2^2\nu + \sigma_2^4\nu^2x^2}{8(\sigma_1^2 + \sigma_2^2x^2)}$$

which satisfies

$$U(\pm\infty) = \frac{\sigma_2^2\nu^2}{8}.$$

Thus both endpoints $\pm\infty$ are oscillatory with the same cut-off $-\Lambda = -\sigma_2^2\nu^2/8$ and the process belongs to Linetsky's spectral category III. The spectrum consists of a (possibly empty) set of simple eigenvalues in $[-\sigma_2^2\nu^2/8, 0]$ together with an essential component contained in $(-\infty, -\sigma_2^2\nu^2/8]$. The essential part of the spectrum has multiplicity 2. Finally, because

$$\lim_{x\to\pm\infty} z^2(x)(U(x) - \Lambda) = 0 > -\frac{1}{4}$$

the SL equation is nonoscillatory for $\lambda = -\Lambda$, which means that A has only finitely many eigenvalues, all found in $[-\sigma_2^2\nu/8, 0]$. We note that 0 is an eigenvalue in precisely those cases when an invariant distribution π exists, i.e. when the speed measure is finite. This is equivalent, in this model, to the condition $\nu > 0$.

Writing $x(z) = x_0 \sinh(z)$ in $Q(z) = U(x(z))$ leads us to the Schrödinger potential

$$Q(z) = c_0 + \frac{c_1}{\cosh^2(\sigma_2 z)},$$

$$c_0 = \frac{\sigma_2^2\nu^2}{8}, \quad c_1 = -\frac{(2\mu_2 + 3\sigma_2^2)\nu}{8}$$

This is a special case of the *modified Pöschl-Teller* potential, see Grosche & Steiner [56], pp244-245, especially equation (6.6.10).

5.6 Dynamic probabilities

Dynamic moment evolution

Writing $m_n(t) = E[X_t^n]$, $t \geq 0$, $n \in \mathbf{N}$, we have from Shaw & Schofield [102]

$$\dot{m}_n(t) + (\mu_2 n - \frac{n(n-1)}{2}\sigma_2^2)m_n(t) = \frac{n(n-1)}{2}\sigma_1^2 m_{n-2}(t) + (\mu_1 n + n(n-1)r\sigma_1\sigma_2)m_{n-1}(t).$$

with $m_0 = 1$, and $m_n(0) = 0$, $n \geq 1$. In particular,

$$EX_t = \frac{\mu_1}{\mu_2}(1 - e^{-\mu_2 t}).$$

The variance is easily written down in the case $\mu_1 = r = 0$:

$$\text{var}\,(X_t) = \frac{\sigma_1^2}{(\nu - 2)\sigma_2^2}(1 - e^{-\sigma_2^2(\nu-2)t}). \tag{5.9}$$

For $\nu = 2$, the variance equals $\sigma_1^2 t$. This formula for the variance in the $\nu = 2$ case was confirmed in Shaw & Schofield [102].

Knowing the conditional moments in full is useful for constructing estimating functions for estimation of the parameters from discretely sampled data. However, we must be careful in this model whether these moments exist. Using the first moment involves tacitly assuming that $\nu > 1$, and using second moments, $\nu > 2$. Historical data appear to support the latter assumption in some, but not all, cases. If $X_0 = x$, we have

$$m(x) = E^x X_t = \frac{\mu_1}{\mu_2} + \left(x - \frac{\mu_1}{\mu_2}\right) e^{-\mu_2 t} \tag{5.10}$$

and, *assuming that* $\nu > 2$,

$$q(x) = E^x X_t^2 = h(x) + (x^2 - h(x))e^{-(2\mu_2 - \sigma_2^2)t}, \tag{5.11}$$

where

$$h(x) = \frac{2\mu_1^2/\mu_2 + \sigma_1^2}{2\mu_2 - \sigma_2^2} - \frac{2(\mu_1 - \mu_2 x)}{\mu_2 - \sigma_2^2}.$$

Towards the transition densities

We consider only the symmetric case $\mu_1 = r = 0$. Thus, scale and speed densities simplify to

$$s(x) = \left(1 + \left(\frac{x}{x_0}\right)^2\right)^{\frac{\nu-1}{2}}, \quad m(x) = \frac{2}{\sigma_1^2}\left(1 + \left(\frac{x}{x_0}\right)^2\right)^{-\frac{\nu+1}{2}}. \tag{5.12}$$

We have already noted that the Liouville transformation

$$z(x) = \frac{1}{\sigma_2}\text{arsinh}\left(\frac{x}{x_0}\right), \quad v(z) = u(x)/\sqrt{\sigma(x)s(x)}, \tag{5.13}$$

takes the SL equation

$$\frac{1}{2}\sigma^2(x)u''(x) + b(x)u'(x) - \lambda u(x) = 0$$

to its *Liouville normal form*

$$\frac{1}{2}v''(z) - \left(\lambda + c_0 + c_1\mathrm{sech}^2(\sigma z)\right)v(z) = 0 \tag{5.14}$$

where

$$c_0 = \frac{\sigma_2^2\nu^2}{8}, \quad c_1 = -\frac{(2\mu_2 + 3\sigma_2^2)\nu}{8}.$$

The solution of the equation in this form is well-known: the increasing and descreasing solutions of this equation are, respectively

$$\tilde{\psi}_\lambda(z) = P_{\nu/2}^{-\delta(\lambda)}(\tanh(-\sigma_2 z)), \quad \tilde{\chi}_\lambda(z) = P_{\nu/2}^{-\delta(\lambda)}(\tanh(\sigma_2 z)), \tag{5.15}$$

where we have used the notation

$$\delta(\lambda) = \sqrt{\frac{2\lambda}{\sigma_2^2} + \frac{\nu^2}{4}} \tag{5.16}$$

for the discriminant. It is easily checked that these are solutions; the difficulty, that $\tilde{\chi}_\lambda$ represents the decreasing solution, is the same as the result in section 1 of Vagurina [108] (where he uses the notation $\tilde{\varphi}$).

The corresponding solutions of the original SL equation are then found by reversing the procedure of making these substitutions: $u(x) = v(z(x))\sqrt{\sigma(x)s(x)}$. We observe first that

$$\xi(x) = \tanh(\sigma_2 z(x)) = \tanh(\mathrm{arsinh}(x/x_0)) = \frac{x/x_0}{\sqrt{1 + (x/x_0)^2}} \in (-1, 1)$$

and

$$\sigma(x)s(x) = \sigma_1\left(1 + \left(\frac{x}{x_0}\right)^2\right)^{\nu/2}.$$

Note that ξ is strictly increasing, and is a bijection from $\mathbf{R}$ to $(-1, 1)$. It now follows that the fundamental increasing and decreasing solutions of the SL equation $(A - \lambda)u = 0$ are

$$\psi_\lambda(x) = \left(1 + \left(\frac{x}{x_0}\right)^2\right)^{\nu/4} P_{\nu/2}^{-\delta(\lambda)}\left(-\frac{x/x_0}{\sqrt{1 + (x/x_0)^2}}\right)$$

and

$$\chi_\lambda(x) = \left(1 + \left(\frac{x}{x_0}\right)^2\right)^{\nu/4} P_{\nu/2}^{-\delta(\lambda)}\left(\frac{x/x_0}{\sqrt{1+(x/x_0)^2}}\right)$$

respectively. The Wronskian is

$$w_\lambda = \frac{2+2\delta+\nu}{2x_0}\left[P_{\nu/2}^{-\delta(\lambda)}(\xi)P_{\nu/2+1}^{-\delta(\lambda)}(-\xi) + P_{\nu/2}^{-\delta(\lambda)}(-\xi)P_{\nu/2+1}^{-\delta(\lambda)}(\xi)\right].$$

The general SL theory tells us that the Wronskian does not depend on x, hence ξ, though this is not obvious from the formula. Obvious possibilities for us are to consider $\xi = 0$ and $\xi \to 1$. Taking $\xi = 0$ gives useful cancellations for expressions for transitions from $x = 0$:

$$\begin{aligned} G_\lambda(0,y)m(y) &= \frac{2^{\delta(\lambda)-1}\Gamma\left(\frac{\delta(\lambda)-\nu/2}{2}\right)\Gamma\left(\frac{1+\delta(\lambda)+\nu/2}{2}\right)}{\sqrt{\pi}\sigma_1\sigma_2} \times \\ &\quad \times \left(1+\left(\frac{y}{x_0}\right)^2\right)^{-\frac{\nu/2+1}{2}} P_{\nu/2}^{-\delta(\lambda)}(|\xi(y)|), \end{aligned}$$

which agrees with formula (110) in Shaw & Schofield [102]. Using instead from Erdélyi et al. table 3.9.2 formulae (8) and (14) for the behaviour as $\xi \to 1$ of the Wronskian, we obtain

$$w_\lambda = \frac{2}{x_0\Gamma\left(\delta(\lambda)-\frac{\nu}{2}\right)\Gamma\left(\delta(\lambda)+\frac{\nu}{2}+1\right)} \tag{5.17}$$

which gives an expression for the Green function

$$G_\lambda(x,y) = \left\{ \begin{array}{l} \frac{x_0}{2}\Gamma\left(\frac{\nu}{2}+\delta(\lambda)+1\right)\Gamma\left(\delta(\lambda)-\frac{\nu}{2}\right)\left(1+\left(\frac{x}{x_0}\right)^2\right)^{\nu/4} \quad \times \\ \times \quad \left(1+\left(\frac{y}{x_0}\right)^2\right)^{\nu/4} P_{\nu/2}^{-\delta(\lambda)}(-\xi(x\wedge))P_{\nu/2}^{-\delta(\lambda)}(\xi(x\vee y)). \end{array} \right\} \tag{5.18}$$

Since the Green function is the integral kernel of the resolvent with respect to the speed measure, we are actually interested in finding expressions for

$$G_\lambda(x,y)m(y) = \left\{ \begin{array}{l} \frac{1}{\sigma_1\sigma_2}\Gamma(\delta(\lambda)+1+\nu/2)\Gamma(\delta(\lambda)-\nu/2)\left(1+\left(\frac{y}{x_0}\right)^2\right)^{-\nu/4-1/2} \times \\ \times \quad \left(1+\left(\frac{x}{x_0}\right)^2\right)^{\nu/4} P_{\nu/2}^{-\delta(\lambda)}(-\xi(x_\wedge y))P_{\nu/2}^{-\delta(\lambda)}(\xi(x_\vee y)). \end{array} \right\} \tag{5.19}$$

The Legendre functions $P_{\nu/2}^{-\delta}(\xi)$ may be expressed in terms of Gauss' hypergeometric function (as indeed was the chosen representation in Shaw &

Schofield [102]) via formula 3.4(6) in Erdélyi et al. [43]

$$P^{-\delta}_{\nu/2}(\xi) = \frac{1}{\Gamma(1+\delta)}\left(\frac{1-\xi}{1+\xi}\right)^{\delta/2} {}_2F_1\left(-\nu/2, \nu/2+1; 1+\delta; \frac{1-\xi}{2}\right).$$

We note first that we may write

$$\frac{1-\xi(x)}{1+\xi(x)} = e^{-2\text{arsinh}(x/x_0)}.$$

Now certain special cases present themselves: the hypergeometric function reduces to a polynomial if $\nu/2$ is any integer. Its degree will then be

$$\begin{cases} \nu/2 & \text{if } \nu \geq 0 \\ -\nu/2 - 1 & \text{if } \nu < 0. \end{cases}$$

We now conduct an analysis of these special cases.

(a) **The case $\nu = 0$.**

This case could be solved directly from the Liouville normal form (5.14) of the SL equation, because in this case $c_0 = c_1 = 0$ and

$$\frac{1}{2}v''(z) - \lambda v(z) = 0.$$

We have

$$\psi_\lambda(x) = e^{\sqrt{2\lambda}z(x)}, \qquad \chi_\lambda(x) = e^{-\sqrt{2\lambda}z(x)}$$

and

$$G_\lambda(x,y)m(y) = \frac{1}{\sqrt{2\lambda(\sigma_1^2 + \sigma_2^2 y^2)}} \exp\left(-\sqrt{2\lambda}\frac{|\text{arsinh}(y/x_0) - \text{arsinh}(x/x_0)|}{\sigma_2}\right).$$

Laplace transform inversion is easy[3] in this case and gives us the closed-form transition density

$$p(t;x,y) = \frac{1}{\sqrt{2\pi t(\sigma_1^2 + \sigma_2^2 y^2)}} \exp\left(-\frac{|\text{arsinh}(y/x_0) - \text{arsinh}(x/x_0)|^2}{2\sigma_2^2 t}\right).$$

We can just check that our more general formulae give the same thing when the special case $\nu = 0$ is considered. Since

$$P_0^{-\delta(\lambda)}(\xi(x)) = \frac{\exp\left(-\delta(\lambda)\text{arsinh}(x/x_0)\right)}{\Gamma(1+\delta(\lambda))},$$

[3] Formula **29.3.84** in Abramowitz & Stegun [2]

(5.18) becomes in this case

$$G_\lambda(x,y) = \frac{x_0}{2}\frac{\Gamma(1+\delta(\lambda))\Gamma(\delta(\lambda))}{\Gamma(\delta(\lambda)+1)\Gamma(\delta(\lambda)+1)}\exp\left(-\delta(\lambda)|\text{arsinh}(y/x_0)-\text{arsinh}(x/x_0)|\right)$$

which reduces to the same Green function as above.

In terms of the Lamperti transform z and its inverse z^{-1} (see (5.8)), we may write this

$$p(t;x,y) = p_W(t;z_x,z_y)(z^{-1})'(z_y)$$

where z_x denotes the position of the process $z(X)$ when X is at x and p_W is the transition density of the standard Wiener process

$$p_W(t;z_x,z_y) = \frac{1}{\sqrt{t}}\varphi\left(\frac{z_y-z_x}{\sqrt{t}}\right).$$

So $X = x_0\sinh(\sigma_2 Z)$ for a Wiener process Z.

This leads us to some useful interpretations by Bougerol's [22] identity. First, X_t has the same distribution as $x_0 B(A_t(Z))$ for an independent Brownian motion B and an Asian time-change

$$A_t(Z) = \int_0^t e^{2\sigma_2 Z_u}du.$$

Furthermore, using Vakeroudis [109] **Corolloary 2.2** with $\mu=\nu=0$ (due to Alili et al. [3]) we can see that the process X has the form

$$X_t = \sigma_1\int_0^t e^{\sigma_2 Z_u}dB_u.$$

In fact, by direct solution of the hybrid SDE (4.5) Shaw & Schofield were able to replace the two Brownian motions B and Z in this formula with the original ones W^1 and W^2 driving, respectively, the fundamentally and technically motivated noise in the price process.

(b) **The case $\nu = 2$.**

We begin by recalling that the speed measure is finite on $\mathbf{R}$, and we have an invariant distribution π which is Student's t, with 2 degrees of freedom and scale parameter $x_0/\sqrt{2}$ (the variance is infinite). Now the Legendre function of interest is

$$P_1^{-\delta}(\xi(x)) \;=\; \frac{\exp\left(-\delta\text{arsinh}(x/x_0)\right)}{\Gamma(1+\delta)}{}_2F_1\left(-1,2;1+\delta;\frac{1-\xi}{2}\right)$$

$$= \frac{\exp\left(-\delta \text{arsinh}(x/x_0)\right)}{\Gamma(1+\delta)} \frac{(\delta + \xi(x))}{(1+\delta)}.$$

Put

$$\delta(\lambda) = \sqrt{\frac{2\lambda}{\sigma_2^2} + 1}.$$

We obtain

$$G_\lambda(x,y)m(y) = \frac{\sqrt{1+(x/x_0)^2}\, e^{-\delta(\lambda)\frac{|z(x)-z(y)|}{\sigma_2}}}{2\, x_0\, \lambda\, \delta(\lambda)(1+(y/x_0)^2)}[\delta(\lambda) - \xi(x_\wedge y)][\delta(\lambda) + \xi(x_\vee y)].$$

With $x = 0$, this becomes

$$G_\lambda(0,y)m(y) = \frac{\exp\Big(-\delta(\lambda)|\text{arsinh}(y/x_0)|\Big)[\delta(\lambda) + |\xi(y)|]}{2x_0\lambda\left(1+\left(\frac{y}{x_0}\right)^2\right)}$$

This is invertible explicitly. From this form of the integrand, we can perform the inversion by writing $\lambda = (1/2)\sigma_2^2(\delta-1)(\delta+1)$ in the denominator, splitting the expression into two separate fractions and inverting using formulae **29.3.88** and **29.2.14** from Abramowitz & Stegun [2]. The result is

$$\begin{aligned} p(t;0,y) &= \frac{e^{-\frac{z^2(y)}{2t} - \frac{1}{2}\sigma_2^2 t}}{\sqrt{2\pi\sigma_1^2 t}\Big(1+(y/x_0)^2\Big)} \\ &\quad + \frac{1}{2x_0\Big(1+(y/x_0)^2\Big)^{3/2}}\left[\Phi\left(\frac{z(y)}{\sigma_2\sqrt{t}} + \sigma_2\sqrt{t}\right) - \Phi\left(\frac{z(y)}{\sigma_2\sqrt{t}} - \sigma_2\sqrt{t}\right)\right] \end{aligned} \tag{5.20}$$

where z is given in (5.13) and Φ is the standard normal distribution function.

(c) **The case $\nu = 4$.**

The invariant distribution is Student's t with four degrees of freedom and scale parameter $x_0/2$, which has mean 0, variance $x_0^2/2$ and skewness 0. In this case we define

$$\delta(\lambda) = \sqrt{\frac{2\lambda}{\sigma_2^2} + 4}.$$

Suppressing the dependence on λ of δ in our notation, we have

$$G_\lambda(x,y)m(y) = \left\{ \begin{array}{l} \frac{\left(1+\left(\frac{x}{x_0}\right)^2\right)\exp\left(-\delta|\mathrm{arsinh}(x/x_0)-\mathrm{arsinh}(y/x_0)|\right)}{\sigma_1\sigma_2(\delta-2)_5\left(1+\left(\frac{y}{x_0}\right)^2\right)^{3/2}}\times \\ \times \quad [(\delta^2-1)-3\delta\xi(x)+3\xi^2(x)][(\delta^2-1)+3\delta\xi(y)+3\xi^2(y)] \end{array} \right\}$$

for $x, y \in \mathbf{R}$. Here, by $(\delta-2)_5$, we meant the rising factorial

$$(\delta-2)_5 = (\delta-2)(\delta-1)\,\delta\,(\delta+1)(\delta+2) = \frac{2\lambda}{\sigma_2^2}\left(\frac{2\lambda}{\sigma_2^2}+3\right)\sqrt{\frac{2\lambda}{\sigma_2^2}+4}.$$

In the case $x = 0$ we have a cancellation of the second term on the right with the polynomial in $\xi(x)$ in the Green function. Thus,

$$\begin{aligned} G_\lambda(0,y)m(y) \;\; = \;\; & \frac{\sigma_2\exp\Big(-\delta|\mathrm{arsinh}(x/x_0)-\mathrm{arsinh}(y/x_0)\Big)}{2\sigma_1\lambda\delta(\lambda)\left(1+\left(\frac{y}{x_0}\right)^2\right)^{3/2}}\times \\ & \times[(\delta^2-1)+3\delta|\xi(y)|+3\xi^2(y)]. \end{aligned}$$

(d) **The case $\nu = -2$.**

In the case $\nu = -2$, the hypergeometric function equals constant 1, and

$$G_\lambda(x,y) = \frac{\exp\Big(-\delta(\lambda)|\mathrm{arsinh}(x/x_0)-\mathrm{arsinh}(y/x_0)|\Big)}{\sigma_1\sigma_2\delta(\lambda)\sqrt{1+\left(\frac{x}{x_0}\right)^2}}$$

This formula is easily invertible in closed form, and leads us to the transition density

$$p(t;x,y) = \frac{e^{-\frac{\sigma_2^2}{2}t}}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2x^2)t}}\exp\left(-\frac{\Big(\mathrm{arsinh}(x/x_0)-\mathrm{arsinh}(y/x_0)\Big)^2}{2\sigma_2^2t}\right)$$

An exercise in elementary calculus shows that, as a function of y, this is a pdf.

In terms of the Lamperti transform, we may write this

$$p(t;x,y) = \frac{1}{\sqrt{2\pi t(\sigma_1^2+\sigma_2^2x^2)}}e^{-\frac{1}{4t}\left[\left((z_y-z_x)-\sigma_2t\right)^2+\left((z_y-z_x)+\sigma_2t\right)^2\right]}$$

Looking again at the result of Alili et al.[3] (Vakeroudis **Corollary 2.2**, this time with $\mu = 1$ and $\nu = 0$) we recognise

$$X_t \sim x_0 \sinh(W_{\sigma_2 t} + \varepsilon\sigma_2 t)$$

where ε is a symmetric Bernoulli random variable taking values in $\{\pm 1\}$. The interpretation is then that the Bernoulli random variable decides which way the momentum effects of the technical trading are going to push the process X. The unconditional density is a mixture of two densities moving apart at constant speed, one rightwards and the other leftwards, as t increases.

(e) **The case $\nu = -4$.**

Via the Legendre symmetry, we see that this case mirrors the $\nu = 2$ case, hence an explicit inversion is available for $x = 0$. This time,

$$\delta(\lambda) = \sqrt{\frac{2\lambda}{\sigma_2^2} + 4}\,.$$

A little algebra leads us to the formula

$$G_\lambda(0, y)m(y) = u(y)e^{-k(y)\delta(\lambda)} \left(\frac{1 + |\xi(y)|}{\delta(\lambda) - 1} + \frac{1 - |\xi(y)|}{\delta(\lambda) + 1} \right)$$

where $u(y) = \sqrt{1 + (y/x_0)^2}/(2\sigma_1\sigma_2)$ and $k(y) = |\text{arsinh}(y/x_0)| = \sigma_2|z(y)|$ (z begin given by (5.13)). Noting that

$$e^{-k(y)}(1 + |\xi(y)|) = e^{k(y)}(1 - |\xi(y)|) = (\cosh^2(\text{arsinh}(|y|/x_0)))^{1/2},$$

direct inversion (using again formulae **29.3.88** and **29.2.14** from Abramowitz & Stegun [2]) leads us to the explicit formula

$$\begin{aligned} p(t; 0, y) = \quad & e^{-2\sigma_2^2 t} \left(\frac{\sqrt{1+(y/x_0)^2}}{\sqrt{2\pi\sigma_1^2 t}} e^{-\frac{z^2(y)}{2t}} \right. \\ & + \left. \frac{e^{\frac{1}{2}\sigma_2^2 t}}{2x_0} \left[\Phi\left(\frac{z(y)+\sigma_2 t}{\sqrt{t}} \right) - \Phi\left(\frac{z(y)-\sigma_2 t}{\sqrt{t}} \right) \right] \right), \end{aligned} \tag{5.21}$$

where Φ is the standard normal distribution function and z is given by (5.13).

Chapter 6

Arithmetic-Cox-Ingersoll-Ross hybrid

The arithmetic-Cox-Ingersoll-Ross hybrid model is the solution of the SDE

$$dX_t = (\mu_1 - \mu_2 X_t)dt + \sigma_1 dW_t^1 + \sqrt{|X_t|}dW_t^2, \qquad t > 0, \quad X_0 = x, \qquad (6.1)$$

for two standard Wiener processes W^1 and W^2. For tractability, we consider only the case in which W^1 and W^2 are uncorrelated, in which case we are investigating the equivalent SDE

$$dX_t = (\mu_1 - \mu_2 X_t)dt + \sqrt{(\sigma_1^2 + \sigma_2^2|X_t|)}dW_t, \qquad t > 0, \quad X_0 = x, \qquad (6.2)$$

where W is a standard Brownian motion by Lévy's characterisation (see the discussion at the beginning of chapter 5). This process is again a regular diffusion on the whole real line $I = (-\infty, \infty)$. By Theorem[1] 5, equation (6.2) has a unique strong solution. The difficulties involved here stem from the fact that the diffusion coefficient $a(x) = \sigma_1^2 + \sigma_2^2|x|$ is not smooth at zero, and solutions must be written down separately on $(0, \infty)$ and $(-\infty, 0)$, and junction conditions imposed at zero. We shall not find closed-form expressions for the transition density, but we can write down the Green function easily enough in terms of confluent hypergeometric functions (section 2.6). The transition operators and densities may therefore be approximated by means of (2.23) with contour of integration (2.26).

Again, the parameters μ_1 and μ_2 are real, whilst σ_1 and σ_2 must be taken

[1]Note that the Hölder condition (2.14) is only just weak enough to be applicable here. The theorems from other sources mentions in the introduction, subsection 1.5.2, are not sharp enough for this diffusion coefficient.

positive (nonzero to avoid trivialities). We have

$$B(x) = -\rho|x| + \tilde{c}\log\left(1 + \frac{|x|}{x_0}\right)$$

where the parameters associated with this process are

$$\rho = \tfrac{2\mu_2}{\sigma_2^2}, \quad x_0 = \tfrac{\sigma_1^2}{\sigma_2^2},$$

$$\tilde{c} = \begin{cases} c = \frac{2(\mu_2\sigma_1^2 + \mu_1\sigma_2^2)}{\sigma_2^4}, & x > 0, \\ \hat{c} = \frac{2(\mu_2\sigma_1^2 - \mu_1\sigma_2^2)}{\sigma_2^4}, & x < 0, \end{cases} \tag{6.3}$$

$$z_0 = \rho x_0 = \tfrac{1}{2}(c + \hat{c}).$$

The symmetry condition $\mu_1 = 0$ results in the simplifications $z_0 = c = \hat{c} = \tilde{c}$. Observe that the rate parameter ρ always has the same sign as the technical drift μ_2.

6.1 Scale

The scale density is

$$s(x) = e^{\rho|x|}\left(1 + \frac{|x|}{x_0}\right)^{-\tilde{c}}. \tag{6.4}$$

The associated scale measure S on $\mathbf{R}$ is

- finite when $\mu_2 < 0$.
- infinite at both endpoints when $\mu_2 > 0$.

When $\mu_2 = 0$, we have

$$s(x) = \begin{cases} \left(1 + \frac{x}{x_0}\right)^{-c} & x > 0 \\ \\ \left(1 + \frac{|x|}{x_0}\right)^{c} & x < 0. \end{cases}$$

In these cases, the scale measure is always infinite at one of the endpoints; however we note that it is:

- finite at the left endpoint ($l = -\infty$) precisely if $c < -1$, i.e. $\mu_1 < -\sigma_2^2/2$.
- finite at the right endpoint ($r = \infty$) precisely if $c > 1$, i.e. $\mu_1 > \sigma_2^2/2$.

When $\mu_2 = 0$ and $c \in [-1, 1]$, the scale measure is infinite at both endpoints. For $\mu_1 = 0$, the scale measure equals Lebesgue measure on $\mathbf{R}$. This we should

expect since our process is then just a stochastic integral with respect to a standard Wiener process, making it a continuous local martingale.

6.2 Speed

From the formula

$$m(x) = \frac{2}{\sigma^2(x)s(x)}$$

we obtain the speed density of the hybrid arithmetic-CIR process

$$m(x) = \frac{2}{\sigma_1^2} e^{-\rho|x|} \left(1 + \frac{|x|}{x_0}\right)^{\tilde{c}-1}. \tag{6.5}$$

The speed measure is finite if $\mu_2 > 0$ and infinite if $\mu_2 < 0$. If $\mu_2 = 0$, then the speed measure is

- finite at $+\infty$ if and only if $c < 0$ i.e. $\mu_1 < 0$.
- finite at $-\infty$ if and only if $c > 0$, i.e., $\mu_1 > 0$.

6.3 Boundary classification

We have seen that both boundaries are attracting if $\mu_2 < 0$ and neither are if $\mu_2 > 0$. This is intuitive. $\mu_2 < 0$ refers to the momentum case, so that once the returns start moving upwards towards $+\infty$, it is then driven further that way, and similarly if it starts moving towards $-\infty$ its momentum forces it downwards. In the case $\mu_2 = 0$, the overall direction of movement ceases to be state space dependent and the drift term μ_1 comes into play. The upper boundary $+\infty$ is attracting if and only if $\mu_1 > \sigma_2^2/2$ and the lower boundary $-\infty$ is attracting if and only if $\mu_1 < -\sigma_2^2/2$. We note that both boundaries are natural boundaries in either case. In particular, even though the boundaries can be attracting, they are never attainable.

6.4 Equilibria

Normalising the speed density in the case $\mu_2 > 0$, and keeping it continuous at zero, the invariant distribution π has density

$$\pi(x) = \begin{cases} p\frac{\rho^c}{\Gamma(c,\rho x_0)} e^{-\rho(x_0+x)}(x_0+x)^{c-1}, & x > 0, \\ q\frac{\rho^{\hat{c}}}{\Gamma(\hat{c},\rho x_0)} e^{-\rho(x_0+|x|)}(x_0+|x|)^{\hat{c}-1}, & x < 0. \end{cases} \tag{6.6}$$

where the probabilities attributed to $(0,\infty)$ and $(-\infty,0)$ respectively are

$$p = \frac{\Gamma(c,\rho x_0)}{\Gamma(c,\rho x_0) + (\rho x_0)^{c-\hat{c}}\Gamma(\hat{c},\rho x_0)} \quad \text{and} \quad q = \frac{(\rho x_0)^{c-\hat{c}}\Gamma(\hat{c},\rho x_0)}{\Gamma(c,\rho x_0) + (\rho x_0)^{c-\hat{c}}\Gamma(\hat{c},\rho x_0)}.$$

In the symmetric case $c = \hat{c}$, one has equal probabilities $p = q = 1/2$. p is increasing with μ_1.

To clarify the nature of the general invariant distribution, we make some observations. The gamma distribution Gamma(c,ρ) has density

$$g(x;c,\rho) = \frac{\rho^c}{\Gamma(c)} e^{-\rho x} x^{c-1}, \quad x > 0,$$

where $\rho > 0$ is the *rate* parameter and $c > 0$ is the *shape* parameter. It has survival function

$$Q(c,\rho x) = \frac{\Gamma(c,\rho x)}{\Gamma(c)}, \quad x > 0$$

giving the probability of the interval (x,∞). If this distribution is then truncated, by chopping off the interval $(0,x_0)$, shifting its left endpoint back to zero and renormalising, one obtains the density

$$g(x;c,\rho,x_0) = \frac{\rho^c}{\Gamma(c,\rho x_0)} e^{-\rho(x_0+x)}(x_0+x)^{c-1}.$$

where the parameter is now a triple $(c,\rho,x_0) \in \mathbf{R} \times (0,\infty) \times (0,\infty)$. Observe that by truncating the distribution at x_0, we no longer require the condition $c > 0$ of integrability at zero. The invariant distribution π of the present model is then a mixture of this truncated shifted Gamma(c,ρ,x_0) distribution on $(0,\infty)$, and a truncated shifted Gamma$(\hat{c},-\rho,x_0)$, which is, by definition, concentrated on $(-\infty,0)$. (We define $g(x;c,-\rho,x_0) = g(-x;c,\rho,x_0)$.)

From these considerations, the raw moments $m_k = \int_{\mathbf{R}} x^k \pi(x)dx$ may be easily calculated as

$$m_k = \sum_{r=0}^{k} \binom{k}{r} \frac{(-x_0)^{k-r}}{\rho^r} \left[\frac{(\rho x_0)^{\hat{c}-c}\Gamma(c+r,\rho x_0) + (-1)^k \Gamma(\hat{c}+r,\rho x_0)}{(\rho x_0)^{\hat{c}-c}\Gamma(c,\rho x_0) + \Gamma(\hat{c},\rho x_0)} \right]. \tag{6.7}$$

In particular we note that when we are in the symmetric case $\mu_1 = 0$, then $c = \hat{c}$ and $m_k = 0$ when k is odd and

$$m_k = \sum_{r=0}^{k} \binom{k}{r} \frac{(-x_0)^{k-r}\Gamma(c+r,\rho x_0)}{\rho^r \Gamma(c,\rho x_0)}.$$

6.5 Liouville transformation

We begin by finding the Lamperti transform of X, so that the transformed process $Z = z(X)$ has unit diffusion coefficient. The appropriate transform is

$$z(x) = \begin{cases} \frac{2}{\sigma_2^2}(\sqrt{(\sigma_1^2 + \sigma_2^2 x)} - \sigma_1), & x \geq 0, \\ \\ -\frac{2}{\sigma_2^2}(\sqrt{(\sigma_1^2 - \sigma_2^2 x)} - \sigma_1), & x \leq 0. \end{cases}$$

The inverse transform is given by

$$x(z) = \begin{cases} \sigma_1 z + \frac{1}{4}\sigma_2^2 z^2, & z \geq 0, \\ \\ \sigma_1 z - \frac{1}{4}\sigma_2^2 z^2, & z \leq 0. \end{cases}$$

Note that $\mathrm{sgn}(z(x)) = \mathrm{sgn}(x)$ for all $x \in \mathbf{R}$. The drift coefficient of Z in

$$dZ_t = b_Z(Z_t)dt + dW_t$$

is given by

$$b_Z(z) = \begin{cases} \frac{(\mu_1 - \frac{1}{4}\sigma_2^2) - \mu_2\sigma_1 z - \frac{1}{4}\mu_2\sigma_2^2 z^2}{\sigma_1 + \frac{1}{2}\sigma_2^2 z}, & z \geq 0 \\ \\ \frac{(\mu_1 + \frac{1}{4}\sigma_2^2) - \mu_2\sigma_1 z + \frac{1}{4}\mu_2\sigma_2^2 z^2}{\sigma_1 - \frac{1}{2}\sigma_2^2 z}, & z \leq 0 \end{cases}$$

The Schrödinger potential in this case is $Q(z) = U(x(z))$, where

$$U(x) = \frac{1}{32\sigma^2(x)} \begin{cases} A + Bx + Cx^2, & x \geq 0, \\ \\ \hat{A} + Bx + Cx^2, & x \leq 0, \end{cases}$$

$$A = 16\mu_1^2 - 16\mu_2\sigma_1^2 - 16\mu_1\sigma_2^2 + 3\sigma_2^4, \quad \hat{A} = 16\mu_1^2 - 16\mu_2\sigma_1^2 + 16\mu_1\sigma_2^2 + 3\sigma_2^4,$$

$$B = -32\mu_1\mu_2, \qquad\qquad C = 16\mu_2^2.$$

Assume that $\mu_2 \neq 0$. Then, although $z(\mathbf{R}) = \mathbf{R}$, we have $U(\pm\infty) = \infty$ which means that the boundaries $\pm\infty$ are both non-oscillatory and our process falls into Linetsky's spectral category I. The spectrum is discrete, with a strictly decreasing sequence of eigenvalues of A accumulating at $-\infty$.

In the special case $\mu_2 = 0$, we have $U(\pm\infty) = 0$, which means that the process is spectral category III. The spectrum is purely continuous, and contained in $(-\infty, 0]$. The SL equation is oscillatory for $\lambda \in (-\infty, 0)$ and non-oscillatory in $[0, \infty)$. We shall find that, as a function of the spectral parameter λ, the

Green function has a branch cut along the negative half-line.

6.6 Dynamic probabilities

This time the dynamic moment evolution is not so tractable as it was for the arithmetic-geometric hybrid. However, we may still write down the first moment

$$EX_t = \frac{\mu_1}{\mu_2}(1 - e^{-\mu_2 t}).$$

We move on to obtaining expressions for the transition densities. The case $\mu_1 = 0$ is notationally simplest, and for this reason we shall study it in detail. Expressions for the general case are obtained in essentially the same way, but the formulae are lengthier, because the transformation[2]

$$z(x) = \rho(x_0 + |x|)$$

works differently on the separate intervals $(-\infty, 0)$ and $(0, \infty)$.

Mean-reverting case $\mu_2 > 0$.

The SL equation $(A - \lambda)u = 0$ is solved by the functions

$$M\left(\frac{\lambda}{\mu_2}, c, \rho(x_0 + |x|)\right) \quad \text{and} \quad U\left(\frac{\lambda}{\mu_2}, c, \rho(x_0 + |x|)\right)$$

where $M(a, c, z)$ and $U(a, c, z)$ are Kummer's and Tricomi's confluent hypergeometric functions respectively (see Abramowitz & Stegun [2] for this standard notation). For abbreviation, we shall write

$$M(\lambda, x) = M\left(\frac{\lambda}{\mu_2}, c, \rho(x_0 + x)\right), \quad U(\lambda, x) = U\left(\frac{\lambda}{\mu_2}, c, \rho(x_0 + x)\right).$$

The function $U(a, c, z)$ then equals 0 at $z = +\infty$ (i.e. $x = \pm\infty$), and we choose the fundamental solutions of the SL equation as follows:

$$\psi_\lambda(x) = \begin{cases} U(\lambda, -x), & x \le 0 \\ C_1(\lambda)M(\lambda, x) + C_2(\lambda)U(\lambda, x), & x \ge 0 \end{cases}$$

where the constants $C_1(\lambda)$ and $C_2(\lambda)$ are chosen so that ψ_λ is a continuously differentiable function on $\mathbf{R}$ (i.e. choose the constants so that ψ_λ and ψ'_λ are

[2]This is *not* the Lamperti transform.

continuous at zero). Because of the symmetry $\mu_1 = 0$, we may take also

$$\chi_\lambda(x) = \psi_\lambda(-x), \quad x \in \mathbf{R}.$$

The formulae given above must be modified in the general case $\mu_1 \in \mathbf{R}$, since the parameter c must be exchanged for $\hat{c}$ when constructing these solutions on $(-\infty, 0)$.

Next, we denote the classical Wronskian at zero

$$W_\lambda = M'(\lambda,0)U(\lambda,0) - M(\lambda,0)U'(\lambda,0).$$

Since, with our normalisation of the scale density, $s(0) = 1$, we have $w_\lambda = W_\lambda C_1(\lambda)$ for the scaled Wronskian (2.21). The constants are all given by

$$\begin{aligned} C_1(\lambda) &= W_\lambda^{-1}[-2U(\lambda,0)U'(\lambda,0)], && (6.8)\\ C_2(\lambda) &= W_\lambda^{-1}[M'(\lambda,0)U(\lambda,0) + M(\lambda,0)U'(\lambda,0)], && (6.9)\\ W_\lambda &= \rho\frac{\Gamma(c)}{\Gamma\left(\frac{\lambda}{\mu_2}\right)}e^{\rho x_0}(\rho x_0)^{-c}. && (6.10) \end{aligned}$$

For the Green function, there are six cases, dependent upon the relative positions on the real line of the three points x, y and 0, though because of the symmetries $G_\lambda(x,y) = G_\lambda(y,x)$ and $G_\lambda(x,y) = G_\lambda(x,-y)$, there are only really two cases that need to be computed, when 0 lies between x and y and when both x and y are to one side of 0. For $x > 0$, we have

$$G_\lambda(x,y) = \begin{cases} \frac{U(\lambda,x)U(\lambda,-y)}{-2U(\lambda,0)U'(\lambda,0)}, & y \le 0, \\ \frac{[C_1(\lambda)M(\lambda,y)+C_2(\lambda)U(\lambda,y)]U(\lambda,x)}{-2U(\lambda,0)U'(\lambda,0)}, & 0 \le y \le x, \\ \frac{[C_1(\lambda)M(\lambda,x)+C_2(\lambda)U(\lambda,x)]U(\lambda,y)}{-2U(\lambda,0)U'(\lambda,0)}, & x \le y, \end{cases} \qquad (6.11)$$

and for $x < 0$,

$$G_\lambda(x,y) = \begin{cases} \frac{[C_1(\lambda)M(\lambda,-x)+C_2(\lambda)U(\lambda,-x)]U(\lambda,-y)}{-2U(\lambda,0)U'(\lambda,0)}, & y \le x, \\ \frac{[C_1(\lambda)M(\lambda,-y)+C_2(\lambda)U(\lambda,-y)]U(\lambda,-x)}{-2U(\lambda,0)U'(\lambda,0)}, & x \le y \le 0, \\ \frac{U(\lambda,-x)U(\lambda,y)}{-2U(\lambda,0)U'(\lambda,0)}, & y \ge 0, \end{cases} \qquad (6.12)$$

where the expression in the denominator is

$$-2U(\lambda,0)U'(\lambda,0) = C_1(\lambda)W_\lambda = \frac{2\rho}{\mu_2}\lambda U\left(\frac{\lambda}{\mu_2}, c, c\right) U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right). \tag{6.13}$$

These formulae are also all valid for $x = 0$, though since this case leads to considerable simplification in the formulae, we quote it separately. We have

$$G_\lambda(0,y) = \frac{\mu_2}{2\rho}\frac{U\left(\frac{\lambda}{\mu_2}, c, \rho(x_0+|y|)\right)}{\lambda U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}. \tag{6.14}$$

or

$$G_\lambda(0,y)m(y) = \frac{1}{2x_0^c}\frac{e^{-\rho|y|}(x_0+|y|)^{c-1}U\left(\frac{\lambda}{\mu_2}, c, \rho(x_0+|y|)\right)}{\lambda U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}$$

From this, the eigenfunction expansion of the conditional density of X_t given $X_0 = 0$ is

$$p(t;0,y) = \sum_n r_n e^{-\lambda_n t} U\left(\frac{-\lambda_n}{\mu_2}, c, \rho(x_0+|y|)\right) e^{-\rho|y|}(x_0+|y|)^{c-1}. \tag{6.15}$$

Here $-\lambda_n$ are the zeros of the function $\lambda \mapsto \lambda U(\lambda/\mu_2+1, c+1, c)$ and r_n are constants which can be determined by contour integration around the eigenvalue $-\lambda_n$:

$$r_n = \frac{1}{2\pi i}\int_{\partial B(-\lambda_n,\varepsilon)} \frac{1}{2x_0^c \lambda U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}\, d\lambda$$

As $t \to \infty$, the contributions of higher terms decreases fast, and we are left mainly with the invariant distribution, which corresponds to the $n = 0$ term.

Momentum case $\mu_2 < 0$

In the momentum case, we introduce the positive rate parameter $\beta = -\rho$ and the variable $\zeta = -z = \beta(x_0 + |x|)$. Again assume $\mu_1 = 0$, so that $c = -\beta x_0 = -\zeta_0$. The analysis can proceed analogously to the mean-reverting case using instead the functions

$$\hat{M}(\lambda,x) = e^{-\beta|x|}M(c-\lambda/\mu_2, c, \zeta), \quad \hat{U}(\lambda,x) = e^{-\beta|x|}U(c-\lambda/\mu_2, c, \zeta).$$

The expression for the Green function in this case is

(a) for $x > 0$,

$$G_\lambda(x,y) = \begin{cases} \frac{\hat{U}(\lambda,x)\hat{U}(\lambda,-y)}{-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0)} & y \le 0 \\ \frac{\hat{U}(\lambda,x)[\hat{C}_1(\lambda)\hat{M}(\lambda,y)+\hat{C}_2(\lambda)\hat{U}(\lambda,y)]}{-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0)}, & 0 \le y \le x \\ \frac{[\hat{C}_1(\lambda)\hat{M}(\lambda,x)+\hat{C}_2\hat{U}(\lambda,x)]\hat{U}(\lambda,y)}{-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0)} & x \le y; \end{cases} \tag{6.16}$$

(b) for $x < 0$, similarly

$$G_\lambda(x,y) = \begin{cases} \frac{[\hat{C}_1(\lambda)\hat{M}(\lambda,-x)+\hat{C}_2\hat{U}(\lambda,-x)]\hat{U}(\lambda,-y)}{-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0)} & y \le x \\ \frac{\hat{U}(\lambda,-x)[\hat{C}_1(\lambda)\hat{M}(\lambda,-y)+\hat{C}_2(\lambda)\hat{U}(\lambda,-y)]}{-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0)}, & x \le y \le 0 \\ \frac{\hat{U}(\lambda,-x)\hat{U}(\lambda,y)}{-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0)} & y \ge 0; \end{cases} \tag{6.17}$$

where this time, (note: $\beta x_0 = -c$)

$$-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0) = 2\beta U\left(c - \frac{\lambda}{\mu_2}, c, -c\right) U\left(c - \frac{\lambda}{\mu_2}, c+1, -c\right) \tag{6.18}$$

$$\hat{C}_1(\lambda) = \hat{W}_\lambda^{-1}[-2\hat{U}(\lambda,0)\hat{U}'(\lambda,0)] \tag{6.19}$$

$$\hat{C}_2(\lambda) = \hat{W}_\lambda^{-1}[\hat{M}'(\lambda,0)\hat{U}(\lambda,0) + \hat{M}(\lambda,0)\hat{U}'(\lambda,0)] \tag{6.20}$$

$$\hat{W}_\lambda = \beta\frac{\Gamma(c)}{\Gamma\left(\frac{\lambda}{\mu_2}\right)} e^{\beta x_0}(\beta x_0)^{-c} \tag{6.21}$$

Comparing with (6.13), we are missing the factor λ in (6.18), so $0 \notin \sigma(A)$. The spectrum is $\sigma(A) = (-\lambda_n)_{n\ge1}$, where $0 < \lambda_1 < \lambda_2 < \cdots$, consisting of the zeros of the functions $U(c-\lambda/\mu_2, c, -c)$ and $U(c-\lambda/\mu_2, c+1, -c)$ of λ. Note that, in the case $x = 0$, these expressions simplify to

$$G_\lambda(0,y) = \frac{e^{-\beta|y|}U\left(c - \frac{\lambda}{\mu_2}, c, \beta(x_0+|y|)\right)}{2\beta U\left(c - \frac{\lambda}{\mu_2}, c+1, -c\right)} \tag{6.22}$$

or

$$G_\lambda(0,y)m(y) = \frac{(x_0+|y|)^{c-1}U\left(c - \frac{\lambda}{\mu_2}, c, \beta(x_0+|y|)\right)}{2|\mu_2|x_0^c\, U\left(c - \frac{\lambda}{\mu_2}, c+1, -c\right)}.$$

The eigenfunction expansion of the conditional density of X_t given $X_0 = 0$ is then

$$p(t;0,y) = \sum\nolimits_n r_n e^{-\lambda_n t} U\left(c + \frac{\lambda_n}{\mu_2}, c, \beta(x_0 + |y|)\right)(x_0 + |y|)^{-\beta-1} \qquad (6.23)$$

where the λ_n are zeros of the function $\lambda \mapsto U(c - \lambda/\mu_2, c+1, -c)$ and r_n are constants which can be determined by contour integration around each eigenvalue $-\lambda_n$. Zero is not an eigenvalue: as t increases, the contribution of all terms for fixed $y \in \mathbf{R}$ converges to zero because, eventually, all probability mass escapes any bounded interval.

Borderline case $\mu_2 = 0$

In the case $\mu_2 = 0$, the solution of the SL equation is more conveniently expressed in terms of Bessel functions, rather than the confluent hypergeometric functions. For this case, we remove our symmetry assumption $\mu_1 = 0$. Recall that

$$c = \frac{2\mu_1}{\sigma_2^2}, \quad s(x) = \left(1 + \frac{x}{x_0}\right)^{-c\,\mathrm{sgn}(x)}, \quad m(x) = \frac{2}{\sigma_1^2}\left(1 + \frac{x}{x_0}\right)^{c\,\mathrm{sgn}(x)-1}.$$

The first fundamental solution of the equation

$$Au(x) := \frac{1}{2}(\sigma_1^2 + \sigma_2^2|x|)u''(x) + \mu_1 u'(x) = \lambda u(x)$$

is given by

$$\psi_\lambda(x) = \begin{cases} (\sigma(x))^{1+c} K_{c+1}(z_x(\lambda)) & x \le 0 \\ (\sigma(x))^{1-c}[C_1(\lambda) I_{c-1}(z_x(\lambda)) + C_2(\lambda) K_{c-1}(z_x(\lambda))] & x \ge 0 \end{cases}$$

where

$$z_x(\lambda) = \frac{2}{\sigma_2^2}\sqrt{2\lambda}\,\sigma(x), \qquad (6.24)$$

$$C_1(\lambda) = \frac{2}{\sigma_2^2}\sqrt{2\lambda}\,\sigma_1^{2c+1}[K_{c-1}(z_0(\lambda)) + K_{c+1}(z_0(\lambda))]K_c(z_0(\lambda)), \qquad (6.25)$$

$$C_2(\lambda) = \frac{2}{\sigma_2^2}\sqrt{2\lambda}\,\sigma_1^{2c+1}[I_c(z_0(\lambda))K_{c+1}(z_0(\lambda)) - I_{c-1}(z_0(\lambda))K_c(z_0(\lambda))]. \qquad (6.26)$$

and where $I_\nu(z)$ and $K_\nu(z)$ refer respectively to the modified Bessel functions of the first and second kinds respectively: see Abramowitz & Stegun [2] for this standard notation. The second fundamental solution of the SL equation

is

$$\chi_\lambda(x) = \psi_\lambda(-x), \quad x \in \mathbf{R}.$$

In this case, the scaled Wronskian is the function of only λ given by

$$w_\lambda = 2\sigma_1^{2c+1}\sqrt{2\lambda}\; K_c(z_0(\lambda))K_{c+1}(z_0(\lambda)). \tag{6.27}$$

We then obtain a Green function in the cases

(a) $x \le 0$,

$$G_\lambda(x,y) = \begin{cases} \dfrac{\left(1-\frac{x}{x_0}\right)^{\frac{1-c}{2}}[C_1(\lambda)I_{c-1}(z_x(\lambda))+C_2(\lambda)K_{c-1}(z_x(\lambda))]\left(1-\frac{y}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_y(\lambda))}{2\sigma_1^{2c-1}\sqrt{2\lambda}\; K_c(z_0(\lambda))K_{c+1}(z_0(\lambda))}, & \text{(i)} \\ \dfrac{\left(1-\frac{x}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_x(\lambda))\left(1-\frac{y}{x_0}\right)^{\frac{1-c}{2}}[C_1(\lambda)I_{c-1}(z_y(\lambda))+C_2(\lambda)K_{c-1}(z_y(\lambda))]}{2\sigma_1^{2c-1}\sqrt{2\lambda}\; K_c(z_0(\lambda))K_{c+1}(z_0(\lambda))}, & \text{(ii)} \\ \dfrac{\sigma_1}{2}\dfrac{\left(1-\frac{x}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_x(\lambda))\left(1+\frac{y}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_y(\lambda))}{\sqrt{2\lambda}\; K_c(z_0(\lambda))K_{c+1}(z_0(\lambda))}, & \text{(iii)}; \end{cases} \tag{6.28}$$

in the cases

(i) $y \le x$,

(ii) $x \le y \le 0$,

(iii) $y \ge 0$.

(b) $x \ge 0$

$$G_\lambda(x,y) = \begin{cases} \dfrac{\left(1+\frac{x}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_x(\lambda))\left(1-\frac{y}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_y(\lambda))}{2\sigma_1^{2c-1}\sqrt{2\lambda}\; K_c(z_0(\lambda))K_{c+1}(z_0(\lambda))}, & \text{(i)} \\ \dfrac{\left(1+\frac{x}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_x(\lambda))\left(1+\frac{y}{x_0}\right)^{\frac{1-c}{2}}[C_1(\lambda)I_{c-1}(z_y(\lambda))+C_2(\lambda)K_{c-1}(z_y(\lambda))]}{2\sigma_1^{2c-1}\sqrt{2\lambda}\; K_c(z_0(\lambda))K_{c+1}(z_0(\lambda))}, & \text{(ii)} \\ \dfrac{\left(1+\frac{x}{x_0}\right)^{\frac{1-c}{2}}[C_1(\lambda)I_{c-1}(z_x(\lambda))+C_2(\lambda)K_{c-1}(z_x(\lambda))]\left(1+\frac{y}{x_0}\right)^{\frac{1+c}{2}}K_{1+c}(z_y(\lambda))}{2\sigma_1^{2c+2}\sqrt{2\lambda}\; K_c(z_0(\lambda))K_{c+1}(z_0(\lambda))}, & \text{(iii)} \end{cases}$$

in each of the cases

(i) $y \le 0$,

(ii) $0 \le y \le x$,

(iii) $y \ge x$.

(c) In particular, for $x = 0$, we can write

$$G_\lambda(0,y) = \frac{1}{2}\sigma_1 \frac{\left(1+\frac{|y|}{x_0}\right)^{\frac{1+c}{2}} K_{1+c}(z_y(\lambda))}{\sqrt{2\lambda}\, K_c(z_0(\lambda))} \tag{6.29}$$

and

$$G_\lambda(0,y)m(y) = \frac{x_0^{\frac{c+1}{2}}(x_0+|y|)^{-\frac{c+1}{2}} K_{1+c}(2\delta(\lambda)\sqrt{x_0+|y|})}{\sigma_1\sigma_2\delta(\lambda)K_c(2\delta(\lambda)\sqrt{x_0})} \tag{6.30}$$

where

$$\delta(\lambda) = \sqrt{\frac{2\lambda}{\sigma_2^2}}$$

6.7 Risk measures

6.7.1 Stationary VaR and ES

To calculate the VaR and ES for the invariant distribution (6.6), we work out first how to calculate them for a gamma distribution. Let γ be Gamma(c,ρ) distributed. The tail probabilities are by definition

$$\overline{G}(x;c,\rho) = Q(c,\rho x) = \frac{\Gamma(c,\rho x)}{\Gamma(c)}, \tag{6.31}$$

and the quantile function is therefore

$$\mathrm{VaR}_\alpha(\gamma) = G^{-1}(\alpha;c,\rho) = \rho^{-1}Q^{-1}(c,1-\alpha), \tag{6.32}$$

where the inverse function Q^{-1} is known, and implementable for example in *Mathematica* as InverseGammaRegularized. Denoting this quantity ξ_α, we have[3]

$$\mathrm{ES}_\alpha(\gamma) = \frac{1}{\rho}\left[c + \frac{\xi_\alpha g(\xi_\alpha;c,\rho)}{1-\alpha}\right], \tag{6.33}$$

where $g(\cdot;c,\rho)$ is the pdf of γ.

Suppose now that X is a random variable with the density $x \mapsto \pi(-x)$, where π is given by (6.6), representing the loss made on an investment (with a profit made if $X < 0$). Conditional on making a loss (an event with probability q), the distribution of X is a left shifted gamma distribution with parameters $\hat{c}$, λ and x_0. Therefore, at a given probability level α, the VaR of X is just the VaR of the left truncated gamma distribution at the level α/q, then relocated

[3] Klugman *et al.* [78]

by subtracting x_0. We have

$$\mathrm{VaR}_\alpha(X) = \xi_\alpha = \rho^{-1} Q^{-1}\left(\hat{c}, \left(\frac{1-\alpha}{q}\right) Q(\hat{c}, \rho x_0)\right) - x_0. \tag{6.34}$$

A calculation then shows that ES for this distribution is given by

$$\mathrm{ES}_\alpha(X) = \frac{1}{\rho}\left[\hat{c} + \frac{(x_0+\xi_\alpha) g(x_0+\xi_\alpha; \hat{c}, \rho)}{Q(\hat{c}, \rho(x_0+\xi_\alpha))} - \rho x_0\right]. \tag{6.35}$$

6.7.2 Dynamic VaR and ES

We would like to know (2.23) for

$$u_k(x) = 1_{(-\infty,-\xi)}(x) \cdot (-x)^k, \quad \text{for } \xi > 0 \text{ and small } k \in \mathbf{Z}_+. \tag{6.36}$$

Finding the VaR is the problem of calculating the value of ξ for which $P_t u_0(x) = \alpha$. Once that is calculated, $ES_\alpha(X_t) = P_t u_1(x)/(1-\alpha)$ when ξ in (6.36) is the VaR. Higher values of k give higher conditional moments, which can be used to give more detailed information about the behaviour of the distribution in the left tail.

Mean-reverting case $\mu_2 > 0$

The first step is to find $R_\lambda u(x)$. From (2.29), a calculation gives

$$R_\lambda u_0(x) = \frac{1}{2} e^{-\rho\xi}\left(1 + \frac{\xi}{x_0}\right)^c \frac{U(\lambda, x) U\left(\frac{\lambda}{\mu_2}+1, c+1, \rho(x_0+\xi)\right)}{\lambda U(\lambda, 0) U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}. \tag{6.37}$$

The probability $P^x\{X_t < -\xi\}$ is then found by (2.23), and when the probability α is given the VaR is the inversion problem of finding ξ so that $P_t u(x) = \alpha$.

An integration by parts, again using (2.30) yields the relevant integral for the expected shortfall:

$$\begin{aligned} &\int_{-\infty}^{-\xi} (-y) \frac{1}{2x_0^c} e^{\rho y} (x_0 - y)^{c-1} U\left(\frac{\lambda}{\mu_2}, c, \rho(x_0 - y)\right) dy \\ &= \frac{1}{2} e^{-\rho\xi}\left(1 + \frac{\xi}{x_0}\right)^c \left[\xi U\left(\frac{\lambda}{\mu_2}+1, c+1, \eta\right) + (x_0+\xi) U\left(\frac{\lambda}{\mu_2}+2, c+2, \eta\right)\right] \end{aligned}$$

where $\eta = \rho(x_0 + \xi)$. If $\xi > 0$ is the value-at-risk at a certain level α, then the ES is calculated as $P_t u_1(x)$. The resolvent $R_\lambda u_1(x)$ is found now by multiplying

the above expression by

$$\frac{U\left(\frac{\lambda}{\mu_2}, c, \rho(x_0+x)\right)}{\lambda U\left(\frac{\lambda}{\mu_2}, c, c\right) U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}.$$

Then the ES may be evaluated numerically via (2.23). A little simplification occurs when $x = 0$. It is the function

$$\begin{aligned}\lambda \;\mapsto\; & \frac{1}{2}e^{-\rho\xi}\left(1+\frac{\xi}{x_0}\right)^c \times \\ & \times \frac{\left[\xi U\left(\frac{\lambda}{\mu_2}, c+1, \rho(x_0+\xi)\right) + (x_0+\xi)U\left(\frac{\lambda}{\mu_2}+2, c+2, \rho(x_0+\xi)\right)\right]}{\lambda U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}\end{aligned}$$

on which we must perform the Laplace transform inversion (2.23). The expected shortfall is then found by dividing by $(1-\alpha)$.

For higher powers k, one may use

$$\begin{aligned}\int_{-\infty}^{-\xi} (-y)^k \; & \frac{1}{2x_0^c}e^{\rho y}(x_0-y)^{c-1}U\left(\frac{\lambda}{\mu_2}, c, \rho(x_0-y)\right)dy \\ = \; & \frac{e^{\rho x_0}}{2x_0^c\rho^{c+k}}\int_{\rho(x_0+\xi)}^{\infty}(z-\rho x_0)^k e^{-z}z^{c-1}U\left(\frac{\lambda}{\mu_2}, c, z\right)dz. \quad (6.38)\end{aligned}$$

which can be integrated by parts.

As a final note, (6.38) may be used to find the variance of this distribution. Taking $\xi = 0$ and $k = 2$, repeated integration by parts leads to

$$\int_{-\infty}^{0} y^k G_\lambda(0,y)\, m(y)\, dy = \frac{x_0^2 U\left(\frac{\lambda}{\mu_2}+3, c+3, c\right)}{\lambda U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}. \quad (6.39)$$

Because of the symmetry, the same integral over $(0,\infty)$ is the same. Therefore, the variance may be found as

$$\mathrm{var}\,(X_t) = E^0[X_t^2] = 2x_0^2\,\frac{1}{2\pi i}\int_{\mathcal{C}} e^{\lambda t}\frac{U\left(\frac{\lambda}{\mu_2}+3, c+3, c\right)}{\lambda U\left(\frac{\lambda}{\mu_2}+1, c+1, c\right)}\,d\lambda. \quad (6.40)$$

Momentum trading $\mu_2 < 0$

We have not found useful expressions for VaR and ES in this case. The reason for this is that the speed measure is not finite in any neighbourhood of $-\infty$ and, therefore, the functions u_k defined in (6.36) do not belong to the Hilbert

space $L^2(I,m)$ in which we are working, so the resolvent operators R_λ may not be applied to u_k. This does not mean that the VaR or ES are not defined. It means that our approach of performing the y-integration first, followed by the numerical Laplace transform is not valid. Instead the transition density must be obtained first, via the numerical integration (2.24) using (6.29). The VaR and ES are finally computed by integration in the y-domain.

Borderline case $\mu_2 = 0$

The same difficulty is present in the borderline cases with $\mu_1 \leq 0$, but when $\mu_1 > 0$, the speed measure is finite in every neighbourhood of $-\infty$ which is not also a neighbourhood of $+\infty$. We recall that

$$m(x) = \frac{2}{\sigma_2^2} \begin{cases} \left(1 + \frac{x}{x_0}\right)^{c-1}, & x \leq 0, \\ \left(1 - \frac{x}{x_0}\right)^{-c-1}, & x \leq 0. \end{cases}$$

Therefore, $u_0 \in L^2(\mathbf{R}, m)$ and we may calculate the probability of such a neighbourhood by first integrating the Green function over it, then performing the numerical Laplace transform inversion. If in addition $c > 1$, i.e. $\mu_1 > \sigma_2^2/2$, then also $u_1 \in L^2(\mathbf{R}, m)$.

With the notation

$$\delta(\lambda) = \sqrt{\frac{2\lambda}{\sigma_2^2}}$$

that we introduced in (6.30), we have

$$\int_{-\infty}^{-\xi} G_\lambda(0,y) m(y) dy$$

$$= \frac{x_0^{\frac{1+c}{2}} (x_0+\xi)^{-c/2} \left\{ 2\Gamma(1+c) K_c\big(2\delta(\lambda)(x_0+\xi)\big) + \frac{\pi}{\sin(\pi c)} \big[\delta(\lambda)(x_0+\xi)\big]^c \right\}}{2\sigma_1\sigma_2\Gamma(1+c)\delta(\lambda) K_c\big(2\delta(\lambda)\sqrt{x_0}\big)}$$

and

$$\int_{-\infty}^{-\xi} y G_\lambda(0,y) m(y) dy$$

$$= \frac{x_0^{\frac{c+1}{2}} \left\{ \begin{array}{l} \frac{(1+c)\pi}{\sin(\pi c)} \left(\delta(\lambda)\sqrt{x_0+\xi}\right)^c \left(c + x_0\delta^2(\lambda)\right) \\ -2\Gamma(c+2)\delta(\lambda) \Big[\sqrt{x_0+\xi} K_{c-1}(2\delta(\lambda)\sqrt{x_0+\xi}) \\ \qquad + \xi\delta(\lambda) K_c(2\delta(\lambda)\sqrt{x_0+\xi}) \Big] \end{array} \right\}}{2(x_0+\xi)^{c/2}\sigma_1\sigma_2\Gamma(c+2)\delta^4(\lambda) K_c(2\delta(\lambda)\sqrt{x_0})}$$

Note that, in this case, the function $y \mapsto y^2$ never belongs to $L^2(\mathbf{R}, m)$, so the variance cannot be evaluated in this way.

Chapter 7

Model fitting and parameter estimation

Now it is time to consider the problem of estimating the parameters of these models from historical data. We begin by performing a classical maximum likelihood estimation on daily returns, treating them as i.i.d. This allows a comparison between competing models, although it is limited in that it ignores dependency between successive observations. We fit the data to the invariant distributions from either model (given by (5.6) and (6.6)), also comparing with a fitted Gaussian distribution and also a fitted Gaussian mixture with random volatility. In the latter model, the volatility may take one of two values, σ_1 or σ_2 and we denote this mixture distribution

$$\mathrm{GMix}(\mu, \sigma_1^2, \sigma_2^2, \theta) = \theta N(\mu, \sigma_1^2) + (1-\theta) N(\mu, \sigma_2^2).$$

The data are the same seven world indices used in the introduction for the Jarque-Bera test. This time, the observations are the daily returns between 2nd January 2001 and 31st December 2012. To understand why we might like to fit non-constant volatility, look at figure 7.1. The absolute returns of three of these indices have been plotted against time and we see in this figure periods of high volatility and low volatility, the low volatility appearing to occur between 2004 and 2007.

The findings of the likelihood test are presented in table 7.1. The likelihood functions are denoted L_T, L_Γ, L_G and L_{GM} respectively for the aBm-gBm invariant distribution (5.6), the aBm-CIR invariant distribution (6.6), the Gaussian model $N(\mu, \sigma^2)$ and the mixture of Gaussians $\mathrm{GMix}(\mu, \sigma_1^2, \sigma_2^2, \theta)$. We observe that the Gaussian mixture gives a much better fit than the constant-vol Gaussian, but is in every given case beaten by at least one of the Student's

Figure 7.1: Absolute returns of seven world indices between 2001 and 2013

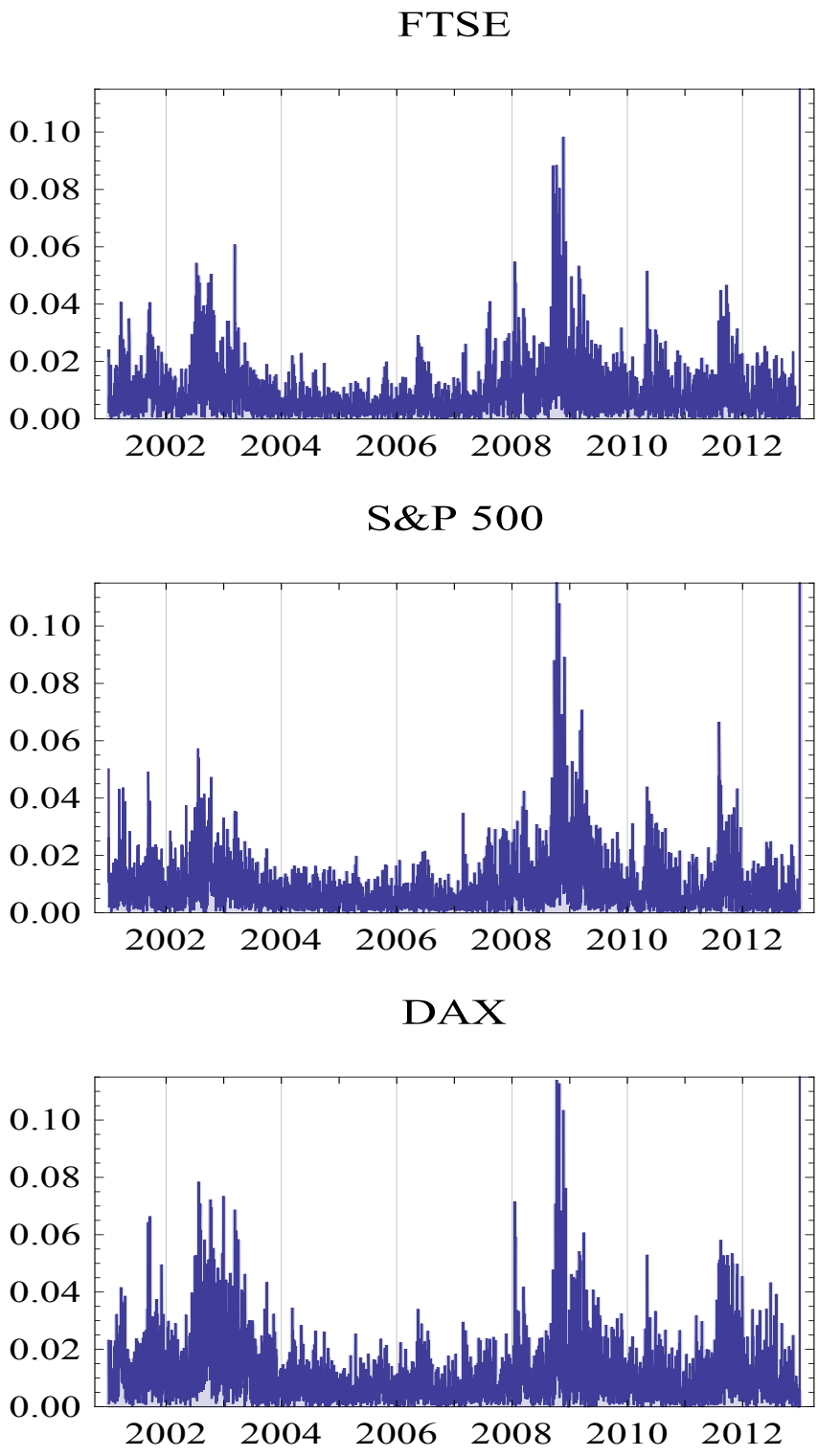


Table 7.1: Maximum log-likelihood, AIC and BIC per observation of Gaussian, Student's t, hybrid gamma and Gaussian mixture distributions fitted to seven financial time-series, running between 01/01/2001 and 31/12/2012. The best fit for each series is shown in bold. (Numbers are shown to 3 d.p. - no two of the numbers are actually equal.)

	Criterion (quantity per	**Distribution**			
Index	observation)	Gaussian	t	Hybrid gamma	Gaussian mixture
	Log-likelihood	2.922	**3.027**	3.022	3.022
FTSE 100	AIC	-5.842	-6.052	-6.042	-6.042
($N = 3029$)	BIC	-5.838	-6.046	-6.036	-6.034
	Log-likelihood	2.889	**3.023**	3.022	3.014
S&P 5000	AIC	-5.777	-6.044	-6.043	-6.025
($N = 3016$)	BIC	-5.773	-6.038	-6.037	-6.017
	Log-likelihood	2.699	2.785	**2.787**	2.783
DAX	AIC	-5.397	-5.567	-5.573	-5.563
($N = 3060$)	BIC	-5.393	-5.562	-5.567	-5.555
	Log-likelihood	2.732	**2.820**	2.817	2.817
Cac 40	AIC	-5.462	-5.638	-5.631	-5.632
($N = 3070$)	BIC	-5.459	-5.632	-5.626	-5.624
	Log-likelihood	2.725	2.838	**2.842**	2.826
Hang Seng	AIC	-5.449	-5.673	-5.682	-5.648
($N = 2996$)	BIC	-5.445	-5.667	-5.676	-5.640
	Log-likelihood	2.731	**2.807**	2.777	2.804
Nikkei 225	AIC	-5.461	-5.612	-5.551	-5.605
($N = 2943$)	BIC	-5.457	-5.606	-5.545	-5.597
	Log-likelihood	2.550	**2.603**	2.598	2.601
BVSP	AIC	-5.098	-5.205	-5.193	-5.200
($N = 2965$)	BIC	-5.094	-5.199	-5.187	-5.191

t-distribution and the hybrid gamma distribution.

We should note that the usual Gaussian model with constant volatility has only two parameters, our invariant distributions (5.6) and (6.6) have three parameters, while the Gaussian mixture has four. We quote the values of the information criteria defined in (2.39) also for these models fitted to these datasets, but note that with such a large sample size relative to the number of parameters, the conclusions we reach are the same as if we just use maximum likelihood. In most cases, the equilibrium distribution arising in the hybrid aBm-gBm fits best, in the other cases, it is that arising from the hybrid aBm-CIR process.

These likelihoods were maximized over the whole parameter space, so the values of the parameters for which they were largest provide us with an estimate of the actual parameter values. However, this does not complete the parameter estimation problem for these models because, while the dynamic models (5.1) and (6.1) are four-parameter model (take correlation $r = 0$ in

(5.1)), their equilibria have only three free parameters, the other one from the dynamic model describing the rate at which the process approaches its equilibrium. (Note also that fitting the invariant distribution from the diffusion model automatically assumes $\mu_2 > 0$.)

To complete the parameter estimation, I therefore suggest using a linear estimating function to obtain the parameters in the drift coefficient as follows. Recall from section 2.8.2 that the usual construction of a martingale estimating function is

$$F_N(\theta) = \sum_{n=1}^{N} f(Y_{n-1}, Y_n, \theta), \qquad \text{where} \qquad f(x, y, \theta) = a(x, \theta)h(x, y, \theta),$$

a being a (matrix-valued) weight function, h a function satisfying the conditional moment condition

$$\int_{\mathbf{R}} h(x, y, \theta)p(\Delta; x, y, \theta)dy = 0.$$

Normally one chooses $h(x, y, \theta) = g(y, \theta) - E_\theta^x[g(Y_1, \theta)]$, to ensure that the conditional moment condition is satisfied. If we choose g linear in y, we call F_N a *linear estimating function*; this method is applicable if the first conditional moment is known in closed form. Indeed, we choose

$$h(x, y, \theta) = y - m(x, \theta)$$

where

$$m(x, \theta) = E_\theta^x X_\Delta = \frac{\mu_1}{\mu_2} + \left(x - \frac{\mu_1}{\mu_2}\right) e^{-\mu_2 t}$$

is the conditional mean of X_Δ. We know $m(x, \theta)$ in closed form just because of the simple form of the drift coefficient $b(x, \theta)$ in (5.1) and(6.1). Here $\Delta > 0$ is the fixed time-lag between observations, taken in our data as $1/252$ for daily returns. To simplify matters, the weight matrix may be taken as $a(x, \theta) = (1, -x, 0, 0)^T$, which means that it does not depend on θ, and that we are only considering estimating the parameters occurring in the top two lines of the equation $F_N(\theta) = 0$. (This is a system of four equations in four unknowns, but we have reduced it to a system of two equations in two unknowns.)

Consider the time-series $Y = (Y_n) = (X_{n\Delta})$ for the fixed time-lag Δ. We may estimate the parameters μ_1 and μ_2 in the drift coefficient for Y. We may also relocate the time-series, and perform the estimation for the time-series $Y - b = (Y_0 - b, Y_1 - b, \ldots, Y_N - b)$ where b is a location parameter. b then determines μ_1, leaving μ_2 unaffected. This allows us to justify using a

Index	Model Parameters				Descriptive Parameters	
	b	μ_2	σ_1	σ_2	ν	x_0
FTSE 100	-0.18	0.97	0.01	0.98	3.03	0.01
S & P 500	-0.08	0.87	0.01	1.03	2.63	0.01
DAX	-0.13	0.4	0.01	0.62	3.07	0.02
Cac 40	-0.42	0.84	0.02	0.86	3.27	0.02
Hang Seng	0.21	0.29	0.01	0.54	3.01	0.02
N225	-0.24	0.64	0.01	0.63	4.26	0.02
BVSP	1.65	0.13	0.01	0.27	4.66	0.03

Table 7.2: Preliminary estimates of the parameters in the aBm-gBm model

Index	Model Parameters				Descriptive Parameters		
	b	μ_2	σ_1	σ_2	ρ	c	x_0
FTSE 100	-0.18	0.97	0.01	0.13	110.21	0.98	0.01
S& P 500	-0.08	0.87	0.01	0.14	84.19	0.50	0.01
DAX	-0.13	0.4	0.01	0.10	86.30	0.95	0.01
Cac 40	-0.42	0.84	0.01	0.13	95.76	1.14	0.01
Hang Seng	0.21	0.29	0.01	0.09	80.88	0.70	0.01
N225	-0.24	0.64	0.02	0.09	162.45	6.08	0.04
BVSP	1.65	0.13	0.01	0.05	111.56	2.78	0.02

Table 7.3: Preliminary estimates of the parameters in the aBm-CIR model

model we have formulae for, by choosing b to fix $\mu_1 = 0$. (We have invariably found that estimates of the parameter $\mu = \mu_1/\mu_2$ in the static estimation are extremely small, which is consistent with this modification.) Combining this method for the drift and the static estimation for the remaining parameters, we can come up with the preliminary estimates of all the parameters shown in tables 7.2 and 7.3.

Since the parameter estimation has presented difficulties that have not so far been overcome, I suggest that the estimates so obtained must next be adjusted in order to use the model. I envisage this being achieved either by further statistical analysis of the time-series in question, or by adjusting to the current market conditions. One adjustment that can easily be made when the model implies a finite variance is to scale the parameters so as to match the model and observed standard deviations.

In the mean-reverting cases of the aBm-CIR model, the variance is obtained by (6.40). If the original model volatility was s_1 and the observed volatility is s_2, then the mappings

$$\rho \mapsto \rho\frac{s_1}{s_2}, \quad x_0 \mapsto x_0\frac{s_2}{s_1}$$

yield a model with volatility s_2. This adjusts the scale of the distribution without affecting the shape parameter c. The final estimates of the parameters this leads to are given in table 7.4.

In known cases of the aBm-gBm model ($\nu \in \{-4, -2, 0, 2\}$) the variance

Index	Model Parameters				Descriptive Parameters		
	b	μ_2	σ_1	σ_2	ρ	c	x_0
FTSE 100	-0.18	0.9677	0.2024	0.5337	6.7938	0.9771	0.1438
S& P 500	-0.08	0.8651	0.2075	0.6226	4.4638	0.4959	0.1111
DAX	-0.13	0.3991	0.2547	0.4830	3.4222	0.9518	0.2781
Cac 40	-0.42	0.8445	0.0517	0.2511	26.7913	1.1365	0.0424
Hang Seng	0.21	0.2936	0.2481	0.4765	2.5859	0.7010	0.2711
N225	-0.24	0.6444	0.2487	0.3384	11.2551	6.0783	0.5400
BVSP	1.65	0.1325	0.2985	0.3037	2.8728	2.7763	0.9664

Table 7.4: Adjusted estimates of the parameters in the aBm-CIR model

can be found by direct integration if it exists. This is computationally less demanding than an integration in the λ-domain. Simply increasing the x_0 parameter in this model by a factor of σ, while keeping the other parameters fixed, results in an increase in the standard deviation of the model by σ, hence the variance is scaled by the factor σ^2. In terms of the model parameters, this corresponds precisely to multiplying σ_1 by a factor of σ and leaving μ_2 and σ_2 untouched. We feel it is unnecessary to illustrate this simple adjustment to the values in table 7.2 with another table.

Chapter 8

Numerical illustrations

Our aim in this chapter is to illustrate the use of the theory developed so far. We shall use the formulae obtained for some of the Green functions in the aBm-CIR model and the explicit formulae we have found for the dynamic densities in the aBm-gBm models to plot the dynamic densities and make comparisons. For this approach to the aBm-CIR model we note that the dynamic density may be evaluated by a numerical integration in the λ-space. We have not yet actually quoted this formula:

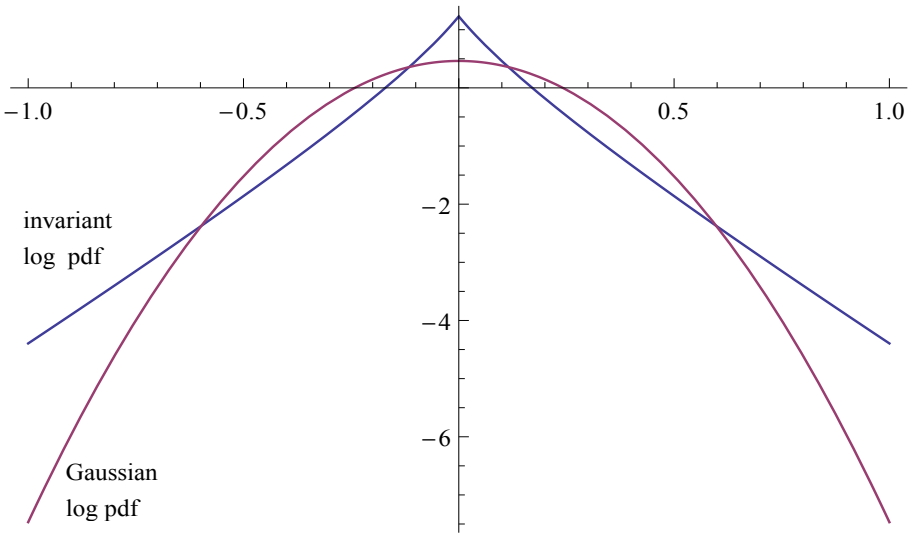
$$p_t(y) = p(t;0,y) = \frac{1}{2\pi i}\int_{\mathcal{C}} e^{\lambda t} G_\lambda(0,y)\, d\lambda\, m(y) \tag{8.1}$$

where $\mathcal{C}$ is taken from (2.26), though it is obvious from (2.23). Our choice of integration contour is simply for the speed of convergence: either of the contours (2.9) or (2.25) would do. Having considered the transition densities, we would also like to compute VaR and ES in each of the models. The results of the numerical integrations to find the risk measures are illustrated graphically in this chapter and tabulated in appendix A.

8.1 Arithmetic-CIR model

We begin with the aBm-CIR process. To make use of our formulae, we shall take $\mu_1 = 0$ and use the parameters obtained in fitting this model to the S & P 500 data. Thus we take $\mu_2 = 0.865103$, $\sigma_1 = 0.207513497$ and $\sigma_2 = 0.622578927$. The invariant distribution (6.6) has a mean of 0, and a variance which is easily calculated (see (6.7)) to be 0.0630402. We shall aim to compare the true pdf for time-increments $t = 1/252$ (daily returns), $t = 1/12$ (monthly returns) and $t = 1$ (annual returns) with each other and with the Gaussian density fitted to the same data, and the invariant distribution (6.6)

Figure 8.1: Log-pdf for Gaussian and aBm-CIR invariant distribution.



(equivalently obtained by setting $t = \infty$ in this model).

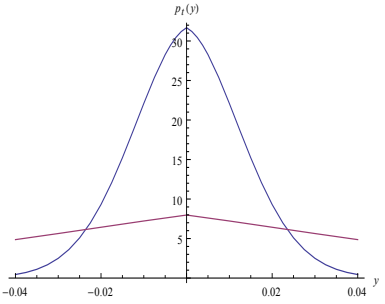
Before doing any computation of the dynamic densities, let us compare the invariant density with the Gaussian density of the same mean and variance. Figure 8.1 shows the logarithm of these pdfs. The invariant distribution attributes much more mass than the Gaussian does to the tails of the support. This is, of course, a desirable feature of asset return models.

Direct computation of dynamic pdf

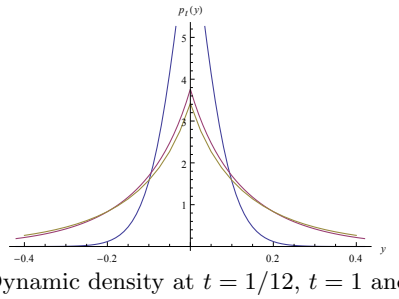
First, we wish to obtain an idea of the dynamic behaviour of the pdf, simply by directly evaluating (8.1) for $x = 0$ and for successive values of t. As we know, at $t = 0$, the distribution is just a spike at zero. And as t increases, the probability mass diffuses outwards along the line, converging with $t \to \infty$ to the invariant distribution. Figure 8.2 shows this happening.

Two plots are necessary to illustrate this development of the dynamic pdf

Figure 8.2: Plots of the pdf of X_t in the aBm-CIR model.

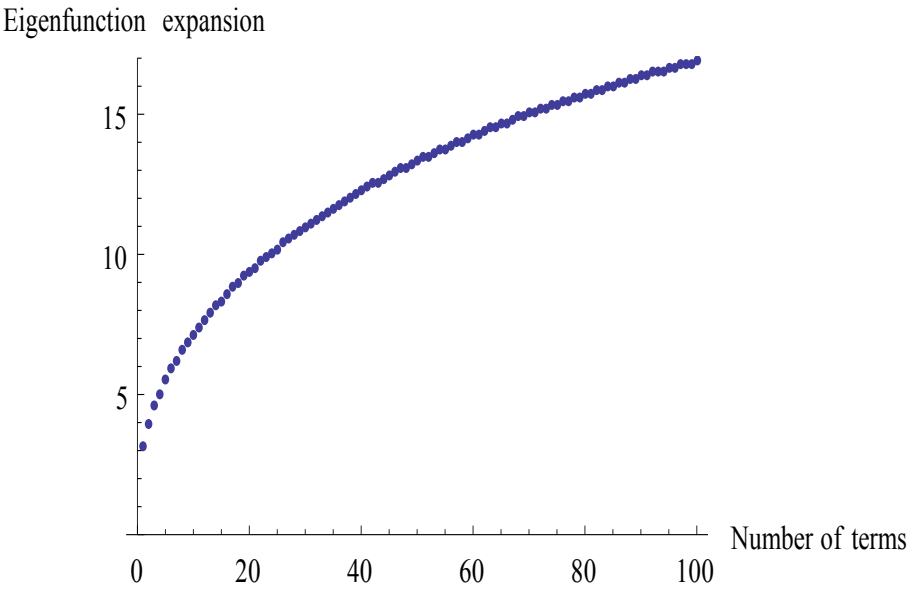


(a) Dynamic density at $t = 1/252$ and $t = 1/12$.



(b) Dynamic density at $t = 1/12$, $t = 1$ and the invariant density

Figure 8.3: Eigenfunction approximation against number of terms at $(t, y) = (1/252, 0.1)$.



because so much probability mass escapes the domain that the plot is capable of showing. This can be explained by the increase of the standard deviation of X_t with t, given in table A.1.

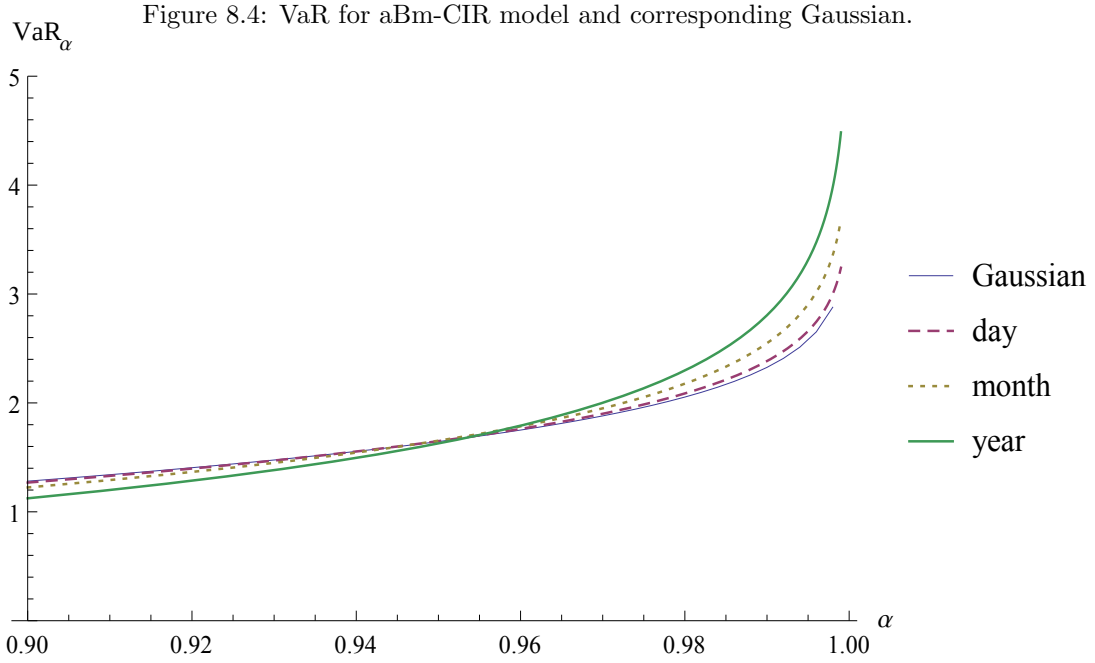
Eigenfunction expansion as approximation

The invariant distribution (6.6) is the 0th-order eigenfunction expansion. That is, from (6.15), we take only the $n = 0$ term. If this approximation is not good enough, then it is natural to try adding further terms. The convergence is actually extremely slow and this method of approximation is not useful. Figure 8.3 shows the convergence of the approximation at y=0.01. The actual value is $p_{1/252}(0.01) = 22.0938$. For the approximation to be of any use, a huge number of terms would be needed in the approximation.

Value-at-risk

In order to be able to compare like-for-like distributions, we standardise the distributions, by setting $\mathrm{var}\,(X_t) = 1$. This corresponds simply to dividing by the model standard deviation. To obtain actual values, these are then multiplied by the actual observed standard deviation. To give us an idea of the sort of numbers that we are dealing with, it is very straightforward to calculate the VaR implied by the invariant distribution π. From (6.34), we can calculate the values in the last column in table A.2.

Figure 8.4: VaR for aBm-CIR model and corresponding Gaussian.



For calculating the dynamic VaR, we first calculate the distribution function, which is found from the formula

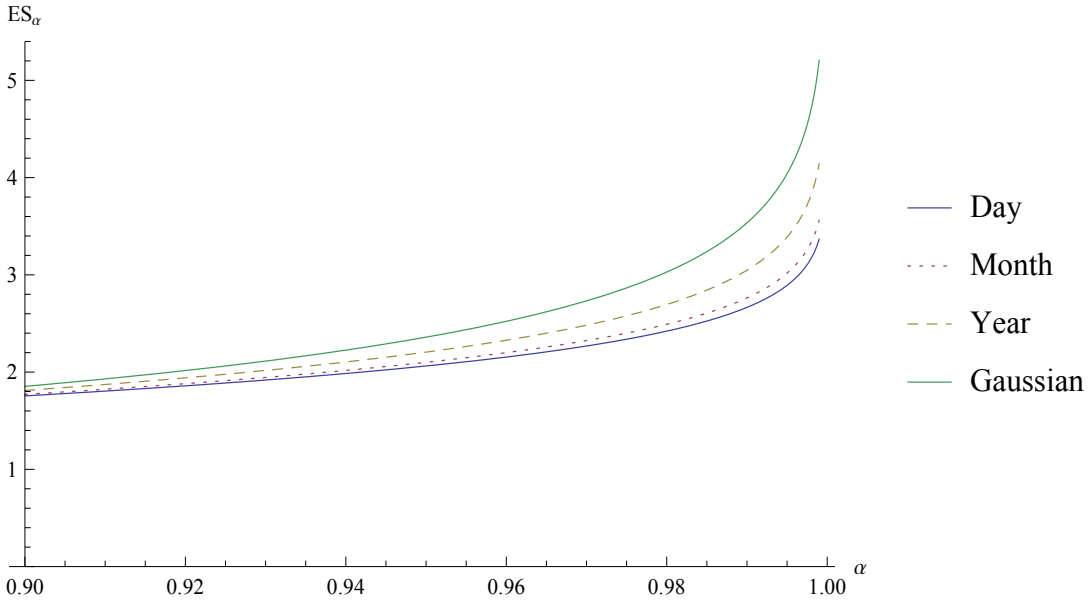
$$P^x\{X_t \leq \xi\} = P_t u_0(x) = 1 - \alpha$$

where u_0 is given by (6.36). We know the resolvent from (6.37), so it remains to invert the Laplace transform. We illustrate the values obtained in figure 8.4. It shows the VaR plotted as a function of the confidence level, for $t = 1/252$, $t = 1/12$ and $t = 1$, and also shows the corresponding Gaussian model for comparison as a benchmark. Some of the values we obtained are also presented in the third, fourth and fifth columns in table A.2. Finally they are compared also, in the second column, with the values obtained from a standard Gaussian distribution. Note that the thinner tailed distributions give greater VaR for the lower confidence levels. Even compensating for the increased variances, the distributions are fatter-tailed for larger time horizons t and the invariant distribution has the fattest tail of all.

Expected Shortfall

Using the steps in section 6.7.2 it is also possible to calculate the Expected Shortfalls in the model. The standardised values illustrated in figure 8.5 and compared in table A.3. Observe again that the distribution is fatter tailed for longer time horizons.

Figure 8.5: ES for aBm-CIR model and corresponding Gaussian.



Computational considerations

The numbers we obtained came from numerical integration along the Talbot contour (2.26) with respect to the complex-valued parameter λ, which makes them necessarily computationally expensive. However, there is one easy way of making the calculations run more quickly. Recall that both VaR and ES satisfy the property known as *positive homogeneity* in the theory on risk measures, which means that they scale linearly with the standard deviation (volatility). Given the specific value of the parameters μ_2 and c, the standardised VaR or ES is then fully determined, and the actual value to be used is found by multiplying by the volatility which fits the situation best. VaR calculations run more quickly for lower volatilities, while ES is more easily computed for higher volatilities (because the corresponding VaR is then larger). Particularly for the ES in the $t = 1/252$, I found it much quicker to scale the volatility down for the purpose of calculating the expected shortfall. Take care to note that the calculation of ES relies on knowing the VaR, so if ES is calculated for a different volatility, the VaR must be scaled accordingly first.

Approximations

Tables A.2 and A.3 allow us to measure the error of using approximations, such as the Gaussian for the one-day VaR, or the invariant distribution for the year VaR. The errors are quite large as tabulated in table A.4. Negative

numbers indicate that the approximating distribution underestimates the risk, while it is overestimated if the error is positive.

8.2 Arithmetic-geometric model

The numerical Laplace transform inversion, needed for the calculations in the aBm-CIR model, is computationally demanding, so in the aBm-gBm model, we shall restrict our attention to the cases in which we have been able to perform the inversion symbolically. These are the cases $\nu \in \{-4, -2, 0, 2\}$. Given the choice of model (i.e. ν), we shall then make a choice of parameters. We continue to use the S & P 500 daily returns, between 2001 and 2012 inclusive, to fit our model and the parameter choices are given in table 8.1. Unfortunately, we are not yet quite sure how to perform the full estimation. We have estimated $\mu_2 = 0.865103$, using the linear estimating function, as in the aBm-CIR model, but for $\nu \leq 1$ we need to take μ_2 negative, so we simply take $\mu_2 = -0.865103$ for the sake of these illustrations. Given our choice of ν, σ_2 may then be calculated. σ_1 simply determines the standard deviation, so we can fit it to any daily volatility, which is calculated from (5.9).

Note that a stationary distribution exists only in the case $\nu = 2$, and it does not have finite variance.

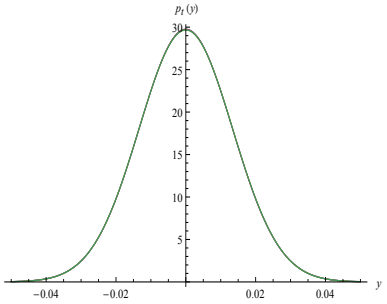
Direct computation of the dynamic pdf

We have observed that the time-dependent behaviour of the one-dimensional distributions of the aBm-CIR model is to start off as a spike at zero and gradually spread its mass along the line, the tails becoming fatter as t increases. This is also the case for the aBm-gBm models. The three plots in figure 8.6 show the probability mass dissipating towards $\pm\infty$ as t increases, and also that this process is quickest in the case $\nu = -4$ and slowest for $\nu = 2$. Take care to observe that the scales on both axes change between the plots.

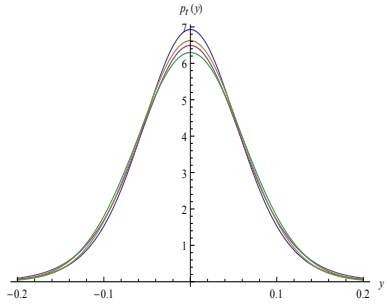
Table 8.1: Parameter choices for hybrid aBm-gBm models.

ν	μ_2	σ_1	σ_2
2	0.8651	0.2136	1.3154
0	-0.8651	0.2129	1.3154
-2	-0.8651	0.2136	0.7594
-4	-0.8651	0.2132	0.5883

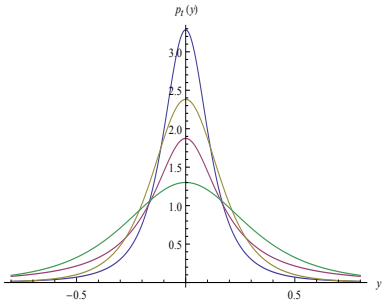
Figure 8.6: Dynamic densities for aBm-gBm models $\nu \in \{2, 0, -2, -4\}$.



(a) $t = 1/252$.



(b) $t = 1/12$.

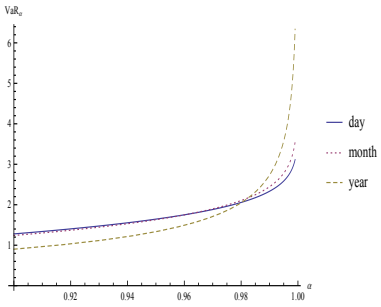


(c) $t = 1$.

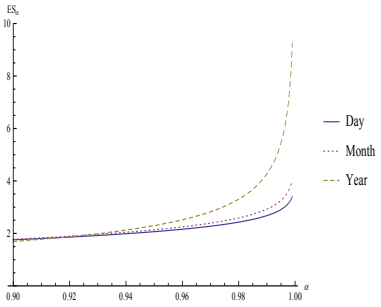
VaR and ES

VaR and ES are found by direct integration along the real line with respect to the y-variable, and are consequently very quick calculations. We present the results of these integrations in tables A.5 to A.8, and show the VaR and ES of the $\nu = 2$ case in figure 8.7. We have therefore shown how these quantities are computable, and that the method of computation that we have suggested give reasonable results. Again, the standardised distributions have fatter tails for the longer time horizons.

Figure 8.7: VaR and ES in the aBm-gBm model $\nu = 2$.



(a) VaR.



(b) ES.

Unfortunately at this stage, a comparison between the different models (i.e. differing ν) is not particularly meaningful, because our choice of parameters has been somewhat arbitrary, even though we have made a vague attempt to fit the model to the S& P data between 2001 and 2012. The reader will notice that the risk measures are greatest in the the $\nu = 0$ model, since the parameters σ_1 and σ_2 we have used to force $\nu = 0$ are greatest.

Finally note that when the model is fitted, the variance is also known, via (5.9).

Chapter 9

Conclusions and further work

In this final chapter, we review the progress we have made with this investigation, then summarise the main lines of research that it leaves open. In order to place the discussion of our progress into its context, let us recall our original aims with this investigation.

- We aimed to specify elementary modelling assumptions that might be satisfied by the returns process of a single financial asset. These assumptions would lead under certain conditions to the standard Gaussian models ubiquitous in the world of financial modelling, but could also be used for us to arrive at more exotic models which are hopefully able to explain some of the probabilistic properties observed in actual financial asset returns. In particular, we hope that some assumptions in our elementary model could lead to fatter-tailed financial models.
- We intended to study the resulting processes probabilistically, especially to be able to give a comparison of the tails of their one-dimensional distribution in comparison with the Gaussian distributions. We hoped to find that the distributions have thicker tails than the Gaussian distribution, in keeping with the tails observed for financial asset returns.
- We finally hoped to be able to demonstrate some usability of the new models.

We refer to these components respectively as *motivation* of our new models, their *implications* and *implementation*.

Motivation

The motivation section is itself divided into two separate sections. The first of these is the microstructure model, which aim to capture aspects of the

price formation process and convert it into mathematics. The microstructure model that we used ended up being a semi-Markov process. This was built forwards in time: given the present state of the market, two random quantities were of interest, namely the waiting time until the market state next changes and the actual quantitative change in the market state at that point in time. An important aspect of our models was that the distributions of these two quantities were able, for some portion of the traders in the market, to change conditionally upon the present market state.

Our development of this model has been quite successful. We have developed it in a very general form and shown that, with our other later assumptions, it leads to the hybrid aBm-CIR and aBm-gBm models without specification of these two (conditional) distributions.

The second part of the motivation section is the process approximation. We began by using Donsker's Theorem to show that the spatially homogeneous renewal-reward processes can be made to converge (by increasing the market liquidity) to the aBm model. By making the distributions of the relevant random quantities depend on the current market state, we moved from renewal-reward processes to semi-Markov processes. We then showed that, in the special cases when this price formation process is actually Markovian, we can completely approximate by a continuous-time Markov process, which may or may not include jumps. When it came to the more realistic semi-Markov model, however, we were not quite able to complete the argument by checking the conditions for application of convergence theorems in Jacod & Shiryaev [68].

Implications

We have used Linetsky's spectral classification to obtain aspects of the one-dimensional distributions of these processes via spectral methods. Another interpretation of our approach is that we have solved the Laplace transform of Kolmogorov's forward equation, which has yielded expressions for the associated probabilities. We have then attempted to find the actual probabilities associated with the process by Laplace transform inversion.

A full explicit inversion has been possible only in the aBm-gBm hybrid, in the cases $\nu \in \{-4, -2, 0, 2\}$. In fact these were already known in Shaw & Schofield [102], although the expression in the case $\nu = -4$ was not given there. In the other cases, the inversion in the aBm-gBm model presents challenges.

We have been able to find general expressions for the Laplace transform

of the transition probabilities in the hybrid aBm-CIR process. Furthermore, in cases in which the speed measure is finite in neighbourhoods of the left endpoint $-\infty$ of the state space, it has been possible to find expressions for the Laplace transform of the VaR and ES. The reason that this is important is that it looks like the Laplace transform inversion integral has to be performed numerically. Risk measures are then obtained by integrating over a domain in the state space, but when the speed measure is finite on that domain, an interchange of the order of performing these two integrations is valid, and the one in the state space can be done symbolically. This leaves the risk measure as simply a numerical integration with respect to the spectral parameter λ.

We have been unable to find usable expressions for these risk measures in the remaining cases, and this is the most important area in which further work is required in this section. In the aBm-gBm models, we would suggest that it would be useful at first to find the expressions for $\nu \in \{-4, -3, \ldots, 3, 4\}$. There is also scope for investigation into other hybrid models that might arise when the model of the behaviour of the technical traders in the motivation section is altered. All these Markov processes are associated with a linear operator, namely the generator. Diffusion models are the easiest of all, because the generator is local, so solutions to the Sturm-Liouville equation at least are likely to be known. But jump diffusions are also important for applications, and an understanding of the distributions can be built from the corresponding diffusion without jumps, and the probabilistic behaviour of the jumps if it is known.

Implementation

We have attempted to investigate a method of estimating the parameters of the models from historical data. We used a linear estimating function to estimate the (technical) drift parameter, which appears to have been successful. We have also performed a static maximum likelihood investigation on daily return data, to fit it to the invariant distributions in the various models, then used these to infer the modelling parameters. This method is itself only really valid to provide a first guess at the model parameters which best fit the data, and we found that the model then seriously underestimated the volatility. The parameters were then adjusted to fit the model to the correct daily volatility.

This section of the investigation is in need of attention, to produce a method of fitting the models to data with a much greater degree of accuracy. We then envisage that it should be possible to adjust the parameters on, say, a

daily basis, according to the particular market conditions. The latter topic of calibration has not been explored at all in this study.

Given the parameters we obtained, which may be described as nothing more than an educated guess, we were then able to utilise the expressions found in the implications section to plot graphs of the dynamic pdfs. We then obtained standardised VaR and ES at various confidence levels, demonstrating the ease of use of these expressions. It turns out that the numerical integrations in the λ domain are somewhat computationally intensive, but this can be adjusted by moving the volatility of the model, downwards for VaR, or upwards for ES, and then rescaling the number so obtained.

Further work

The main remaining research to be completed in the motivation section the approximation of the hybrid diffusion process by semi-Markov jump processes which are not fully Markovian. In this case, the route to a convergence theorem is by characterising the semi-Markov processes via their characteristics as semimartingales, and checking the conditions of an appropriate convergence theorem. If these characteristics converge in an appropriate manner to the characteristics of our candidate for the limiting process then a straightforward application of the appropriate theorem will complete the argument. We think that it will be possible to show convergence to a diffusion or a jump diffusion process, and the form of semimartingale characteristics will inform us under which conditions jumps arise. As an extension of the present investigation, one might also attempt to model the behaviour of the technical traders using market statistic processes, leading not to a one-dimensional Markov process, but to a Markov-reducible one, or even model more explicitly the information flow available to the market participants, as in the BHM framework developed by Brody et al. [24].

In investigating the implications of these models, the most important further work to be done is the derivation of expressions for the risk measures in those cases where we have not yet found them. These are the cases $\nu \notin \{-4, -2, 0, 2\}$ in the aBm-gBm model, and cases when the speed measure is infinite in the aBm-CIR model. There is then further investigation to be conducted into extensions of our models which contain jumps. Recall from section 4.5, in particular equation (4.15), that the distributions of the trade orders can be decomposed into small and large trades, which correspond to diffusion and jumps after taking the limit in the approximation. This means that, by only

adjusting our approach to the large trades, we can study processes which have the same diffusion behaviour but different jumps. Such an investigation would necessarily start from the knowledge we have developed of the stochastic behaviour between the jumps (if there are only fintely many jumps per bounded time-interval).

In cases where the practitioner is aware of a reliable method of inferring the optimal parameters to use, the models are ready to use in the cases that we have illustrated. In terms of applying these models in practice then, the most important lines for further study are (1) the development of a standard method of parameter estimation, followed by methods of calibration for daily use, and (2) the implementation of these models in the remaining cases. For the aBm-gBm model, it would perhaps suffice to have an implementation for integers $\nu \in \{-4, -3, \ldots, 3, 4\}$, but at present, we have to hope that one of the values -4, -2, 0 or 2 is the best fit. It is quite realistic to expect that one can find it for $\nu = 4$ (hence $\nu = -6$ via the Legendre symmetry) but the odd integers might be more difficult. For the aBm-CIR model, it is most important to be able to obtain methods of calculating VaR and ES when the speed measure is infinite in neighbourhoods of $-\infty$. Together with a method of parameter estimation, these models will then be fully ready to use.

Appendix A

VaR and ES for the various models

Here we present the results of the numerical integrations which have been performed to calculate the value-at-risk and expected shortfall in the various models. This shows that the calculations are possible, and studying the results shows that they give reasonable results. First, we note that working in the models with the parameters that we have used, we know or can calculate the variances implied by the models. The models have been fitted to the daily volatility in the data, which means that the standard deviation is the same over a day, regardless of the model. The choice of model then determines volatility development. In the aBm-gBm model, these can be obtained by the formula (5.9). These values have been cross-checked with those arising from direct numerical integration for obtaining the variances.

To obtain the VaR or ES for using in the given model, one simply multiplies the standard deviation given in table A.1 by the standardised VaR or ES given in the tables below.

Table A.1: Volatility development in all models

Model		s.d.(X_t)		
		$t = 1/252$	$t = 1/12$	$t = 1$
aBm-CIR	$c = 0.4959\ldots$	0.0135	0.0660	0.2058
aBm-gBm	$\nu = 2$	0.0135	0.0617	0.2136
	$\nu = 0$	0.0135	0.0662	0.6355
	$\nu = -2$	0.0135	0.0646	0.4220
	$\nu = -4$	0.0135	0.0643	0.3907

Table A.2: Standardised VaR for aBm-CIR models.

$\alpha =$ Confidence level	$\mathrm{VaR}_\alpha(X_t)/\mathrm{s.d.}(X_t)$				
	Gaussian $(t \downarrow 0)$	$t = 1/252$	$t = 1/12$	$t = 1$	$t = \infty$
90%	1.2816	1.2684	1.2237	1.1238	1.0871
95%	1.6449	1.6487	1.6530	1.6279	1.6074
99%	2.3263	2.3840	2.5481	2.8052	2.8711
99.5%	2.5758	2.6600	2.9044	3.3117	3.4308
99.9%	3.0902	3.2407	3.6868	4.4838	4.7531

Table A.3: Standardised ES for aBm-CIR models.

$\alpha =$ Confidence level	$\mathrm{ES}_\alpha(X_t)/\mathrm{s.d.}(X_t)$				
	Gaussian $(t \downarrow 0)$	$t = 1/252$	$t = 1/12$	$t = 1$	$t = \infty$
90%	1.7550	1.7708	1.8109	1.8527	1.8552
95%	2.0627	2.1004	2.2052	2.3587	2.3945
99%	2.6652	2.7627	3.0475	3.5341	3.6859
99.5%	2.8919	3.0176	3.3876	4.0393	4.2535
99.9%	3.3671	3.5616	4.1422	5.2069	5.5892

Table A.4: Error of the approximating distribution in the aBm-CIR model for risk measures.

Confidence level	Error of Gaussian distribution for one-day VaR	Error of invariant distribution for one-year VaR	Error of Gaussian distribution for one-day ES	Error of invariant distribution for one-year ES
90%	1.04%	-11.16%	-3.09%	0.13%
95%	-0.24%	-2.76%	-6.46%	1.52%
99%	-2.42%	12.68%	-12.54%	4.30%
99.5%	-3.17%	18.12%	-14.63%	5.30%
99.9%	-4.64%	28.92%	-18.71%	7.34%

Table A.5: Standardised VaR and ES for aBm-gBm model, $\nu = 2$.

α	$\mathrm{VaR}_\alpha(X_t)/\mathrm{s.d.}(X_t)$			$\mathrm{ES}_\alpha(X_t)/\mathrm{s.d.}(X_t)$		
	$t = 1/252$	$t = 1/12$	$t = 1$	$t = 1/252$	$t = 1/12$	$t = 1$
90%	1.2796	1.2407	0.9015	1.7563	1.7764	1.6911
95%	1.6443	1.6301	1.3346	2.0667	2.1367	2.2950
99%	2.3327	2.4452	2.7324	2.6787	2.9257	4.2693
99.5%	2.5865	2.7793	3.5748	2.9105	3.2581	5.4477
99.9%	3.1133	3.5466	6.3353	3.4002	4.0313	9.2423

Table A.6: Standardised VaR and ES for aBm-gBm model, $\nu = 0$.

α	$\mathrm{VaR}_\alpha(X_t)/\mathrm{s.d.}(X_t)$			$\mathrm{ES}_\alpha(X_t)/\mathrm{s.d.}(X_t)$		
	$t = 1/252$	$t = 1/12$	$t = 1$	$t = 1/252$	$t = 1/12$	$t = 1$
90%	1.2796	1.2379	0.6636	1.7563	1.7777	1.5390
95%	1.6443	1.6291	1.0936	2.0667	2.1415	2.2345
99%	2.3327	2.4530	2.7101	2.6787	2.9425	4.7152
99.5%	2.5865	2.7927	3.7668	2.9106	3.2813	6.2737
99.9%	3.1134	3.5756	7.4154	3.3998	4.0748	11.4835

Table A.7: Standardised VaR and ES for aBm-gBm model, $\nu = -2$.

α	$\mathrm{VaR}_\alpha(X_t)/\mathrm{s.d.}(X_t)$			$\mathrm{ES}_\alpha(X_t)/\mathrm{s.d.}(X_t)$		
	$t = 1/252$	$t = 1/12$	$t = 1$	$t = 1/252$	$t = 1/12$	$t = 1$
90%	1.2809	1.2671	1.0516	1.7554	1.7637	1.7943
95%	1.6447	1.6404	1.5137	2.0640	2.0909	2.3359
99%	2.3285	2.3714	2.7848	2.6697	2.7610	3.8366
99.5%	2.5794	2.6516	3.4380	2.8981	3.0253	4.6050
99.9%	3.0979	3.2561	5.2496	3.3782	3.6042	6.7214

Table A.8: Standardised VaR and ES for aBm-gBm model, $\nu = -4$.

α	$\mathrm{VaR}_\alpha(X_t)/\mathrm{s.d.}(X_t)$			$\mathrm{ES}_\alpha(X_t)/\mathrm{s.d.}(X_t)$		
	$t = 1/252$	$t = 1/12$	$t = 1$	$t = 1/252$	$t = 1/12$	$t = 1$
90%	1.2812	1.2729	1.1413	1.7552	1.7604	1.8016
95%	1.6447	1.6423	1.5795	2.0635	2.0799	2.2686
99%	2.3276	2.3537	2.6646	2.6679	2.7231	3.4531
99.5%	2.5780	2.6218	3.1773	2.8958	2.9722	4.0172
99.9%	3.0949	3.1903	4.5002	3.3739	3.5097	5.4744

Bibliography

[1] Y. A. Abramovich and C. D. Aliprantis. *An Invitation to Operator Theory*. American Mathematical Society, 2002.

[2] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. Dover, 1965.

[3] L. Alili, D. Dufresne, and M. Yor. *Exponential Functional and Principal Value related to Brownian Motion. A collection of research papers*, chapter Sur l"identite de Bougerol pour les fonctionnelles exponentielles du mouvement brownien avec drift, pages 3–14. Biblioteca de la Revista matemtica iberoamericana, 1997.

[4] Álvaro Cartea and Thilo Meyer-Brandis. How duration between trades of underlying securities affects option prices. *Review of Finance*, 14:749–785, 2010. doi:10.1093/rof/rfp013.

[5] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Thinking coherently. *Risk*, 10:33–49, 1997.

[6] P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–28, 1999.

[7] F. Avram, A. E. Kyprianou, and M. R. Pistorius. Exit problems for spectrally negative Lévy processes and applications to (canadized) russian options. *Annals of Applied Probability*, 14(1):215–238, 2004.

[8] F. Avram, Z. Palmowski, and M. R. Pistorius. On the optimal dividend problem for a spectrally negative Lévy process. *Annals of Applied Probability*, 17(1):156–180, 2007.

[9] O. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 353(1674):pp. 401–419, 1977. URL: http://www.jstor.org/stable/79167.

[10] O. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2:41–68, 1998.

[11] O. E. Barndorff-Nielsen and S. Z. Levendorskiĭ. Feller processes of normal inverse Gaussian type. *Quantitative Finance*, 1(3):318–331, 2001. URL: `http://www.informaworld.com/10.1088/1469-7688/1/3/303`.

[12] A. Berkelaarand and R. Kouwenberg. Dynamic asset allocation and downside-risk aversion. 2000.

[13] P. Billingsley. *Convergence of Probability Measures*. Wiley. First edition, 1968; second edition, 1999.

[14] P. Billingsley. *Probability and Measure*. Wiley, 2012. Anniversary Edition.

[15] T. Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, third edition, 2009.

[16] F. Black, E. Derman, and W. Toy. A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal*, 46(1):33–39, 1990.

[17] F. Black and P. Karasinski. Bond and option pricing when short rates are lognormal. *Financial Analysts Journal*, 47(4):pp. 52–59. URL: `http://www.jstor.org/stable/4479456`.

[18] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81(3):637–654, 1973.

[19] R. M. Blumenthal and R. K. Getoor. *Markov Processes and Potential Theory*. Academic Press, 1968.

[20] A. Bobrowski. *Functional Analysis for Probability and Stochastic Processes*. Cambridge University Press, 2005.

[21] A. N. Borodin and P. Salminen. *Handbook of Brownian Motion - Facts and Formulae*. Birkhäuser Verlag, second edition, 2002.

[22] Bougerol. Exemples de théorèmes locaux sur les groupes résolubles. *Ann. Inst. H. Poincaré*, 19(4):369–391, 1983.

[23] S. I. Boyarchenko and S. Z. Levendorskiĭ. *Non-Gaussian Merton-Black-Scholes Theory*. World Scientific Publishing Co. Pte. Ltd., 2002.

[24] D. C. Brody, L. P. Hughston, and A. Macrina. Information-based asset pricing. *International Journal of Theoretical and Applied Finance*, 11(1):107–142, 2008. doi:10.1142/9789814407892_0005.

[25] J. Y. Campbell, A. W. Lo, and A. C. MacKinlay. *The Econometrics of Financial Markets*. Princeton University Press, 1997.

[26] P. Carr, H. Geman, D. B. Madan, and M. Yor. The fine structure of asset returns: An empirical investigation. *Journal of Business*, 75(2):305–332, 2002.

[27] P. Carr and L. Wu. The finite moment log stable process and option pricing. *The Journal of Finance*, LVIII(2):753–778, 2003.

[28] E. Çinlar. *Introduction to Stochastic Processes*. Prentice-Hall, 1975.

[29] E. Çinlar. Markov renewal theory: A survey. *Management Science*, 21(7):727–752, 1975.

[30] E. Çinlar. *Probability and Stochastics*. Number 261 in Graduate Texts in Mathematics. Springer, 2011.

[31] A. M. Cohen. *Numerical Methods for Laplace Transform Inversion*. Springer, 2007.

[32] P. H. Cootner. *The random character of stock market prices*. M.I.T. Press, 1964.

[33] J. C. Cox, J. E. Ingersoll, and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53:385–407, 1985.

[34] J. Daníelsson. *Financial risk forecasting*. Wiley, 2011.

[35] B. Davies. *Integral Transforms and Their Applications*. Springer Verlag, 2002.

[36] M. D. Donsker. An invariance pinciple for certain probability limit theorems. *Memoirs of the American Mathematical Society*, 6, 1951.

[37] N. Dunford and J. T. Schwartz. *Linear Operators Part 1: General Theory*. Interscience, 1958.

[38] E. B. Dynkin. *Markov Processes*. Academic Press, 1965. Volumes 1 & 2.

[39] E. Eberlein and U. Keller. Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299, 1995.

[40] E. Eberlein and K. Prause. The generalized hyperbolic model: Financial derivatives and risk measures. In *Mathematical Finance Bachelier Congress 2000, Geman*, pages 245–267. Springer, 1998.

[41] P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling extremal events for insurance and finance*. Springer-Verlag, 1997.

[42] K.-J. Engel and R. Nagel. *One-Parameter Semigroups for Linear Evolution Equations*. Springer, 2000.

[43] A. Erdélyi, W. Magnus, F. Oberhettinger, and F. G. Tricomi. *Higher Transcendental Functions*, volume I. McGraw-Hill, 1953. Bateman Manuscript Project.

[44] B. Eriksson and M Pistorius. A method of moments approach to pricing double barrier contracts driven by a general class of jump diffusions. *International Journal of Theoretical and Applied Finance*, 14(7):1139–1158, 2011.

[45] A. Etheridge. *A Course in Financial Calculus*. Cambridge University Press, 2002.

[46] S. N. Ethier and T. G. Kurtz. *Markov Processes: Characterization and Convergence*. Wiley, 1986.

[47] E. F. Fama. Mandelbrot and the stable Paretian hypothesis. *Journal of Business*, 36(4):420–429, 1963.

[48] W. Feller. Two singular diffusion problems. *Annals of Mathematics*, 54(1):pp. 173–182, 1951. URL: `http://www.jstor.org/stable/1969318`.

[49] W. Feller. The parabolic differentail equations and the associated semigroups of transformations. *Annals of Mathematics*, 55:468–519, 1952.

[50] W. Feller. Diffusion processes in one dimension. *Transactions of the American Mathematical Society*, 77:1–31, 1954.

[51] W. Feller. On boundaries and lateral conditions for the kolmogorov differential equations. *Annals of Mathematics*, 65(3):527–570, 1957.

[52] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume I. Wiley, third edition, 1968. Revised Printing.

[53] W. Feller. *An Introduction to Probability Theory and Its Applications*, volume II. Wiley, second edition, 1971.

[54] C. T. Fulton, S. Pruess, and Y. Xie. The automatic classification of sturm-liouville problems. *J. Appld. Math. and Comp.*

[55] I. I. Gihman and A. V. Skorohod. *The Theory of Stochastic Processes*, volume II. Springer, 2004. Reprint of the 1974 edition.

[56] C. Grosche and F. Steiner. *Handbook of Feynman Path Integrals*. Springer, 1998.

[57] Jr. H. P. McKean. Elementary solutions for certain parabolic partial differential equations. *Transactions of the American Mathematical Society*, 82(2):519–548, 1956.

[58] J. Heinrich. A guide to the pearson IV distribution, 2004. URL: `http://www-cdf.fnal.gov/physics/statistics/notes/cdf6820_pearson4.pdf`.

[59] E. Hille and R. S. Phillips. *Functional Analysis and Semi-Groups*. American Mathematical Society, 1957.

[60] T. S. Y. Ho and S.-B. Lee. Term structure movements and pricing interest rate contingent claims. *The Journal of Finance*, 41:1011–1029, 1986.

[61] L. P. Hughston and A. Macrina. Pricing fixed-income securities in an information-based framework. *Applied Mathematical Finance*, 19(4):361–379, 2012. `doi:10.1080/1350486X.2011.631757`.

[62] J. Hull and A. White. Pricing interest-rate-derivative securities. *The Review of Financial Studies*, 3(4):573–592, 1990.

[63] J. C. Hull. *Options, futures and other derivatives*. Prentice-Hall, 7th edition, 2008.

[64] S. M. Iacus. *Simulation and Inference for Stochastic Differential Equations*. Springer, 2008.

[65] D. L. Iglehart. Diffusion approximations in collective risk theory. *Journal of Applied Probability*, 6:285–292, 1969.

[66] K. Itô and Jr. H. P. McKean. *Diffusion Processes and their Sample Paths*. Springer, 1965.

[67] N. Jacob. *Pseudo-Differential Operators and Markov Processes: Fourier Analysis and Semigroups*, volume I: Fourier Analysis and Semigroups. Imperial College Press, 2001.

[68] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, second edition, 2003.

[69] H. Jeffreys. *Theory of Probability*. Oxford University Press, third edition, 1961.

[70] Z.-Q. Jiang, W. Chen, and W.-X. Zhou. Scaling in the distribution of intertrade durations of chinese stocks. *Physica A*, 387:58185825, 2008.

[71] E. Jondeau, S.-H. Poon, and M. Rockinger. *Financial modeling under non-Gaussian distributions*. Springer-Verlag, 2007.

[72] O. Kallenberg. *Foundations of Modern Probability*. Springer, second edition, 2002.

[73] S. Karlin and H. M. Taylor. *A Second Course in Stochastic Processes*. Academic Press, 1981.

[74] T. Kato. *Perturbation Theory for Linear Operators*. Springer, second edition, 1980. Corrected Printing.

[75] A. Kempf and O. Korn. Market depth and order size. *Journal of Financial Markets*, 2:29–48, 1999.

[76] C. D. Kirkpatrick and J. R. Dalquist. *Technical analysis: the complete resource for financial market technicians*. FT Press Financial Times, 2007.

[77] F. Klebaner. *Introduction to Stochastic Calculus with Applications*. Imperial College Press, 2005.

[78] S. A. Klugman, H. H. Panjer, and G. E. Willmot. *Loss Models: From Data to Decisions*. Wiley, third edition, 2008.

[79] U. Küchler and S. Tappe. Bilateral gamma distributions and processes in financial mathematics. *Stochastic Processes and their Applications*, 118:261–283, 2008.

[80] W. R. LePage. *Complex Variables and the Laplace Transform for Engineers*. Dover, 1980.

[81] F. Lillo, J. Doyne Farmer, and R. N. Mantegna. Single Curve Collapse of the Price Impact Function for the New York Stock Exchange. *eprint arXiv:cond-mat/0207428*, July 2002. `arXiv:arXiv:cond-mat/0207428`.

[82] V. Linetsky. The spectral decomposition of the option value. *International Journal of Theoretical and Applied Finance*, 7(3):337–384, 2004.

[83] V. Linetsky. *Handbooks in Operations Research and Management Science, Volume 15: Financial Engineering*, chapter 6: Spectral Methods in Options Pricing. Elsevier B. V., 2008.

[84] D. B. Madan and E. Seneta. The variance gamma (v.g.) model for share market returns. *The Journal of Business*, 63(4):511–524, 1990.

[85] F. Mainardi, M. Raberto, R. Gorenflo, and E. Scalas. Fractional calculus and continuous time finance ii: the waiting-time distribution. *Physica A-Statistical Mechanics and its Applications*, 287:53–64, 2000.

[86] B. B. Mandelbrot. The variation of certain speculative prices. *Journal of Business*, 36:394–419, 1963.

[87] M. Merkle. Topics in weak convergence of probability measures. *Zb. radova Mat. Inst. Beograd*, 9(17):235–274, 2000.

[88] R. C. Merton. Theory of fational option pricing. *The Bell Journal of Economics and Management Science*, 4:141–183, 1973.

[89] G.-H. Mu, W. Chen, J. Kertész, and W.-X. Zhou. Preferred numbers and the distribution of trade sizes and trading volumes in the chinese stock market. *The European Physical Journal B*, 68, Issue 1:145–152, 2009.

[90] J. Murphy. *Technical Analysis of the Futures Markets*. New York Institute of Finance, 1986.

[91] M. Politi and E. Scalas. Fitting the empirical distribution of intertrade durations. *Physica A-Statistical Mechanics and its Applications*, 387:2025–2034, 2009.

[92] S. J. Press. A compound events model for security prices. *The Journal of Business*, 40(3):317–335, 1967.

[93] Y. V. Prohorov and Y. A. Rozanov. *Probability theory: basic concepts, limit theorems, random processes*. Springer, 1969.

[94] S. M. D. Queirós. On the emergence of a generalised gamma distribution. application to traded volume in financial markets. *Europhysics Letters*, 71(3):339–345, 2005.

[95] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2012. ISBN 3-900051-07-0. URL: `http://www.R-project.org/`.

[96] S. I. Resnick. *Adventures in Stochastic Processes*. Birkhäuser Boston, 1992.

[97] L. C. G. Rogers and D. Williams. *Diffusions, Markov Processes and Martingales*, volume 1: Foundations; 2: Itô Calculus. Cambridge University Press, 2nd edition, 2000.

[98] F. J. Rohlf and R. R. Sokal. *Statistical Tables*. W. H. Freeman and Company, second edition, 1981.

[99] K. Sato. *Lévy processes and infinitely divisible distributions*. Cambridge University Press, 1999.

[100] N. Sazuka, J. Inoue, and E. Scalas. The distribution of first-passage times and durations in forex and future markets. *Physica A-Statistical Mechanics and its Applications*, 388(14):2839, 2009.

[101] E. Scalas, R. Gorenflo, H. Luckock, F. Mainardi, M. Mantelli, and M. Raberto. On the intertrade waiting-time distribution. *Finance Letters*, 3:38–43, 2005.

[102] W. T. Shaw and M. Schofield. A model of returns for the post-credit-crunch reality: hybrid brownian motion with price feedback. *Quantitative Finance*, pages 1–24. `doi:10.1080/14697688.2011.642810`.

[103] S. E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. Springer, 2010.

[104] L. Slater. *Confluent Hypergeometric Functions*. Cambridge University Press, 1960.

[105] H. Sørensen. Parametric inference for diffusion processes observed at discrete points in time: a survey. *Int. Stat. Rev.*, 72(3):337–354, 2004.

[106] D. W. Stroock. *Probability Theory, An analytic view*. Cambridge University Press, 1993.

[107] H. M. Taylor and S. Karlin. *An Introduction to Stochastic Modeling*. Academic Press, third edition, 1998.

[108] I. V. Vagurina. Diffusion processes that correspond to the hypergeometric equation. *Journal of Mathematical Sciences*, 133(3):1249–1256, 2006.

[109] S. Vakeroudis. Bougerol's identity in law and extensions. *Probability Surveys*, 9:411–437, 2012.

[110] O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177 – 188, 1977. URL: `http://www.sciencedirect.com/science/article/pii/0304405X77900162`, `doi:10.1016/0304-405X(77)90016-2`.

[111] P. Wilmott, S. Howison, and J. Dewynne. *The Mathematics of Financial Derivatives: A Student Introduction*. Cambridge University Press, 1995.

[112] Wolfram. Wolfram functions site. URL: `http://functions.wolfram.com/`.

[113] K. Yosida. *Functional Analysis*. Springer, sixth edition, 1980.

[114] Y.-C. Zhang. Toward a theory of marginally efficient markets. *Physica A*, 269(1):30–44, 1999.