American Monte Carlo Option Pricing under Pure Jump Lévy Models

Lydia West

Thesis presented in partial fulfilment of the requirements for the degree of Master of Science in the Faculty of Science at Stellenbosch University

Supervisor: Dr. P. W. Ouwehand

November 2012

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

November 2012

Abstract

We study Monte Carlo methods for pricing American options where the stock price dynamics follow exponential pure jump Lévy models. Only stock price dynamics for a single underlying are considered.

The thesis begins with a general introduction to American Monte Carlo methods. We then consider two classes of these methods. The first class involves regression — we briefly consider the regression method of Tsitsiklis and Van Roy [2001] and analyse in detail the least squares Monte Carlo method of Longstaff and Schwartz [2001]. The variance reduction techniques of Rasmussen [2005] applicable to the least squares Monte Carlo method, are also considered. The stochastic mesh method of Broadie and Glasserman [2004] falls into the second class we study. Furthermore, we consider the dual method, independently studied by Andersen and Broadie [2004], Rogers [2002] and Haugh and Kogan [March 2004] which generates a high bias estimate from a stopping rule. The rules we consider are estimates of the boundary between the continuation and exercise regions of the option. We analyse in detail how to obtain such an estimate in the least squares Monte Carlo and stochastic mesh methods.

These models are implemented using both a pseudo-random number generator, and the preferred choice of a quasi-random number generator with bridge sampling. As a base case, these methods are implemented where the stock price process follows geometric Brownian motion.

However the focus of the thesis is to implement the Monte Carlo methods for two pure jump Lévy models, namely the variance gamma and the normal inverse Gaussian models. We first provide a broad discussion on some of the properties of Lévy processes, followed by a study of the variance gamma model of Madan et al. [1998] and the normal inverse Gaussian model of Barndorff-Nielsen [1995]. We also provide an implementation of a variation of the calibration procedure of Cont and Tankov [2004b] for these models. We conclude with an analysis of results obtained from pricing American options using these models.

Uittreksel

Ons bestudeer Monte Carlo metodes wat Amerikaanse opsies, waar die aandeleprys dinamika die patroon van die eksponensiële suiwer sprong Lévy modelle volg, prys. Ons neem slegs aandeleprys dinamika vir 'n enkele aandeel in ag.

Die tesis begin met 'n algemene inleiding tot Amerikaanse Monte Carlo metodes. Daarna bestudeer ons twee klasse metodes. Die eerste behels regressie — ons bestudeer die regressiemetode van Tsitsiklis and Van Roy [2001] vlugtig en analiseer die least squares Monte Carlo metode van Longstaff and Schwartz [2001] in detail. Ons gee ook aandag aan die variansie reduksie tegnieke van Rasmussen [2005] wat van toepassing is op die least squares Monte Carlo metodes. Die stochastic mesh metode van Broadie and Glasserman [2004] val in die tweede klas wat ons onder oë neem. Ons sal ook aandag gee aan die dual metode, wat 'n hoë bias skatting van 'n stop reël skep, en afsonderlik deur Andersen and Broadie [2004], Rogers [2002] and Haugh and Kogan [March 2004] bestudeer is. Die reëls wat ons bestudeer is skattings van die grense tussen die voortsettings- en oefenareas van die opsie. Ons analiseer in detail hoe om so 'n benadering in die least squares Monte Carlo en stochastic mesh metodes te verkry.

Hierdie modelle word geïmplementeer deur beide die pseudo kansgetalgenerator en die verkose beste quasi kansgetalgenerator met brug steekproefneming te gebruik. As 'n basisgeval word hierdie metodes geïmplimenteer wanneer die aandeleprysproses 'n geometriese Browniese beweging volg.

Die fokus van die tesis is om die Monte Carlo metodes vir twee suiwer sprong Lévy modelle, naamlik die variance gamma en die normal inverse Gaussian modelle, te implimenteer. Eers bespreek ons in breë trekke sommige van die eienskappe van Lévy prossesse en vervolgens bestudeer ons die variance gamma model soos in Madan et al. [1998] en die normal inverse Gaussian model soos in Barndorff-Nielsen [1995]. Ons gee ook 'n implimentering van 'n variasie van die kalibreringsprosedure deur Cont and Tankov [2004b] vir hierdie modelle. Ons sluit af met die resultate wat verkry is, deur Amerikaanse opsies met behulp van hierdie modelle te prys.

Acknowledgements

I thank my supervisor Dr. Peter Ouwehand for his guidance throughout the thesis. His suggestions, detailed explanations and thorough review are sincerely appreciated.

A special thanks to Dr. Thomas McWalter for assisting me in setting up the Stellenbosch University LaTeX document, in creating high-quality graphs, as well as his observations and suggestions, in particular concerning part III.

Finally, I thank my dear husband and best friend, Dr. Graeme West, for his frequent comments, his patience and encouragement. Without his support, I would not have been able to start the thesis; and without his love I would not have been able to finish.

In loving memory of my husband Graeme.

Contents

List of Notation

Random Variables	
$(\Omega, \mathcal{F}, \mathbb{P})$	Probability space where Ω is the set of all possible realisations, $\mathcal{F}$ is a σ-algebra and $\mathbb{P}$ is a probability measure on $\mathcal{F}$.
$\{\mathcal{F}_t\}_{t\geq 0}$	Filtration on $(\Omega, \mathcal{F}, \mathbb{P})$.
$\mathbb{Q}$	Risk-neutral probability measure equivalent to $\mathbb{P}$.
$\mathcal{B}(\mathbb{R})$	Borel algebra, i.e. σ-algebra generated by all open subsets of $\mathbb{R}$.
$F_X(\cdot)$	Cumulative distribution function of random variable X.
$f_X(\cdot)$	Probability density function of random variable X.
$\Phi_X(\cdot)$	Characteristic function of random variable X.
$M_X(\cdot)$	Moment generating function of random variable X.
$m_k(X)$	k^{th} moment of random variable X.
$\bar{m}_k(X)$	k^{th} central moment of random variable X.
$\Psi_X(\cdot)$	Cumulant generating function of random variable X.
$c_k(X)$	k^{th} cumulant of random variable X.
$\mathbb{E}[X]$	Expectation of random variable X.
$\mathbb{V}\text{ar}[X]$	Variance of random variable X, also denoted by Σ.
$\mathbb{C}\text{ov}[X, Y]$	Covariance between random variables X and Y.
$\mathbb{C}\text{orr}[X, Y]$	Correlation between random variables X and Y.
$s(X)$	Skewness coefficient of random variable X.
$\varkappa(X)$	Kurtosis of random variable X.
$\bar{\varkappa}(X)$	Excess kurtosis of random variable X.
$X \overset{\mathcal{D}}{=} Y$	X is equal to Y in distribution.
i.i.d.	Independent and identically distributed.
a.s.	Almost surely.

General American Monte Carlo Pricing	
r	Continuously compounded risk-free rate.
q	Continuously compounded dividend yield.
K	Strike.
T	Maturity or expiration date.
S_t	Underlying asset price process at time $t \in [0, T]$.
I_t	Intrinsic value or payoff process at time $t \in [0, T]$.
V_t	American option value process at time $t \in [0, T]$.
Z_t	Continuation or holding value process at time $t \in [0, T]$.
$\bar{X}$	Approximation of X.
$\hat{X}$	Approximation of X that is biased high or used in the calculation of a high biased estimate.
$\check{X}$	Approximation of X that is biased low or used in the calculation of a low biased estimate.
$0 = t_0 < t_1 < \ldots < t_M = T$	Discretised exercise times.
Δt_j	Time periods $t_j - t_{j-1}$, $j = 1,\ 2,\ \ldots,\ M$ between discretised exercise times.
X_j^i	Value of discretised process X at time t_j, $j = 0,\ 1,\ \ldots,\ M$ in sample path i, $i = 1,\ 2,\ \ldots,\ N$.
$\mathcal{T}_{j,M}$	Set of all stopping times with values in $\{t_j,\ t_{j+1},\ \ldots,\ t_M\}$, $j = 0,\ 1,\ \ldots,\ M$.
TV	*Test value* — vector of length N, where the i^{th} entry TV^i, $i = 1,\ 2,\ \ldots,\ N$, is equal to the modelled continuation value at a specific time. This vector is overwritten at every time point t_j, $j = 0,\ 1,\ \ldots,\ M$.

Regression, LSM and LSM-Rasmussen Methods	
$\bar{\tau}_j^i$	Modelled stopping time of sample path i, $i = 1,\ 2,\ \ldots,\ N$, determined at time t_j, $j = 0,\ 1,\ \ldots,\ M$.
${}_\kappa\bar{Z}_j\left(S_j^i\right)$	Modelled continuation value time t_j, $j = 1,\ 2,\ \ldots,\ M-1$, given the stock price is S_j^i, $i = 1,\ 2,\ \ldots,\ N$.
E	*European value* — vector of length N, where the i^{th} entry E^i, $i = 1,\ 2,\ \ldots,\ N$, is equal to the value of a European option with term $t_M - t_j$ and spot S_j^i, $j = 0,\ 1,\ \ldots,\ M$. The strike and style are the same as the American option under consideration. This vector is overwritten at every time point t_j.
EE	*Eventual exercise* — vector of length N, where the i^{th} entry EE^i, $i = 1,\ 2,\ \ldots,\ N$, is equal to the payoff value at time $\bar{\tau}_j^i$ discounted to time t_j, $j = 0,\ 1,\ \ldots,\ M$. This vector is overwritten at every time point t_j.
EEE	*Eventual European value* — vector of length i, where the i^{th} entry EEE^i, $i = 1,\ 2,\ \ldots,\ N$, is equal to the value of the European option E^i at time $\bar{\tau}_j^i$ discounted to time t_j, $j = 0,\ 1,\ \ldots,\ M$. This vector is overwritten at every time point t_j.

Stochastic Mesh Method	
$w_j^{i,k}$	Weight connecting node i at time t_j (that is, S_j^i) to node k at time t_{j+1} (that is, S_{j+1}^k), $j = 0,\ 1,\ \ldots,\ M-1$ and $i, k = 1,\ 2,\ \ldots,\ N$.
$\hat{Z}_j^i$	Modelled continuation value at time t_j, $j = 1,\ 2,\ \ldots,\ M-1$, given the stock price is S_j^i, $i = 1,\ 2,\ \ldots,\ N$.
$f_{j,j+1}(\cdot)$	Probability density function of S_{j+1}, conditional on $S_j = x$, that is $\mathbb{P}\left(S_{j+1} \middle\vert S_j = x\right) = \int_{-\infty}^{\alpha} f_{j,j+1}(x,y)\,dy$.

Dual Method	
η	$= \begin{cases} 1, & \text{if option is a call.} \\ -1, & \text{if option is a put.} \end{cases}$
CSP	*Critical stock price* — vector of length $M+1$, where the j^{th} entry CSP_j $j = 0,\ 1,\ \ldots,\ M$ is the modelled optimal stopping or free boundary at time t_j.

Lévy Processes	
W_t	Standard Brownian motion at time t, distributed Normal$(0,t)$.
X_t^{G}	Gamma process at time t, distributed Gamma$(\alpha t, \beta)$ where $\alpha, \beta \in \mathbb{R}$ and $\alpha, \beta > 0$.
X_t^{VG}	Variance gamma process at time t, distributed $\mathrm{VG}\left(\theta t, \sigma^2 t, \frac{\nu}{t}\right)$ where $\theta, \sigma, \nu \in \mathbb{R}$ and $\sigma, \nu > 0$.
X_t^{IG}	Inverse Gaussian process at time t, distributed IG$(\eta t, \gamma)$ where $\eta, \gamma \in \mathbb{R}$ and $\eta, \gamma > 0$.
X_t^{NIG}	Normal inverse Gaussian process at time t, distributed NIG$(\alpha, \beta, \delta t)$ where $\alpha, \beta, \delta \in \mathbb{R}$ and $\alpha > 0$, $-\alpha < \beta < \alpha$, $\delta > 0$.
$\psi(\cdot)$	Characteristic exponent, Lévy symbol or Lévy exponent.
$l(\cdot)$	Laplace exponent.
J_X	Random jump measure of càdlàg process X.
$\left(\left(\sigma^X\right)^2, \nu^X, \gamma^X\right)$	Lévy triplet of Lévy process X where $\sigma^X \in \mathbb{R}$ and $\sigma^X > 0$, ν^X is the Lévy measure of X and $\gamma^X \in \mathbb{R}$.
γ_0^X	Drift of Lévy process X equal to $\gamma^X - \int_{\lvert x\rvert \leq 1} x\, \nu^X(dx)$.
$K_a(\cdot)$	Modified Bessel function of the second kind with index a.
$\Gamma(\cdot)$	Gamma function.
$\Re(z)$	Real part of complex number z.

Introduction

In this thesis we consider the pricing of vanilla American options where the stock price dynamics follow exponential pure jump Lévy models, using Monte Carlo methods.

Several methods for pricing American options that do not rely on simulation exist for single underlying assets, such as finite difference methods, binomial trees or other lattice methods; and the QUAD [Andricopoulos et al., 2003], CONV [Lord et al., 2008] or COS methods [Fang and Oosterlee, 2009]. Generally these methods are computationally much faster than methods requiring simulation for a single underlying, but are not feasible for multidimensional problems. However, the methods and models we consider are extendable to more than one underlying — as noted by Fu et al. [2001] —"Since the convergence rate of Monte Carlo methods is generally independent of the number of state variables, it is clear that they become viable as the underlying models (asset prices and volatilities, interest rates) and derivative contracts themselves (defined on path-dependent functions or multiple assets) become more complicated."

This thesis presents a comparison between Monte Carlo methods for American options. Despite the fact that these Monte Carlo methods are able to compete with the methodologies mentioned above only in the multidimensional case, we consider only the one-dimensional vanilla American case. Our aim is to present a clear layout of the essential workings of the models and how they compare against each other. We only consider a single underlying, even though all these models can be extended to multiple underlyings; and we apply these models to vanilla options, even though they can be implemented for a variety of exotic options.

A general introduction to American Monte Carlo methods is provided in Chapter 1. In general, American Monte Carlo methods produce estimates whose expectation is lower or higher than the true American option price. We refer to these estimates as low or high bias estimates. A formal definition for bias is given in Chapter 1.

This is not an exhaustive study of American Monte Carlo methods. We do not consider methods such as random tree methods (see Broadie and Glasserman [1997b] and also Glasserman [2004, §8.3]), state-space partitioning (see Barraquand and Martineau [1995], Bally and Pagès [2003] and also Glasserman [2004, §8.4]) or policy iteration (see Kolodko and Schoenmakers [2006] and Bender and Schoenmakers [2006]).

The American Monte Carlo methods we will study are

- the *least squares Monte Carlo method* in Chapter 2. The least squares Monte Carlo method was introduced by Longstaff and Schwartz [2001]. This algorithm is from a class of methods that make use of regression to approximate the continuation value from simulated paths. Other methods

combining regression and Monte Carlo that have been proposed include those given by Carrière [1996] and Tsitsiklis and Van Roy [1999, 2001]. We briefly mention differences between the least squares Monte Carlo method and the method proposed by Tsitsiklis and Van Roy [2001] which we will refer to as the *regression method*. The regression method produces estimates of the option price which have a high bias, whereas the algorithm proposed by Longstaff and Schwartz [2001] produces an approximation of the option price where the bias cannot be quantified — the methodology includes both high and low biasing factors. The convergence of the least squares Monte Carlo method to the exact solution was proved by Clément et al. [2002] in the Brownian motion case.

Furthermore, in Chapter 3 we consider some variance reduction techniques discussed in Rasmussen [2002, 2005] applicable to the least squares Monte Carlo method. Here, we consider a particular control variate applicable to the least squares Monte Carlo method, as well as a method for finding a smooth modelled exercise boundary which is referred to as dispersion. These methods improve the convergence of the least squares Monte Carlo method considerably. Thus, subsequent to this chapter, whenever we use the least squares Monte Carlo method, it will be in conjunction with these techniques.

- the *stochastic mesh method* in Chapter 4. The stochastic mesh method first appeared in Broadie and Glasserman [1997a]. We will refer to Broadie and Glasserman [2004], which is the revised version of this working paper, as well as Glasserman [2004, §8.5]. The stochastic mesh method finds high and low bias estimates of the actual option price. Furthermore, conditions under which the method converges, as the computational effort increases, are given by Broadie and Glasserman [2004]. The computational effort required for this method is linear in the number of exercise dates and quadratic in the number of points in the mesh. This is in contrast with the random tree method shown by Broadie and Glasserman [1997b] which has an exponential dependence on the number of exercise dates.

- the *dual method* in Chapter 5. The dual method, independently studied by Andersen and Broadie [2004], Rogers [2002] and Haugh and Kogan [March 2004], finds a high bias estimate of an American option price by extracting a martingale from an existing stopping rule. In this chapter we first provide an original algorithm that determines an approximation to the free boundary from a stopping rule. We call this approximation the critical stock price function, and determine a low bias estimate using this function. Then we discuss how the dual method derives the high bias estimate using this function. We frequently refer to Glasserman [2004, §8.7] when studying this method.

  We applied the dual method to both the least squares Monte Carlo and stochastic mesh methods and found that it performed very poorly for the stochastic mesh method. On the other hand, when applying the dual method to the least squares Monte Carlo method (along with the variance reduction techniques of Rasmussen [2002, 2005]) we obtained better results. We believe this is because the critical stock price function obtained from the stochastic mesh method is unacceptable, whereas the critical stock price function of the LSM method is much better.

In the literature on the methods discussed above, the underlying process follows geometric Brownian motion. In Chapter 7 we review the generation of stock price paths that follow geometric Brownian motion. We will consider two generators that will help achieve this

- a pseudo-random number generator called *Mersenne Twister* introduced by Matsumoto and Nishimura [1998] in §7.3.
- the quasi-random number generator of Sobol' [1967] in §7.4. Approaches shown by Jäckel [2002], Joe and Kuo [2003] and Glasserman [2004] on the implementation of the Sobol' generator are considered in Appendix B. We also study bridge sampling introduced by Caflisch and Moskowitz [1995] and Moskowitz and Caflisch [1996] which should always be implemented whenever Sobol' random numbers are used.

Our focus in this thesis is to implement the least squares Monte Carlo (including the variance reduction techniques and high bias from the dual method) and stochastic mesh methods for pure jump Lévy models. In Chapter 6 we present an introduction to Lévy processes and consider specific properties that apply to the models we consider. We discuss two pure jump Lévy models, namely

- the *variance gamma model* in Chapter 8. The variance gamma model was introduced into finance by Madan and Seneta [1987]. Variance gamma processes form a special case of the CGMY processes which were first considered in finance by Geman et al. [2001] and Carr et al. [2002]. Further important references on the variance gamma model include Madan and Seneta [1990] and Madan and Milne [1991]. We will however be concerned with the variance gamma model as presented in Madan et al. [1998]. Here, the asymmetric variance gamma model is introduced and shown to be equivalent to a gamma time-changed Brownian motion with drift. Madan et al. [1998] also deduce formulae for the variance gamma density in terms of Bessel functions.
- the *normal inverse Gaussian model* in Chapter 9. The normal inverse Gaussian, along with hyperbolic processes, forms a subclass of generalised hyperbolic processes which were proposed by Barndorff-Nielsen [1977]. The normal inverse Gaussian model was originally applied in finance by Barndorff-Nielsen [1995] and the hyperbolic processes by Eberlein and Keller [1995] and Eberlein et al. [1998]. Further references on the normal inverse Gaussian model include Barndorff-Nielsen [1997, 1998] and Rydberg [1996a,b, 1997]. Barndorff-Nielsen [1997] notes that the log of returns of asset prices can often be fitted extremely well by the normal inverse Gaussian distribution.

In Chapter 10 we discuss the risk-neutral modelling of the variance gamma and normal inverse Gaussian processes. We consider the calibration of these models in Chapter 11. This calibration is a variation of the method of Cont and Tankov [2004b]. We propose a modification which requires moment matching of the variance gamma and normal inverse Gaussian models. The results obtained from the moment matching then serve as a prior to the calibration procedure. Some of the material in this chapter is original and some is joint work with Graeme West.

Finally, in Chapter 12, we implement all the Monte Carlo pricing methods previously considered, where the underlying follows exponential variance gamma and normal inverse Gaussian processes. Here, we discuss the results obtained from this implementation and make concluding remarks.

Several appendices which provide technical tools that are required, but are not central to the theme of this thesis, are included, and will be referred to as they are needed. All of the models considered in this thesis were implemented in c++ using an x64-bit Intel® Core™ i7 CPU M 620 @ 2.67GHz and 8.00GB RAM.

Part I

American Monte Carlo Methods

Chapter 1

Introduction to American Monte Carlo Methods

An *option* is a derivative security, that is, a financial contract whose value is derived from that of a more basic security, such as a bond or stock [Karatzas and Shreve, 1988, §2.1]. A *vanilla call (put) option* is an option in which the holder has the right to buy (sell) the underlying security for a contractually specified price also referred to as the strike price. A *European option* can only be exercised at its expiration date. For example, we define a European call option on an underlying security $\{S_t\}_{t\in[0,T]}$ to be a contract with a payoff $I_T = \max\{S_T - K, 0\}$ where $K > 0$ is the strike price and T its expiration date.

Unlike a European option, an *American option* can be exercised any time up to its expiration. Thus the option holder is continually faced with the choice to either exercise or hold the option. If we consider the example of a call option again, this means that at any exercise time $t \in [0, T]$, the option holder determines whether $I_t = \max\{S_t - K, 0\}$ [1] is worth more or less than the value of holding the option to exercise later. This value of holding is referred to as the *continuation value*. However the continuation value is not available in closed form, and so it is estimated using one of several models.

The form of the model is often an estimate of the *optimal stopping boundary*. This is a function of time, $f(t)$, which determines exercise behaviour. If $S_t > f(t)$, then one exercises (holds) if the option under consideration is a call (put), and if $S_t < f(t)$, then one holds (exercises) if the option is a call (put). The existence of such an optimal stopping boundary requires certain technical conditions which we will briefly mention in §1.2.

The earliest investigation into the pricing of American options is given by McKean [1965] who writes the American option price explicitly up to knowing the optimal stopping boundary. The study of properties of the optimal stopping boundary is done further by van Moerbeke [1976]. Later, Bensoussan [1984] and Karatzas [1988] provide arbitrage arguments which show that the price of an American option is the solution to the optimal stopping problem.

It is this optimisation problem which causes difficulty when pricing options using Monte Carlo simulation. In particular, substantial computational effort is required. As noted in Fu et al. [2001], standard

[1] For any $t \leq T$, I_t is referred to as the *intrinsic value* of the option.

simulation techniques generate sample paths for the underlying forward in time and the majority of path-dependent options are easily priced by simulating these paths. On the other hand, Fu et al. [2001] observe that pricing American options requires a backward algorithm (which we discuss in §1.3). Essentially the complication arises when applying forward simulation to a problem that requires a backward algorithm. We study two methods which address this problem, namely the least squares Monte Carlo method in Chapter 2 and the stochastic mesh method in Chapter 4.

In this chapter we will give a general formulation of the problem, notation and concepts of American Monte Carlo simulation that will be applicable to both methods. We found the introduction given by Glasserman [2004, §8.1] very informative on this topic.

1.1 Assumptions

We begin our discussion with the well-known assumptions, namely that of efficient markets, that no transaction costs are incurred when buying or selling assets, the ability to buy and sell fractional parts of assets, the ability to buy and sell assets as much and as often as one wishes, and that there are no restrictions on short-selling of assets in the market.

We will consider an American option written on a single underlying stock price process $S = \{S_t\}_{t\in[0,T]}$. We will also assume the stock price process follows a Markov process, that is, the evolution of the stock price only depends on its present state and is independent of its history (see for example Shreve [2004, Definition 2.3.6] for a formal definition). This assumption is an important requirement in the American Monte Carlo methods which we will consider.

Furthermore, we will assume that the option can only be exercised at discrete times $0 = t_0 < t_1 < t_2 < \ldots < t_M = T$. The time periods between exercise times are not necessarily equal in length, and we denote these time intervals by $\Delta t_j := t_j - t_{j-1}$ for $j = 1,\ 2,\ \ldots,\ M$. Options in which exercise can only occur at discrete times are known as *Bermudan options*. However, the value of an option with a finite set of exercise dates can be viewed as an approximation to an option which allows for continuous exercise; the greater M (assuming some roughly uniform distribution of the t_j), the better the approximation. See DuPuis and Wang [2005] for a detailed presentation on the convergence of Bermudan prices to American prices. Also, as noted by Glasserman [2004, §8.1], the valuation of these Bermudan options already poses a significant challenge to Monte Carlo. Hence we will only consider the valuation of these finite-exercise options, and from now on refer to them as American options.

Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space. Here Ω is the set of all possible realisations of the financial market, that is, the possible realisations of stock price paths. $\mathcal{F}$ is a σ-algebra and $\mathbb{P}$ is a probability measure defined on $\mathcal{F}$. We assume there exists a risk-neutral measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ under which all asset prices are martingales relative to a numéraire. This numéraire is the bank account, which we denote by A, and is given by $A_t = e^{rt}$ where r is the constant continuously compounded risk-free rate. For the discrete times $t_0 < t_1 < t_2 < \ldots < t_M$ mentioned above, let $\{S_j\}_{j=0,\ 1,\ \ldots,\ M}$, with $S_j := S_{t_j}$, indicate the stochastic process that models the underlying asset price; and let $\{A_j\}_{j=0,\ 1,\ \ldots,\ M}$ denote the numéraire process, with $A_j := e^{rt_j}$. Furthermore, suppose $\{S_j\}$ is adapted to the filtration given by $\{\mathcal{F}_j\}_{j=0,\ 1,\ \ldots,\ M}$, with $\mathcal{F}_j := \mathcal{F}_{t_j}$, where $\mathcal{F}_j$ models the information available at time t_j. Then we have that the tuple $(\Omega, \mathcal{F}, \mathbb{P}, \{S_j\}, \{A_j\}, \{\mathcal{F}_j\})$ represents the *securities market model*.

Unless otherwise specified, when considering a time t_j, the index j will always be for $j = 0,\ 1,\ \ldots,\ M$.

Also, the number of the sample path i under consideration will always be for $i = 1,\ 2,\ \ldots,\ N$ unless we explicitly state otherwise. We refer to a discretised process X_j^i; thus X_j^i indicates the value of the process at time t_j in sample path i. In particular, we refer to the stock price S_j^i and the intrinsic value process $I_j^i := I_j\left(S_j^i\right)$, where $I_j^i = \max\left\{S_j^i - K, 0\right\}$ for a call and $I_j^i = \max\left\{K - S_j^i, 0\right\}$ for a put.

1.2 Optimal Stopping Problem

As we have discussed previously, the value of an American option is found by determining the value achieved from exercising optimally. If, at maturity t_M, the option is in the money, the option holder exercises; otherwise lets the option expire. At any exercise time before time t_M, the option holder exercises if the intrinsic value at that time is greater than their model of the continuation value, or waits until the next exercise time if not, and checks again. It is this continuation value which is approximated by the least squares Monte Carlo method in Chapter 2 and the stochastic mesh method in Chapter 4.

Thus, in the most general sense, an American contingent claim V is governed by a payoff process $\{I_t\}_{t\in[0,T]}$. The holder has the right to exercise at any time t and so $V_t \geq I_t$. The value at time t of such a claim, V_t, is determined by optimal exercise. Therefore the holder must choose an exercise time, τ^+, at which he or she expects to receive the greatest discounted payoff, and so, as we show below, $V_t = \sup_{\tau^+\in[0,T]} \mathbb{E}\left[e^{-r\tau^+} I_{\tau^+}\right]$.

In our discussion of the optimal stopping problem, we will refer to stopping times. Therefore we briefly introduce stopping times. A positive random variable $\tau : \Omega \to \mathbb{R}^+$ that represents the time at which some event is going to take place is known as a *random time*. Some examples of random times are the first time a stock price reaches 100 or the time when a stock price reaches its maximum in some time period. The value of τ is dependent on the path $\omega \in \Omega$ and so the times in the examples above will be different for different ω's.

Definition 1.2.1 Stopping Time
Let $\{\mathcal{F}_t\}_{t\geq 0}$ be a filtration on $(\Omega, \mathcal{F}, \mathbb{P})$. A random time τ is an $\mathcal{F}_t$-stopping time *if*

$$\{\omega \in \Omega : \tau(\omega) \leq t\} \in \mathcal{F}_t.$$

for all $t \geq 0$ [Varadhan, 2007, Definition 1.4].

Suppose τ is a random time, that is, a random variable giving the time at which some event occurs. Then τ is a stopping time if and only if for all $t \geq 0$, it is known at time t whether or not $\tau \leq t$, that is, whether or not that event has occurred. So the first time the stock price reaches a 100 is a stopping time. However, the time when the stock price reaches its maximum in some time period is not a stopping time, because only at the end of the time period it is possible to say when the maximum was reached.

The value of the American option at time t_0 is given by the solution to the *optimal stopping problem*

$$V_0 = \sup_{\tau\in\mathcal{T}_{0,M}} \mathbb{E}\left[e^{-r\tau} I_\tau\right] \tag{1.1}$$

(see for example Björk [2004, §21.2]) where $\mathcal{T}_{j,M}$ is the set of all stopping times with values in $\{t_j,\ t_{j+1},\ \ldots,\ t_M\}$, $\{I_j\}_{j=0,\ 1,\ \ldots,\ M}$, with $I_j := I_{t_j}$, is the adaptive payoff (intrinsic value) process; and the expectation is taken under the risk-neutral measure. Consider Figure 1.1 where we assume a known

optimal stopping boundary and simulate a single path of the underlying process hitting this boundary at the optimal stopping time τ.

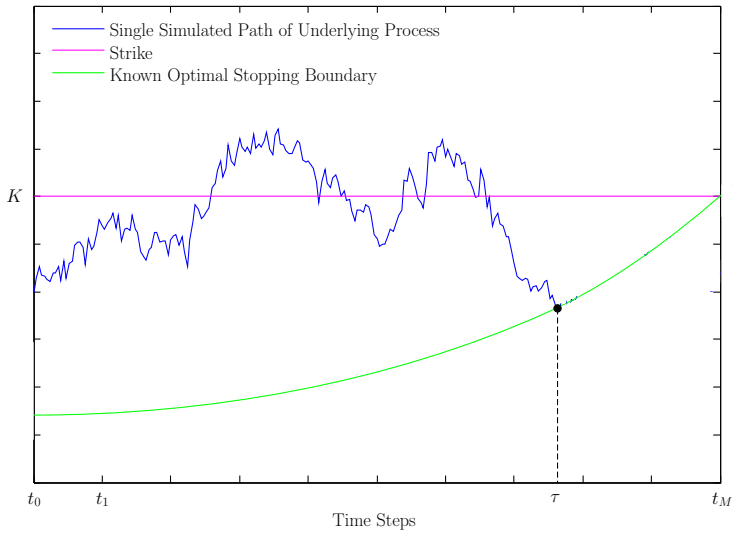

Figure 1.1: *Consider an American put option with strike K (pink) and suppose the optimal stopping boundary (green) is known. If the simulated path (blue) for the underlying process is followed, then the stopping time τ is the optimal time to exercise the option.*

As mentioned before, the proof, which allows us to call (1.1) the price of an American option, was first given in Bensoussan [1984] and Karatzas [1988]. See also Duffie [2001]. Furthermore, if we ignore discounting, in order to guarantee the existence of $\mathbb{E}[I_\tau]$ in (1.1), the following condition is required [see Peskir and Shiryaev, 2006, §1.1]

$$\mathbb{E}\left[\sup_{\tau\in\mathcal{T}_{0,M}} |I_\tau|\right] < \infty.$$

Throughout this thesis, we will assume that the above condition holds, that is the existence of optimal stopping times. For further details see Peskir and Shiryaev [2006, §1.1].

1.3 Dynamic Programming Formulation

The American Monte Carlo methods we will consider all make use of an algorithm called *dynamic programming* [see Glasserman, 2004, §8.1] to find the value of an American option. This algorithm is a recursive representation of the Snell envelope (see for example Lamberton and Lapeyre [2008, §2.2], Elliott and Kopp [2005, §5.4] or Peskir and Shiryaev [2006, §1.1.3]):

Let $\{V_j\}_{j=0,\,1,\,\ldots,\,M}$, with $V_j := V_{t_j}$, be the value process of the American option, then

$$V_j = \sup_{\tau\in\mathcal{T}_{j,M}} \mathbb{E}\left[e^{-r(\tau-t_j)} I_\tau \middle| \mathcal{F}_j\right]. \tag{1.2}$$

$\{V_j\}_{j=0,\,1,\,\ldots,\,M}$ is called the *Snell envelope* of $\{I_j\}_{j=0,\,1,\,\ldots,\,M}$, with $I_j := I_{t_j}$. Moreover, if we let

$$Z_j := \sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[e^{-r(\tau - t_j)} I_\tau \middle| \mathcal{F}_j\right] \tag{1.3}$$

then

$$\begin{aligned} V_j &= \sup_{\tau \in \mathcal{T}_{j,M}} \mathbb{E}\left[e^{-r(\tau - t_j)} I_\tau \middle| \mathcal{F}_j\right] \\ &= \max\left\{I_j, \sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[e^{-r(\tau - t_j)} I_\tau \middle| \mathcal{F}_j\right]\right\} \\ &= \max\left\{I_j, Z_j\right\}. \end{aligned}$$

This is the *dynamic programming principle*:

$$V_M = I_M \tag{1.4}$$

$$V_j = \max\left\{I_j, Z_j\right\} \text{ for } j = 0,\ 1,\ \ldots,\ M-1. \tag{1.5}$$

Z_j is known as the *continuation* or *holding value*. Using the tower property of conditional expectations, we may rewrite (1.3) as:

$$\begin{aligned} Z_j &= \sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[e^{-r(\tau - t_{j+1} + t_{j+1} - t_j)} I_\tau \middle| \mathcal{F}_j\right] \\ &= e^{-r\Delta t_{j+1}} \sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[e^{-r(\tau - t_{j+1})} I_\tau \middle| \mathcal{F}_j\right] \\ &= e^{-r\Delta t_{j+1}} \sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[\mathbb{E}\left[e^{-r(\tau - t_{j+1})} I_\tau \middle| \mathcal{F}_{j+1}\right] \middle| \mathcal{F}_j\right] \\ &\leq e^{-r\Delta t_{j+1}} \mathbb{E}\left[\sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[e^{-r(\tau - t_{j+1})} I_\tau \middle| \mathcal{F}_{j+1}\right] \middle| \mathcal{F}_j\right] \\ &= e^{-r\Delta t_{j+1}} \mathbb{E}\left[V_{j+1} \middle| \mathcal{F}_j\right]. \end{aligned} \tag{1.6}$$

Let us assume, as in §1.2, the existence of an optimal stopping time. Now if $\tau^* \geq t_{j+1}$ is the optimal time in the time period $[t_{j+1}, t_M]$, that is, $V_{j+1} = \mathbb{E}\left[e^{-r(\tau^* - t_{j+1})} I_{\tau^*} \middle| \mathcal{F}_{j+1}\right]$, then

$$\begin{aligned} e^{-r\Delta t_{j+1}} \mathbb{E}\left[V_{j+1} \middle| \mathcal{F}_j\right] &= e^{-r\Delta t_{j+1}} \mathbb{E}\left[\sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[e^{-r(\tau - t_{j+1})} I_\tau \middle| \mathcal{F}_{j+1}\right] \middle| \mathcal{F}_j\right] \\ &= e^{-r\Delta t_{j+1}} \mathbb{E}\left[\mathbb{E}\left[e^{-r(\tau^* - t_{j+1})} I_{\tau^*} \middle| \mathcal{F}_{j+1}\right] \middle| \mathcal{F}_j\right] \\ &= e^{-r\Delta t_{j+1}} \mathbb{E}\left[e^{-r(\tau^* - t_{j+1})} I_{\tau^*} \middle| \mathcal{F}_j\right] \\ &\leq \sup_{\tau \in \mathcal{T}_{j+1,M}} \mathbb{E}\left[e^{-r(\tau - t_{j+1} + t_{j+1} - t_j)} I_\tau \middle| \mathcal{F}_j\right] \\ &= Z_j \end{aligned}$$

Thus, it follows that $Z_j = e^{-r\Delta t_{j+1}} \mathbb{E}\left[V_{j+1} \middle| \mathcal{F}_j\right]$.

When calculating the American option value using Monte Carlo methods, we will begin by setting the value of the option at time t_M equal to the intrinsic value at time t_M as in (1.4). In this thesis we will

consider several models for estimating the continuation value Z_j: in Chapter 2 the least squares Monte Carlo method, in Chapter 3 the least squares Monte Carlo method combined with a control variate and in Chapter 4 the stochastic mesh method. Then, proceeding backwards in time, the modelled value of the option at time t_j is calculated as the maximum of the intrinsic value and this estimate of Z_j at time t_j for $j = M-1,\ M-2,\ \ldots,\ 0$ as in (1.5).

1.4 Bias

The *bias* of an estimator is defined to be the difference between the expected value of this estimator and the true value of the variable that is being approximated. We will refer to a *high bias estimator* if its expected value is greater than or equal to the true value of what is being approximated and a *low bias estimator* if the expected value of the estimator is less than or equal to the true value of the quantity being approximated.

Sources of high and low bias affect all methods for pricing American options by simulation. High bias results from the *max* operator used in (1.5), whereas low bias results from following a suboptimal exercise rule. Separating the sources of bias produces a pair of estimates which in expectation straddles the optimal value.

We will denote the true value of the American option and continuation value by V and Z respectively (and thus V_j^i would indicate the option value at time t_j given the stock price is S_j^i etc.). Estimated values will be denoted by an accented letter, e.g. $\bar{V}$ would indicate a modelled option value. In particular, we will denote high bias approximations by using the *hat* accent, thus $\hat{V}$, $\hat{Z}$ etc. and low bias approximations using the *breve* accent, $\breve{V}$, $\breve{Z}$ etc. The high (or low) bias accents will be used either when we have already established that the approximation has a high (or low) bias, or as a promise that this will be established subsequently.

1.4.1 High Bias

Let $\hat{V} = \left\{\hat{V}_j\right\}_{t=t_j}$ denote an estimator which is found when the decision to exercise or not, and calculating the continuation estimate, are based on the same finite sample. This is achieved when the decision is made using an approximation of the continuation value.

Definition 1.4.1 High Bias Dynamic Programming Formulation
Let $\hat{Z}_j$, $j = 0,\ 1,\ \ldots,\ M-1$, denote a model of the continuation value $Z_j = e^{-r\Delta t_{j+1}}\mathbb{E}\left[V_{j+1}|\,\mathcal{F}_j\right]$ which is obtained from a finite number of samples. Then the high bias dynamic programming formulation *is given by*

$$\hat{V}_M = I_M \tag{1.7}$$

$$\hat{V}_j = \max\left\{I_j, \hat{Z}_j\right\} \text{ for } j = 0,\ 1,\ \ldots,\ M-1 \tag{1.8}$$

(see Glasserman [2004, §8.1] for example).

An argument relying on Jensen's inequality shows that $\hat{V}$ has a high bias at every exercise date of the option. It should be noted that the high bias does not result from the use of future information in making

the decision to exercise. If this was the case, a low bias value would result from estimating the price of an option with a concave payoff.

Theorem 1.4.2
For every t_j

$$\mathbb{E}\left[\hat{V}_j \middle| \mathcal{F}_j\right] \geq V_j. \tag{1.9}$$

Proof.
At time t_M (1.9) holds trivially as we have from (1.7) that $\hat{V}_M = I_M$ and from (1.4) that $I_M = V_M$. Next for any j, $j = 0,\ 1,\ \ldots,\ M-1$, we have that

$$\begin{aligned}
\mathbb{E}\left[\hat{V}_j \middle| \mathcal{F}_j\right] &= \mathbb{E}\left[\max\left\{I_j, \hat{Z}_j\right\} \middle| \mathcal{F}_j\right] \\
&\geq \max\left\{I_j, \mathbb{E}\left[\hat{Z}_j \middle| \mathcal{F}_j\right]\right\} \\
&= \max\{I_j, Z_j\} \\
&= V_j
\end{aligned}$$

where the inequality follows from Jensen's inequality and the second equality follows from the law of large numbers. □

1.4.2 Low Bias

A low bias estimator results from suboptimal exercise. Thus, by following some exercise policy, one obtains a low bias estimator since no policy is better than an optimal policy.

Such an estimator, $\breve{V} = \left\{\breve{V}_j\right\}_{t=t_j}$, can be created by splitting up sample information into two disjoint groups independent of each other, as in Glasserman [2004, §8.3.2]. One set of information will be used to determine the exercise policy, while the other set will be used to estimate the continuation value. We will now show that the estimator obtained in this way results in a low bias.

Suppose that at time t_j we want to calculate the low bias estimator $\breve{V}_j$. We set at time t_M, as usual, $\breve{V}_M = I_M$. Also, let ${}_1\breve{Z} := \left\{{}_1\breve{Z}_j\right\}_{t=t_j}$ and ${}_2\breve{Z} := \left\{{}_2\breve{Z}_j\right\}_{t=t_j}$ denote the estimated continuation value determined from the first and second sample sets respectively. Then

$$\breve{V}_j = \begin{cases} I_j & \text{if } {}_1\breve{Z}_j \leq I_j \\ {}_2\breve{Z}_j & \text{otherwise.} \end{cases} \tag{1.10}$$

A general argument given by Glasserman [2004, §8.3.2] shows that $\breve{V}$ is biased low at every exercise date of the option, that is:

Theorem 1.4.3
For every t_j

$$\mathbb{E}\left[\breve{V}_j \middle| \mathcal{F}_j\right] \leq V_j. \tag{1.11}$$

Proof.
As in Theorem 1.4.2, at time t_M (1.11) holds trivially as we have by definition that $\breve{V}_M = I_M$ and from (1.4) that $I_M = V_M$. From (1.10) it follows that for every $j = 0,\ 1,\ \ldots,\ M-1$

$$\begin{aligned}\breve{V}_j &= \max\left\{I_j, {}_2\breve{Z}_j\right\} \\ &= \mathbb{1}_{\left\{{}_1\breve{Z}_j \leq I_j\right\}} I_j + \mathbb{1}_{\left\{{}_1\breve{Z}_j > I_j\right\}} {}_2\breve{Z}_j.\end{aligned}$$

Hence, from the law of large numbers, we have that

$$\mathbb{E}\left[\breve{V}_j \middle| \mathcal{F}_j\right] = \mathbb{P}\left({}_1\breve{Z}_j \leq I_j\right) I_j + \left(1 - \mathbb{P}\left({}_1\breve{Z}_j \leq I_j\right)\right) Z_j.$$

Hence,

$$\begin{aligned}\mathbb{E}\left[\breve{V}_j \middle| \mathcal{F}_j\right] &\leq \max\{I_j, Z_j\} \\ &= V_j.\end{aligned}$$

□

${}_1\breve{Z}$ and ${}_2\breve{Z}$ can be created by splitting information in many different ways. Broadie and Glasserman [1997a] make several splits and then the average of the resulting $\breve{V}_j$'s is taken to be the value at time t_j.

Chapter 2

The Least Squares Monte Carlo Method

Regression-based methods for pricing American options have been proposed in various papers, in particular Carrière [1996], Tsitsiklis and Van Roy [1999, 2001] and Longstaff and Schwartz [2001].

Tsitsiklis and Van Roy [1999, 2001] make use of regression to approximate the continuation value from simulated paths. Furthermore, Tsitsiklis and Van Roy [2001] provide some convergence results on this method which we will refer to as the *regression method*. Results obtained by Glasserman [2004, Table 8.1] show that this method can have a very high bias.

In this chapter we will focus on the *least squares Monte Carlo (LSM) method* as introduced by Longstaff and Schwartz [2001]. Longstaff and Schwartz [2001] combine the approximation of the continuation value obtained by regression with what Glasserman [2004, p.449] refers to as an *interleaving estimator*. This estimator mingles the sources of high and low bias and so hopefully the bias of this method (which could be in either direction) is not too severe. Clément et al. [2002] prove the convergence of the LSM method to the exact solution.

2.1 The Least Squares Approach of Longstaff and Schwartz [2001]

The LSM approach approximates the continuation value Z_j by using least squares. In fact, the LSM approach consists of two approximations:

2.1.1 Truncated Series Approximation

As in §1.1, suppose that the underlying stock price process $S = \{S_j\}_{t=t_j}$ follows a Markov process. Let L_k, $k = 0,\ 1,\ \ldots$ be functions that form a total system[1] of orthogonal polynomials in $L^2\left(\mathbb{R}^+\right)$ (e.g. Laguerre or

[1] Recall that a total set in a space is one whose linear span is dense. Thus, implying that every square-integrable function can be approximated arbitrarily closely by a linear combination of L_k.

Hermite polynomials[2]), where $L^2(\mathbb{R}^+)$ is the set of all square-integrable real-valued functions $f : \mathbb{R}^+ \to \mathbb{R}$.

Longstaff and Schwartz [2001] propose that the continuation value at S_j be written as a series in terms of the L_k (see also Glasserman [2004, §8.6.1] or Clément et al. [2002])

$$Z_j(S_j) = \sum_{k=0}^{\infty} {}_k\beta_j L_k(S_j) \tag{2.1}$$

where the ${}_k\beta_j$ are the coefficients in the L^2 expansion of $Z_j(\cdot)$. Note that here we assume the continuation value is $\mathcal{F}_j$-measurable, that is, it only depends on S_j at time t_j. We also assume that Z_j is square-integrable. Now from the fact that L_k are total in $L^2(\mathbb{R}^+)$ and using the Doob-Dynkin lemma, we are able to write the continuation value in this form. Since the L_k are orthogonal, (2.1) may then be approximated by the projection onto the subspace of $L^2(\mathbb{R}^+)$ spanned by $L_0,\ L_1,\ \ldots,\ L_\kappa$

$${}_\kappa Z_j(S_j) := \sum_{k=0}^{\kappa} {}_k\beta_j L_k(S_j). \tag{2.2}$$

Zastawniak [February 2009] calls this approximation the *truncated series approximation*.

According to Longstaff and Schwartz [2001] the choice of basis functions makes little difference in the resulting option value. The number of polynomials will typically be about 10. Moreno and Navas [2004] report that the pricing of a standard American put is very robust with respect to the choice and number of basis functions. However they show that this is not the case for more complex options. Glasserman [2004, §8.6] notes that the choice of basis functions of any regression-based method undoubtedly affects how well the method performs. We found this to be the case in the examples we were considering — see Figure 2.1 which include the value given by a binomial tree[3]. Here, and in the rest of the thesis, S_0 denotes the spot price of the underlying, K will indicate the strike, as before r denotes the constant continuously compounded risk-free rate of interest and q the constant continuously compounded dividend yield.

2.1.2 Least Squares Regression Monte Carlo Approximation

In this section, estimators are indicated by a *bar* accent, since the approximations given here have a mixed bias, as we will discuss in §2.2.

For the second approximation, Monte Carlo simulations and least squares are used to approximate the coefficients ${}_0\beta_j,\ {}_1\beta_j,\ \ldots,\ {}_\kappa\beta_j$ appearing in (2.2). This is achieved as follows:

(i) Generate a matrix of stock prices S_j^i.

(ii) At the maturity t_M of the option under consideration, define the model option value and stopping time to be

$$\bar{V}_M^i = I_M^i \text{ and } \bar{\tau}_M^i = \begin{cases} t_M & \text{if } I_M^i > 0, \\ \infty & \text{otherwise} \end{cases} \tag{2.3}$$

[2]Orthogonality is defined with respect to an inner product which depends on a measure of integration. Laguerre polynomials are orthogonal over $[0, \infty)$ with respect to the exponential distribution and Hermite polynomials are orthogonal over $(-\infty, \infty)$ with respect to the normal distribution.

[3]Here and later the binomial tree is constructed to have the same number of exercise possibilities as the given Bermudan option, but with steps in between these exercise dates where early exercise is not permitted. The number of steps in-between exercise opportunities will typically be something like 100, so that the lognormal distribution is well modelled by the binomial tree. Thus the option price found by using this binomial tree should be very accurate.

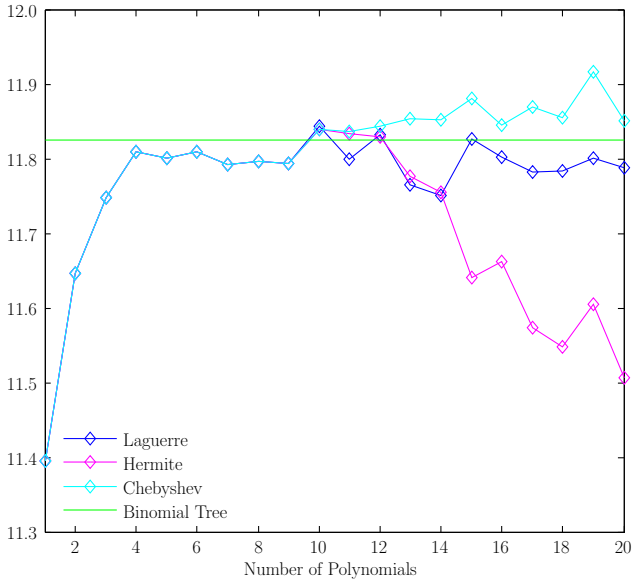

Figure 2.1: *Performance of the LSM method using different types and number of polynomials.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $S_0 = 135$, $\sigma = 30\%$, $r = 10\%$ and $q = 2\%$; and we are considering a 1 year put with $K = 135$. We generated 4096 sample paths with 20 time steps using techniques we discuss in §7.4 and §7.5. We also include the approximation given by the binomial tree with 20 exercise opportunities.

In this example we see that the Hermite polynomial produces unacceptable answers and we also observe that at least 4 polynomials are required for a good approximation. Similar results were obtained when varying the volatility, term, number of time steps and number of sample paths — in most cases the Laguerre polynomials performed well, whereas the Hermite polynomials performed poorly. We found that far out-the-money options produced poor results for all three types of polynomials.

respectively, where we set $I^i_M = 0$ if $\bar{\tau}^i_M = \infty$.

(iii) Step back to time t_{M-1}.

- For each path i, calculate the discounted payoff, that is $e^{-r\left(\bar{\tau}^i_M - t_{M-1}\right)} I^i_{\bar{\tau}^i_M}$. Then use least squares to find the best approximation spanned by the L_k to this discounted payoff. We will refer to this best approximation as the *model continuation value* and denote it by ${}_\kappa\bar{Z}_{M-1}$. Thus, find the coefficients ${}_k\bar{\beta}_{M-1} \in \mathbb{R}$ that minimise

$$\sum_{i=1}^{N} \left| e^{-r\left(\bar{\tau}^i_M - t_{M-1}\right)} I^i_{\bar{\tau}^i_M} - \sum_{k=0}^{\kappa} {}_k\bar{\beta}_{M-1} L_k\left(S^i_{M-1}\right) \right|^2 \tag{2.4}$$

and set

$${}_\kappa\bar{Z}_{M-1}\left(S^i_{M-1}\right) := \sum_{k=0}^{\kappa} {}_k\bar{\beta}_{M-1} L_k\left(S^i_{M-1}\right). \tag{2.5}$$

As is well known, the coefficients in this expansion are found using least squares. Zastawniak [February 2009] refers to this second approximation as the *least squares regression Monte Carlo estimator* and Clément et al. [2002] refer to this as the *Monte Carlo procedure*.

- Now use the model continuation value to decide at time t_{M-1} whether to exercise or hold the option: if the model continuation value is greater than the intrinsic value, then hold, otherwise exercise. Thus, define the model stopping time $\bar{\tau}^i_{M-1}$ and the value of the option (as determined by the LSM method) $\bar{V}^i_{M-1}$ as follows: if $I^i_{M-1} >_\kappa \bar{Z}_{M-1}\left(S^i_{M-1}\right)$, then set

$$\bar{V}^i_{M-1} = I^i_{M-1} \text{ and } \bar{\tau}^i_{M-1} = t_{M-1} \tag{2.6}$$

otherwise set

$$\bar{V}^i_{M-1} = e^{-r\left(\bar{\tau}^i_M - t_{M-1}\right)} I^i_{\bar{\tau}^i_M} \text{ and } \bar{\tau}^i_{M-1} = \bar{\tau}^i_M \tag{2.7}$$

and hence we may write

$$\bar{V}^i_{M-1} = e^{-r\left(\bar{\tau}^i_{M-1} - t_{M-1}\right)} I^i_{\bar{\tau}^i_{M-1}}.$$

(iv) In general, at time step t_j, $j = 1,\ 2,\ \ldots,\ M-2$ given the values $\bar{V}^i_{j+1}$ and $\bar{\tau}^i_{j+1}$:

- Determine, as in (2.4), the ${}_k\bar{\beta}_j \in \mathbb{R}$ that minimises

$$\sum_{i=1}^{N} \left| e^{-r\left(\bar{\tau}^i_{j+1} - t_j\right)} I^i_{\bar{\tau}^i_{j+1}} - \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S^i_j\right) \right|^2$$

and set, as in (2.5),

$${}_\kappa\bar{Z}_j\left(S^i_j\right) := \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S^i_j\right).$$

- Again, as in (2.6) and (2.7) define the model stopping time and option value: if $I^i_j >_\kappa \bar{Z}_j\left(S^i_j\right)$, then set

$$\bar{V}^i_j = I^i_j \text{ and } \bar{\tau}^i_j = t_j$$

otherwise set

$$\bar{V}^i_j = e^{-r\left(\bar{\tau}^i_{j+1} - t_j\right)} I^i_{\bar{\tau}^i_{j+1}} \text{ and } \bar{\tau}^i_j = \bar{\tau}^i_{j+1}$$

and hence

$$\bar{V}^i_j = e^{-r\left(\bar{\tau}^i_j - t_j\right)} I^i_{\bar{\tau}^i_j}.$$

Note that at time t_j, $\bar{\tau}^i_j$ will denote the first time t_s, $j \le s \le \infty$ where we exercise the option on the i^{th} path. Consider Figure 2.2 where we plot $\bar{V}_j$, I_j and ${}_\kappa Z_j$ at a particular time step.

(v) Continue in this way until we reach time step t_0 where we let

$$\bar{V}_0 = \max\left\{ I_0, e^{-r\Delta t_1} \frac{1}{N} \sum_{i=1}^{N} \bar{V}^i_1 \right\}. \tag{2.8}$$

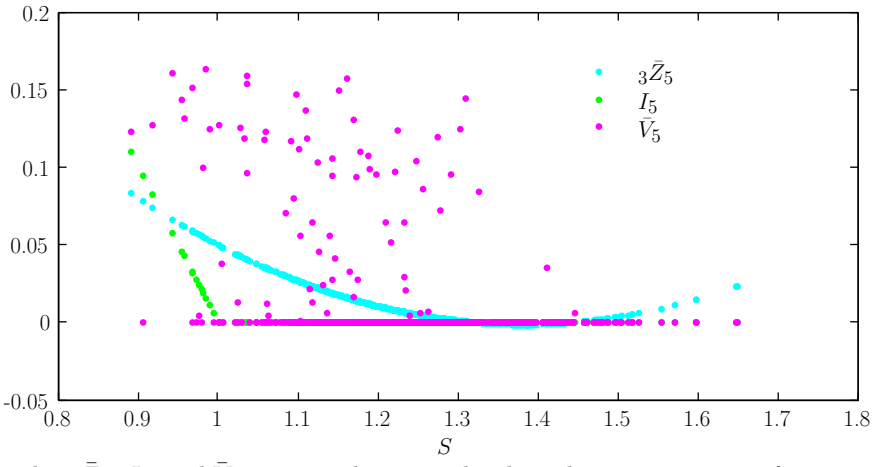

Figure 2.2: *We plot ${}_3\bar{Z}_5$, I_5 and $\bar{V}_5$ against the normalised stock at time step 5 of a 20 step method.*

The option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 20\%$, $S_0 = 110$, $r = 5\%$ and $q = 2\%$; and we are considering a 1 year put option with $K = 90$. Here the stock prices have been normalised by the strike. We simulated 512 paths using simulation techniques we discuss in §7.4 and §7.5.

The model continuation value ${}_3\bar{Z}_5$ is fitted, using least squares, to the first 3 Laguerre polynomials; hence the parabolic form of ${}_3\bar{Z}_5$. Thus ${}_3\bar{Z}_5$ is the parabola closest to the set of the $\bar{V}_5$'s in the least squares sense.

This inductive procedure is presented in pseudo-code in §2.4.

Longstaff and Schwartz [2001] consider only those paths that are in-the-money at a particular time step and thus, in their regression equation, do not sum over all N paths. According to Longstaff and Schwartz [2001] numerical experiments show that more than two or three times as many basis functions are required when using all the paths to reach the same level of accuracy by the approximated value obtained from in-the-money paths[4]. Furthermore, Jonen [2009] notes that for the same number of basis functions the accuracy of the approximation of the continuation value is higher when regressing at time t_j over in-the-money paths only. Rasmussen [2005] also only includes in-the-money paths at each time step. However Clément et al. [2002] and Zastawniak [February 2009] consider all N paths as in (2.11). In a particular example, Glasserman [2004, p.463, 464] notes that using in-the-money paths only led to inferior results. We found that using in-the-money paths only often had little impact on the resulting prices and sometimes produced inferior results.

The Regression Method of Tsitsiklis and Van Roy [2001]

Thus far we have considered the LSM method only. We briefly compare this method to the regression method of Tsitsiklis and Van Roy [2001]. At time t_M the model option value is the same as that of the

[4]Here and in the rest of this thesis, the phrase *in-the-money paths*, means those paths that are in-the-money at a particular time step.

LSM method in (2.3). Stepping back in time, at time t_j we find the coefficients ${}_k\bar{\beta}_j \in \mathbb{R}$ that minimises

$$\sum_{i=1}^{N} \left| e^{-r\Delta t_{j+1}} \bar{V}_{j+1}^i - \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S_j^i\right) \right|^2 . \tag{2.9}$$

This is the first difference between the regression method and the LSM method. Longstaff and Schwartz [2001] note that if one follows the above approach, the option value may have an upward bias. As we mentioned before, Glasserman [2004, Table 8.1] reports significant high bias obtained using this method in a particular example.

After obtaining the regression coefficients, we calculate ${}_\kappa\bar{Z}_j\left(S_j\right)$ as in (2.5). Instead of having (2.6) and (2.7) as in the LSM method, the next step in the regression method is to set

$$\bar{V}_j^i = \max\left\{I_j^i,\ {}_\kappa\bar{Z}_j\left(S_j^i\right)\right\}.$$

As in the LSM method, once we have obtained the $\bar{V}_j^i$, we step back to the next time step t_j and repeat the above procedure. Finally, at time t_0 we find the model of the option value under the regression method as in (2.8).

In Figure 2.3 we plot the performance of the regression method and the LSM method as a function of the number of sample paths. In Table 2.1, we compare values (for in-the-money, at-the-money and out-the-money options) and the time taken (in seconds) of these methods. In these examples we see that the LSM method clearly outperforms the regression method.

2.2 Mixed Bias

In §1.4 we observed that American Monte Carlo methods include sources of either high or low bias. The LSM method, however, has a combination of both. Glasserman [2004, p.49] notes that this combination in the LSM method may lead to a more accurate approximation of the American option value since the biases may offset somewhat. From Definition 1.4.1 and Theorem 1.4.2 we observe that the high bias factor results from following backward recursion in (2.6) and (2.7). Since the time at which we exercise is determined by considering ${}_\kappa\bar{Z}_j\left(S_j\right)$, that is, we exercise at time t_j if $I_j^i > {}_\kappa\bar{Z}_j\left(S_j^i\right)$, we are using an exercise strategy. Thus, as we have shown in §1.4.2, this is the source of the low bias factor.

Immediately after their Proposition 1, Longstaff and Schwartz [2001] imply that for sufficiently large N, estimates increase as κ increases, and because this estimate is bounded above by the true value one can have a convergence criterion in κ. However, Moreno and Navas [2004] observe that when considering a particular polynomial basis, estimates do not increase monotonically with κ, and their results show the difficulty of implementing any intuitively appealing convergence criterion.

Longstaff and Schwartz [2001] suggest a test of convergence for the LSM method in which a different set of paths are used to calculate the conditional expectation approximation than the set of paths to which these approximations are applied. We have an accurate estimate if the option value, obtained as in §2.1, is close to the option value obtained by using the same paths for both estimating the conditional value and applying them.

Recall that the LSM requires two approximations for the continuation value: first the approximation using a finite number of basis functions and secondly the approximation by least squares. Therefore the

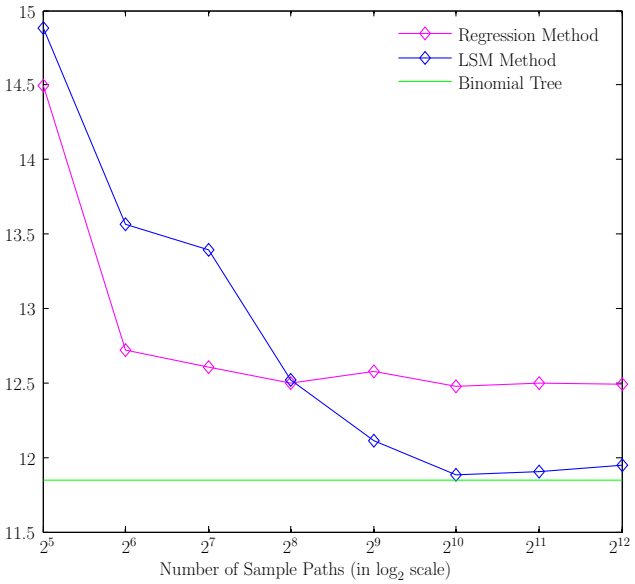

Figure 2.3: *The performance of the regression and LSM methods as a function of the number of sample paths using 8 Laguerre polynomials.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put with* $K = 135$. *The sample paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5. We also include the value given by the binomial tree with 30 exercise opportunities.*

In this example we see that the LSM method clearly outperforms the regression method.

convergence of the LSM method is more involved than other American Monte Carlo methods. Clément et al. [2002] provide an in-depth technical study of the of the LSM method. In particular they prove two theorems with respect to the convergence of the LSM method.

They prove convergence of the projection of the continuation value to the true continuation value as the number of basis functions goes to infinity [Clément et al., 2002, Theorem 3.1], that is,

$$\lim_{\kappa \to \infty} \mathbb{E}\left[{}_{\kappa}Z_j \middle| \mathcal{F}_j \right] = \mathbb{E}\left[Z_j \middle| \mathcal{F}_j \right]$$

in L^2 for $j = 1,\ 2,\ \ldots,\ M-1$. Furthermore, they also prove that for a given number of basis functions κ, the approximated projection found by regression converges to the 'true' projection as $N \to \infty$ [Clément et al., 2002, Theorem 3.2], that is, as $N \to \infty$

$$\frac{1}{N} \sum_{i=1}^{N} {}_{\kappa}\bar{Z}_j \left(S_j^i \right) \to \mathbb{E}\left[{}_{\kappa}Z_j \right]$$

almost surely for $j = 1,\ 2,\ \ldots,\ M-1$. Results with regard to the rate of convergence of the LSM method are also given by Clément et al. [2002], but we will not discuss these here.

	Value		
K	Binomial Tree	Regression	LSM
85	0.4391	0.9252	0.4434
130	9.6090	10.2102	9.6489
135	11.8490	12.4912	11.9490
140	14.3803	15.0600	14.5279
185	50.0000	50.0000	50.0000
	Time (seconds)		
K	Binomial Tree	Regression	LSM
85	0.234	244.767	225.706
130	0.234	222.500	207.975
135	0.292	248.287	232.428
140	0.260	229.498	216.415
185	0.234	308.680	273.419

Table 2.1: *The performance of the regression and LSM methods for various strikes, 4096 sample paths and 8 Laguerre polynomials.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ and $q = 2\%$; and we are considering a 1 year put. The sample paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5. We also include the value given by the binomial tree with 30 exercise opportunities.

In this example we see that the LSM method clearly outperforms the regression method, in particular, when comparing the far out-the-money put estimates.

2.3 Low Bias

In this section we will show how to find $\breve{V}$, that is, a low biased American option value obtained by following the exercise strategy determined by the LSM method or the regression method. Let S_j^i denote the stock prices used to calculate the option value $\bar{V}$.

We simulate another set of stock price paths with the stock price in the i^{th} path at time t_j denoted by $\breve{S}_j^i$. Let $I_j^i := I_j\left(\breve{S}_j^i\right)$ indicate the intrinsic value of the simulated path, given the stock price is $\breve{S}_j^i$. Now for each path i define the time

$$\breve{\tau}^i = \min\left\{j : I_j^i \geq \breve{Z}_j^i\right\}$$

where

$$\breve{Z}_j^i :=_{\kappa}\breve{Z}_j\left(\breve{S}_j^i\right) = \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(\breve{S}_j^i\right) \tag{2.10}$$

and the coefficients ${}_k\bar{\beta}_j \in \mathbb{R}$ are those determined in (2.4) for the LSM method or in (2.9) for the regression

method (using the original stock price paths S_j^i). Once $\breve{\tau}^i$ is obtained we set

$$\breve{V}^i = e^{-r\left(\breve{\tau}^i - t_0\right)} I^i_{\breve{\tau}^i}.$$

When we have moved through time for all paths i, the low bias option value is given by

$$\breve{V} = \frac{1}{N}\sum_{i=1}^{N} \breve{V}^i.$$

$\breve{V}$ is a low bias approximation since the stopping time $\breve{\tau}$ is not necessarily the optimal stopping time. Thus we may write

$$\mathbb{E}\left[\breve{V}\right] \leq V.$$

2.4 Pseudo-Code for the Least Squares Monte Carlo Method

- Suppose that we are given the stock price matrix S_j^i and intrinsic value matrix I_j^i. We now define several vectors of length N at time point t_j for $j = M,\ M-1,\ \ldots,\ 1$ which we will use in the algorithm. Let
  - E indicate the vector where the entry E^i is the *European value*[5] whose term is given by $t_M - t_j$ and spot by S_j^i. The strike and style (call or put) of this European option is equal to that of the American option under consideration;
  - EE denote the vector with entry EE^i equal to $e^{-r\left(\bar{\tau}^i_{j+1} - t_j\right)} I^i_{\bar{\tau}^i_{j+1}}$, that is, the discounted *eventual exercise* value of path i and;
  - TV indicate the vector where the entry TV^i is the modelled continuation value ${}_{\kappa}\bar{Z}_j$ in (2.5), that is, the *test value* used to decide whether we exercise at node S_j^i.

  Note that the dependence of these vectors on the time index j is suppressed because we will be overwriting their entries at every time step.

- At maturity t_M we initialise EE:

  For $i = 1$ **To** N

  Set $EE^i = I^i$.

  Next i

- Now we step backwards in time:

  For $j = M - 1$ **To** 1 **Step** -1

  - Set $EE^i := e^{-r\Delta t_{j+1}} EE^i$.

[5] Here and in the rest of this thesis, we will refer to the value of a European option as a *European value*.

– Calculate the European option value E^i [6]. Note that this step is not included in the original algorithm. However, when the exercise decision is made, the intrinsic value should not only be greater than the modelled continuation value, but also greater than the European value.

– Use least squares regression to fit the basis functions to EE^i so that we can find β. Then calculate TV^i as in (2.5). Thus we determine ${}_0\bar{\beta}_j,\ {}_1\bar{\beta}_j,\ \ldots,\ {}_\kappa\bar{\beta}_j$ in order to realise

$$\min_{{}_0\bar{\beta}_j,\ {}_1\bar{\beta}_j,\ \ldots,\ {}_\kappa\bar{\beta}_j} \sum_{\substack{i=1 \\ I_j^i>0}}^{N} \left| EE^i - \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S_j^i\right) \right|^2 .$$

Here $I_j^i > 0$ indicates of course that the summation is only taken over in-the-money paths. $\beta := \left[{}_0\bar{\beta}_j,\ {}_1\bar{\beta}_j,\ \ldots,\ {}_\kappa\bar{\beta}_j\right]'$ is found as the regression solution to what are referred to as the *normal equations*

$$X'X\beta = X'Y \tag{2.11}$$

where $[x_{i,k}] = L_k\left(S_j^i\right)$ and $[y_i] = EE^i$. Here $X'X$ is a $(\kappa+1)\times(\kappa+1)$ matrix and $X'Y$ is a $(\kappa+1)\times 1$ vector with entries

$$[X'X]_{kk'} = \sum_{\substack{i=1 \\ I_j^i>0}}^{N} x_{i,k}x_{i,k'} \text{ and } [X'Y]_k = \sum_{\substack{i=1 \\ I_j^i>0}}^{N} x_{i,k}y_i$$

respectively[7]. In order to determine β, Longstaff and Schwartz [2001] use the double precision DLSBRR algorithm in IMSL. The approach we will follow is to use singular value decomposition (SVD) which we discuss in Appendix A.

– We then set $TV^i := \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S_j^i\right)$.

– **If** $I^i > \max\left\{E^i, TV^i\right\}$ **Then**

Set $EE^i := I^i$.

End If

Next j

[6] If the underlying stock price follows geometric Brownian motion and we are considering a vanilla payoff, one can use the Black-Scholes formula. When the underlying stock price follows one of the Lévy models we will consider later, one could make use of the *COS method* (see Appendix E.1 for a short discussion on this method).

[7] The normal equations in (2.11) are derived as follows: let time t_j be fixed, then we solve for β such that the sum of the squared differences is minimised

$$\sum_{i=1}^{N}\left(y_i - \sum_{k=1}^{\kappa} {}_k\beta\, x_{i,k}\right)^2 .$$

This is achieved by differentiating the above with respect to ${}_l\beta$, $0 \le l \le \kappa$ and setting the result equal to 0

$$\sum_{i=1}^{N}\left(y_i - \sum_{k=1}^{\kappa} {}_k\beta\, x_{i,k}\right)x_{i,l} = 0 \Rightarrow \sum_{i=1}^{N} y_i x_{i,l} = \sum_{k=1}^{\kappa} x_{i,l}x_{i,k}\ {}_k\beta$$

and so we have that $[X'Y]_l = [X'X\beta]_l$ implying that $X'Y = X'X\beta$.

- The option value as determined by the LSM method at time t_0 is then given by

$$\max\left\{I_0, \frac{e^{-rt_1}}{N}\sum_{i=1}^{N} EE^i\right\}.$$

Longstaff and Schwartz [2001] note that to prevent a computational underflow they normalise all cashflows and stock prices by the strike prices. This is also implemented in Moreno and Navas [2004] with double precision variables. In Figure 2.4 we plot the performance of the LSM method where we have, and have not, normalised the spot. In this example we see that by normalising the spot the results produced are far better than otherwise. We found many examples where by not normalising the spot performed just as well as by normalising, but it was never worse. Including the normalisation is hardly complicated and does not add to the computation time. Hence we normalise the spot whenever we apply the LSM method.

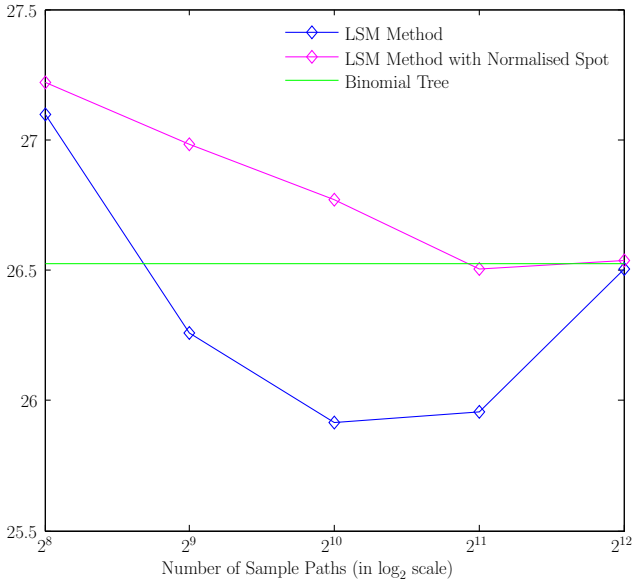

Figure 2.4: *The performance of the LSM method where we have, and have not, normalised the spot as a function of the number of sample paths and 10 Laguerre polynomials.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 28\%$ *and* $S_0 = 110$, $r = 6\%$ *and* $q = 1\%$*; and we are considering a 1 year put with* $K = 135$. *The sample paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5. We also include the value given by a 30 step binomial tree.*

In this example we see that by normalising the spot, better results are obtained than by not normalising.

Chapter 3

The Variance Reduction Techniques Suggested by Rasmussen [2005]

Variance reduction techniques are employed to reduce the variance in the estimate obtained from Monte Carlo simulation. Thus these techniques increase the efficiency of Monte Carlo simulation. There are several such techniques such as antithetic variates, stratified sampling, importance sampling and control variates (see for example Glasserman [2004, Chapter 4] or Jäckel [2002, Chapter 10]).

As discussed in §B.3 we will not consider antithetic variates in our implementation. Rasmussen [2005] proposes an improvement over the standard method of using control variates which we discuss in §3.1. Furthermore, [Rasmussen, 2005, §6] proposes an 'initial dispersion' method which replaces the importance sampling and stratified sampling techniques. This method is discussed in §3.2. Thus we discuss the improved control variate and the initial dispersion methods. In addition to Rasmussen [2005], we will also refer to Rasmussen [2002] in the following sections.

In this chapter we will continue to use the notation as introduced in Chapter 1.

3.1 Least Squares Monte Carlo Control Variates

Traditionally (see Glasserman [2004, §4.1.1] for various examples on employing control variate techniques on exotic options), when applying the control variate technique, one would perform the complete Monte Carlo simulation to obtain the price of the American option, say A_{MC} .

One then uses the same random numbers to determine the European option value, E_{MC}, and finally make use of the closed-form formula for the European option, E_{CF}. The value of the option is then found by using the control variate technique

$$\bar{V} = A_{\text{MC}} + \theta \left(E_{\text{MC}} - E_{\text{CF}} \right).$$

Observe that when E_{MC} is unbiased, it follows that $\mathbb{E}\left[\bar{V}\right] = \mathbb{E}\left[A_{\text{MC}}\right]$ for any value of θ. It is easily shown (see Glasserman [2004, §4.1.1] or Jäckel [2002, §10.3] for example) that the value of θ, that minimises the variance, is equal to minus the regression coefficient. Here the American estimate is the dependent

variable and the European estimate is the independent variable i.e.

$$\theta = -\frac{\mathbb{C}\mathrm{ov}\left[A_{\mathrm{MC}}, E_{\mathrm{MC}}\right]}{\mathbb{V}\mathrm{ar}\left[E_{\mathrm{MC}}\right]} = -\frac{\mathbb{E}\left[A_{\mathrm{MC}}E_{\mathrm{MC}}\right] - \mathbb{E}\left[A_{\mathrm{MC}}\right]\mathbb{E}\left[E_{\mathrm{MC}}\right]}{\mathbb{E}\left[E_{\mathrm{MC}}^2\right] - \left(\mathbb{E}\left[E_{\mathrm{MC}}\right]\right)^2}. \tag{3.1}$$

θ is almost always unknown, but it is typical to approximate it using the same simulation that determined $\bar{V}$, as the bias it introduces is deemed to be immaterial.

However, Rasmussen [2005] proposes a more sophisticated control variate technique when pricing American options using the LSM method. Here the control variate is applied to every simulation path individually, at its exercise time. Rasmussen [2005] suggests sampling the discounted value process at the time of exercise of the option, instead of sampling the discounted payoff process at the option's terminal time. Thus, if we let $E = \{E_j\}_{t=t_j}$ indicate the European option price process, then the model of the continuation value given in (2.5) becomes

$$_{\kappa}\bar{Z}_j\left(S_j\right) := \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S_j\right) + \theta_j\left(e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}} - E^i_j\right) \tag{3.2}$$

where, as before, $\bar{\tau}$ indicates the stopping time at which the American option has been exercised as determined by the LSM method (see §2.1 for notation). θ_j will be defined in due course. Since $\bar{\tau}_{j+1}$ is a bounded stopping time it follows from the Optional Sampling Theorem that the expectation of $e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}} - E^i_j$ is 0.

The motivation for this approach becomes apparent when considering the following result, which is a slight modification of [Rasmussen, 2005, Theorem 1].

Theorem 3.1.1

Let $\tau_1, \tau_2 \in \mathcal{T}_{j,M}$ be two stopping times such that $\tau_1 \leq \tau_2$. Suppose that $X = \{X_j\}_{t=t_j}$ is an adaptive process and $Y = \{Y_j\}_{t=t_j}$ is a martingale process. Then

$$\mathbb{C}orr\left[X_{\tau_1},\, Y_{\tau_1} \middle|\, \mathcal{F}_j\right] \geq \mathbb{C}orr\left[X_{\tau_1},\, Y_{\tau_2} \middle|\, \mathcal{F}_j\right].$$

Proof.

Assume, without loss of generality, that $Y_0 = 0$. Then we have that

$$\begin{aligned}
\mathbb{C}\mathrm{ov}\left[X_{\tau_1}, Y_{\tau_2} \middle|\, \mathcal{F}_j\right] &= \mathbb{E}\left[X_{\tau_1}Y_{\tau_2} \middle|\, \mathcal{F}_j\right] \\
&= \mathbb{E}\left[X_{\tau_1}\mathbb{E}\left[Y_{\tau_2} \middle|\, \mathcal{F}_{\tau_1}\right] \middle|\, \mathcal{F}_j\right] \\
&= \mathbb{E}\left[X_{\tau_1}Y_{\tau_1} \middle|\, \mathcal{F}_j\right] \\
&= \mathbb{C}\mathrm{ov}\left[X_{\tau_1}, Y_{\tau_1} \middle|\, \mathcal{F}_j\right]
\end{aligned}$$

where the second equality follows from the Tower Property of expectations and the third from the Optional Sampling theorem. Furthermore Y^2 is a submartingale, because Y is a martingale. That is $\mathbb{E}\left[Y^2_{\tau_1} \middle|\, \mathcal{F}_j\right] \leq \mathbb{E}\left[Y^2_{\tau_2} \middle|\, \mathcal{F}_j\right]$ and so $\mathbb{V}\mathrm{ar}\left[Y_{\tau_1} \middle|\, \mathcal{F}_j\right] \leq \mathbb{V}\mathrm{ar}\left[Y_{\tau_2} \middle|\, \mathcal{F}_j\right]$. Therefore

$$\mathbb{C}\mathrm{orr}\left[X_{\tau_1}, Y_{\tau_2} \middle|\, \mathcal{F}_j\right] = \frac{\mathbb{C}\mathrm{ov}\left[X_{\tau_1}, Y_{\tau_2} \middle|\, \mathcal{F}_j\right]}{\sqrt{\mathbb{V}\mathrm{ar}\left[X_{\tau_1} \middle|\, \mathcal{F}_j\right]\mathbb{V}\mathrm{ar}\left[Y_{\tau_2} \middle|\, \mathcal{F}_j\right]}} \leq \frac{\mathbb{C}\mathrm{ov}\left[X_{\tau_1}, Y_{\tau_1} \middle|\, \mathcal{F}_j\right]}{\sqrt{\mathbb{V}\mathrm{ar}\left[X_{\tau_1} \middle|\, \mathcal{F}_j\right]\mathbb{V}\mathrm{ar}\left[Y_{\tau_1} \middle|\, \mathcal{F}_j\right]}} = \mathbb{C}\mathrm{orr}\left[X_{\tau_1}, Y_{\tau_1} \middle|\, \mathcal{F}_j\right].$$

□

A good control variate is one that is highly correlated with the variable which one would like to estimate. Now set X equal to the discounted intrinsic value process $I = \{I_j\}_{t=t_j}$ and Y equal to the chosen control variate. Then the theorem above shows that applying the control variate at the stopping time τ at which the option is exercised instead of at maturity of the option, yields a higher correlation and hence a better control variate.

3.1.1 θ using Simple Linear Regression

In this section we consider the approximation of θ as suggested by Rasmussen [2002, Algorithm 2]. Here θ is found by performing simple linear regression between the eventual exercise $e^{-r(\bar{\tau}_{j+1}-t_j)}I^i_{\bar{\tau}_{j+1}}$ and the European eventual exercise $e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}$. Thus European eventual exercise means the value of the corresponding European option at the exercise time $\bar{\tau}^i_{j+1}$ of the American option, discounted to time t_j. Using (3.1) we set

$$\bar{\theta} := -\frac{\sum_{i=1}^N A^i_{\text{MC}}E^i_{\text{MC}} - \sum_{i=1}^N A^i_{\text{MC}}\sum_{i=1}^N E^i_{\text{MC}}}{\sum_{i=1}^N \left(E^i_{\text{MC}}\right)^2 - \left(\sum_{i=1}^N E^i_{\text{MC}}\right)^2}.$$

Calculating $\bar{\theta}$ using the same simulation as that used to calculate $\bar{V}$ introduces a bias. This may be overcome by running a separate simulation (much smaller than the main simulation used to calculate $\bar{V}$) to calculate $\bar{\theta}$. However the magnitude of this bias is negligible in most cases [see Glasserman, 2004, §4.1.3]. Hence as suggested by Rasmussen [2005] we use the same set of paths to determine $\bar{\theta}$ as $\bar{V}$ in favour of efficiency.

We now provide a modified version of the LSM algorithm in §2.4. Here changes to the original algorithm, that is, the inclusion of the control variate discussed where the calculation of $\bar{\theta}$ is as above, are in red.

- Suppose that we are given the stock price matrix S^i_j and intrinsic value matrix I^i_j. We now define several vectors of length N at time point t_j for $j = M,\ M-1,\ \ldots,\ 1$ which we will use in the algorithm. Let
  - E indicate the vector where the entry E^i is the *European value* whose term is given by $t_M - t_j$ and spot by S^i_j. The strike and style (call or put) of this European option is equal to that of the American option under consideration;
  - EE denote the vector with entry EE^i equal to $e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}I^i_{\bar{\tau}^i_{j+1}}$, that is, the discounted *eventual exercise* value of path i;
  - TV indicate the vector where the entry TV^i is the modelled continuation value ${}_\kappa\bar{Z}_j$ in (2.5), that is, the *test value* used to decide whether we exercise at node S^i_j;
  - EEE denote the vector with entry EEE^i equal to the *European eventual exercise*, that is, (as mentioned above) the value of the European option at exercise time $\bar{\tau}^i_{j+1}$ discounted to t_j for path i or $e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}$.

  Note that the dependence of these vectors on the time index j is suppressed because we will be overwriting their entries at every time step.

- At maturity t_M we initialise EE and EEE:

  For $i = 1$ **To** N

  Set $EE^i = I^i$ and $EEE^i = I^i$.

  Next i

- Now we step backwards in time:

  For $j = M - 1$ **To** 1 **Step** -1

  – Set $EE^i := e^{-r\Delta t_{j+1}} EE^i$ and $EEE := e^{-r\Delta t_{j+1}} EEE^i$.

  – Calculate the European option value E^i .

  – Use least squares regression to fit the basis functions to EE^i so that we can find β. Then calculate TV^i as in (2.5). Thus we determine ${}_0\bar{\beta}_j,\ {}_1\bar{\beta}_j,\ \ldots,\ {}_\kappa\bar{\beta}_j$ in order to realise

  $$\min_{{}_0\bar{\beta}_j,\ {}_1\bar{\beta}_j,\ \ldots,\ {}_\kappa\bar{\beta}_j} \sum_{\substack{i=1 \\ I_j^i>0}}^{N} \left| EE^i - \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S_j^i\right) \right|^2 .$$

  Here $I_j^i > 0$ indicates of course that the summation is only taken over in-the-money paths. $\beta := \left[{}_0\bar{\beta}_j,\ {}_1\bar{\beta}_j,\ \ldots,\ {}_\kappa\bar{\beta}_j\right]'$ is found as the regression solution to

  $$X'X\beta = X'Y$$

  where $[x_{i,k}] = L_k\left(S_j^i\right)$ and $[y_i] = EE^i$. Here $X'X$ is a $(\kappa+1) \times (\kappa+1)$ matrix and $X'Y$ is a $(\kappa+1) \times 1$ vector with entries

  $$[X'X]_{kk'} = \sum_{\substack{i=1 \\ I_j^i>0}}^{N} x_{i,k}x_{i,k'} \text{ and } [X'Y]_k = \sum_{\substack{i=1 \\ I_j^i>0}}^{N} x_{i,k}y_i$$

  respectively.

  – Calculate $\bar{\theta}$ as

  $$\bar{\theta} := -\frac{N\sum_{i=1}^{N} EE^i\ EEE^i - \sum_{i=1}^{N} EE^i \sum_{i=1}^{N} EEE^i}{N\sum_{i=1}^{N}\left(EEE^i\right)^2 - \left(\sum_{i=1}^{N} EEE^i\right)^2}. \tag{3.3}$$

  Thus $\bar{\theta}$ is obtained by performing simple linear regression between EE and EEE.

  – We then set $TV^i := \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S_j^i\right) + \bar{\theta}\left(EEE^i - E^i\right)$.

  – **If** $I^i > \max\left\{E^i, TV^i\right\}$ **Then**

  Set $EE^i := I^i$ and $EEE^i := E^i$.

  End If

  Next j

- The option value as determined by the LSM method combined with the control variate suggested by Rasmussen [2005] at time t_0 is then given by

$$\max\left\{I_0, \frac{e^{-rt_1}}{N}\sum_{i=1}^{N} EE^i\right\}.$$

This algorithm is demonstrated by Rasmussen [2002] using three time steps. However, we found when implementing it for a larger number of time steps, results produced by this method were worse than those produced by the original LSM method. Probably this is the reason why this algorithm does not appear in Rasmussen [2005]. Nevertheless, part of this algorithm does have a favourable application in §3.1.2.

3.1.2 θ as a Functional Form

In §3.1.1, θ is calculated as a number, however Rasmussen proposes an improvement where θ is a function of spot (Rasmussen [2002, Algorithm 3], see also Rasmussen [2005, §5]).

Recall from §2.4 that at a given time step t_j, $j = 1,\ 2,\ \ldots,\ M-1$, β is found as the regression solution in (2.4) to normal equations given in (2.11)

$$X'X\beta = X'Y.$$

Here Y is a column vector of length N containing the discounted eventual exercise, that is $e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}I_{\bar{\tau}^i_{j+1}}$, X is a matrix of size $N \times (\kappa+1)$ with $[x_{i,k}] = L_k\left(S^i_j\right)$ and $\beta = \left[{}_0\bar{\beta}_j,\ {}_1\bar{\beta}_j,\ \ldots,\ {}_\kappa\bar{\beta}_j\right]'$ is a column vector of length $\kappa+1$ where the model of continuation value is given by

$${}_\kappa\bar{Z}_j\left(S_j\right) := \sum_{k=0}^{\kappa} {}_k\bar{\beta}_j L_k\left(S_j\right)$$

as we have seen in §2.1.2.

Now in order to approximate the θ in (3.2), Rasmussen [2005] proposes that it is calculated as

$$\begin{aligned}\bar{\theta}^i_j &= -\frac{\mathbb{C}\mathrm{ov}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}I^i_{\bar{\tau}^i_{j+1}}, e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right]}{\mathbb{V}\mathrm{ar}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right]}\\ &= -\frac{\mathbb{E}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}I^i_{\bar{\tau}^i_{j+1}}e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right] - \mathbb{E}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}I^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right]\mathbb{E}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right]}{\mathbb{E}\left[\left(e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\right)^2\middle|\mathcal{F}_j\right] - \left(\mathbb{E}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right]\right)^2}\end{aligned}$$

where all expectations above are with respect to the risk-neutral measure. Note that $\bar{\theta}^i_j$ is not a number, but a functional form, i.e. $\bar{\theta}^i_j = \theta\left(S^i_j\right)$ with

$$\mathbb{E}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}I^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right],\ \mathbb{E}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right],$$

$$\mathbb{E}\left[e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}I^i_{\bar{\tau}^i_{j+1}}e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\middle|\mathcal{F}_j\right] \text{ and } \mathbb{E}\left[\left(e^{-r\left(\bar{\tau}^i_{j+1}-t_j\right)}E^i_{\bar{\tau}^i_{j+1}}\right)^2\middle|\mathcal{F}_j\right]$$

approximated using basis functions L_k in the same way as ${}_\kappa\bar{Z}_j$ is approximated. θ_j^i is then found as part of the solution to the regression equation. This is achieved by replacing the vector Y in (2.11) with a matrix of size $N \times 4$ where the columns of the matrix are given by

$$e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} I_{\bar{\tau}_{j+1}^i}^i,\ e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} E_{\bar{\tau}_{j+1}^i}^i,\ e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} I_{\bar{\tau}_{j+1}^i}^i e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} E_{\bar{\tau}_{j+1}^i}^i \text{ and } \left(e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} E_{\bar{\tau}_{j+1}^i}^i\right)^2.$$

Note that the first column is exactly the column vector Y used in the regression without a control variate. As before, β is then found using SVD, but this time as a matrix of size $(\kappa + 1) \times 4$ where each column contains the coefficients of the approximation of

$$\mathbb{E}\left[e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} I_{\bar{\tau}_{j+1}^i}^i \middle| \mathcal{F}_j\right],\ \mathbb{E}\left[e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} E_{\bar{\tau}_{j+1}^i}^i \middle| \mathcal{F}_j\right],$$
$$\mathbb{E}\left[e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} I_{\bar{\tau}_{j+1}^i}^i e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} E_{\bar{\tau}_{j+1}^i}^i \middle| \mathcal{F}_j\right] \text{ and } \mathbb{E}\left[\left(e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} E_{\bar{\tau}_{j+1}^i}^i\right)^2 \middle| \mathcal{F}_j\right]$$

respectively.

We provide a modified version of the algorithm given in §3.1.1 with the changed calculation of $\bar{\theta}^i$ indicated in green.

- Suppose that we are given the stock price matrix S_j^i and intrinsic value matrix I_j^i. We now define several vectors of length N at time point t_j for $j = M,\ M-1,\ \ldots,\ 1$ which we will use in the algorithm. Let
  - E indicate the vector where the entry E^i is the *European value* whose term is given by $t_M - t_j$ and spot by S_j^i. The strike and style (call or put) of this European option is equal to that of the American option under consideration;
  - EE denote the vector with entry EE^i equal to $e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} I_{\bar{\tau}_{j+1}^i}^i$, that is, the discounted *eventual exercise* value of path i;
  - TV indicate the vector where the entry TV^i is the modelled continuation value ${}_\kappa\bar{Z}_j$ in (2.5), that is, the *test value* used to decide whether we exercise at node S_j^i;
  - EEE denote the vector with entry EEE^i equal to the *European eventual exercise*, that is, the value of the European option at exercise time $\bar{\tau}_{j+1}^i$ discounted to t_j for path i or $e^{-r\left(\bar{\tau}_{j+1}^i - t_j\right)} E_{\bar{\tau}_{j+1}^i}^i$.

  Note that the dependence of these vectors on the time index j is suppressed because we will be overwriting their entries at every time step.

- At maturity t_M we initialise EE and EEE:

  For $i = 1$ **To** N

  Set $EE^i = I^i$ and $EEE^i = I^i$.

  Next i

- Now we step backwards in time:

  For $j = M - 1$ **To** 1 **Step** -1

  - Set $EE^i := e^{-r\Delta t_{j+1}} EE^i$ and $EEE := e^{-r\Delta t_{j+1}} EEE^i$.

– Calculate the European option value E^i .

– Use least squares regression to fit the basis functions to $Y^i = \left[EE^i \; EEE^i \; EE^i EEE^i \; \left(EEE^i\right)^2\right]$ so that we can find β. Then calculate TV^i as in (2.5) and $\bar{\theta}^i$ which we show below. Thus we determine ${}^{l}_{0}\bar{\beta}_j, \; {}^{l}_{1}\bar{\beta}_j, \; \ldots, \; {}^{l}_{\kappa}\bar{\beta}_j$ for $l = 1, \; 2, \; \ldots, \; 4$ in order to realise

$$\min_{{}^{l}_{0}\bar{\beta}_j, \; {}^{l}_{1}\bar{\beta}_j, \; \ldots, \; {}^{l}_{\kappa}\bar{\beta}_j} \sum_{\substack{i=1 \\ I^i_j>0}}^{N} \left| y_{i,l} - \sum_{k=0}^{\kappa} {}^{l}_{k}\bar{\beta}_j L_k\left(S^i_j\right) \right|^2 .$$

Here $I^i_j > 0$ indicates of course that the summation is only taken over in-the-money paths. $\beta := \left[{}^{l}_{0}\bar{\beta}_j, \; {}^{l}_{1}\bar{\beta}_j, \; \ldots, \; {}^{l}_{\kappa}\bar{\beta}_j\right]'$ is found as the regression solution to

$$X'X\beta = X'Y$$

where $[x_{i,k}] = L_k\left(S^i_j\right)$. Here $X'X$ is a $(\kappa+1)\times(\kappa+1)$ matrix and $X'Y$ is a $(\kappa+1)\times 4$ matrix with entries

$$[X'X]_{kk'} = \sum_{\substack{i=1 \\ I^i_j>0}}^{N} x_{i,k}x_{i,k'} \text{ and } [X'Y]_{k,l} = \sum_{\substack{i=1 \\ I^i_j>0}}^{N} x_{i,k}y_{i,l}$$

for $l = 1, \; 2, \; 3, \; 4$ respectively.
Thus $\sum_{k=0}^{\kappa} {}^{1}_{k}\bar{\beta}_j L_k\left(S^i_j\right)$ approximates EE^i, $\sum_{k=0}^{\kappa} {}^{2}_{k}\bar{\beta}_j L_k\left(S^i_j\right)$ approximates EEE^i, $\sum_{k=0}^{\kappa} {}^{3}_{k}\bar{\beta}_j L_k\left(S^i_j\right)$ approximates EE^iEEE^i and $\sum_{k=0}^{\kappa} {}^{4}_{k}\bar{\beta}_j L_k\left(S^i_j\right)$ approximates $\left(EEE^i\right)^2$.

– Calculate $\bar{\theta}^i$ as

$$\bar{\theta}^i := -\frac{\sum_{k=0}^{\kappa} {}^{3}_{k}\bar{\beta}_j L_k\left(S^i_j\right) - \sum_{k=0}^{\kappa} {}^{1}_{k}\bar{\beta}_j L_k\left(S^i_j\right) \sum_{k=0}^{\kappa} {}^{2}_{k}\bar{\beta}_j L_k\left(S^i_j\right)}{\sum_{k=0}^{\kappa} {}^{4}_{k}\bar{\beta}_j L_k\left(S^i_j\right) - \left(\sum_{k=0}^{\kappa} {}^{2}_{k}\bar{\beta}_j L_k\left(S^i_j\right)\right)^2}.$$

– We then set $TV^i := \sum_{k=0}^{\kappa} {}^{1}_{k}\bar{\beta}_j L_k\left(S^i_j\right) + \bar{\theta}^i\left(\sum_{k=0}^{\kappa} {}^{2}_{k}\bar{\beta}_j L_k\left(S^i_j\right) - E^i\right)$.

– **If** $I^i > \max\left\{E^i, TV^i\right\}$ **Then**
Set $EE^i := I^i$ and $EEE^i := E^i$.
End If

Next j

- The option value as determined by the LSM method combined with the control variate suggested by Rasmussen [2005] at time t_0 is then given by

$$\max\left\{I_0, \frac{e^{-rt_1}}{N}\sum_{i=1}^{N} EE^i\right\}.$$

A different version is given in Rasmussen [2002]. Here at time t_0 we calculate $\bar{\theta}$ as in (3.3) using the vectors EE and EEE obtained at time t_1. The model option value is then given by

$$\max\left\{I_0, \frac{e^{-rt_1}}{N}\sum_{i=1}^{N} EE^i + \bar{\theta}\left(\frac{e^{-rt_1}}{N}\sum_{i=1}^{N} EEE^i - E_0\right)\right\}.$$

We found that by including this technique, a remarkable improvement in the pricing performance was observed. Hence we believe that the omission of this step in Rasmussen [2005] is anomalous.

Time step t_{M-1} is treated differently in [Rasmussen, 2002, Algorithm 3]. Here, instead of discounting $EE^i_M = I^i_M$ back one step to t_{M-1}, EE^i_{M-1} is set to be the maximum of I^i_{M-1} and E^i_{M-1}. The reason for this is that because of the discretisation of the American option, there are no exercise opportunities between time step t_{M-1} and time t_M and thus the value of the option is either intrinsic or the value of the European option that terminates at time t_M. This special treatment is actually unnecessary and coincides with the given algorithm. To see this note that the approximations for EE^i_{M-1} and EEE^i_{M-1} are equal for all i, that is

$$\sum_{k=0}^{\kappa} {}^1_k\bar{\beta}_{M-1} L_k\left(S^i_{M-1}\right) = \sum_{k=0}^{\kappa} {}^2_k\bar{\beta}_{M-1} L_k\left(S^i_{M-1}\right)$$

and hence

$$\sum_{k=0}^{\kappa} {}^3_k\bar{\beta}_{M-1} L_k\left(S^i_{M-1}\right) = \sum_{k=0}^{\kappa} {}^4_k\bar{\beta}_{M-1} L_k\left(S^i_{M-1}\right).$$

Therefore $\bar{\theta}^i_{M-1} = -1$ for every i and so $TV^i_{M-1} = E^i_{M-1}$.

We call the model's estimate of the optimal stopping boundary the *critical stock price function*. We would hope that the critical stock price for a particular time step occurs at the point where $TV - I$ is equal to 0. Because of its functional form θ is a rational function, and so has vertical asymptotes. Therefore $TV - I$ also has asymptotes as we can see in Figure 3.1, and so potentially has several zeros. This prevents us from obtaining the critical stock price at a specific time by simple zero-search methods. A method for finding the critical stock price at every time step, independent of the Monte Carlo method, is discussed later in §5.1.

Whenever we refer to the method discussed in this section, we will call it the *LSM-Rasmussen method*.

3.2 Least Squares Monte Carlo with Initial Dispersion

Recall that when approximating the optimal stopping boundary using the LSM method, we may choose to only make use of in-the-money paths. Furthermore as noted in Rasmussen [2005, §6], given the variance of $e^{-r(\bar{\tau}-t_j)} I_{\bar{\tau}}$, the accuracy of ${}_\kappa\bar{Z}_j\left(S_j\right)$ in (2.5) can only increase if we use more paths in the least squares regression. These facts suggest that increasing the number of in-the-money paths may improve the approximation found using the LSM method. Rasmussen [2005, §6] makes this observation and provides a method which generates enough in-the-money paths for all possible exercise points, in particular for the longest expiry, to estimate the optimal stopping boundary.

Of course the exercise boundary is independent of the current level of the stock price.

Create an artificial initial time point $t_{-1} < t_0$ and generate paths from this time point. This point is chosen so that the risk-neutral drift of S_t is given by $r - q$, implying that $S_t e^{-(r-q)t}$ is a martingale. Thus we have that $S_0 = S_{-1} e^{(r-q)\Delta t_0}$ and hence, instead of generating paths from S_0, we generate paths from

$$S_{-1} = S_0 e^{-(r-q)\Delta t_0}$$

which obviously is model independent. Since the discounted stock price process is a martingale under the risk-neutral measure $\mathbb{Q}$, the distribution at time t_0 will reflect the volatility of the stock price process while

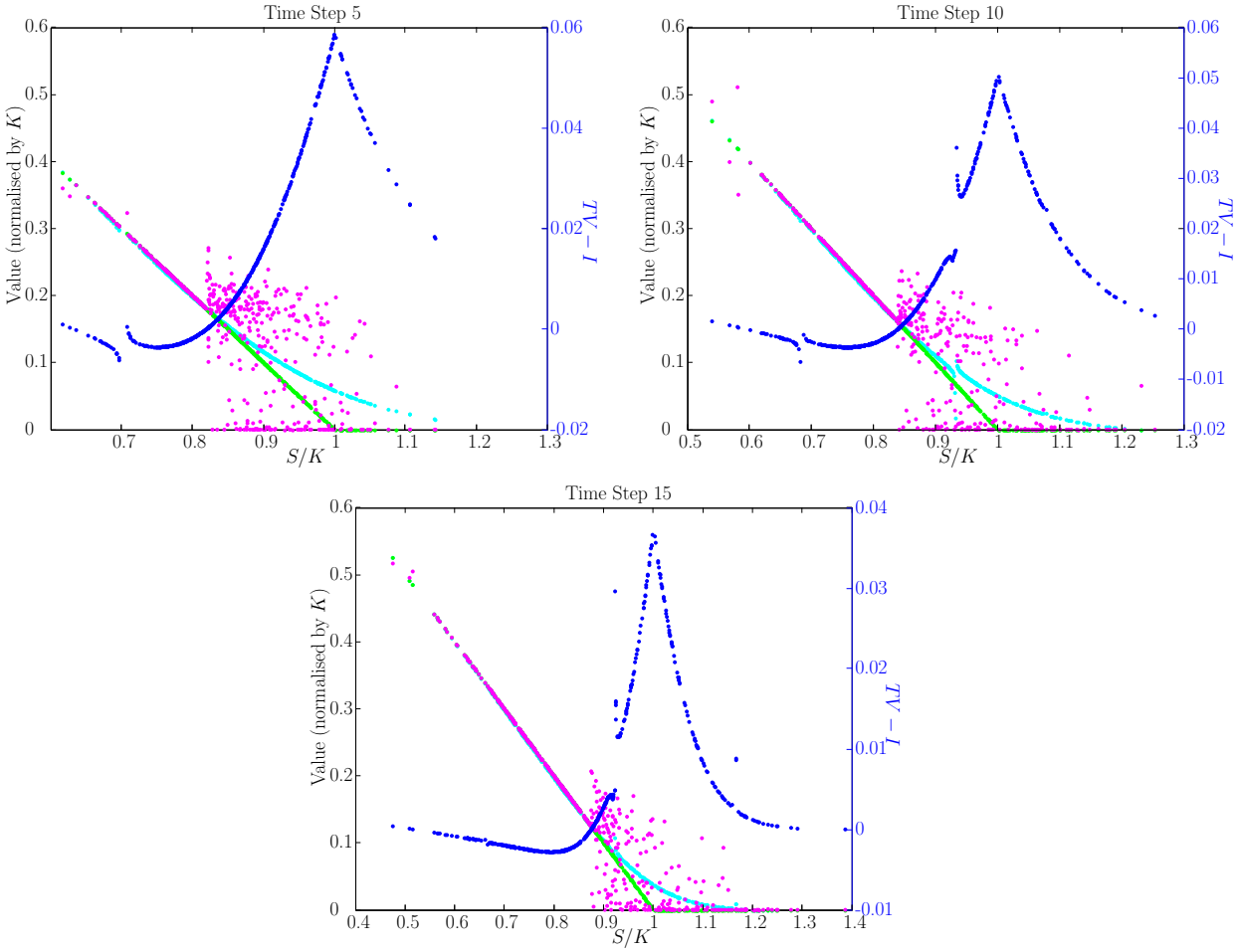

Figure 3.1: *We plot the test value (light blue), intrinsic (green), discounted eventual exercise (pink) and difference between the test value and intrinsic (dark blue) against the normalised stock price at time step 5, 10 and 15 of a 20 step method. All values are plotted against the left vertical axis, except for $TV - I$ which is plotted against the right vertical axis in dark blue.*

The option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 20\%$, $S_0 = 110$, $r = 5\%$ and $q = 2\%$; and we are considering a 1 year put option with $K = 130$. Here the stock prices have been normalised by the strike. We simulated 512 paths using simulation techniques we discuss in §7.4 and §7.5.

As mentioned, note the vertical asymptotes of $TV - I$. Similar results were obtained for an out-the-money put with $K = 90$. However even though there were asymptotes present, they did not seem as problematic as in the case of the in-the-money put.

centered around the spot price for which the option value is sought. By inspecting plots of the exercise boundary obtained under geometric Brownian motion, Rasmussen [2005] notes that irregularities of the exercise boundaries occur when the time to expiry is less than half the maturity. Hence Rasmussen [2005] chooses this initial time point to be $t_{-1} := \frac{t_M}{2}$. We found this to be adequate when the underlying follows

geometric Brownian motion. However, as we shall see later, in the case where the underlying follows an exponential variance gamma or normal inverse Gaussian process, t_{-1} needs to be adjusted.

When applying the importance sampling technique, the distribution of the stock price is shifted so that it covers the region of interest. However when considering American options the region of importance is around the exercise boundary, which is unknown. As noted by Rasmussen [2005, §6] the method proposed above serves as a replacement for importance sampling, in the sense that the distribution of the stock price will include paths containing critical points around the exercise boundary.

One should note that this method of initial dispersion does not produce an approximation for the American option value, but an approximation of the optimal stopping boundary. Thus one would first use initial dispersion combined with the LSM method to determine the critical stock price function, and then run another Monte Carlo simulation to calculate the value of the option using the critical stock price function determined in the first.

3.3 Results

The improvement to the LSM method by introducing the control variate is remarkable. Consider Figure 3.2 where we have plotted the performance of the LSM and LSM-Rasmussen methods for various number of Laguerre polynomials. Note that the convergence of the LSM-Rasmussen method is achieved using very few polynomials.

Rasmussen [2005] suggests using basis functions which require the strike, stock price, European option price[1] and some combination of these, in particular Rasmussen [2005] uses K, S_j, $v(S_j; K, T)$ and $S_j v(S_j; K, T)$. However, we found that the choice of basis functions hardly had any effect on the results we obtained.

Figure 3.3 illustrates again how much better the LSM-Rasmussen method performs compared to the LSM method. Here we applied the two methods for a varying number of sample paths using the same inputs as in Figure 2.3. In Table 3.1, we compare values (for in-the-money, at-the-money and out-the-money options) and the time taken (in seconds) of these methods. We used the same inputs as in Table 2.1, except for the number of sample paths, which was decreased.

[1] The European option considered here has the same parameters as the American option we are pricing, but with a term equal to $T - t_j$ and spot price equal to S_j.

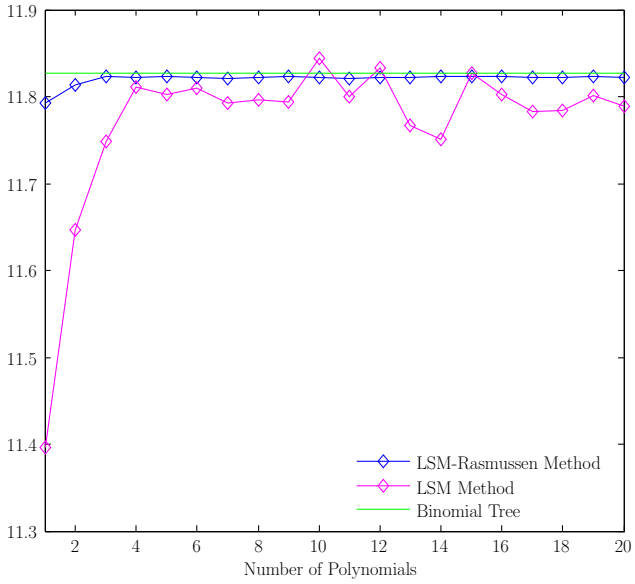

Figure 3.2: *The performance of the LSM and LSM-Rasmussen methods for various number of Laguerre polynomials.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put with* $K = 135$. *We used 4096 sample paths with 20 time steps using simulation techniques we discuss in §7.4 and §7.5. We also plot the approximation of a 20 step binomial tree.*

In this example we see how much better the LSM-Rasmussen method performs compared to the LSM method. Similar results were obtained when varying volatility, term, strike, number of time steps and number of sample paths. Compared to the LSM method, the LSM-Rasmussen method performed particularly well for far out-the-money options.

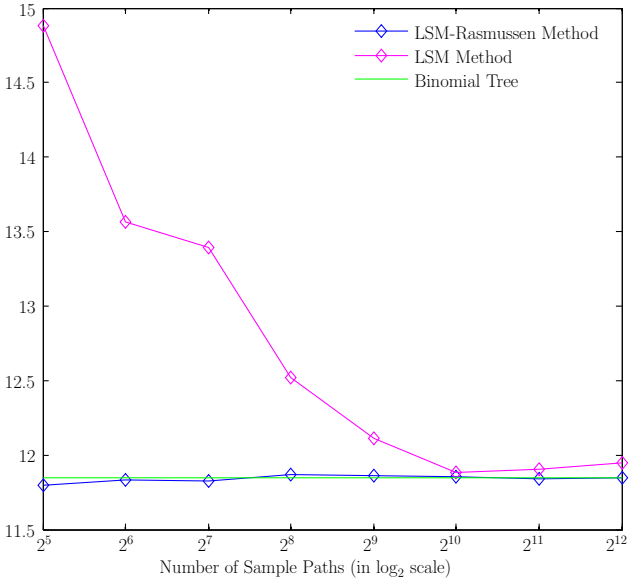

Figure 3.3: *The performance of the LSM and LSM-Rasmussen methods as a function of the number sample paths with 8 Laguerre polynomials.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put with* $K = 135$. *We simulated paths with 30 time steps using simulation techniques we discuss in §7.4 and §7.5. We also include the value given by a 30 step binomial tree.*

Again we see that the LSM-Rasmussen method performs much better than the ordinary LSM method.

	Value		
K	Binomial Tree	LSM	LSM-Rasmussen
85	0.4391	0.4549	0.4364
130	9.6090	10.1323	9.6218
135	11.8490	12.5220	11.8671
140	14.3803	15.0654	14.3891
185	50.0000	50.8100	50.0000
	Time (seconds)		
K	Binomial Tree	LSM	LSM-Rasmussen
85	0.345	6.004	5.331
130	0.326	4.468	3.941
135	0.234	4.657	4.058
140	0.265	5.112	4.487
185	0.275	5.455	4.849

Table 3.1: *The performance of the LSM and LSM-Rasmussen methods for 256 sample paths with 8 Laguerre polynomials and varied strikes.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put. We simulated paths with 30 time steps using simulation techniques we discuss in §7.4 and §7.5. We also include the value given by a 30 step binomial tree.*

Again we see that the LSM-Rasmussen method performs much better than the ordinary LSM method.

Chapter 4

The Stochastic Mesh Method

The stochastic mesh method was originally introduced in Broadie and Glasserman [1997a] (preprint) and Broadie and Glasserman [2004] (published). We refer to the latter, as well as Glasserman [2004, §8.5] where a more general version than the original is provided. Beyond what we deal with here, Broadie et al. [2000] provide a different approach to the previous references and a more recent reference on the stochastic mesh method is given by Liu and Hong [2009].

4.1 General Methodology

The stochastic mesh method requires the generation of N independent sample paths of a Markov process $\mathbf{S}_j = \left\{S_j^1, S_j^2, \ldots, S_j^N\right\}$ for times t_j, $j = 1,\ 2,\ \ldots,\ M$ (see Figure 4.1) where $\ln \frac{S_j^i}{S_{j-1}^i}$ are i.i.d. random samples from the same density for all i.

Once these random vectors have been generated, all nodes at time step t_j are connected to all nodes at time step t_{j+1} (see Figure 4.2) and so the individual original paths are forgotten.

In the *random tree method* of Broadie and Glasserman [1997b] (also discussed by Glasserman [2004, §8.3]) one simulates a Markovian non-recombining tree of paths of the underlying stock price process $\{S_t\}_{t\geq 0}$. At each node a small number of paths are simulated to successor nodes. Even though the number of immediate successor nodes are small, the number of nodes of the tree increases exponentially as the number of exercise dates increases and hence so does the computational effort.

In contrast, in the stochastic mesh method, the number of nodes is fixed at every time step after time t_0. This is achieved by connecting every node at time t_j to every node at time t_{j+1}. In the random tree method, the immediate successor nodes are generated at random, so we may regard them as equally likely. However, given a node i at time t_j in the stochastic mesh, clearly not all nodes at time t_{j+1} are equally likely. Hence we need to assign a weight to each connecting edge. It is reasonable to expect that even though we have to compute these weights, the computational cost would be linear, thus making the method computationally feasible.

Let us denote the weight connecting node i at time t_j (that is S_j^i) to node k at time t_{j+1} (that is S_{j+1}^k) by $w_j^{i,k}$ with $i, k = 1,\ 2,\ \ldots,\ N$ (see Figure 4.3).

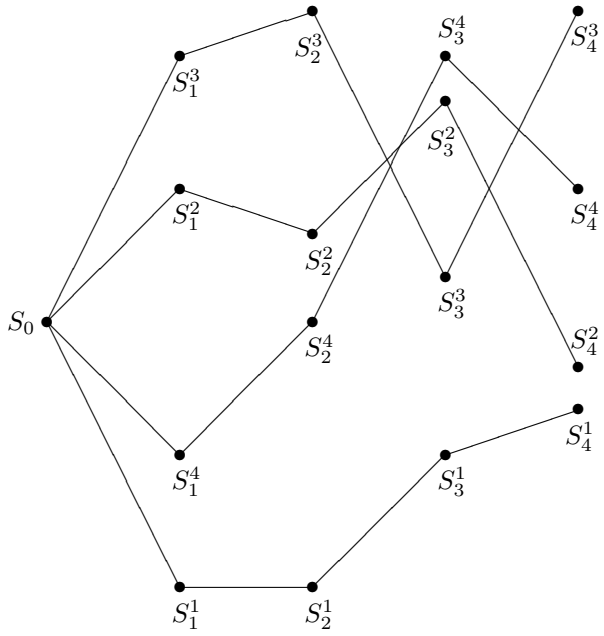

Figure 4.1: *Four independent paths where the nodes S_j^i are generated from the same returns distribution for $j = 1,\ 2,\ \ldots,\ 4$ and $i = 1,\ 2,\ \ldots,\ 4$.*

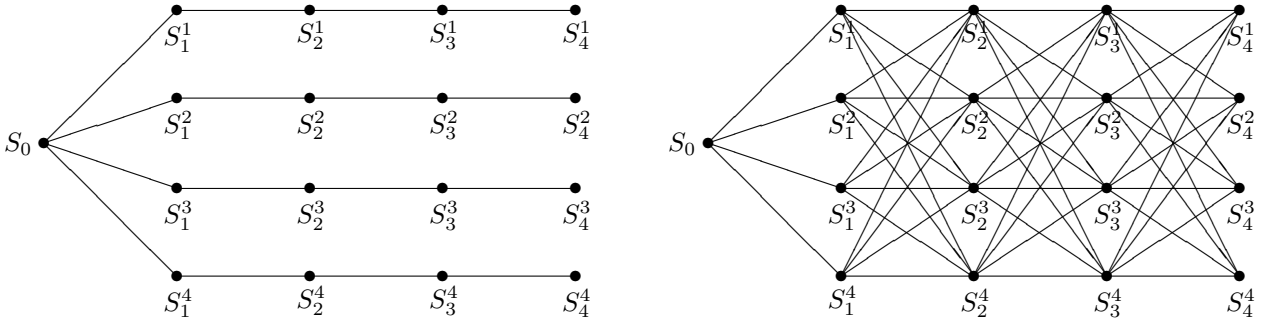

Figure 4.2: *The figure on the left is a schematic representation of Figure 4.1 of nodes generated from independent paths. The figure on the right shows how the mesh is constructed by connecting all nodes from one time step to another. See Glasserman [2004, Fig. 8.7] for similar figures.*

The stochastic mesh method provides a way of estimating the value of the American option by solving a randomly sampled dynamic programming problem. Here the continuation value is estimated by using the set of weights that connects the stock price nodes. That is, with $\bar{Z}_j^i$ denoting the modelled continuation value at time t_j,

$$\bar{Z}_j^i = e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} \bar{V}_{j+1}^k$$

if we are at node S_j^i, where $\bar{V}_{j+1}^k$ indicates the modelled value of the option at time step t_{j+1} at node k,

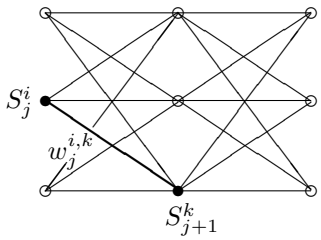

Figure 4.3: *Connecting node S_j^i to node S_{j+1}^k using weight $w_j^{i,k}$ indicated in bold.*

that is at S_{j+1}^k.

Thus when applying the stochastic mesh method, calculating the continuation value at time step t_j requires the use of all nodes at time step t_{j+1} and their weights, and not just the original successor of the current node as in the random tree method.

As noted by [Glasserman, 2004, §8.5.1], the main difficulty regarding the stochastic mesh method is in determining these weights.

4.2 Mesh Density Weights

We give here the construction of weights as presented in Broadie and Glasserman [2004] and Glasserman [2004, §8.5.2] which requires the conditional density of the underlying process to be known. This is not the only way of constructing weights — an alternative approach in Broadie et al. [2000] entirely avoids densities.

Glasserman [2004, §8.5.1] imposes three conditions on the mesh construction and the weights. Denote by $\mathbf{S}_j$ the vector of N nodes $\left[S_j^1,\ S_j^2,\ \ldots,\ S_j^N\right]$ at time t_j, $j = 1,\ 2,\ \ldots,\ M$ and let $\mathbf{S}_0 = S_0$.

(i) The first condition requires that, conditional on $\mathbf{S}_j$, the sets $\{\mathbf{S}_0,\ \mathbf{S}_1,\ \ldots,\ \mathbf{S}_{j-1}\}$ and $\{\mathbf{S}_{j+1},\ \mathbf{S}_{j+2},\ \ldots,\ \mathbf{S}_M\}$ are independent for every $j = 1,\ 2,\ \ldots,\ M-1$. This clearly holds as the stock price process $\mathbf{S}_j$ follows a Markov process.

(ii) The second condition requires that the weights $w_j^{i,k}$ are a deterministic function of $\mathbf{S}_j$ and $\mathbf{S}_{j+1}$.

(iii) The third condition is more restrictive. It requires that the weights chosen on average yields the correct continuation value. That is, for every node i at every time step t_j, $j = 1,\ 2,\ \ldots,\ M-1$

$$Z_j^i = e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} \mathbb{E}\left[w_j^{i,k} V_{j+1}^k \middle| S_j^i \right] \tag{4.1}$$

where Z_j^i and V_{j+1}^k indicates the true continuation value and value of the option respectively.

Broadie and Glasserman [2004] do not state these conditions explicitly. However, they do make three moment assumptions when proving the convergence of the high and low bias. Although the true values of Z_j^i and V_{j+1}^k are not known, we will show abstractly in the next section that for a particular choice of $w_j^{i,k}$ this condition is satisfied.

In order to simplify our presentation we will assume that the risk-free rate $r = 0$ in the rest of this section. We will however reintroduce a non-zero risk-free rate in our implementation in §4.3.

4.2.1 Deriving Abstract Weights

Suppose that the stock price S_{j+1}, conditional on $S_j = x$, has density $f_{j,j+1}(x,\cdot)$ [1], that is,

$$\mathbb{P}\left(S_{j+1} \leq \alpha | \, S_j = x\right) = \int_{-\infty}^{\alpha} f_{j,j+1}(x,y)\, dy.$$

Let g_j be any such density function of S_j at t_j, $j = 1,\ 2,\ \ldots,\ M$ conditional on S_0; the g_j's are generated independently of each other.

For now we derive weights using abstract g_j that satisfy conditions (i), (ii) and (iii) mentioned above.

Consider the following expectation of the option continuation value with respect to $f_{j,j+1}(x,\cdot)$:

$$\begin{aligned} Z_j(x) = \mathbb{E}_{f_{j,j+1}}\left[V_{j+1}\left(S_{j+1}\right) | \, S_j = x\right] &= \int V_{j+1}(y) f_{j,j+1}(x,y)\, dy \\ &= \int V_{j+1}(y) \frac{f_{j,j+1}(x,y)}{g_{j+1}(y)} g_{j+1}(y)\, dy \\ &= \mathbb{E}_{g_{j+1}}\left[V_{j+1}\left(S_{j+1}\right) \frac{f_{j,j+1}\left(x, S_{j+1}\right)}{g_{j+1}\left(S_{j+1}\right)}\right]. \end{aligned}$$

Note that the first expectation is with respect to $f_{j,j+1}(x,\cdot)$ whereas the last expectation is with respect to $g_{j+1}(\cdot)$. Now, if the mesh points $\mathbf{S}_j$ have density $g_j(\cdot)$ for $j = 1,\ 2,\ \ldots,\ M$, then we may approximate the last expectation by

$$\frac{1}{N}\sum_{k=1}^{N} V_{j+1}\left(S_{j+1}^k\right) \frac{f_{j,j+1}\left(x, S_{j+1}^k\right)}{g_{j+1}\left(S_{j+1}^k\right)}$$

and so let this be our model of the continuation value:

$$\bar{Z}_j(x) := \frac{1}{N}\sum_{k=1}^{N} \bar{V}_{j+1}\left(S_{j+1}^k\right) \frac{f_{j,j+1}\left(x, S_{j+1}^k\right)}{g_{j+1}\left(S_{j+1}^k\right)} \tag{4.2}$$

where $\bar{V}_{j+1}\left(S_{j+1}^k\right)$ is the approximated value of the American option at node S_{j+1}^k at time t_{j+1} already defined by backwards induction. Let us define the weights by

$$w_j^{i,k} := \frac{f_{j,j+1}\left(S_j^i, S_{j+1}^k\right)}{g_{j+1}\left(S_{j+1}^k\right)} \tag{4.3}$$

for $j = 0,\ 1,\ \ldots,\ M-1$. Then since the convergence

$$\frac{1}{N}\sum_{k=1}^{N} V_{j+1}^k w_j^{i,k} \to \mathbb{E}_{g_{j+1}}\left[V_{j+1}\left(S_{j+1}\right) \frac{f_{j,j+1}\left(S_j^i, S_{j+1}\right)}{g_{j+1}\left(S_{j+1}\right)}\right] \text{ as } N \to \infty$$

is unbiased, we have that

$$\mathbb{E}\left[\frac{1}{N}\sum_{k=1}^{N} V_{j+1}^k w_j^{i,k}\right] = \mathbb{E}_{g_{j+1}}\left[V_{j+1}\left(S_{j+1}\right) \frac{f_{j,j+1}\left(S_j^i, S_{j+1}\right)}{g_{j+1}\left(S_{j+1}\right)}\right]$$

[1] A general method for finding stock price densities from return densities is given in Appendix F.2.

and hence we may write

$$\begin{aligned} Z_j^i &= \mathbb{E}_{g_{j+1}}\left[V_{j+1}\left(S_{j+1}\right)\frac{f_{j,j+1}\left(S_j^i, S_{j+1}\right)}{g_{j+1}\left(S_{j+1}\right)}\right] \\ &= \frac{1}{N}\sum_{k=1}^{N}\mathbb{E}_{g_{j+1}}\left[V_{j+1}^k\frac{f_{j,j+1}\left(S_j^i, S_{j+1}^k\right)}{g_{j+1}\left(S_{j+1}^k\right)}\right] \\ &= \frac{1}{N}\sum_{k=1}^{N}\mathbb{E}_{g_{j+1}}\left[V_{j+1}^k w_j^{i,k}\middle| S_j^i\right] \end{aligned}$$

which is what we have in (4.1). Therefore, the conditions (i), (ii) and (iii) given in §4.2 holds for g_{j+1}.

4.2.2 Choosing an Appropriate g_j

We now provide some derivations which give guidance to making a suitable choice of g_j. In particular, we show that choosing g_j to be the seemingly natural choice, $f_{0,j}(S_0, \cdot)$, will not be suitable.

If we approximate the value of a *European option*, that is, $\bar{Z}_j = \bar{V}_j$ then

$$\begin{aligned} \bar{V}_0\left(S_0\right) &= \frac{1}{N}\sum_{i_1=1}^{N}\bar{V}_1\left(S_1^{i_1}\right)\frac{f_{0,1}\left(S_0, S_1^{i_1}\right)}{g_1\left(S_1^{i_1}\right)} \\ &= \frac{1}{N}\sum_{i_1=1}^{N}\frac{f_{0,1}\left(S_0, S_1^{i_1}\right)}{g_1\left(S_1^{i_1}\right)}\left[\frac{1}{N}\sum_{i_2=1}^{N}\frac{f_{1,2}\left(S_1^{i_1}, S_2^{i_2}\right)}{g_2\left(S_2^{i_2}\right)}\bar{V}_2\left(S_2^{i_2}\right)\right] \\ &= \frac{1}{N}\sum_{i_1=1}\frac{f_{0,1}\left(S_0, S_1^{i_1}\right)}{g_1\left(S_1^{i_1}\right)}\left[\frac{1}{N}\sum_{i_2=1}^{N}\frac{f_{1,2}\left(S_1^{i_1}, S_2^{i_2}\right)}{g_2\left(S_2^{i_2}\right)}\left[\frac{1}{N}\sum_{i_3=1}^{N}\frac{f_{2,3}\left(S_2^{i_2}, S_3^{i_3}\right)}{g_3\left(S_3^{i_3}\right)}\bar{V}_3\left(S_3^{i_3}\right)\right]\right] \end{aligned} \tag{4.4}$$

where $\bar{V}_j\left(S_j^{i_j}\right)$ indicates the European option value at the mesh point $S_j^{i_j}$ at time t_j. We can write (4.4) as

$$\begin{aligned} \bar{V}_0\left(S_0\right) &= \frac{1}{N}\sum_{i_1=1}^{N}\frac{1}{N}\sum_{i_2=1}^{N}\frac{1}{N}\sum_{i_3=1}^{N}\frac{f_{0,1}\left(S_0, S_1^{i_1}\right)f_{1,2}\left(S_1^{i_1}, S_2^{i_2}\right)f_{2,3}\left(S_2^{i_2}, S_3^{i_3}\right)}{g_1\left(S_1^{i_1}\right)g_2\left(S_2^{i_2}\right)g_3\left(S_3^{i_3}\right)}\bar{V}_3\left(S_3^{i_3}\right) \\ &= \frac{1}{N}\sum_{i_3=1}^{N}\frac{1}{N}\sum_{i_2=1}^{N}\frac{f_{2,3}\left(S_2^{i_2}, S_3^{i_3}\right)}{g_3\left(S_3^{i_3}\right)}\frac{1}{N}\sum_{i_1=1}^{N}\frac{f_{1,2}\left(S_1^{i_1}, S_2^{i_2}\right)}{g_2\left(S_2^{i_2}\right)}\frac{f_{0,1}\left(S_0, S_1^{i_1}\right)}{g_1\left(S_1^{i_1}\right)}\bar{V}_3\left(S_3^{i_3}\right) \end{aligned} \tag{4.5}$$

where the last equation has been rearranged so that the sum is reversed.

We show by induction that that $\bar{V}_0(S_0)$ may be written in terms of $\bar{V}_j$ for every $j = 1,\ 2,\ \ldots,\ M$: the approximation of the European option value at time 0 may be written as

$$\bar{V}_0\left(S_0\right) = \frac{1}{N}\sum_{i_j=1}^{N}\frac{1}{N}\sum_{i_{j-1}=1}^{N}\cdots\frac{1}{N}\sum_{i_1=1}^{N}\prod_{m=1}^{j}\frac{f_{m-1,m}\left(S_{m-1}^{i_{m-1}}, S_m^{i_m}\right)}{g_m\left(S_m^{i_m}\right)}\bar{V}_j\left(S_j^{i_j}\right) \tag{4.6}$$

for every $j = 1,\ 2,\ \ldots,\ M$.

We have already seen by (4.5) that (4.6) holds for $j = 1$ (and 2 and 3) and let us assume that

$$\bar{V}_0\left(S_0\right) = \frac{1}{N}\sum_{i_j=1}^{N}\frac{1}{N}\sum_{i_{j-1}=1}^{N}\cdots\frac{1}{N}\sum_{i_1=1}^{N}\prod_{m=1}^{j}\frac{f_{m-1,m}\left(S_{m-1}^{i_{m-1}}, S_m^{i_m}\right)}{g_m\left(S_m^{i_m}\right)}\bar{V}_j\left(S_j^{i_j}\right).$$

From the above and the fact that $\bar{V}_j\left(S_j^{i_j}\right)$ is given by (4.2) we have that

$$\begin{aligned}\bar{V}_0(S_0) &= \frac{1}{N}\sum_{i_j=1}^{N}\frac{1}{N}\sum_{i_{j-1}=1}^{N}\cdots\frac{1}{N}\sum_{i_1=1}^{N}\prod_{m=1}^{j}\frac{f_{m-1,m}\left(S_{m-1}^{i_{m-1}},S_m^{i_m}\right)}{g_m\left(S_m^{i_m}\right)}\left[\frac{1}{N}\sum_{i_{j+1}=1}^{N}\bar{V}_{j+1}\left(S_{j+1}^{i}\right)\frac{f_{j,j+1}\left(S_j^{i_j},S_{j+1}^{i_{j+1}}\right)}{g_{j+1}\left(S_{j+1}^{i_{j+1}}\right)}\right]\\ &= \sum_{i_{j+1}=1}^{N}\frac{1}{N}\sum_{i_j=1}^{N}\frac{1}{N}\sum_{i_{j-1}=1}^{N}\cdots\frac{1}{N}\sum_{i_1=1}^{N}\prod_{m=1}^{j+1}\frac{f_{m-1,m}\left(S_{m-1}^{i_{m-1}},S_m^{i_m}\right)}{g_m\left(S_m^{i_m}\right)}\bar{V}_{j+1}\left(S_{j+1}^{i_{j+1}}\right)\end{aligned}$$

where the last equation follows from rearranging the sum.

Similarly to Broadie and Glasserman [2004], let us denote by

$$L(1,i_1) := \frac{f_{0,1}\left(S_0,S_1^{i_1}\right)}{g_1\left(S_1^{i_1}\right)}$$

and hence

$$\bar{V}_0(S_0) = \frac{1}{N}\sum_{i_1=1}^{N}L(1,i_1)\,\bar{V}_1\left(S_1^{i_1}\right). \tag{4.7}$$

Inductively define

$$L(j,i_j) = \frac{1}{N}\sum_{i_{j-1}=1}^{N}\frac{f_{j-1,j}\left(S_{j-1}^{i_{j-1}},S_j^{i_j}\right)}{g_j\left(S_j^{i_j}\right)}L(j-1,i_{j-1}) \tag{4.8}$$

for $j = 1,\ 2,\ \ldots,\ M$. Again using induction, we show that $\bar{V}_0$ may be written in terms of $L(j,i_j)$:

The approximation of the European option value $\bar{V}_0$ may be written as

$$\bar{V}_0(S_0) = \frac{1}{N}\sum_{i_j=1}^{N}L(j,i_j)\,\bar{V}_j\left(S_j^{i_j}\right) \tag{4.9}$$

for every $j = 1,\ 2,\ \ldots,\ M$.

We have already seen in (4.7) that (4.9) holds when $j = 1$. Now assume that

$$\bar{V}_0(S_0) = \frac{1}{N}\sum_{i_j=1}^{N}L(j,i_j)\,\bar{V}_j\left(S_j^{i_j}\right).$$

Then using (4.2) we may write the above as

$$\begin{aligned}\bar{V}_0(S_0) &= \frac{1}{N}\sum_{i_j=1}^{N}L(j,i_j)\left[\frac{1}{N}\sum_{i_{j+1}=1}^{N}\bar{V}_{j+1}\left(S_{j+1}^{i_{j+1}}\right)\frac{f_{j,j+1}\left(S_j^{i_j},S_{j+1}^{i_{j+1}}\right)}{g_{j+1}\left(S_{j+1}^{i_{j+1}}\right)}\right]\\ &= \frac{1}{N}\sum_{i_{j+1}=1}^{N}\frac{1}{N}\sum_{i_j=1}^{N}L(j,i_j)\frac{f_{j,j+1}\left(S_j^{i_j},S_{j+1}^{i_{j+1}}\right)}{g_{j+1}\left(S_{j+1}^{i_{j+1}}\right)}\bar{V}_{j+1}\left(S_{j+1}^{i_{j+1}}\right)\\ &= \frac{1}{N}\sum_{i_{j+1}=1}^{N}L(j+1,i_{j+1})\,\bar{V}_{j+1}\left(S_{j+1}^{i_{j+1}}\right)\end{aligned}$$

where the last equation follows from the definition of $L\left(j+1, i_{j+1}\right)$ in (4.8).

A complicated proof in Broadie and Glasserman [2004, Proposition 1] shows that under modest technical assumptions, the variance of $L\left(j, i_j\right)$ increases exponentially for an arbitrary g_j. This will cause the variance of the Monte Carlo estimator to grow. Broadie and Glasserman [2004] note that this is true in particular for $f_{0,j}\left(S_0, \cdot\right)$.

However Broadie and Glasserman [2004] make the inspired choice

$$g_1(y) = f_{0,1}\left(S_0, y\right), \tag{4.10}$$

$$g_j(y) = \frac{1}{N}\sum_{i=1}^{N} f_{j-1,j}\left(S_{j-1}^i, y\right) \text{ for } j = 2,\ 3,\ \ldots,\ M. \tag{4.11}$$

Observe that g_j is a density since the average of densities is a density and is referred to as the *average density function* in Broadie and Glasserman [2004].

If we choose g_j as in (4.10) and (4.11) (with g_j conditional only on S_0 inductively) then each $L\left(j, i_j\right) = 1$, $j = 1,\ 2,\ \ldots,\ M$, and therefore no explosion in variance occurs for this choice of g_j. This follows by induction: note that from (4.10) $L(1, i_1) = 1$ and suppose that $L\left(j-1, i_{j-1}\right) = 1$. Then

$$\begin{aligned} L\left(j, i_j\right) &= \frac{1}{N}\sum_{i_{j-1}=1}^{N} \frac{f_{j-1,j}\left(S_{j-1}^{i_{j-1}}, S_j^{i_j}\right)}{g_j\left(S_j^{i_j}\right)} L\left(j-1, i_{j-1}\right) \\ &= \frac{1}{N}\sum_{i_{j-1}=1}^{N} \frac{f_{j-1,j}\left(S_{j-1}^{i_{j-1}}, S_j^{i_j}\right)}{g_j\left(S_j^{i_j}\right)} \end{aligned}$$

which follows from the induction hypothesis. Furthermore, we have from (4.11) that

$$g_j(y) = \frac{1}{N}\sum_{i=1}^{N} f_{j-1,j}\left(S_{j-1}^i, y\right)$$

and hence $L\left(j, i_j\right) = 1$.

Now suppose that $h_j(\cdot)$ is the unconditional risk-neutral density used to generate the mesh points $\mathbf{S}_j$ at time t_j for each j, then

$$h_{j+1}(y) = \int h_j(x) f_{j,j+1}\left(x, y\right)\, dx$$

and from (4.10) $g_1(\cdot) = f_{0,1}\left(S_0, \cdot\right) = h_1(\cdot)$. Furthermore for $j = 1,\ 2,\ \ldots,\ M-1$, $g_{j+1}(y)$ is approximately equal to $\int g_j(x) f_{j,j+1}(x, y)\, dx$, that is from (4.11) for $j = 1,\ 2,\ \ldots,\ M-1$

$$g_{j+1}(y) = \frac{1}{N}\sum_{i=1}^{N} f_{j,j+1}\left(S_j^i, y\right) \approx \int g_j(x) f_{j,j+1}(x, y)\, dx.$$

So, without being too precise but rather retaining intuition, we see that if $g_j \approx h_j$, then by induction $g_{j+1} \approx h_{j+1}$ and we would have equality in the limit as $N \to \infty$.

4.3 The Stochastic Mesh Method

As in the case of the LSM method in Chapter 2, we determine the American option value $\hat{V}$ using the stochastic mesh method by backward induction

$$\begin{aligned}\hat{V}_M^i &= I_M^i \\ \hat{V}_j^i &= \max\left\{I_j^i, e^{-r\Delta t_{j+1}}\frac{1}{N}\sum_{k=1}^{N}\hat{V}_{j+1}^k w_j^{i,k}\right\}\end{aligned} \tag{4.12}$$

where we have used (4.2) and (4.3) in §4.2 with a non-zero risk-free rate r to write the approximation of the American option value at time t_j at node i. The weights $w_j^{i,k}$ are given by

$$\begin{aligned}w_j^{i,k} &= \frac{f_{j,j+1}\left(S_j^i, S_{j+1}^k\right)}{g_{j+1}\left(S_{j+1}^k\right)} \\ &= \frac{f_{j,j+1}\left(S_j^i, S_{j+1}^k\right)}{\frac{1}{N}\sum_{l=1}^{N} f_{j,j+1}\left(S_j^l, S_{j+1}^k\right)}.\end{aligned} \tag{4.13}$$

Below we provide an algorithm for calculating (4.12).

- Given the stock price matrix S_j^i and intrinsic value matrix, let NV (new value) and OV (old value) denote vectors of length N. We will suppress the time index j in these vectors because their entries will be overwritten at every time step. We begin by setting $NV^i := I_M^i$ for every i at time t_M.
- Next we step backwards in time from t_{M-1} until t_1:

  For $j = M-1$ **To** 1 **Step** -1

  - Set $OV := NV$.
  - For each $k = 1,\ 2,\ \ldots,\ N$, we would like to calculate the denominator in (4.13). In order to do this, we make use of a helper vector D of length N, with the k^{th} entry of D denoted byD^k. Set $D^k := 0$.

    For $i = 1$ **To** N

    * Let $w^{i,k}$ be the $(i,k)^{\text{th}}$ entry of an $N \times N$ matrix. This $w^{i,k}$ will eventually be equal to (4.13) (again suppressing the time index j for the same reason as before). For now we set $w^{i,k} := f_j\left(S_j^i, S_{j+1}^k\right)$.
    * Set $D^k = D^k + w_j^{i,k}$.

    Next i

    Set $D^k := D^k/N$ and hence D^k is equal to the denominator in (4.13).
  - Calculate the weights between all nodes i at time t_j and all nodes k at time t_{j+1}, that is, we find (4.13) by setting $w^{i,k} := w^{i,k}/D^k$.
  - For each i, calculate $\hat{V}_j^i$ as in (4.12): First set $NV^i := 0$ and calculate the model continuation value at node i in NV^i by setting $NV^i := e^{-r\Delta t_{j+1}}\frac{1}{N}\sum_{k=1}^{N} OV^k w^{i,k}$. Then set $NV^i = \max\left\{I_j^i, NV^i\right\}$.

Next j

- As in the case of the LSM method, we find the approximated value at time 0 as

$$\max\left\{I_0, e^{-r\Delta t_1}\frac{1}{N}\sum_{i=1}^{N} NV^i\right\}.$$

4.4 High Bias

Let the estimated continuation value be given by

$$\hat{Z}_j^i = e^{-r\Delta t_{j+1}}\frac{1}{N}\sum_{k=1}^{N} w_j^{i,k}\hat{V}_{j+1}^k \tag{4.14}$$

with $w_j^{i,k}$ indicating the weight which connects node i at time t_j with node k at time t_{j+1}. For the S_{j+1}^i increasing, the values of the weights will increase and then decrease (these values will start decreasing at the mode of the distribution of S_{j+1}). Thus there is no reason to expect that even if the $\hat{V}_{j+1}^i$ are monotone, that the model of the continuation value is monotone (even though the real continuation value is monotone). This undesired feature indeed can occur as is shown in Figure 4.4.

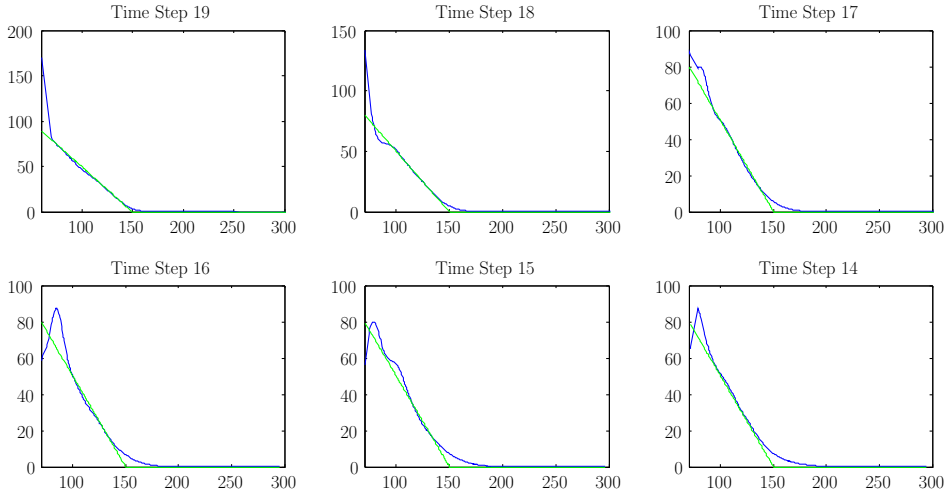

Figure 4.4: *The continuation value (in blue) and intrinsic value (in green) plotted against various stock prices as determined by the high bias estimator at time steps 14 to 19 of a 20 step grid.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 150$. *The stock price paths were generated using simulation techniques we discuss in §7.4 and §7.5.*

As discussed, notice that the model of the continuation value is not monotone.

We define by backward induction [Glasserman, 2004, 8.5.1]

$$\hat{V}_M^i = I_M^i \tag{4.15}$$

$$\hat{V}_j^i = \max\left\{I_j^i, \hat{Z}_j^i\right\} \text{ for } j = 1,\ 2,\ \ldots,\ M-1 \tag{4.16}$$

and

$$\hat{V}_0 = e^{-r\Delta t_1}\frac{1}{N}\sum_{i=1}^{N}\hat{V}_1^i.$$

Thus $\hat{V}_j^i$ results from applying dynamic programming to the stochastic mesh. Glasserman [2004, §8.5.1] and Broadie and Glasserman [2004] refer to $\hat{V}_j^i$ as the *mesh estimator*.

We show in a proof similar to that of Theorem 1.4.2 that for every t_j and i, $\hat{V}_j^i$ has a high bias. We proceed by induction: first observe that this statement trivially holds at t_M since $\hat{V}_M^i = I_M^i = V_M^i$ for every i. Next we assume that $\mathbb{E}\left[\hat{V}_{j+1}^i \middle| \mathbf{S}_{j+1}\right] \geq V_{j+1}^i$. From Jensen's inequality we have that

$$\mathbb{E}\left[\hat{V}_j^i \middle| \mathbf{S}_j\right] \geq \max\left\{I_j^i, e^{-r\Delta t_{j+1}}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\left[w_j^{i,k}\hat{V}_{j+1}^k \middle| \mathbf{S}_j\right]\right\}.$$

Using the tower property and condition (ii) we may write

$$\begin{aligned}\mathbb{E}\left[w_j^{i,k}\hat{V}_{j+1}^i \middle| \mathbf{S}_j\right] &= \mathbb{E}\left[\mathbb{E}\left[w_j^{i,k}\hat{V}_{j+1}^k \middle| \mathbf{S}_{j+1}\right] \middle| \mathbf{S}_j\right] \\ &= \mathbb{E}\left[w_j^{i,k}\mathbb{E}\left[\hat{V}_{j+1}^k \middle| \mathbf{S}_{j+1}\right] \middle| \mathbf{S}_j\right]. \end{aligned} \tag{4.17}$$

Now from the induction hypothesis we have that

$$\begin{aligned}\mathbb{E}\left[\hat{V}_{j+1}^k \middle| \mathbf{S}_{j+1}\right] &\geq V_{j+1}^k \\ \Rightarrow w_j^{i,k}\mathbb{E}\left[\hat{V}_{j+1}^k \middle| \mathbf{S}_{j+1}\right] &\geq w_j^{i,k}V_{j+1}^k \\ \Rightarrow \mathbb{E}\left[w_j^{i,k}\mathbb{E}\left[\hat{V}_{j+1}^k \middle| \mathbf{S}_{j+1}\right] \middle| \mathbf{S}_j\right] &\geq \mathbb{E}\left[w_j^{i,k}V_{j+1}^k \middle| \mathbf{S}_j\right]\end{aligned}$$

and hence from (4.17) and the above it follows that

$$\begin{aligned}\max\left\{I_j^i, e^{-r\Delta t_{j+1}}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\left[w_j^{i,k}\hat{V}_{j+1}^k \middle| \mathbf{S}_j\right]\right\} &= \max\left\{I_j^i, e^{-r\Delta t_{j+1}}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\left[w_j^{i,k}\mathbb{E}\left[\hat{V}_{j+1}^k \middle| \mathbf{S}_{j+1}\right] \middle| \mathbf{S}_j\right]\right\} \\ &\geq \max\left\{I_j^i, e^{-r\Delta t_{j+1}}\frac{1}{N}\sum_{k=1}^{N}\mathbb{E}\left[w_j^{i,k}V_{j+1}^k \middle| \mathbf{S}_j\right]\right\} \\ &= \max\left\{I_j^i, Z_j^i\right\} \\ &= V_j^i\end{aligned}$$

where the second last equality follows from condition (iii). This completes the proof.

In a similar proof Broadie and Glasserman [2004, Theorem 1] show the mesh estimator is biased high without the conditions we gave in §4.1, but there the weights $w_j^{i,k}$ are defined in terms of the likelihood ratio as in (4.3).

Broadie and Glasserman [1997a] show under some conditions, that the high bias estimator $\hat{V}_j^i$ converges in probability to the true option value V_j^i given the stock price is S_j^i as $N \to \infty$. Glasserman [2004, §8.5.1] provides some intuition why these conditions were imposed. By applying the *contraction property*[2] and

[2]The *contraction property* is given by

$$|\max\{a, c_1\} - \max\{a, c_2\}| \leq |c_1 - c_2|.$$

condition (iii) in §4.1 to the dynamic programming formulation of the high bias estimator given in §4.4 we have that

$$
\begin{aligned}
&\left|\hat{V}_j^i - V_j^i\right| \\
&\leq \left| e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} \hat{V}_{j+1}^k - e^{-r\Delta t_{j+1}} \mathbb{E}\left[w_j^{i,k} V_{j+1}^i \middle| \mathcal{F}_j \right] \right| \\
&= \left| e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} \hat{V}_{j+1}^k - e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} V_{j+1} + e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} V_{j+1} - e^{-r\Delta t_{j+1}} \mathbb{E}\left[w_j^{i,k} V_{j+1}^i \middle| \mathcal{F}_j \right] \right| \\
&\leq \left| e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} \hat{V}_{j+1}^k - e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} V_{j+1} \right| + \left| e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} V_{j+1} - e^{-r\Delta t_{j+1}} \mathbb{E}\left[w_j^{i,k} V_{j+1}^i \middle| \mathcal{F}_j \right] \right| \\
&= \left| e^{-r\Delta t_{j+1}} \frac{1}{N} \sum_{k=1}^{N} w_j^{i,k} \left(\hat{V}_{j+1}^k - V_{j+1} \right) \right| + \left| e^{-r\Delta t_{j+1}} \sum_{k=1}^{N} w_j^{i,k} V_{j+1} - e^{-r\Delta t_{j+1}} \mathbb{E}\left[w_j^{i,k} V_{j+1}^i \middle| \mathcal{F}_j \right] \right|.
\end{aligned}
$$

In order for the term on the left in the last equality to go to zero, a sufficiently strong induction hypothesis for convergence of $\hat{V}_{j+1}^k$ to V_{j+1}^k is required. For the term on the right to go to zero, it is required that the sum satisfy the law of large numbers. Broadie and Glasserman [1997a] use these observations to prove convergence of the high bias estimator.

Glasserman [2004] notes that Avramidis and Matzinger [2004] derive a probabilistic upper bound on the error in the high bias estimator with a dependence structure that adheres to conditions (i) and (ii) in §4.1 and then use this bound to prove convergence of the estimator as $N \to \infty$.

Assuming conditions on the moments of payoffs, weights and likelihood ratios, Broadie and Glasserman [2004] prove the convergence of the high bias estimator in the p-norm, that is, for any path i and time t_j

$$\left\| \hat{V}_j^i - V_j^i \right\|_p \to 0$$

as $N \to \infty$. This convergence implies convergence of $\hat{V}_0 \to V_0$ in probability

$$\mathbb{P}\left(\lim_{N \to \infty} \hat{V}_0 = V_0 \right) = 1.$$

We also have that the high bias estimator is asymptotically unbiased since convergence in probability implies that as $N \to \infty$

$$\mathbb{E}\left[\hat{V}_0 \right] \to V_0.$$

As in the case where the high bias is shown by Broadie and Glasserman [2004], $w_j^{i,k}$ is defined in terms of the likelihood ratio as in (4.3).

4.5 Low Bias

In this section we will show how to find an estimator which has a low bias in the stochastic mesh method first suggested by Broadie and Glasserman [1997a]. Glasserman [2004, §8.5.1] and Broadie and Glasserman [2004] also consider this estimator and refer to it as the *path estimator* because of the way in which it is

found. The path estimator is found in a very similar way the low bias approximation is found in §2.3 for the regression or LSM methods.

The path estimator, which we will denote by $\breve{V}_j$ is obtained by using a stopping rule on the existing mesh — we denote the stock prices from this mesh by S_j^i. We begin by simulating paths at times t_j of the stock price $\breve{S}_j^i$. As in §2.3 we define stopping times

$$\breve{\tau}^i = \min\left\{j : I_j^i \geq \breve{Z}_j^i\right\}$$

where $I_j^i := I_j\left(\breve{S}_j^i\right)$ indicates the intrinsic value of the simulated path given the stock price is $\breve{S}_j^i$. Analogous to $\breve{Z}_j^i$ in (2.10), $\breve{Z}_j^i$ depends on the existing mesh as well as the newly generated mesh and is given by

$$\breve{Z}_j^i := \breve{Z}_j\left(\breve{S}_j^i\right) = \frac{1}{N}\sum_{k=1}^{N} w_j\left(\breve{S}_j^i, S_{j+1}^k\right)\hat{V}_{j+1}^k.$$

Observe that all values from time step t_{j+1} are found from the existing mesh, that is, the option values $\hat{V}_{j+1}^k$ and the stock prices S_{j+1}^k for the paths $k = 1,\ 2,\ \ldots,\ N$. The low bias estimator for path i is then given by

$$\breve{V}_j^i = e^{-r\left(\breve{\tau}^i - t_j\right)} I_{\breve{\tau}^i}^i.$$

Taking the average of these then gives the path estimator

$$\breve{V}_j = \frac{1}{N}\sum_{i=1}^{N} \breve{V}_j^i.$$

In Figure 4.5 we plot the performance of the stochastic mesh method as a function of the number of sample paths used in the high and low bias estimate. We used the same inputs as in Figure 2.3 and Figure 3.3. These calculations were very slow compared to the LSM-Rasmussen method.

In Table 4.1 we compare the performance of the stochastic mesh high and low bias prices to those generated by the LSM-Ramsussen method (for in-the-money, at-the-money and out-the-money options) and the time taken (in seconds) of these methods. We used the same inputs as in Table 3.1. In these examples we see that the LSM-Rasmussen method clearly outperforms the stochastic mesh method in computation time and accuracy.

Imposing the same conditions as when proving the convergence of the high bias estimator, Broadie and Glasserman [2004, Theorem 4] prove that the low bias estimator is asymptotically unbiased, that is, as $N \to \infty$

$$\mathbb{E}\left[\breve{V}_0\right] \to V_0.$$

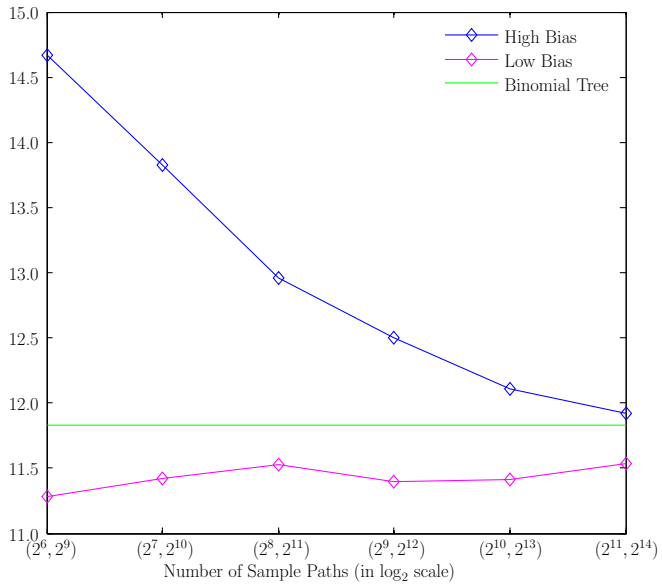

Figure 4.5: *Performance of the high and low bias prices generated using the stochastic mesh method.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$. *The simulation paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5. The number of sample paths used in the high bias estimate is given by the first entry in the bracket and that of the low bias estimate in the second on the horizontal axis. We also include the value given by a 30 step binomial tree.*

Similar convergence results were obtained for in-the-money and out-the-money options.

	Value			
K	Binomial Tree	LSM-Rasmussen	High Bias	Low Bias
85	0.4391	0.4364	0.4528	0.4405
130	9.6090	9.6218	10.6110	9.4038
135	11.8490	11.8671	12.9573	11.5308
140	14.3803	14.3891	15.4947	13.9470
185	50.0000	50.0000	52.0216	48.1773
	Time (seconds)			
K	Binomial Tree	LSM-Rasmussen	High Bias	Low Bias
85	0.345	5.331	241.149	0.005
130	0.326	3.941	219.393	0.005
135	0.234	4.058	188.121	0.000
140	0.265	4.487	222.395	0.007
185	0.275	4.849	258.797	0.008

Table 4.1: *The performance of the high and low bias prices generated using the stochastic mesh method compared to the LSM-Rasmussen method for varied strikes.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$. *The simulation paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5. The high bias and LSM-Rasmussen values were generated using 256 sample paths, whereas the low bias values were generated using 2048 sample paths. We also include the value given by a 30 step binomial tree.*

Clearly the LSM-Rasmussen method outperforms the stochastic mesh method in computation time and accuracy.

Chapter 5

Duality

In this chapter we consider the *dual method* for generating a high bias estimate when pricing American options using Monte Carlo simulation, independently suggested by Andersen and Broadie [2004], Rogers [2002] and Haugh and Kogan [March 2004]. This is achieved by making use of a given stopping rule which produces a lower bias estimate of the true value of an American option, and then a dual value is defined by extracting a martingale from the existing exercise rule which complements this low bias with a high bias estimate. We will frequently refer to Glasserman [2004, §8.7], which provides an excellent summary of this method.

In the next section we will provide our own algorithm for determining a low bias estimate using an existing exercise policy. Following this we will discuss the dual method which will make use of the exercise policy to produce the high bias estimate.

5.1 An Approximation of the Optimal Stopping Boundary

Thus far we have considered the LSM (where we may or may not have incorporated the control variate by Rasmussen) and regression methods in Chapters 2 and 3, and the stochastic mesh method in Chapter 4 from which we will obtain a stopping rule. Recall from §1.4.2 that if we follow an exercise policy, a low bias estimate of the American option price is produced.

We provide an algorithm that can be applied to any of these methods that finds an approximation of the free boundary between the exercise and continuation regions. This algorithm is our own.

We continue to make use of the notation introduced in Chapter 1. In addition to this notation, let η denote the style of the option under consideration. That is, $\eta = 1$ if we are considering a call and $\eta = -1$ if we are considering a put.

Let *CSP* indicate a vector of length $M + 1$ which eventually will contain the *critical stock price* for each time t_j, $j = 0,\ 1,\ \ldots,\ M$. CSP_j is the approximation of the free boundary at time t_j. It is clear that *CSP* is a function of the stock price process, but it is not a function of the initial spot price of the option we are considering, and we will explicitly make use of this fact in due course.

- In the stochastic mesh case, we begin by setting $CSP_M := K$. In the regression or LSM method, we

set $CSP_M := 1$, because we normalise the stock price by the strike of the option in order to prevent computational underflow (see §2.4).

- Suppose for definiteness that we are considering a put. If, at time t_j, we take into account all the stock prices $\mathbf{S}_j = \left[S_j^1,\ S_j^2,\ \ldots,\ S_j^N\right]$, then there is a particular index i^* such that we exercise when $S_j^i \leq S_j^{i^*}$ and hold if $S_j^i > S_j^{i^*}$. We will likewise find the index for the smallest hold price, that is we find i' such that we hold when $S_j^i \geq S_j^{i'}$ and exercise when $S_j^i < S_j^{i'}$. We then approximate the critical stock price CSP_j as a weighted average of $S_j^{i^*}$ and $S_j^{i'}$ [1]. This is achieved by using a weight λ that depends on $I_j - TV_j$ [2] at the nodes $S_j^{i^*}$ and $S_j^{i'}$. Furthermore, to ensure that we have a monotone exercise boundary, we compare the critical stock price at the previous time step t_{j+1}, CSP_{j+1} with CSP_j.

  A similar procedure is followed if we are considering a call.

  We now consider the technicalities of finding i^* and i'. We determine the critical stock price function by performing backward recursion. We will require the following tool: given a vector V, let W be a vector with the same components as V, but arranged in increasing order. Then, assuming the elements in V are unique, let *rank* be a function such that

  $$\text{rank}(i) = k \Leftrightarrow V^i = W^k \Leftrightarrow i = \text{rank}(k).$$

  Furthermore, let *index* be the inverse function of *rank*. In actual fact a numerical algorithm usually determines *index* first: see `indexx` in Press et al. [2004, §8.4] for calculating *index*, and hence *rank* as the inverse function.

  If we have identified a certain element with index i^*, then the nearest element when ordered (above or below) has index i' where $\text{rank}(i') = \text{rank}(i^*) - \eta$. Hence because *index* and *rank* are inverse functions we have $i' = \text{index}(\text{rank}(i^*) - \eta)$.

  For $j = M - 1$ **To** 0 **Step** -1

  - We start the search for the critical stock price at time t_j at CSP_{j+1}. Append to the array of stock prices $\mathbf{S}_j = \left[S_j^1,\ S_j^2,\ \ldots,\ S_j^N\right]$ at time t_j the previous critical value CSP_{j+1}. Now we define $i^* := \text{index}(N+1)$ so that i^* indicates the position of CSP_{j+1} in the order (see Figure 5.1). The functions *rank* and *index* are based on the expanded array of stock prices containing the previous critical value.
  - To ensure a monotone optimal stopping boundary, we search for the first point (smaller in the case of a put or larger in the case of a call) outside the model's continuation region. Thus, in the case of a put, we will decrease i^* until the first point inside the model's exercise region is found, that is, until $I_j^{i^*} \geq TV_j^{i^*}$. Note that we may decrease (or increase in the case of a call) the index, since we are considering an ordered array. In order to find this point, we proceed as follows:
    - We begin by setting $i^* = \text{index}(\text{rank}(i^*) + \eta)$, then i^* is the index of the stock price which is the first candidate to be inside the model's continuation region in the augmented array (see Figure 5.2).

[1] Note that $S_j^{i'}$ is the smallest S_j^i such that $S_j^i > S_j^{i^*}$.
[2] Recall that TV_j indicates the modelled continuation value at time step t_j.

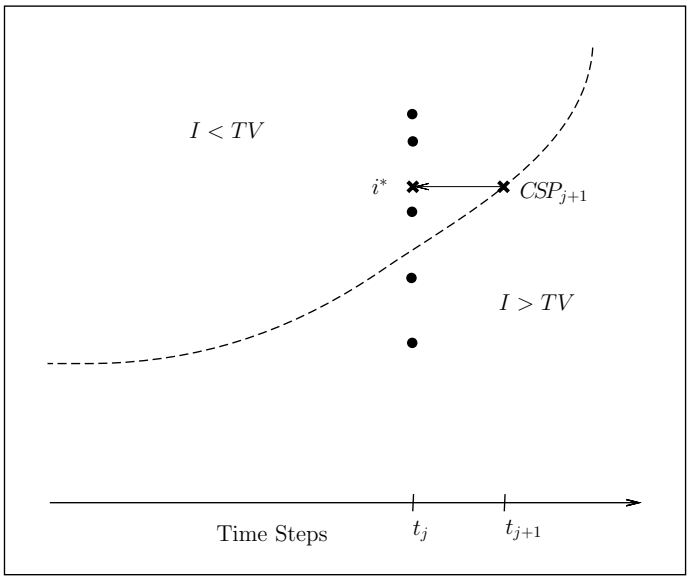

Figure 5.1: *Set i^* at time t_j equal to the index of CSP_{j+1} in the augmented stock price array. The dotted line indicates the critical stock price function that we are trying to find (a put is illustrated here, but the process carries over for calls).*

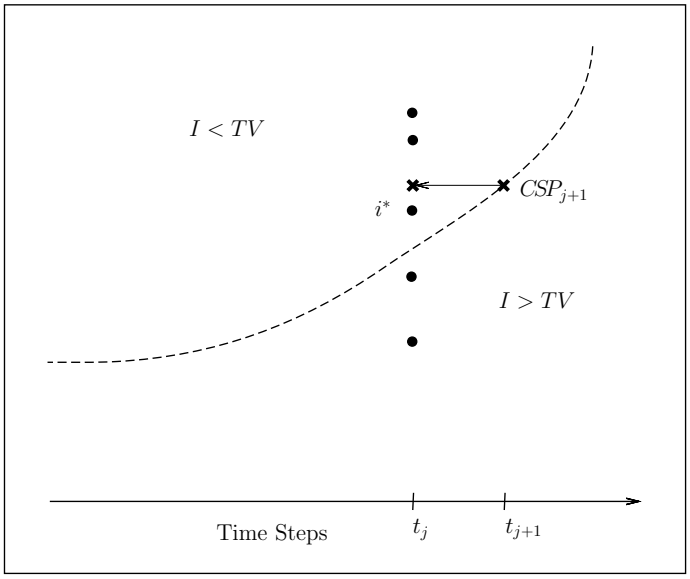

Figure 5.2: *Increase (or decrease) i^* so that it indicates the index of the stock price which is the first candidate to be inside the modelled continuation region.*

* Once i^* is obtained, discard the expanded array of stock prices, thus reverting to the actual array of stock prices. Crucially i^* in this array is still the index of the stock price found above (see Figure 5.3).
* We now loop until we find a point at which we exercise, that is, the first point in the

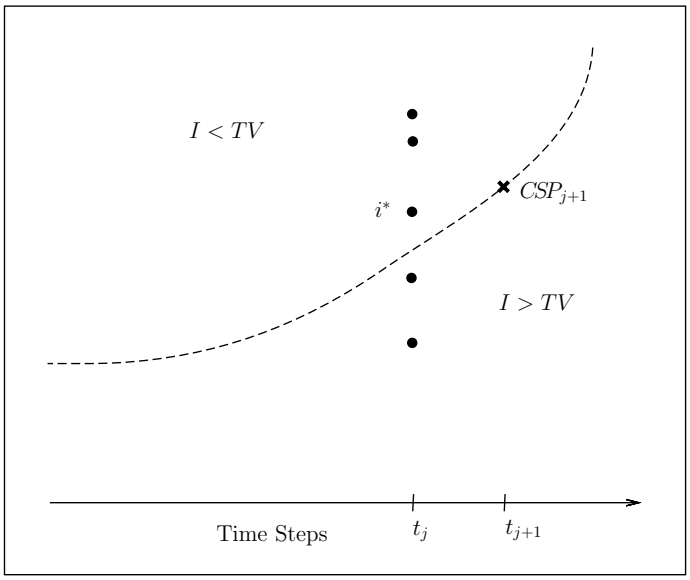

Figure 5.3: *We discard the expanded array of stock prices, i.e. we remove the index of CSP_{j+1}. Note that i^* is still the index of the stock price we found in Figure 5.2.*

modelled exercise region (see Figure 5.4):

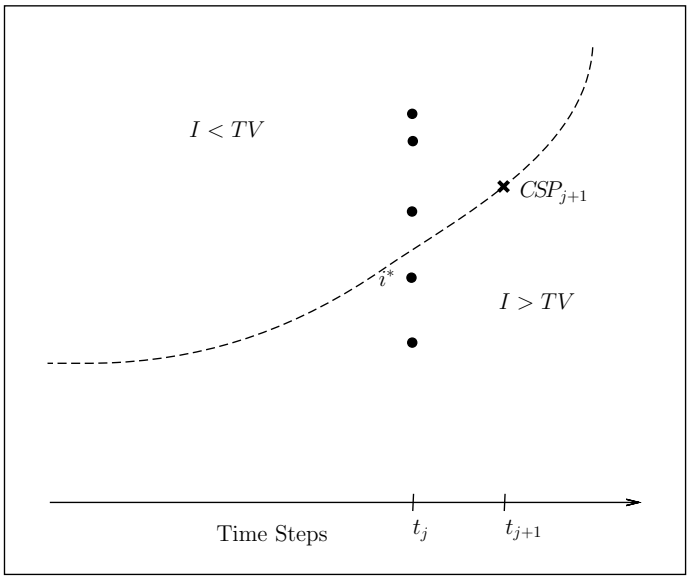

Figure 5.4: *Set i^* to be the index of the first stock price in the modelled exercise region.*

While $I_j^{i^*} < TV_j^{i^*}$ **Do**
Set $i^* := \text{index}(\text{rank}(i^*) + \eta)$
End While

TV_j^i indicates the modelled continuation value given by ${}_\kappa\bar{Z}_j\left(S_j^i\right)$ in (2.5) when considering the regression or LSM methods, or (3.2) when considering the LSM-Rasmussen method;

and $\hat{Z}_j^i$ in (4.14) when considering the stochastic mesh method.

– Let i' denote the index of the last stock price where we continued, thus we set $i' := \text{index}(\text{rank}(i^*) - \eta)$ (see Figure 5.5).

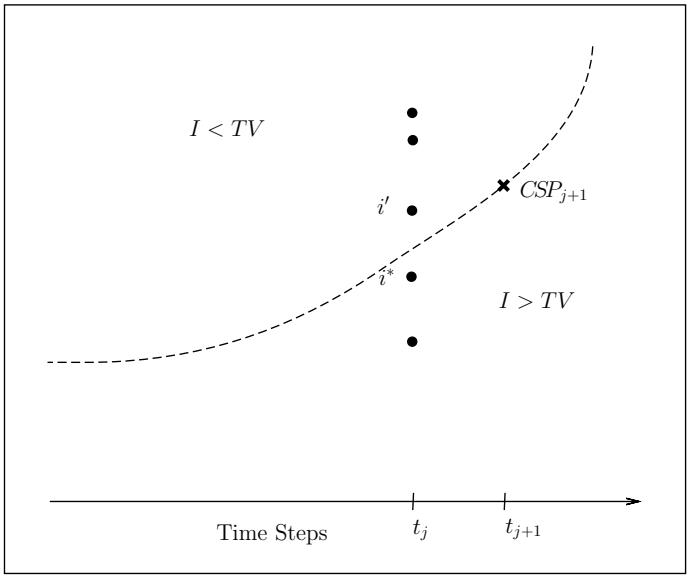

Figure 5.5: *Set i' to be the index of the first stock price in the modelled continuation region.*

We then find the critical stock price for time step t_j as the linearly interpolated value at the point $(CSP_j, 0)$ between $\left(S_j^{i^*}, I_j^{i^*} - TV_j^{i^*}\right)$ and $\left(S_j^{i'}, I_j^{i'} - TV_j^{i'}\right)$. So let

$$\lambda := \frac{I_j^{i'} - TV_j^{i'}}{I_j^{i'} - TV_j^{i'} - I_j^{i^*} + TV_j^{i^*}}.$$

Then we set

$$CSP_j := \lambda S_j^{i^*} + (1-\lambda) S_j^{i'}.$$

In some cases, in particular when implementing a small number of sample paths, anomalies may occur. The interpolated value CSP_j may be higher than CSP_{j+1} and hence produces a modelled free boundary which is not monotone. To ensure this is not the case we actually set

$$CSP_j := \eta \max\left\{\eta CSP_{j+1}, \eta CSP_j\right\}.$$

Further, if $\lambda \notin [0,1]$, then we set

$$CSP_j = CSP_{j+1}.$$

Next j

Recall the initial dispersion technique suggested by Rasmussen [2005] which we discussed in §3.2. If we do not implement this technique, the indices i^* and i' can very often not be found (the code fails) close

to time t_0. This is because the initial spot price S_0 is not close enough to the optimal stopping boundary and hence none of the simulated paths will cross this boundary. Even when S_0 is close to what appears to be the probable boundary, this problem can still occur. Also, even if it doesn't, the modelled optimal stopping boundary is not as smooth as when using the initial dispersion technique. This can be seen in Figure 5.6. Finally, when we implement the initial dispersion technique, the modelled exercise boundary starts at t_0 and not at t_1.

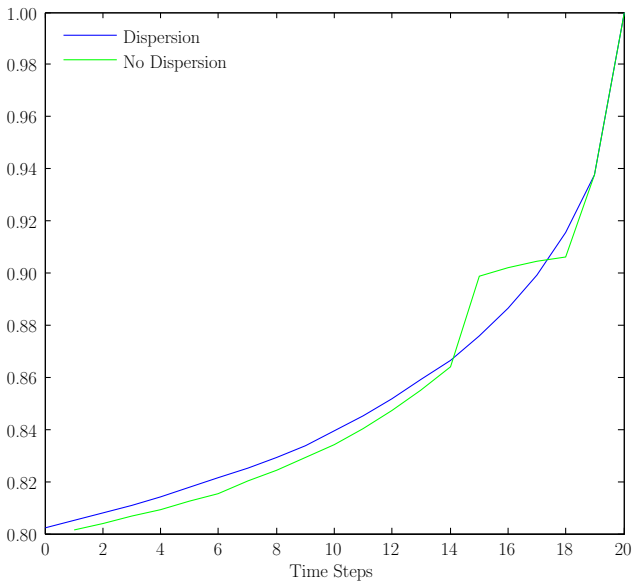

Figure 5.6: *The modelled optimal stopping boundary using the LSM-Rasmussen method. The boundary in blue is found using the initial dispersion technique whereas the other in green is found by choosing a suitable S_0/K.*

The option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 20\%$, $S_0 = 104.25$, $r = 5\%$ and $q = 2\%$; and we are considering a 1 year put option with $K = 130$. We generated 4096 sample paths with 20 time steps using techniques we consider in §7.4 and §7.5.

Note the smoother boundary, found using the initial dispersion technique, compared to the one found by choosing a suitable S_0/K.

In Figure 5.7 we see how the modelled optimal stopping boundary improves when increasing the number of sample paths. When considering both figures we note that even the modelled optimal stopping boundary, created with dispersion with the lowest number of sample paths, performs as well (if not better) as the boundary created without it.

Once we have the critical stock prices, we may calculate the low bias estimate $\breve{V}$ as follows. We simply run samples until we hit the boundary and discount the payoff. This is computationally very fast.

- Generate a new sample of N' stock price paths $\breve{S}_j^i$ for times t_j. Here the sample size $N' \gg N$.
- Set $\breve{V} := 0$.
- **For** $i = 1$ **To** N'

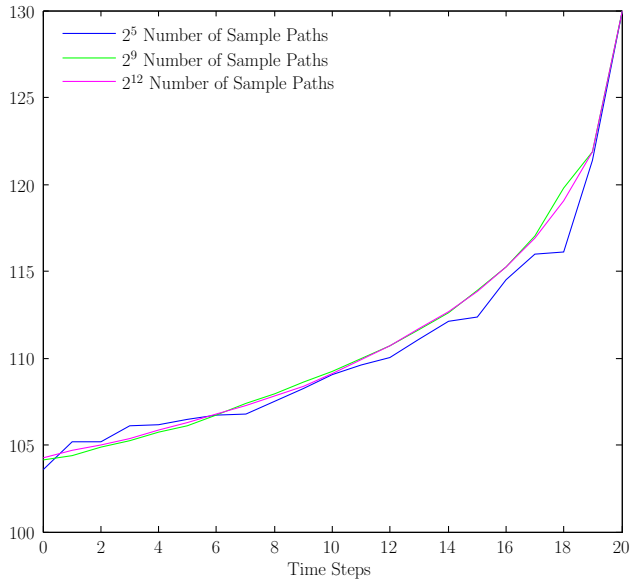

Figure 5.7: *The modelled optimal stopping boundary using the LSM-Rasmussen method as a function of the number of sample paths. These boundaries are found using the initial dispersion technique.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 20\%$, $r = 5\%$ *and* $q = 2\%$*; and we are considering a 1 year American put option with* $K = 130$. *We generated simulation paths with 20 time steps using techniques we discuss in §7.4 and §7.5.*

Note how the modelled boundary improves when increasing the number of sample paths.

- **While** $\eta\left(\breve{S}_j^i - CSP_j\right) < 0$ **Do**
  Set $j := j + 1$.
  End While
- Set $\breve{V} := \breve{V} + e^{-r(t_j - t_0)} I_j^i$, where $I_j^i := I_j\left(\breve{S}_j^i\right)$.

Next i

- Set $\breve{V} := \frac{1}{N'}\breve{V}$.

5.2 The Dual Method

In this section we refer to Glasserman [2004, §8.7] and Andersen and Broadie [2004]. Suppose that $\tilde{M} = \left\{\tilde{M}_j\right\}_{t=t_j}$ is a martingale process with $\tilde{M}_0 = 0$. Furthermore, let τ be a bounded stopping time in $\mathcal{T}_{0,M}$. Then, from the Optional Sampling Theorem

$$\begin{aligned}\mathbb{E}\left[e^{-r(\tau - t_0)} I_\tau \middle| \mathcal{F}_0\right] &= \mathbb{E}\left[e^{-r(\tau - t_0)} I_\tau - \tilde{M}_\tau \middle| \mathcal{F}_0\right] \\ &\leq \mathbb{E}\left[\max_{j=0,\ 1,\ \ldots,\ M}\left\{e^{-r(t_j - t_0)} I_j - \tilde{M}_j\right\} \middle| \mathcal{F}_0\right].\end{aligned}$$

Therefore by taking the infimum over all martingales $\tilde{M}$ with initial value 0 we have

$$\mathbb{E}\left[e^{-r(\tau-t_0)}I_\tau\middle|\mathcal{F}_0\right] \leq \inf_{\tilde{M}}\mathbb{E}\left[\max_j\left\{e^{-r(t_j-t_0)}I_j-\tilde{M}_j\right\}\middle|\mathcal{F}_0\right].$$

Since the above inequality holds for an arbitrary stopping time τ, it also holds for the supremum over all stopping times and hence from (1.1) it follows that

$$V_0=\sup_{\tau\in\mathcal{T}_{0,M}}\mathbb{E}\left[e^{-r(\tau-t_0)}I_\tau\middle|\mathcal{F}_0\right] \leq \inf_{\tilde{M}}\mathbb{E}\left[\max_j\left\{e^{-r(t_j-t_0)}I_j-\tilde{M}_j\right\}\middle|\mathcal{F}_0\right]. \tag{5.1}$$

The expression on the right hand side of the inequality in (5.1) is the *dual problem* and the inequality itself is referred to as the *duality gap* [see Andersen and Broadie, 2004].

It is possible to find a martingale using the Doob decomposition[3] [Shiryaev, 1996, §1.7, Theorem 2] such that (5.1) holds with equality; thus the duality gap is 0: Note that since $V_j = \max\left\{I_j, e^{-r\Delta t_{j+1}}\mathbb{E}\left[V_{j+1}\middle|\mathcal{F}_j\right]\right\}$ it follows that $V_j \geq e^{-r\Delta t_{j+1}}\mathbb{E}\left[V_{j+1}\middle|\mathcal{F}_j\right]$. Thus the discounted process $\left\{e^{-r(t_j-t_0)}V_j\right\}_{t=t_j}$ is a supermartingale. By the Doob decomposition there exist a martingale $\tilde{M}^*$ and a decreasing adapted process A such that $e^{-r(t_j-t_0)}V_j = V_0+\tilde{M}^*_j+A_j$ with $\tilde{M}^*_0=0=A_0$. Since $A_j\leq 0$ and $I_j\leq V_j$ for all j, it follows that $e^{-r(t_j-t_0)}I_j-\tilde{M}^*_j = V_0+e^{-r(t_j-t_0)}(I_j-V_j)+A_j\leq V_0$ for all j and hence

$$\inf_{\tilde{M}}\mathbb{E}\left[\max_j\left\{e^{-r(t_j-t_0)}I_j-\tilde{M}_j\right\}\middle|\mathcal{F}_0\right] \leq \mathbb{E}\left[\max_j\left\{e^{-r(t_j-t_0)}I_j-\tilde{M}^*_j\right\}\middle|\mathcal{F}_0\right] \leq V_0.$$

Glasserman [2004, §8.7] remarks that obtaining this martingale, which is referred to as the *optimal martingale*, is as difficult to find as the original optimal stopping time. However, if one can find a martingale which is close to the optimal martingale, we have an estimate of an upper bound of the American option price [Glasserman, 2004, §8.7].

Suppose that the stopping times $\tau_0,\ \tau_1,\ \ldots,\ \tau_M$ are specified via a critical stock price function. Here τ_j means the exercise time which is optimal amongst those times τ satisfying $\tau\geq t_j$ [4]. Then define

$$\Delta_j := \mathbb{E}\left[e^{-r(\tau_j-t_j)}I_{\tau_j}\middle|\mathcal{F}_j\right]-\mathbb{E}\left[e^{-r(\tau_j-t_j)}I_{\tau_j}\middle|\mathcal{F}_{j-1}\right]. \tag{5.2}$$

By the tower property $\mathbb{E}\left[\Delta_j\middle|\mathcal{F}_{j-1}\right]=0$ and hence we can define a martingale $\tilde{M}$ by $\tilde{M}_0:=0$ and $\tilde{M}_j:=\tilde{M}_{j-1}+\Delta_j$. Now

$$\mathbb{E}\left[e^{-r(\tau_j-t_j)}I_{\tau_j}\middle|\mathcal{F}_j\right] = \begin{cases} I_j & \text{if we stop at time } t_j, \text{ that is, } \tau_j=t_j;\\ \mathbb{E}\left[e^{-r(\tau_{j+1}-t_j)}I_{\tau_{j+1}}\middle|\mathcal{F}_j\right] & \text{if we do not stop at } t_j, \text{ that is, } \tau_j>t_j.\end{cases} \tag{5.3}$$

Now if $\tau_j>t_j$, then $\tau_j=\tau_{j+1}$ and hence if we can estimate the expression

$$\mathbb{E}\left[e^{-r(\tau_{j+1}-t_j)}I_{\tau_{j+1}}\middle|\mathcal{F}_j\right]$$

for $j=0,\ 1,\ \ldots,\ M-1$, then using (5.3) we can estimate Δ_j as in (5.2).

[3] Note that one requires an application of the the more complicated Doob-Meyer decomposition in the continuous setting.

[4] Clearly the optimal τ_j is unknown, but may be approximated by $TV_j=\min\{k\geq j: I_k\geq TV_k\}$, where TV_k indicates the modelled continuation value.

We provide an algorithm for calculating an estimation $\hat{V}$ of the dual value $\inf_{\tilde{M}} \mathbb{E}\left[\max_j \left\{e^{-r(t_j-t_0)}I_j - \tilde{M}_j\right\} \middle| \mathcal{F}_0\right]$:

Suppose we have already determined a vector of critical stock prices *CSP* from a sample of N as in §5.1. Generate a new sample of stock prices $\hat{S}_j^i$, $i = 1, 2, \ldots, N''$ where $N'' \ll N$. The algorithm will require the generation of subpaths from each $\hat{S}_j^i$ [5] and hence we choose N'' to be small, and the number of such subpaths to be a very small number P. Glasserman [2004, §8.7, p.476] tests the dual method for $P = 10$ and $P = 100$. In the following section in Figure 5.8, we plot the performance of the dual method for $P = 10$, $P = 100$ and $P = 1000$.

- Set $\hat{V} := 0$.
- **For** $i = 1$ **To** N''
  - $V_{\max}$ will be the running maximum of the $e^{-r(t_j-t_0)}I_j - \tilde{M}_j$, so initialise $V_{\max} := 0$.
  - **For** $j = 0$ **To** $M - 1$
    - Simulate P subpaths from $\hat{S}_j^i$. Each subpath runs until we reach the stopping time τ_j, that is, follow the exercise policy τ_j determined by the critical stock price function[6].
    - Record the payoffs and calculate the average, that is, calculate the estimate $\mathbb{E}\left[e^{-r(\tau_{j+1}-t_j)}I_{\tau_{j+1}} \middle| \mathcal{F}_j\right]$.
    - Also determine $\mathbb{E}\left[e^{-r(\tau_j-t_j)}I_{\tau_j} \middle| \mathcal{F}_j\right]$ using (5.3).
    - Calculate Δ_j using (5.2). Also set $\tilde{M}_j := 0$ if $j = 0$ and $\tilde{M}_j := \tilde{M}_{j-1} + \Delta_j$ for $j > 0$.
    - Set $V_{\max} := \max\left\{V_{\max}, e^{-r(t_j-t_0)}I_j - \tilde{M}_j\right\}$.

    Next j
  - Set $\hat{V} := \hat{V} + V_{\max}$.

  Next i
- Set $\hat{V} := \frac{1}{N''}\hat{V}$.

5.3 Results

In Figure 5.8 we plot the performance of the dual method for varying number of subpaths. The results improve as the number of subpaths increases. However, the increase in computational time from using $P = 100$ to $P = 1000$ subpaths and only a small improvement in the outcome suggest that using $P = 100$ subpaths is sufficient. These dual values and their corresponding times taken are provided Table 5.1 where the number of samples paths is 512. In this example we obtained the critical stock price function from the LSM-Rasmussen method. Similar results were obtained when deriving the critical stock price function from the stochastic mesh method.

[5] By *subpath* we mean a path starting at $\hat{S}_j^i$ at time t_j and stopped according to τ_j.

[6] We cannot make use of quasi-random numbers (where bridging is included), since the number of time steps until we reach the stopping time is unknown in advance. The reason for this will become clearer once Sobol' random numbers have been discussed in §7.5. Thus, even if we use quasi-random numbers to generate $\hat{S}_j^i$, these subpaths will be generated using pseudo-random numbers.

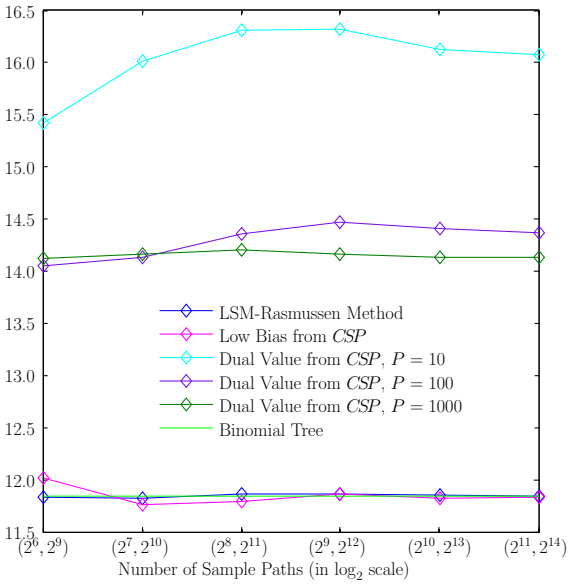

Figure 5.8: *We plot the performance of the LSM-Rasmussen method as a function of the number of sample paths using 8 Laguerre polynomials. We also plot the low bias obtained from the critical stock price function calculated in the LSM method. Furthermore, we plot the performance of the dual method with subpaths the number of subpaths* $P = 10$, $P = 100$ *and* $P = 1000$ *which is calculated using this critical stock price function. We also include the value given by a 30 step binomial tree. The number of sample paths used in the high bias calculations and LSM method is given by the first entry in the bracket and that of the low bias estimate in the second on the horizontal axis.*

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$. *Sample paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5.*

Clearly increasing the number of subpaths improves the results produced by the dual method. However, given the increase in computation time as P *increases and the fact that the difference between the outcome for* $P = 100$ *and* $P = 1000$ *is small, we conclude that choosing* $P = 100$ *is sufficient.*

P	Dual Value	Time (seconds)
10	16.324	17.289
100	14.471	173.895
1000	14.159	1488.003

Table 5.1: *Values and time taken when applying the dual method for varying subpaths. These values correspond to those presented in Figure 5.8 where the number of sample paths is given by 512.*

In Figures 5.9 and 5.10, on the left, we plot the performance of the dual method where we have obtained a critical stock price function from the LSM-Rasmussen method and the stochastic mesh method. We also

include the low bias estimate obtained from the critical stock price function and the value given by a binomial tree.

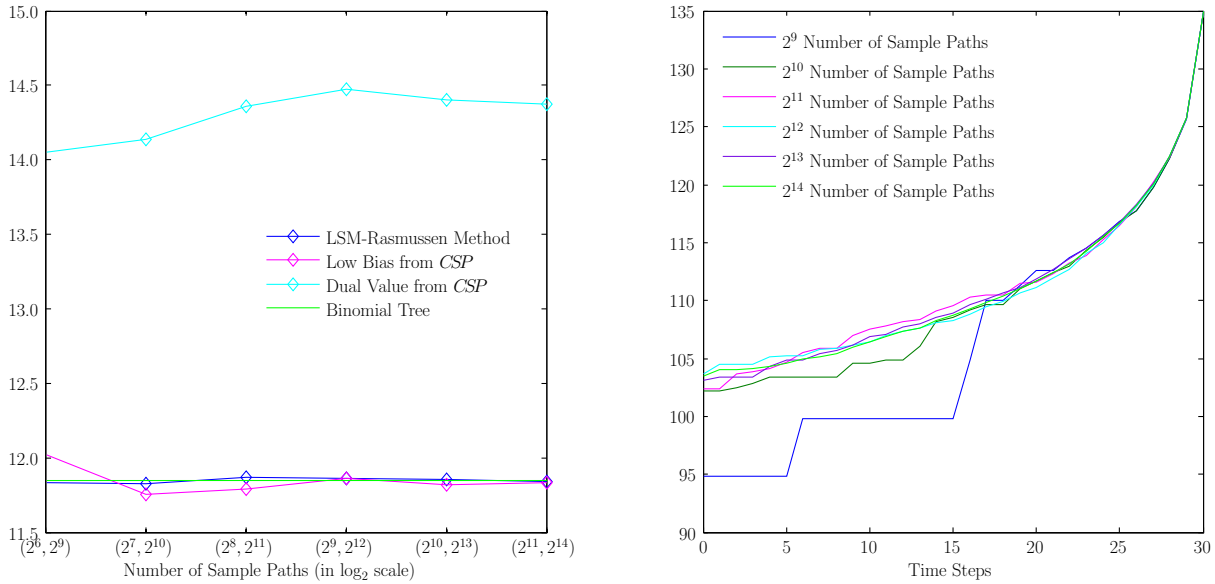

Figure 5.9: *On the left we plot the performance of the LSM-Rasmussen method as a function of the number of sample paths using 8 Laguerre polynomials. We also plot the low bias obtained from the critical stock price function calculated in the LSM method. Furthermore, we plot the performance of the dual method which is calculated using this critical stock price function with* $P = 100$. *We also include the value given by a 30 step binomial tree. The number of sample paths used in the high bias calculations and LSM method is given by the first entry in the bracket and that of the low bias estimate in the second on the horizontal axis.*

On the right we plot the corresponding critical stock price function determined from the LSM-Rasmussen method for a varying number of sample paths.

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$; *and we are considering a 1 year put option with* $K = 135$. *Sample paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5.*

Furthermore, in Figures 5.9 and 5.10, on the right, we plot the corresponding critical stock price functions for a varying number of sample paths. Note how the critical stock price function becomes smoother as the number of sample paths increases in the LSM case. However, the critical stock price function does not perform well in the stochastic mesh method.

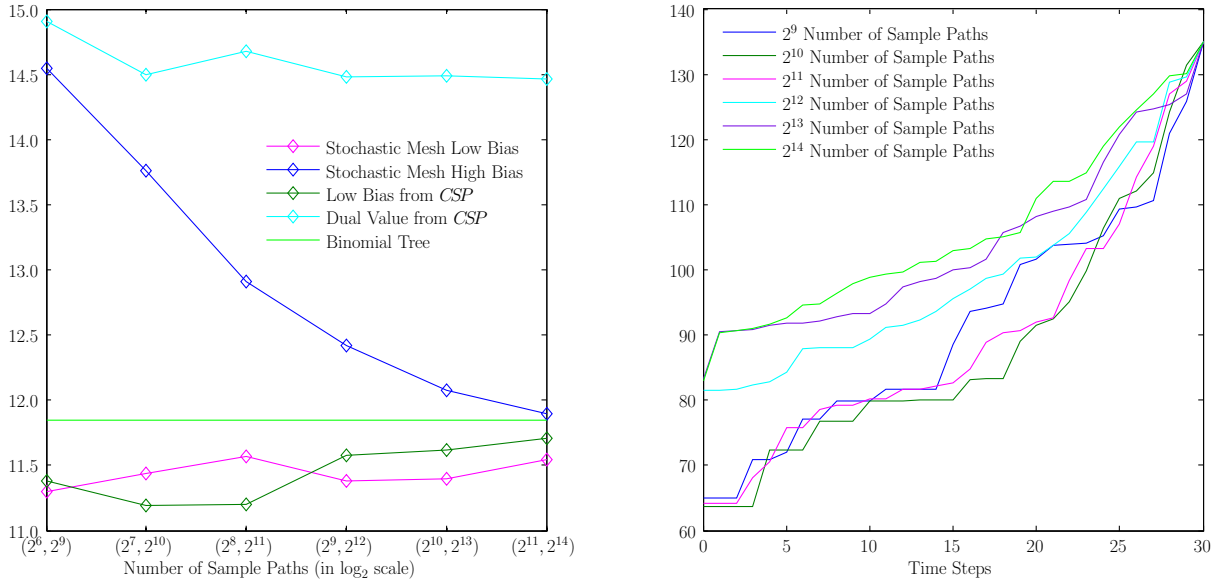

Figure 5.10: *On the left we plot the performance of the high and low bias values using the stochastic mesh method. We also plot the low bias obtained from the critical stock price function determined by using the stochastic mesh. Furthermore, we plot the performance of the dual method which is calculated using this critical stock price function with* $P = 100$. *We also include the value given by a 30 step binomial tree. Clearly the low and high bias determined from the critical stock price function and dual method perform poorly compared to the original stochastic mesh method. The number of sample paths used in the high bias calculations is given by the first entry in the bracket and that of the low bias estimate in the second on the horizontal axis.*

On the right we plot the corresponding critical stock price function determined from the stochastic mesh method for a varying number of sample paths.

Here the option details are as follows: the underlying stock price process follows geometric Brownian motion with $\sigma = 30\%$, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$; *and we are considering a 1 year put option with* $K = 135$. *Sample paths were generated with 30 time steps using techniques we discuss in §7.4 and §7.5.*

Part II

Price Processes under Exponential Lévy Models

Chapter 6

Introduction to Lévy Processes

Lévy processes, named after the French mathematician Paul Lévy, form an important class of stochastic processes. Lévy processes allow for jumps: as shown in the Lévy-Itô decomposition, Lévy processes can be decomposed into drift, diffusion and jump parts. Brownian motion falls under the class of Lévy processes — any continuous Lévy process must be a Brownian motion with drift. Lévy processes are both semi-martingales and Markovian, and they possess many well understood probabilistic and analytical properties, making them attractive as mathematical tools. There are several references on Lévy processes — Sato [1999], Applebaum [2004], Kyprianou [August 2007] and Cont and Tankov [2004a] to name but a few.

Lévy models were introduced into mathematics of finance in the 1980s and 1990s and have become increasingly popular in this field. This is because when using Lévy processes, one is able to capture distributional characteristics in the stock returns such as fat tails and asymmetry. Furthermore, they allow for jumps in the price process, which can be interpreted as shocks in the market, and effects due to trading taking place in 'business' time rather than 'real' time. Thus Lévy processes might describe the observed reality of financial markets more accurately than Brownian motion (the traditional Black-Scholes world)[1].

Further arguments in favour of the application of Lévy processes in finance result from the fact that they are stationary with independent increments[2]. When modelling price processes in finance, one usually requires the corresponding returns processes to be stationary. Certainly occasions arise in which stationarity is undesirable (such as business cycles or changes in risk aversion), but this requirement serves as a good first approximation. Also, in order to incorporate the efficient market hypothesis, returns processes are required to have future returns independent of past returns. Thus returns must have independent increments. Note that, if in addition to the above conditions, we require the price process to be continuous, then the returns process must be arithmetic Brownian motion (by the Lévy-Itô decomposition which is discussed in §6.5).

For the purpose of this thesis, we will only consider one-dimensional Lévy processes, even though in many cases the multidimensional process can be defined. However, when a definition only applies to one-dimensional Lévy processes, we will explicitly state this. We will present an intuitive account of Lévy

[1] This is the opinion of the school of Madan and Carr, however not of Dupire for example.
[2] That is, if we require Lévy processes to be almost surely continuous, as in Definition 6.1.1.

processes, focussing on the meaning and importance of results, but omitting technical details.

6.1 Definition of Lévy Processes

A Lévy process is any càdlàg continuous-time stochastic process that has stationary, independent increments (see Sato [1999, Definition 1.6], Applebaum [2004, §1.3] and Cont and Tankov [2004a, Definition 3.1] for example).

Definition 6.1.1 Lévy Process
Let $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space. Then a càdlàg stochastic process* $X = \{X_t\}_{t\geq 0}$ *on* $(\Omega, \mathcal{F}, \mathbb{P})$ *with values in* $\mathbb{R}$ *is called a* Lévy process *if*

(i) $X_0 = 0$ *a.s.;*

(ii) for any $0 \leq s < t \leq T$, $X_t - X_s$ *is independent of* $\mathcal{F}_s$ *(*independent increments*);*

(iii) for any $s, t \geq 0$, *the law of* $X_{t+s} - X_t$ *is independent of* t *(*stationary increments*);*

(iv) for every $\varepsilon > 0$ *and* $s, t \geq 0$, $\lim_{s\to t} \mathbb{P}(|X_t - X_s| \geq \varepsilon) = 0$ *(*convergence in probability*).*

Brownian motion and Poisson processes are probably the most well-known examples of Lévy processes; there are many other examples such as arithmetic Brownian motion, compound Poisson, variance gamma and normal inverse Gaussian processes.

In the definition above, the càdlàg property is a technical requirement which ensures that the paths of X do not explode. Conditions (ii) and (iii) characterise Lévy processes from a modelling standpoint. They show that for any $t > s$, the distribution of $X_t - X_s$ is dependent only on the time interval $t - s$ and that $X_t - X_s$ is independent of $\{X_u\}_{u\leq s}$. Furthermore, conditions (ii) and (iii) enable us to derive the infinite divisibility of Lévy processes which we discuss in the following section. Condition (iv) in no way implies that the individual sample paths of X are necessarily continuous. Stochastic continuity does not preclude large jumps. That is, the probability of knowing that a large jump occurs at a given time t is 0, and thus processes whose large jumps occur at fixed times are excluded. Note that if stochastic continuity is extended to path continuity, then the resulting process is arithmetic Brownian motion.

As noted by Kou [2001], "The main empirical motivation of using Lévy processes in finance comes from fitting asset return distributions." He mentions that Lévy processes are able to capture the leptokurtic feature so often seen in financial asset prices. Furthermore, when using Lévy processes to model returns (as we will later in this thesis), the independent increments can be interpreted as future returns independent of the past. Thus these conditions could be seen as implying market efficiency, i.e. that current asset prices reflect all publicly available information. However, independent increments, along with the stationarity property, prevent Lévy processes from modelling volatility clustering which is another phenomenon of return distributions. Volatility clustering may be addressed by considering models that combine Lévy processes with others, but we will not consider such models here.

6.2 Infinite Divisibility

In this section we frequently refer to characteristic functions — see Appendix C.1 for definitions, properties and examples.

There exists a one-to-one correspondence between the class of infinitely divisible distributions and the class of Lévy processes.

Definition 6.2.1 Infinite Divisibility
A probability distribution μ on $\mathbb{R}$ is called infinitely divisible *if for any $n \in \mathbb{N}$, there exists n i.i.d. random variables $X_1,\ X_2,\ \ldots,\ X_n$ such that $X_1 + X_2 + \ldots + X_n$ has distribution μ (see Applebaum [2004, §1.2.2] or Cont and Tankov [2004a, Definition 3.2]).*

Equivalently, the distribution μ_X of a random variable X is infinitely divisible if for every $n \in \mathbb{N}$ there exist n i.i.d. random variables $X_1,\ X_2,\ \ldots,\ X_n$ such that

$$X \overset{\mathcal{D}}{=} X_1 + X_2 + \ldots + X_n.$$

If the distribution of a random variable X is infinitely divisible, we will sometimes say X is infinitely divisible for short. Note that the distribution of X is given by the convolution of the distributions of $X_1,\ X_2,\ \ldots,\ X_n$ [see Varadhan, 2001, §3.1, p.36].

Every Lévy process is infinitely divisible. To see this, observe that a Lévy process is a continuous-time analogue of a random walk (see eg. Cont and Tankov [2004a, §3.1] or Kyprianou [August 2007]): suppose we sample a Lévy process X at intervals $0,\ \Delta,\ 2\Delta,\ \ldots$ and let $S_n(\Delta) = X_{n\Delta}$. Then

$$S_n(\Delta) = \sum_{i=1}^{n} \left(X_{i\Delta} - X_{(i-1)\Delta} \right),$$

where, for $i = 1,\ 2,\ \ldots,\ n$ $X_{i\Delta} - X_{(i-1)\Delta}$, are i.i.d. random variables having the same distribution as the random variable X_Δ. So if we sample a Lévy process at different intervals Δ, then we get a family of random walks $S_n(\Delta)$. Now, let $t = n\Delta$, then for every $t > 0$ and $n \geq 1$ we have that $X_t = S_n(\Delta)$. Then as before

$$S_n(\Delta) = \sum_{i=1}^{n} \left(X_{i\frac{t}{n}} - X_{(i-1)\frac{t}{n}} \right)$$

where, for $i = 1,\ 2,\ \ldots,\ n$ $X_{i\frac{t}{n}} - X_{(i-1)\frac{t}{n}}$, are i.i.d. random variables whose distribution is equal to that of $X_{\frac{t}{n}}$. Therefore, if X is a Lévy process, then for any $t > 0$, X_t is infinitely divisible. We have shown that every Lévy process is infinitely divisible. Conversely, given an infinitely divisible distribution μ there exists a unique Lévy process [see Sato, 1999, Corollary 11.6] X_1 with the property that $X_1 \overset{\mathcal{D}}{=} \mu$.

The characteristic function of an infinitely divisible distribution, and hence a Lévy process, has a useful form, as is shown in the next proposition. Here, and in the rest of the thesis, we denote the characteristic function of a random variable X by Φ_X.

Proposition 6.2.2 Characteristic Function of a Lévy Process
Suppose $\{X_t\}_{t\geq 0}$ is a Lévy process on $\mathbb{R}$. Then there exists a continuous function $\psi : \mathbb{R} \to \mathbb{C}$ such that

$$\Phi_{X_t}(z) = e^{t\psi(z)}$$

for $z \in \mathbb{R}$ [Sato, 1999, Lemma 7.6].

The function ψ is called the *characteristic exponent*, the *Lévy symbol* or the *Lévy exponent* of X. As we will see when discussing the Lévy-Khintchine representation (see Corollary 6.5.2), much can be said about the form of ψ. The law of X is characterised by ψ in the sense that if two Lévy processes have the same characteristic exponent, they have the same law. Thus we conclude from Proposition 6.2.2 that the law of X_1 determines the law of X_t.

An outline of the proof for Proposition 6.2.2 is as follows: let X be a Lévy process. Then X_1 is infinitely divisible and hence $\Phi_{X_1}(z) \neq 0$ for any $z \in \mathbb{R}$ [see Sato, 1999, Lemma 7.5]. Since every characteristic function is continuous, there exists a unique continuous function ψ such that $\Phi_{X_1}(z) = e^{\psi(z)}$ [see Sato, 1999, Lemma 7.6]. Suppose that $p, q \in \mathbb{N}$, then

$$X_1 = \sum_{j=1}^{q} \left(X_{\frac{j}{q}} - X_{\frac{j-1}{q}} \right)$$

which shows that $\Phi_{X_{\frac{1}{q}}}(z)^q = \Phi_{X_1}(z)$ and hence $\Phi_{X_{\frac{1}{q}}}(z) = e^{\frac{1}{q}\psi(z)}$ if we take q^{th} roots. Now given a rational $t = \frac{p}{q}$,

$$X_{\frac{p}{q}} = \sum_{k=1}^{p} \left(X_{\frac{k}{q}} - X_{\frac{k-1}{q}} \right)$$

so that $\Phi_{X_{\frac{p}{q}}}(z) = \Phi_{X_{\frac{1}{q}}}(z)^p = e^{\frac{p}{q}\psi(z)}$. This then shows that for a rational t, $\Phi_{X_t}(z) = e^{t\psi(z)}$. If we let $t \in \mathbb{R}^+$ and t_n be a sequence of rational numbers such that $t_n \to t$, then $X_{t_n} \to X_t$ in probability and hence in distribution. So we have that $\Phi_{X_{t_n}}(z) \to \Phi_{X_t}(z)$ and thus $\Phi_{X_t}(z) = e^{t\psi(z)}$. More rigorous arguments for the above can be found in Sato [1999, §7].

6.3 Poisson and Compound Poisson Processes

Poisson and compound Poisson processes are crucial to the understanding of Lévy processes. In particular, there is a close relationship between compound Poisson processes and Lévy processes as we will see in Proposition 6.3.5. Compound Poisson processes can be seen as a superposition of independent Poisson processes. In turn, the Poisson distribution is closely connected to the exponential distribution which we consider next.

Exponential random variables are used to describe the times between events in a Poisson process. An exponentially distributed random variable τ can take any nonnegative real value and has probability density function, with parameter $\lambda \in \mathbb{R}^+$ given by

$$f(t) = \lambda e^{-\lambda t} \mathbb{1}_{[0,\infty)}(t). \tag{6.1}$$

τ is called the *first arrival time*. Exponential random variables posses a property called the *absence of memory*. That is, knowing the time the last event occurred is in no way helpful in predicting the time of the next event. If τ is an exponential random variable, then using Bayes' rule

$$\mathbb{P}(\tau > t + s | \tau > t) = \frac{\int_{t+s}^{\infty} \lambda e^{-\lambda y} dy}{\int_{t}^{\infty} \lambda e^{-\lambda y} dy} = \frac{e^{-\lambda(t-s)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbb{P}(\tau > s),$$

for every $t, s > 0$. It is this memoryless property that allows the exponential distribution to be a favourable model for inter-arrival times of events. A distribution with this property must be an exponential distribution as shown in the next proposition [Cont and Tankov, 2004a, Proposition 2.8]:

Proposition 6.3.1 Characterisation of Exponential Distributions
Let $\tau \geq 0$ be a random variable. If for all $s, t > 0$

$$\mathbb{P}(\tau > t + s | \tau > t) = \mathbb{P}(\tau > s)$$

then τ has an exponential distribution.

In the following we define a Poisson process N_t that counts the number of events or random times $\{T_k\}_{k\geq 1}$ occurring between 0 and t. In this definition we define the Poisson process as a counting process of random times where each random time is the sum of i.i.d. exponential random variables [Cont and Tankov, 2004a, Definition 2.17].

Definition 6.3.2 Poisson Process
Suppose that $\{\tau_i\}_{i\geq 1}$ are independent exponential random variables with parameter λ and that $T_k = \sum_{i=1}^{k} \tau_i$. Then the process $\{N_t\}_{t\geq 0}$ defined by

$$N_t = \sum_{k=1}^{\infty} \mathbb{1}_{\{T_k \leq t\}}$$

is called a Poisson process *with intensity λ.*

In the definition above, the $\{\tau_i\}_{i\geq 1}$ are the inter-arrival times, whereas $\{T_k\}_{k\geq 1}$ are the arrival times. Thus N_t counts the total number of arrivals that have occurred in time t. Furthermore, as shown in the next proposition, the random variable N_t follows a Poisson distribution [Cont and Tankov, 2004a, Proposition 2.12].

Proposition 6.3.3
A Poisson process N_t as defined above follows a Poisson distribution with parameter λt, that is,

$$\mathbb{P}(N_t = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

Note that a Poisson process $N = \{N_t\}_{t\geq 0}$ is not a martingale. However, the process $\widetilde{N} = \left\{\widetilde{N}_t\right\}_{t\geq 0}$ defined by

$$\widetilde{N}_t := N_t - \lambda t \tag{6.2}$$

is a martingale. $\widetilde{N}$ is called a *compensated Poisson process* and $\{\lambda t\}_{t\geq 0}$ is called the *compensator* of N.

Instead of having jumps of size 1, as in the case of the Poisson process, the compound Poisson process has jumps of random size which are independent random variables with the same distribution [see Cont and Tankov, 2004a, Definition 3.3].

Definition 6.3.4 Compound Poisson Process
A stochastic process $X = \{X_t\}_{t\geq 0}$, with intensity $\lambda > 0$ and jump size distribution f, defined by

$$X_t = \sum_{k=1}^{N_t} Y_k = \sum_{k=1}^{\infty} Y_k \mathbb{1}_{\{T_k \leq t\}},$$

where jump sizes Y_k are i.i.d. random variables with distribution f and $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ independent of $\{Y_k\}_{k\geq 1}$, is called a compound Poisson process.

As noted by Cont and Tankov [2004a, p.71], the following conclusions can be made from the above:

(i) Each time N_t jumps by 1, X_t jumps by a random size which has distribution f. The jumps at different instances are independent and X_t inherits the independent and stationary increments property from N_t. X_t jumps only when N_t does and hence the sample paths of X_t are càdlàg since the sample paths of N_t are. Also, X_t is piecewise constant because N_t is.

(ii) The jump times $\{T_k\}_{k\geq 1}$ have the same distribution as the jump times of N_t.

A compound Poisson process is a Lévy process with a.s. piecewise constant sample paths as shown in the next proposition:

Proposition 6.3.5
A process X is a compound Poisson process if and only if it is a Lévy process and its sample paths are piecewise constant functions [Cont and Tankov, 2004a, Proposition 3.3].

As a consequence Lévy processes, in general, can be adequately approximated by compound Poisson processes.

6.4 Jump and Lévy Measures

In this section we discuss jump and Lévy measures and how they relate to each other. In order to aid us in our discussion on jump measures, we consider random measures first.

Let $(E, \mathcal{E})$ be a measurable space and $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Then the function $M : \Omega \times \mathcal{E} \to \mathbb{R}$ is called a *random measure* on $(E, \mathcal{E})$ if and only if

(i) for almost all $\omega \in \Omega$, $M(\omega, \cdot)$ is a Radon measure[3] on $(E, \mathcal{E})$.

(ii) for each measurable set $A \in \mathcal{E}$, $M(\cdot, A)$ is a random variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

A random measure is called a *point process* if it is integer-valued. Now, a Poisson random measure is a point process that has independent increments. We say that a random measure M has *independent increments* if for disjoint measurable sets $A_1, A_2, \ldots, A_n$, then $M(\cdot, A_1), M(\cdot, A_2), \ldots, M(\cdot, A_n)$ are independent random variables.

Definition 6.4.1 Poisson Random Measure
Suppose that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and μ is a positive Radon measure on $(E, \mathcal{E})$, where $E \subseteq \mathbb{R}$. A Poisson random measure *on E with intensity measure μ is a function*

$$M : \Omega \times \mathcal{E} \to \mathbb{N}$$

such that

[3] If $E \subset \mathbb{R}$, then a Radon measure on $(E, \mathcal{E})$ is a measure μ such that for every bounded closed subset $A \in \mathcal{E}$, $\mu(A) < \infty$.

(i) for almost all $\omega \in \Omega$, $M(\omega, \cdot)$ is an integer-valued Radon measure on E.

(ii) for each measurable set $A \in \mathcal{E}$, $M(\cdot, A)$ is a Poisson random variable with parameter $\mu(A)$. Thus for every $A \in \mathcal{E}$ and every $n \in \mathbb{N}$

$$\mathbb{P}\left(M(\cdot, A) = n\right) = e^{-\mu(A)} \frac{\mu(A)^n}{n!}.$$

(iii) for disjoint measurable sets $A_1, A_2, \ldots, A_n \in \mathcal{E}$ variables $M(\cdot, A_1), M(\cdot, A_2), \ldots, M(\cdot, A_n)$ are independent.

[see Cont and Tankov, 2004a, Definition 2.18].

A random measure can be associated with every càdlàg process, in particular a compound Poisson process. Let $X = \{X_t\}_{t \geq 0}$ be a càdlàg process with values in $\mathbb{R}$ and let the jumps of X be denoted by

$$\Delta X_t = X_t - X_{t_-} \neq 0$$

where $X_{t_-} = \lim_{s \to t, s<t} X_s$ exists by the càdlàg property. Then define the *random jump measure* $J_X : \Omega \times \mathcal{B}\left(\mathbb{R}^+\right) \times \mathcal{B}\left(\mathbb{R}\right) \to \mathbb{N}$ associated with X as

$$\begin{aligned} J_X(\omega, (0,t], A) &= \#\left\{s : (s, \Delta X_s(\omega)) \in (0,t] \times A\right\} \\ &= \sum_{s>0} \mathbb{1}_{((0,t],A)}(s, \Delta X_s(\omega)) \end{aligned}$$

where $\omega \in \Omega$ and $A \in \mathcal{B}\left(\mathbb{R}\right)$ [4] is bounded away from 0 [5,6]. Thus J_X is a counting measure and $J_X(\cdot, (0,t], A)$ is the number of jumps of X, by time t, whose size is in A. So for any measurable set $A \subset \mathbb{R}$, the jump measure $J_X(\cdot, (t_1, t_2], A)$ of a process X, counts the number of jumps of X in the interval $(t_1, t_2]$ with jump sizes in A. The jump measure of a compound Poisson process is a Poisson random measure as shown in the next proposition.

Proposition 6.4.2 Jump Measure of a Compound Poisson Process
Suppose X is a compound Poisson process with intensity λ and jump size distribution f. Then its jump measure J_X is a Poisson random measure on $[0, \infty) \times \mathbb{R}$ with intensity measure $\mu(dt \times dx) = \nu(dx)dt$ where $\nu(dx) = \lambda f(dx)$ [Cont and Tankov, 2004a, Proposition 3.5].

Proposition 6.4.2 implies that any compound Poisson process X can be represented as

$$X_t = \sum_{s \in [0,t]} \Delta X_s = \int_0^t \int_{-\infty}^{\infty} x J_X(\cdot, ds, dx) \tag{6.3}$$

where J_X is a Poisson random measure with intensity $\nu(dx)dt$. There are no convergence problems with the integral in (6.3). This is because compound Poisson processes only have a finite number of jumps in a

[4] $\mathcal{B}\left(\mathbb{R}\right)$ denotes the Borel algebra, i.e. the σ-algebra generated by all the open subsets of $\mathbb{R}$.
[5] $A \in \mathcal{B}\left(\mathbb{R}\right)$ is bounded away from 0 if and only if 0 is not in the closure of A.
[6] The requirement for A to be bounded away from 0 is explained as follows: since the process X is càdlàg, it can have at most finitely many jumps where $|\Delta X_\cdot| > \varepsilon$ for any $\varepsilon > 0$. It may, however, have infinitely many jumps where $|\Delta X_\cdot| < \varepsilon$, in any non-zero time interval.

finite interval and hence the integral above is a finite sum. Thus compound Poisson processes are of finite activity as we shall see later in Proposition 6.6.1.

From Proposition 6.3.5, it follows that all piecewise constant Lévy processes can be defined using the jump measure. Furthermore, the ν in Proposition 6.4.2 is called the *Lévy measure* of the process X, and can be interpreted as the average number of jumps per unit time.

Definition 6.4.3
Let $X = \{X_t\}_{t\geq 0}$ be a Lévy process on $\mathbb{R}$. Then for $A \in \mathcal{B}(\mathbb{R})$ define the measure ν on $\mathbb{R}$ by

$$\nu(A) = \mathbb{E}\left[\#\{t \in [0,1] : \Delta X_t \neq 0, \Delta X_t \in A\}\right]. \tag{6.4}$$

ν is called the Lévy measure *of X. $\nu(A)$ is the expected number of jumps, per unit time, where the jump sizes are in A [Cont and Tankov, 2004a, Definition 3.4].*

From the above and Proposition 6.3.5, associated with every piecewise constant Lévy process, is a Poisson process $M(\cdot, A)$ counting the number of jumps with size in A. Thus there is a sequence of independent exponential random variables $\{T_A^n\}$, where T_A^n is the n^{th} time X has a jump of size in A and has a mean of $\frac{1}{\nu(A)}$.

One can extract useful information regarding the structure of the jumps of a Lévy process from its Lévy measure. By considering the Lévy measure of a process we are able to determine whether it has only a finite number of jumps on every time interval or infinitely many (see Proposition 6.6.1). Also, whether a Lévy process has finite variation or not is partly dependent on the Lévy measure (see Sato [1999, Theorem 21.9] or Papapantoleon [2008, Proposition 6.2]).

As we have seen, the Lévy measure describes the expected number of jumps of a certain size per unit time. More formally, a Lévy measure can be defined as a measure ν on $\mathcal{B}(\mathbb{R})$ such that

$$\nu(\{0\}) = 0 \tag{6.5}$$

and

$$\int_{\mathbb{R}\setminus\{0\}} \left(x^2 \wedge 1\right) \nu(dx) < \infty \tag{6.6}$$

(see Papapantoleon [2008] for example). ν has zero mass at the origin, but could have infinite mass near the origin. Thus infinitely many small jumps can occur close to the origin. On the other hand note that for $0 < \varepsilon \leq 1$

$$\begin{aligned}
\nu\left((-\infty, -\varepsilon] \cup [\varepsilon, \infty)\right) &= \frac{1}{\varepsilon^2} \int_{(-\infty,-\varepsilon]\cup[\varepsilon,\infty)} \varepsilon^2 \, \nu(dx) \\
&\leq \frac{1}{\varepsilon^2} \int_{(-\infty,-\varepsilon]\cup[\varepsilon,\infty)} \left(x^2 \wedge 1\right) \, \nu(dx) \\
&< \infty
\end{aligned}$$

which shows that the mass away from the origin is bounded, that is, there is only a finite number of big jumps. Some texts, such as Schoutens [2003, §5.1] or Applebaum [2004, §1.2.4], define ν on the σ-algebra $\mathcal{B}(\mathbb{R}\setminus\{0\})$ without condition (6.5). For convenience we have defined ν on $\mathcal{B}(\mathbb{R})$ by including this condition[7].

[7] This is practical, since a 'jump' of size 0, is not really a jump and therefore cannot be counted.

6.5 Lévy-Itô Decomposition

In this section, we present the Lévy-Itô decomposition, the Lévy-Khintchine representation for Lévy processes and finally the Lévy-Khintchine formula. The Lévy-Khintchine formula is frequently introduced before the Lévy-Itô decomposition (see Papapantoleon [2008], Sato [1999] or Applebaum [2004] for example). However, we follow the order in which the theorems appear in Cont and Tankov [2004a] so that the Lévy-Khintchine representation and formula is a consequence of the Lévy-Itô decomposition.

Suppose we have a standard Brownian motion with drift $\gamma t + W_t$ which is independent of a compound Poisson process (this process is a piecewise constant Lévy process by Proposition 6.3.5) $\left\{X_t^0\right\}_{t\geq 0}$, then

$$X_t = \gamma t + W_t + X_t^0$$

is another Lévy process and can be written as

$$X_t = \gamma t + W_t + \sum_{s\in[0,t]} \Delta X_s^0 = \gamma t + W_t + \int_0^t \int_{-\infty}^{\infty} x J_{X^0}(\cdot, ds, dx), \tag{6.7}$$

where J_{X^0} is a Poisson random measure on $[0,\infty)\times\mathbb{R}$ with intensity $\nu(dx)dt$ (see (6.3)) and ν is the Lévy measure given in Definition 6.4.3.

A similar decomposition as in (6.7) can be found if we begin with some Lévy process X. The Lévy-Itô decomposition implies that every Lévy process can be written as a combination of standard Brownian motion with drift and a, possibly infinite, sum of independent compound Poisson processes (see Cont and Tankov [2004a, Proposition 3.7] for example).

Theorem 6.5.1 Lévy-Itô Decomposition
Let $X = \{X_t\}_{t\geq 0}$ be a Lévy process on $\mathbb{R}$ and ν its Lévy measure as in (6.4). Then X has decomposition

$$X_t = \gamma t + \sigma W_t + \int_{|x|\geq 1}\int_0^t x J_X(\cdot, ds, dx) + \lim_{\epsilon\downarrow 0}\int_{\epsilon\leq|x|<1}\int_0^t x\left[J_X(\cdot, ds, dx) - \nu(dx)ds\right], \tag{6.8}$$

where

(i) the terms are independent;

(ii) $\gamma\in\mathbb{R}$, $\sigma\in\mathbb{R}$ with $\sigma\geq 0$ and W_t is standard Brownian motion;

(iii) the convergence in the last term is a.s. and uniform in t on $[0,T]$;

(iv) J_X is a Poisson random measure on $[0,\infty)\times\mathbb{R}$ with intensity measure $\nu(dx)dt$;

Thus for every Lévy process, there exist γ, $\sigma\in\mathbb{R}$, with $\sigma\geq 0$, and a positive measure ν that uniquely determine its distribution. The triplet (σ^2,ν,γ) is called the *characteristic triplet* or *Lévy triplet* of X.

Theorem 6.5.1 states that a Lévy process $\{X_t\}_{t\geq 0}$ can be written as

$$X_t = X_t^c + X_t^d,$$

where

$$X_t^c = \gamma t + \sigma W_t$$

is the continuous Gaussian Lévy part, and the discontinuous part

$$X_t^d = \int_{|x|\geq 1}\int_0^t xJ_X(\cdot, ds, dx) + \int_{|x|<1}\int_0^t x\left[J_X(\cdot, ds, dx) - \nu(dx)ds\right]$$

gives the sum of the jumps, where the jumps are described by the Lévy measure ν. Now as we mentioned in §6.4, ν is not necessarily finite and there can be infinitely many small jumps. Therefore, the discontinuous part X_t^d is split into two integrals. First, since $\int_{|x|\geq 1}\nu(dx) < \infty$, there is a finite number of jumps with size greater or equal to 1 [8]. As in (6.3) we may write the sum of these jumps as

$$\sum_{0\leq s\leq t}^{|\Delta X_s|\geq 1} \Delta X_s = \int_{|x|\geq 1}\int_0^t xJ_X(\cdot, ds, dx)$$

which is a compound Poisson process of finite variation. Secondly, for every $\epsilon > 0$ the integral

$$\sum_{0\leq s\leq t}^{\epsilon\leq|\Delta X_s|<1} \Delta X_s = \int_{\epsilon\leq|x|<1}\int_0^t xJ_X(\cdot, ds, dx)$$

is convergent. However to obtain convergence as $\epsilon \to 0$, we need to centre the process

$$\int_{\epsilon\leq|x|<1}\int_0^t x\left[J_X(\cdot, ds, dx) - \nu(dx)ds\right]$$

which is a compensated Poisson processes as defined in (6.2). Thus

$$\lim_{\epsilon\downarrow 0}\int_{\epsilon\leq|x|<1}\int_0^t x\left[J_X(\cdot, ds, dx) - \nu(dx)ds\right]$$

is convergent and is a martingale [see Cont and Tankov, 2004a, Proposition 2.16].

The Lévy-Khintchine representation, which gives a form of the characteristic function in terms of the characteristic triplet (σ^2, ν, γ), now follows easily [Cont and Tankov, 2004a, Theorem 3.1].

Corollary 6.5.2 Lévy-Khintchine Representation
Suppose $\{X_t\}_{t\geq 0}$ is a Lévy process on $\mathbb{R}$ with Lévy triplet (σ^2, ν, γ). Then for all $z \in \mathbb{R}$

$$\Phi_{X_t}(z) = e^{t\psi(z)} \tag{6.9}$$

with

$$\psi(z) = -\tfrac{1}{2}z^2\sigma^2 + iz\gamma + \int_{-\infty}^{\infty} e^{izx} - 1 - izx\mathbb{1}_{\{|x|\leq 1\}}\,\nu(dx). \tag{6.10}$$

In § 6.2 we saw that an infinitely divisible distribution is the distribution at time $t = 1$ of some Lévy process. This, together with Theorem 6.5.1 and Corollary 6.5.2 gives us the Lévy-Khintchine formula:

Theorem 6.5.3 Lévy-Khintchine Formula
Let μ be an infinitely divisible distribution on $\mathbb{R}$. Then for $z \in \mathbb{R}$ its characteristic function is given by

$$\Phi_\mu(z) = e^{\psi(z)} \tag{6.11}$$

[8] Note that the jump size 1 is arbitrary and can be set to any positive constant.

with

$$\psi(z) = -\tfrac{1}{2}z^2\sigma^2 + iz\gamma + \int_{-\infty}^{\infty} e^{izx} - 1 - izx\mathbb{1}_{\{|x|\leq 1\}}\,\nu(dx), \tag{6.12}$$

where $\sigma, \gamma \in \mathbb{R}$ *with* $\sigma \geq 0$ *and* ν *is a Lévy measure on* $\mathbb{R}$. *Here* ν *is called the* Lévy measure *of the distribution* μ.

Let X be a Lévy process with Lévy triplet (σ^2, ν, γ) and consider the compensated term in (6.8)

$$\int_{|x|\leq 1}\int_0^t x\,[J_X(\cdot, ds, dx) - \nu(dx)ds] = \int_{|x|\leq 1}\int_0^t xJ_X(\cdot, ds, dx) - t\int_{|x|\leq 1} x\,\nu(dx).$$

Recall that we have centred this term because the integral $\int_{|x|\leq 1}\int_0^t xJ_X(\cdot, ds, dx)$ might be infinite. Now J_X is a Poisson random measure with intensity measure $\nu(dx)dt$ and hence

$$\mathbb{E}\left[\int_{|x|\leq 1}\int_0^t xJ_X(\cdot, ds, dx)\right] = \int_{|x|\leq 1}\int_0^t x\,\nu(dx)ds.$$

Therefore, if $\int_{|x|\leq 1}|x|\,\nu(dx) < \infty$ we have that $\int_{|x|\leq 1}\int_0^t xJ_X(\cdot, ds, dx) < \infty$ and thus we do not need a compensation term in (6.8). So, in this case the Lévy-Itô decomposition of X becomes

$$\begin{aligned} X_t &= \gamma t + \sigma W_t + \int_0^t\int_{-\infty}^{\infty} xJ_X(\cdot, ds, dx) - t\int_{|x|\leq 1} x\,\nu(dx) \\ &=: \gamma_0 t + \sigma W_t + \int_0^t\int_{-\infty}^{\infty} xJ_X(\cdot, ds, dx) \end{aligned}$$

and its characteristic function is given by

$$\Phi_{X_t}(z) = e^{t\psi(z)}$$

with

$$\psi(t) = -\tfrac{1}{2}z^2\sigma^2 + i\gamma_0 z + \int_{-\infty}^{\infty} e^{izx} - 1\,\nu(dx).$$

Here $\gamma_0 = \gamma - \int_{|x|\leq 1} x\,\nu(dx)$ is called the *drift* of the Lévy process X.

6.6 Infinite Activity and Pure Jump Lévy Processes

In this section we consider infinite activity Lévy models, that is, models that have an infinite number of jumps in every interval. These models may contain a Brownian component. If they don't, they are called pure jump Lévy models. As noted by Cont and Tankov [2004a, §4.1.1], several authors [Madan, 2001b, Carr et al., 2003, Geman, 2002] have considered infinite activity models to be ideal as they are able to describe the price process at various time scales more realistically. A pure jump process is of finite (infinite) activity if the number of price jumps in any interval of time is finite (infinite).

Consider the Lévy measure ν as given in Definition 6.4.3 of a Lévy process X. If $\nu(\mathbb{R}) = \infty$, then an infinite number of small jumps are expected and X is called an infinite activity process (see Sato [1999, Theorem 21.3] or Papapantoleon [2008] for example).

Proposition 6.6.1
Let X be a Lévy process with triplet (σ^2, ν, γ).

(i) If $\nu(\mathbb{R}) < \infty$ then X has a.s. a finite number of jumps on every closed and bounded interval and we say X has finite activity.

(ii) If $\nu(\mathbb{R}) = \infty$ then X has a.s. an infinite number of jumps on every closed and bounded interval and we say X has infinite activity.

For a proof see Sato [1999, Theorem 21.3]. Some remarks regarding the proposition above should be made: If X is of infinite activity, then X has an infinite number of jumps within any finite interval and the expected number of jumps in $[0, 1]$ is not finite. Consider the following: let T_ε be the first time that X has a jump of size bigger or equal to ε. Then T_ε is an exponential random variable with mean $\frac{1}{\lambda_\varepsilon}$ where $\lambda_\varepsilon = \nu(\{x : |x| \geq \varepsilon\})$. It is clear that if $\varepsilon_1 \leq \varepsilon_2$, then $T_{\varepsilon_1} \leq T_{\varepsilon_2}$ and so $\mathbb{P}(T_\varepsilon \leq t) = 1 - e^{-\lambda_\varepsilon t}$. Since $\nu(\mathbb{R}) = \infty$, $\lambda_\varepsilon \to \infty$ as $\varepsilon \to 0$. Therefore, for any $t > 0$

$$\lim_{\varepsilon \to 0} \mathbb{P}(T_\varepsilon \leq t) = 1$$

and hence $\lim_{\varepsilon \to 0} T_\varepsilon = 0$ a.s. Therefore, X jumps a.s. before any time $t > 0$. Examples of infinite activity processes are the variance gamma and normal inverse Gaussian processes, which will be discussed in Chapters 8 and 9 respectively.

6.7 Subordinators

'Subordination' was first referred to by Bochner [1955], who introduced the notion of time-changing a Markov process by an independent Lévy process which results in another Markov process [see Bertoin, 1998]. Subordinators are increasing pure jump Lévy processes; they can be used to construct Lévy processes by performing a time-change on Brownian motion. Examples of subordinators include Poisson, gamma and inverse Gaussian processes.

In finance, popular models of time-changed Brownian motion are the variance gamma (Madan and Seneta [1990] and Madan et al. [1998]) and normal inverse Gaussian (Barndorff-Nielsen [1997, 1998]) processes, where the subordinators are the gamma and inverse Gaussian processes, respectively. These subordinators can be seen as replacing the 'calendar' time in Brownian motion with what is interpreted in mathematical finance as 'business' or 'market' time.

As noted in Madan and Yor [2005] the variance gamma and normal inverse Gaussian processes are constructed as time-changed Brownian motions. However, some Lévy processes, such as CGMY (see Koponen [1995] and Boyarchenko and Levendorskiĭ [1999], and Carr et al. [2002]) or Meixner processes (see Schoutens and Teugels [1998], Grigelionis [1999], Schoutens [2000] and Pitman and Yor [2003]), are defined directly by their Lévy measures and it is not known in advance whether they can be written as time-changed Brownian motion.

We give a formal definition of subordinators (see Sato [1999, Definition 21.4] or Applebaum [2004, §1.3.2]).

Definition 6.7.1 Subordinator
A subordinator *is a one-dimensional a.s. nondecreasing Lévy process. Thus a subordinator is a Lévy process $\{X_t\}_{t \geq 0}$, where for $t \geq s$, $X_t \geq X_s$ a.s.*

Since subordinators are Lévy processes, they possess stationary and independent increments. Furthermore, as we shall see in the next proposition, they are pure jump processes (possibly having infinite activity) with an added deterministic drift.

Proposition 6.7.2
Let $X = \{X_t\}_{t\geq 0}$ be a Lévy process on $\mathbb{R}$. Then X is a subordinator if and only if the characteristic triplet of X satisfies $\sigma = 0$, $\nu((-\infty, 0]) = 0$, $\int_0^\infty (x \wedge 1)\, \nu(dx) < \infty$ and $\gamma_0 \geq 0$ where $\gamma_0 = \gamma - \int_{|x|\leq 1} x\, \nu(dx)$ [Cont and Tankov, 2004a, Proposition 3.10].

The intuition behind Proposition 6.7.2 is as follows: if $\sigma \neq 0$, then X has a Brownian component and hence can have downward moves. Since the Brownian and jump components of X are independent, downward moves originating from the Brownian component will not be offset by positive moves of the jump component. Again X will be able to decrease. Furthermore, if γ_0 is allowed negative values, then X can have negative drift and will be able to decrease. Finally, if we had $\nu((-\infty, 0]) > 0$, then X could have downward jumps and hence could decrease.

Since a subordinator is an a.s. nondecreasing process, it is of finite variation[9] [see Schoutens, 2003, §2.2.3]. Furthermore, a Lévy process with $\sigma = 0$ and $\int_0^\infty (x \wedge 1)\, \nu(dx) < \infty$ is of finite variation:

Proposition 6.7.3
A Lévy process is of finite variation if and only if its characteristic triplet (σ^2, ν, γ) satisfies

$$\sigma = 0 \text{ and } \int_{|x|\leq 1} |x|\, \nu(dx) < \infty$$

[Cont and Tankov, 2004a, Proposition 3.9].

Hence, the characteristic function of the subordinator X takes the form

$$\Phi_{X_t}(z) = e^{t\psi(z)},$$

where

$$\psi(z) = iz\gamma_0 + \int_0^\infty e^{izx} - 1\, \nu(dx) \tag{6.13}$$

with $z \in \mathbb{R}$. Since the subordinator X_t is positive for all t, we can describe it using a Laplace transform instead of a Fourier transform. The moment generating function of X_t is

$$M_{X_t}(u) = e^{tl(u)}$$

where

$$l(u) = u\gamma_0 + \int_0^\infty e^{ux} - 1\, \nu(dx) \tag{6.14}$$

with $u \in \mathbb{R}$. Here $l(u)$ is called the *Laplace exponent* of X.

The following theorem shows that a Lévy process, time-changed by an independent subordinator, always results in another Lévy process [Cont and Tankov, 2004a, Theorem 4.2].

[9] A Lévy process X is of finite variation if every path of X is of finite variation with probability 1. Brownian motion is a well-known example of a Lévy process that is not of finite variation [Kallenberg, 2001, Corollary 13.10].

Theorem 6.7.4 Subordination of a Lévy Process
Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and suppose $X = \{X_t\}_{t\geq 0}$ is a Lévy process on $\mathbb{R}$ with characteristic exponent $\psi(u)$ and triplet $\left((\sigma^X)^2, \nu^X, \gamma^X\right)$. Furthermore, let $S = \{S_t\}_{t\geq 0}$ be a subordinator with Laplace exponent $l(u)$ and triplet $(0, \nu^S, \gamma^S)$. Then the process $Y = \{Y_t\}_{t\geq 0}$ defined for each $\omega \in \Omega$ by

$$Y_t(\omega) = X_{S_t(\omega)}(\omega)$$

is a Lévy process and has Lévy triplet $\left((\sigma^Y)^2, \nu^Y, \gamma^Y\right)$ where

$$\begin{aligned}
(\sigma^Y)^2 &= \gamma_0^S (\sigma^X)^2 \\
\nu^Y(B) &= \gamma_0^S \nu^X(B) + \int_0^\infty f_{X_s}(B)\nu^S(ds), \text{ for all } B \in \mathcal{B}(\mathbb{R}) \\
\gamma^Y &= \gamma_0^S \gamma^X + \int_0^\infty \nu^S(ds) \int_{|x|\leq 1} x\, f_{X_s}(dx).
\end{aligned}$$

Here f_{X_t} is the probability distribution of X_t and $\gamma_0^S = \gamma^S - \int_{|x|\leq 1} x\nu^S(dx)$.

For a proof, see Cont and Tankov [2004a, Theorem 4.2] or Sato [1999, Theorem 30.1]. The time-changed process Y is said to be *subordinate* to the process X.

Chapter 7

Generating Geometric Brownian Motion Paths

The purpose of this chapter is to serve as a basic reference for path simulation in Chapters 8 and 9. We discuss the generation of paths where the dynamics of the stock price is assumed to obey geometric Brownian motion (briefly considered in §7.1). However, unlike Chapters 8 and 9, we do not discuss the particular properties related to Lévy processes.

Generating geometric Brownian paths requires normal random numbers. We will generate uniform random numbers and then transform them into normal random numbers using a method we discuss in §7.2. In order to generate uniform random numbers, we will offer two options, namely

- a pseudo-random number generator. We will make use of the Mersenne Twister of Matsumoto and Nishimura [1998] (see §7.3).
- a quasi-random number generator. Here we will consider Sobol' sequences by Sobol' [1967] which we introduce in §7.4 and develop in greater detail in Appendix B. Furthermore, as we will see later, Sobol' sequences should always be used in conjunction with bridges. Therefore, we will look at bridge sampling in §7.5 which was introduced by Caflisch and Moskowitz [1995] and Moskowitz and Caflisch [1996].

7.1 Geometric Brownian Motion

We will assume the existence of an equivalent martingale measure $\mathbb{Q}$ equivalent to the real-world measure $\mathbb{P}$ under which discounted asset prices are martingales. In finance, such an equivalent martingale measure is also known as a *risk-neutral measure*. Thus, according to the First Fundamental Theorem of Asset Pricing we have an arbitrage-free model [Shreve, 2004, Theorem 5.4.7].

When we consider Brownian motion, the risk-neutral measure is unique because the Martingale Representation Property holds. Thus if the number of underlying assets is greater or equal to the dimension of the Brownian motion, it follows from the Second Fundamental Theorem of Asset Pricing, that the market

is complete [Shreve, 2004, Theorem 5.4.9]. We construct the unique equivalent martingale measure using Girsanov's Theorem by changing the drift so that discounted asset prices become martingales under $\mathbb{Q}$.

When considering the Black-Scholes model, the stock price dynamics of the stochastic process $\{S_t\}_{t\geq 0}$ under the real-world measure $\mathbb{P}$ is given by

$$dS_t = \mu S_t dt + \sigma S_t dW_t^{\mathbb{P}} \tag{7.1}$$

where $W_t^{\mathbb{P}}$ indicates standard Brownian motion and $\mu,\ \sigma > 0$ are real numbers. If we change to the risk-neutral measure $\mathbb{Q}$ using Girsanov's Theorem (see Etheridge [2002, Theorem 4.5.1] for example) (7.1) becomes

$$dS_t = (r-q)S_t dt + \sigma S_t dW_t \tag{7.2}$$

where, as before, r denotes the constant continuously compounded risk-free rate and q the constant continuously compounded dividend yield. Using Itô's lemma in (7.2) gives

$$\ln \frac{S_{t+\Delta t}}{S_t} = \left(r - q - \tfrac{1}{2}\sigma^2\right)\Delta t + \sigma W_{\Delta t}. \tag{7.3}$$

Thus

$$\begin{aligned} S_{t+\Delta t} &= S_t \exp\left[\left(r - q - \tfrac{1}{2}\sigma^2\right)\Delta t + \sigma W_{\Delta t}\right] \\ &= S_t \exp\left[\left(r - q - \tfrac{1}{2}\sigma^2\right)\Delta t + \sigma\sqrt{\Delta t} n\right] \end{aligned} \tag{7.4}$$

for $n \sim \text{Normal}(0,1)$.

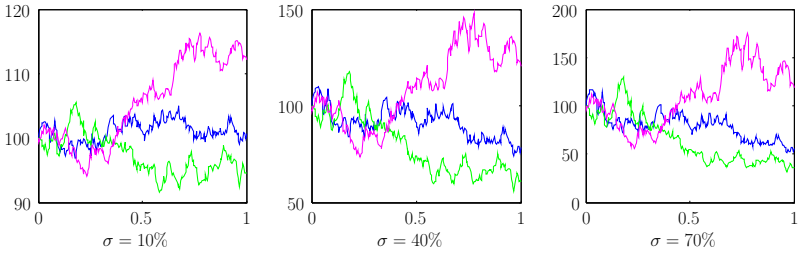

Figure 7.1: *Simulation of three geometric Brownian motion paths with* $S_0 = 100$, $r = 7\%$, $q = 0\%$ *where the values of* σ *are varied. Note that the scales of the vertical axes are different in each figure.*

For later reference, we note that using the probability density function of a normal random variable X_t with parameters μ and $\sigma > 0$

$$f_{X_t}(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu t}{\sigma\sqrt{t}}\right)^2\right] \tag{7.5}$$

for $x \in \mathbb{R}$ and from (7.3) we find the density of $Y_t = \ln \frac{S_t}{S_0}$ under $\mathbb{Q}$

$$f_{Y_t}^{\mathbb{Q}}(x) = \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{1}{2}\left(\frac{x-\left(r-q-\frac{1}{2}\sigma^2\right)t}{\sigma\sqrt{t}}\right)^2\right]. \tag{7.6}$$

7.2 Generating Normal Random Numbers from Uniform Random Numbers

In order to generate random numbers from the standard normal distribution, we will apply the method given by Moro [1995]. This method provides an approximation to the cumulative inverse of the normal distribution and so is an inverse transform method (see Appendix H). Moro's approximation is accurate to within 12 decimal places and only requires one uniform random number to generate a standard normal random number.

Other methods for generating standard normal random numbers exist such as those given by Box and Muller [1958] and Acklam [2004]. However, using Moro's method seems to be standard practice and is a fast and accurate approximation. Given a uniform random variable, we will always generate a standard normal random variable by applying Moro's method.

7.3 Generating Pseudo-Random Numbers using Mersenne Twister

Pseudo-random numbers are generated by some deterministic algorithm and are designed to look like random numbers. These numbers are not random at all as they are completely determined by a relatively small set of initial values. When given a particular seed a second time, the same sequence will be generated. Furthermore, these numbers have finite periodicity.

Linear congruential generators, lagged Fibonacci generators, linear feedback shift registers and generalised feedback shift registers are common examples of pseudo-random number generators. More recent pseudo random algorithms include Blum Blum Shub [Blum et al., 1986] and Fortuna [Ferguson and Schneier, 2003]. However these generators seem to be focussed on cryptography problems. In fact, Blum Blum Shub is not appropriate for use in simulations as it is very slow.

We will use the Mersenne Twister which is a pseudo-random number generator with an extremely long period, is faster than other statistically reasonable generators and has some attractive statistical properties. The *Mersenne Twister (MT)*, developed in Matsumoto and Nishimura [1998], is a pseudo-random number generating algorithm based on a matrix linear recurrence over a finite binary field F_2. The period length is chosen to be a Mersenne prime, hence the name *Mersenne Twister*. There are several variants of the algorithm. A more commonly used variation is the *MT19937* which has a 32-bit word length. Another is the *MT19937-64* which has a 64-bit word length and generates a different sequence.

MT is widely used as it has a far longer period and a higher order of dimensional equidistribution[1] than other implemented generators: it has a period of $2^{19937} - 1$ and 623-dimensional equidistribution property. Several c++ and c implementations of MT are available for download at `http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/VERSIONS/C-LANG/c-lang.html`. We have implemented the version of Richard Wagner available at `http://www-personal.umich.edu/~wagnerr/MersenneTwister.html`.

[1] A high order of dimensional equidistribution means that the serial correlation between successive values in the resulting sequence is negligible.

7.3.1 Simulating Sequential Geometric Brownian Motion

We provide an algorithm that generates sequential stock price paths where the stock price follows geometric Brownian as shown in (7.3). Let σ indicate the volatility of the log returns and let S be a matrix of stock prices where S_j^i indicates the stock price at time t_j of the i^{th} sample path.

- For each simulation path i, $i = 1,\ 2,\ \ldots,\ N$, we loop through the time steps Δt_j, $j = 1,\ 2,\ \ldots,\ M$.

  For $i = 1$ **To** N

  - **For** $j = 1$ **To** M
    - Generate a uniform random number u_j^i using the pseudo-random number generator discussed in §7.3.
    - Using Moro's method, calculate n_j^i, the cumulative inverse standard normal random number of u_j^i.
    - Calculate S_j^i using (7.4):

$$S_j^i = S_{j-1}^i \exp\left[\left(r - q - \tfrac{1}{2}\sigma^2\right)\Delta t_j + \sigma\sqrt{\Delta t_j}\, n_j^i\right].$$

  Next j

Next i

7.4 Generating Quasi-Random numbers using Sobol' Numbers

Like pseudo-random numbers, quasi-random numbers are not random and are created deterministically. Pseudo-random numbers suffer from the lack of uniformity which produces inefficient convergence in Monte Carlo integration. However quasi-random numbers are generated in such a way that they form uniformly distributed sequences which allow for better convergence in Monte Carlo integration. Quasi-random sequences are also called *low discrepancy sequences* as they are designed to minimise the discrepancy between sampled points.

According to Acworth et al. [1996], for option pricing, quasi-Monte Carlo methods surpass pseudo-Monte Carlo methods with and without antithetics, with the exception of cases where the dimension is very high and the number of points is very small. Dimension can be roughly understood as the number of different uniform random numbers required to generate an entire sample path. When the process follows geometric Brownian motion, the dimension will be equal to M, that is, the number of time steps in the option under consideration (see §1.1). Other processes, such as exponential variance gamma as we will see in §8.3.1 or normal inverse Gaussian as we will see in §9.3.1, require two or three uniform random numbers respectively at each time step and hence the dimension in these cases will be $2M$ or $3M$, etc.

There are many examples of quasi-random sequences, such as Halton, Niederreiter, Faure and Sobol'. It is mentioned in Acworth et al. [1996] that among the various quasi-Monte Carlo methods there is no uniformly superior method. However, as stated in Acworth et al. [1996], Sobol' numbers in conjunction with bridge construction most often has the lowest root mean square relative error and is never noticeably worse than the other methods. We will only discuss Sobol' sequences and we will consider bridge construction in §7.5.

In Haug [2007], Peter Jäckel points out that of the many types of low discrepancy numbers none of them are usable in high dimensions apart from properly initialised Sobol' numbers. When asked about combining pseudo-random numbers and low discrepancy numbers in Haug [2007], Peter Jäckel notes that since it became more widely known that properly initialised Sobol' numbers work extremely well, publications on mixing approaches have diminished.

The generation of Sobol' sequences are considered in detail in Appendix B. There we will also discuss approaches shown by Jäckel [2002], Joe and Kuo [2003] and Glasserman [2004] on the implementation of the Sobol' generator. Furthermore, in §B.4 we perform various tests in order to check that our implementation of Sobol' sequences is reasonable.

7.5 Bridges and Effective Dimension

When quasi-Monte Carlo techniques are used to perform numerical integration, it is possible to apply some additional techniques to reduce the variance of the resulting approximation to the integrand. We use a sampling algorithm called *bridge sampling* which was introduced in the Brownian motion case in Caflisch and Moskowitz [1995] and Moskowitz and Caflisch [1996].

Here, together with quasi-Monte Carlo, sample paths are structured as follows: first we simulate in one step to terminal time (thus a requirement to apply this bridging technique is that the terminal distribution of the process is known). Discrete paths are then sampled by recursively subdividing the sampling time period, conditional on the already generated values of the process (assuming that the mathematical properties of the process are sufficiently well-known to be able to do this). Variations on this type of restructuring of sampling paths with quasi-Monte Carlo have been suggested in Acworth et al. [1996] and Åkesson and Lehoczky [2000]. It is noted in Avramidis et al. [2004] that the conclusions made in Acworth et al. [1996], Åkesson and Lehoczky [2000] and Caflisch et al. [1997] indicate that these path sampling algorithms outperform pseudo-Monte Carlo very often and sometimes by orders of magnitude.

As noted in Tavella [2002, Chapter 4] the total variance of all the increments of the path does not depend on the way the path is constructed. Also, in Jäckel [2002, §10.8] it is remarked that the specific path construction technique does not directly influence the variance of any pseudo-Monte Carlo simulation, but a good choice of a path construction technique can significantly improve the convergence behaviour when quasi-random numbers are used rather than pseudo-random numbers. In a personal correspondence Mark Joshi highlighted this fact.

The reason is as follows: low discrepancy sequences for lower coordinates are better distributed compared with higher coordinates, unlike pseudo-random Monte Carlo which is equally (good or bad) at sampling in any dimension. Thus it would be sensible to have more efficient coverage of dimensions — here dimension means the dimension of the integration problem and not that of the low discrepancy sequence — with larger variance, and allow the random numbers in which we have less confidence to cover dimensions of smaller variance.

Bridge sampling more or less redefines the dimensions so that they are in order of decreasing variance, as opposed to merely in order of occurrence. As noted in Tavella [2002, Chapter 4], the number of dimensions is not reduced here. The idea of *effective dimension* first appeared in Paskov and Traub [1995] and is discussed in detail in Caflisch et al. [1997]. Effective dimension can roughly be understood as the number of dimensions required to explain a sufficiently large proportion of the variance in the output. The

proportion used in Caflisch et al. [1997] is 99%, but this is arbitrary.

Quasi-Monte Carlo methods without this restructuring of path sampling will outperform pseudo-Monte Carlo only by chance. They may also underperform [see Papageorgiou, 2002] even with bridges, although this type of construction is probably pathological. Joshi [2011, §19.1] notes that quasi-Monte Carlo without bridges may even converge to the incorrect solution.

We will consider the bridging technique for Brownian motion in §7.5.2, and later for gamma processes in §8.3.1 and inverse Gaussian processes in §9.3.1.

7.5.1 Bridge Implementation

Let us consider our own general algorithm for constructing a bridge. The algorithms in Glasserman [2004, §3.1.1, p.85] and Jäckel [2002, §10.8.3] do not allow for unequal time steps, a volatility term structure, or where the number of time steps is chosen not to be a power of 2; the algorithm we develop here will allow for these possibilities.

Suppose that we would like to simulate the Wiener process $\{W_t\}_{t\geq 0}$ for discrete times $t_1,\ t_2,\ \ldots,\ t_M$. Assume we have drawn M i.i.d. standard normal random numbers $z_1,\ z_2,\ \ldots,\ z_M$ in order to perform the simulation (in the case of *sequential sampling*, the z_i correspond to the t_i as $W_{t_i} = W_{t_{i-1}} + \sqrt{t_i - t_{i-1}}z_i$). First consider the simplest case where M is a power of 2 and the time points are equally spaced.

Generate a single step from t_0 to t_M using z_1. Then, using the fact that we know the end point, we perform a type of 'random interpolation' using the conditional distribution to determine the value of the process at time $t_{\frac{M}{2}}$ using z_2. Next we find the value of the process at time $t_{\frac{M}{4}}$ by performing the 'interpolation' between times t_0 and $t_{\frac{M}{2}}$ using z_3, and similarly for time $t_{\frac{3M}{4}}$ using z_4. We continue in this way, finding the values of the process by halving the time steps and performing the 'interpolation' using the conditional distribution until the total number of points is found. This algorithm is implemented in Jäckel [2002, §10.8.3].

Now suppose the number of time intervals are not necessarily a power of 2, or points are not equally spaced. One would like bridging points to be chosen in such a way that the maximum of the *outstanding variances* is reduced as quickly as possible.

Definition 7.5.1 Outstanding Variance
The outstanding variance *of an interval (s,t), where the values of the process at the end points s and t are known, the value of the process at all interior points is unknown and this set of points is not empty, is given by*

$$\int_s^t \sigma^2(u)du$$

where σ is the instantaneous volatility of the stock price process and is a deterministic function of time.

In the constant volatility case the outstanding variance is proportional to the length of the time interval.

An example is given in Jäckel [2002, §10.8.3] in which 14 time periods are equally spaced. However, the algorithm which works fine for the case where the number of points is a power of 2, is no longer optimal. The order in which the bridge points are chosen is given by 14, 7, 3, 10, *1*, ... whereas a minimising sequence would be 14, 7, 3, 10, *5* or *12*, ... Furthermore, this implementation does not allow for varying time interval lengths.

Consider Figure 7.2 where the observation dates of an Asian out option (an option where the averaging dates are the last few business days before maturity). Clearly, creating a bridge point at time t_1 would minimise outstanding variance. In fact, it is clear, that just knowing the stock price at times t_8 and t_1, gives us a pretty good idea of the option payoff. However, the implementation in Jäckel [2002, §10.8.3] would not choose t_1 as the first position at which a bridge point is created, but t_4.

Figure 7.2: *Asian out option observation dates.*

The algorithm that finds the bridging point which reduces outstanding variance as quickly as possible is the solution to a *minimax problem*. Consider an example in which this algorithm is illustrated: Figure 7.3 indicates a flat volatility structure with 8 varying time periods. Suppose that bridge points have been created at times t_1, t_4 and t_7 in previous inductive steps. Here we have 4 time periods namely (t_0, t_1), (t_1, t_4), (t_4, t_7) and (t_7, t_8). We will only consider the intervals (t_1, t_4) and (t_4, t_7) as these are the only intervals containing points for which a bridge point must still be created. Of these two intervals, (t_4, t_7) has the greatest variance and thus we would choose a point from this interval to minimise the outstanding variance.

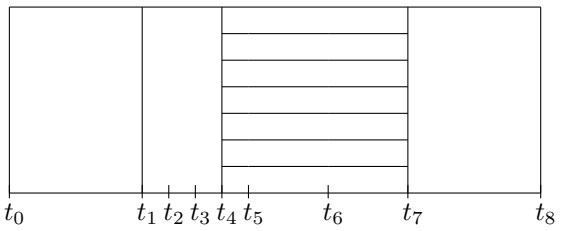

Figure 7.3: *A flat volatility term structure with 8 varying time periods. Vertical lines indicate that bridging points have already been created inductively. Since interval (t_4, t_7) is of greatest variance, we will choose the next bridging point from this interval.*

Once we have determined the interval of *maximum* variance, we find the point in that interval which, if chosen, will *minimise* the outstanding variance.

When considering Figure 7.3, we see that if we choose the next bridging point to be at time t_5, the left Figure 7.4 is produced; whereas choosing to create a bridging point at t_6, the right Figure 7.4 results. Clearly a bridge point at time t_6 would be more desirable as the resulting greatest outstanding variance for the newly created intervals (t_4, t_6) and (t_6, t_7) would be smaller than for the intervals (t_4, t_5) and (t_5, t_7) resulting from choosing to create a bridge point at time t_5.

Thus we choose the next bridging point to be at time t_6, and induct.

Let M be the dimension of the problem under consideration. We now give two algorithms which will enable us to find a vector of integers BI (bridge index) of length M where BI_k, $k = 1, 2, \ldots, M$ will contain the index at which the bridge point is created in step k. Suppose inductively that BI is partially completed, as well as the vector B of length $M + 1$ containing boolean entries which indicate whether in an earlier time step a bridge point has been created for a position or not. Observe that $BI_1 = M$, since

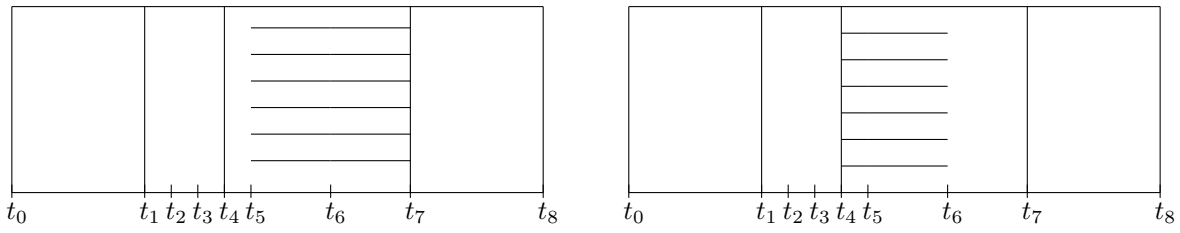

Figure 7.4: *The horizontal lines indicate the resulting outstanding variance in the interval* (t_4, t_7) *and when choosing to create a bridge point at time* t_5 *and* t_6 *on the left and right respectively.*

the first bridge point is always created at the last time point. Also, note that $B_0 = 1$ and $B_M = 1$, that is, bridge points exist at the first and last time point.

The first algorithm finds the left index LL (largest left) of the interval of largest outstanding variance. Suppose Σ is a vector of length M whose entries are given by $\Sigma_j = \int_{t_0}^{t_j} \sigma^2(s)ds$ for $j = 1, 2, \ldots, M$. Furthermore, let L and R be integer variables which will denote left and right indices of a test interval. Let LV (largest variance) be a variable which will contain the largest outstanding variance that has been found thus far.

- Set $LL := 0$, $LV := 0$, $L := 0$ and $R := 0$.
- **While** $L < M$ **Do**
  - Set $R := L + 1$.
  - Increment R until B_R is equal to 1, that is, until we reach a point at which a bridge point has been created.
  - If R is equal to $L + 1$ then set $L := R$.
  - **If** $\Sigma_R - \Sigma_L > LV$ [2] **Then**
    Set $LV := \Sigma_R - \Sigma_L$ and $LL := L$.
    End If
  - Set $L := R$.

  End While

Our second algorithm finds the point that will minimise the outstanding variance of the chosen interval. Let SV (smallest variance) be a variable containing the smallest variance thus far and let RLV (resulting largest variance) be a variable that is equal to the largest variance if a bridge point should be created. Finally, let NBI (next bridge index) be an integer variable which will eventually be equal to the index at which the next bridge point will be created.

- Set $SV := \Sigma_R - \Sigma_{LL}$.

[2] Several such variance differences can be equal but can have differing decimal representations due to rounding. Thus it is necessary to replace here LV with $LV + \epsilon$, for some ϵ just larger than the level of machine precision.

- **For** $j = LL + 1$ **To** $R - 1$
  - Set $RLV := \max\{\Sigma_R - \Sigma_j, \Sigma_j - \Sigma_{LL}\}$.
  - **If** $SV > RLV$ **Then**
    Set $SV := RLV$ and $NBI := j$.
    End If

Next j

Once we have completed this algorithm, we update the vectors B and BI by setting $B_{NBI} := 1$ and $BI_k := NBI$. Then we return to the first algorithm to find the next bridging point.

7.5.2 Brownian Bridge Sampling

Suppose we want to simulate a discretised path of standard Brownian motion $\{W_{t_i}\}_{i=0,\ 1,\ ...,\ M}$. In order to simulate the path of Brownian motion, we first simulate Brownian motion at the end of the term t_M as follows

$$W_{t_M} = \sqrt{t_M} n_1 \tag{7.7}$$

where n_1 is the first standard normal random variable in the sequence. Then, using the results of the algorithm as shown in §7.5.1, we choose the first point t_j at which a bridge must be created with bridge points $W_{t_0} = W_0 = 0$ and W_{t_M} at times t_0 and t_M. We determine the conditional distribution of the Brownian motion increment $W_{t_j} - W_{t_0}$, using the next standard normal random variable in the sequence n_2, given that we know $W_{t_M} - W_{t_0}$. We proceed by induction.

The general inductive step will be as follows: consider three consecutive times $s < t < u$ and suppose that x, y and z are realisations of the Brownian motion increments $W_t - W_s$, $W_u - W_t$ and $W_u - W_s$ respectively (see Figure 7.5). Here W_s and W_u are points which have already been determined and t was determined to be the point at which the next bridge is to be created.

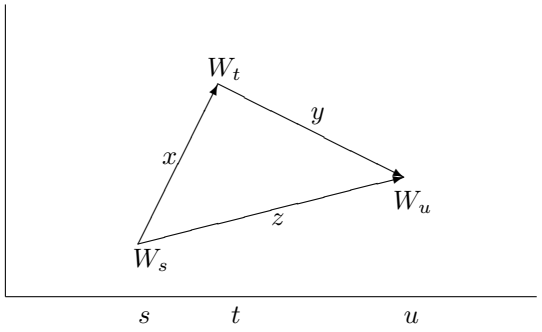

Figure 7.5: *Brownian Bridge*

In order to find the conditional distribution of the Brownian motion increment $W_t - W_s = x$ given that we know the increment $W_u - W_s = z$, we will require Bayes' Theorem. Suppose that X and Z are

the distributions from which x and z are chosen, then

$$f_{X|Z=z}(x) = \frac{f_{Z|X=x}(z) f_X(x)}{f_Z(z)}. \tag{7.8}$$

Applying (7.8) gives

$$\begin{aligned}
f_{(W_t - W_s | W_u - W_s = z)}(x) &= \frac{f_{(W_u - W_s | W_t - W_s = x)}(z) f_{W_t - W_s}(x)}{f_{W_u - W_s}(z)} \\
&= \frac{f_{W_u - W_t + x}(z) f_{W_t - W_s}(x)}{f_{W_u - W_s}(z)} \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{(u-t)(t-s)}{(u-s)}}} \exp\left[-\frac{1}{2} \left(\frac{(z-x)^2}{u-t} + \frac{x^2}{t-s} - \frac{z^2}{u-s} \right) \right] \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\frac{(u-t)(t-s)}{(u-s)}}} \exp\left[-\frac{1}{2} \left(\frac{x - \frac{t-s}{u-s} z}{\sqrt{\frac{(u-t)(t-s)}{u-s}}} \right)^2 \right].
\end{aligned}$$

Thus, conditional on knowing the increment $W_u - W_s = z$, $W_t - W_s$ is normally distributed with mean $\frac{t-s}{u-s} z$ and standard deviation $\sqrt{\frac{(u-t)(t-s)}{u-s}}$. Therefore, for a standard normal random variable n, we may write (see Jäckel [2002, §10.9.2] for example)

$$\begin{aligned}
W_t - W_s &= \frac{t-s}{u-s} z + \sqrt{\frac{(u-t)(t-s)}{u-s}} n \\
&= \frac{t-s}{u-s} (W_u - W_s) + \sqrt{\frac{(u-t)(t-s)}{u-s}} n \\
\Rightarrow W_t &= \left(1 - \frac{t-s}{u-s} \right) W_s + \frac{t-s}{u-s} W_u + \sqrt{\frac{(u-t)(t-s)}{u-s}} n \\
&= \frac{u-t}{u-s} W_s + \frac{t-s}{u-s} W_u + \sqrt{\frac{(u-t)(t-s)}{u-s}} n.
\end{aligned} \tag{7.9}$$

Thus we have completed the inductive step.

Suppose that we want to simulate arithmetic Brownian motion $X = \{X_t\}_{t \geq 0}$ with drift μ and variance σ^2. Given the Brownian bridge W_t in (7.9), we find its corresponding arithmetic Brownian motion X_t by multiplying W_t with σ and adding μt

$$\begin{aligned}
X_t &= \mu t + \sigma W_t \\
&= \mu t + \sigma \left(\frac{u-t}{u-s} W_s + \frac{t-s}{u-s} W_u + \sqrt{\frac{(u-t)(t-s)}{u-s}} n \right) \\
&= \frac{u-t}{u-s} \mu s + \frac{t-s}{u-s} \mu u + \sigma \left(\frac{u-t}{u-s} W_s + \frac{t-s}{u-s} W_u + \sqrt{\frac{(u-t)(t-s)}{u-s}} n \right) \\
&= \frac{u-t}{u-s} X_s + \frac{t-s}{u-s} X_u + \sigma \sqrt{\frac{(u-t)(t-s)}{u-s}} n.
\end{aligned} \tag{7.10}$$

Thus we may construct our algorithm for creating arithmetic Brownian motion in either of two ways: first generate Brownian bridges for all times and then find their corresponding arithmetic Brownian motion, or for each Brownian bridge W_t calculate its corresponding arithmetic Brownian motion X_t immediately.

Chapter 8

The Variance Gamma Model

The class of variance gamma distributions was originally introduced as a model of stock returns in Madan and Seneta [1987]. The symmetric variance gamma process was considered in the seminal paper of Madan and Seneta [1990], showing its applicability as a model for stock returns. The risk-neutral variance gamma process with skewness was studied in Madan and Milne [1991]. It was shown in Madan et al. [1998] that the asymmetric risk-neutral process is equivalent to arithmetic Brownian motion which has been subordinated by a gamma process. Furthermore, the formula for the variance gamma density in terms of Bessel functions is deduced in Madan et al. [1998], where it is also shown that a variance gamma process can be written as the difference of two gamma processes. Hence variance gamma processes are of finite variation.

In Geman et al. [2001] several arguments supporting the use of time-changed Brownian motion as a model for price processes are presented. Geman et al. [2001] show how the price process may be viewed as Brownian motion, but only in 'business' time which is modelled by a subordinator. We will consider the asymmetric variance gamma process in Madan et al. [1998], defined as the arithmetic Brownian motion time-changed by a gamma process, and begin by considering gamma processes.

8.1 Gamma Processes

The gamma distribution was originally referred to as the Pearson Type III distribution and originated from the work of Pearson (Pearson [1893, 1895]). Only later, in the 1930s and 1940s (see for example Weatherburn [1946]), it became known as the gamma distribution. In its most general form (see Johnson et al. [1994, Chapter 17] for example), the gamma distribution is dependent on three parameters and its probability density is of the form

$$f_{\mathrm{G}}(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}(x-\gamma)^{\alpha-1}e^{-\beta(x-\gamma)} \tag{8.1}$$

where α, $\beta > 0$ and $x > \gamma$ and $\Gamma(\cdot)$ indicates the gamma function. The most common definition of the gamma function is given by Euler's integral for $z \in \mathbb{C}$

$$\Gamma(z) = \int_0^{\infty} t^{z-1}e^{-t}dt,$$

where $\Re(z) > 0$. A consequence of the above is [Abramowitz and Stegun, 1974, 6.1.1]

$$\Gamma(z) = k^z \int_0^\infty t^{z-1} e^{-kt} dt \tag{8.2}$$

for $\Re(k) > 0$.

We consider the two-parameter form of the gamma distribution only: if $\gamma = 0$, (8.1) becomes

$$f_{\text{G}}(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \mathbb{1}_{\{x \geq 0\}}. \tag{8.3}$$

From now on, we refer to the two-parameter gamma distribution as the gamma distribution.

The gamma distribution is closely related to the exponential distribution: if $\alpha = 1$, (8.3) becomes the density function of the exponential distribution with parameter β (see (6.1)). Since the exponential distribution is infinitely divisible, it has a Lévy process associated with it (see §6.2). In the case of the exponential distribution, the associated Lévy process is given by the gamma process. The exponential distribution has often been used to model arrival times of events. In particular, gamma processes have been used to model aggregate insurance claims [see Embrechts et al., 2001, §2]. In a similar way, gamma processes can be used to model the arrival of information in the market at discrete points in time. Thus, as a subordinator, a gamma process can be used to model 'business' time.

We denote a random variable X that follows a gamma distribution with parameters $\alpha > 0$ and $\beta > 0$ as

$$X \sim \text{Gamma}(\alpha, \beta)$$

where α is referred to as the *shape parameter* and β as the *rate parameter*.

Definition 8.1.1 Gamma Process
A gamma process *with parameters* α, $\beta > 0$ *is a Lévy process* $X^{\text{G}} = \{X_t^{\text{G}}\}_{t \geq 0}$, *where the increments* $X_{t+\Delta t}^{\text{G}} - X_t^{\text{G}} \stackrel{\mathcal{D}}{=} X_{\Delta t}^{\text{G}}$ *are distributed* Gamma$(\alpha \Delta t, \beta)$ *[Schoutens, 2003, §5.3.3].*

In Table 8.1 we list some results concerning gamma processes. These results will be discussed subsequently.

Property	Expression/Value
Probability Density Function	$f_{X_t^{\text{G}}}(x) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} x^{\alpha t - 1} e^{-x\beta} \mathbb{1}_{\{x>0\}}$
Characteristic Function	$\Phi_{X_t^{\text{G}}}(z) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)(\beta - iz)^{\alpha t}} (\beta - iz)^{\alpha t} \int_0^\infty x^{\alpha t - 1} e^{-(\beta - iz)x}\, dx$
Activity	Infinite Activity
Lévy Triplet	$\sigma^{\text{G}} = 0$
	$\gamma_0^{\text{G}} = 0$ and $\gamma^{\text{G}} = \frac{\alpha}{\beta}\left(1 - e^{-\beta}\right)$
	$\nu^{\text{G}}(x) = \frac{\alpha e^{-\beta x}}{x} \mathbb{1}_{\{x>0\}}$

Table 8.1: *Results for a gamma process* X^{G} *with* $x, z, \alpha, \beta \in \mathbb{R}$ *and* α, $\beta > 0$.

If X^{G} is a gamma process with parameters $\alpha > 0$ and $\beta > 0$, then X_t^{G} has density

$$f_{X_t^{\text{G}}}(x) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} x^{\alpha t - 1} e^{-x\beta} \mathbb{1}_{\{x>0\}}.$$

The characteristic function of X_t^{G} is calculated for $z \in \mathbb{R}$ as follows

$$\begin{aligned}\Phi_{X_t^{\mathrm{G}}}(z) &= \mathbb{E}\left[e^{izX_t^{\mathrm{G}}}\right] \\ &= \frac{\beta^{\alpha t}}{\Gamma(\alpha t)} \int_0^\infty x^{\alpha t-1} e^{-(\beta-iz)x}\, dx \\ &= \frac{\beta^{\alpha t}}{\Gamma(\alpha t)\,(\beta-iz)^{\alpha t}} (\beta-iz)^{\alpha t} \int_0^\infty x^{\alpha t-1} e^{-(\beta-iz)x}\, dx.\end{aligned}$$

Noting that $\Re(\beta - iz) = \beta > 0$ and $\Re(\alpha t) = \alpha t > 0$ we apply (8.2) with $k := \beta - iz$ and $z := \alpha t$ to get

$$(\beta-iz)^{\alpha t} \int_0^\infty x^{\alpha t-1} e^{-(\beta-iz)x}\, dx = \Gamma(\alpha t)$$

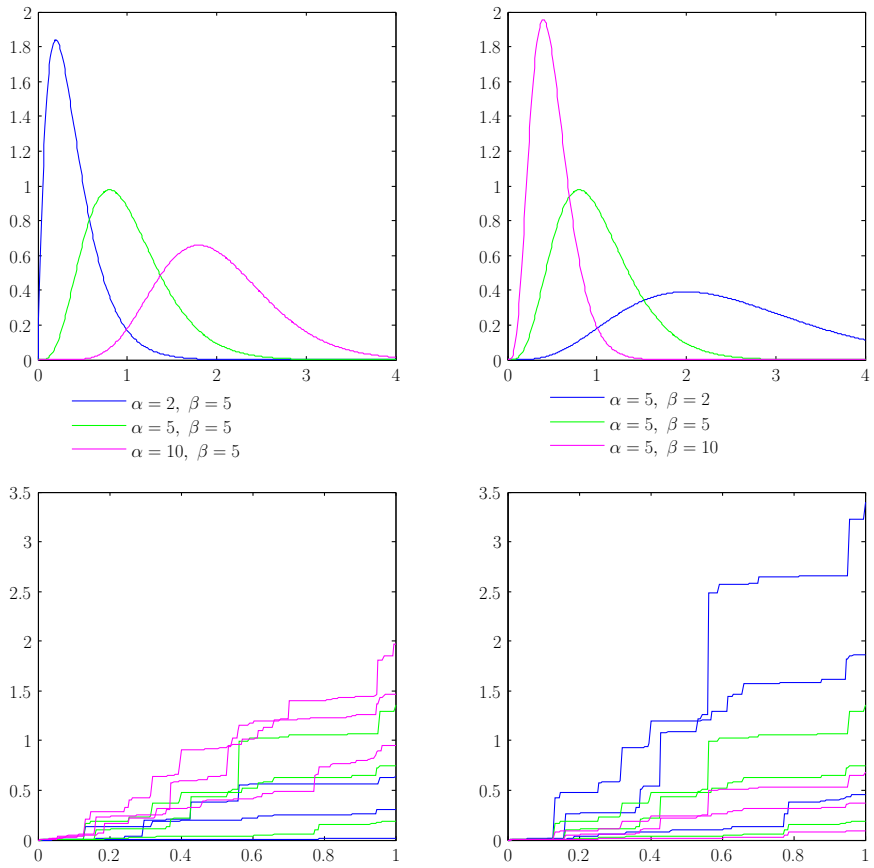

Figure 8.1: *In the first row we plot the gamma probability density function for various values of α and β where one parameter is fixed and the other is varied. The graphs shown in green are the same in both figures with parameters used in common.*

In the second row, we simulate three realisations of three gamma Processes with parameter values corresponding to the densities in the first row. We see that as the values of α increase the jump frequency increases. Furthermore, note how the jump sizes decrease as the values of β increase.

and hence we have

$$\Phi_{X_t^{\mathrm{G}}}(z) = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)\,(\beta - iz)^{\alpha t}}\Gamma(\alpha t) = \left(\frac{\beta}{\beta - iz}\right)^{\alpha t} = \left(1 - i\frac{z}{\beta}\right)^{-\alpha t}. \tag{8.4}$$

Observe that at $t = \frac{1}{\alpha}$, we have that X_t^{G} is exponential with parameter β. Later in §10.2.1 we will make use of the characteristic function of X_t^{G} in the form[1]

$$\Phi_{X_t^{\mathrm{G}}}(z) = \exp\left[-\alpha t \ln\left(1 - i\frac{z}{\beta}\right)\right]. \tag{8.5}$$

By inspecting the characteristic function of X_t^{G}, we observe that X_t^{G} is infinitely divisible. Hence, as we have seen in §6.2, there exists a Lévy process that has characteristic function as in (8.5) with $t = 1$. Furthermore, from the stationarity and independence of increments of the gamma process we have that for $0 \leq s < t < \infty$ $X_{t-s}^{\mathrm{G}} \overset{\mathcal{D}}{=} X_t^{\mathrm{G}} - X_s^{\mathrm{G}}$. Now, since X_{t-s}^{G} is gamma distributed, it is strictly positive with probability 1 (which follows from (8.3)) and hence $X_t^{\mathrm{G}} > X_s^{\mathrm{G}}$ a.s. A gamma process is therefore a Lévy process with a.s. strictly increasing paths and hence, by Definition 6.7.1, is a subordinator.

The moment generating function of X_t^{G} is calculated similarly to its characteristic function

$$M_{X_t^{\mathrm{G}}}(u) = \mathbb{E}\left[e^{uX_t^{\mathrm{G}}}\right] = \frac{\beta^{\alpha t}}{\Gamma(\alpha t)}\int_0^\infty x^{\alpha t - 1}e^{-(\beta - u)x}\,dx.$$

As before, we use (8.2) with $k := \beta - u$ and $z := \alpha t$, but here we are restricted to having $u < \beta$. Thus, for $u < \beta$

$$M_{X_t^{\mathrm{G}}}(u) = \left(1 - \frac{u}{\beta}\right)^{-\alpha t} = \exp\left[-\alpha t \ln\left(1 - \frac{u}{\beta}\right)\right]. \tag{8.6}$$

The Laplace exponent of X_t^{G} for $u < \beta$ is then given by

$$l(u) = -\alpha \ln\left(1 - \frac{u}{\beta}\right). \tag{8.7}$$

The following lemma enables us to write the Laplace exponent of X_t^{G} given above in the form of (6.14) [see Kyprianou, 2006, Lemma 1.7].

Lemma 8.1.2 Frullani Integral
For every α, $\beta > 0$ and $z \in \mathbb{C}$ such that $\Re(z) \leq 0$ we have

$$\left(1 - \frac{z}{\beta}\right)^{-\alpha} = \exp\left[-\int_0^\infty (1 - e^{zx})\,\frac{\alpha e^{-\beta x}}{x}\,dx\right].$$

[1] The form in which the $\Phi_{X_t^{\mathrm{G}}}(z)$ is expressed in (8.4) may seem ambiguous since we are taking a real-valued power of a complex number which results in a multi-valued function. This however is incorrect, since the characteristic function of a Lévy process is unique (see §6.2). The correct single-valued form of (8.12) is given by

$$\Phi_{X_t^{\mathrm{G}}}(z) = \exp\left[-\alpha t \mathrm{Ln}\left(1 - i\frac{z}{\beta}\right)\right]$$

where Ln denotes the *principal value* of the natural logarithm. It is, however, a nontrivial exercise to show that Ln may indeed be used here (see Lord and Kahl [2010, Theorem 4.1] where this is shown for the variance gamma characteristic function).

Thus, for $u \leq 0 < \beta$, (8.7) becomes

$$l(u) = \int_0^\infty (e^{ux} - 1) \frac{\alpha e^{-\beta x}}{x} dx.$$

Comparing the above with (6.14) we find that $\gamma_0^{\mathrm{G}} = 0$ and the Lévy density of X^{G} is given by

$$\nu^{\mathrm{G}}(dx) = \frac{\alpha e^{-\beta x}}{x} \mathbb{1}_{\{x \geq 0\}} dx. \tag{8.8}$$

Since we have

$$\begin{aligned} \int_0^\infty \nu^{\mathrm{G}}(dx) &> \alpha \int_0^1 \frac{e^{-\beta x}}{x} \, dx \\ &> \alpha \int_0^1 \frac{e^{-\beta}}{x} \, dx \\ &= \alpha e^{-\beta} \int_0^1 \frac{1}{x} \, dx \\ &= \infty \end{aligned} \tag{8.9}$$

we see from Proposition 6.6.1 that gamma processes are infinite activity processes, that is, they have an infinite number of jumps arriving per unit time. We may calculate γ^{G} by making use of Proposition 6.7.2 as follows

$$\gamma^{\mathrm{G}} = \gamma_0^{\mathrm{G}} + \int_0^1 x \nu^{\mathrm{G}}(dx) = \int_0^1 \alpha e^{-\beta x} dx = \frac{\alpha}{\beta} \left(1 - e^{-\beta}\right). \tag{8.10}$$

Again using Proposition 6.7.2 we find that $\sigma^{\mathrm{G}} = 0$. Therefore, we conclude that the Lévy triplet of X^{G} is given by $\left((\sigma^{\mathrm{G}})^2, \nu^{\mathrm{G}}, \gamma^{\mathrm{G}}\right)$ where

$$\sigma^{\mathrm{G}} = 0, \quad \nu^{\mathrm{G}}(x) = \frac{\alpha e^{-\beta x}}{x} \mathbb{1}_{\{x > 0\}} \quad \text{and} \quad \gamma^{\mathrm{G}} = \frac{\alpha}{\beta} \left(1 - e^{-\beta}\right).$$

8.2 Variance Gamma Processes

As shown in Madan et al. [1998], a variance gamma process is constructed from Brownian motion and a gamma process. Suppose that $X = \{X_t\}_{t \geq 0}$ is an arithmetic Brownian motion that has drift θ and variance σ^2, then if W_t is standard Brownian motion, we have

$$X_t = \theta t + \sigma W_t.$$

If we evaluate the Brownian motion X_t at an independent random time which is a gamma process, we obtain the variance gamma process (see Madan [2001a, (7)], Schoutens [2003, p.58] or Applebaum [2004, Example 1.3.31] for example).

Definition 8.2.1 Variance Gamma Process
Let $\{X_t\}_{t \geq 0}$ be arithmetic Brownian motion with drift θ and variance σ^2. Also, let $\{X_t^G\}_{t \geq 0}$ be a gamma

process with shape parameter $\frac{1}{\nu}$ [2] *and rate parameter* $\frac{1}{\nu}$ [3]. *Then a* variance gamma (VG) process $X^{VG} = \{X_t^{VG}\}_{t\geq 0}$ *is defined by*

$$\begin{aligned} X_t^{VG} &= X_{X_t^G} \\ &= \theta X_t^G + \sigma W_{X_t^G}. \end{aligned}$$

The name 'variance gamma' comes from the fact that the X_t^{VG} results from replacing the variance of the normal random variable by a gamma random variable.

X_t^{VG} can also be expressed as the difference of two independent gamma processes (for details see Madan et al. [1998]). We will denote the random variable X_t^{VG} distributed VG with parameters θ, σ and ν as

$$X_t^{\text{VG}} \sim \text{VG}\left(\theta t, \sigma^2 t, \frac{\nu}{t}\right).$$

As in the gamma case, we list some results for VG processes in Table 8.2 which we will discuss subsequently.

Property	Expression/Value
Probability Density Function	$f_{X_t^{\text{VG}}}(x) = \frac{2\exp\left[\frac{\theta}{\sigma^2}x\right]}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)\sqrt{2\pi\sigma^2}}\left(\frac{x^2}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{\sqrt{x^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}}{\sigma^2}\right)$
Characteristic Function	$\Phi_{X_t^{\text{VG}}}(z) = \left(1-\nu\left(iz\theta-\frac{1}{2}\sigma^2 z^2\right)\right)^{-\frac{t}{\nu}}$
Activity	Infinite Activity
Variation	Finite Variation
Lévy Triplet	$\sigma^{\text{VG}} = 0$
	$\gamma_0^{\text{VG}} = 0$ and $\gamma^{\text{VG}} = \int_{\lvert x\rvert\leq 1} x\nu^{\text{VG}}(dx)$
	$\nu^{\text{VG}}(x) = \frac{\exp\left[\frac{\theta}{\sigma^2}x - \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}\lvert x\rvert\right]}{\nu\lvert x\rvert}$

Table 8.2: *Results for a VG process* X^{VG} *with* $x, z, \theta, \sigma, \nu \in \mathbb{R}$, $\sigma > 0$ *and* $\nu > 0$.

The probability density function of the VG process X^{VG} is derived by Madan et al. [1998].

Theorem 8.2.2 Density Function of X^{VG}
The density function of X_t^{VG} *is given by*

$$f_{X_t^{VG}}(x) = \frac{2\exp\left[\frac{\theta}{\sigma^2}x\right]}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)\sqrt{2\pi\sigma^2}}\left(\frac{x^2}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{\sqrt{x^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}}{\sigma^2}\right) \tag{8.11}$$

where $K_\cdot(\cdot)$ *is the modified Bessel function of the second kind (see Appendix D for a brief discussion on the modified Bessel function of the second kind).*

[2] Note that ν is a real number and is unrelated to the Lévy measure mentioned before. This notation, however, seems to be standard.

[3] It is desirable that the subordinator, on average, coincide with 'calendar' time t. Therefore, the shape and rate parameters of the gamma process are chosen to be equal, i.e. $\alpha = \frac{1}{\nu} = \beta$ and so $\mathbb{E}\left[X_t^{\text{G}}\right] = \frac{\alpha t}{\beta} = t$.

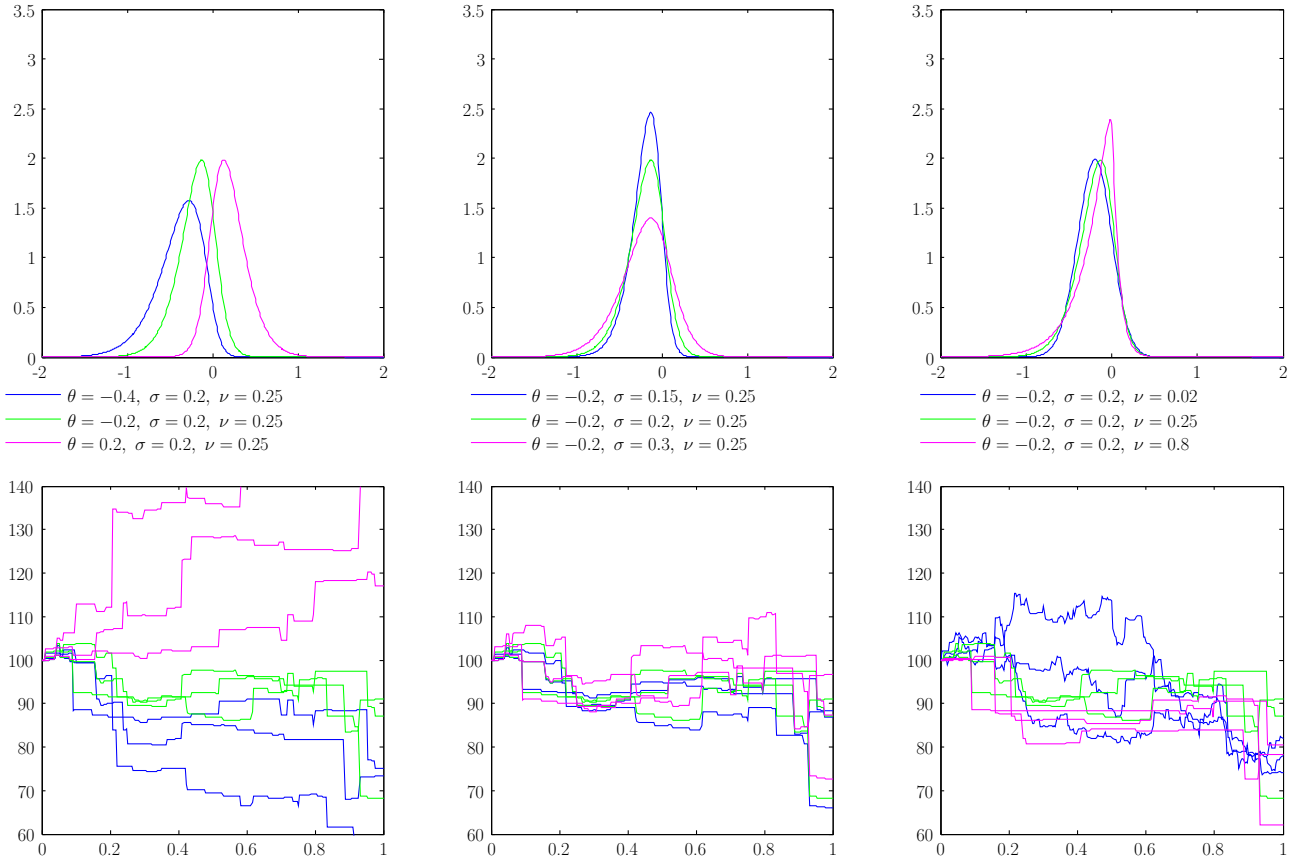

Figure 8.2: *VG probability density functions (top row) and corresponding process paths (bottom row) where one parameter is varied, while the other two are fixed. The density functions shown in green are the same in all three figures with parameters we used in common.*

We derive the characteristic function of X_t^{VG} for $z \in \mathbb{R}$ using conditional expectation:

$$\begin{aligned}
\Phi_{X_t^{\mathrm{VG}}}(z) := \mathbb{E}\left[e^{izX_t^{\mathrm{VG}}}\right] &= \mathbb{E}\left[e^{iz\left(\theta X_t^{\mathrm{G}} + \sigma W_{X_t^{\mathrm{G}}}\right)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{iz\left(\theta X_t^{\mathrm{G}} + \sigma W_{X_t^{\mathrm{G}}}\right)} \,\middle|\, X_t^{\mathrm{G}} = g\right]\right] \\
&= \int \mathbb{E}\left[e^{iz\left(\theta X_t^{\mathrm{G}} + \sigma W_{X_t^{\mathrm{G}}}\right)} \,\middle|\, X_t^{\mathrm{G}} = g\right] \mathbb{P}\left(X_t^{\mathrm{G}} \in dg\right) \\
&= \int e^{g\left(iz\theta - \frac{1}{2}z^2\sigma^2\right)} f_{X_t^{\mathrm{G}}}(g)\, dg \\
&= \mathbb{E}\left[e^{X_t^{\mathrm{G}}\left(iz\theta - \frac{1}{2}z^2\sigma^2\right)}\right] \\
&= \left(1 - \nu\left(iz\theta - \tfrac{1}{2}\sigma^2 z^2\right)\right)^{-\frac{t}{\nu}} \quad [4]. \qquad (8.12)
\end{aligned}$$

The third equality follows from taking the expectation of e^{izX_t}, where X_t is arithmetic Brownian motion, that is, the characteristic function of arithmetic Brownian motion given in (C.4). The last equality is found using the characteristic function of the gamma process X_t^{G}, as shown in (8.4) with $\alpha = \frac{1}{\nu} = \beta$. As expected, the VG distribution is infinitely divisible — this is clear from the linearity of the log of the characteristic function in the time variable.

Using Theorem 6.7.4 we see that $\sigma^{\mathrm{VG}} = 0$, since $\gamma_0^{\mathrm{G}} = 0$, and hence the VG process has no diffusion component. The VG process X^{VG} has Lévy measure (see [Madan et al., 1998, (14)])

$$\nu^{\mathrm{VG}}(x) = \frac{\exp\left[\frac{\theta}{\sigma^2}x - \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}|x|\right]}{\nu|x|}.$$

This is obtained using Theorem 6.7.4 as follows

$$\begin{aligned}
\nu^{\mathrm{VG}}(x) &= \int_0^\infty f_{X_g}(x)\nu^{\mathrm{G}}(dg) \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 g}} \exp\left[-\frac{(x-\theta g)^2}{2\sigma^2 g}\right] \frac{1}{\nu g} \exp\left[-\frac{g}{\nu}\right] dg \qquad (8.13)\\
&= \int_0^\infty \frac{1}{\sigma\nu\sqrt{2\pi g^3}} \exp\left[-\frac{x^2}{2\sigma^2 g} + \frac{x\theta g}{2\sigma^2 g} - \frac{\theta^2 g^2}{2\sigma^2 g} - \frac{g}{\nu}\right] dg \\
&= \frac{\exp\left[\frac{\theta}{2\sigma^2}x\right]}{\nu} \frac{|x|}{|x|} \int_0^\infty \frac{1}{\sigma\sqrt{2\pi g^3}} \exp\left[-\frac{x^2}{2\sigma^2 g} - \frac{\theta^2 g}{2\sigma^2} - \frac{g}{\nu}\right] dg \\
&= \frac{\exp\left[\frac{\theta}{2\sigma^2}x\right]}{\nu|x|} \frac{\exp\left[-\frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}|x|\right]}{\exp\left[-\frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}|x|\right]} \int_0^\infty \frac{x}{\sigma\sqrt{2\pi g^3}} \exp\left[-\frac{x^2}{2\sigma^2 g} - \frac{\theta^2 g}{2\sigma^2} - \frac{g}{\nu}\right] dg \\
&= \frac{\exp\left[\frac{\theta}{2\sigma^2}x - \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}|x|\right]}{\nu|x|} \int_0^\infty \frac{x}{\sigma\sqrt{2\pi g^3}} \exp\left[-\frac{x^2}{2\sigma^2 g} - \frac{\theta^2 g}{2\sigma^2} - \frac{g}{\nu} + \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}x\right] dg \\
&= \frac{\exp\left[\frac{\theta}{2\sigma^2}x - \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}|x|\right]}{\nu|x|},
\end{aligned}$$

since we may write

$$\begin{aligned}
&\frac{x}{\sigma\sqrt{2\pi g^3}} \exp\left[-\frac{x^2}{2\sigma^2 g} - \frac{\theta^2 g}{2\sigma^2} - \frac{g}{\nu} + \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}x\right] \mathbb{1}_{\{x>0\}} \\
&= \frac{x/\sigma}{\sqrt{2\pi g^3}} \exp\left[-\frac{1}{2}\left(\frac{x/\sigma - \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma}g}{\sqrt{g}}\right)\right] \mathbb{1}_{\{x>0\}},
\end{aligned}$$

[4] Again, as in the gamma case, the form in which $\Phi_{X_t^{\mathrm{VG}}}(z)$ is expressed in (8.12) may seem ambiguous since we are taking a real-valued power of a complex number which results in a multi-valued function. The single-valued form (see [Lord and Kahl, 2010, Theorem 4.1]) of (8.12) is given by

$$\Phi_{X_t^{\mathrm{VG}}}(z) = \exp\left[-\frac{t}{\nu}\mathrm{Ln}\left(1 - \nu\left(iz\theta - \tfrac{1}{2}\sigma^2 z^2\right)\right)\right].$$

which we see from (G.6) is the inverse Gaussian density with $x := g$, $\eta := \frac{x}{\sigma}$ and $\gamma := \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma}$. The Lévy measure has infinite mass and hence from Proposition 6.6.1 we have that a VG process has infinitely many jumps in any finite time interval. From Proposition 6.7.3 we see that the VG process is of finite variation since $\sigma^{\text{VG}} = 0$ and

$$\begin{aligned}
\int_{-1}^{1} |x|\nu^{\text{VG}}(dx) &= \frac{1}{\nu}\int_{-1}^{1} \exp\left[\frac{\theta}{\sigma^2}x - \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}|x|\right] dx \\
&= \frac{1}{\nu}\int_{-1}^{0} \exp\left[\frac{\theta}{\sigma^2}x + \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}x\right] dx + \frac{1}{\nu}\int_{0}^{1} \exp\left[\frac{\theta}{\sigma^2}x - \frac{\sqrt{\theta^2+2\sigma^2/\nu}}{\sigma^2}x\right] dx \\
&= \frac{1}{\nu}\int_{-1}^{0} \exp\left[\varphi_+ x\right] dx + \frac{1}{\nu}\int_{0}^{1} \exp\left[\varphi_- x\right] dx \\
&= -\frac{1}{\nu}\int_{1}^{0} \exp\left[-\varphi_+ y\right] dy + \frac{1}{\nu}\int_{0}^{1} \exp\left[\varphi_- x\right] dx \\
&= \frac{1}{\nu}\int_{0}^{1} \exp\left[-\varphi_+ y\right] dy + \frac{1}{\nu}\int_{0}^{1} \exp\left[\varphi_- x\right] dx \\
&< \infty,
\end{aligned}$$

where we have made the same change of variables as above. γ^{VG} in the Lévy triplet $\left(\sigma^2, \nu, \gamma\right)$ is found using Theorem 6.7.4 as follows

$$\begin{aligned}
\gamma^{\text{VG}} &= \int_0^\infty \nu^{\text{G}}(dg) \int_{|x|\leq 1} x f_{X_g}(dx) \\
&= \int_0^\infty \frac{\exp\left[-\frac{g}{\nu}\right]}{\nu g} \int_{|x|\leq 1} x \frac{1}{\sqrt{2\pi\sigma^2 g}} \exp\left[-\frac{(x-\theta g)^2}{\sigma^2 g}\right] dx\, dg \\
&= \int_{|x|\leq 1} x \int_0^\infty \frac{1}{\sqrt{2\pi\sigma^2 g}} \exp\left[-\frac{(x-\theta g)^2}{2\sigma^2 g}\right] \frac{1}{\nu g} \exp\left[-\frac{g}{\nu}\right] dg\, dx \\
&= \int_{|x|\leq 1} x\nu^{\text{VG}}(dx),
\end{aligned}$$

where the third equality is obtained by using Fubini's Theorem and the last equality by making use of (8.13). Thus, the Lévy triplet of a VG process is given by $\left((\sigma^{\text{VG}})^2, \nu^{\text{VG}}, \gamma^{\text{VG}}\right)$, where

$$\sigma^{\text{VG}} = 0, \quad \nu^{\text{VG}}(x) = \frac{\exp\left[\frac{\theta}{\sigma^2}x - \frac{\sqrt{\theta^2+\frac{2\sigma^2}{\nu}}}{\sigma^2}|x|\right]}{\nu|x|} \quad \text{and} \quad \gamma^{\text{VG}} = \int_{|x|\leq 1} x\nu^{\text{VG}}(dx).$$

8.3 Bridge Sampling

As in §7.5.2, where we considered Brownian bridge sampling, we now discuss similar bridge sampling for the gamma distribution. This section requires the generation of standard normal, gamma and beta random variables. Given a uniform random number, we obtain:

- a standard normal random number by applying Moro's method, as mentioned in §7.2;

- a gamma variate by applying the method discussed in Appendix H to the cumulative gamma function discussed in §G.1; and
- a beta variate by again applying the method discussed in Appendix H but this time to the cumulative beta function discussed in §G.2.

As noted in §8.2, the VG process can be expressed in two forms — as a gamma time-changed Brownian motion and as the difference of two gamma processes. And so, when simulating a VG process, either of these two forms can be used. We will only consider the time-changed Brownian motion case.

8.3.1 Gamma Bridge Sampling

In a similar manner to §7.5.2 we derive the conditional distribution of gamma increments.

As before, consider three consecutive times $s < t < u$ and suppose that x, y and z are realisations of the gamma increments $X_t^{\text{G}} - X_s^{\text{G}}$, $X_u^{\text{G}} - X_t^{\text{G}}$ and $X_u^{\text{G}} - X_s^{\text{G}}$ respectively (see Figure 8.3). Here X_s^{G} and X_u^{G} are points which have already been determined and t is the position at which the next bridge point is to be created.

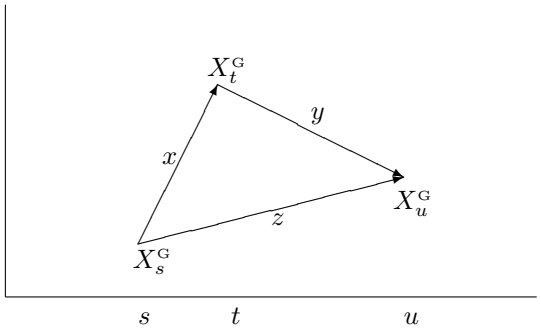

Figure 8.3: *Gamma Bridge*

In order to find the conditional distribution of the gamma increment $X_t^{\text{G}} - X_s^{\text{G}} = x$ given that we know the increment $X_u^{\text{G}} - X_s^{\text{G}} = z$, we make use of the density of the gamma process $X^{\text{G}} = \{X_t^{\text{G}}\}_{t \geq 0}$

$$f_{X_t^{\text{G}}}(x) = \frac{x^{\frac{t}{\nu}-1} e^{-\frac{1}{\nu}x}}{\nu^{\frac{t}{\nu}} \Gamma\left(\frac{t}{\nu}\right)} \mathbb{1}_{\{x>0\}}$$

and again apply Bayes' Theorem as in §7.5.2

$$\begin{aligned}
f_{\left(X^{\mathrm{G}}_t - X^{\mathrm{G}}_s | X^{\mathrm{G}}_u - X^{\mathrm{G}}_s = z\right)}(x) &= \frac{f_{(X^{\mathrm{G}}_u - X^{\mathrm{G}}_s | X^{\mathrm{G}}_t - X^{\mathrm{G}}_s = x)}(z) f_{X^{\mathrm{G}}_t - X^{\mathrm{G}}_s}(x)}{f_{X^{\mathrm{G}}_u - X^{\mathrm{G}}_s}(z)} \\
&= \frac{f_{X^{\mathrm{G}}_u - X^{\mathrm{G}}_t + x}(z) f_{X^{\mathrm{G}}_t - X^{\mathrm{G}}_s}(x)}{f_{X^{\mathrm{G}}_u - X^{\mathrm{G}}_s}(z)} \\
&= \frac{\frac{(z-x)^{\frac{u-t}{\nu}-1} e^{-\frac{1}{\nu}(z-x)}}{\nu^{\frac{u-t}{\nu}} \Gamma\left(\frac{u-t}{\nu}\right)} \mathbb{1}_{\{z-x>0\}} \frac{x^{\frac{t-s}{\nu}-1} e^{-\frac{1}{\nu}x}}{\nu^{\frac{t-s}{\nu}} \Gamma\left(\frac{t-s}{\nu}\right)} \mathbb{1}_{\{x>0\}}}{\frac{z^{\frac{u-s}{\nu}-1} e^{-\frac{1}{\nu}z}}{\nu^{\frac{u-s}{\nu}} \Gamma\left(\frac{u-s}{\nu}\right)} \mathbb{1}_{\{z>0\}}} \\
&= \frac{1}{z} \frac{\Gamma\left(\frac{u-t}{\nu} + \frac{t-s}{\nu}\right)}{\Gamma\left(\frac{u-t}{\nu}\right) \Gamma\left(\frac{t-s}{\nu}\right)} \left(\frac{x}{z}\right)^{\frac{t-s}{\nu}-1} \left(1 - \frac{x}{z}\right)^{\frac{u-t}{\nu}-1} \mathbb{1}_{\{0<\frac{x}{z}<1\}}.
\end{aligned}$$

We have derived the expression in Ribeiro and Webber [2004, (19)].

The inspired step is to now apply Example F.1.2 by setting $\lambda = \frac{1}{z}$ and $g(X) = B = \frac{X}{z}$, then $f_B(b) = f_X(zx)z$, and

$$f_B(b) = \frac{\Gamma\left(\frac{u-t}{\nu} + \frac{t-s}{\nu}\right)}{\Gamma\left(\frac{u-t}{\nu}\right) \Gamma\left(\frac{t-s}{\nu}\right)} b^{\frac{t-s}{\nu}-1} (1-b)^{\frac{u-t}{\nu}-1} \mathbb{1}_{\{0<b<1\}}.$$

That is, B has a beta distribution with parameters $\alpha := \frac{t-s}{\nu}$ and $\beta := \frac{u-t}{\nu}$. Thus if we draw a random variable

$$b \sim \text{Beta}\left(\frac{t-s}{\nu}, \frac{u-t}{\nu}\right)$$

then we may write

$$\frac{X^{\mathrm{G}}_t - X^{\mathrm{G}}_s}{X^{\mathrm{G}}_u - X^{\mathrm{G}}_s} = \frac{x}{z} = b \tag{8.14}$$

$$\Rightarrow X^{\mathrm{G}}_t = bX^{\mathrm{G}}_u + (1-b)X^{\mathrm{G}}_s. \tag{8.15}$$

Thus, X^{G}_t is interpolated between X^{G}_s and X^{G}_u in a random manner.

8.3.2 Time-Changed Brownian Motion

We begin the gamma bridge sampling by simulating the gamma process at terminal time. This is achieved by generating a gamma variate

$$X^{\mathrm{G}}_{t_M} \sim \text{Gamma}\left(\frac{t_M}{\nu}, \frac{1}{\nu}\right)$$

with the first uniform random number from our quasi random sequence and then generating a time-changed Wiener process

$$W_{X^{\mathrm{G}}_{t_M}} \sim \text{Normal}\left(0, X^{\mathrm{G}}_{t_M}\right)$$

with the second. Next, in a manner similar to Brownian bridge sampling, we use the algorithm of §7.5.1 to choose the first point t_j at which a bridge point must be created with bridge points $W_{X^{\mathrm{G}}_0} = 0$ and $W_{X^{\mathrm{G}}_{t_M}}$ at times t_0 and t_M.

We next proceed as in the Brownian motion case, except that instead of using one random number from the quasi random sequence, we use two random numbers at each step. For the general inductive step, let $s < t < u$ be consecutive time steps. Suppose $W_{X_s^{\text{G}}}$ and $W_{X_u^{\text{G}}}$ are known and t is the point at which the next bridge is to be created. Then, in order to find $W_{X_t^{\text{G}}}$, we substitute the gamma process at the appropriate times in (7.9) and obtain

$$W_{X_t^{\text{G}}} = \frac{X_t^{\text{G}} - X_s^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} W_{X_u^{\text{G}}} + \frac{X_u^{\text{G}} - X_t^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} W_{X_s^{\text{G}}} + \sqrt{\frac{(X_t^{\text{G}} - X_s^{\text{G}})\,(X_u^{\text{G}} - X_t^{\text{G}})}{X_u^{\text{G}} - X_s^{\text{G}}}}\, n, \tag{8.16}$$

where n is a standard normal random variable.

Now, we use the result (8.15) obtained from deriving the conditional density of gamma increments in order to write X_t^{G} in (8.16) as a function of a beta variate. We write (8.16) as

$$W_{X_t^{\text{G}}} = bW_{X_u^{\text{G}}} + (1-b)W_{X_s^{\text{G}}} + \sqrt{(1-b)b\,(X_u^{\text{G}} - X_s^{\text{G}})}\, n, \tag{8.17}$$

where b is a draw from a beta distribution with parameters $\alpha := \frac{t-s}{\nu}$ and $\beta := \frac{u-t}{\nu}$. Thus, using the next two quasi random numbers from our sequence we find b and n in order to calculate $W_{X_t^{\text{G}}}$. This completes the inductive step.

The coefficient of n, namely $\sqrt{(1-b)b\,(X_u^{\text{G}} - X_s^{\text{G}})}$, can also be written as $\sqrt{b\,(X_u^{\text{G}} - X_t^{\text{G}})}$, but it is not in the correct mathematical form. After unpacking the notation, the latter form is that of Fu [2007, Fig. 2] and Avramidis et al. [2004, Figure 3]. An alternative form is $\sqrt{\frac{(X_t^{\text{G}} - X_s^{\text{G}})(X_u^{\text{G}} - X_t^{\text{G}})}{X_u^{\text{G}} - X_s^{\text{G}}}}$, however if $X_u^{\text{G}} - X_s^{\text{G}} = 0$ (which can occur to machine precision) there will be a division by 0.

In a similar way to finding arithmetic Brownian motion in (7.10), we simulate a VG process $X^{\text{VG}} = \{X_t^{\text{VG}}\}_{t \geq 0}$, where $X_t^{\text{VG}} \sim \text{VG}\left(\theta t, \sigma^2 t, \frac{\nu}{t}\right)$, by multiplying $W_{X_t^{\text{G}}}$ in (8.16) with σ and adding θX_t^{G}

$$\begin{aligned}
X_t^{\text{VG}} &= \theta X_t^{\text{G}} + \sigma W_{X_t^{\text{G}}} \\
&= \theta X_t^{\text{G}} + \sigma \left(\frac{X_t^{\text{G}} - X_s^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} W_{X_u^{\text{G}}} + \frac{X_u^{\text{G}} - X_t^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} W_{X_s^{\text{G}}} + \sqrt{\frac{(X_t^{\text{G}} - X_s^{\text{G}})\,(X_u^{\text{G}} - X_t^{\text{G}})}{X_u^{\text{G}} - X_s^{\text{G}}}}\, n \right) \\
&= \frac{X_t^{\text{G}} - X_s^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} \theta X_u^{\text{G}} + \frac{X_u^{\text{G}} - X_t^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} \theta X_s^{\text{G}} \\
&\quad + \sigma \left(\frac{X_t^{\text{G}} - X_s^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} W_{X_u^{\text{G}}} + \frac{X_u^{\text{G}} - X_t^{\text{G}}}{X_u^{\text{G}} - X_s^{\text{G}}} W_{X_s^{\text{G}}} + \sqrt{\frac{(X_t^{\text{G}} - X_s^{\text{G}})\,(X_u^{\text{G}} - X_t^{\text{G}})}{X_u^{\text{G}} - X_s^{\text{G}}}}\, n \right) \\
&= bX_u^{\text{VG}} + (1-b)X_s^{\text{VG}} + \sigma\sqrt{(1-b)b\,(X_u^{\text{G}} - X_s^{\text{G}})}\, n.
\end{aligned} \tag{8.18}$$

As before we may construct our algorithm for creating a VG process in either of two ways: generate gamma bridges for all times and then find their corresponding VG process, or for each gamma bridge $W_{X_t^{\text{G}}}$ calculate its corresponding VG process X_t^{VG} immediately.

Chapter 9

The Normal Inverse Gaussian Model

The normal inverse Gaussian distribution is a type of generalised hyperbolic distribution which was introduced by Barndorff-Nielsen [1977]. The hyperbolic distribution, which was introduced into finance by Eberlein and Keller [1995] (see also Eberlein et al. [1998]), is another well-known type of generalised hyperbolic distribution. The use of normal inverse Gaussian processes in financial modelling was first proposed by Barndorff-Nielsen [1995]. Further studies on the normal inverse Gaussian distribution can be found in Barndorff-Nielsen [1997, 1998] and Rydberg [1996a,b, 1997].

Several authors on the subject comment on how well the normal inverse Gaussian distribution fit the log returns of stocks. In particular, Rydberg [1997] shows that Danish and German financial data fit excellently to the normal inverse Gaussian distribution. Korn et al. [2010] mention that the reason for this is because the normal inverse Gaussian distribution is more flexible than the normal distribution. Moreover, Korn et al. [2010] note that, while having the same mean and variance, the normal inverse Gaussian distribution can generate higher peaks and at the same time heavier tails than the normal distribution.

Like the variance gamma process, the normal inverse Gaussian process satisfies the general property of being time-changed Brownian motion [see Barndorff-Nielsen, 1998]. Here the time-change may be chosen as an inverse Gaussian process independent of the directing Brownian motion.

9.1 Inverse Gaussian Processes

The name 'inverse Gaussian' was first used in Tweedie [1947] as the inverse relationship between the cumulant generating functions of these distributions, and those of Gaussian distributions. The density of the Wald distribution which is a special case of the inverse Gaussian distribution can be seen in Johnson et al. [1994, Chapter 15, §2]. By performing some substitutions, the standard form of the two-parameter inverse Gaussian distribution is obtained. This is the distribution we will consider in this section and we will refer to it as the inverse Gaussian distribution. The inverse Gaussian process describes the distribution of the time Brownian motion with positive drift takes to reach a fixed positive level (see Applebaum [2004, Example 1.3.21] or Kyprianou [August 2007, §2.5] for example).

Definition 9.1.1 Inverse Gaussian Processes
An inverse Gaussian process $X^{IG} = \{X_t^{IG}\}_{t\geq 0}$ *is defined by*

$$X_t^{IG} = \inf_{s>0}\{\gamma s + W_s = \eta t\} \tag{9.1}$$

where η, $\gamma > 0$ *and* $\{W_s\}_{s\geq 0}$ *is standard Brownian motion.*

In Table 9.1 we list some of the results concerning inverse Gaussian processes which will be discussed subsequently.

Property	Expression/Value
Probability Density Function	$f_{X_t^{\text{IG}}}(x) = \frac{\eta t}{\sqrt{2\pi x^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta t - \gamma x}{\sqrt{x}}\right)^2\right] \mathbb{1}_{\{x>0\}}$
Characteristic Function	$\Phi_{X_t^{\text{IG}}}(z) = \exp\left[\eta t\left(\gamma - \sqrt{\gamma^2 - 2iz}\right)\right]$
Activity	Infinite Activity
Lévy Triplet	$\sigma^{\text{IG}} = 0$
	$\gamma_0^{\text{IG}} = 0$ and $\gamma^{\text{IG}} = \frac{\eta}{\gamma}(2F_{\text{N}}(\gamma) - 1)$
	$\nu^{\text{IG}}(x) = \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x} \mathbb{1}_{\{x>0\}}$

Table 9.1: *Results for an inverse Gaussian process* X^{IG} *with* $x, z, \eta, \gamma \in \mathbb{R}$ *and* η, $\gamma > 0$*; and* $F_N(\cdot)$ *indicating the standard normal distribution function.*

If X^{IG} is an inverse Gaussian process with parameters η, $\gamma > 0$, then we write $X_t^{\text{IG}} \sim \text{IG}(\eta t, \gamma)$ and X_t^{IG} has density

$$f_{X_t^{\text{IG}}}(x) = \frac{\eta t}{\sqrt{2\pi x^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta t - \gamma x}{\sqrt{x}}\right)^2\right] \mathbb{1}_{\{x>0\}} \tag{9.2}$$

(see Appendix G.3 for a brief discussion on the inverse Gaussian probability density function).

We calculate the moment generating function of the inverse Gaussian process X_t^{IG} as follows

$$\begin{aligned} M_{X_t^{\text{IG}}}(u) &= \mathbb{E}\left[e^{uX_t^{\text{IG}}}\right] \\ &= \frac{\eta t}{\sqrt{2\pi}} \int_0^\infty e^{ux} \frac{1}{\sqrt{x^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta t - \gamma x}{\sqrt{x}}\right)^2\right] dx \\ &= \frac{\eta t}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{x^3}} \exp\left[-\frac{1}{2}\frac{\eta^2 t^2 - 2\gamma\eta t x + \gamma^2 x^2 - 2ux^2}{x}\right] dx. \end{aligned}$$

Completing the square in the exponent for $\sqrt{\gamma^2 - 2u} > 0$ gives

$$\begin{aligned} -\frac{1}{2}\frac{\eta^2 t^2 - 2\gamma\eta t x + \gamma^2 x^2 - 2ux^2}{x} &= -\frac{1}{2}\frac{\eta^2 t^2 - 2\gamma\eta t x + \left(\gamma^2 - 2u\right)x^2}{x} \\ &= -\frac{1}{2}\left(\frac{\eta t - \sqrt{\gamma^2 - 2uz}x}{\sqrt{x}}\right)^2 + \eta t\gamma - \eta t\sqrt{\gamma^2 - 2u} \end{aligned}$$

and hence we may write

$$M_{X_t^{\text{IG}}}(u) = \exp\left[\eta t\left(\gamma - \sqrt{\gamma^2 - 2u}\right)\right] \frac{\eta t}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{x^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta t - \sqrt{\gamma^2 - 2ux}}{\sqrt{x}}\right)^2\right] dx.$$

Now,

$$\frac{\eta t}{\sqrt{2\pi}} \int_0^\infty \frac{1}{\sqrt{x^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta t - \sqrt{\gamma^2 - 2ux}}{\sqrt{x}}\right)^2\right] dx = 1$$

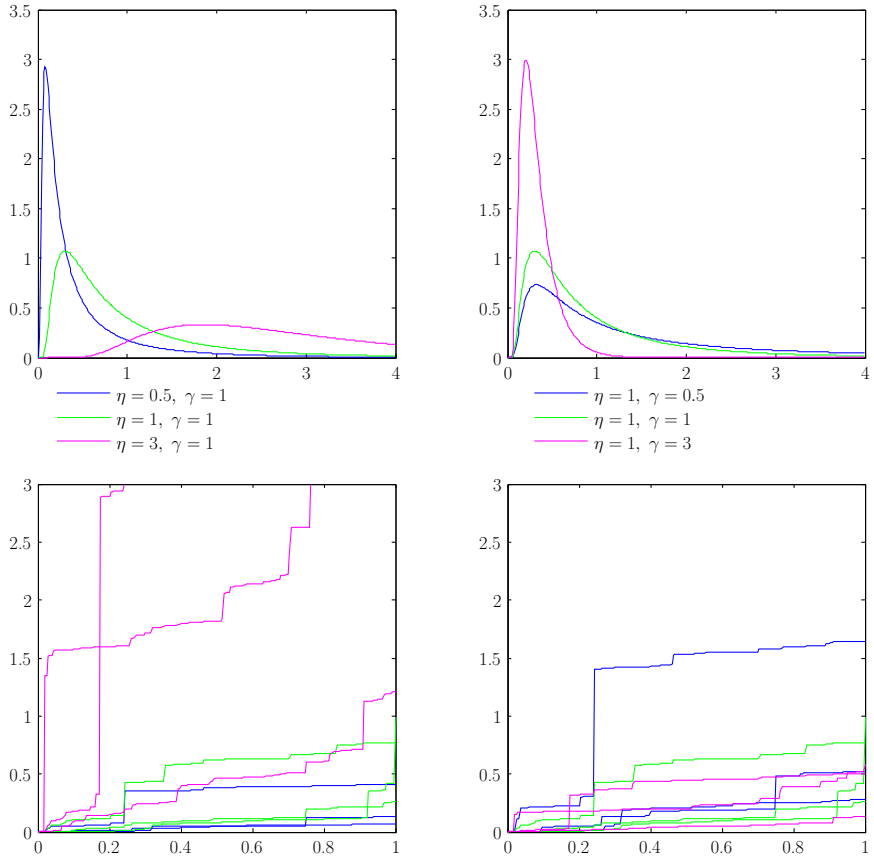

Figure 9.1: *In the first row we plot the inverse Gaussian probability density functions for various values of η and γ where one parameter is fixed and the other is varied. The graphs shown in green are the same in both figures with parameters used in common.*

In the second row, we simulate three realisations of three inverse Gaussian processes with parameter values corresponding to the densities in the first row. We see that as the values of η increase, the frequency of the jumps increase. Furthermore, the jump sizes decrease and the jump frequencies increase as the values of γ increase.

since it is the density function of a random variable distributed $\text{IG}\left(\eta t, \sqrt{\gamma^2 - 2u}\right)$. Thus we have that

$$M_{X_t^{\text{IG}}}(u) = \exp\left[\eta t\left(\gamma - \sqrt{\gamma^2 - 2u}\right)\right] \tag{9.3}$$

for $u < \frac{\gamma^2}{2}$. From (9.3) we get the Laplace exponent of X_t^{IG} for $u < \frac{\gamma^2}{2}$

$$l(u) = \eta\left(\gamma - \sqrt{\gamma^2 - 2u}\right). \tag{9.4}$$

Kyprianou [August 2007, Exercise 6] notes that the characteristic function of the inverse Gaussian process X_t^{IG} can be found by considering

$$\mathbb{E}\left[e^{-uX_t^{\text{IG}}}\right] = \exp\left[-\eta t\left(-\gamma + \sqrt{\gamma^2 + 2u}\right)\right].$$

If we replace u with $a - iz$ where $a > 0$ and $z \in \mathbb{R}$, then both sides of the equation can be shown to be analytical functions[1] and therefore they agree on the parameter range. Taking limits as a tends to 0 shows that both functions agree when we replace u with iz, $z \in \mathbb{R}$. Thus the characteristic function of X_t^{IG} is given by

$$\Phi_{X_t^{\text{IG}}}(z) = \exp\left[\eta t\left(\gamma - \sqrt{\gamma^2 - 2iz}\right)\right], \tag{9.5}$$

where $z \in \mathbb{R}$.

As in the case of the gamma process, the linearity of the log of the characteristic function with respect to t, shows that X^{IG} is infinitely divisible. Thus, as we have seen in §6.2, there exists a Lévy process $\{X_t\}_{t\geq 0}$ such that the distribution of X_1 is determined by the distribution of X_1^{IG}. It is clear that X^{IG} has a.s. non-decreasing paths and is therefore by Definition 6.7.1 a subordinator.

The following two results enable us to write the Lévy exponent of X_t^{IG}, that is, $\eta\left(\gamma - \sqrt{\frac{1}{2}\gamma^2 - iz}\right)$ in the form of (6.13): For $u > 0$ and $0 < \alpha < 1$

$$\Gamma(-\alpha)\, u^{\alpha} = \int_0^{\infty} \left(e^{-ux} - 1\right) x^{-\alpha - 1}\, dx \tag{9.6}$$

[see Kyprianou, August 2007, Exercise 4]. This equation holds when $-u$ is replaced by any complex number $w \neq 0$ where $\Re(w) \leq 0$ [see Kyprianou, August 2007, Exercise 4].

Consider the Lévy exponent of X_t^{IG}

$$\begin{aligned}
\eta\left(\gamma - \sqrt{\gamma^2 - 2iz}\right) &= \eta\frac{\Gamma\left(-\frac{1}{2}\right)}{\Gamma\left(-\frac{1}{2}\right)}\gamma - \eta\frac{\Gamma\left(-\frac{1}{2}\right)}{\Gamma\left(-\frac{1}{2}\right)}\sqrt{\gamma^2 - 2iz} \\
&= -\eta\frac{\Gamma\left(-\frac{1}{2}\right)}{2\sqrt{\pi}}\gamma + \eta\frac{\Gamma\left(-\frac{1}{2}\right)}{2\sqrt{\pi}}\sqrt{\gamma^2 - 2iz} \\
&= -\eta\frac{\Gamma\left(-\frac{1}{2}\right)}{\sqrt{2\pi}}\frac{\gamma^2}{\sqrt{2}} + \eta\frac{\Gamma\left(-\frac{1}{2}\right)}{\sqrt{2\pi}}\sqrt{\frac{\gamma^2 - 2iz}{2}} \\
&= -\frac{\eta}{\sqrt{2\pi}}\Gamma\left(-\tfrac{1}{2}\right)\left(\tfrac{1}{2}\gamma^2\right)^{1/2} + \frac{\eta}{\sqrt{2\pi}}\Gamma\left(-\tfrac{1}{2}\right)\left(\tfrac{1}{2}\gamma^2 - iz\right)^{1/2} \\
&= -\frac{\eta}{\sqrt{2\pi}}\Gamma\left(-\tfrac{1}{2}\right)\left(\tfrac{1}{2}\gamma^2\right)^{1/2} + \frac{\eta}{\sqrt{2\pi}}\Gamma\left(-\tfrac{1}{2}\right)\left(\tfrac{1}{2}\gamma^2 - iz\right)^{1/2},
\end{aligned} \tag{9.7}$$

[1] A complex function is said to be analytic on a region A if it is complex differentiable at every point in A.

where we have made use of the identity [Abramowitz and Stegun, 1974, 6.1.8]

$$\Gamma\left(\tfrac{1}{2}\right) = \sqrt{\pi} \Rightarrow \left(\tfrac{1}{2}-1\right)\Gamma\left(\tfrac{1}{2}-1\right) = \sqrt{\pi} \Rightarrow \Gamma\left(-\tfrac{1}{2}\right) = -2\sqrt{\pi}.$$

We now apply (9.6) by setting $\alpha := \frac{1}{2}$, $u := \frac{1}{2}\gamma^2$ for the first term and $u := \frac{1}{2}\gamma^2 - iz$ for the last term in (9.7) and obtain

$$\begin{aligned}
&\eta\left(\gamma - \sqrt{\gamma^2 - 2iz}\right) \\
&= -\int_0^\infty \frac{\eta}{\sqrt{2\pi}}\left(e^{-\gamma^2/2x} - 1\right) x^{-3/2}\, dx + \int_0^\infty \frac{\eta}{\sqrt{2\pi}}\left(e^{-(\gamma^2/2 - iz)x} - 1\right) x^{-3/2}\, dx \\
&= -\int_0^\infty \frac{\eta}{\sqrt{2\pi}} e^{-\frac{1}{2}\gamma^2 x} x^{-3/2}\, dx + \int_0^\infty \frac{\eta}{\sqrt{2\pi}} x^{-3/2}\, dx + \int_0^\infty \frac{\eta}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}\gamma^2 - iz\right)x} x^{-3/2}\, dx - \int_0^\infty \frac{\eta}{\sqrt{2\pi}} x^{-3/2}\, dx \\
&= -\int_0^\infty \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x}\, dx + \int_0^\infty \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x + izx}\, dx \\
&= \int_0^\infty \left(e^{izx} - 1\right) \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x}\, dx.
\end{aligned}$$

When comparing the above with (6.13) we find the $\gamma_0^{\text{IG}} = 0$ and that the Lévy measure of X_t^{IG} is given by

$$\nu^{\text{IG}}(dx) = \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x} \mathbb{1}_{\{x>0\}}\, dx.$$

From Proposition 6.6.1 we see that the inverse Gaussian process is of infinite activity, since we have that

$$\begin{aligned}
\int_0^\infty \nu^{\text{IG}}(dx) &= \int_0^\infty \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x}\, dx \\
&> \int_0^1 \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2}\, dx \\
&= \frac{\eta}{\sqrt{2\pi}} e^{-\frac{1}{2}\gamma^2}\left[-\frac{2}{\sqrt{x}}\right]_0^1 \\
&= \infty.
\end{aligned}$$

The last component of the Lévy triplet γ^{IG} is found using Proposition 6.7.2 as follows

$$\begin{aligned}
\gamma^{\text{IG}} &= \gamma_0^{\text{IG}} + \int_0^1 x\nu^{\text{IG}}(dx) \\
&= \int_0^1 x \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x}\, dx \\
&= \frac{\eta}{\sqrt{2\pi}} \int_0^1 \frac{1}{\sqrt{x}} e^{-\frac{1}{2}\gamma^2 x}\, dx \\
&= \frac{\eta}{\sqrt{2\pi}} \int_0^\gamma \frac{\gamma}{y} e^{-\frac{1}{2}y^2} \frac{2y}{\gamma^2}\, dy \\
&= \frac{2\eta}{\gamma} \int_0^\gamma \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2}\, dy \\
&= \frac{2\eta}{\gamma}\left(F_{\text{N}}(\gamma) - F_{\text{N}}(0)\right) \\
&= \frac{\eta}{\gamma}\left(2F_{\text{N}}(\gamma) - 1\right),
\end{aligned}$$

where we have made a change of variables in the fourth equality by setting $x := \frac{y^2}{\gamma^2}$ and $F_{\text{N}}(\cdot)$ indicates the standard normal distribution function. Finally, from Proposition 6.7.2 we see that $\sigma^{\text{IG}} = 0$ and hence the Lévy triplet is given by $\left((\sigma^{\text{IG}})^2, \nu^{\text{IG}}, \gamma^{\text{IG}}\right)$, where

$$\sigma^{\text{IG}} = 0, \quad \nu^{\text{IG}}(x) = \frac{\eta}{\sqrt{2\pi x^3}} e^{-\frac{1}{2}\gamma^2 x} \mathbb{1}_{\{x>0\}} \quad \text{and} \quad \gamma^{\text{IG}} = \frac{\eta}{\gamma}\left(2F_{\text{N}}(\gamma) - 1\right).$$

9.2 Normal Inverse Gaussian Processes

The normal inverse Gaussian process can be defined as an inverse Gaussian time-changed Brownian motion (see Barndorff-Nielsen [1998] or Applebaum [2004, Example 1.3.32]).

Definition 9.2.1 Normal Inverse Gaussian Process
If we let $\eta = 1$ [2] *in Definition 9.1.1, then we obtain*

$$X_t^{IG} = \inf_{s>0}\{X_s = t\}$$

with $X_s = \gamma s + W_s$ *for* $\gamma \in \mathbb{R}$. *Then the* normal inverse Gaussian (NIG) process $X^{NIG} = \{X_t^{NIG}\}_{t\geq 0}$ *is obtained from an inverse Gaussian time-changed Brownian motion with drift* μ *and volatility* σ. *That is,*

$$X_t^{NIG} = \mu X_t^{IG} + \sigma W_{X_t^{IG}}.$$

If we let $\beta = \frac{\mu}{\sigma^2}$, $\alpha^2 = \frac{\gamma^2}{\sigma^2} + \frac{\mu^2}{\sigma^4}$ and $\delta = \sigma$, X_t^{NIG} can be written as

$$X_t^{\text{NIG}} = \beta\delta^2 X_t^{\text{IG}} + \delta W_{X_t^{\text{IG}}} \tag{9.8}$$

with X_t^{IG} having parameters $\eta = 1$ and $\gamma = \delta\sqrt{\alpha^2 - \beta^2}$. We denote an NIG process $X^{\text{NIG}} = \{X_t^{\text{NIG}}\}_{t\geq 0}$ with parameters α, β and δ by

$$X_t^{\text{NIG}} \sim \text{NIG}\left(\alpha, \beta, \delta t\right)$$

where $\alpha > 0$, $-\alpha < \beta < \alpha$ and $\delta > 0$. The parameter α indicates the tail heaviness of steepness, β indicates symmetry and δ is a scale parameter.

As in the inverse Gaussian case, we list some results concerning NIG processes in Table 9.2 which will be discussed later.

The density function of X_t^{NIG} is given by Barndorff-Nielsen [1998, 2.2]

$$f_{X_t^{\text{NIG}}}(x) = \frac{\alpha\delta t}{\pi} \exp\left[\delta t\sqrt{\alpha^2 - \beta^2} + \beta x\right] \frac{K_1\left(\alpha\sqrt{\delta^2 t^2 + x^2}\right)}{\sqrt{\delta^2 t^2 + x^2}}, \tag{9.9}$$

where $K_\cdot(\cdot)$ indicates the modified Bessel function of the second kind (see Appendix D for a brief discussion on modified Bessel functions of the second kind).

[2] As in the VG case, we would like the subordinator X^{IG}, on average, to coincide with 'calendar' time t. Thus we choose $\eta = 1$, so that $\mathbb{E}\left[X_t^{\text{IG}}\right] = \eta t = t$.

Property	Expression/Value
Probability Density Function	$f_{X_t^{\text{NIG}}}(x) = \frac{\alpha\delta t}{\pi} \exp\left[\delta t\sqrt{\alpha^2-\beta^2} + \beta x\right] \frac{K_1\left(\alpha\sqrt{\delta^2 t^2 + x^2}\right)}{\sqrt{\delta^2 t^2 + x^2}}$
Characteristic Function	$\Phi_{X_t^{\text{NIG}}}(z) = \exp\left[-\delta t\left(\sqrt{\alpha^2-(\beta+iz)^2} - \sqrt{\alpha^2-\beta^2}\right)\right]$
Activity	Infinite Activity
Variation	Infinite Variation
Lévy Triplet	$\sigma^{\text{NIG}} = 0$
	$\gamma_0^{\text{NIG}} = 0$ and $\gamma^{\text{NIG}} = \frac{2\delta\alpha}{\pi}\int_0^1 \sinh(\beta x) K_1(\alpha x)\, dx$
	$\nu^{\text{NIG}}(x) = \frac{\delta\alpha e^{\beta x} K_1(\alpha\lvert x\rvert)}{\pi\lvert x\rvert}$

Table 9.2: *Results for an NIG process X^{NIG} with $x, z, \alpha, \beta, \delta \in \mathbb{R}$, $\alpha > 0$, $-\alpha < \beta < \alpha$ and $\delta > 0$.*

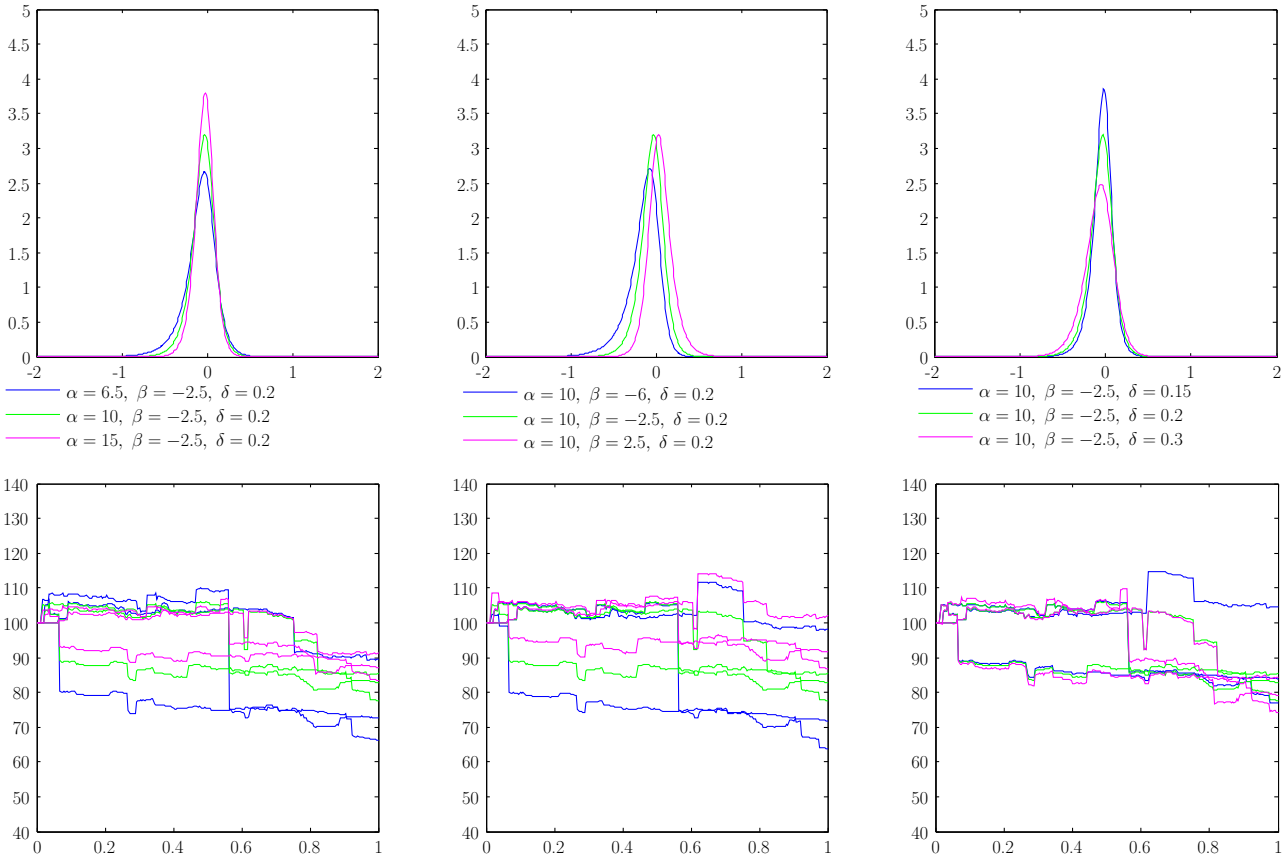

Figure 9.2: *NIG probability density functions (top row) and corresponding process paths (bottom row) where one parameter is varied while the other two are fixed. The density functions shown in green are the same in all three figures with parameters we used in common.*

As in the VG case, we derive the characteristic function using conditional expectation

$$\begin{aligned}
\Phi_{X_t^{\text{NIG}}}(z) := \mathbb{E}\left[e^{izX_t^{\text{NIG}}}\right] &= \mathbb{E}\left[e^{iz\left(\mu X_t^{\text{IG}}+\sigma W_{X_t^{\text{IG}}}\right)}\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[e^{iz\left(\mu X_t^{\text{IG}}+\sigma W_{X_t^{\text{IG}}}\right)}\middle| X_t^{\text{IG}}=g\right]\right] \\
&= \int \mathbb{E}\left[e^{iz\left(\mu X_t^{\text{IG}}+\sigma W_{X_t^{\text{IG}}}\right)}\middle| X_t^{\text{IG}}=g\right]\mathbb{P}\left(X_t^{\text{IG}}\in dg\right) \\
&= \int e^{g\left(iz\mu-\frac{1}{2}z^2\sigma^2\right)} f_{X_t^{\text{IG}}}(g)\,dg \\
&= \mathbb{E}\left[e^{X_t^{\text{IG}}\left(iz\mu-\frac{1}{2}z^2\sigma^2\right)}\right] \\
&= \exp\left[-\delta t\left(\sqrt{\alpha^2-(\beta+iz)^2}-\sqrt{\alpha^2-\beta^2}\right)\right] \text{ [3]}.
\end{aligned} \tag{9.10}$$

Here we applied the same arguments we used when calculating the VG characteristic function in (8.12), except that we have used the characteristic function of the inverse Gaussian process X_t^{IG} as shown in (9.5) with $\eta = 1$ and $\gamma = \delta\sqrt{\alpha^2-\beta^2}$ for the last equality. From the linearity of the log of the characteristic function in the time variable, we see that this is an infinitely divisible process with stationary independent increments.

Again, using Theorem 6.7.4 we observe that $\sigma^{\text{NIG}} = 0$ since $\gamma_0^{\text{IG}} = 0$ and therefore has no diffusion component. The Lévy measure of the NIG process is given by Barndorff-Nielsen [1997, (3.15)]

$$\nu^{\text{NIG}}(dx) = \frac{\delta\alpha e^{\beta x}K_1(\alpha|x|)}{\pi|x|}dx,$$

where $K_\cdot(\cdot)$ indicates the modified Bessel function of the second kind. As in the VG case, this can be found using Theorem 6.7.4 as follows

$$\begin{aligned}
\nu^{\text{NIG}}(x) &= \int_0^\infty f_{X_g}(x)\nu^{\text{IG}}(dg) \\
&= \int_0^\infty \frac{1}{\sqrt{2\pi\delta^2 g}}\exp\left[-\frac{\left(x-\beta\delta^2 g\right)^2}{2\delta^2 g}\right]\frac{1}{\sqrt{2\pi g^3}}\exp\left[-\frac{\delta^2\left(\alpha^2-\beta^2\right)g}{2}\right]dg \qquad (9.11) \\
&= \int_0^\infty \frac{1}{2\pi\delta g^2}\exp\left[-\frac{x^2}{2\delta^2 g}+\frac{\beta\delta^2 xg}{2\delta^2 g}-\frac{\beta^2\delta^4 g^2}{2\delta^2 g}-\frac{\delta^2\alpha^2 g}{2}+\frac{\delta^2\beta^2 g}{2}2\right]dg \\
&= \frac{e^{\beta x}}{\pi}\frac{1}{2\delta}\int_0^\infty \frac{1}{g^2}\exp\left[-\frac{(x/\delta)^2}{2g}-\frac{(\delta\alpha)^2 g}{2}\right]dg. \qquad (9.12)
\end{aligned}$$

Using the identity [see Cont and Tankov, 2004a, (A.2)]

$$2\left(\frac{c}{b}\right)^a K_a(bc) = \int_0^\infty \frac{1}{g^{1+a}}\exp\left[-\frac{b^2}{2g}-\frac{c^2 g}{2}\right]dg$$

[3] As in the gamma and VG cases, the form in which the $\Phi_{X_t^{\text{NIG}}}(z)$ is expressed in (9.10) may seem ambiguous since the square-root function is multi-valued. The single-valued form of (9.10) is given by

$$\Phi_{X_t^{\text{NIG}}}(z) = \exp\left[-\delta t\left(\exp\left[\tfrac{1}{2}\text{Ln}\left(\alpha^2-(\beta+iz)^2\right)\right]-\sqrt{\alpha^2-\beta^2}\right)\right].$$

This relies on the fact that the term $\alpha^2-(\beta+iz)^2 = \left(\alpha^2-\beta^2+z^2\right)-i2\beta z$ never crosses the negative real axis (recall that $-\alpha<\beta<\alpha$).

for $bc > 0$ and $a \in \mathbb{Z}$ with $a := 1$, $b := |x|/\delta$ and $c := \delta\alpha$, then

$$\begin{aligned}\nu^{\text{NIG}}(x) &= \frac{e^{\beta x}}{\pi}\frac{1}{2\delta}2\frac{\delta^2\alpha}{|x|}K_1\left(\delta\alpha\frac{|x|}{\delta}\right)\\ &= \frac{\delta\alpha e^{\beta x}K_1(\alpha|x|)}{\pi|x|}.\end{aligned}$$

As in the VG case, the Lévy measure has infinite mass and hence from Proposition 6.6.1 an NIG process has infinitely many jumps in any finite time interval.

Unlike the VG process, the NIG process is of infinite variation since

$$\int_{-1}^{1}|x|\nu^{\text{NIG}}(dx) = \infty$$

(see Sato [1999, Theorem 21.9] or Papapantoleon [2008] for example). To see this we again make use of the integral form of the Lévy measure given in (9.12) and Fubini's Theorem

$$\begin{aligned}\int_{-1}^{1}|x|\nu^{\text{NIG}}(dx) &= \int_{-1}^{1}|x|\int_0^\infty \frac{1}{g^2}\exp\left[-\frac{(x/\delta)^2}{2g} - \frac{(\delta\alpha)^2 g}{2}\right]dg\,dx\\ &= \frac{1}{2\pi\delta}\int_0^\infty\frac{1}{g^2}\exp\left[\frac{1}{2}\delta^2\left(\alpha^2-\beta^2\right)g\right]\int_{-1}^{1}|x|\exp\left[-\frac{1}{2}\frac{\left(x-\delta^2\beta g\right)^2}{\delta^2 g}\right]dx\,dg.\end{aligned}$$

The above will be equal to ∞ if the integrand is of order $\mathcal{O}(k)$ for some $k \leq -1$ as $g \to 0$, since for $\epsilon > 0$, $\int_0^\epsilon \frac{1}{g^k}\,dg = \infty$ if $k \leq -1$. Since $\exp\left[\frac{1}{2}\delta^2\left(\alpha^2-\beta^2\right)g\right] \to 1$ as $g \to 0$, we only require the integral with respect to x to have order $\mathcal{O}(k)$ for some $k \leq 1$. Consider the indeterminant integral

$$\int x\exp\left[-\frac{1}{2}\frac{\left(x-\delta^2\beta g\right)^2}{\delta^2 g}\right]dx$$

and make the substitution $w := \frac{x-\delta^2\beta g}{\delta\sqrt{g}}$, then the anti-derivative of the above is given by

$$G(w) := -\delta^2 g e^{-\frac{1}{2}w^2} + \delta^3 g^{3/2}\beta\sqrt{2\pi}F_{\text{N}}(w),$$

where $F_{\text{N}}(\cdot)$ denotes the cumulative normal distribution function. Therefore

$$\begin{aligned}&\int_{-1}^{1}|x|\exp\left[-\frac{1}{2}\frac{\left(x-\delta^2\beta g\right)^2}{\delta^2 g}\right]dx\\ &= -\int_{-1}^{0}x\exp\left[-\frac{1}{2}\frac{\left(x-\delta^2\beta g\right)^2}{\delta^2 g}\right]dx + \int_0^1 x\exp\left[-\frac{1}{2}\frac{\left(x-\delta^2\beta g\right)^2}{\delta^2 g}\right]dx\\ &= -G(x=0)+G(x=-1)+G(x=1)-G(x=0)\\ &= G(x=1)+G(x=-1)-2G(x=0)\\ &= G\left(\frac{1}{\delta\sqrt{g}}-\delta\beta\sqrt{g}\right)+G\left(-\frac{1}{\delta\sqrt{g}}-\delta\beta\sqrt{g}\right)-2G\left(-\delta\beta\sqrt{g}\right)\\ &= \delta^2 g\exp\left[-\frac{1}{2}\delta^2\beta^2 g\right]\left(2-\exp\left[-\frac{1}{2}\left(\frac{1}{\delta^2 g}-2\beta\right)\right]-\exp\left[-\frac{1}{2}\left(\frac{1}{\delta^2 g}+2\beta\right)\right]\right).\end{aligned}$$

Since $\exp\left[-\frac{1}{2}\delta^2\beta^2 g\right] \to 1$, $\exp\left[-\frac{1}{2}\left(\frac{1}{\delta^2 g} - 2\beta\right)\right] \to 0$ and $\exp\left[-\frac{1}{2}\left(\frac{1}{\delta^2 g} + 2\beta\right)\right] \to 0$ as $g \to 0$, the integral above is of order $\mathcal{O}(1)$ as $g \to 0$, and we are done.

The last component of the Lévy triplet is given by Barndorff-Nielsen [1997, 3.13]

$$\gamma^{\text{NIG}} = \frac{2\delta\alpha}{\pi}\int_0^1 \sinh(\beta x)K_1(\alpha x)dx$$

and is found using Theorem 6.7.4 as follows

$$\begin{aligned}
\gamma^{\text{NIG}} &= \int_0^\infty \nu^{\text{IG}}(dg)\int_{|x|\leq 1} x f_{X_g}(dx) \\
&= \int_0^\infty \nu^{\text{IG}}(dg)\int_{|x|\leq 1} x f_{X_g}(dx) \\
&= \int_{|x|\leq 1} x \int_0^\infty \frac{1}{\sqrt{2\pi\delta^2 g}}\exp\left[-\frac{\left(x-\beta\delta^2 g\right)^2}{2\delta^2 g}\right]\frac{1}{\sqrt{2\pi g^3}}\exp\left[-\frac{\delta^2\left(\alpha^2-\beta^2\right)g}{2}\right]\,dg\,dx \\
&= \int_{|x|\leq 1} x\nu^{\text{NIG}}(dx),
\end{aligned}$$

where the third equality follows from Fubini's Theorem and the last from (9.11). This can be simplified further

$$\begin{aligned}
\gamma^{\text{NIG}} &= \int_{|x|\leq 1} x\frac{\delta\alpha e^{\beta x}K_1(\alpha|x|)}{\pi|x|}\,dx \\
&= \frac{\delta\alpha}{\pi}\left[\int_{-1}^0 \frac{x}{-x}e^{\beta x}K_1\left(-\alpha x\right)\,dx + \int_0^1 e^{\beta x}K_1\left(\alpha x\right)\,dx\right] \\
&= \frac{\delta\alpha}{\pi}\left[\int_1^0 e^{-\beta y}K_1\left(\alpha y\right)\,dy + \int_0^1 e^{\beta x}K_1\left(\alpha x\right)\,dx\right] \\
&= \frac{\delta\alpha}{\pi}\left[-\int_0^1 e^{-\beta y}K_1\left(\alpha y\right)\,dy + \int_0^1 e^{\beta x}K_1\left(\alpha x\right)\,dx\right] \\
&= \frac{\delta\alpha}{\pi}\int_0^1 \left(e^{\beta x}-e^{-\beta x}\right)K_1\left(\alpha x\right)\,dx \\
&= \frac{2\delta\alpha}{\pi}\int_0^1 \sinh(\beta x)K_1(\alpha x)\,dx,
\end{aligned}$$

where we have made a change of variables in the third equality by setting $x := -y$ and made use of the identity $\sinh(\beta x) = \frac{1}{2}\left(e^{\beta x} - e^{-\beta x}\right)$ in the last equality. Hence the Lévy triplet is given by $\left(\left(\sigma^{\text{NIG}}\right)^2, \nu^{\text{NIG}}, \gamma^{\text{NIG}}\right)$ with

$$\sigma^{\text{NIG}} = 0, \quad \nu^{\text{NIG}}(x) = \frac{\delta\alpha e^{\beta x}K_1(\alpha|x|)}{\pi|x|} \quad \text{and} \quad \gamma^{\text{NIG}} = \frac{2\delta\alpha}{\pi}\int_0^1 \sinh(\beta x)K_1(\alpha x)\,dx.$$

9.3 Bridge Sampling

As in §7.5.2 and §8.3.1, we consider bridge sampling for the inverse Gaussian distribution. In this section, we require standard normal and Inverse Gaussian random variables. Given a uniform random number,

- a standard normal random variable is found by applying Moro's method (briefly discussed in §7.2) and

- an inverse Gaussian random variable is obtained by applying the method given in Appendix H to the cumulative inverse Gaussian function which we discuss in §G.3.

9.3.1 Inverse Gaussian Bridge Sampling

As we have done in §7.5.2 and §8.3.1 we derive the conditional distribution of inverse Gaussian increments. We follow the derivation given by Ribeiro and Webber [2003].

As before, consider three consecutive times $s < t < u$ and suppose that x, y and z are realisations of the inverse Gaussian increments $X_t^{\text{IG}} - X_s^{\text{IG}}$, $X_u^{\text{IG}} - X_t^{\text{IG}}$ and $X_u^{\text{IG}} - X_s^{\text{IG}}$ respectively (see Figure 9.3). Here X_s^{IG} and X_u^{IG} are points which have already been determined and t the position at which the next bridge point is to be created.

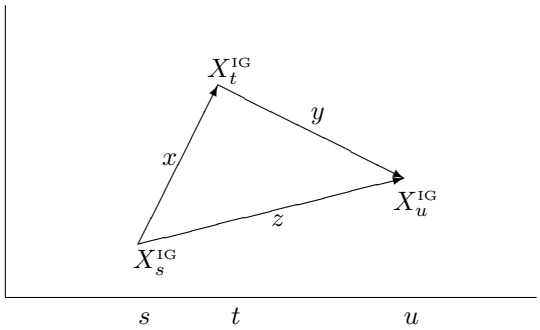

Figure 9.3: *Inverse Gaussian Bridge*

Recall from (9.2) the density of an inverse Gaussian process $X^{\text{IG}} = \{X_t^{\text{IG}}\}_{t\geq 0}$ with parameters $\eta, \gamma > 0$ is given by

$$f_{X_t^{\text{IG}}}(x) = \frac{\eta t}{\sqrt{2\pi x^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta t - \gamma x}{\sqrt{x}}\right)^2\right] \mathbb{1}_{\{x>0\}}.$$

In order to find the conditional distribution of the inverse Gaussian increment $X_t^{\text{IG}} - X_s^{\text{IG}} = x$ given that we know the increment $X_u^{\text{IG}} - X_s^{\text{IG}} = z$, we again apply Bayes' Theorem as in §7.5.2 and §8.3.1 and make use of the density of an inverse Gaussian process.

$$\begin{aligned}
&f_{(X_t^{\text{IG}} - X_s^{\text{IG}} | X_u^{\text{IG}} - X_s^{\text{IG}} = z)}(x) \\
&= \frac{f_{(X_u^{\text{IG}} - X_s^{\text{IG}} | X_t^{\text{IG}} - X_s^{\text{IG}} = x)}(z) f_{X_t^{\text{IG}} - X_s^{\text{IG}}}(x)}{f_{X_u^{\text{IG}} - X_s^{\text{IG}}}(z)} \\
&= \frac{f_{X_u^{\text{IG}} - X_t^{\text{IG}} + x}(z) f_{X_t^{\text{IG}} - X_s^{\text{IG}}}(x)}{f_{X_u^{\text{IG}} - X_s^{\text{IG}}}(z)} \\
&= \frac{\frac{\eta(u-t)}{\sqrt{2\pi(z-x)^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta(u-t)-\gamma(z-x)}{\sqrt{z-x}}\right)^2\right] \mathbb{1}_{\{z-x>0\}} \frac{\eta(t-s)}{\sqrt{2\pi x^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta(t-s)-\gamma x}{\sqrt{x}}\right)^2\right] \mathbb{1}_{\{x>0\}}}{\frac{\eta(u-s)}{\sqrt{2\pi z^3}} \exp\left[-\frac{1}{2}\left(\frac{\eta(u-s)-\gamma z}{\sqrt{z}}\right)^2\right] \mathbb{1}_{\{z>0\}}} \\
&= \frac{\eta}{\sqrt{2\pi}} \frac{(u-t)(t-s)}{u-s} \sqrt{\frac{z^3}{(z-x)^3 x^3}} \exp\left[-\frac{\eta^2}{2}\left(\frac{(u-t)^2}{z-x} + \frac{(t-s)^2}{x} - \frac{(u-s)^2}{z}\right)\right] \mathbb{1}_{\{\frac{x}{z}<1\}}. \qquad (9.13)
\end{aligned}$$

A method for sampling from the distribution above is given in Ribeiro and Webber [2003]. This method requires the use of a version of Tweedie's Theorem due to Seshadri [1993] and a result in Michael et al. [1976]. In Ribeiro and Webber [2003] it is remarked that if X and Z are inverse Gaussian increments as above, then

$$Q = \eta^2 \left(\frac{(u-t)^2}{Y} + \frac{(t-s)^2}{X} - \frac{(u-s)^2}{Z} \right) \sim \chi_1^2,$$

where χ_1^2 is the chi-squared distribution with one degree of freedom. The theorem given in Seshadri [1993] requires that $X \sim \text{IG}\left(\eta(t-s), \eta^{-1}\right)$, $Y \sim \text{IG}\left(\eta(u-t), \eta^{-1}\right)$ and $Z \sim \text{IG}\left(\eta(u-s), \eta^{-1}\right)$, however as noted in Ribeiro and Webber [2003], we are able to apply this theorem since the parameter γ does not appear in (9.13).

Consider the exponential in (9.13) and suppose we let $q(x) = \eta^2 \left(\frac{(u-t)^2}{z-x} + \frac{(t-s)^2}{x} - \frac{(u-s)^2}{z} \right)$. Then q can be rearranged so that

$$\begin{aligned} q(x) = \eta^2 \frac{(u-t)^2}{z} \frac{x}{z-x} \frac{(t-s)^2}{(u-t)^2} \times \\ \left(\frac{z-x+x}{x} \frac{(u-t)^2}{(t-s)^2} + \frac{(z-x+x)(z-x)}{x^2} - \frac{(u-t)^2 + 2(u-t)(t-s) + (t-s)^2}{(t-s)^2} \frac{z-x}{x} \right). \end{aligned}$$

If we let $a = \frac{z-x}{x}$, $\lambda = \frac{\eta^2 (u-t)^2}{z}$ and $\mu = \frac{u-t}{t-s}$, then substituting a, λ and μ in the above equation yields

$$\begin{aligned} q(x) &= \lambda \frac{1}{a\mu^2} \left(a\mu^2 + \mu^2 + (a+1)a - \mu^2 a - 2\mu a - a \right) \\ &= \lambda \frac{1}{a\mu^2} \left(a^2 - 2\mu a + \mu^2 \right) \\ &= \lambda \frac{(a-\mu)^2}{a\mu^2} =: g(a). \end{aligned} \tag{9.14}$$

We may rewrite (9.13) as follows

$$\begin{aligned} & f_{\left(X_t^{\text{IG}} - X_s^{\text{IG}} \mid X_u^{\text{IG}} - X_s^{\text{IG}} = z\right)}(x) \\ &= \frac{\eta}{\sqrt{2\pi}} \frac{(u-t)(t-s)}{u-s} \sqrt{\frac{z^3}{(z-x)^3 x^3}} \exp\left[-\frac{q(x)}{2} \right] \mathbb{1}_{\{\frac{x}{z} < 1\}} \\ &= \frac{1}{\sqrt{2\pi}} \frac{\eta(u-t)}{\sqrt{z}} \frac{1}{\frac{u-t}{t-s} + 1} \sqrt{\frac{x^3}{(z-x)^3}} \frac{z}{x^2} \left(\frac{z-x}{x} + 1 \right) \exp\left[-\frac{q(x)}{2} \right] \mathbb{1}_{\{\frac{z-x}{x} > 0\}}. \end{aligned}$$

If we let $\lambda = z$ and $g(X) =: A = \frac{z-X}{X}$ in Example F.1.3, then $f_A(a) = f_X(x)\frac{x^2}{z}$. Hence

$$f_A(a) = \frac{a+1}{\mu+1} \sqrt{\frac{\lambda}{2\pi a^3}} \exp\left[-\frac{\lambda}{2\mu^2 a}(a-\mu)^2 \right] \mathbb{1}_{\{a>0\}}, \tag{9.15}$$

where we have made the substitutions leading up to (9.14).

So, if we have a draw $a = \frac{z-x}{x}$ from the distribution given in (9.15), then we may write

$$\begin{aligned} a &= \frac{z-x}{x} = \frac{X_u^{\mathrm{IG}} - X_t^{\mathrm{IG}}}{X_t^{\mathrm{IG}} - X_s^{\mathrm{IG}}} \\ \Rightarrow X_t^{\mathrm{IG}} &= \frac{X_u^{\mathrm{IG}} + aX_s^{\mathrm{IG}}}{a+1} \\ &= \frac{X_u^{\mathrm{IG}} - X_s^{\mathrm{IG}} + X_s^{\mathrm{IG}} + aX_s^{\mathrm{IG}}}{a+1} \\ &= \frac{1}{a+1} X_u^{\mathrm{IG}} + \frac{a}{a+1} X_s^{\mathrm{IG}}. \end{aligned} \tag{9.16}$$

Again, as with gamma bridge sampling, we have a 'random interpolation'.

Now in order to obtain the draw a, we begin by solving for a in (9.14). There are exactly two solutions a_1 and a_2 for any q, namely

$$a_1 = \mu + \frac{\mu^2 q}{2\lambda} - \frac{\mu}{2\lambda}\sqrt{4\mu\lambda q + \mu^2 q^2} \tag{9.17}$$

$$a_2 = \frac{\mu^2}{a_1}. \tag{9.18}$$

As noted in Ribeiro and Webber [2003], a result given in Michael et al. [1976] can be used to sample from the density of A by sampling a chi-squared random number $q \sim \chi_1^2$ and then choosing root a_1 (calculated from q) with probability

$$\frac{1}{1 + \left|\frac{g'(a_1)}{g'(a_2)}\right| \frac{f_A(a_2)}{f_A(a_1)}},$$

where g is defined in (9.14) and f_A is the density of A given in (9.15). Using (9.18) we obtain

$$\frac{g'(a_1)}{g'(a_2)} = -\frac{\mu^2}{a_1^2} \text{ and } \frac{f_A(a_2)}{f_A(a_1)} = \frac{a_1^2}{\mu^2}\frac{\mu^2 + a_1}{1 + a_1}$$

and hence the root a_1 must be chosen with probability

$$\frac{\mu(1+a_1)}{(1+\mu)(\mu+a_1)}. \tag{9.19}$$

Therefore, to obtain a sample a from the bridge distribution given in (9.15) we proceed as follows:

- Draw a standard normal number $n_1 \sim \text{Normal}(0,1)$. Then $q := n_1^2$ is a variate from the χ_1^2-distribution.
- Next, calculate the two roots a_1 and a_2 using (9.17) and (9.18).
- Draw a uniform random number $u \sim \text{Uniform}(0,1)$. If $u < \frac{\mu(1+a_1)}{1+\mu(\mu+a_1)}$, then we set $a = a_1$, else we set $a = a_2$.

9.3.2 Time-Changed Brownian Motion

We begin the inverse Gaussian bridge sampling by simulating the inverse Gaussian process at terminal time. This is achieved by generating an inverse Gaussian variate

$$X_{t_M}^{\mathrm{IG}} \sim \mathrm{IG}(t_M, \gamma) \tag{9.20}$$

with the first number from our quasi random sequence and then generating a time-changed Wiener process

$$W_{X^{\text{IG}}_{t_M}} \sim \text{Normal}\left(0, X^{\text{IG}}_{t_M}\right)$$

with the second.

Next, similarly to Brownian or gamma bridge sampling, we use the algorithm of §7.5.1 to choose the first point t_j at which a bridge point must be created, with bridge points $W_{X^{\text{IG}}_0} = 0$ and $W_{X^{\text{IG}}_{t_M}}$ at times t_0 and t_M.

We next proceed as in the Brownian motion and gamma case, except that instead of using one random number or two (as in the gamma case) from the quasi random sequence, we use three random numbers at each step. For the general inductive step, let $s < t < u$ be consecutive time steps. Suppose $W_{X^{\text{IG}}_s}$ and $W_{X^{\text{IG}}_u}$ are known and t was determined to be the point at which the next bridge is to be created. Then, in order to find $W_{X^{\text{IG}}_t}$, we substitute the inverse Gaussian process at the appropriate times in (7.9) and obtain

$$W_{X^{\text{IG}}_t} = \frac{X^{\text{IG}}_t - X^{\text{IG}}_s}{X^{\text{IG}}_u - X^{\text{IG}}_s} W_{X^{\text{IG}}_u} + \frac{X^{\text{IG}}_u - X^{\text{IG}}_t}{X^{\text{IG}}_u - X^{\text{IG}}_s} W_{X^{\text{IG}}_s} + \sqrt{\frac{\left(X^{\text{IG}}_t - X^{\text{IG}}_s\right)\left(X^{\text{IG}}_u - X^{\text{IG}}_t\right)}{X^{\text{IG}}_u - X^{\text{IG}}_s}} n_2, \tag{9.21}$$

where n_2 is a standard normal random variable. Thus, using the next three quasi random numbers from our sequence, we first find a using the algorithm as described in §9.3.1 (and hence two random numbers) and n in order to calculate $W_{X^{\text{IG}}_t}$. This completes the inductive step.

In a similar way to finding arithmetic Brownian motion in (7.10) and a VG process in (8.18), we simulate an NIG process $X^{\text{NIG}} = \{X^{\text{NIG}}_t\}_{t \geq 0}$ where $X^{\text{NIG}}_t \sim \text{NIG}(\alpha, \beta, \delta t)$ by multiplying $W_{X^{\text{IG}}_t}$ in (9.21) with δ and adding $\beta\delta^2 X^{\text{IG}}_t$

$$\begin{aligned}
&X^{\text{NIG}}_t \\
&= \beta\delta^2 X^{\text{IG}}_t + \delta W_{X^{\text{IG}}_t} \\
&= \beta\delta^2 X^{\text{IG}}_t + \delta\left(\frac{X^{\text{IG}}_t - X^{\text{IG}}_s}{X^{\text{IG}}_u - X^{\text{IG}}_s} W_{X^{\text{IG}}_u} + \frac{X^{\text{IG}}_u - X^{\text{IG}}_t}{X^{\text{IG}}_u - X^{\text{IG}}_s} W_{X^{\text{IG}}_s} + \sqrt{\frac{\left(X^{\text{IG}}_t - X^{\text{IG}}_s\right)\left(X^{\text{IG}}_u - X^{\text{IG}}_t\right)}{X^{\text{IG}}_u - X^{\text{IG}}_s}} n_2\right) \\
&= \frac{X^{\text{IG}}_t - X^{\text{IG}}_s}{X^{\text{IG}}_u - X^{\text{IG}}_s}\beta\delta^2 X^{\text{IG}}_u + \frac{X^{\text{IG}}_u - X^{\text{IG}}_t}{X^{\text{IG}}_u - X^{\text{IG}}_s}\beta\delta^2 X^{\text{IG}}_s \\
&\quad + \delta\left(\frac{X^{\text{IG}}_t - X^{\text{IG}}_s}{X^{\text{IG}}_u - X^{\text{IG}}_s} W_{X^{\text{IG}}_u} + \frac{X^{\text{IG}}_u - X^{\text{IG}}_t}{X^{\text{IG}}_u - X^{\text{IG}}_s} W_{X^{\text{IG}}_s} + \sqrt{\frac{\left(X^{\text{IG}}_t - X^{\text{IG}}_s\right)\left(X^{\text{IG}}_u - X^{\text{IG}}_t\right)}{X^{\text{IG}}_u - X^{\text{IG}}_s}} n_2\right) \\
&= \frac{X^{\text{IG}}_t - X^{\text{IG}}_s}{X^{\text{IG}}_u - X^{\text{IG}}_s} X^{\text{NIG}}_u + \frac{X^{\text{IG}}_u - X^{\text{IG}}_t}{X^{\text{IG}}_u - X^{\text{IG}}_s} X^{\text{NIG}}_s + \delta\sqrt{\frac{\left(X^{\text{IG}}_t - X^{\text{IG}}_s\right)\left(X^{\text{IG}}_u - X^{\text{IG}}_t\right)}{X^{\text{IG}}_u - X^{\text{IG}}_s}} n_2.
\end{aligned}$$

As before, we may construct our algorithm for creating an NIG process in either of two ways: first generate inverse Gaussian bridges for all times and then find their corresponding NIG process, or for each inverse Gaussian bridge $W_{X^{\text{IG}}_t}$ calculate its corresponding NIG process X^{NIG}_t immediately.

Part III

American Monte Carlo Option Pricing under Exponential Lévy Models

Chapter 10

Risk-Neutral Modelling using the Variance Gamma and Normal Inverse Gaussian Models

For any Lévy process other than Brownian motion and Poisson processes — in particular the VG and NIG models as seen in §8.2 and §9.2 respectively — the Lévy market is incomplete. This follows from the martingale representation property for Lévy processes as proved by Nualart and Schoutens [2000], as the Brownian and Poisson cases are the only ones whose representation simplifies to the classical martingale representation [see Schoutens, 2003, Chapter 5, p.46]. From the Second Fundamental Theorem of Asset Pricing, it follows that the risk-neutral measure $\mathbb{Q}$ is not unique in all other cases.

There are many ways of finding an equivalent martingale measure $\mathbb{Q}$. One way is to use the Esscher transform. However, according to Schoutens [2003], it is not clear that the market chooses this kind of (exponential) transform, even though it is sometimes easy to find. More details can be found in Schoutens [2003, §6.2.2]. We will not discuss this method further. Another way of finding an equivalent martingale measure, is by changing the drift, in a manner very similar to the way the drift is changed when using *Girsanov's Theorem*. We will take this approach.

10.1 The Drift Term

In a similar way the unique $\mathbb{Q}$ is found with Brownian motion, we can construct a risk-neutral measure for VG or NIG processes. This is done by assuming that the risk-neutral dynamics of the asset have the same form as in the real-world, except that we adjust the drift so that the expected returns are the riskless rate. This means that statistical samples of returns will not be used to estimate the drift — the observed drift is irrelevant.

When we add a drift, say $m \in \mathbb{R}$ to the VG or the NIG process, the distribution of the processes is being translated by the value m. This translation does not change the fact that these distributions are infinitely divisible and is done similar to the way we transform a normal random variable $X \sim \text{Normal}\left(0, \sigma^2\right)$ to

$\widetilde{X} \sim \text{Normal}\left(m, \sigma^2\right)$.

Thus adding a drift $m \in \mathbb{R}$ to a process X_t results in the process

$$\widetilde{X}_t = X_t + mt$$

and its new characteristic function is found using (C.2)

$$\Phi_{\widetilde{X}_t}(z) = e^{izmt}\Phi_{X_t}(z). \tag{10.1}$$

Only the γ term in the Lévy triplet of $\widetilde{X}_t$ is affected by the translation and the triplet is calculated as

$$\widetilde{\gamma} = \gamma + m, \ \widetilde{\sigma}^2 = \sigma^2 \text{ and } \widetilde{\nu}(x) = \nu(x).$$

The density function of $\widetilde{X}_t$ is given by

$$f_{\widetilde{X}_t}(x) = f_{X_t}(x - mt). \tag{10.2}$$

This follows from

$$\begin{aligned} F_{\widetilde{X}_t}(x) &= \mathbb{P}\left(\widetilde{X}_t \leq x\right) \\ &= \mathbb{P}\left(X_t + mt \leq x\right) \\ &= \mathbb{P}\left(X_t \leq x - mt\right) \\ &= F_{X_t}(x - mt), \end{aligned}$$

where $F_{\widetilde{X}_t}(\cdot)$ and $F_{X_t}(\cdot)$ indicate the cumulative distribution functions of $\widetilde{X}_t$ and X_t respectively.

10.2 Risk-Neutral Modelling with Exponential Lévy Processes

Suppose we wish to construct a risk-neutral model using a Lévy process $X = \{X_t\}_{t\geq 0}$ such that

$$S_t = S_0 e^{Y_t} \text{ with } Y_t := (r - q)t + X_t \tag{10.3}$$

under a risk-neutral measure $\mathbb{Q}$, where r is the constant risk-free rate and q the constant continuous dividend yield. Such a model is called an *exponential Lévy model*. In order to guarantee that the discounted stock price is a martingale, one must have that [Cont and Tankov, 2004a, Proposition 3.18]

(i) $\mathbb{E}\left[e^{X_t}\right] < \infty$ which is equivalent to $\int_{|x|\geq 1} e^x \nu(dx) < 0$ [Cont and Tankov, 2004a, Proposition 3.14]; and

(ii)

$$\tfrac{1}{2}\sigma^2 + \gamma + \int_{-\infty}^{\infty} e^x - 1 - x\mathbb{1}_{\{|x|\leq 1\}}\, \nu(dx) = 0 \tag{10.4}$$

which can be found by applying Itô's formula to e^{X_t} and setting the resulting drift to 0. Alternatively, one can show that the process

$$\left\{\frac{e^{X_t}}{\mathbb{E}\left[e^{X_t}\right]}\right\}_{t\geq 0}$$

is a martingale [Cont and Tankov, 2004a, Proposition 3.17]. This implies that e^{X_t} is a martingale if and only if $\mathbb{E}\left[e^{X_t}\right] = 1$. However,

$$\mathbb{E}\left[e^{X_t}\right] = \exp\left[t\left(\tfrac{1}{2}\sigma^2 + \gamma + \int_{-\infty}^{\infty} e^x - 1 - x\mathbb{1}_{\{|x|\leq 1\}}\,\nu(dx)\right)\right]$$

[Eberlein, 2009, §3] and hence (10.4) results.

For more details on martingales relating to Lévy processes, see Cont and Tankov [2004a, §3.9, §8.4.1] or Eberlein [2009, §3].

In the next sections we derive risk-neutral stock price dynamics where X_t is a VG or NIG process. Here the equivalent martingale measure $\mathbb{Q}$ is determined by following the method discussed in Schoutens [2003, §6.2.2], where the derived $\mathbb{Q}$ is referred to as the *mean-correcting martingale measure* .

10.2.1 Variance Gamma

Let us assume that the stock price process $\{S_t\}_{t\geq 0}$ has risk-neutral dynamics

$$\begin{aligned} S_t &= S_0 e^{mt+\theta X_t^{\mathrm{G}}+\sigma W_{X_t^{\mathrm{G}}}} \\ \Rightarrow \ln\frac{S_t}{S_0} &= mt + \theta X_t^{\mathrm{G}} + \sigma W_{X_t^{\mathrm{G}}}. \end{aligned} \tag{10.5}$$

We want the expected value of S_t to be equal to $S_0 e^{(r-q)t}$ and therefore we set

$$\begin{aligned} S_0 e^{(r-q)t} &= \mathbb{E}\left[S_0 e^{mt+\theta X_t^{\mathrm{G}}+\sigma W_{X_t^{\mathrm{G}}}}\right] \\ \Rightarrow e^{(r-q)t} &= e^{mt}\mathbb{E}\left[\mathbb{E}\left[e^{\theta X_t^{\mathrm{G}}+\sigma W_{X_t^{\mathrm{G}}}}\,\middle|\, X_t^{\mathrm{G}} = g\right]\right] \\ &= e^{mt}\int \mathbb{E}\left[e^{\theta X_t^{\mathrm{G}}+\sigma W_{X_t^{\mathrm{G}}}}\,\middle|\, X_t^{\mathrm{G}} = g\right]\mathbb{P}\left(X_t^{\mathrm{G}} \in dg\right) \\ &= e^{mt}\int e^{g\left(\theta+\frac{1}{2}\sigma^2\right)} f_{X_t^{\mathrm{G}}}(g)\,dg \\ &= e^{mt}\mathbb{E}\left[e^{X_t^{\mathrm{G}}\left(\theta+\frac{1}{2}\sigma^2\right)}\right] \\ &= e^{mt}\exp\left[-\frac{t}{\nu}\ln\left(1-\left(\theta+\tfrac{1}{2}\sigma^2\right)\nu\right)\right]. \end{aligned}$$

This is similar to the calculation of the VG characteristic function in (8.12), except this time we use the moment generating functions of arithmetic Brownian motion given in (C.7) and the gamma process given in (8.6) with $\alpha = \frac{1}{\nu} = \beta$. Solving for m gives

$$m = r - q + \tfrac{1}{\nu}\ln\left[1-\left(\theta+\tfrac{1}{2}\sigma^2\right)\nu\right]. \tag{10.6}$$

Using (8.11) and (10.2) we find the density of $Y_t = \ln\frac{S_t}{S_0}$ under $\mathbb{Q}$

$$\begin{aligned} f_{Y_t}^{\mathbb{Q}}(x) &= f_{Y_t}(x-mt) \\ &= \frac{2\exp\left[(x-mt)\frac{\theta}{\sigma^2}\right]}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)\sqrt{2\pi\sigma^2}}\left[\frac{(x-mt)^2}{\frac{2\sigma^2}{\nu}+\theta^2}\right]^{\frac{t}{2\nu}-\frac{1}{4}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{\sqrt{(x-mt)^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}}{\sigma^2}\right). \end{aligned} \tag{10.7}$$

Using (8.12) and (10.1) we obtain the risk-neutral characteristic function of $\widetilde{X}_t^{\text{VG}} = mt + X_t^{\text{VG}}$

$$\begin{aligned}\Phi_{\widetilde{X}_t^{\text{VG}}}(z) &= e^{izmt}\Phi_{X_t^{\text{VG}}}(z) \\ &= e^{izmt}\left(1 - \nu\left(iz\theta - \tfrac{1}{2}\sigma^2 z^2\right)\right)^{-\frac{t}{\nu}} \end{aligned} \tag{10.8}$$

In Figures 10.1, 10.2 and 10.3 we consider the probability density functions (top left), implied volatility skews (top right) and a small sample of stock paths (bottom) for various sets of parameters. The graphs shown in green are the same in all three figures with parameters used in common. The relative strike considered (in the figure on the right) is with respect to the forward level of the spot. We considered 1 year options with $r = 8\%$ and $q = 2\%$ and initial stock price $S_0 = 100$.

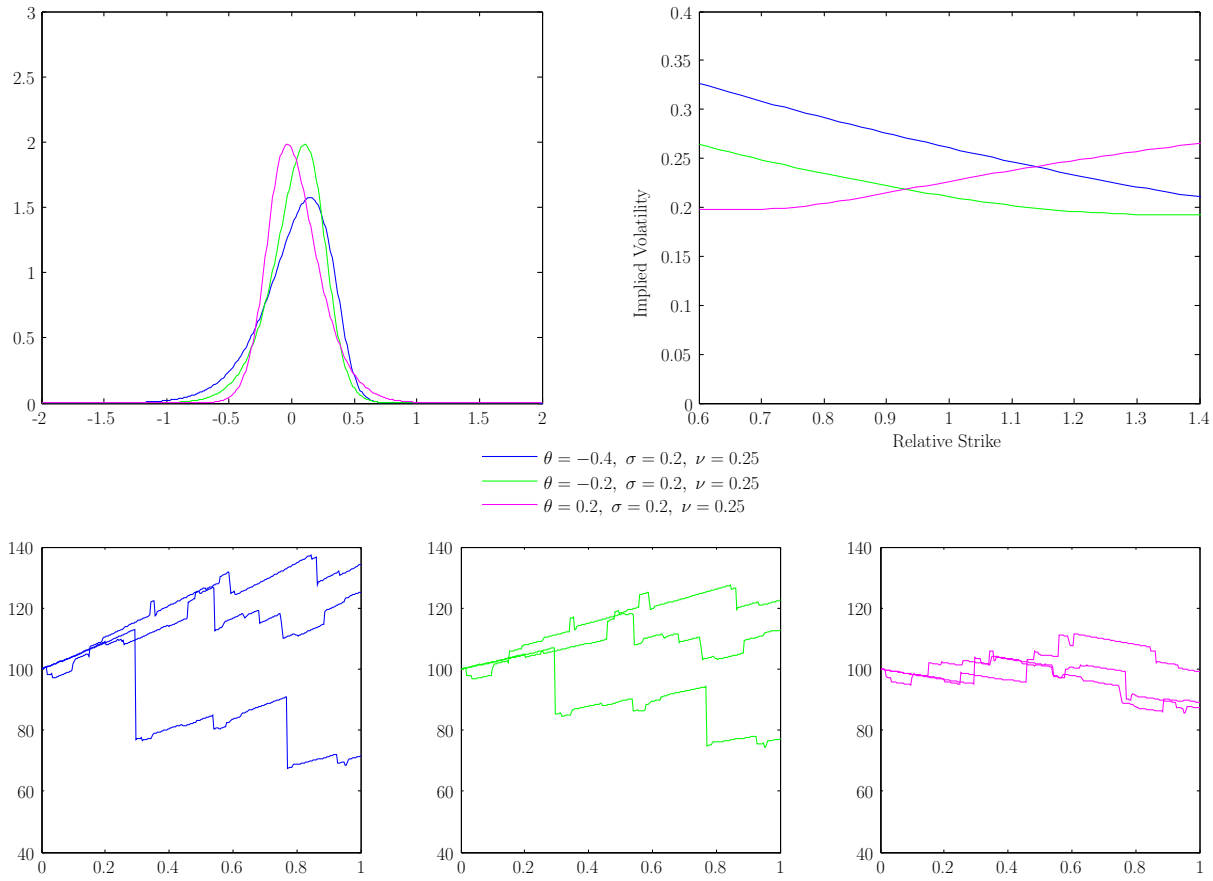

Figure 10.1: *VG probability density functions, implied volatility skews and a sample of three stock paths where we vary θ, while σ and ν are fixed. Observe that when θ is negative we have a downward skew, whereas a positive θ produces an upward sloping skew. When θ is negative, the jumps are mostly downward, but we have a positive drift. However, when θ is positive, jumps are mostly upward and we have a negative drift.*

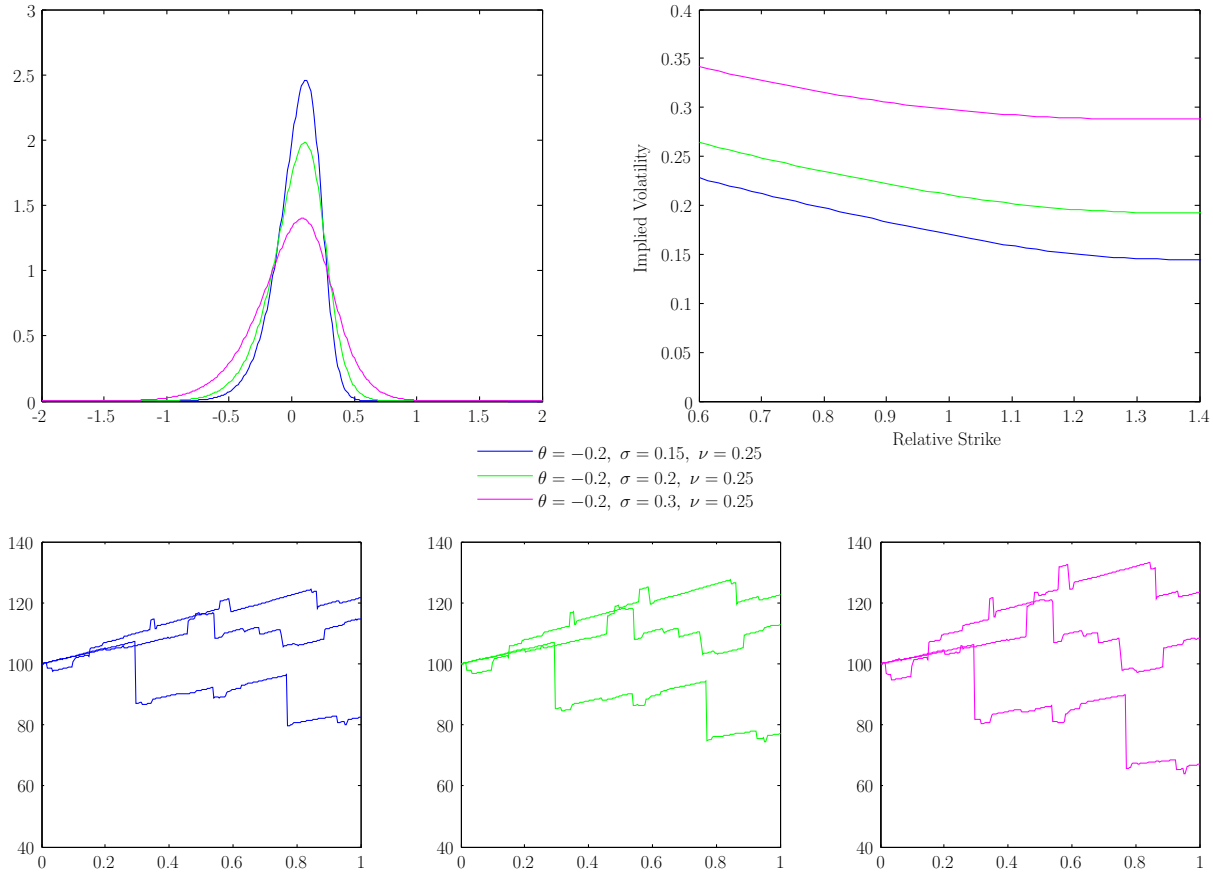

Figure 10.2: *VG probability density functions, implied volatility skews and a sample of three stock paths where the σ parameter is varied, while θ and ν are fixed. σ is more or less proportional to the general level of volatility.*

10.2.2 Normal Inverse Gaussian

Suppose that the stock price process $\{S_t\}_{t\geq 0}$ has risk-neutral dynamics

$$S_t = S_0 e^{mt+\beta\delta^2 X_t^{\mathrm{IG}}+\delta W_{X_t^{\mathrm{IG}}}}$$
$$\Rightarrow \ln\frac{S_t}{S_0} = mt + \beta\delta^2 X_t^{\mathrm{IG}} + \delta W_{X_t^{\mathrm{IG}}}. \tag{10.9}$$

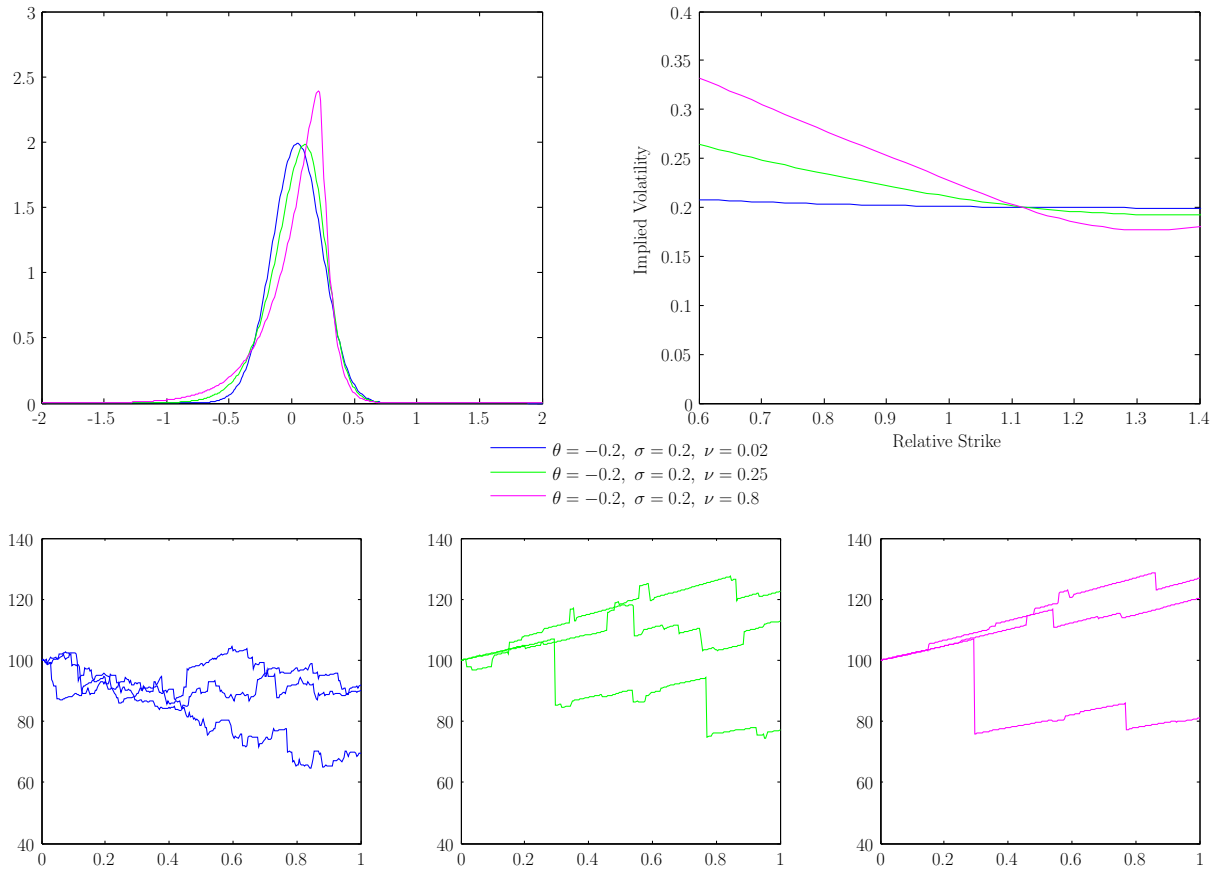

Figure 10.3: *VG probability density functions, implied volatility skews and a sample of three stock paths where ν is varied, while θ and σ are fixed. Note that as ν increases, a steeper skew is produced. As ν increases, the frequency of the large jumps decreases but their magnitude increases.*

Again, we require that the expected value of S_t to be equal $S_0 e^{(r-q)t}$

$$
\begin{aligned}
S_0 e^{(r-q)t} &= \mathbb{E}^{\mathbb{Q}}\left[S_0 e^{mt+\beta\delta^2 X_t^{\mathrm{IG}}+\delta W_{X_t^{\mathrm{IG}}}}\right] \\
\Rightarrow e^{(r-q)t} &= e^{mt}\mathbb{E}\left[\mathbb{E}\left[e^{\beta\delta^2 X_t^{\mathrm{IG}}+\delta W_{X_t^{\mathrm{IG}}}}\,\middle|\, X_t^{\mathrm{IG}}=g\right]\right] \\
\Rightarrow e^{(r-q)t} &= e^{mt}\int \mathbb{E}\left[e^{\beta\delta^2 X_t^{\mathrm{IG}}+\delta W_{X_t^{\mathrm{IG}}}}\,\middle|\, X_t^{\mathrm{IG}}=g\right]\mathbb{P}\left(X_t^{\mathrm{IG}}\in dg\right) \\
&= e^{mt}\int e^{g\left(\beta\delta^2+\frac{1}{2}\delta^2\right)} f_{X_t^{\mathrm{IG}}}(g)\,dg \\
&= e^{mt}\mathbb{E}\left[e^{X_t^{\mathrm{IG}}\left(\beta\delta^2+\frac{1}{2}\delta^2\right)}\right] \\
&= e^{mt}\exp\left[t\left(\delta\sqrt{\alpha^2-\beta^2}-\sqrt{\delta^2\left(\alpha^2-\beta^2\right)-2\left(\beta\delta^2+\tfrac{1}{2}\delta^2\right)}\right)\right].
\end{aligned}
$$

In a manner similar to the VG case, we have used the moment generating function of arithmetic Brownian

motion given in (C.7) and the moment generating function of the inverse Gaussian process in (9.3) with $\eta = 1$ and $\gamma = \delta\sqrt{\alpha^2 - \beta^2}$. Then solving for m gives

$$m = r - q + \delta\left(\sqrt{\alpha^2 - (\beta+1)^2} - \sqrt{\alpha^2 - \beta^2}\right). \tag{10.10}$$

Note that in the above we require $-\alpha < \beta + 1 < \alpha$. Thus, the final constraint is $-\alpha < \beta < \alpha - 1$.

Using (9.9) and (10.2) we find the density of $Y_t = \ln \frac{S_t}{S_0}$ under $\mathbb{Q}$

$$\begin{aligned} f^{\mathbb{Q}}_{Y_t}(\alpha, \beta, \delta; x) &= f_{Y_t}(x - mt) \\ &= \frac{\alpha\delta t}{\pi} \exp\left[\delta t\sqrt{\alpha^2 - \beta^2} + \beta(x - mt)\right] \frac{K_1\left(\alpha\sqrt{\delta^2 t^2 + (x - mt)^2}\right)}{\sqrt{\delta^2 t^2 + (x - mt)^2}}. \end{aligned} \tag{10.11}$$

The risk-neutral characteristic function of $\widetilde{X}^{\text{NIG}}_t = mt + X^{\text{NIG}}_t$ is found using (9.10) and (10.1)

$$\begin{aligned} \Phi_{\widetilde{X}^{\text{NIG}}_t}(z) &= e^{izmt}\Phi_{X^{\text{NIG}}_t}(z) \\ &= \exp\left[izmt - \delta t\left(\sqrt{\alpha^2 - (\beta + iz)^2} - \sqrt{\alpha^2 - \beta^2}\right)\right]. \end{aligned} \tag{10.12}$$

In Figures 10.4, 10.5 and 10.6 we consider the probability density functions (top left), implied volatility skews (top right) and a small sample of stock paths (bottom) for various sets of parameters. The graphs shown in green is the same in all three figures with parameters used in common. The relative strike considered (in the figure on the right) is with respect to the forward level of the spot. We considered 1 year options with $r = 8\%$ and $q = 2\%$ and initial stock price $S_0 = 100$. Unlike the case of the VG distribution, the meaning of the parameters in the NIG distribution is not always clear.

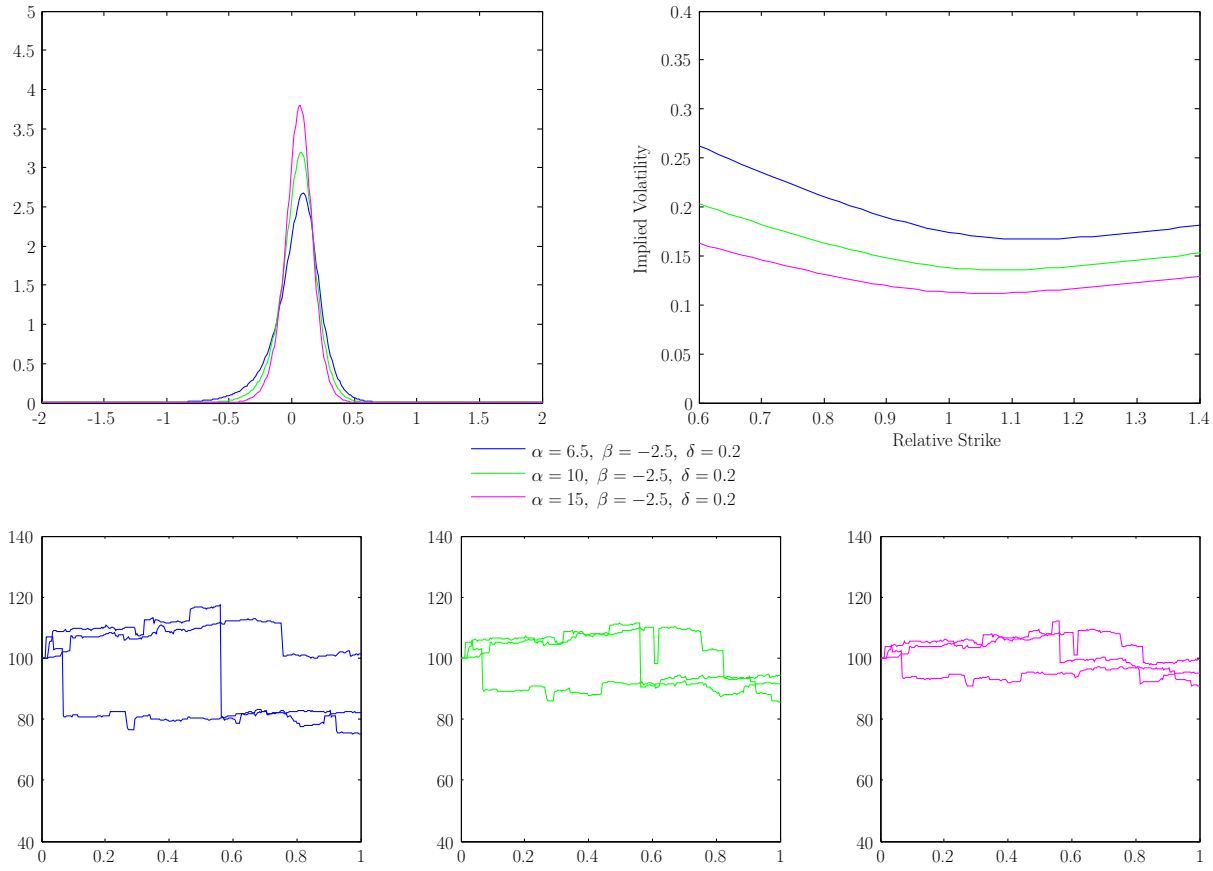

Figure 10.4: *NIG probability density functions, implied volatility skews and a sample of three stock paths where we vary α, while β and δ are fixed. Note that as the values of α increase, the size of the bigger jumps decreases.*

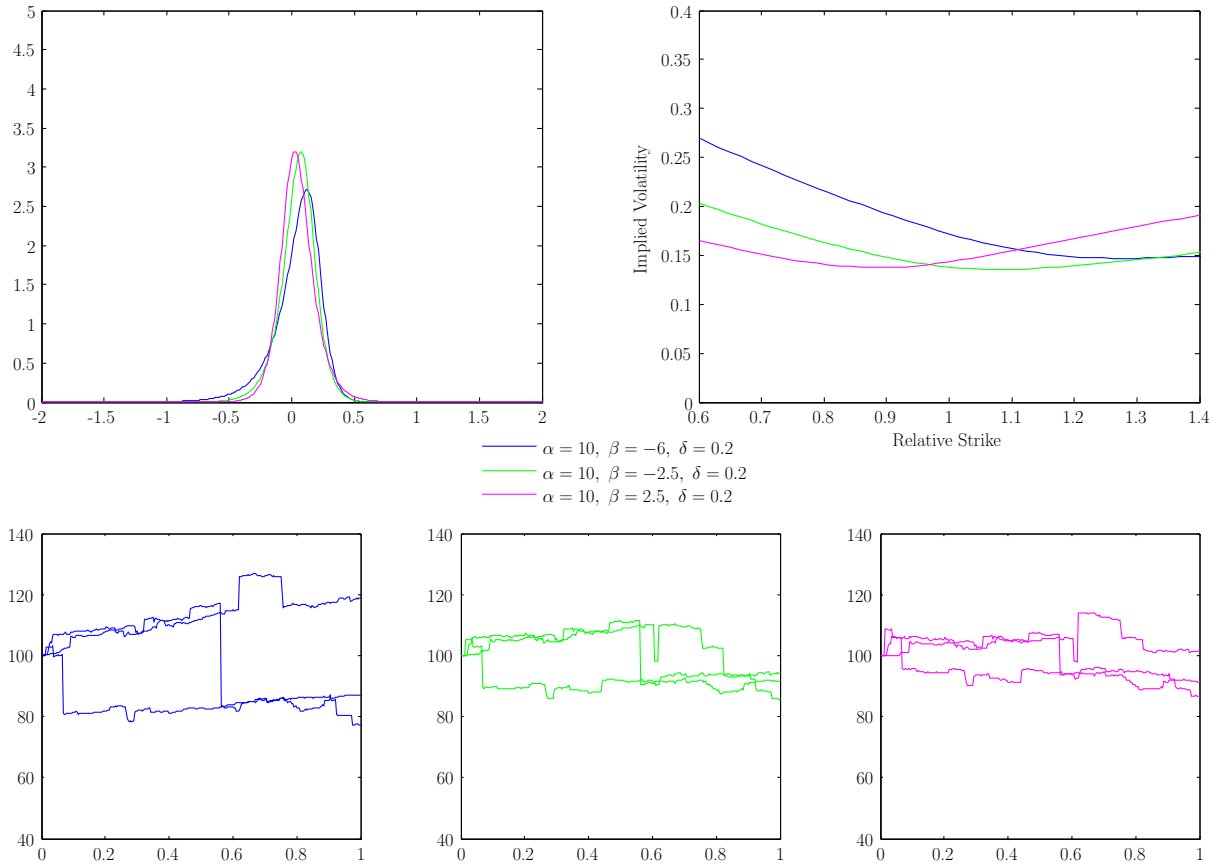

Figure 10.5: *NIG probability density functions, implied volatility skews and a sample of three stock paths where β is varied, while α and δ are fixed. Observe that when β is negative (positive), we have a downward (upward) sloping skew. As in the previous figure, as the values of β decrease, the size of the bigger jumps decreases.*

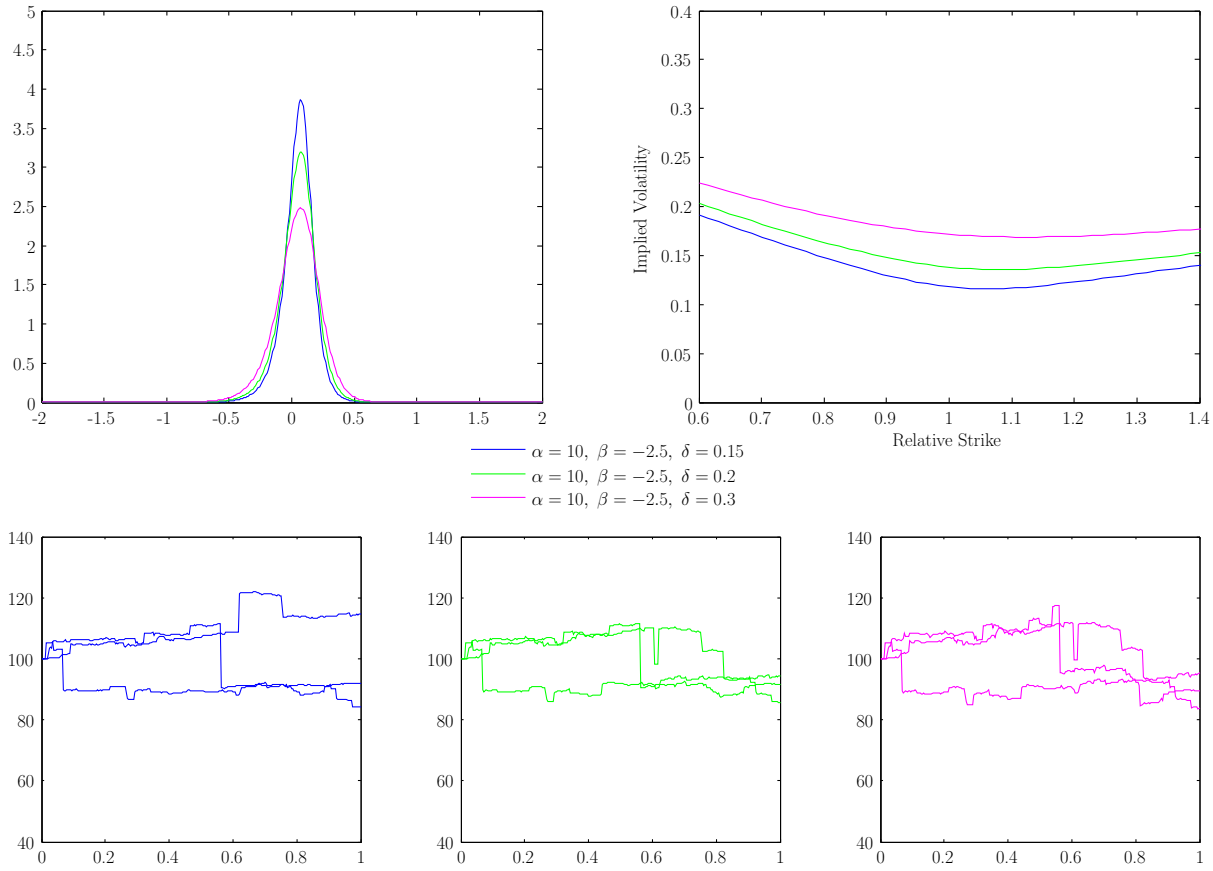

Figure 10.6: *NIG probability density functions, implied volatility skews and a sample of three stock paths where δ is varied, while α and β are fixed.*

Chapter 11

Calibration of the Variance Gamma and Normal Inverse Gaussian Models

The focus of the thesis is the pricing of American options under the VG and NIG models. The thesis would, however, be incomplete if we did not consider a calibration methodology that can be applied to these models.

We now consider the calibration of the VG and NIG models using a variation of the method suggested by Cont and Tankov [2004b]. They consider the calibration within the class of all Lévy models, but we will restrict the calibration to the VG and NIG models, which will yield a far more tractable calibration procedure. Here we will make use of a historical prior and market implied volatilities, and then find the model parameters by minimising a weighted sum of a measure of the distance between the model and the prior, and a measure of the distance between the model and the market. The order in which we proceed with the calibration process is as follows:

(i) Using historical data we perform a technique called *block bootstrap* (see the references in Lima and Tabak [2007]), which generates a sample of returns for the period of the option, where some of the serial properties of the daily data are preserved. We will consider this technique in §11.1. We then apply kernel smoothing to the new set of data obtained from the block bootstrap and extract its moments. This is discussed in §11.2.

(ii) We then generate the prior mentioned above. In §11.3 we use simultaneous equation solving to find the parameters of the Lévy model that has the same second, third and fourth moments as those found in the previous step. This model is made risk-neutral (see §10.2) and we denote the measure associated with it by $\mathbb{Q}_0$. A continuous dividend yield is required as input here.

(iii) We then turn our attention to the market implied model. Using the Black-Scholes option pricing formula, we price the options and calculate the Black-Scholes vega using the skew volatilities. A continuous dividend yield is required as input here as well.

(iv) Using the risk-neutral Lévy model with measure $\mathbb{Q}$, we then price the options on the skew and find some measure (the vega-weighted ℓ^2-distance for example) between these prices and the market

prices. That is, we define

$$\delta\left(\mathbb{Q}, \text{Market}\right) := \sum_i \mathcal{V}_i^{\text{Market}} \left| P_i^{\mathbb{Q}} - P_i^{\text{Market}} \right|^2 \tag{11.1}$$

where P_i^{Market} indicates the premium obtained from the market using the i^{th} point on the skew, $\mathcal{V}_i^{\text{Market}}$ indicates the corresponding Black-Scholes vega, and $P_i^{\mathbb{Q}}$ indicates the premium determined by the model with measure $\mathbb{Q}$ using the COS method as discussed in §E.1.

(v) We estimate the *Kullback-Leibler discrepancy* [Cont and Tankov, 2004b] (also referred to as the *relative entropy* of $f^{\mathbb{Q}}$ with respect to $f^{\mathbb{Q}_0}$) between $\mathbb{Q}$ and $\mathbb{Q}_0$ given by

$$\varepsilon\left(\mathbb{Q}, \mathbb{Q}_0\right) := \int_{-\infty}^{\infty} f^{\mathbb{Q}}(x) \ln \frac{f^{\mathbb{Q}}(x)}{f^{\mathbb{Q}_0}(x)}\, dx. \tag{11.2}$$

The technique for estimating this integral is discussed in §11.4.

The Kullback-Leibler discrepancy in (11.2) provides a type of measure of the distance between $\mathbb{Q}$ and $\mathbb{Q}_0$. Note that even though the discrepancy possesses some properties of a metric, it is not a metric. The motivation for including the Kullback-Leibler discrepancy in our calibration is that otherwise a solution to the calibration may be very difficult to find: solutions to the same calibration problem might produce parameter values that are dramatically different but skews generated from these parameter values are very similar.

(vi) The error is a weighted sum of $\delta\left(\mathbb{Q}, \text{Market}\right)$ and $\varepsilon\left(\mathbb{Q}, \mathbb{Q}_0\right)$ determined in (iv) and (v) respectively. The relative importance of the two factors is determined by a technique called the *Morozov discrepancy principle* Morozov [1966]. A detailed account of this last step is provided in §11.5.

11.1 Block Bootstrap

Sample statistics are often calculated using daily returns, which then need to be annualised. In this case, annualisation would be achieved by applying (C.24) and (C.25) to the sample mean, variance, skewness and kurtosis.

However, by considering daily returns, one finds that the resulting returns distribution is like that of the normal (with a skewness and excess kurtosis of 0). As an example we considered the South African Top40 index over 10 years and obtained an annual variance of 0.0577, skewness of -0.0076 and excess kurtosis of 0.0107: the distribution has similar properties to the normal distribution and hence there is no skew. The reason for this is that the way in which we have obtained the annual parameters creates normal behaviour in the data because of the law of large numbers: the serial features of the data are lost.

Thus in order to preserve the serial properties of the data, we employ the technique of block bootstrap. We divide the log returns into blocks of a specified size (we chose the number to be about 60). Then we sample blocks, with replacement, at random from our data until we have enough points with which we can perform our analysis, e.g. for a 1 year option we required 250 days. Note that this may involve only sampling a part of a terminal block, which is often referred to as a stub. We then sum the returns of the total points sampled. The entire procedure is repeated (with replacement) a large number of times and the annual variance, skewness and excess kurtosis are determined from this new set of data. We found by

performing this procedure on the Top40 index mentioned above, that the annual variance was given by 0.0569, the skewness by -1.0540 and the excess kurtosis by 3.5295 which is much more reasonable than the moments obtained directly from the daily data.

11.2 Kernel Smoothing

As noted in Schoutens [2003, §4.1.2], when approximating an empirical density, we make use of kernel density estimators. Kernel density estimation is a data smoothing technique where we make conclusions about the empirical density based on a finite set of data. Here we follow the kernel density estimation method as in Schoutens [2003, §4.1.2].

Suppose that the returns of the stock for the option period are calculated and denoted as $x_1, x_2, \ldots, x_N$. We imagine that the return x_i actually observed is a draw from a normal distribution with mean x_i, with the variance to be determined and denoted by h^2. Thus, the probability density function for the i^{th} draw is the Gaussian kernel,

$$k_i(x) = \frac{1}{\sqrt{2\pi}h} \exp\left[-\tfrac{1}{2}\left(\frac{x - x_i}{h}\right)^2\right]. \tag{11.3}$$

Let the distribution of interest be X. The kernel density estimator for the probability density function f_X of X at the point x is given by

$$\hat{f}_X(x) = \frac{1}{N}\sum_{i=1}^{N} k_i(x), \tag{11.4}$$

where N is the sample size, and $h = 1.06\sigma N^{-1/5}$ is the bandwidth (here σ is the sample standard deviation of the log price ratios).

Given this, we can calculate any number of raw moments of the distribution whose probability density function is $\hat{f}_X$ exactly. Let this distribution be $\hat{X}$. Then

$$\begin{aligned}
\mathbb{E}\left[\hat{X}^n\right] &= \frac{1}{\sqrt{2\pi}hN}\sum_{i=1}^{N}\int_{-\infty}^{\infty} x^n \exp\left[-\tfrac{1}{2}\frac{(x-x_i)^2}{h}\right] dx \\
&= \frac{1}{\sqrt{2\pi}N}\sum_{i=1}^{N}\int_{-\infty}^{\infty} (hw + x_i)^n e^{-\frac{1}{2}w^2}\, dw \\
&= \frac{1}{\sqrt{2\pi}N}\sum_{i=1}^{N}\sum_{j=0}^{n} h^j x_i^{n-j}\int_{-\infty}^{\infty} w^j e^{-\frac{1}{2}w^2}\, dw \\
&= \frac{1}{N}\sum_{i=1}^{N}\sum_{j=0}^{n} h^j x_i^{n-j} a_j,
\end{aligned}$$

where $a_0 = 1$, $a_1 = 0$, $a_j = (j-1)a_{j-2}$ for $j > 1$.

Here we use the key fact (from integration by parts) that, for $n \geq 2$,

$$\int w^n e^{-\frac{1}{2}w^2}\, dw = -w^{n-1}e^{-\frac{1}{2}w^2} + (n-1)\int w^{n-2}e^{-\frac{1}{2}w^2}\, dw.$$

We now transform from these raw moments to the mean, variance, skewness and kurtosis. This routine procedure is described in Appendix C.3.

11.3 Moment Matching

We now show how to calculate the parameters of the VG and NIG models from a given variance, skewness and excess kurtosis.

11.3.1 Variance Gamma

We believe the solution provided in this section is new.

As we have shown in Appendix C.3, we may calculate the mean, variance, skewness and excess kurtosis of a VG process X_t^{VG} using the moment generating function

$$\begin{aligned}
\mathbb{E}\left[X_t^{\text{VG}}\right] &= \theta t \\
\mathbb{V}\text{ar}\left[X_t^{\text{VG}}\right] &= \left(\theta^2\nu + \sigma^2\right) t \qquad (11.5) \\
s\left(X_t^{\text{VG}}\right) &= \frac{\theta\nu\left(2\theta^2\nu + 3\sigma^2\right)}{\left(\theta^2\nu + \sigma^2\right)^{\frac{3}{2}} t^{\frac{1}{2}}} \qquad (11.6) \\
\bar{\varkappa}\left(X_t^{\text{VG}}\right) &= 3\left(2\frac{\nu}{t} - \frac{\nu\sigma^4}{\left(\theta^2\nu + \sigma^2\right)^2 t}\right) \qquad (11.7)
\end{aligned}$$

where $\theta \in \mathbb{R}$, $\nu > 0$ and $\sigma > 0$. Note that

$$\bar{\varkappa}\left(X_t^{\text{VG}}\right) = \frac{3\nu}{t}\left[2 - \frac{\sigma^4}{\left(\theta^2\nu + \sigma^2\right)^2}\right] \geq \frac{3\nu}{t} > 0,$$

i.e. the excess kurtosis is positive. It is a stylised fact that market data displays excess kurtosis.

We derive the parameters θ, σ and ν in terms of the model variance, skewness and excess kurtosis using moment matching. For simplicity in the derivations below we set $t = 1$ in the above equations which means that the parameters we extract are annual. For ease of notation we denote $\mathbb{V}\text{ar}\left[X_t^{\text{VG}}\right]$ by Σ and suppress the symbol of the process X_t^{VG} when considering $s\left(X_t^{\text{VG}}\right)$ and $\bar{\varkappa}\left(X_t^{\text{VG}}\right)$.

Observe that we now ignore the mean of X_t^{VG} as we discussed in §10.1. In the above (11.5), (11.6) and (11.7) provide three equations in the unknowns θ, σ and ν. We solve for these unknowns as follows: firstly using (11.5) we may write

$$\theta^2\nu = \Sigma - \sigma^2$$

and then using (11.6) and the above gives

$$\theta\nu = \frac{s\left(\theta^2\nu + \sigma^2\right)^{\frac{3}{2}}}{\left(2\theta^2\nu + 3\sigma^2\right)} = \frac{s\Sigma^{\frac{3}{2}}}{\left(2\Sigma + \sigma^2\right)}. \qquad (11.8)$$

Then we have

$$\nu = \frac{(\theta\nu)^2}{\theta^2\nu} = \frac{s^2\Sigma^3}{\left(2\Sigma + \sigma^2\right)^2\left(\Sigma - \sigma^2\right)}. \qquad (11.9)$$

Now let

$$\begin{aligned} L &:= \frac{\bar{\varkappa}}{3} \\ &= 2\nu - \frac{\nu\sigma^4}{\left(\theta^2\nu + \sigma^2\right)^2} \\ &= \frac{s^2\Sigma^3}{\left(2\Sigma + \sigma^2\right)^2\left(\Sigma - \sigma^2\right)}\left(2 - \frac{\sigma^4}{\Sigma^2}\right) \\ &= \frac{s^2\Sigma\left(2\Sigma^2 - \sigma^4\right)}{\left(2\Sigma + \sigma^2\right)^2\left(\Sigma - \sigma^2\right)}. \end{aligned}$$

If we set $x := \sigma^2$, then we may write

$$L = \frac{s^2\Sigma\left(2\Sigma^2 - x^2\right)}{\left(2\Sigma + x\right)^2\left(\Sigma - x\right)}$$

and rewriting this we obtain a cubic in x

$$\begin{aligned} -Lx^3 + \left(-3L\Sigma + \Sigma s^2\right)x^2 + \left(4L\Sigma^3 - 2\Sigma^3 s^2\right) &= 0 \\ x^3 + \left(3\Sigma - \frac{\Sigma s^2}{L}\right)x^2 + \left(-4\Sigma^3 + 2\frac{\Sigma^3 s^2}{L}\right) &= 0. \end{aligned} \tag{11.10}$$

We write this in the form $p(x) = x^3 + ax^2 + bx + c = 0$ where the coefficients a, b and c of the cubic are given by

$$a = 3\Sigma - \frac{\Sigma s^2}{L},\ b = 0 \text{ and } c = -4\Sigma^3 + 2\frac{\Sigma^3 s^2}{L}.$$

We see that $\frac{s^2}{L}$ is a key quantity in this equation and consider three cases:

(i) Suppose $0 \leq \frac{s^2}{L} < 2$. In this case $a > 0$, $c < 0$, so the polynomial $p(x)$ has one coefficient sign change. Hence using Descarte's rule of signs, the polynomial has exactly one positive root.

(ii) Suppose $2 \leq \frac{s^2}{L} \leq 3$. In this case $a > 0$, $c > 0$, so there is no coefficient sign change and no positive root.

(iii) Suppose $3 < \frac{s^2}{L}$. In this case $a < 0$, $c > 0$, so there are two coefficient sign changes and hence either two positive roots or none.

Thus for definiteness we restrict attention to the case where $\frac{s^2}{L} < 2$. When considering historical data, we found it is possible that this ratio is larger than 2, depending on the block bootstrap parameters chosen. Under this condition it might be impossible to find a VG distribution that matches these moments, or there may be exactly two different distributions which match these moments.

We determine the roots x (if they exist) of the cubic (11.10) using the algorithm in Press et al. [2004,

§5.6]. Then we are able to calculate σ, ν and θ

$$\begin{aligned}
\sigma &= \sqrt{x} \\
\nu &= \frac{s^2\Sigma^3}{(2\Sigma+x)^2(\Sigma-x)} \\
\theta &= \frac{1}{\nu}\frac{s\Sigma^{\frac{3}{2}}}{(2\Sigma+x)} \\
&= \frac{(2\Sigma+x)^2(\Sigma-x)}{s^2\Sigma^3}\frac{s\Sigma^{\frac{3}{2}}}{(2\Sigma+x)} \\
&= \frac{(2\Sigma+x)(\Sigma-x)}{s\Sigma^{\frac{3}{2}}},
\end{aligned}$$

where we have made use of (11.8) and (11.9).

11.3.2 Normal Inverse Gaussian

Subsequent to performing this derivation, we found that similar results appear in Eriksson et al. [2004].

Recall from §10.2.2 that the parameters of the NIG process have the following constraints

$$\begin{aligned}
\alpha &> 0 \\
-\alpha < \beta &< \alpha - 1 \\
\delta &> 0.
\end{aligned}$$

The parameters α and β are scale invariant, whereas δ is not. Thus it is necessary to indicate whether the δ parameter has been annualised or not. For example, the δ values provided in Rydberg [1997, Table 2] are per period, and hence to annualise these numbers we need to multiply them by the number of observations per year which we have taken to be 250. In our presentation all parameters will always be the annualised version.

As shown in Appendix C.3 we may determine the mean, variance, skewness and excess kurtosis of an NIG process X_t^{NIG}

$$\begin{aligned}
\mathbb{E}\left[X_t^{\text{NIG}}\right] &= \frac{\beta\delta t}{\sqrt{\alpha^2-\beta^2}} \\
\mathbb{V}\text{ar}\left[X_t^{\text{NIG}}\right] &= \frac{\alpha^2\delta t}{(\alpha^2-\beta^2)^{\frac{3}{2}}} && (11.11) \\
s\left(X_t^{\text{NIG}}\right) &= \frac{3\beta}{\alpha(\alpha^2-\beta^2)^{\frac{1}{4}}(\delta t)^{\frac{1}{2}}} && (11.12) \\
\bar{\varkappa}\left(X_t^{\text{NIG}}\right) &= 3\frac{\alpha^2+4\beta^2}{\alpha^2\sqrt{\alpha^2-\beta^2}\delta t}.
\end{aligned}$$

Observe that the excess kurtosis $\bar{\varkappa}\left(X_t^{\text{NIG}}\right)$ of the NIG process is always positive. As in the VG case, this is empirically not a problem.

We would like to write these expressions so that the parameters α, β and δt are the subjects of the equations using moment matching. As in the VG case we denote $\mathbb{V}\text{ar}\left[X_t^{\text{NIG}}\right]$ with Σ, and when referring

to $s\left(X_t^{\text{NIG}}\right)$ and $\bar{\varkappa}\left(X_t^{\text{NIG}}\right)$ we suppress the symbol of the process X_t^{NIG}. Also we set t to be equal to 1. We derive these equations as follows:

Consider

$$\begin{aligned} s^2 &= \frac{9\beta^2}{\alpha^2\sqrt{\alpha^2-\beta^2}\delta} \\ \Rightarrow \Sigma s^2 &= \frac{9\beta^2}{\left(\alpha^2-\beta^2\right)^2} \\ \frac{\bar{\varkappa}}{3} &= \frac{\alpha^2+4\beta^2}{\alpha^2\sqrt{\alpha^2-\beta^2}\delta} \\ \Rightarrow \frac{\bar{\varkappa}}{s^2} &= \frac{\alpha+4\beta^2}{9\beta^2}. \end{aligned} \tag{11.13}$$

Furthermore, let

$$A := \frac{\bar{\varkappa}}{3s^2} - \frac{5}{9} = \frac{\alpha^2-\beta^2}{9\beta^2},$$

then using the above and (11.13) we obtain

$$\begin{aligned} A^2\Sigma s^2 &= \frac{\left(\alpha^2-\beta^2\right)^2}{\left(9\beta^2\right)^2}\frac{9\beta^2}{\left(\alpha^2-\beta^2\right)^2} \\ &= \frac{1}{9\beta^2} \\ \Rightarrow \beta^2 &= \frac{1}{9A^2\Sigma s^2}. \end{aligned}$$

From (11.12) we observe that the sign of β is the same as the sign of s. Thus

$$\beta = \frac{1}{3A\sqrt{\Sigma}s}.$$

Having obtained β we can then determine α using (11.13)

$$\begin{aligned} \left(\alpha^2-\beta^2\right)^2 &= \frac{9\beta^2}{\Sigma s^2} \\ \Rightarrow \alpha^2-\beta^2 &= \frac{3\beta}{\sqrt{\Sigma}s} \\ \Rightarrow \alpha &= \sqrt{\frac{3\beta}{\sqrt{\Sigma}s}+\beta^2}. \end{aligned}$$

Finally, using (11.11) we calculate δ as

$$\delta = \frac{\Sigma(\alpha^2-\beta^2)^{\frac{3}{2}}}{\alpha^2}.$$

11.4 The Kullback-Leibler Discrepancy Approximation

We determine the Kullback-Leibler discrepancy in (11.2) by applying the following approximations:

(i) truncate the domain of integration using the interval $[a, b]$ in (E.2) (which is applied in the COS Method in Appendix E.1). We only calculate one interval $[a, b]$ for all choices of $\mathbb{Q}$ and use $\mathbb{Q}_0$ for this calculation with an at-the-money option. Typical parameter sets produced intervals more or less equal to $[-6, 6]$.

(ii) we attempted to apply *Gauss-Kronrod quadrature* Kronrod [1965] to this interval.

However, only using Gauss-Kronrod quadrature fails because the function $f^{\mathbb{Q}}(x) \ln \frac{f^{\mathbb{Q}}(x)}{f^{\mathbb{Q}_0}(x)}$ oscillates rapidly in the area concentrated around 0 for typical parameter sets (see Figure 11.1). Since this function has turning points which occur in unknown positions, the Gauss-Kronrod quadrature produces spurious results. If we knew where these turning points occurred, we could apply the Gauss-Kronrod quadrature to intervals between the turning points.

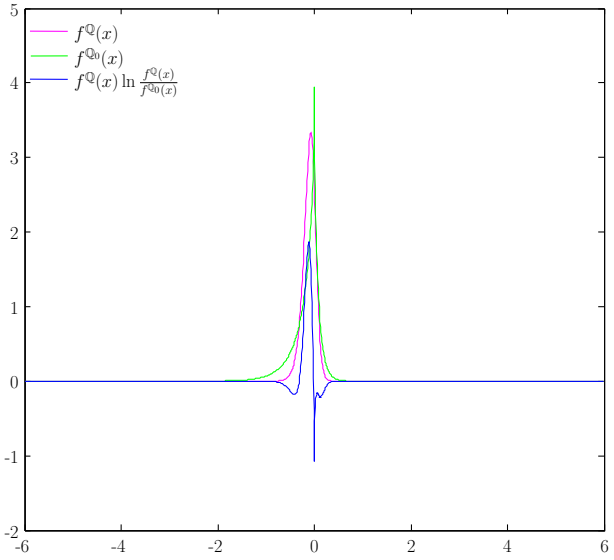

Figure 11.1: *A candidate density function $f^{\mathbb{Q}}(x)$ in our calibration procedure, along with the prior density $f^{\mathbb{Q}_0}(x)$ and the function $f^{\mathbb{Q}}(x) \ln \frac{f^{\mathbb{Q}}(x)}{f^{\mathbb{Q}_0}(x)}$. Observe the rapid oscillation of $f^{\mathbb{Q}}(x) \ln \frac{f^{\mathbb{Q}}(x)}{f^{\mathbb{Q}_0}(x)}$ in the area around 0.*

Therefore, we changed this approximation by estimating the integral using the *Trapezoid Rule* for the interval $[-1, 1]$ and the Gauss-Kronrod quadrature separately for the intervals $[-6, -1]$ and $[1, 6]$. Although this increases the duration of the calibration procedure, it is still acceptable.

11.5 The Morozov Discrepancy Principle

In this section we discuss how the Morozov discrepancy principle is applied to obtain the relative weights of the two distance functions given in (11.1) and (11.2); and use this to determine the parameters of the calibrating Lévy model.

Let $\mathcal{M}$ indicate all possible realisations of the Lévy model under consideration with $\mathbb{Q}^m$ denoting the measure of model $m \in \mathcal{M}$. We begin by defining the intrinsic model error

$$\varepsilon_0 := \inf_{m \in \mathcal{M}} \delta\left(\mathbb{Q}^m, \text{Market}\right),$$

where $\delta\left(\mathbb{Q}^m, \text{Market}\right)$ is the vega-weighted ℓ^2 distance between the Lévy model with measure $\mathbb{Q}$ and the market prices as shown in (11.1). Furthermore, define

$$g(\lambda) := \inf_{m \in \mathcal{M}} \lambda\varepsilon\left(\mathbb{Q}^m, \mathbb{Q}_0\right) + \delta\left(\mathbb{Q}^m, \text{Market}\right), \tag{11.14}$$

where $\varepsilon\left(\mathbb{Q}^m, \mathbb{Q}_0\right)$ is the Kullback-Leibler discrepancy between $\mathbb{Q}^m$ and $\mathbb{Q}_0$ as in (11.2). In the equation above, $\delta\left(\mathbb{Q}^m, \text{Market}\right)$ becomes a function of λ since it falls under the infimum that is being taken and hence we may write $h(\lambda) := \delta\left(\mathbb{Q}^m, \text{Market}\right)$. The Morozov discrepancy principle says that λ should be chosen in such a way that $h(\lambda) \approx 1.1\varepsilon_0$. If we let

$$f(\lambda) := h(\lambda) - 1.1\varepsilon_0 \tag{11.15}$$

then the Morozov discrepancy principle requires that $f(\lambda) = 0$. Note that even though $g(\lambda)$ is increasing, $h(\lambda)$ (and hence $f(\lambda)$) is not necessarily increasing (see Figure 11.2). However they will be more or less increasing and so a standard root-finding algorithm should suffice. In order to find the λ which solves this equation, we may use some root-finding algorithm such as *Brent's method* which we discuss in §H.2. Also, since $h(0) = \varepsilon_0$, we know what $f(0) = -0.1\varepsilon_0$. Thus we may use a modified version of the algorithm `BrentWithGuess` (discussed in §H.2) which takes the known left point of the interval containing the root and guesses the right point.

Observe that for each iteration of this root-finding algorithm, we minimise the error, i.e. determine $g(\lambda)$ by searching over the space of all $\mathbb{Q}^m$'s using methods such as Nelder and Mead [1965] or *Particle Swarm Optimisation* Kennedy and Eberhart [1995]. Once we have found λ we return to minimising (11.14), but with λ fixed (thus this expression is no longer a function of λ). The model parameters obtained from this last minimisation yield the final calibrated model we require.

Note that if, in the definition of $f(\lambda)$ (11.15), we increase the number 1.1, then the λ for which $f(\lambda) = 0$ is higher. We will refer to this number as the *perturbation factor*. We found that setting the perturbation factor to 1.1 produced λ's very close to 0. In such a case the market information given in the form of $\delta\left(\mathbb{Q}^m, \text{Market}\right)$ dominates the outcome of the calibration, since λ determines how important a role the Kullback-Leibler discrepancy plays in our calibration (as we see in (11.14)). Thus, when applying the Morozov discrepancy principle, this perturbation factor needs to be adjusted according to how close one believes one should be to the market. Consider Figure 11.2 where we have plotted $f(\lambda)$, $g(\lambda)$ and $h(\lambda)$ for various values of λ.

Consider the skews obtained from the market, from the prior distribution obtained from historical data, and the calibrated skew for the the NIG model in Figure 11.3. Here we plot calibrated skews for various values of λ. Note how steep the market skew is compared to the skew obtained from the prior distribution. The market data implied an annual variance of 0.0954, skewness of -2.5161 and excess kurtosis of 12.2083. As we have noted in §11.1, after performing the block bootstrap and kernel smoothing of the historical data we obtained an annual variance of 0.0569, skewness of -1.0540 and excess kurtosis of 3.5295. The higher skewness obtained from the market compared to that of the historical data, implies that the market prices in more downside protection. Furthermore, the market prices in much fatter tails than is found in historical data.

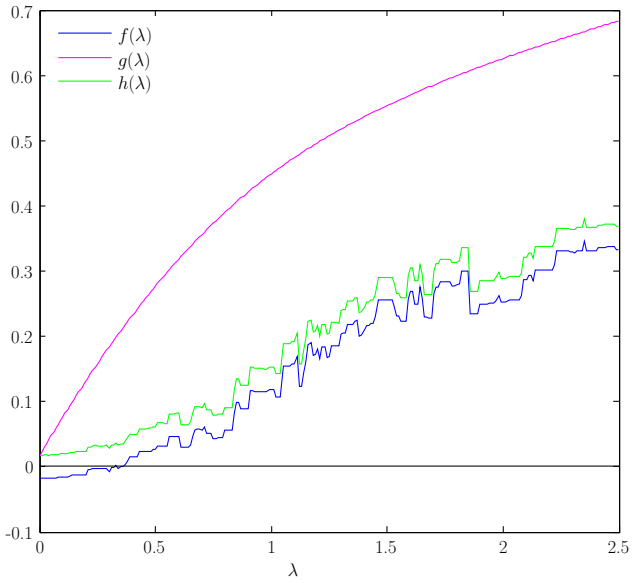

Figure 11.2: *$f(\lambda)$, $g(\lambda)$ and $h(\lambda)$ for various values of λ. Here the perturbation factor is set to 2.1.*

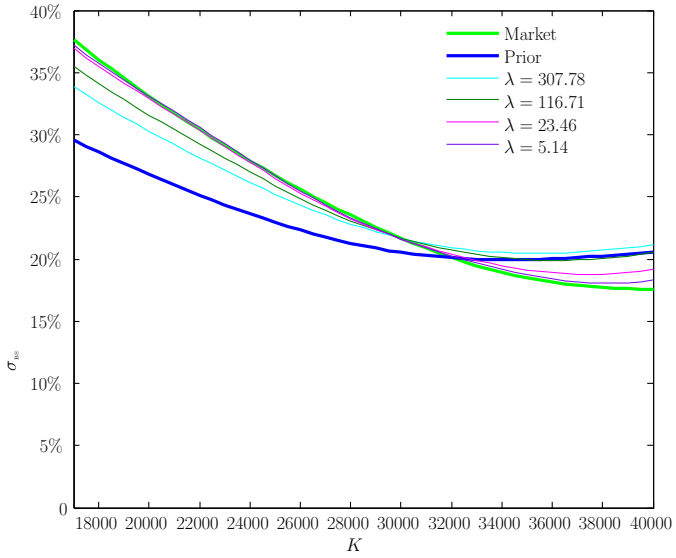

Figure 11.3: *We plot the market, prior and several calibrated skews for the South African Top40 index, where the λ's are varied for the calibrated skews. Here the market parameters are given by $\alpha = 6.5690,\ \beta = -4.9166,\ \delta = 0.1828$ and the prior parameters by $\alpha = 6.1928,\ \beta = -2.6331,\ \delta = 0.2613$.*

Chapter 12

Results and Conclusion

In this chapter we consider the results produced by the Monte Carlo methods presented in Chapters 2, 3, 4 and 5, where the underlying follows an exponential VG or NIG process. We also provide a summary of our findings in §12.2 and conclude the thesis.

12.1 Results

When considering the VG and NIG models, we compare the approximations obtained from our Monte Carlo methods with the value produced by the Bermudan COS method [see Fang and Oosterlee, 2009].

12.1.1 LSM and LSM-Rasmussen Methods

We begin by comparing the LSM method with the LSM-Rasmussen method as a function of the number of polynomials used as in Figure 3.2; and as a function of the number of sample paths as in Figure 3.3. In Figures 12.1 and 12.2 we give the results obtained using the VG model for two parameter sets. The results obtained from the NIG model appear in Figures 12.3 and 12.4.

We also compare values (for in-the-money, at-the-money and out-the-money options) and the time taken (in seconds) of the LSM and LSM-Rasmussen methods as we have done in Table 3.1. The VG results are given in Table 12.1 and the NIG results in Table 12.2. We used the same inputs as in Table 2.1.

As in the GBM case, the LSM-Rasmussen method performs remarkably better than the ordinary LSM method in both the VG and NIG cases. Note that when using the second set of parameters, the Monte Carlo methods perform much better than when the first set of parameters is used. We found that when parameter sets produced steep skews (which is the case of the first set of parameters), the LSM method and LSM-Rasmussen method did not perform as well as parameter sets which produced less steep skews (which is the case of the second set of parameters). However, the error is not severe.

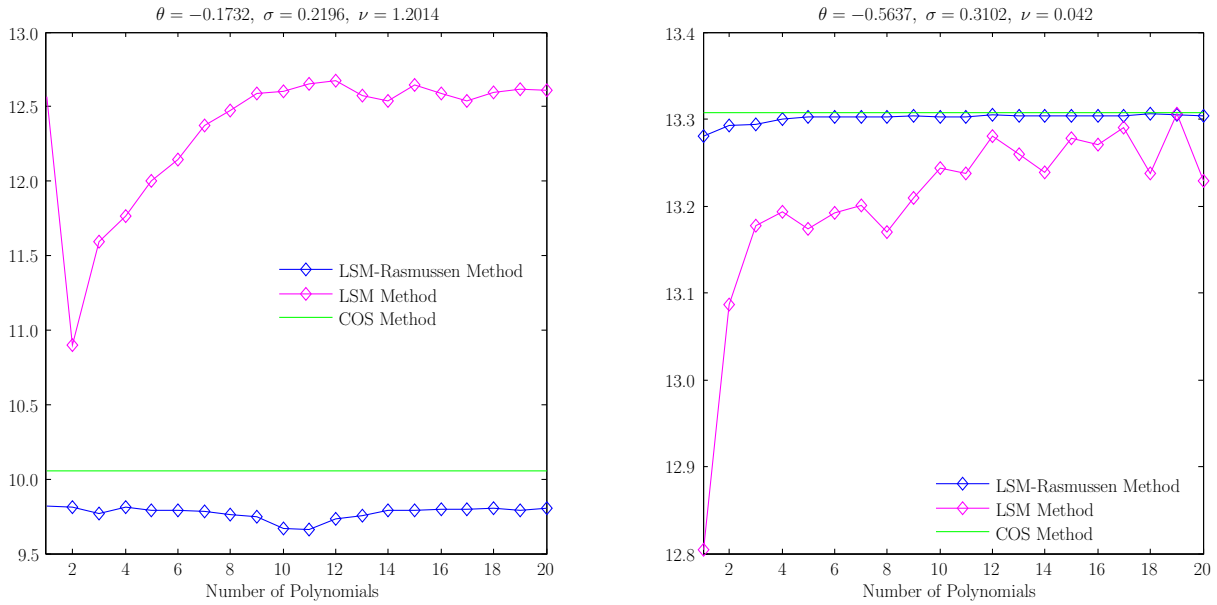

Figure 12.1: *The performance of the LSM and LSM-Rasmussen methods as a function of the number of Laguerre polynomials used.*

The option details are as follows: the underlying stock price process follows an exponential VG process where the parameters are indicated in the titles of the figures, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put with* $K = 135$. *We used 4096 sample paths with 20 time steps using Sobol' random numbers with bridging. We include the value given by the COS method with 20 time steps.*

Compared to the GBM case in Figure 3.2, the LSM method performs worse, in particular for the parameter set on the left (a parameter set that produces a steeper skew). As in the GBM case, the LSM-Rasmussen method clearly outperforms the LSM method. We obtained similar results for both parameter sets when varying strike, number of time steps and number of sample paths.

12.1.2 Stochastic Mesh Method

Next, we plot the performance of the high and low bias estimates obtained from the stochastic mesh method as we did in Figure 4.5 for the GBM case. We only consider the NIG case, since the computation time of these estimates in the VG case was infeasible. This is because the weights that need to be calculated in the stochastic mesh method given in (4.13) involves the calculation of the probability density function of the VG process given in (10.7). The calculation of the modified Bessel function of the second kind in the VG probability density function is very slow, because it is of fractional order and not of integer order as in the NIG case (see Appendix D for a brief discussion on modified Bessel functions of the second kind). A possible improvement in speed of this calculation would be to construct the weights without using densities as suggested by Broadie et al. [2000].

As in the GBM case in Table 4.1, we compare in Table 12.3 the performance of the stochastic mesh high and low bias prices to those generated by the LSM-Ramsussen method (for in-the-money, at-the-money and out-the-money options) and the time taken (in seconds) of these methods. We used the same inputs

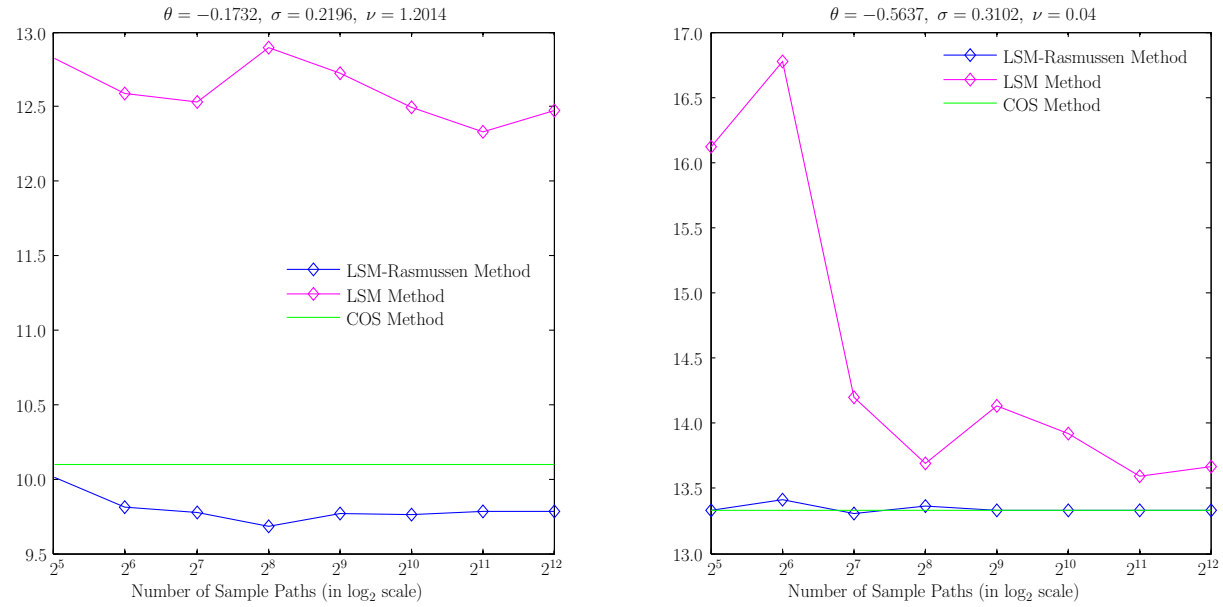

Figure 12.2: *The performance of the LSM and LSM-Rasmussen methods as a function of the number of sample paths used, with 8 Laguerre polynomials.*

The option details are as follows: the underlying stock price process follows an exponential VG process where the parameters are indicated in the titles of the two figures, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put with* $K = 135$. *We simulated paths with 30 time steps using Sobol' random numbers with bridging. We include the value given by the COS method with 30 time steps.*

As before, when comparing to the GBM case in Figure 3.3, the LSM method performs worse, in particular for the parameter set on the left (a parameter set that produces a steeper skew). As in the GBM case, the LSM-Rasmussen method clearly outperforms the LSM method.

as in Table 3.1. Again, we see that the LSM-Rasmussen method clearly outperforms the stochastic mesh method in computation time and accuracy.

12.1.3 Initial Dispersion

Recall from §3.2, that when considering the initial dispersion technique, t_{-1} was set to $\frac{t_M}{2}$. In §5.3 we saw that this choice of t_{-1} was adequate in the geometric Brownian motion case. However, as we see in Figures 12.6 and 12.7, this number needs to be adjusted. In the case of the VG model in Figure 12.6 we see that only when t_{-1} is set to about 2.0 years for the first parameter set and 1.4 years for the second parameter set, the modelled optimal stopping boundary becomes smooth. Note that we are considering a 1 year option. When considering the NIG model in Figure 12.7 we see that t_{-1} needs to be set to about 0.9 years for the first parameter set and about 0.3 for the second parameter set.

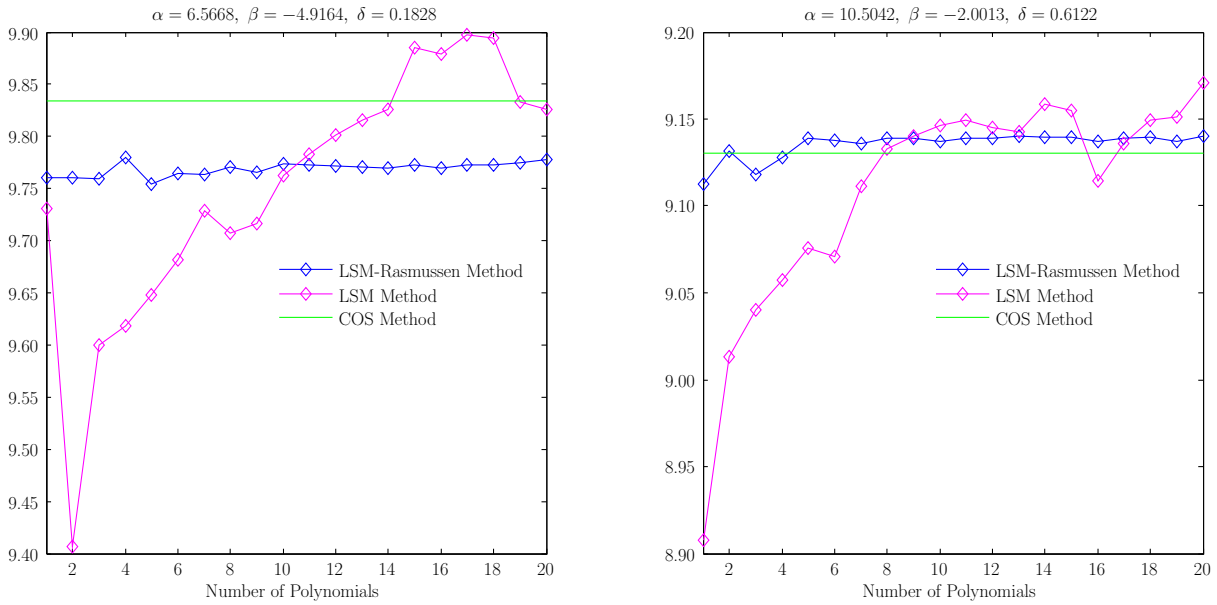

Figure 12.3: *The performance of the LSM and LSM-Rasmussen methods as a function of the number of Laguerre polynomials used.*

The option details are as follows: the underlying stock price process follows an exponential NIG process where the parameters are indicated in the titles of the figures, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put with* $K = 135$. *We used 4096 sample paths with 20 time steps using Sobol' random numbers with bridging. We include the value given by the COS method with 20 time steps.*

Compared to the GBM case in Figure 3.2, the LSM method performs worse, in particular for the parameter set on the left (a parameter set that produces a steeper skew). As in the GBM case, the LSM-Rasmussen method clearly outperforms the LSM method. We obtained similar results for both parameter sets when varying strike, number of time steps and number of sample paths.

12.1.4 Dual Method

Finally, we consider the performance of the dual method. We plot similar figures to those given in §5.3 for the VG and NIG models. Here we only consider the LSM-Rasmussen method as the results obtained from the stochastic mesh method were as poor for the VG and NIG models as those for the GBM case in §5.3. As before, we calculate the critical stock price function for each model using the LSM-Rasmussen method and then calculate the dual value from the critical stock price function (see the first row of Figures 12.8 and 12.9). We also plot the critical stock price functions as a function of the number of sample paths (see the second row of Figures 12.8 and 12.9). When considering VG parameter sets that produce steep skews, the critical stock price function produced poor results as we observe in Figure 12.8 for the first parameter set.

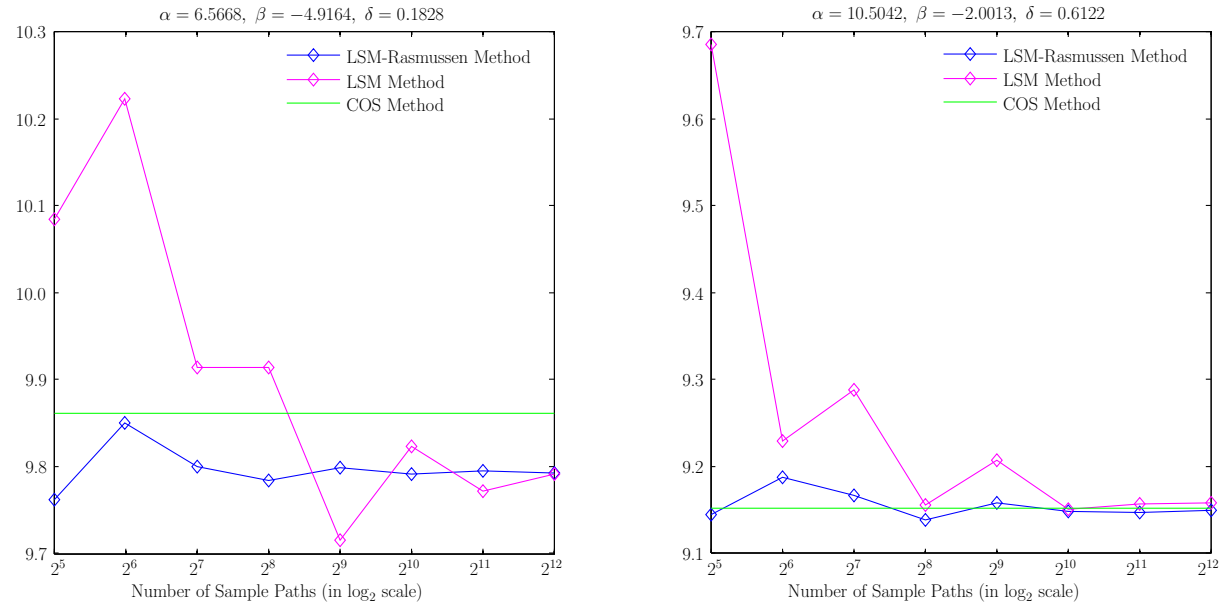

Figure 12.4: *The performance of the LSM and LSM-Rasmussen methods as a function of the number of sample paths used with 8 Laguerre polynomials.*

The option details are as follows: the underlying stock price process follows an exponential NIG process where the parameters are indicated in the titles of the two figures, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put with* $K = 135$. *We simulated paths with 30 time steps using Sobol' random numbers with bridging. We include the value given by the COS method with 30 time steps.*

As before, when comparing to the GBM case in Figure 3.3, the LSM method performs worse, in particular for the parameter set on the left (a parameter set that produces a steeper skew). As in the GBM case, the LSM-Rasmussen method clearly outperforms the LSM method.

12.2 Conclusion

We have considered various American Monte Carlo methods for the VG and NIG models, namely the LSM method in Chapter 2, the LSM-Rasmussen method in Chapter 3 (here variance reduction techniques are applied to the LSM method), and the stochastic mesh method in Chapter 4. We also considered the dual method in Chapter 5, which was applied to both the LSM and stochastic mesh methods.

From the results provided above, it is clear that the LSM-Rasmussen method not only outperforms the LSM method remarkably well under GBM, but also under the VG and NIG models. Furthermore, when applying the initial dispersion technique to approximate the optimal stopping boundary, satisfactory results were obtained for both the VG and NIG models by adjusting t_{-1} accordingly.

The stochastic mesh method did not perform as well as the LSM-Rasmussen method. In fact, the computation time of this method under the VG model was found infeasible and only results for the NIG model were reported. As noted before, a possible improvement to the tractability of this method is to construct the weights without considering densities [see Broadie et al., 2000]. A further improvement can be achieved by incorporating variance reduction techniques [see Broadie and Glasserman, 1997a].

$\theta = -0.1732,\ \sigma = 0.2196,\ \nu = 1.2014$

	Value		
K	COS	LSM	LSM-Rasmussen
85	1.2035	1.6825	1.2364
130	8.3693	10.8044	8.2288
135	10.1000	12.8954	9.6792
140	12.0543	15.4617	11.4235
185	50.0000	50.0000	50.0000
	Time		
K	COS	LSM	LSM-Rasmussen
85	0.000	12.415	12.725
130	0.000	13.730	13.979
135	0.000	10.090	10.239
140	0.000	9.745	9.719
185	0.000	10.621	10.687

$\theta = -5637,\ \sigma = 0.3102,\ \nu = 0.042$

	Value		
K	COS	LSM	LSM-Rasmussen
85	0.8927	0.9051	0.8850
130	11.0823	12.0281	11.0549
135	13.3349	13.6872	13.3642
140	15.8426	16.7008	15.9186
185	50.0000	52.7276	50.0000
	Time		
K	COS	LSM	LSM-Rasmussen
85	0.000	10.732	10.774
130	0.000	10.415	10.432
135	0.000	10.754	10.929
140	0.000	10.882	10.794
185	0.000	10.550	10.847

Table 12.1: *The performance of the LSM and LSM-Rasmussen methods for 256 sample paths with 8 Laguerre polynomials and varied strikes.*

Here the option details are as follows: the underlying stock price process follows an exponential VG process where the parameters are indicated in the table headings with $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put. We simulated paths with 30 time steps using Sobol' random numbers with bridging. We include the value given by the COS method with 30 time steps.*

Again we see that the LSM-Rasmussen method performs much better than the ordinary LSM method. Note that both methods perform better for the second parameter set.

The results obtained from applying the dual method to the stochastic mesh method were poor. Hence we only reported the results obtained when applying the dual method to the LSM-Rasmussen method. Both the VG and NIG models produced satisfactory results.

Thus we conclude that when pricing American options using Monte Carlo simulation, the LSM-Rasmussen method outperforms the other methods considered here; and can be used under both the VG and NIG models, producing excellent results.

$\alpha = 6.5668,\ \beta = -4.9164,\ \delta = 0.1828$			
	Value		
K	COS	LSM	LSM-Rasmussen
85	1.4239	1.5860	1.4064
130	8.2535	8.3071	8.2398
135	9.8616	10.0844	9.7617
140	11.8139	12.3991	11.6192
185	50.0000	50.0000	50.0000
	Time (seconds)		
K	COS	LSM	LSM-Rasmussen
85	0.000	12.747	12.764
130	0.000	13.178	13.117
135	0.000	12.417	12.487
140	0.000	13.615	13.643
185	0.000	12.411	12.441

$\alpha = 10.5042,\ \beta = -2.0013,\ \delta = 0.6122$			
	Value		
K	COS	LSM	LSM-Rasmussen
85	0.2497	0.3038	0.2478
130	7.1248	7.2374	7.1155
135	9.1524	9.6856	9.1450
140	11.5451	12.0767	11.5207
185	50.0000	50.0000	50.0000
	Time (seconds)		
K	COS	LSM	LSM-Rasmussen
85	0.000	13.535	13.716
130	0.000	12.446	12.556
135	0.000	12.277	12.519
140	0.000	12.369	12.565
185	0.000	12.466	12.368

Table 12.2: *The performance of the LSM and LSM-Rasmussen methods for 256 sample paths with 8 Laguerre polynomials and varied strikes.*

Here the option details are as follows: the underlying stock price process follows an exponential NIG process where the parameters are indicated in the table headings with $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put. We simulated paths with 30 time steps using Sobol' random numbers with bridging. We include the value given by the COS method with 30 time steps.*

Again we see that the LSM-Rasmussen method performs much better than the ordinary LSM method. We also note that the LSM-Rasmussen is more accurate for in-the-money options and that both methods perform better for the second parameter set.

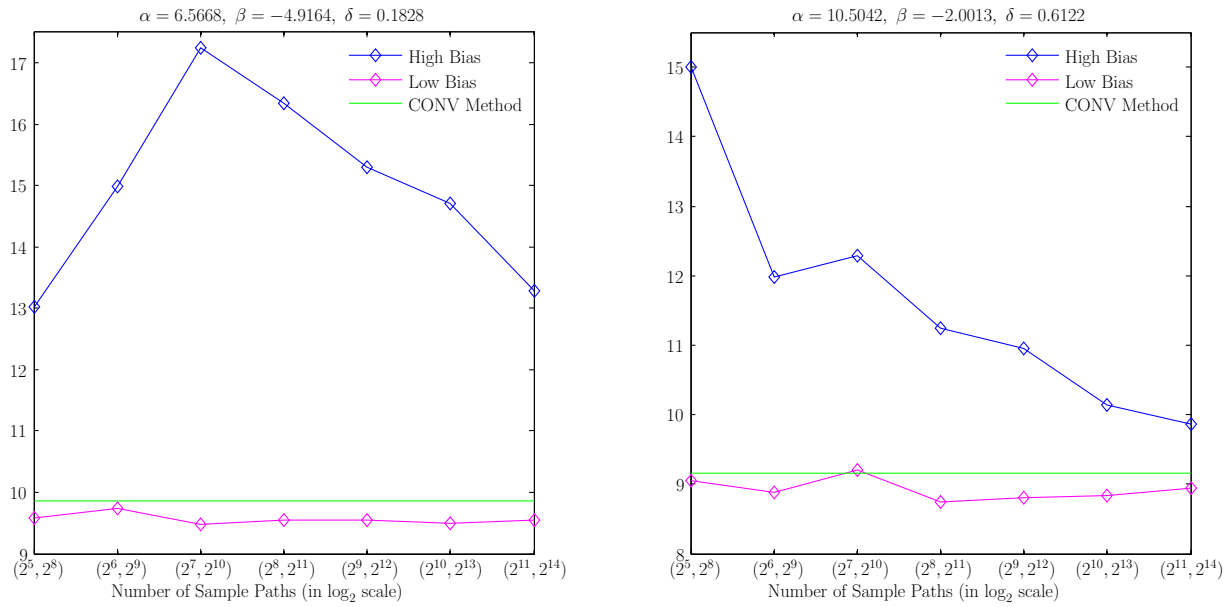

Figure 12.5: *Performance of the high and low bias prices generated using the stochastic mesh method.*

The option details are as follows: the underlying stock price process follows an exponential NIG process with the parameters indicated in the titles of the two figures, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$*. The sample paths were generated with 30 time steps using Sobol' random numbers with bridging. The number of sample paths used in the high bias estimate is given by the first entry in the bracket and that of the low bias estimate in the second on the horizontal axis. We also include the value given by the COS method with 30 time steps.*

Both parameter sets produced similar convergence results for in-the-money and out-the-money options.

$\alpha = 6.5668,\ \beta = -4.9164,\ \delta = 0.1828$

	Value			
K	COS	LSM-Rasmussen	High Bias	Low Bias
85	1.4239	1.4064	2.3920	1.4401
130	8.2535	8.2398	13.8563	8.1096
135	9.8616	9.7617	16.3485	9.5407
140	11.8139	11.6192	19.1730	11.4801
185	50.0000	50.0000	62.8713	42.8717
	Time (seconds)			
K	COS	LSM-Rasmussen	High Bias	Low Bias
85	0.000	12.764	639.289	0.158
130	0.000	13.117	651.895	0.143
135	0.000	12.487	645.484	0.140
140	0.000	13.643	640.927	0.141
185	0.000	12.441	649.041	0.141

$\alpha = 10.5042,\ \beta = -2.0013,\ \delta = 0.6122$

	Value			
K	COS	LSM-Rasmussen	High Bias	Low Bias
85	0.2497	0.2478	0.2859	0.2415
130	7.1248	7.1155	8.8234	6.8159
135	9.1524	9.1450	11.2420	8.7407
140	11.5451	11.5207	13.9842	10.9040
185	50.0000	50.0000	53.8459	43.9904
	Time (seconds)			
K	COS	LSM-Rasmussen	High Bias	Low Bias
85	0.000	13.716	649.275	0.157
130	0.000	12.556	636.825	0.157
135	0.000	12.519	633.814	0.160
140	0.000	12.565	651.025	0.156
185	0.000	12.368	646.432	0.157

Table 12.3: *The performance of the high and low bias prices generated using the stochastic mesh method compared to the LSM-Rasmussen method for varied strikes.*

Here the option details are as follows: the underlying stock price process follows an exponential NIG process with parameters as indicated in the titles of the table, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$. *The sample paths were generated with 30 time steps using Sobol' random numbers with bridging.. The high bias and LSM-Rasmussen values were generated using 256 sample paths, whereas the low bias values were generated using 2048 sample paths. We also include the value given by the COS method with 30 time steps.*

Clearly the LSM-Rasmussen method outperforms the stochastic mesh method in computation time and accuracy.

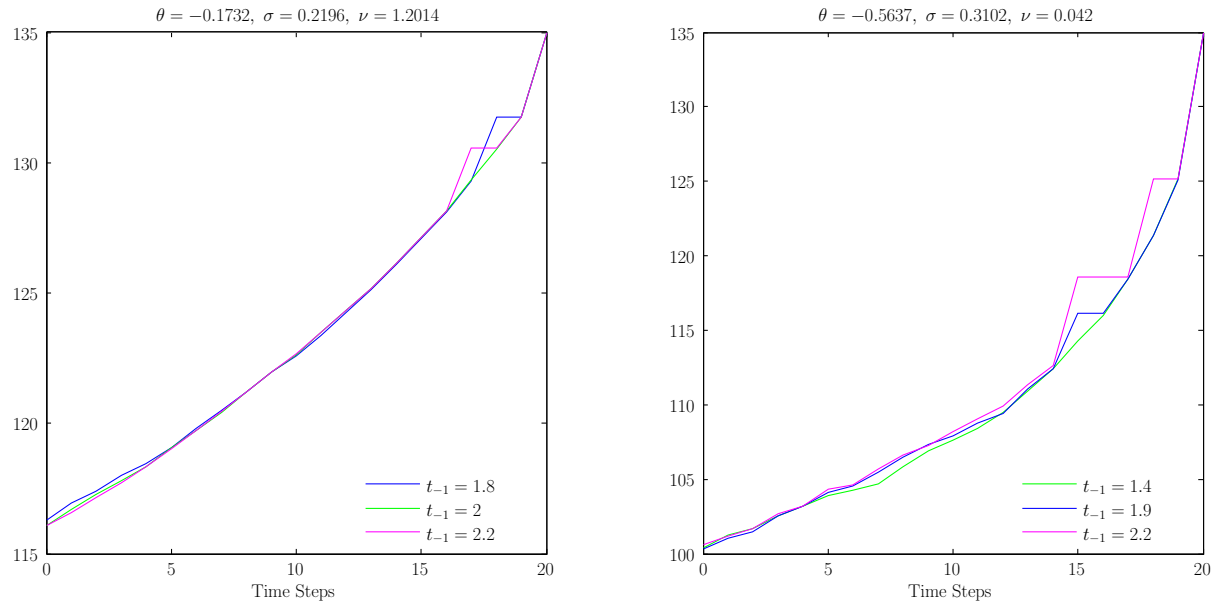

Figure 12.6: *The critical stock price function determined from the LSM-Rasmussen method with 8 Laguerre polynomials for various dispersion terms.*

The option details are as follows: stock price process follows exponential VG with the parameter sets indicated in the title of the graphs. Also, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$. *Here the number of sample paths used is 4096 with 20 time steps using Sobol' random numbers with bridging an.*

We see that to obtain a smooth model of the optimal stopping boundary, we require t_{-1} *to be about* 2.0 *years for the parameter set on the left and about* 1.4 *years for the parameter set on the right.*

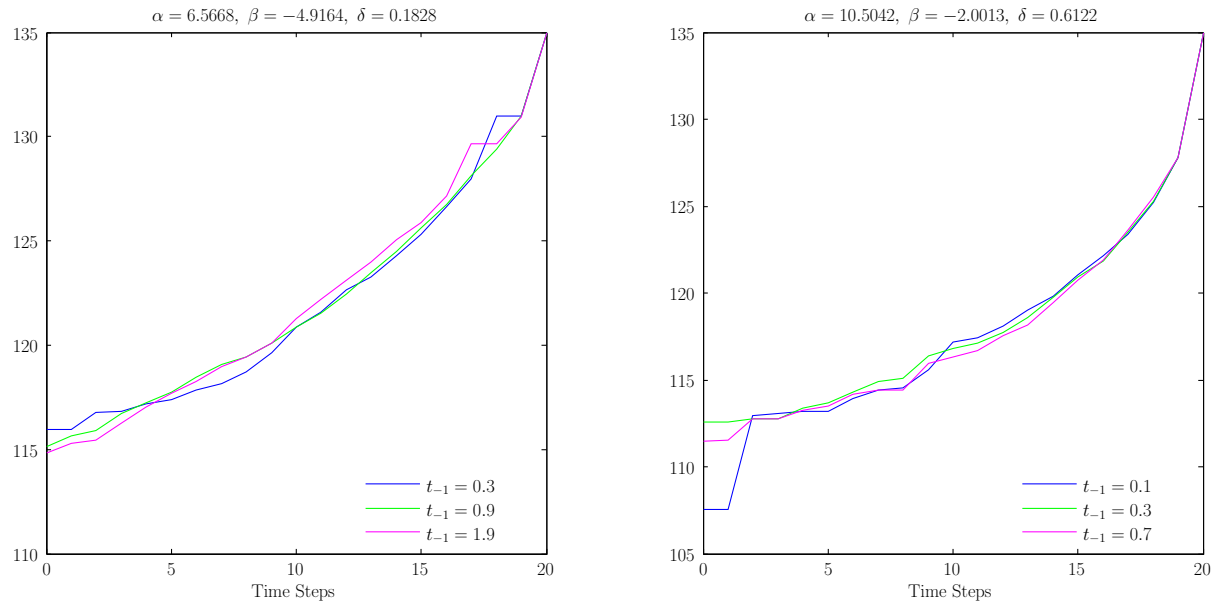

Figure 12.7: *The critical stock price function determined from the LSM-Rasmussen method with 8 Laguerre polynomials for various dispersion terms.*

The option details are as follows: stock price process follows exponential NIG with the parameter sets indicated in the title of the graphs. Also, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$. *Here the number of sample paths used is 4096 with 20 time steps using Sobol' random numbers with bridging an.*

We see that to obtain a smooth model of the optimal stopping boundary, we require t_{-1} *to be about* 0.9 *years for the parameter set on the left and about* 0.3 *years for the parameter set on the right.*

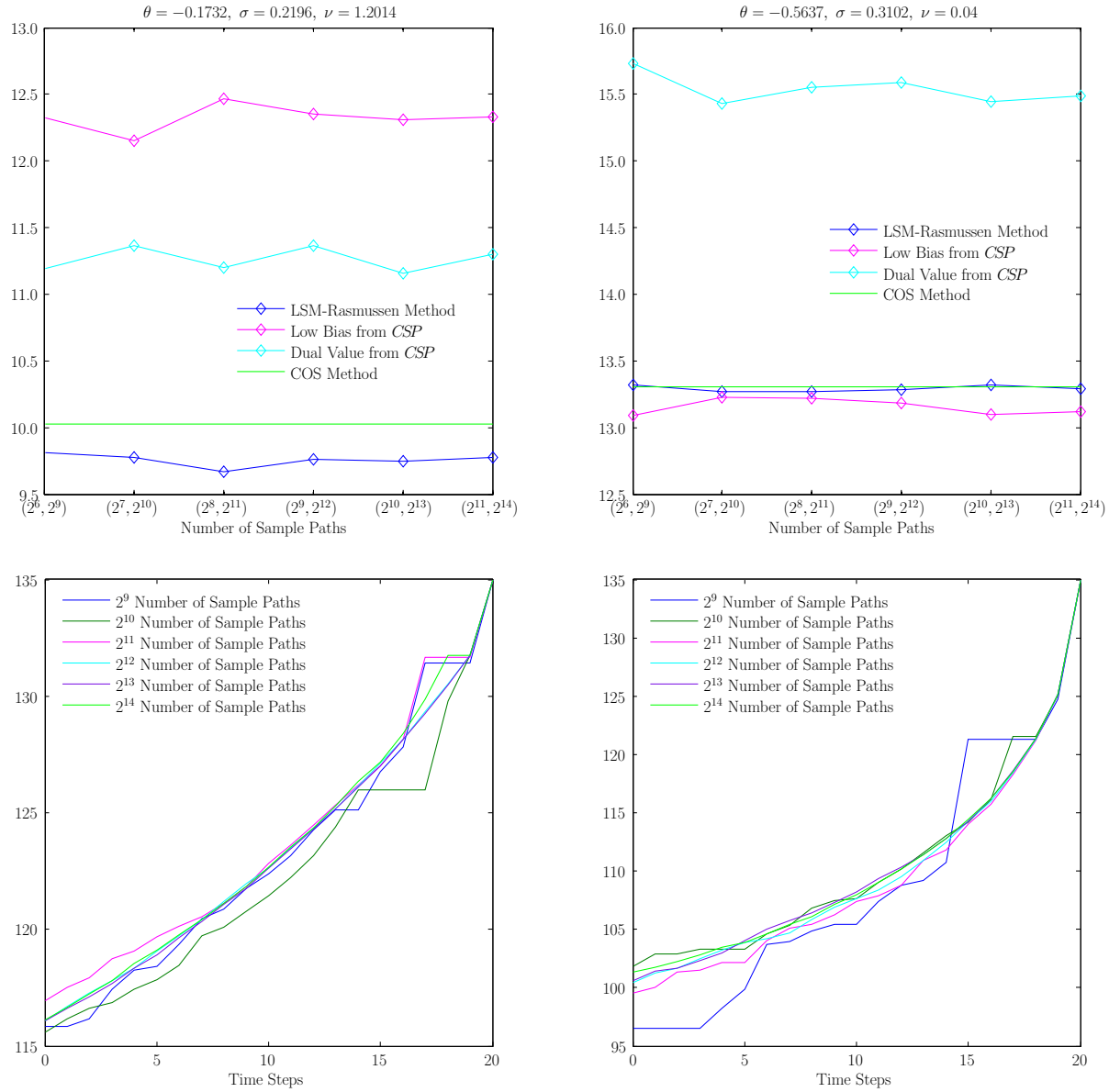

Figure 12.8: *In the first row, we plot the performance of the LSM-Rasmussen method as a function of the number of sample paths used with 8 Laguerre polynomials. We also plot the low bias obtained from the critical stock price function calculated in the LSM method. Furthermore, we plot the performance of the dual method which is calculated using this critical stock price function. The number of sample paths used in the high bias calculations and LSM method is given by the first entry in the bracket and that of the low bias estimate in the second on the horizontal axis. We also include the value given by the COS method with 20 time steps.*

In the second row, we plot the corresponding critical stock price function determined from the LSM-Rasmussen method for a varying number of sample paths.

The option details are as follows: the underlying stock price process follows exponential VG with parameters indicated in the titles of the figures, $S_0 = 135$, $r = 10\%$ and $q = 2\%$; and we are considering a 1 year put option with $K = 135$. Here, sample paths were generated with 20 time steps using Sobol' numbers with bridging.

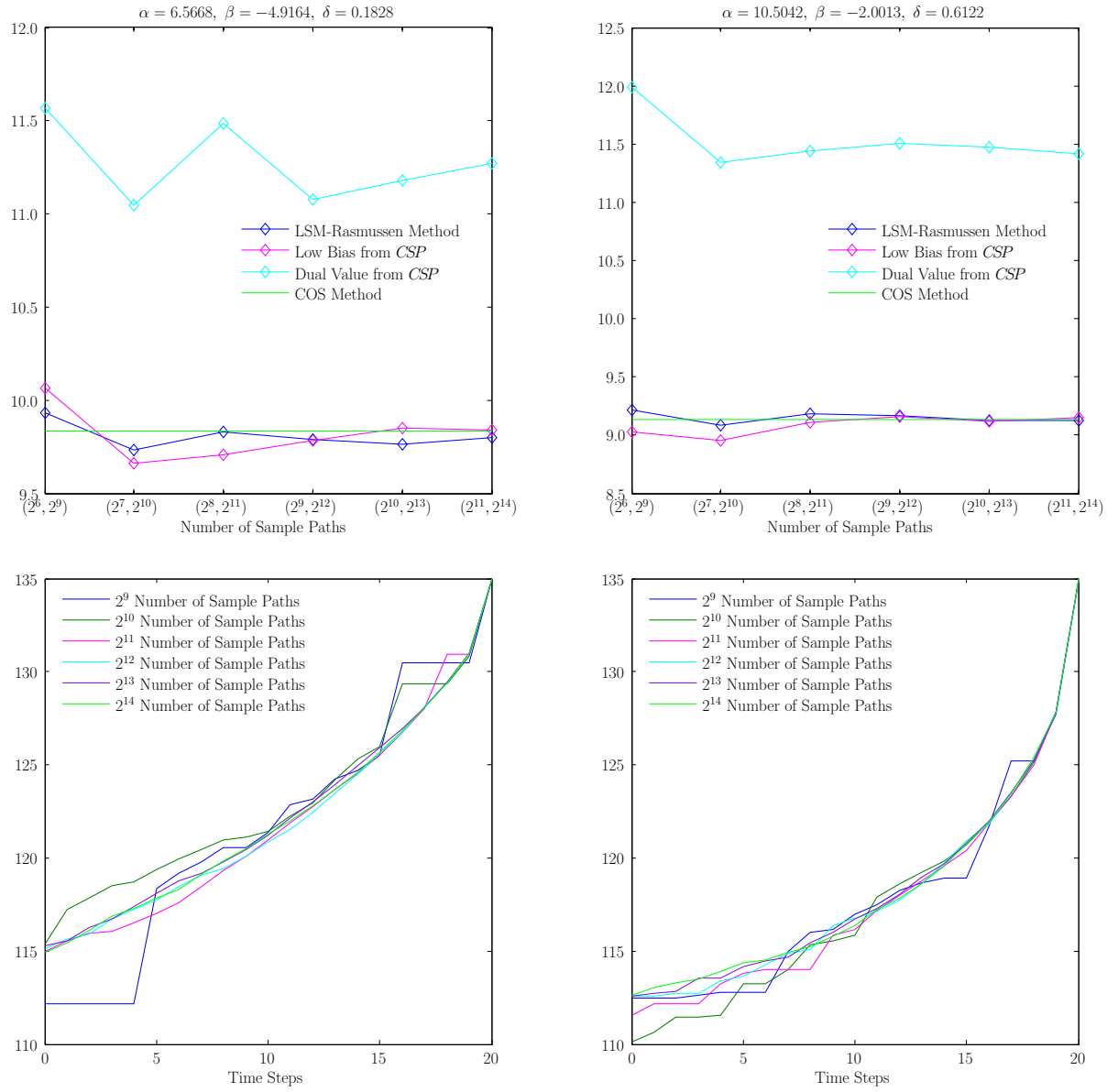

Figure 12.9: *In the first row, we plot the performance of the LSM-Rasmussen method as a function of the number of sample paths used with 8 Laguerre polynomials. We also plot the low bias obtained from the critical stock price function calculated in the LSM method. Furthermore, we plot the performance of the dual method which is calculated using this critical stock price function. The number of sample paths used in the high bias calculations and LSM method is given by the first entry in the bracket and that of the low bias estimate in the second on the horizontal axis. We also include the value given by the COS method with 20 time steps.*

In the second row, we plot the corresponding critical stock price function determined from the LSM-Rasmussen method for a varying number of sample paths.

The option details are as follows: the underlying stock price process follows exponential NIG with parameters indicated in the titles of the figures, $S_0 = 135$, $r = 10\%$ *and* $q = 2\%$*; and we are considering a 1 year put option with* $K = 135$. *Sample paths were generated with 20 time steps using Sobol' numbers with bridging.*

Appendix A

Singular Value Decomposition

A naïve attempt at solving for β in a regression equation such as (2.11) is to try to calculate the inverse of $X'X$ so that

$$\beta = (X'X)^{-1} X'Y.$$

One can also solve for β by performing LU decomposition and then forward reduction and back substitution (see Press et al. [2004, §2.3] for example) without explicitly calculating the inverse of $X'X$. This will be faster and produce a more accurate result. But this will work well only if the number of columns of X is small (which is the number of basis functions chosen).

However, the number of basis functions will typically not be small. As noted in Longstaff and Schwartz [2001], some basis functions are highly correlated with each other, which will result in a cross-moment matrix that is almost singular as soon as the number of basis functions is not small. That is, the columns of X are close to being colinear and so $X'X$ is close to being singular. In this case either of the above methods will be numerically unstable because the matrix manipulation involves numbers which might be flirting with machine precision issues. However, a technique known as *singular value decomposition (SVD)* may be used to solve this problem. According to Press et al. [2004, §2.6], SVD is the method of choice for solving most linear least-squares problems.

Suppose that X is an $N \times \kappa$ matrix with $\kappa < N$, then X has at most κ nonzero singular values and these are real and positive. Let these *singular values* be denoted by λ_k, $k = 1,\ 2,\ \ldots,\ \kappa$ which without loss of generality can be arranged in decreasing order. According to a theorem in linear algebra (see Mehrmann [2003] or Shores [2007, §5.6] for example) X may be written as

$$X = U\Sigma V'$$

where U and V are $N \times N$ and $\kappa \times \kappa$ unitary matrices respectively. Here Σ is an $N \times \kappa$ matrix where nonzero entries only occur on the 'diagonal':

$$\Sigma_{ik} = \begin{cases} \lambda_i & \text{for } 1 \leq i = k \leq \kappa \\ 0 & \text{otherwise.} \end{cases}$$

We then calculate the Moore-Penrose pseudoinverse of X

$$X^+ = V\Sigma^+ U'$$

where Σ^+ is an $\kappa \times N$ matrix with

$$\Sigma^+_{ki} = \begin{cases} \frac{1}{\lambda_i} & \text{for } 1 \leq i = k \leq \kappa,\ \lambda_i \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

Thus, X^+ is $\kappa \times N$. The Moore-Penrose pseudoinverse satisfies the following four properties

$$XX^+X = X \tag{A.1}$$
$$X^+XX^+ = X^+$$
$$(XX^+)' = XX^+ \tag{A.2}$$
$$(X^+X)' = X^+X.$$

A matrix with these properties is unique[1]. This characterisation and uniqueness was proved in Penrose [1955].

Consider (2.11) again and choose $\beta = X^+Y$, then

$$\begin{aligned} X'X\beta &= X'XX^+Y \\ &= X'\left(XX^+\right)'Y \\ &= \left(XX^+X\right)'Y \\ &= X'Y \end{aligned}$$

where the second equality follows from (A.2) and the last from (A.1). It can be shown with this choice the ℓ^2-norm of $X\beta - Y$ is minimised (see [Gentle, 2003, §6.7.3]).

Note that calculating β requires the multiplication of four matrices, i.e.

$$\beta = X^+Y = V\Sigma^+U'Y.$$

Since matrix multiplication is associative, we can choose the order in which these matrices are multiplied together. However the order of multiplication chosen will affect the efficiency of our code. The problem of choosing the most efficient way to multiply matrices together is in general called *matrix chain multiplication*. The problem here is slightly different because of the presence of many zeros in the Σ^+ matrix. To see this, suppose that Σ^+ was fully populated, then calculating Σ^+U' results in KN^2 multiplicative operations. However, since Σ^+ only has entries on its 'diagonal', calculating Σ^+U' can be done with only KN multiplicative operations.

We determined the five different ways in which these matrices may be multiplied together:

(i) $V\Sigma^+,\ U'Y \to V\Sigma^+U'Y$ which gives $2\kappa^2 + N^2$ multiplicative operations;

(ii) $V\Sigma^+ \to V\Sigma^+U' \to V\Sigma^+U'Y$ which gives $\kappa^2 + \kappa^2 N + KN$ multiplicative operations;

[1] A *pseudoinverse* or *generalised inverse* of a matrix is one that only satisfies (A.1). A generalised inverse is not unique.

(iii) $\Sigma^+U' \to V\Sigma^+U' \to V\Sigma^+U'Y$ which gives $2KN + \kappa^2 N$ multiplicative operations;

(iv) $U'Y \to \Sigma^+U'Y \to V\Sigma^+U'Y$ which gives $N^2 + \kappa + \kappa^2$ multiplicative operations;

(v) $\Sigma^+U' \to \Sigma^+U'Y \to V\Sigma^+U'Y$ which gives $2KN + \kappa^2$ multiplicative operations.

Clearly (v) results in the most efficient algorithm. Furthermore, since Y is a vector of size N and Σ^+ can be represented by a vector of size κ we may calculate

$$\left[\Sigma^+U'Y\right]_k = \left[\sum_{i=1}^{N} \frac{1}{\lambda_k} u_{i,k} y_i\right]_k .$$

Appendix B

Sobol' Sequences

Sobol' sequences were originally introduced in Sobol' [1967]. We present here the algorithm appearing in Bratley and Fox [1988] using the modification as given in Antonov and Saleev [1979]. We find Šelić [2006, §4.2.1], who prices using Monte Carlo in high dimensions — in the Libor Market Model — to have a very instructive presentation on these matters. Another useful account of Sobol' sequences is given in Glasserman [2004, §5.2.3].

B.1 Sobol' Sequences in Multiple Dimensions

Generating Sobol' numbers for more than one dimension requires the selection of a different primitive polynomial modulo 2 for each dimension.

Definition B.1.1 Primitive Polynomial of Degree d Over $\mathbb{Z}_2$
A primitive polynomial of degree d over $\mathbb{Z}_2$ *is irreducible and is given by*

$$p(x) = x^d + a_1 x^{d-1} + \ldots + a_{d-1}x + 1 \tag{B.1}$$

where $a_k \in \mathbb{Z}_2 = \{0, 1\}$ *for* $k = 1,\ 2,\ \ldots,\ d-1$ *and* d *is the smallest integer such that* $p(x) | \, x^{2^d-1} + 1$.

Clearly the leftmost bit of a primitive polynomial is always 1. However, note that this is also the case for the rightmost it, otherwise the polynomial is divisible by x, and therefore not irreducible.

A list of all the primitive polynomials modulo 2 up to degree 27 (which is equal to $8,129,334$ primitive polynomials) is provided in Jäckel [2002, accompanying CD] (see Table B.1 for a list of polynomials up to degree 8). Here each polynomial is encoded as a decimal number whose binary representation indicates the coefficients of the specific polynomial but with the leftmost and rightmost bits removed. The rightmost bit is removed simply by shifting the binary representation right.

The list in Jäckel [2002, accompanying CD] is ordered so that for each degree, the numbers appearing in decimal representation are in increasing order. To pick a primitive polynomial, we choose a decimal number from the list, and simply convert it into its binary representation which is in the form of $a_1 a_2 \ldots a_{d-1}$ as in (B.1).

Example B.1.2.
Consider the polynomial $x^3 + x^2 + 1$. Its binary representation is given by 1101_2. Removing the leftmost bit gives 101_2. Shifting this to the right removes the rightmost bit which yields 10_2 and this is the decimal number 2 in the list. □

It is recommended in Jäckel [2002, Accompanying CD] to only compile as many polynomials as ever will be needed since 8,129,334 longs compile into an object file of at least 32,517,336 byte size. In fact, just having this c file open caused our fairly powerful computer to hang frequently and therefore we have deleted all polynomials of degree 15 and higher. Thus we are left with all polynomials up to degree 14 which gives 1867 polynomials.

A list of the 53 polynomials of lowest degree is given in Glasserman [2004, Table 5.2] which is shown in Table B.2. These polynomials are from Bratley and Fox [1988]. Unlike the polynomials provided by Jäckel [2002, Accompanying CD], the first polynomial is given by 1 not $x + 1$. Also, the leftmost and rightmost bits have not been removed[1]. One should note carefully that the ordering of the polynomials in Glasserman [2004, Table 5.2] is different to that of Jäckel [2002, Accompanying CD]. Consider, for example, the second polynomial of degree 5 from Table B.1, that is 4. Its representation in Table B.2 is given by

$$2^5 + 2.4 + 1 = 41$$

which is the last polynomial of degree 5 in Table B.2. The order of these polynomials plays an important role, especially when using specific initialisation numbers.

Degree	Decimal Encoding of Primitive Polynomials Jäckel [2002, Accompanying CD]
0	N/A
1	0
2	1
3	1 2
4	1 4
5	2 4 7 11 13 14
6	1 13 16 19 22 25
7	1 4 7 8 14 19 21 28 31 32 37 41 42 50 55 56 59 62
8	14 21 22 38 47 49 50 52 56 67 70 84 97 103 115 122

Table B.1: *Polynomials up to degree 8 as provided in Jäckel [2002, Accompanying CD].*

A list of 1111 polynomials used in Joe and Kuo [2003] can be found at `http://web.maths.unsw.edu.au/~fkuo/sobol/joe-kuo-old.1111`. Here the polynomial encoding is given in the form provided

[1]Thus to translate between Table B.1 and Table B.2 we have the following formulae

$$g = 2^d + 2j + 1 \text{ and } j = (g - 2^d - 1)/2. \tag{B.2}$$

Here g and j indicates the decimal encoding of the polynomials given in Glasserman [2004, Table 5.2] and Jäckel [2002, Accompanying CD] respectively, and d denotes the polynomial's corresponding degree.

Degree	Decimal Encoding of Primitive Polynomials Glasserman [2004, Table 5.2]																	
0	1																	
1	3																	
2	7																	
3	11	13																
4	19	25																
5	37	59	47	61	55	41												
6	67	97	91	109	103	115												
7	131	193	137	145	143	241	157	185	167	229	171	213	191	253	203	211	239	247
8	285	369	299	425	301	361	333	357	351	501	355	397	391	451	463	487		

Table B.2: *Polynomials up to degree 8 as found in Glasserman [2004, Table 5.2].*

in Jäckel [2002, Accompanying CD]. Also, the ordering of the polynomials up to dimension 40 agree with that of Glasserman [2004, Table 5.2], but for higher dimensions primitive polynomials of the same degree are arranged in increasing order of the decimal representation of these polynomials.

A list of polynomials up to dimension 19000 used in Joe and Kuo [2008] can also be found at `http://web.maths.unsw.edu.au/~fkuo/sobol/joe-kuo-6.19000`. These primitive polynomials coincide with Jäckel [2002, Accompanying CD].

As we have mentioned before, each dimension requires the selection of a primitive polynomial. We have warned that the order in which polynomials are chosen is important. It is suggested in Joe and Kuo [2003] that the chosen polynomial is of the lowest degree possible. Thus, for each dimension, we select the polynomial of lowest degree, but which polynomial with a specific degree shall we choose? The selection within a degree depends on the specific initialisation numbers implemented and will be discussed later.

Once our choice of polynomial has been made, generating Sobol' sequences for multidimensional problems is reduced to generating Sobol' sequences, for single dimensions, in each dimension.

B.2 Generating a Sobol' Sequence in 1 Dimension

Given the primitive polynomial the algorithm generates W *direction numbers* $v_1, v_2, \ldots, v_W \in (0, 1)$. The number W represents the number of bits in an unsigned integer on the given computer and will typically be 32 [2]. In order to obtain these direction numbers, we will generate a sequence $m_k \in \mathbb{Z}$ for

[2] In our c++ code we have set W to 32. It should be noted that if an integer instead of an unsigned integer is used, then W needs to be set to 31 as the first bit is used for the sign and so the 32^{nd} number will be some nonsense resulting from overflow.

$k = 1,\ 2,\ \ldots,\ W$, and put[3]

$$v_k = \frac{m_k}{2^k}. \tag{B.3}$$

The way in which these m_k are generated is split into two groups. For $k \leq d$ there are many alternative approaches to determine the m_k. Furthermore it is important to carefully choose the approach and hence we spend some time in discussing the various options. For $k > d$, there is only one way in which the m_k are obtained — see the recursion formula (B.4).

B.2.1 Finding m_k when $k \leq d$

Let d indicate the degree of the polynomial. Then for $k = 1,\ 2,\ \ldots,\ d$, choose m_k to be odd integers such that $0 < m_k < 2^k$. In Jäckel [2002, §8.3.4] the m_k are referred to as *initialisation numbers* since the direction numbers constructed from the m_k initialise the entire construction of the sequence.

There is some freedom in the choice of the initialisation numbers. However care should be taken when selecting the initialisation numbers as the performance of the Sobol' sequence is very dependent on the choice of these numbers.

Additional uniformity properties, called *Property A* and *Property A′*, provided in Sobol' [1976] in order to choose the appropriate initialisation numbers, are discussed in Bratley and Fox [1988], Joe and Kuo [2003] and Jäckel [2002, §8.3.4].

We give here the definitions of Properties A and A′ as presented in Glasserman [2004, §5.2.3].

Definition B.2.1 Property A
An M-dimensional sequence $s_0,\ s_1,\ \ldots,\ s_{M-1}$ satisfies Property A *if for every $j = 0,\ 1,\ \ldots,\ M-1$ precisely one of the s_i where $j2^M \leq i < (j+1)2^d$ is located in each of the 2^M cubes given as*

$$\prod_{j=1}^{M} \left[\frac{a_j}{2}, \frac{a_j+1}{2}\right)$$

where the $a_j \in \{0, 1\}$.

Definition B.2.2 Property A′
An M-dimensional sequence $s_0,\ s_1,\ \ldots,\ s_{M-1}$ satisfies Property A′ *if for every $j = 0,\ 1,\ \ldots,\ M-1$ precisely one of the s_i where $j4^M \leq i < (j+1)4^d$ is located in each of the 4^M cubes given as*

$$\prod_{j=1}^{M} \left[\frac{a_j}{4}, \frac{a_j+1}{4}\right)$$

where the $a_j \in \{0,\ 1,\ 2,\ 3\}$.

[3]Dividing by 2^k implies that in the binary representation of m_k, the bits are shifted k places to the right. In our c++ code, the vectors containing the values of m_k and v_k are unsigned integers as the bitwise XOR does not apply to floating point numbers. Now in c++, when applying the shift operators $\ll$ and $\gg$ on unsigned integers, the result is a *logical shift*. This means that bits which are discarded, are shifted out and zeros are shifted in (the direction depends on which way the shift occurs). Therefore, in order to prevent the loss of information when shifting, we have also multiplied the entries of the vector containing m_k by 2^W as well as dividing by 2^k.

A table of these initialisation numbers appears in Press et al. [2004, §7.7] (not the numbers in brackets which are calculated later using (B.4), but the numbers before them) and are named *starting values*. It is not shown how these numbers were chosen, and it is noted in Glasserman [2004, §5.2.3] that these numbers fail the test for Property A in dimension 3, 5 and 6.

An implementation for the initialisation numbers up to 40 dimensions is provided in Bratley and Fox [1988]. However, it is remarked in Glasserman [2004, §5.2.3] that Property A does not consistently hold for these numbers for dimensions greater than 19. These initialisation numbers (up to dimension 19) are also given in Glasserman [2004, Table 5.3] which we present in Table B.3. It is noted in Glasserman [2004,

Dimension	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
0	1	(1)	(1)	(1)	(1)	(1)	(1)	(1)
1	1	(3)	(5)	(15)	(17)	(51)	(85)	(255)
2	1	1	(7)	(11)	(13)	(61)	(67)	(79)
3	1	3	7	(5)	(7)	(43)	(49)	(147)
4	1	1	5	(3)	(15)	(51)	(125)	(141)
5	1	3	1	1	(9)	(59)	(25)	(89)
6	1	1	3	7	(31)	(47)	(109)	(173)
7	1	3	3	9	9	(57)	(43)	(43)
8	1	3	7	13	3	(35)	(89)	(9)
9	1	1	5	11	27	(53)	(69)	(25)
10	1	3	5	1	15	(19)	(113)	(115)
11	1	1	7	3	29	(51)	(47)	(97)
12	1	3	7	7	21	(61)	(55)	(19)
13	1	1	1	9	23	37	(97)	(97)
14	1	3	3	5	19	33	(3)	(197)
15	1	1	3	13	11	7	(37)	(101)
16	1	1	7	13	25	5	(83)	(255)
17	1	3	5	11	7	11	(103)	(29)
18	1	1	1	3	13	39	(27)	(203)
19	1	3	1	15	17	63	13	(65)

Table B.3: *Initialisation values provided in Glasserman [2004, Table 5.3]. Values in brackets are obtained from (B.4).*

§5.2.3], whether or not Property A holds for a set of initialisation numbers depends on whether m_k is equal to 1 for all k in dimension 0.

In Lemieux et al. [2002], an implementation for initialisation numbers for up to 360 dimensions is provided. In Glasserman [2004, §5.2.3.] it is noted that their values do not necessarily satisfy Property A, but they are the result of a search for good values based on a resolution criterion used in design of random number generators.

Initialisation numbers for up to 1111 dimensions used in Joe and Kuo [2003] are provided at `http://web.maths.unsw.edu.au/~fkuo/sobol/joe-kuo-old.1111`, satisfying Property A. The initiali-

sation numbers of the first 20 dimensions agree with the values given in Glasserman [2004, Table 5.3].

Furthermore, it is advised in Joe and Kuo [2003] (see also Glasserman [2004, §5.2.3]) that the values for these initialisation numbers should be chosen differently for any two primitive polynomials of the same degree.

The use of a separate pseudo-random number generator for generating the initialisation numbers is suggested in Jäckel [2002, §8.3.4]: for $k = 1,\ 2,\ \ldots,\ d$, draw a uniform random number u from a pseudo-random number generator and set

$$m_k := \lfloor u 2^k \rfloor$$

if $\lfloor u 2^k \rfloor$ is odd, otherwise keep drawing until $\lfloor u 2^k \rfloor$ is odd.

It is noted in Joe and Kuo [2003] that Property A alone does not ensure the absence of bad correlations between pairs of dimensions. For details see Morokoff and Caflisch [1994]. A new set of initialisation numbers to help improve the problem of poor two-dimensional projections can be found at `http://web.maths.unsw.edu.au/~fkuo/sobol/joe-kuo-6.19000`. These initialisation numbers, given in Table B.4 are found using the search algorithm given in Joe and Kuo [2008] and satisfy Property A up to dimension 1111.

Dimension	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8
0	1	(1)	(1)	(1)	(1)	(1)	(1)	(1)
1	1	(3)	(5)	(15)	(17)	(51)	(85)	(255)
2	1	3	(3)	(9)	(29)	(23)	(71)	(197)
3	1	3	1	(5)	(31)	(29)	(81)	(147)
4	1	1	1	(11)	(31)	(55)	(61)	(157)
5	1	1	3	3	(25)	(9)	(43)	(251)
6	1	3	5	13	(11)	(37)	(31)	(227)
7	1	1	5	5	17	(9)	(9)	(45)
8	1	1	5	5	5	(53)	(53)	(113)
9	1	1	7	11	19	(37)	(69)	(91)
10	1	1	5	1	1	(27)	(79)	(35)
11	1	1	1	3	11	(43)	(75)	(43)
12	1	3	5	5	31	(35)	(113)	(51)
13	1	3	3	9	7	49	(33)	(163)
14	1	1	1	15	21	21	(77)	(157)
15	1	3	1	13	27	49	(35)	(133)
16	1	1	1	15	7	5	(123)	(103)
17	1	3	1	15	13	25	(27)	(109)
18	1	1	5	5	19	61	(87)	(187)
19	1	3	7	11	23	15	103	(65)

Table B.4: *Initialisation values found using the search algorithm given in Joe and Kuo [2008]. Values in brackets are obtained from (B.4).*

B.2.2 Finding m_k when $k > d$

Now for $d < k \leq W$, the m_k are found by using the a_k in (B.1) as follows

$$m_k = 2a_1 m_{k-1} \oplus 2^2 a_2 m_{k-2} \oplus \ldots \oplus 2^{d-1} a_{d-1} m_{k-d+1} \oplus 2^d m_{k-d} \oplus m_{k-d} \tag{B.4}$$

where $\oplus$ indicates the bitwise *exclusive-or (XOR)* [4].

B.2.3 Generating the Sobol' Sequence

Finally, the Sobol' sequence $x_1, x_2, \ldots, x_N$ is generated recursively according to the method suggested in Antonov and Saleev [1979] that makes use of *Gray codes* which enable the generation of a new unique integer for each new draw (see Press et al. [2004, 20.2], Glasserman [2004, §5.2.3] or Jäckel [2002, §8.3.3] for more on Gray codes):

$$x_0 = 0 \tag{B.5}$$

$$x_{i+1} = x_i \oplus v_{c(i)} \tag{B.6}$$

for $0 \leq i \leq N-1$ and $c(i)$ is the position of the least significant zero in the binary expansion of i [5]. The construction in (B.5) and (B.6) ensures that the zero$^{\text{th}}$ draw is the only draw that can be zero.

B.3 Our Implementation

Source code to construct Sobol' sequences in 6 dimensions is provided in Press et al. [2004]. Besides the fact that this code is difficult to understand and hence to extend or modify, it appears to be appropriate only for Monte Carlo simulation for European options as it generates random numbers for Monte Carlo simulation path by path. This is not suitable for American Monte Carlo methods: here we require dimension by dimension. Therefore we have written our own code (not using Press et al. [2004] at all) which, given the sample size and dimension, generates Sobol' numbers dimension by dimension.

It is mentioned in Joe and Kuo [2003] that in order to avoid the problems caused by a bad choice of initialisation numbers, one should skip the initial part of the Sobol' sequence. In Acworth et al. [1996] it is noted that standard low discrepancy sequences are known to perform better if an initial portion of the sequence is removed. A strategy is given in Acworth et al. [1996], also implemented in Joe and Kuo [2003], which drops the largest power of two smaller than N number of points, that is, $2^{\lfloor \ln N / \ln 2 \rfloor}$ points.

As noted in our introduction, quasi-Monte Carlo methods usually outperform pseudo-Monte Carlo methods with antithetics. We do not use antithetics in conjunction with Sobol' numbers — Mark Joshi comments (on the Wilmott forum, July 2008) that it would be unwise because "antithetic gives you certain properties that are already present when doing Sobol". Similarly, Jäckel [2002, §20.1] notes that using the antithetic method along with low discrepancy numbers is unlikely to improve the accuracy and may lead to incorrect results.

In order to make full use of the 'antithetic property' that comes free of charge with Sobol' sequences, one should choose N and the number of points to discard carefully. What is meant by the 'antithetic

[4] That is, $0 \oplus 0 = 1 \oplus 1 = 0$ and $0 \oplus 1 = 1 \oplus 0 = 1$. Thus $12 \oplus 10 = 1100_2 \oplus 1010_2 = 110_2 = 6$.
[5] For example, $c(5) = c(101_2) = 2$ and $c(7) = c(111_2) = 4$.

property' is that a Sobol' sequence consists of consecutive sets where for every number u occurring in the set, $1-u$ also occurs. Such a set has a size which is a power of two. These sets start in positions 1, 2, 4, 8, 16, etc. (Counting starts at the 0^{th} index, with initialisation number 0.)

Thus we should insist that N is a power of two, say $N = 2^T$. We choose the set $x_{2^T}, x_{2^T+1}, \ldots, x_{2^{T+1}-1}$ which is antithetic and is found by discarding $x_0, x_1, \ldots, x_{2^T-1}$, that is we discard the first 2^T entries from the original sequence[6]. This justifies the unexplained strategy in Acworth et al. [1996]. On the other hand, if one does not discard any Sobol' numbers, then one takes a union of the first occurring antithetic sets, so the Sobol' sequence that will be used is given by $x_1, x_2, \ldots, x_{2^T-1}$, that is, the first $2^T - 1$ numbers excluding 0.

B.4 Tests

We compare the Sobol' numbers generated using polynomials and initialisation numbers found in three sources, namely Jäckel [2002, §8.3], Joe and Kuo [2003] and Joe and Kuo [2008]. We do not use the Sobol' numbers generated from the polynomials and initialisation numbers provided in Glasserman [2004, §5.2.3] as these are identical to those in Joe and Kuo [2003].

We first consider plots of 3 dimension pairs from each source in Figures B.1 and B.2. Observe the grouping of Sobol' numbers and white space, or 'chessboard'-pattern occurring for at least one dimension pair from each source.

In the following sections we perform 2 tests to determine which source produces the best quasi random numbers.

B.4.1 Integral Test

As in Joe and Kuo [2003, §4], we approximate the integral taken from Wang and Fang [2003]

$$\int_{[0,1]^M} \prod_{j=1}^{M} \frac{|4x_j - 2| + c_j}{1 + c_j} d\mathbf{x} \tag{B.7}$$

for large M which makes the integration problem harder. Note that this integral has value 1 since dimensions are independent and $\int_0^1 |4u-2|du = 1$ for a uniform $(0,1)$ random number u. The c_j in the above integral determine the difficulty of the integration problem. In Bratley and Fox [1988], the c_j were all set to 0. However, as noted in [Joe and Kuo, 2003, §4], the results obtained in Bratley and Fox [1988] and the discussion in Fox [1986] show that difficulty already occurs with this integration by dimension 40. In Wang and Fang [2003], c_j were chosen to be j or j^2, but integrals there had a much lower effective dimension. In Joe and Kuo [2003, §4] $\{c_j\}$ is chosen to be a slower growing sequence and the c_j is set to $j^{1/3}$ in order to obtain a test integral which is reasonable in high dimensions. We have chosen to take this approach.

A comparison between Sobol' sequences, Faure sequences and shifted lattice rules of the approximation of (B.7) for various dimensions and number of simulations is given in a table in Joe and Kuo [2003, §4]. However no integral test is given in Joe and Kuo [2008] for the new numbers appearing here.

[6] Since $1 + \sum_{i=0}^{T-1} 2^i = 2^T$.

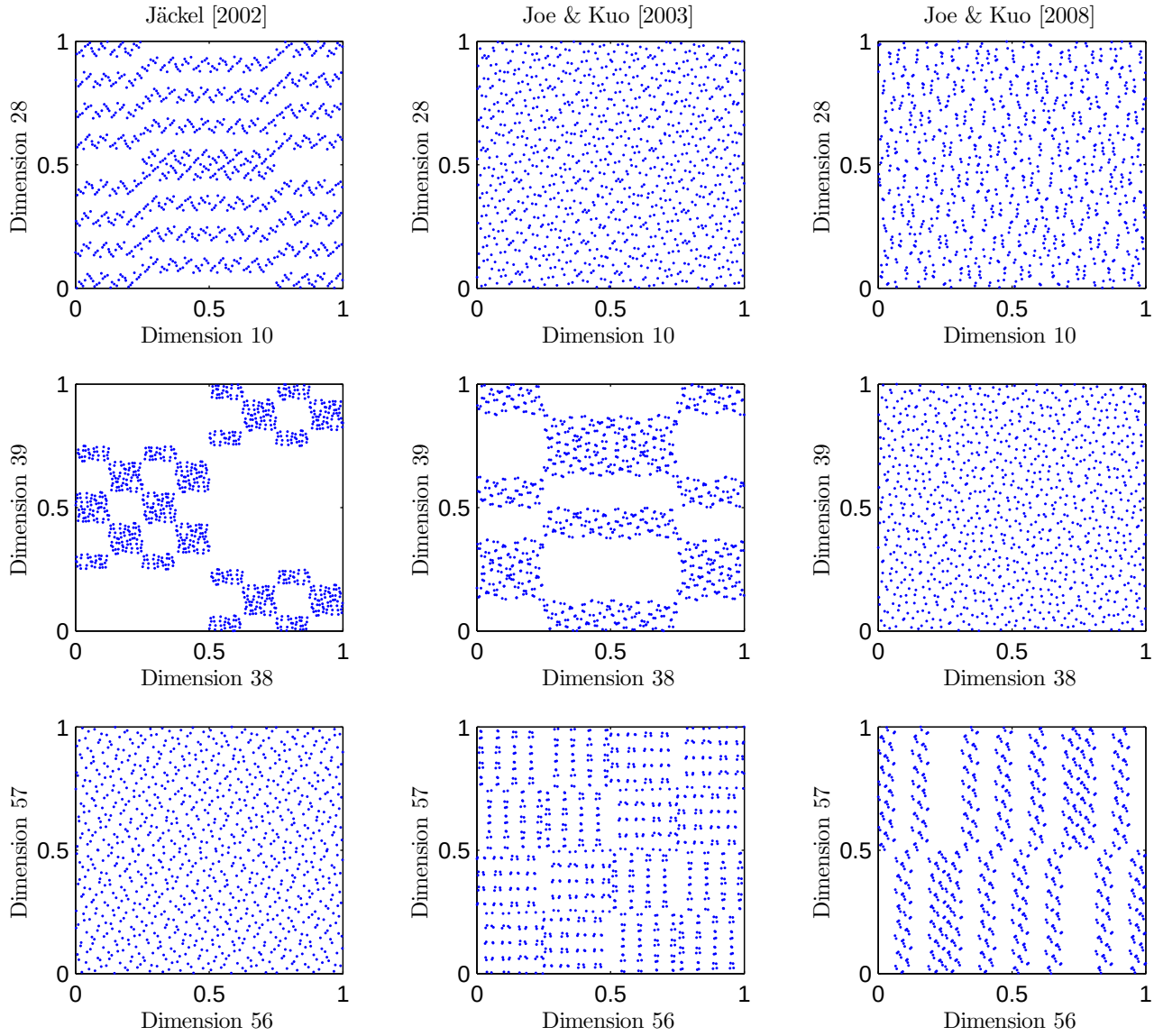

Figure B.1: *Dimension pairs of* $2^{10} = 1024$ *Sobol' numbers generated using the polynomials and initialisation numbers from Jäckel [2002, §8.3], Joe and Kuo [2003] and Joe and Kuo [2008] with discard.*

In Table B.5 we have approximated (B.7) using the Sobol' numbers generated from the polynomials and initialisation numbers as discussed in Jäckel [2002, §8.3], Joe and Kuo [2003] and Joe and Kuo [2008] respectively. We approximate the integral where the number of simulations, N, are powers of 2: $2^{10} = 1024$, $2^{11} = 2048$, $\ldots$, $2^{15} = 32768$; and the dimensions, M, vary from 50 to 1100 in steps of 50. The results of Joe and Kuo [2008] appears to be the best. The results in all cases without the discard were significantly worse.

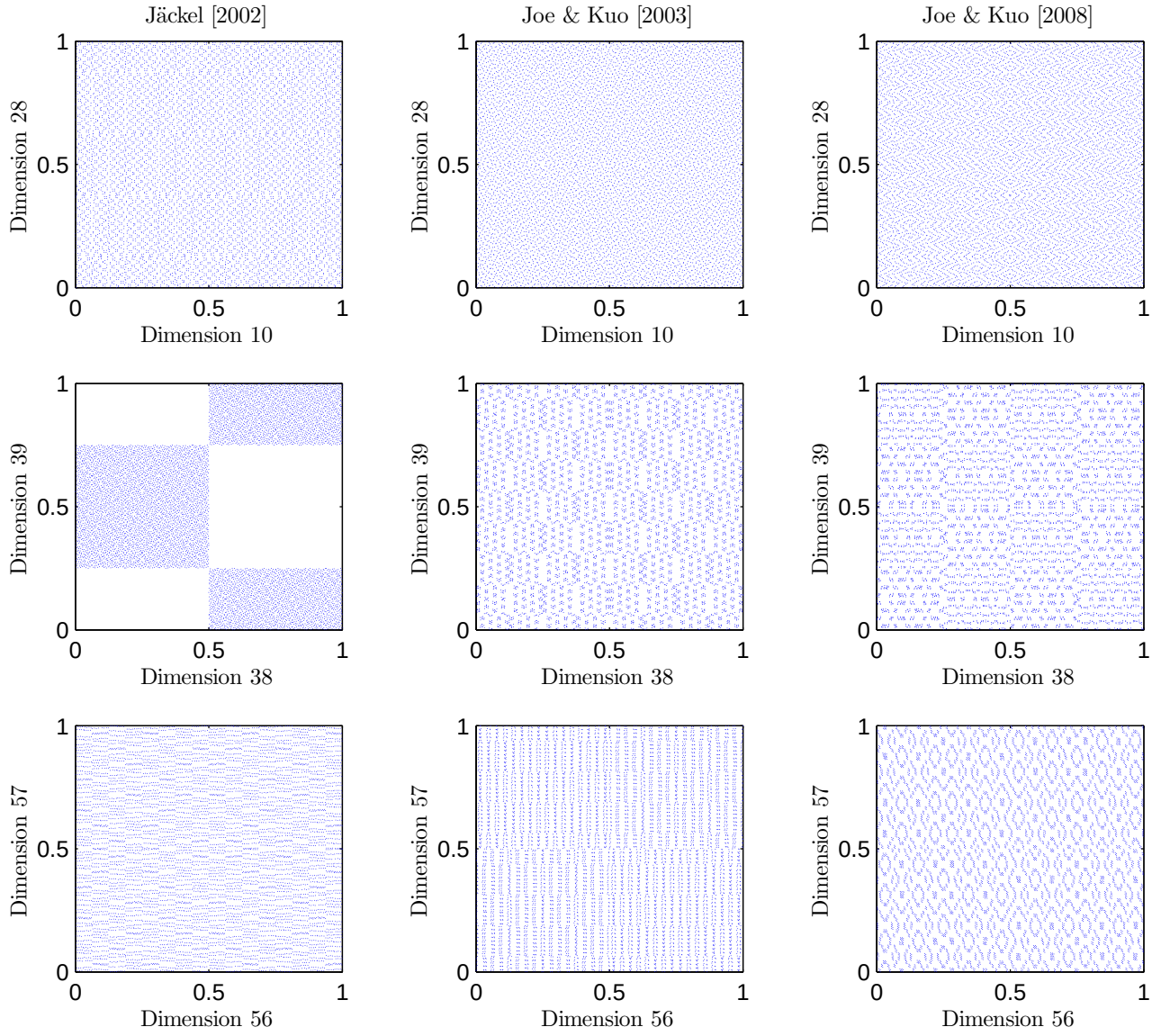

Figure B.2: *Dimension pairs of* $2^{13} = 8192$ *Sobol' numbers generated using the polynomials and initialisation numbers from Jäckel [2002, §8.3], Joe and Kuo [2003] and Joe and Kuo [2008] with discard.*

Even though the integral (B.7) provides a good idea of how well the various generated Sobol' numbers perform, it can fail to pick up gross errors. Suppose we have a uniform random number $U \sim \text{Uniform}\left(0, \frac{1}{2}\right)$,

then (B.7) over 1 dimension is given by

$$\begin{aligned}\mathbb{E}\left[\frac{|4U-2|+c}{1+c}\right] &= \frac{\mathbb{E}\left[|4U-2|\right]+c}{1+c} \\ &= \frac{2\int_0^{\frac{1}{2}}|4u-2|du+c}{1+c} \\ &= \frac{1+c}{1+c} \\ &= 1.\end{aligned}$$

Thus the integral using uniform random numbers that have an incorrect distribution can give the correct answer.

	Jäckel [2002, §8.3]						Joe and Kuo [2003]						Joe and Kuo [2008]					
M \ N	1024	2048	4096	8192	16384	32768	1024	2048	4096	8192	16384	32768	1024	2048	4096	8192	16384	32768
50	0.9904	0.9946	0.9902	0.9923	0.9977	0.9989	1.0161	0.9994	0.9939	0.9983	0.9972	0.9974	1.0179	1.0012	1.0024	1.0012	1.0034	0.9997
100	0.9441	0.9959	0.9907	1.0067	1.0110	1.0088	1.0083	0.9794	0.9675	0.9658	0.9931	0.9958	1.0700	1.0266	1.0346	1.0097	1.0148	1.0102
150	0.9381	0.9358	0.9706	0.9921	1.0080	1.0080	1.0914	0.9904	0.9800	0.9567	0.9878	0.9834	0.9540	0.9836	1.0290	1.0170	1.0352	1.0136
200	0.8770	0.8831	0.9261	0.9522	0.9845	0.9914	1.0983	0.9710	0.9454	0.9512	0.9745	0.9726	0.9421	0.9980	1.0542	1.0180	1.0308	1.0112
250	0.8305	0.8762	0.9641	0.9710	0.9848	0.9838	1.2467	1.0480	0.9437	0.9633	0.9782	0.9741	0.8619	1.0290	1.0576	1.0247	1.0188	1.0076
300	0.8576	0.8833	0.9801	0.9824	0.9934	0.9982	1.1034	0.9456	0.8789	0.9156	0.9534	0.9645	0.8297	1.0947	1.0748	1.0453	1.0502	1.0252
350	0.9540	0.9642	1.0091	1.0192	1.0011	0.9938	1.0569	0.9153	0.8469	0.9005	0.9327	0.9598	0.8368	1.1238	1.2031	1.1394	1.0840	1.0432
400	0.9213	0.9641	1.0266	0.9955	1.0273	1.0047	1.1944	0.9588	0.8843	0.9260	0.9346	0.9660	0.7885	1.0590	1.1930	1.1046	1.0631	1.0332
450	0.8186	0.8874	0.9893	0.9523	1.0182	1.0006	1.2661	1.0318	0.9101	0.9111	0.9164	0.9575	0.8049	0.9965	1.0523	1.0472	1.0487	1.0363
500	0.8797	0.9239	0.9521	0.9250	1.0026	0.9877	1.3290	1.0762	0.9177	0.9332	0.9463	0.9738	0.8370	1.0240	1.1109	1.0425	1.0492	1.0232
550	0.8638	0.9302	0.9617	0.9494	1.0181	0.9945	1.1427	0.9706	0.8530	0.9181	0.9564	0.9781	0.9848	1.2200	1.1437	1.0635	1.0282	1.0128
600	0.8185	0.9614	0.9664	0.9366	1.0219	0.9954	0.9866	0.8887	0.7979	0.8724	0.9058	0.9390	0.8981	1.1905	1.0858	1.0655	1.0695	1.0434
650	0.8185	0.9597	1.0042	0.9667	1.0415	0.9908	1.0255	0.9489	0.8308	0.9092	0.9275	0.9473	0.8374	1.2770	1.0662	1.0291	1.0780	1.0609
700	0.7853	0.9248	0.9857	0.9447	1.0090	0.9890	0.9906	0.8921	0.7754	0.8510	0.8890	0.9340	0.8607	1.1901	1.0393	1.0154	1.1114	1.0984
750	0.7812	0.9848	0.9354	0.9316	0.9761	0.9582	0.9761	0.8876	0.7778	0.8697	0.8876	0.9259	0.9269	1.2404	1.0390	1.0433	1.1535	1.1604
800	0.6964	0.9334	0.9036	0.9079	0.9801	0.9532	0.8740	0.8082	0.7578	0.8481	0.8762	0.9266	0.8516	1.2196	0.9965	1.0460	1.0891	1.1497
850	0.6385	0.9949	0.8901	0.8951	0.9690	0.9284	0.9348	0.8545	0.7747	0.8511	0.8575	0.9190	0.8519	1.0878	0.9219	1.0088	1.1253	1.2072
900	0.6723	1.0935	0.9274	0.8884	0.9472	0.9521	1.1839	0.9615	0.8326	0.8497	0.8589	0.9258	0.7815	1.0048	0.9211	1.0562	1.0990	1.1485
950	0.6347	1.2583	0.9784	0.9096	0.9192	0.9444	1.1837	0.9348	0.8177	0.8266	0.8338	0.9246	0.7822	1.1317	1.0198	1.1903	1.1174	1.1232
1000	0.6635	1.3945	1.0511	0.9614	0.9390	0.9676	1.2447	0.9881	0.8468	0.8876	0.8571	0.9301	0.7403	0.9893	0.9633	1.1711	1.0962	1.0620
1050	0.6474	1.2901	1.0590	0.9639	0.9268	0.9267	1.1112	0.9200	0.8028	0.9080	0.8696	0.9493	0.8184	1.1086	1.0487	1.1844	1.1013	1.1318
1100	0.6418	0.9705	0.8794	0.8568	0.8655	0.8799	1.0568	0.8885	0.7855	0.8635	0.8376	0.9499	0.6897	0.8790	0.9552	1.0961	1.0801	1.1159

Table B.5: *Approximating (B.7) with Sobol' numbers generated using the methods of Jäckel [2002, §8.3], Joe and Kuo [2003] and Joe and Kuo [2008] with discard.*

B.4.2 Covariance Test

In this section we consider a test of Joshi [2011, 22.2].

We begin by generating uniform random numbers using the polynomials and initialisation numbers of Jäckel [2002], Joe and Kuo [2003] and Joe and Kuo [2008] with discard. In addition we also generate uniform random numbers using the Mersenne Twister pseudo-random number generator. We then calculate the cumulative normal inverses of these numbers and calculate the covariances of the dimension pairs. We have chosen to calculate these covariances for dimensions up to 50, 100 and 150 even though the various methods above are able to produce many more dimensions. This is because we will at most be considering 50 time steps in our Monte Carlo simulation (where in some cases we will require 3 random draws per time step per simulation). We then determine the minimum and maximum covariances as shown in Table B.6 and Table B.7.

	Minimum			Maximum		
Source \ Dimensions	50	100	150	50	100	150
Jäckel [2002]	-0.6369	-0.6385	-0.6385	0.2316	0.6383	0.6383
Joe and Kuo [2003]	-0.6415	-0.6415	-0.6415	0.2241	0.6398	0.6440
Joe and Kuo [2008]	-0.0748	-0.2276	-0.2276	0.0750	0.0767	0.0972
Mersenne Twister	-0.1070	-0.1070	-0.1198	0.1152	0.1152	0.1297

Table B.6: *The maximum and minimum covariances of dimension pairs using* $2^{10} = 1024$ *number of simulations with discard for the Sobol' random numbers.*

	Minimum			Maximum		
Source \ Dimensions	50	100	150	50	100	150
Jäckel [2002]	-0.0832	-0.6367	-0.6367	0.0175	0.6370	0.6370
Joe and Kuo [2003]	-0.0735	-0.0735	-0.2236	0.0075	0.6371	0.6371
Joe and Kuo [2008]	-0.0232	-0.0233	-0.0736	0.0063	0.0229	0.0247
Mersenne Twister	-0.0293	-0.0393	-0.0393	0.0344	0.0405	0.0405

Table B.7: *The maximum and minimum covariances of dimension pairs using* $2^{13} = 8192$ *number of simulations with discard for the Sobol' numbers.*

Observe that from the results shown in Table B.6 and Table B.7 it clearly follows that the Sobol' numbers generated using Joe and Kuo [2008] are superior. In Figure B.3 we have plotted the 20 minimum and 20 maximum covariances along with their corresponding dimensions when considering 150 dimensions. Here the results obtained from using the Mersenne Twister and Joe and Kuo [2008] look much better than the results obtained using Jäckel [2002] or Joe and Kuo [2003]. Joe and Kuo [2008] ultimately produce the best results compared to the other three sources.

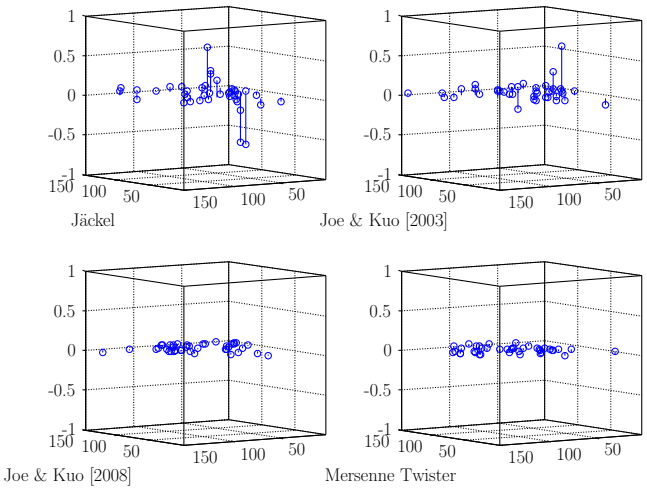

Figure B.3: *The 20 minimum and 20 maximum covariances from the various sources where we have considered 150 dimensions. The results produced by Mersenne Twister and Joe and Kuo [2008] are much better compared to the other sources. The results from Joe and Kuo [2008] are superior to Mersenne Twister — note that the results from the Mersenne Twister are more clustered than that produced by Joe and Kuo [2008].*

Appendix C

Characteristic and Generating Functions

C.1 Characteristic Functions

Characteristic functions play an important role in the analysis of distributions. We shall see in this section that the characteristic function of a random variable completely defines its distribution function. It provides an alternative (and often simpler way) to working directly with probability density or cumulative distribution functions when deriving properties of the distribution.

Definition C.1.1 Characteristic Function
The characteristic function *of a real-valued random variable X is the Fourier transform of μ_X, i.e. the function $\Phi_X : \mathbb{R} \to \mathbb{C}$ defined by*

$$\Phi_X(z) = \int_{-\infty}^{\infty} e^{izx} \mu_X(dx) = \mathbb{E}\left[e^{izX}\right] \tag{C.1}$$

for all $z \in \mathbb{R}$ (see Sato [1999, §1.2], Applebaum [2004, §1.1.6] or Varadhan [2001, §2.1]).

When considering a distribution μ with the random variable X not explicitly stated, we will denote the characteristic function as Φ_μ. If we let $(E, \mathcal{E}, \mu)$ be a measure space with $\mu \geq 0$ and $f : E \to \mathbb{C}$ a measurable complex-valued function, then

$$\left| \int f d\mu \right| \leq \int |f| d\mu$$

where $|\cdot|$ indicates the complex modulus [Cohn, 1980, Proposition 2.6.4]. Now for $\theta \in \mathbb{R}$, $\left|e^{i\theta}\right| = 1$ so that e^{izx} is bounded for every $z \in \mathbb{R}$ and hence

$$|\Phi_\mu(z)| = \left| \int e^{izx} \mu(dx) \right| \leq \int \left|e^{izx}\right| \mu(dx) = \int \mu(dx) = 1$$

if μ is a probability measure. Thus the the integral in (C.1) always exists, that is, the characteristic function is always well-defined.

We determine the characteristic functions of the standard normal and Poisson distributions in the following examples.

Example C.1.2 Characteristic Function of the Poisson Distribution.
An integer valued random variable N is said to follow a *Poisson distribution* with parameter $\lambda > 0$, $\lambda \in \mathbb{R}$ if

$$\mathbb{P}(N = n) = e^{-\lambda}\frac{\lambda^n}{n!}.$$

The characteristic function of N for every $z \in \mathbb{R}$ is calculated as

$$\begin{aligned}\mathbb{E}\left[e^{izN}\right] &= \sum_{n=0}^{\infty} e^{izn}e^{-\lambda}\frac{\lambda^n}{n!} \\ &= e^{-\lambda}\sum_{n=0}^{\infty}\frac{(e^{iz}\lambda)^n}{n!} \\ &= e^{-\lambda}e^{e^{iz}\lambda} \\ &= e^{\lambda(e^{iz}-1)}.\end{aligned}$$

□

Example C.1.3 Characteristic Function of the Standard Normal Distribution.
Let X be a standard normal random variable, then

$$\mu_X(dx) = \mathbb{P}(X \in dx) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx.$$

Then the characteristic function for $z \in \mathbb{R}$ is given by

$$\begin{aligned}\Phi_X(z) &= \int_{-\infty}^{\infty} e^{izx}\,\mu_X(dx) \\ &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{izx}e^{-x^2/2}\,dx \\ &= \frac{1}{\sqrt{2\pi}}\left(\int_{-\infty}^{\infty} e^{-x^2/2}\cos zx\,dx + i\int_{-\infty}^{\infty} e^{-x^2/2}\sin zx\,dx\right).\end{aligned}$$

But $\int_{-\infty}^{\infty} e^{-x^2/2}\sin zx\,dx = 0$ because $e^{-x^2/2}\sin zx$ is an odd function and hence

$$\Phi_X(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\cos zx\,dx.$$

Differentiating with respect to z gives

$$\Phi'_X(z) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} -xe^{-x^2/2}\sin zx dx$$

(we may differentiate inside the integral sign as the function has zero limit at $\pm\infty$). Then integration by parts yield

$$\Phi'_X(z) = -\frac{z}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}\cos zx dx = -z\Phi_X(z)$$

which is a first-order separable differential equation with solution

$$\Phi_X(z) = \Phi_X(0)e^{-z^2/2}.$$

Since $\Phi_X(0) = 1$, we have that

$$\Phi_X(z) = e^{-z^2/2}. \qquad \square$$

Note that the characteristic function of $a + bX$ is given by $e^{iza}\Phi_X(bz)$, $z \in \mathbb{R}$:

$$\begin{aligned}\Phi_{a+bX}(z) &= \mathbb{E}\left[e^{iz(a+bX)}\right] \\ &= \mathbb{E}\left[e^{iza}e^{izbX}\right] \\ &= e^{iza}\mathbb{E}\left[e^{izbX}\right] \\ &= e^{iza}\Phi_X(zb). \end{aligned} \tag{C.2}$$

Example C.1.4 Characteristic Function of Arithmetic Brownian Motion.
Let $X = \{X_t\}_{t\geq 0}$ denote arithmetic Brownian motion, then

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

where $\mu, \sigma > 0$ are real numbers and W_t is standard Brownian motion. Since $W_t \sim \text{Normal}(0, t)$ we have from Example C.1.3 that

$$\Phi_{W_t}(z) = e^{-\frac{1}{2}z^2 t} \tag{C.3}$$

where $z \in \mathbb{R}$. Then from (C.3) and using (C.2) the characteristic function of X is given by

$$\Phi_{X_t}(z) = \exp\left[iz\mu t - \tfrac{1}{2}z^2\sigma^2 t\right] \tag{C.4}$$

for $z \in \mathbb{R}$. In particular, the risk-neutral characteristic function of $\widetilde{X}_t = \left(r - q - \frac{1}{2}\sigma^2\right)t + \sigma W_t$ is given by

$$\Phi_{\widetilde{X}_t}(z) = \exp\left[iz\left(r - q - \frac{1}{2}\sigma^2\right)t - \tfrac{1}{2}z^2\sigma^2 t\right] \tag{C.5}$$

where r indicates the risk-free rate and q the continuous dividend yield. $\square$

Finally, we state an important property of characteristic functions, namely that if the characteristic function of the random variable is known, we also have its distribution function.

Theorem C.1.5
The distribution of a random variable is uniquely determined by its characteristic function.

The above theorem is a consequence of Lévy's inversion formula which provides the link between the characteristic function and distribution function of a random variable (see Williams [1991, §16.6] or Varadhan [2001, §2.1]).

C.2 Moment Generating Functions

The k^{th} *moment about the origin* or *raw moment*, if the expectation exists, of a random variable X on $\mathbb{R}$ is defined as

$$m_k(X) = \mathbb{E}\left[X^k\right], \tag{C.6}$$

and its k^{th} *central moment* by

$$\bar{m}_k(X) = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)^k\right]$$

(see Grimmett and Stirzaker [2001, §3.3] or Casella and Berger [1990, Definition 2.3.1]).

Certain distributions do not possess a moment generating function since some of their moments do not exist. For example the Student's t distribution with n degrees of freedom has only $n-1$ moments and moments of the F distribution up to twice the degrees of freedom of the denominator do not exist. We give a formal definition of the moment generating function next (see Varadhan [2001] or Casella and Berger [1990, Definition 2.3.6]).

Definition C.2.1 Moment Generating Function
The moment generating function *of a real-valued random variable X is the Laplace transform of μ_X, i.e. the function $M : \mathbb{R} \to \mathbb{R}$ defined by*

$$M_X(u) = \int_{-\infty}^{\infty} e^{ux}\, \mu_X(dx) = \mathbb{E}\left[e^{uX}\right]$$

for those $u \in \mathbb{R}$ for which this integral exists. Note that $M_X(0) = 1$.

Example C.2.2 Moment Generating Function of Arithmetic Brownian Motion.
Let $X = \{X_t\}_{t\geq 0}$ denote arithmetic Brownian motion with $X_t = \mu t + \sigma W_t$ where $\mu, \sigma > 0$ are real numbers and W_t is standard Brownian motion. Then we calculate the moment generating function as

$$\begin{aligned} M_{X_t}(u) = \mathbb{E}\left[e^{uX_t}\right] &= \frac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{+\infty} e^{ux} e^{-(x-\mu t)^2/\left(2\sigma^2 t\right)}\, dx \\ &= \frac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{+\infty} \exp\left[\frac{-(x^2 - 2\mu t x + \mu^2 t^2 - 2\sigma^2 t u x)}{2\sigma^2 t}\right] dx. \end{aligned}$$

Completing the square yields

$$\begin{aligned} x^2 - 2\mu t x + \mu^2 t^2 - 2u\sigma^2 t x &= x^2 - 2(\mu t + u\sigma^2 t)x + \mu^2 t^2 \\ &= (x - (\mu t + u\sigma^2 t))^2 - (\mu t + u\sigma^2 t)^2 + \mu^2 t^2 \\ &= (x - (\mu t + u\sigma^2 t))^2 - 2\mu t u\sigma^2 t - u^2\sigma^4 t^2. \end{aligned}$$

Hence

$$\begin{aligned} M_{X_t}(u) &= \frac{1}{\sqrt{2\pi\sigma^2 t}} \int_{-\infty}^{+\infty} \exp\left[\frac{-((x - (\mu t + u\sigma^2 t))^2 - 2\mu t u\sigma^2 t - u^2\sigma^4 t^2)}{2\sigma^2 t}\right] dx \\ &= \exp\left[\frac{u^2\sigma^4 t^2 + 2\mu t u\sigma^2 t}{2\sigma^2 t}\right] \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[\frac{-(x - (\mu t + u\sigma^2 t))^2}{2\sigma^2 t}\right] dx \\ &= \exp\left[\mu t u + \tfrac{1}{2}u^2\sigma^2 t^2\right] \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[\frac{-(x - (\mu t + u\sigma^2 t))^2}{2\sigma^2 t}\right] dx. \end{aligned}$$

But $\frac{1}{\sqrt{2\pi\sigma^2 t}} \exp\left[-(x-(\mu t+u\sigma^2 t))^2\right]$ is the probability density function of a continuous normal random variable with mean $\mu t + u\sigma^2 t$ and variance $\sigma^2 t$. Therefore the moment generating function of X_t is given by

$$M_{X_t}(u) = \exp\left[\mu t u + \tfrac{1}{2}u^2\sigma^2 t\right] \tag{C.7}$$

for $u \in \mathbb{R}$. □

Note that one may write the integral above (if it exists) as

$$\begin{aligned} M_X(u) &= \int_{-\infty}^{\infty} e^{ux}\, \mu_X(dx) \\ &= \int_{-\infty}^{\infty} \sum_{k=0}^{\infty} \frac{(ux)^k}{k!}\, \mu_X(dx) \\ &= \sum_{k=0}^{\infty} \frac{u^k}{k!} \int_{-\infty}^{\infty} x^k\, \mu_X(dx) \\ &= \sum_{k=0}^{\infty} \frac{u^k}{k!} m_k(X) \end{aligned}$$

where the last equation follows from (C.6). If M_X is defined on a neighbourhood of 0, then all moments of X are finite and can be found as follows

$$m_k(X) = \frac{\partial^k M_X}{\partial u^k}(0). \tag{C.8}$$

The central moments may obviously be written in terms of the raw moments using the binomial expansion

$$\bar{m}_k(X) = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} m_k(X) m_1^{n-k}(X).$$

For convenience, the second, third and fourth central moments in terms of the raw moments are given below:

$$\bar{m}_2(X) = m_2(X) - m_1^2(X) \tag{C.9}$$

$$\bar{m}_3(X) = m_3(X) - 3m_2(X)m_1(X) + 2m_1^3(X) \tag{C.10}$$

$$\bar{m}_4(X) = m_4(X) - 4m_3(X)m_1(X) + 6m_2(X)m_1^2(X)^2 - 3m_1^4(X). \tag{C.11}$$

C.3 Cumulant Generating Functions

We define the cumulant generating function in terms of the moment generating function as follows (see Casella and Berger [1990, §2.6.2] or Spanos [1999, §3.7]):

Definition C.3.1 Cumulant Generating Function
The cumulant generating function *of a random variable X is defined as*

$$\Psi_X(u) = \ln M_X(u)$$

whenever the moment generating function M_X is well-defined at $u \in \mathbb{R}$. Note that $\Psi_X(0) = 0$ since $M_X(0) = 1$.

The *cumulants* or *semi-invariants* of X are obtained by partially differentiating the cumulant generating function with respect to u and setting u equal to 0:

$$c_k(X) = \frac{\partial^k \Psi_X}{\partial u^k}(0). \tag{C.12}$$

Then the cumulant generating function can be written in terms of the cumulants as follows

$$\Psi_X(u) = \sum_{k=1}^{\infty} \frac{u^k}{k!} c_k(X).$$

From the definition of the cumulant generating function we have that for a random variable X and $u \in \mathbb{R}$

$$M_X(u) = e^{\Psi_X(u)}.$$

Differentiating the above partially with respect to u yields

$$M'_X(u) = \Psi'_X(u) e^{\Psi_X(u)} = \Psi'_X(u) M_X(u). \tag{C.13}$$

Then setting $u = 0$ gives

$$M'_X(0) = \Psi'_X(0) M_X(0) \tag{C.14}$$

$$\Rightarrow m_1(X) = c_1(X) = \mathbb{E}[X] \tag{C.15}$$

using (C.8). In order to obtain the relationship between the second moment and cumulants we differentiate (C.13) partially with respect to u as before

$$M''_X(u) = \Psi'_X(u) M'_X(u) + \Psi''_X(u) M_X(u) \tag{C.16}$$

and setting $u = 0$ gives

$$\begin{aligned} M''_X(0) &= \Psi'_X(0) M'_X(0) + \Psi''_X(0) M_X(0) \\ \Rightarrow m_2(X) &= c_1(X) m_1(X) + c_2(X) = m_1^2(X) + c_2(X) \\ \Rightarrow c_2(X) &= m_2(X) - m_1^2(X) = \bar{m}_2(X) =: \mathbb{V}\mathrm{ar}[X] \end{aligned} \tag{C.17}$$

which follows from (C.15) and (C.9). Similar to before we differentiate partially (C.16) with respect to u

$$M'''_X(u) = \Psi'_X(u) M''_X(u) + \Psi''_X(u) M'_X(u) + \Psi''_X(u) M'_X(u) + \Psi'''_X(u) M_X(u) \tag{C.18}$$

and setting u to 0 yields

$$\begin{aligned} m_3(X) &= c_1(X) m_2(X) + 2 c_2(X) m_1(X) + c_3(X) \\ &= m_1(X) m_2(X) + 2\left(m_2(X) - m_1^2(X)\right) m_1(X) + c_3(X) \\ \Rightarrow c_3(X) &= m_3(X) - 3 m_2(X) m_1(X) + 2 m_1^3(X) = \bar{m}_3(X) \end{aligned} \tag{C.19}$$

using (C.15), (C.17) and (C.10). Finally we differentiate (C.18) partially with respect to u and obtain

$$\begin{aligned} M_X^{(4)}(u) &= \Psi_X'(u)M_X'''(u) + \Psi_X''(u)M_X''(u) + \Psi_X''(u)M_X''(u) + \Psi_X'''(u)M_X'(u) \\ &\quad + \Psi_X''(u)M_X''(u) + \Psi_X'''(u)M_X'(u) + \Psi_X'''(u)M_X'(u) + \Psi_X^{(4)}(u)M_X(u) \\ &= \Psi_X'(u)M_X'''(u) + 3\Psi_X''(u)M_X''(u) + 3\Psi_X'''(u)M_X'(u) + \Psi_X^{(4)}(u)M_X(u). \end{aligned}$$

Then setting u equal to 0 in the above leads to

$$\begin{aligned} M_X^{(4)}(0) &= \Psi_X'(0)M_X'''(0) + 3\Psi_X''(0)M_X''(0) + 3\Psi_X'''(0)M_X'(0) + \Psi_X^{(4)}(0)M_X(0) \\ \Rightarrow m_4(X) &= c_1(X)m_3(X) + 3c_2(X)m_2(X) + 3c_3(X)m_1(X) + c_4(X) \\ &= m_3(X)m_1(X) + 3\left(m_2(X) - m_1^2(X)\right)m_2(X) + 3\left(m_3(X) - 3m_2(X)m_1(X) + 2m_1^3(X)\right)m_1(X) + c_4(X) \\ \Rightarrow c_4(X) &= m_4(X) - 4m_3(X)m_1(X) - 3m_2^2(X) + 12m_2(X)m_1^2(X) - 6m_1^4(X) \\ &= \bar{m}_4(X) - 3\bar{m}_2^2(X) \end{aligned} \tag{C.20}$$

using (C.15), (C.17), (C.19) and (C.11).

The *skewness* coefficient of X is defined as

$$s(X) = \frac{c_3(X)}{c_2(X)^{3/2}} = \frac{\bar{m}_3(X)}{\bar{m}_2(X)^{3/2}}. \tag{C.21}$$

If $s(X) > 0$ then X is *positively skewed*. The *excess kurtosis* of X is defined by

$$\bar{\varkappa}(X) = \frac{c_4(X)}{c_2(X)^2} = \frac{\bar{m}_4(X) - 3\bar{m}_2(X)^2}{\bar{m}_2(X)^2} = \frac{\bar{m}_4(X)}{\bar{m}_2(X)^2} - 3. \tag{C.22}$$

X is called *leptokurtic* if $\bar{\varkappa}(X) > 0$. The *kurtosis* of X is defined in terms of the central moments by

$$\varkappa(X) = \frac{\bar{m}_4(X)}{\bar{m}_2(X)^2}. \tag{C.23}$$

C.4 Changing Time Frames

Suppose that we have the theoretical mean, variance, skewness and kurtosis of a Lévy process X_t for a time $t_1 > 0$ and would like to know these values for a time $t_2 > 0$. We may write the cumulant generating function of X_t for $u \in \mathbb{R}$ as

$$\Psi_{X_t}(u) = t\Psi_{X_1}(u).$$

Furthermore, recall that we obtain cumulants from the cumulant generating function as in (C.12). Using these two facts we have that

$$c_k(X_t) = \frac{\partial^k \Psi_{X_t}}{\partial u^k}(0) = t\frac{\partial^k \Psi_{X_1}}{\partial u^k}(0) = tc_k(X_1)$$

which implies that for $t_1, t_2 > 0$

$$\begin{aligned} \frac{1}{t_2}c_k(X_{t_2}) = c_k(X_1) = \frac{1}{t_1}c_k(X_{t_1}) \\ c_k(X_{t_2}) = \frac{t_2}{t_1}c_k(X_{t_1}). \end{aligned}$$

Thus, from (C.15) and (C.17) we have that

$$\mathbb{E}\left[X_{t_2}\right] = \frac{t_2}{t_1}\mathbb{E}\left[X_{t_1}\right] \text{ and } \mathbb{V}\text{ar}\left[X_{t_2}\right] = \frac{t_2}{t_1}\mathbb{V}\text{ar}\left[X_{t_1}\right] \tag{C.24}$$

and from (C.21) and (C.22)

$$s\left(X_{t_2}\right) = \sqrt{\frac{t_1}{t_2}}s\left(X_{t_1}\right) \text{ and } \bar{\varkappa}\left(X_{t_2}\right) = \frac{t_1}{t_2}\bar{\varkappa}\left(X_{t_1}\right). \tag{C.25}$$

Appendix D

Modified Bessel Functions of the Second Kind

Modified Bessel functions of the second kind have also been referred to as Basset functions, modified Bessel functions of the third kind or Macdonald functions. The *modified Bessel functions of the second kind* $K_\nu(z)$, is one of the solutions to the modified Bessel differential equation

$$z^2\frac{d^2w}{dz^2} + z\frac{dw}{dz} - \left(z^2 + \nu^2\right)w = 0$$

[National Institute of Standards and Technology, 2010, 10.25.1] with $z \in \mathbb{C}$ and $\nu \in \mathbb{R}$. $K_\nu(x)$ can be expressed in terms of the *modified Bessel function of the first kind* $I_\cdot(\cdot)$

$$K_\nu(z) = \frac{\pi}{2}\frac{I_{-\nu}(z) - I_\nu(z)}{\sin(\nu\pi)}$$

[Abramowitz and Stegun, 1974, 9.6.2] where $I_{-\nu}(z)$ and $I_\nu(z)$ form a fundamental set of solutions of the modified Bessel's equation for noninteger ν.

Modified Bessel functions are equivalent to the ordinary Bessel functions evaluated for purely imaginary arguments. Unlike the usual Bessel functions that have sinusoidal behaviour, the modified functions have exponential behaviour.

In the case of the NIG probability density function given in (9.9) we will consider the modified Bessel function of the second kind and index 1, that is, $K_1(\cdot)$. As noted in Press et al. [2004, §6.6], once the exponential factor is removed, the smoothness of the modified Bessel functions enables one to approximate $I_0(\cdot)$, $I_1(\cdot)$, $K_0(\cdot)$ and $K_1(\cdot)$ suitably with a simple polynomial of a few terms. In our implementation of the NIG density function we will be using the function `bessk1` as in Press et al. [2004, §6.6] in order to calculate $K_1(\cdot)$.

When considering the VG probability density function in (8.11) we will require an algorithm for computing the modified Bessel function of the second kind of fractional order. Calculating these functions requires a complicated algorithm which is discussed in Press et al. [2004, §6.7]. We have implemented the function `bessik` in Press et al. [2004, §6.7] in order to calculate the modified Bessel function of the second kind and fractional order namely $K_{\frac{t}{\nu}-\frac{1}{2}}(\cdot)$ which appears in the VG density function. Complications arise

in the calculation of the probability density function when $x = 0$. Thus when $|x| < \epsilon$ for some $\epsilon > 0$ we will implement the limit as x tends to 0 instead of (8.11). Consider the limit

$$\begin{aligned}\lim_{x\to 0} f_{X_t^{\mathrm{VG}}}(x) &= \lim_{x\to 0} \frac{2\exp\left[\frac{\theta}{\sigma^2}x\right]}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)\sqrt{2\pi\sigma^2}}\left(\frac{x^2}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{\sqrt{x^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}}{\sigma^2}\right)\\ &= \frac{2}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)\sqrt{2\pi\sigma^2}}\left(\frac{1}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} \lim_{x\to 0} x^{\frac{t}{\nu}-\frac{1}{2}} K_{\frac{t}{\nu}-\frac{1}{2}}\left(\frac{\sqrt{x^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}}{\sigma^2}\right) \qquad \text{(D.1)}\end{aligned}$$

Now from National Institute of Standards and Technology [2010, 10.25.1] we have that

$$\begin{aligned}\lim_{x\to 0} x^a K_a(bx) &= \lim_{x\to 0} x^a \frac{1}{2}\Gamma(a)\left(\frac{1}{2}bx\right)^{-a}\\ &= \frac{1}{2}\Gamma(a)\left(\frac{2}{b}\right)^a\end{aligned}$$

for $a, b \in \mathbb{R}$. Thus if we set $a = \frac{t}{\nu} - \frac{1}{2}$ and $b = \frac{\sqrt{\frac{2\sigma^2 t}{\nu}+\theta^2}}{\sigma^2}$, then (D.1) becomes

$$\begin{aligned}&\frac{2}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)\sqrt{2\pi\sigma^2}}\left(\frac{1}{\frac{2\sigma^2}{\nu}+\theta^2}\right)^{\frac{t}{2\nu}-\frac{1}{4}} \frac{1}{2}\Gamma\left(\frac{t}{\nu}-\frac{1}{2}\right)\left(\frac{2\sigma^2}{\sqrt{\frac{2\sigma^2}{\nu}+\theta^2}}\right)^{\frac{t}{\nu}-\frac{1}{2}}\\ &= \frac{\Gamma\left(\frac{t}{\nu}-\frac{1}{2}\right)}{\Gamma\left(\frac{t}{\nu}\right)}\frac{1}{\nu^{\frac{t}{\nu}}\sqrt{2\pi\sigma^2}}\left(\frac{1}{\nu}+\frac{\theta^2}{2\sigma^2}\right)^{\frac{1}{2}-\frac{t}{\nu}}.\end{aligned}$$

Appendix E

Pricing European Options under Lévy Models

The LSM method in Chapter 2 and Rasmussen's modification of the LSM method in Chapter 3, requires the European option price for a given stock price at each exercise date of the Bermudan discretisation of the American option. When the underlying follows geometric Brownian motion, we may use the Black-Scholes formula. In the case of the VG model, Madan et al. [1998] derive a closed-form formula in terms of the modified Bessel function of the second kind and the confluent hypergeometric function of two variables. However as noted in http://demonstrations.wolfram.com/OptionPricesInTheVarianceGammaModel/ these special functions require complicated implementations due to the presence of a singularity in one of the special functions. Furthermore, we were unable to find any implementation for the confluent hypergeometric function of two variables. There are no known closed-form formulae if the underlying follows an exponential NIG process.

The VG process (or NIG process respectively), conditional on knowing the random time change g (which has an independent gamma (or inverse Gaussian) distribution), is normally distributed as we have seen in §8.2 (or §9.2). This suggests another way of calculating the European option price under VG (or NIG) risk-neutral dynamics — integrate a conditional Black-Scholes type formula over g with respect to the gamma density (inverse Gaussian density). Madan and Milne [1991] obtained the option price in the VG case in this way by using numerical integration. Joshi [2003, §17.2] shows how the option price is found in the VG case using an integral over Black-Scholes prices.

The *QUAD method* by Andricopoulos et al. [2003] approximates option prices by considering the risk-neutral valuation formula

$$v\left(x;K,t\right)=e^{-r\Delta t}\int_{-\infty}^{\infty}v\left(y;K,t+\Delta t\right)f_{X_t}\left(y\middle|\,x\right)\,dy$$

where K indicates the strike of the option, $X_t=\ln S_t$, $f_{X_t}(\cdot|\cdot)$ indicates the conditional density of X_t and $v\left(x;K,t\right)$ indicates the option value at time t. In the QUAD method $v\left(y;K,t+\Delta t\right)$ is found by backward recursion and the integral is then approximated using a particular quadrature technique. If the conditional density of the model under consideration is known, this method may be used to price a wide variety of options.

Another way of pricing options makes use of Fourier transform methods. Carr and Madan [1999] first applied Fourier transform methods to price options in the VG model. Lewis [2001] provides a Fourier method allowing for a much easier approximation of the integral at infinity than the method given by Carr and Madan [1999]. Lord et al. [2008] provide yet another Fourier method called the *CONV method*. A Fourier method, called the *COS method*, shows the strongest convergence results and is given by Fang and Oosterlee [2008]. We briefly discuss the last method.

E.1 The COS Method

The main idea of the COS method is to replace the density function with its Fourier Cosine series expansion. As before, let K be the strike, $p(\cdot, K)$ indicate the payoff function and $v(S_0; K, t)$ be the value of the European option at time $t_0 = 0$ with maturity t, then

$$v(S_0; K, t) = e^{-rt} \int_{-\infty}^{\infty} p(x, K) f_{X_t}(x | S_0) \, dx \tag{E.1}$$

where $X_t = \ln S_t$.

As noted by Fang and Oosterlee [2008], since the density rapidly decreases to 0 as $x \to \pm\infty$ in (E.1), a truncation of the infinite integration range can be made without losing significant accuracy. The truncation range, which according to Fang and Oosterlee [2008] is accurate with a truncation error of 10^{-12} for maturities between 0.1 to 10 years, is given by

$$[a, b] = \left[c_1 - L\sqrt{c_2 + \sqrt{c_4}}, c_1 + L\sqrt{c_2 + \sqrt{c_4}} \right] \tag{E.2}$$

where $L = 10$ and c_1, c_2 and c_4 are the cumulants of $\ln \frac{S_t}{K}$. In Table E.1 we provide the cumulants $\bar{c}_1$ $\bar{c}_2$ and $\bar{c}_4$ when the underlying follows geometric Brownian motion, exponential VG and exponential NIG. These may be obtained by using (C.15), (C.17) and (C.20). Then we find the cumulants of $\ln \frac{S_t}{K}$ [1] from

Cumulant	X_1	X_1^{VG}	X_1^{NIG}
$\bar{c}_1$	μ	θ	$\frac{\beta\delta}{\sqrt{\alpha^2-\beta^2}}$
$\bar{c}_2$	σ^2	$\sigma^2 + \nu\theta^2$	$\frac{\alpha^2\delta}{(\alpha^2-\beta^2)^{3/2}}$
$\bar{c}_4$	0	$3\left(\sigma^4\nu + 2\theta^4\nu^3 + 4\sigma^2\theta^2\nu^2\right)$	$\frac{3\alpha^2\left(\alpha^2+4\beta^2\right)\delta}{(\alpha^2-\beta^2)^{7/2}}$

Table E.1: *Cumulants of arithmetic Brownian motion, VG and NIG processes at time 1.*

those in the table as follows

$$c_1 = \bar{c}_1 t + mt + \ln \frac{S_0}{K},\ c_2 = \bar{c}_2 t \text{ and } c_4 = \bar{c}_4 t$$

where m is given by $r - q - \frac{1}{2}\sigma^2$ and $\mu = 0$ in the geometric Brownian case and by (10.6) and (10.10) in the exponential VG and NIG cases respectively.

Thus (E.1) may be approximated by

$$v\left(S_0; K, t\right) \approx e^{-rt} \int_a^b p\left(x, K\right) f_{X_t}\left(x \middle| S_0\right)\, dx. \tag{E.3}$$

As noted by Fang and Oosterlee [2008], $f_{X_t}\left(x \middle| S_0\right)$ is often unknown[2] and therefore Fang and Oosterlee [2008] replace the density function by its Cosine expansion

$$f_{X_t}\left(x \middle| S_0\right) = {\sum_{k=0}^{\infty}}' A_k\left(S_0\right) \cos\left(k\pi \frac{x-a}{b-a}\right) \tag{E.4}$$

where

$$A_k\left(S_0\right) := \frac{2}{b-a} \int_a^b \cos\left(k\pi \frac{x-a}{b-a}\right) f_{X_t}\left(x \middle| S_0\right)\, dx$$

and as in Fang and Oosterlee [2008] $\sum'$ indicates that the 0^{th} term in the summation is weighted by a factor of $\frac{1}{2}$. Thus (E.3) becomes

$$\begin{aligned} v\left(S_0; K, t\right) &\approx e^{-rt} \int_a^b p\left(x, K\right) {\sum_{k=0}^{\infty}}' A_k\left(S_0\right) \cos\left(k\pi \frac{x-a}{b-a}\right)\, dx \\ &= \frac{b-a}{2} e^{-rt} {\sum_{k=0}^{\infty}}' A_k\left(S_0\right) \frac{2}{b-a} \int_a^b p\left(x, K\right) \cos\left(k\pi \frac{x-a}{b-a}\right)\, dx \end{aligned}$$

where the summation and integration has been interchanged in the equality [see Fang and Oosterlee, 2008, (16)].

[1] Recall from (C.15) that

$$\begin{aligned} m_1 = c_1 &= \mathbb{E}\left[\ln \frac{S_t}{K}\right] \\ &= \mathbb{E}\left[\ln S_0 e^{mt + X_t}\right] - \ln K \\ &= \mathbb{E}\left[\ln e^{mt + X_t}\right] + \ln \frac{S_0}{K} \\ &= \mathbb{E}\left[X_t\right] + mt + \ln \frac{S_0}{K} \\ &= \bar{c}_1 + mt + \ln \frac{S_0}{K} \end{aligned}$$

where X_t indicates a VG or NIG process.

[2] Even though this is not strictly true in the VG or NIG case, the probability density functions (see (10.7) and (10.11) respectively) in these cases are in terms of modified Bessel functions of the second kind which can be difficult to implement and are computationally inefficient.

Now A_k may be approximated as

$$A_k \approx \frac{2}{b-a}\Re\left(\Phi_{X_t}\left(\frac{k\pi}{b-a}\right)\exp\left[-i\frac{ka\pi}{b-a}\right]\right)$$

[Fang and Oosterlee, 2008, (9)]. If the characteristic function $\Phi_{X_t}(\cdot)$ is known, as in the case where the stock price process follows geometric Brownian motion (C.4), exponential VG (10.8) or exponential NIG (10.12), this approximation allows for another approximation which will enable us to compute an approximation of the European option price. Looking further afield, we note that the characteristic function of the Heston model is known [see Heston, 1993] and hence it is possible to calculate an approximation of European options under this model as well using the COS method [Fang and Oosterlee, 2008].

Since the coefficients in the Fourier Cosine series decay rapidly, a further approximation involves truncating the series in (E.4). This approximation along with the approximation of A_k produces

$$v\left(S_0;K,t\right) \approx e^{-rt}\sum_{k=0}^{N-1}{}'\Re\left(\Phi_{X_t}\left(\frac{k\pi}{b-a}\exp\left[-i\frac{ka\pi}{b-a}\right]\right)\right)\frac{2}{b-a}\int_a^b p\left(x,K\right)\cos\left(k\pi\frac{x-a}{b-a}\right)dx. \tag{E.5}$$

$\frac{2}{b-a}\int_a^b p\left(x,K\right)\cos\left(k\pi\frac{x-a}{b-a}\right)dx$ can be obtained analytically for vanilla and digital options [see Fang and Oosterlee, 2008, §3.1 & §3.2]. When considering Lévy models, (E.5) may be written as

$$v\left(S_0;K,t\right) \approx Ke^{-rt}\Re\left(\sum_{k=0}^{N-1}{}'\Phi_{X_t}\left(\frac{k\pi}{b-a}\right)U_k\exp\left[ik\pi\frac{x-a}{b-a}\right]\right)$$

[Fang and Oosterlee, 2008, (30)] where

$$U_k = \frac{2}{b-a}\eta\left(\chi_k(0,b)-\psi_k(0,b)\right)$$

[Fang and Oosterlee, 2008, (29)] where $\eta = 1$ indicates a call and $\eta = -1$ indicates a put; and

$$\begin{aligned}\chi_k(c,d) = \frac{1}{1+\left(\frac{k\pi}{b-a}\right)^2}&\left[\cos\left(k\pi\frac{d-a}{b-a}\right)e^d-\cos\left(k\pi\frac{c-a}{b-a}\right)e^c\right.\\&\left.+\frac{k\pi}{b-a}\sin\left(k\pi\frac{d-a}{b-a}\right)e^d-\frac{k\pi}{b-a}\sin\left(k\pi\frac{c-a}{b-a}\right)e^c\right]\\ \psi_k(c,d) = &\begin{cases}\left[\sin\left(k\pi\frac{d-a}{b-a}\right)-\sin\left(k\pi\frac{c-a}{b-a}\right)\right]\frac{b-a}{k\pi} & \text{for } k\neq 0\\ d-c & \text{for } k=0\end{cases}\end{aligned}$$

[Fang and Oosterlee, 2008, (22) & (23)].

Appendix F

Change of Variables

The following result is well known and is often stated in the literature. However proofs are not as plentiful, so we give one.

Proposition F.1.1
Let X be a random variable with density function $f_X(\cdot)$ and $g(\cdot)$ a strictly monotone function that is differentiable. Then the random variable $Y := g(X)$ has density function

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \frac{1}{\left|g'\left(g^{-1}(y)\right)\right|}. \tag{F.1}$$

Proof.
First consider the case where $g(\cdot)$ is increasing, then the cumulative distribution function of Y

$$\begin{aligned} F_Y(y) &= \mathbb{P}\left(Y \leq y\right) \\ &= \mathbb{P}\left(g(X) \leq y\right) \\ &= \mathbb{P}\left(X \leq g^{-1}(y)\right) \\ &= \int_{-\infty}^{g^{-1}(y)} f_X(x)\, dx. \end{aligned}$$

If we let $z = g(x)$, then $dz = g'(x)dx$ and $x = g^{-1}(z)$. Furthermore, if $x \leq g^{-1}(y)$ then since g is increasing $z = g(x) \leq y$. Performing the substitution $z = g(x)$ in the above we obtain

$$F_Y(y) = \int_{-\infty}^{y} f_X\left(g^{-1}(z)\right) \frac{1}{g'\left(g^{-1}(z)\right)}\, dz.$$

By the Fundamental Theorem of Calculus $F_Y(\cdot)$ is differentiable at y and

$$f_Y(y) = F_Y'(y) = f_X\left(g^{-1}(y)\right) \frac{1}{g'\left(g^{-1}(y)\right)}.$$

If g is decreasing then

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) \\ &= \mathbb{P}(g(X) \leq y) \\ &= \mathbb{P}\left(X \geq g^{-1}(y)\right) \\ &= \int_{g^{-1}(y)}^{\infty} f_X(x)\, dx. \end{aligned}$$

A similar substitution as before gives

$$\begin{aligned} F_Y(y) &= \int_{y}^{\infty} f_X\left(g^{-1}(z)\right) \frac{1}{g'\left(g^{-1}(z)\right)}\, dz \\ &= 1 - \int_{-\infty}^{y} f_X\left(g^{-1}(z)\right) \frac{1}{g'\left(g^{-1}(z)\right)}\, dz. \end{aligned}$$

Differentiating yields

$$f_Y(y) = F_Y'(y) = -f_X\left(g^{-1}(y)\right) \frac{1}{g'\left(g^{-1}(y)\right)}.$$

Therefore since g is decreasing g' will be negative and hence we may write for any strictly monotone g

$$f_Y(y) = f_X\left(g^{-1}(y)\right) \frac{1}{|g'\left(g^{-1}(y)\right)|}.$$ □

We give three examples:

Example F.1.2.
Consider the random variable $Y = \lambda X$. If we let $g(x) = \lambda x$, then $g'(x) = \lambda$ and $g^{-1}(y) = \frac{y}{\lambda}$. From (F.1) we have

$$f_Y(y) = f_X\left(\frac{y}{\lambda}\right) \frac{1}{\lambda}$$

since $g'\left(g^{-1}(y)\right) = \lambda$. □

Example F.1.3.
Consider the random variable $Y = \frac{\lambda - X}{X}$. If we let $g(x) = \frac{\lambda - x}{x}$, then $g'(x) = -\frac{\lambda}{x^2}$ and $g^{-1}(y) = \frac{\lambda}{y+1}$. From (F.1) it follows that

$$f_Y(y) = f_X\left(\frac{\lambda}{y+1}\right) \frac{\lambda}{(y+1)^2}$$

since $g'\left(g^{-1}(y)\right) = -\frac{(y+1)^2}{\lambda}$. □

Example F.1.4.
Consider the random variable $Y = \lambda e^X$. If we let $g(x) = \lambda e^x$, then $g'(x) = \lambda e^x$ and $g^{-1}(y) = \ln \frac{y}{\lambda}$. Therefore from (F.1) we have

$$f_Y(y) = f_X\left(\ln \frac{y}{\lambda}\right) \frac{1}{y}$$

since $g'\left(g^{-1}(y)\right) = y$. □

F.2 Returns vs. Prices

The density of the stock price process is found using the transformation

$$S = se^{Y} := g(Y) \tag{F.2}$$

where s is a known stock price at time t_1, S indicates the stock price at some future time $t_2 > t_1$ and Y the return for the period $t_2 - t_1$. The function $g(\cdot)$, as well as its inverse $g^{-1}(\cdot)$, is differentiable and increasing.

In §7.1, §10.2.1 and §10.2.2 we provide the risk-neutral density functions of the returns at a time t that follow arithmetic Brownian motion, VG and NIG processes respectively. However, in some cases we might require the density of the stock price process at t. Here all density functions are risk-neutral, but we suppress the superscript $\mathbb{Q}$ for ease of notation.

Consider (F.2) again where $S = S_{t_2}$ is a random variable representing the stock price at a future time t_2, S_{t_1} is a realised value of the stock price at time $t_1 < t_2$ and let $t := t_2 - t_1$. We can find the density of S from the density of Y using Example F.1.4, as

$$f_S\left(S_{t_2}\right) = f_Y\left(\ln \frac{S_{t_2}}{S_{t_1}}\right) \Big/ S_{t_2}. \tag{F.3}$$

Using (F.3) along with the risk-neutral density of arithmetic Brownian motion (7.6) we obtain the risk-neutral density of the stock price process under geometric Brownian motion

$$f_S(S_{t_2}) = \frac{1}{S_{t_2}\sqrt{2\pi\sigma^2 t}} \exp\left[-\frac{1}{2}\left(\frac{\ln \frac{S_{t_2}}{S_{t_1}} - \left(r - q - \frac{1}{2}\sigma^2\right)t}{\sigma\sqrt{t}}\right)^2\right].$$

Similarly, using the risk-neutral density of VG (10.7) we obtain the risk-neutral density of the stock price process under an exponential VG process

$$f_S(S_{t_2}) = \frac{2\exp\left[\left(\ln \frac{S_{t_2}}{S_{t_1}} - mt\right)\frac{\theta}{\sigma^2}\right]}{S_{t_2}\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)\sqrt{2\pi\sigma^2}} \left[\frac{\left(\ln \frac{S_{t_2}}{S_{t_1}} - mt\right)^2}{\frac{2\sigma^2}{\nu} + \theta^2}\right]^{\frac{t}{2\nu} - \frac{1}{4}} K_{\frac{t}{\nu} - \frac{1}{2}}\left(\frac{\sqrt{\left(\ln \frac{S_{t_2}}{S_{t_1}} - mt\right)^2\left(\frac{2\sigma^2}{\nu} + \theta^2\right)}}{\sigma^2}\right)$$

where m is given by (10.6). Finally from (F.3) and the risk-neutral density of the NIG (10.11) we find the risk-neutral density of the stock price process under an exponential NIG process

$$f_S(S_{t_2}) = \frac{\alpha\delta t}{S_{t_2}\pi} \exp\left[\delta t\sqrt{\alpha^2 - \beta^2} + \beta\left(\ln \frac{S_{t_2}}{S_{t_1}} - mt\right)\right] \frac{K_1\left(\alpha\sqrt{\delta^2 t^2 + \left(\ln \frac{S_{t_2}}{S_{t_1}} - mt\right)^2}\right)}{\sqrt{\delta^2 t^2 + \left(\ln \frac{S_{t_2}}{S_{t_1}} - mt\right)^2}}$$

where m is given by (10.10).

Appendix G

Cumulative Distribution Functions

In the next sections we will discuss the cumulative distribution functions of the gamma and inverse Gaussian and beta distributions and how we will implement them.

We will denote the cumulative distribution functions of the normal, gamma, inverse Gaussian and beta distributions by $F_{\text{N}}(\cdot)$, $F_{\text{G}}(\cdot)$, F_{IG} and $F_{\text{B}}(\cdot)$ respectively. When we refer to the cumulative distribution function of a specific random variable X we will also denote it by $F_X(\cdot)$.

G.1 The Gamma Distribution

Let $X \sim \text{Gamma}(\alpha, \beta)$ with $\alpha > 0$ and $\beta > 0$, then the cumulative gamma distribution function is given by

$$\begin{aligned} F_{\text{G}}(x) &= \mathbb{P}(X \leq x) \\ &= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^x t^{\alpha-1} e^{-\beta t} dt \\ &= \frac{\int_0^{\beta x} t^{\alpha-1} e^{-t} dt}{\Gamma(\alpha)} \\ &= P(\alpha, \beta x). \end{aligned}$$

Here $P(\cdot, \cdot)$ is known as a *regularised gamma function* and is given by

$$P(\alpha, x) = \frac{\gamma(\alpha; x)}{\Gamma(\alpha)}$$

[see Press et al., 2004, 6.2.1] where for $\alpha > 0$ the numerator $\gamma(\alpha; x) = \int_0^x e^{-t} t^{\alpha-1} dt$ is called the *lower incomplete gamma function* and can be expressed as a series

$$\gamma(\alpha; x) = e^{-x} x^\alpha \sum_{n=0}^{\infty} \frac{\Gamma(\alpha)}{\Gamma(\alpha + 1 + n)} x^n \tag{G.1}$$

as shown in Press et al. [2004, 6.2.5]. As noted in Press et al. [2004], $\Gamma(\alpha + 1 + n)$ does not need to be computed for each n. Instead we can use the recurrence relation $\Gamma(z + 1) = z\Gamma(z)$ for $z \in \mathbb{C}$. Thus we

may write (G.1) as

$$\gamma(\alpha;x) = e^{-x}x^{\alpha}\left[\frac{1}{\alpha} + \frac{1}{\alpha(\alpha+1)}x + \frac{1}{\alpha(\alpha+1)(\alpha+2)}x^2 + \cdots + \frac{1}{\alpha(\alpha+1)\cdots(\alpha+n)}x^n + \cdots\right]. \quad \text{(G.2)}$$

Now (G.2) converges rapidly for x less than about $\alpha+1$ [Press et al., 2004, §6.2]. When x is greater than $\alpha+1$, we will rather use a rapidly converging continued fraction instead of the series given in (G.2) [Press et al., 2004, §6.2]. Continued fractions often converge much faster than power series expansions [Press et al., 2004, §5.2]. We first note that the complement of $P(\cdot,\cdot)$, also known as a regularised gamma function, is given by

$$\begin{aligned} Q(\alpha,x) &:= 1 - P(\alpha,x) \\ &= \frac{\Gamma(\alpha;x)}{\Gamma(\alpha)} \end{aligned}$$

as found in Press et al. [2004, 6.2.3]. Here $\Gamma(\alpha;x) = \int_x^{\infty} e^{-t}t^{\alpha-1}dt$ with $\alpha > 0$ is known as the *upper incomplete gamma function* and can be written as the continued fraction[1] given below

$$\Gamma(\alpha;x) = x^{\alpha}e^{-x}\frac{1}{x+}\frac{1-\alpha}{1+}\frac{1}{x+}\frac{2-\alpha}{1+}\frac{2}{x+}\frac{3-\alpha}{1+}\cdots \quad \text{(G.3)}$$

To evaluate this continued fraction we will be using the *modified Lentz method* [Lentz, 1976] as shown in Press et al. [2004, §5.2]. Furthermore, instead of using (G.3), we will be implementing the continued fraction given by

$$\Gamma(\alpha;x) = e^{-x}x^{\alpha}\frac{1}{x+1-\alpha-}\frac{1-\alpha}{x+3-\alpha-}\frac{2(2-\alpha)}{x+5-\alpha-}\frac{3(3-\alpha)}{x+7-\alpha-}\frac{4(4-\alpha)}{x+7-\alpha-}\cdots \quad \text{(G.4)}$$

which is the even part of (G.3) and converges twice as fast as (G.3) ([Press et al., 2004, §5.2]).

According to Press et al. [2004, §6.2], (G.2) and (G.4) each requires at most $k\sqrt{\alpha}$ terms to converge for some k, and this many only when x is close to $\alpha+1$, where the incomplete gamma functions are varying most rapidly. Thus when x is far away from $\alpha+1$, we have very fast convergence and when x is close to $\alpha+1$ there is a bound on the number of terms required for convergence. We will use (G.2) when x is less than $\alpha+1$, that is, compute $P(\alpha,x)$ for $x < \alpha+1$. When x is greater or equal to $\alpha+1$ we will use (G.4), that is, compute $1-Q(\alpha,x)$ for $x \geq \alpha+1$.

[1] A continued fraction has the form

$$f(x) = b_0 + \cfrac{a_1}{b_1 + \cfrac{a_2}{b_2 + \cfrac{a_3}{b_3 + \cfrac{a_4}{b_4 + \cfrac{a_5}{b_5 + \ldots}}}}}$$

and is denoted

$$f(x) = b_0 + \frac{a_1}{b_1+}\frac{a_2}{b_2+}\frac{a_3}{b_3+}\frac{a_4}{b_4+}\frac{a_5}{b_5+}\cdots$$

[Press et al., 2004, §5.2].

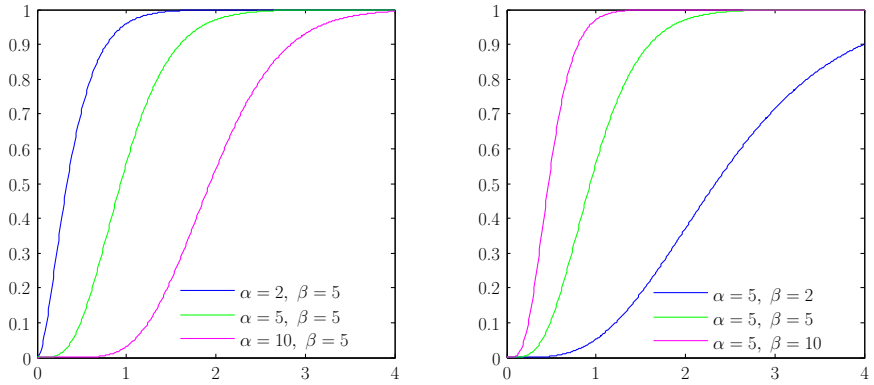

Figure G.1: *Gamma cumulative distribution functions for various values of α and β where one parameter is fixed and the other is varied. The graphs shown in green is the same in both figures with parameters used in common. These figures corresponds to the density functions shown in Figure 8.1.*

G.2 The Beta Distribution

The beta distribution has real parameters α, $\beta > 0$ which are known as the *shape parameters*. We will denote a random variable X that follows a beta distribution with parameters α, β as $X \sim \text{Beta}(\alpha, \beta)$. The beta probability density function and cumulative distribution function can be written in terms of the incomplete beta function.

Definition G.2.1 Incomplete Beta Function
The incomplete beta function *is defined by*

$$B(\alpha, \beta; x) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}dt$$

for $\alpha, \beta > 0$, $\alpha, \beta \in \mathbb{R}$ *and* $0 \leq x \leq 1$. *The* beta function*, which is also known as the* Euler integral of the first kind*, can be written in terms of the incomplete beta function as follows*

$$B(\alpha, \beta) := B(\alpha, \beta; 1).$$

From the definition of the beta function, it is clear that the following defines a probability density function:

Definition G.2.2 Beta Probability Density Function
The beta probability density function *with parameters* α, $\beta > 0$ *is defined as*

$$f_B(x) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)}$$

for $0 \leq x \leq 1$.

Observe that the distribution is symmetric when α and β are equal and when $\alpha = \beta = 1$ we have the density of the uniform distribution.

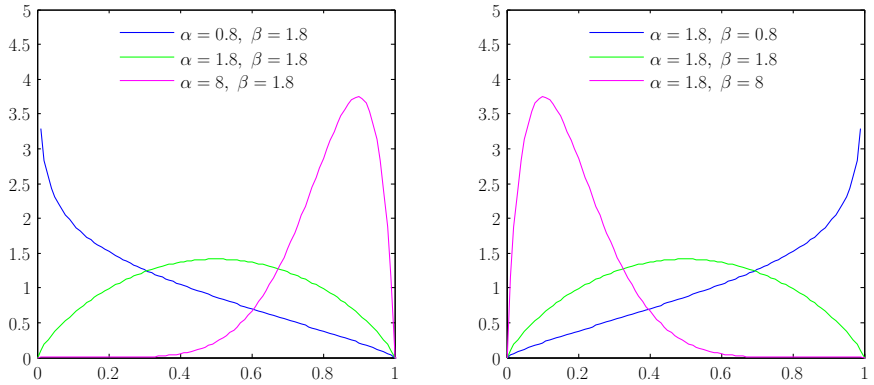

Figure G.2: *Beta probability density functions for various values of α and β where one parameter is fixed and the other is varied. The graphs shown in green is the same in both figures with parameters used in common.*

Definition G.2.3 Cumulative Beta Distribution Function
The cumulative beta distribution function *is given by*

$$F_B(x) = \frac{B(\alpha, \beta; x)}{B(\alpha, \beta)}$$

where $\alpha, \beta > 0$ *and* $0 \leq x \leq 1$.

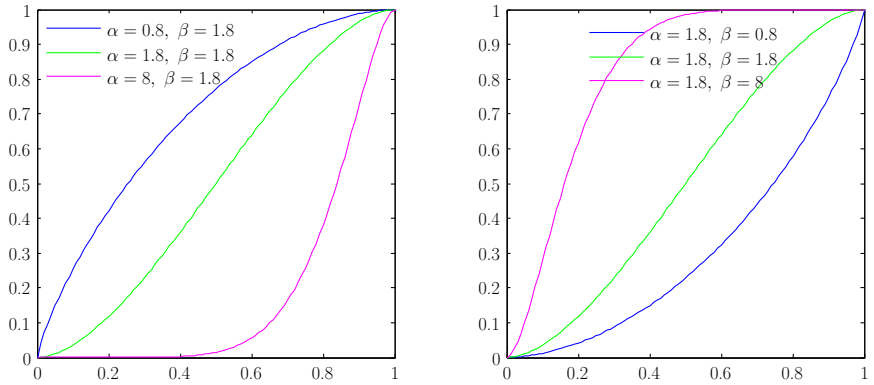

Figure G.3: *Beta cumulative distribution functions for various values of α and β where one parameter is fixed and the other is varied. The graphs shown in green is the same in both figures with parameters used in common. These figures correspond to the probability density functions shown in Figure G.2.*

In our implementation of the beta function we have made use of the expression

$$B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x + y)}$$

as given by National Institute of Standards and Technology [2010, 5.12.1] where $\Gamma(\cdot)$ is the gamma function. We will determine the value of the gamma function using Press et al. [2004, §6.1]. This is an implementation

of the approximation derived in Lanczos [1964]. Furthermore, as noted and implemented in Press et al. [2004, §6.1], we will calculate the log of the gamma function instead of the gamma function itself since the latter will overflow many computers' floating-point representation at quite modest values of z.

When implementing the incomplete beta function we will make use of its continued fraction representation [Press et al., 2004, §6.4]

$$B(\alpha,\beta;x)=\frac{x^{\alpha}(1-x)^{\beta}}{\alpha}\left(\frac{1}{1+}\frac{d_1}{1+}\frac{d_2}{1+}\cdots\right)$$

where for $m=0,\ 1,\ \ldots$

$$d_{2m+1}=-\frac{(\alpha+m)(\alpha+\beta+m)x}{(\alpha+2m)\,(\alpha+2m+1)}\text{ and }d_{2m}=\frac{m(\beta-m)x}{(\alpha+2m-1)(\alpha+2m)}.$$

As noted in Press et al. [2004, §6.4], this continued fraction converges rapidly for $x\leq\frac{\alpha+1}{\alpha+\beta+2}$ [2]. However for $x>\frac{\alpha+1}{\alpha+\beta+2}$ we can use the symmetry relation as shown by Press et al. [2004, §6.4]

$$\frac{B(\alpha,\beta;x)}{B(\alpha,\beta)}=1-\frac{B(\beta,\alpha;1-x)}{B(\beta,\alpha)}.$$

As in §G.1 we will evaluate the continued fraction for incomplete beta function using the *modified Lentz method* as it appears in Press et al. [2004, §5.2].

G.3 The Inverse Gaussian Distribution

Similarly to an inverse Gaussian process (see §9.1), an inverse Gaussian random variable τ with parameters $\eta,\gamma\in\mathbb{R}$ follows the hitting time distribution of a Brownian motion with drift

$$\tau=\inf_{t>0}\{\gamma t+W_t=\eta\}$$

where $\eta>0$, $\gamma>0$ and $W=\{W_t\}_{t\geq0}$ is standard Brownian motion. We will denote an inverse Gaussian random variable with parameters η and γ by $\tau\sim\mathrm{IG}(\eta,\gamma)$.

The probability density and cumulative distribution functions of an inverse Gaussian random variable is used to find the value of a digital option in the Black-Scholes economy. The derivation of this probability density and cumulative distribution functions can be found in West [2011, Chapter 5] which are lecture notes expanding Wystup [2002]. Thus the density function of an inverse Gaussian random variable τ with parameters $\eta>0$ and $\gamma>0$ is given by

$$f_\tau(t)=\frac{\eta}{\sqrt{2\pi t^3}}\exp\left[-\frac{1}{2}\left(\frac{\eta-\gamma t}{\sqrt{t}}\right)^2\right].\tag{G.5}$$

This parameterisation was given in Barndorff-Nielsen [1998][3,4].

[2] Actually Press et al. [2004, §6.4] specifies that $x<\frac{\alpha+1}{\alpha+\beta+2}$, but then the case where $x=\frac{\alpha+1}{\alpha+\beta+2}$ is not dealt with in the symmetry argument which follows. This case certainly arises eg. when $\alpha=\beta$ and $x=0.5$.

[3] We have used the symbol η instead of δ as in Barndorff-Nielsen [1998], because δ is used in §9.2.

[4] Using another parameterisation, the density of an inverse Gaussian variable can be written as

$$f_{\mathrm{IG}}(\lambda,\mu;t)=\sqrt{\frac{\lambda}{2\pi t^3}}\exp\left[-\frac{\lambda}{2\mu^2 t}(t-\mu)^2\right]\mathbb{1}_{\{t>0\}}\tag{G.6}$$

where $\mu>0$ is called the *mean* and $\lambda>0$ is called the *shape parameter* with $\mu,\lambda\in\mathbb{R}$. The density given in (G.5) can be obtained from (G.6) by substituting $\mu=\frac{\eta}{\gamma}$ and $\lambda=\eta^2$.

By using the density function in (G.5) and letting

$$e_{\pm}(t) = \frac{\mp\eta - \gamma t}{\sqrt{t}}$$

we find the cumulative inverse Gaussian distribution

$$F_{\tau}(t) = e^{2\eta\gamma} F_{\mathrm{N}}\left(e_{+}(t)\right) + F_{\mathrm{N}}\left(-e_{-}(t)\right).$$

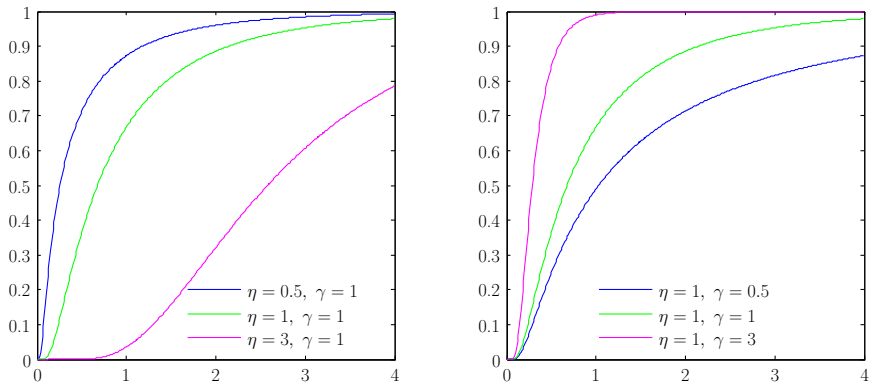

Figure G.4: *Inverse Gaussian cumulative distribution functions for various values of η and γ where one parameter is fixed and the other is varied. The graphs shown in green is the same in both figures with parameters used in common. These figures correspond to the probability density functions shown in Figure 9.1.*

Appendix H

Inverse Transformation Methods

In §7.3 and §7.4 we have investigated the generation of uniform random numbers. In this appendix we will show how to use these uniform random numbers in order to produce random numbers from the various distributions we will consider.

For random draws from distributions other than the normal distribution, we will apply a method called *inverse transform sampling*. Here we will apply the *inverse probability integral transform* or the *inverse transformation method* which generates numbers from any probability distribution given its cumulative distribution function.

There are other methods with which one can generate samples from a given distribution. However, these methods will require more than one uniform random number. For example the Box-Muller method uses two uniform random numbers to generate one normal random number. Thus the Sobol' dimensions are used up sooner, and we run into the problem of the deterioration of the random numbers that much quicker. When the number of random numbers needed is not known in advance, we have an even bigger problem when using Sobol' random numbers. In this case, Joshi [2003] notes that the particular structure of the sequences is destroyed.

In the §H.2 we will make use of a root-finding algorithm called *Brent's method*. This method is a generic method and can be applied to find any random number from a single uniform sample, as long as its cumulative distribution function can be calculated. As we have seen in Appendix G, this is the case for the cumulative distribution functions of the gamma, beta and inverse Gaussian distributions.

H.1 Evaluating the Inverse Transform

Definition H.1.1 Inverse Transform
Let U be a Uniform$(0,1)$ random number. Then the inverse transform *of U is defined by*

$$X = F_X^{-1}(U)$$

where $F_X(\cdot)$ is the cumulative distribution function of X.

If we let u be a draw from Uniform$(0,1)$ and x a draw from the distribution of X, then $x = F_X^{-1}(u)$ and so $F_X(x) = u$. In order to calculate the cumulative inverse of a distribution, we will apply Brent's

method to

$$F_X(x) - u = 0. \tag{H.1}$$

Brent's method finds the root of (H.1). When calculating the cumulative inverse of a distribution we will, with the exception of the normal distribution, always make use of Brent's method.

H.2 Brent's Method

Brent's method [Brent, 1973] is an improvement on an algorithm developed in the 1960's by Dekker [1969]. It is a root-finding algorithm that combines the bisection and secant methods along with inverse quadratic interpolation. The algorithm makes use of the secant method or inverse quadratic interpolation because of faster convergence, but switches to the bisection method when necessary. Brent guarantees that the method will converge as long as the function can be evaluated within the initial interval known to contain a root.

In our implementation of Brent's method we have made use of the algorithm provided in Press et al. [2004, §9.3] and created a templated function in a header file as shown in Joshi [2004, §9.3] (the advantages and disadvantages of this approach are also discussed in Joshi [2004, §9.6]). As discussed in Joshi [2004, §9.3], templatisation allows code that handles many different classes simultaneously, where these classes have to contain certain operations defined with the same syntax. In this case, classes that make use of `Brent` should have the overloaded operator `()` defined:

```
double operator()(double ) const
```

Thus the syntax `f(u)` is well-defined for an instantiated object `f` of a class (below it is indicated by `T`) and a double `u`. Since the type of the object `T` is unknown one cannot precompile the template code in a source file and the function must be in a header file.

```
template<class T>
double Brent(T myFunction, double x1, double x2)
```

The function root is known to lie in the interval [`x1`, `x2`]; if it does not, the code will fail. Here the function of which we want to find the root is `myFunction`.

We have also created an improved version `BrentWithGuess` of the above function in which the user provides a guess of the root `guess`; the guess does not have to be particularly good. The code creates an interval in which the root lies; the interval is found by suitable multiplicative scaling near the guess. This method relies on the assumption that `myFunction` is monotone at least in the area of interest — this is certainly the case for most financial applications. Also, x cannot be 0 in this method.

```
template<class T>
double BrentWithGuess(T myFunction, double guess)
```

H.3 Implementation

When computing the cumulative inverse, we will make use of the version of Brent's method that requires an interval containing the root as input. The interval given will be [`DOMAINMINIMUM`, `DOMAINMAXIMUM`],

where DOMAINMINIMUM and DOMAINMAXIMUM are double values representing the practical lower and upper domain bounds respectively of the distribution under consideration. A way of speeding up the calculation of the cumulative inverse is by performing some precalculation.

At the end of this section we will provide a numerical example, Example H.3.1, to illustrate our implementation.

We use an array for the lookup table, called table, containing MAXIMUMINDEX + 1 entries. The element with index 0 is the cumulative inverse of u = 0 and the entry with index MAXIMUMINDEX contains DOMAINMAXIMUM. The remaining entries are the cumulative inverses of u's which have been evenly divided between 0 and MAXIMUMINDEX.

```
void calculateTable()
{
  double u;
```

Instantiate the Distribution class, with parameters parameter1, parameter2, ... and false which indicates that the lookup table is not in use for this object.

```
  Distribution distributionObject(parameter1, parameter2, ..., false);
  for (int i = 0; i < MAXIMUMINDEX; i++)
  {
```

Find MAXIMUMINDEX number of evenly divided u's.

```
    u = i/static_cast<double>(MAXIMUMINDEX);
```

Find the cumulative inverse of u.

```
    table[i] = distributionObject.cumulativeInverse(u);
  }
  table[MAXIMUMINDEX] = DOMAINMAXIMUM;
```

Set the boolean flag which indicates the table has been created.

```
  tableCalculated = true;
}
```

Since we are making use of the templated function Brent as discussed in the previous section, we will require a class that has the () operator as one of its functions [see Joshi, 2004, §9.2]. This function will allow us to write cumulativeInverseFunction(u) for an object cumulativeInverseFunction of the class ErrorInCumInverseGuess shown below and a double u in the function Brent. We create a class and not just a function because the class is able to contain extra information. In this case, the extra information is in the form of the parameters of the distribution under consideration, as well as u for which the cumulative inverse is to be calculated.

```
ErrorInCumInverseGuess::ErrorInCumInverseGuess
                        (double aParameter1, double aParameter2, ..., double aU):
```

```
parameter1(aParameter1), parameter2(aParameter2), ..., u(aU)

{
```

This is the constructor for the class `ErrorInCumInverseGuess` that has private members `parameter1`, `parameter2`, ... and `u`.

```
}

double ErrorInCumInverseGuess::operator()(double x)
{
  Distribution distributionObject(parameter1, parameter2, ..., false);
  return distributionObject.cumulativeFunction(x) - u;
```

Effectively this is the function for which a zero is found using Brent's method.

```
}
```

Note that in the `()` operator function we need to be able to calculate the cumulative distribution function of the given distribution.

Finally, we present the code which calls the function `Brent` to calculate the cumulative inverse for a given `u`.

```
double cumulativeInverse(double u)
{
```

The cumulative distribution is truncated at the `DOMAINMINIMUM` and `DOMAINMAXIMUM`.

```
  if (u == 0) return DOMAINMINIMUM;
  else if (u == 1) return DOMAINMAXIMUM;
  else
  {
    ErrorInCumInverseGuess
        cumulativeInverseFunction(parameter1, parameter2, ..., u);
```

If the lookup table exists, use it.

```
    if (lookup == true)
    {
```

If the lookup table has not been calculated, do so.

```
      if (tableCalculated == false) calculateTable();
```

Find the indices containing entries for the smaller interval.

```
      int low = static_cast<int>(u*MAXIMUMINDEX);
      int high = low + 1;
      return Brent(cumulativeInverseFunction, table[low], table[high]);
    }
```

The lookup table does not exist or is not used, so we use extreme (and inefficient) lower and upper bounds.

```
    else
      return Brent(cumulativeInverseFunction, DOMAINMINIMUM, DOMAINMAXIMUM);
  }
}
```

Example H.3.1.
Consider the beta distribution (a discussion on this distribution can be found in §G.2). Suppose that $\alpha = 2$ and $\beta = 2$. Thus, `parameter1` and `parameter2` are set to 2 when we create the `Distribution` object `distributionObject`. Since the beta distribution has domain $[0, 1]$, we set `DOMAINMINIMUM` to 0 and `DOMAINMAXIMUM` to 1. In our example, we will set `MAXIMUMINDEX` equal to 10, but in our implementation we have set it to 100 or even 1000 depending on the specific distribution.

Consider Table H.1 where we have calculated the cumulative beta inverses for values $u = 0,\ 0.1,\ \ldots,\ 1$.

Index	Cumulative Inverse
0	0.0000
1	0.1958
2	0.2871
3	0.3633
4	0.4329
5	0.5000
6	0.5671
7	0.6367
8	0.8667
9	0.9333
10	1.0000

Table H.1: *The lookup table with cumulative beta inverses calculated for values* 0, 0.1, ..., 1.

This is achieved when the subroutine `calculateTable` is run. Each entry in the table is found by calling `Brent` with the $[0, 1]$ as the interval containing the root. Now suppose we would like to calculate the inverse of $u = 0.43$ after the table has been created. This requires the function `cumulativeInverse` which calculates `low` to be the floor of $0.43 \times 10 = 4.3$, that is 4, and `high` to be $4 + 1 = 5$. Therefore the routine `Brent` is called with $[0.4329, 0.5000]$ as the interval containing the root. □

Appendix I

Application

The application *American Monte Carlo Pricer*, developed during the course of this thesis, has the following fields which are to be completed:

(i) In the *Stock Price Process Details* section

- **Stock Price Process** — choose one of *GBM*, *VG* or *NIG*.
- Enter the values of the parameters corresponding to the selected stock price process. Percentages should be entered in decimal form, e.g. enter 0.3 and not 30 if the input is 30%. If the selected stock price process is *GBM*, the only parameter to be entered is σ, which must be positive. If the selected stock price process is *VG*, enter values for θ, σ and ν. Both σ and ν must be positive. Finally, if the selected stock price process is *NIG*, enter values for α, β and δ. δ must be positive and $-\alpha < \beta < \alpha - 1$.
- **Spot** — enter the current stock price.
- **Risk-Free Rate** — enter the constant continuously compounded risk-free rate. As before, percentages should be entered in decimal form.
- **Dividend Yield** — enter the constant continuously compounded dividend yield. Again, percentages should be entered in decimal form.

(ii) In the *Option Details* section

- **Term** — enter the term of the option in years.
- **Strike** — enter the strike of the option under consideration.
- **Style** — choose either *Call* or *Put*.

(iii) In the *Monte Carlo Details* section

- **Monte Carlo Method** — choose one of *Regression*, *LSM*, *Rasmussen*, *Rasmussen Low Bias*, *Rasmussen Dual*, *Stochastic Mesh Low Bias* or *Stochastic Mesh High Bias*. Note that computation time for the *Stochastic Mesh Low Bias* and *Stochastic Mesh High Bias* is much higher than

the other methods. The computation becomes increasingly expensive as the number of time steps and simulations are increased. If the selected stock price process is *VG*, the *Stochastic Mesh Low Bias* option is removed from the Monte Carlo method list in order to avoid slow computation time. Likewise the selected Monte Carlo method is *Stochastic Mesh Low Bias*, then the *VG* option is removed from the stock price process list.

- **Polynomial** — select the type of basis functions used in the regression type Monte Carlo methods. This option is greyed out when the selected Monte Carlo method is either *Stochastic Mesh Low Bias* or *Stochastic Mesh High Bias*. Choose one of *Laguerre*, *Hermite*, *Chebyshev* or *Rasmussen*. *Rasmussen* indicates the polynomials we mention in §3.3. Note that when selecting the *Hermite* polynomial as basis, erratic results for varying number of time steps and number of polynomials may be obtained.
- **Number of Polynomials** — select the number of basis functions used in the regression type Monte Carlo methods. In this application we have restricted the number of polynomials to 10 in order to avoid long computation times. Again, this option is greyed out when the selected Monte Carlo method is either *Stochastic Mesh Low Bias* or *Stochastic Mesh High Bias*.
- **Number of Time Steps** — select the number of time steps as a multiple of 2.
- **Random Number Generator** — select either the pseudo-random number generator *Mersenne Twister* or the quasi-random number generator *Sobol' with Bridging*.
- **Seed** — enter an integer which is used as the seed if the random number generator selected is *Mersenne Twister*, the initial dispersion technique is applied or the dual method is employed. If this field is not populated, an integer produced by the pseudo-random number generator native to C# is used.
- **Number of Sample Paths** — choose a power of 2 for the number of sample paths. In order to avoid long computation times, several restrictions have been made. If the selected Monte Carlo method is *Regression*, *LSM*, *Rasmussen*, *Rasmussen Dual* or *Stochastic Mesh Low Bias*, the number of simulations range from 2^5 to 2^{12}. If the selected Monte Carlo method is *Rasmussen Low Bias*, the number of simulations range from 2^5 to 2^{15}. Finally, if the selected Monte Carlo method is *Stochastic Mesh High Bias*, the number of simulations range from 2^5 to 2^{10}.

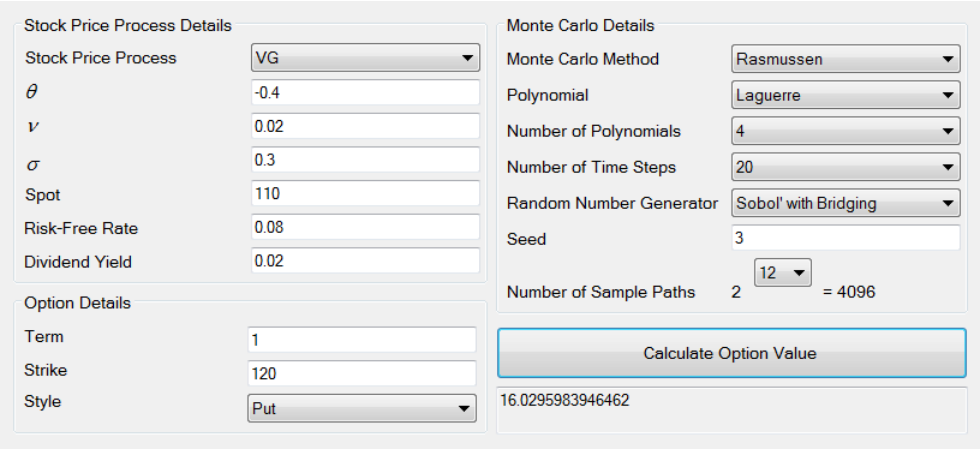

Figure I.1: *A screen shot of the application.*

Bibliography

Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions, With Formulas, Graphs, and Mathematical Tables*. Dover, 1974. 89, 104, 173

Peter John Acklam. An algorithm for computing the inverse normal cumulative distribution function, 2004. URL http://home.online.no/~pjacklam. 80

Peter A. Acworth, Mark Broadie, and Paul Glasserman. A comparison of some Monte Carlo and quasi-Monte Carlo techniques for option pricing. In H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, editors, *Monte Carlo and Quasi-Monte Carlo Methods 1996*, volume 127 of *Lecture Notes in Statistics*, pages 1–18. Springer, July 1996. Proceedings of a Conference at the University of Salzburg, Austria. 81, 82, 157, 158

Fredrik Åkesson and John P. Lehoczky. Path Generation for Quasi-Monte Carlo Simulation of Mortgage-Backed Securities. *Management Science*, 46(9):1171–1187, 2000. 82

Leif Andersen and Mark Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222–1234, 2004. ii, iii, 2, 51, 57, 58

Ari D. Andricopoulos, Martin Widdicks, Peter W. Duck, and David P . Newton. Universal option valuation using quadrature methods. *Journal of Financial Economics*, 67:447–471, 2003. 1, 175

I.A. Antonov and V.M. Saleev. An Economic Method of Computing LP_τ-sequences. *USSR Computational Mathematics Mathematical Physics*, 19(1):252–256, 1979. 151, 157

David Applebaum. *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2004. 64, 65, 66, 71, 72, 75, 92, 100, 105, 165

Athanassios N. Avramidis and Heinrich Matzinger. Convergence of the stochastic mesh estimator for pricing Bermudan options. *Journal of Computational Finance*, 7(4):73–91, 2004. 47

Athanassios N. Avramidis, Pierre L'Ecuyer, and Pierre-Alexandre Tremblay. Efficient Simulation of Gamma and Variance-Gamma Processes. In S. Chick, P.J. Sánchez, D. Ferrin, and D.J. Morrice, editors, *Proceedings of the 2003 Winter Simulation Conference*, pages 319–326, 2004. 82, 99

Vlad Bally and Gilles Pagès. A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6):1003–1049, 2003. 1

Ole E. Barndorff-Nielsen. Exponentially decreasing distributions for the logarithm of particle size. *Proceedings of the Royal Society of London A*, 353(1674):401–419, 1977. 3, 100

Ole E. Barndorff-Nielsen. Normal inverse Gaussian distributions and the modeling of stock returns. Research Report 300, Department of Theoretical Statistics, Aarhus University, 1995. ii, iii, 3, 100

Ole E. Barndorff-Nielsen. Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of Statistics*, 24(1):1–13, 1997. 3, 75, 100, 107, 109

Ole E. Barndorff-Nielsen. Processes of normal inverse Gaussian type. *Finance and Stochastics*, 2(1):41–68, 1998. 3, 75, 100, 105, 186

Jérôme Barraquand and Didier Martineau. Numerical Valuation of High Dimensional Multivariate American Securities. *Journal of Financial and Quantitative Analysis*, 30(3):383–405, 1995. 1

Christian Bender and John Schoenmakers. An iterative method for multiple stopping: convergence and stability. *Advances in Applied Probability*, 38(3):729–749, 2006. 1

Alain Bensoussan. On the theory of option pricing. *Acta Applicandae Mathematicae*, 2(2):139–158, 1984. 5, 8

Jean Bertoin. *Lévy Processes*. Cambridge University Press, 1998. 75

Tomas Björk. *Arbitrage Theory in Continuous Time*. Oxford University Press, second edition, 2004. 7

Lenore Blum, Manuel Blum, and Michael Shub. A simple unpredictable pseudo-random number generator. *SIAM Journal on Computing*, 15(2):364–383, 1986. 80

Salomon Bochner. *Harmonic Analysis and the Theory of Probability*. University of California Press, 1955. 75

George E. P. Box and Mervin E. Muller. A Note on the Generation of Random Normal Deviates. *Annals of Mathematical Statistics*, 29(2):610–611, 1958. 80

Svetlana I. Boyarchenko and Sergei Z. Levendorskiĭ. Generalizations of the Black-Scholes equation for truncated Lévy processes. Working paper, 1999. 75

Paul Bratley and Bennett L. Fox. Algorithm 659 implementing Sobol's quasirandom sequence generator. *ACM Transactions on Mathematical Software*, 14(1):88–100, 1988. 151, 152, 154, 155, 158

Richard P. Brent. *Algorithms for Minimization Without Derivatives*, chapter 3 & 4. Prentice-Hall, Englewood Cliffs, NJ, 1973. 189

Mark Broadie and Paul Glasserman. A stochastic mesh method for pricing high-dimensional American options. PaineWebber Working Papers in Money, Economics and Finance #PW9804, Columbia Business School, New York, 1997a. 2, 12, 37, 46, 47, 139

Mark Broadie and Paul Glasserman. Pricing American-style securities by simulation. *Journal of Economic Dynamics and Control*, 21:1323–1352, 1997b. 1, 2, 37

Mark Broadie and Paul Glasserman. A stochastic mesh method for pricing high-dimensional American options. *Journal of Computational Finance*, 7(4):35–72, 2004. ii, iii, 2, 37, 39, 42, 43, 46, 47, 48

Mark Broadie, Paul Glasserman, and Zachary Ha. Pricing American Options by Simulation Using a Stochastic Mesh with Optimized Weights. In *Probabilistic Constrained Optimization: Methodology and Applications*, pages 32–50. Kluwer, 2000. 37, 39, 136, 139

Russel E. Caflisch and Bradley Moskowitz. Modified Monte Carlo Methods Using Quasi-Random Sequences. In *Lecture Notes in Statistics 106*, pages 1–16. Springer-Verlag, 1995. 3, 78, 82

Russel E. Caflisch, William Morokoff, and Art Owen. Valuation of Mortgage Backed Securities Using Brownian Bridges to Reduce Effective Dimension. *The Journal of Computational Finance*, 1(1):27–46, 1997. 82, 83

Peter Carr and Dilip B. Madan. Option valuation using the fast Fourier transform. *Journal of Computational Finance*, 2:61–73, 1999. 176

Peter Carr, Hélyette Geman, Dilip B. Madan, and Marc Yor. The fine structure of asset returns: An empirical investigation. *Journal of Business*, 75(2):305–332, 2002. 3, 75

Peter Carr, Hélyette Geman, Dilip B. Madan, and Marc Yor. Stochastic Volatility for Lévy Processes. *Mathematical Finance*, 13(3):345–382, 2003. 74

Jacques F. Carrière. Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics*, 19(1):19–30, 1996. 2, 13

George Casella and Roger L. Berger. *Statistical Inference*. Wadsworth and Brooks/Cole, 1990. 168, 169

Emmanuelle Clément, Damien Lamberton, and Philip Protter. An analysis of a least squares regression method for American option pricing. *Finance and Stochastics*, 6:449–471, 2002. 2, 13, 14, 16, 17, 19

Donald L. Cohn. *Measure Theory*. Birkhäuser, 1980. 165

Rama Cont and Peter Tankov. *Financial Modelling with Jump Processes*. Chapman & Hall/CRC, 2004a. 64, 65, 66, 68, 69, 70, 71, 72, 73, 74, 76, 77, 107, 116, 117

Rama Cont and Peter Tankov. Non-parametric calibration of jump-diffusion option pricing models. *Journal of Computational Finance*, 7(3):1–49, 2004b. ii, iii, 3, 125, 126

Theodorus J. Dekker. Finding a zero by means of successive linear interpolation. In Bruno Dejon and Peter Henrici, editors, *Constructive Aspects of the Fundamental Theorem of Algebra*, pages 37–48, New York, 1969. Wiley-Interscience. 189

Darrell Duffie. *Dynamic Asset Pricing Theory*. Princeton University Press, third edition, 2001. 8

Paul DuPuis and Hui Wang. On the Convergence from Discrete to Continuous Time in an Optimal Stopping Problem. *The Annals of Applied Probability*, 15(2):1339–1366, 2005. 6

Ernst Eberlein. Jump-type Lévy processes. In T.G. Andersen, R.A. Davis, J.-P. Kreiß, and T.Mikosch, editors, *Financial Time Series*, pages 439–455. Springer Verlag, 2009. 117

Ernst Eberlein and Ulrich Keller. Hyperbolic distributions in finance. *Bernoulli*, 1(3):281–299, 1995. 3, 100

Ernst Eberlein, Ulrich Keller, and Karsten Prause. New Insights into Smile, Mispricing, and Value at Risk: The Hyperbolic Model. *Journal of Business*, 71(3):371–405, 1998. 3, 100

Robert J. Elliott and Ekkehard P. Kopp. *Mathematics of Financial Markets*. Springer Finance. Springer, second edition, 2005. 8

Paul Embrechts, Rüdiger Frey, and Hansjörg Furrer. Stochastic Processes in Insurance and Finance. In D.N. Shanbhag and C.R. Rao, editors, *Handbook of Statistics "Stochastic Processes: Theory and Methods"*, volume 19, chapter 12, pages 365–412. Elsevier Science, 2001. 89

Anders Eriksson, Lars Forsberg, and Eric Ghysels. Approximating the Probability Distribution of Functions of Random Variables: A New Approach. Technical Report s-21, CIRANO, 2004. URL http://www.cirano.qc.ca/pdf/publication/2004s-21.pdf. 130

Alison Etheridge. *A Course in Financial Calculus*. Cambridge University Press, 2002. 79

Fang Fang and Cornelis W. Oosterlee. A Novel Pricing Method for European Options Based on Fourier-Cosine Series Expansions. *SIAM Journal on Scientific Computing*, 31:826–848, 2008. 176, 177, 178

Fang Fang and Cornelis W. Oosterlee. Pricing early-exercise and discrete barrier options by Fourier-Cosine Series Expansions. *Numerische Mathematik*, 114(1):27–62, 2009. URL http://ta.twi.tudelft.nl/mf/users/oosterle/oosterlee/bermCOS.pdf. 1, 135

Niels Ferguson and Bruce Schneier. *Practical Cryptography*. John Wiley & Sons, 2003. 80

Bennett L. Fox. Algorithm 647: implementation and relative efficiency of quasirandom sequence generators. *ACM Transactions on Mathematical Software*, 12(4):362–376, 1986. 158

Michael C. Fu. Variance-Gamma and Monte Carlo. In M.C. Fu, R.A. Jarrow, J.-Y.J. Yen, and R.J. Elliott, editors, *Advances in Mathematical Finance*, number 28 in Applied and Numerical Harmonic Analysis, pages 21–35. Birkhäuser, 2007. URL http://www.rhsmith.umd.edu/faculty/mfu/fu_files/Fu07.pdf. 99

Michael C. Fu, Scott B. Laprise, Dilip B. Madan, Yi Su, and Rongwen Wu. Pricing American Options: A Comparison of Monte Carlo Simulation Approaches. *Journal of Computational Finance*, 4(3):39–88, 2001. 1, 5, 6

Hélyette Geman. Pure Jump Lévy Processes for Asset Price Modelling. *Journal of Banking and Finance*, 26(7):1297–1316, 2002. 74

Hélyette Geman, Dilip B. Madan, and Marc Yor. Time changes for Lévy processes. *Mathematical Finance*, 11(1):79–96, 2001. 3, 88

James E. Gentle. *Random Number Generation and Monte Carlo Methods*. Springer, 2003. 149

Paul Glasserman. *Monte Carlo Methods in Financial Engineering*. Springer, 2004. 1, 2, 3, 6, 8, 10, 11, 13, 14, 17, 18, 24, 26, 37, 38, 39, 45, 46, 47, 51, 57, 58, 59, 82, 83, 151, 152, 153, 154, 155, 156, 157, 158

Bronius Grigelionis. Processes of Meixner type. *Lithuanian Mathematics Journal*, 39(1):33–41, 1999. 75

Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes*. Oxford University Press, third edition, 2001. 168

Espen Gaarder Haug. *Derivatives Models on Models*. Wiley, 2007. 82

Martin B. Haugh and Leonid Kogan. Pricing American Options: A Duality Approach. *Operations Research*, 52(2):258–270, March 2004. ii, iii, 2, 51

Steven L. Heston. A closed form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, 6(2):327–343, 1993. 178

Peter Jäckel. *Monte Carlo Methods in Finance*. Wiley, 2002. 3, 24, 82, 83, 84, 87, 151, 152, 153, 154, 156, 157, 158, 159, 160, 162, 163

Stephen Joe and Frances Y. Kuo. Remark on Algorithm 659: Implementing Sobol's quasirandom sequence genrator. *ACM Transactions on Mathematical Software*, 29(1):49–57, 2003. 3, 82, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, 163

Stephen Joe and Frances Y. Kuo. Constructing Sobol' Sequences with Better Two-Dimensional Projections. *SIAM Journal on Scientific Computing*, 30(5):2635–2654, 2008. 153, 156, 158, 159, 160, 162, 163, 164

Norman L. Johnson, Samuel Kotz, and N. Balakrishnan. *Continuous Univariate Distributions: Volume 1*. Wiley Series in Probability and Mathematical Statistics, second edition, 1994. 88, 100

Christian Jonen. An efficient implementation of a least squares Monte Carlo method for valuing American-style options. *International Journal of Computer Mathematics*, 86(6):1024–1039, 2009. 17

Mark S. Joshi. *The Concepts and Practice of Mathematical Finance*. Mathematics, Finance and Risk. Cambridge University Press, 2003. 175, 188

Mark S. Joshi. *C++ Design Patterns and Derivatives Pricing*. Cambridge University Press, 2004. 189, 190

Mark S. Joshi. More Mathematical Finance. Forthcoming, 2011. 83, 163

Olav Kallenberg. *Foundations of modern probability*. Springer-Verlag, 2001. 76

Ioannis Karatzas. On the pricing of American Options. *Applied Mathematics and Optimization*, 17:37–60, 1988. 5, 8

Ioannis Karatzas and Steven E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer-Verlag, 1988. 5

J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks*, volume 4, pages 1942–1948, 1995. 133

Anastasia Kolodko and John Schoenmakers. Iterative construction of the optimal Bermudan stopping time. *Finance and Stochastics*, 10:27–49, 2006. 1

Ismo Koponen. Analytic approach to the problem of convergence of truncated Lévy flights towards the Gaussian stochastic process. *Physical Review E*, 52(1):1197–1199, 1995. 75

Ralf Korn, Elke Korn, and Gerald Kroissandt. *Monte Carlo methods and models in finance and insurance*. Chapman & Hall, 2010. 100

S. Kou. On pricing of discrete barrier options, 2001. Columbia University Preprint. 65

Aleksandr S. Kronrod. *Nodes and weights of quadrature formulas. Sixteen-place tables*. Consultants Bureau, New York, 1965. Authorised translation from the Russian. 132

Andreas E. Kyprianou. *Introductory lectures on fluctuations of Lévy processes with applications*. Springer, 2006. 91

Andreas E. Kyprianou. An Introduction to the Theory of Lévy Processes, August 2007. URL http://www.maths.bath.ac.uk/~ak257/Levy-sonderborg.pdf. Sønderborg Denmark Lecture notes. 64, 66, 100, 103

Damien Lamberton and Bernard Lapeyre. *Introduction to Stochastic Calculus Applied to Finance*. Chapman & Hall, second edition, 2008. 8

Cornelius Lanczos. A Precision Approximation of the Gamma Function. *SIAM Journal on Numerical Analysis: Series B*, 1:86–96, 1964. 186

Christiane Lemieux, Mikolaj Cieslak, and Kristopher Luttmer. RandQMC user's guide: a package for randomized quasi-Monte Carlo methods in C, 2002. Technical Report 2002-712-15, Department of Computer Science, University of Calgary, Calgary, Canada. 155

William J. Lentz. Generating Bessel Functions in Mie Scattering Calculations using Continued Fractions. *Applied Optics*, 15(3):668–671, 1976. 183

Alan L. Lewis. A Simple Option Formula for General Jump-Diffusion and other Exponential Levy Processes. Working paper, 2001. URL http://www.optioncity.net/pubs/ExpLevy.pdf. 176

Eduardo J. A. Lima and Benjamin M. Tabak. Building Confidence Intervals with Block Bootstraps for the Variance Ratio Test of Predictability. Working Paper Series 151, 2007. URL http://www.bcb.gov.br/pec/wps/ingl/wps151.pdf. 125

Guangwu Liu and L. Jeff Hong. Revisit of stochastic mesh method for pricing american options. *Operations Research Letters*, 37(6):411–414, 2009. 37

Francis A. Longstaff and Eduardo S. Schwartz. Valuing American options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113–147, 2001. ii, iii, 1, 2, 13, 14, 17, 18, 22, 23, 148

R. Lord, F. Fang, F. Bervoets, and C. W. Oosterlee. A Fast And Accurate FFT-Based Method For Pricing Early-Exercise Options Under Lévy Processes. *SIAM Journal on Scientific Computing*, 30(4): 1678–1705, 2008. 1, 176

Roger Lord and Christian Kahl. Complex Logarithms In Heston-Like Models. *Mathematical Finance*, 20 (4):671–694, 2010. 91, 95

Dilip B. Madan. Purely discontinuous asset price processes. In Elyès Jouini, Jaksa Cvitanić, and Marek Musiela, editors, *Option Pricing, Interest Rates and Risk Management*. Cambridge University Press, 2001a. URL http://www.imub.ub.es/events/sssf/pdappjc.pdf. 92

Dilip B. Madan. Financial modeling with discontinuous price processes. In Ole E. Barndorff-Nielsen, Thomas Mikosch, and Sidney I. Resnick, editors, *Lévy Processes — Theory and Application*. Birkhäuser, Boston, 2001b. 74

Dilip B. Madan and Frank Milne. Option pricing with VG martingale components. *Mathematical Finance*, 1(4):39–55, 1991. 3, 88, 175

Dilip B. Madan and Eugene Seneta. Chebyshev Polynomial Approximations and Characteristic Function Estimation. *Journal of the Royal Statistical Society B*, 49(2):163–169, 1987. 3, 88

Dilip B. Madan and Eugene Seneta. The Variance Gamma (V.G.) model for share market returns. *Journal of Business*, 63(4):511–524, 1990. 3, 75, 88

Dilip B. Madan and Marc Yor. CGMY and Meixner Subordinators are Absolutely Continuous with respect to One Sided Stable Subordinators. Unpublished, 2005. URL http://arxiv.org/PS_cache/math/pdf/0601/0601173v2.pdf. 75

Dilip B. Madan, Peter P. Carr, and Eric C. Chang. The Variance Gamma Process and Option Pricing. *European Finance Review*, 2:79–105, 1998. URL http://www.math.nyu.edu/research/carrp/papers/pdf/VGEFRpub.pdf. ii, iii, 3, 75, 88, 92, 93, 95, 175

Makoto Matsumoto and Takuji Nishimura. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation*, 8(1): 3–30, 1998. URL http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/ARTICLES/mt.pdf. 3, 78, 80

Henry P. McKean. Appendix: A Free Boundary Problem for the Heat Equation Arising from a Problem in Mathematical Economics. *Industrial Management Review*, 6(2):32–39, 1965. 5

Volker Mehrmann. *Frontiers in Numerical Analysis*, pages 303–349. Universitext. Springer-Verlag, 2003. Papers presented at the 10th LMS-EPSRC Numerical Analysis Summer School, held at the University of Durham. 148

John R. Michael, William R. Schucany, and Roy W. Haas. Generating random variates using transformations with multiple roots. *The American Statistician*, 30(2):88–89, 1976. 111, 112

Manuel Moreno and Javier F. Navas. On the Robustness of Least-Squares Monte Carlo (LSM) for Pricing American Derivatives. *Review of Derivatives Research*, 6(2):107–128, 2004. 14, 18, 23

Boris Moro. The full Monte. *Risk*, 8(2):57–58, 1995. 80

William J. Morokoff and Russel E. Caflisch. Quasi-random sequences and their discrepancies. *SIAM Journal on Scientific Computing*, 15(6):1251–1279, 1994. 156

V.A. Morozov. On the solution of functional equations by the method of regularization. *Soviet Mathematics Doklady*, 7:414–417, 1966. 126

Bradley Moskowitz and Russel E. Caflisch. Smoothness and Dimension Reduction in Quasi-Monte Carlo Methods. *Mathematical and Computer Modelling*, 23(8):37–54, 1996. 3, 78, 82

National Institute of Standards and Technology. Digital library of mathematical functions, 2010. URL http://dlmf.nist.gov/. 173, 174, 185

J. A. Nelder and R. Mead. A simplex method for function minimization. *The Computer Journal*, 7: 308–313, 1965. 133

David Nualart and Wim Schoutens. Chaotic and predictable representations for Lévy processes. *Stochastic Processes and their Applications*, 90(1):109–122, 2000. 115

Anargyros Papageorgiou. The Brownian Bridge Does Not Offer a Consistent Advantage in Quasi-Monte Carlo Integration. *Journal of Complexity*, 18(1):171–186, 2002. 83

Antonis Papapantoleon. An Introduction to Lévy Processes with Applications in Finance, 2008. URL http://www.fam.tuwien.ac.at/~papapan/papers/introduction.pdf. Lecture notes, TU Vienna. 71, 72, 74, 108

Spassimir H. Paskov and Joseph F. Traub. Faster valuation of financial derivatives. *Journal of Portfolio Management*, 22(1):113–120, 1995. 82

Karl Pearson. Contributions to the mathematical theory of evolution. *Proceedings of the Royal Society of London*, 54:329–333, 1893. 88

Karl Pearson. Contributions to the mathematical theory of evolution, II: Skew variation in homogeneous material. *Philosophical Transactions of the Royal Society of London A*, 186:343–414, 1895. 88

Roger Penrose. A generalized inverse for matrices. *Mathematical Proceedings of the Cambridge Philosophical Society*, 51:406–413, 1955. 149

Goran Peskir and Albert Shiryaev. *Optimal Stopping and Free-Boundary Problems*. Birkhäuser, 2006. 8

Jim Pitman and Marc Yor. Infinitely divisible laws associated with hyperbolic functions. *Canadian Journal of Mathematics*, 55(2):292–330, 2003. 75

William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge University Press, 2004. 52, 129, 148, 155, 157, 173, 182, 183, 185, 186, 189

Nicki Søndergaard Rasmussen. Improving the Least-Squares Monte-Carlo Approach, 2002. URL http://www.hha.dk/afl/wp/fin/glfin/D02_18.PDF. Finance Working Papers. 2, 24, 26, 28, 30, 31

Nicki Søndergaard Rasmussen. Control Variates for Monte Carlo Valuation of American Options. *Journal of Computational Finance*, 9(1):84—102, 2005. ii, iii, 2, 17, 24, 25, 26, 28, 30, 31, 32, 33, 55

Claudia Ribeiro and Nick J. Webber. A Monte Carlo Method for the Normal Inverse Gaussian Option Valuation Model using an Inverse Gaussian Bridge. Working paper, February 2003. 110, 111, 112

Claudia Ribeiro and Nick J. Webber. Valuing Path Dependent Options in the Variance-Gamma Model by Monte Carlo with a Gamma Bridge. *Journal Of Computational Finance*, 7(2):81–100, 2004. 98

L.C.G. Rogers. Monte Carlo valuation of American options. *Mathematical Finance*, 12:271–286, 2002. ii, iii, 2, 51

T. H. Rydberg. The normal inverse Gaussian Lévy process: simulations and approximation. Research Report 344, Department of Theoretical Statistics, Aarhus University, 1996a. 3, 100

T. H. Rydberg. Generalized hyperbolic diffusions with applications towards finance. Research Report 342, Department of Theoretical Statistics, Aarhus University, 1996b. 3, 100

T. H. Rydberg. The normal inverse Gaussian Lévy process: simulations and approximation. *Communications in Statistics: Stochastic Models*, 13(4):887–910, 1997. 3, 100, 130

Ken-Iti Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999. 64, 65, 66, 67, 71, 72, 74, 75, 77, 108, 165

Wim Schoutens. *Stochastic Processes and Orthogonal Polynomials*, volume 146 of *Lecture Notes in Statistics*. Springer, 2000. 75

Wim Schoutens. *Lévy Processes in Finance*. Wiley Series in Probability and Statistics. Wiley, 2003. 71, 76, 89, 92, 115, 117, 127

Wim Schoutens and Jozef L. Teugels. Lévy processes, polynomials and martingales. *Communications in Statistics: Stochastic Models*, 14(1 & 2):335–349, 1998. 75

V. Seshadri. *The Inverse Gaussian Distribution*. Oxford science publications. Oxford University Press, 1993. 111

Albert N. Shiryaev. *Probability*. Springer, 1996. 58

Thomas S. Shores. *Applied Linear Algebra and Matrix Analysis*. Springer, 2007. 148

Steven E. Shreve. *Stochastic Calculus for Finance II: continuous time models*. Springer, 2004. 6, 78, 79

Ilya M. Sobol'. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, 7(4):86–112, 1967. 3, 78, 151

Ilya M. Sobol'. Uniformly distributed sequences with additional uniform properties. *USSR Computational Mathematics and Mathematical Physics*, 16(5):236–242, 1976. 154

Aris Spanos. *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data*. Cambridge University Press, 1999. 169

Domingo Tavella. *Quantitative Methods in Derivatives Pricing: An Introduction to Computational Finance*. Wiley Finance, 2002. 82

John N. Tsitsiklis and Benjamin Van Roy. Optimal Stopping of Markov Processes: Hilbert Space Theory, Approximation Algorithms, and an Application to Pricing High-Dimensional Financial Derivatives. *IEEE Transactions on Automatic Control*, 44(10):1840–1851, 1999. URL http://www.mit.edu/~jnt/Papers/J074-99-bvr-stop.pdf. 2, 13

John N. Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks*, 12(4):694–703, 2001. ii, iii, 2, 13, 17

M.C.K. Tweedie. Functions of a statistical variate with given means, with special reference to Laplacian distributions. *Proceedings of the Cambridge Philosophical Society*, 43:41–49, 1947. 100

P. L. J. van Moerbeke. On Optimal Stopping and Free Boundary Problems. *Archive for Rational Mechanics and Analysis*, 60:101–148, 1976. 5

Srinivasa R. S. Varadhan. *Probability Theory*, volume 7 of *Courant Lecture Notes in Mathematics*. American Mathematical Society, 2001. 66, 165, 167, 168

Srinivasa R. S. Varadhan. *Stochastic Processes*, volume 16 of *Courant Lecture Notes in Mathematics*. American Mathematical Society, 2007. 7

Nevena Šelić. The LIBOR market model. Master's thesis, University of the Witwatersrand, 2006. URL http://www.cam.wits.ac.za/mfinance/papers/NevenaMSc.pdf. 151

Xiaoqun Wang and Kai-Tai Fang. The effective dimension and quasi-Monte Carlo integration. *Journal of Complexity*, 19(2):101–124, 2003. 158

Charles E. Weatherburn. *A first course in mathematical statistics*. Cambridge University Press, 1946. 88

Graeme West. Exotic Equity Options: Lecture notes, 2011. URL http://www.finmod.co.za/exotics.pdf. 186

David Williams. *Probability with Martingales*. Cambridge University Press, 1991. 167

Uwe Wystup. Rebates, 2002. URL http://www.mathfinance.de. 186

Tomasz Zastawniak. Overview of Monte Carlo methods for American options, February 2009. AIMS Summer School in Mathematical Finance. 14, 16, 17