Efficient pricing algorithms for exotic derivatives

ISBN 978 90 5170 909 4

This book is no. 437 of the Tinbergen Institute Research Series, established through cooperation between Thela Thesis and the Tinbergen Institute. A list of books which already appeared in the series can be found in the back.

Efficient pricing algorithms for exotic derivatives

Efficiënte waarderingsalgoritmen voor exotische derivaten

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de
rector magnificus

Prof. dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties

De openbare verdediging zal plaatsvinden op

vrijdag 21 november 2008 om 16.00 uur

door

ROGER LORD
geboren te Geleen

Promotiecommissie

Promotor:	Prof.dr. A.A.J. Pelsser
Overige leden:	Prof.dr. D.J.C. van Dijk Prof.dr. A.C.F. Vorst Prof.dr. P.P. Carr

Preface

When I started thinking about pursuing a PhD in mathematical finance, it was soon clear to me that I wanted to combine academia with the financial industry. The combination of Erasmus University Rotterdam, where Antoon Pelsser was my supervisor, and the Derivatives Research & Validation Team at Rabobank International enabled me to do this. I would like to take this opportunity to thank Antoon for his guidance.

At Rabobank International I am particulary indebted to Sacha van Weeren and Maarten Rosenberg for creating my position. I would like to thank Sacha for the stimulating environment his team provided, the pointers to many good articles, and especially for the fact that, while annoying at the time, he made me explain every step in every derivation I made.

I will now chronologically walk through the various chapters of this thesis, and thank everyone where thanks are due. Without a doubt I will in the process forget some people – they are thanked too. Initially the focus of my thesis was meant to be on exotic interest rate derivatives, but my interests drifted when Jeroen van der Hoek, a former colleague from Cardano Risk Management, pointed me towards Curran's approximation for Asian options. This led to the foundation of Chapter 7, for which I thank Jeroen.

The next chapter was inspired by a presentation Antoon Pelsser gave at the Unfinished Manuscripts seminar in Rotterdam. This eventually led to our joint publication in Chapter 8. I would like to thank Antoon for his ideas for this paper and the opportunity he gave me to present out work at the Quantitative Methods in Finance conference in Sydney.

Somewhere along the line I came into contact with Christian Kahl, with whom it was an honour and great pleasure to work. I would like to take this opportunity to thank Christian for the many ideas we generated in Amsterdam and Wuppertal, which resulted in Chapters 3 and 4, and also for the invaluable feedback he has given on Chapter 6.

The final two chapters, chronologically that is, Chapters 5 and 6, would certainly have been different had I not been allowed to supervise students, both at the university and at the bank. Thanks go out to all my students. Three of them I will mention by name. Firstly ManWo Ng, whose Master's thesis was an inspiring first look at topics covered in Chapters 3 and 4. Secondly Frank Bervoets, whose work inspired the CONV method in Chapter 5. And finally, Remmert Koekkoek, without whom the full truncation method from Chapter 6 would never have existed. I am also grateful to my other co-authors, Kees Oosterlee and Fang Fang in Chapter 5, and Dick van Dijk in Chapter 6, who I also thank for the great opportunity he gave me to attend the Fourth World Congress of the Bachelier Finance Society in Tokyo.

In general I am grateful to the Econometric Institute and the Tinbergen Institute for their financial support, which enabled me to attend several conferences. I am much obliged to several people at Erasmus University Rotterdam: Antoon Pelsser, Martin Martens and Dick van Dijk for allowing me to teach several courses on risk management and option pricing, Michiel de Pooter and Francesco Ravazzolo for creating a pleasant atmosphere when I was in Rotterdam, and finally the secretarial staff from both the Econometric Institute and the Tinbergen Institute for their assistance.

At Rabobank International I would like to thank all colleagues for creating an enjoyable working environment, in particular past and present colleagues from the Derivatives Research & Validation Team, the Risk Research Team (now Quantitative Risk Analytics) and the Credit Risk

Modelling team. My former direct colleagues at the Derivatives Research & Validation Team are thanked for enduring my many talks and technical reports, and for their feedback: Natalia Borovykh, Vladimir Brodski, Richard Dagg, Freddy van Dijk, Dejan Janković, Abdel Lantere, Maurice Lutterot, Herwald Naaktgeboren, Thomas Pignard and Erik van Raaij.

My friends are thanked for all the necessary distractions they provided. A special thank you is in place for Christian Kahl and Peter van de Ven for agreeing to be my paranymphs and therefore agreeing to wear a dress suit at my viva voce.

Last, but not least, I would like to thank my parents for their constant and invaluable support throughout my entire education. And of course Hellen, who I am certain is cherishing the additional time I will now be able to spend with her and our beloved Sam. The saying is true, no man succeeds without a good woman behind him.

Roger Lord
24 September 2008
Richmond, Greater London

Contents

Chapter 1

Introduction[1]

Options and derivatives came into existence long before the Nobel-prize winning papers of Black and Scholes [1973] and Merton [1973] launched the field of financial engineering. Option contracts had already been mentioned in ancient Babylonian and Greek times, though the first recorded case of organised futures trading occurred in Japan during the 1600s, when Japanese feudal lords sold their rice for future delivery in a market called *cho-ai-mai*, literally "rice trade on book". Around the same time, the Dutch started trading futures contracts and options on tulip bulbs at the Amsterdam stock exchange during the tulip mania of the early 1600s. The first formal futures and options exchange, the Chicago Board of Trade opened in 1848, but it was not until the opening of the Chicago Board Options Exchange in 1973, one month prior to the publication of the Black-Scholes paper, that option trading really took off.

Prior to the discovery of the Black-Scholes formula, investors and speculators would have had to use heuristic methods and their projections of the future to arrive at a price for a derivative. Attempts had been made to arrive at an option pricing formula, starting with Bachelier [1900], but all lacked the crucial insight of Black, Scholes and Merton that, under certain assumptions, the risk of an option can be fully hedged by dynamically investing in the underlying asset of that option. If one assumes that no arbitrage opportunities exist in financial markets, the price of any option must therefore be equal to the price of its replicating portfolio. This discovery, together with the arrival of hand-held calculators and, later, personal computers, made the derivatives market into the large industry it is today.

With the replication argument in hand, more exotic structures could be priced. One of the necessary requirements for such a price to make sense is that within the option pricing model, the prices of simpler, actively traded instruments, such as forward contracts and European options, coincide with their market price. It became apparent that this was not the case in the Black-Scholes model, and that the assumption that the underlying asset follows a geometric Brownian motion with constant, possibly time-dependent, drift and volatility was inappropriate. If this assumption were true, inverting the Black-Scholes formula with respect to the volatility for options with different strikes, but the same maturity, should yield approximately the same implied volatility. This is not the case. Much of the research within mathematical finance has therefore focused on alternative stochastic processes for the underlying asset, such that the prices of traded European options are more closely, if not perfectly, matched. To price an exotic option one then:

1. Chooses a model which is both economically plausible and analytically tractable;
2. Calibrates the model to the prices of traded vanilla options;
3. Prices the exotic option with the calibrated model, using appropriate numerical techniques.

This thesis is mainly concerned with the second and third steps in this process. Practitioners demand fast and accurate prices and sensitivities. As the financial models and option contracts used in practice are becoming increasingly complex, efficient methods have to be developed to

[1] Background information for this chapter has been used from Bernstein [1996], Dunbar [2000] and Teweles and Jones [1999].

cope with such models. All but one chapter of this thesis are therefore dedicated to the efficient pricing of options within so-called affine models, using methods ranging from analytical approximations to Monte Carlo methods and numerical integration.

The analytically tractable class of affine models, encompassing the Black-Scholes model, many stochastic volatility models including Heston's [1993] model, and exponential Lévy models, is described in Chapter 2. In order for a model to be practically relevant for the pricing of exotic options, it must allow for a fast and accurate calibration to plain vanilla option prices. This is indeed one of the appealing features of affine models. As for many affine models the characteristic function is known in closed-form, European option prices can be computed efficiently via Fourier inversion techniques.

Chapter 3 deals with a problem that occurs in the evaluation of the characteristic function of many affine models, among which the stochastic volatility model of Heston [1993]. Its characteristic function involves the evaluation of the complex logarithm, which is a multivalued function. If we restrict the logarithm to its principal branch, as is done in most software packages, the characteristic function can become discontinuous, leading to completely wrong option prices if options are priced by Fourier inversion. In this chapter we prove there is a formulation of the characteristic function in which the principal branch is the correct one. Similar problems in other models are also discussed.

Chapter 4 deals with the Fourier inversion that is used to price European options within the class of affine models. In Chapter 2 it is already shown that shifting the contour of integration along the complex plane allows for different representations of the inverse Fourier integral. In this chapter, we present the optimal contour of the Fourier integral, taking into account numerical issues such as cancellation and explosion of the characteristic function. This allows for fast and robust option pricing for virtually all levels of strikes and maturities, as demonstrated in several numerical examples.

The next three chapters are concerned with the actual pricing of exotic options, the third step we outlined above. Chapter 5 is mainly concerned with the pricing of early exercise options, though the presented algorithm can also be used for certain path-dependent options. The method is based on a quadrature technique and relies heavily on Fourier transformations. The main idea is to reformulate the well-known risk-neutral valuation formula by recognising that it is a convolution. The resulting convolution is dealt with numerically by using the Fast Fourier Transform (FFT). This novel pricing method, which we dub the Convolution method, CONV for short, is applicable to a wide variety of payoffs and only requires the knowledge of the characteristic function of the model. As such the method is applicable within many affine models.

Chapter 6 focuses on the simulation of square root processes, in particular within the Heston stochastic volatility model. Using an Euler discretisation to simulate a mean-reverting square root process gives rise to the problem that while the process itself is guaranteed to be nonnegative, the discretisation is not. Although an exact and efficient simulation algorithm exists for this process, at present this is not the case for the Heston stochastic volatility model, where the variance is modelled as a square root process. Consequently, when using an Euler discretisation, one must carefully think about how to fix negative variances. Our contribution is threefold. Firstly, we unify all Euler fixes into a single general framework. Secondly, we introduce the new full truncation scheme, tailored to minimise the upward bias found when pricing European options. Thirdly and finally, we numerically compare all Euler fixes to other recent schemes. The choice of fix is found to be extremely important.

Chapter 7 is a slight departure from the previous chapters in that we focus solely on the Black-Scholes model. Some methods from this chapter can however be applied to the whole class of affine models, as shown in Lord [2006]. This chapter focuses on the pricing of European Asian options, though all approaches considered are readily extendable to the case of an Asian basket option. We consider several methods for evaluating the price of an Asian option, and contribute to them all. Firstly, we show the link between several PDE methods. Secondly, we show how a

closed-form expression can be derived for Curran's and Rogers and Shi's lower bound for the general case of multiple underlyings. Thirdly, we considerably sharpen Thompson's [1999a,b] upper bound such that it is tighter than all known upper bounds. Finally, we consider analytical approximations and combine the traditional moment matching approximations with Curran's conditioning approach. The resulting class of partially exact and bounded approximations can be proven to lie between a sharp lower and upper bound. In numerical examples we demonstrate that they outperform all current state-of-the-art bounds and approximations.

Finally, in Chapter 8 we consider a completely different topic, namely the properties of correlation matrices of term structure data, which can be used as inputs within term structure option pricing models, such as interest rate models. The first three factors resulting from a principal components analysis of term structure data are in the literature typically interpreted as driving the level, slope and curvature of the term structure. Using slight generalisations of theorems from total positivity, we present sufficient conditions under which level, slope and curvature are present. These conditions have the nice interpretation of restricting the level, slope and curvature of the correlation surface. It is proven that the Schoenmakers-Coffey correlation matrix also brings along such factors. Finally, we formulate and corroborate our conjecture that the order present in correlation matrices causes slope.

Chapter 2

Affine models

In this section we discuss the class of affine models which will be used frequently in the next chapters. This class of models was pioneered by Duffie and Kan [1996] in a term structure context, and subsequently analysed in great detail in Duffie, Pan and Singleton [2000] and Duffie, Filipović and Schachermayer [2003]. The popularity of this class can be explained by both its modelling flexibility and its analytical tractability, which greatly facilitates the estimation and calibration of such models. Prominent members of this class are the Black-Scholes model [1973], the term structure models of Vašiček [1977] and Cox, Ingersoll and Ross [1985], as well as Heston's [1993] stochastic volatility model.

Our exposition of affine processes will be based on Duffie, Pan and Singleton, who consider affine jump-diffusion processes. We slightly relax their assumptions as we allow the jump processes to have infinite activity. This setup is not quite as general as the regular affine processes considered by Duffie, Filipović and Schachermayer, but is sufficient for our purposes. In Section 2.1 we briefly describe Lévy processes, the building blocks of the affine models we consider in Section 2.2. The motivation behind using more general Lévy processes than the Brownian motion with drift is the fact that the Black-Scholes model is not able to reproduce the volatility skew or smile present in most financial markets. Over the past few years it has been shown that several exponential Lévy models are, at least to some extent, able to reproduce the skew or smile. In Section 2.3 we demonstrate how the discounted characteristic function can be derived for affine processes, and conclude in Section 2.4 by showing how European options can be priced by inverting the characteristic function.

2.1. Lévy processes

A Lévy process, named after the French mathematician Paul Lévy, is a continuous-time stochastic process with stationary independent increments. Its most well-known examples are Wiener processes or Brownian motions, and Poisson processes. To be precise, a càdlàg, adapted, real-valued process L(t) on the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$, with L(0) = 0, is a Lévy process if:

1. it has independent increments;
2. it has stationary increments;
3. it is stochastically continuous, i.e. for any $t \geq 0$ and $\varepsilon > 0$ we have:

$$\lim_{s \to t} \mathbb{P}\big(|L(t) - L(s)| > \varepsilon\big) = 0 \tag{2.1}$$

Each Lévy process can be characterised by a triplet (μ, σ, ν), representing the drift, diffusion and jump component of the process. We have $\mu \in \mathbb{R}$, $\sigma \geq 0$ and ν a measure satisfying $\nu(0) = 0$ and:

$$\int_{\mathbb{R}} \min(1, |x|^2)\, \nu(dx) < \infty \tag{2.2}$$

The Lévy measure carries useful information about the path properties of the process. For example, if $\nu(\mathbb{R}) < \infty$, then almost all paths have a finite number of jumps on every compact interval. In this case the process is said to have finite activity. This is the case for e.g. compound Poisson processes. In contrast, if $\nu(\mathbb{R}) = \infty$, the process is said to have infinite activity.

In terms of the Lévy triplet the characteristic function of the Lévy process for $u \in \mathbb{R}$ equals:

$$\begin{aligned}\phi(u) &= \mathbb{E}\left[\exp\left(iu\,L(t)\right)\right] = \exp(\ln\psi(u))\\ &= \exp\left(t\left(i\mu u - \tfrac{1}{2}\sigma^2u^2 + \int_{\mathbb{R}}(e^{iux} - 1 - iux1_{[|x|<1]})\,\nu(dx)\right)\right)\end{aligned} \tag{2.3}$$

the celebrated Lévy-Khinchin formula. The exponent $\psi(u)$ in (2.3) is referred to as the Lévy or characteristic exponent.

Lévy processes are strongly connected to infinitely divisible distributions. A random variable X is said to have an infinitely divisible distribution, if for all $n \in \mathbb{N}$ there exist i.i.d. random variables $X_1^{(1/n)}, \ldots, X_n^{(1/n)}$ such that:

$$X \overset{d}{=} X_1^{(1/n)} + \ldots + X_n^{(1/n)} \tag{2.4}$$

Equivalently, the characteristic function of X then satisfies:

$$\phi_X(u) = \left(\phi_{X^{(1/n)}}(u)\right)^n \tag{2.5}$$

The probability law of a random variable X is infinitely divisible if and only if its characteristic function can be written in the form of the Lévy-Khinchin formula.

In the subsequent chapters we will often deal with exponential Lévy models, where the asset price is modelled as an exponential function of a Lévy process L(t):

$$S(t) = S(0)\,\exp\left(L(t)\right) \tag{2.6}$$

Let us assume the existence of a bank account M(t) which evolves according to $dM(t) = r\,M(t)\,dt$, r being the deterministic and constant risk-free rate. As is common in most models nowadays we assume that (2.6) is formulated under the risk-neutral measure. For ease of exposure we also assume the asset pays a a deterministic and continuous stream of dividends, measured by the dividend rate q. To ensure the reinvested relative price $e^{qt}S(t) / M(t)$ is a martingale under the risk-neutral measure, we require:

$$\phi(-i) = \mathbb{E}\left[\exp\left(L(t)\right)\right] = e^{(r-q)t} \tag{2.7}$$

which is satisfied if we choose the drift μ as:

$$\mu = r - q - \tfrac{1}{2}\sigma^2 - \int_{\mathbb{R}}\left(e^x - 1 - x1_{[|x|<1]}\right)\nu(dx) \tag{2.8}$$

For more background information we refer the interested reader to Cont and Tankov [2004] for an extensive manuscript on the usage of Lévy processes in a financial context and to Sato [1999, 2001] for a detailed analysis of Lévy processes in general. Papapantoleon [2006] provides a good short introduction to the applications of Lévy processes in mathematical finance.

2.2. Affine processes

Let[2] $\mathbf{X}$ be a Markov process in the domain $D \subset \mathbb{R}^n$ satisfying:

$$d\mathbf{X}(t) = \boldsymbol{\mu}(\mathbf{X}(t))dt + \boldsymbol{\sigma}(\mathbf{X}(t))d\mathbf{W}(t) + d\mathbf{Z}(t) \tag{2.9}$$

Here $\mathbf{W}$ is a standard Brownian motion in $\mathbb{R}^n$, $\boldsymbol{\mu}$: $D \to \mathbb{R}^n$, $\boldsymbol{\sigma}$: $D \to \mathbb{R}^{n\times n}$ and $\mathbf{Z}$ is a pure jump Lévy process of infinite activity, whose jumps are described by the Lévy measure $\boldsymbol{\nu}$ on $\mathbb{R}^n$ and arrive with intensity $\{\lambda(\mathbf{X}(t)) : t \geq 0\}$ for some λ: $D \to \mathbb{R}^+$. Most models encountered in the literature either combine a Brownian motion with a jump process of finite activity, or leave out the Brownian motion and use a jump process of infinite activity. However, it is obviously possible to combine both. For notational convenience we do not allow μ, σ, λ and ν to depend on time. All results are easily extended to accommodate for time dependency in these functions.

In addition to the process in (2.9) we specify the short rate as a function r: $D \to \mathbb{R}$. The money market or bank account is then defined as:

$$dM(t) = r(\mathbf{X}(t))M(t)\,dt \tag{2.10}$$

The process $\mathbf{X}$ is affine if and only if:

- $\boldsymbol{\mu}(\mathbf{x}) = \mathbf{m}_0 + \mathbf{m}_1\mathbf{x}$, for $(\mathbf{m}_0, \mathbf{m}_1) \in \mathbb{R}^n \times \mathbb{R}^{n\times n}$;
- $\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}(\mathbf{x})^T = \boldsymbol{\Sigma}_0 + \sum_{i=1}^n \boldsymbol{\Sigma}_{1i} x_i = \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1\mathbf{x}$, for $(\boldsymbol{\Sigma}_0, \boldsymbol{\Sigma}_1) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n\times n}$;
- $\lambda(\mathbf{x}) = \ell_0 + \boldsymbol{\ell}_1^T\mathbf{x}$, for $(\ell_0, \boldsymbol{\ell}_1) \in \mathbb{R} \times \mathbb{R}^n$.

where $\boldsymbol{\Sigma}_1 = (\boldsymbol{\Sigma}_{11} \quad \cdots \quad \boldsymbol{\Sigma}_{1n})$ and $\boldsymbol{\Sigma}_1\mathbf{x}$ is interpreted as a vector inner product. Finally, we also assume the short rate or discount function is affine:

- $r(\mathbf{x}) = r_0 + \mathbf{r}_1^T\mathbf{x}$, for $(r_0, r_1) \in \mathbb{R} \times \mathbb{R}^n$.

In words, for the process to be affine, we require both its instantaneous drift, variance and jump intensity to be at most affine combination of the factors. Finally, we also assume the short rate is an affine combination of the factors.

2.2.1. Heston's stochastic volatility model

As an example, let us consider a model that will feature prominently in the rest of this thesis - the Heston stochastic volatility model. Heston [1993] proposed the following model, where the stochastic volatility is modelled by the same square-root process that is used for the short rate in the Cox-Ingersoll-Ross model:

$$\begin{aligned} dS(t) &= \mu(t)S(t)dt + \sqrt{v(t)}\,S(t)\,dW_S(t) \\ dv(t) &= -\kappa(v(t) - \theta)dt + \omega\sqrt{v(t)}\,dW_v(t) \end{aligned} \tag{2.11}$$

[2] As a matter of notation, vectors and matrices will be typeset in bold.

In this stochastic differential equation (SDE) S represents the asset, whose stochastic variance v is modelled as a mean-reverting square root process. The Brownian motions are correlated with correlation coefficient ρ. Leaving the interpretation of the parameters to later chapters, we focus on the characterisation of (2.11) as an affine process. Clearly, if we take (S,v) as the state variables, the Heston model is not affine. The drift will be affine in these state variables, the variance will however not be, as we can see by calculating the instantaneous variance of the stock price, which is equal to $v(t)S(t)^2$. If we however consider (x,v), with x = ln S, as the state variables, we do obtain an affine process. Applying Itō's lemma, and rewriting the SDE in terms of independent Brownian motions, we obtain:

$$\begin{aligned} dx(t) &= \left(\mu - \tfrac{1}{2}v(t)\right)dt + \sqrt{v(t)}\, dW_1(t) \\ dv(t) &= -\kappa(v(t)-\theta)dt + \rho\omega\sqrt{v(t)}\, dW_1(t) + \omega\sqrt{(1-\rho^2)\, v(t)}\, dW_2(t) \end{aligned} \tag{2.12}$$

Using the notation of the previous section for the drift matrix, we obtain:

$$\mathbf{m}_0 = \begin{pmatrix} \mu \\ \kappa\theta \end{pmatrix} \quad \mathbf{m}_1 = \begin{pmatrix} 0 & -\frac{1}{2} \\ 0 & -\kappa \end{pmatrix} \tag{2.13}$$

The instantaneous variance-covariance matrix is equal to:

$$\boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}(\mathbf{x})^T = \begin{pmatrix} \sqrt{v} & 0 \\ \rho\omega\sqrt{v} & \omega\sqrt{(1-\rho^2)v} \end{pmatrix}\begin{pmatrix} \sqrt{v} & 0 \\ \rho\omega\sqrt{v} & \omega\sqrt{(1-\rho^2)v} \end{pmatrix}^T = \begin{pmatrix} v & \rho\omega v \\ \rho\omega v & \omega^2 v \end{pmatrix} \tag{2.14}$$

leading to:

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \boldsymbol{\Sigma}_1 = \left(\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 1 & \rho\omega \\ \rho\omega & \omega^2 \end{pmatrix}\right) \tag{2.15}$$

The interest rate process is taken constant and since there are no jumps, there is no need to specify the jump intensity and measure.

2.3. The discounted characteristic function

Duffie et al. [2000] studied the discounted extended characteristic function of affine processes, which is defined as:

$$\phi\left(\mathbf{u}, t, T, \mathbf{X}(t)\right) \equiv M(t)\cdot \mathbb{E}_t\left[\frac{\exp\left(i\mathbf{u}^T\mathbf{X}(T)\right)}{M(T)}\right] = P(t,T)\cdot \mathbb{E}_t^T\left[\exp\left(i\mathbf{u}^T\mathbf{X}(T)\right)\right] \tag{2.16}$$

where $\mathbf{u} \in \mathbb{C}^n$. We attached a subscript t to the expectation operator to indicate that the expectation is being taken with respect to the information set at time t. The second expectation is taken under the T-forward measure, induced by taking P(·,T), the zero-coupon bond maturing at time T, as the numeraire asset. The characteristic function in (2.16) divided by P(t,T) will be

referred to as the forward characteristic function, though we will omit the term forward if the context is clear. Note that when $\mathbf{u} = 0$, we obtain P(t,T).

Contrary to the usual characteristic function, which takes a real-valued argument, the extended characteristic function is not always defined. We have, in shorthand notation:

$$|\phi(\mathbf{u})| = |\mathbb{E}[e^{i\mathbf{u}^T\mathbf{X}(T)}]| \leq \mathbb{E}[|e^{i\mathbf{u}^T\mathbf{X}(T)}|] = \varphi(i\,\mathrm{Im}(\mathbf{u})) \tag{2.17}$$

so that the strip of regularity of the extended characteristic function, $\Lambda_{\mathbf{X}}$, is defined by:

$$\Lambda_{\mathbf{X}} = \left\{\mathbf{u} \in \mathbb{C}^n \,\middle|\, \phi(i\,\mathrm{Im}(\mathbf{u})) < \infty\right\} \tag{2.18}$$

In the following we will only deal with extended characteristic functions, so that we drop the term extended.

The question we consider here is how to solve the characteristic function of an affine process. From the Feynman-Kac formula for jump processes (see e.g. Cont and Tankov [2004, Proposition 12.5]), we know that ϕ should obey the following partial integro-differential equation (PIDE):

$$\begin{aligned}&\frac{\partial\phi}{\partial t} + \left(\frac{\partial\phi}{\partial \mathbf{x}}\right)^T \boldsymbol{\mu} + \tfrac{1}{2}\mathrm{tr}\left(\frac{\partial^2\phi}{\partial \mathbf{x}\partial \mathbf{x}^T}\boldsymbol{\sigma}\boldsymbol{\sigma}^T\right)\\ &+ \lambda\int_{\mathbb{R}^n}\left(\phi(t,\mathbf{x}+\mathbf{z}) - \phi(t,\mathbf{x}) - \mathbf{z}^T 1_{[|\mathbf{z}|\leq 1]}\frac{\partial\phi}{\partial \mathbf{x}}\right)\upsilon(d\mathbf{z}) = r\phi\end{aligned} \tag{2.19}$$

with the obvious boundary condition

$$\phi(\mathbf{u}, T, T, \mathbf{X}(t)) = \exp(i\mathbf{u}^T\mathbf{X}(T)) \tag{2.20}$$

The solution to (2.19)-(2.20), under certain technical regularity conditions (see Duffie et al. [2000, Proposition 1]), is given by:

$$\phi(\mathbf{u}, t, T, \mathbf{x}) = \exp(A(\mathbf{u}, t, T) + \mathbf{B}(\mathbf{u}, t, T)^T\mathbf{x}) \tag{2.21}$$

where A and $\mathbf{B}$ are respectively $\mathbb{R}$ and $\mathbb{R}^n$-valued functions. For the proof and regularity conditions we refer the interested reader to the article itself. From (2.21) we can deduce that:

$$\frac{\partial\phi}{\partial t} = \left(\frac{dA}{dt} + \left(\frac{d\mathbf{B}}{dt}\right)^T\mathbf{x}\right)\phi \qquad \frac{\partial\phi}{\partial \mathbf{x}} = \mathbf{B}\phi \qquad \frac{\partial^2\phi}{\partial \mathbf{x}\mathbf{x}^T} = \mathbf{B}\mathbf{B}^T\phi \tag{2.22}$$

Using the well-known identity that $\mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA})$ if $\mathbf{AB}$ and $\mathbf{BA}$ are well-defined, we deduce:

$$\mathrm{tr}\left(\frac{\partial^2\phi}{\partial \mathbf{x}\mathbf{x}^T}\boldsymbol{\sigma}\boldsymbol{\sigma}^T\right) = \phi\cdot\mathrm{tr}\left(\mathbf{B}\mathbf{B}^T(\boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_1\mathbf{x})\right) = \phi\cdot\mathbf{B}^T\boldsymbol{\Sigma}_0\mathbf{B} + \phi\cdot\mathbf{B}^T\boldsymbol{\Sigma}_1\mathbf{x}\mathbf{B} \tag{2.23}$$

Inserting the functional form (2.21) into the PIDE in (2.19), and dividing by ϕ yields:

$$\begin{aligned}&\frac{dA}{dt}+\left(\frac{d\mathbf{B}}{dt}\right)^{T}\mathbf{x}+\mathbf{B}^{T}(\mathbf{m}_0+\mathbf{m}_1\mathbf{x})+\tfrac{1}{2}\left(\mathbf{B}^{T}\boldsymbol{\Sigma}_0\mathbf{B}+\mathbf{B}^{T}\boldsymbol{\Sigma}_1\mathbf{x}\mathbf{B}\right)\\&+(\ell_0+\boldsymbol{\ell}_1^{T}\mathbf{x})\int_{\mathbb{R}^n}(e^{\mathbf{B}^{T}z}-1-\mathbf{z}^{T}1_{[|\mathbf{z}|\le 1]}\mathbf{B})\nu(d\mathbf{z})=r_0+\mathbf{r}_1^{T}\mathbf{x}\end{aligned} \tag{2.24}$$

Let us denote $\theta(\mathbf{u})\equiv\int_{\mathbb{R}^n}(e^{i\mathbf{u}^{T}z}-1-i\mathbf{u}^{T}\mathbf{z}1_{[|\mathbf{z}|\le 1]})\nu(d\mathbf{z})$ for $\mathbf{u}\in\mathbb{C}^n$, the jump-transform, provided it is well-defined. Now, since (2.24) should hold for every $\mathbf{x}$, it must also hold for $\mathbf{x}=0$. This yields an ordinary differential equation (ODE) for A. Using this relation, we can find an additional n ODEs for each element of $\mathbf{B}$ by setting $\mathbf{x}=(1,0,\ldots,0)^{T}$, …, $\mathbf{x}=(0,0,\ldots,1)^{T}$:

$$\begin{aligned}&\frac{dA}{dt}+\mathbf{B}^{T}\mathbf{m}_0+\tfrac{1}{2}\mathbf{B}^{T}\boldsymbol{\Sigma}_0\mathbf{B}+\ell_0\theta(-i\mathbf{B})=r_0\\&\frac{d\mathbf{B}}{dt}+\mathbf{m}_1^{T}\mathbf{B}+\tfrac{1}{2}\mathbf{B}^{T}\boldsymbol{\Sigma}_1\mathbf{B}+\boldsymbol{\ell}_1\theta(-i\mathbf{B})=\mathbf{r}_1\end{aligned} \tag{2.25}$$

where with a slight abuse of notation we denote $\mathbf{B}^{T}\boldsymbol{\Sigma}_1\mathbf{B}$ for the vector in $\mathbb{C}^n$ with i^{th} element equal to $\mathbf{B}^{T}\boldsymbol{\Sigma}_{1i}\mathbf{B}$. The boundary conditions on ϕ carry over to A and $\mathbf{B}$: $A(\mathbf{u},T,T)=0$ and $\mathbf{B}(\mathbf{u},T,T)=i\mathbf{u}$. In general, the solutions to A and $\mathbf{B}$ will have to be found by numerically solving the ODEs in (2.25), for example by using the Runge-Kutta method, see e.g. Press, Teukolsky, Vetterling and Flannery [2007]. In these cases it can be very advantageous to choose a jump measure such that θ can easily be computed. Models for which the characteristic function can be solved in closed-form of course of course have a large advantage over models for which the ODEs in (2.25) have to be solved numerically. The characteristic functions of the models we consider will be supplied when we start using the models in later chapters. For example, the characteristic function for the Heston model from Section 2.2.1 is derived in Chapter 3.

2.4. Characteristic functions and option pricing

The first option pricing model in the literature that provided semi-analytical option prices by means of inverting the characteristic function of the underlying asset was the stochastic volatility model of Heston [1993]. Prior to Heston, Stein and Stein [1991] had utilised Fourier inversion techniques to calculate the stock price distribution in their stochastic volatility model. Whereas Heston's approach is directly applicable to any model where the characteristic function of the logarithm of the asset is known, Stein and Stein's approach relies heavily on the independence of the stochastic volatility process and the asset itself. Since Heston's seminal paper, the pricing of European options by means of Fourier inversion has become more and more commonplace.

Before demonstrating how European options can be priced by means of inversion techniques, we recall that the risk-neutral valuation theorem states that the forward price of a European call option on a single asset S can be written as:

$$C(S(t),K,T)=\mathbb{E}[(S(T)-K)^{+}] \tag{2.26}$$

where V denotes the value, T the maturity and K the strike price of the call. The expectation is taken under the T-forward probability measure. As (2.26) is an expectation, it can be calculated

via numerical integration provided that the probability density is known in closed-form. This is not the case for many models which do however have a closed-form characteristic function[3].

Starting from Heston [1993], many papers have solved the problem differently. We note that equation (2.26) can be written very generally as:

$$C\big(S(t),K,T\big) = F(t,T)\cdot \mathbb{S}\big(S(T)>K\big) - K\cdot \mathbb{P}\big(S(T)>K\big) \tag{2.27}$$

with F(t,T) the forward price of the underlying asset at time T, as seen from t. Finally, $\mathbb{P}$ and $\mathbb{S}$ indicate respectively the T-forward probability measure and the stock price measure, induced by taking the asset price itself as the numeraire asset. Note that (2.27) has the same form as the celebrated Black-Scholes formula. Both cumulative probabilities can be found by inverting the forward characteristic function:

$$\begin{aligned}\mathbb{P}\big(S(T)>K\big) &= \tfrac{1}{2} + \tfrac{1}{\pi}\int_0^\infty \mathrm{Re}\; e^{-iuk}\,\frac{\phi(u)}{iu}\,du \\ \mathbb{S}\big(S(T)>K\big) &= \tfrac{1}{2} + \tfrac{1}{\pi}\int_0^\infty \mathrm{Re}\; e^{-iuk}\,\frac{\phi(u-i)}{iu\phi(-i)}\,du\end{aligned} \tag{2.28}$$

where k is the logarithm of the strike price K. This approach dates back to Lévy [1925]. His inversion theorem is restricted to cases where the random variable is strictly positive. Gurland [1948] and Gil-Pelaez [1951] derived this more general inversion theorem. We refer the interested reader to Lukacs [1970] for a detailed account of the history of these approaches.

To allow for greater flexibility, Carr and Madan [1999] found an alternative representation for the European call price. Note that L^1-integrability is a sufficient condition for the Fourier transform of a function to exist. A call option is not L^1-integrable with respect to the logarithm of the strike price, as:

$$\lim_{k\to-\infty} C\big(S(t),e^k,T\big) = F(t,T) \tag{2.29}$$

Damping the option price with $\exp(\alpha k)$ for $\alpha > 0$ solves this however. Let us consider the case where the asset price is an exponential function of a stochastic process X:

$$S(t) = \exp\big(X(t)\big) \tag{2.30}$$

Following Carr and Madan, we consider the Fourier transform of $\exp(\alpha k)\cdot C(k)$, the damped forward call price. It is equal to:

$$\begin{aligned}\psi(v,\alpha) &\equiv \int_{-\infty}^{\infty} e^{ivk}e^{\alpha k}\,C(k)\,dk = \int_{-\infty}^{\infty} e^{(iv+\alpha)k}\,\mathbb{E}[(e^{X(T)}-e^k)^+]\,dk \\ &= \mathbb{E}\left[\int_{-\infty}^{X(T)} e^{(iv+\alpha)k}(e^{X(T)}-e^k)\,dk\right] = \mathbb{E}\left[\frac{e^{i(v-i(\alpha+1))X(T)}}{-(v-i\alpha)(v-i(\alpha+1))}\right] \\ &= \mathbb{E}\left[\frac{e^{i(v-i(\alpha+1))X(T)}}{-(v-i\alpha)(v-i(\alpha+1))}\right] = \frac{\phi\big(v-i(\alpha+1)\big)}{-(v-i\alpha)(v-i(\alpha+1))}\end{aligned} \tag{2.31}$$

[3] Or, their probability density involves complicated special functions whereas their characteristic function is comparatively easier.

If X is an affine process, the characteristic function ϕ can easily be found via the numerical solution of the ODEs in (2.25). Inverting the Fourier transform and undamping yields:

$$C(S(t),K,T) = \frac{e^{-\alpha k}}{\pi}\int_0^\infty \mathrm{Re}\left(e^{-ivk}\psi(v,\alpha)\right) dv \tag{2.32}$$

The Fourier transform was taken with respect to the log-strike price in view of utilising the fast Fourier transform (FFT) to retrieve option values for a whole grid of strikes with just one evaluation of the FFT. The focus of Carr and Madan was purely on call options, which can be retrieved by using $\alpha > 0$. Aside from this condition, a necessary and sufficient condition for the damped option price and the option price itself to be well-defined is:

$$\left|\phi(v-(\alpha+1)i)\right| \le \phi(-(\alpha+1)i) = \mathbb{E}[S(T)^{\alpha+1}] < \infty \tag{2.33}$$

i.e. that the $(\alpha+1)^{st}$ moment of the asset exists. Although this approach was new to the area of option pricing, the idea of damping functions on the positive real line in order to be able to find their Fourier transform is an idea that goes back to at least Dubner and Abate [1968].

Raible [2000] and Lewis [2001][4] considered similar approaches, except that their transforms were taken with respect to the log-forward and log-spot price, respectively. Both authors demonstrated that the formula in (2.32) is quite general in that it can be adapted to a wide variety of European payoff functions, provided we can analytically calculate the Fourier transform of the damped payoff function. In Lewis [2001], however, an important step was made by considering the resulting integral as a contour integral in the complex plane. By shifting the contour (effectively changing α in (2.32)), various parity relations are obtained. As we will stick with the formula obtained by taking the Fourier transform of the log-strike price, the result of Lewis here comes down to:

$$C(S(t),K,T,\alpha) = R(F(t,T),K,\alpha) + \frac{1}{2\pi}\int_{-\infty-i\alpha}^{\infty-i\alpha} e^{-izk}\,\frac{\phi(z-i)}{-z(z-i)}\,dz \tag{2.34}$$

where the residue term equals:

$$R(F,K,\alpha) = F\cdot 1_{[\alpha\le 0]} - K\cdot 1_{[\alpha\le -1]} - \tfrac{1}{2}\left(F\cdot 1_{[\alpha=0]} - K\cdot 1_{[\alpha=-1]}\right) \tag{2.35}$$

These parity relations remove the restriction that $\alpha > 0$, leaving (2.33) as the only real restriction. The representation in (2.31)-(2.35) has two distinct advantages over (2.28). Firstly, it only requires a single numerical integration. Secondly, whereas (2.28) can suffer from cancellation errors, the numerical stability of (2.32) can be controlled by means of the damping coefficient α, as we will discuss at great lengths in Chapter 4.

[4] We thank Ariel Almendral Vázquez for pointing out the equivalence of these two approaches to us.

Chapter 3

Complex logarithms in Heston-like models[5]

In this chapter we will analyse the complex discontinuities one finds when evaluating the characteristic function of several popular option pricing models, in particular the seminal stochastic volatility model of Heston [1993]. Since the initial breakthrough by Heston a whole range of models have appeared that allow for closed-form characteristic functions. As demonstrated in the previous chapter, in such models, the prices of European options can be calculated semi-analytically by means of Fourier inversion.

This approach has in recent years been refined by Carr and Madan [1999], Lewis [2001] and Lee [2004], leading to the option pricing equations in (2.31)-(2.35). We restate some of the equations here, as they are crucial to this chapter. Recall that the forward price of a European call on an underlying asset S can be written as:

$$\mathbb{E}\left[(S(T)-K)^{+}\right] = R(F,K,\alpha) + \frac{1}{\pi}\mathrm{Re}\int_0^{\infty} e^{-i(v-i\alpha)k}\,\frac{\phi\left(v-i(\alpha+1)\right)}{-\left(v-i(\alpha+1)\right)(v-i\alpha)}\,dv \tag{3.1}$$

where ϕ is the forward characteristic function of the log-stock price, $\phi(u) = \mathbb{E}\left[e^{iu\ln S(T)}\right]$, and:

$$R(F,K,\alpha) = F\cdot 1_{[\alpha\le 0]} - K\cdot 1_{[\alpha\le -1]} - \tfrac{1}{2}\left(F\cdot 1_{[\alpha=0]} - K\cdot 1_{[\alpha=-1]}\right) \tag{3.2}$$

is a residue term, arising from the poles of the integrand in (3.1). This representation only holds for values of the damping coefficient α satisfying $\phi(-i(\alpha+1)) < \infty$, i.e. for those values of α where the $(\alpha+1)^{th}$ moment of S(T) is finite. In theory, the option price is independent of the parameter α; in practice however, α affects the behaviour of the integrand and choosing the right value for it is crucial. In Chapter 4 we outline the optimal choice of α.

If we use equations (3.1)-(3.2) to evaluate the option price, discontinuities in a characteristic function will clearly lead to discontinuities in the integrand of (3.1). In turn, this will produce completely wrong option prices, as we will see shortly. The problem at hand is not unique to option pricing, but will occur in any application where one requires an evaluation of a characteristic function. Another example of this is the calculation of density functions via inversion of the characteristic function, see the exact simulation algorithm of Broadie and Kaya [2006] for an application of this in the Heston model.

To fix ideas, we will focus on the Heston stochastic volatility model throughout this chapter, though we explore other models in the penultimate section. Two formulations of the characteristic function are prevalent in the literature. The first, which is the original formulation of Heston [1993], is known to suffer from discontinuities when the complex logarithm is restricted to its principal branch. Schöbel and Zhu [1999] first mentioned such problems in the literature, albeit for their own stochastic volatility model. They proposed an ad-hoc workaround by letting the

[5] An abridged version of this chapter has been accepted for publication in Mathematical Finance. Its precursor is Lord, R. and C. Kahl [2006], "Why the rotation count algorithm works", Tinbergen Institute Discussion Paper No. TI 2006-065/2, available at: http://ssrn.com/abstract=921335.

integration algorithm used to calculate (3.1) pick up any discontinuities and correct for them. As this is not foolproof, Kahl and Jäckel [2005] considered the same problem recently and came up with the rotation count algorithm, an easily implementable algorithm that keeps the complex logarithm in the Heston model continuous. Though the algorithm seems to work perfectly well, a formal proof is required as one does not want to be caught by a counterexample.

The second and less widespread formulation has to our knowledge first appeared in Bakshi, Cao and Chen [1997, eq. A.11] and later in Duffie, Pan and Singleton [2000] and Gatheral [2006] among others. Interestingly most of the authors using this formulation never mentioned any problems with regard to the complex logarithm. In this chapter we prove that indeed no complex discontinuities arise in this formulation if the complex logarithm is restricted to its principal branch, something that has already been conjectured by Lord and Kahl [2006] and Gatheral [2006]. In Lord and Kahl [2006] we proved this result under a mild constraint on the correlation coefficient ρ. In particular, the result certainly holds true when ρ is non-positive, which seems to be the practically most relevant case. In addition we verified the correctness of the rotation count algorithm under this same constraint. Subsequently, Albrecher, Mayer, Schoutens and Tistaert [2007] proved the continuity of this second formulation without any restrictions on the parameters. However, they do restrict α in $\phi(v-(\alpha+1)i)$ to be positive. As results in Lee [2004] and Chapter 4 demonstratively show that a robust option pricer would need to able to choose α freely, this restriction would hamper the implementation hereof. A final paper to appear on this topic, Fahrner [2007], considers the special case when $\alpha = -1/2$. The restriction on the correlation coefficient is exactly the same as in Lord and Kahl.

As the union of the results from Lord and Kahl and Albrecher et al. do not fully prove the aforementioned conjecture, we set out to finally prove this result without any restrictions on the parameters, and more importantly, without any restrictions on α. The remainder of this chapter is structured as follows. In Section 3.1 we state the two formulations of the Heston characteristic function and show what problems can be caused by using the wrong branch of the complex logarithm. Section 3.2 proves that the second formulation is continuous when the complex logarithm is restricted to its principal branch. Section 3.3 considers the rotation count algorithm and analyses in which parameter region we can guarantee that it works. Finally, in Section 3.4 we consider other models with similar problems and show how to avoid the complex discontinuities there. The models considered are the stochastic volatility models of Schöbel and Zhu [1999] and Duffie, Pan and Singleton [2000], as well as the recent exact simulation algorithm that Broadie and Kaya [2006] developed for the Heston model and extensions thereof. Section 3.5 concludes.

3.1. Complex discontinuities in the Heston model

In this section we will first derive the characteristic function of the underlying asset in the Heston model. In the second subsection we discuss the complex discontinuities that are present in the Heston model, and present several examples of the impact this could have on option prices.

3.1.1. Derivation of the characteristic function

Under the risk-neutral pricing measure the Heston stochastic volatility model is specified by the following set of stochastic differential equations (cf. (2.11)):

$$\begin{aligned} dS(t) &= \mu(t)S(t)dt + \sqrt{v(t)}\,S(t)\,dW_S(t) \\ dv(t) &= -\kappa\big(v(t)-\theta\big)dt + \omega\sqrt{v(t)}\,dW_v(t) \end{aligned} \tag{3.3}$$

where the Brownian motions satisfy $dW_S(t) \cdot dW_v(t) = \rho\, dt$. The underlying asset S has a stochastic variance v, which is modelled as a mean-reverting square root process. The parameter κ is the rate of mean reversion of the variance, θ is the long-term level of variance and ω is the volatility of variance. Finally, the drift $\mu(t)$ is used to fit to the forward curve of the underlying. We make the following assumption on the parameters:

Assumption:

$$\kappa > 0,\ \omega > 0,\ |\rho| < 1 \tag{3.4}$$

If $\omega = 0$, the model in (3.3) collapses to the Black-Scholes model with a time-dependent volatility. Similarly, if $|\rho| = 1$, the model is a special case of the local volatility model. The assumption that $\kappa > 0$ is not essential, though it will facilitate the analysis. Though the Heston model as postulated here is typically used for asset classes such as equity and foreign exchange, the mean-reverting square root process can be used as a stochastic volatility driver in any asset class, see e.g. Andersen and Andreasen [2002] and Andersen and Brotherton-Ratcliffe [2005] for applications in an interest rate context, and Mercurio and Moreni [2006] for an inflation context.

Although the model in (3.3) is not affine in the underlying asset S and the stochastic variance v, it is affine in ln S and v, as demonstrated in Section 2.2.1. Furthermore, note that the only time-inhomogeneous part of the model is the forward curve of the underlying asset. Hence, following Duffie et al. [2000] we know the characteristic function of the logarithm of the underlying will be exponentially affine in the logarithm of the forward, and the stochastic variance:

$$\phi(u) = \mathbb{E}\left[e^{iu \ln S(T)}\right] = \exp\left(iuf + A(u,\tau) + B_v(u,\tau) \cdot v(0)\right) \tag{3.5}$$

Here $u \in \mathbb{C}$ and f is shorthand for ln F(T), the logarithm of the forward price. Following Section 2.2 we know that the functions A and B_v satisfy the following system of ODEs:

$$\begin{aligned} \frac{dB_v}{d\tau} &= \hat{\alpha}(u) - \beta(u)B_v + \gamma B_v^2 \\ \frac{dA}{d\tau} &= \kappa\theta B_v \end{aligned} \tag{3.6}$$

subject to the initial conditions A(u,0) = 0 and B(u,0) = 0. The auxiliary variables we introduced are $\hat{\alpha}(u) = -\frac{1}{2}u(i+u)$, $\beta(u) = \kappa - \rho\omega ui$ and $\gamma = \frac{1}{2}\omega^2$. Recasting the first Riccati equation as:

$$\frac{dB_v}{d\tau} = \gamma(B_v - a)(B_v - b) \tag{3.7}$$

immediately leads to the following solution:

$$B_v(u,\tau) = ab\frac{1 - e^{(b-a)\gamma\tau}}{a - be^{(b-a)\gamma\tau}} \tag{3.8}$$

The roots of the Riccati equation for B_v are $a = (\beta+D)/\omega^2$ and $b = (\beta-D)/\omega^2$ with $D(u) = \sqrt{\beta(u)^2 - 4\hat{\alpha}(u)\gamma}$. The solution for B_v thus equals:

$$B_v(u,\tau) = \frac{\beta(u) - D(u)}{\omega^2} \frac{1 - e^{-D(u)\tau}}{1 - G(u)e^{-D(u)\tau}} \tag{3.9}$$

where $G(u) = \dfrac{\beta(u) - D(u)}{\beta(u) + D(u)}$ represents the ratio of the two roots b and a. This quantity will play a great role in the remainder of the chapter. The solution to A now follows from:

$$\begin{aligned} \int_0^\tau B_v(s)ds &= \frac{\beta - D}{\omega^2} \int_0^\tau \frac{1 - e^{-Ds}}{1 - Ge^{-Ds}} ds = \frac{\beta - D}{\omega^2 D} \int_G^{Ge^{-D\tau}} \frac{z/G - 1}{z(1-z)} dz \\ &= \frac{\beta - D}{\omega^2 D} \cdot \left[\frac{(G-1)\ln(z-1) - G\ln z}{G} \right]_{z=G}^{z=Ge^{-D\tau}} \\ &= \omega^{-2} \left((\beta - D)\tau - 2\ln\left(\frac{Ge^{-D\tau} - 1}{G - 1} \right) \right) \end{aligned} \tag{3.10}$$

Clearly there are numerous ways of writing the characteristic function, and over the years many different formulations have been used. It turns out to be fundamentally important which formulation one uses for the function $A(u,\tau)$, in particular what one keeps under the logarithm in (3.10) and what one takes out of it. The formulation we have now derived is what we will refer to as the second formulation, as it is different to the original formulation in Heston [1993]. To the best of our knowledge this formulation has first appeared in Bakshi, Cao and Chen [1997] [6], and later in e.g. Duffie, Pan and Singleton [2000] and Gatheral [2006]:

Formulation 2:

$$A(u,\tau) = \kappa\theta\omega^{-2}\left((\beta(u) - D(u))\tau - 2\ln\psi_2(u,\tau) \right), \quad \psi_2(u,\tau) = \frac{G(u)e^{-D(u)\tau} - 1}{G(u) - 1} \tag{3.11}$$

The first formulation is the original one used by Heston, and also appears in e.g. the articles of Lee [2004] and Kahl and Jäckel [2005]:

Formulation 1:

$$A(u,\tau) = \kappa\theta\omega^{-2}\left((\beta(u) + D(u))\tau - 2\ln\psi_1(u,\tau) \right), \quad \psi_1(u,\tau) = \frac{c(u)e^{D(u)\tau} - 1}{c(u) - 1} \tag{3.12}$$

where we introduced $c(u) = 1/G(u)$. Though both formulations are algebraically equivalent, it is well-known that formulation 1 causes discontinuities when the principal branch of the complex logarithm is used, whereas for formulation 2 this turns out not to be the case. As an example of yet another formulation we mention Zhu [2000]. As this formulation too causes discontinuities when the principal branch of the logarithm is used, we restrict ourselves to these two formulations. The next subsection discusses the complex discontinuities caused by formulation 1.

[6] Although our formulation appears slightly different than that of Bakshi et al. and Duffie et al., the term under the logarithm is actually equivalent.

3.1.2. Complex discontinuities

An easy way to avoid any complex discontinuities is to integrate the ODEs in (3.6) numerically, as this would automatically lead to the correct and continuous solution. A comparative advantage of the Heston model is however that it has a closed-form characteristic function, something which significantly reduces the computational effort and would be forsaken if we proceeded in this way. From a computational point of view it is therefore certainly worthwhile to investigate how to avoid discontinuities in the closed-form solution.

By taking a closer look at the characteristic function, it is clear that two multivalued functions are present, both of which could cause complex discontinuities. The first candidate is the square root used in D(u). It turns out that the characteristic function is even in D, so that we will from hereon use the common convention that the real part of the square root is nonnegative. The complex logarithm used in (3.11) and (3.12) is the second candidate. Let us recall that the logarithm of a complex variable z can be written as:

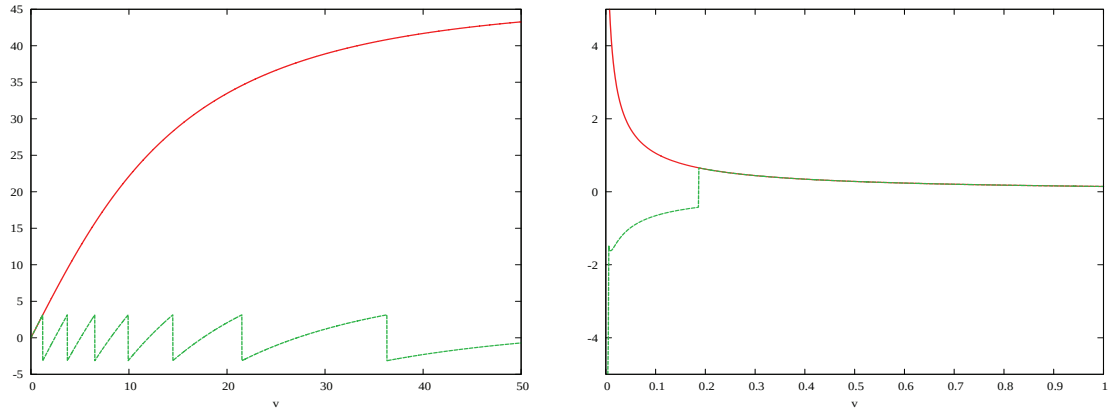
$$\ln z = \ln |z| + i(\arg(z) + 2\pi n) \tag{3.13}$$

where arg(z) is the argument of the complex number, |z| is its radius and $n \in \mathbb{Z}$. A typical choice used by most software packages is to restrict the logarithm to its principal branch, by letting arg(z)[7] be the principal argument, $\arg(z) \in [-\pi,\pi)$, and setting n = 0. In this case the branch cut of the complex logarithm is $(-\infty, 0]$, and the complex logarithm is discontinuous along it. It is well-documented that by restricting the complex logarithm to its principal branch formulation 1 will yield discontinuities in the characteristic function and hence in the option price. For any set of parameters, unless of course $2\kappa\theta/\omega^2 \in \mathbb{N}$, problems will arise if τ is sufficiently large, and in fact the number of discontinuities will grow with the time to maturity. A first mention of this problem was made by Schöbel and Zhu [1999], who encountered the same problem in the implementation of their own stochastic volatility model. They mention: "*Therefore we implemented our formula carefully keeping track of the complex logarithm along the integration path. This leads to a smooth CF...*". It seems that over the past years this approach was, and perhaps still is, best market practice when it comes to the implementation of stochastic volatility models. The problem with this approach is that it requires a very fine integration grid in order to be sure that no discontinuities can arise. Even then, one is not sure that the integration routine has singled out and corrected for all discontinuities, as one does not know a priori how many discontinuities there are.

A significant improvement on this ad-hoc approach has been made by Kahl and Jäckel [2005], who came up with the rotation count algorithm. This algorithm ensures that the principal argument of ψ_1 in (3.12) is continuous. Most importantly, this algorithm is easily implementable and allows for any numerical integration scheme to be used to evaluate the option price in (3.1), hereby opening up myriads of possibilities for improving the efficiency of implementations of stochastic volatility models. We return to this algorithm in Section 3.3.

Though Kahl and Jäckel have in their paper documented extensively what can and will go wrong if one uses the principal argument of the complex logarithm in conjunction with formulation 1, it is good to stress this again in a realistic example. For this we take the parameters that Duffie et al. [2000] implied from market data of S&P500 index options. The left panel of Figure 3.1 shows the discontinuities we would get in the argument or phase of $\psi_1(v - i(\alpha+1))$ if we would use the principal branch only, compared with the argument obtained by applying the rotation count algorithm. In the right panel the impact of these discontinuities on the integrand in equation (3.1) are shown. Here we used the optimal α and transformed the integrand to the unit

[7] In the remainder arg(z) will denote the principal argument of z.



(A) (B)

Figure 3.1: Complex discontinuities in the Heston model
Parameters: $\kappa = 6.21$, $\omega = 0.61$, $\rho = -0.7$, $\theta = 0.019$, $v(0) = 0.010201$, $F = K = 1$, $\tau = 10$, $\alpha = 3.35861$
(A) The principal argument/phase of ψ_1 with (red solid line) and without (green dashed line) correction
(B) Integrand in equation (3.1) in formulation 1 (green dashed line) vs. formulation 2 (red solid line)

interval with a logarithmic transformation, as will be outlined in Chapter 4. Note that the values in the figures do not include the scaling by $1/\pi$. The chosen example is a 10 year at-the-forward European call, whose true price is 0.1676. The option price found from the green dashed line would have been 0.0119, a marked difference that will certainly not go unnoticed. At smaller maturities however, the differences may not stand out so clearly. With a maturity of 2.5 years there is only one discontinuity; the true option price is 0.0816 whereas the option price found by bluntly using the principal argument in formulation 1 yields 0.0839.

3.2. Why the principal branch can be used

In this section we prove the conjecture of Lord and Kahl [2006] that formulation 2 in (3.11) remains continuous when the complex logarithm is restricted to its principal branch. As mentioned, both Lord and Kahl [2006] and Albrecher et al. [2007] have provided partial proofs for this conjecture, though the combination of their results involves restrictions on both the parameters as well as on Im(u).

In order to discover which part of the puzzle is still missing, we will need to analyse the proofs used in both papers. Let us first introduce some notation. Throughout this chapter we will write $u = x + yi$ for $u \in \mathbb{C}$, with $x = \mathrm{Re}(u)$ and $y = \mathrm{Im}(u)$. For any complex valued number, say $z \in \mathbb{C}$, we will sometimes use the shorthand notation $z_r = \mathrm{Re}(z)$ and $z_i = \mathrm{Im}(z)$.

We only need to consider $u \in \Lambda_x$, the strip of regularity of the characteristic function for which $|\phi(u)| < \infty$, see also (2.17) and (2.18). Analysing the strip of regularity, or moment stability, entails analysing the range of $\zeta \in \mathbb{R}$ for which $\phi(-\zeta i) < \infty$. Clearly this range will be of the form (ζ_-, ζ_+). It can be seen that $\Lambda_x = \{u \in \mathbb{C} \mid -\mathrm{Im}(u) \in (\zeta_-, \zeta_+)\}$, as we then immediately have:

$$|\phi(u)| = \left|\mathbb{E}[e^{iu \ln S(T)}]\right| \le \mathbb{E}[|e^{iu \ln S(T)}|] = \phi(-\mathrm{Im}(u)\,i) < \infty \tag{3.14}$$

Though moment stability in the Heston model is dealt with in great detail in Andersen and Piterbarg [2007], we will require the following result in the proof of our conjecture.

Theorem 3.1
The characteristic function of the Heston model is analytic for $u \in \Lambda \subset \mathbb{C}$, where $u \notin \Lambda$ if $\zeta = -\mathrm{Im}(u)$ satisfies one of the two following conditions:

1. $G(-\zeta i)e^{-D(-\zeta i)\tau} = 1 \wedge D(-\zeta i) \neq 0$ (3.15)
2. $D(-\zeta i) = 0 \wedge \beta(-\zeta i) \neq 0 \wedge \tau = -2\beta(-\zeta i)^{-1}$ (3.16)

Proof:
Analysing the strip of regularity boils down to analysing the stability of the system of ODEs in equation (3.6) for $u = -\zeta i$. As the function A will simply be an integral over B_v, the stability of B_v is what matters. The solution in (3.9) is clearly stable if and only if the denominator is not equal to zero, modulo those cases where both the numerator and the denominator are zero and (3.9) remains well-defined. Let us first check those cases when the numerator equals zero, namely $\beta = D$ or $D = 0$. In the first case, the denominator will be equal to 1 and we have $B_v(u,\tau) = 0$. In the last case we need one application of l'Hôpital's rule to find:

$$\lim_{D(u)\to 0} B_v(u,\tau) = \lim_{D\to 0} \frac{e^{-D\tau}\tau(\beta - D) - (1 - e^{-D\tau})}{\omega^2 e^{-D\tau}\frac{2\beta+\beta^2\tau-\tau D^2}{(\beta+D)^2}} = \frac{\beta^2\tau}{2\gamma(2+\beta\tau)} \tag{3.17}$$

so that (3.9) remains well-defined, provided that $\tau \neq -2\beta(u)^{-1}$ whenever $D(u) = 0$. This can only happen when $\mathrm{Re}(u) = 0$, the situation we have here. Combined with the condition following from the denominator of (3.9) we arrive at the conditions (3.15)-(3.16). Note that $\Lambda_x \subset \Lambda$.

Remark 3.1
For $\zeta \in \Lambda_z$ where $D(-\zeta i) = 0$ and $\beta(-\zeta i) < 0$ we have $\tau \leq -2\beta(-\zeta i)^{-1}$. When $\tau = -2\beta(-\zeta i)^{-1}$, Theorem 3.1 states that $\zeta \notin \Lambda$, so that certainly $\phi(-\zeta i)$ is infinite for $\tau > -2\beta(-\zeta i)^{-1}$.

Remark 3.2
Condition (3.15) is sufficient when $\rho \leq \kappa/\omega$, as is proven in Lord and Kahl [2006]. We note that this condition, together with the restriction that $\rho \leq \kappa/\omega$ also appears in Lee [2004, Appendix A.2], however without any hint as to how it can be derived.

Finally, we note that solving ζ_+ and ζ_- from (3.15) is not a well-posed problem, as there are an infinite number of solutions. To this end it is convenient to use the critical time analysed in Andersen and Piterbarg [2007]. In Chapter 4 good starting solutions are provided.

3.2.1. The proof of Lord and Kahl

The proof used by Lord and Kahl [2006] heavily hinges on the observation that $|G| \leq 1$ for a large range of parameter values. If this is the case, it is not difficult to show that ψ_2 cannot cross the negative real line. Before turning to the main result, we state the following lemmas.

Lemma 3.1
When $G(u) = 1$, ψ_1 and ψ_2 will never cross the negative real line.

Proof:
When $G(u) = 1$, we have $D(u) = 0$ and hence:

$$\lim_{D(u)\to 0} \psi_1(u,\tau) = \lim_{D(u)\to 0} \psi_2(u,\tau) = 1 + \tfrac{1}{2}\beta(u)\tau \tag{3.18}$$

One can check that $D(u) = 0$ can only occur when $Re(u) = 0$. If $\beta(u) \geq 0$ the result is immediately clear. For $\beta(u) < 0$ we can appeal to Remark 3.1 to conclude that $\tau \leq -2\beta(-\zeta i)^{-1}$ which precludes ψ_1 and ψ_2 from crossing the negative real line.

Lemma 3.2
If $\rho \leq \kappa/\omega$, or $Im(u) \geq y_2$ and $\kappa/\omega \leq \rho \leq 2\kappa/\omega$, we have:

$$\left| G(u) \right| = \left| \frac{\beta(u) - D(u)}{\beta(u) + D(u)} \right| \leq 1 \tag{3.19}$$

Proof: The proof for $\rho \leq 0$ is stated in the appendix. For a full proof we refer the interested reader to Lord and Kahl [2006].

The main result from Lord and Kahl now follows. It uses the result from Lemma 3.2 and proves that under these conditions ψ_2 can never cross the negative real line.

Theorem 3.2
Let $u \in \Lambda_x$, $\rho \leq \kappa/\omega$, or $Im(u) \geq -\kappa/(\rho\omega)$ and $\kappa/\omega \leq \rho < 2\kappa/\omega$. If we are evaluating the characteristic function by means of formulation 2 in (3.11), the principal branch of the complex logarithm is the right one.

Proof:
We have to prove that $\psi_2(u,\tau)$, defined in (3.11), never crosses the negative real line for the parameter combinations under consideration. Suppose that it does, i.e. that:

$$\psi_2(u,\tau) = \frac{G(u)e^{-D(u)\tau} - 1}{G(u) - 1} = -\xi \tag{3.20}$$

for some $\xi \geq 0$. We can assume that $G \neq 1$ as this case is covered by Lemma 3.1. Also, as $u \in \Lambda_x$ the numerator cannot equal zero when $D(u) \neq 0$, due to condition (3.15) in Theorem 3.1. Hence, we can assume that ξ is strictly larger than zero. Rearranging (3.20) yields:

$$G(\xi + e^{-D\tau}) = \xi + 1 \tag{3.21}$$

In view of ξ, τ, $D_r \geq 0$ and $|G| \leq 1$ we can take the modulus of the left-hand side:

$$|G(\xi + e^{-D\tau})| \leq |G| \cdot (\xi + e^{-D_r\tau}) \leq \xi + 1 \tag{3.22}$$

The first inequality is strict unless $D_i\tau = 2n\pi$ with $n \in \mathbb{Z}$, and the second inequality is strict unless $|G| = 1$ and $D_r = 0$. Equality in (3.21) can therefore only occur if $|G| = 1$, $D_r = 0$ and $D_i\tau = 2n\pi$. First, one can check that $|G| = 1$ implies $\beta_i = 0$ or $n = 0$. If $n = 0$, $G = 1$, which we excluded a priori. Second, D_r can only be zero when $u = yi$. This shows that $\psi_2 = 1$, contradicting (3.20).

The restrictions on the parameters in Lemma 3.2 and Theorem 3.2 are mainly formulated in terms of the correlation coefficient ρ as it characterises the main difference between stochastic volatility

models among various asset classes. Whereas in an equity or FX context ρ is used to fit to the skew or smile present in that market, the correlation parameter ρ is often set to zero when used in a term structure context. Though there is empirical evidence for this, see references in e.g. Andersen and Brotherton-Ratcliffe [2005], the main reason is of a practical nature. If ρ would be unequal to zero, a change of probability measure would change the structure of the stochastic volatility driver and in a term structure context would cause forward rates to appear in its drift. By setting ρ equal to zero this is avoided. A displacement coefficient is often added to the underlying order to be able to fit the implied volatility skew.

Negative values of ρ are typically required to be able to fit to a skew. Even if a positive value of ρ would be required, one typically finds $\kappa > \omega$ in an implied calibration, so that the condition on ρ in Lemma 3.2 and Theorem 3.2 is not in the least bit restrictive. For evidence of this in the literature we refer the reader to e.g. Bakshi, Cao and Chen [1997], Duffie, Pan and Singleton [2000] and Jäckel [2004].

3.2.2. Filling in the missing gaps

Whereas the proof of Lord and Kahl uses the observation that $|G| \leq 1$ for a large range of parameter values, Albrecher et al. [2007] assume that the characteristic function $\phi(u)$ is being evaluated in $u \in \mathbb{C}$ with $u = x + yi$ and $y < -1$, corresponding to using a positive α in (3.1). Not being able to choose Im(u) freely when valuing options via Fourier inversion would restrict the implementation of a robust option pricer, as we will show in Chapter 4.

With $y_2 = -\kappa/(\rho\omega)$ and $D(x + yi) = \sqrt{p(x, y) + q(x, y) \cdot i}$, where both p and q are functions on the real line, it can be shown that the proof of Albrecher et al. is split into five cases:

1. $\rho \leq 0$ and $q^{(1,0)}(x,y) \geq 0$
2. $\rho \leq 0$ and $q^{(1,0)}(x,y) < 0$
3. $\rho > 0$ and $y \geq y_2$
4. $\rho > 0$, $y < y_2$ and $q^{(1,0)}(x,y) < 0$
5. $\rho > 0$, $y < y_2$ and $q^{(1,0)}(x,y) \geq 0$

The proofs of Cases 1 and 3 use similar arguments to those used in Lemma 3.2. In the remaining cases it is proven that ψ_2 cannot be in either the second or third quadrant, and henceforth can never cross the negative real line.

Though the authors assume that $y < -1$, this assumption is not explicitly made clear in their proofs of the above cases. In fact, a closer look at their proof reveals that only Cases 1 and 3 use this assumption. First, in the proof of Case 1 it is stated that $\rho \leq 0$ and $q^{(1,0)}(x,y) \geq 0$ imply $y \leq y_2$. Though this is certainly true when $y < -1$, it is not true in general. Since the case where $\rho \leq 0$ is dealt with in full generality by our Theorem 3.2, encompassing Cases 1 and 2, we need not worry about this.

Second, in the proof of Case 3 the authors state that $\rho > 0$ and $y \geq y_2$ implies $q^{(1,0)}(x,y) \leq 0$. One can check that this is only true when $y \leq -1/2$. Hence we have $y_2 \leq y \leq -1/2$, implying that $\rho \leq 2\kappa/\omega$. Once again this case is dealt with in our Theorem 3.2. The case where $q^{(1,0)}(x,y)$ is larger than zero is an open problem, and is dealt with in the following lemma.

Lemma 3.3
When $\rho > 0$, $y \geq y_2$ and $q^{(1,0)}(x,y) > 0$, ψ_2 does not cross the negative real line.

Proof: See the appendix.

Finally, though Case 4 has a minor overlap with Theorem 3.2, both Cases 4 and 5 deal with the situation where $\rho > 2\kappa/\omega$, which is not at all covered by our previous work. Since the only open problem has been dealt with in Lemma 3.3, we are ready to prove the main theorem.

Theorem 3.3
Evaluating the Heston characteristic function by means of formulation 2, where we restrict the complex logarithm to its principal branch, ensuring that the characteristic function of the Heston model remains continuous.

Proof:
This immediately follows by combining the results of Theorem 3.2, Lemma 3.3 and Cases 4 and 5 of Albrecher et al. [2007].

3.3. Why the rotation count algorithm works

Having proven that no branch switching of the complex logarithm is required within formulation 2 of the Heston characteristic function, it seems that there is no longer a need for the rotation count algorithm of Kahl and Jäckel [2005]. In addition to its easier implementation, a further distinct advantage of formulation 2 is its numerical stability. Since D(x+yi) tends to $\omega\sqrt{1-\rho^2}\,x$ as x tends to infinity, its real part becomes quite large when calculating option prices. While this leads to numerical instabilities in the calculation of ψ_1 (formulation 1), as mentioned in Kahl and Jäckel's paper, ψ_2 (formulation 2) is much better behaved.

Nevertheless, from a theoretical perspective it is worthwhile to take a closer look at the rotation count algorithm, and show why it works. The rotation count algorithm automatically adapts the branch of the complex logarithm such that the characteristic function in formulation 1 is continuous, as it should be. For $a, b \in \mathbb{C}$ and $c \in \mathbb{R}$ the algorithm is basically concerned with the evaluation of d in:

$$d = ae^{b} + c \tag{3.23}$$

such that its argument is kept continuous if both the complex function b and the argument of a are continuous functions. In the algorithm we will use both the classical and polar representations of a complex number, i.e. for $z \in \mathbb{C}$ we will write $z = z_r + z_i i$ with $z_r = \mathrm{Re}(z)$ and $z_i = \mathrm{Im}(z)$, as well as $z = |z|\, e^{iz_\theta}$ with $|z|$ being its radius and z_θ its argument[8]. The algorithm follows as:

1. Calculate the phase interval of ae^b as $n = \left\lfloor \frac{1}{2\pi}(a_\theta + b_i + \pi) \right\rfloor$, equal to 0 if it is in $[-\pi, \pi)$;
2. $|d| = |ae^{b} + c|$
3. $d_\theta = \arg(ae^{b} + c) + 2\pi n$ where arg denotes the principal argument.

Algorithm 3.1: The rotation count algorithm of Kahl and Jäckel

The premise under which Algorithm 3.1 is valid is that the addition of the real number c does not change the phase interval of the resulting complex number. If this is indeed the case, we can use two successive applications of the rotation count algorithm to evaluate ψ_1 in formulation 1, see equation (3.12). First we evaluate $d_1(u,\tau) = c(u)e^{D(u)\tau} - 1$, subsequently $d_2(u) = c(u) - 1$, and

[8] Note that this is not necessarily the principal argument.

finally we write:

$$\psi_1(u,\tau) = \frac{|d_1(u,\tau)|}{|d_2(u)|} e^{i(d_{1\theta}(u,\tau) - d_{2\theta}(u))} \tag{3.24}$$

The following lemma shows that evaluating ψ_1 in this way leads to a continuous characteristic function, if the premise of the rotation count algorithm is valid.

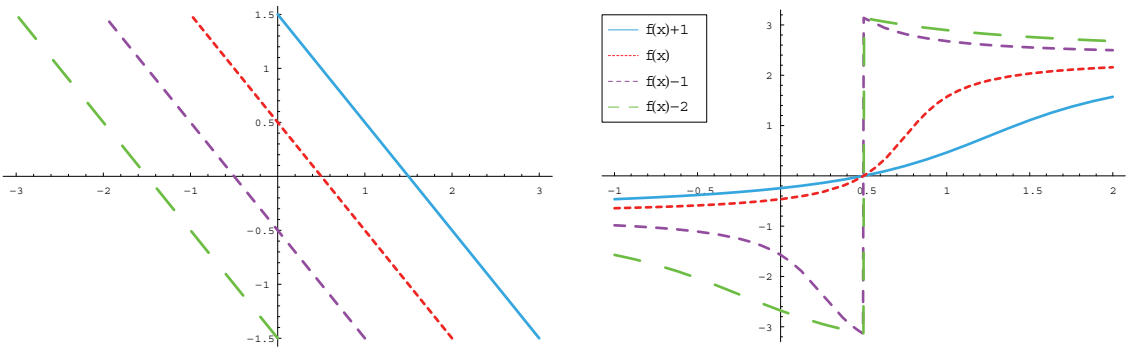
Lemma 3.4
Assume that the premise of the rotation count algorithm is valid and that u: $\mathbb{R} \to \Lambda_x$ describes a continuous path in the complex plane. Under these conditions applying the rotation count algorithm to evaluate the argument of $\psi_1(u,\tau)$ as $d_{1\theta}(u,\tau) - d_{2\theta}(u)$ yields a continuous characteristic function.

Proof:
Though the arguments $d_{1\theta}$ and $d_{2\theta}$ are not necessarily continuous, any discontinuities caused by arg(c) are cancelled out as they appear in both terms. The remaining term that could cause a discontinuity is D(u). If we write u = x + yi, one can check there are values of y where $\lim_{x\uparrow 0} D(x+yi) = \overline{\lim_{x\downarrow 0} D(x+yi)}$. Nevertheless this is not a problem either, as for u ∈ Λ_x the characteristic function is real on the imaginary axis, so that this discontinuity in D(u) does not have an impact on the characteristic function. Under the premise of the rotation count algorithm the assertion is thus true.

The remaining problem is now to check that the premise of the rotation count algorithm is true. Consider the following complex function:

$$f(x) = (1-x) - (\tfrac{1}{2} - x)i \tag{3.25}$$

where x ∈ $\mathbb{R}$. In the following figure we have drawn f(x)+1, f(x), f(x)-1 and f(x)-2 in the complex plane, as well as their principal arguments:



(A) (B)

Figure 3.2: Addition of a real number to a complex number can change the phase interval
(A) f(x)+1, f(x), f(x)-1 and f(x)-2 in the complex plane
(B) Argument of all four functions as a function of x

While f(x)+1 and f(x) do have a continuous principal argument, f(x)-1 and f(x)-2 clearly do not. The discontinuities here are caused by the fact that these functions make a transition from the second to the third quadrant, implying that the principal argument changes from the interval $[\frac{1}{2}\pi, \pi)$ to $[-\pi, -\frac{1}{2}\pi)$, clearly causing a jump. The following lemma formalises this observation.

Lemma 3.5
Consider a continuous function $z: \mathbb{R} \to \mathbb{C}$, where both the real and complex part of z are strictly monotone. Assume that z never passes through the origin[9]. Adding $y \in \mathbb{R}$ where $y \neq 0$ to z(x) does not make the principal argument of z(x) + y discontinuous as a function of x when compared to the principal argument of z(x), if and only if:

- $\mathrm{Re}(z(x)) \notin (-y,0)$ for $y > 0$ whenever $\mathrm{Im}(z(x))$ changes sign;
- $\mathrm{Re}(z(x)) \notin (0,-y)$ for $y < 0$ whenever $\mathrm{Im}(z(x))$ changes sign.

Proof: See the appendix.

Lemma 3.5 immediately gives necessary and sufficient conditions under which Algorithm 3.1 will work. Unfortunately, proving the rotation count algorithm works is more involved than the proof of Theorem 3.3. Whereas in the proof of Theorem 3.3 it was sufficient to check that ψ_2 never crossed the negative real line, we here have to simultaneously check that:

- whenever the imaginary parts of $c(u)$ or $c(u)e^{D(u)\tau}$ are zero, their real parts are not in [0,1];
- if the imaginary part of $c(u)e^{D(u)\tau}$ is zero and its real part is in [0,1], this must also be the case for $c(u)$, and vice versa.

Given that formulation 2 should in any case be preferred over the rotation count algorithm, we do not venture to provide a full proof. Nevertheless, it turns out that Lemmas 3.1 and 3.2 have given us enough machinery to prove that the rotation count algorithm works for almost all relevant parameter values. The following theorem is akin to Theorem 3.2.

Theorem 3.4
The rotation count algorithm can safely be applied to the Heston model as long as $u \in \Lambda_x$, $\rho \leq \kappa/\omega$, or $\mathrm{Im}(u) \geq -\kappa/(\rho\omega)$ and $\kappa/\omega \leq \rho < 2\kappa/\omega$.

Proof:
First note that $c = 1/G$. For the parameter combinations considered here we have $|c| \geq 1$ by virtue of Lemma 3.2. Furthermore:

$$|ce^{D\tau}| = |c|e^{D_r\tau} \geq |c| \geq 1 \tag{3.26}$$

since we use the convention that the real part of the square root is nonnegative and $\tau \geq 0$. If the inequality is strict, we can immediately conclude that whenever the imaginary parts of $c(u)$ or $c(u)e^{D(u)\tau}$ are zero their real parts are not in the interval [0,1]. When the inequality is an equality, we need only worry about the cases where G = -1 or G = 1. The first case is not a problem, whereas in the second case we need to evaluate ψ_1 as indicated in Lemma 3.1.

[9] Note that the principal argument of zero is undefined. As we only consider the characteristic function on its strip of regularity, it will always be well-defined and finite by virtue of (3.14).

3.4. Related issues in other models

Having analysed the Heston characteristic function in great detail, it is time to turn to other models. Firstly, we show that the Variance Gamma model does not suffer from any complex discontinuities, even though it contains the multivalued complex power function. Secondly, using our results from the Heston model, we show how to avoid any complex discontinuities in both the the Schöbel-Zhu model and the exact simulation algorithm of the Heston model. Potential issues in other extensions of the Heston model are deferred till Section 3.5.

3.4.1. The Variance Gamma model

To demonstrate that other models besides stochastic volatility models may have complex discontinuities, we turn to a model of the exponential Lévy class, the Variance Gamma (VG) model, first introduced by Madan and Seneta [1990]. Here the underlying asset is modelled as:

$$S(t) = F(t)\exp\big(\omega t + \theta G(t) + \sigma W\big(G(t)\big)\big) \tag{3.27}$$

where W(t) is a standard Brownian motion, G(t) is a Gamma process with parameter $\nu > 0$ and F(t) is the forward price of the underlying stock at time t. Without loss of generality we assume that σ is strictly positive. The parameter ω is chosen such that the expectation of (3.27) is F(t):

$$\omega = \tfrac{1}{\nu}\ln(1 - \theta\nu - \tfrac{1}{2}\sigma^2\nu) \tag{3.28}$$

To simplify notation we introduce f(t) = ln F(t) and $\widetilde{f}(t) = f(t) + \omega t$. If we denote τ as the time to maturity, the conditional characteristic function of the VG model is specified as:

$$\phi(u) = \mathbb{E}\big[e^{iu\ln S(T)}\big] = \frac{\exp\big(iu\,\widetilde{f}(T)\big)}{\big(1 - iu(\theta + \tfrac{1}{2}i\sigma^2 u)\nu\big)^{\tau/\nu}} \tag{3.29}$$

The ζ^{th} moment of the underlying asset exists as long as $\zeta \in (\zeta_-, \zeta_+)$, defined by:

$$\zeta_\pm = -\frac{\theta}{\sigma^2} \pm \sqrt{\frac{\theta^2}{\sigma^4} + \frac{2}{\nu\sigma^2}} \tag{3.30}$$

so that the extended characteristic function in (3.29) is well-defined for $u \in \Lambda_x$, its strip of regularity. As the VG model is fully time-homogeneous, the maximum and minimum allowed moments do not depend on the maturity T, in contrast with the situation in the Heston model. In the denominator of (3.29) we are using the complex power function, again a multivalued function. The complex discontinuities in the Heston model were in fact also caused by the branch switching of the complex power function. Although the characteristic exponent of the Heston model contains a complex logarithm, this term is multiplied by $-2\kappa\theta/\omega^2$ and subsequently its exponent is taken. In essence we are thus raising ψ_1 or ψ_2 (depending on which formulation we use) to the power of $-2\kappa\theta/\omega^2$, so that the branch switching of the complex power function is the cause of our complex discontinuities. As mentioned in Section 3.1.2, if $2\kappa\theta/\omega^2 \in \mathbb{N}$, there will be no discontinuities in either formulation if we restrict the logarithm to its principal branch.

If we here restrict the complex power function to its principal branch, i.e. if we evaluate z^{α} for $z \in \mathbb{C}$ and $\alpha \in \mathbb{R}$ as $|z|^{\alpha}\, e^{i\alpha \arg(z)}$, with arg(z) being the principal argument of z, the characteristic function of the VG model will only be continuous if $1 - iu(\theta + \frac{1}{2}i\sigma^2 u)\nu$ does not cross the negative real line. In the following theorem it is proven that this never occurs.

Theorem 3.5
When evaluating the characteristic function of the VG model in $u \in \Lambda_x$, we can safely restrict the complex power function to its principal branch.

Proof:
If we write u = x + yi, the imaginary part of $1 - iu(\theta + \frac{1}{2}i\sigma^2 u)\nu$ can only equal zero when either x = 0 or $y = \theta/\sigma^2$. For x = 0 we have:

$$1 - iu(\theta + \tfrac{1}{2}i\sigma^2 u)\nu = 1 + \nu\theta y - \tfrac{1}{2}\nu\sigma^2 y^2 \tag{3.31}$$

which for $u \in \Lambda_x$, or here $-y \in (\zeta_-, \zeta_+)$, is strictly positive. For $y = \theta/\sigma^2$ we find:

$$1 - iu(\theta + \tfrac{1}{2}i\sigma^2 u)\nu = 1 + \frac{\nu\theta^2}{2\sigma^2} + \tfrac{1}{2}\nu\sigma^2 x^2 \geq 1 \tag{3.32}$$

so that $1 - iu(\theta + \frac{1}{2}i\sigma^2 u)\nu$ can clearly never cross the negative real line. The principal branch of the complex power function is the correct one, as this is the only one that leads to real values for the moment generating function.

One can similarly check that the popular CGMY model, also known as the KoBoL or generalised tempered stable model, which contains the Variance Gamma model as a special case, also does not suffer from complex discontinuities in its original formulation.

3.4.2. The Schöbel-Zhu model

The first mention of discontinuities in characteristic functions caused by the branch switching of the complex logarithm or indeed the complex power function was in the article of Schöbel and Zhu [1999], who encountered these problems when implementing their extension of the Stein and Stein model to allow for non-zero correlation between the underlying asset and the stochastic volatility process. We will investigate whether, as in the Heston model, we can recast its characteristic function into a form suitable for the principal branch of the complex logarithm.

Under the risk-neutral pricing measure the underlying asset in the Schöbel-Zhu model evolves according to the following set of SDEs:

$$\begin{aligned} dS(t) &= \mu(t)S(t)dt + \sigma(t)S(t)\,dW_S(t) \\ d\sigma(t) &= -\kappa\big(\sigma(t) - \theta\big)dt + \omega\, dW_\sigma(t) \end{aligned} \tag{3.33}$$

where the Brownian motions satisfy $dW_S(t) \cdot dW_\sigma(t) = \rho\, dt$. The difference with the Heston model is that instead of the stochastic variance, now the stochastic volatility itself follows an Ornstein-Uhlenbeck process. A problem with the Schöbel-Zhu model, as noted by e.g. Jäckel [2004], is

that when the volatility process becomes negative, the sign of the instantaneous correlation between S and σ effectively changes. This is economically implausible.

Using the classification of Gaspar [2004] and Cheng and Scaillet [2007] one can conclude that the Schöbel-Zhu model is a linear-quadratic model in ln S and σ, and by the latter paper therefore equivalent to an affine model once we add the coordinate $v(t) = \sigma^2(t)$:

$$dv(t) = 2\sigma(t)d\sigma(t) + \omega^2 dt = \left(-2\kappa v(t) + 2\kappa\theta\sigma(t) + \omega^2\right)dt + 2\omega\sigma(t)dW_\sigma(t) \tag{3.34}$$

One can check that the model is certainly affine in ln S, σ and v, and its characteristic function will therefore have the same exponentially affine form as (3.5):

$$\phi(u) = \mathbb{E}\left[e^{iu\ln S(T)}\right] = \exp\left(iuf + A(u,\tau) + B_\sigma(u,\tau)\cdot\sigma(0) + B_v(u,\tau)\cdot v(0)\right) \tag{3.35}$$

where A, B_σ and B_v can be solved from the following system of ODEs:

$$\begin{aligned}
\frac{dB_v}{d\tau} &= \hat{\alpha}(u) - \beta(u)B_v + \gamma B_v^2 \\
\frac{dB_\sigma}{d\tau} &= 2\kappa\theta B_v + (-\tfrac{1}{2}\beta(u) + \gamma B_v)\cdot B_\sigma \\
\frac{dA}{d\tau} &= \kappa\theta B_\sigma + \tfrac{1}{2}\omega^2 B_\sigma^2 + \omega^2 B_v
\end{aligned} \tag{3.36}$$

subject to the initial conditions $B_v(u,0) = B_\sigma(u,0) = A(u,0) = 0$. The auxiliary variables are similar to the ones defined in the Heston model, namely $\hat{\alpha}(u) = -\tfrac{1}{2}u(i+u)$, $\beta(u) = 2(\kappa - \rho\omega ui)$ and $\gamma = 2\omega^2$. Indeed, there are more similarities with the Heston model. Following remarks by both Heston and Schöbel-Zhu, we know that when $\theta = 0$, the Schöbel-Zhu model collapses to a particular instance of the Heston model as can be seen from equation (3.34) – the variance then has a mean-reversion speed of 2κ, a volatility of variance equal to 2ω and a mean-reversion level of $\omega^2/2\kappa$. If we denote the Heston characteristic function as $\phi_H(u, S(0), v(0), \kappa, \omega, \theta, \rho, \tau)$, the Schöbel-Zhu characteristic function becomes:

$$\begin{aligned}
\phi_{SZ}(u, S(0), \sigma(0), \kappa, \omega, \theta, \rho, \tau) = \phi_H(u, S(0), \sigma(0)^2, 2\kappa, 2\omega, \omega^2/2\kappa, \rho, \tau) \\
\cdot \exp\left(A_\sigma(\tau) + B_\sigma(\tau)\cdot\sigma(0)\right)
\end{aligned} \tag{3.37}$$

where A_σ follows the ODE $\dfrac{dA_\sigma}{d\tau} = \kappa\theta B_\sigma + \tfrac{1}{2}\omega^2 B_\sigma^2$. By recognising that the characteristic function of the Schöbel-Zhu model can be expressed as an add-on on top of a special case of the Heston model, it is immediately clear that the discontinuities can be avoided in the same way as in the Heston model. Note that Schöbel and Zhu's original formulation of the characteristic function is different: the term under the complex logarithm is different to that found by using (3.37) in conjunction with either formulation 1 or 2. This explains why they would have had to correct for complex discontinuities when restricting the complex logarithm to its principal branch.

For completeness we provide the remainder of the characteristic function here. Tedious though straightforward manipulations show that A_σ and B_σ can be solved in closed-form as:

$$B_\sigma(u,\tau) = \kappa\theta\frac{\beta - D}{D\omega^2}\frac{(1-e^{-\frac{1}{2}D\tau})^2}{1-Ge^{-D\tau}}$$
$$A_\sigma(\tau) = \frac{(\beta-D)\kappa^2\theta^2}{2D^3\omega^2}\left(\beta(D\tau-4)+D(D\tau-2)+\frac{4e^{-\frac{1}{2}D\tau}\left(\frac{D^2-2\beta^2}{\beta+D}e^{-\frac{1}{2}D\tau}+2\beta\right)}{1-Ge^{-D\tau}}\right) \tag{3.38}$$

with D defined as before in the Heston model.

3.4.3. The exact simulation algorithm of the Heston model

Though the Heston model was originally proposed in 1993, an exact simulation algorithm for the SDEs in (3.3) was not published until recently by Broadie and Kaya [2006]. It goes too far to outline the full algorithm in this chapter, though we return to it later on in Chapter 6. The crucial step of the algorithm is the simulation of the integrated square root process conditional upon its start and endpoint. As this distribution is not known in closed-form, Broadie and Kaya chose to simulate from it by inverting its cumulative distribution function, which is itself found by inversion of the characteristic function:

$$\phi(u) = \mathbb{E}\left[\exp\left(iu\int_s^t v(u)\,du\right)\,\middle|\,v(s),v(t)\right] = \frac{D(u)e^{-\frac{1}{2}(D(u)-\kappa)\tau}(1-e^{-\kappa\tau})}{\kappa(1-e^{-D(u)\tau})}$$
$$\cdot\exp\left(\frac{v(s)+v(t)}{\omega^2}\cdot\left[\frac{\kappa(1+e^{-\kappa\tau})}{1-e^{-\kappa\tau}}-\frac{D(u)(1+e^{-D(u)\tau})}{1-e^{-D(u)\tau}}\right]\right) \tag{3.39}$$
$$\cdot\frac{I_{\frac{1}{2}\nu-1}\left(\sqrt{v(s)\,v(t)}\cdot\frac{4D(u)e^{-\frac{1}{2}D(u)\tau}}{\omega^2(1-e^{-D(u)\tau})}\right)}{I_{\frac{1}{2}\nu-1}\left(\sqrt{v(s)\,v(t)}\cdot\frac{4\kappa e^{-\frac{1}{2}\kappa\tau}}{\omega^2(1-e^{-\kappa\tau})}\right)}$$

with $D(u) = \sqrt{\kappa^2 - 2\omega^2 iu}$, the degrees of freedom $\nu = 4\kappa\theta/\omega^2$ and I_ν representing the modified Bessel function of the first kind. Finally, τ equals $t - s$. As the characteristic function in (3.39) depends non-trivially on the two realisations v(s) and v(t), it is not an easy task to precompute major parts of the calculations. As a result this step of the algorithm will be highly time-consuming. It is therefore not surprising that one of our findings in Chapter 6 is that several biased simulation schemes outperform the exact simulation scheme in terms of both speed and accuracy, even when the asset value is only required at one time instance. Nevertheless, the exact simulation method can be very useful as a benchmark.

Turning to (3.39), we note that its numerator contains a complex-valued modified Bessel function. Broadie and Kaya carefully tracked arg(z) when evaluating $I_\nu(z)$ in (3.39) , and changed the branch when necessary by means of the following continuation formula, cf. Abramowitz and Stegun [1972]:

$$I_\nu(ze^{m\pi i}) = e^{m\nu\pi i}I_\nu(z) \tag{3.40}$$

with m an integer value. To demonstrate it really is necessary to track the branch of I_ν, we define:

$$z(u) = \frac{D(u)e^{-\frac{1}{2}D(u)\tau}}{1 - e^{-D(u)\tau}} \tag{3.41}$$

which up to a scaling by $\frac{4}{\omega^2}\sqrt{v(s)\,v(t)}$ is the complex-valued argument of I_ν in (3.39). In Figure 3.3 we take the second parameter set Broadie and Kaya considered, and graph $\gamma(u) = z(u) / |z(u)|^{9/10}$ as a function of u. From Figure 3.3 it is clear that arg(z(u)) is discontinuous, as (the rescaled version of) z(u) repeatedly crosses the negative real line in this plot. The key issue is therefore to find a way to keep track of the correct branch of z(u). The insights of the Heston model allow us to do exactly this, as the following lemma shows.

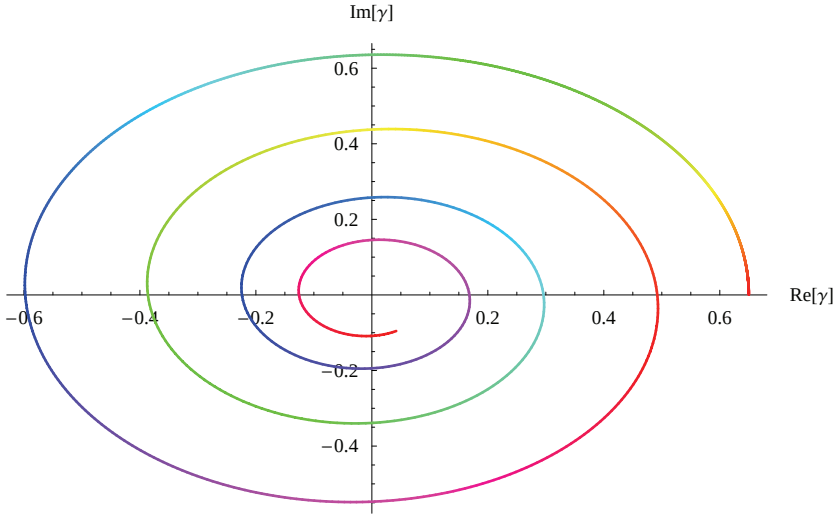


Figure 3.3: Plot of $\gamma(u) = z(u) / |z(u)|^{9/10}$ with hue function $\log_{10}(u+1)$ for $u \in [0, 100]$
Parameters from Broadie and Kaya [2006]: $\kappa = \omega = 1$, $\theta = 0.09$, $\tau = 5$

Lemma 3.6

The function $-\frac{1}{2}\tau\,\mathrm{Im}(D(u)) + \arg(f(u))$ with $f(u) = \frac{D(u)}{1 - e^{-D(u)\tau}}$, corresponding to the argument of z(u) in (3.41), is continuous for $u \in \Lambda_x = \{u \in \mathbb{C} \mid -\mathrm{Im}(u) \in (-\infty, (\kappa^2+4\pi^2)/2\omega^2)\}$.

Proof:
The strip of regularity can easily be checked from (3.39). That Im(D(u)) is continuous should be clear, so that it is sufficient to prove that f(u) never crosses the negative real line. We once again write u = x + yi. As $f(-\bar{u}) = \overline{f(u)}$ and f(yi) > 0 since $u \in \Lambda_x$, it is sufficient to focus on the case where x > 0. We will prove that the imaginary part of f(u) can never be positive, so that it will never cross the negative real line. First of all note that the sign of Im(D) coincides with the sign of $\mathrm{Im}(D^2) = -2\omega^2 x < 0$. To show that $\mathrm{Im}(f) \le 0$ is equivalent to proving:

$$d_i\tau e^{d_r\tau} - d_i\tau\cos(d_i\tau) - d_r\tau\sin(d_i\tau) \le 0 \tag{3.42}$$

Replacing $d_r\tau$ by $a \ge 0$ and $d_i\tau$ by $-b \le 0$ for notational convenience we find:

$$-be^{a}+b\cos(b)+a\sin(b)\le b\big(\cos(b)-1\big)+a\big(\sin(b)-b\big)\le 0 \tag{3.43}$$

where the first inequality followed by a first order expansion of the exponent, and the second inequality follows by noting that cos(b) – 1 ≤ 0 and sin(b) - b ≤ 0. Finally, note that when D(u) = 0, which happens only when Re(u) = 0 and Im(u) = $-\kappa^2/2\omega^2$, we have f(u) = exp(-τ). ∈

This is sufficient information if we use the power series to evaluate the modified Bessel function, as we can then evaluate $z(u)^k$ as exp(k ln z(u)), where the logarithm is kept continuous if it is evaluated as $\ln z(u) = -\frac{1}{2}D(u)\tau + \ln f(u)$, and ln f(u) is restricted to its principal branch. Nevertheless, there are alternative numerical methods available for evaluating the modified Bessel function, and we need to come up with a formulation that is independent of the chosen numerical algorithm. Theorem 3.6 provides us with such a formulation.

Theorem 3.6
The characteristic function is continuous for $u \in \Lambda_x$ if we evaluate it as:

$$\phi(u)\cdot\frac{\exp\big(\nu \ln z(u)\big)}{z(u)^{\nu}} \tag{3.44}$$

where φ(u) is evaluated by using the principal branch for the modified Bessel function, ln z(u) in the numerator is evaluated as sketched above, and the denominator uses the principal branch of the complex power function.

Proof:
Immediately follows by realising that the additional term in (3.44) is exactly the correction term in the continuation formula (3.40). □

To monitor the discontinuity of the characteristic function Broadie and Kaya would have had to use a very fine discretisation of the Fourier integral leading to the cumulative density function. The method we propose opens up the possibility of using arbitrary quadrature schemes, hereby speeding up their exact simulation algorithm considerably. Nonetheless, we expect that biased simulation schemes will remain the simulation schemes of choice, certainly after the arrival of the highly accurate schemes recently introduced by Andersen [2008].

3.5. Conclusions

In this chapter we have analysed the complex discontinuities which are found when evaluating the closed-form characteristic function of several popular option pricing models. Such discontinuities have first been documented in option pricing by Schöbel and Zhu [1999], and are, at least in the Heston and Schöbel-Zhu stochastic volatility models and their extensions, caused by the branch switching of the complex logarithm. Being unaware of these issues can lead to completely wrong option prices if we price European options by means of Fourier inversion.

When pricing options via Fourier inversion, the method which most practitioners seem to use to correct for these discontinuities is to carefully monitor the imaginary part of the complex logarithm and change its branch if a discontinuity is detected. Clearly this method is not foolproof, and moreover, highly inefficient. The only method to this date to guarantee a continuous characteristic function is to bypass the closed-form solution and numerically integrate the ordinary differential equation that gives rise to the complex logarithm. Unfortunately this

approach forsakes the comparative advantage of these option pricing models, precisely the fact that their characteristic function can be calculated in closed-form.

As a foolproof alternative, Kahl and Jäckel [2005] recently proposed their rotation count algorithm, which is an easily implementable algorithm that claims to be able to keep the complex logarithm in the Heston model continuous. In this article we have rigorously proven, under non-restrictive conditions on the parameters, that this is indeed the case. Under the same conditions we had already proven in Lord and Kahl [2006] that in an alternative formulation of the Heston characteristic function, which has appeared in e.g. Bakshi, Cao and Chen [1997], Duffie, Pan and Singleton [2000] and Gatheral [2006], the principal branch of the complex logarithm is the correct one. Since then other papers have appeared on this issue. Most notably Albrecher et al. [2007] have considered the second formulation under the restriction that the imaginary argument of the characteristic function is smaller than minus 1, corresponding to positive values of the damping coefficient α in Carr and Madan's option pricing formula, and have proven that in this case the principal branch is the correct one. While our proof and theirs do not overlap entirely, the union of both proofs does not cover all possible configurations of the Heston model. We analysed the cases that were still open, and filled in the missing gaps. This proves that with the second formulation we do not have to worry about the branch switching of the complex logarithm, and can stick with its principal branch. As this formulation is easier to implement than the rotation count algorithm, and in addition more numerically stable, it should be the preferred formulation.

With the lessons from the Heston model in hand, the remainder of this article investigates the complex discontinuities that arise in a selection of other models. First of all we show that although the Variance Gamma model involves the complex power function, its characteristic function can be evaluated by restricting the complex power function to its principal branch. Secondly, we have shown how to avoid complex discontinuities in both the Schöbel-Zhu model and the exact simulation algorithm of the Heston model proposed by Broadie and Kaya [2006].

Many other models may suffer from similar problems, but it is clearly beyond the scope of this chapter to consider them all here. In conclusion we will merely mention several extensions of Heston's model that are practically relevant. The first example we mention is the pricing of forward starting options in Heston's model, not via the bivariate integral of Kruse and Nögel [2005], but via the equivalent univariate integral of Hong [2004] and Lucic [2004], whose work demonstrates that we can use Carr-Madan's pricing formula to price these options, since the characteristic function of $\ln S(T) / S(t)$ for $t \leq T$ can be derived in closed-form. It appears that one can use our findings from Section 3.2 to construct a formulation in which we can restrict the complex logarithm to its principal branch. The remaining examples we discuss appear to be harder to analyse. One such example is, strangely enough, Heston's model with piecewise constant parameters, considered in e.g. Mikhailov and Nögel [2004]. Though it allows for a much greater flexibility when calibrating to market data, the fact that its characteristic function is solved by repeated application of the tower law of conditional expectation does not facilitate the analysis. Other examples are Matytsin's [1999] model, which we considered in Lord and Kahl [2006], and Duffie, Pan and Singleton's [2000] SVJJ model, which allows for correlated jumps in the asset and stochastic variance. Finally, in the joint characteristic function in the Heston model we have so far not been able to find a formulation in which we can safely restrict the complex logarithm to its principal branch. Though there are no options which directly depend on the latent stochastic volatility, the joint characteristic function may be of importance when pricing exotic options via lattice-based algorithms such as the CONV algorithm, as demonstrated in Chapter 5.

Appendix 3.A – Proofs

In this appendix we provide the proofs for lemmas 3.2, 3.3 and 3.5.

Lemma 3A.1
If $x, y \in \mathbb{C}$ and $y = \sqrt{x}$, with $y_r > 0$, we can write:

$$y_r = \sqrt{\tfrac{1}{2}x_r + \tfrac{1}{2}\sqrt{x_r^2 + x_i^2}} \qquad y_i = \frac{x_i}{2y_r} \tag{3A.1}$$

if we use the convention that the real part of the square root is positive.

Lemma 3A.2
For $x, y \in \mathbb{C}$ we have that if $\mathrm{Re}(x)\,\mathrm{Re}(\sqrt{y}) - \mathrm{Im}(x)\,\mathrm{Im}(\sqrt{y}) \geq 0$, then $x\sqrt{y} = \sqrt{x^2 y}$. If the condition is not satisfied, we have $x\sqrt{y} = -\sqrt{x^2 y}$.

Lemma 3A.3
For $x, y \in \mathbb{C}$ consider the following complex number:

$$z = \frac{x - \sqrt{x^2 + y}}{x + \sqrt{x^2 + y}} \tag{3A.2}$$

Its modulus is equal to 1 if $x = 0$ or $\mathrm{Re}(y/x^2) \leq -1$ and $\mathrm{Im}(y/x^2) = 0$, and less than 1 if:

$$\mathrm{Re}(\tfrac{1}{x})\,\mathrm{Re}(\sqrt{x^2 + y}) - \mathrm{Im}(\tfrac{1}{x})\,\mathrm{Im}(\sqrt{x^2 + y}) \geq 0 \tag{3A.3}$$

Proof:
If $x = 0$ we immediately have $z = -1$. Assume that (3A.3) holds true and that $x \neq 0$. Then:

$$z = \frac{x - \sqrt{x^2 + y}}{x + \sqrt{x^2 + y}} = \frac{1 - \frac{1}{x}\sqrt{x^2 + y}}{1 + \frac{1}{x}\sqrt{x^2 + y}} = \frac{1 - \sqrt{1 + \frac{y}{x^2}}}{1 + \sqrt{1 + \frac{y}{x^2}}} \tag{3A.4}$$

by virtue of lemma 3A.2. We can write this as $z = (1-u)/(1+u)$ with $u_r \geq 0$. Its modulus satisfies:

$$|z|^2 = \left|\frac{1-u}{1+u}\right|^2 = \frac{(1-u_r)^2 + u_i^2}{(1+u_r)^2 + u_i^2} \leq 1 \tag{3A.5}$$

with equality attained only when $u_r = 0$, i.e. if and only if $\mathrm{Re}(y/x^2) \leq -1$ and $\mathrm{Im}(y/x^2) = 0$. Clearly the modulus of z is larger than or equal to 1 if (3A.3) does not hold true.

The following proposition collects many properties of some functions that we require hereafter. All properties can be proven by using basic algebra, so that we omit the proof.

Proposition
Consider the following two functions:

$$\begin{aligned} p(x,y) &= \mathrm{Re}\left(D(u)^2\right) = \kappa^2 + \omega^2(1-\rho^2)x^2 - \omega(\omega - 2\kappa\rho)y - \omega^2(1-\rho^2)y^2 \\ q(x,y) &= \mathrm{Im}\left(D(u)^2\right) = \omega(\omega - 2\kappa\rho)x + 2\omega^2(1-\rho^2)xy \end{aligned} \tag{3A.6}$$

where we have introduced the convention that $u = x + yi$, for $x, y \in \mathbb{R}$. Furthermore, define:

$$y_1 = -\frac{\omega - 2\kappa\rho}{2\omega(1-\rho^2)} \qquad y_2 = -\frac{\kappa}{\rho\omega} \tag{3A.7}$$

Note that:
- when $\rho < 0$ or $\rho \geq 2\kappa/\omega$, $y_2 \geq y_1$;
- when $0 < \rho \leq 2\kappa/\omega$, $y_2 \leq y_1$.

For $x \geq 0$ the function p has the following properties:
- p is maximal w.r.t. y in y_1;
- For $y < y_1$, p is strictly increasing in y, for $y > y_1$, p is strictly decreasing in y;
- p is always strictly increasing in x;

Similarly, we can show that for $x \geq 0$ the function q has the following properties:
- $q(x,y_1) = 0$;
- q is positive and strictly increasing in x for $y > y_1$;
- q is negative and strictly decreasing in x for $y < y_1$;
- q is strictly increasing in y.

The following lemma is key to proving lemma 3.2. Its proof uses many of the previous properties.

Lemma 3A.4
For $x \geq 0$ the functions p and q defined in the previous proposition satisfy:

$$(\kappa + \omega\rho y)\left(p(x,y) + \sqrt{p(x,y)^2 + q(x,y)^2}\right) - \omega\rho x q(x,y) \geq 0 \tag{3A.8}$$

if in addition $\rho \leq \kappa/\omega$, or $y \geq y_2$ and $\kappa/\omega \leq \rho \leq 2\kappa/\omega$.

Proof:
The full proof is given in Lord and Kahl [2006], we only provide the proof for $\rho < 0$ and $\rho = 0$ here. Each case is divided into several sub cases, based on ranges for the variable y.

<u>Case 1: $\rho < 0$</u>

1a) $y \leq y_1 \leq y_2$
We can reshuffle:

$$(\kappa + \omega\rho y)\sqrt{p(x,y)^2 + q(x,y)^2} \geq -(\kappa + \omega\rho y)p(x,y) + \omega\rho x q(x,y) \tag{3A.9}$$

Since $y \leq y_2$ implies $\kappa+\omega\rho y \geq 0$, it is sufficient to square both sides and prove the resulting inequality. We obtain:

$$\begin{aligned} &\omega^2 x q(x,y) f(x,y) \geq 0 \\ &f(x,y) = \omega\rho(2\kappa - \omega\rho)(x^2+y^2) + \kappa^2(1+2y) \end{aligned} \tag{3A.10}$$

Since q is negative for $y \leq y_1$, we have to prove that $f(x,y) \leq 0$. The function f is maximal with respect to x for x = 0, and maximal w.r.t. y in:

$$y_3 \equiv \frac{\kappa^2}{-\omega\rho(2\kappa-\omega\rho)} \tag{3A.11}$$

Here we have $y_3 \geq 0$. Since $y_1 \leq 0$, it suffices to show that $f(0,y_1) < 0$. We have:

$$\frac{4\omega(1-\rho^2)^2}{\rho} f(0,y_1) = (2\kappa-\omega\rho)\left((2\kappa-\omega\rho)^2 + \omega^2(1-\rho^2)\right) \equiv g(\rho) \tag{3A.12}$$

Clearly, $g(\rho) > 0$ for $\rho < \min(2\kappa/\omega, 1)$. Since the left-hand side is negative for $\rho < 0$, $f(0,y_1) < 0$.

1b) $y_1 \leq y \leq y_2$

We still have $\kappa+\omega\rho y \geq 0$. If p is positive, the inequality is immediately seen to be true. If p is negative, we have $\sqrt{p(x,y)^2 + q(x,y)^2} \geq -p(x,y)$, so that the inequality clearly also holds.

1c) $y_1 \leq y_2 \leq y$

Here $\kappa+\omega\rho y \leq 0$. First note that in this region:

$$(\kappa+\omega\rho y)p(x,y) - \omega\rho x q(x,y) \geq 0 \tag{3A.13}$$

If p is negative, the proof is easy, since $\rho \leq 0$ and q is nonnegative. So let us assume that p is positive. Working out the function shows that x^2 is the only power of x in it, and its coefficient is:

$$\omega^2\left(\kappa(1+\rho^2) - \omega\rho(1+(1-\rho^2)y)\right) \tag{3A.14}$$

which is increasing in y, and positive for $y \geq 0$. Since $y_2 \geq 0$ in this region, the coefficient of x^2 is positive. Because p too is strictly increasing in x, it suffices to check the inequality for that x where p(x,y) = 0. But then the remaining inequality is $-\omega\rho x q(x,y) \geq 0$ which is immediately seen to be true. Hence, if we can prove the inequality from 1a) but now in reverse, we are done:

$$\omega^2 x q(x,y) f(x,y) \leq 0 \tag{3A.15}$$

Since q is here positive, it remains to show that f(x,y) is negative. As in 1a), f is maximal w.r.t. x in x = 0 and maximal w.r.t. y in y_3. In this region $y_2 \geq y_3$, so that it is sufficient to check that $f(0,y_2) \leq 0$. It turns out that $f(0,y_2) = 0$, which concludes the proof of case 1.

Case 2: ρ = 0

When ρ = 0, the inequality can be reduced to $p(x,y)+\sqrt{p(x,y)^2+q(x,y)^2}\geq 0$. Using the rationale of 1b) it is clear that this is true.

We now have the necessary machinery to prove lemma 3.2.

Lemma 3.2
If $\rho \leq \kappa/\omega$, or $Im(u) \geq y_2$ and $\kappa/\omega \leq \rho \leq 2\kappa/\omega$, we have:

$$\left|G(u)\right| = \left|\frac{\beta(u)-D(u)}{\beta(u)+D(u)}\right| \leq 1 \tag{3A.16}$$

Proof:
As before, we will write u = x + yi here. It is fairly easy to show that $G(-\bar{u}) = \overline{G(u)}$, so that it suffices to focus on the case $x \geq 0$. Since $D^2 = \beta^2 + \omega^2 u(i+u)$, G(u) is of the form treated in lemma 3A.3. Let us therefore first assume that $\beta \neq 0$. The condition from lemma 3A.3 which guarantees that (3A.16) holds, is then:

$$Re(\beta)\,Re(D) + Im(\beta)\,Im(D) \geq 0 \tag{3A.17}$$

which we obtain from (3A.3) by multiplying with $|\beta|^2$.

Case 1: x ≥ 0, β ≠ 0, Re(D) = 0
Lemma 3A.1 shows us that:

$$Re\big(D(u)\big) = \sqrt{\tfrac{1}{2}p(x,y)^2 + \tfrac{1}{2}\sqrt{p(x,y)^2+q(x,y)^2}} \tag{3A.18}$$

If Re(D) = 0 we must therefore have $q = Im(D^2) = 0$ and $p = Re(D^2) \leq 0$. Since $D = \sqrt{p+qi}$, it is clear that we then have $D = i\sqrt{-p}$, and (3A.17) becomes:

$$-\omega\rho x\sqrt{-p(x,y)} \geq 0 \tag{3A.19}$$

Clearly q can only be zero if x = 0 or if $y = y_1$. Tedious but straightforward algebra shows that p is strictly positive when $y = y_1$ and x > 0, so that we can conclude that x = 0, and that (3A.17) always holds in case 1.

Case 2: x ≥ 0, β ≠ 0, Re(D) > 0
Since Re(D) > 0, we can multiply (3A.17) by 2Re(D) and apply lemma 3A.1 to obtain:

$$(\kappa+\omega\rho y)\Big(p(x,y)+\sqrt{p(x,y)^2+q(x,y)^2}\Big) - \omega\rho x q(x,y) \geq 0 \tag{3A.20}$$

This inequality is proven in lemma 3A.4 under the conditions which we impose on ρ and y.

Case 3: $x \geq 0$, $\beta = 0$
When $\beta = 0$, lemma 3A.3 states $|G| = 1$. We can only have $\beta = 0$ when $\rho \neq 0$, $x = 0$ and $y = y_2$.

This concludes the proof.

Lemma 3.3
When $\rho > 0$, $y \geq y_2$ and $q^{(1,0)}(x,y) > 0$, ψ_2 does not cross the negative real line.

Proof:
We must have $x + yi \in \Lambda_x$, so that $\psi_2(yi)$ will not lie on the negative real line by construction. Since $\psi_2(-\bar{u}) = \overline{\psi_2(u)}$, it suffices to prove that $\psi_2(x+yi)$ will not cross the negative real line for values of $x \geq 0$. In addition we can impose $|G| > 1$ w.l.o.g., as for $|G| \leq 1$ we already know that ψ_2 can never cross the negative real line from Theorem 3.2. Since $q^{(1,0)}(x,y) > 0$ we have $y \geq y_1$, and one can check that both d_r and d_i are strictly positive. Also, $y \geq y_2$ implies $\kappa + \omega\rho y \geq 0$. Note that:

$$2\psi_2(u) = \beta \cdot \frac{1 - e^{-d\tau}}{d} + 1 + e^{-d\tau} \tag{3A.21}$$

We will prove that ψ_2 cannot lie in the second quadrant, which implies the negative real line can never be crossed. Let us define the following positive constants:

$$\begin{aligned} A &= d_i \omega\rho x - d_r(\kappa + \omega\rho y) \\ B &= d_r^2 + d_i^2 \\ C &= d_r \omega\rho x + d_i(\kappa + \omega\rho y) \end{aligned} \tag{3A.22}$$

Positivity of A follows from the fact that $|G| > 1$, so that the reverse of (3A.17) is true. First of all note that the imaginary part of ψ_2 equals (up to a positive scaling, namely $2|d|^2\exp(d_r\tau)$):

$$\mathrm{Im}(\psi_2) \propto -\sin(d_i\tau) \cdot (A + B) + C \cdot \left(\cos(d_i\tau) - e^{d_r\tau}\right) \tag{3A.23}$$

If the sine would be positive, the whole term would obviously be negative, and we would be finished. Hence, we can assume that the sine is negative. Let us suppose that ψ_2 lies in the second quadrant, implying that (3A.23) is positive, and hence:

$$B > \frac{C \cdot \left(\cos(d_i\tau) - e^{d_r\tau}\right)}{\sin(d_i\tau)} - A \tag{3A.24}$$

We will try to arrive at a contradiction by proving that $\mathrm{Re}(\psi_2) > 0$. Up to the same scaling:

$$\mathrm{Re}(\psi_2) \propto \sin(d_i\tau) \cdot C + A\left(\cos(d_i\tau) - e^{d_r\tau}\right) + B\left(e^{d_r\tau} + \cos(d_i\tau)\right) \tag{3A.25}$$

Since the coefficient of B is positive, we can invoke (3A.24) to bound this from below by:

$$\frac{-2Ae^{d_r\tau}\sin(d_i\tau) - C(e^{2d_r\tau} - 1)}{\sin(d_i\tau)} \tag{3A.26}$$

We claim that this is positive, which amounts to proving that:

$$2A\sin(d_i\tau) + C(e^{d_r\tau} - e^{-d_r\tau}) \geq 0 \tag{3A.27}$$

The first derivative of the left-hand side w.r.t. τ is:

$$2Ad_i\cos(d_i\tau) + Cd_r(e^{d_r\tau} + e^{-d_r\tau}) > Cd_r(e^{d_r\tau} + e^{-d_r\tau}) - 2Ad_i > 2Cd_r - 2Ad_i \tag{3A.28}$$

so that it is sufficient to prove that:

$$Cd_r - Ad_i = (d_r^2 - d_i^2)\omega\rho x + 2d_i d_r(\kappa + \rho\omega y) > 0 \tag{3A.29}$$

as then the left-hand side of (3A.27) is increasing in τ, and it is zero for $\tau = 0$, so that we can conclude that (3A.27) is true. When $d_r \geq d_i$ (3A.29) is obviously true. Dividing (3A.29) by $d_i^2 > 0$:

$$\frac{Cd_r - Ad_i}{d_i^2} = \left(\frac{d_r^2}{d_i^2} - 1\right)\omega\rho x + 2\frac{d_r}{d_i}(\kappa + \rho\omega y) > 0 \tag{3A.30}$$

Now, note that since $q(x,y) > 0$ we can write:

$$\frac{d_r}{d_i} = \frac{p(x,y)}{q(x,y)} + \sqrt{\frac{p(x,y)^2}{q(x,y)^2} + 1} \Rightarrow \frac{d_r}{d_i} = z_d \Leftrightarrow \frac{p(x,y)}{q(x,y)} = \frac{z_d^2 - 1}{2z_d} \tag{3A.31}$$

In order for d_r to be smaller than d_i, p/q must be negative. As for $y > y_1$ q is positive, this implies p must be negative. Now, let us solve the equation $p/q = z_{pq}$ for x. Note that $z_{pq} < 0$ whereas $0 \leq z_d < 1$. This yields two solutions:

$$x_{1,2} = z_{pq}(y - y_1) \pm \frac{\sqrt{h(y)}}{2\omega(1-\rho^2)} \tag{3A.32}$$

with:

$$h(y) \equiv \left(2z_{pq}\omega(1-\rho^2)(y - y_1)\right)^2 - 4(1-\rho^2)p(0,y) \tag{3A.33}$$

As $p(0,y) = p(x,y) - \omega^2(1-\rho^2)x^2$ and $p(x,y) < 0$, we must have $p(0,y) < 0$, so that h(y) is always positive. Note that in (3A.32) $z_{pq} < 0$ and $y > y_1$, implying that the first term is negative. Secondly, it is obvious that the smallest solution (x_1) is negative, whereas the largest solution (x_2) is positive. Disregarding the negative solution, we can rewrite inequality (3A.30) as:

$$\left(z_d^2 - 1\right)\omega\rho x_2 + 2z_d(\kappa + \rho\omega y) > 0 \tag{3A.34}$$

Now, consider $p(x_2, y)$. For which values of y is this negative? We have:

$$p(x_2, y) = p(0,y) + \frac{\left(2z_{pq}\omega(1-\rho^2)(y - y_1) + \sqrt{h(y)}\right)^2}{4(1-\rho^2)} \tag{3A.35}$$

Tedious algebra shows that the only two real zeroes of this equation coincide with those of p(0,y):

$$p(0,y) = 0 \Leftrightarrow y = y_1 \pm \frac{\sqrt{(2\kappa - \omega\rho)^2 + \omega^2(1-\rho^2)}}{2\omega(1-\rho^2)} \tag{3A.36}$$

Obviously only the positive zero has to be considered here. Let us call this zero y_3. It can be shown that $y_3 > y_2$. Also note that h(y) is minimal in y_1, and since $h(y_3) > 0$ and $y_3 > y_1$, h(y) is positive for $y > y_3$. Finally, since $\lim_{y \to \pm\infty} p(x_2, y) = \mp\infty$, we conclude that $p(x_2, y)$ is decreasing and negative on this domain.

We now have to investigate (3A.34) for $y \geq y_3$ and $0 < z_d < 1$ (or $z_{pq} < 0$). It turns out that (3A.34) is increasing in y. First of all, note that to show that (3A.34) is increasing in y is equivalent to proving that:

$$x_2'(y) < -z_{pq}^{-1} \tag{3A.37}$$

Expanding the left-hand side and rearranging yields:

$$\frac{2\omega(1-\rho^2)(1+z_{pq}^2)(y-y_1)}{\sqrt{h(y)}} < -z_{pq} - \frac{1}{z_{pq}} \tag{3A.38}$$

Since both sides are positive and h(y) > 0, we can bound h(y) in (3A.33) as:

$$h(y) > \left(2z_{pq}\omega(1-\rho^2)(y-y_1)\right)^2 \tag{3A.39}$$

and hence:

$$\frac{2\omega(1-\rho^2)(1+z_{pq}^2)(y-y_1)}{\sqrt{h(y)}} < \frac{2\omega(1-\rho^2)(1+z_{pq}^2)(y-y_1)}{-2z_{pq}\omega(1-\rho^2)(y-y_1)} = -z_{pq} - \frac{1}{z_{pq}} \tag{3A.40}$$

as we wanted to prove. Hence, it is sufficient to prove (3A.34) for $y = y_3$. Equating y to y_3 and multiplying by $(1-\rho^2)/z_d > 0$ yields:

$$\rho\sqrt{(2\kappa - \omega\rho)^2 + \omega^2(1-\rho^2)} > -(2\kappa - \rho\omega) \tag{3A.41}$$

Clearly, if $\rho < 2\kappa/\omega$ this is true. If the reverse is true, we can square both sides and from the assumption that $\rho > 2\kappa/\omega$ it follows that (3A.41) is true. Finally, we can conclude that (3A.34) and hence (3A.29) holds true for $d_r < d_i$ as well. This shows that if $\mathrm{Im}(\psi_2) > 0$, ψ_2 must lie in the first quadrant, so that ψ_2 can indeed never lie in the second quadrant. The negative real line can therefore never be crossed.

Lemma 3.5
Consider a continuous function z: $\mathbb{R} \rightarrow \mathbb{C}$, where both the real and complex part of z are strictly monotone. Assume that z never passes through the origin. Adding $y \in \mathbb{R}$ where $y \neq 0$ to z(x) does not add any discontinuities to the principal argument of z(x) + y when compared to the principal argument of z(x), if and only if:

- $\mathrm{Re}(z(x)) \notin (-y,0)$ for $y > 0$ whenever Im(z(x)) changes sign;
- $\mathrm{Re}(z(x)) \notin (0,-y)$ for $y < 0$ whenever Im(z(x)) changes sign.

Proof:
Let us first define the difference of the principal arguments of z(x) and z(x) + y as f(x):

$$f(x) = \arg(z(x)) - \arg(z(x) + y) \tag{3A.42}$$

Let a trajectory from quadrant i to quadrant j, without crossing any quadrants inbetween, be denoted as a tuple (i,j). The direction in which the trajectory is traversed does not matter here. Trajectories of z that do not cause any discontinuities are (1,2), (1,4), (3,4). A trajectory of z that does cause discontinuities is (2,3). Clearly the horizontal trajectories (1,2), (3,4) and vice versa can be neglected here, as Im(z(x)) does not change sign here. The diagonal trajectories (1,3) and (2,4) can also be excluded, as we assumed that z never passes through the origin. Finally, let x* be that x on the trajectory such that $\mathrm{Im}(z(x^*)) = 0$.
Let us start with the trajectory (1,4). Evidently the trajectory of z(x) + y remains (1,4), provided that $\mathrm{Re}(z(x^*)) + y > 0$. However, if $\mathrm{Re}(z(x^*)) + y < 0$, the trajectory will pass through the origin in an infinitesimal neighbourhood of x*. This will lead to a discontinuity in the principal argument, so we have to exclude this case. The same happens when $\mathrm{Re}(z(x^*)) + y = 0$.
If we start out with (2,3) as the trajectory of z(x), the same analysis leads to the requirement that $\mathrm{Re}(z(x^*)) + y < 0$, if we want to keep f(x) continuous. Collecting the results we find that f(x) remains continuous provided that:

- If $\mathrm{Re}(z(x^*)) < 0$, y must satisfy $\mathrm{Re}(z(x^*)) + y < 0$;
- If $\mathrm{Re}(z(x^*)) > 0$, y must satisfy $\mathrm{Re}(z(x^*)) + y > 0$.

This result is slightly rephrased in the lemma, so this concludes the proof.

Chapter 4

Optimal Fourier inversion in semi-analytical option pricing[10]

In recent years Fourier inversion has become the computational method of choice for plain vanilla option pricing in models with closed-form characteristic functions, such as the affine models from Chapter 2. As mentioned, if a model is to be used for the pricing of exotic derivatives, it is highly desirable if a fast and accurate calibration to plain vanilla option prices is possible. Numerical problems often arise when the time to maturity is short or the option is far out-of or in-the-money, as the integrands are then highly oscillating or strongly peaked. In this chapter we take into account such numerical issues and propose a solution, allowing for fast and robust option pricing for virtually all levels of strikes and maturities.

In this chapter we focus on the Carr-Madan/Lewis representation of the forward option price as a Fourier integral, as discussed in Section 2.4, cf. (2.31)-(2.35) and also (3.1)-(3.2). We repeat it here for convenience:

$$\mathbb{E}\left[(S(T)-K)^{+}\right] = R(F,K,\alpha) + \frac{1}{\pi}\mathrm{Re}\int_0^{\infty} e^{-i(v-i\alpha)k}\,\frac{\phi\left(v-i(\alpha+1)\right)}{-\left(v-i(\alpha+1)\right)(v-i\alpha)}\,dv \tag{4.1}$$

The residue term follows from Cauchy's residue theorem:

$$R(F,K,\alpha) = F\cdot 1_{[\alpha\le 0]} - K\cdot 1_{[\alpha\le -1]} - \tfrac{1}{2}\left(F\cdot 1_{[\alpha=0]} - K\cdot 1_{[\alpha=-1]}\right) \tag{4.2}$$

To simplify further notation, we define:

$$\psi(v,\alpha) = \mathrm{Re}\left(\frac{e^{-ivk}\varphi(v-i(\alpha+1))}{-(v-i\alpha)(v-i(\alpha+1))}\right) \tag{4.3}$$

As shown in (2.34), the integral in (4.1) can be written as a contour integral in the complex plane. By shifting the contour, or, if you will, changing the damping coefficient α in (4.1), various parity relations are obtained. The restriction that α must satisfy is (cf. (2.33)):

$$\left|\phi(v-(\alpha+1)i)\right| \le \phi(-(\alpha+1)i) = \mathbb{E}[S(T)^{\alpha+1}] < \infty \tag{4.4}$$

i.e. $v-(\alpha+1)i$ for $v \in \mathbb{R}$ should lie in the strip of regularity Λ_x.

In their seminal article Carr and Madan, who only used $\alpha > 0$, already reported numerical problems for strike prices far from the at-the-money level (ATM) and for short maturities. The reason is that the integrand in (4.1) becomes highly oscillatory or peaked, and hence difficult to

[10] This chapter has appeared as Lord, R. and C. Kahl [2007]. "Optimal Fourier inversion in semi-analytical option pricing", *Journal of Computational Finance*, vol. 10, no. 4, pp. 1-30.

integrate numerically. As an alternative they considered taking a Fourier transform of out-of-the-money option prices, though this is also prone to numerical difficulties for short maturities. Andersen and Andreasen [2002] suggested another approach to stabilise the numerical Fourier inversion. They used the Black-Scholes model as a control variate by subtracting the Black-Scholes characteristic function from the integrand and adding the Black-Scholes price back to the equation. This approach would work perfectly well if the characteristic functions of both models would be close and if we would know an appropriate volatility level for the Black-Scholes model. Nonetheless, this approach may yield better results than using (4.1) with a default value of α. Lee [2004] intensively discussed Carr and Madan's approach, and proposed an algorithm to arrive at an optimal α in the situation where the Fourier integral in (4.1) is approximated by the discrete Fourier transform (DFT). The algorithm consists of maximising the sum of the truncation error and the discretisation error with respect to the parameters of the discretisation, as well as α. Although this approach seems to work quite well for the examples Lee considered, it is quite specifically tailored towards the use of the DFT. Secondly, for most characteristic functions the estimated truncation error is typically a very conservative estimate of the true truncation error, which will certainly affect the resulting α.

The truncation error can actually be completely avoided by transforming the infinite integral to a finite domain using the limiting behaviour of the characteristic function as shown by Kahl and Jäckel [2005] for the Heston model. This reduces the sources of error from two to one, just leaving the discretisation error. Transforming the range of integration in this way precludes the use of the FFT algorithm. Therefore we are not able to use the power of the FFT, but this is not a real issue here. Firstly, using a different value of α for each strike/maturity pair would also already preclude the use of the FFT. Secondly, when calibrating a model to quoted option prices one typically has quotes for just a couple of strikes and maturities. Using the FFT would require a uniform grid in the log-strike direction in (4.1). The strikes of the options to which we calibrate will typically not lie on this grid, so that an additional source of error is introduced when using the FFT: interpolation error. Combined with the fact that the FFT binds us to a uniform grid usually makes it favourable to use a direct integration of (4.1). In terms of practical reliability and robustness it therefore seems more appropriate to follow the lead by Kahl and Jäckel [2005] use an adaptive numerical integration scheme such as the adaptive Gauss-Lobatto scheme developed by Gander and Gautschi [2000]. This does not necessarily minimise the overall computational workload, but it certainly ensures that the results are sufficiently accurate.

The outline of this article is as follows. In Section 4.1, we summarise the characteristic functions of the models we want to discuss in the following. Their analytical features are analysed in order to be able to transform the integration domain to a finite interval. The very heart of this article follows, namely the appropriate choice of α in Section 4.2. It is shown that the optimal choice of α is related to the contour shift used in saddle-point approximations. The penultimate section gives some numerical results underlining the previous results. Finally, we conclude.

4.1. Characteristic functions and domain transformation

In this section we introduce the affine models we investigate in the remainder of this chapter. We provide their characteristic functions and analyse their limiting behavior, in order to be able to transform the integration domain to a finite one. To keep things general, we analyse an affine jump-diffusion (AJD) stochastic volatility model, as well as a model of the exponential Lévy class: the variance gamma (VG) model. The model encompasses as special cases the models of Black–Scholes, Heston [1993], Stein and Stein [1991] and Schöbel and Zhu [1999], as well as their respective extensions to include jumps in the asset price.

4.1.1. Affine diffusion stochastic volatility model

The affine diffusion stochastic volatility model we consider here is characterised by the following two-dimensional system of stochastic differential equations (SDEs):

$$\begin{aligned} dS(t) &= rS(t)\,dt + \hat{\eta}\sigma(t)^p S(t)\,dW_S(t) \\ d\sigma(t) &= \kappa(\theta - \sigma(t))dt + \omega\sigma(t)^{1-p}\,dW_V(t) \end{aligned} \tag{4.5}$$

with correlated Brownian motions $dW_S(t)\,dW_V(t) = \rho\,dt$. For general values of p the model is not affine, though it is for p equal to 0, ½ or 1. For p = 0 we have the standard Black-Scholes model, whilst for p = ½ we obtain the Heston stochastic volatility model. In a model with a single underlying asset the effect of $\hat{\eta}$ can be fully subsumed by the parameters of the variance process, so that it is safe to assume that $\hat{\eta} = 1$. If the same stochastic volatility driver is used for multiple assets, $\hat{\eta}$ serves as a relative scaling to indicate how volatile each asset is compared to the other. An example of such a model is the stochastic volatility extension of the BGM/J market models due to Andersen and Andreasen [2002]. To simplify the notation in the following we assume $\hat{\eta} = 1$ for $p \neq 0$, though of course all results remain valid if $\hat{\eta}$ is unequal to 1. Finally, for p = 1 the model is equivalent to the Schöbel-Zhu model, itself a generalisation of the Stein and Stein model to allow for non-zero correlation between the volatility and the spot price. Though the Schöbel-Zhu model is often referred to as being a non-affine stochastic volatility model, it is actually affine in ln S(t), $\sigma(t)$ and $\sigma(t)^2$, as demonstrated in Section 3.4.2. As the characteristic function of affine models is exponentially affine in its state variables, we have:

$$\varphi_{\text{Affine}}(u) = e^{iuf + A(u,\tau) + B_\sigma(u,\tau)\,\sigma(0) + B_v(u,\tau)\,\sigma(0)^2} \tag{4.6}$$

where f = ln F, the logarithm of the forward price of the underlying asset. The functions A, B_σ and B_v satisfy the system of Ricatti equations in (3.36), leading to the solution in (3.37)-(3.38). Transformation of the integration domain requires the asymptotics of the different components which are given in the following proposition.

Proposition 4.1
Assuming that $\kappa, \theta, \omega, \tau > 0$ and $\rho \in (-1,1)$ we obtain the following asymptotics for the integrand of the Schöbel-Zhu model:

$$\lim_{u\to\infty} \psi(u,\alpha) \approx \psi(0,\alpha)\cdot e^{-uC_\infty} \cdot \text{Re}\left(-\frac{e^{iut_\infty}}{u^2}\right) = \psi(0,\alpha)\cdot e^{-uC_\infty} \cdot \frac{\cos(ut_\infty)}{-u^2} \tag{4.7}$$

with:

$$C_\infty = \tfrac{1}{4} D_\infty(\tau + V(0)/\omega^2) \qquad t_\infty = \tfrac{1}{4}\beta_\infty(\tau + V(0)/\omega^2) + \ln\tfrac{F}{K} \tag{4.8}$$

and auxiliary variables $D_\infty = 2\omega\sqrt{1-\rho^2}$ and $\beta_\infty = -2\omega\rho$. The proof is in Appendix 4A. □

Remark 4.1
Using the limiting behaviour of the Schöbel–Zhu model we can transform the integration domain using the transformation function $g(x) = - \ln x / C_\infty$, see Kahl and Jäckel [2005, equation (41)] for more details.

The characteristic function of the Heston model was derived in Section 3.1.1, see the preferred Formulation 2 in equations (3.5)-(3.11). Its asymptotics are due to Kahl and Jäckel [2005, proposition 3.1], which we repeat here.

Proposition 4.2
Assuming that $\kappa, \theta, \omega, \tau > 0$ and $\rho \in (-1,1)$ we obtain the following asymptotics for the integrand of the Heston model:

$$\lim_{u\to\infty} \psi(u,\alpha) \approx \psi(0,\alpha) \cdot e^{-uC_\infty} \cdot \frac{\cos(ut_\infty)}{-u^2} \tag{4.9}$$

These are the same asymptotics we found in (4.7), apart from:

$$C_\infty = \frac{\sqrt{1-\rho^2}}{\omega}\left(V(0) + \kappa\theta\tau\right) \qquad t_\infty = -\frac{\rho\left(\kappa\theta\tau + V(0)\right)}{\omega} + \ln\tfrac{F}{K} \tag{4.10}$$

Remark 4.2
The asymptotic behaviour of the Heston characteristic function is equivalent to that of the Schöbel–Zhu model, once we equate θ to ω^2/κ, and double κ and ω, a relation already pointed out by Lord and Kahl [2006] and in Chapter 3.

We conclude this section with an analysis of the easiest affine diffusion: the Black–Scholes model. Its characteristic function is given by:

$$\varphi(u) = \exp(iuf - \tfrac{1}{2}\hat{\eta}^2 u(u+i)\tau) \tag{4.11}$$

A closer look reveals that the characteristic function of the Black-Scholes model decays faster than the characteristic function of the stochastic volatility models.

Lemma 4.1
Assuming that the volatility $\hat{\eta} > 0$, the asymptotics of the Black-Scholes model are:

$$\lim_{u\to\infty} \psi(u,\alpha) \approx \psi(0,\alpha) \cdot e^{-u^2 C_\infty} \cdot \frac{\cos(ut_\infty)}{-u^2} \tag{4.12}$$

with:

$$C_\infty = \tfrac{1}{2}\hat{\eta}^2\tau \qquad t_\infty = \ln\tfrac{F}{K} + \tfrac{1}{2}\hat{\eta}^2\tau \tag{4.13}$$

When integrating the Black-Scholes model, the result in Lemma 4.1 tempts one to choose $g(x) = \sqrt{-C_\infty^{-1} \ln x}$ as the interval transforming function. However, using this transformation leads to instabilities for $x \to 1$. Luckily it turns out that $g(x) = -C_\infty^{-\frac{1}{2}} \ln x$ works surprisingly well

to transform the integration domain. From the tail behaviour of the stochastic volatility models it is already clear that the rate of decay of the characteristic function will increase when ω is decreased. In fact, for $\omega \to 0$ the affine diffusion model degenerates to the Black-Scholes model and we obtain the following.

Lemma 4.2
Consider the characteristic function of the Heston or Schöbel-Zhu model with limiting behaviour given by equation (4.7) of Proposition 4.1 and equation (4.9) of Proposition 4.2. We then obtain:

$$\lim_{\omega \to 0} e^{-uC_\infty} \frac{\cos(ut_\infty)}{-u^2} \tag{4.14}$$

4.1.2. AJD stochastic volatility model

To show how the analysis changes when jumps are included, we add jumps to the underlying asset. The analysis is much the same when jumps are added to the stochastic volatility driver. When jumps are added the SDE in (4.5) changes to:

$$dS(t)/S(t) = (r - \lambda\mu)\, dt + \hat{\eta} \cdot \sigma(t)^p\, dW_S(t) + J_{N(t)}\, dN(t) \tag{4.15}$$

where N is a Poisson process independent of the Wiener process W_S with intensity parameter λ such that $E[N(t)] = \lambda t$. The random variable J_i describes the size of the i^{th} jump. We will here only consider lognormally distributed jumps, although of course any jump size distribution could be used. With lognormal jump sizes:

$$\ln(1 + J_t) \sim N\left(\ln(1+\mu) - \tfrac{1}{2}\eta^2, \eta\right) \tag{4.16}$$

the model collapses to Merton's jump-diffusion model [1976] when p = 0, and to Bates' [1996] model when p = ½. As the Poisson process and the jump sizes are independent of the Brownian motions in the model, the characteristic function of the full model $\phi_{AJD}(u)$ with jumps can be found as the product of the characteristic function of the affine diffusion part and $\phi_{Affine}(u)$ and the characteristic function of the jump part $\phi_{Jump}(u)$:

$$\phi_{AJD}(u) = \phi_{Affine}(u) \cdot \phi_{Jump}(u) \tag{4.17}$$

For lognormal jump sizes, $\phi_{Jump}(u)$ is equal to:

$$\phi_{Jump}(u) = \exp\left(-\lambda\mu iu\tau + \lambda\tau\left((1+\mu)^{iu} \exp(\tfrac{1}{2}\eta^2 iu(iu-1)) - 1\right)\right) \tag{4.18}$$

It is straightforward to deduce that for u = x + iy, we have:

$$\lim_{x \to \infty} \phi_{Jump}(x + iy) \approx \exp(-\lambda\mu ix\tau) \tag{4.19}$$

which does not influence the asymptotic behaviour of $\phi_{AJD}(x+iy)$. All AJD models considered therefore allow us to transform the inverse Fourier integral to a finite domain which strongly simplifies the numerical computation of the semi-analytical option price.

4.1.3. VG model

In contrast to the AJD stochastic volatility models analysed thus far, the characteristic function of the VG model, cf. (3.29) in Section 3.4.1, decays only polynomially, also leading to a polynomial decay for the integrand in (4.1). For all transformation functions we tried, this polynomial decay, combined with the oscillatory nature of the characteristic function, causes the oscillations to bunch up at one end of the finite interval. Finding an appropriate transformation of the integration domain is therefore still an unsolved problem at the time of writing this thesis. Fortunately the easy analytical structure of (3.29) allows us to bound the characteristic function quite sharply, as has been done by Lee [2004]. Combining this approach with a suitable choice of α will still reduce the numerical difficulties of the Fourier inversion significantly.

4.2. On the choice of α

Now we come to the very heart of this chapter: the optimal choice of α. Taking a closer look at the representation of the option price in (2.27)-(2.28), we recognise that when using the Lévy inversion approach the range of option prices that can be calculated numerically is limited by cancellation errors. Indeed the crucial point for cancellation is the addition of $\frac{1}{2}(F - K)$ to the integral. Assuming that we use a highly sophisticated numerical integration scheme such as the suggested adaptive Gauss–Lobatto method, we can approximate the integrand up to a relative accuracy of 16 digits[11], which is the machine precision in IEEE 64-bit floating point arithmetic. Without loss of generality we set the forward value to $F(T) = 1$. Calculating OTM options with small times to maturity thus requires the subtraction of two values of almost the same size to obtain a much smaller option price. Cancellation is the consequence. Hence, within the original Heston parametrisation we are not able to compute option prices below the machine size precision, independent of the integration scheme that is used. Clearly this problem is not specific to the Heston model when using Equations (2.27)-(2.28) to recover option prices. Moreover, this problem will occur in almost every situation where a Fourier transform has to be numerically inverted.

The Carr–Madan.Lewis representation in (4.1), has several advantages. First of all, the number of numerical integrals is reduced from two to one. As an added benefit the denominator of the integrand is now a quadratic function in the integrating variable v, and as such decays faster than the integrands in (2.28). Finally, and most importantly for our purposes in this chapter, the Carr-Madan/Lewis representation allows us to split the problem of tiny option prices from the problem of restricted machine size precision since $\exp(-\alpha k)$ serves as a scaling factor. An appropriate choice of α enables us to find a scaling which allows us to calculate arbitrarily small option prices. Unfortunately, the situation is not as simple as that, since by changing α, the integrand can become either strongly peaked when getting close to the poles of the integrand, or highly oscillatory when reaching the maximum allowed α. It is therefore of substantial importance to have an appropriate choice of α.

Although the literature on mathematical finance has recognised the need for an appropriate choice of α, there has not been much work in this direction. Before discussing some recent studies, we mention a related article by Levendorskiĭ and Zherder [2002], which considers a different contour shift than that considered here. Although they successfully applied their technique to the KoBoL/CGMY (Koponen–Boyarchenko–Levendorskiĭ/Carr–Geman–Madan–Yor) model of order smaller than one, it turns out that their technique is not directly applicable to

[11] The relative accuracy is provided by the macro DBL_EPSILON in the C header <float.h>, which is the smallest positive number x such that $1 + x \neq 1$ in the computer's floating point representation.

the stochastic volatility models considered here. We return to this in the beginning of Section 4.2.2. As far as the damping factor α is concerned, Carr and Madan [1999] suggested the use of one quarter of the maximal allowed α whilst Raible [2000] recommended that, for the models he considered, choosing α equal to 25 works best. Schoutens et al. [2004] suggested using an α equal to 0.75. Clearly, only the suggestion of Carr and Madan will work for any given model, due to the restriction that the $(\alpha + 1)^{th}$ moment is finite in (4.4). For the ad-hoc choices of $\alpha = 25$ or 0.75 this need not be the case. The only rigorous study into an optimal choice of α has been performed by Lee [2004]. Lee suggests to minimise the sum of the discretisation error and the truncation error with respect to α and the quadrature parameters. In our approach, we transform the integration interval to a finite one, thus avoiding truncation error. Furthermore, by using adaptive quadrature we virtually avoid any discretisation error, so that there is nothing really left to minimise.

Nonetheless, choosing the right α can be very important, in particular for short maturities and/or strikes that are away from the ATM level. Figure 4.1 shows the relative error for different values of α. To generate this figure we used the adaptive Gauss–Lobatto scheme on the finite interval for different relative and absolute tolerance levels. It is obvious from Figure 4.1 that for certain contract parameters the outcome of the numerical integration is highly sensitive to the choice of α. Luckily there is a range of α which leads to the correct value, although the width of this range becomes increasingly narrow as the maturity of the option decreases and the strikes move away from the ATM level. Figure 4.2 compares the integrand of (4.1), i.e. $\exp(-\alpha k)\,\psi(v, \alpha)$, in the Heston model for various values of α. The integration domain $[0, \infty)$ has been transformed to the unit interval using the transformation advocated in Section 4.1. Furthermore, all integrands have been scaled by a constant such that their values in $x = 0$ coincide. Whilst choosing an α below the optimal one leads to oscillation at the left end of the interval, the overestimation close to the maximum allowed α makes the characteristic function oscillatory at the other end. Vice versa the optimal choice of α leads to an integrand which is neither peaked nor oscillating at all. Since the remaining shape of the optimal characteristic function differs depending on the parameter configuration, a standard saddlepoint approximation[12] would be too inaccurate. For that reason we suggest to estimate the optimal α and, in addition, use an adaptive quadrature scheme to obtain robust and accurate option prices. When possible, we also advocate transforming the integration domain to a finite one so that an analytical estimation of an appropriate upper limit of integration in (4.1) can be avoided.

4.2.1. Minimum and maximum allowed α

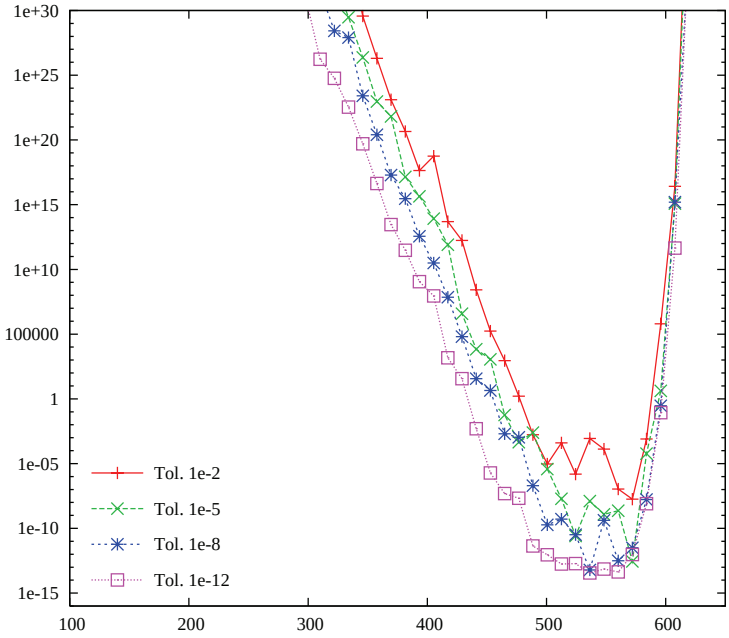
This section examines how to determine the strip of regularity for the characteristic function, a problem which has already been investigated at great lengths in Andersen and Piterbarg [2007]. Whilst Andersen and Piterbarg concentrate on finding the critical time T for which the ζ^{th} moment:

$$\mu(\zeta, \tau) = \mathbb{E}[\,S(T)^{\zeta}\,] \tag{4.20}$$

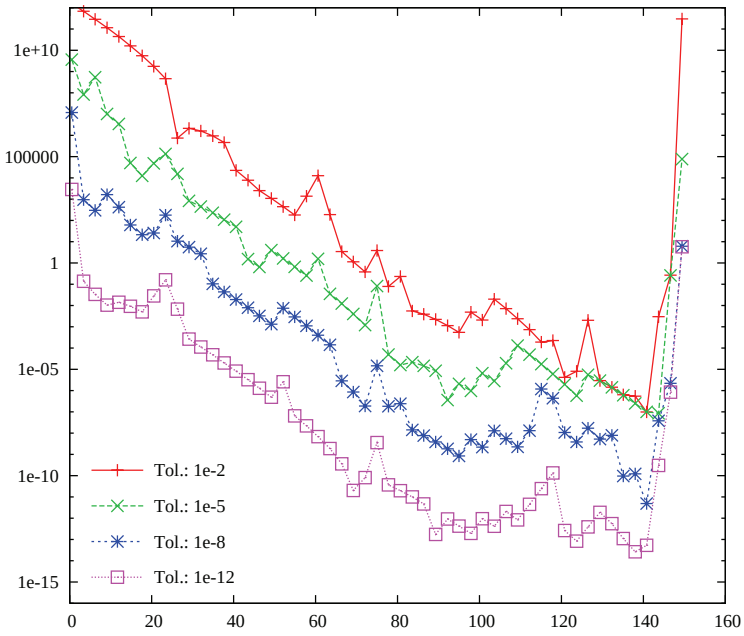
is still finite, we are interested in the whole strip of regularity, thus fixing T and finding all $\zeta \in \mathbb{R}$ such that $\mu(\zeta, \tau) < \infty$. Clearly this range will be of the form (ζ_-, ζ_+). The strip of regularity from equation (2.18) can therefore be written as:

[12] The reader who is not familiar with saddlepoint approximations can find an introduction to this topic in Section 4.2.3.

Figure 4.1: Relative pricing errors for the Heston model using the adaptive Gauss-Lobatto scheme for different absolute and relative tolerance levels over varying values of α.
Underlying: $dS(t)/S(t) = \mu\,dt + \sqrt{V(t)}\,dW_S(t)$ with S = F = 1 and $\mu = 0$.
Variance: $dV(t) = \kappa(\theta - V(t))dt + \omega\sqrt{V(t)}\,dW_V(t)$ with $V(0) = \theta = 0.1$, $\kappa = \omega = 1$ and $\rho = -0.9$.

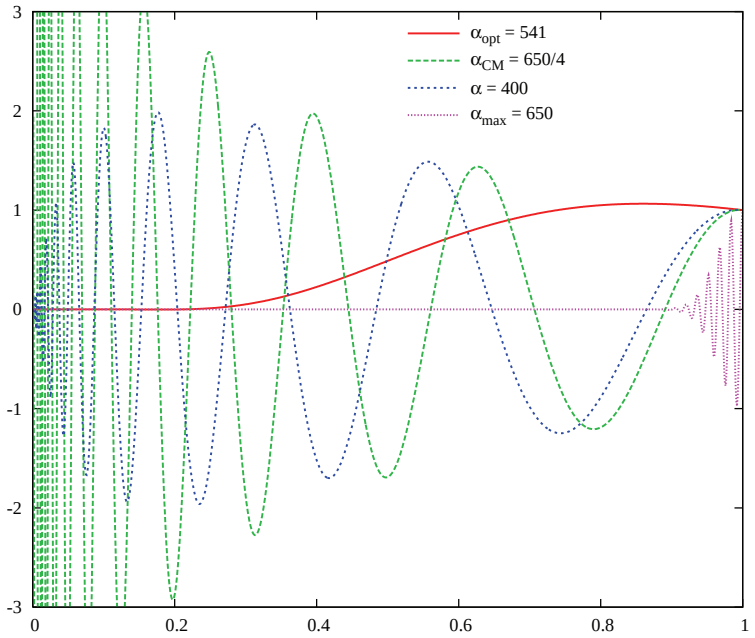


(A) $\tau = 1/52$ and K = 2, call value = $3.25 \cdot 10^{-126}$, optimal $\alpha = 541.93$;

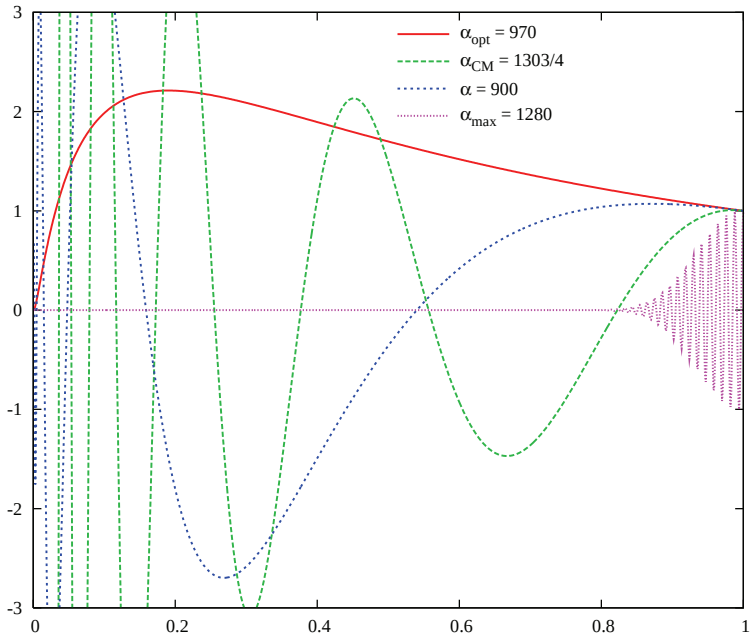


(B) $\tau = 1/12$ and K =1.5, call value = $1.1802 \cdot 10^{-17}$, optimal $\alpha = 121.24$.

Figure 4.2: Integrand in (4.1) for different values of α, where we used $g(x) = -\ln x / C_\infty$ as the transformation function and the asymptotics from Proposition 4.2 in Equations (4.9)-(4.10).
Underlying: $dS(t)/S(t) = \mu\,dt + \sqrt{V(t)}\,dW_S(t)$ with $S = F = 1$ and $\mu = 0$.
Variance: $dV(t) = \kappa(\theta - V(t))dt + \omega\sqrt{V(t)}\,dW_V(t)$ with $V(0) = \theta = 0.1$ and $\kappa = \omega = 1$.



(A) $\tau = 1/52$, $\rho = -0.9$ and $K = 2$



(B) $\tau = 1/360$, $\rho = -0.2$ and $K = 2.4$

$$\Lambda_x = \{u = x + iy \in \mathbb{C} \,|\, \zeta_- < -y < \zeta_+\} \tag{4.21}$$

For all $u \in \Lambda_x$ the finiteness of the characteristic function is guaranteed, since:

$$|\phi(u)| = |\mathbb{E}[e^{iu \ln S(\tau)}]| \leq \mathbb{E}[|e^{iu \ln S(\tau)}|] = \phi(iy) \tag{4.22}$$

The range $(\alpha_{min}, \alpha_{max})$ which contains all allowed values of α corresponds to $(\zeta_- - 1, \zeta_+ - 1)$, as the integrand $\psi(v, \alpha)$ requires an evaluation of $\phi(v - i(\alpha+1))$, see (4.1). In the remaining sections we give the strip of regularity for the models considered in Section 4.1.

4.2.1.1. Affine diffusion stochastic volatility model

The crucial point for explosions in the characteristic function of an affine diffusion stochastic volatility model is the Ricatti equation for $B_v(u, \tau)$ given by:

$$\frac{\partial B_v}{\partial \tau} = \hat{\alpha}(u) - \beta(u) B_v + \gamma B_v^2 \tag{4.23}$$

This equation holds for both the Heston and the Schöbel–Zhu model, with $\hat{\alpha}$, β and γ defined just below (3.36) in the Schöbel–Zhu model, whereas for the Heston model they are defined just below (3.6). When considering the moment stability of the Schöbel–Zhu model, we can therefore immediately use the results of the Heston model, as long as we double κ and ω. Following the lead by Andersen and Piterbarg [2007], who analysed moment explosions within the Heston model[13], one can estimate the critical value of ζ with almost the same argumentation they use to determine the critical time to maturity. It can be shown that $\mu(\zeta, T)$ given by Equation (4.20) is finite if $T < T^*$ and infinite for $T \geq T^*$, where T^* is given by one of three possibilities:

1. $D(-\zeta i)^2 \geq 0$, $\beta(-\zeta i) \geq 0$ or $\zeta \in [0, 1]$: $T^* = 0$ (4.24)

2. $D(-\zeta i)^2 \geq 0$, $\beta(-\zeta i) < 0$: $T^* = \frac{1}{c} \ln \frac{\beta - c}{\beta + c}$ (4.25)

3. $D(-\zeta i)^2 < 0$: $T^* = \frac{2}{c}\left(1_{[\beta>0]}\pi + \arctan(-\frac{c}{\beta})\right)$ (4.26)

where β is shorthand notation for $\beta(-\zeta i)$ and we introduced:

$$c = \tfrac{1}{2} |D(-\zeta i)| \tag{4.27}$$

As mentioned above, Andersen and Piterbarg analysed the case where $\zeta > 1$. For $\zeta \in [0, 1]$ it is clear that all moments will be finite, as we can use the inequality $S(T)^\zeta < 1 + S(T)$, so that the moment will be bounded from above by the finite first moment, plus 1. It can be shown that for $\zeta < 0$ the results of Andersen and Piterbarg for $\zeta > 1$ still remain valid. The strip of regularity Λ_x can then be calculated by fixing the critical time $T^* = \tau$ and solving the non-linear equation for α.

[13] In fact, they analysed the moment stability of stochastic volatility models to be able to indicate when the values of certain contracts requiring the existence of higher moments are finite. Examples of such contracts are Eurodollar futures, LIBOR-in-arrears and constant maturity swap (CMS) payments.

The efficiency of a numerical solution for a non-linear equation strongly depends on a good approximation for the starting value. Here we only discuss the approximation of the starting values for some practically relevant parameter configurations such as $\rho \approx -1$ for an equity model and $\rho = 0$ in the interest rate case. The critical time in the notation of Andersen and Piterbarg is closely related to the sign of $D(-\zeta i)^2$ which is a second-order polynomial in ζ:

$$D(-\zeta i)^2 = \kappa^2 + \omega(\omega - 2\kappa\rho)\zeta - \omega^2(1-\rho^2)\zeta^2 \tag{4.28}$$

with roots equal to:

$$\zeta_{D\pm} = \frac{\omega - 2\kappa\rho \pm \sqrt{(\omega - 2\kappa\rho)^2 + 4(1-\rho^2)\kappa^2}}{2\omega(1-\rho^2)} \tag{4.29}$$

Lemma 4.3
For $\rho \leq 0$, we have $\zeta_+ \geq \zeta_{D+} \geq 1$ and $\zeta_- \leq \zeta_{D-} < 0$.

Proof:
For $\rho \leq 0$ the zeroes of $D(-\zeta i)^2$ are not the critical moments, as is evident from the analysis of Andersen and Piterbarg [2007] and Lord and Kahl [2006]. Furthermore, $D(-i)^2 = (\kappa - \omega\rho)^2 \geq 0$ and $D(0)^2 = \kappa^2 > 0$.

Proposition 4.3
The maximum allowed ζ for a highly negative correlation $\rho \approx -1$ can be approximated by:

$$\zeta_+ \approx \frac{\omega - 2\kappa\rho \pm \sqrt{(\omega - 2\kappa\rho)^2 + 4(1-\rho^2)(\kappa^2 + 4\pi^2\tau^{-2})}}{2\omega(1-\rho^2)} \tag{4.30}$$

Proof:
For $\rho \to -1$, ζ_{D+} tends to infinity so that we can approximate (4.26) as $\tau = 2\pi/c$, leading to the analytical solution stated in (4.30).

As a final remark, when we have $\rho = 0$, as in an interest rate setting, symmetry is introduced into the problem: the critical time of the $(\zeta_{D+} + z)^{th}$ moment for $z > 0$ is equal to the critical time of the $(\zeta_{D-} - z)^{th}$ moment. Furthermore $\zeta_{D+} + \zeta_{D-} = 1$ here so that $\zeta_+ = 1 - \zeta_-$.

4.2.1.2. AJD stochastic volatility model

When independent jumps are added to the underlying asset we have two possible sources of explosions: the affine part of the model requires $\phi_{Affine}(x + iy)$ to be finite, whereas the jump part requires $\varphi_{Jump}(x + iy)$ to be finite. The second is only a practical problem since the jump component is always theoretically finite, although it may exceed the largest number representable on a finite computer system. Since we already discussed the affine part in the last section, we concentrate on the latter here. As before we focus on lognormally distributed jumps which are added to the underlying asset. For that reason we introduce:

$$f_{Jump}(\zeta) = \ln \phi_{Jump}(-i\zeta) = -\lambda\mu\tau\zeta + \lambda\tau\left((1+\mu)^{\zeta} \exp(\tfrac{1}{2}\eta^2\zeta(\zeta-1)) - 1\right) \tag{4.31}$$

τ	$\hat{\zeta}_{Max}$	$\zeta_{Jump,Max}$	Steps	$\hat{\zeta}_{Min}$	$\zeta_{Jump,Min}$	Steps
1/52	39.6266	39.6265	2	-57.6884	-57.6886	2
1	30.6901	30.6857	3	-48.7408	-48.7478	3
10	24.4450	24.4042	3	-42.3941	-42.4663	3

Table 4.1: Jump-diffusion: $\mu = \lambda = \eta = 0.1$.
Number of Newton-Rhapson iteration steps to obtain an accuracy of 10^{-8}.

Our aim is to find a range for ζ such that the characteristic function can be calculated using standard double precision. We therefore impose an upper bound on (4.31) equal to ¼ ln(DBL_MAX) ≈ 177 $= d_{Max}$. The motivation for this choice is that we still want to leave enough room for the remaining terms of the characteristic function. In view of the optimal choice of α discussed in Section 4.2.2 this upper bound is large enough to guarantee that the critical α of the jump part is greater than the optimal α. The next step is to calculate the critical values of ζ. This requires the solution of a non-linear equation which can be done by using the standard Newton–Raphson method. To find a good starting point, let us define the function g_{Jump} as a slight modification of f_{Jump}:

$$g_{Jump}(\zeta) = f_{Jump}(\zeta) + \lambda\mu\tau\zeta \tag{4.32}$$

This enables us to compute the solution of the non-linear equation $g_{Jump,Max}(\hat{\zeta}) = d_{Max}$ analytically:

$$\hat{\zeta}_{Max,Min} = \frac{\eta^2 - 2\ln(\mu+1) \pm \sqrt{8\ln((d_{Max}+\lambda\tau)/\lambda\tau)\eta^2 + (2\ln(\mu+1) - \eta^2)^2}}{2\eta^2} \tag{4.33}$$

Now one can use this ζ as a starting point for the Newton–Raphson method on $f_{Jump}(\zeta_{Jump}) = d_{Max}$. Table 4.1 shows that the initial guess ζ is a good approximation for the solution of the non-linear equation. Concluding, although the strip of regularity of the AJD stochastic volatility model is equivalent to that of the Heston model, numerical explosions of the characteristic function suggest that we should restrict the strip of regularity (ζ_-, ζ_+) to a strip of numerical regularity:

$$\zeta_{Min} = \max(\zeta_{Jump,Min}, \zeta_-) \qquad \zeta_{Max} = \min(\zeta_{Jump,Max}, \zeta_+) \tag{4.34}$$

Although we have focused on lognormally distributed jumps here, the same caveats apply to models with other jump size distributions.

4.2.1.3. VG model

Due to the comparative simplicity of the characteristic function of the VG model given by Equation (3.29), the strip of regularity (ζ_-, ζ_+) can straightforwardly be deduced as in (3.30). Since the VG model is fully time-homogeneous, the maximum and minimum allowed ζ, and therefore also α, do not depend on the time to maturity, contrary to what we find in the AJD stochastic volatility models.

4.2.2. Optimal α

Before we embark upon our quest for the optimal α, we return to the method considered by Levendorskiĭ and Zherder [2002], which we mentioned in the beginning of this section. Their integration-along-cut (IAC) method is specifically tailored towards KoBoL/CGMY processes of order smaller than one. For these processes it turns out that the characteristic function φ(u) is analytic for $u \in \mathbb{C}$, with cuts $[i\zeta_+, i\infty)$ and $(-i\infty, i\zeta_-]$. In words, this means that the characteristic function is well defined on the whole complex plane, apart from being discontinuous along the mentioned cuts. Instead of integrating along $(-\infty - i\alpha, \infty - i\alpha)$ in the complex plane to obtain the option price via Equation (2.34), they integrate along one of the two aforementioned cuts. Unfortunately this technique is not generally applicable to any process. For example, it was shown by Lord and Kahl [2006] that for the stochastic volatility models we consider here, the characteristic function has an infinite number of singularities along the imaginary axis. As we are seeking a method that is fully general and applicable to any model, we therefore do not consider this method any further in this article.

The literature on option pricing has recognised the need for a good choice of α. Choosing α too small or too big leads to problems with either cancellation errors or highly oscillating integrands as shown in Figures 4.1 and 4.2. The best α ensures that the absolute value of the integrand is as constant as possible over the whole integration area. The oscillation of a function $f: \mathbb{R} \to \mathbb{R}$ on the finite interval $[-1, 1]$ can be measured by the total variation:

$$TV(f) = \int_{-1}^{1} \left| \frac{\partial f(x)}{\partial x} \right| dx \tag{4.35}$$

As shown by Förster and Petras [1991] a small total variation reduces the approximation error of the applied numerical integration scheme. Ideally one should choose α such that the total variation of the integrand is minimised:

$$\alpha^* = \underset{\alpha \in (\alpha_{Min}, \alpha_{Max})}{\arg\min} \; e^{-\alpha k} \int_0^{\infty} \left| \frac{\partial}{\partial v} \psi(v, \alpha) \right| dv \tag{4.36}$$

Clearly this optimisation problem is of a rather theoretical nature as it has to be solved again for each option price and would therefore require many more function evaluations than the original problem. A simplification of (4.36) can be found by assuming that φ(v, α) is monotone in v on $[0, \infty)$. In this case it can be shown that:

$$e^{-\alpha k} \int_0^{\infty} \left| \frac{\partial}{\partial v} \psi(v, \alpha) \right| dv = e^{-\alpha k} \left| \psi(0, \alpha) - \psi(\infty, \alpha) \right| = \left| e^{-\alpha k} \psi(0, \alpha) \right| \tag{4.37}$$

since the characteristic function of an affine jump-diffusion model vanishes at infinity. Another motivation for can be found be minimising the absolute error:

$$\alpha^* = \underset{\alpha \in (\alpha_{Min}, \alpha_{Max})}{\arg\min} \; \left| C(k, \alpha) - C(k) \right| \tag{4.38}$$

where C(k, α) denotes the numerical approximation to the option price. This minimisation coincides with the stable region of α's we find in Figure 4.1. It can be shown that (4.37) is the first order condition for optimality in (4.38). We therefore suggest to choose α according to:

$$\alpha^* = \underset{\alpha \in (\alpha_{\text{Min}}, \alpha_{\text{Max}})}{\arg\min} \left| e^{-\alpha k} \psi(0,\alpha) \right| \tag{4.39}$$

or, equivalently:

$$\alpha^* = \underset{\alpha \in (\alpha_{\text{Min}}, \alpha_{\text{Max}})}{\arg\min} \; -\alpha k + \tfrac{1}{2} \ln\left(\psi(0,\alpha)^2 \right) \equiv \Psi(\alpha, k) \tag{4.40}$$

rendering the calculation more stable since the function ψ is not necessarily positive. Finding the optimal α thus requires to find the minimum of Ψ. The optimal α resulting from this optimization problem will be referred to as the payoff-dependent α, as ψ depends on the particular payoff function we are considering. In the following, we also consider a payoff-independent alternative:

$$\alpha^* = \underset{\alpha \in (\alpha_{\text{Min}}, \alpha_{\text{Max}})}{\arg\min} \; -\alpha k + \ln\left(\phi(-(\alpha+1)i) \right) \equiv \Phi(\alpha, k) \tag{4.41}$$

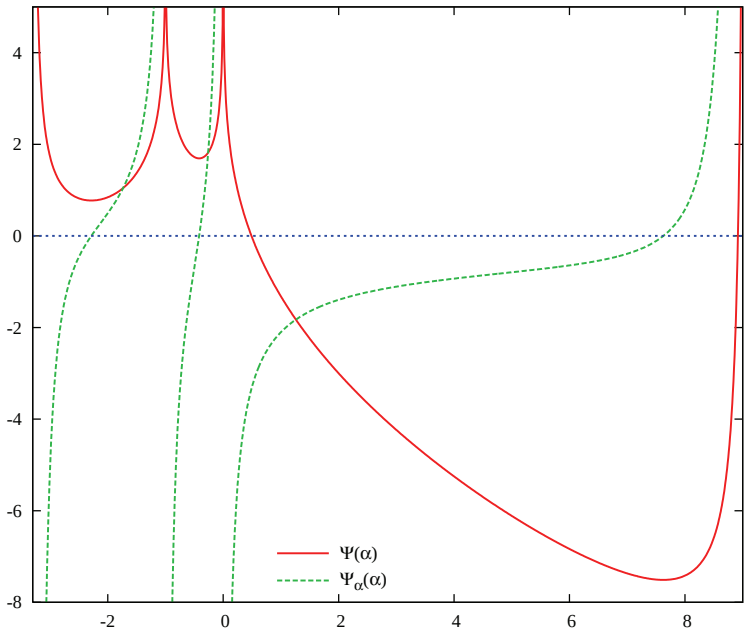
Coincidentally, this payoff-independent way of choosing α has a close link to how the optimal contour is chosen in the area of saddlepoint approximations, something we discuss in the following section. A related article by Choudhury and Whitt [1997] considers the contour shift following from (4.40) to avoid numerical problems in the numerical inversion of non-probability transforms. To the best of our knowledge this choice of α has not been picked up in the area of option pricing.

Let us for a moment consider whether solving the optimal α from (4.40) is a well-posed problem. From the theory of saddlepoint approximations (see Daniels [1954]) we know that the minimum of the damped characteristic exponent in (4.41) is unique under relatively mild conditions on the first derivative of the characteristic exponent (cf. Daniels [1954, Equation (6.1)]), which are satisfied here. This analysis leads us to conclude that the payoff-dependent function Ψ will have a local minimum in three ranges: $\alpha \in (\alpha_{\text{Min}}, -1)$, $\alpha \in (-1, 0)$ and $\alpha \in (0, \alpha_{\text{Max}})$, since Ψ is infinite when $\alpha = -1$ or 0. To find the optimal payoff-dependent α^* we have to solve the non-linear equation:

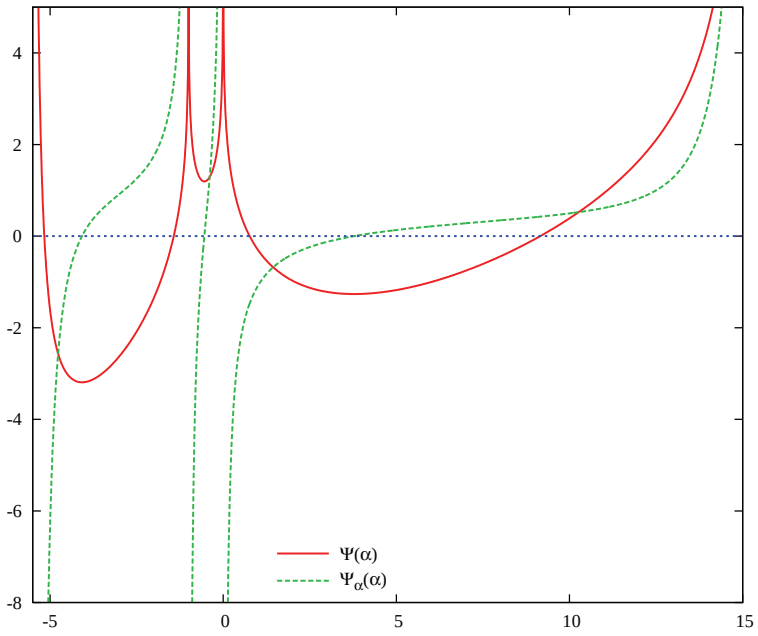
$$\frac{\partial \Psi(\alpha, k)}{\partial \alpha} = 0 \tag{4.42}$$

Its typical shape can be found in Figure 3, which indicates that for $F < K$, $\alpha^* \in (0, \alpha_{\text{Max}})$ and for $F > K$ we will have $\alpha^* \in (\alpha_{\text{Min}}, -1)$. Although this is no rigorous proof, more often than not we find that this rule of thumb holds true. This allows us to restrict the search for the optimal α to just one of the three aforementioned ranges, if computational time is of the essence. In other words the rule of thumb states that it is always preferable to price OTM options. The following two sections explore the payoff-independent choice of α for two models in which it can be calculated analytically: the Black–Scholes model and the VG model.

Figure 4.3: Function Ψ given by Equation (4.40) and its first derivative for the Heston model w.r.t. α.
Underlying: $dS(t)/S(t) = \mu\,dt + \sqrt{V(t)}\,dW_S(t)$ with S = F = 1 and $\mu = 0$.
Variance: $dV(t) = \kappa(\theta - V(t))dt + \omega\sqrt{V(t)}\,dW_V(t)$ with $V(0) = \theta = 0.1$, $\kappa = \omega = 1$ and $\rho = -0.7$.



(A) $\tau = 1$ and K = 1.2



(B) $\tau = 1/2$ and K = 0.7

4.2.2.1. Black-Scholes model

As the characteristic function of the Black–Scholes formula has a convenient analytical structure, the optimal payoff-independent α we propose in Equation (4.41) can be calculated analytically. We find:

$$\alpha^* = \underset{\alpha \in (\alpha_{\text{Min}}, \alpha_{\text{Max}})}{\arg\min} \; \Phi(\alpha, k) = -\frac{f - k + \frac{1}{2}\hat{\eta}\tau}{\hat{\eta}^2 \tau} = -\frac{d_1}{\hat{\eta}\sqrt{\tau}} \tag{4.43}$$

where d_1 is a well-known part of the Black–Scholes option pricing formula. Note that the payoff-independent α also roughly obeys the rule of thumb we stated earlier. The payoff-dependent α^* can also be solved in closed-form, although this does not lead to nice analytical expressions as (4.42) becomes a fourth-order polynomial in α.

There is one more thing to be said about the payoff-independent α^* in the Black-Scholes model. Suppose we write:

$$\phi(v - (\alpha + 1)i) = \exp\big(\phi_r(v, \alpha) + i\phi_i(v, \alpha) \big) \tag{4.44}$$

with both φ_r and φ_i being real-valued functions. The Carr-Madan representation of the European call option price can then be written as:

$$\begin{aligned} C(\alpha) = \frac{e^{-\alpha k}}{\pi} \int_0^\infty & e^{\varphi_r(v,\alpha)} \cdot \\ & \cdot \frac{(\alpha(\alpha+1) - v^2)\cos(kv - \phi_i(v,\alpha)) - (1+2\alpha)v\sin(kv - \phi_i(v,\alpha))}{(v - i(\alpha+1))(v + i(\alpha+1))(v - i\alpha)(v + i\alpha)} \, dv \end{aligned} \tag{4.45}$$

which is a real-valued integral. So far this discussion has been model-independent. Upon inspection it seems logical to choose α such that all oscillations due to the sine and cosine are removed completely, by setting:

$$kv - \varphi_i(v, \alpha) = 0 \tag{4.46}$$

Typically, the solution to this equation will depend on v, although it does not in the Black-Scholes case. In this case the solution for α in (4.46) is exactly the payoff-independent α^* we derived earlier in (4.43).

4.2.2.2. VG model

Standard yet tedious calculations yield:

$$\alpha^* = \underset{\alpha \in \mathbb{R}}{\arg\min} \; \Phi(\alpha, k) = -\frac{\theta}{\sigma^2} - 1 + \frac{\tau}{\nu \tilde{m}} - \text{sgn}(\tilde{m}) \sqrt{\frac{\theta^2}{\sigma^4} + \frac{2}{\nu\sigma^2} + \frac{\tau^2}{\nu^2 \tilde{m}^2}} \tag{4.47}$$

where we introduced $\tilde{m} = \tilde{f} - k = f - k + \omega\tau$, a quantity related to the log-moneyness of the option. One can easily check that $\alpha^* \in (\alpha_-, \alpha_+)$, as should be the case. We mention that Aït-Sahalia and Yu [2006] provide the saddlepoint of the VG model and several other models where it can be calculated analytically. As we will show in the next section, the saddlepoint, minus one, coincides exactly with our payoff-independent α^*.

4.2.3. Saddlepoint approximations

The concept of saddlepoint approximations dates back to Daniels [1954] in the context of calculating probability densities via Fourier inversion:

$$p(x) = \frac{1}{2\pi}\int_{-\infty-i\alpha}^{\infty-i\alpha} e^{-izx}\phi_p(z)\,dz \tag{4.48}$$

where φ_p is the characteristic function associated with the density p. Although the standard inversion formula would set α equal to zero, we can shift the contour of integration by choosing α different from zero. Saddlepoint approximations typically continue along the following lines. Let $M(z) = \ln \varphi_p(z)$, the characteristic exponent, then Daniels advocated to choose α as the minimum of the damped characteristic exponent (cf. (4.41)):

$$\alpha^* = \underset{\alpha\in(\alpha_{Min},\alpha_{Max})}{\arg\min} \; -\alpha x + M(-i\alpha) \tag{4.49}$$

Thus, we obtain $M'(-i\alpha) = ix$. Note that the payoff-independent α coincides with the saddlepoint minus one. Applying a Taylor expansion around its minimum leads to:

$$M(z) - izx = M(-i\alpha^*) - \alpha^* x + \tfrac{1}{2}M''(-i\alpha^*)(z+i\alpha^*)^2 + O(z^3) \tag{4.50}$$

so that the density can be approximated via:

$$\begin{aligned} p(x) &= \frac{1}{2\pi}\int_{-\infty-i\alpha^*}^{\infty-i\alpha^*} e^{-izx}\phi_p(z)\,dz \approx \frac{\phi_p(-i\alpha^*)e^{-\alpha^* x}}{2\pi}\int_{-\infty-i\alpha^*}^{\infty-i\alpha^*} e^{\frac{1}{2}M''(-i\alpha^*)z^2}\,dz \\ &= \frac{\phi_p(-i\alpha^*)e^{-\alpha^* x}}{\sqrt{-2\pi M''(-i\alpha^*)}} \equiv p_{simple}(x) \end{aligned} \tag{4.51}$$

Instead of numerically evaluating an integral over an infinite domain, this "simple" saddlepoint approximation requires only a few function evaluations[14]. As Aït-Sahalia and Yu [2006] mention, the name "saddlepoint" stems from the shape of the right-hand side of (4.50) in a neighbourhood of its minimum, which can be seen as a saddle. Based on this saddlepoint approximation to the density function, an approximation can be derived for the cumulative density function. The most famous of these is the Lugannani–Rice formula (see Lugannani and Rice [1980]). All saddlepoint approximations are found to work remarkably well in the tails of the distribution. As such, saddlepoint approximations in finance have mainly been used for VaR and expected loss calculations. Although option pricing mainly deals with the bulk of the distribution, Rogers and Zane [1999] applied the Lugannani–Rice formula to compute the two probabilities in Equations (2.27) and (2.28). As expected they obtained accurate results for options close to maturity and away from the ATM level, although the accuracy was lower around the ATM level. For small option prices this approach will lead to cancellation errors, so that we suggest to apply the saddlepoint approximation directly to the Carr–Madan/Lewis representation in (4.1). Let us define $\psi(z) = \psi(v, \alpha)$ when $z = v - i\alpha$ for $v, \alpha \in \mathbb{R}$, i.e. $\psi(z) = \varphi(z - i) / (-z(z - i))$. Instead of

[14] The number of function evaluations depends on the approximation of the second-order derivative.

expanding the characteristic exponent around its minimum, we set $M(z) = \frac{1}{2} \ln \psi(z)^2$, leading to the "simple" saddlepoint approximation of the option price:

$$\begin{aligned} \mathbb{E}\left[(S(T)-K)^+\right] &= R(S,K,\alpha^*) + \frac{1}{2\pi}\int_{-\infty-i\alpha^*}^{\infty-i\alpha^*} e^{-izk}\psi(z)\,dz \\ &\approx \frac{\psi(-i\alpha^*)e^{-\alpha^* k}}{\sqrt{-2\pi M''(-i\alpha^*)}} \equiv C_{simple}(k) \end{aligned} \tag{4.52}$$

where α^* here is the payoff-dependent optimal α.

Naturally, higher order expansions of the density can also be derived. Aït-Sahalia and Yu [2006] gave a higher-order expansion derived by expanding the left-hand side of (4.50) to fourth order in z:

$$p(x) = p_{simple}(x)\left(1 + \frac{1}{8}\frac{M^{(4)}(-i\alpha^*)}{M^{(2)}(-i\alpha^*)^2} - \frac{5}{24}\frac{M^{(3)}(-i\alpha^*)^2}{M^{(2)}(-i\alpha^*)^3}\right) \equiv p_{Sophisticated}(x) \tag{4.53}$$

We will refer to the analogue of (4.53) in the context of option pricing as the "sophisticated" saddlepoint approximation of the option price. The numerical effort for this approximation has increased considerably since we have to approximate the fourth-order derivative numerically[15].

As a final note, we mention that the saddlepoint approximations considered here can be seen as an expansion around a Gaussian base (cf. (71)). Aït-Sahalia and Yu [2006] showed how to derive saddlepoint approximations which expand the characteristic exponent around a non-Gaussian base, and demonstrate that this improves the accuracy of the approximations considerably for jump-diffusion and Lévy models. Although this route is viable, the problem of determining which base is most suitable for a given model is obviously a very model-dependent problem. As we would like to keep our pricing methods as model-independent as possible, the only model-dependent part being of course the specification of the characteristic function and a transformation function[16], we only consider the saddlepoint approximations around a Gaussian base here.

4.3. Numerical results

In this final section we present some numerical tests comparing the different numerical methods for semi-analytical option pricing with regard to their approximation quality and the computational effort. In Table 2 we compare the number of function evaluations and the computational error for the standard Heston formulation in (2.27) and (2.28) to the Carr-Madan/Lewis representation in (4.1) using the optimal payoff-dependent α from (4.40)).

One can recognise that the optimal choice of α decreases both the necessary number of function evaluations and the computational error. The parameter configuration was taken from Kahl and Jäckel [2005, Figure 7] and is such that the standard Heston implementation leads to a smooth implied volatility surface when using a relative tolerance of 10^{-12} in the adaptive Gauss–Lobatto scheme. Another great advantage is that the optimal choice of α allows us to calculate

[15] As we have closed-form expressions for all characteristic functions considered in this article, one could theoretically derive all the required derivatives in closed-form.

[16] In case the infinite domain of integration cannot be transformed to a finite one we require an appropriate upper limit of integration.

	Heston			Optimal α		
Tolerance	**avg. FE**	**max. error**	**avg. error**	**avg. FE**	**max. error**	**avg. error**
Minimum	24	173.63%	4.8071%	12	3.17%	0.048%
10^{-3}	212.9	32.91%	0.2724%	73.9	0.47%	0.0038%
10^{-5}	540.4	2.5%	0.0044%	179.8	0.03%	0.0001%
10^{-7}	1097.4	0.09%	9.7×10^{-5} %	373.5	0.0001%	7.7×10^{-7} %

Table 4.2: Average number of function evaluations (FE) and error measured in implied volatility for the Heston model calculated over a whole range of strikes and maturities.
Underlying: $dS(t)/S(t) = \mu\,dt + \sqrt{V(t)}\,dW_S(t)$ with S = F = 1 and μ = 1.
Variance: $dV(t) = \kappa(\theta - V(t))dt + \omega\sqrt{V(t)}\,dW_V(t)$ with V(0) = θ = 0.16, κ = 1, ω = 2, ρ = -0.8, $\tau \in$ [1, 5/4, 6/4, …, 15] and K $\in$ [1/10, 2/10, …, 4].

τ – K	9.5	9.6	9.7	9.8	9.9	10.0
1/52	6.4232×10^{-260}	2.6773×10^{-261}	1.1522×10^{-262}	5.1158×10^{-264}	2.3423×10^{-265}	1.1052×10^{-266}
2/52	3.4710×10^{-133}	6.9920×10^{-261}	1.4313×10^{-134}	2.9768×10^{-135}	6.2873×10^{-136}	1.3483×10^{-136}
3/52	76.979×10^{-91}	26.221×10^{-91}	9.0293×10^{-92}	3.1424×10^{-92}	1.1051×10^{-92}	3.9263×10^{-93}
4/52	1.2869×10^{-69}	5.7020×10^{-70}	2.5472×10^{-70}	1.1471×10^{-70}	5.2069×10^{-71}	2.3818×10^{-71}

Table 4.3: Option prices within the Heston model with the same parameter configuration as in Figure 4.3

option prices for all ranges of strikes and maturities. Figure 4.4 shows the implied volatility surface smooth implied volatility surface when using a relative tolerance of 10^{-12} in the adaptive Gauss–Lobatto scheme. Another great advantage is that the optimal choice of α allows us to calculate option prices for all ranges of strikes and maturities. Figure 4.4 shows the implied volatility surface for the Heston model calculated with the optimal α. This surface is perfectly smooth. In the upper right corner we have options which are close to maturity and decisively OTM. In Table 4.3 we present the corresponding option prices. Using the optimal α in conjunction with the adaptive Gauss–Lobatto scheme, and transforming the integration domain to a finite interval, allows us to compute option prices down to the lowest number[17] representable in floating point precision. Moreover, this method can be used as a black-box algorithm to semi-analytical option pricing since we avoid numerical instabilities for small option prices.

As a final assessment of how well our method performs, we compare the optimal α method with the saddlepoint approximations discussed in Section 4.2.3. In Figure 4. we present the error made in the implied volatility surface using the saddlepoint approximation to approximate the inverse Fourier integral. As mentioned these saddlepoint approximations are different to the Lugannani–Rice based approximation Rogers and Zane [1999] considered in their article, although we find their accuracy to be comparable to ours. For this reason we focus on the simple and sophisticated saddlepoint approximation we suggested in Section 4.2.3. Both saddlepoint approximations work extremely well in the tails of the probability density but fail for strikes that are around the ATM level. The method is most efficient for small times to maturity. Comparing the different saddlepoint methods we see that including higher-order terms of the Taylor expansion certainly leads to an improved accuracy, at the cost of having to compute fourth-order derivatives numerically.

As a fair comparison of the saddlepoint methods to the optimal α method we finally present the error in the implied volatility surface for the optimal α method in Figure 4.6, using a maximum of 12 function evaluations per option price. This is more than competitive with the sophisticated saddlepoint approximation, as just using central-difference formulas of order $O(h^2)$,

[17] This information can be found again in the C header <float.h> in the macro DBL_MIN.

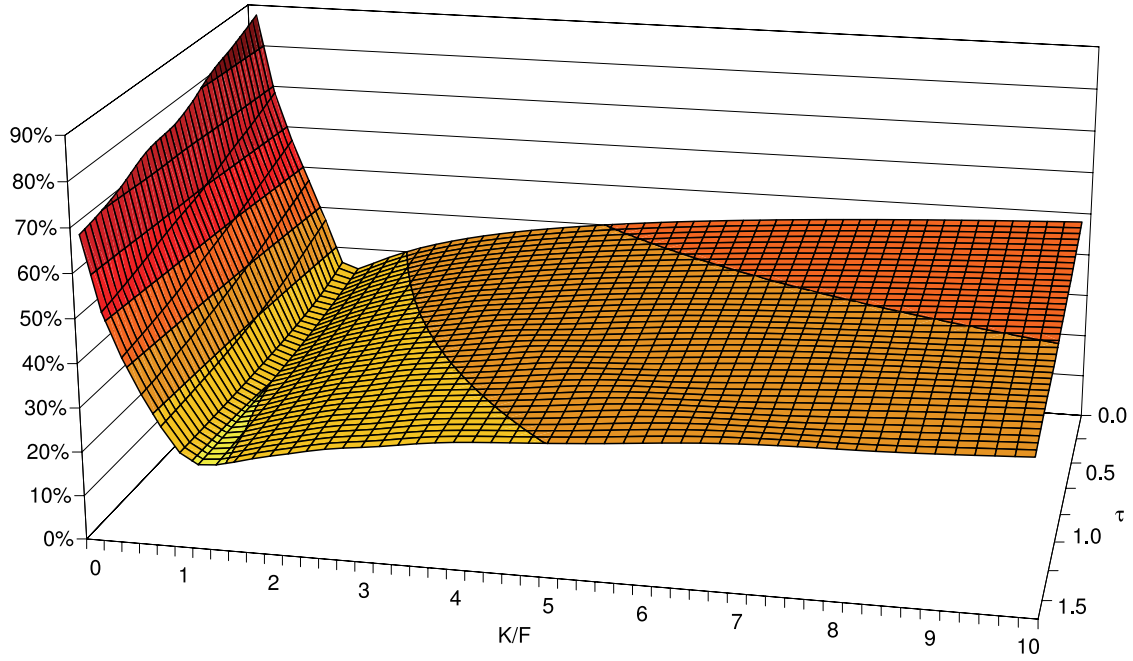
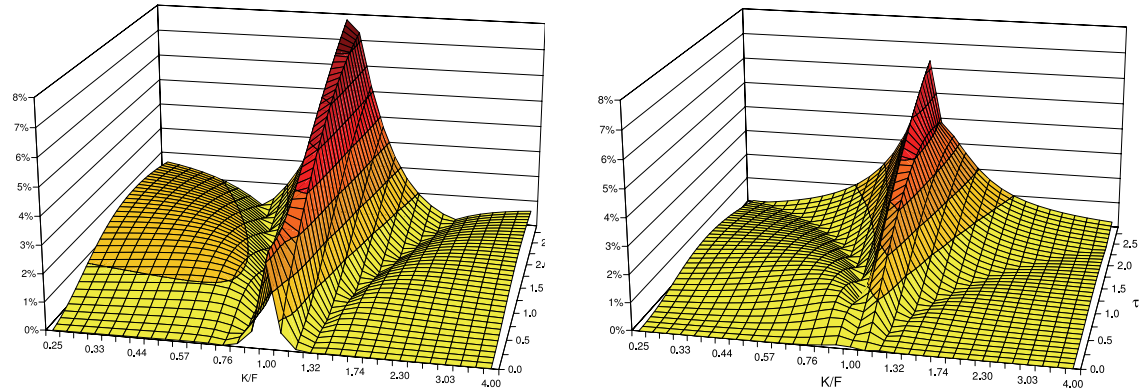


Figure 4.4: Black implied volatilities from the Heston model with parameter configuration as in Figure 4.3. Maturities τ range from 1/52 till 1.5 years and strikes K range from 1/10 to 10.



(A) (B)

Figure 4.5: Error in the implied volatility surface for the saddlepoint approximation in the Heston model. Parameter configuration as in Figure 4.3, with $\rho = -0.5$. (A) Simple saddlepoint approximation from (4.51) (B) Sophisticated saddlepoint approximation from (4.53)

with h being the stepsize used for the numerical differentiation, would already require five function evaluations. Calculating the derivatives numerically with two different values of h and subsequently applying Richardson extrapolation would require 7 function evaluations. Clearly, using a more sophisticated scheme to calculate the required derivatives numerically will require many more function evaluations. We find that the stripped-down optimal α method produces an

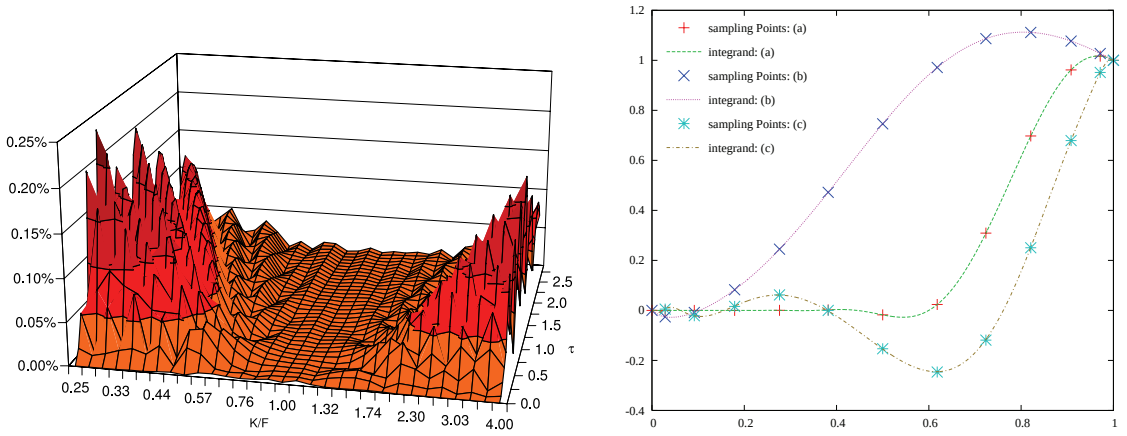


Figure 4.6: (A) Error in the implied volatility surface for the optimal α method in the Heston model using 12 function evaluations in the adaptive Gauss-Lobatto method for each price. Parameter configuration as in Figure 4.3, with ρ = -0.5. (B) Sampling points for the adaptive Gauss-Lobatto quadrature scheme for (a) K = 1 and τ = ½, (b) K = 4 and τ = ½ and (c) K = 0.4 and τ = 3.

Method	(A)	(B)	(C)	(D)	(E)
Correct price	0.13989525	0.07588180	0.00198142	0.01292888	1.011027×10^{-14}
Stripped-down optimal α	0.13987989	0.07587201	0.00197497	0.01290560	1.011313×10^{-14}
Simple saddle	0.10638964	0.06665373	0.00215159	0.01507162	1.027944×10^{-14}
Advanced saddle	0.11499332	0.07440537	0.00206302	0.01402231	1.011445×10^{-14}

Table 4.4: Option prices within the Heston model for different strike and maturity scenarios (A)-(E) and different pricing methodologies. The parameters of the Heston model are the same as in Figure 4.6. (A) τ = 2, K = 1, (B) τ = ½, K = 1, (C) τ = ½, K = ½, (D) τ = 3/2, K = ½ and (E) τ = 1/12, K = ¼.

error of less than half a per cent over the whole surface, which is more than sufficient for all practical purposes. Furthermore, the stripped-down optimal α method only seems to produce marginally worse results than the sophisticated saddlepoint method for options with low strikes and high maturities. In order to give further evidence we directly compare the different methods in Table 4.4. It is particularly interesting to note that both saddlepoint approximations fail to give sufficiently accurate results if used ATM whilst the stripped down optimal α approach is equally applicable over the whole grid of different strikes and maturities. From a computational point of view the optimal α method should therefore certainly be preferred, as numerical integration of a function is a much more stable operation than the numerical differentiation required in the calculation of the saddlepoint approximations.

4.4. Conclusions

In this chapter we have investigated the problem of handling the Fourier inversion required in semi-analytical option pricing. Using the Carr–Madan representation of the option price and shifting the contour of integration along the complex plane allows for different representations of the European call price. The key point is to choose the contour of integration best suited for numerical quadrature, which can be changed by varying the damping parameter α. Changing the contour of integration alters the behaviour of the integrand considerably, so that we have to ensure that the integrand does not become too peaked or too oscillatory. Furthermore,

cancellation errors also have to be avoided at all cost. By transforming the integration domain as discussed in Section 4.1, as well as choosing the optimal α as in Section 4.2, we obtain a call price formula which allows for a robust implementation enabling us to price options down to machine size precision. The optimal choice of the damping parameter α is the only way to overcome any numerical instabilities and guarantee accurate results.

Appendix 4.A – Proof of Proposition 4.1

In this appendix we present the missing proof.

Proposition 4.1
Assuming that $\kappa, \theta, \omega, \tau > 0$ and $\rho \in (-1, 1)$ we obtain the following asymptotics for the integrand of the Schöbel-Zhu model:

$$\lim_{u\to\infty}\psi(u,\alpha) \approx \psi(0,\alpha)\cdot e^{-uC_\infty}\cdot \mathrm{Re}\left(-\frac{e^{iut_\infty}}{u^2}\right) = \psi(0,\alpha)\cdot e^{-uC_\infty}\cdot\frac{\cos(ut_\infty)}{-u^2} \tag{4A.1}$$

with:

$$C_\infty = \tfrac{1}{4}D_\infty(\tau + V(0)/\omega^2) \qquad t_\infty = \tfrac{1}{4}\beta_\infty(\tau + V(0)/\omega^2) + \ln\tfrac{F}{K} \tag{4A.2}$$

and auxiliary variables $D_\infty = 2\omega\sqrt{1-\rho^2}$ and $\beta_\infty = -2\omega\rho$.

Proof:
One can succesively deduce that:

$$\begin{aligned}
&\lim_{u\to\infty}\frac{\beta(u)}{u} = -2i\rho\omega \equiv i\beta_\infty \\
&\lim_{u\to\infty}\frac{D(u)}{u} = 2\omega\sqrt{1-\rho^2} \equiv D_\infty \\
&\lim_{u\to\infty} G(u) = \text{constant}
\end{aligned} \tag{4A.3}$$

Using the limiting behaviour of D we can further show that:

$$\lim_{u\to\infty} A_\sigma(u) = 0 \qquad \lim_{u\to\infty} B_\sigma(u) = 0 \tag{4A.4}$$

which finally leads to:

$$\lim_{u\to\infty} A_\sigma(u) = \tfrac{1}{4}(\beta_\infty - D_\infty)\tau \qquad \lim_{u\to\infty} B_\sigma(u) = 0 \tag{4A.5}$$

Combining all of the results we obtain:

$$C_\infty = \tfrac{1}{4}D_\infty(\tau + V(0)/\omega^2) \tag{4A.6}$$

completing the proof.

Chapter 5

A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes[18]

When valuing and risk-managing exotic derivatives, practitioners demand fast and accurate prices and sensitivities. As the financial models and option contracts used in practice are becoming increasingly complex, efficient methods have to be developed to cope with such models. Aside from non-standard exotic derivatives, plain vanilla options in many stock markets are actually of the American type. As any pricing and risk management system has to calibrate to these plain vanilla options, it is important to be able to value them quickly and accurately.

By means of the risk-neutral valuation formula the price of any option without early exercise features can be written as an expectation of the discounted payoff of this option. Starting from this representation one can apply several numerical techniques to calculate the price itself: Monte Carlo simulation, numerical solution of the corresponding partial-(integro) differential equation (P(I)DE) and numerical integration. While the treatment of early exercise features within the first two techniques is relatively standard, the pricing of such contracts via quadrature pricing techniques has not been considered until recently, see Andricopoulos, Widdicks, Duck and Newton [2003] and O'Sullivan [2005]. Each of these methods has its merits and demerits, though for the pricing of American options the PIDE approach currently seems to be the clear favourite, see e.g. Higham [2004] and Wilmott, Dewinne and Howison [1993].

In the past couple of years a vast body of literature has considered the modelling of asset returns as infinite activity Lévy processes, due to the ability of such processes to adequately describe the empirical features of asset returns and at the same time provide a reasonable fit to the implied volatility surfaces observed in option markets. Valuing American options in such models is however far from trivial, due to the weakly singular kernels of the integral terms appearing in the PIDE, as reported in, e.g., Almendral and Oosterlee [2007a, b], Cont and Tankov [2004], Hirsa and Madan [2004], Matache, Nitsche and Schwab [2005] and Wang, Wan and Forsyth [2007].

In this chapter we present a quadrature-based method for pricing options with early exercise features. The method combines the recent quadrature pricing methods of Andricopoulos et al. [2003] and O'Sullivan [2005] with the methods based on Fourier transformation pioneered by Carr and Madan [1999], Raible [2000] and Lewis [2001], as considered in Chapters 2 and 4.

Though the transform methods so far have mainly been used for the pricing of European options, we show how early exercise features can be incorporated naturally. The requirements of the method are that the increments of the driving processes are independent of each other, and that the conditional characteristic function of the underlying asset is known. This is the case for many of the affine models considered in the previous chapters. In contrast to the PIDE methods, processes of infinite activity, such as the Variance Gamma (VG) or CGMY models can be handled with relative ease.

[18] This chapter has appeared as Lord, R., Fang, F., Bervoets, F. and C.W. Oosterlee [2008]. "A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes", *SIAM Journal on Scientific Computing*, vol. 30, no. 4, pp. 1678-1705.

The present chapter is organised as follows. As this thesis has so far focused on the pricing of European options, we start with an introduction to the pricing of options with early exercise features. Subsequently we introduce the novel method called Convolution (CONV) method for such options. Its high accuracy and speed are demonstrated by pricing several Bermudan and American options under the Black-Scholes, VG, CGMY and Kou models.

5.1. Options with early exercise features

The best known examples of options with early exercise features are American and Bermudan options. American options can be exercised at any time prior to the option's expiry, whereas Bermudan options can only be exercised at certain dates in the future. In the algorithm we consider in this chapter, American options are priced by approximation – either as Bermudans with many exercise opportunities, or via Richardson extrapolation. For this reason we focus on Bermudan options in this section.

Let us first introduce some notation. We define the set of exercise dates as $\mathcal{T} = \{t_1, \ldots, t_M\}$ where $0 = t_0 \leq t_1 < \ldots < t_M$. For ease of exposure we assume the exercise dates are equally spaced, so that $t_{m+1} - t_m = \Delta t$. If the option is exercised at some time $t \in \mathcal{T}$, the holder of the option obtains the exercise payoff $E(t, S(t))$. The valuation of any option starts from the risk-neutral valuation formula (cf. (2.26)), that for a European option reads:

$$V(t,S(t)) = P(t,T) \cdot \mathbb{E}_t\left[V(T,S(T))\right] \tag{5.1}$$

where V denotes the value of the option, T is the maturity of the option and P(t,T) denotes the time t price of a zero-coupon bond maturing at T. The variable S denotes the asset on which the option contract is based. To simplify the exposition, we assume that the interest rate is deterministic. At the cost of an increased dimensionality of the pricing problem the interest rate can be made stochastic.

The Bermudan option price can then be found via backward induction as:

$$\begin{cases} V(t_M, S(t_M)) = E(t_M, S(t_M)) \\ C(t_M, S(t_M)) = P(t,T) \cdot \mathbb{E}_{t_m}\left[V(t_{m+1}, S(t_{m+1})) \right] \\ V(t_m, S(t_m)) = \max\{ C(t_m, S(t_m)), E(t_m, S(t_m)) \} \\ V(t_0, S(t_0)) = C(t_0, S(t_0)) \end{cases} \tag{5.2}$$

where C is the continuation value of the option and V the value of the option immediately prior to the exercise opportunity. The dynamic programming problem in (5.2) is a successive application of the risk-neutral valuation formula, and we can write the continuation value as:

$$C(t_m, S(t_m)) = P(t,T) \cdot \int_{-\infty}^{\infty} V(t_{m+1}, y)\, f(y \mid S(t_m))\, dy \tag{5.3}$$

where $f(y|x)$ represents the probability density describing the transition from $S(t_m)$ at t_m to y at t_{m+1}. Based on (5.2) and (5.3) the QUAD method was introduced in Andricopoulos et al. [2003]. The method requires the transition density to be known in closed-form, which is the case in e.g. the Black-Scholes model and Merton's jump-diffusion model. This requirement is relaxed in O'Sullivan [2005], where the QUAD-FFT method is introduced. The underlying idea is that the

transition density can be recovered by Lévy inversion, so that the QUAD method can be used for a wider range of models. As such the QUAD-FFT method, also applied in Chourdakis [2005], effectively combines the QUAD method with the early transform methods. The overall complexity of both methods is $O(MN^2)$ for an M-times exercisable Bermudan option with N grid points used to discretise the price of the underlying asset.

The complexity of this method can be improved to $O(MN \log_2 N)$ if the underlying is a monotone function of a Lévy process. We will demonstrate this shortly. In the remainder we assume, as is common, that the underlying process is modelled as an exponential of a Lévy process. Let $x_1, \ldots, x_N$ be a uniform grid for the log-asset price. If we discretise (5.3) by the trapezoidal rule we can write the continuation value in matrix form as:

$$\mathbf{C}(t_m) \approx P(t_m, t_{m+1}) \cdot \Delta x \cdot \left[\mathbf{F}\mathbf{V} - \tfrac{1}{2}\left(V(t_{m+1}, x_1)\mathbf{f}_1 + V(t_{m+1}, x_N)\mathbf{f}_N\right)\right] \tag{5.4}$$

where:

$$\mathbf{f}_i = \begin{pmatrix} f(x_i \mid x_1) \\ \vdots \\ f(x_i \mid x_N) \end{pmatrix} \qquad \mathbf{F} = \begin{pmatrix} \mathbf{f}_1 & \cdots & \mathbf{f}_N \end{pmatrix} \qquad \mathbf{V} = \begin{pmatrix} V(t_{m+1}, x_1) \\ \vdots \\ V(t_{m+1}, x_N) \end{pmatrix} \tag{5.5}$$

and $f(y|x)$ now denotes the transition density in logarithmic coordinates. The key observation is that the increments of Lévy processes are independent, so that due to the uniform grid:

$$\mathbf{F}_{j,\ell} = f(y_i \mid y_\ell) = f(y_{j+1} \mid y_{\ell+1}) = \mathbf{F}_{\mathbf{j+1},\ell+1} \tag{5.6}$$

The matrix **F** is hence a Toeplitz matrix. A Toeplitz matrix can easily be represented as a circulant matrix, which has the property that the FFT algorithm can be employed to efficiently calculate matrix-vector multiplications. Therefore, an overall computational complexity of $O(MN \log_2 N)$ can be achieved. Though this method is significantly faster than the QUAD or QUAD-FFT methods, we do not pursue it in this chapter as the method we develop in the next section has the same complexity, yet requires fewer operations.

The previous literature does not seem to have picked up on a presentation by Reiner [2001], where it was recognised that for the Black-Scholes model the risk neutral valuation formula in (5.3) can be seen as a convolution or correlation of the continuation value with the transition density. As convolutions can be handled very efficiently by means of the FFT, an overall complexity of $O(MN \log_2 N)$ can be achieved. By working forward instead of backward in time a number of discrete path-dependent options can also be treated, such as lookbacks, barriers, Asian options and cliquets. Building on Reiner's idea, Broadie and Yamamoto [2005] reduced the complexity to $O(MN)$ for the Black-Scholes model by combining the double exponential integration formula and the Fast Gauss Transform. Their technique is applicable to any model in which the transition density can be written as a weighted sum of Gaussian densities, which is the case in e.g. Merton's jump-diffusion model.

As one of the defining properties of a Lévy process is that its increments are independent of each other, the insight of Reiner has a much wider applicability than only to the Black-Scholes model. This is especially appealing since the usage of Lévy processes in finance has become more established nowadays. By combining Reiner's ideas with the work of Carr and Madan, we introduce the Convolution method, or CONV method for short. The complexity of the method is $O(MN \log_2 N)$ for an M-times exercisable Bermudan option. Our method has similarities with both the quadrature pricing and the PIDE methods. Though the complexity of our method is

smaller than that of the QUAD variants, we share the construct that time steps are only required on the exercise dates of the product. However, our application of the FFT to approximate convolution integrals bears more resemblance to the approximation of the integral term in the numerical solution of a PIDE. Here Andersen and Andreasen [2000] were the first to suggest that for jump-diffusion models the integral term in the PIDE can be calculated efficiently via use of the FFT, rendering the complexity $O(MN \log_2 N)$ instead of $O(MN^2)$. Since then similar ideas have been applied to various jump-diffusion and infinite activity Lévy models, see d'Halluin, Forsyth and Labahan [2004], Almendral and Oosterlee [2007a,b] and Wang et al. [2007]. We will compare our method in terms of accuracy and speed to two PIDE methods in Section 5.5. Alternative methods for valuing options in Lévy models are the lattice-based approach of K¨ellezi and Webber [2004], which is $O(MN^2)$ and the multinomial tree of Maller, Solomon and Szimayer [2006] which is $O(M^2)$.

5.2. The CONV method

The main premise of the CONV method is that the conditional probability density f(y|x) in (5.3) only depends on x and y via their difference:

$$f(y \mid x) = f(y - x) \tag{5.7}$$

Note that x and y do not have to represent the asset price directly, they could be monotone functions of the asset price. The assumption made in (5.7) therefore certainly holds when the asset price is modelled as a monotone function of a Lévy process, since one of the defining properties of a Lévy process is that its increments are independent of each other as we know from Section 2.1. We choose to work with exponential Lévy models in the remainder of this chapter. In this case x and y in (5.7) represent the log-spot price. By including (5.7) in (5.3) and changing variables $z = y - x$ the continuation value can be expressed as:

$$C(t_m, x) = P(t_m, t_{m+1}) \cdot \int_{-\infty}^{\infty} V(t_{m+1}, x + z) f(z)\, dz \tag{5.8}$$

which is a cross-correlation[19] of the option value at time t_{m+1} and the density f(z), or equivalently, a convolution of $V(t_{m+1})$ and the conjugate of f(z). If the density function has a closed-form expression, it may be beneficial to proceed along the lines of (5.4). However, for many exponential Lévy models we either do not have a closed form expression for the density (e.g. the CGMY/KoBoL model of Boyarchenko and Levendorskiĭ [2002] and Carr, Geman, Madan and Yor [2002] and many affine models), or if we have, it involves one or more special functions (e.g. the VG model). In contrast, the characteristic function of the log-return can typically be obtained in closed-form or, in case of affine models, via the solution of a system of OIDEs.

We therefore take the Fourier transform of (5.8). The insight that the continuation value can be seen as a convolution is useful here, as the Fourier transform of a convolution is the product of the Fourier transforms of the two functions being convolved. In the remainder we will employ the following definitions for the continuous Fourier transform and its inverse:

[19] The cross-correlation of two functions f and g, denoted by $f \star g$, is defined as:

$$f(t) \star g(t) = \bar{f}(-t) * g(t) = \int_{-\infty}^{\infty} f(\tau) g(t + \tau)\, d\tau$$

where $*$ denotes the convolution operator.

$$\begin{aligned}\hat{h}(u) &= \mathcal{F}\{h\}(u) = \int_{-\infty}^{\infty} e^{iut} h(t)\,dt \\ h(t) &= \mathcal{F}^{-1}\{\hat{h}\}(t) = \tfrac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iut}\hat{h}(u)\,du\end{aligned} \tag{5.9}$$

If we dampen the continuation value in (5.8) by a factor $\exp(\alpha x)$ and subsequently take its Fourier transform, we obtain:

$$\begin{aligned}\mathcal{F}\{c(t_m,\cdot)\}(u) &= P(t_m,t_{m+1})\cdot\int_{-\infty}^{\infty} e^{iux} e^{\alpha x}\int_{-\infty}^{\infty} V(t_{m+1},x+z)\,f(z)\,dz\,dx \\ &= P(t_m,t_{m+1})\cdot\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{iu(x+z)} v(t_{m+1},x+z) e^{-i(u-i\alpha)z}\,f(z)\,dz\,dx\end{aligned} \tag{5.10}$$

where in the first step we introduced the risk-neutral valuation formula from (5.8). We introduced the convention that small letters indicate damped quantities, i.e. $c(t_m,x) = e^{\alpha x}C(t_m,x)$ and similarly for v. Changing the order of integration and remembering that x = y – z, we obtain:

$$\begin{aligned}\mathcal{F}\{c(t_m,\cdot)\}(u) &= P(t_m,t_{m+1})\cdot\int_{-\infty}^{\infty} e^{iuy} v(t_{m+1},y)\,dy\cdot\int_{-\infty}^{\infty} e^{-i(u-i\alpha)z}\,f(z)\,dz \\ &= P(t_m,t_{m+1})\cdot\mathcal{F}\{v(t_m,\cdot)\}(u)\cdot\phi(-(u-i\alpha))\end{aligned} \tag{5.11}$$

In the last step we used the fact that the complex-valued Fourier transform of the density is the extended characteristic function:

$$\phi(x+yi) = \int_{-\infty}^{\infty} e^{i(x+yi)z}\,f(z)\,dz \tag{5.12}$$

which is well-defined when $\phi(yi) < \infty$, as $|\phi(x+yi)| \le |\phi(yi)|$. As such (5.11) puts a restriction on the damping coefficient α, as it requires $\phi(\alpha i)$ to be finite.

The difference with the Carr-Madan approach in (2.31)-(2.35) is that we take a transform with respect to the log-spot price instead of the log-strike price, something which Lewis [2001] and Raible [2000] also consider for European option prices. The damping factor is again necessary when considering e.g. a Bermudan put, as then V (t_{m+1}, x) tends to a constant when $x \to -\infty$, and as such is not L^1-integrable. For the Bermudan put we must choose $\alpha > 0$. Though other values of α are allowed in principle, we need to know the payoff-transform itself in order to apply Cauchy's residue theorem, see Lewis [2001], Lee [2004] and Raible [2000]. This restriction on α will disappear when we switch to a discretised version of (5.11) in the next section. The Fourier transform of the damped continuation value can thus be calculated as the product of two functions, one of which, the extended characteristic function, is readily available in exponential Lévy models. We now recover the continuation value by taking the inverse Fourier transform of the right-hand side of (5.11), and calculate $V(t_m)$ as the maximum of the continuation and the exercise value at t_m. This procedure, as outlined in (5.2) is repeated recursively until we obtain the option price at time t_0. In pseudo-code the CONV algorithm is presented in Algorithm 5.1.

In Appendix 5.A we demonstrate how the hedge parameters can be calculated in the CONV method. As differentiation is exact in Fourier space, they will be more stable than when calculated via finite-difference based approximations. The following section deals with the implementation of the CONV algorithm. In particular we employ the FFT to approximate the continuous Fourier transforms that are involved.

$V(t_M, x) = E(t_M, x)$ for all x
$E(t_0, x) = 0$ for all x
For m = M-1 to 0
  Dampen $V(t_{m+1}, x)$ with $\exp(\alpha x)$ and take its Fourier transform
  Calculate the right-hand side of (5.11)
  Calculate $C(t_m, x)$ by applying Fourier inversion to (5.11) and undamping
  $V(t_m, x) = \max\{ E(t_m, x) , C(t_m, x) \}$
Next m

Algorithm 5.1: The CONV algorithm for Bermudan options

5.3. Implementation details and error analysis

The essence of the CONV method is the calculation of a convolution:

$$c(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-iux}\hat{v}(u)\phi(-(u-i\alpha))du \tag{5.13}$$

where $\hat{v}(u)$ is the Fourier transform of v:

$$\hat{v}(u) = \int_{-\infty}^{\infty} e^{iuy}v(y)\,dy \tag{5.14}$$

In the remainder of this section we will just focus on equations (5.13) and (5.14) for notational ease. To be able to use the FFT means we have to switch to logarithmic coordinates. For this reason the state variables x and y will represent $\ln S(t_m)$ and $\ln S(t_{m+1})$, up to a constant shift. Section 5.3.1 deals with the discretisation of the convolution in (5.13)-(5.14). Section 5.3.2 analyses the error made by one step of the CONV method and provides guidelines on choosing the grids for u, x and y. Section 5.3.3 considers the choice of grid further and investigates how to deal with points of discontinuity. This will prove to be important if we want to guarantee a smooth convergence of the algorithm. Finally, Sections 5.3.4 and 5.3.5 deal with the pricing of Bermudan and American options with the CONV method.

5.3.1. Discretising the convolution

We approximate both integrals in (5.13)-(5.14) by a discrete sum, so that the FFT algorithm can be employed for their computation. This necessitates the use of uniform grids for u, x and y:

$$u_0 = u_0 + j\Delta u \qquad x_j = x_0 + j\Delta x \qquad y_j = y_0 + j\Delta y \tag{5.15}$$

where j = 0, …, N-1. Though they may be centered around a different point, the x- and y-grids have the same mesh size: $\Delta x = \Delta y$. Further, the Nyquist relation must be satisfied:

$$\Delta u \cdot \Delta y = \frac{2\pi}{N} \tag{5.16}$$

In principle we could use the Fractional FFT algorithm (FrFT), which does not require the Nyquist relation to be satisfied. Numerical tests indicated however that this advantage of the FrFT does not outweigh the speed of the FFT, so we use the FFT throughout. Details about the location

of x_0 and y_0 will be given in Section 5.3.3. Inserting (5.14) into (5.13), and approximating (5.14) with a general Newton-Côtes rule and (5.13) with the left-rectangle rule yields:

$$c(x_p) \approx \Delta u \cdot \Delta y \cdot \tfrac{1}{2\pi} \sum_{j=0}^{N-1} e^{-iu_j x_p} \phi(-(u_j - i\alpha)) \sum_{n=0}^{N-1} w_n e^{iu_j y_n} v(y_n) \tag{5.17}$$

for p = 0, …, N-1. When using the trapezoidal rule we choose the weights w_n as:

$$w_0 = \tfrac{1}{2} \quad w_{N-1} = \tfrac{1}{2} \quad w_n = 1 \text{ for } n = 1, ..., N-2 \tag{5.18}$$

Though it may seem that using the left-rectangle rule in (5.17) would cause the leading error term to be $O(\Delta u)$, the error analysis will show that the Newton-Côtes rule one uses to approximate (5.14) is the determining factor. Inserting the definitions of our grids into (5.17) yields:

$$c(x_p) \approx e^{-iu_0(x_0 + p\Delta y)} \tfrac{1}{2\pi} \Delta u \sum_{j=0}^{N-1} e^{-ijp\frac{2\pi}{N}} e^{ij(y_0 - x_0)\Delta u} \phi\left(-(u_j - i\alpha)\right) \hat{v}(u_j) \tag{5.19}$$

where the Fourier transform of v is approximated by:

$$\hat{v}(u_j) \approx e^{iu_0 y_0} \Delta y \sum_{n=0}^{N-1} e^{ijn\frac{2\pi}{N}} e^{inu_0 \Delta y} w_n v(y_n) \tag{5.20}$$

Let us now define the DFT and its inverse of a sequence x_p, p = 0, …, N-1 as:

$$\mathcal{D}_j\{x_n\} = \sum_{n=0}^{N-1} e^{ijn\frac{2\pi}{N}} x_n \qquad \mathcal{D}_n^{-1}\{x_j\} = \tfrac{1}{N} \sum_{j=0}^{N-1} e^{-ijn\frac{2\pi}{N}} x_j \tag{5.21}$$

Though the reason why will become clear later, let us set $u_0 = -\tfrac{N}{2}\Delta u$. As $e^{inu_0\Delta y} = (-1)^n$ this finally leads us to write (5.19)-(5.20) as:

$$c(x_p) \approx e^{iu_0(y_0 - x_0)} (-1)^p \tfrac{1}{N} \mathcal{D}_p^{-1}\left\{ e^{ij(y_0 - x_0)\Delta u} \phi\left(-(u_j - i\alpha)\right) \cdot \mathcal{D}_j\left\{(-1)^n w_n v(x_n)\right\}\right\} \tag{5.22}$$

5.3.2. Error analysis for Bermudan options

A first inspection of (5.22) suggests that errors will arise from two[20] sources:

- Discretisation of both integrals in (5.13)-(5.14);
- Truncation of these integrals.

We will now consider both integrals in (5.13), (5.14) separately, and estimate both discretisation and truncation errors by applying the error analysis of Abate and Whitt [1992]. Lee [2004] recently combined an analysis similar to theirs with sharp upper bounds on European plain vanilla option prices to find a sharp error bound for the discretised Carr-Madan formula. Though it is possible to use parts of their analysis, we found that the resulting error bounds overestimated the true error of the discretised CONV formula. To be precise, the discretisation of (5.13) does not

[20] If the spot price for which we want to calculate our option price does not lie on the grid, another error source will be added as we will have to interpolate between option prices.

contribute to the error of (5.22) which is why we can use the left-rectangle rule to approximate (5.13). Based on a Fourier series expansion of the damped continuation value c(x), we will show why this is the case. This is natural, as the Fourier transform itself is generalised from Fourier series of periodic functions by letting their period approach infinity. We start from the risk-neutral valuation formula with damping and, for notational convenience, without discounting:

$$c(x) = \int_{-\infty}^{\infty} v(x+z)\, e^{-\alpha z} f(z)\, dz \tag{5.23}$$

Suppose that the density f(z) is negligible outside [-A/2, A/2], and that we are only interested in values of x in [-B/2, B/2]. According to (5.23), we require knowledge of v(x) for x in [-(A+B)/2, (A+B)/2]. Truncating the integration range in (5.23) leads to:

$$c(x) \approx \tilde{c}_1(x) = \int_{-A/2}^{A/2} v(x+z)\, e^{-\alpha z} f(z)\, dz \tag{5.24}$$

We can replace v by its Fourier series expansion on [-L/2, L/2], where we defined L = A+B:

$$\begin{aligned} \tilde{c}_1(x) &= \int_{-A/2}^{A/2} \sum_{j=-\infty}^{\infty} v_j e^{-ij\frac{2\pi}{L}(x+z)}\, e^{-\alpha z} f(z)\, dz \\ &= \sum_{j=-\infty}^{\infty} v_j e^{-ij\frac{2\pi}{L}x} \int_{-A/2}^{A/2} e^{-(\alpha + ij\frac{2\pi}{L})z} f(z)\, dz \end{aligned} \tag{5.25}$$

and the Fourier series coefficients of v are given by:

$$v_j = \frac{1}{L} \int_{-L/2}^{L/2} v(y) e^{ij\frac{2\pi}{L}y}\, dy \tag{5.26}$$

Secondly, we can replace the integral in (5.25) by the known characteristic function:

$$\tilde{c}_1(x) \approx \tilde{c}_2(x) = \sum_{j=-\infty}^{\infty} v_j e^{-ij\frac{2\pi}{L}x} \phi\left(-\left(j\frac{2\pi}{L} - i\alpha\right)\right) \tag{5.27}$$

The sum of both truncation errors now equals:

$$\begin{aligned} e_1(L) + e_2(L) &= \tilde{c}_2(x) - c(x) \\ &= \int_{\mathbb{R}\setminus[-\frac{A}{2}, \frac{A}{2}]} \left(v(x+z) - \sum_{j=-\infty}^{\infty} v_j e^{-ik\frac{2\pi}{L}(x+z)} \right) \cdot e^{-\alpha z} f(z)\, dz \end{aligned} \tag{5.28}$$

Note that only the parameter L will appear in the final discretisation. A general guideline for choosing L is to ensure that the mass of the density outside [-L/2, L/2] is negligible. The function $\tilde{c}_2$ can, at least on this interval, be interpreted as an approximate Fourier series expansion of c(x).

The third error, e_3, arises by truncating the summation from –N/2 to N/2-1, leading to $\tilde{c}_3$:

$$\begin{aligned} \tilde{c}_3(x) &= \sum_{j=-N/2}^{N/2-1} v_j e^{-ij\frac{2\pi}{L}x} \phi\left(-\left(j\frac{2\pi}{L} - i\alpha\right)\right) \\ \left| e_3(L,N) \right| &= \left| \tilde{c}_3(x) - \tilde{c}_2(x) \right| \le \sum_{|j|=N/2}^{\infty} |v_j| \cdot \left| \phi\left(-\left(j\frac{2\pi}{L} - i\alpha\right)\right) \right| \end{aligned} \tag{5.29}$$

To further bound this error we require knowledge about the rate of decay of Fourier coefficients. It is well known that even if v is only piecewise C^1 on [-L/2, L/2] its Fourier series coefficients v_j tend to zero as $j \to \pm\infty$. The modulus of v_j can therefore be bounded as:

$$| v_j | \leq \eta_1(L) \cdot | j |^{-\beta_1} \tag{5.30}$$

By $\eta_i(\cdot)$ we denote a constant, depending only on the quantities between its brackets. For functions that are piecewise continuous on [-L/2, L/2] but whose L-periodic extension is discontinuous, we have $\beta_1 = 1$. The following example demonstrates this is for a European put.

Example 5.1 – European put
Suppose that we have a European put payoff and that $y = \ln S(t) - \ln K$. Then the payoff function equals $v(y) = e^{\alpha y} K(1 - e^y)^+$ and its Fourier series coefficients equal:

$$v_j = K\left(e^{-\alpha L/2}(-1)^j \frac{e^{-L/2} - 1}{L(\alpha+1) + 2\pi i j} - L \frac{e^{-\alpha L/2}(-1)^j - 1}{(L(\alpha+1) + 2\pi i j)(L\alpha + 2\pi i j)} \right) \tag{5.31}$$

Clearly, $\beta_j = 1$ in (5.30), though when $L \to \infty$ and $j\frac{2\pi}{L} \to u$ it can be shown that the Fourier series coefficient converges to the Fourier transform of the payoff function, which can be seen to be $O(u^{-2})$ from (2.31).

The characteristic function can be assumed to have power decay:

$$| \phi(x + yi) | \leq \frac{\eta_2(y)}{| x |^{\beta_2}} \tag{5.32}$$

This is overly conservative for e.g. the Black-Scholes model, where the characteristic function of the log-return $\phi(x+yi)$ decays as $\exp(-cx^2)$, or the Heston model where the characteristic function has exponential decay. For the most popular Lévy models however the power decay assumption is appropriate. The VG model for example, cf. (3.29), has $\beta_2 = 2\tau/\nu$ with τ being the timestep between two exercise dates.

Remark 5.1
It should be noted that the error analysis here is valid for Bermudan options and *not* for American options in the limit when $\tau \to 0$. In Section 5.3.5 we will price American options by Richardson extrapolation on the prices of Bermudan options with a varying number of exercise opportunities. For problems where the time intervals are small and the characteristic function decays slowly, we may encounter some numerical problems due to the oscillatoriness of the integrand. These problems are however well-known and can in part be avoided by choosing a proper value of α.

Combining (5.30) and (5.32) yields:

$$\begin{aligned} \left| e_3(L,N) \right| &\leq \sum_{|j|=\frac{1}{2}N}^{\infty} \frac{\eta_1(L)}{|j|^{\beta_1}} \frac{\eta_2(\alpha)}{(\frac{2\pi}{L})^{\beta_2} |j|^{\beta_2}} \leq \eta_3(\alpha, L) \cdot \int_{\frac{1}{2}N-1}^{\infty} x^{-\beta_1-\beta_2} dx \\ &= \eta_3(\alpha, L) \cdot \frac{(\frac{1}{2}N - 1)^{1-\beta_1-\beta_2}}{\beta_1 + \beta_2 - 1} \end{aligned} \tag{5.33}$$

where $\eta_3(\alpha, L) = 2\eta_1(L)\eta_2(\alpha)(\frac{2\pi}{L})^{-\beta_2}$. We finally arrive at the discretised CONV formula in (5.22) by approximating the Fourier series coefficients of v in (5.29) with a Newton-Côtes rule:

$$\tilde{v}(u_j) = \tfrac{1}{L}\Delta y \sum_{n=0}^{N-1} w_n e^{iu_j y_n} v(y_n) \tag{5.34}$$

This is equal to the right-hand side of (5.19) divided by L. It becomes clear that we can set $\Delta y = L/N$ and $y_0 = -L/2$. Inserting (5.34) in $\tilde{c}_3$ results in the final approximation:

$$\tilde{c}_4(x) = \sum_{j=-N/2}^{N/2-1} \tilde{v}(u_j) e^{-ij\frac{2\pi}{L}x} \phi\left(-(j\tfrac{2\pi}{L} - i\alpha)\right) \tag{5.35}$$

Assuming that the chosen Newton-Côtes rule is of $O(N^{-\beta_3})$, one can bound:

$$\left| v_j - \tilde{v}(u_j) \right| \le \frac{\eta_4(\alpha, L)}{N^{\beta_3}} \tag{5.36}$$

leading to the following error estimate for $\beta_2 \neq 1$:

$$\begin{aligned} \left| e_4(L,N) \right| = \left| \tilde{c}_3(x) - \tilde{c}_4(x) \right| &\le \frac{\eta_4(\alpha, L)}{N^{\beta_3}} \sum_{j=-N/2}^{N/2-1} \left| \phi\left(-(j\tfrac{2\pi}{L} - i\alpha)\right) \right| \\ &\le \frac{\eta_4(\alpha, L)}{N^{\beta_3}} \left(3\phi(i\alpha) + 2\eta_2(\alpha)\left(\tfrac{2\pi}{L}\right)^{-\beta_2} \sum_{|j|=2}^{N/2} \frac{1}{|j|^{\beta_2}} \right) \\ &= \frac{\eta_5(\alpha, L)}{N^{\beta_3}} + \frac{\eta_6(\alpha, L)}{(1-\beta_2)N^{\beta_3}} \left(\frac{1}{N^{\beta_2 - 1}} - 1 \right) \end{aligned} \tag{5.37}$$

with $\eta_5(\alpha, L) = 3\eta_4(\alpha, L)\phi(i\alpha)$ and $\eta_6(\alpha, L) = 2\eta_2(\alpha)\eta_4(\alpha, L)\left(\frac{2\pi}{L}\right)^{-\beta_2}$. For $\beta_2 = 1$ the second error term should of course equal $\frac{\eta_6(\alpha, L) \cdot \ln(\frac{N}{2})}{N^{\beta_3}}$.

Summarising, if we use a Newton-Côtes to discretised the Fourier transform of the payoff function v, the error in the discretised CONV formula can be bounded as:

$$\begin{aligned} \left| c(x) - \tilde{c}_4(x) \right| &\le e_1(L) + e_2(L) + e_3(L,N) + e_4(L,N) \\ &= e_1(L) + e_2(L) + O\left(N^{-\beta_3 + \min(1-\beta_2, 0)}\right) \end{aligned} \tag{5.38}$$

As demonstrated, in most applications we will have $\beta_1 = 1$. The magnitude of β_3 will depend on the interplay between the chosen Newton-Côtes rule and the nature of the payoff function, something we investigate in the next section. However, let us assume that $\beta_3 \ge 2$, which we may expect if we use the trapezoidal rule or more sophisticated Newton-Côtes rules. This implies that, aside from truncation error the order of convergence will be:

- $O(N^{-\beta_3})$ for characteristic functions decaying faster than a polynomial;
- $O\left(N^{-\beta_3 - \min(0, \beta_2 - 1)}\right)$ for characteristic functions having power decay.

For the Black-Scholes model this implies that the order of convergence will be fully dictated by the chosen Newton-Côtes rule, whereas in the VG model, where $\beta_2 = 2\tau/\nu$, we can lose up to an order for sufficiently small timesteps.

One final word should be mentioned on the damping coefficient α. In the continuous version of the algorithm in Section 5.2 α was chosen such that the damped continuation value was L^1-integrable. The direct construction of the discretised CONV formula in Section 5.3.2 via a Fourier series expansion of the continuation value replaces L^1-integrability on $(-\infty,\infty)$ with L^1-summability on $[-L/2,L/2]$, so that the restriction on α is removed. In principle any value of α is allowed as long as $\varphi(i\alpha)$ is finite. Nevertheless it makes sense to adhere to the guidelines stated before, as the function will resemble its continuous counterpart more and more as L increases. The impact of α on the accuracy of the CONV algorithm is investigated in Section 5.4.1.

This concludes the error analysis of one step of the CONV algorithm. It is easy to show that the error is not magnified further in the remaining time steps. The leading error of our algorithm is therefore dictated by the time step where the order of convergence in (5.38) is the smallest.

Remark 5.2
We explicitly mention that aliasing, a commonly observed feature when dealing with a convolution of sampled signals by means of the FFT, is not a problem in our application. We encounter a convolution of the characteristic function and the DFT of a vector with option values. The DFT is periodical but this would make the convolution circular only if the characteristic function would also be obtained by a DFT. We can however work with the analytical characteristic function, which is not periodic.

5.3.3. Dealing with discontinuities

Our focus in this section lies on achieving smooth convergence for the CONV algorithm. As numerical experiments have shown that it is difficult to achieve smooth convergence with higher order Newton-Côtes rules, we will from here on focus on the second order trapezoidal rule in (5.18). Smooth convergence is desirable as we will be using extrapolation techniques later on to price American options in Section 5.3.5.

The previous section analysed the error in the discretised CONV formula when we use a Newton-Côtes rule to integrate the function V, the maximum of the continuation value and the exercise value. If we focus on a simple Bermudan put it is clear that already at the last time step this function will have a discontinuous first derivative. Certainly it is also possible that V itself is discontinuous, think of contracts with a barrier clause. This will affect the order of convergence.

It is well-known that if we numerically integrate a function with (a finite number of) discontinuities, we should split up the integration domain such that we are only integrating continuous functions. Appendix 5.B demonstrates this for the trapezoidal rule. In particular, we show that the trapezoidal rule remains second-order if only the first derivative of the integrand is discontinuous, at the cost of non-smooth convergence. If the integrand itself is discontinuous, the trapezoidal rule loses an order. Smooth second-order convergence can be restored by placing the discontinuities on the grid. This notion has often been utilised in lattice-based techniques, though the solutions have more often than not been payoff-specific. An approach that is more or less payoff-independent was recently proposed in Hu, Kerkhof, McCloud and Wackertapp [2006] generalising previous work by Hunt, Kennedy and Pelsser [2000]. Unfortunately, we cannot use their methodology here, as our desire to use the FFT binds us to a uniform grid.

Before investigating how to handle discontinuities in the CONV algorithm, we collect the results from the previous sections and restate the grid choice for the basic CONV algorithm. Equating the grids for x and y for now we have:

$$u_j = (j - \tfrac{n}{2})\Delta u \qquad x_j = y_j = (j - \tfrac{1}{2})\Delta y \qquad j = 0,..., N-1 \tag{5.39}$$

Here x and y represent, up to a constant shift, ln $S(t_m)$ and ln $S(t_{m+1})$, respectively. If in particular $x = \ln S(t_m) - \ln S(0)$ and $y = \ln S(t_{m+1}) - \ln S(0)$, so that x and y represent total log-returns, we will refer to this discretisation as Discretisation I. A convenient property of this discretisation is that the spot price always lies on the grid, so that no costly interpolation is required to determine the desired option value. Note that we need to ensure that the mass of the density of x and y outside [−L/2, L/2] is negligible. Though more sophisticated approximations can be devised, we use a rule of thumb from O'Sullivan which chooses L as a multiple of the standard deviation of ln $S(t_M)$, i.e.:

$$L = \delta \cdot \sqrt{-\frac{\partial^2 \phi(t_M, u)}{\partial u^2}\bigg|_{u=0} + \left(\frac{\partial \phi(t_M, u)}{\partial u}\bigg|_{u=0}\right)^2} \tag{5.40}$$

where $\phi(t_M, u)$ is the characteristic function of ln $S(t_M)$ (the terminal value of S) conditional upon ln S(0), and δ is a proportionality constant. Note that there is a trade-off in the choice of L: as we set $\Delta y = L/N$, the Nyquist relation implies $\Delta u = 2\pi/L$ and hence $[u_0, u_{N-1}] = [-N\pi/L, (N-2)\pi/L]$. While larger values of L imply smaller truncation errors, they also cause the range of the grid in the Fourier domain to be smaller, so that the error in turn will be larger initially.

A choice of grid that allows us to place one discontinuity on the grid is described here. Suppose that at time t_m the discontinuity we would like to place on the grid is d_m. We then shift our grid by a small amount to get:

$$x_j = \varepsilon_x + (j - \tfrac{L}{2})\Delta x \qquad y_j = \varepsilon_y + (j - \tfrac{L}{2})\Delta x \tag{5.41}$$

where $\varepsilon_x = d_m - \lceil d_m / \Delta x \rceil \cdot \Delta x$ and ε_y is chosen in a similar fashion. This discretisation will be referred to as Discretisation II. Even for plain vanilla European options where only one time step is required this is useful. By choosing $\varepsilon_y = \ln K/S(0)$ and $\varepsilon_x = 0$ we ensure that the discontinuity of the call or put payoff lies on the y-grid, and the spot price lies on the x-grid. When more discontinuities are present it seems impossible to guarantee smooth convergence while keeping the restriction of a uniform grid. In order to still be able to use the computational speed of the FFT we will then have to resort to e.g. the discontinuous FFT algorithm of Fan and Liu [2004] or a recent transform inversion technique in Den Iseger [2006]. These directions are left for further research. Discretisation II is however well-suited for the pricing of Bermudan and American options, as we will show in the following sections.

5.3.4. Pricing Bermudan options

It is well-known that in the case of American options under Black-Scholes dynamics the derivative of the value function is continuous (smooth fit principle). This is however not the case anymore when pricing Bermudan options, for which the function V in (5.2) will have a discontinuous first derivative. Though at the final exercise time t_M the location of this discontinuity is known, this is not the case at previous exercise times. All we know after approximating V is that the discontinuity is contained in an interval of width Δx, say $[x_\ell, x_{\ell+1}]$.

If we proceed with the CONV algorithm without placing the discontinuity on the grid, the algorithm will show a non-smooth convergence. In the QUAD method of Andricopoulos et al. this is overcome by equating the exercise payoff and the continuation value, and solving

numerically for the location of the discontinuity. In our framework this can be quite costly, so that we propose an effective alternative. We can use a simple linear interpolation to locate the discontinuity, say d_m:

$$d_m \approx \frac{x_{\ell+1}\left(C(t_m, x_\ell) - E(t_m, x_\ell)\right) - x_\ell\left(C(t_m, x_{\ell+1}) - E(t_m, x_{\ell+1})\right)}{\left(C(t_m, x_\ell) - E(t_m, x_\ell)\right) - \left(C(t_m, x_{\ell+1}) - E(t_m, x_{\ell+1})\right)} \tag{5.42}$$

We assume that the error made in determining in (5.42) is negligible compared to the other error terms appearing (see also the discussion in Appendix 5.B).

As in Discretisation II we can now shift the grid such that d_m lies on it, and recalculate both the continuation and the exercise value. In particular, note that the inner DFT of (5.22) does not have to be recalculated, the only term that is affected is the outer inverse DFT. Moreover, calculating dm automatically gives us an approximation of the exercise boundary.

It is demonstrated in Appendix 5.B that if we choose the trapezoidal rule a linear interpolation is sufficient to guarantee a smooth convergence. Obviously, if higher-order Newton-Côtes rules are used, higher order interpolation schemes will have to be employed to locate the discontinuity. The resulting algorithm we use to value Bermudan call or put options with a fixed strike K is presented below in pseudo-code.

```
Ensure that the strike price K lies on the grid by setting ε_y = ln K / S(0)
For m = M-1 to 1
   Equate the x-grid at t_m to the y-grid at t_{m+1}
   Compute C(t_m, x) through (5.22)
   Locate x_ℓ and x_{ℓ+1} and approximate d_m, e.g. via (5.42)
   Set ε_x = d_m and recomputed C(t_m, x)
   Calculate V(t_m, x) = max{ E(t_m, x) , C(t_m, x) }
   Set the y-grid at t_m to be equal to the x-grid at t_m
Next m
Set ε_x = 0 such that the initial spot price lies on the grid
Compute V(0, x) = C(0, x) using (5.22)
```

Algorithm 5.2: Details of the algorithm for valuing Bermudan options

5.3.5. Pricing American options

Within the CONV algorithm there are basically two approaches to value an American option. One way is to approximate an American option by a Bermudan option with many exercise opportunities, the other is to use Richardson extrapolation on a series of Bermudan options with an increasing number of exercise opportunities. The method we use has been described in detail by Chang, Chung, and Stapleton [2007], though the approach in finance dates back to Geske and Johnson [1984]. The QUAD method in Andricopoulos et al. [2003] also uses the same technique to price American options. We restrict ourselves to the essentials here. Let $V(\Delta t)$ be the price of a Bermudan option with a maturity of T years where the exercise dates are Δt years apart. It is assumed that $V(\Delta t)$ can be expanded as:

$$V(\Delta t) = V(0) + \sum_{i=1}^{\infty} a_i (\Delta t)^{\gamma_i} \tag{5.43}$$

with γ a positive and increasing sequence. V(0) is the price of the American option. Classical extrapolation procedures assume that the exponents γ_i are known, which means that we can use

n+1 Bermudan prices with varying Δt in order to eliminate n of the leading order terms in (5.43). The only paper we are aware of that considers an expansion of the Bermudan option price in terms of Δt is Howison [2007], who shows that $\gamma_1 = 1$ for the Black-Scholes model. Nevertheless, numerical tests indicate that the assumption $\gamma_i = i$ produces satisfactory results for the Lévy models we consider.

5.4. Numerical experiments

By various experiments we show the accuracy and speed of the CONV method. The method's flexibility is presented by showing results for three asset price processes, geometric Brownian motion (GBM, used in the Black-Scholes model), VG, and CGMY. In addition, we value a multi-asset option to give an impression of the CPU times required to value a basket option of moderate dimension. The pricing problems considered are of European, Bermudan and American style. We typically present the (positive or negative) error $V(0, S(0)) - V_{ref}(0, S(0))$, where the reference value $V_{ref}(0, S(0))$ is either obtained via another numerical scheme, or via the CONV algorithm with 2^{20} grid points. In the tables to follow we will also present the error convergence defined as the absolute value of the ratio between two consecutive errors. A factor of 4 then denotes second order convergence. All single-asset tests were performed in C++ on an Intel Xeon CPU 5160, 3.00GHz with 2 GB RAM. The multi-asset calculations were programmed in C on an Intel Core 2 CPU 6700, 2.66 GHz and 8 GB RAM. To assess the performance of the CONV method, we compare its speed and accuracy to that of two PIDE methods, one for the VG model, a Lévy process with infinite activity, and a recent PIDE scheme for Kou's jump-diffusion model.

5.4.1. The test setup

The CONV method, as outlined in Section 5.2, is particularly well-suited for exponentially affine models with closed-form characteristic functions. In this chapter we will use the extended CGMY/KoBoL model (from hereon extended CGMY model) of Boyarchenko and Levendorskiĭ [2002] and Carr et al. [2002], and Kou's model to analyse the performance of the CONV method.

The CGMY model is a univariate exponential Lévy model. Its underlying Lévy process is characterised by the triple $(\mu, \sigma, \nu_{CGMY})$, where the Lévy density is specified as:

$$\nu_{CGMY}(x) = \begin{cases} C\dfrac{\exp(-G|x|)}{|x|^{1+Y}} \\ C\dfrac{\exp(-G|x|)}{|x|^{1+Y}} \end{cases} \tag{5.44}$$

The parameters satisfy $C \geq 0$, $G \geq 0$, $M \geq 0$, and $Y < 2$. The condition $Y < 2$ is induced by the requirement that Lévy densities integrate x^2 in the neighbourhood of 0. Conveniently, the characteristic function of the log-asset price can be found in closed-form as:

$$\phi(u) = \exp\Big(t\left(i\mu u - \tfrac{1}{2}\sigma^2 u^2 + C\Gamma(-Y)\Big((M-iu)^Y - M^Y + (G+iu)^Y - G^Y\Big)\right)\Big) \tag{5.45}$$

where $\Gamma(x)$ is the gamma function. The drift μ follows from (2.7) as:

$$\mu = r - q - \tfrac{1}{2}\sigma^2 - G\Gamma(-Y)\Big((M-1)^Y - M^Y + (G+1)^Y - G^Y\Big) \tag{5.46}$$

One can verify that the parameters G and M represent respectively the smallest and largest finite moment in the model, as $\varphi(-iu) = E[S(t)^u]$ is infinite for $u < -G$ and for $u > M$. The model encompasses several models. When $\sigma = 0$ and $Y = 0$ we obtain the Variance Gamma (VG) model, which is often parameterised slightly differently with parameters[21] σ, θ and ν, as in Section 3.4.1. These are related to C, G and M through:

$$C = \frac{1}{\nu} \qquad G = \frac{\theta}{\sigma^2} + \sqrt{\frac{\theta^2}{\sigma^4} + \frac{2}{\nu\sigma^2}} \qquad M = -\frac{\theta}{\sigma^2} + \sqrt{\frac{\theta^2}{\sigma^4} + \frac{2}{\nu\sigma^2}} \tag{5.47}$$

Finally, when $C = 0$ the model collapses to the Black-Scholes model.

The other model we consider is Kou's model [2002]. Its characteristic function equals:

$$\phi(u) = \exp\left(t \cdot \left(iu\mu - \tfrac{1}{2}u^2\sigma^2 + iu\lambda\left(\frac{p}{\eta_+ - iu} - \frac{1-p}{\eta_- + iu} \right) \right) \right) \tag{5.48}$$

In Kou's jump-diffusion model jumps arrive via a Poisson process with intensity λ. The logarithm of each jump follows a double-exponential density.

The various parameter sets we will use throughout this section are supplied in Table 5.1. The two parameters that control the accuracy of the CONV method are δ from (5.40), which determines the range of the grid, and the damping coefficient α. For all GBM tests we set $\delta = 20$; for the other CGMY tests we use $\delta = 40$, as these models have fatter tails. The test set for Kou's model required a very wide grid, we used $\delta = 100$.

Test set	S	r	q	σ	C	G	M	Y	Other
T1-GBM	100	10%	0%	25%					
T2-VG[22]	100	10%	0%	0%	5	18.3663	37.8108	0	
T3-CGMY	1	10%	0%	0%	1	5	5	0.5	
T4-CGMY	90	6%	0%	0%	0.42	4.37	191.2	1.0102	
T5-GBM	40	6%	4%	20%					$\rho_{ij} = 25\%$
T6-VG[23]	1	10%	0%	0%	1	5.00001	5.00001	0	

Test set	S	r	q	σ	λ	P	η_-	η_+
T7-Kou	100	5%	0%	15%	10%	34.45%	3.0775	3.0465

Table 5.1: Parameter sets in the numerical experiments

Regarding the choice of α, we have demonstrated in Chapter 4 how to approximate the optimal damping coefficient when the payoff-transform is known, hereby increasing the numerical stability of the Carr-Madan formula. This is particularly effective for in/out-of-the-money options and options with short maturities. Though this rationale can to some extent be carried over to the pricing of European plain vanilla options in the CONV method (the difference being that now the payoff-transform is also approximated numerically), the problem becomes much more opaque when dealing with Bermudan options. To see this, note that the continuation value of the Bermudan option at the penultimate exercise date equals that of a European option. At each grid point, the European option will have a different degree of moneyness, calling for a

[21] The parameters σ and ν should not be confused with the volatility and Lévy density of the Lévy triplet.

[22] The actual parameters used were, in the VG parametrisation, for T2-VG: $\sigma = 0.12$, $\theta = -0.14$, $\nu = 0.2$, and for T6-VG: $\sigma = 0.282842$, $\theta = 0$ and $\nu = 1$.

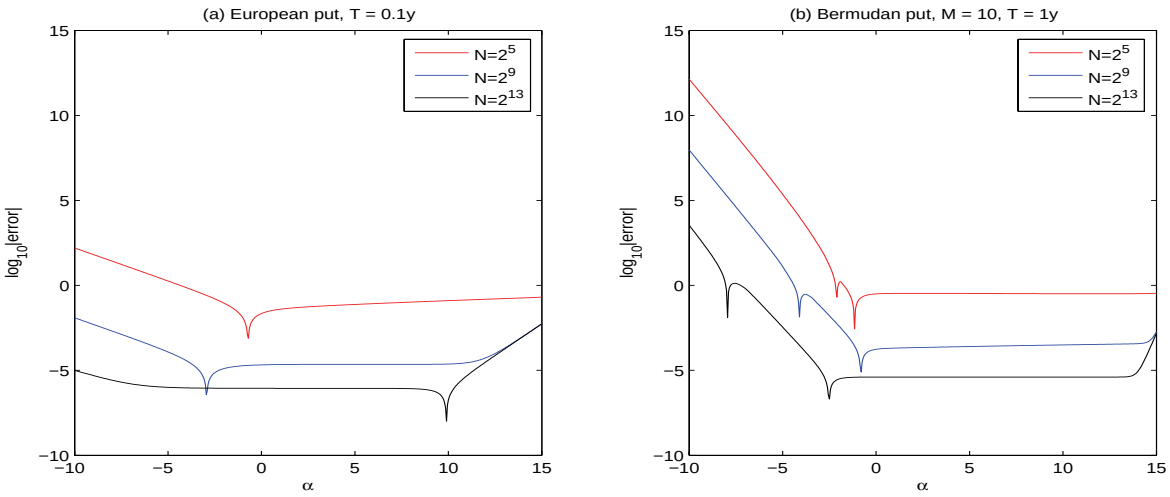


Figure 5.1: Error of CONV method under T2-VG and K = 110 for a European and Bermudan put in dependence of parameter α

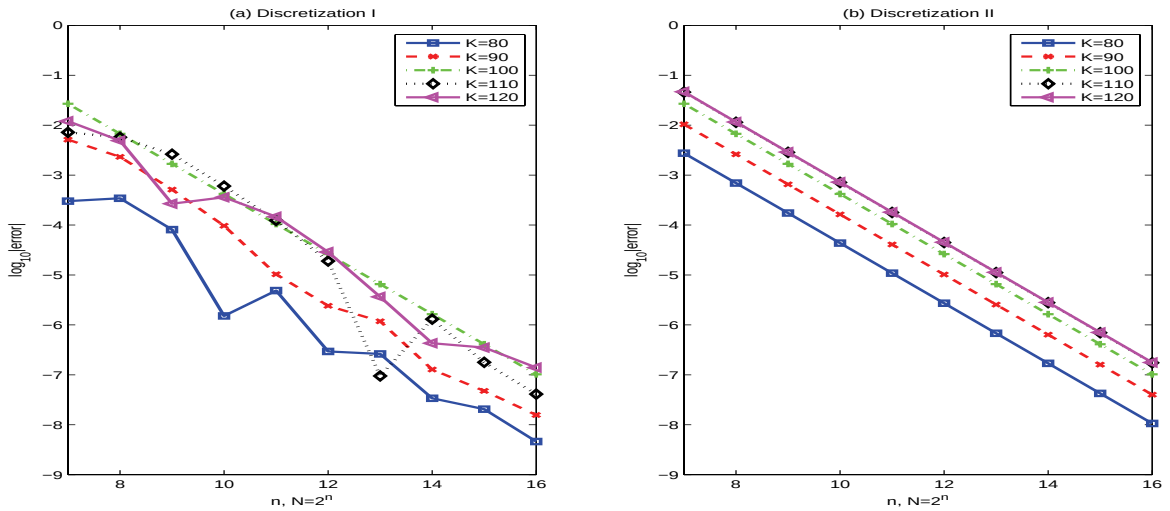


Figure 5.2: Convergence of Discretisation I and II for pricing European call options under T2-VG

different value of α per grid point. Which single choice for α will be optimal is not clear at all, a problem that becomes more complex as the number of exercise dates increases. What is evident from Figure 5.1, where we present the error of the CONV algorithm as a function of α for a European and a Bermudan put under T2-VG, is that there is a relatively large range for which the error is stable. In all numerical experiments we will set α = 0 which, at least for our examples, produces satisfactory results.

5.4.2. European call under GBM and VG

First of all, we evaluate the CONV method for pricing European options under VG. The parameters for the first test are from T2-VG with T = 1. Figure 5.2 shows that Discretisations I

and II generate results of similar accuracy. What we notice from Figure 5.2 is that the only option with a stable convergence in Discretisation I is the at-the-money option with K = 100. It is clear that placing the strike on the y-grid in Discretisation II ensures a regular second order convergence. The results are obtained in comparable CPU time. From the error analysis in Section 5.3.2 it became clear that for short maturities in the VG model, the slow decay of the characteristic function ($\beta_2 = 2\tau/\nu$) might impair the second order convergence. To demonstrate this, we choose a call option with a maturity of 0.1 years, and K = 90. Table 5.2 presents the error of Discretisation II for this option in models T1-GBM and T2-VG. The convergence under GBM is clearly of a regular second order. From the error analysis we expect the convergence under VG to be of first order. The non-smooth convergence observed in Table 5.2 is caused by the highly oscillatory integrand. Note that all reference values are based on an adaptive integration of the Carr-Madan formula, using the techniques from Chapter 4. All CPU times, in milliseconds, are determined after averaging the times of 1000 experiments. The Greeks of the GBM call from Table 5.2 are computed in Appendix 5.A.

n ($N = 2^n$)	**T1-GBM** Reference value: 11.1352431			**T2-VG** Reference value: 10.9937032		
	Time (msec)	**Error**	**Conv.**	**Time (msec)**	**Error**	**Conv.**
7	0.095	-2.08e-3	-	0.15	-2.91e-4	-
8	0.20	-5.22e-4	4.0	0.29	-1.42e-4	2.1
9	0.34	-1.30e-4	4.0	0.35	-4.61e-5	3.1
10	0.58	-3.26e-5	4.0	1.04	-9.49e-6	4.9
11	1.08	-8.15e-6	4.0	2.04	-8.55e-7	11.1
12	2.15	-2.04e-6	4.0	4.19	7.97e-7	1.1

Table 5.2: CPU time, error and convergence rate for European call options under T1-GBM and T2-VG, K = 90, T = 0.1 (using Discretisation II)

5.4.3. Bermudan option under GBM and VG

Turning to Bermudan options, we compare Discretisations I and II for 10-times exercisable Bermudan put options under both T1-GBM and T2-VG. The reference values reported in Table 5.3 and 4 are found by the CONV method with 220 grid points. It is shown in Tables 3 and 4 that both Discretisation I and II give results of similar accuracy. Discretisation I uses somewhat less CPU time, but Discretisation II shows a regular second order convergence, enabling the use of extrapolation. The computational speed of both discretisations is highly satisfactory.

n ($N = 2^n$)	**Discretisation I**			**Discretisation II**		
	Time (msec)	**Error**	**Conv.**	**Time (msec)**	**Error**	**Conv.**
7	0.13	9.09e-3	-	0.23	-2.72e-2	-
8	0.25	-1.29e-3	7.0	0.46	-7.36e-3	3.7
9	0.48	1.80e-6	717.8	0.90	-2.00e-3	3.7
10	1.09	2.71e-5	0.1	2.00	-5.22e-4	3.8
11	2.00	-9.31e-6	2.9	3.85	-1.32e-4	4.0
12	3.98	-1.31e-5	0.7	7.84	-3.31e-5	4.0

Table 5.3: CPU time, error and convergence rate for a 10-times exercisable Bermudan put under T1-GBM; K = 110, T = 1 with reference value 11.98745352

5.4.4. American options under GBM, VG and CGMY

Because Discretisation II yields a regular convergence, we choose it in this section to price American options. We compare the accuracy and CPU time of the two approximation methods

n ($N = 2^n$)	Discretisation I			Discretisation II		
	Time (msec)	Error	Conv.	Time (msec)	Error	Conv.
7	0.18	-8.45e-2	-	0.28	-9.63e-2	-
8	0.35	-9.02e-3	9.4	0.55	-1.07e-2	9.0
9	0.68	1.70e-4	53.1	1.09	-2.27e-3	4.7
10	1.33	2.04e-4	0.8	2.15	-6.06e-4	3.8
11	2.67	4.28e-5	4.8	4.38	-1.59e-4	3.8
12	5.64	1.11e-5	3.8	9.29	-4.08e-5	3.9

Table 5.4: CPU time, error and convergence rate for a 10-times exercisable Bermudan put under T2-VG; K =110, T = 1 with reference value 9.040646119

mentioned in Section 5.3.5, i.e. the direct approximation via a Bermudan option, and the repeated Richardson extrapolation technique. For the latter we opted for 2 extrapolations on 3 Bermudan options with 128, 64 and 32 exercise opportunities, which gave robust results. In our first test we price an American put under T1-GBM. The reference value was obtained by solving the Black-Scholes PDE on a very fine grid. The performance of both approximation methods is summarised in Table 5.5, where 'P(N/2)' denotes that the American option is approximated by an N/2-times exercisable Bermudan option. 'Richardson' denotes the results obtained by the 2-times repeated Richardson extrapolation scheme. It is evident that the extrapolation-based method converges fastest and costs far less CPU time than the direct approximation approach (e.g. to reach an accuracy of 10^{-4}, the extrapolation method is approximately 50 times faster). In Appendix 5.A the Greeks of the American put from Table 5.5 are computed.

n ($N = 2^n$)	Discretisation I			Discretisation II		
	Time (msec)	Error	Conv.	Time (msec)	Error	Conv.
7	0.18	-8.45e-2	-	0.28	-9.63e-2	-
8	0.35	-9.02e-3	9.4	0.55	-1.07e-2	9.0
9	0.68	1.70e-4	53.1	1.09	-2.27e-3	4.7
10	1.33	2.04e-4	0.8	2.15	-6.06e-4	3.8
11	2.67	4.28e-5	4.8	4.38	-1.59e-4	3.8
12	5.64	1.11e-5	3.8	9.29	-4.08e-5	3.9

Table 5.5: CPU time and errors for an American put under T1-GBM; K = 110, T = 1 with reference value 12.169417

n ($N = 2^n$)	T2-VG		T3-CGMY		T4-CGMY	
	Time (msec)	Error	Time (msec)	Error	Time (msec)	Error
7	3.42	-4.53e-2	3.82	4.58e-5	3.83	3.38e-2
8	6.85	4.26e-2	7.60	9.52e-5	7.68	6.63e-3
9	14.29	1.34e-2	15.87	-1.03e-4	15.78	-1.94e-3
10	28.99	-5.00e-3	32.21	-1.58e-5	33.37	-5.41e-6
11	61.67	-1.88e-2	68.16	-1.09e-5	68.59	-1.72e-4
12	135.09	1.31e-3	148.16	3.73e-6	147.96	-7.94e-5

Table 5.6: CPU time and errors for American puts under VG and CGMY

In the remaining tests we demonstrate the ability of the CONV method to price American options accurately under alternative dynamics, using the VG and both CGMY test sets. All reported reference values were generated with the CONV method on a mesh with 2^{20} points and 2-times Richardson extrapolation on 512-, 256- and 128-times exercisable Bermudans. We have included one CGMY test with $Y < 1$, and one with $Y > 1$, as the latter is considered a hard test case when numerically solving the corresponding PIDE. Both CGMY tests stem from the PIDE literature,

where reference values for the same American puts were reported as 0.112171 for T3-CGMY in Almendral and Oosterlee [2007a], and 9.2254842 for T4-CGMY in Wang et al. [2007]. The VG parameter set originally stems from Madan, Carr and Chang [1998], and is used in examples in Këllezi and Webber [2004], O'Sullivan [2005] and Maller et al. [2006]. In the latter paper the American option price is reported as 10. The reference values we use are calculated with the CONV method (using 2^{20} grid points and Richardson extrapolation on Bermudans with 512, 256 and 128 exercise opportunities) and agree up to four digits with the values from the literature. Though the convergence in Table 5.6 is less stable than for Bermudan options, the results in this section indicate that the CONV method is able to price American options under a wide variety of Lévy processes. A reasonable accuracy can be obtained quite quickly, so that it might be possible to calibrate a model to the prices of American options[23].

5.4.5. 4D basket options under GBM

The CONV method can easily be generalised to higher dimensions. The only assumption that the multi-dimensional model is required to satisfy is the independent increments assumption in (5.7). We do not state the multi-dimensional version of Algorithm 5.1 here as it is a trivial generalisation of the univariate case. Its ability to price options of a moderate dimension is demonstrated by considering a 4-asset basket put option. Upon exercise at time t_i, the payoff is:

$$V(t_i, S(t_i)) = \left(K - \tfrac{1}{4}\sum_{j=1}^{4} S_j(t_i)\right)^+ \tag{5.49}$$

The results of pricing a European and a 10-times exercisable Bermudan put under T5-GBM are summarised in Table 5.7. The CPU times on the tensor-product grids are very satisfactory,

N	European option		10-times exercisable Bermudan	
	Approximation	**Time (sec)**	**Approximation**	**Time (sec)**
16^4	1.6428	0.02	1.7721	0.15
32^4	1.6537	0.51	1.7390	3.12
64^4	1.6539	7.0	1.7394	61.6
128^4	1.6538	159.2	1.7393	1511.7

Table 5.7: CPU time and prices for multi-asset European and 10-times exercisable Bermudan basket put options under T5-GBM, K = 40 and T = 1

especially as the results on the coarse grids obtained in only a few seconds seem to have converged within practical tolerance levels. In order to be able to price higher-dimensional problems the multi-dimensional CONV method is combined with sparse grids in Leentvaar and Oosterlee [2007].

5.5. Comparison with PIDE methods

In this section we will compare the speed and accuracy of the CONV method to two PIDE schemes, one for the VG model which was used in Almendral and Oosterlee [2007b] and one recent scheme for Kou's model from Toivanen [2008]. An advantage of the CONV method over various PIDE schemes is that it is flexible with respect to the choice of model, whereas the integral term in PIDEs typically requires a very careful treatment, for example due to its weakly singular kernel for infinite activity Lévy models. Furthermore PIDE methods require a relatively

[23] The majority of exchange-traded options in the equity markets are American.

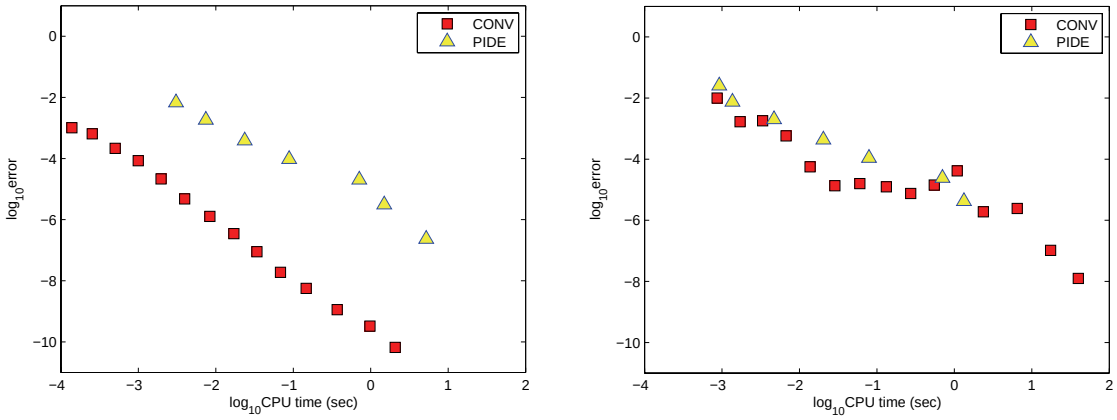


Figure 5.3: Comparison of error/CPU time of the CONV method with the PIDE solver of Almendral and Oosterlee [2007b] for Bermudan and American options

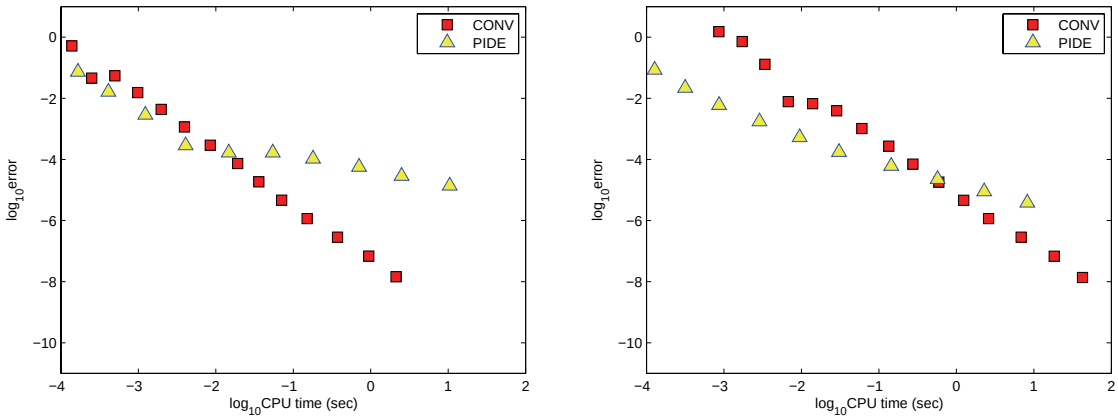


Figure 5.4: Comparison of error/CPU time of the CONV method with the PIDE solver of Toivanen [2008] for Bermudan and American options

fine discretisation in the time direction to guarantee an accurate representation of the solution, whereas in the CONV method we only require as many time steps as exercise dates. For Bermudan options with few exercise dates this is advantageous, though for American options it works in our disadvantage. A strong advantage of the PIDE methods is the ability to use coarse grids in the asset direction of the asset; in its present form the CONV method is restricted to using log-uniform grids in the asset direction.

At the end of the day the only fair comparison is to compare two implementations in the same computer language, on the same CPU, in terms of speed and accuracy. First we compare the PIDE scheme from Almendral and Oosterlee for the VG model to our method. Parameters are given by T6-VG in Table 5.1. Code for the PIDE scheme of Almendral and Oosterlee was available in Matlab. As we wrote the CONV code in both Matlab and C++, we were able to conduct a fair comparison. We found the CONV code, for large values of N, to be roughly three times as fast as the Matlab code, so we scaled CPU times in Figure 5.3 accordingly. For a

Bermudan option with relatively few exercise dates the CONV method is a clear winner. The advantage is reduced when pricing American options, as we price these by extrapolating the values of Bermudan options with a relatively large number of exercise dates. Nevertheless, in case of the VG model the CONV method still reaches a higher accuracy given the same computational budget as the PIDE scheme.

The second comparison is based on Kou's model [2002]. Toivanen's [2008] recent schemes for Kou's model utilises the log-double-exponential form of the jump density to derive efficient recursion formulae for evaluating the integral term in the PIDE. The benefits are clear: the complexity is reduced to O(MN), and in addition his schemes are no longer bound to uniform grids. Therefore it is to be expected that this method outperforms ours, which is not tailored to any specific model. Code for Toivanen's penalty method was made available to us in C++. We use parameter set T7-Kou from Table 5.1. In both examples in Figure 5.4 the PIDE scheme reaches a higher accuracy than the CONV method for small computational budgets. This is partially due to the fact that for this example we required a very wide grid for the CONV method ($\delta = 100$) in order to converge to the right solution, which implies a lower accuracy for small values of N. For the Bermudan option the CONV method is still competitive, yet for the American option the PIDE scheme is the clear favourite. Though the CONV method appears to converge faster than the penalty method, the PIDE scheme would be the method of choice for practical levels of accuracy.

Although we have only compared to two methods we believe it is fair to say that the CONV method will compare favourable to most PIDE schemes for the pricing of Bermudan options under Lévy models. There will however always be special cases, such as Kou's jump-diffusion model, for which one can design highly efficient PIDE schemes that are faster and more accurate than the CONV method.

5.6. Conclusions

In this chapter we have presented a novel FFT-based method for pricing options with early-exercise features, the CONV method. Like other FFT-based methods, it is flexible with respect to the choice of asset price process and the type of option contract, which has been demonstrated in numerical examples for European, Bermudan and American options. Path-dependent exotics can in principle also be valued by a forward propagation in time, though this has not been demonstrated here. The crucial assumption of the method is that the underlying assets are driven by processes with independent increments, whose characteristic function is readily available. Though we have mainly focused on univariate exponential Lévy models, the techniques presented here certainly also extend to multivariate models, as Section 5.4.5 has shown. The main strengths of the method are its flexibility and computational speed. By using the FFT to calculate convolutions we achieve a complexity of $O(MN \log_2 N)$, where N is the number of grid points andM is the number of exercise opportunities of the option contract. In comparison, the QUAD method of Andricopoulos et al. [2003] is $O(MN^2)$. We have compared the CONV method to two PIDE schemes. The conclusion of this experiment is that we expect the CONV method to have an edge over PIDE schemes for the pricing of Bermudan options, in particular in exponential Lévy models with infinite activity. However, there will always be special cases, such as the Black-Scholes model and Kou's jump-diffusion model for which highly efficient P(I)DE schemes can be designed. The speed of the method may make it possible to calibrate models to the prices of American options, as exchange-traded options are mainly of the American type. Future research will focus on the usage of more advanced quadrature rules, combined with speeding up the method for high-dimensional problems.

Appendix 5.A – The hedge parameters

Here we present the CONV formulae for two important hedge parameters, Δ and Γ:

$$\Delta = \frac{\partial V}{\partial S} = \frac{1}{S}\frac{\partial V}{\partial x} \qquad \Gamma = \frac{\partial^2 V}{\partial S^2} = \frac{1}{S^2}\left(-\frac{\partial V}{\partial x} + \frac{\partial^2 V}{\partial x^2}\right) \tag{5A.1}$$

As it is relatively easy to derive the corresponding CONV formulae, we merely present them here. For notational convenience we define:

$$\mathcal{F}\left\{e^{\alpha \cdot} V(t_0, \cdot)\right\}(u) = e^{-r\Delta t} A(u) \tag{5A.2}$$

where $A(u) = \mathcal{F}\left\{e^{\alpha \cdot} V(t_1, \cdot)\right\}(u) \cdot \phi(-u + i\alpha)$, and we assume $t_1 > 0$. We now obtain the CONV formula for Δ as:

$$\Delta = \frac{e^{-\alpha x - r\Delta t}}{S} \cdot \left(\mathcal{F}^{-1}\left\{-i \cdot A(\cdot)\right\}(x) - \alpha \mathcal{F}^{-1}\left\{A\right\}(x)\right) \tag{5A.3}$$

For Γ one can derive:

$$\Gamma = \frac{e^{-\alpha x - r\Delta t}}{S^2} \cdot \left(\mathcal{F}^{-1}\left\{-(i\cdot)^2 A(\cdot)\right\}(x) - (1 + 2\alpha)\mathcal{F}^{-1}\left\{-i \cdot A(\cdot)\right\}(x)\right) \tag{5A.4}$$

Note that the only additional calculations occur at the final step of the CONV algorithm, where we calculate the value of the option given the continuation and exercise values at time t_1. Since differentiation is exact in Fourier space the rate of convergence of the Greeks will be the same as that of the value. To demonstrate this we evaluate the delta and gamma under T1-GBM of the European call from Table 5.2 and the American put from Table 5.5. For both tests we choose Discretisation II. Tables A.1 and A.2 present the results. The reference values for the European call option are analytic solutions, for the American call these were found by numerically solving the Black-Scholes PDE on a very fine grid. Note that the delta and gamma of the American put converge to a slightly different value - this is due to our approximation of the American option via 2 Richardson extrapolations on 128-, 64- and 32-times exercisable Bermudans. If we would increase the number of exercise opportunities of the Bermudan options, the delta and gamma would, at the cost of a longer computation time, converge to their true values.

n ($N = 2^n$)	**T1-GBM: European call**			
	$\Delta_{ref} = 0.933029$		$\Gamma_{ref.} = 0.01641389$	
	Error	**Conv.**	**Error**	**Conv.**
7	-3.75e-4	-	3.79e-5	-
8	-9.37e-5	4.0	9.43e-6	4.0
9	-2.34e-5	4.0	2.35e-6	4.0
10	-5.86e-6	4.0	5.88e-7	4.0
11	-1.46e-6	4.0	1.47e-7	4.0
12	-3.66e-7	4.0	3.68e-8	4.0

Table 5A.1: Accuracy of hedge parameters for a European call under T1-GBM; K = 90, T = 0.1

n ($N = 2^n$)	**T1-GBM: American put**	
	$\Delta_{ref} = -0.62052$	$\Gamma_{ref.} = 0.0284400$
7	-0.62170	0.028498
8	-0.62035	0.028687
9	-0.62050	0.028464
10	-0.62053	0.028463
11	-0.62054	0.028463
12	-0.62055	0.028463

Table 5A.2: Values of hedge parameters for an American put under T1-GBM; K = 110, T = 0.1

Appendix 5.B – Error analysis of the trapezoidal rule

Suppose we are integrating a function $f \in \mathbb{C}^\infty$ over an interval [a,b]. The discretisation error induced by approximating this integral with the trapezoidal rule follows from the Euler-Maclaurin summation formula:

$$\int_a^b f(x)\,dx - T(a,b,f,\Delta x) = \sum_{j=1}^{\infty} (\Delta x)^{2j} \frac{B_{2j}}{(2j)!}\left(f^{(2j-1)}(b) - f^{(2j-1)}(a)\right) \tag{5A.5}$$

where B_j is the j-th Bernoulli number and $T(a,b,\Delta x)$ is the trapezoidal sum:

$$T(a,b,f,\Delta x) = \Delta x \sum_{j=0}^{N-1} f(x_j) - \tfrac{1}{2}\Delta x\left(f(a) + f(b)\right) \tag{5A.6}$$

with $\Delta x = \frac{b-a}{N-1}$ and $x_j = a + j\Delta x$. From (5A.5) it is clear that if the value of the first derivative is not the same in a and b, the trapezoidal rule is of order $1/N^2$.

The trapezoidal rule can obviously also be applied to functions that are piecewise continuously differentiable. The convergence may however be less stable if we do not know the exact location of the discontinuities. To see this, suppose that f can be written as:

$$f(x) = \begin{cases} g(x) & x \le z \\ h(x) & x > z \end{cases} \tag{5A.7}$$

Further, we define:

$$\ell = \max\left\{ j \mid x_j \le z, j = 0,\ldots,N-1 \right\} \tag{5A.8}$$

so that the interval $[x_\ell, x_{\ell+1}]$ contains z. Placing the discontinuity on the grid would result in the same order of convergence as the trapezoidal rule itself:

$$\begin{aligned} \int_a^b f(x)\,dx \approx\ & T(a,x_\ell,g,\Delta x) + T(x_{\ell+1},b,h,\Delta x) \\ & + \tfrac{1}{2}(z - x_\ell)\left(g(x_\ell) + g(z)\right) + \tfrac{1}{2}(x_{\ell+1} - z)\left(h(z) + h(x_{\ell+1})\right) \end{aligned} \tag{5A.9}$$

A straightforward application of the trapezoidal rule would lead to $T(a,b,f,\Delta x)$. The difference with (5A.9) is:

$$\begin{aligned}&\tfrac{1}{2}\Delta x g(x_\ell)+\tfrac{1}{2}\Delta x h(x_{\ell+1})-\tfrac{1}{2}(z-x_\ell)\big(g(x_\ell)+g(z)\big)\\&-\tfrac{1}{2}(x_{\ell+1}-z)\big(h(z)+h(x_{\ell+1})\big)\end{aligned} \tag{5A.10}$$

Expanding both g and h around the point of discontinuity z yields:

$$\begin{aligned}&\tfrac{1}{2}\big(x_{\ell+1}+x_\ell-2z\big)\big(g(z)-h(z)\big)+\tfrac{1}{2}(x_{\ell+1}-z)(z-x_\ell)\big(g^{(1)}(z)-h^{(1)}(z)\big)\\&+\tfrac{1}{2}(x_{\ell+1}-z)\sum_{j=1}^{\infty}\tfrac{1}{j!}(z-x_\ell)^j g^{(j)}(z)+\tfrac{1}{2}\big(z-x_\ell\big)\sum_{j=1}^{\infty}\tfrac{1}{j!}(x_{\ell+1}-z)^j h^{(j)}(z)\end{aligned} \tag{5A.11}$$

If f is continuous, but the first derivatives of g and h do not match at z, the order of convergence is still $1/N^2$ since $(x_{\ell+1}-z)(z-x_\ell)\le(\Delta x)^2$. It is clear that as N changes, the ratio of $(x_{\ell+1}-z)(z-x_\ell)$ to $(\Delta x)^2$ may vary strongly, leading to non-smooth convergence. If f is discontinuous, i.e. if the values of g and h in z disagree, the order of convergence is of order 1/N.

Now suppose that we have computed g and h at grid points x_j, j = 0, …, N-1. We know that g(z) = h(z), though we do not know the exact location of z. All we know is that it is contained in the interval $[x_\ell, x_{\ell+1}]$. This is a situation we encounter in the pricing of Bermudan options, as outlined in Section 5.3.4. If we proceed to integrate f on this grid, we will not obtain a smooth convergence. A simple approximation of the discontinuity can however be found by assuming a linear relationship between x and g(x) – h(x). This leads to:

$$z\approx\frac{x_{\ell+1}\big(g(x_\ell)-h(x_\ell)\big)-x_\ell\big(g(x_{\ell+1})-h(x_{\ell+1})\big)}{\big(g(x_\ell)-h(x_\ell)\big)-\big(g(x_{\ell+1})-h(x_{\ell+1})\big)}+O(\Delta x^2) \tag{5A.12}$$

where the error estimate follows from linear interpolation. Now suppose that we shift our grid (and recalculate g and h) such that either x_ℓ or $x_{\ell+1}$ coincide with this approximation of z, and redo the numerical integration. It is easy to see that smooth convergence will be restored, as the contribution of the error term in (5A.12) to the error term in (5A.10) will be of $O(\Delta x^3)$. Note that if we use higher-order Newton-Côtes rules, a higher order interpolation step will be required.

Chapter 6

A comparison of biased simulation schemes for stochastic volatility models[24]

Having dealt with quadrature-based pricing techniques in the previous chapters, we turn to simulation in this chapter. Most models used for the pricing of derivatives start from a set of stochastic differential equations (SDEs) that describe the evolution of certain financial variables, such as the stock price, interest rate or volatility of an asset. Since Monte Carlo simulation is often the method of choice for the valuation of exotic derivatives due to its ability to handle both early exercise and path dependent features with relative ease, it is important to know exactly how to simulate the evolution of the variables of interest. Obviously, if the SDEs can be solved such that the relevant variables can be expressed as a function of a finite set of state variables for which we know the joint distribution, the problem is reduced to sampling from this distribution. This is for example the case with the Black-Scholes model.

Unfortunately not all models allow for such simple representations. For these models the conceptually straightforward Euler-Maruyama (Euler for short) discretisation can be used, see e.g. Kloeden and Platen [1999], Jäckel [2002] or Glasserman [2003]. The Euler scheme discretises the time interval of interest, such that the financial variables are simulated on this discrete time grid. Under certain conditions it can be proven that the Euler scheme converges to the true process as the time discretisation is made finer and finer. Nevertheless, the disadvantages of such a discretisation are clear. Firstly, the magnitude of the bias is unknown for a certain time discretisation, so that one will have to rerun the same simulation with a finer discretisation to check whether the result is sufficiently accurate. Secondly, the time grid required for a certain accuracy may be much finer than is strictly necessary for the derivative under consideration – many trades only depend on the realisation of the processes at a small number of dates. Clearly, if exact and efficient simulation methods can be devised for a model, they should be preferred.

The model, or class of models we consider in this chapter is referred to as the CEV-SV model, see e.g. Andersen and Brotherton-Ratcliffe [2005] and Andersen and Piterbarg [2007]. The asset price process (S) and the variance process (V) evolve according to the following SDEs, specified under the risk-neutral probability measure:

$$\begin{aligned} dS(t) &= \mu S(t)dt + \lambda\sqrt{V(t)}\,S(t)^{\beta}\,dW_S(t) \\ dV(t) &= -\kappa\big(V(t)-\theta\big)dt + \omega V(t)^{\alpha}\,dW_V(t) \end{aligned} \tag{6.1}$$

Here μ is the risk neutral drift of the asset price, κ is the speed of mean-reversion of the variance, θ is the long-term average variance, and ω is the so-called volatility of variance or volatility of volatility. Finally, λ is a scaling constant and W_S and W_V are correlated Brownian motions, with instantaneous correlation coefficient ρ.

[24] This chapter will appear as Lord, R., Koekkoek R. and D. van Dijk [2008]. "A comparison of biased simulation schemes for stochastic volatility models", *Quantitative Finance*.

To simplify the exposition, we will mainly concentrate on the special affine case $\alpha = ½$ and $\beta = 1$, leading to the popular Heston [1993] model which already featured prominently in the previous chapters. The best performing schemes will however also be tested in a more general example. The Heston model was heavily inspired by the interest rate model of Cox, Ingersoll and Ross [1985], who used the same mean-reverting square root process to model the spot interest rate. It is well known that, given an initial nonnegative value, a square root process cannot become negative, see e.g. Feller [1951], giving the process some intuitive appeal for the modelling of interest rates or variances. The Heston model is often used as an extension of the Black-Scholes model to incorporate stochastic volatility, and is often used for product classes such as equity and foreign exchange, although extensions to an interest rate context also exist, see e.g. Andersen and Andreasen [2002] and Andersen and Brotherton-Ratcliffe [2005].

Although pricing in the Cox-Ingersoll-Ross (CIR) and Heston models is a well-documented topic, most textbooks seem to avoid the issue of how to simulate these models. If we focus purely on the mean-reverting square-root component of (6.1), there is not a real problem, as Cox et al. [1985] found that the conditional distribution of V(t) given V(s) is noncentral chi-squared. Both Glasserman [2003] and Broadie and Kaya [2006] provide a detailed description of how to simulate from such a process. Combining this algorithm with recent advances on the simulation of gamma random variables by Marsaglia and Tsang [2000] will lead to a fast and efficient simulation of the mean-reverting square root process.

Complications arise, however, when we superimpose a correlated asset price, as in (6.1). As there is no straightforward way to simulate a noncentral chi-squared increment together with a correlated normal increment for the asset price process, the next idea that springs to mind is an Euler discretisation. This involves two problems, the first of which is of a practical nature. Despite the domain of the square root process being the nonnegative real line, for any choice of the time grid the probability of the variance becoming negative at the next time step is strictly greater than zero. As we will see, this is much more of an issue in a stochastic volatility context than in the CIR interest rate model, due to the much higher values typically found for the volatility of variance ω. Practitioners have therefore often opted for a quick "fix" by either setting the process equal to zero whenever it attains a negative value, or by reflecting it in the origin, and continuing from there on. These fixes are often referred to as *absorption* or *reflection*, see e.g. Gatheral [2006]. Interestingly this problem also arises in a discrete time setting, a lead we follow up on in the final section.

The second problem is of both a theoretical and practical nature. The usual theorems leading to strong or weak convergence in Kloeden and Platen [1999] require the drift and diffusion coefficients to satisfy a linear growth condition, as well as being globally Lipschitz. Since the square root is not globally Lipschitz, convergence of the Euler scheme is not guaranteed. Although recently Albin et al. [2005] relax the global Lipschitz condition somewhat, their results are only applicable to processes on an open interval, whereas the domain of the square root process is $[0,\infty)$, with 0 being an attainable boundary. For this reason, various alternative methods have been used to prove convergence of particular discretisations for the square root process. We mention Deelstra and Delbaen [1998], Diop [2003], Bossy and Diop [2004], Berkaoui, Bossy and Diop [2008] and Alfonsi [2005], who deal with the square root process in isolation.

It is only recently that papers dealing with the simulation of the Heston model in its full glory have started appearing. Andersen and Brotherton-Ratcliffe [2005] were among the first to suggest an approximation scheme for (6.1) which preserves the positivity of both S and V for general values of α and β. In Broadie and Kaya [2004,2006] an exact simulation algorithm has been devised for the Heston model. In numerical comparisons of their algorithm to an Euler discretisation with the absorption fix, they find that for the pricing of European options in the Heston model and variations thereof, the exact algorithm compares favourably in terms of root-mean-squared (RMS) error. Their algorithm is however highly time-consuming, as we will see,

and therefore certainly not recommendable for the pricing of strongly path dependent options that require the value of the asset price on a large number of time instants. Higham and Mao [2005] considered an Euler discretisation of the Heston model with a novel fix, for which they prove strong convergence. To the best of our knowledge they are the first to rigorously prove that using an Euler discretisation in the Heston model is theoretically correct, by proving that the sample averages of certain options converge to the true values. Unfortunately they do not provide numerical results on the convergence of their fix compared to other Euler fixes. The recent paper of Kahl and Jäckel [2006] considers a number of discretisation methods for a wide range of stochastic volatility models. For the Heston model they find that their IJK-IMM scheme, a quasi-second order scheme tailored specifically toward stochastic volatility models, gives the best results. Their numerical results are however not comparable to those of Broadie and Kaya, as they use a strong convergence measure which cannot directly be related to an RMS error. Finally we should mention the simulation schemes recently constructed by Andersen [2008]. As this paper compares to our full truncation scheme and as it postdates an initial version of this chapter, we chose not to include these schemes in our comparison. The schemes, specifically tailored for the Heston model, seem to produce a smaller bias than any scheme considered in this chapter, at the cost of a more complex implementation.

The contribution of this article is threefold. Firstly, we unify all Euler discretisations corresponding to the different fixes for the problem of negative variance known thus far under a single framework. Secondly, we propose a new fix, called the full truncation scheme. Full truncation is a modification of the Euler scheme of Deelstra and Delbaen [1998], which we will refer to as the partial truncation method. The difference between both methods lies in the treatment of the drift. Whereas partial truncation only truncates terms involving the variance in the diffusion of the variance, full truncation also truncates within the drift. In both schemes however the variance process itself remains negative. Both schemes are extended to (6.1). Following the train of thought of Higham and Mao, we are able to prove strong convergence for both of these fixes. With this proof in hand the pricing of plain vanilla options and certain exotics via Monte Carlo is justified, as we can then appeal to the results of Higham and Mao. Thirdly and finally, we numerically compare all Euler fixes to the other schemes mentioned above in terms of the size of the bias, as well as RMS error given a certain computational budget.

The article is set up as follows. Section 6.1 deals with the CEV-SV model and its properties. Section 6.2 considers simulation schemes for the Heston model. In Section 6.3 we consider Euler schemes for the CEV-SV model and introduce the full truncation scheme, for which we prove strong convergence. Section 6.4 provides numerical results, whereas Section 6.5 concludes.

6.1. The CEV-SV model and its properties

For reasons of clarity, we repeat equation (6.1) here, which specifies the dynamics of the asset price and variance process in the CEV-SV model under the risk neutral probability measure:

$$\begin{aligned} dS(t) &= \mu S(t)dt + \lambda\sqrt{V(t)}\,S(t)^{\beta}\,dW_S(t) \\ dV(t) &= -\kappa\big(V(t)-\theta\big)dt + \omega V(t)^{\alpha}\,dW_V(t) \end{aligned} \tag{6.2}$$

We restrict β to be lie in (0,1] and α to be positive. This model is analysed in great detail in Andersen and Piterbarg [2007]. Before turning to the issue of the simulation of (6.2) in general and the Heston model in particular, we briefly mention some well-known properties of the process V(t) and S(t) that we require in the remainder of this chapter. The mean-reverting CEV process V(t) has the following properties:

i) 0 is always an attainable boundary for $0 < \alpha < ½$;
ii) 0 is an attainable boundary when $\alpha = ½$ and $\omega^2 > 2\kappa\theta$. The boundary is strongly reflecting;
iii) 0 is unattainable for $\alpha > ½$;
iv) ∞ is an unattainable boundary.

Via the Yamada condition it can be verified that the SDE for V(t) has a unique strong solution when $\alpha \geq ½$. For $\alpha < ½$ we impose that the process for V(t) is reflected in the origin. All properties follow from the classical Feller boundary classification criteria (see e.g. Karlin and Taylor [1981]). Turning to the condition $\omega^2 > 2\kappa\theta$, we mention that to calibrate the Heston model to the skew observed in equity or FX markets, one often requires large values for the volatility of variance ω, see e.g. the calibration results in Duffie, Pan and Singleton [2000] where $\omega \approx 60\%$. In the CIR model ω, then representing the volatility of interest rates, is markedly lower, see e.g. the calibration results in Brigo and Mercurio [2001, p. 115] where this parameter is around 5%. Moreover, the product $\kappa\theta$ is usually of the same magnitude in both models if we use a deterministic shift extension to fit the initial term structure in the CIR model, so that it is safe to say that for typical parameter values the origin will be attainable within the Heston model, whereas in the CIR interest rate model it will not. Concerning ii) we mention that strongly reflecting here means that the time spent in the origin is zero - V(t) can touch zero, but will leave it immediately. The interested reader is referred to Revuz and Yor [1991] for more details.

Turning to the asset price process in the CEV-SV model, Andersen and Piterbarg [2007] prove that the process S can reach 0 with a positive probability. To ensure that the SDE in (6.2) has a unique solution, they impose the natural boundary condition that:

v) S(t) has an absorbing barrier at 0.

We do the same here, and mention that v) seems to be consistent with the asymptotic expansion derived for the SABR model in Hagan, Kumar, Lesniewski and Woodward [2002]. The SABR model is a special case of an CEV-SV model with $\theta = 0$, $\kappa = -\omega^2/4$ and $\alpha = 1$.

The following section specifically considers the simulation of the Heston model as this model is of great practical importance.

6.2. Simulation schemes for the Heston model

We now turn to the simulation of (6.2) when $\alpha = ½$ and $\beta = 1$, i.e. the Heston model. Obviously there are myriads of schemes one could use to simulate the Heston model. Though we by no means aim to be complete, we outline some schemes here that yield promising results or are frequently cited. We postpone the treatment of Euler schemes to the next section. Firstly, we demonstrate why in case of the Heston model it is not wise to change coordinates to the volatility, i.e. the square root of V. Secondly, we briefly discuss the exact simulation method of Broadie and Kaya [2006]. Finally, we take a look at alternative discretisations, in particular the quasi-second order schemes of Ninomiya and Victoir [2004] and Kahl and Jäckel [2006].

Apart from the schemes considered in this section, lately a number of papers have appeared in which splitting schemes are considered for mean-reverting CEV processes, see e.g. Moro [2004] and Dornic, Chaté and Muñoz [2005] and Moro and Schurz [2007]. The schemes in these papers heavily rely on an exact solution being known for a subsystem of the original SDE. Whilst this is certainly the case for univariate mean-reverting CEV processes, it does not seem likely that such a splitting can be found for the full-blown CEV-SV model. For this reason we do not further consider these schemes here, though the topic does warrant further study.

6.2.1. Changing coordinates

For reasons of increased speed of convergence it is often preferable to transform an SDE in such a way that it obtains a constant volatility term, see e.g. Jäckel [2002, section 4.2.3]. If we do this for the process V(t) in (6.2) with $\alpha = ½$, we can achieve this by considering volatility itself:

$$d\sqrt{V(t)} = \left(\frac{\kappa\theta - \frac{1}{2}\omega^2}{2\sqrt{V(t)}} - \tfrac{1}{2}\kappa\sqrt{V(t)} \right) dt + \tfrac{1}{2}\omega dW_V(t) \tag{6.3}$$

Although this transformation is seemingly correct, we are only allowed to apply Itō's lemma if the square root is twice differentiable on the domain of V(t). However, since the origin is attainable for $\omega^2 > 2\kappa\theta$, and the square root is not differentiable in zero, the process obtained by incorrectly applying Itō's lemma is structurally different, as is also mentioned in Jäckel [2004]. Even when the origin is inaccessible, the numerical behaviour of the transformed equation is rather unstable. Unless $\omega^2 = 2\kappa\theta$, when V(t) is sufficiently small, the drift term in (6.3) will blow up, temporarily assigning a much too high volatility to the stock price, in turn greatly distorting the sample average of the Monte Carlo simulation. Luckily, anyone trying to implement (6.3) will pick up this feature rather quickly, as will be illustrated in the numerical results in Section 6.3. We mention that similar issues arise with other coordinate transformations, such as switching to the logarithm of V(t).

6.2.2. Exact simulation of the Heston model

As mentioned, Broadie and Kaya [2004, 2006] have recently derived a method to simulate without bias from the Heston stochastic volatility model in (6.2). Part of this algorithm was already considered in Section 3.4.3. Although we refer to their papers for the exact details, we outline their algorithm here to motivate why it is highly time-consuming. First of all a large part of their algorithm relies on the result that for $s \le t$, V(t) conditional upon V(s) is, up to a constant scaling factor, noncentral chi-squared:

$$V(t) \sim \frac{\omega^2(1-e^{-\kappa(t-s)})}{4\kappa} \chi_\nu^2\left(\frac{4\kappa e^{-\kappa(t-s)} V(s)}{\omega^2(1-e^{-\kappa(t-s)})} \right) \tag{6.4}$$

where $\chi_\nu^2(\xi)$ is a noncentral chi-squared random variable with ν degrees of freedom and non-centrality parameter ξ. The degrees of freedom are equal to $\nu = 4\kappa\theta\omega^{-2}$. Glasserman [2003] as well as Broadie and Kaya show how to simulate from a noncentral chi-squared distribution. Combining this with recent advances by Marsaglia and Tsang [2000] on the simulation of gamma random variables (the chi-squared distribution is a special case of the gamma distribution), leads to a fast and efficient simulation of V(t) conditional upon V(s).

Secondly, let us define $V(s,t) = \int_s^t V(u)\,du$ and $V_a(s,t) = \int_s^t \sqrt{V(u)}\,dW_a(u)$ for a = S,V. First of all Broadie and Kaya recognized that integrating the equation for the variance yields:

$$V(t) = V(s) - \kappa V(s,t) + \kappa\theta(t-s) + \omega V_V(s,t) \tag{6.5}$$

so that we can calculate $V_V(s,t)$ if we know V(s), V(t) and V(s,t). Knowing all these terms, and

solving for ln S(t) conditional upon ln S(s) yields the final step:

$$\ln S(t) \sim N\left(\ln S(s) + \mu(t-s) - \tfrac{1}{2}V(s,t) + \rho V_V(s,t), (1-\rho^2)V(s,t)\right) \quad (6.6)$$

where N indicates the normal distribution. The algorithm can thus be summarised by:

1. Simulate V(t), conditional upon V(s) from (6.4)
2. Simulate V(s,t) conditional upon V(t) and V(s)
3. Calculate $V_V(s,t)$ from (6.5)
4. Simulate S(t) given V(s,t), $V_V(s,t)$ and S(s), by means of (6.6)

Algorithm 6.1: Exact simulation of the Heston model by Broadie and Kaya

The crucial and time-consuming step is the one we skipped over for a reason – step 2. Broadie and Kaya show how to derive the characteristic function of V(s,t) conditional upon V(t) and V(s). We considered this characteristic function in Section 3.4.3, equation (3.39), and showed how to evaluate it in order to avoid complex discontinuities due to the complex logarithm in it. Step 2 utilises the transform method, so that one has to numerically invert the cumulative distribution function, itself found by the numerical Fourier inversion of the characteristic function. Since the characteristic function non-trivially depends on the two realisations V(s) and V(t) via e.g. modified Bessel functions of the first kind, it is not trivial to cache a major part of the calculations. Hence we must repeat this step at each path and date that is relevant for the derivative at hand. It suffices to say that this makes step 2 very time-consuming and unsuitable for highly path-dependent exotics.

6.2.3. Quasi-second order schemes

In Glasserman [2003, pp. 356-358], a quasi-second order[25] Taylor scheme is considered. Its convergence is found to be rather erratic, which is one of the reasons why Broadie and Kaya [2006] chose not to compare their exact scheme to second order Taylor schemes. A closer look at Glasserman's scheme shows the probable cause of this erratic convergence – the discretisation contains terms which are very similar to the drift term in (6.3), and can therefore become quite large when V(t) is small. Since then, two papers have applied second order schemes to either the mean-reverting square root process or the Heston model in its full-fledged form, namely Alfonsi [2005] and Kahl and Jäckel [2006]. We start with the latter. After comparing a variety of schemes, Kahl and Jäckel conclude that at least for the Heston model applying the implicit Milstein method[26] (IMM) to the variance, combined with their bespoke IJK scheme for the logarithm of the stock price, yields the best results as measured by a strong convergence measure. Their results indicate that their scheme by far outperforms the Euler schemes with the absorption fix. The IMM method discretises the variance as follows:

$$V(t+\Delta t) = V(t) - \kappa\Delta t\left(V(t+\Delta t) - \overline{V}\right) + \omega\sqrt{V(t)}\cdot\Delta W_V(t) + \tfrac{1}{4}\omega^2\cdot\left(\Delta W_V(t)^2 - \Delta t\right) \quad (6.7)$$

The IMM method actually preserves positivity for the mean-reverting square root process, provided that $\omega^2 < 4\kappa\theta$, see Kahl [2004]. Unfortunately, this condition is not frequently satisfied

[25] By quasi-second order we mean schemes that do not simulate the double Wiener integral.

[26] Though they consider the balanced Milstein method (BMM), for the square root process their control functions (see their figure 6) coincide with the implicit Milstein method. From now on we will therefore refer to their scheme as the IJK-IMM scheme.

in an implied calibration of the Heston model. For values outside this range, a fix is again required. The best scheme for the logarithm of the stock price is their IJK scheme:

$$\begin{aligned}\ln S(t+\Delta t) = \ln S(t) + \mu\Delta t - \tfrac{1}{4}\Delta t\big(V(t)+V(t+\Delta t)\big) + \rho\sqrt{V(t)}\cdot\Delta W_V(t) \\ + \tfrac{1}{2}\Big(\sqrt{V(t)}+\sqrt{V(t+\Delta t)}\Big)\cdot\big(\Delta W_S(t)-\rho\Delta W_V(t)\big) + \tfrac{1}{4}\omega\rho\left(\Delta W_V(t)^2-\Delta t\right)\end{aligned} \tag{6.8}$$

which is specifically tailored to stochastic volatility models, where typically ρ is highly negative. For more details on both discretisations, we refer the interested reader to Kahl [2004] and Kahl and Jäckel [2006]. In the remainder we will refer to (6.7)-(6.8) as the IJK-IMM scheme.

Alfonsi [2005] deals with the mean-reverting square root process in isolation, and develops an implicit scheme that also preserves positivity by considering the transformed equation (6.3). The range of parameters for which the scheme works is again $\omega^2 < 4\kappa\theta$. He also considers Taylor expansions of this implicit scheme, the best of which (his E(0) scheme) is equivalent to (6.7) to first order in Δt. We therefore purely focus on Kahl and Jäckel's scheme in our numerical results. As an interesting sidenote, the E(0) scheme coincides exactly with a special case of the variance equation in the Heston and Nandi [2000, Appendix B] model, which they show converges to the mean-reverting square-root process as the time step tends to zero.

Finally, we consider a second-order scheme proposed in Ninomiya and Victoir [2004] for SDEs whose drift and diffusion coefficients are smooth functions with bounded derivatives of any order. Though the scheme converges weakly with order 2, it does not seem applicable to the Heston model – the first derivative of the square root function is already not bounded. The example the authors consider however is based in the Heston model, and does, for their choice of parameters, seem to have a second order convergence. Nevertheless, as the technical conditions on the drift and diffusion coefficients are not satisfied, we will refer to the scheme as a quasi-second order scheme.

Let us first describe their scheme for a fully general SDE in Stratonovich form:

$$d\mathbf{Y}(t) = \mathbf{g}_0\big(t,\mathbf{Y}(t)\big)dt + \sum_{i=1}^{d}\mathbf{g}_i\big(t,\mathbf{Y}(t)\big)\circ dW_i(t) \tag{6.9}$$

where $\mathbf{Y} \in \mathbb{R}^n$ and $\mathbf{g}_i : \mathbb{R}^n \to \mathbb{R}^n$ for i = 0, ... d are smooth functions whose derivatives of any order are bounded. Starting from **y**(t), a discretisation of **Y**(t), the value at the next time step is:

$$\mathbf{y}(t+\Delta t) = \mathbf{y}_{d+1}\big(\tfrac{1}{2}\Delta t\big) \tag{6.10}$$

which is found by solving the following d+2 ordinary differential equations (ODEs):

$$\frac{d\mathbf{y}_i}{dt} = \begin{cases}\mathbf{g}_i & \text{if } \Lambda(t) = -1 \\ \mathbf{g}_{d+1-i} & \text{if } \Lambda(t) = 1\end{cases} \quad \text{subject to } \mathbf{y}_i(0) = \mathbf{y}_{i-1}\big(Z_{i-1}(t)\sqrt{t}\big) \tag{6.11}$$

for i = 0, ..., d+1. With the exception of $Z_0 = \tfrac{1}{2}\sqrt{\Delta t}$, all $Z_i(t)$'s for i = 1, ..., d are i.i.d. standard normal random variables. Further, Λ(t) is an independent Bernoulli random variable, and the initial condition of the last ODE is $\mathbf{y}_0(0) = \mathbf{y}(t)$. Finally, $\mathbf{g}_{d+1} = \mathbf{g}_0$. If closed-form solutions to the ODEs can be found these should be preferred, otherwise one can turn to approximations.

Ninomiya and Victoir's example dealt with the Heston model for $\rho = 0$ and considered the system $\mathbf{Y}(t) = (S(t), V(t))^T$. We consider their scheme for $\mathbf{Y}(t) = (X(t), V(t))^T$, where X(t) is ln S(t), for general values of ρ. The Stratonovich SDE for this system is:

$$\begin{aligned} dX(t) &= (\mu - \tfrac{1}{2}V(t) - \tfrac{1}{4}\omega\rho)\,dt + \sqrt{V(t)} \circ dW_1(t) \\ dV(t) &= \left(-\kappa(V(t)-\theta) - \tfrac{1}{4}\omega^2\right)dt + \omega\sqrt{V(t)} \cdot \left(\rho \circ dW_1(t) + \sqrt{1-\rho^2} \circ dW_2(t)\right) \end{aligned} \tag{6.12}$$

Before stating the NV scheme, we first need to deal with one problematic ODE.

Lemma 6.1:
The solution to the ODE $v'(t) = \alpha\sqrt{v(t)}$, with $v(0) \geq 0$ a known constant, is:

$$v(t) = f(t, \alpha, v(0)) = \max(\tfrac{1}{2}\alpha t + \sqrt{v(0)}, 0)^2 \tag{6.13}$$

if we make the choice that v(t) immediately leaves the origin when $v(0) = 0$ and $\alpha, t \geq 0$.

Proof:
Let us assume that $t \geq 0$ as by symmetry the solution for $t < 0$ is the same as that for v(-t) from the above ODE with $-\alpha$. The general solution is:

$$v(t) = (\tfrac{1}{2}\alpha t + \tfrac{1}{2}C)^2 \tag{6.14}$$

with C an arbitrary constant. In order to satisfy the initial condition, C has to equal $\pm 2\sqrt{v(0)}$. It is clear that v(t) must be monotonically decreasing when $\alpha < 0$, and increasing when $\alpha > 0$. As $v'(0) = \tfrac{1}{2}\alpha C$, C must be positive and thus $C = 2\sqrt{v(0)}$. The solution for $\alpha < 0$ needs to be adapted slightly. The time at which v reaches zero follows as the solution to $v(t^*) = 0$ in (6.14):

$$t^* = -\frac{2\sqrt{v(0)}}{\alpha} \tag{6.15}$$

Hereafter, v(t) must be absorbed in zero, as v(t) must remain nonnegative and its derivative cannot be positive. The only problematic case is when $\alpha > 0$ and $v(0) = 0$. As the square root is not Lipschitz in 0, it follows that the solution to the ODE with $v(0) = 0$ is not guaranteed to be unique. Indeed, both $v(t) = 0$ and $v(t) = \tfrac{1}{4}\alpha^2 t^2$ are valid solutions, and can be combined to create an infinite number of solutions. As the origin is strongly reflecting for the square root process, we choose the latter to remain as close to the SDE as possible. This leads to (6.13). □

We remark that the ODE in lemma 6.1 is incorrectly solved in Ninomiya and Victoir's paper. We expect this to be less important in their example, as ω is there 10%. With the aid of lemma 6.1, the solutions to the ODEs in (6.11) now follow as:

$$\begin{aligned} x_0(t) &= x_0(0) + (\mu - \tfrac{1}{4}\omega\rho)t - \tfrac{1}{2}v_0(0,t) & v_0(t) &= e^{-\kappa t}v_0(0) + (1-e^{-\kappa t})(\theta - \tfrac{\omega^2}{4\kappa}) \\ x_1(t) &= x_1(0) + \frac{v_1(t) - v_1(0)}{\omega\rho} & v_1(t) &= f(t, \omega\rho, v_1(0)) \\ x_2(t) &= x_2(0) & v_2(t) &= f(t, \omega\sqrt{1-\rho^2}, v_2(0)) \end{aligned} \tag{6.16}$$

where f is the solution in (6.13), and:

$$v_0(0,t) = \int_0^t v_0(u)\,du = \tfrac{1}{\kappa}(1-e^{-\kappa t})\left(\tfrac{\omega^2}{4\kappa} - \theta + v_0(0)\right) + (\theta - \tfrac{\omega^2}{4\kappa})t \tag{6.17}$$

We trust the reader can grasp how the scheme works. As in the schemes of Kahl and Jäckel and Alfonsi, the condition $\omega^2 < 4\kappa\theta$ ensures the variance remains positive, as otherwise $v_0(t)$ becomes negative for $t > -\tfrac{1}{\kappa}\ln\frac{4\kappa\theta-\omega^2}{4\kappa\theta-\omega^2-4\kappa v} \equiv t^*(v)$. When $\omega^2 > 4\kappa\theta$ we fix this by using $v_0(\tau)$ instead of $v_0(t)$, and $v_0(0,\tau)$ in $x_0(t)$ instead of $v_0(0,t)$, where $\tau = \min(t^*(v_0(0)), t)$.

As a final remark, it should be clear that not absorbing v in zero is the right choice. If we would absorb, consider the situation where $\omega^2 < 4\kappa\theta$ and $v(0) = 0$. Then $v(t) = 0$, and:

$$\lim_{\Delta t\to 0} S(T) = S(0)\exp\left((\mu - \tfrac{1}{4}\omega\rho)T\right) \tag{6.18}$$

which clearly is undesirable. As we will see the forward asset price is still far from the correct one, even if we impose that v(t) leaves zero immediately. For this reason we omit numerical results for those configurations where $\omega^2 < 4\kappa\theta$ is violated.

6.3. Euler schemes for the CEV-SV model

Given that the exact simulation method of Broadie and Kaya can be rather time-consuming, as well as the fact that no exact scheme is likely to be devised for the non-affine CEV-SV model, a simple Euler discretisation is certainly not without merit. Even if in future a more efficient exact simulation method for the Heston model would be developed, Euler and higher-order discretisations will remain useful for strongly path-dependent options and stochastic volatility extensions of the LIBOR market model, see e.g. Andersen and Andreasen [2002] and Andersen and Brotherton-Ratcliffe [2005], as it is unlikely that the complicated drift terms in such models will allow for exact simulation methods to be devised.

In Section 6.3.1 we firstly unify all presently known Euler discretisations for the CEV-SV model into one framework. Section 6.3.2 compares all schemes and makes a case for a new scheme – the full truncation scheme. In Section 6.3.3 we prove strong convergence of this scheme. Finally, Section 6.3.4 takes a look at the Euler scheme of Andersen and Brotherton-Ratcliffe [2005], which preserves positivity of the variance process in an alternative way.

6.3.1. Euler discretisations - unification

Turning to Euler discretisations, a naïve Euler discretisation for V in (6.1) would read:

$$V(t+\Delta t) = (1-\kappa\Delta t)V(t) + \kappa\theta\Delta t + \omega V(t)^{\alpha}\cdot\Delta W_V(t) \tag{6.19}$$

with $\Delta W_V(t) = W_V(t+\Delta t) - W_V(t)$. When $V(t) > 0$, the probability of $V(t+\Delta t)$ going negative is:

$$\mathbb{P}(V(t+\Delta t) < 0) = N\left(\frac{-(1-\kappa\Delta t)V(t) - \kappa\theta\Delta t}{\omega V(t)^{\alpha}\sqrt{\Delta t}}\right) \tag{6.20}$$

where N is the standard normal cumulative distribution function. Although the probability decays as a function of the time step Δt, it will be strictly positive for any choice hereof. Furthermore,

since ω typically is much higher in a stochastic volatility setting than in an interest rate setting, the problem will be much more pronounced for the Heston model. To prevent V from crossing over to the complex domain we will have to decide what to do in case V turns negative. Practitioners have often opted for a quick "fix" by either setting the process equal to zero whenever it attains a negative value, or by reflecting it in the origin, and continuing from there on. These fixes are often referred to as *absorption* and *reflection* respectively, see e.g. Gatheral [2006]. We note that this terminology is somewhat at odds with the terminology used to classify the boundary behaviour of stochastic processes, see Karlin and Taylor [1981]. In that respect the absorption fix is much more similar to reflection in the origin for a continuous stochastic process, whereas absorption as a boundary classification means that the process stays in the absorbed state for the rest of time. Deelstra and Delbaen [1998] and Higham and Mao [2005] have considered other approaches for fixing the variance when it becomes negative. These are discussed below.

All of these Euler schemes can be unified in a single general framework:

$$\begin{aligned} \widetilde{V}(t+\Delta t) &= f_1\left(\widetilde{V}(t)\right) - \kappa\Delta t \cdot \left(f_2\left(\widetilde{V}(t)\right) - \overline{V}\right) + \omega \cdot f_3\left(\widetilde{V}(t)\right)^{\alpha} \cdot \Delta W_V(t) \\ V(t+\Delta t) &= f_3\left(\widetilde{V}(t+\Delta t)\right) \end{aligned} \tag{6.21}$$

where $\widetilde{V}(0) = V(0)$ and the functions f_i, i = 0 through 3 have to satisfy:

- $f_i(x) = x$ for $x \geq 0$ and i = 1, 2, 3;
- $f_i(x) \geq 0$ for $x \in \mathbb{R}$ and i = 0, 3.

The second condition is a strict requirement for any scheme: we have to fix the volatility term when the variance becomes negative. The first condition seems quite a natural thing to ask from a simulation scheme: if the volatility is not negative, the "fixing" functions f_1 through f_3 should collapse to the identity function in order not to distort the results. In the remainder we use the identity function x, the absolute value function |x| and $x^+ = \max(x,0)$ as fixing functions. Obviously only the last two are suitable choices for f_3. The schemes considered thus far in the literature, as well as our new scheme that is introduced below, are summarised in Table 6.1.

Scheme	**Paper**	**$f_1(x)$**	**$f_2(x)$**	**$f_3(x)$**
Absorption	Unknown	x^+	x^+	x^+
Reflection	Diop [2003], Bossy and Diop [2004], Berkaoui et al. [2008]	|x|	|x|	|x|
Higham and Mao	Higham and Mao [2005]	x	x	|x|
Partial truncation	Deelstra and Delbaen [1998]	x	x	x^+
Full truncation	Lord, Koekkoek and Van Dijk [2008]	x	x^+	x^+

Table 6.1: Overview of Euler schemes known in the literature

While the mentioned papers, apart from Higham and Mao, have dealt with the mean-reverting CEV process in isolation, we also have the asset price S to simulate. For the asset price we switch to logarithms, as in Andersen and Brotherton-Ratcliffe [2005]. This guarantees non-negativity:

$$\ln S(t+\Delta t) = \ln S(t) + \left(\mu - \tfrac{1}{2}\lambda^2 S(t)^{2(\beta-1)} V(t)\right)\Delta t + \lambda S(t)^{\beta-1}\sqrt{V(t)} \cdot \Delta W_S(t) \tag{6.22}$$

and automatically ensures that the first moment of the asset is matched exactly. In an implementation of (6.22) one would use the Cholesky decomposition to arrive at

$\Delta W_S(t) = \rho \Delta W_V(t) + \sqrt{1-\rho^2}\,\Delta Z(t)$, with Z(t) independent of $W_V(t)$. Note that special care has to be taken when S(t) drops to zero, due to property v).

6.3.2. Euler discretisations – a comparison and a new scheme

One thing to keep in mind when fixing negative variances is the behaviour of the true process. At the beginning of this section we mentioned that the origin is strongly reflecting if it is attainable, in the sense that when the variance touches zero, it leaves again immediately. If we think of both the reflection and the absorption fixes in a discretisation context, the absorption fix seems to capture this behaviour as closely as possible. To analyse the behaviour of all fixes, it is worthwhile to consider the case where an Euler discretisation causes the variance to go negative, say $\widetilde{V}(t) = -\delta < 0$, whereas the true process would stay positive and close to zero, $V(t) = \varepsilon \geq 0$. In Table 6.2 we have depicted the new starting point $f_1\left(\widetilde{V}(t)\right)$, the effective variance[27] $f_3\left(\widetilde{V}(t)\right)$ and the drift for all fixes as well for the true process.

Scheme	New starting point	Effective variance	Drift
True process	ε	ε	$\kappa(\theta - \varepsilon)$
Absorption	0	0	$\kappa\theta$
Reflection	δ	δ	$\kappa(\theta - \delta)$
Higham and Mao	$-\delta$	δ	$\kappa(\theta + \delta)$
Partial truncation	$-\delta$	0	$\kappa(\theta + \delta)$
Full truncation	$-\delta$	0	$\kappa\theta$

Table 6.2: Analysis of the dynamics when $V(t) = \varepsilon \geq 0$, but the Euler discretisation equals $-\delta < 0$

A priori we expect that the effect of a misspecified effective variance will be the largest, as this directly affects the stock price on which the options we are pricing depend. From Table 6.2 it seems that reflection has the closest resemblance to the true scheme. However, if $\delta > \varepsilon$, which often is the case, it can be expected that the misspecified variance will cause a larger positive bias than absorption. It is worthwhile to note that in the context of the Heston model it has been numerically demonstrated by Broadie and Kaya [2006] that the absorption fix induces a positive bias in the price of a plain vanilla European call. The Higham and Mao fix tries to lower the bias in the reflection scheme by letting the auxiliary process $\widetilde{V}(t)$ remain negative. This however has an undesirable side-effect when at the same time reflecting the variance in the origin to obtain the effective volatility. If $\widetilde{V}(t)$ drops even further, the effective variance $f_3\left(\widetilde{V}(t)\right)$ will be much too high, in turn causing larger than intended moves in the stock price.

Both the schemes by Deelstra and Delbaen and ourselves can be interpreted as corrections to the absorption scheme. As in the Higham and Mao scheme, both schemes aim to achieve this by allowing the auxiliary process to attain negative values. Contrary to the Higham and Mao scheme, the side-effect of leaving the auxiliary variance negative is not present here, as the effective variance is set equal to zero. We dub the scheme by Deelstra and Delbaen the partial truncation scheme, as only terms involving V in the diffusion of V are truncated at zero. Note that Glasserman [2003, eq. (6.3.66)] also uses this scheme for the CIR process. As will be demonstrated in the numerical results, partial truncation still causes a positive bias. With a view to lowering the bias, we introduce a new Euler scheme, called full truncation, where the drift of V

[27] By effective variance we mean the instantaneous variance of the stock price.

is truncated as well. By doing this the auxiliary process remains negative for longer periods of time, effectively lowering the volatility of the stock, which helps in reducing the bias.

Though this argumentation is heuristic and hard to prove rigorously, the first moment of all "fixed" Euler schemes matches the pattern we described above.

Lemma 6.2:
When $\Delta t < 1/\kappa$ the first moments of $\tilde{V}(t)$ in the various "fixed" Euler schemes in Table 6.1 satisfy the following ordering:

Reflection > Absorption > Higham-Mao = Partial truncation > Full truncation

Proof:
We consider a finite time horizon [0,T], discretised on a uniform grid $t_n = n\Delta t$, n = 1, …, T/Δt. Let us denote all discretisations as:

$$\tilde{v}_{n+1} = f_1(\tilde{v}_n) - \kappa\Delta t\big(f_2(\tilde{v}_n) - \theta\big) + \omega f_3(\tilde{v}_n)^{\alpha}\,\Delta W_{Vn} \tag{6.23}$$

with $\tilde{v}_n$ indicating the value of the discretisation at t_n and $\Delta W_{Vn} = W_V(t_{n+1}) - W_V(t_n)$. Let us define the first moment as $x_n = \mathbb{E}[\tilde{v}_n]$, where the expectation is taken at time 0. The first moment of the Higham-Mao scheme can be shown to satisfy the difference equation $x_{n+1} = (1-\kappa\Delta t)x_n + \kappa\Delta t\theta$, which by noting that $x_0 = v_0$ can be solved as:

$$x_n = (1-\kappa\Delta t)^n (v_0 - \theta) + \theta \tag{6.24}$$

The result holds regardless of the chosen function f_3, and therefore also holds for the partial truncation scheme. This is an accurate approximation of the first moment of the continuous process V(t), as it is a well-known result that $\mathbb{E}[V(t)] = (1-e^{-\kappa t})(V(0)-\theta)+\theta$. Since we initially have $x_0 = v_0$ for all schemes, the remaining results can be found by noting that:

$$(1-\kappa\Delta t)\cdot|\tilde{v}_n| \geq (1-\kappa\Delta t)\cdot\tilde{v}_n^{+} \geq (1-\kappa\Delta t)\cdot\tilde{v}_n \geq v_n - \kappa\Delta t\cdot\tilde{v}_n^{+} \tag{6.25}$$

which are the drift terms of, from left to right, the reflection, absorption, Higham-Mao, partial and full truncation schemes. As x_{n+1} is exactly the expectation of these terms, the statement follows by induction, starting with n = 0. In the second step (n = 1) the inequality already becomes strict, as in each of the schemes v_1 can become negative.

Certainly the first moment is not all that matters, but the above lemma does demonstrate that both the Higham-Mao and truncation fixes adjust respectively the reflection and absorption fixes such that the first moment is lowered. Both the partial truncation and the Higham-Mao scheme already obtain an accurate approximation of the true first moment. By truncating the drift, full truncation pulls the first moment down even further, with a view to adjust any remaining bias of the partial truncation scheme.

6.3.3. Strong convergence of the full truncation scheme

As it is our final goal to price derivatives in the Heston model, we have to be absolutely sure that the sample averages of the realised payoffs converge to the option prices as the time step

used in the discretisation tends to zero. For European options weak convergence is typically enough to prove this result for Euler discretisations, see e.g. Kloeden and Platen [1999], although for more complex path-dependent derivatives strong convergence may be required. As mentioned earlier though, the non-Lipschitzian dynamics of the CEV-SV model preclude us from invoking the usual theorems on weak and strong convergence of Euler discretisations. Focusing on mean-reverting CEV processes, many authors have proven convergence of their particular discretisation. Recently, Diop [2003] and Bossy and Diop [2004] have proven that an Euler discretisation with the reflection fix converges weakly for a variety of mean-reverting CEV processes. For the special case of the mean-reverting square root process, weak convergence of order 1 in the time step is proven, provided that $\omega^2 < \frac{1}{2}\kappa\theta$. This certainly ensures that the origin is not attainable. As the proof may carry over to the general case, we mention that the order of convergence derived is $\min\left(\kappa\theta\omega^{-2}, 1\right)$. Diop proves strong convergence in the L^p ($p \geq 2$) sense of order ½ under a very restrictive condition, which is relaxed somewhat in Berkaoui et al. [2008]. For p = 2 the condition becomes:

$$\kappa\theta \geq \tfrac{1}{2}\omega^2 + \max\left\{\omega\sqrt{14\kappa}, 6\sqrt{2}\omega^2\right\} \tag{6.26}$$

One can easily check that, unfortunately, this condition is hardly ever satisfied for any practical values of the parameters. Both Higham and Mao and Deelstra and Delbaen prove strong convergence of order ½ for their discretisation, without any restrictions on the parameters. As for the absorption scheme, to the best of our knowledge there is no paper dealing with the convergence properties of the absorption fix, although its use in practice is widespread, see e.g. Broadie and Kaya [2004,2006] and Gatheral [2006].

For the mean-reverting CEV process in isolation, following Deelstra and Delbaen and Higham and Mao, we use Yamada's [1978] method to find the order of strong convergence. In the proof we restrict α to lie in [½, 1]. This seems to be the case for most practical applications so that the restriction is not that severe. The big picture of our proof is identical to that of Higham and Mao, but the truncated drift complicates the proofs considerably. The full proof is given in the Appendix, here we merely report the main findings.

First let us introduce some notation. The discretisation has already been introduced in equation (6.23) of lemma 6.2. For the full truncation scheme we have $f_1(x) = x$ and $f_2(x) = f_3(x) = x^+$. To distinguish between the discretisation of the variance and the true process, we will denote the discretisation with lowercase letters and the true process with uppercase letters. Following Higham and Mao [2005] we also require the continuous-time approximation of (6.23):

$$\tilde{v}(t) \equiv \tilde{v}_n - \kappa(t - t_n)(\tilde{v}_n^+ - \theta) + \omega\sqrt{\tilde{v}_n^+}\cdot\left(W_V(t) - W_V(t_n)\right) \tag{6.27}$$

The convergence of the full truncation scheme is proven in the following theorem.

Theorem 6.1 – Strong convergence of v(t) in the L^1 sense
The full truncation scheme converges strongly in the L^1 sense, i.e. for sufficiently small values of the time step Δt we have:

$$\lim_{\Delta t \to 0} \sup_{t\in[0,T]} \mathbb{E}\left[\left|V(t) - v(t)\right|\right] = 0 \tag{6.28}$$

Proof: See the appendix.

Although the above theorem is only proven for the full truncation scheme, it also holds for the partial truncation scheme, albeit with a slightly easier proof. As the proof of strong convergence for the full CEV-SV process and the proof of convergence for plain-vanilla and barrier option prices are quite similar to those provided by Higham and Mao, we omit them here.

6.3.4. Euler schemes with moment matching

Before comparing all schemes to each other, we finally mention a moment-matching Euler scheme suggested by Andersen and Brotherton-Ratcliffe [2005]. In their discretisation, the variance V is locally lognormal, where the parameters are determined such that the first two moments of the discretisation coincide with the theoretical moments:

$$\begin{aligned} V(t+\Delta t) &= \left(e^{-\kappa\Delta t}V(t) + (1-e^{-\kappa\Delta t})\theta\right)\cdot e^{-\frac{1}{2}\Gamma(t)^2\Delta t+\Gamma(t)\cdot\Delta W_V(t)} \\ \Gamma(t)^2 &= \Delta t^{-1}\cdot\ln\left(1+\frac{\frac{1}{2}\omega^2\kappa^{-1}V(t)^{2\alpha}(1-e^{-2\kappa\Delta t})}{\left(e^{-\kappa\Delta t}V(t)+(1-e^{-\kappa\Delta t})\theta\right)^2}\right) \end{aligned} \tag{6.29}$$

The advantage of this scheme is that no "fixes" have to be used to prevent the variance from becoming negative. As mentioned earlier, Andersen [2008] constructs more discretisations for the Heston model along the lines of (6.29), taking the shape of the Heston density function into account. We only compare to (6.29) and show that it is already much more effective than many of the Euler fixes mentioned in Section 6.3.1.

6.4. Numerical results

The previous section established the strong convergence of the full truncation scheme. Though it is certainly useful to theoretically establish the convergence of a scheme, at the end of the day we should be interested in what practitioners really care about: the size of the mispricing given a certain computational budget. It is our goal in this section to compare all mentioned schemes to each other. In our comparisons we take into account both the bias and RMS error, as well as the computation time required. To be clear, if α is the true price of a European call, and $\hat{\alpha}$ is its Monte Carlo estimator, the bias of the estimator equals $\mathbb{E}[\hat{\alpha}]-\alpha$, the variance of the estimator is $\mathrm{Var}(\hat{\alpha})$, and finally the root-mean-squared error (RMS error or RMSE) is defined as $(\text{bias}^2+\text{variance})^{1/2}$. This fills an important gap in the literature as far as the Euler fixes are concerned, as we do not know of a numerical study that compares the various fixes to one another. In the context of the Heston model, Broadie and Kaya only consider the absorption scheme, and estimate its order of weak convergence to be about ½. Alfonsi [2005] compares both reflection and partial truncation to his scheme, but only for the mean-reverting square root process in isolation.

Example	κ	ω	ρ	θ	V(0)	α
SV-I	2	1	-0.3	0.09	0.09	0.5
SV-II	0.5	1	-0.9	0.04	0.04	0.5
SV-III	0.5	1	0	0.04	0.04	0.5
SVJ	3.99	0.27	-0.79	0.014	0.08836	0.5
CEV-SV	1	1.4	0	1	1	0.75

Table 6.3: Parameter configurations of the examples used

The parameter configurations we consider for the variance process are given in Table 6.3. We first focus on the Heston (SV) model, and next consider the Bates (SVJ) model. The latter is an extension of the Heston model to include jumps in the asset price. Clearly all results readily carry over to further extensions of the Heston model, such as the models by Duffie, Pan and Singleton [2000] and Matytsin [1999], both of which add jumps to the stochastic variance process. The final subsection considers a non-Heston CEV-SV model.

6.4.1. Results for the Heston model

In this subsection we investigate the performance of the various simulation schemes for the Heston model. As Heston [1993] solved the characteristic function of the logarithm of the stock price, European plain vanilla options can be valued efficiently using the Fourier inversion approach of Carr and Madan [1999]. For very recent developments with regard to the evaluation of the multi-valued complex logarithm in the Heston model we refer the interested reader to Lord and Kahl [2007a]. Among other things, this paper proves how to keep the characteristic function in both the Heston model and Broadie and Kaya's exact simulation algorithm continuous for all possible inputs. Finally, for a very efficient Fourier inversion technique which works for virtually all strike prices and maturities we point the reader to Lord and Kahl [2007b].

For the Heston model we consider three parameter configurations, which can be found in Table 6.3. In all three examples $\omega^2 >> 2\kappa\theta$, implying that the origin of the mean-reverting square root process is attainable. An example where the origin is not attainable is deferred to Section 6.4.2. For both quasi-second order schemes this means we have to use a fix. We opted for the absorption fix, which is the one Kahl and Jäckel also use in their examples. The probability of a particular discretisation yielding a negative value for V(t) is magnified via the large value of ω, cf. equation (6.20), so that the way in which each discretisation treats the boundary condition will be put to the test. The first example stems from Broadie and Kaya [2006], and is the harder of the two examples they consider. Conveniently, using the example of Broadie and Kaya allows us to compare all biased schemes to their exact scheme. The second example stems from Andersen [2008], where it is used to represent the market for long-dated FX options. The lower level of mean-reversion should make the example more challenging than the first. The third example finally is used to price a double-no-touch option. The correlation of example SV-II is changed to zero here, as this allows us to use reference values from the literature.

As Broadie and Kaya report computation times for both the Euler scheme with absorption and their exact scheme, we scaled our computation times to match their results. Their results were generated on a desktop PC with an AMD Athlon 1.66 GhZ processor, 624 Mb RAM, using Microsoft Visual C++ 6.0 in a Windows XP environment. Relative to the Euler schemes from section 6.3.2, the IJK-IMM scheme, the Andersen and Brotherton-Ratcliffe (ABR) scheme and the Ninomiya and Victoir (NV) scheme take respectively 14%, 16% and 17.5% longer to value a European option. One final word should be mentioned on the implementation of the biased simulation schemes. Clearly, the efficiency of the simulations could be improved greatly by using the conditional Monte Carlo techniques of Willard [1997]. As Broadie and Kaya point out, this only affects the standard error and the computation time, not the size of the bias, which arises mainly due to the integration of the variance process. We therefore chose to keep the implementation as straightforward as possible.

Starting with the first example, Table 6.4 reports the biases of all biased schemes for an at-the-money (ATM) call. To obtain accurate estimates of the bias we used 10 million simulation paths. If a bias is not significantly different from zero at the 95% confidence level, it is marked bold. The first thing one notices is the enormous difference in the magnitude of the bias, demonstrating the need for an appropriate fix. To relate the size of the bias to implied volatilities, one can glance at Figure 6.1. Even with twenty time steps per year the bias of the full truncation scheme is only 7

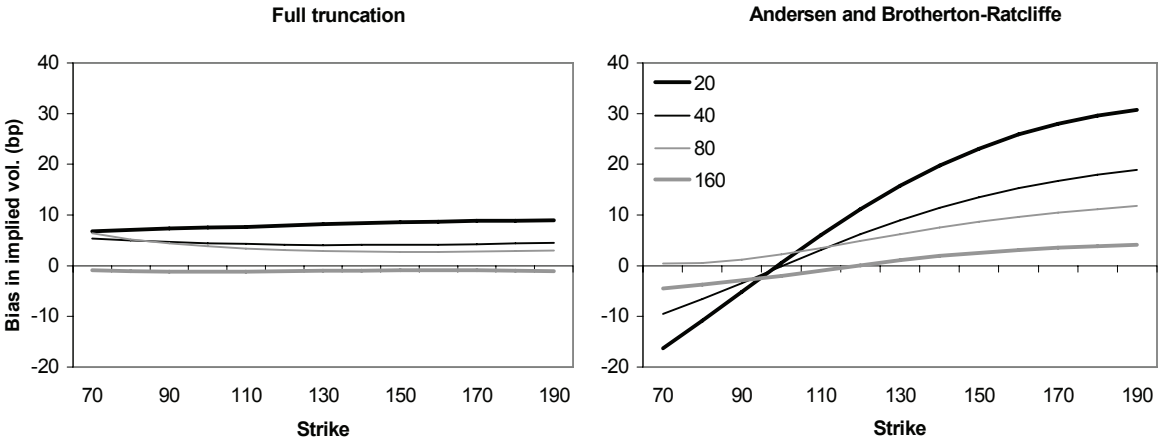


Figure 6.1: Bias as a function of the strike and the time step in example SV-I

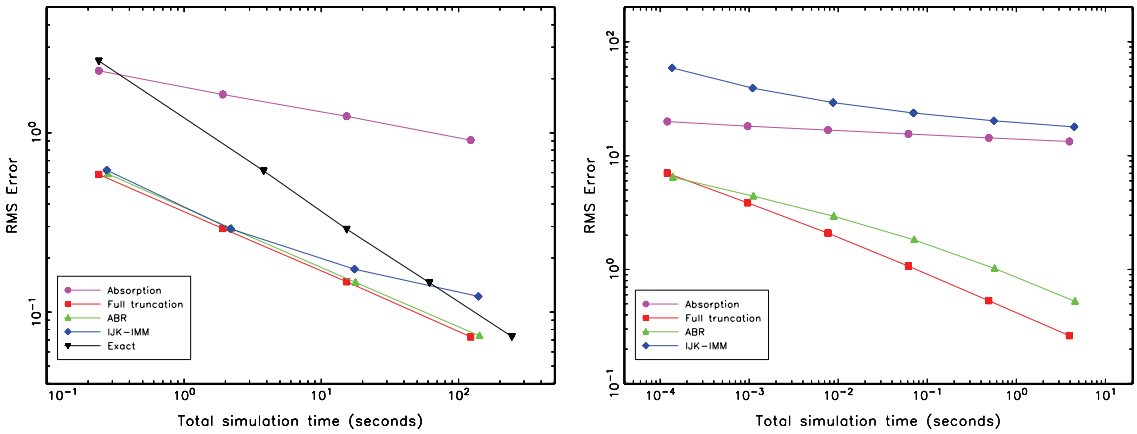


Figure 6.2: Convergence of the RMS error in the Heston model for an ATM call
Left panel: SV-I example, Right panel: SV-II example

Steps/yr.	A	R	HM	PT	FT	ABR	IJK-IMM
20	2.114	4.385	2.732	0.424	0.052	**0.004**	-0.223
40	1.602	3.207	1.680	0.197	**0.031**	**-0.001**	**-0.016**
80	1.225	2.388	1.046	0.096	**0.027**	**0.015**	0.094
160	0.906	1.759	0.615	**0.020**	**-0.008**	**-0.014**	0.098
$O(\Delta t^p)$	0.41	0.44	0.71	1.42	0.82	-0.94	-0.63

Table 6.4: Bias when pricing an ATM call in example SV-I
Asset price process: $S(0) = 100$, $\mu = r = 0.05$, $\lambda = 1$, $\beta = 1$
Deal specification: European call option, Maturity 5 yrs. True option price: 34.9998.

Paths	Steps/yr.	Full truncation			ABR			Exact scheme	
		Bias	RMSE	CPU	Bias	RMSE	CPU	RMSE	CPU
10,000	20	0.052	0.585	0.2	**0.004**	0.590	0.3	0.613	3.8
40,000	40	**0.031**	0.292	1.9	**-0.001**	0.293	2.2	0.290	15.3
160,000	80	**0.027**	0.147	15.4	**0.015**	0.146	17.8	0.146	61.3
640,000	160	**-0.008**	0.073	122.6	**-0.014**	0.074	142.1	0.073	244.5
$O(\Delta t^p)$		0.95	1.02		0.10	0.74		1.02	

Table 6.5: Bias, RMS error and CPU time (in sec.) in the example SV-I for an ATM call

basispoints (bp) for the ATM call, i.e. the option has an implied volatility of 28.69% instead of 28.62%. This is already accurate enough for practical purposes. In contrast, the bias for the absorption scheme is 3.02%, and 6.28% for the reflection scheme. The ABR scheme seems to yield the best results for the ATM case, though Figure 6.1 demonstrates that considered over all strikes the bias of the full truncation scheme is much lower and more stable.

For the order of weak convergence, it is worthwhile to note that under suitable regularity conditions, see e.g. Theorem 14.5.2. of Kloeden and Platen [1999], the Euler scheme converges weakly with order 1 in the time step. Though the SDE for the mean-reverting square root process does not satisfy these conditions, and it is quite hard to properly estimate the weak order[28] of convergence with only 10 million paths, both truncation schemes seem to regain this weak order. In contrast, absorption and reflection have a weak order of convergence slightly under ½.

For the quasi-second-order IJK-IMM scheme we note the convergence is somewhat erratic, similar to the aforementioned findings of Glasserman [2003, pp. 356-358]. The bias seems to increase when increasing the number of time steps per year from 40 to 80. In contrast, the absolute value of the bias decreases uniformly for all Euler schemes, neglecting those cases where the bias is statistically indistinguishable from zero.

Finally, let us examine the RMS error and computation time. These are reported in Table 6.5 for full truncation, ABR and the exact scheme. In the left panel of Figure 6.2 the RMSE is plotted as a function of the time step for all schemes. The choice of the number of paths is an important issue here. Duffie and Glynn [1995] have proven that if the weak order of convergence is p, one should increase the number of paths proportional to $(\Delta t)^{-p}$. When p = 1, this means that if the time step is halved, we should quadruple the number of paths. Obviously, a priori we often do not have an exact value for p, nor do we know the optimal constant of proportionality. We refer the interested reader to the discussion in Broadie and Kaya for the rationale behind the choice of the number of paths in this example. The convergence of the exact scheme is clearly the best. The method produces no bias and hence has $O(N^{-1/2})$ convergence[29], N being the number of paths. For a scheme that converges weakly with order p, Duffie and Glynn have proven that for the optimal allocation the RMSE has $O(N^{-p/(6.2p+1)})$ convergence. Indeed, all biased schemes show a lower rate of convergence than the exact scheme. However, due to the fact that the full truncation scheme already produces virtually no bias with only twenty time steps per year, the RMSEs of both schemes are roughly the same.

For the SV-II example we only report the bias in Table 6.6 as results from the exact scheme are not available to us for this parameter configuration. Again, the truncation schemes outperform the simple Euler schemes by far. Though the ABR scheme initially has a lower bias, it converges considerably slower than the full truncation scheme. Considered over all strikes the full truncation again generates the least bias, making it the clear winner. Interestingly, the IJK-IMM scheme performs much worse than in the SV-I example – the bias is too large for any practical application. As mentioned in Section 3.3 we do not consider the NV scheme for the parameter configurations where $\omega^2 > 2\kappa\theta$, as even the forward is already far from correct. This is particularly evident in this example. If we take e.g. 32 steps per year, the forward price of the asset in the NV scheme equals roughly 179. Considering the fact that the reflection scheme, which at 32 steps per year has the highest bias of the schemes considered, produces a forward price of 101 (the correct answer is 100), it should be clear that the NV scheme is unsuitable when the origin of the square root process is attainable.

So far we have only considered the bias present in European option prices, which reflects the terminal distribution of the underlying asset. As a measure of how well these schemes

[28] The order of weak convergence was estimated here by regressing ln(|bias|) on a constant plus ln(Δt).

[29] The discussion here clearly only holds true when using pseudo random numbers, as we do in this chapter. In a Quasi-Monte Carlo setting the convergence would be $O((\ln N)^2/N)$.

approximate the joint distribution of the asset at various times, we will investigate the bias in double-no-touch prices, which are path-dependent options. A double-no-touch option pays 1 unit

Steps/yr.	A	R	HM	PT	FT	ABR	IJK-IMM
1	18.962	48.472	32.332	12.219	6.371	5.438	57.924
2	17.959	43.321	32.433	8.503	3.710	4.136	38.866
4	16.720	37.842	24.983	5.682	2.041	2.863	29.176
8	15.481	33.161	22.163	3.596	1.055	1.801	23.683
16	14.321	29.200	17.508	2.148	0.525	1.016	20.218
32	13.305	25.987	13.988	1.205	0.259	0.523	17.859
$O(\Delta t^p)$	0.10	0.18	0.25	0.67	0.93	0.68	0.33

Table 6.6: Bias when pricing an ATM call in example SV-II
Asset price process: $S(0) = 100$, $\mu = r = 0$, $\lambda = 1$, $\beta = 1$
Deal specification: European call option, Maturity 10 yrs. True option price: 13.0847.

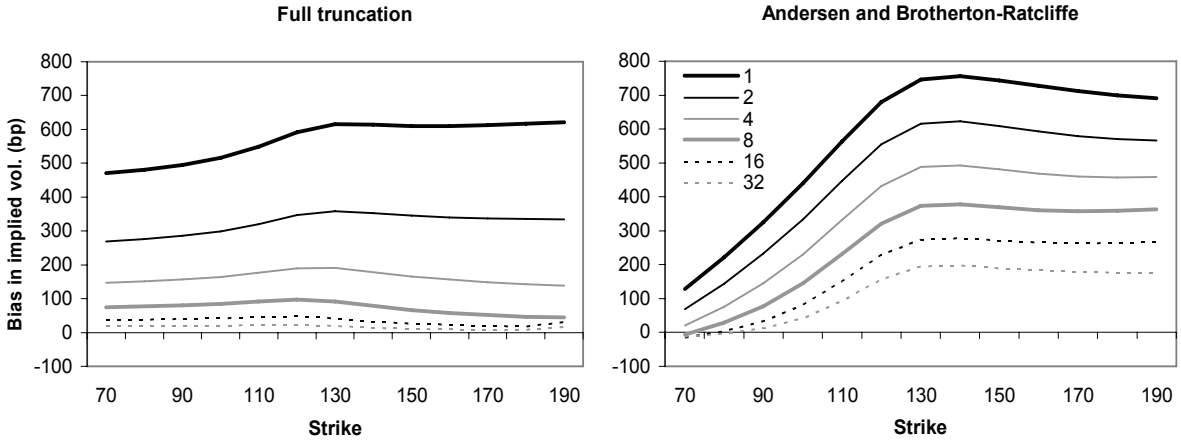


Figure 6.3: Bias as a function of the strike and the time step in example SV-II

of currency if the spot price never hits one of the two barriers. Such options are not uncommon in FX option markets. One reason why we consider them here is that Faulhaber [2002] has shown[30] how to modify Lipton's [2001] eigenfunction expansion approach in order to price double-no-touch options when $\rho = 0$ and the underlying has no drift. This conveniently allows us to generate a reference value with which the simulated values can be compared. Note that both barriers are continuously monitored.

Steps/yr.	A	R	HM	PT	FT	ABR	IJK-IMM
250	-0.190	-0.372	-0.358	0.020	0.022	0.017	-0.235
500	-0.182	-0.346	-0.329	0.016	0.017	0.015	-0.228
1000	-0.174	-0.321	-0.301	0.012	0.013	0.012	-0.218
2000	-0.165	-0.298	-0.275	0.009	0.010	0.009	-0.207

Table 6.7: Bias when pricing a double-no-touch option in example SV-III
Asset price process: $S(0) = 100$, $\mu = r = 0$, $\lambda = 1$, $\beta = 1$
Deal specification: 1 yr. double-no-touch option, barriers at 90 and 110. True price: 0.5011.

[30] The author has provided an implementation at http://www.oliverfaulhaber.de.

In Table 6.7 the bias of the various schemes is reported. The number of time steps per year coincides with the number of monitoring dates used in the simulation. Though both truncation schemes and the ABR scheme do quite a good job, all other schemes produce a completely wrong price, even for an option with a maturity of 1 year. The need for a scheme which correctly treats the boundary behaviour of the variance process is apparent.

6.4.2. Results for the Bates model

In the Bates (SVJ) model [1996], the Heston model is extended with lognormal jumps for the stock price process, where the jumps arrive via a Poisson process:

$$\begin{aligned} dS(t) &= (\mu - \xi\bar{\mu}_J)S(t)dt + \sqrt{\lambda V(t)}\, S(t)dW_S(t) + J_{N(t)}\, S(t)dN(t) \\ dV(t) &= -\kappa\big(V(t)-\theta\big)dt + \omega\sqrt{V(t)}\, dW_V(t) \end{aligned} \tag{6.30}$$

where N is a Poisson process with intensity ξ, independent of the Brownian motions. The random variable J_i denotes the i^{th} relative jump size and is lognormally distributed, $\ln J_i \sim N(\mu_J, \sigma_J^2)$. If the i^{th} jump occurs at time t, the stock price right after the jump equals $S(t+) = (6.1+J_i)\, S(t-)$. To ensure no arbitrage, $\bar{\mu}_J$ in (6.30) has to be the expected relative jump size:

$$1 + \bar{\mu}_J = \mathbb{E}[J_i] = \exp(\mu_J + \tfrac{1}{2}\sigma_J^2) \tag{6.31}$$

The Bates model is often used in an equity or FX context, where the jumps mainly serve to fit the model to the short term skew. Since the jump process is specified independently from the remainder of the model, the same simulation procedure as for the Heston model can be used. If a time step of length T is made till the next relevant date, we draw a random Poisson variable with mean ξT, representing the number of jumps. Subsequently the jump sizes are drawn from the lognormal distribution, and the stock price is adjusted accordingly. In this way the addition of jumps does not add to the discretisation error.

The SVJ example stems from Duffie, Pan and Singleton [2000], where parameters resulted from a calibration to S&P500 index options. Broadie and Kaya [2006] also use this example, which again allows us to compare the various biased simulation schemes to their exact scheme. We note that the example under consideration satisfies $\omega^2 << 2\kappa\theta$, which firstly means that the origin of the square root process is not attainable. Secondly, the low level of ω implies that the probability of any discretisation yielding a negative value for V is significantly smaller than in the Heston example. Hence we may expect that the biases are lower than in the previous example. Thirdly and finally, this combination of parameters is such that the quasi-second order schemes preserve positivity. Contrary to the previous examples this means that both the IJK-IMM scheme and the NV scheme do not require assumptions about the treatment of V at the boundary.

The bias and RMSE of all schemes, now also including the Euler scheme where we transformed coordinates of the variance as in (6.3), are reported in Table 6.8 and Figure 6.4 respectively. The overall picture is the same as before – the full truncation scheme yields the lowest bias, followed by the ABR scheme and the partial truncation scheme. As the level of bias is so low here, given a fixed computational budget the full truncation scheme by far outperforms the exact scheme. Turning to the transformed scheme, we see its bias is huge compared to the other schemes. Its standard deviation is also much larger, due to the fact that the drift in (6.3) blows up when V becomes small. Finally, though the quasi-second order schemes automatically preserve positivity for this parameter configuration, they are outperformed in terms of bias and order of weak convergence by the full truncation scheme.

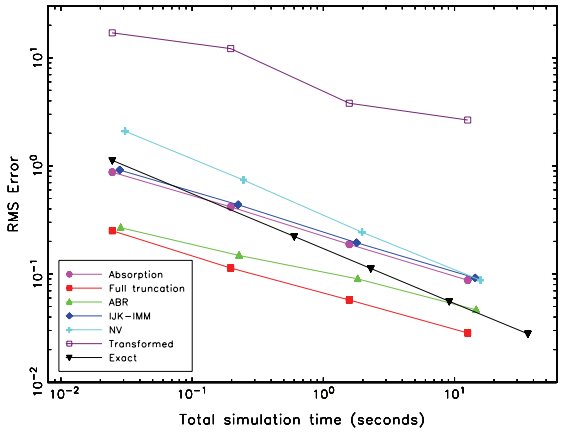


Figure 6.4: Convergence of the RMS error in the SVJ example for an ATM call

Steps/yr.	A	R	HM	PT	FT	ABR	IJK-IMM	NV	Trans
2	0.836	2.489	5.774	2.790	0.106	-0.146	0.887	2.081	9.043
4	0.400	0.900	0.898	0.399	0.016	-0.096	0.423	0.733	6.844
8	0.179	0.396	0.239	0.083	**-0.013**	-0.070	0.186	0.237	3.725
16	0.083	0.175	0.065	0.019	**-0.005**	-0.037	0.088	0.078	2.518
$O(\Delta t^p)$	1.12	1.27	2.13	2.38	1.36	0.64	1.12	1.58	0.64

Table 6.8: Bias when pricing an ATM call in the SVJ example
Asset price process: $S(0) = 100$, $\mu = r = 0.0319$, $\lambda = 1$, $\beta = 1$
Jump process: $\xi = 0.11$, $\overline{\mu}_J = -0.12$, $\sigma_J = 0.15$
Deal specification: European call option, Maturity 5 yrs. True option price: 20.1642.

6.4.3. Results for a non-Heston CEV-SV model

To conclude our extensive numerical analysis, we consider a non-Heston example. The CEV-SV example from Table 6.3 stems from Andersen and Brotherton-Ratcliffe [2005, Appendix A], where their moment-matching Euler scheme is benchmarked to a solution found by solving the corresponding partial differential equation via finite differences. Note that $\alpha = 0.75$, so the origin of the variance process is certainly not attainable.

Steps/yr.	A	R	HM	PT	FT	ABR
1	5.462	13.007	13.007	5.462	1.278	0.460
2	3.097	6.637	4.887	1.821	0.405	0.273
4	1.381	2.824	1.424	0.513	0.092	0.141
8	0.421	0.844	0.249	0.088	**0.012**	0.073
16	0.062	0.132	**0.010**	**-0.002**	**-0.009**	**0.033**
32	**-0.028**	**-0.023**	**-0.033**	**-0.033**	**-0.033**	**-0.011**
$O(\Delta t^p)$	1.62	1.84	2.07	1.94	1.30	1.07

Table 6.9: Bias when pricing an ATM call in the CEV-SV example
Asset price process: $S(0) = 100$, $\mu = 0$, $\lambda = 0.04899$, $\beta = 0.5$, discount factor: 2687.74
Deal specification: European call option, Maturity 10 yrs. True option price: 39.22.

Table 6.9 reports the biases of all Euler schemes. Though the schemes in Kahl and Jäckel [2006] and Ninomiya and Victoir [2004] can be used for the more general CEV-SV process, we chose to focus on the Euler schemes as many of them outperformed the quasi-second order schemes in the previous tests. Once again we conclude that all Euler schemes arrive at the correct answer sooner or later, though the truncation and ABR schemes require much less time steps to do so.

6.5. Conclusions and further research

In this chapter we have considered the simulation of the CEV-SV stochastic volatility model and varieties thereof, focusing largely on the Heston model. In the CEV-SV model, the stochastic variance is modelled as a mean-reverting CEV process. When discretising this process one immediately runs into the problem that although the process itself is guaranteed to be nonnegative, any Euler discretisation has a non-zero probability of becoming negative in the next time step, regardless of the size of the time step. Hence, one has to "fix" these negative variances.

Our contribution is threefold. Firstly, we unify all "fixes" appearing in the literature in a single general framework. Secondly, by analysing the rationale behind the known fixes, we are led up to propose a new scheme, the full truncation scheme, designed specifically to minimise the positive bias one finds when pricing European options using the traditional fixes. Strong convergence is proven for this scheme.

Thirdly and finally, we numerically compare the various Euler schemes to each other, as well as to the quasi-second order schemes by Kahl and Jäckel [2006] and Ninomiya and Victoir [2004], and finally the exact scheme of Broadie and Kaya [2006]. All three of these papers compare their schemes to the Euler scheme with an absorption fix and find their scheme to be superior. Our numerical results demonstrate that using the correct fix at the boundary is extremely important, and significantly impacts the magnitude of the bias. In our examples, we find the full truncation scheme produces the smallest bias, closely followed by the moment-matching Euler scheme of Andersen and Brotherton-Ratcliffe [2005] and the partial truncation scheme. The order of weak convergence of the full truncation scheme appears to be close to 1 in the time step, bringing back the order of weak convergence convergence to the theoretical level for an Euler discretisation of an SDE with Lipschitzian dynamics. The performance of the quasi-second order schemes is found to be somewhat disappointing. In particular, we demonstrated the NV scheme is only suitable for parameter configurations where $\omega^2 < 2\kappa\theta$, often not the case in practice.

When the volatility of volatility is not too high, the full truncation scheme has relatively small levels of bias and is able to generate a smaller RMS error given a certain computational budget than any other biased or exact scheme considered here. This holds true for both European and path-dependent options. Since an initial version of this chapter, Andersen [2008] has specifically designed simulation schemes for the Heston model which mimic its distribution quite closely. These schemes have negligible bias, at the cost of a more complex implementation. On the other hand the full truncation scheme, or indeed that of Andersen and Brotherton-Ratcliffe, is very easy to implement and appears to work fine for a wide variety of processes.

As a final note, we return to the lead mentioned in the introduction, namely that the issues considered here in a continuous time setting can also arise in a discrete time setting. Examples of models where such problems can arise are the model of Heston and Nandi [2000] and the Box-Cox model of Christoffersen and Jacobs [2004]. Let us be more specific and look at the first-order version of the Heston and Nandi model. Here the log-stock price is modelled as:

$$\begin{aligned} \ln S(t) &= \ln S(t-\Delta t) + r + \lambda h(t) + \sqrt{h(t)}z(t) \\ h(t+\Delta t) &= \tilde{\omega} + \beta h(t) + \alpha\left(z(t) - \gamma\sqrt{h(t)}\right)^2 \end{aligned} \tag{6.32}$$

where z(t) is a standard normal random variable and h(t) is the conditional variance of the log-return between t-Δt and t. In this setup h(t) is known at time t-Δt. Note that all the model parameters will depend on the chosen time step Δt. The process remains stationary with finite first two moments if $\beta + \alpha\gamma^2 < 1$. Without further restrictions on the parameters, h(t+Δt) can become negative. In their estimates however ω, β and α are positive and significant at the 95% confidence level, so that there does not seem to be a problem. Turning to their appendix B however, where they prove convergence of (6.32) to the Heston model with $\rho = -1$ as the time step tends to zero, we see that in their proof they choose[31] $\tilde{\omega} = (\kappa\theta - \frac{1}{4}\omega^2)(\Delta t)^2$, $\beta = 0$ and $\alpha = \frac{1}{4}\omega^2(\Delta t)^2$. Positivity of the conditional variance h(t+Δt) can thus only be guaranteed provided that $\kappa\theta \geq \frac{1}{4}\omega^2$. This is the same condition under which the schemes of Alfonsi [2005] and Kahl and Jäckel [2006] preserve positivity, and not surprisingly so as we already remarked the equivalence of these three schemes to first order in Δt in section 6.2.3. Looking in closer detail at their estimation procedure, we see that they only included options with an absolute moneyness less than or equal to ten percent, i.e. at or around at-the-money options. In the Heston model $\kappa\theta$ can certainly be smaller than $\frac{1}{4}\omega^2$ when the skew is quite pronounced. This would not be noticed if only options with strikes at or around the at-the-money level would be included in the calibration procedure. Concluding, it may be necessary to introduce restrictions on the parameters in a discrete time setting in order to ensure that the conditional variance process remains positive.

[31] It seems to us that there are different ways to prove this; the conclusion here will however be the same.

Appendix 6.A – Proof of strong convergence

In this appendix we prove strong convergence of the full truncation scheme applied to the mean-reverting CEV process with ½ ≤ α ≤ 1. We use the same style of proof as Higham and Mao [2005]. As the proof of convergence for the full CEV-SV process follows along the same lines, we only focus on the strong L^1 convergence for the stochastic variance here. Though lemmas 1 and 2 also hold when $0 < \alpha < ½$, the proof used for the main theorem no longer seems applicable. Nevertheless, all practical applications seem to use $\alpha \geq ½$, so that this is no restriction.

For ease of exposure the discretisation over a finite time horizon [0,T] is performed on a uniform grid $t_n = n\Delta t$, n = 1, …, T/Δt. The discretisation of the auxiliary process at t_n is given by:

$$\tilde{v}_{n+1} = \tilde{v}_n - \kappa\Delta t(\tilde{v}_n^+ - \theta) + \omega\tilde{v}_n^{+\alpha}\Delta W_{Vn} \tag{6A.1}$$

where $\Delta W_{Vn} = W_V(t_{n+1}) - W_V(t_n)$. The effective variance is $v_n = \tilde{v}_n^+$. To distinguish between the discretisation of the variance and the true process, we will denote the discretisation with small letters and the true process with capital letters. Following Higham and Mao [2005] we will consider the continuous-time approximation of (6A.1):

$$\tilde{v}(t) \equiv \tilde{v}_n - \kappa(t - t_n)(\tilde{v}_n^+ - \theta) + \omega\tilde{v}_n^{+\alpha} \cdot \left(W_V(t) - W_V(t_n)\right) \tag{6A.2}$$

or, in integral notation:

$$\tilde{v}(t) = \tilde{v}(0) - \kappa\int_0^t (\tilde{v}_\tau(u)^+ - \theta)\,du + \omega\int_0^t \tilde{v}_\tau(u)^{+\alpha}\,dW_V(u) \tag{6A.3}$$

where $\tilde{v}(0) = v_0$, $\tilde{v}_\tau(0) = \tilde{v}(\tau(t))$ and $\tau(t)$ equals t_n if $t_n \leq t \leq t_{n+1}$. Obviously $\tilde{v}_\tau(t)$ coincides with $\tilde{v}(t)$ at the gridpoints of the discretisation.

One of the elements required in proving strong convergence of the full truncation scheme, are bounds on the first and second moments of the effective variance v_n. In the remainder we denote the first and second moments by $x_n \equiv \mathbb{E}[\tilde{v}_n]$ and $y_n \equiv \mathbb{E}[\tilde{v}_n^2]$ respectively. In the main text lemma 6.2 already supplied the following inequality:

$$x_n = \mathbb{E}[\tilde{v}_n] \leq (1 - \kappa\Delta t)^n (v_0 - \theta) + \theta \tag{6A.4}$$

As we do not require sharp bounds, we will use the following corollary which follows directly.

Corollary 6A.1:
For $\Delta t < 2/\kappa$ the first moment of $\tilde{v}_n$ in the full truncation scheme is bounded from above by:

$$x_n \leq |v_0 - \theta| + \theta \tag{6A.5}$$

Proof:
Follows immediately from lemma 6.2.

Secondly, we will find an upper bound on the second moment of $\tilde{v}_n$.

Lemma 6A.1 – Bounding the second moment of the full truncation scheme
For any n = 0, …, N where NΔt = T, and Δt < 2/κ, the second moment of $\widetilde{v}_n$ in the full truncation scheme is bounded by:

$$y_n \le \gamma^N v_0^2 + \frac{\gamma^N - 1}{\gamma - 1} \cdot \left(2\kappa\theta\Delta t U_x + (\kappa\theta\Delta t)^2 + \omega^2\Delta t\right) \equiv U_y(\Delta t) \tag{6A.6}$$

where $\gamma \equiv \max\left\{1, (1-\kappa\Delta t)^2 + 2\alpha\omega^2\Delta t\right\}$.

Proof:
Clearly, $y_0 = v_0^2$ so that the assertion is true for n = 0. Suppose the lemma now holds true for some n. Using (6A.1) we can then write:

$$y_{n+1} = (\kappa\theta\Delta t)^2 + \mathbb{E}\left[(\widetilde{v}_n - \kappa\Delta t\widetilde{v}_n^+)^2\right] + 2\kappa\theta\Delta t \cdot \mathbb{E}\left[(\widetilde{v}_n - \kappa\Delta t\widetilde{v}_n^+)\right] + \omega^2\Delta t\,\mathbb{E}[\widetilde{v}_n^{+2\alpha}] \tag{6A.7}$$

To bound this expression, we note that, apart from the first constant, the right-hand side can be written as the expectation of the following function:

$$f(\widetilde{v}_n) = \begin{cases} \widetilde{v}_n^2 + 2\kappa\theta\Delta t\,\widetilde{v}_n & \widetilde{v}_n \le 0 \\ \widetilde{v}_n^2(1-\kappa\Delta t)^2 + 2\kappa\theta\Delta t(1-\kappa\Delta t)\widetilde{v}_n + \omega^2\Delta t\widetilde{v}_n^{+2\alpha} & \widetilde{v}_n \ge 0 \end{cases} \tag{6A.8}$$

Since $\widetilde{v}_n^{+2\alpha} \le 1 + 2\alpha\widetilde{v}_n^2$ as long as $\alpha \le 1$, (6A.8) can be bounded from above by:

$$f(\widetilde{v}_n) \le \gamma\widetilde{v}_n^2 + 2\kappa\theta\Delta t\widetilde{v}_n + \omega^2\Delta t \tag{6A.9}$$

where γ is as defined above. Returning to (6A.7) we then have:

$$y_{n+1} \le \gamma y_n + 2\kappa\theta\Delta t \cdot x_n + (\kappa\theta\Delta t)^2 + \omega^2\Delta t \tag{6A.10}$$

Repeated use of (6A.10) and our corollary immediately yields (6A.6).

It is important to note that:

$$\lim_{\Delta t \to 0} U_y(\Delta t) = \max\left\{1, e^{2(\alpha\omega^2-\kappa)T}\right\} v_0^2 + \frac{e^{2(\alpha\omega^2-\kappa)T} - 1}{2(\alpha\omega^2 - \kappa)}\left(2\kappa\theta U_x + \omega^2\right) < \infty \tag{6A.11}$$

so that the second moment of the discretisation does not blow up in finite time. Before addressing the strong L^1 error we need a bound on the L^2 difference between the two continuous-time approximations $v_\tau(t)$ and $v(t)$. The proof entirely depends on lemmas 6.1 and 6A.1.

Lemma 6A.2 - The L^2 difference between $v_\tau(t)$ and $v(t)$
For $\Delta t < 2/\kappa$ we have:

$$\sup_{t\in[0,T]} \mathbb{E}\left[\left(v(t) - v_\tau(t)\right)^2\right] \le (\kappa\Delta t)^2 \cdot \left(\theta + U_y(\Delta t)\right) + \omega^2\Delta t \cdot U_y(\Delta t)^\alpha \equiv U_{cont}(\Delta t) \tag{6A.12}$$

Proof:
First of all note that $\mathbb{E}\left[\left(v(t)-v_\tau(t)\right)^2\right] \le \mathbb{E}\left[\left(\tilde{v}(t)-\tilde{v}_\tau(t)\right)^2\right]$. For $t \in [t_n,t_{n+1})$ we have:

$$\mathbb{E}\left[\left(\tilde{v}(t)-\tilde{v}_\tau(t)\right)^2\right] = \kappa^2(t-t_n)^2 \cdot \mathbb{E}[(\tilde{v}_n^+ - \theta)^2] + \omega^2(t-t_n)\cdot \mathbb{E}[\tilde{v}_n^{+2\alpha}] \tag{6A.13}$$

The first term can be bounded from above by:

$$\mathbb{E}[(\tilde{v}_n^+ - \theta)^2] = \theta^2 - 2\theta\mathbb{E}[\tilde{v}_n^+] + \mathbb{E}[\tilde{v}_n^+\tilde{v}_n^+] \le \theta^2 + y_n \tag{6A.14}$$

so that (6A.14) becomes:

$$\begin{aligned}\mathbb{E}\left[\left(\tilde{v}(t)-\tilde{v}_\tau(t)\right)^2\right] &\le \kappa^2(t-t_n)^2\cdot\left(\theta + y_n\right) + \omega^2(t-t_n)\cdot y_n^\alpha \\ &\le \kappa^2(t-t_n)^2\cdot\left(\theta + U_y(\Delta t)\right) + \omega^2(t-t_n)\cdot U_y(\Delta t)^\alpha\end{aligned} \tag{6A.15}$$

The supremum on [0,T] is then bounded from above by (6A.12), which completes the proof.

Clearly $U_{cont}(\Delta t)$ is of $O(\Delta t)$, so that the difference between the discrete-time approximation and its continuous extension vanishes when the time step tends to zero. We are now ready to prove strong convergence in the L^1 sense.

Theorem 6.1 – Strong convergence of v(t) in the L^1 sense
The full truncation scheme converges strongly in the L^1 sense:

$$\lim_{\Delta t\to 0}\sup_{t\in[0,T]}\mathbb{E}\left[\left|V(t)-v(t)\right|\right] = 0 \tag{6A.16}$$

Proof:
First note that $\mathbb{E}\left[\left|V(t)-v(t)\right|\right] \le \mathbb{E}\left[\left|V(t)-\tilde{v}(t)\right|\right]$, so that it is sufficient to show (6A.16) for the latter expression. We will bound it from above in a function of the time step, so that we can prove that this L^1 norm tends to zero as the time step tends to zero. As in Yamada [1978], this is achieved by bounding $\mathbb{E}\left[\phi_k\left(V(t)-\tilde{v}(t)\right)\right]$ for a series of $C^2(\mathbb{R},\mathbb{R})$ functions ϕ_k which tend to the absolute function. Here we use the same notation as in Higham and Mao [2005]. First of all let $a_k = e^{-k(k+1)/2}$ for $k \ge 0$, so that $\int_{a_k}^{a_{k-1}} u^{-1}du = k$. For each integer $k \ge 1$ there exists a continuous function ψ_k with support in (a_{k-1}, a_k) such that $0 \le \psi_k(u) \le 2k^{-1}u^{-1}$ and $\int_{a_{k-1}}^{a_k}\psi_k(u)du = 1$. Defining $\phi_k(x) = \int_0^{|x|}\int_0^y \psi_k(u)\,du\,dy$, it follows that $\phi_k \in C^2(\mathbb{R},\mathbb{R})$, $\phi_k(0) = 0$, and:

$$\begin{aligned}&\left|\phi_k'(x)\right| \le 1 \\ &\left|\phi_k''(x)\right| = 2k^{-1}\,|x|^{-1}\,1_{[a_k<|x|<a_{k-1}]} \\ &|x| - a_{k-1} \le \phi_k(x) \le |x|\end{aligned} \tag{6A.17}$$

Consider $\phi_k\left(V(t)-\tilde{v}(t)\right)$. Using Itō's lemma and taking expectations yields:

$$\mathbb{E}\left[\phi_k\left(V(t)-\tilde{v}(t)\right)\right] = -\kappa M(t) + \tfrac{1}{2}\omega^2 I(t) \tag{6A.18}$$

where we defined:

$$\begin{aligned} M(t) &\equiv \mathbb{E}\left[\int_0^t \phi_k'\left(V(u)-\tilde{v}(u)\right)\cdot\left(V(u)-\tilde{v}_\tau(u)^+\right)du\right] \\ I(t) &\equiv \mathbb{E}\left[\int_0^t \phi_k''\left(V(u)-\tilde{v}(u)\right)\cdot\left(V(u)^\alpha-\tilde{v}_\tau(u)^{+\alpha}\right)^2 du\right] \end{aligned} \tag{6A.19}$$

Note that for ½ ≤ α ≤ 1 we can bound:

$$\left(V(u)^\alpha-\tilde{v}_\tau(u)^{+\alpha}\right)^2 \le \left|V(u)-\tilde{v}_\tau(u)\right|\cdot\left(1+(2\alpha-1)\cdot\left|V(u)-\tilde{v}_\tau(u)\right|\right) \tag{6A.20}$$

and furthermore we have $\left|V(u)-\tilde{v}_\tau(u)\right| \le \left|V(u)-\tilde{v}(u)\right| + \left|\tilde{v}(u)-\tilde{v}_\tau(u)\right|$. Using the property of the second derivative of ϕ in (6A.17) it follows that:

$$I(t) \le \frac{2t}{k}\cdot\left(1+2\tilde{\alpha}\sqrt{U_{cont}(\Delta t)}+\tilde{\alpha}a_{k-1}\right)+\frac{2t}{ka_k}\left(\sqrt{U_{cont}(\Delta t)}+\tilde{\alpha}U_{cont}(\Delta t)\right) \equiv U_I(t,\Delta t) \tag{6A.21}$$

where we used $\mathbb{E}[|X|] \le \sqrt{\mathbb{E}[X^2]}$ for any random variable X and lemma 6A.2. Turning to M(t), we use the property of the first derivative of ϕ from (6A.17) to obtain:

$$\begin{aligned} M(t) &\le \mathbb{E}\left[\int_0^t \left|V(u)-\tilde{v}_\tau(u)^+\right|du\right] \le \mathbb{E}\left[\int_0^t \left|V(u)-\tilde{v}_\tau(u)\right|du\right] \\ &\le \mathbb{E}\left[\int_0^t \left|V(u)-\tilde{v}(u)\right|du\right]+\mathbb{E}\left[\int_0^t \left|\tilde{v}(u)-\tilde{v}_\tau(u)\right|du\right] \\ &\le \mathbb{E}\left[\int_0^t \left|V(u)-\tilde{v}(u)\right|du\right]+t\sqrt{U_{cont}(\Delta t)} \end{aligned} \tag{6A.22}$$

Combining the bounds on I(t) and M(t) in (6A.18) with the third property in (6A.17) yields:

$$\mathbb{E}\left[\phi_k\left(V(t)-\tilde{v}(t)\right)\right] \le \kappa\mathbb{E}\left[\int_0^T \left|V(u)-\tilde{v}(u)\right|du\right]+T\sqrt{U_{cont}(\Delta t)}+\tfrac{1}{2}\omega^2 U_I(T,\Delta t) \tag{6A.23}$$

where we also bounded t from above by T. This gives an upper bound of the same form as in Higham and Mao, and allows us to apply Gronwall's inequality:

$$\sup_{t\in[0,T]}\mathbb{E}\left[\left|V(t)-\tilde{v}(t)\right|\right] \le e^{\kappa T}\left[a_{k-1}+T\sqrt{U_{cont}(\Delta t)}+\tfrac{1}{2}\omega^2 U_I(T,\Delta t)\right] \tag{6A.24}$$

Since (6A.24) holds for any value of k, it is easy to show that $\lim_{\Delta t\to 0}\sup_{t\in[0,T]}\mathbb{E}\left[\left|V(t)-\tilde{v}(t)\right|\right]=0$ as in corollary 3.1 of Higham and Mao. This immediately implies (6A.16). The order of convergence unfortunately does not follow from this proof.

Chapter 7

Partially exact and bounded approximations for arithmetic Asian options[32]

In a departure from the previous chapters, this chapter focuses solely on the Black-Scholes model, and deals with the pricing of arithmetic European Asian options. Though some of the techniques considered here can be extended to the class of exponentially affine models, see Lord [2006], we omit these results for reasons of brevity. Asian options[33], also referred to as average rate or average price options, are financial derivatives depending on the average of a certain underlying asset over a prespecified time interval. Interest rates, exchange rates, bonds, commodities, stocks or indices act as the underlying asset.

Asian options come in numerous flavours. If the strike price depends on a fixed quantity, the option is referred to as a fixed strike Asian option or an average price or rate option. If instead the strike price is proportional to the asset price itself, the contract is called a floating strike Asian option or average strike option. A further distinction can be made on basis of the nature of the average. This can be either arithmetic or geometric, both with possibly varying weights for past observations. The average itself can be discretely sampled, i.e. based on a finite number of past realisations, or continuously sampled. In practice all contracts are based on the discretely sampled arithmetic average, although a vast amount of papers deal with the continuously sampled variety.

There are many reasons for the existence of Asian options. Whereas contracts depending only on one snapshot of an asset price are vulnerable to sudden large shocks or price manipulation, Asian options are much more robust against such phenomena. End-users may prefer Asian options as hedging instruments as they may be exposed to the average performance of the underlying over time. In addition, Asian options are cheaper than their plain vanilla counterparts and are easier to hedge. The latter can easily be seen if we consider the volatility of the average: it will typically be lower than that of the underlying asset. In addition, the closer we are to the maturity date, the smaller the uncertainty in the average will be. This implies a lesser dependence of the option on the spot price than a plain vanilla option with the same maturity.

Asian options can be embedded in more complex financial structures. To counter the price manipulation issue, many exotic options contain so-called "Asian tails". This entails nothing else than that the final payoff is based on the average price of the underlying over a certain interval prior to expiry. Another example of such a structure is unit-linked insurance, as mentioned by e.g. Nielsen and Sandmann [2003] and Schrager and Pelsser [2004]. In its most basic form this can be described as an investment plan with a long maturity where periodic payments are invested in risky investment funds. An Asian option on the average return can be used as a rate of return guarantee. Schrager and Pelsser mention that many insurance companies have supplied these forms of insurance without realising the risk attached to the embedded options. With the fair

[32] This chapter appears as Lord, R. [2006]. "Partially exact and bounded approximations for arithmetic Asian options", *Journal of Computational Finance*, vol. 10, no. 2, pp. 1-52.

[33] The only reason that they are referred to as "Asian" options is that the first known transaction occurred in Tokyo, as is noted in Falloon and Turner [1999].

value calculations of insurance contracts currently at the center of attention, quantifying the risk attached to these embedded options is of the utmost importance.

Within the Black-Scholes framework already no closed-form solutions are available for arithmetic Asian options. In this framework the underlying is assumed to follow a geometric Brownian motion, which amounts to a lognormal distribution for each asset price. Unlike the geometric average, which as a product of lognormal random variables is itself lognormally distributed, the arithmetic average is a sum of correlated lognormal random variables. Unfortunately no closed-form expression is available for the probability law of this sum. The exact same problem occurs when pricing a basket option, whose price depends on the arithmetic average of several assets.

Consequently, the literature has explored a large variety of ways to find a price for the value of an arithmetic Asian option, and research is still ongoing. One can broadly distinguish between methods based on the solution of a partial differential equation (PDE), analytical approximations, lower and upper bounds, tree methods, Monte Carlo methods and transform methods. In the following we by no means aim to give a complete overview. In particular, as this chapter contributes to the first three areas of research, the subsequent sections will deal with these methods in much greater detail than we do here.

For now it suffices to say that PDE methods are most probably the most flexible and efficient way of valuing Asian options, especially if more exotic features are included in the financial structure. Over the past years several authors, starting with Rogers and Shi [1995], have demonstrated that the price of an Asian option can be found by solving a PDE in one space dimension, we mention Andreasen [1998], Hoogland and Neumann [2000a,b] and Večeř [2001]. Though the PDE methods can be extended to the case where the option depends on a basket of underlyings, this is less advisable if the number of factors is large. Our contribution is to relate all mentioned PDE approaches to each other by a change of variable. For Večeř's formulation, which seems to be the most stable in practice, we propose two reductions that increase the numerical stability and reduce the calculation time.

Analytical approximations can be very useful to quickly generate a hopefully accurate estimate of the option value and its sensitivities. In contrast with the PDE methods, once we have an analytical approximation for an Asian option it is usually relatively straightforward to come up with an approximation for a basket option that has the same computational cost. The approximations we will consider in this chaper combine the moment matching approaches, which date back to Levy [1992], and the conditioning approaches of Curran [1994]. We introduce the class of partially exact and bounded (PEB) approximations, which apply conditional moment matching. The class is a logical extension of Curran's approximation. The advantage of conditional moment matching is that, in contrast with the traditional moment matching approaches, the size of the error made lies between a sharp lower and upper bound. We show that the error tends to zero when the strike price tends to zero or to infinity. Though the latter is a very natural criterion, we show that it is not satisfied by Curran's approximation; there the call price diverges as the strike price tends to infinity.

The final line of research that we contribute to is concerned with deriving lower and upper bounds for the value of an Asian option. Bounds can themselves serve as an approximation if they are sufficiently tight, with the advantage that the sign of the error is known a priori. The sharpest lower bound that is currently known is due to Curran [1992, 1994] and Rogers and Shi [1995]. Though a closed-form expression is already available for the lower bound in case of an Asian option, Deelstra, Liinev and Vanmaele [2004] already noted that this expression does not always carry over when there are multiple underlyings. In general one therefore has to resort to a numerical integral over a discontinuous integrand, something which is undesirable. We derive a closed-form expression for the general case, which requires at most three numerical searches. As the previous lower bound was already very sharp, subsequent research efforts have mainly focused on deriving a sharp upper bound. We revisit an upper bound proposed by Thompson

[1999a], and show how his upper bound incorporates many of the upper bounds considered later in the literature. More importantly, we enhance Thompson's upper bound and produce a new bound which is much tighter than all currently reported upper bounds. The new upper bound still performs well when the total volatility is high, i.e. for high volatility environments and/or for long maturities. As the previously mentioned unit-linked insurance typically has long maturities, this is an important finding.

Clearly, the above methods are not all one can use. Tree methods (see Hull and White [1993] and more recently Klassen [2001]) make it possible to value certain path-dependent contracts in a tree. Monte Carlo methods can be very useful, as one can combine control variates (see e.g. Kemna and Vorst [1990], who demonstrate that the geometric average option is highly correlated with the arithmetic average option), importance sampling and stratification (see Glasserman, Heidelberger and Shahabuddin [1990] for an application to Asian options) to significantly reduce the variance. Finally, we should also mention transform methods. Whereas Geman and Yor [1993], using Laplace inversion, derive a closed-form expression for the value of a continuously sampled Asian option, Carverhill and Clewlow [1990] consider the discretely sampled case. Their work, revisited more recently by Benhamou [2002] and Den Iseger and Oldenkamp [2006], shows how to write the arithmetic average as a product of independent random variables, so that its distribution can be found by convolution techniques. Unfortunately these approaches heavily hinge on the assumption that the stock increments are independent, which clearly is only true in the Black-Scholes model when there is only one underlying asset. As such the algorithm cannot easily be adapted to the more general case of a basket option.

The remainder of this chapter is organised as follows. In Section 7.1 we briefly describe the model of the financial market that will be used throughout this document. Section 7.2 shows the link between Rogers and Shi's, Andreasen's and Večeř's PDE, and proposes two reductions for the latter PDE. In Section 7.3 we consider lower bounds, and show how to arrive at a closed-form expression for Curran's and Rogers and Shi's lower bound, which remains valid when multiple underlyings are present. Section 7.4 deals with Thompson's upper bound, and sharpens it considerably. In Section 7.5 the class of PEB approximations are introduced. Section 7.6 concludes with the numerical comparison of the new upper bound and some elements of the class of PEB approximations to state-of-the-art bounds and approximations. Večeř's PDE is used to generate the "exact" prices for the numerical examples. Deltas, gammas and vegas are calculated to demonstrate the accuracy of the PEB approximations.

Finally, it cannot be stressed enough that the approximations and bounds considered here extend to more general situations where the underlying can be expressed as a sum of lognormal random variables. Basket options and the pricing of interest rate swaptions in a Gaussian term structure model are hence two other examples of financial products for which the approximations and bounds will be valid.

7.1. The model

Throughout this document we will constrain ourselves to the Black-Scholes framework. As mentioned earlier, this simple framework already does not yield closed-form expressions for the value of an Asian option. For ease of exposure we will work with a constant and deterministic interest rate, volatility and growth rate of the asset. All results remain valid when the interest rate, volatility and growth rate are deterministic functions of time. The results could even be extended to the case where the term structure of interest rates is Gaussian.

In the Black-Scholes framework the underlying asset and the money market account evolve according to the following stochastic differential equation:

$$\begin{aligned} dS(t) &= \mu S(t)dt + \sigma S(t)dW(t) \\ dB(t) &= rB(t)\,dt \end{aligned} \tag{7.1}$$

where μ is the growth rate, r the interest rate, σ the volatility of the stock, and W(t) is a Brownian motion under the risk-neutral probability measure. For dividend protected assets no arbitrage restrictions enforce the growth rate to be equal to the risk-free rate, whereas when the asset under consideration has a constant dividend yield equal to q, the growth rate μ must be equal to r-q. Throughout the document we will assume, without loss of generality, that the current date is 0. The arithmetic average at the maturity date T will be defined as:

$$A(T) = \int_0^T S(t)\,\rho(t)\,dt \tag{7.2}$$

where ρ is a non-negative weighting function, which integrates to 1 over the interval [0,T]. As an example, the continuously sampled arithmetic average with equal weights is represented by $\rho(t) = \frac{1}{T}$, whereas the discretely sampled arithmetic average with fixing dates $0 < t_1 \leq ... \leq t_N = T$ and equal weights is obtained when $\rho(t) = \frac{1}{N}\sum_{i=1}^{N}\delta(t_i - t)$. Here δ is Dirac's delta function.

In our analysis we will only consider newly issued, non-forward-starting, fixed strike arithmetic Eurasian calls. This is no loss of generality. Put options can be priced via the Asian put-call parity, as we will show. In Hull [2005, pp. 538-540] it is shown how to treat running average Eurasian options as newly issued ones. Similarly, when interest rates are deterministic, forward-starting options can be dealt with easily as well. As for floating strike options, symmetry results between floating and fixed strike Asian options were first shown to exist in Hoogland and Neumann [2000a], and later in Henderson and Wojakowski [2002] and Vanmaele et al. [2006].

For ease of exposure we will mostly deal with forward prices in our analysis. The forward price of the arithmetic Eurasian fixed strike option is equal to its expected value under the risk-neutral probability measure $\mathbb{Q}$, conditional upon all information known at time 0:

$$c_A(T,K) = \mathbb{E}_0^{\mathbb{Q}}[(A(T) - K)^+] \tag{7.3}$$

If no confusion can arise, we will leave out the superscript indicating the measure and the subscript indicating at which time the expectation is evaluated. Current prices can easily be obtained by discounting with the risk-free interest rate. The Asian put-call parity (in terms of forward prices) states that:

$$p_A(T,K) = c_A(T,K) + K - \mathbb{E}[A(T)] \tag{7.4}$$

From (7.4) it is evident that lower and upper bounds for calls also translate to lower and upper bounds for puts. A final quantity of interest is the geometric average. This is defined as:

$$G(T) = \exp\left(\int_0^T \ln S(t)\,\rho(t)\,dt\right) \tag{7.5}$$

An application of the weighted Jensen's inequality shows that $A(T) \geq G(T)$, with equality attained if and only if all components of the average (where $\rho(t) \neq 0$) are equal.

7.2. The partial differential equation approach

In general the price of an Asian option can be found by solving a PDE in two space dimensions, see Wilmott [2006]. In the seminal paper of Rogers and Shi [1995], a variable reduction was used to find a PDE in one space dimension for the value of an Asian claim, for both fixed and floating strikes. A problem associated with this PDE is that for both discretely sampled Asians and floating strike options the Dirac delta function appears as a coefficient in the PDE. Zvan, Forsyth and Vetzal [1997/98] applied techniques from the field of computational fluid dynamics to this PDE to improve the numerical accuracy. Recently, Hoogland and Neumann [2000a,b] and Večeř [2001] arrive at a different one-dimensional PDE for the value of an Asian claim, which does not have the problems associated with the Rogers and Shi PDE. Hoogland and Neumann use the notion of local scale invariance, whereas Večeř considers options on a traded account, and demonstrates that Asian options are a special case hereof.

Basing ourselves on Večeř's derivation of the PDE we will demonstrate, as Hoogland and Neumann [2000a] briefly mention, that it is related to the Rogers and Shi's PDE by a simple change of variable, hereby eliminating the Dirac delta function from the PDE. Similarly it can be shown that Andreasen's [1998] PDE, which was applied only for discretely sampled Asians, is also related to both PDE's. In the second section we propose two reductions for this PDE formulation, which increase the numerical stability and reduce the calculation time required.

7.2.1. The equivalence of various PDE approaches

We will start by introducing some concepts related to Večeř's PDE, as this will be insightful for the reductions to follow in the next section. This PDE is formulated for options on a traded account. A traded account can be viewed as a bank account in which we are allowed to invest in stocks during the life of the option, within certain restrictions. The remaining cash position is invested against a constant interest rate. An option on this so-called traded account is a contract that promises to pay the value of the traded account at maturity if this is positive. If the trading strategy has not been prosperous, and the value of the traded account is negative, the holder receives nothing. Options on a traded account generalise the concept of many options. European, American, passport and vacation options can be shown to be special cases of options on a traded account. For the remainder we will assume that the trading strategy is known a priori and that the interest rate earned on the remaining cash position is zero. This is sufficient for our purposes. If we denote the trading strategy at time t by q(t) and the value of the traded account by X(t), then the terminal value of the traded account equals:

$$X(T) = X(0) + \int_0^T q(t)dS(t) \tag{7.6}$$

The payoff of the option on the traded account is equal to $X(T)^+$ at maturity. Note that q(t) = 1 and initial wealth equal to S(0) – K will yield a final payoff equal to that of a European call option, a result we will use later. The link between Asian options and options on a traded account is found by relating the trading strategy q(t) to the weighting function ρ(t) in (7.2):

$$q(s) = \int_s^T \rho(t)\,dt \qquad 0 \le s \le T \tag{7.7}$$

As ρ is non-negative and integrates to 1, q(0) = 1 and q(T) = 0. Partial integration then yields:

$$\begin{aligned} X(T) &= X(0) + q(T)S(T) - q(0)S(0) + \int_0^T S(t)\,\rho(t)\,dt \\ &= X(0) - S(0) + \int_0^T S(t)\,\rho(t)\,dt \end{aligned} \tag{7.8}$$

Setting the initial value of the traded account to be equal to X(0) = S(0) – K, we see that the payoff of the option on the traded account is equal to that of an arithmetic Eurasian call.

Let us introduce the value of the traded account relative to the stock price, Z(t) = X(t)/S(t). The value of the option at time t is now given by:

$$\begin{aligned} V\big(t,S(t),X(t)\big) &= e^{-r(T-t)}\,\mathbb{E}_t^{\mathbb{Q}}[X(T)^+] \\ &= S(t)\,e^{-(r-\mu)(T-t)}\,\mathbb{E}_t^{\mathbb{S}}\left[\tfrac{X(T)^+}{S(T)}\right] = S(t)\,e^{-(r-\mu)(T-t)}\,\mathbb{E}_t^{\mathbb{S}}[Z(T)^+] \end{aligned} \tag{7.9}$$

where $\mathbb{S}$ is the probability measure associated with taking the stock price (including all accumulated dividends) as the numeraire asset. By deriving the dynamics of Z(t) and applying the Feynman-Kač theorem one finally obtains the following PDE:

$$\frac{\partial u}{\partial t} + \mu\big(q(t) - Z(t)\big)\frac{\partial u}{\partial Z} + \tfrac{1}{2}\sigma^2\big(q(t) - Z(t)\big)^2\,\frac{\partial^2 u}{\partial Z^2} = 0 \tag{7.10}$$

which has to be solved subject to the terminal condition $u\big(T, Z(T)\big) = Z(T)^+$. The option price follows from (7.9) after solving the PDE. The link with Rogers and Shi's PDE will now be obtained by a change of variable. To this end we introduce:

$$Y(t) = q(t) - Z(t) \tag{7.11}$$

The reparametrised option price is denoted as $f\big(t, Y(t)\big) = u\big(t, Z(t)\big)$. The PDE then becomes:

$$\frac{\partial f}{\partial t} - \big(\mu Y(t) + \rho(t)\big)\frac{\partial f}{\partial Y} + \tfrac{1}{2}\sigma^2 Y(t)^2\,\frac{\partial^2 f}{\partial Y^2} = 0 \tag{7.12}$$

with the boundary condition being $f\big(T, Y(T)\big) = Y(T)^-$ for the fixed strike option. Rogers and Shi only considered the case μ = r; in this case the PDE in (7.12) coincides completely with their PDE for the option price divided by the spot price.

Due to the appearance of ρ(t), the Dirac delta function will enter the coefficients of the PDE for discretely sampled and floating strike options. In contrast, these disappear in the formulations of Hoogland and Neumann, and Večeř, simply by integrating over time. This does not leave Rogers and Shi's formulation without any merit. From a computational point of view it is clear that for continuously sampled arithmetic Asians with equal weights the PDE of Rogers and Shi may be preferred, as it has constant coefficients. This leads to a much quicker solution when using finite differences.

Finally, Andreasen's PDE is formulated for the option price divided by the stock price. Realising this, it is easy to see that we can obtain this PDE from Večeř's formulation by simply defining a new variable which is equal to –Z(t). We leave this for the interested reader.

7.2.2. Reducing calculations

Having shown the relation between the various PDE approaches, we will now turn to its numerical implementation. We will again work from Večeř's PDE. Two reductions are proposed which will increase the numerical stability and the calculation time required. The reductions are derived with a finite difference solution method in mind, as demonstrated in Večeř's article. The first reduction is the equivalent of the fact that the value of an Asian with the average taken over only one observation is equal to that of its plain vanilla counterpart. The second reduction, which can be found in a slightly different form for discretely sampled options in Andreasen [1998], is based on the fact that a large number of values are already known beforehand.

Reducing calculations: elimination of final fixing date

It can be shown that in the case of discretely sampled Eurasians, when N fixing dates are present we only have to build a grid for N-1 fixing dates. Assume we have fixing dates t_i, where as before $0 < t_1 \leq ... \leq t_N = T$. Suppose that q_j is the trading strategy[34] on the interval $[t_{j-1}, t_j)$. If we focus on the final interval $[t_{N-1},t_N)$ and use the change of variable $Y(t) = \frac{1}{q_N} Z(t)$, we have to solve the following PDE to calculate the value of the option here:

$$\frac{\partial u}{\partial t} + \mu(1 - Y(t))\frac{\partial u}{\partial Y} + \tfrac{1}{2}\sigma^2 (1 - Y(t))^2 \frac{\partial^2 u}{\partial Y^2} = 0 \tag{7.13}$$

whereas the constraint changes into: $u(T, Y(T)) = q_N \cdot Y(T)^+$. As mentioned earlier, when the trading strategy is equal to 1, we are dealing with a European call. Here we thus have q_N European calls. The risk-free rate and volatility are specified, so all that we have to figure out are the correct spot and strike price. As Z(0) = X(0)/S(0) and X(0) = S(0) – K, it follows that the strike price must equal K = S(0) · (1-Z(0)). Fixing the initial spot price at 1, an initial wealth of Z(t) corresponds to a strike price of 1-Z(t). We end up with:

$$u(t_{N-1}, Z(t_{N-1})) = q_N e^{-r(t_N - t_{N-1})} \mathbb{E}^{\mathbb{Q}}_{t_{N-1}} \left[\left(\frac{S(t_N)}{S(t_{N-1})} - \left(1 - \tfrac{1}{q_N} Z(t_{N-1})\right) \right)^+ \right] \tag{7.14}$$

This formula can be calculated straightforwardly using the Black-Scholes formula. Concluding, it is evident that we only need to build a grid for the first N-1 fixing dates. Formula (7.14) takes the role of the boundary condition. Although this may not speed up calculations, the boundary condition now no longer has any discontinuities, which increases the numerical stability. Unfortunately, using this approach for the remaining fixing dates will not yield a closed-form formula for discretely sampled Eurasians. The result in this section is therefore no different from the fact that an Asian option with one fixing is equal to its plain vanilla counterpart. In the next subsection we show how to reduce the calculations significantly by using known values on the grid.

[34] Note that the trading strategy is always positive for a discrete arithmetic Eurasian fixed strike call.

Reducing calculations: known values on the grid

The reductions in this subsection apply to all types of arithmetic Eurasians. Consider that we have arrived at time s in the grid. Assuming the trading-strategy is non-zero, we propose the following change of variable:

$$Y(t) = \tfrac{1}{q(s)} Z(t) \qquad s \le t \le T \tag{7.15}$$

As in the previous subsection it is easy to see that only two things change:

- The boundary condition becomes $u(T, Y(T)) = q(s) \cdot Y(T)^+$;
- The trading strategy at time t becomes $q(t)/q(s)$.

The boundary condition is no problem, as this only implies we are pricing q(s) options on a traded account. Via the change of variable the weighting function on [s,T] turns into $\rho(t)/q(s)$, which by definition is non-negative and integrates to 1. We can thus consider the value of the option at time s as that of a newly issued option with time to maturity equal to T-s, in conjunction with the adjusted boundary condition and trading strategy.

How is this advantageous? We are going to use the fact that the price of an arithmetic Eurasian call is known analytically for strike prices equal to zero. When the strike price is zero, the contract is simply a forward on the arithmetic average, for which closed-form expressions are readily available. For strike prices smaller than zero, we only need to add a zero-coupon bond with a notional equal to the absolute value of the strike price. In formulas:

$$e^{-r(T-t)} \mathbb{E}_t^Q[(A(T) - K)^+] = e^{-r(T-t)} \cdot \left(\mathbb{E}_t^Q[A(T)] - K\right) \qquad K \le 0 \tag{7.16}$$

As we showed in the previous subsection, an initial wealth of Y(t) corresponds to a strike price of 1-Y(t), if we keep the spot price fixed at 1. This is smaller than or equal to zero when:

$$1 - Y(t) = 1 - \tfrac{1}{q(s)} Z(t) \le 0 \Leftrightarrow Z(t) \ge q(s) \tag{7.17}$$

For the time interval on which the trading strategy is equal to q(s) we only need to build a grid for those points not satisfying (7.17). All grid points that do satisfy this inequality can simply be assigned the known value. The latter only requires the knowledge of the forward price of the arithmetic average.

To quantify the reduction, suppose we have a uniform grid in space and time for the PDE, i.e. $z_i = z_0 + i\,dz$ and $t_j = j\,dt$, for $0 \le i \le M$ and $0 \le j \le N$. The space step and timestep are represented by dz and dt, respectively. For the final timepoint we have $t_N = T$. Note that the choice $z_M = 1$ corresponds to an option with a strike price equal to zero. For the continuously sampled call we have $q(s) = \frac{T-s}{T}$ for $0 \le s \le T$, so that the reduction for the continuously sampled Eurasian call with equal weights is approximately equal to:

$$\tfrac{1}{T} \int_0^T \frac{z_M - q(t)}{z_M (z_M - z_0)} dt = \frac{z_M - \frac{1}{2}}{z_M (z_M - z_0)} \tag{7.18}$$

Assuming that z_M is larger than or equal to 1, and z_0 is smaller than zero, $(z_M - q(t))/z_M$ is approximately the fraction of positive grid points that do not have to be calculated. We divide by $z_M - z_0$ to obtain the fraction of all grid points that are known. Note that e.g. the Crank-Nicolson scheme requires O(M) calculations at each timestep, where M is the amount of space points, so

that this is an appropriate measure of the reduction. Subsequently we average over all timepoints by integrating over time, and divide by the maturity. Večeř chooses[35] $z_0 = -1$ and $z_M = 1$, so that the reduction in this case is approximately 25% and thus very significant.
In addition to speeding up calculations, this reduction also increases the numerical stability. Typically, one would assume that for large values of z, i.e. for high positions of the normalised traded account, the option value will be linear in z. Estimating the point from where onwards this will be approximately valid is already very difficult, as this requires prior knowledge of the function we are trying to solve. Using a known value as the boundary condition bypasses this problem, making the solution more numerically stable.

More on the implementation of Večeř's PDE will follow later in the section on numerical results, where we numerically solve the PDE in order to compare the exact prices of Asian options to upper bounds and approximations. We now continue with lower bounds for the value of an Asian option.

7.3. Lower bounds via conditioning

As mentioned a line of research on Asian options has dealt with deriving lower and upper bounds for the value of these options. In this section we will focus on lower bounds. The first article to our knowledge to derive a lower bound for the value of an Asian option was that of Vorst [1992]. Vorst uses the knowledge that the geometric average is always smaller than or equal to the corresponding average. This is all that is required to see that the value of an arithmetic Eurasian fixed strike call is bounded below by that of its geometric counterpart. Subsequently Curran [1992, 1994] and Rogers and Shi [1995] managed to derive a very tight lower bound by conditioning and applying Jensen's inequality. The resulting lower bound is very tight and in most cases is in fact a better estimate of the option value than a large number of analytical approximations. As the underlying ideas will feature prominently in the remainder of this chapter, we provide its derivation in section 7.3.2. Prior to this we give some preliminary results on conditioning in a Gaussian setting. Finally, section 7.3.3 gives a closed-form expression for the lower bound when applied to an arbitrary sum of lognormal random variables. Though a closed-form expression was already available for the case of arithmetic Asian options, this was not the case for basket options.

7.3.1. Some preliminary results

Let Z be an arbitrary Gaussian conditioning variable. We introduce the following notation for ease of exposure:

$$\mu_Z = \mathbb{E}[Z] \quad \sigma_Z^2 = \mathrm{Var}(Z) \quad \sigma_Z(t) = \mathrm{Cov}\left(\ln S(t), Z\right) \tag{7.19}$$

All expectations are taken under $\mathbb{Q}$, conditional upon all information known at time 0. Using standard results about Gaussian random variables, the distribution of ln S(t) given Z equals:

$$\ln S(t) \mid Z \sim N\left(\ln S(0) + \tfrac{\sigma_Z(t)}{\sigma_Z^2}(Z - \mu_S), \sigma^2 t - \tfrac{\sigma_Z^2(t)}{\sigma_Z^2}\right) \tag{7.20}$$

The conditional expectation of S(t) given Z thus equals:

[35] Note that this choice will not be suitable for an arbitrary choice of parameter values, but is sufficient for the parameters chosen by Večeř.

$$\mathbb{E}[S(t) \mid Z] = S(0)\exp\left(\mu t + \frac{\sigma_Z(t)}{\sigma_Z^2}(Z-\mu_Z) + \frac{1}{2}(\sigma^2 t - \frac{\sigma_Z^2(t)}{\sigma_Z^2})\right) \tag{7.21}$$

With this result in hand we can easily calculate the expectation of A(T) given Z. In the remainder of this section we will require the following expectation:

$$\mathbb{E}[S(t)1_{[Z\geq z]}] = \int_z^\infty \mathbb{E}[S(t)\mid Z=u]\, dF_Z(u) \tag{7.22}$$

where F_Z is the cdf of the Gaussian Z. To find a closed-form expression for this integral we will use the following result:

$$\mathbb{E}[\exp(tZ)1_{[Z\geq z]}] = \exp(\tfrac{1}{2}t^2)\cdot N(t-z) \tag{7.23}$$

where N is the normal cdf. Applying this to (7.22):

$$\begin{aligned}\frac{\mathbb{E}[S(t)1_{[Z\geq z]}]}{\mathbb{E}[S(t)\mid Z=0]} &= \int_{\lambda(K)}^\infty \exp\left(\frac{\sigma_Z(t)}{\sigma_Z^2}u\right) dF_Z(u) \\ &= \int_{\frac{z-\mu_Z}{\sigma_Z}}^\infty \exp\left(\frac{\sigma_Z(t)}{\sigma_Z^2}(\mu_Z+\sigma_Z u)\right)\varphi(u)\, du \\ &= \exp\left(\frac{\sigma_\Lambda(t)}{\sigma_\Lambda^2}\left(\mu_Z + \tfrac{1}{2}\sigma_Z(t)\right)\right)\cdot N\left(\frac{\sigma_Z(t)-(z-\mu_Z)}{\sigma_Z}\right)\end{aligned} \tag{7.24}$$

where φ is the normal pdf. Using this closed-form solution it is trivial to find $\mathbb{E}[A(T)1_{[Z\geq z]}]$. In the following section it will become clear how the formulae derived here can be used to find a lower bound for arithmetic Asian options.

7.3.2. Derivation of the lower bound

Though Curran's and Rogers and Shi's lower bounds are derived in a different manner, they coincide when the same conditioning variable is used. Rogers and Shi's lower bound is more general in that it allows for an arbitrary conditioning variable. Curran's lower bound, as mentioned by Vanmaele et al., only works for Gaussian[36] variables Λ which have the convenient property that there is a threshold value λ(K) for which Λ ≥ λ(K) implies A(T) ≥ K. To distinguish general conditioning variables from such special conditioning variables, we introduce the following convention.

Notation:

Λ will denote a Gaussian random variable for which Λ ≥ λ(K) implies A(T) ≥ K
Z will denote a general Gaussian random variable

Clearly, if we have a random variable that is a lower bound for the arithmetic average, such a random variable Λ can be constructed. Two such examples follow.

[36] There is no loss of generality in this. If Λ is not Gaussian and has F as its cumulative distribution function (cdf), $N^{-1}(F(\Lambda))$ is Gaussian and the corresponding threshold is $N^{-1}(F(\lambda(K)))$. Here N is the normal cdf.

Example 7.1:
As the geometric average G(T) in (7.5) is a lower bound for A(T), its logarithm obviously satisfies the previous criterion. We will denote $\Lambda_{GA} = \ln G(T)$, and the corresponding threshold is $\lambda_{GA}(K) = \ln K$. A second conditioning variable which can be used follows from a first order approximation of A(T), as shown by Vanmaele et al. For $t \geq 0$ the solution to the SDE in (7.1) easily follows as $S(t) = S(0)e^{(\mu-\frac{1}{2}\sigma^2)t+\sigma W(t)}$. Due to the convexity of the exponential, it is clear that the following first order approximation, which we will denote by Λ_{FA}, also acts as a lower bound for the arithmetic average:

$$A(T) \geq \int_0^T S(0)e^{(\mu-\frac{1}{2}\sigma^2)t}\left(1+\sigma W(t)\right)\rho(t)\,dt \equiv \Lambda_{FA} \tag{7.25}$$

The corresponding threshold value for this first order approximation is $\lambda_{FA}(K) = K$.

We start with Rogers and Shi's lower bound, which is based on the following application of Jensen's inequality:

$$\mathbb{E}\left[X^+\right] = \mathbb{E}\left[\mathbb{E}[X^+ \mid Y]\right] \geq \mathbb{E}\left[\mathbb{E}[X \mid Y]^+\right] \tag{7.26}$$

Rogers and Shi's lower bound then follows as:

$$\begin{aligned}\mathbb{E}[(A(T)-K)^+] &\geq \mathbb{E}[\mathbb{E}[(A(T)-K)\mid Z]^+] \\ &= \int_{-\infty}^{\infty}\left(\mathbb{E}[A(T)\mid Z=z]-K\right)^+ dF_Z(z) \equiv LB(Z)\end{aligned} \tag{7.27}$$

where F_Z is the cdf of the Gaussian Z. Curran's lower bound differs in that it conditions on Λ, and splits the expectation in two parts, depending on whether the conditioning variable is above or below its threshold:

$$\begin{aligned}c_A(T,K) &= \mathbb{E}[(A(T)-K)^+ 1_{[\Lambda<\lambda(K)]} + (A(T)-K)^+ 1_{[\Lambda\geq\lambda(K)]}] \\ &= \mathbb{E}[(A(T)-K)^+ 1_{[\Lambda<\lambda(K)]} + (A(T)-K) 1_{[\Lambda\geq\lambda(K)]}] \\ &\equiv c_1(T,K,\Lambda) + c_2(T,K,\Lambda)\end{aligned} \tag{7.28}$$

The special structure of Λ allows us to split the option price into two parts – c_1 and c_2. The second part can be calculated in closed-form by using the results from Section 7.3.1. To this end define:

$$\begin{aligned}c_2(T,K,Z,z) &= \mathbb{E}\left[(A(T)-K)1_{[Z\geq z]}\right] \\ &= \int_0^T \mathbb{E}[S(t)\mid Z=0]\cdot\exp\left(\frac{\sigma_Z(t)\cdot(\mu_Z+\frac{1}{2}\sigma_Z(t))}{\sigma_Z^2}\right)\cdot N\left(\frac{\sigma_Z(t)-(z-\mu_Z)}{\sigma_Z}\right)\rho(t)dt \\ &\quad - KN\left(\frac{\mu_Z-z}{\sigma_Z}\right)\end{aligned} \tag{7.29}$$

It follows that $c_2(T,K,\Lambda) = c_2\left(T,K,\Lambda,\lambda(K)\right)$. A lower bound for c_1 is found once again by applying (7.26):

$$\begin{aligned} c_1(T,K,\Lambda) &= \mathbb{E}[(A(T)-K)^+ 1_{[\Lambda<\lambda(K)]}] = \mathbb{E}[\mathbb{E}[(A(T)-K)^+ 1_{[\Lambda<\lambda(K)]} \mid \Lambda]] \\ &\geq \mathbb{E}[\mathbb{E}[(A(T)-K)1_{[\Lambda<\lambda(K)]} \mid \Lambda]^+] \\ &= \int_{-\infty}^{\lambda(K)} (\mathbb{E}[A(T) \mid \Lambda=\lambda]-K)^+ dF_\Lambda(\lambda) \end{aligned} \tag{7.30}$$

The resulting lower bound will be denoted by LB(Λ). Rogers and Shi's and Curran's lower bound coincide when a conditioning variable Λ is used. Both approaches have in common that eventually we must numerically evaluate an integral with a discontinuous integrand, something which is undesirable. This problem is dealt with in the following section.

The reason why LB(Λ) works so well is that the proposed choices of Λ resemble A(T) closely. The more information the conditioning variable contains about A(T), the lower the contribution of c_1 will be to the option price, and hence the smaller the error will be which we make in the lower bound. An estimate of this error can be made, and leads to Rogers and Shi's upper bound, recently sharpened by Vanmaele et al. [2006].

As a final note, it can be shown that LB(Λ_{GA}) is tighter than Vorst's lower bound. Vorst's lower bound, the value of the geometric average call, can be rewritten as:

$$\mathbb{E}[(G(T)-K)^+] = \mathbb{E}[(G(T)-K)1_{[\Lambda_{GA}\geq\lambda_{GA}(K)]}] \tag{7.31}$$

as $\{\Lambda_{GA} \geq \lambda_{GA}(K)\} = \{G(T) \geq K\}$. Hence, Vorst's lower bound is already smaller than $c_2(T,K)$ and therefore is clearly strictly smaller than LB(Λ_{GA}).

7.3.3. Closed-form expression for the lower bound

It has already been pointed out in the literature that when using Λ_{FA} or Λ_{GA} as conditioning variables the lower bound of Curran and Rogers and Shi can be calculated in closed form, see e.g. Thompson [1999a], Nielsen and Sandmann [2003] and Vanmaele et al. [2006]. The result on which it is based, stated generally in Lemma 7.1, is similar to the result used in Jamshidian's [1989] decomposition used for the pricing of options on a coupon-bearing bond.

Lemma 7.1:
Provided that $\sigma_Z(t)$, the covariance between the underlying asset and the conditioning variable Z, is strictly positive, there is a unique z'(K) such that:

$$\mathbb{E}[A(T) \mid Z = z'(K)] = K \tag{7.32}$$

In case $Z = \Lambda$, we know that $\lambda'(K) \in [-\infty, \lambda(K)]$.

Proof:
Due to the positivity of $\sigma_\Lambda(t)$ the conditional expectation of A(T), built up of terms like (7.21), is strictly increasing and convex in z, and takes values in $[0, \infty)$. This implies there is a z'(K) such that (7.32) holds. If $Z = \Lambda$, the structure of Λ implies that the threshold $\lambda'(K) \in (-\infty, \lambda(K)]$.

Remark 7.1:
For the conditioning variables Λ_{FA} and Λ_{GA} we indeed have that $\sigma_\Lambda(t)$ is strictly positive. This implies that the lower bound can be found by numerically searching for λ'(K) in (7.32), and subsequently evaluating $LB(\Lambda) = c_2(T,K,\Lambda,\lambda'(K))$.

Though the result of Lemma 7.1 is not restricted only to arithmetic Asian options, it is no longer applicable when there are negative correlations between the random variables and the conditioning variable. Therefore there is no closed-form solution for the lower bound when e.g.:

- we are using an arbitrary conditioning variable Z;
- we are considering a basket option with an arbitrary correlation structure between the assets.

In general one must therefore resort to (7.27) or (7.30), a problem Deelstra et al. [2004] also ran into. This involves a discontinuous integrand and as such is undesirable.

Luckily it turns out that even in the more general setup, the conditional expectation can only have a restricted number of forms. Though the derivations here will be based on the arithmetic Asian case, we stress that all results are fully general and apply to the situation where we have multiple underlyings as well. From (7.21) we can write the expectation of A(T), conditional upon a general Gaussian random variable Z, as:

$$f(z) \equiv \mathbb{E}[A(T) \mid Z = z] = \int_0^T \alpha(t) e^{\beta(t) \cdot z} dt \tag{7.33}$$

where $\alpha(t) \geq 0$ and $\beta(t) = \sigma_Z(t) / \sigma_Z^2$. We can write $Z = \int_0^T \eta(t) \ln S(t) \, dt$, where $\eta(t) \neq 0$ only if $\alpha(t) > 0$, and $\mathrm{Var}(Z) > 0$. It makes no sense for Z to have a component independent of all ln S(t) where $\alpha(t) > 0$. Let $\gamma(t) = \alpha(t)\beta(t)$. There are three possible situations:

1. $\gamma(t) \geq 0$: (7.33) will be monotone increasing;
2. $\gamma(t) \leq 0$: (7.33) will be monotone decreasing;
3. γ takes on both strictly positive and negative values.

The monotone increasing situation has been dealt with by Lemma 7.1. The monotone decreasing can be dealt with similarly. The situation in the third case is less clear. To analyse this assume that $\beta(t)$ is non-decreasing (otherwise we can rearrange the time-interval so that this holds true). Suppose s is the largest s for which $\beta(s) \leq 0$, so that for $0 \leq t \leq s$ we have $\beta(t) \leq 0$ (and $\beta(0) < 0$). For $s < t \leq T$ we have $\beta(t) > 0$. We now write:

$$f(z) = \int_0^s \alpha(t) e^{\beta(t) \cdot z} dt + \int_s^T \alpha(t) e^{\beta(t) \cdot z} dt \tag{7.34}$$

Given that γ takes on both strictly positive and negative values, f(z) is the sum of a monotone decreasing and a monotone increasing function. Furthermore, f(z) possesses a special structure for which we can prove that solving f(z) = K yields at most two solutions, making it particularly easy to calculate the lower bound numerically. Before showing this, we require the following lemma.

Lemma 7.2:
Consider f(z) in (7.33). In case there is a $t \in [0,s]$ for which $\gamma(t) < 0$, and there is a $t \in (s,T]$ for which $\gamma(t) > 0$, the first derivative of f(z) has only one zero.

Proof: See the appendix.
In the case the lemma deals with, the conditional expectation f(z) has the property that:

$$\lim_{|z| \to \infty} f(z) = \infty \tag{7.35}$$

Since the derivative has only one zero (say z^*), z^* is the unique minimum of the function, so that f(z) will be decreasing until z^* and increasing afterwards. From a practical point of view it is perfectly valid to ask when this situation will occur. Given the general form of Z, assume there is a t such that $\eta(t)\,\mathrm{Cov}(Z,\ln S(t)) < 0$. Since $\mathrm{Cov}(Z,Z) = \mathrm{Var}(Z) > 0$, there must be an s such that $\eta(s)\,\mathrm{Cov}(Z,\ln S(s)) > 0$. Now, if we have either $\eta(u)$ nonnegative or nonpositive for all u, which is the case for the conditioning variables Λ_{FA} or Λ_{GA}[37], the existence of a negative covariance of Z with some ln S(t) implies the existence of a positive covariance of Z with some ln S(s), where both $\eta(t)$ and $\eta(s)$ are nonzero. We therefore may run into this situation when using Λ_{FA} or Λ_{GA} in a situation with multiple underlyings with some negative correlations between the underlyings. In this situation the conditional expectation will have the shape which is indicated below.

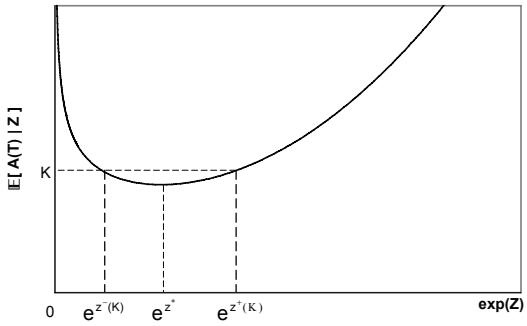


Figure 7.1: Possible shape of conditional expectation when negative correlations are present

Similarly, if η is not everywhere of the same sign, it is also possible to end up with a conditional expectation that has the shape indicated in Figure 7.1. Applying the knowledge gained from Lemmas 1 and 2 yields the following theorem, which provides a closed-form expression for the lower bound in case of an arbitrary conditioning variable Z.

Theorem 7.1:
Let f(z) be the conditional expectation of A(T) given Z, as defined in (7.33). If $\sigma_Z(t)$ is nonnegative or nonpositive for all t, let z'(K) denote the unique solution to f(z) = K. The lower bound can then be found as:

$$\mathrm{LB}(Z) = \begin{cases} c_2\left(T,K,Z,z'(K)\right) & \forall_t\, \sigma_Z(t) \geq 0 \\ \mathbb{E}[A(T)] - K - c_2\left(T,K,Z,z'(K)\right) & \forall_t\, \sigma_Z(t) \leq 0 \end{cases} \tag{7.36}$$

If $\sigma_Z(t)$ is not of one sign, let z^* be the unique minimum of f(z). The equation f(z) = K has at most two solutions. If it has two solutions, we will denote the smallest as $z^-(K)$ and the largest as $z^+(K)$. The lower bound can then be found as:

$$\mathrm{LB}(Z) = \begin{cases} \mathbb{E}[A(T)] - K & f(z^*) > K \\ c_2\left(T,K,Z,z^*\right) & f(z^*) = K \\ \mathbb{E}[A(T)] - K - c_2\left(T,K,Z,z^-(K)\right) + c_2\left(T,K,Z,z^+(K)\right) & f(z^*) < K \end{cases} \tag{7.37}$$

[37] For these conditioning variables we have $\eta(u) = \rho(u)$.

Proof:
Follows immediately from Lemmas 7.1 and 7.2 and the results in Section 7.3.1.

Remark 7.2:
If we condition on a Λ we in addition know that we can restrict our searches for the solutions to the interval $(\infty, \lambda(K)]$, as $f(\lambda(K)) \geq K$ by definition of Λ.

Theorem 7.1 allows us to write the lower bound for any option on a sum of correlated lognormal random variables as a closed-form expression, requiring at most three numerical searches. As the lower bounds in the arithmetic Asian case are already very tight, we do not consider optimising over the conditioning variable, something Deelstra et al. do in the case of multiple underlyings. The closed-form expression derived here would certainly facilitate such an optimisation greatly.

7.4. Thompson's upper bound revisited

As the lower bound of Curran and Rogers and Shi was found to be a very accurate approximation of the arithmetic Asian option, subsequent research efforts have focused on upper bounds. As with the lower bounds, the first article to our knowledge which derives an upper bound for the value of an Asian option is again that of Vorst [1992]. It is found by adding the difference in forward value of the arithmetic average and the geometric average to the price of a geometric call. In Rogers and Shi [1995] an upper bound was derived via estimation of the error made in their lower bound. Nielsen and Sandmann [2003] sharpened this bound considerably when the conditioning variable is the geometric average. The analysis is however valid for any Λ which has the property that $\Lambda \geq \lambda(K)$ implies $A(T) \geq K$, as pointed out in Vanmaele et al. [2006]. We return to this upper bound in the next section.

Most other upper bounds can be reduced to an application of the following idea:

$$\mathbb{E}\left[(A(T)-K)^+\right] = \mathbb{E}\left[\left(\int_0^T (S(t)-Kf(t))\rho(t)\,dt\right)^+\right] \tag{7.38}$$

provided $\int_0^T f(t)\rho(t)\,dt = 1$, where f(t) may be stochastic. The upper bound is then derived as:

$$\begin{aligned}\mathbb{E}\left[\left(\int_0^T (S(t)-Kf(t))\rho(t)\,dt\right)^+\right] &\leq \mathbb{E}\left[\int_0^T \left((S(t)-Kf(t))\rho(t)\right)^+ dt\right] \\ &= \int_0^T \mathbb{E}\left[(S(t)-Kf(t))^+\right]\rho(t)\,dt\end{aligned} \tag{7.39}$$

All upper bounds based on this form assume some function for f(t), be it deterministic or stochastic, and minimise the upper bound with respect to f(t). Note that the optimal[38] choice for f(t) is $1+\frac{1}{K}\left(S(t)-\int_0^T S(u)\rho(u)\,du\right)$, as this yields the same value as the Asian call. This does not help us, since we are searching for an upper bound which is easily calculable.

[38] The author would like to thank Antoon Pelsser for pointing this out. This is similar to importance sampling, where a zero variance estimator can be derived if the value of the integral is known.

Thompson [1999a, 1999b] was the first to give this upper bound for stochastic f(t). In particular, he assumed a Gaussian form for f(t). It is this upper bound which will be the topic of the next few sections. Other upper bounds that appear in the literature are essentially either special cases of (7.38)-(7.39), or combine the above ideas with the conditioning techniques from the previous section. These bounds are documented in the following table.

Upper bound	**Idea**	**Paper**
Comonotonic upper bound (CUB)	Deterministic f(t)	Simon et al. [2000], Nielsen and Sandmann [2003]
Improved CUB (ICUB)	Condition on a Gaussian Z, and then apply the CUB	Dhaene et al. [2002], Vanmaele et al. [2006]
Partially exact CUB (PECUB)	Use Curran's idea and apply the ICUB to the c_1-part	Vanmaele et al. [2006]

Table 7.1: Various related upper bounds in the literature based on (7.38)-(7.39)

Though the results in Simon et al. [2000], Dhaene et al. [2002] and Vanmaele et al. [2006] rely heavily on comonotonicity theory, their upper bounds can be derived alternatively by using the calculus of variations, as Thompson does. Note that if we use the same conditioning variable Λ, we have $\text{PECUB}(\Lambda) \leq \text{ICUB}(\Lambda) \leq \text{CUB}$. As far as the ICUB goes, Vanmaele et al. found it to work best with Z = W(T) as a conditioning variable, at least when applied to an arithmetic Asian option. We should mention that Nielsen and Sandmann [2003] also considered this upper bound when $\Lambda = \Lambda_{GA}$, and refer to it as $C_A^{u,G}$. However they did not succeed in finding an analytical expression for the optimal function f and resorted to numerical optimisation.

The following sections deal with Thompson's upper bound when f(t) has a general Gaussian form. Section 7.4.1 will address the issue of calculating the upper bound, given the parameters that enter f(t). Section 7.4.2 deals with the optimality conditions on f(t). The true optimal choice with respect to the deterministic part of f(t) can in general already not be found analytically, as this requires an analytic expression for the distribution of the sum of a lognormal and a normal random variable. The way Thompson approximates this is described in Section 7.4.3, whereas Section 7.4.4 discusses our proposition: a shifted lognormal approximation. Finally, Section 7.4.5 considers a heuristic to approximate the optimal volatility parameter of f(t).

7.4.1. Calculation of the upper bound

In Thompson [1999a, 1999b] the following form is proposed for the strike-scaling f(t) which enters the upper bound in (7.38)-(7.39):

$$\begin{aligned} f(t) &= \mu(t) - \overline{\sigma} X(t) \\ X(t) &= \xi(t) \int_0^T W(u)\eta(u)\rho(u)\,du - W(t)\eta(t) \end{aligned} \tag{7.40}$$

The functions η, μ and ξ are deterministic, restricted to:

$$\int_0^T \mu(t)\rho(t)\,dt = \int_0^T \xi(t)\rho(t)\,dt = 1 \tag{7.41}$$

so that the restriction on f is satisfied. We will refer to this setup as Thompson's generalised Gaussian upper bound. It is slightly more general than the bound considered for Asian options in Thompson [1999a], although Thompson [1999b] suggests this setup for basket options. In Thompson [1999a] the scaled volatility $\bar{\sigma}$ was chosen equal to σ, and $\eta(t) = \xi(t) = 1$. Note that up to a different mean and scaling, f(t) is then equal to ln S(t) – ln G(T), hereby resembling the optimal choice. These restrictions will be relaxed in the following section, where we discuss Thompson's upper bound more thoroughly.

Given any choice of functions η, μ and ξ and the constant $\bar{\sigma}$, the upper bound can be written as a double integral. To evaluate the integral we must resort to numerical integration. We will now derive the formula for the upper bound. If we condition on W(t), we find that S(t) – Kf(t) is distributed as:

$$S(t) - K\mu(t) + K\bar{\sigma} \cdot \left(\mathbb{E}[X(t) \mid W(t)] + \sqrt{\mathrm{Var}(X(t) \mid W(t))} \cdot Z \right) \tag{7.42}$$

where $Z \sim N(0,1)$. Secondly, we know that the following holds:

$$\mathbb{E}[(a + bZ)^+] = aN\left(\tfrac{a}{b}\right) + b\,\varphi\left(\tfrac{a}{b}\right) \tag{7.43}$$

These are all the results we need to write down the expression for the upper bound. The expectation within the integral in (7.39) can be calculated as:

$$\begin{aligned} \mathbb{E}\left[(S(t) - Kf(t))^+\right] &= \int_{-\infty}^{\infty} \mathbb{E}\left[(S(t) - Kf(t))^+ \mid W(t) = \sqrt{t} \cdot z\right] \varphi(z)\,dz \\ &= \int_{-\infty}^{\infty} \left\{ a(t,z) N\left(\tfrac{a(t,z)}{b(t,z)}\right) + b(t,z)\,\varphi\left(\tfrac{a(t.z)}{b(t,z)}\right) \right\} \varphi(z)\,dz \end{aligned} \tag{7.44}$$

where we defined:

$$\begin{aligned} a(t,z) &= S(0) e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma\sqrt{t} \cdot z} - K\mu(t) + K\bar{\sigma}\,\mathbb{E}\left[X(t) \mid W(t) = \sqrt{t} \cdot z\right] \\ b(t,z) &= K\bar{\sigma}\sqrt{\mathrm{Var}(X(t) \mid W(t))} \end{aligned} \tag{7.45}$$

Simply substituting (7.44) into the expression for the upper bound in (7.39) yields the final formula for the upper bound:

$$\int_0^T \int_{-\infty}^{\infty} \left\{ a(t,z) N\left(\tfrac{a(t.z)}{b(t,z)}\right) + b(t,z)\,\varphi\left(\tfrac{a(t.z)}{b(t,z)}\right) \right\} \varphi(z)\,\rho(t)\,dz\,dt \tag{7.46}$$

One must resort to numerical integration to evaluate this integral. In the next section we consider the optimal choice for μ. From (7.44) it is clear that we can write Thompson's upper bound as a portfolio of delayed payment options, although we must create an artificial asset in order to achieve this. As we do not find a delayed payment structure in terms of options on the actual underlying itself, we do not pursue this further.

7.4.2. Optimality conditions

In the rest of this section we will take the functions η and ξ as given, and consider optimality conditions with respect to μ and the parameter $\overline{\sigma}$. Although it is possible to consider varying η and ξ, their optimality conditions are much more involved. The Lagrangian w.r.t. μ and $\overline{\sigma}$ is:

$$L(\lambda, \overline{\sigma}, \{\mu(t)\}) = \int_0^T \mathbb{E}\left[(S(t) - Kf(t))^+\right] \rho(t)\, dt - \lambda\left(\int_0^T \mu(t)\rho(t)dt - 1\right) \tag{7.47}$$

The first order condition for the optimality of $\{\mu(t)\}$ follows from the calculus of variations, by means of the Euler-Lagrange equation. The condition can be rewritten as:

$$\mathbb{Q}\left(S(t) - Kf(t) \geq 0\right) = -\tfrac{\lambda}{K} \equiv \tilde{\lambda} \tag{7.48}$$

for all t where $\rho(t)$ is unequal to zero. Here $\tilde{\lambda}$ is merely a constant. Defining Y(t) as:

$$Y(t) = S(t) + K\overline{\sigma}X(t) \tag{7.49}$$

we see that the condition in (7.48) can be rewritten as:

$$\mathbb{Q}\left(Y(t) \leq K\mu(t)\right) = 1 - \tilde{\lambda} \tag{7.50}$$

i.e. $K\mu(t)$ is equal to the $1-\tilde{\lambda}$ quantile of the probability distribution of Y(t). Note that the random variable Y(t) is equal to the sum of a lognormal and a normal random variable.

The optimal μ can be determined exactly when $\overline{\sigma}$ equals zero, for this we refer the reader to Thompson [1999a]. As mentioned earlier, this situation coincides with the CUB discussed in the previous section. For $\overline{\sigma}$ unequal to zero, μ is harder to determine analytically. This is due to the fact that there is no closed-form expression for the probability law of Y(t). By approximating the probability law of Y(t) we can obtain approximate solutions for the optimal choice of μ. In the next section we consider the particular approximation that Thompson chose, and improve upon it. The first order condition with respect to $\overline{\sigma}$ will be investigated at a later stage.

7.4.3. A shifted lognormal approximation

To circumvent the difficult probability law in condition (7.50), Thompson used the following first order approximation for S(t):

$$S(t) = S(0)e^{(\mu - \frac{1}{2}\sigma^2)t + \sigma W(t)} \approx S(t)e^{(\mu - \frac{1}{2}\sigma^2)t}\left(1 + \sigma W(t)\right) \tag{7.51}$$

which is valid for small values of $\sigma W(t)$. This approximation will typically only work well when $\sigma W(t)$ has a small variance (i.e. when $\sigma^2 t$ is small) or when the variance of the normally distributed component of Y(t) is much larger than the variance of the lognormally distributed component. The approximation has the desired effect: the calculation of the quantile in (7.50) becomes easy. We refer the interested reader to Thompson [1999a] for the remaining details regarding the calculation of the upper bound.

Clearly, a better approximation for the probability law of Y(t) will yield a tighter upper bound. We propose to approximate Y(t) by a shifted lognormal random variable. A shifted lognormal random variable X can be written as follows:

$$X = \alpha + \exp(\mu + \sigma Z) \tag{7.52}$$

where Z ~ N(0,1). This is a special case of the Johnson-I distribution, see Johnson [1949]. What we propose here is to fit a shifted lognormal random variable to the sum of a normal and a lognormal random variable by matching the first three moments. A reason for this is the following. Firstly, both the normal and the lognormal distributions are special cases of (7.52). When $\alpha = 0$ we have a lognormal distribution. To recognise the normal distribution, we reparametrise:

$$X = \beta - \tfrac{1}{q} + \tfrac{1}{q}\exp\big(q(\mu + \sigma Z)\big) \tag{7.53}$$

where q is restricted to be positive. This is also known as the q-model, considered e.g. by Khong-Huu [1999]. By letting q tend to zero, we find $\lim_{q \downarrow 0} X = \beta + \mu + \sigma Z$ so that X indeed approaches a normal random variable. Intuitively we may thus expect that the behaviour of X will be inbetween these two extremes.

Hill, Hill and Holder [1976] already considered how to fit a shifted lognormal random variable to an arbitrary distribution of which we know the first three moments. The parameters α, μ and σ can be determined analytically given these first three moments. The only condition required for it to be possible that a shifted lognormal random variable as in (7.52) is fit to an arbitrary distribution is that the distribution has a positive skewness or central third moment. This is proven in the following lemma.

Lemma 7.3:
A sum of a normal and a lognormal random variable has a positive third central moment.

Proof: See the appendix.

We conclude that a shifted lognormal variable can be fitted to the sum of a normal and a lognormal random variable. Returning to the problem at hand, we suggest to approximate Y(t) as:

$$Y_{SLN}(t) = \alpha(t) + \exp\big(\nu(t) + \omega(t) \cdot Z\big) \tag{7.54}$$

The shift, mean and volatility functions $\alpha(t)$, $\nu(t)$ and $\omega(t)$ will be determined for each point where $\rho(t)$ is unequal to zero, by equating the first three moments of (7.54) to those of the original random variable Y(t). The same reasoning as before shows that condition (7.50) implies:

$$\frac{\ln\big(K\mu_{SLN}(t) - \alpha(t)\big) - \nu(t)}{\omega(t)} = \gamma_{SLN} \tag{7.55}$$

in turn yielding the following expression for $\mu_{SLN}(t)$:

$$\mu_{SLN}(t) = \tfrac{1}{K}\big(\alpha(t) + \exp\big(\nu(t) + \gamma_{SLN}\,\omega(t)\big)\big) \tag{7.56}$$

Finally, the constant γ_{SLN} is again determined by the condition in (7.41), but must now be determined via a numerical search.

Concluding, we expect that approximating Y(t) by a shifted lognormal random variable will yield better results than Thompson's approximation, at approximately the same computational cost. Thompson's approximation may be preferred when the variance of the normally distributed component is much higher than that of the lognormal component of the sum. This will be the case for high strike values or low volatilities. We now return to the Lagrangian and consider the optimality conditions for the scaled volatility $\bar{\sigma}$.

7.4.4. Optimal value of $\bar{\sigma}$

As stated before, in Thompson [1999a] only the choice of $\bar{\sigma}$ equal to σ and $\eta(x) = \xi(x) = 1$ was considered. This already gave great results for the situations he considered. For long maturities, high volatilities and high strike values, the effects of different choices for $\bar{\sigma}$, η and ξ will be considerable, as we will see.

We saw in the previous sections that an approximately optimal choice for μ, given $\bar{\sigma}$, η and ξ, can be determined with relative ease. The first order conditions for η and ξ are more difficult and would typically create a large optimisation problem. For example, when we have N fixings we will have to determine 2N optimal values, subject to one constraint. Finding the optimal values will take a long time when the amount of fixings is large. One way to get rid of this problem is to parameterise η and ξ as functions of a small set of parameters. For our purposes however it seems that by only varying $\bar{\sigma}$, in conjunction with the shifted lognormal approximation to the optimal function μ, already yields results which outperforms all known upper bounds from the literature.

From the Lagrangian in (7.47) we see that the first order condition with respect to $\bar{\sigma}$ equals:

$$\begin{aligned}\frac{\partial L(\lambda,\bar{\sigma},\{\mu(t)\})}{\partial\bar{\sigma}} &= \int_0^T \mathbb{E}\Big[\Big(-K\tfrac{\partial}{\partial\bar{\sigma}}\mu(t)+KX(t)\Big)\cdot 1_{[S(t)-Kf(t)\geq 0]}\Big]\rho(t)\,dt \\ &= \int_0^T K\,\mathbb{E}\big[X(t)\cdot 1_{[S(t)-Kf(t)\geq 0]}\big]\rho(t)\,dt = 0\end{aligned} \tag{7.57}$$

The Lagrange multiplier does not appear in this equation, since the condition in (7.41) implies:

$$\int_0^T \tfrac{\partial}{\partial\bar{\sigma}}\mu(t)\rho(t)\,dt = 0 \tag{7.58}$$

This property is also used in (7.57), in conjunction with the fact that the probability of each option being in-the-money is equal to $\tilde{\lambda}$. As we use an approximation to derive the optimal function for μ, but subsequently calculate the upper bound exactly, equation (7.58) will not represent the optimality condition for our approximate model.

In practice we find that the upper bound is quadratic in a wide range around the optimal value of the scaled volatility. Outside this range the upper bound increases approximately linearly in the scaled volatility, the further we stray from the optimal value. Using this observation we can find a fast and accurate alternative to numerical optimisation for determining the optimal value of $\bar{\sigma}$. We propose the following algorithm:

1. Calculate the upper bound using μ_{SLN} for three carefully chosen values of $\overline{\sigma}$
2. Fit a quadratic function in $\overline{\sigma}$ to these values
3. Determine the value of $\overline{\sigma}$ in which the upper bound attains its minimum
4. Recalculate the upper bound in the approximately optimal $\overline{\sigma}$

Algorithm 7.1: Calculation of the SLNQuad upper bound

One could of course skip the last step and determine the minimum value of the upper bound by directly substituting the optimal value of $\overline{\sigma}$ in the quadratic function, though this procedure is not guaranteed to yield an upper bound. For this reason we recalculate the upper bound in step 4. In the first step the calculations can be sped up considerably by calculating the upper bound with less accuracy than in the last step. We will refer to the resulting upper bound as SLNQuad. In the final section it will be shown that this approximation gives results which are very close to the optimal value.

7.5. Partially exact and bounded approximations

A much-heard criticism of many analytical approximations is that the size of the error is not known beforehand. The majority of approximations are based on approximating the probability law of the arithmetic average by a distribution that is analytically tractable. The parameters of the approximating distribution are usually determined by matching the first couple of moments to that of the arithmetic average. Via an Edgeworth expansion it can be shown that the error made by approximating the pdf tends to zero when the number of moments that are matched tends to infinity, see Jarrow and Rudd [1982]. When the number of moments that are matched is finite, the error made will have to be determined via a numerical analysis. The approximations of e.g. Turnbull and Wakeman [1991], Levy [1992], Curran[39] [1994], Posner and Milevsky [1998] and Milevsky and Posner [1998] have this shortcoming.

Two approximations for which the size of the error can be estimated are those of Vorst [1992] and more recently Vyncke, Goovaerts and Dhaene [2004]. Vorst's approximation essentially approximates the arithmetic average as a constant plus the geometric average, where the constant is determined such that the first moment coincides with that of the arithmetic average. As such it can be seen to lie between his lower bound and his upper bound.

Vyncke et al. [2004] propose two analytical approximations which by construction lie between a lower and an upper bound. Their idea is to take a convex combination of the LB(Λ) lower bound and the CUB or the ICUB(Z) upper bound. Interestingly, they show that there is a convex combination such that the first two moments of the approximation and the arithmetic average coincide. Therefore, though their approximation is a two-moment matching approximation, its accuracy can be gauged by the difference between the upper bound and the lower bound.

Inspired by these approximations, we will in the next section consider sufficient conditions for an approximation to lie between LB(Λ) and a recent sharpening of Rogers and Shi's upper bound due to Nielsen and Sandmann [2003] and Vanmaele et al. [2006]. The conditions give lead to a class of approximations, which we will name the class of partially exact and bounded (PEB) approximations. The error of these approximations also goes to zero when the strike price tends to zero or to infinity. The proof of this property will lead us to believe that Curran's approximation

[39] Note that this is only true for Curran's so-called "sophisticated" approximation. Curran's "naïve" approximation coincides with the lower bound of Rogers and Shi and as such lies between this lower bound and any upper bound. Throughout we will refer to the "sophisticated" approximation as Curran's approximation.

in fact diverges when the strike price tends to infinity. We show that this is indeed the case, before considering elements of the PEB approximations in the last section.

7.5.1. The class of partially exact and bounded approximations

This section will deal with the construction of an approximation which lies between LB(Λ) and $UB_1(\Lambda)$, a sharpening of Rogers and Shi's upper bound due to Nielsen and Sandmann [2003] and Vanmaele et al. [2006]. The upper bound is based on the following inequality:

$$\begin{aligned} 0 \le \mathbb{E}[X^+] - \mathbb{E}[X]^+ &= \tfrac{1}{2}\left(\mathbb{E}\left[\left|X\right|\right] - \left|\mathbb{E}[X]\right|\right) \\ &\le \tfrac{1}{2}\mathbb{E}\left[\left|X - \mathbb{E}[X]\right|\right] \le \tfrac{1}{2}\sqrt{\mathrm{Var}(X)} \end{aligned} \tag{7.59}$$

The error made in the lower bound LB(Λ) can then be bounded from above by:

$$\begin{aligned} 0 &\le c_A(T,K) - LB(\Lambda) \\ &= \mathbb{E}\left[\mathbb{E}[(A(T)-K)^+ 1_{[\Lambda<\lambda(K)]} \mid \Lambda] - \mathbb{E}[(A(T)-K)1_{[\Lambda<\lambda(K)]} \mid \Lambda]^+\right] \\ &\le \tfrac{1}{2}\mathbb{E}\left[\mathrm{Var}\left(A(T)1_{[\Lambda<\lambda(K)]} \mid \Lambda\right)^{1/2}\right] \equiv \varepsilon_1(\Lambda) \end{aligned} \tag{7.60}$$

leading to the upper bound:

$$UB_1(\Lambda) = LB(\Lambda) + \varepsilon_1(\Lambda) \tag{7.61}$$

A bound that is slightly less tight, though easier to evaluate, can be found by bounding (7.60) further. We do not consider this bound here however. Note that Rogers and Shi's upper bound corresponds to the limit of $UB_1(\Lambda)$ when λ(K) tends to infinity. As a consequence their original upper bound is independent of the strike.

To end up with an approximation that lies between LB(Λ) and $UB_1(\Lambda)$ we could of course take a convex combination of both. A similar approach using the PECUB upper bound is employed in Vanmaele, Deelstra and Liinev [2004] to arrive at a bounded approximation. We will not follow this route, but will employ Curran's idea of decomposing the option price into an exact part and a part that has to be approximated. Recall that the part that has to be approximated is:

$$c_1(T,K,\Lambda) = \mathbb{E}[(A(T)-K)^+ 1_{[\Lambda<\lambda(K)]}] \tag{7.62}$$

The idea is to approximate A(T) by Ã(T), which has an analytically tractable law:

$$\tilde{c}_1(T,K,\Lambda) = \mathbb{E}[(\tilde{A}(T)-K)^+ 1_{[\Lambda<\lambda(K)]}] \tag{7.63}$$

so that the final approximation[40] of the forward option price is:

$$\tilde{c}(T,K,\Lambda) \equiv \tilde{c}_1(T,K,\Lambda) + c_2(T,K,\Lambda) \tag{7.64}$$

[40] Typically the approximation will have to be calculated via numerical integration.

The following theorem supplies a sufficient condition for the resulting forward option price to be bounded below by LB(Λ).

Theorem 7.2:
A sufficient condition that the approximating random variable $\tilde{A}(T)$ must satisfy in order for the resulting approximation in (7.65) to be greater than or equal to LB(Λ) is:

$$\mathbb{E}[\tilde{A}(T) \mid \Lambda = \lambda] \geq \mathbb{E}[A(T) \mid \Lambda = \lambda] \tag{7.65}$$

for $\lambda \in (-\infty, \lambda(K))$, i.e. its conditional mean given Λ must be greater than or equal to that of the arithmetic average.

Proof:
The following set of equations constitutes the proof:

$$\begin{aligned} \tilde{c}_1(t, K, \Lambda) &= \int_{-\infty}^{\lambda(K)} \mathbb{E}[\left(\tilde{A}(T) - K\right)^+ \mid \Lambda = \lambda] dF_\Lambda(\lambda) \\ &\geq \int_{-\infty}^{\lambda(K)} \left(\mathbb{E}[\tilde{A}(T) \mid \Lambda = \lambda] - K\right)^+ dF_\Lambda(\lambda) \\ &\geq \int_{-\infty}^{\lambda(K)} \left(\mathbb{E}[A(T) \mid \Lambda = \lambda] - K\right)^+ dF_\Lambda(\lambda) \end{aligned} \tag{7.66}$$

The first equation is merely another representation of (7.63). The first inequality follows from Jensen's inequality, whereas the second is an application of the condition imposed in (7.65). The last expression is simply the lower bound constructed for c_1 in LB(Λ). This concludes the proof. □

Theorem 7.2 supplies a nice condition to ensure that the resulting approximation in (7.64) is bounded below by a sharp lower bound. Any approximating random variable satisfying (7.65) can be used. Note that Curran's approximation also satisfies this condition. There are however other problems with this approximation, as we will see in the next section. We now supply sufficient conditions for the approximation to lie between LB(Λ) and $UB_1(\Lambda)$.

Theorem 7.3:
If we impose the following two conditions on the approximation random variable $\tilde{A}(T)$:

$$\begin{aligned} &\mathbb{E}[\tilde{A}(T) \mid \Lambda = \lambda] = \mathbb{E}[A(T) \mid \Lambda = \lambda] \\ &\mathrm{Var}[\tilde{A}(T) \mid \Lambda = \lambda] \leq \mathrm{Var}[A(T) \mid \Lambda = \lambda] \end{aligned} \tag{7.67}$$

for $\lambda \in (-\infty, \lambda(K))$, the resulting approximation in (7.64) lies between LB(Λ) and $UB_1(\Lambda)$.

Proof:
The proof follows along the same lines as (7.60), which led to the construction of $UB_1(\Lambda)$:

$$
\begin{aligned}
&0 \le \tilde{c}(T,K,\Lambda) - LB(\Lambda) \\
&= \int_{-\infty}^{\lambda(K)} \left(\mathbb{E}[\left(\tilde{A}(T) - K\right)^+ \mid \Lambda = \lambda] - \left(\mathbb{E}[A(T) \mid \Lambda = \lambda] - K\right)^+ \right) dF_\Lambda(\lambda) \\
&= \int_{-\infty}^{\lambda(K)} \left(\mathbb{E}[\left(\tilde{A}(T) - K\right)^+ \mid \Lambda = \lambda] - \left(\mathbb{E}[\tilde{A}(T) \mid \Lambda = \lambda] - K\right)^+ \right) dF_\Lambda(\lambda) \\
&\le \tfrac{1}{2} \int_{-\infty}^{\lambda(K)} \sqrt{Var\left(\tilde{A}(T) \mid \Lambda = \lambda\right)}\, dF_\Lambda(\lambda) \\
&\le \tfrac{1}{2} \int_{-\infty}^{\lambda(K)} \sqrt{Var\left(A(T) \mid \Lambda = \lambda\right)}\, dF_\Lambda(\lambda) = \varepsilon_1(\Lambda)
\end{aligned}
\tag{7.68}
$$

It is clear that the first inequality holds, as the first condition in (7.67) implies via theorem 7.2 that the resulting approximation is greater than or equal to LB(Λ). The rest of the derivation is similar to the derivation in (7.60). It immediately follows that:

$$LB(\Lambda) \le \tilde{c}(T,K,\Lambda) \le UB_1(\Lambda) \tag{7.69}$$

which concludes the proof of theorem 7.3. □

We call the class of approximations for which condition (7.67) holds, the class of partially exact and bounded (PEB) approximations. Partially exact refers to the fact that they are constructed out of a part that is exact, whereas bounded refers to the fact that the approximation is known to lie between a sharp lower and an upper bound. By construction LB(Λ) is an element of this class of approximations. Other elements of this class will be considered in section 7.5.3. In addition to being bounded, we find that the PEB class of approximations has some desirable properties which are stated in the following theorem.

Theorem 7.4:
The error made in the PEB approximations approaches zero when the strike price K approaches infinity. Furthermore, if K equals zero, the error is also equal to zero.

Proof:
Since the upper bound $UB_1(\Lambda)$ approaches a constant when K tends to infinity, we have to prove the first part in a different manner. Let us extend the approximation $\tilde{A}(T)$ in (7.67) to hold for all $\lambda \in \mathbb{R}$. For the approximating part of the PEB approximation we can then write:

$$
\begin{aligned}
\tilde{c}_1(T,K,\Lambda) &= \mathbb{E}[\left(\tilde{A}(T) - K\right)^+ 1_{[\Lambda < \lambda(K)]}] \\
&\le \mathbb{E}[\left(\tilde{A}(T) - K\right)^+] = \int_K^\infty Q\left(\tilde{A}(T) > x\right) dx \\
&\le \int_K^\infty Q\left(\left| \tilde{A}(T) \right| > x \right) dx = \int_K^\infty Q\left(\tilde{A}(T)^2 > x^2\right) dx \\
&\le \int_K^\infty \tfrac{1}{x^2} \mathbb{E}[\tilde{A}(T)^2] dx = \tfrac{1}{K} \mathbb{E}[\tilde{A}(T)^2]
\end{aligned}
\tag{7.70}
$$

where the Markov inequality is applied at the end. Since $\tilde{A}(T)$ has the same conditional expectation as the arithmetic average, and its conditional variance is smaller or equal to that of the arithmetic average, the same holds for the unconditional expectation and variance. Note that the first two moments of the arithmetic average are bounded; this in turn implies that the second

unconditional moment of the approximating distribution is bounded. Finally, since the approximation $\widetilde{c}_1(T,K,\Lambda) \geq 0$ and $\lim_{K\to\infty} \frac{1}{K}\mathbb{E}[\widetilde{A}(T)^2] = 0$ we can invoke the sandwich theorem:

$$\lim_{K\to\infty} \widetilde{c}_1(T,K,\Lambda) = 0 \tag{7.71}$$

The true value of the call option on the arithmetic average also approaches zero when the strike price approaches infinity. Since $0 \leq c_2(T,K,\Lambda) \leq c_A(T,K)$, we also have that c_2 approaches zero when K approaches infinity. Hence, the error made in the PEB approximation approaches zero when the strike price approaches infinity. We will now demonstrate that the error is zero when the strike price is zero. The Gaussian conditioning variable has the property that $\Lambda \geq \lambda(K)$ implies $A(T) \geq K$. The threshold value $\lambda(K)$ is the smallest value for which this holds. Since the arithmetic average is always larger than or equal to zero, we must have $\lambda(0) = -\infty$. Then:

$$\widetilde{c}(T,0,\Lambda) = c_2(T,0,\Lambda) = \mathbb{E}[A(T)^+ 1_{[\Lambda>-\infty]}] = \mathbb{E}[A(T)] \tag{7.72}$$

which coincides with the true value of the call option when K = 0. □

Although this property seems rather trivial, it is not satisfied by all moment matching approximations. Let us first consider the case where the strike is zero. Since for distributions that can take on negative values $\mathbb{E}[\widetilde{A}(T)^+] \neq \mathbb{E}[\widetilde{A}(T)]$, the error does not equal zero in these cases. An example where this occurs is Posner and Milevsky's [1998] 4-moment matching approximation, which fits a Johnson Type-II distribution to the arithmetic average. The support of this distribution is $\mathbb{R}$, so that negative values can indeed be attained. Similar things can happen with their 3-moment matching approximation, which fits a shifted lognormal random variable to the arithmetic average.

7.5.2. Curran's approximation

Moment matching approximations that fit a distribution with a finite expectation and variance to the arithmetic average, satisfy the property that the error tends to zero when the strike price tends to infinity. Curran's approximation does not satisfy these properties. As Curran's approximation yields very accurate results for moderate strike prices, it is worthwhile to investigate why the approximation diverges for large strike values. In section 7.5.3 we cure this undesirable property and obtain two very accurate PEB approximations.

From Curran [1994] it is clear that like our PEB approximations his approximation consists of an exact part (namely c_2 with the geometric average as the conditioning variable) and an approximating part, which satisfies:

Curran2M approximation:

$$\begin{aligned}
&\widetilde{A}(T) \mid G(T) = G(T) + \exp\left(\mu_{A|G}(K) + \sigma_{A|G}(K)\cdot Z\right) \\
&\mathbb{E}\left[\widetilde{A}(T) \mid G(T) = K\right] = \mathbb{E}\left[A(T) \mid G(T) = K\right] \\
&\mathrm{Var}\left(\widetilde{A}(T) \mid G(T) = K\right) = \mathrm{Var}\left(A(T) \mid G(T) = K\right)
\end{aligned} \tag{7.73}$$

for $G(T) \leq K$ and $Z \sim N(0,1)$. In words, the arithmetic average is approximated by a shifted lognormal random variable where the shift is equal to the geometric average. This ensures that the approximating quantity has the same natural restriction as the arithmetic average, namely that it is always larger than or equal to G(T). The parameters of the lognormal component, $\mu_{A|G}(K)$ and $\sigma_{A|G}(K)$, are constant and chosen such that the two moments of the approximating distribution coincide with that of the arithmetic average, conditional upon G(T) being equal to K. The rationale behind this assumption is that the largest contribution of c_1 to the option value will come from paths where the geometric mean is close to the strike price. Therefore, it is particularly important to fit the mean and variance of A(T) given G(T) close to the strike price. As such the approximation is a local two-moment fit at the strike price. To emphasise this fact we will refer to the approximation as the Curran2M approximation.

The Curran2M approximation is not a member of the PEB class, due to the fact that the conditional mean is not fit exactly in all points below the strike price. The fact that the approximation is strike-price dependent is what causes the approximation to diverge for large values of the strike price, which is proven in the next theorem.

Theorem 7.5:
The Curran2M approximation diverges when the strike tends to infinity.

Proof:
Consider the approximating part of the Curran2M approximation:

$$\begin{aligned} &\mathbb{E}[\left(\widetilde{A}(T)-K\right)^{+}1_{[G(T)<K]}] \geq \mathbb{E}[\left(\widetilde{A}(T)-K\right)1_{[G(T)<K]}]^{+} \\ &= \left(\mathbb{E}[G(T)1_{[G(T)<K]}]+\mathbb{Q}\left(G(T)<K\right)\cdot\left(\mathbb{E}[A(T)|\,G(T)=K]-2K\right)\right)^{+} \end{aligned} \tag{7.74}$$

where we used Jensen's inequality and property (7.73) to derive the last expression. For large values of K this lower bound will tend towards:

$$\left(\mathbb{E}[G(T)]+\mathbb{E}[A(T)|\,G(T)=K]-2K\right)^{+} \tag{7.75}$$

All that remains to be proven is that the conditional expectation of the arithmetic average at the strike price is superlinear in K, since then (7.75) clearly diverges for large K. To show that this indeed is the case, consider the covariance of ln S(t) with the geometric average G(T):

$$\sigma_{\Lambda_{GA}}(t) = \mathrm{Cov}\left(\ln S(t), \Lambda_{GA}\right) = \int_0^T \sigma^2 \min(t,u)\,\rho(u)\,du \tag{7.76}$$

This is non-decreasing in t. Suppose that s is the largest value in [0,T] for which $\rho(s) > 0$. The maximum covariance is then attained when t = s. In particular, using (7.21) we find:

$$\lim_{K\to\infty} \frac{\mathbb{E}\left[A(T)\,|\,G(T)=K\right]}{\mathbb{E}\left[S(s)\,|\,G(s)=K\right]\,\rho(s)} = 1 \tag{7.77}$$

The conditional expectation in the denominator is equal to:

$$\mathbb{E}\left[S(s)\,|\,G(s)=K\right] = \mathbb{E}\left[S(s)\,|\,G(s)=1\right]\cdot K^{\sigma_{\Lambda_{GA}}(s)/\sigma^2_{\Lambda_{GA}}} \tag{7.78}$$

It is clear that the variance of Λ_{GA}, an average of the covariances $\sigma_{\Lambda_{GA}}(t)$, will be smaller than the largest covariance $\sigma_{\Lambda_{GA}}(s)$. Therefore (7.78), and also the conditional expectation in (7.75) is superlinear in K. This in turn causes the lower bound, and therefore also the Curran2M approximation itself, to diverge when the strike price tends to infinity.

In practice we find that this is more noticeable for high volatility/long maturity situations. Though these situations may not be so important for options that just have an "Asian tail", it certainly is important when pricing rate-of-return guarantees which are embedded in many unit-linked insurance policies. These tend to have long maturities, and are in a Black-Scholes world equivalent to arithmetic Asian options, as has been shown by Schrager and Pelsser [2004]. Numerical examples of our finding will be shown in the last section. We finally note that the above analysis also holds for the approximations of Deelstra et al. [2004] that use constant coefficients.

7.5.3. Suggestions for partially exact and bounded approximations

From the manner in which the Curran2M approximation is constructed it should immediately be clear that we can cure the divergence issue by matching all two conditional moments exactly for those cases when $\Lambda < \lambda(K)$. Hence it seems wise to consider the following approximations:

Conditional two-moment matching:

$$\widetilde{A}(T) \mid \Lambda = \Psi(\Lambda) \quad \mathbb{E}[\widetilde{A}(T) \mid \Lambda] = \mathbb{E}[A(T) \mid \Lambda] \quad \mathrm{Var}\left(\widetilde{A}(T) \mid \Lambda\right) = \mathrm{Var}\left(A(T) \mid \Lambda\right) \quad (7.79)$$

for $\Lambda < \lambda(K)$. Here $\Psi(\Lambda)$ is a non-negative random variable belonging to a distribution with at least two parameters, so that the first two moments can be fitted according to (7.79). Matching all first two conditional moments as in (7.79) implies that the first two unconditional moments are also matched. By doing this we may intuitively expect the resulting error to be smaller than when we only match the first two unconditional moments. Theorem 7.3 shows that a major benefit of doing this lies in the fact that we obtain a sharp analytical bound on the size of the error. Also, by virtue of theorem 7.4, we know that the error goes to zero when the strike price tends to zero or infinity. Regardless of the choice of Ψ, calculating $\widetilde{c}_1$ is straightforward:

$$\begin{aligned} \widetilde{c}_1(T,K,\Lambda) &= \mathbb{E}[\left(\widetilde{A}(T) - K\right)^+ 1_{[\Lambda<\lambda(K)]}] \\ &= \int_{-\infty}^{\lambda(K)} \mathbb{E}\left[\left(\Psi(\lambda) - K\right)^+\right] dF_\Lambda(\lambda) \end{aligned} \quad (7.80)$$

In general we will have to resort to numerical integration to evaluate (7.80). We therefore need to calculate the conditional mean and variance of the arithmetic average in all points in which we evaluate the integrand. As Turnbull and Wakeman [1991] pointed out, the moments of the discretely sampled arithmetic average can be calculated very efficiently by using the property that the returns of the underlying are independent. Calculating the first m moments becomes a process of order O(mN), where N is the number of fixings. When we condition on Λ, this property is unfortunately forsaken; calculating the first m moments is now a process of order $O(N^m)$. Calculating the variance in each point of the algorithm may become too costly when N is very large. Theorem 7.3 provides a nice alternative: we can choose just to evaluate the variance in a number of points. As long as we ensure that the approximating conditional variance is lower than

or equal to the true conditional variance, (7.68) is satisfied and we still have a PEB approximation. This also shows that if we can derive an analytical lower bound[41] on the variance of $A(T) \mid \Lambda$, we can use this expression as the fitted variance. In practice it may however be just as efficient to approximate the numerical integral in (7.80) with a smaller accuracy, requiring less evaluations of the conditional moments.

It is not immediately clear how to choose Ψ, as the arithmetic average, conditional upon Λ, remains to be a sum of correlated lognormal random variables. For low to moderate strike values, the conditional variance of A(T) given G(T), when $G(T) \leq K$, will be smaller than the unconditional variance for low to moderate strike values. As such we expect the choice of Ψ may be less of an issue than when we are only matching the conditional moments. Since Curran's approximation works quite well in practice, we adapt his idea to obtain two PEB approximations:

Curran2M+ approximation:

$$\Psi(G(T)) = G(T) + \exp\left(\mu_{A|G}(G(T)) + \sigma_{A|G}(G(T)) \cdot Z\right) \tag{7.81}$$

Curran3M+ approximation:

$$\Psi(G(T)) = \alpha_{A|G}(G(T)) + \exp\left(\mu_{A|G}(G(T)) + \sigma_{A|G}(G(T)) \cdot Z\right) \tag{7.82}$$

The Curran2M+ functional form retains the natural restriction that $A(T) \geq G(T)$. By imposing (7.79) we extend the Curran2M approximation from a local two-moment fit at the strike price to a global two-moment fit. We note this approximation was also considered in Deelstra et al. [2004]. The Curran3M+ approximation sacrifices the natural restriction that $A(T) \geq G(T)$. It may however still be an improvement upon the Curran2M+ approximation, as the conditional skewness is now also matched exactly. The Curran3M+ approximation can also be seen as an extension of the 3-moment matching approximation of Posner and Milevsky [1998]. Here, and also in Hill, Hill and Holder [1976] it is shown how to determine the coefficients α, μ and σ.

Both approximations ensure that the expectation within the integrand in (7.80) will be a Black-Scholes like expression, so that the evaluation of the integral is fairly straightforward. Note that when the amount of observations in the average is large, the extra computational effort required in the Curran3M+ approximation will be large, as we need to evaluate the third moment, a process of order $O(N^3)$, in each integration point.

The PEB approximations are found to be quite robust with regard to the specification of the conditioning variable and the approximation we use for the conditional law of the arithmetic average. Deelstra et al. [2004] have already demonstrated the first point, as they considered the Curran2M+ and Curran2M approximations for basket options. Using either Λ_{FA} or Λ_{GA} did not affect their results much. For the second point we mention that the Curran2M+ and 3M+ approximations seemed to yield the best results over a number of other choices we considered, though results for other distributions were close. To demonstrate the robustness with regard to the specification of the conditional law we include the 2M+Uniform approximation in our numerical results. This is obtained by replacing the lognormal random variable in (7.82) with a uniform random variable.

[41] We were not able to derive a sharp lower bound.

7.6. Numerical results and conclusions

In this final section we compare the new upper bound and the partially exact and bounded approximations to other bounds and approximations we mentioned in this document. We will focus on equally weighted, discretely sampled arithmetic Eurasian fixed strike calls. The stock underlying the arithmetic average will be dividend protected. In the first section we briefly give implementation details for the bounds and approximations we reviewed, and compare calculation times. The next section compares SLNQuad, and other choices one can make in Thompson's generalised Gaussian upper bound framework, to all other upper bounds. In the third section we compare the performance of SLNQuad and the PEB approximations suggested in the previous section (Curran2M+ and Curran3M+) to the lower bounds and approximations we mentioned in Sections 7.3 and 7.5 of this document. Several Greeks are calculated in order to assess whether the new bounds or approximations can be used to determine hedging positions. We finish with conclusions and recommendations.

Before starting the comparison, it is worthwhile to briefly review the outcomes of other studies. This will show us on what kind of examples we should focus our attention. The two studies we look at are those of Turnbull and Wakeman [1991] and those of Nielsen and Sandmann [2003]. Their examples also feature in other papers, e.g. in Vanmaele et al. [2006], which uses the examples from both studies to assess the performance of their upper bounds. In Table 7.2 on the next page the details are given about the examples in these studies. The first six columns describe the settings of the parameters for each of the examples. As before, N indicates the number of samplings in the average and r is the riskfree rate. The next three columns supply the spread between the smallest upper bound and the largest lower bound, for the at-the-money point. This is the point where the price of a call is equal to that of a put, i.e. when the strike is equal to the forward price of the arithmetic average. We measure the spread in this point as the at-the-money point is where many approximations already display large errors:

$$\frac{e^{-rT}(\tilde{c}_A - c_A)}{S} \cdot 10000 \tag{7.83}$$

i.e. the discounted difference between the true forward price and the approximation (or bound), divided by the spot price, in basispoints (bp). It is a measure of the absolute error made. In the following we will mostly use this as a measure of error, as relative errors tend to explode for out-of-the-money calls. The spread between the upper bound and lower bound is simply the difference in their pricing error. In brackets the relative errors (in percentages) are supplied, where the Curran3M+ price acts as the true price. As we will see shortly, this price is very close to the true price; the error introduced hereby is thus rather small. The final column gives the pricing error of the Levy approximation.

The examples used by Turnbull and Wakeman have a relatively short maturity and a large number of averagings within this period. The volatility of the arithmetic average is therefore quite small, which is reflected in the spreads. There is little room for improvement left in this example, since even the simple (but widely used and quite effective in low volatility scenarios) Levy approximation prices the option with virtually no error. Nielsen and Sandmann already used two examples which are better test cases for the various bounds and approximations. Had they considered Thompson's upper bound, the error at the at-the-money point would have been reduced quite drastically. The SLNQuad bound performs even better. It is clear that the Levy approximation already displays larger errors for these examples.

For our own examples we chose to stress test the bounds and approximations even further than is done in Nielsen and Sandmann. In the first example we consider an Asian with yearly averaging and a maturity of 5 years. The volatility is an extremely high 50%. For our second

Table 7.2: Spread between min. upper bound and max. lower bound, pricing error in Levy approximation for various studies

Studies	**Averaging frequency**	Maturity	N	r	σ	**Spread in bp. between min. UB and max. LB (relative error)**			**Pricing error (bp) in Levy approximation (Relative error)**
						Without Thompson and SLNQuad	*With Thompson, without SLNQuad*	*With Thompson, SLNQuad*	
Turnbull-Wakeman[42]	Daily	120d	30	8.62%	20%	0.57 (0.14%)	0.00 (0.00%)	0.00 (0.00%)	0.01 (0.00%)
	Daily	120d	30	8.62%	30%	1.27 (0.20%)	0.01 (0.00%)	0.01 (0.00%)	0.02 (0.00%)
	Daily	120d	30	8.62%	40%	2.22 (0.27%)	0.03 (0.00%)	0.03 (0.00%)	0.05 (0.01%)
Nielsen-Sandmann	Monthly	3y	36	4%	25%	20.63 (2.13%)	1.97 (0.20%)	1.72 (0.18%)	8.94 (0.92%)
	Monthly	10y	120	4%	25%	53.11 (3.40%)	14.53 (0.93%)	10.14 (0.65%)	47.96 (3.07%)
Our examples	Yearly	5y	5	5%	50%	134.59 (5.06%)	78.94 (2.97%)	34.20 (1.29%)	124.88 (4.70%)
	Yearly	30y	30	5%	25%	110.54 (5.78%)	73.74 (3.86%)	36.46 (1.91%)	135.69 (7.09%)

Table 7.3: Number of integration points required for an accuracy (in terms of pricing error) of $1 \cdot 10^{-3}$ bp; calculation times in brackets (in seconds, per 100 prices)

Maturity	**σ**	**UB_1**	**PECUB**	**ICUB**	**Thompson**	**SLNQuad[43]**	**Curran2M**	**Curran2M+**	**2M+Uniform**	**Curran3M+**
5y	50%	17 (0.08)	15 (0.38)	69 (0.77)	81 (0.29)	98 (0.64)	15 (0.03)	16 (0.09)	94 (0.33)	12 (0.16)
30y	25%	20 (2.00)	16 (2.86)	39 (3.22)	99 (2.01)	144 (4.62)	25 (0.17)	12 (1.10)	119 (10.34)	12 (14.21)

Table 7.4: Calculation times for other bounds and approximations (in seconds, per 100 prices)

Maturity	**σ**	**LB_{GA}**	**LB_{FA}/CUB**	**LB_{FA}/ICUB**	**2M**	**3M**	**4M**
5y	50%	0.02	0.05	0.81	0.01	0.01	0.02
30y	25%	0.12	0.81	4.18	0.03	0.04	0.06

[42] Note that the riskfree rate is equal to ln(1+0.09). Furthermore, the averaging occurs at the end of the contract, whereas in all other examples the averaging occurs throughout the life of the contract.

[43] To determine the optimal value of the scaled volatility we used 40 integration points.

example we lower the volatility to a moderate 25%, but increase the maturity to 30 years. The aforementioned unit-linked insurance policies, which contain embedded rate-of-return guarantees, tend to have long maturities. Although one is not advised to use a deterministic riskfree rate over such a long maturity (see e.g. Schrager and Pelsser [2004]), this example can still serve as a test case of how the bounds and approximations perform for long maturities.

7.6.1. Implementation details

Before letting the numerical results speak for themselves, we briefly discuss some implementation issues[44] for the various approximations and bounds. All calculations were carried out on a Pentium 4 2.66 GHz PC, in Visual Basic for Excel. Care was taken to optimise the code to allow for the fastest possible execution, given the programming language and processor. One should however not focus on the absolute time it took to calculate the various approximations and bounds, but on the relative difference in calculation times between the various methods.

To generate the true prices we numerically solved Večeř's PDE in (7.17) using a Crank-Nicolson finite-difference discretisation. Since in both examples the variance of the arithmetic average is quite high, due to either the high volatility or the long maturity, we found that a logarithmic transformation of the state space variable increased the numerical stability of the solutions. We transformed coordinates by defining $Y(t) = -\ln(1 - Z(t))$, where $Z(t)$ is the value of the traded account divided by the stock price. Remember from Section 7.2 that $Z(t)$ can be related one-to-one to the strike of the Asian option under consideration. For the 50% volatility example we built a grid for strike prices ranging from 2.5 to 35000 (when S = 100). For the 25% volatility example the strike prices ranged from 0.5 to 55000. At the high strike prices we set the value of the option equal to zero; similarly, at low strike prices[45] we set the value equal to $e^{-rT}(\mathbb{E}[A(T)] - K)$. The ranges were chosen such that the error made was negligible.

For both examples we solved the PDE with three space/time points combinations; in the 5 year example the combinations were 500/5000, 1000/10000 and 2000/20000. We performed three Richardson extrapolations on the various outcomes. The largest spread (in bp) between the prices resulting from the second and third extrapolation was smaller than $5 \cdot 10^{-4}$ bp for both examples. The first reduction mentioned in Section 7.2.2 indeed made the results more stable. The second reduction notably decreased the calculation time. Whereas the calculation time for the 1000/10000 space/time points combination was 47.41 seconds prior to using the reduction, it decreased to 38.08 seconds when using it, a reduction of about 20%.

We aim to compare the calculation times of the various methods in a fair manner. Since most approximations and bounds are based on a discretisation, we chose to find that discretisation where the pricing error (when compared to the true value for that approximation/bound) was just below $1 \cdot 10^{-3}$ bp. In formulae, if the approximation uses a discretisation based on N points, we searched for the smallest N for which:

$$\frac{e^{-rT}\left|\widetilde{c}_A(N) - \widetilde{c}_A(\infty)\right|}{S} \cdot 10000 \leq 1 \cdot 10^{-3} \tag{7.84}$$

For the finite-difference solution of the PDE in the 50% volatility example, we found that performing one Richardson extrapolation on the 500/5000 and 1000/10000 space/time point

[44] Naturally it goes too far to discuss all implementation details here. More details are available on request.
[45] Note that this condition is usually not invoked when the second reduction of section 3.2 is used.

combinations already yielded this accuracy. The combined calculation time for these two grids was 48.66 seconds. Calculation times were roughly the same for the 25% volatility example.

For the approximations and bounds that required numerical integration, we used Gauss-Legendre quadratures, see e.g. Press et al. [1996]. All upper bounds we considered require numerical integration. We must mention that the integral displayed for the ICUB and the PECUB in Vanmaele et al. [2006] was not numerically stable. Whereas their integrals were obtained by conditioning on a uniform random variable (i.e. $N(\frac{\Lambda-\mu_\Lambda}{\sigma_\Lambda})$ instead of Λ), we found it numerically much more stable to condition directly on a standardised version of Λ. The approximations requiring numerical integration are the recent approximations of Vyncke et al. [2004] (which we will refer to as the LB_{FA}/CUB and LB_{FA}/ICUB approximations), and the PEB approximations.

The search for the smallest number of integration points for which the error made in the discretisation was smaller than $1 \cdot 10^{-3}$ bp was done over the set of strike prices we used to generate the results in the following sections. The results and corresponding calculation times are displayed in Table 7.3. For the UB_1 and PECUB upper bounds the results displayed are only for the geometric average as conditioning variable. The calculation times and number of integration points required are however comparable when we use Λ_{FA} as the conditioning variable. In Table 7.4 calculation times are supplied for other bounds and approximations, namely the $LB(\Lambda_{GA})$ bound, which does not require any numerical integration, the approximations of Vyncke et al. and the moment matching approximations 2M, 3M and 4M. The 2M approximation is Levy's approximation; the other approximations stem from Posner and Milevsky's [1998] article. The 3M approximation fits a shifted lognormal distribution to the arithmetic average, whereas the 4M approximation fits a Johnson Type-II distribution to the first four moments of the average.

The performance of the various bounds and approximations will be the topic of the next two sections. When discussing their performance, we will reflect on their calculation times.

7.6.2. Comparison of all bounds

In this section we will discuss the performance of various parameterisations of Thompson's generalised Gaussian upper bound (7.39)-(7.41), relative to the other upper bounds known in the literature. The comparison will be on basis of our first example from Table 7.2, where we consider pricing a discretely sampled arithmetic Eurasian fixed strike call. The average is based on five yearly observations, σ = 50%, and r is equal to 5%. Finally, S(0) = 100. The following varieties of Thompson's upper bound are considered:

- *Thompson*: Thompson's [1999a] upper bound, with $\eta(t) = \xi(t) = 1$ and $\overline{\sigma} = \sigma$;
- *Thompson($\overline{\sigma}$)*: same as *Thompson*, except that we numerically optimise over $\overline{\sigma}$;
- *SLNQuad*: same as *Thompson*, using a shifted-lognormal approximation to approximate the optimal μ; the optimal $\overline{\sigma}$ is approximated via a quadratic[46] approximation to the upper bound.
- *SLN*: same as *SLNQuad*, except that the optimal $\overline{\sigma}$ was found via numerical optimisation;
- $\mu, \overline{\sigma}$: Optimal values for μ and $\overline{\sigma}$, $\xi(t) = 1$;
- μ, η: Optimal values for μ and η, $\xi(t) = 1$;
- $\mu, \xi, \overline{\sigma}$: Optimal values for μ, ξ and $\overline{\sigma}$;
- μ, ξ, η: Optimal values for μ, ξ and η; optimal upper bound of Thompson's generalised form;

[46] We evaluated the upper bound in $\overline{\sigma} = 0.5\sigma$, 0.75σ and σ and fit a quadratic function through these values to approximate the optimal scaled volatility.

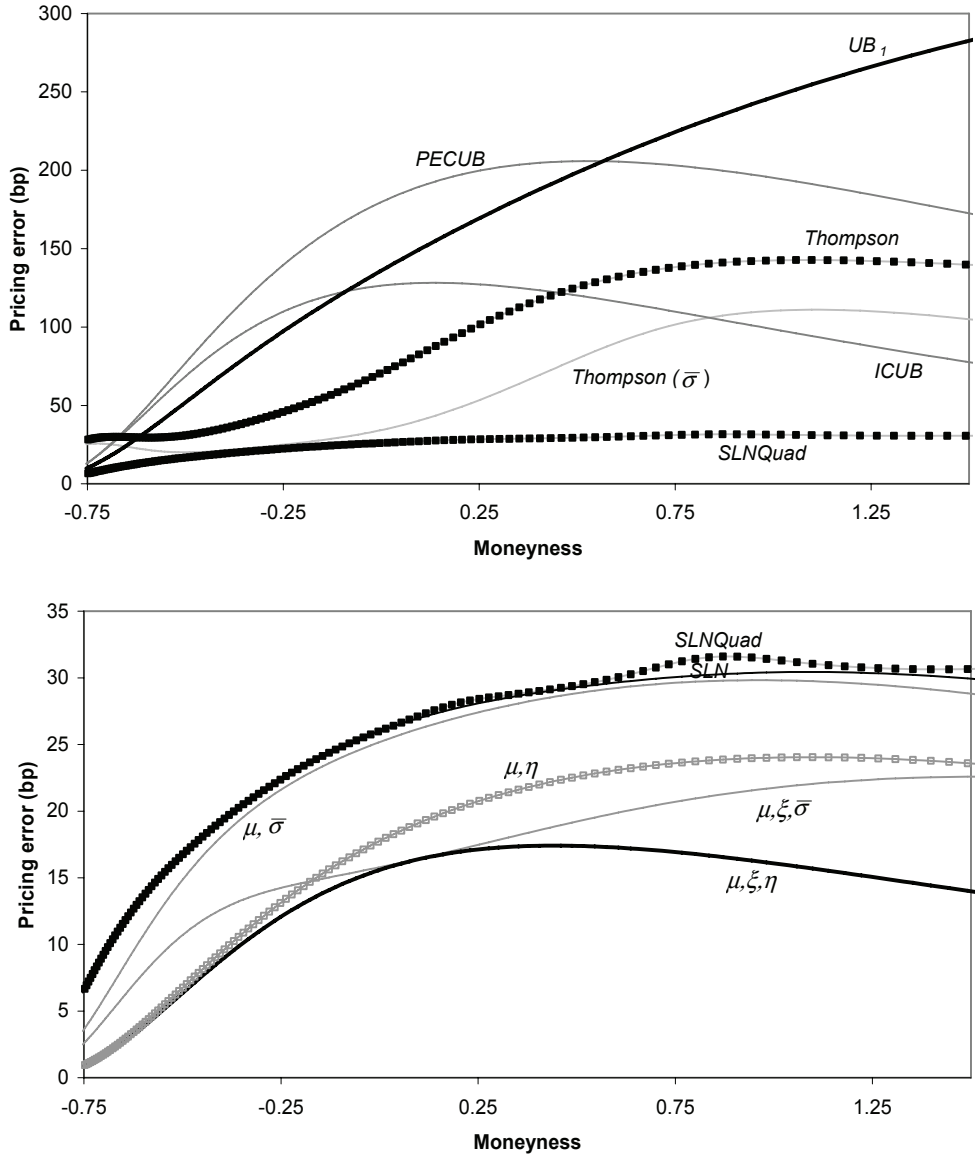


Figure 7.2: Parameterisations of Thompson's generalised Gaussian upper bound, σ = 50%, 5y maturity

The results are displayed in Figure 7.2. Note that the second picture is an enlargement of the first, to make it easier to distinguish between the numerous parameterisations. The horizontal axis denotes the moneyness of the option, which is an affine transformation of the strike price, defined as $K/\mathbb{E}[A(T)]-1$, i.e. the quotient of the strike and the forward of the arithmetic average, minus 1. Negative moneyness corresponds with in-the-money Asian calls, positive moneyness with out-of-the-money calls. A moneyness of zero indicates that the option is at-the-money. The range of strikes over which the graph is plotted was chosen such that the probability of the arithmetic average exceeding either side of the range was approximately equal to 5%.

Firstly considering Thompson's original parameterisation, it is clear that choosing $\overline{\sigma}$ optimally tightens the upper bound considerably. Using the shifted-lognormal approximation (cf. Section 7.4.3) to approximate the optimal function μ has an even bigger impact, in particular for large strikes. The SLNQuad upper bound, which uses a quadratic approximation to approximate

the optimal scaled volatility, is very close to the SLN upper bound, which uses the true value of the optimal scaled volatility. Furthermore, the upper bound obtained by optimally choosing both the function μ and $\overline{\sigma}$, is very close to both SLN and SLNQuad. These facts combined make that the SLNQuad upper bound is a fast and astonishingly accurate approximation to the upper bound found by optimally choosing μ and $\overline{\sigma}$. As far as the other parameterisations go, it is clear that by optimally choosing η and/or ξ, we can obtain even tighter upper bounds. Although this is theoretically interesting, it is far from practical as it involves numerically optimising over a large set of parameters when the number of fixings is large.

Comparing the various formulations of Thompson's upper bound to the other bounds included in Figure 7.2, we find that SLNQuad already outperforms all known upper bounds. This remains true when varying the parameters and the contract specifications. Reflecting on the calculation times required (Table 7.3), we notice that although the SLNQuad takes longer to evaluate than most bounds, it is not that much slower than the ICUB, which for large strike values is the second-best upper bound (and in fact faster for the 50% volatility example). Concluding, the SLNQuad upper bound is considerably sharper than all currently known upper bounds, and does not require a drastic increase in calculation time when compared to other upper bounds.

7.6.3. Comparison of all approximations

Having concluded that the SLNQuad upper bound outperforms all known upper bounds, and can be computed much faster than the even tighter variations of Thompson's generalised Gaussian upper bound, we now turn to testing the performance of the PEB approximations we introduced in Section 7.5, and set them off against various other approximations. We also include the lower bound LB(Λ) in the comparison, with both Λ_{FA} and Λ_{GA} as conditioning variables. The approximations under consideration are Curran's approximation (Curran2M), the three PEB approximations (Curran2M+, 2M+Uniform and Curran3M+), as well as Levy's two-moment matching approximation (2M) and Posner and Milevsky's three and four-moment matching approximations (3M and 4M). Though we could have included many more approximations in this test, we chose to focus on the popular moment matching approximations. An approximation by Ju [2002], based on a Taylor expansion of the characteristic function of the arithmetic average, that is reported to work very well was found to be less accurate than Curran's and Rogers and Shi's lower bound in our examples, so that we excluded it from the comparison. Finally, we also include the SLNQuad upper bound in the numerical results, to set off its performance against the various approximations and bounds. The results are displayed on the following two pages.

Like in Section 7.6.2 the range of strikes over which the comparison is made was chosen such that the probability of A(T) exceeding either side of the interval was equal to 5%. Although this may seem like a very wide range of strike prices, we will show that the strike prices considered are not that extreme for the second example, where we consider an arithmetic average over 30 years. Suppose that we have a financial contract in which at each time t_0 (= 0) through t_{N-1} a fixed premium P is invested in the stock for the user; at the maturity date (t_N = T) the owner of the contract obtains the value of his investments at the maturity date. If this is lower than a guaranteed payment of K, the buyer of the contract receives this amount. The forward price of such a contract is equal to:

$$\mathbb{E}\left[\max\left(\sum_{i=0}^{N-1} P\frac{S(T)}{S(t_i)}, K\right)\right] = K + \mathbb{E}\left[\left(\sum_{i=0}^{N-1} P\frac{S(T)}{S(t_i)} - K\right)^+\right] \tag{7.85}$$

i.e. a guaranteed amount K, increased with a call option which gives the buyer the value of his investments above the guaranteed amount K. It is shown in Schrager and Pelsser [2004] that

when we are in the Black-Scholes framework, the option in (7.85) is equivalent to an Asian option. Indeed, in this case we have:

$$\mathbb{E}\left[\left(\sum_{i=0}^{N-1} P\frac{S(T)}{S(t_i)} - K\right)^+\right] = \frac{NP}{S(0)}\mathbb{E}\left[\left(\sum_{i=1}^{N}\frac{1}{N}S(t_i) - \frac{S(0)}{NP}K\right)^+\right] \tag{7.86}$$

Suppose that we want to guarantee the user a (continuously compounded) return equal to g. We would then choose the strike such that $K = \sum_{i=0}^{N-1} Pe^{g(t_N - t_i)}$. In case of yearly averaging the strike of the Asian option in (7.86) would equal $\frac{S(0)e^g(e^{gN}-1)}{N(e^g-1)}$. Returning to our example with 30 years maturity, a guaranteed return of g = 10% would mean setting the strike of the Asian option equal to 668.5, a moneyness of 1.8. Although a rate-of-return guarantee of 10% is a bit on the high side, it is not completely unrealistic – the chosen guaranteed level will depend on the real-world behaviour of the underlying.

Returning to the numerical results, we first note that the SLNQuad upper bound and also the optimal choice of Thompson's generalised Gaussian upper bound are still not as tight as LB(Λ), which is just an application of something as seemingly straightforward as Jensen's inequality.

As for the approximations, let us first consider the Curran2M approximation. Although it performs very well for the 50% volatility example, the 25% volatility example shows that care should be taken when using this approximation. The effect we noticed in Section 7.5.2, that the approximation diverges for large strike prices, is very noticeable in this example. As for the two, three and four-moment matching approximations, we notice that fitting more moments does not necessarily imply a smaller error. As anticipated earlier, we notice in both examples that indeed the error for the 4M approximation does not tend to zero when the strike price tends to zero. The second picture in both Figures 7.3 and 7.4 is again an enlargement of the previous picture. Note that the LB_{FA}/CUB, LB_{FA}/ICUB, Curran2M+ and Curran3M+ approximations were not included in the first picture, in order to not make the graph completely illegible. The results of the 2M+Uniform approximation were not included in any of the graphs, as its pricing error was quite close to that of the Curran2M+ approximation; it would therefore only clutter the graph.

The LB_{FA}/CUB approximation performs slightly better than the LB_{FA}/ICUB approximation. Comparing their calculation times in Table 7.2, we see that indeed the LB_{FA}/CUB should certainly be preferred, something which Vyncke et al. already concluded. These two approximations are quite a significant improvement over the 2M, 3M and 4M approximations, even though at heart they are also two-moment approximations.

With little extra calculation time however, the Curran2M+ method is virtually indistinguishable from a zero pricing error. For respectively the 50% and 25% volatility examples, its largest absolute pricing error (over the range of strike prices we considered) was 0.46 and 0.49 bp. The Curran3M+ method is even more accurate, with the largest absolute pricing errors equal to respectively 0.05 and 0.14 bp. This goes hand in hand with a significantly increased calculation time. As mentioned, calculating the third moment of the arithmetic average, conditional upon some Gaussian random variable Λ, is a process of the order $O(N^3)$, where N is the amount of fixings of the arithmetic average. This is particularly noticeable when N is large, as we see from Table 7.2.

As far as the robustness of the PEB approximations against the specification of the conditional law of the arithmetic average is concerned, we see that using a uniform distribution, which is totally different than the lognormal one, still yields a very reasonable approximation. The 2M+Uniform approximation still by far outperforms the LB_{FA}/ICUB and LB_{FA}/CUB approximations. Concluding, provided the conditioning variable is highly correlated with the arithmetic average, matching all first two conditional moments ensures that the approximation

Figure 7.3: Performance of bounds and approximations, σ = 50%, 5y maturity, yearly averaging

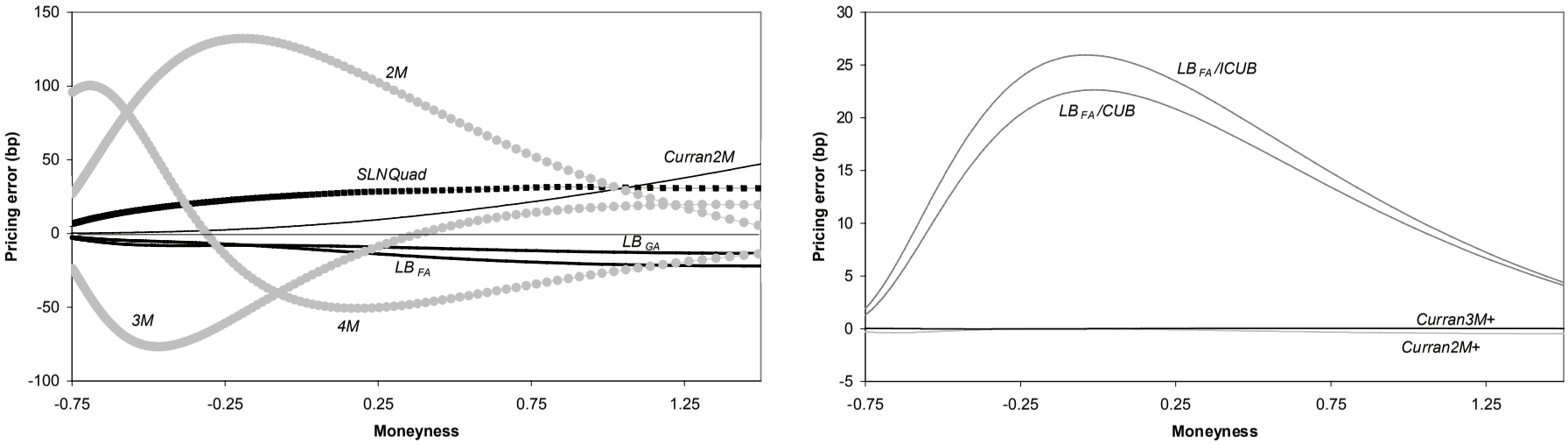


Table 7.5a: True price of Asian call and various approximations for three strike, σ = 50%, 5y maturity, yearly averaging

Strike	Moneyness	PDE	Curran2M+	2M+Uniform	Curran3M+	LB_{FA}/CUB	LB_{GA}	SLNQuad	Levy/2M
58.2370	-0.5	49.3944	49.3920	49.3967	49.3943	49.5103	49.3151	49.5617	50.3965
116.4741	0	26.5780	26.5778	26.5812	26.5781	26.8044	26.4962	26.8382	27.8268
174.7111	0.5	15.5342	15.5321	15.5436	15.5347	15.7073	15.4301	15.8286	16.3011

Table 7.5b: Pricing error (in bp)

Moneyness	Curran2M+	2M+Uniform	Curran3M+	LB_{FA}/CUB	LB_{GA}	SLNQuad	Levy/2M
-0.5	-0.24	0.23	-0.01	11.59	-7.93	16.73	100.21
0	-0.02	0.32	0.01	22.64	-8.18	26.02	124.88
0.5	-0.21	0.94	0.05	17.31	-10.41	29.44	76.69

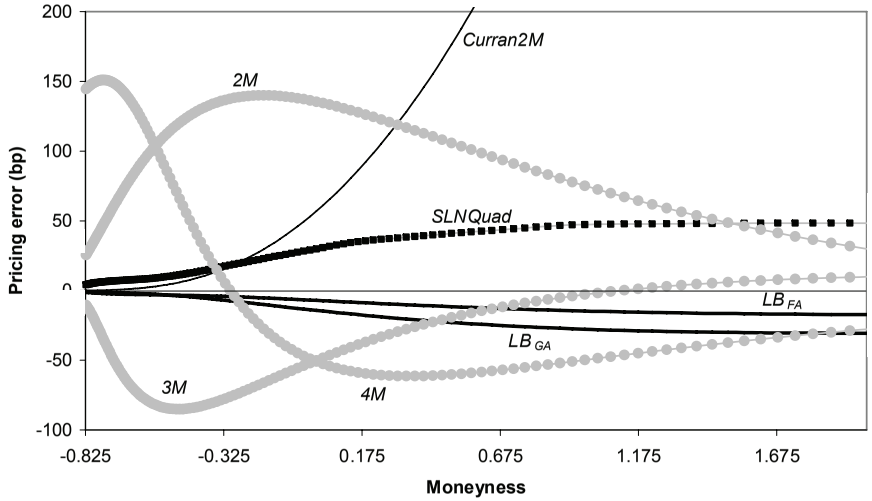
Table 7.5c: Implied volatility (in bp)

Moneyness	Curran2M+	2M+Uniform	Curran3M+	LB_{FA}/CUB	LB_{GA}	SLNQuad	Levy/2M
-0.5	-1.10	1.06	-0.06	53.32	-36.70	76.83	449.36
0	-0.05	0.65	0.01	45.50	-16.43	52.29	251.98
0.5	-0.39	1.75	0.09	32.07	-19.29	54.53	142.00

Table 7.5d: True Delta, Gamma and Vega and those resulting from bounds and approximations; Vega is in percentages

	PDE			LB_{GA}			Curran3M+			SLNQuad		
Moneyness	*Delta*	*Gamma*	*Vega*	*Delta*	*Gamma*	*Vega*	*Delta*	*Gamma*	*Vega*	*Delta*	*Gamma*	*Vega*
-0.5	0.8164	0.0023	0.2166	0.8159	0.0023	0.2112	0.8164	0.0023	0.2166	0.8167	0.0022	0.2287
0	0.5733	0.0045	0.4981	0.5727	0.0045	0.4934	0.5733	0.0045	0.4980	0.5747	0.0045	0.5158
0.5	0.3876	0.0045	0.5396	0.3873	0.0045	0.5351	0.3876	0.0045	0.5396	0.3898	0.0045	0.5591

Figure 7.4: Performance of bounds and approximations, σ = 25%, 30y maturity, yearly averaging



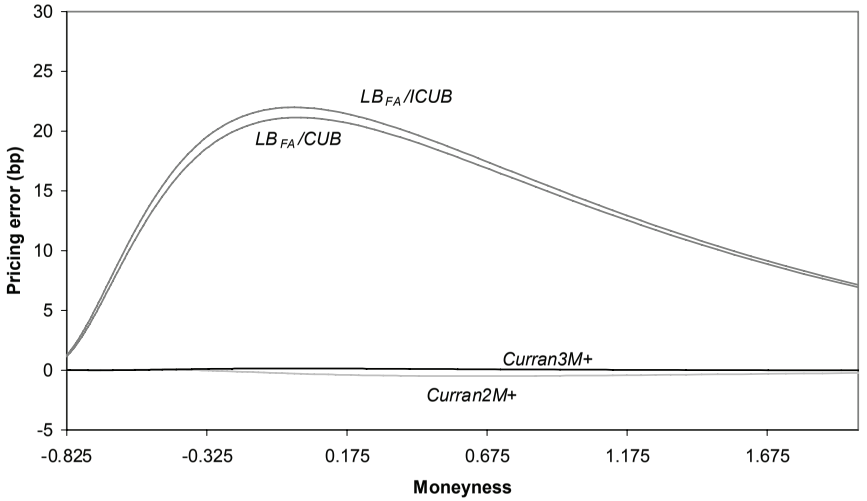


Table 7.6a: True price of Asian call and various approximations for three strike, σ = 25%, 30y maturity, yearly averaging

Strike	Moneyness	PDE	Curran2M+	2M+Uniform	Curran3M+	LB_{FA}/CUB	LB_{GA}	SLNQuad	Levy/2M
118.9819	-0.5	30.5153	30.5158	30.5246	30.5158	30.6554	30.4791	10.7417	31.6930
237.9638	0	19.1249	19.1220	19.1614	19.1263	19.3363	18.9845	29.4680	20.4832
356.9457	0.5	13.1168	13.1120	13.1676	13.1178	13.3010	12.8881	40.9490	14.1745

Table 7.6b: Pricing error (in bp)

Moneyness	Curran2M+	2M+Uniform	Curran3M+	LB_{FA}/CUB	LB_{GA}	SLNQuad	Levy/2M
-0.5	0.05	0.93	0.05	14.01	-3.62	10.74	117.77
0	-0.29	3.65	0.14	21.14	-14.04	29.47	135.83
0.5	-0.48	5.08	0.10	18.42	-22.87	40.95	105.77

Table 7.6c: Error in implied volatility (in bp)

Moneyness	Curran2M+	2M+Uniform	Curran3M+	LB_{FA}/CUB	LB_{GA}	SLNQuad	Levy/2M
-0.5	0.14	2.67	0.14	40.31	-10.45	30.93	334.94
0	-0.42	5.30	0.21	30.76	-20.36	42.91	199.75
0.5	-0.61	6.47	0.13	23.45	-29.07	52.18	135.13

Table 7.6d: True Delta, Gamma and Vega and those resulting from bounds and approximations; Vega is in percentages

	PDE			LB_{GA}			Curran3M+			SLNQuad		
Moneyness	*Delta*	*Gamma*	*Vega*	*Delta*	*Gamma*	*Vega*	*Delta*	*Gamma*	*Vega*	*Delta*	*Gamma*	*Vega*
-0.5	0.4655	0.0012	0.3467	0.4659	0.0012	0.3431	0.4655	0.0012	0.3467	0.4651	0.0013	0.3625
0	0.3503	0.0020	0.6884	0.3510	0.0020	0.6843	0.3503	0.0020	0.6886	0.3496	0.0020	0.7116
0.5	0.2663	0.0021	0.7858	0.2661	0.0021	0.7791	0.2663	0.0021	0.7860	0.2681	0.0020	0.8122

is very close to the true value of the Asian option, regardless of the distribution we choose to approximate the conditional law of the arithmetic average with. From Table 7.2 we see that the calculation time of the 2M+Uniform approximation is significantly higher than that of the Curran2M+ approximation. This is caused by the nature of the uniform distribution: above a certain strike price, the option price is equal to zero, making the integrand in (7.80) discontinuous. Therefore we needed a much larger amount of integration points to obtain the same accuracy as the Curran2M+ approximation.

Finally, we note that UB_1 (see Figure 7.2) is the theoretical upper bound on the PEB approximations. For low to moderate volatilities and maturities this is one of the best upper bounds available, as is concluded in Vanmaele et al. For higher volatilities and maturities our examples show that it is outperformed by other upper bounds. It is clear that the error in the PEB approximations is much smaller than the theoretical error estimate we derived. Further research can hopefully derive a tighter upper bound for the PEB approximations.

In Tables 7.5a and 7.6a we supplied the current prices the best four approximations, the $LB(\Lambda_{GA})$ lower bound, the SLNQuad upper bound and Levy's approximation, so that the reader can judge the accuracy for herself. Pricing errors for these approximations and bounds are supplied in Tables 7.5b and 7.6b. An alternative measure of error is used in Tables 7.5c and 7.6c, where we show the error in implied volatility. The displayed results are deviations from the true volatility in bp. For example, the volatility we should have used in the PDE to obtain the same at-the-money price as the LB_{FA}/CUB approximation is equal to 25.4031% in the 25% volatility example. Whether this is an acceptable error will depend on the liquidity of the market.

In a trading environment it is not only important to have a good estimate of the price of a financial derivative. Calculating the correct hedging positions is equally, if not more, important. Although we have shown that the PEB approximations are extremely close to the true price, this in principle does not say anything about the accuracy of their Greeks. Nevertheless, as the approximations we propose are smooth and non-oscillating functions, it may be expected that the accuracy of the Greeks is comparable. Nielsen and Sandmann [2003] and Vanmaele et al. [2006] supplied formulae for Delta, Gamma and Vega for their bounds. We will not do this here, as the Greeks of the PEB approximations and the SLNQuad upper bound are quite involved.

One of Nielsen and Sandmann's findings is that there are significant differences between the Vega of the $LB(\Lambda_{GA})$, the CUB and the $UB_2(\Lambda_{GA})$ bounds. As they supply the three Greeks for three volatility levels, the size of the Vega can be approximated numerically from their tables using e.g. a central difference. Doing this shows that their results for Vega cannot be correct, something which is also verified when recalculating their example. Furthermore, the Gammas they display are also incorrect. Recalculating their results shows that the three bounds they considered do in fact yield a reasonable estimate of Vega, invalidating their conclusions. This is also clear from our examples (Tables 7.5d and 7.6d). These examples show that the lower bound and the new SLNQuad upper bound already yield a good estimate on the size of the Delta, Gamma and Vega. As expected, the Curran3M+ approximation almost returns the true Greeks.

7.6.4. Conclusions and recommendations

In this chapter we deal with the pricing of arithmetic Eurasian fixed strike options in the Black-Scholes framework. Since the underlying in the Black-Scholes framework is modelled as a geometric Brownian motion, the arithmetic average is a sum (or integral) of correlated lognormal random variables. As there is no closed-form expression for the probability of this sum, pricing these options is not trivial. Many research efforts to date have focused on deriving an accurate value for these types of options. We contribute to three areas of research.

Our first contribution is in the area of research that deals with the pricing of Asian options via the numerical solution of a PDE. We showed the link between the one-dimensional PDE of

Rogers and Shi [1995], Andreasen [1999] and the recently derived PDEs by Hoogland and Neumann [2000a,b] and Večeř [2001]. For the latter PDEs, which Večeř found to be more numerically stable than the PDE of Rogers and Shi, we propose two reductions, which increase the numerical stability and reduce the calculation time.

Both the second and third contribution lie in the area of research that derives bounds on the value of an Asian option. We first show how Rogers and Shi's lower bound can be evaluated in closed form for basket options, using at most three numerical searches. This replaces a numerical integral over a discontinuous integrand by a closed-form expression, which is practically relevant. Hereafter we considerably sharpened Thompson's [1999a,b] upper bound. This is important for the practically relevant case of options with long maturities. Numerical results show that the resulting upper bound (SLNQuad) is considerably tighter than recently introduced upper bounds in studies by Nielsen and Sandmann [2003] and Vanmaele et al. [2006].

Our final contribution deals with analytical approximations. We introduce a new class of analytical approximations, the partially exact and bounded (PEB) approximations, which can be proven to lie between Rogers and Shi's lower bound, and an upper bound recently derived by Nielsen and Sandmann. The error made in these approximations tends to zero when the strike price tends to zero or to infinity. We show that the latter property is violated by Curran's approximation; it in fact diverges when the strike price tends to infinity. Adapting Curran's idea to the class of PEB approximations however yields two approximations that almost return the true price, even for high volatilities and long maturities. Furthermore, the class of PEB approximations seems to be very robust to both the choice of approximating distribution for the conditional law of the arithmetic average, as well as to the conditioning variable, provided of course it is highly correlated with the arithmetic average. The approximations are found to outperform all of the current state-of-the-art bounds and approximations.

Appendix 7.A – Proofs

Lemma 7.1:
Consider f(z) in (7.33). In case there is a $t \in [0,s]$ for which $\gamma(t) < 0$, and there is a $t \in (s,T]$ for which $\gamma(t) > 0$, the first derivative of f(z) has only one zero.

Proof:
The first derivative of f(z) can be written as:

$$f'(z) = \int_0^s \gamma(t) e^{\beta(t)\cdot z} dt + \int_s^T \gamma(t) e^{\beta(t)\cdot z} dt \tag{7A.1}$$

where $\gamma(t) \leq 0$ for $0 \leq t \leq s$ and $\gamma(t) > 0$ for $s < t \leq T$. There is a $t \in [0,s]$ so that $\gamma(t) < 0$. Hence we have $\lim_{z\to-\infty} f'(z) = -\infty$ and $\lim_{z\to\infty} f'(z) \geq 0$. There has to be at least one zero. Now write:

$$f'(z) = 0 \Leftrightarrow -e^{\beta(s)\cdot z} \int_0^s \gamma(t) e^{(\beta(t)-\beta(s))\cdot z} dt = e^{\beta(s)\cdot z} \int_s^T \gamma(t) e^{(\beta(t)-\beta(s))\cdot z} dt \tag{7A.2}$$

After dividing $e^{\beta(s)\cdot z}$ out, let us define the function on the left-hand side as g(z) and the function on the right-hand side as h(z). It is easy to verify that $\lim_{z\to-\infty} h(z) = 0$ and $\lim_{z\to\infty} h(z) = \infty$. Furthermore, since $\beta(t) - \beta(s) > 0$ on (s,T], h'(z) will be strictly positive for $z > 0$. For g(z) we consider two situations. If $\beta(t) = \beta(s)$ for all $t \in [0,s]$ where $\gamma(t) < 0$, then g(z) is constant and larger than zero, and h(z) will intersect g(z) exactly once. Otherwise we have $\lim_{z\to-\infty} g(z) = \infty$ and $\lim_{z\to\infty} g(z) < \infty$. Due to g'(z) being negative we again only have one solution to g(z) = h(z). This in turn implies f'(z) has only one zero. □

Lemma 7.3:
A sum of a normal and a lognormal random variable has a positive third central moment.

Proof:
Let $X = \mu_1 + \sigma_1 Z_1 + \exp(\mu_2 + \sigma_2 Z_2)$ with Z_1 and Z_2 standard normal random variates with correlation ρ. Tedious calculations show that its third central moment equals:

$$\mathbb{E}\left[(X - \mathbb{E}[X])^3\right] = F^3\left(e^{3\sigma_2^2} - 3e^{\sigma_2^2} + 2\right) + 6\rho\sigma_1\sigma_2 F^2\left(e^{\sigma_2^2} - 1\right) + 3F\rho^2\sigma_1^2\sigma_2^2 \tag{7A.3}$$

where $F = e^{\mu_2 + \frac{1}{2}\sigma_2^2}$, the expectation of the lognormally distributed part. As this is a quadratic equation in ρ, we find that the minimum is attained for a value of ρ equal to $-\sigma_1^{-1}\sigma_2^{-1}F(e^{\sigma_2^2} - 1)$. Substituting this in (7A.3) yields:

$$\mathbb{E}\left[(X - \mathbb{E}[X])^3\right] \geq F^3(e^{\sigma_2^2} - 1)^3 \tag{7A.4}$$

which is positive for all possible parameter values.

Chapter 8

Level-slope-curvature – fact or artefact? [47]

In stark contrast to the previous chapters, this chapter deals with a topic not related to the numerical techniques behind option pricing, but with the structure of term structure correlations. In an attempt to parsimoniously model the behaviour of the interest rate term structure, many studies find that using the first three principal components of the covariance or correlation matrix already accounts for 95-99% of the variability, a result first noted for interest rate term structures by Steeley [1990] and Litterman and Scheinkman [1991]. These results were also found to hold for the term structure of copper futures prices by Cortazar and Schwartz [1994], and also for the multiple-curve case, as shown by Hindanov and Tolmasky [2002].

This chapter does not deal with the question of how many factors one should use to model the interest rate term structure, or any term structure for that matter, but addresses the shape of the first three factors. The shape hereof is such that many authors, starting from Litterman and Scheinkman, have attached an interpretation to each of these three factors. The first factor, or indeed eigenvector of the covariance or correlation matrix, is usually relatively flat. As such it is said to determine the *level* or *trend* of the term structure. The second, which has opposite signs at both ends of the term structure, can be interpreted as determining the *slope* or *tilt*. The third factor finally, having equal signs at both ends of the maturity spectrum, but an opposite sign in the middle, is said to determine the *curvature*, *twist* or *butterfly* of the term structure.

A question that comes to mind is whether the observed pattern is caused by some fundamental structure within term structures, or whether it is merely an artefact of principal components analysis (PCA). Alexander [2003] in fact claims that "… the interpretation of eigenvectors as trend, tilt and curvature components is one of the stylised facts of all term structures, particularly when they are highly correlated". In this chapter we investigate sufficient conditions under which the level-slope-curvature effect occurs. To the best of our knowledge only one article has so far tried to mathematically explain this level-slope-curvature effect in the context of a PCA of term structures, namely that of Forzani and Tolmasky [2003]. They demonstrate that when the correlation between two contracts maturing at times t and s is of the form $\rho^{|t-s|}$, where ρ is a fixed positive correlation, the observed factors are perturbations of cosine waves with a period which is decreasing in the number of the factor under consideration. This correlation function is widely used as a parametric correlation function in e.g. the LIBOR market model, see Rebonato [2002]. In fact, Joshi [2000,2003] analyses a stylised example with three interest rates, which sheds some light on the conditions for the occurrence of an exponentially decaying correlation function; the same analysis is included in Rebonato [2002].

We formulate the level-slope-curvature effect differently than Forzani and Tolmasky. As noted, the first factor is quite flat, the second has opposite signs at both ends of the maturity spectrum, and the third finally has the same sign at both ends, but has an opposite sign in the middle. This observation leads us to consider the number of sign changes of each factor or eigenvector. If the first three factors have respectively zero, one and two sign changes, we say that we observe level, slope and curvature. Obviously this is only a partial description of the level-slope-curvature effect, as the sign-change pattern does not necessarily say anything about

[47] This chapter appears as Lord, R. [2007] and A.A.J. Pelsser. "Level-slope-curvature – fact or artefact?", *Applied Mathematical Finance*, vol. 14, no. 2, pp. 105-130.

the shape of the eigenvectors. However, if we want to analyse a general correlation matrix, choices have to be made.

Using a concept named total positivity, Gantmacher and Kreĭn considered the spectral properties of totally positive matrices in the first half of the twentieth century. One of the properties of a sub-class of these matrices, so-called oscillation matrices, is indeed that the n^{th} eigenvector of such a matrix has exactly n-1 sign changes. These results can be found e.g. in their book [1960, 2002]. With a minor generalisation of their theorems, we find sufficient conditions under which a term structure indeed displays the level-slope-curvature effect. The conditions have the nice interpretation of placing restrictions on the level, slope and curvature of the correlation curves.

Subsequently we turn to a correlation parameterisation which was recently proposed by Schoenmakers and Coffey [2003]. In matrix theory the resulting correlation matrix is known as a *Green's* matrix. The exponentially decreasing correlation function considered by Forzani and Tolmasky is contained as a special case of the Schoenmakers-Coffey parameterisation. The resulting correlation matrix has the nice properties that correlations decrease when moving away from the diagonal term along a row or a column. Furthermore, the correlation between equally spaced rates rises as their expiries increase. These properties are observed empirically in correlation matrices of term structures. Gantmacher and Kreĭn derived necessary and sufficient conditions for a Green's matrix to be an oscillation matrix, and hence to display level, slope and curvature. The Schoenmakers-Coffey parameterisation satisfies these restrictions, and hence also displays this effect. This actually confirms and proves a statement by Lekkos [2000], who numerically showed that when continuously compounded forward rates are independent, the resulting correlation matrix of zero yields displays level, slope and curvature.

Unfortunately total positivity and related concepts only provide a partial explanation of the level, slope and curvature phenomenon. We therefore end the chapter with a conjecture that an ordered correlation matrix with positive elements will always display level and slope. This conjecture is not proven, but is corroborated by results from a simulation study.

The chapter is organised as follows. In Section 8.1 we first introduce the terminology used in principal components analysis, and perform an empirical analysis of Bundesbank data[48], which contains interest rate data for the Euro market from 1972 onwards. Observing the same empirical pattern as in other studies, we mathematically formulate our criteria for the level-slope-curvature effect. In Section 8.2 we present and slightly modify some theorems of total positivity, leading to sufficient conditions for level, slope and curvature. We also provide an interpretation of these conditions. In Section 8.3 we turn to the Green's or Schoenmakers-Coffey correlation matrices, and show that they satisfy the conditions stated in Section 8.2. In Section 8.4 we consider sign regularity, a concept extending total positivity and end with our conjecture that positive and ordered correlation matrices will always display level and slope. Section 8.5 concludes.

8.1. Problem formulation

As stated before, we will in this section investigate conditions under which we observe the level-slope-curvature effect. Before mathematically formulating the problem, we will, for the purpose of clarity, briefly review some concepts of principal components analysis in the first subsection. For a good introduction to PCA we refer the reader to Jackson [2003]. In the second subsection we will review some empirical studies and conduct a PCA on historical data obtained from the Bundesbank database to illustrate the level-slope-curvature effect we will be analysing. Finally, the third and final subsection we will formulate our problem mathematically.

[48] This data can be obtained by selecting the daily term structure of interest rates from the time series database, subsection capital market, at http://www.bundesbank.de/statistik/statistik_zeitreihen.en.php.

8.1.1. Principal components analysis

Suppose we are based in a model with N random variables, in our case prices of contracts within the term structure. These random variables will be contained in a column vector **X**. For notational purposes we will assume that these random variables are centered. The goal of PCA is to describe the data we have with K << N orthogonal random variables, so-called principal components, which will be linear combinations of the original stochastic variables. We denote the k^{th} principal component as:

$$Y_k = \mathbf{X}^T \mathbf{w}_k \text{ for } k = 1, \ldots, N \tag{8.1}$$

Having determined all weight vectors $\mathbf{w}_i$ for i = 1, …, k, the weight vector w_k follows from:

$$\begin{aligned} &\max_{\mathbf{w}_k \in \mathbb{R}^N} \operatorname{Var}(\mathbf{X}^T \mathbf{w}_k) \\ &\text{s.t. } \forall_{i \le k}\ \mathbf{w}_i^T \mathbf{w}_k = 1_{[i=k]} \end{aligned} \tag{8.2}$$

We maximise the variance of each principal component, so that each component describes as large a part of the total variability as possible. The restriction that each weight vector must have length 1 only serves to remove the indeterminacy. Since the vectors $\mathbf{w}_k$ form an orthonormal system, it should not surprise the reader that the solution to $\mathbf{w}_k$ in (8.2) is the k^{th} eigenvector, i.e. the eigenvector associated with the k^{th} largest eigenvalue λ_k of the covariance matrix $\mathbf{\Sigma}$. The variance of the k^{th} principal component is therefore equal to $\operatorname{Var}(\mathbf{X}^T \mathbf{w}_k) = \mathbf{w}_k^T \mathbf{\Sigma} \mathbf{w}_k = \lambda_k$. A quantity often used in PCA is the proportion of variance explained by the k^{th} factor, which then simply equals the ratio of λ_k to the sum of all the eigenvalues. Note that all eigenvalues of a covariance (or correlation) matrix obtained from data will be positive, since any proper covariance matrix will be positive definite[49].

The final step is to determine which linear combination of the K principal components we have to use to describe the original data. One can show that the least squares estimate of these weights is in fact $\mathbf{W}_{(K)}$, the matrix with the first K eigenvectors as its columns. Then:

$$\mathbf{X} = \mathbf{W}_{(K)} \mathbf{Y}_{(K)} + \boldsymbol{\varepsilon} \tag{8.3}$$

where $\mathbf{Y}_{(K)}$ denotes the first K principal components. As a final note, we know by definition from (8.1) that the j^{th} entry of a weight vector $\mathbf{w}_k$ contains the weight with which X_j is embedded within the k^{th} principal component. Within PCA, the scaled eigenvectors $\sqrt{\lambda_k}\,\mathbf{w}_k$ are called *factors*, and its entries are referred to as *factor loadings*.

8.1.2. Empirical results

As mentioned in the introduction, many studies have dealt with a PCA of term structures, in particular term structures of interest rates. Although in this chapter we will mainly focus on the level-slope-curvature effect for an arbitrary covariance matrix, and the work will be more

[49] There are situations where one can obtain an estimate for a covariance matrix that is not positive definite, e.g. when one has missing data for one or more of the observed variables. However, any proper covariance matrix must be positive definite, since otherwise we can construct a linear combination of our random variables that has a negative variance. This clearly cannot be the case.

mathematical than empirical, it is nevertheless worthwhile to review a number of results from recent empirical studies. Hereafter we will investigate whether we find the level-slope-curvature pattern in the Bundesbank dataset.

We first mention a recent study by Lardic, Priaulet and Priaulet [2003]. Noticing that many studies use quite different methodologies, they pose a number of questions in their paper. The first question is whether one should use interest rate changes or levels as an input to a PCA. Naturally, interest rate levels are much more correlated than interest rate changes. They find that interest rate changes are stationary and conclude that therefore a PCA should be implemented with interest rate changes. Secondly, they investigate whether one should use centered changes, or standardised changes, i.e. whether a PCA should be conducted on a covariance or a correlation. Since the volatility term structure is typically not flat, but either hump-shaped or hockey stick shaped, there certainly is a difference between both methods. They conclude that a PCA with a covariance matrix will overweight the influence of the more volatile short term rates, and hence that one should use a PCA only with correlation matrices. Later on we will show that, under certain restrictions, our definition of level, slope and curvature will be such that it is irrelevant whether we use a covariance or a correlation matrix. Their final questions address whether the results of a PCA are dependent on the rates that are included in the analysis, and on the data frequency. Both aspects certainly affect the results one obtains, but we feel these questions are less important, as they depend on the application under consideration.

The second study we mention is that of Lekkos [2000]. He criticises the conclusion of many authors, starting from Steeley [1990] and Litterman and Scheinkman [1991], that three factors, representing the level, slope and curvature of the term structure, are sufficient to describe the evolution of interest rates. He claims that the results are mainly caused by the fact that most studies focus on zero yields, as opposed to (continuously compounded) forward rates. We will explain this now. In mathematical models the price of a zero-coupon bond is often written as:

$$\begin{aligned} P(t,T) &= \exp\big(-R(t,T)\cdot(T-t)\big) \\ &= \exp\left(-\int_0^T f(t,u)\,du\right) \\ &= \exp\big(-\alpha\big(f(t,t,t+\alpha)+\ldots+f(t,T-\alpha,T)\big)\big) \\ &= \big(1+\alpha F(t,t,t+\alpha)\big)^{-1}\cdot\ldots\cdot\big(1+\alpha F(t,T-\alpha,T)\big)^{-1} \end{aligned} \tag{8.4}$$

where P(t,T) is the time t price of a zero-coupon bond paying 1 unit of currency at time T. The first formulation uses the zero yield R(t,T) of the zero-coupon bond. The second through fourth formulations are in terms of forward rates. The second uses instantaneous forward rates, typically only used in mathematical models such as the Heath, Jarrow and Morton framework. The third is in terms of continuously compounded forward rates, where f(t,T,S) indicates the time t forward rate over [T,S]. Finally, the fourth formulation uses discretely compounded forward rates, which is the way interest rates are typically quoted in the market. Lekkos works with the third formulation. Relating the zero yields to these forward rates, where we use a fixed tenor equal to α, we find that the zero yields are averages of these continuously compounded forward rates:

$$R(t,T) = \tfrac{\alpha}{T-t}\big(f(t,t,t+\alpha)+\ldots+f(t,T-\alpha,T)\big) \tag{8.5}$$

Lekkos claims that the high correlation found for interest rate changes is mainly caused by this averaging effect in (8.5), and that we should therefore analyse the spectral structure of α-forward rates instead. In a numerical example he shows that when these α-forward rates are independent, the correlation matrix of the zero yields still displays the level-slope-curvature effect. We will in

fact prove this result later on, in Section 8.3. Although forward rates are not found to be independent in his empirical analysis, the spectral structure for α-forward rates he finds is quite different than that of the zero yields. The second and third factors cannot be interpreted as driving the slope and curvature of the term structure, and furthermore up to five factors are required to account for 95% of the total variation.

The final study we consider is that of Alexander and Lvov [2003]. One of the things considered in their paper are the statistical properties of a time series of discretely compounded forward rates. The time series are obtained from quoted rates via three different yield curve fitting techniques, namely two spline methods, and the Svensson[50] [1994] method. The functional form of an instantaneous forward rate with time to maturity T in the Svensson model is given by:

$$f(T) = \beta_0 + \beta_1 \exp(-\tfrac{T}{\tau_1}) + \beta_2 \tfrac{T}{\tau_1} \exp(-\tfrac{T}{\tau_1}) + \beta_3 \tfrac{T}{\tau_2} \exp(-\tfrac{T}{\tau_2}) \tag{8.6}$$

where the six parameters β_0 through β_3 and τ_1 and τ_2 have to be estimated from the data. The equation (8.6) is an addition of an asymptotic value and several negative exponentials, which are able to create humps or U-shapes. This model is able to capture several facts found empirically in the term structure of forward rates. Alexander and Lvov conclude that the choice of the yield curve fitting technique affects the correlation matrix much more than the choice of sample size. In their study they find that the Svensson curve gives the best overall sample fit, and through its parametric form it also yields the smoothest correlation matrices. As an interesting note, the first three factors from their PCA can all be interpreted as driving the level, slope and curvature of the term structure, contrary to the study of Lekkos. Although Alexander and Lvov use discretely compounded forward rates, whereas Lekkos uses continuously compounded rates, we would not expect this to affect the results so markedly. Therefore, we suspect that the differences between Alexander and Lvov's results and those of Lekkos can mainly be attributed to the difference in yield curve fitting technique. Lekkos uses the bootstrap method and linearly interpolates between missing quotes. This is known to cause kinks in the forward rate curve and as such will have quite some impact on the prices of exotic interest rate derivatives. It is therefore best market practice to use smooth curves. We will return to this issue later on in this section.

Using these insights, we will now ourselves conduct a PCA of Bundesbank data, which contains estimated Svensson curves for the Euro market from 1972 onwards. Until 1997 the curves have been estimated on a monthly basis. From August 1997 onwards, the curves are

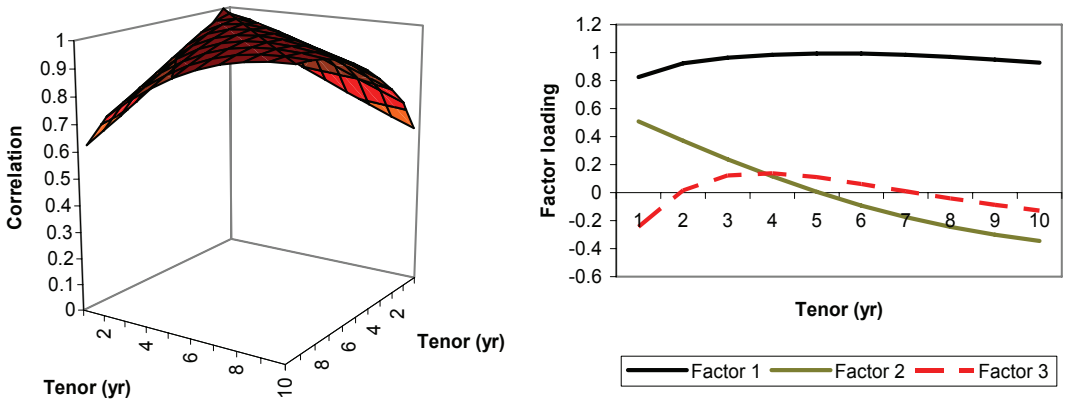


Figure 8.1: Estimated correlations between and first three factors of monthly log-returns on 1-10 year zero yields

[50] The Svensson model is also often referred to as the extended Nelson and Siegel model, as it is an extension of the original model by Nelson and Siegel [1987].

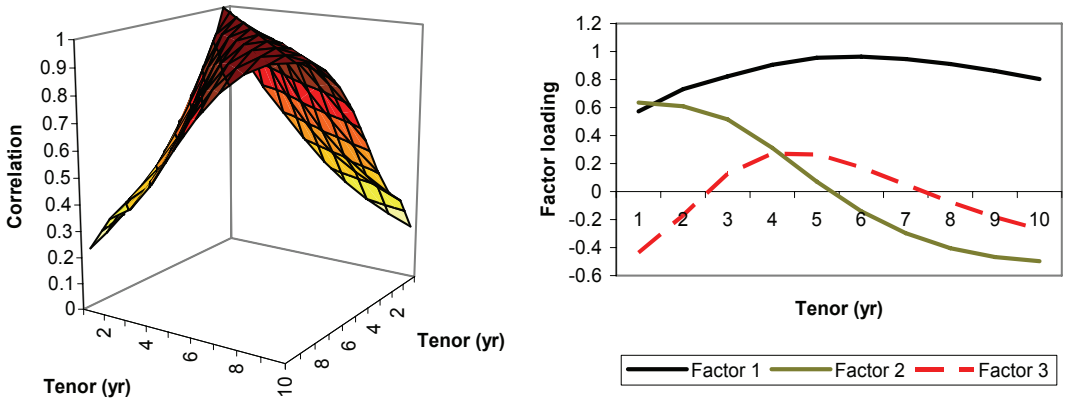


Figure 8.2: Estimated correlations between and first three factors of monthly log-returns on continuously compounded annual forward rates, with maturities ranging from 1-10 years

available on a daily basis. As we are only interested in reproducing the level-slope-curvature effect here, we ignore both the sample size and frequency issues, and use all end-of-month data from January 1980[51] up to and including June 2004. We calculated the correlations between the correlations between logarithmic returns on both zero yields, with tenors from 1 to 10 years, as well as on continuously compounded annual forward rates, with maturities also ranging from 1 to 10 years. The estimated correlation surfaces, as well as the first three factors following from a PCA, can be found in figures 1 and 2 on this and the previous page.

We indeed notice that that the resulting correlation surfaces are quite different for the zero yields than for the forward rates. The relation between zero yields and forward rates in (8.4) indicates that zero yields are averages of the forward rates. This relation by itself causes the correlations between the zero yields (or log-returns hereof) to be higher than those between the forward rates. Also noticeable in figure 8.2 is the well-documented (see e.g. Rebonato [2002]) convexity of the correlation curve when the front forward rates are taken as the reference rate, which changes to concavity when later-expiring forward rates are taken as the reference rates. For the full sample period we find that in the zero yield case the first three factors explain up to 99% of the total variability, which is reduced to 91% in the case of forward rates. Changing the sample period to 1987-1995, similar to the period considered in Lekkos [2000], changes this last number.

As a more extreme example of how non-smooth curves can distort the eigenvector pattern, we left out the observations of the 6, 8 and 9 year rates, and assumed the yield curve was piecewise

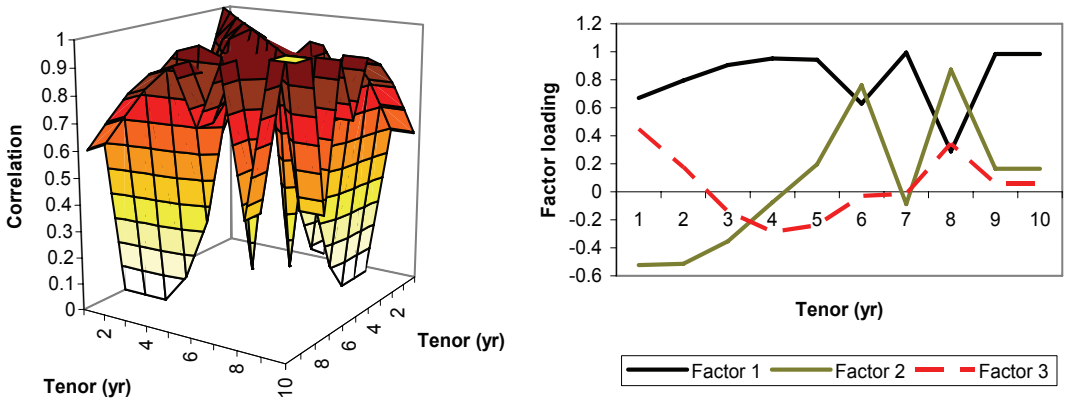


Figure 8.3: Same as **Figure 8.2**, only now with a bootstrapped yield curve

[51] Data from the seventies was not included as it changed the correlation estimates severely.

constant inbetween. Although the results for zero yields are not that different from contains estimated Svensson curves for the Euro market from 1972 onwards. Until 1997 the curves have been estimated on a monthly basis. From August 1997 onwards, the curves are **Figure 8.1**, the results for continuously compounded annual forward rates are markedly different, as shown in e
Figure **8.3**. The kinks in the discretely compounded forward rate curve have clearly distorted the usual pattern. The picture of the factors in e
Figure **8.3** is actually very similar to the factors Lekkos finds for a variety of currencies, and possibly implies that his choice of yield curve fitting technique is what causes the absence of level, slope and curvature in his study.

Since the previous analysis has demonstrated that level, slope and curvature do not always occur in correlation matrices, a natural question to ask is whether the pattern always occurs in the case of highly correlated and ordered stochastic systems. To this end consider the following artificially constructed correlation matrix:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.649 & 0.598 & 0.368 & 0.349 \\ 0.649 & 1 & 0.722 & 0.684 & 0.453 \\ 0.598 & 0.722 & 1 & 0.768 & 0.754 \\ 0.368 & 0.684 & 0.768 & 1 & 0.896 \\ 0.349 & 0.453 & 0.754 & 0.896 & 1 \end{pmatrix} \tag{8.7}$$

The matrix is a proper correlation matrix, and furthermore it satisfies certain properties which are typically found in empirical interest rate correlation matrices:

i) $\rho_{i,j+1} \leq \rho_{ij}$ for $j \geq i$, i.e. correlations decrease when we move away from the diagonal;

ii) $\rho_{i,j-1} \leq \rho_{ij}$ for $j \leq i$, same as i);

iii) $\rho_{i,i+j} \leq \rho_{i+1,i+j+1}$, i.e. the correlations increase when we move from northwest to southeast.

In words property (iii) means that the correlation between two adjacent contracts or rates increases as the tenor of both contracts increases. For instance, the 4 and 5 year rate are more correlated than the 1 and 2 year rate. Hence, the matrix in (8.7) is a correlation matrix of an ordered and highly correlated system, and could well be the correlation matrix of a term structure.

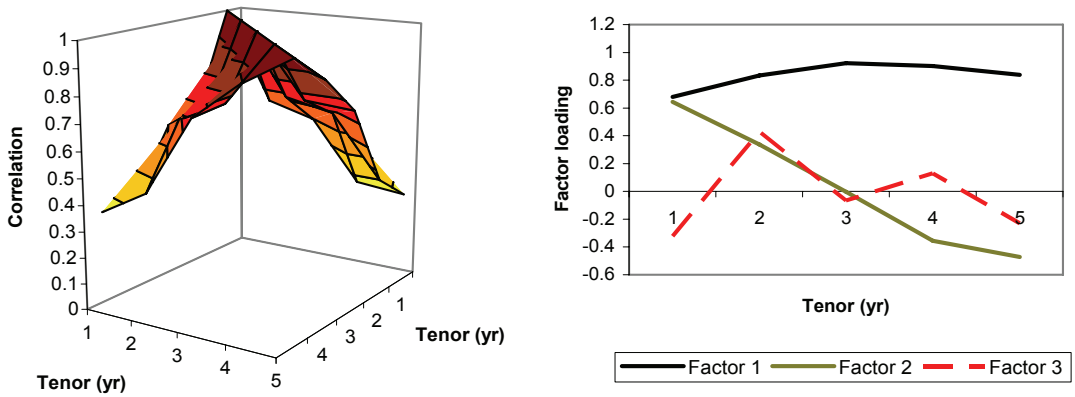


Figure 8.4: Artificially constructed correlation matrix which does not display curvature

The above figure, in which its correlation surface and first three factors are depicted, demonstrates however that conditions (i)-(iii) (i.e. (i), (ii) and (iii)) are insufficient for a matrix to display level, slope and curvature. Although the first two eigenvectors can certainly be interpreted as level and slope, the third eigenvector displays a different pattern than we usually find.

Concluding, although the correlation structure between either consecutive zero yields or forward rates is quite different, we find the level-slope-curvature effect in both cases, provided we use a smooth enough yield curve fitting technique. Finally, the fact that we have a highly correlated system, in combination with certain properties that empirical interest rate correlation matrices satisfy, is not enough for the correlation matrix to display the observed pattern. Additional or different conditions are required, something we will investigate in the next section. Using these empirical findings we will first mathematically formulate level, slope and curvature in the next section.

8.1.3. Mathematical formulation of level, slope and curvature

Regardless of whether we consider correlations between (returns of) zero yields or forward rates, we have seen the presence of level, slope and curvature. Before analysing this effect, we have to find a proper mathematical description. Forzani and Tolmasky [2003] analysed the effect in case the correlation structure between contracts maturing at times t and s is equal to $\rho^{|t-s|}$. Working with a continuum of tenors on [0,T], they analyse the eigensystem of:

$$\int_0^T \rho^{|y-x|} f(y)\,dy = \lambda f(x) \tag{8.8}$$

This problem is analogous to determining the eigenvectors of the correlation matrix, when we consider a discrete set of tenors. When ρ approaches 1, they find that the n^{th} eigenfunction (associated with the n^{th} largest eigenvalue), approaches the following function:

$$f(x) = \begin{cases} \cos\left(\frac{n\pi x}{T}\right) - \frac{2T\ln\rho}{n^2\pi^2 + T^2(\ln\rho)^2} & n \text{ even} \\ \cos\left(\frac{n\pi x}{T}\right) & n \text{ odd} \end{cases} \tag{8.9}$$

We notice that the first factor, corresponding to n = 0, approaches a constant, and hence will be relatively flat when the contracts in the term structure are highly correlated. Similarly, we notice that the n^{th} eigenfunction has a period equal to 2T/n. Hence, the second factor (n = 1) will have half a period on [0,T], and the third factor (n = 2) will have a full period on [0,T]. In the following figure we display the functions 1, $\cos\left(\frac{\pi x}{T}\right)$ and $\cos\left(\frac{2\pi x}{T}\right)$ on [0,T] where T = 10:

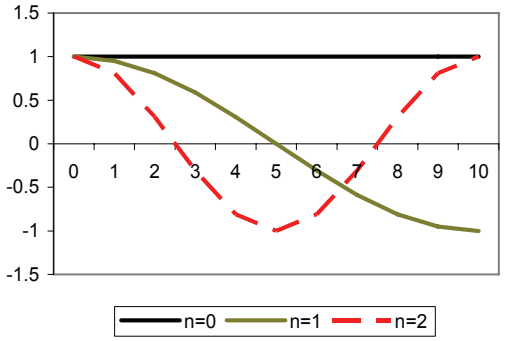


Figure 8.5: Limits of eigenfunctions for $\rho^{|t-s|}$ when $\rho \to 1$

Indeed, these limiting functions do resemble our notion of level, slope and curvature. The true eigenfunctions are perturbations of these cosine waves.

For the exponentially decaying correlation function the analysis is much facilitated, as the eigenfunctions can be calculated explicitly. We are not able to do this in general. Therefore we use another definition of level, slope and curvature, which will not require the knowledge of the explicit form of the eigenvectors or eigenfunctions. We notice in Figures 8.1 and 8.2 that the first factor is quite flat, and in fact has equal sign for all tenors. The second factor has opposite signs at both ends of the maturity range. Finally, the third has equal signs at both extremes, but has an opposite sign in the middle. If we therefore look at the number of times each factor or eigenvector changes sign, we notice that the first factor has zero sign changes, the second has one, and the third has two. This does not give a full description of what we perceive as level, slope and curvature. For instance, if in Figure 8.4 the third factor would be shifted slightly upwards, it would only have two sign changes, although it would still be dissimilar from the usual pattern. In all empirical studies we have seen however, our definition correctly signals the presence of level, slope and curvature, so that we expect it to be sufficient.

For a continuous eigenfunction, the number of sign changes is easily defined as the number of zeroes of this function. We will however mainly be working with a discrete set of tenors, which calls for a slightly different definition. For an $N \times 1$ vector $\mathbf{x}$ we mathematically define the number of sign changes as follows:

- $S^{-}(\mathbf{x})$ - the number of sign changes in $x_1, ..., x_N$ with zero terms discarded;
- $S^{+}(\mathbf{x})$ - the maximum number of sign changes in $x_1, ..., x_N$, with zero terms arbitrarily assigned either +1 or –1.

Both functions will only give a different number when the eigenvector contains zeroes and the non-zero elements at either side of a sequence of zeroes have the same sign. In the next section the distinction between both definitions will ultimately not be that important, as the sufficient conditions under which we will find the level-slope-curvature effect will imply that both definitions will give the same result when applied to the eigenvectors at hand. Ignoring zero terms within an eigenvector, we therefore define level, slope and curvature as the following sign-change pattern within the first three eigenvectors:

- Level: $S^{-}(\mathbf{x}^1) = 0$
- Slope: $S^{-}(\mathbf{x}^2) = 1$
- Curvature: $S^{-}(\mathbf{x}^3) = 2$

where $\mathbf{x}^i$ is the i^{th} eigenvector. In the next section we consider total positivity theory, which will provide us with sufficient conditions under which we find level-slope-curvature.

8.2. Sufficient conditions for level, slope and curvature

In this section we turn to theory on total positivity, which, for our formulation of the level-slope-curvature effect, will yield the right tools to clarify its occurrence. First we ntroduce some notation and concepts we will require in the remainder of this chapter. Second we review some results from total positivity theory. Minor generalisations hereof will yield sufficient conditions under which level, slope and curvature occur. Third, we rewrite these conditions, and show we can interpret them as being conditions on the level, slope and curvature of the correlation surface. We will mainly work with a discrete set of tenors, although we also touch upon the case where

we have a continuum of tenors. The continuous case will greatly facilitate the interpretation of the conditions we find.

8.2.1. Notation and concepts

Before turning to some theorems from total positivity theory, we need to introduce some notation and concepts. First of all we will be dealing with covariance or correlation matrices. A covariance matrix $\mathbf{\Sigma}$ of size $N \times N$ satisfies the following properties:

1. $\mathbf{\Sigma}$ is symmetric, that is $\mathbf{\Sigma} = \mathbf{\Sigma}^T$;
2. $\mathbf{\Sigma}$ is positive definite, i.e. for any non-zero vector $\mathbf{x} \in \mathbb{R}^N$ we have $\mathbf{x}^T\mathbf{\Sigma}\mathbf{x} > 0$.

Any matrix satisfying these properties is invertible and can be diagonalized as $\mathbf{\Sigma} = \mathbf{X}\mathbf{\Lambda}\mathbf{X}^T$, where the eigenvectors of the matrix are contained in $\mathbf{X}$, and the eigenvalues in $\mathbf{\Lambda}$. All eigenvalues are furthermore strictly positive. The correlation matrix $\mathbf{R}$ associated with $\mathbf{\Sigma}$ is obtained as:

$$\mathbf{R} = \text{diag}(\mathbf{\Sigma})^{-1/2}\,\mathbf{\Sigma}\,\text{diag}(\mathbf{\Sigma})^{-1/2} \tag{8.10}$$

where diag($\mathbf{\Sigma}$) is a matrix of the same dimensions as $\mathbf{\Sigma}$, containing its diagonal and zeroes everywhere else. Naturally $\mathbf{R}$ is also a covariance matrix.

The theorems in the next section will require the following concepts. For a given positive integer N we define:

$$I_{p,N} = \left\{\mathbf{i} = (i_1,...,i_p) \mid 1 \le i_1 < ... < i_p \le N\right\} \tag{8.11}$$

where of course $1 \le p \le N$. When $\mathbf{\Sigma}$ is an $N \times N$ matrix, we define for $\mathbf{i}, \mathbf{j} \in I_{p,N}$:

$$\mathbf{\Sigma}_{[p]}(\mathbf{i},\mathbf{j}) = \mathbf{\Sigma}\begin{pmatrix} i_1,...,i_p \\ j_1,...,j_p \end{pmatrix} = \det(a_{i_k j_\ell})_{k,\ell=1}^{p} \tag{8.12}$$

In terms of covariance matrices, definition (8.12) means we are taking the determinant of the covariance matrix between the interest rates indexed by vector $\mathbf{i}$, and those indexed by vector $\mathbf{j}$. The p^{th} compound matrix $\mathbf{\Sigma}_{[p]}$ is defined as the $\binom{N}{p} \times \binom{N}{p}$ matrix with entries equal to $\left(\mathbf{\Sigma}_{[p]}(\mathbf{i},\mathbf{j})\right)_{\mathbf{i},\mathbf{j} \in I_{p,N}}$, where the $\mathbf{i} \in I_{p,N}$ are arranged in lexicographical order, i.e. $\mathbf{i} \ge \mathbf{j}$ ($\mathbf{i} \ne \mathbf{j}$) if the first non-zero term in the sequence $i_1 - j_1$, ..., $i_p - j_p$ is positive.

8.2.2. Sufficient conditions via total positivity

Before turning to the theory of total positivity, we will solve the level problem. Perron's theorem, which can be found in most matrix algebra textbooks, deals with the sign pattern of the first eigenvector.

Theorem 8.1 – Perron's theorem
Let A be an $N \times N$ matrix, all of whose elements are strictly positive. Then A has a positive eigenvalue of algebraic multiplicity equal to 1, which is strictly greater in modulus than all other eigenvalues of A. Furthermore, the unique (up to multiplication by a non-zero constant) associated eigenvector may be chosen so that all its components are strictly positive. □

The result of the theorem only applies to matrices with strictly positive elements. Since the term structures we are investigating are highly correlated, this is certainly not a restriction for our purposes. The result is valid for any square matrix, not only for symmetric positive definite matrices. As long as all correlations between the interest rates are positive, this means that the first eigenvector will have no sign changes.

This has solved the level problem. For the sign-change pattern of other eigenvectors we have to turn to the theory of total positivity. The results in this section mainly stem from a paper by Gantmacher and Kreĭn [1937], which, in an expanded form, can be found in Gantmacher and Kreĭn [1960, 2002]. Most results can also be found in the monograph on total positivity by Karlin [1968]. For a good and concise overview of the theory of total positivity we refer the reader to Ando [1987] and Pinkus [1995]. The latter paper gives a good picture of the historical developments in this field, and the differences between the matrix and the kernel case.

A square matrix $\mathbf{A}$ is said to be totally positive (sometimes totally non-negative, TP), when for all $\mathbf{i}, \mathbf{j} \in I_{p,N}$ and $p \leq N$, we have:

$$\mathbf{A}_{[p]}(\mathbf{i}, \mathbf{j}) \geq 0 \tag{8.13}$$

In the case of covariance matrices, this means that we require the covariance matrix between $\mathbf{i}$ and $\mathbf{j}$ to have a non-negative determinant. When $\mathbf{i} = \mathbf{j}$ this will clearly be the case, as the resulting matrix is itself a covariance matrix, and will be positive definite. In the other cases the meaning of this condition is less clear. In the next section we will spend some time on interpreting these conditions. If strict inequality holds then we say that the matrix is strictly totally positive (STP). Furthermore, we say that a matrix is TP_k if (8.13) holds for $p = 1, ..., k$, and we define STP_k in a similar fashion. Hence, an $N \times N$ matrix is TP when it is TP_N, and STP when it is STP_N. Gantmacher and Kreĭn proved the following theorem for general STP matrices. A full version of their theorem also considers the so-called variation-diminishing property of such matrices, but we will here only deal with the sign-change pattern of such matrices. We reformulate their theorem for covariance matrices that are not necessarily STP, but only STP_k. Reading their proof shows that it can be altered straightforwardly to cover this case. For completeness we have included the proof in the appendix.

Theorem 8.2 – Sign-change pattern in STP_k matrices
Assume $\mathbf{\Sigma}$ is an $N \times N$ positive definite symmetric matrix (i.e. a valid covariance matrix) that is STP_k. Then we have $\lambda_1 > \lambda_2 > \ldots > \lambda_k > \lambda_{k+1} \geq \ldots \lambda_N > 0$, i.e. at least the first k eigenvalues are simple. Denoting the j^{th} eigenvector by $\mathbf{x}^j$, we have $S^-(\mathbf{x}^j) = S^+(\mathbf{x}^j) = j\text{-}1$, for $j = 1, ..., k$.

Proof: See the appendix. □

A consequence of theorem 8.2 is that a sufficient condition for a correlation matrix to display level, slope and curvature, is for it to be STP_3. Naturally all principal minors of a covariance matrix are determinants of a covariance matrix, and hence will be strictly positive. It is however not immediately clear what the remaining conditions mean – we will find an interpretation hereof in the following section. The conditions in theorem 8.2 can be relaxed somewhat further via the concept of an oscillation or oscillatory matrix, again due to Gantmacher and Kreĭn. The name oscillation matrix arises from the study of small oscillations of a linear elastic continuum, e.g. a string or a rod. An $N \times N$ matrix $\mathbf{A}$ is an oscillation matrix if it is TP and some power of it is STP. As in theorem 8.2, we slightly alter the original theorem by using the concept of an oscillation matrix of order k.

Theorem 8.3 – Oscillation matrix of order k
Akin to the concept of an oscillation matrix, we define an oscillation matrix of order k. An $N \times N$ matrix $\mathbf{A}$ is oscillatory of order k if:
1. $\mathbf{A}$ is TP_k;
2. $\mathbf{A}$ is non-singular;
3. For all i = 1, ..., N-1 we have $a_{i,i+1} > 0$ and $a_{i+1,i} > 0$.

For oscillatory matrices of the order k, we have that $\mathbf{A}^{N-1}$ is STP_k.

Proof: See the appendix. □

Gantmacher and Kreĭn proved theorem 8.3 and its converse for the STP case. As we are only interested in sufficient conditions for level, slope and curvature, we do not consider the converse. The proof of theorem 8.3 is included in the appendix for completeness, although the original proof carries over almost immediately.

Corollary 8.1
In theorem 8.2 we can replace the condition that the matrix is STP_k with the requirement that some finite power of it is oscillatory of order k.

Proof:
Suppose $\mathbf{\Sigma}$ is a positive definite symmetric N x N matrix, for which $\mathbf{\Sigma}^i$ is oscillatory of order k. As the matrix is invertible, we can write $\mathbf{\Sigma} = \mathbf{X\Lambda X}^T$, and hence:

$$\mathbf{\Sigma}^{i(N-1)} = \mathbf{X\Lambda}^{i(N-1)}\mathbf{X}^T \tag{8.14}$$

so that $\mathbf{\Sigma}^{i(N-1)}$ has the same eigenvectors as $\mathbf{A}$. Since $\mathbf{\Sigma}^{i(N-1)}$ is STP_k, we can apply theorem 8.2 to first find that $S^-(\mathbf{x}^j) = S^+(\mathbf{x}^j) = j\text{-}1$, for j = 1, ..., k. In other words, we have the same sign-change pattern for matrices of which a finite power is oscillatory of order k. Finally, the eigenvalues can be ordered as $\lambda_1^{i(N-1)} > \ldots > \lambda_k^{i(N-1)} \geq \ldots \lambda_N^{i(N-1)} > 0$. This directly implies that the first k eigenvalues are simple. □

With this corollary the sufficient conditions from theorem 8.2 have been relaxed somewhat. Instead of requiring that the covariance or correlation is STP_3, we now only need some finite power of it to be TP_3, invertible, and to have a strictly positive super- and subdiagonal. The following corollary states that multiplying an oscillatory matrix by a totally positive and invertible matrix (both of the same order), yields a matrix which is again oscillatory.

Corollary 8.2
Let $\mathbf{A}$ and $\mathbf{B}$ be a square $N \times N$ matrices, where $\mathbf{A}$ is oscillatory of order k, and $\mathbf{B}$ is invertible and TP_k. Then $\mathbf{AB}$ and $\mathbf{BA}$ are oscillatory of order k.

Proof:
We can verify whether a matrix is oscillatory of order k by checking its three defining properties. Obviously the first and second properties are satisfied for both matrices. We only have to check the third criterium, concerning the positivity of the super- and subdiagonal elements. For the superdiagonal we basically have:

$$(\mathbf{AB})_{i,i+1} = \sum_{j=1}^{N} a_{ij} b_{j,i+1} \tag{8.15}$$

which is certainly non-negative, due to the fact that both matrices are TP_k. One element contained in (8.15) is $a_{i,i+1}b_{i+1,i+1}$. For **A** we know that all superdiagonal elements are positive. Furthermore, since **B** is invertible, all its diagonal elements must be strictly positive, so that (8.15) is clearly strictly positive. The proof is identical for the subdiagonal. □

This corollary directly implies the following, so that when analysing the sign change pattern of oscillatory matrices, it does not matter whether we analyse covariance or correlation matrices.

Corollary 8.3
A valid covariance matrix is oscillatory if and only if its correlation matrix is oscillatory.

Proof:
Suppose we have a valid covariance matrix which can be written as $\boldsymbol{\Sigma} = \mathbf{SRS}$, where **S** is a diagonal matrix containing the (strictly positive) standard deviations on its diagonal, and **R** is the correlation matrix. The "if" part now follows. Since $\boldsymbol{\Sigma}$ is invertible, so is **S**. An invertible diagonal matrix with strictly positive diagonal elements is clearly totally positive. Hence, if **R** is oscillatory, so will **SRS** by virtue of corollary 8.2. The "only if" part follows similarly. □

Corollary 8.3 states that the sign change pattern in the eigenvectors will be the same in covariance and correlation matrices. A graph of the eigenvectors will however look quite different in both matrices, due to the fact that the term structure of volatilities is typically not flat. As argued in section 8.1.3, the actual shape of the eigenvectors, e.g. that the first eigenvector is relatively flat, is caused by the fact that the term structure is highly correlated.

Having derived sufficient conditions under which a matrix displays level, slope and curvature, we try to interpret these conditions in the next section.

8.2.3. Interpretation of the conditions

As we saw in the previous section, a sufficient condition for a covariance or correlation matrix to display level, slope and curvature, is for it to be oscillatory of order 3. We will here try to interpret these conditions. Remember that corollary 8.3 showed that our definition is invariant to whether we use a covariance or a correlation matrix, so that we opt to use correlation matrices for ease of exposure. For an $N \times N$ correlation matrix **R** to be oscillatory of order 3, we require that:

1. **R** is TP_3;
2. **R** is non-singular;
3. For all i = 1, ..., N-1 we have $\rho_{i,i+1} > 0$ and $\rho_{i+1,i} > 0$.

As any proper covariance or correlation matrix will be invertible, condition ii) is irrelevant. In the term structures we will be analysing, it seems natural to expect that all correlations ρ_{ij} are strictly positive. Condition iii) is immediately fulfilled, as is the case for the order 1 determinants from i). Under this mild condition we can already interpret the first eigenvector as driving the level of the term structure. Hence, the level of the correlations determines whether or not we have level.

Now we turn to the second order determinants. As the usual interpretation of a second order determinant as the signed area of a parallelogram is not very useful here, we need to find another one. Given that **R** is TP_1, it is also TP_2 if for $i \le j$ and $k \le \ell$:

$$\left|\begin{pmatrix} \rho_{ik} & \rho_{i\ell} \\ \rho_{jk} & \rho_{j\ell} \end{pmatrix}\right| = \rho_{ik}\rho_{j\ell} - \rho_{i\ell}\rho_{jk} \ge 0 \Leftrightarrow \rho_{ik}\rho_{j\ell} \ge \rho_{i\ell}\rho_{jk} \tag{8.16}$$

It is not immediately clear how this condition should be interpreted. However, since all correlations were assumed to be positive, we can rearrange (8.16) to find the following condition:

$$\frac{\rho_{ik}}{\rho_{i\ell}} \geq \frac{\rho_{jk}}{\rho_{j\ell}} \Leftrightarrow \frac{\rho_{j\ell} - \rho_{jk}}{\rho_{j\ell}} \geq \frac{\rho_{i\ell} - \rho_{ik}}{\rho_{i\ell}} \tag{8.17}$$

In words, condition (8.17) states that the relative change from moving from k to ℓ ($k \leq \ell$), relative to the correlation with ℓ, should be larger on the correlation curve of j than on the curve of i, where $i \leq j$. This says that on the right-hand side of the diagonal the relative change on correlation curves for larger tenors should be flatter than for shorter tenors, as is depicted in the Figure on the following page. On the left-hand side of the diagonal this is reversed – the relative change there should be larger for shorter than for larger tenors. The derived condition clearly puts a condition on the slopes of the correlation curves.

In practice we usually have a continuous function from which we generate our correlation matrix. With a continuum of tenors we do not analyse the eigensystem of a covariance matrix, but of a symmetric and positive definite kernel $K \in C([0,T] \times [0,T])$. The eigenfunctions and eigenvalues satisfy the following integral equation:

$$\int_0^T K(x,y)\phi(y)\,dy = \lambda\phi(x) \tag{8.18}$$

This setting is also analysed in Forzani and Tolmasky [2003] for a specific choice of K.

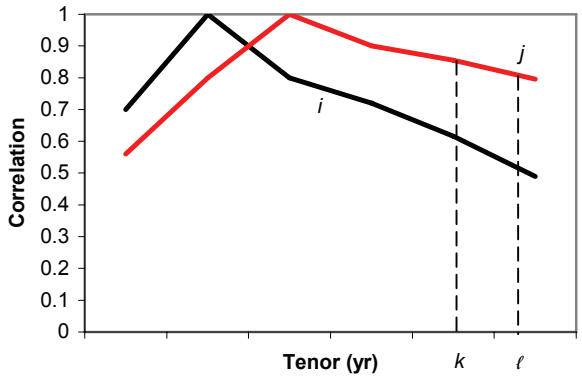


Figure 8.6: Two correlation curves from a TP_2 matrix

Analysing a continuous problem sometimes makes life easier, but surprisingly the analysis here remains essentially the same. The kernel case was historically studied prior to the matrix case, by O.D. Kellogg. Kellogg [1918] noticed that sets of orthogonal functions often have the property "that each changes sign in the interior of the interval on which they are orthogonal once more than its predecessor". He noted that this property does not only depend on the fact that the functions are orthogonal. As in the discrete case, total positivity of order n is equivalent to:

$$K\begin{pmatrix} x_1,\ldots,x_n \\ y_1,\ldots,y_n \end{pmatrix} = \det\left(K(x_i,y_j)\right)_{i,j=1}^{n} \geq 0 \tag{8.19}$$

for all x,y ∈ [0,T]. When n = 2 we regain condition (8.16): $K(x_1,y_1)K(x_2,y_2) \geq K(x_1,y_2)K(x_2,y_1)$. If we in addition assume that K is twice differentiable, one can show that an equivalent condition is:

$$K(x,y)\frac{\partial^2 K(x,y)}{\partial x \partial y} - \frac{\partial K(x,y)}{\partial x}\frac{\partial K(x,y)}{\partial y} = K(x,y)^2 \frac{\partial^2 \ln K(x,y)}{\partial x \partial y} \geq 0 \tag{8.20}$$

Note that if we have a kernel that only depends on the difference of the two arguments, in other words if K(x,y) = f(x-y), (8.20) states that f should be log-concave. A slightly stronger condition than (8.20) is obtained by considering the empirical properties of correlation matrices of term structures we mentioned in section 8.1.2. Typically correlations are positive, i.e. K(x,y) > 0. Secondly, correlations decrease if we move away from the diagonal along a row or a column, implying that $\frac{\partial K(x,y)}{\partial x}\frac{\partial K(x,y)}{\partial y} < 0$. From (8.20) we then see that K is TP_2 if $\frac{\partial^2 K(x,y)}{\partial x \partial y} \geq 0$. Again, if K only depends on the difference of its two arguments, this property requires f to be concave.

Although the condition for slope allows for a clear interpretation, the condition for curvature is much more cumbersome. We just present the final result as the intermediate steps again just follow from rewriting the determinant inequality in (8.13) for p = 3. We first define the relative change from moving from k to ℓ ($k \leq \ell$), along correlation curve i as:

$$\Delta_i(k,\ell) = \frac{\rho_{i\ell} - \rho_{ik}}{\rho_{i\ell}} \tag{8.21}$$

Using this definition, the matrix is obviously TP_2 if and only if $\Delta_j(k,\ell) \geq \Delta_i(k,\ell)$ for all i < j and k < ℓ. The additional condition we must impose for the matrix to be TP_3 is then:

$$\frac{\left(\Delta_j(\ell,n) - \Delta_j(m,n)\right) - \left(\Delta_i(\ell,n) - \Delta_i(m,n)\right)}{\Delta_j(m,n) - \Delta_i(m,n)} \geq \frac{\left(\Delta_k(\ell,n) - \Delta_k(m,n)\right) - \left(\Delta_j(\ell,n) - \Delta_j(m,n)\right)}{\Delta_k(m,n) - \Delta_j(m,n)} \tag{8.22}$$

The terms $\Delta_j(\ell,n) - \Delta_j(m,n)$ are changes in relative slopes, and hence are a measure of curvature of correlation curve j. Although it is harder to visualise (8.22) than (8.17), the condition states that this (weighted) "curvature" is allowed to change more from i to j than from j to k.

Summarising we find that the derived sufficient conditions for level, slope and curvature are in fact conditions on the level, slope and curvature of the correlation surface. It seems that, provided the term structure is properly ordered, the conditions do not state much more than that the correlation curves should be flatter and less curved for larger tenors, and steeper and more curved for shorter tenors.

8.3. Parametric correlation surfaces

Many articles have proposed various parametric correlation matrices, either to facilitate the empirical estimation of correlation matrices or the calibration to market data. One example of this we have seen already is the exponentially decaying correlation function which features in many articles as a simple, but somewhat realistic correlation function. Other examples are the

correlation parameterisations by Rebonato [2002], De Jong, Driessen and Pelsser [2004] and Alexander [2003]. The latter parameterisation is a rank three correlation matrix, defined by restricting the first three "eigenvectors" to be flat, linear and quadratic. We say "eigenvectors" because the constructed vectors are not chosen to be orthogonal, so that these vectors will not be the true eigenvectors. Since the resulting matrix is not of full rank, we will not consider it here. The first two are formulated from an economically plausible perspective, but are unfortunately not always guaranteed to be positive definite – this is only the case for the first two formulations of Rebonato [2002, Section 7.4.3].

The correlation matrices we consider in this section will be based on Green's matrices, which in the finance literature are probably better known as Schoenmakers-Coffey correlation matrices. In a continuous setting they already feature in Santa-Clara and Sornette [2001]. Schoenmakers and Coffey [2003] analysed the properties of its discrete analog and proposed various subparameterisations which they claim allow for a stable calibration to market swaption and caplet volatilities. A motivation for their matrix follows directly from the following construction. We will here take a slightly more general route than Schoenmakers and Coffey, leading to a more general correlation matrix. Let b_i, i = 1, …, N be an arbitrary sequence which is increasing in absolute value. We set $b_0 = b_1 = 1$ and $a_1 = 1$, $a_i = \sqrt{b_i^2 - b_{i-1}^2}$. Finally, let Z_i, i = 1, …, N be uncorrelated random variables, with unit variance. We now define:

$$Y_i = \mathrm{sgn}(b_i) \cdot \sum_{k=1}^{i} a_k Z_k \tag{8.23}$$

The covariance between Y_i and Y_j for for $i \le j$ is equal to:

$$\mathrm{Cov}(Y_i, Y_j) = \mathrm{sgn}(b_i b_j) \sum_{k=1}^{i} a_j^2 = \mathrm{sgn}(b_i b_j) b_i^2 \tag{8.24}$$

implying that their correlation is equal to:

$$\mathrm{Corr}(Y_i, Y_j) = \frac{b_i}{b_j} = \mathrm{sgn}(b_i b_j) \cdot \frac{\min(|b_i|, |b_j|)}{\max(|b_i|, |b_j|)} \tag{8.25}$$

It is easy to see that we obtain the same correlation structure if the Z_i's do not have unit variance, and also when each Y_i is premultiplied with a non-zero constant c_i. The difference with the approach of Schoenmakers and Coffey is that we here allow the sequence b_i to take negative values, whereas they only considered non-negative correlations. Furthermore, they restricted the sequence b_i/b_{i+1} to be strictly increasing, which has a nice consequence as we will see shortly. Even without these restrictions, the above construction always yields a valid correlation matrix.

We note that an N × N correlation matrix of the above form, say $\mathbf{R} = (\rho_{ij})_{i,j=1}^{N}$, can also be written in the following form:

$$\rho_{ij} = \prod_{k=i}^{j-1} \rho_{k,k+1} \tag{8.26}$$

i.e. we can view it as a parameterisation in terms of super- or alternatively subdiagonal elements. Schoenmakers and Coffey showed that the above parameterisation of the correlation matrix (with positive b_i's and with the restriction that $\rho_{i,i+1} = b_i/b_{i+1}$ is increasing) satisfies properties (i) – (iii) from section 8.1.2, properties that are commonly found in empirical correlation matrices of term structures. Sometimes it may be necessary to have a more flexible correlation structure at our disposal, in which case we can relax the restriction that b_i/b_{i+1} is to be increasing. This sacrifices

property (iii), the property that the correlation between two adjacent contracts or rates increases as the tenor increases. Properties (i) – (ii) will however still hold.

Returning to the level-slope-curvature pattern, Gantmacher and Kreĭn [1960] prove total positivity for certain special matrices. One of these matrices is a Green's matrix, in which category the above correlation matrix falls.

Theorem 8.4 – Total positivity of a Green's matrix
An $N \times N$ Green's matrix A with elements:

$$a_{ij} = \begin{cases} u_i v_j & i \geq j \\ u_j v_i & i \leq j \end{cases} \tag{8.27}$$

where all u_i and v_j are different from zero, is totally nonnegative if and only if all u_i and v_j have the same sign and:

$$\frac{v_1}{u_1} \leq \ldots \leq \frac{v_N}{u_N} \tag{8.28}$$

The rank of **A** is equal to the number of times where the inequality in (8.28) is strict, plus one. □

We note that in correlation form the concept of a Green's matrix is not more general than the extended Schoenmakers-Coffey matrix in (8.25) or (8.26). The corresponding correlation matrix **R** of the Green's matrix **A** from Theorem 8.4 has elements equal to:

$$r_{ij} = \begin{cases} \dfrac{u_i v_j}{\sqrt{u_i v_i u_j v_j}} = \dfrac{\sqrt{u_i v_j}}{\sqrt{u_j v_i}} & i \geq j \\ \dfrac{u_j v_i}{\sqrt{u_i v_i u_j v_j}} = \dfrac{\sqrt{u_j v_i}}{\sqrt{u_i v_j}} & i \leq j \end{cases} \tag{8.29}$$

Indeed, setting $b_i = v_i/u_i$ shows that a Green correlation matrix and the extended Schoenmakers-Coffey correlation matrix are equivalent. This observation combined with Theorem 8.4 leads to the following corollary.

Corollary 8.4 – Oscillatoriness of the Schoenmakers-Coffey matrix
The Schoenmakers-Coffey correlation matrix, and its more general formulation in (8.25) or (8.26), is oscillatory provided that all correlations on the superdiagonal are positive and smaller than 1. Hence, the matrix displays level, slope and curvature.

Proof:
The requirement that all correlations on the superdiagonal are positive amounts to requiring the sequence b_i to be strictly positive. The requirement that all entries on the superdiagonal are smaller than 1 implies the sequence b_i should be strictly increasing. Setting $b_i = v_i/u_i$ as mentioned, and substituting it into (8.28) yields:

$$b_1^2 \leq \ldots \leq b_N^2 \tag{8.30}$$

which is true due to the fact that the sequence b_i is strictly increasing. Furthermore, since the inequalities are strict, the correlation matrix is of full rank. The latter result still remains true if we allow the b_i's to take negative numbers, but still require that the sequence is strictly increasing in absolute value. Since all entries on the super- and subdiagonal are strictly positive, the matrices are oscillatory. By virtue of corollary 8.1 this implies that we have level, slope and curvature. □

Hence, if all correlations on the superdiagonal are positive and smaller than 1, the correlation matrix in (8.25) or (8.26) will display level, slope and curvature. We note that property (iii) clearly does not imply or affect level, slope or curvature for these matrices – the extended Schoenmakers-Coffey matrix displays level, slope and curvature regardless of whether property (iii) holds or not. A nice property of a Green's matrix is that its inverse is tridiagonal. Inversion of tridiagonal matrices requires only O(7N) arithmetic operations, and is therefore much more efficient than the $O(N^3/3)$ operations required for arbitrary matrices.

As a final point of interest we return to the claim of Lekkos [2000]. We remind the reader of equation (8.5), where zero yields were expressed as averages of continuously compounded α-forward rates:

$$R(t,T) = \frac{\alpha}{T-t}\left(f(t,t,t+\alpha) + \ldots + f(t,T-\alpha,T)\right) \tag{8.31}$$

In a numerical example Lekkos shows that if these forward rates are statistically independent, the correlation matrix of the zero yields displays level, slope and curvature. The way in which the Schoenmakers-Coffey matrix was constructed in equations (8.23) – (8.25) shows that if all forward rates in (8.31) are independent, the correlation matrix of $R(t,t+\alpha)$, …, $R(t,t+N\alpha)$ will be a Schoenmakers-Coffey correlation matrix, and as such will display level, slope and curvature. Lekkos' claim is therefore true. In fact, using the Schoenmakers-Coffey matrix for consecutive zero yields directly implies that all forward rates must be independent. Similarly, using the Schoenmakers-Coffey correlation matrix for changes in consecutive zero yields implies that the changes in consecutive forward rates are independent. As we have seen in section 8.1.2 forward rates and forward rate changes are far from independent, so that one should be aware of these implications. Schoenmakers and Coffey suggest using their correlation matrix and parameterised versions thereof as an instantaneous correlation matrix within the LIBOR market model, where the above considerations do not apply directly.

8.4. Level, slope and curvature beyond total positivity

In the previous two sections we turned to total positivity theory to provide us with sufficient conditions for level, slope and curvature. Obviously, this is only a partial answer to the question of what drives this phenomenon. In fact, if we look at the empirical correlation matrices from figures 8.1 and 8.2, the theory that we treated up till now is only able to explain level and slope for both figures, as both matrices contain only positive correlations, and the second power of both correlation matrices is oscillatory of order 2. The presence of curvature however still remains unexplained. Clearly there must be a more general theory that allows us to explain the presence of level, slope and curvature. Here we first take a brief look at the concept of sign regularity, which extends the notion of total positivity. However, we demonstrate that the only correlation matrices that were not already captured by the class of totally positive matrices are degenerate in some sense. Finally, we formulate a conjecture which we cannot prove, but which we suspect is true, based on an extensive simulation study. This conjecture directly relates the order present in correlation matrices to level and slope.

8.4.1. Sign regularity

In the literature the concept of total positivity has been extended to the notion of sign regularity. For a square $N \times N$ matrix $\mathbf{A}$ to be sign regular of order k, or SR_k, we require the existence of a sequence ε_1 through ε_k, all $\in \{1,-1\}$, such that for all $p \leq k$ and $\mathbf{i}, \mathbf{j} \in I_{p,N}$, such that:

$$\varepsilon_p \cdot \mathbf{A}_{[p]}(\mathbf{i},\mathbf{j}) \geq 0 \tag{8.32}$$

Analogous to strict total positivity, strict sign regularity can be defined. Sign regularity hence requires all determinants of a certain order to have the same sign, whereas total positivity required them to be positive. The concept of an oscillatory matrix can easily be extended using sign regularity. We can consider a square invertible matrix $\mathbf{A}$ with non-zero diagonal, super- and subdiagonal elements, that is SR. In this case $\mathbf{A}^2$ is oscillatory, and $\mathbf{A}^{2(N-1)}$ will be strictly totally positive, so that we can again apply theorem 8.2 to this matrix. This extension is however not useful for our application, as we will see in the following theorem.

Theorem 8.5 – The set of SR_3 correlation matrices is degenerate
There are no square $N \times N$ (for any $N \geq 3$) invertible correlation matrices, that are not TP_3, but SR_3. Furthermore, if the matrix is not of full rank, the correlation matrices that are SR_3 but not TP_3 are degenerate.

Proof:
The proof is actually very simple. If the matrix is to be SR_3, but not TP_3, there must be a $p \in \{1,2,3\}$ for which the following determinant is negative:

$$\mathbf{A}_{[p]}(\mathbf{i},\mathbf{j}) \leq 0 \tag{8.33}$$

for all $\mathbf{i},\mathbf{j} \in I_{p,N}$. In particular, (8.33) will also hold true when $\mathbf{i} = \mathbf{j}$, which means that the determinant of the correlation matrix of the contracts indexed by the vector $\mathbf{i}$ will not be positive. Since this submatrix is itself a correlation matrix, it must by assumption of the invertibility of the full matrix be invertible, and thus have a positive determinant. Hence, (8.33) cannot hold true, unless the matrix is not of full rank. If the matrix is not of full rank, SR_3, but not TP_3, one can easily show by considering the 3×3 case that all elements have to be in $\{-1,0,1\}$. □

This last theorem shows that the class of SR_3, but not TP_3, invertible correlation matrices is degenerate. As far as we know, no other classes of matrices are known which have the same sign change pattern as oscillatory matrices.

8.4.2. The relation between order, level and slope

As we already mentioned earlier, Alexander [2003] claims that "… the interpretation of eigenvectors as trend, tilt and curvature components is one of the stylised facts of all term structures, particularly when they are highly correlated". Based on an extensive simulation study we come up with a slightly different conjecture, which will follow shortly. The example in section 8.1.2 demonstrated that curvature is not always present, even though we have an ordered and highly correlated system. Similarly we can show that there are matrices for which no finite power is oscillatory of order 3, so that the theory from Section 8.3 cannot be used to prove the presence of slope and curvature for these correlation matrices. One such example follows.

Example 8.1 – Total positivity is not enough
Consider the following correlation matrix:

$$\mathbf{R} = \begin{pmatrix} 1 & 0.8396 & 0.8297 & 0.8204 \\ 0.8396 & 1 & 0.9695 & 0.901 \\ 0.8297 & 0.9695 & 1 & 0.9785 \\ 0.8204 & 0.901 & 0.9785 & 1 \end{pmatrix} \tag{8.34}$$

This correlation matrix itself is clearly not TP_2, consider for example its second order compound matrix $\mathbf{R}(\{1,2\},\{3,4\})$, i.e. the 2×2 matrix in the right-upper corner of $\mathbf{R}$. Its determinant is negative. From matrix theory we know that:

$$\lim_{k\to\infty} \frac{\mathbf{R}^k}{\lambda_1^k} = \mathbf{x}^1(\mathbf{x}^1)^T \tag{8.35}$$

where λ_1 is the largest eigenvalues and $\mathbf{x}^1$ is the corresponding eigenvector. Since higher powers of $\mathbf{R}$ are also not TP_2, and $\mathbf{R}^5$ is almost indistinguishable from the limiting matrix in (8.35), we can be sure that no finite power of $\mathbf{R}$ will be oscillatory of order 2.

Since the matrix in (8.34) satisfies properties (i)-(iii), as most empirical correlation matrices do, it is natural to formulate the following conjecture.

Conjecture 8.1 – Sufficiency of properties (i)-(iii) for level, slope and curvature
A quasi-correlation matrix $\mathbf{R}$ with strictly positive entries, that satisfies:

i) $\rho_{i,j+1} \leq \rho_{ij}$ for $j \geq i$, i.e. correlations decrease when we move away from the diagonal;

ii) $\rho_{i,j-1} \leq \rho_{ij}$ for $j \leq i$, same as i);

iii) $\rho_{i,i+j} \leq \rho_{i+1,i+j+1}$, i.e. the correlations increase when we move from northwest to southeast.

displays level and slope.

By a quasi-correlation matrix we mean a matrix that resembles a correlation matrix, i.e. has ones on the diagonal and off-diagonal elements that are smaller than or equal to 1 in absolute value, but is not necessarily positive definite. We claim that the empirically observed properties (i)-(iii) are sufficient, although still not necessary, for a quasi or proper correlation matrix to display level and slope. The fact that these properties are not necessary is clear from the Green's matrix – certain Green's matrices are still totally positive even though property (iii) is not satisfied, as we saw in the previous section.

We extensively tested this conjecture by simulating random correlation matrices, satisfying properties (i)-(iii). Although several methods exist to simulate random correlation matrices, we are not aware of one that allows the aforementioned properties to be satisfied. Firstly, we present the algorithm we used to simulate a random quasi-correlation matrix with positive entries, that in addition satisfies (i)-(iii). Note that a smoothing factor α is included in the algorithm that essentially ensures that two consecutive elements on a row are at most $100\alpha\%$ apart. Finally, note that each correlation is drawn from a uniform distribution – this is obviously an arbitrary choice.

1. $\rho_{ii} = 1$ for $1 \leq i \leq N$.
2. For $1 < j \leq N$ set LB_{1j} equal to $(\rho_{1,j-1}-\alpha)^+$ and UB_{1j} equal to $\rho_{1,j-1}$. Draw $\rho_{1j} \sim U[LB_{1j},UB_{1j}]$.
3. For $2 \leq i \leq N-1$ and $i < j$ set $LB_{ij} = \max\left((\rho_{i,j-1}-\alpha)^+, \rho_{i-1,j-1}\right)$ and $UB_{ij} = \rho_{i,j-1}$. If $LB_{ij} > UB_{ij}$, a valid matrix cannot be constructed, so that we have to restart our algorithm at step 1. Otherwise we draw $\rho_{ij} \sim U[LB_{ij},UB_{ij}]$.
4. Set $\rho_{ij} = \rho_{ji}$ for $i > j$.

Algorithm 8.1: Simulation of a quasi-correlation matrix with strictly positive entries, satisfying (i)-(iii)

Algorithm 8.1 can easily be adapted to generate a matrix that only satisfies (i)-(ii), by replacing $\rho_{i-1,j-1}$ in step 3 by $\rho_{i-1,j}$. Adapting this algorithm to yield a positive definite matrix can be achieved if we use the angles parameterisation of Rebonato and Jäckel [1999]. They show[52] that any correlation matrix $\mathbf{R} \in \mathbb{R}^{N \times N}$ can be written as $\mathbf{R}=\mathbf{B}\mathbf{B}^T$, where $\mathbf{B} \in \mathbb{R}^{N \times N}$ is lower triangular and has entries equal to $b_{11} = 1$, and:

$$b_{ij} = \cos\theta_{ij}\prod_{k=1}^{j-1}\sin\theta_{ik} \quad b_{ii} = \prod_{k=1}^{i-1}\sin\theta_{ik} \tag{8.36}$$

for $i > j$ and $i > 1$. Using this parameterisation it can be shown that the first row of the correlation matrix follows directly from the first column of the matrix with angles, i.e. $\rho_{1j} = \cos\theta_{j1}$ for $j > 1$. Hence, adapting step 2 is easy: we only have to solve for θ_{j1} in step 2. Adapting step 3 is slightly more involved. For $i < j$ we have:

$$\begin{aligned}\rho_{ij} &= \sum_{\ell=1}^{i} b_{i\ell}b_{j\ell} \\ &= \sum_{\ell=1}^{i-1}\cos\theta_{i\ell}\cos\theta_{j\ell}\prod_{k=1}^{\ell-1}\sin\theta_{ik}\sin\theta_{jk} + \cos\theta_{ji}\prod_{k=1}^{i-1}\sin\theta_{ik}\sin\theta_{jk}\end{aligned} \tag{8.37}$$

At entry (i,j) of the correlation matrix, we have already solved for the angles in columns 1 up to and including i-1, as well as angles θ_{jk} for $k < i$. The only new angle in (8.37) is thus θ_{ji}. Since we necessarily have $-1 \leq \cos\theta_{ji} \leq 1$, (8.37) places a lower and upper bound on ρ_{ij}. All we have to do is incorporate these additional restrictions into step 3 – this ensures that the new algorithm terminates with a positive definite correlation matrix. The algorithm hence becomes:

1. $\rho_{ii} = 1$ for $1 \leq i \leq N$.
2. For $1 < j \leq N$ set LB_{1j} equal to $(\rho_{1,j-1}-\alpha)^+$ and UB_{1j} equal to $\rho_{1,j-1}$. Draw $\rho_{1j} \sim U[LB_{1j},UB_{1j}]$. Solve θ_{j1} from $\rho_{1j} = \cos\theta_{j1}$.
3. For $2 \leq i \leq N-1$ and $i < j$ set $LB_{ij} = \max\left((\rho_{i,j-1}-\alpha)^+, \rho_{i-1,j-1}\right)$ and $UB_{ij} = \rho_{i,j-1}$. Incorporate lower and upper bound from (8.37) into LB_{ij} and UB_{ij}. If we then have $LB_{ij} > UB_{ij}$, a valid matrix cannot be constructed, so we have to restart our algorithm at step 1. Otherwise we draw ρ_{ij} from $U[LB_{ij},UB_{ij}]$ and solve for θ_{ji} from (8.37).
4. Set $\rho_{ij} = \rho_{ji}$ for $i > j$.

Algorithm 8.2: Simulation of a valid correlation matrix with strictly positive entries, satisfying (i)-(iii)

[52] Their article uses N(N-1) angles, but it can be shown that it suffices to use as many angles as correlations, i.e. ½N(N-1).

Using algorithms 8.1 and 8.2 we performed a large amount of simulations, for various sizes of matrices and values of α. In each simulation we kept track of the percentage of matrices without slope and/or curvature. The pattern was the same in each simulation, so that we here only display results for sizes equal to 3, 4 and 5 and α equal to 20%. The results can be found below:

	Properties (i)-(ii)		**Properties (i)-(iii)**	
Size	*No slope*	*No curvature*	*No slope*	*No curvature*
3	0%	0%	0%	0%
4	0.04%	19.1%	0%	23.05%
5	0.01%	27.98%	0%	43.81%

Table 1: Percentage of random quasi-correlation matrices w/o slope and/or curvature
Results based on 10,000 random matrices from algorithm 1, using $\alpha = 20\%$

	Properties (i)-(ii)		**Properties (i)-(iii)**	
Size	*No slope*	*No curvature*	*No slope*	*No curvature*
3	0%	0%	0%	0%
4	0.13%	14.91%	0%	18.31%
5	0.02%	23.1%	0%	35.38%

Table 2: Percentage of random proper correlation matrices w/o slope and/or curvature
Results based on 10,000 random matrices from algorithm 2, using $\alpha = 20\%$

As is clear from the tables, for both types of matrices properties (i)-(iii) seem to imply the presence of slope. Leaving out property (iii) causes some violations of the slope property, albeit in a very small number of cases. The results seem to indicate that our conjecture has some validity, although this is of course far from a formal proof.

8.5. Conclusions

In this article we analysed the so-called level, slope and curvature pattern one frequently observes when conducting a principal components analysis of term structure data. A partial description of the pattern is the number of sign changes of the first three factors, respectively zero, one and two. This characterisation enables us to formulate sufficient conditions for the occurrence of this pattern by means of the theory of total positivity. The conditions can be interpreted as conditions on the level, slope and curvature of the correlation surface. In essence, the conditions roughly state that if correlations are positive, the correlation curves are flatter and less curved for larger tenors, and steeper and more curved for shorter tenors, the observed pattern will occur. As a by-product of these theorems, we prove that if the correlation matrix is a Green's or Schoenmakers-Coffey matrix, level, slope and curvature is guaranteed. An unproven conjecture at the end of this chapter demonstrates that at least slope seems to be caused by two stylised empirical within term structures: the correlation between two contracts or rates decreases as a function of the difference in tenor between both contracts, and the correlation between two equidistant contracts or rates increases as the tenor of both contracts increases.

Furthermore we addressed Lekkos' critique, whose claim it is that the pattern purely arises due to the fact that zero yields are averages of forward rates, backed up by evidence that the pattern does not occur when considering forward rates. We have demonstrated that it could be the non-smoothness of the used curves that causes the absence of level, slope and curvature in his data.

Returning to the title of this chapter, we can conclude that the level, slope and curvature pattern is part fact, and part artefact. It is caused both by the order and positive correlations present in term structures (fact), as well as by the orthogonality of the factors and the smooth input we use to estimate our correlations (artefact).

Appendix 8.A - Proofs of various theorems

In this appendix we have included the proofs of theorem 8.2 and 8.3 for the sake of completeness. Before we present them we will require the following theorems from matrix algebra.

Theorem 8A.1 – Cauchy-Binet formula
Assume A = BC where A, B and C are N × N matrices. The Cauchy-Binet formula states that:

$$\mathbf{A}_{[p]}(\mathbf{i},\mathbf{j}) = \sum_{\mathbf{k}\in\mathbf{I}_{p,N}} \mathbf{B}_{[p]}(\mathbf{i},\mathbf{k})\mathbf{C}_{[p]}(\mathbf{k},\mathbf{j}) \tag{8A.1}$$

In other words, $\mathbf{A}_{[p]} = \mathbf{B}_{[p]}\mathbf{C}_{[p]}$, the operations of matrix multiplication and compound are interchangeable. □

The next theorem is useful when studying the eigensystem of compound matrices.

Theorem 8A.2 – Part of Kronecker's theorem
Let $\boldsymbol{\Sigma}$ be an invertible N × N matrix with eigenvalues λ_1 , ... , λ_N listed to their algebraic multiplicity. The matrix can be decomposed as $\boldsymbol{\Sigma} = \mathbf{X}\boldsymbol{\Lambda}\mathbf{X}^T$, where $\boldsymbol{\Lambda}$ contains the eigenvalues on its diagonal, and $\mathbf{X}$ contains the eigenvectors. In this case, $\boldsymbol{\Lambda}_{[p]}$ and $\mathbf{X}_{[p]}$ contain respectively the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}_{[p]}$. The $\binom{N}{p}$ eigenvalues of $\boldsymbol{\Sigma}_{[p]}$, listed to their algebraic multiplicity, are $\lambda_{i_1}\cdot\ldots\cdot\lambda_{i_p}$, for $\mathbf{i}\in \mathrm{I}_{p,N}$.

Proof:
When $\boldsymbol{\Sigma}$ is a general square matrix, the theorem concerning the eigenvalues is known as Kronecker's theorem. Its proof is easy and can be found in e.g. Karlin [1968], Ando [1987] or Pinkus [1995]. In case $\boldsymbol{\Sigma}$ is invertible everything is simplified even further. By virtue of the Cauchy-Binet formula we have that $\boldsymbol{\Sigma}\mathbf{X} = \boldsymbol{\Lambda}$ and $\mathbf{X}\mathbf{X}^T = \mathbf{I}$ implies that $\boldsymbol{\Sigma}_{[p]}\mathbf{X}_{[p]}^T = \boldsymbol{\Lambda}_{[p]}$ and $\mathbf{X}_{[p]}\mathbf{X}_{[p]}^T = \mathbf{I}_{[p]} = \mathbf{I}$. Indeed, this means that $\boldsymbol{\Lambda}_{[p]}$ and $\mathbf{X}_{[p]}$ respectively contain the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}_{[p]}$. □

We are now ready to prove theorem 8.2.

Theorem 8.2 – Sign-change pattern in STP_k matrices
Assume $\boldsymbol{\Sigma}$ is an N × N positive definite symmetric matrix (i.e. a valid covariance matrix) that is STP_k. Then we have $\lambda_1 > \lambda_2 > \ldots > \lambda_k > \lambda_{k+1} \geq \ldots \lambda_N > 0$, i.e. at least the first k eigenvalues are simple. Denoting the j^{th} eigenvector by $\mathbf{x}^j$, we have $S^-(\mathbf{x}^j) = S^+(\mathbf{x}^j) = j\text{-}1$, for j = 1, ..., k.

Proof:
We will first prove that at least the largest k eigenvalues are distinct. We know that the eigenvalues of $\boldsymbol{\Sigma}$ are strictly positive, so we can write $\lambda_1 \geq \ldots \geq \lambda_N > 0$. Since $\boldsymbol{\Sigma}$ is STP_k, we can apply Perron's theorem to find that $\lambda_1 > \lambda_2 \geq \ldots \geq \lambda_N$. Assume we have proven the statement for the largest j-1 eigenvalues, j-1 < k, this means we know that $\lambda_1 > \lambda_2 > \ldots > \lambda_{j-1} > \lambda_j \geq \ldots \lambda_N > 0$. Since $\boldsymbol{\Sigma}$ is STP_j, each element of $\boldsymbol{\Sigma}_{[j]}$ is strictly positive. Perron's theorem (theorem 8.1) now states that $\boldsymbol{\Sigma}_{[j]}$ has a unique largest eigenvalue. From Kronecker's theorem (theorem 8A.2) we deduce that this eigenvalue is $\lambda_1\cdot\ldots\cdot\lambda_j$. Hence:

$$\lambda_1 \cdot \ldots \cdot \lambda_j > \lambda_1 \cdot \ldots \lambda_{j-1}\lambda_{j+1} \Leftrightarrow \lambda_j > \lambda_{j+1} \tag{8A.2}$$

i.e., since we know that all eigenvalues are positive, this implies the property also holds for j. For j equal to 1 the property has already been proven, so by induction it follows that the largest k eigenvalues are distinct.

Now we turn to the number of sign changes of the eigenvectors associated with the largest k eigenvalues. Since $\mathbf{\Sigma}$ is positive definite, orthogonal eigenvectors $\mathbf{x}^1$ through $\mathbf{x}^N$ exist. Eigenvector $\mathbf{x}^i$ is associated with eigenvalue λ_i. We write $\mathbf{\Sigma} = \mathbf{X\Lambda\Lambda}^T$, where $\mathbf{X}$ is the matrix containing the eigenvectors as its columns, and $\mathbf{\Lambda}$ contains the eigenvalues on its diagonal. Let us first assume that $S^+(\mathbf{x}^j) \geq j$, for $j \leq k$. Then there are $1 \leq i_0 < \ldots < i_j \leq N$ and an $\in = \pm 1$ such that:

$$\varepsilon(-1)^\ell \mathbf{x}^j_{i_\ell} \geq 0 \text{ for } \ell = 0, \ldots, j \tag{8A.3}$$

Now set $\mathbf{x}^0 = \mathbf{x}^j$, and we extend $\mathbf{X}$ as $\overline{\mathbf{X}}$ to include $\mathbf{x}_0$ as its first column. Obviously we must have that $\overline{\mathbf{X}}\begin{pmatrix} i_0,\ldots,i_j \\ 0,\ldots,j \end{pmatrix} = 0$. On the other hand, we can expand this determinant on the first column:

$$\overline{\mathbf{X}}\begin{pmatrix} i_0,\ldots,i_j \\ 0,\ldots,j \end{pmatrix} = \sum\nolimits_{\ell=0}^{j} (-1)^\ell x^0_{i_\ell} \mathbf{X}\begin{pmatrix} i_0,\ldots,i_{\ell-1},i_{\ell+1},\ldots,i_j \\ 1,\ldots,j \end{pmatrix} \tag{8A.4}$$

From (8A.3) we have that the first part of the sum is of one sign. The determinant on the right-hand side is an element of the first column of $\mathbf{X}_{[j]}$. Via theorem 8.2 we know that this column contains the eigenvectors of $\mathbf{\Sigma}_{[j]}$, which contains strictly positive elements by the assumption that $\mathbf{\Sigma}$ is STP_j for $j \leq k$. Perron's theorem then implies that this first eigenvector is either strictly positive or negative, i.e. every element on the right-hand side is of one sign. If the left-hand side is therefore going to be zero, we must have that $\mathbf{x}^0_{i_\ell} = \mathbf{x}^j_{i_\ell}$ is zero for $\ell = 0, \ldots, j$. Note that the determinant on the right-hand side is the determinant of the submatrix formed by only using the rows i_0 through i_j, excluding i_ℓ, and columns 1 through j. If $\mathbf{x}^0_{i_\ell} = \mathbf{x}^j_{i_\ell}$ is zero for $\ell = 0, \ldots, j$, we would be taking the determinant of a matrix which contains a column filled with zeroes. Necessarily this determinant would be equal to zero, which is a contradiction as we just saw. We can therefore conclude that $S^+(\mathbf{x}^j) \leq j\text{-}1$, for $j \leq k$.

The second part of the proof is very similar. From the definitions of S^+ and S^- it is clear that we must have $S^-(\mathbf{x}^j) \leq S^+(\mathbf{x}^j) \leq j\text{-}1$. Let us assume that $S^-(\mathbf{x}^j) = p \leq j\text{-}2$ for $j \leq k$. This implies that there exist $1 \leq i_0 < \ldots < i_p \leq N$ and an $\in = \pm 1$ such that:

$$\varepsilon(-1)^\ell \mathbf{x}^j_{i_\ell} > 0 \text{ for } \ell = 0, \ldots, p \tag{8A.5}$$

Again, we set $\mathbf{x}^0 = \mathbf{x}^j$. Then obviously the determinant $\overline{\mathbf{X}}\begin{pmatrix} i_0,\ldots,i_p \\ 0,\ldots,p \end{pmatrix} = 0$. As before, we find:

$$\overline{\mathbf{X}}\begin{pmatrix} i_0,\ldots,i_p \\ 0,\ldots,p \end{pmatrix} = \sum\nolimits_{\ell=0}^{p} (-1)^\ell x^0_{i_\ell} \mathbf{X}\begin{pmatrix} i_0,\ldots,i_{\ell-1},i_{\ell+1},\ldots,i_p \\ 1,\ldots,p \end{pmatrix} \tag{8A.6}$$

From before, we know that the determinants on the right-hand side can be chosen to be strictly positive. Together with (8A.5) this implies that the right-hand side is positive, which is a

contradiction. Therefore $S^{-}(\mathbf{x}^j)$ must be larger than j-2, and we have proven $S^{-}(\mathbf{x}^j) = S^{+}(\mathbf{x}^j) = j-1$ for $j \leq k$. □

To prove theorem 8.3, we will also require the Hadamard inequality for positive semi-definite matrices. In Gantmacher and Kreĭn [1937] this was originally proven for TP matrices. In Karlin [1968] we can find the following formulation.

Theorem 8A.3 – Hadamard inequality
For $\mathbf{\Sigma}$ an $N \times N$ positive semi-definite matrix we have:

$$\mathbf{\Sigma}\begin{pmatrix}1,...,N\\1,...,N\end{pmatrix} \leq \mathbf{\Sigma}\begin{pmatrix}1,...,k\\1,...,k\end{pmatrix} \cdot \mathbf{\Sigma}\begin{pmatrix}k+1,...,N\\k+1,...,N\end{pmatrix} \tag{8A.7}$$

for k = 1, …, N-1. □

Now we are ready to prove our theorem about oscillation matrices of order k.

Theorem 8.3 – Oscillation matrix of order k
Akin to the concept of an oscillation matrix, we define an oscillation matrix of order k. An $N \times N$ matrix $\mathbf{A}$ is oscillatory of order k if:
1. $\mathbf{A}$ is TP_k;
2. $\mathbf{A}$ is non-singular;
3. For all i = 1, ..., N-1 we have $a_{i,i+1} > 0$ and $a_{i,i+1} > 0$.

For oscillatory matrices of the order k, we have that $\mathbf{A}^{N-1}$ is STP_k.

Proof:
First we prove that for all matrices satisfying i), ii) and iii), we have $\mathbf{A}_{[p]}(\mathbf{i},\mathbf{j}) > 0$, for $p \leq k$ and for all $\mathbf{i}$ and $\mathbf{j} \in I_{p,n}$ satisfying:

$$\left|i_\ell - j_\ell\right| \leq 1 \text{ and } \max(i_\ell, j_\ell) < \min(i_{\ell+1}, j_{\ell+1}) \qquad \ell = 1, \ldots, p \tag{8A.8}$$

where $i_{p+1} = j_{p+1} = \infty$. Gantmacher and Kreĭn dubbed these minors quasi-principal minors. We will prove this by induction on p. For p = 1 all quasi-principal minors are all diagonal and super- and subdiagonal elements. The latter are positive by assumption. Furthermore, from the assumption of non-singularity and the Hadamard inequality (theorem A.3) for totally positive matrices, we have:

$$0 < \det \mathbf{A} \leq \prod_{i=1}^{N} a_{ii} \tag{8A.9}$$

i.e. all diagonal elements are non-zero, and from the assumption of total positivity are hence positive. Now assume that the assertion holds for p-1, but that it does not hold for $p \leq k$. Hence, all quasi-principal minors of order smaller than p are positive, but there are $\mathbf{i}$ and $\mathbf{j} \in I_{p,N}$ satisfying (8A.8) such that:

$$\mathbf{A}\begin{pmatrix}i_1,...,i_p\\j_1,...,j_p\end{pmatrix} = 0 \tag{8A.10}$$

From the induction assumption we have that:

$$\mathbf{A}\begin{pmatrix} i_1,\ldots,i_{p-1} \\ j_1,\ldots,j_{p-1} \end{pmatrix} \cdot \mathbf{A}\begin{pmatrix} i_2,\ldots,i_p \\ j_2,\ldots,j_p \end{pmatrix} > 0 \tag{8A.11}$$

Consider the matrix (a_{ij}) with elements $i = i_1, i_1+1$ through i_p, and $j = j_1, j_1+1$ through j_p. These last two results and corollary 9.2 from Karlin [1968] imply that the rank of this matrix is p-1. Now set $h = \max(i_1,k_1)$. Then it follows from (8A.8) that $h+p-1 \le \min(i_p,j_p)$. This implies that:

$$\mathbf{A}\begin{pmatrix} h, h+1,\ldots,h+p-1 \\ h, h+1,\ldots,h+p-1 \end{pmatrix} \tag{8A.12}$$

is a principal minor of order p of **A**, and hence is equal to zero. Since the matrix **A** is positive definite, so is the square submatrix (a_{ij}) with i,j = h, …, h+p-1. Therefore (8A.10) cannot hold, and since **A** is TP_k, we must have $A(\mathbf{i},\mathbf{j}) > 0$.
Now we will prove that $B=A^{N-1}$ is STP_k. Indeed, for any $\mathbf{i}, \mathbf{j} \in I_{p,N}$, where $p \le k$ we have, from the Cauchy-Binet formula:

$$\mathbf{B}(\mathbf{i},\mathbf{j}) = \sum_{\boldsymbol{\alpha}^{(1)},\ldots,\boldsymbol{\alpha}^{(N-2)}} \prod_{\ell=1}^{N-1} \mathbf{A}\left(\boldsymbol{\alpha}^{(\ell-1)}, \boldsymbol{\alpha}^{(\ell)}\right) \tag{8A.13}$$

where each $\boldsymbol{\alpha}^{(\ell)} \in I_{p,N}$ and we set $\boldsymbol{\alpha}^{(0)} = \mathbf{i}$ and $\boldsymbol{\alpha}^{(N-1)} = \mathbf{j}$. Since **A** is TP_k, $\mathbf{B}(\mathbf{i},\mathbf{j})$ is a sum of nonnegative determinants, and hence is itself nonnegative. Following Gantmacher and Kreĭn we can now construct a series of $\boldsymbol{\alpha}^{(\ell)}$ such that each determinant in (8A.13) is a quasi-principal minor of order smaller than or equal to k, and hence by the previous result is strictly positive. The construction works as follows.

1. Set $\boldsymbol{\alpha}^{(0)} = \mathbf{i}$ and s = 1.
2. Compare **j** with $\boldsymbol{\alpha}^{(s-1)}$. Writing down $\boldsymbol{\alpha}^{(s-1)}$ in lexicographical order, we see it can be divided into consecutive parts, each of which contains elements $\alpha_\ell^{(s-1)}$ that are either smaller than, larger than, or equal to j_ℓ. We will refer to these parts as the positive, negative and zero parts.
3. We now construct $\boldsymbol{\alpha}^{(s)}$ from $\boldsymbol{\alpha}^{(s-1)}$ by adding 1 to the last s elements in each positive part and subtracting 1 from the first s indices in each negative part. Each zero part is left unchanged. If any part has less than s elements, we alter all elements.
4. Repeat 2 and 3 for s = 2, …, N-1.

As an example, consider the vectors (2, 3, 5, 8, 9) and (2, 4, 6, 7, 9). Let the first vector play the role of $\boldsymbol{\alpha}^{(0)}$ and the second the role of **j**. We group $\boldsymbol{\alpha}^{(0)}$ as ((2), (3,5), (8), (9)), and see that it consists of a zero, a positive, a negative, and a zero part, in that order. We are now ready to construct $\boldsymbol{\alpha}^{(1)}$, and find that it is equal to (2, 3, 6, 7, 9).
We will now prove that each pair $\boldsymbol{\alpha}^{(s-1)}$, $\boldsymbol{\alpha}^{(s)}$ for s = 1, …, N-1 satisfies (8A.8). The first part of (8A.8) is obviously satisfied, as each element of $\boldsymbol{\alpha}^{(s)}$ differs by at most 1 from the corresponding entry in $\boldsymbol{\alpha}^{(s-1)}$. That the constructed $\boldsymbol{\alpha}^{(s)} \in I_{p,N}$, i.e. that $1 \le \alpha_1^{(s)} < \ldots < \alpha_p^{(s)} \le N$ is easy to check; we will omit this here. We will now prove that $\alpha_\ell^{(s)} < \alpha_{\ell+1}^{(s-1)}$ for each ℓ. If both entries come from the same part of $\boldsymbol{\alpha}^{(s-1)}$, this obviously holds. Hence, we need to check that this condition holds at the

boundary of two parts. In fact we only have to check those cases where $\alpha_{\ell+1}^{(s-1)}$ belongs to a negative part. This means:

$$\alpha_{\ell+1}^{(s)} = \alpha_{\ell+1}^{(s-1)} - 1 \geq j_{\ell+1} \tag{8A.14}$$

Since $\boldsymbol{\alpha}^{(s)} \in I_{p,N}$, we immediately have: $\alpha_{\ell}^{(s)} < \alpha_{\ell+1}^{(s)} < \alpha_{\ell}^{(s-1)}$. Hence, all $A(\boldsymbol{\alpha}^{(s-1)}, \boldsymbol{\alpha}^{(s)})$ we have constructed for s = 1, …, N-1, are quasi-principal minors. We will now prove that $\boldsymbol{\alpha}^{(N-1)} = \mathbf{j}$. Note that $\ell \leq i_\ell$ and $j_\ell \leq N\text{-}p+\ell$ are true. These inequalities imply:

$$\left| i_\ell - j_\ell \right| \leq N - p \tag{8A.15}$$

From the construction we followed, it is clear that for any $i_\ell \neq j_\ell$, the ℓ^{th} element of $\boldsymbol{\alpha}^{(0)}, \boldsymbol{\alpha}^{(1)}, \ldots$ will have the value i_ℓ up to a certain point, and will then approach j_ℓ with increments of 1. The convergence towards j_ℓ will start when s = p, at the latest. Due to (8A.15) we will have certainly achieved $\boldsymbol{\alpha}^{(N-1)} = \mathbf{j}$, as we have then performed N-1 steps of the algorithm. This implies that we can construct an element of the summation in (8A.13) where each element is a quasi-principal minor of order $p \leq k$. By the previous result, each of these minors is strictly positive, so that indeed the matrix $B=A^{N-1}$ is STP_k. □

Nederlandse samenvatting (Summary in Dutch)

Opties en derivaten ontstonden lang voordat de Nobelprijs winnende artikels van Black en Scholes [1973] en Merton [1973] het gebied van financial engineering tot stand brachten. Het staat vast dat optiecontracten al ten tijde van de Babyloniërs en de Grieken bestonden, hoewel de eerste georganiseerde markt voor futures pas tot stand kwam in de zeventiende eeuw in Japan, waar Japanse feodale heren hun toekomstige rijstopbrengst verkochten in een markt genaamd *cho-ai-mai*, letterlijk "rijsthandel op boek". Rond dezelfde tijd begon men in Nederland tijdens de tulpenmanie op de Amsterdamse beurs met de handel in futures contracten en opties op tulpenbollen. De eerste officiële futures- en optiebeurs, de Chicago Board of Trade, opende in 1848, maar het was pas na de opening van de Chicago Board Options Exchange in 1973, een maand voor de publicatie van het artikel van Black en Scholes, dat de mondiale optiehandel echt op gang kwam.

Voor de ontdekking van de Black-Scholes formule hadden investeerders en speculanten heuristische methoden en hun projecties van de toekomst nodig om tot een prijs voor een derivaat te komen. Er waren al pogingen ondernomen om tot een optiewaarderingsformule te komen, beginnend met Bachelier [1900], maar alle pogingen misten het cruciale inzicht van Black, Scholes en Merton dat, onder bepaalde omstandigheden, het risico van een optie volledig afgedekt kan worden door dynamisch te investeren in de onderliggende waarde van die optie. Als men aanneemt dat er geen arbitrage mogelijkheden bestaan in financiële markten, moet de prijs van een optie derhalve gelijk zijn aan de waarde van de investeringsstrategie waarmee men deze optie repliceert. Deze ontdekking, tezamen met de komst van de zakrekenmachine en, later, de personal computer, heeft ervoor gezorgd dat de derivatenmarkt de grote industrie is geworden die het vandaag de dag is.

Door middel van replicatie konden meer exotische structuren ook worden gewaardeerd. Een van de noodzakelijke eisen waaraan een optiemodel moet voldoen, zodanig dat de prijs die eruit rolt voor een structuur zinnig is, is dat binnen dit model de prijzen van simpelere producten waarin actief gehandeld wordt, zoals bijvoorbeeld forward contracten en Europese opties, overeenkomen met hun marktprijs. Gaandeweg werd het duidelijk dat dit in het Black-Scholes model niet het geval was, en dat de aanname dat de onderliggende beschreven kan worden door een geometrische Brownse beweging met een constante, mogelijk tijdsafhankelijke drift en volatiliteit, niet adequaat was. Als deze aanname waar zou zijn, zou het oplossen van de volatiliteit van de onderliggende uit de marktprijs van een optie niet van de uitoefenprijs of looptijd van de optie af moeten hangen. Dit is niet het geval. Veel van het onderzoek binnen de financiële wiskunde heeft zich derhalve geconcentreerd op alternatieve stochastische processen voor de onderliggende waarde, zodanig dat verhandelde opties beter, zo niet perfect, teruggeprijsd worden. Om een exotische optie te waarden, moet men dan:

1. Een model kiezen dat zowel economisch plausibel als analytisch makkelijk is;
2. Het model calibreren aan de prijzen van verhandelde vanilla opties;
3. De exotische optie waarderen met het gecalibreerde model, gebruikmakend van toereikende numerieke technieken.

Dit proefschrift houdt zich voornamelijk bezig met de tweede en derde stap in dit proces. In de praktijk heeft men behoefte aan nauwkeurige prijzen en gevoeligheden, die bovendien snel berekend kunnen worden. Aangezien de financiële modellen en optiecontracten steeds complexer worden, moeten efficiënte methoden ontwikkeld worden om met zulke modellen om te gaan. Met uitzondering van één hoofdstuk gaat dit proefschrift over het efficiënt waarden van opties in zogenaamde affiene modellen, gebruikmakend van methoden variërend van analytische benaderingen tot Monte Carlo methoden en numerieke integratie.

De analytisch makkelijke klasse van affiene modellen, waarin het Black-Scholes model, evenals veel stochastische volatiliteit modellen waaronder Heston's [1993] model, en exponentiële Lévy modellen, wordt beschreven in Hoofdstuk 2. De mogelijkheid van een model om snel en nauwkeurig te kalibreren aan prijzen van vanilla opties draagt bij aan zijn praktische relevantie. Dit is inderdaad een van de aansprekende eigenschappen van affiene modellen. Aangezien voor veel affiene modellen een gesloten uitdrukking bekend is voor de karakteristieke functie, kunnen Europese opties gewaardeerd worden middels Fourier inversie.

Hoofdstuk 3 behandelt een probleem dat optreedt bij het evalueren van de karakteristieke functie van veel affiene modellen, waaronder het stochastische volatiliteitsmodel van Heston [1993]. Om deze karakteristieke functie uit te rekenen dient men een complexe logaritme uit te rekenen, welke niet eenduidig gedefinieerd is. Als we de hoofdwaarde van de logaritme nemen, zoals in veel software pakketten wordt gedaan, kan de karakteristieke functie discontinu worden, hetgeen leidt tot compleet verkeerde optieprijzen als men de optie middels Fourier inversie waardeert. In dit hoofdstuk bewijzen we dat er een formulering van de karakteristieke functie is waarin de hoofdwaarde de enige juiste is. Soortgelijke problemen worden ook gevonden in andere modellen.

Hoofdstuk 4 houdt zich bezig met de Fourier inversie techniek die gebruikt wordt om Europese opties te waarderen binnen de klasse van affiene modellen. In Hoofdstuk 2 hebben we al laten zien dat het contour van integratie in het complexe vlak leidt tot verschillende representaties van de inverse Fourier integraal. In dit hoofdstuk presenteren we het optimale contour van de Fourier integraal, waarbij we numerieke zaken zoals cancellation en moment explosie van de karakteristieke functie in ogenschouw nemen. Dit leidt tot een snel en robuust waarderingsalgoritme voor vrijwel alle uitoefenprijzen en looptijden, hetgeen in enkele numerieke voorbeelden gedemonstreerd wordt. Soortgelijke problemen in andere modellen worden eveneens bekeken.

De volgende drie hoofdstukken houden zich voornamelijk bezig met het echte waarderen van exotische opties, de laatste van de drie stappen. In Hoofdstuk 5 behandelen we het waarderen van opties met vroege uitoefenmogelijkheden, hoewel het gepresenteerde algoritme ook gebruikt kan worden voor bepaalde padafhankelijke opties. De methode is gebaseerd op een kwadratuurtechniek en hangt sterk af van Fourier transformaties. Het belangrijkste idee is om de bekende risicio-neutrale waarderingsformule te schrijven als een convolutie. Deze convolutie kan efficiënt uitgerekend worden middels de Fast Fourier Transform (FFT). Deze nieuwe waarderingstechniek, die we de Convolution method noemen, afgekort CONV, kan toegepast worden op een groot aantal optie. Het enige dat men van het model moet weten is de karakteristieke functie. Zodoende is de methode van toepassing binnen veel affiene modellen.

Hoofdstuk 6 focust op de simulatie van square root processen, in het bijzonder binnen het stochastische volatiliteit model van Heston. Indien men gebruik maakt van een Euler discretisatie om een square root proces met mean reversion te simuleren, stuit men op het probleem dat hoewel het proces zelf nonnegatief is, de discretisatie dit niet is. Hoewel een exact en efficiënt simulatie algoritme bestaat voor dit proces, is dit momenteel niet het geval voor het Heston model, waar de variantie gemodelleerd wordt als een square root proces. Dientengevolge moet men bij gebruik van een Euler discretisatie stilstaan bij hoe men negatieve waarden van het proces aanpast. Onze bijdrage is drievoudig. Ten eerste verenigen we alle Euler "aanpassingen" in een algemeen raamwerk. Ten tweede introduceren we het nieuwe full truncation schema, dat

als doel heeft de positieve afwijking te verkleinen die gevonden wordt bij het waarderen van Europese opties. Ten derde, en tenslotte, vergelijken we alle Euler aanpassingen met andere recente schema's. Het blijkt erg belangrijk te zijn hoe men de negatieve waarden van het proces aanpast.

Hoofdstuk 7 is anders dan de vorige hoofdstukken, daar de focus puur en alleen op het Black-Scholes model ligt. Sommige methoden uit dit hoofdstuk kunnen hoe dan ook toegepast worden op de hele klasse van affiene modellen, zoals aangetoond in Lord [2006]. Dit hoofdstuk focust op het waarderen van Europese Aziatische opties, hoewel alle bekeken methoden eenvoudig uit te breiden zijn naar het geval van een Aziatische basket optie. We behandelen een aantal verschillende methoden voor het waarderen van een Aziatische optie, en dragen aan ze allemaal bij. Ten eerste laten we de link zien tussen verschillende PDE methoden. Ten tweede, leiden we een gesloten uitdrukking af voor de ondergrens van Curran en Rogers en Shi, voor het geval van meerdere onderliggende waarden. Ten derde scherpen we Thompson's [1999a,b] bovengrens aan, zodanig dat deze scherper is dan alle bekende bovengrenzen. Tenslotte combineren we de traditionele moment matching approximaties met Curran's conditioneringsmethode. Men kan bewijzen dat een benadering uit de resulterende klasse van partially exact and bounded approximations (gedeeltelijk exacte en begrensde benaderingen tussen een scherpe onder- en bovengrens ligt. In numerieke voorbeelden demonstreren we dat deze benadering beter presteert dan alle op het moment van schrijven bekende grenzen en benaderingen.

Tenslotte staan we in Hoofdstuk 8 bij een compleet verschillend onderwerp stil, namelijk de eigenschappen van correlatie matrices van termijnstructuur data. Deze kunnen als invoer gebruikt worden binnen optie waarderingsmodellen voor termijnstructuren, bijvoorbeeld rente termijnstructuren. De eerste drie factoren uit een principale componenten analyse van termijnstructuur data worden in de literatuur in de regel geïnterpreteerd als zijnde level, slope en curvature (niveau, helling en kromheid) van de termijnstructuur. Gebruikmakend van milde generalisaties van theorema's uit de theorie over total positivity, presenteren we voldoende voorwaarden waaronder men inderdaad level, slope en curvature terugziet in een principale componentenanalyse. Het blijkt dat deze voorwaarden gerelateerd kunnen worden aan voorwaarden op het niveu, de helling en de kromheid van de correlatiecurve. Er wordt bewezen dat de Schoenmakers-Coffey correlatie matrix ook level, slope en curvature tentoonspreidt. Tenslotte formuleren en bekrachtigen we ons vermoeden dat de hiërarchie binnen termijnstructuren de factor slope veroorzaakt.

Bibliography

ABATE, J. AND W. WHITT (1992). "The Fourier-series method for inverting transforms of probability distributions", *Queueing systems*, vol. 10, no. 1-2, pp. 5-87.

AÏT-SAHALIA, Y. AND J. YU (2006). "Saddlepoint approximations for continuous-time Markov processes", *Journal of Econometrics*, vol. 134, no. 2, pp. 507-551.

ALBIN, J.M.P., ASTRUP JENSEN, B., MUSZTA, A. AND M. RICHTER (2005). "A note on the Lipschitz condition for Euler approximation of stochastic differential equations", working paper, Chalmers University of Technology and Copenhagen Business School.

ALBRECHER, H., MAYER, P. SCHOUTENS, W. AND J. TISTAERT (2007). "The little Heston trap", *Wilmott Magazine*, January, pp. 83-92.

ALEXANDER, C. (2003). "Common correlation and calibrating LIBOR", *Wilmott Magazine*, March 2003, pp. 68-78.

ALEXANDER, C. AND D. LVOV (2003). "Statistical properties of forward LIBOR rates", working paper, University of Reading, available at: http://www.icmacentre.ac.uk/pdf/discussion/DP2003-03.pdf.

ALFONSI, A. (2005). "On the discretization schemes for the CIR (and Bessel squared) processes", *Monte Carlo Methods and Applications*, vol. 11, no. 4, pp. 355-384.

ALMENDRAL, A. AND C.W. OOSTERLEE (2007A). "Accurate evaluation of European and American options under the CGMY process", *SIAM Journal on Scientific Computing*, vol. 29, no. 1, pp. 93-117.

ALMENDRAL, A. AND C.W. OOSTERLEE (2007B). "On American options under the Variance Gamma process", *Applied Mathematical Finance*, vol. 14, no. 2, pp. 131-152.

ANDO, T. (1987). "Totally positive matrices", *Linear Algebra and Its Applications*, vol. 90, pp. 165-219.

ANDERSEN, L.B.G. (2008). "Simple and efficient simulation of the Heston stochastic volatility model", *Journal of Computational Finance*, vol. 11, no. 3, pp. 1-42.

ANDERSEN, L.B.G. AND J. ANDREASEN (2000). "Jump-diffusion processes: volatility smile fitting and numerical methods for option pricing", *Review of derivatives research*, vol. 4, no. 3, pp. 231-262.

ANDERSEN, L.B.G. AND J. ANDREASEN (2002). "Volatile volatilities", *Risk*, vol. 15, no. 12, December 2002, pp. 163-168.

ANDERSEN, L.B.G. AND R. BROTHERTON-RATCLIFFE (2005). "Extended LIBOR market models with stochastic volatility", *Journal of Computational Finance*, vol. 9, no. 1, pp. 1-40.

ANDERSEN, L.B.G. AND V.V. PITERBARG (2007). "Moment explosions in stochastic volatility models", *Finance and Stochastics*, vol. 11, no. 1, pp. 29-50.

ANDREASEN, J. (1998). "The pricing of discretely sampled Asian and lookback options: a change of numeraire approach", *Journal of Computational Finance*, vol. 1, no. 3, pp. 15-36.

ANDRICOPOULOS, A.D., WIDDICKS, M., DUCK, P.W. AND D.P. NEWTON (2003). "Universal option valuation using quadrature", *Journal of Financial Economics*, vol. 67, no. 3, pp. 447-471.

BACHELIER, L. (1900). *Théorie de la spéculation*, Gauthiers-Villars.

BAKSHI, G., CAO, C. AND Z. CHEN (1997). "Empirical performance of alternative option pricing models", *Journal of Finance*, vol. 52, pp. 2003-2049.

BATES, D.S. (1996). "Jumps and stochastic volatility: exchange rate processes implicit in Deutsche Mark options", *Review of Financial Studies*, vol.9, no.1., pp. 69-107.

BENHAMOU, E. (2002). "Fast Fourier Transform for discrete Asian options", *Journal of Computational Finance*, vol. 6., no. 1, pp. 49-68.

BERKAOUI, A., BOSSY, M. AND A. DIOP (2008). "Euler schemes for SDEs with non-Lipschitz diffusion coefficient: strong convergence", *ESAIM Probability and Statistics*, vol. 12, no. 1, pp. 1-11.

BERNSTEIN, P.L. (1996). *Against the Gods: The remarkable story of risk*, John Wiley and Sons.

BLACK, F. AND M. SCHOLES (1973). "The pricing of options and corporate liabilities", *Journal of Political Economy*, vol. 81, no. 3, pp. 637-654.

BOSSY, M. AND A. DIOP (2004). "An efficient discretization scheme for one dimensional SDEs with a diffusion coefficient function of the form $|x|^{\alpha}$, $\alpha \in [1/2, 1)$", INRIA working paper no. 5396.

BOYARCHENKO, S.I. AND S.Z. LEVENDORSKIĬ (2002). *Non-Gaussian Merton-Black-Scholes theory*, Vol. 9 of Advanced Series on Statistical Science & Applied Probability, World Scientific Publishing Company.

BRIGO, D. AND F. MERCURIO (2001). *Interest Rate Models: Theory and Practice*, Springer.

BROADIE, M. AND Ö. KAYA (2004). "Exact simulation of option greeks under stochastic volatility and jump diffusion models", in R.G. Ingalls, M.D. Rossetti, J.S. Smith and B.A. Peters (eds.), *Proceedings of the 2004 Winter Simulation Conference*.

BROADIE, M. AND Ö. KAYA (2006). "Exact simulation of stochastic volatility and other affine jump diffusion processes", *Operations Research*, vol. 54, no. 2, pp. 217-231.

BROADIE, M. AND Y. YAMAMOTO (2005). "A double-exponential Fast Gauss transform algorithm for pricing discrete path-dependent options", *Operations Research*, vol. 53, no. 5, pp. 764-779.

CARR, P., GEMAN, H. MADAN, D.B. AND M. YOR (2002). "The fine structure of asset returns: an empirical investigation", *Journal of Business*, vol. 75, no. 2, pp. 305-332.

CARR, P. AND D.B. MADAN (1999). "Option valuation using the Fast Fourier Transform", *Journal of Computational Finance*, vol. 2, no. 4, pp. 61-73.

CARVERHILL, A. AND L. CLEWLOW (1990). "Flexible convolution: valuing average rate (Asian) options", *Risk*, vol. 4, no. 3, pp. 25-29.

CHANG, C-C., CHUNG, S-L. AND R.C. STAPLETON (2007). "Richardson extrapolation techniques for the pricing of American-style options", *Journal of Futures Markets*, vol. 27, no. 8, pp. 791-817.

CHENG, P. AND O. SCAILLET (2007). "Linear-quadratic jump-diffusion modelling", *Mathematical Finance*, vol. 17, no. 4, pp. 575-598.

CHOUDHURY, G.L. AND W. WHITT (1997). "Probabilistic scaling for the numerical inversion of non-probability transforms", *INFORMS Journal of Computing*, vol. 9, no. 2, pp. 175-184.

CHOURDAKIS, K. (2005). "Switching Lévy models in continuous time: finite distributions and option pricing", Proceedings of the Quantitative Methods in Finance conference 2005, Sydney, Australia, http://gemini.econ.umd.edu/cgi-bin/conference/download.cgi?db_name=QMF2005&paper_id=81.

CHRISTOFFERSEN, P. AND K. JACOBS (2004). "Which GARCH model for option valuation?", *Management Science*, vol. 50, no. 9, pp. 1204-1221.

CONT, R. AND P. TANKOV (2004). *Financial modelling with jump processes*, Chapman and Hall.

CORTAZAR, G. AND E. SCHWARTZ (1994). "The valuation of commodity-contingent claims", *Journal of Derivatives,* vol. 1, no. 4, pp. 27-39.

COX, J.C., INGERSOLL, J.E. AND S.A. ROSS (1985). "A theory of the term structure of interest rates", *Econometrica*, vol. 53, no. 2, pp. 385-407.

CURRAN, M. (1992). "Beyond average intelligence", *Risk Magazine*, vol. 5, no. 10.

CURRAN, M. (1994). "Valuing Asian and Portfolio Options by Conditioning on the Geometric Mean Price", *Management Science*, vol. 40, no. 12, pp. 1705-1711.

DANIELS, H.E. (1954). "Saddlepoint approximations in statistics", *The Annals of Mathematical Statistics*, vol. 25, no. 4, pp. 631-650.

DEELSTRA, G. AND F. DELBAEN (1998). "Convergence of discretized stochastic (interest rate) processes with stochastic drift term", *Applied Stochastic Models and Data Analysis*, vol. 14, no. 1, pp. 77-84.

DEELSTRA, G., LIINEV, J. AND M. VANMAELE (2004). "Pricing of arithmetic basket options by conditioning", *Insurance: Mathematics and Economics*, vol. 34, no. 1, pp. 55-77.

DHAENE, J., DENUIT. M, GOOVAERTS, M., KAAS, R. AND D. VYNCKE (2002). "The concept of comonotonicity in actuarial science and finance: Theory", *Insurance: Mathematics and Economics*, vol. 31, no. 1, pp. 3-33.

DIOP, A. (2003). "Sur la discrétisation et le comportement à petit bruit d'EDS unidimensionelles dont les coefficients sont à derives singulières", PhD thesis, INRIA.

DORNIC, I., CHATÉ, H. AND M.A. MUÑOZ (2005). "Integration of Langevin equations with multiplicative noise and the viability of field theories for absorbing phase transitions", *Physical Review Letters*, vol. 94, no. 10, pp. 100601-1, 100601-4.

DUBNER, H. AND J. ABATE (1968). "Numerical inversion of Laplace transforms by relating them to the finite Fourier cosine transform", *Journal of the ACM*, vol. 15, no. 1, pp. 115-123.

DUFFIE, D. AND P. GLYNN (1995). "Efficient Monte Carlo simulation of security prices", *Annals of Applied Probability*, vol. 5, no. 4, pp. 897-905.

DUFFIE, D., D. FILIPOVIĆ AND W. SCHACHERMAYER (2003). "Affine processes and applications in finance", *Annals of Applied Probability*, vol. 13, no. 3, pp. 984-1053.

DUFFIE, D. AND R. KAN (1996). "A yield-factor model of interest rates", *Mathematical Finance*, vol. 6, no. 4, pp. 379-406.

DUFFIE, D., PAN, J. AND K. SINGLETON (2000). "Transform analysis and asset pricing for affine jump-diffusions", *Econometrica*, vol. 68, pp. 1343-1376.

DUNBAR, N. (2000). *Investing Money: The story of Long-Term Capital Management and the legends behind it*, John Wiley and Sons.

FAHRNER, I. (2007). "Modern logarithms for the Heston model", *International Journal of Theoretical and Applied Finance*, vol. 10, no. 1, pp. 23-30.

FALLOON, W. AND D. TURNER (1999). "The evolution of a market", *Managing Energy Price Risk* (Second edition), Risk Publications.

FAN, G. AND G.H. LIU (2004). "Fast Fourier Transform for discontinuous functions", *IEEE Transactions on Antennas and Propagation*, vol. 52, no. 2, pp. 461-465.

FAULHABER, O. (2002). "Analytic methods for pricing double barrier options in the presence of stochastic volatility", MSc thesis, University of Kaiserslautern, available at: http://www.oliverfaulhaber.de/diplomathesis/HestonBarrierAnalytic.pdf

FELLER, W. (1951). "Two singular diffusion problems", *Annals of Mathematics*, vol. 54, pp. 173-182.

FÖRSTER, K-J. AND K. PETRAS (1991). "Error estimates in Gaussian quadrature for functions of bounded variations", *SIAM Journal of Numerical Analysis*, vol. 28, no. 3, pp. 880-889.

FORZANI, L. AND C. TOLMASKY (2003). "A family of models explaining the level-slope-curvature effect", *International Journal of Theoretical and Applied Finance*, vol. 6, no. 3, pp. 239-255.

GANDER, W. AND W. GAUTSCHI (2000). "Adaptive quadrature – revisited", *BIT*, vol. 40, no. 1, pp. 84-101, available as CS Technical Report at: ftp://ftp.inf.ethz.ch/pub/publications/tech-reports/3xx/306.ps.gz.

GANTMACHER, F.R. AND M.G. KREĬN (1937). "Sur les matrices complètement non negatives et oscillatoires", *Composito Math*, vol. 4, pp. 445-476.

GANTMACHER, F.R. AND M.G. KREĬN (1960). "Oszillationsmatrizen, Oszillationskerne und Kleine Schwingungen Mechanischer Systeme,", Akademie-Verlag, Berlin.

GANTMACHER, F.R. AND M.G. KREĬN (2002). "Oscillation matrices and kernels and small vibrations of mechanical systems", AMS Chelsea Publishing.

GASPAR, R. (2004). "General quadratic term structures of bond, futures and forward prices", *SSE/EFI Working paper Series in Economics and Finance*, no. 559, available at: http://ssrn.com/abstract=913460.

GATHERAL, J. (2006). *The volatility surface: a practitioner's guide*, Wiley.

GEMAN, H. AND M. YOR (1993). "Bessel processes, Asian options and perpetuities", *Mathematical Finance*, vol. 4, no. 3, pp. 349-375.

GESKE, R. AND H. JOHNSON (1984). "The American put valued analytically", *Journal of Finance*, vol. 39, no. 5, pp. 1511-1524.

GIL-PELAEZ, J. (1951). "Note on the inversion theorem", *Biometrika*, vol. 37., no. 3-4, pp. 481-482.

GLASSERMAN, P. (2003). *Monte Carlo methods in financial engineering*, Springer Verlag, New York.

GLASSERMAN, P., HEIDELBERGER, P. AND P. SHAHABUDDIN (1999). "Asymptotically optimal importance sampling and stratification for pricing path-dependent options", *Mathematical Finance*, vol. 9, no. 2, pp. 117-152.

GURLAND, J. (1948). "Inversion formulae for the distribution of ratios", *Annals of Mathematical Statistics*, vol. 19, no. 2, pp. 228-237.

HAGAN, P.S., KUMAR, D., LESNIEWSKI, A.S. AND D.E. WOODWARD (2002). "Managing Smile Risk", *Wilmott Magazine*, July 2002, pp. 84-108.

D'HALLUIN, Y., FORSYTH, P.A. AND G. LABAHAN (2004). "A penalty method for American options with jump diffusion processes", *Numerische Mathematik*, vol. 97, no. 2, pp. 321-352.

HENDERSON, V. AND R. WOJAKOWSKI (2002). "On the equivalence of floating and fixed-strike Asian options", *Journal of Applied Probability*, vol. 39, no. 2, pp. 391-394.

HESTON, S.L. (1993). "A closed-form solution for options with stochastic volatility with applications to bond and currency options", *Review of Financial Studies*, vol. 6, no. 2, pp. 327-343.

HESTON, S.L. AND S. NANDI (2000). "A closed-form GARCH option valuation model", *Review of Financial Studies*, vol. 13, no. 3, pp. 585-625.

HIGHAM, D.J. (2004). *An introduction to financial option valuation*, Cambridge University Press.

HIGHAM, D.J. AND X. MAO (2005). "Convergence of the Monte Carlo simulations involving the mean-reverting square root process", *Journal of Computational Finance*, vol. 8, no. 3, pp. 35-62.

HILL, I.D., HILL, R. AND R.L. HOLDER (1976). "Fitting Johnson Curves by Moments", *Applied Statistics*, vol. 25, no. 2, pp. 180-189.

HINDANOV, D. AND C. TOLMASKY (2002). "Principal components analysis for correlated curves and seasonal commodities: The case of the petroleum market", *Journal of Futures Markets*, vol. 22, no. 11, pp. 1019-1035.

HIRSA, A. AND D.B. MADAN (2004). "Pricing American options under Variance Gamma", *Journal of Computational Finance*, vol. 7, no. 2, pp. 63-80.

HONG, G. (2004). "Forward smile and derivative pricing", presentation, University Finance Seminar of the Centre for Financial Research at Cambridge, available at: http://www-cfr.jbs.cam.ac.uk/archive/PRESENTATIONS/seminars/2004/hong.pdf.

HOOGLAND, J.K. AND C.D.D. NEUMANN (2000a). "Asian and cash dividends: Exploiting symmetries in pricing theory", CWI report MAS-R0019, available at: http://ssrn.com/abstract_id=234830.

HOOGLAND, J.K. AND C.D.D. NEUMANN (2000b). "Tradable schemes", CWI report MAS-R0024, available at: http://ssrn.com/abstract_id=241128.

HOWISON, S. (2007). "A matched asymptotic expansions approach to continuity corrections for discretely sampled options. Part 2: Bermudan options", *Applied Mathematical Finance*, vol. 14, no. 1, pp. 91-104.

HU, Z., KERKHOF, J., MCCLOUD, P. AND J. WACKERTAPP (2006). "Cutting edges using domain integration", *Risk*, vol. 19, no. 11, pp. 95-99.

HULL, J.C. (2005). *Options, Futures, and Other Derivatives* (Sixth Edition), Prentice Hall.

HULL, J.C. AND A. WHITE (1993). "Efficient procedures for valuing European and American path-dependent options", *Journal of Derivatives*, vol. 1, no. 1, pp. 21-31.

HUNT, P., KENNEDY, J. AND A.A.J. PELSSER (2000). "Markov-functional interest rate models", *Finance and Stochastics*, vol. 4, no. 4, pp. 391-408.

ISEGER, P. DEN (2006). "Numerical transform inversion using Gaussian quadrature", *Probability in the Engineering and Informational Sciences*, vol. 20, no. 1, pp. 1-44.

ISEGER, P. DEN AND E. OLDENKAMP (2006). "Pricing guaranteed return rate products and average price options", *Journal of Computational Finance*, vol. 9, no. 3.

JÄCKEL, P. (2002). *Monte Carlo methods in finance*, John Wiley and Sons.

JÄCKEL, P. (2004). "Stochastic volatility models: past, present and future", pp. 379-390 in: *The Best of Wilmott 1: Incorporating the Quantitative Finance Review*, P. Wilmott (ed.), John Wiley and Sons.

JACKSON, J.E. (2003). "A user's guide to principal components", *Wiley*.

JAMSHIDIAN, F. (1989). "An exact bond option formula", *Journal of Finance*, vol. 44, no. 1, pp. 205-209.

JARROW, R. AND A. RUDD (1982). "Approximate option valuation for arbitrary stochastic processes", *Journal of Financial Economics*, vol. 10, no. 3, pp. 347-369.

JOHNSON, N.L.. (1949). "Systems of frequency curves generated by methods of translation", *Biometrika* vol. 36, pp. 149-176.

JONG, F. DE, DRIESSEN, J. AND A. PELSSER (2004). "On the information in the interest rate term structure and option prices", *Review of Derivatives Research*, vol. 7., no. 2, pp. 99-127.

JOSHI, M. (2000). "A short note on exponential correlation functions", working paper, Royal Bank of Scotland Quantitative Research Centre.

JOSHI, M. (2003). *The concepts and practice of mathematical finance*, Cambridge University Press.

JU, N. (2002). "Pricing Asian and basket options via Taylor expansion", *Journal of Computational Finance*, Spring 2002.

KAHL, C. (2004). "Positive numerical integration of stochastic differential equations", Diploma thesis, University of Wuppertal and ABN·AMRO.

KAHL, C. AND P. JÄCKEL (2005). "Not-so-complex logarithms in the Heston model", *Wilmott Magazine*, September 2005.

KAHL, C. AND P. JÄCKEL (2006). "Fast strong approximation Monte-Carlo schemes for stochastic volatility models", *Quantitative Finance*, vol. 6, no. 6, pp. 513-536.

KARLIN, S. (1968). *Total positivity, Volume I*, Stanford University Press.

KARLIN, S. AND H. TAYLOR (1981). *A second course in stochastic processes*, Academic Press.

KËLLEZI, E. AND N. WEBBER (2004). "Valuing Bermudan options when asset returns are Lévy processes", *Quantitative Finance*, vol. 7, no. 3, pp. 51-86.

KELLOGG, O.D. (1918). "Orthogonal function sets arising from integral equations", *American Journal of Mathematics*, vol. 40, pp. 145-154.

KEMNA, A.G.Z. AND A.C.F. VORST (1990). "A pricing method for options based on average asset values", *Journal of Banking and Finance*, vol. 14, pp. 113-129.

KHUONG-HUU, P. (1999). "Swaptions with a smile", *Risk*, vol. 12, no. 8, pp. 107-111.

KLASSEN, T.R. (2001). "Simple, fast, and flexible pricing of Asian options", *Journal of Computational Finance*, vol. 4, no. 3, pp. 89-124.

KLOEDEN, P.E. AND E. PLATEN (1999). *Numerical solution of stochastic differential equations*, 3rd edition, Springer Verlag, New York.

KOU, S.G. (2002). "A jump-diffusion model for option pricing", *Management Science*, vol. 48, no. 8, pp. 1086-1101.

KRUSE, S. AND U. NÖGEL (2005). "On the pricing of forward starting options in Heston's model on stochastic volatility", *Finance and Stochastics*, vol. 9, no. 2, pp. 233-250.

LARDIC, S., PRIAULET, P. AND S. PRIAULET (2003). "PCA of the yield curve dynamics: questions of methodologies", *Journal of Bond Trading and Management*, vol. 1, no. 4, pp. 327-349.

LEE, R.W. (2004). "Option pricing by transform methods: extensions, unification and error control", *Journal of Computational Finance*, vol. 7, no. 3, pp. 51-86.

LEENTVAAR, C.C.W. AND C.W. OOSTERLEE (2007). "Multi-asset option pricing using a parallel Fourier-based technique", *Technical report 12-07*, Delft University of Technology, http://ta.twi.tudelft.nl/TWA_Reports/07-12.pdf.

LEKKOS, I. (2000). "A critique of factor analysis of interest rates", vol. 8, no. 1, pp. 72-83.

LEVENDORSKIĬ, S. AND V.M. ZHERDER (2002). "Fast option pricing under regular Lévy processes of exponential type", working paper, University of Texas at Austin and Rostov State University of Economics, available at: http://www.eco.utexas.edu/facstaff/Levendorskiy/paper02.pdf.

LEVY, E. (1992). "Pricing European average rate currency options", *Journal of International Money and Finance*, vol. 11, pp. 474-491.

LÉVY, P. (1925). "Calcul des probabilités", *Gauthiers-Villars*, Paris.

LEWIS, A. (2001). "A simple option formula for general jump-diffusion and other exponential Lévy processes", working paper, OptionCity.net, http://ssrn.com/abstract=282110.

LIPTON, A. (2001). *Mathematical methods for foreign exchange,* World Scientific.

LITTERMAN, R. AND J. SCHEINKMAN (1991). "Common factors affecting bond returns", *Journal of Fixed Income*, vol. 1, no. 1, pp. 54-61.

LORD, R. (2006). "Partially exact and bounded approximations for arithmetic Asian options", *Journal of Computational Finance*, vol. 10, no. 2, pp. 1-52.

LORD, R. (2006). "Pricing of baskets, Asians and swaptions in general models", working paper. Presentation given at 5th Winter School on Financial Mathematics available at: http://www.rogerlord.com/condition_and_conquer_lunteren.pdf.

LORD, R. AND C. KAHL (2006). "Why the rotation count algorithm works", working paper, Rabobank International and ABN·AMRO, Tinbergen Institute Discussion Paper 2006-065/2, available at: http://ssrn.com/abstract=921335.

LORD, R. (2007) AND A.A.J. PELSSER. "Level-slope-curvature – fact or artefact?", *Applied Mathematical Finance*, vol 14, no. 2, pp. 105-130.

LORD, R. AND C. KAHL (2007A). "Complex logarithms in Heston-like models", forthcoming in: *Mathematical Finance*, available at: http://ssrn.com/abstract=1105998.

LORD, R. AND C. KAHL (2007B). "Optimal Fourier inversion in semi-analytical option pricing", *Journal of Computational Finance*, vol. 10, no. 4, pp. 1-30.

LORD, R., FANG, F., BERVOETS, F. AND C.W. OOSTERLEE (2008). "A fast and accurate FFT-based method for pricing early-exercise options under Lévy processes", *SIAM Journal on Scientific Computing*, vol. 30, no. 4, pp. 1678-1705.

LORD, R., KOEKKOEK, R. AND D. VAN DIJK (2008). "A comparison of biased simulation schemes for stochastic volatility models", forthcoming in: *Quantitative Finance*, available at: http://ssrn.com/abstract=903116.

LUCIC, V. (2004). "Forward-start Options in Stochastic Volatility Models", pp. 413-420 in P. Wilmott (ed). *The Best of Wilmott 1: Incorporating the Quantitative Finance Review*, P. Wilmott (ed.), John Wiley and Sons.

LUGANNANI, R. AND S. RICE (1980). "Saddlepoint approximation for the distribution of the sum of independent random variables", *Advances in Applied Probability*, vol. 12, pp. 475-490.

LUKACS, E. (1970). *Characteristic functions*, 2nd edition, Griffin, London.

MADAN, D.B. AND E. SENETA (1990). "The variance gamma (V.G.) model for share market returns", *Journal of Business*, vol. 63, no. 4, pp. 511-524.

MADAN, D.B., CARR, P.P. AND E.C. CHANG (1998). "The Variance Gamma process and option pricing", *European Finance Review*, vol. 2, no. 1, pp. 79-105.

MALLER, R.A., SOLOMON, D.H. AND A. SZIMAYER (2006). "A multinomial approximation for American option prices in Lévy process models", *Mathematical Finance*, vol. 16, no. 4, pp. 613-633.

MARSAGLIA, G. AND W.W. TSANG (2000). "A simple method for generating gamma variables", *ACM Transactions on Mathematical Software*, vol. 26, no. 3, pp. 363-372.

MATACHE, A.M., NITSCHE, P.A. AND C. SCHWAB (2005). "Wavelet Galerkin pricing of American options on Lévy driven assets", *Quantitative Finance*, vol. 5, no. 4, pp. 403-424.

MATYTSIN, A. (1999). "Modelling volatility and volatility derivatives", Columbia Practitioners Conference on the Mathematics of Finance, available at: http://www.math.columbia.edu/~smirnov/Matytsin.pdf.

MERCURIO, F. AND N. MORENI (2006). "Inflation with a smile", *Risk*, vol. 19, no. 3, March 2006, pp. 70-75.

MERTON, R.C. (1973). "Theory of rational option pricing", *Bell Journal of Economics and Management Science*, vol. 4, no. 1, pp. 141-183.

MERTON, R.C. (1976). "Option pricing when underlying stock returns are discontinuous", *Journal of Financial Economics*, vol. 3, no. 1-2,, pp. 125-144.

MIKHAILOV, S. AND U. NÖGEL (2004). "Heston's stochastic volatility model: implementation, calibration and some extensions", pp. 401-412 in P. Wilmott (ed). *The Best of Wilmott 1: Incorporating the Quantitative Finance Review*, P. Wilmott (ed.), John Wiley and Sons.

MILEVSKY, M.A. AND S.E. POSNER (1998). "Asian options, the sum of lognormals, and the reciprocal gamma distribution", *Journal of Financial and Quantitative Analysis*, vol. 33, no. 3, pp. 409-422.

MORO, E. (2004). "Numerical schemes for continuum models of reaction-diffusion systems subject to internal noise", *Physical Review E*, vol. 70, no. 4, pp. 045102(R)-1, 045102(R)-4.

MORO, E. AND H. SCHURZ (2007). "Boundary preserving semi-analytic numerical algorithms for stochastic differential equations", *SIAM Journal on Scientific Computing*, vol. 29, no. 4, pp. 1525-1549.

NELSON, C.R. AND A.F. SIEGEL (1987). "Parsimonious modeling of yield curves", *Journal of Business*, vol. 60, no. 3, pp. 473-489.

NIELSEN, J.A. AND K. SANDMANN (2003). "Pricing bounds on Asian options", *Journal of Financial and Quantitative Analysis*, vol. 38, no. 2, pp. 449-473.

NINOMIYA, S. AND N. VICTOIR (2004). "Weak approximation of stochastic differential equations and application to derivative pricing", working paper, Tokyo Institute of Technology and Oxford University.

PAPAPANTOLEON, A. (2006). "An introduction to Lévy processes with applications in finance", lecture notes, University of Freiburg, available at: http://www.stochastik.uni-freiburg.de/~papapan/Papers/introduction.pdf.

PINKUS, A. (1995). "Spectral properties of totally positive kernels and matrices", in: *Total positivity and its applications* (eds.: M. Gasca and C.A. Micchelli), Kluwer Academic Publishers, pp. 477-511.

POSNER. S.E. AND M.A. MILEVSKY (1998). "Valuing exotic options by approximating the spd with higher moments", *Journal of Financial Engineering*, vol. 7, no. 2, pp. 109-125.

PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T. AND B.P. FLANNERY (2007). *Numerical Recipes – The Art of Scientific Computing*, 3rd edition, Cambridge University Press.

RAIBLE, S. (2000). "Lévy processes in finance: theory, numerics and empirical facts", PhD thesis, Institut für Mathematische Stochastik, Albert-Ludwigs-Universität, Freiburg.

REBONATO, R. (2002). *Modern pricing of interest-rate derivatives: The LIBOR market model and beyond*, Princeton University Press.

REBONATO, R. AND P. JÄCKEL (1999/2000). "The most general methodology for creating a valid correlation matrix for risk management and option pricing purposes", *Journal of Risk*, vol. 2, no. 2, pp. 1–16.

REINER, E. (2001). "Convolution methods for path-dependent options", Financial Mathematics workshop, Institute for Pure and Applied Mathematics, UCLA, January 2001, available at: http://www.ipam.ucla.edu/publications/fm2001/fm2001_4272.pdf.

REVUZ, D. AND M. YOR (1991). *Continuous Martingales and Brownian Motion*, Springer Verlag, New York.

ROGERS, L.C.G. AND O. ZANE (1999). "Saddlepoint approximations to option pricing", *Annals of Applied Probability*, vol. 9, no. 2, pp. 493-503.

ROGERS, L.C.G. AND Z. SHI (1995). "The value of an Asian option", *Journal of Applied Probability*, no. 32, pp. 1077-1088.

SANTA-CLARA, P. AND D. SORNETTE (2001). "The dynamics of the forward interest rate curve with stochastic string shocks", *Review of Financial Stu*dies, vol. 14, no. 1, pp. 149-185.

SATO, K.-I. (1999). *Lévy processes and infinitely divisible distributions*, Cambridge University Press.

SATO, K-I. (2001). "Basic results on Lévy processes", in: *Lévy processes: theory and applications*, O.E. Barndorff-Nielsen, T. Mikosch and S.I. Resnick (eds.), Birkhäuser.

SCHÖBEL, R. AND J. ZHU (1999). "Stochastic volatility with an Ornstein-Uhlenbeck process: an extension", *European Finance Review*, vol. 3, pp. 23-46.

SCHOENMAKERS, J.G.M. AND B. COFFEY (2003). "Systematic Generation of Correlation Structures for the Libor Market Model", *International Journal of Theoretical and Applied Finance*, vol. 6, no. 4, pp. 1-13.

SCHOUTENS, W., SIMONS, E. AND J. TISTAERT (2004). "A perfect calibration! Now what?", *Wilmott Magazine*, March 2004.

SCHRAGER, D.F. AND A.A.J. PELSSER (2004). "Pricing rate of return guarantees in regular premium unit linked insurance", *Insurance: Mathematics and Economics*, vol. 35, no. 2, pp. 369-398.

SIMON, S., GOOVAERTS, M.J. AND J. DHAENE (2000). "An easy computable upper bound for the price of an arithmetic Asian option", *Insurance: Mathematics and Economics*, vol. 26, no. 2-3, pp. 175-184.

STEELEY, J.M. (1990). "Modelling the dynamics of the term structure of interest rates", *Economic and Social Review*, vol. 21, no. 4, pp. 337-361.

STEIN, E. AND J. STEIN (1991). "Stock-price distributions with stochastic volatility – an analytic approach", *Review of Financial Studies*, vol. 4, pp. 727-752.

O'SULLIVAN, C. (2005). "Path dependent option pricing under Lévy processes", EFA 2005 Moscow meetings paper, available at: http://ssrn.com/abstract=673424.

SVENSSON, L. (1994). "Estimating and interpreting forward interest rates: Sweden 1992-94", NBER working paper no. w4871.

TEWELES, R.J. AND F.J. JONES (1999). *The Futures Game: Who Wins? Who Loses? Why?*, 3rd edition, edited by B. Warwick, McGraw Hill.

THOMPSON, G.W.P. (1999A). "Fast narrow bounds on the value of Asian options", working paper, University of Cambridge, available at: http://www-cfr.jbs.cam.ac.uk/archive/giles/thesis/asian.pdf.

THOMPSON, G.W.P. (1999B). "Topics in Mathematical Finance", PhD thesis, University of Cambridge, available at: http://www-cfr.jbs.cam.ac.uk/archive/giles/thesis/thesis.pdf.

TOIVANEN, J. (2008). "Numerical valuation of European and American options under Kou's jump-diffusion model", *SIAM Journal on Scientific Computing*, vol. 30, no. 4, pp. 1949-1970.

TURNBULL, S.M. AND L.M. WAKEMAN (1991). "Quick algorithm for pricing European average options", *Journal of Financial and Quantitative Analysis*, vol. 26, no. 3, pp. 377-389.

VANMAELE, M., DEELSTRA, G. AND J. LIINEV (2004). "Approximation of stop-loss premiums involving sums of lognormals by conditioning on two variables", *Insurance: Mathematics and Economics*, vol. 35, no. 2, pp. 343-367.

VANMAELE, M., DEELSTRA, G., LIINEV, J., DHAENE, J. AND M.J. GOOVAERTS (2006). "Bounds for the price of discretely sampled arithmetic Asian options", *Journal of Computational and Applied Mathematics*, vol. 185, no. 1, pp. 51-90.

VEČEŘ, J. (2001). "A new PDE approach for pricing arithmetic average Asian options", *Journal of Computational Finance*, vol. 4, no. 4, pp. 105-113.

VORST, T. (1992). "Prices and Hedge Ratios of Average Exchange Rate Options", *International Review of Financial Analysis*, vol. 1, no. 3, pp. 179-193.

VYNCKE, D., GOOVAERTS, M.J. AND J. DHAENE (2004). "An accurate analytical approximation for the price of a European-style arithmetic Asian option", *Finance*, vol. 25, HS, pp. 121-139.

WANG, I., WAN, J.W. AND P. FORSYTH (2007). "Robust numerical valuation of European and American options under the CGMY process", *Journal of Computational Finance*, vol. 10, no. 4, pp. 31-70.

WILLARD, G.A. (1997). "Calculating prices and sensitivities for path-independent derivative securities in multifactor models", *Journal of Derivatives*, vol. 5, no. 1, pp. 45-61.

WILMOTT, P. (2006). *Paul Wilmott on Quantitative Finance* (Second Edition), John Wiley & Sons.

WILMOTT, P., DEWYNNE, J. AND S. HOWISON (1993). *Option pricing*, Oxford Financial Press.

YAMADA, T. (1978). "Sur une construction des solutions d'équations différentielles stochastiques dans le cas non-Lipschitzien", in *Séminaire de Probabilité*, vol. XII, LNM 649, pp. 114-131, Springer, Berlin.

ZHU, J. (2000). *Modular pricing of options – an application of Fourier analysis*, Lecture notes in economics and mathematical systems, no. 493, Springer Verlag.

ZVAN, R., FORSYTH, P. AND K. VETZAL (1997/98). "Robust numerical methods for PDE models of Asian options", *Journal of Computational Finance*, vol. 1, no. 2, pp. 39-78.

The Tinbergen Institute is the Institute for Economic Research, which was founded in 1987 by the Faculties of Economics and Econometrics of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam. The Institute is named after the late Professor Jan Tinbergen, Dutch Nobel Prize laureate in economics in 1969. The Tinbergen Institute is located in Amsterdam and Rotterdam. The following books recently appeared in the Tinbergen Institute Research Series:

40. JEROEN DE MUNNIK, *The valuation of interest rates derivative securities.*
49. HARRY KAT, *The efficiency of dynamic trading strategies in imperfect markets.*
94. RONALD HEYNEN, *Models for option evaluation in alternative price-movements.*
121. TERRY CHEUK, *Exotic options.*
125. FABIO MERCURIO, *Claim pricing and hedging under market imperfections.*
155. JUAN MORALEDA, *On the pricing of interest-rate options.*
193. BART OLDENKAMP, *Derivatives in Portfolio Management.*
320. PATRICK HOUWELING, *Empirical Studies on Credit Markets.*
360. ROGER LAEVEN, *Essays on Risk Measures and Stochastic Dependence. With Applications To Insurance and Finance.*
366. ANTOINE VAN DER PLOEG, *Stochastic Volatility and the Pricing of Financial Derivatives.*
397. D.F. SCHRAGER, *Essays on asset liability modeling.*
398. R. HUANG, *Three essays on the effects of banking regulations*.
399. C.M. VAN MOURIK, G*lobalisation and the role of financial accounting information in Japan.*
400. S.M.S.N. MAXIMIANO, *Essays in organizational economics.*
401. W. JANSSENS, *Social capital and cooperation: An impact evaluation of a women's empowerment programme in rural India.*
402. J. VAN DER SLUIS, *Successful entrepreneurship and human capital.*
403. S. DOMINGUEZ MARTINEZ, *Decision making with asymmetric information*.
404. H. SUNARTO, *Understanding the role of bank relationships, relationship marketing, and organizational learning in the performance of people's credit bank.*
405. M.Â. DOS REIS PORTELA, *Four essays on education, growth and labour economics.*
406. S.S. FICCO, *Essays on imperfect information-processing in economics.*
407. P.J.P.M. VERSIJP, *Advances in the use of stochastic dominance in asset pricing*.

408. M.R. WILDENBEEST, *Consumer search and oligopolistic pricing: A theoretical and empirical inquiry.*
409. E. GUSTAFSSON-WRIGHT, *Baring the threads: Social capital, vulnerability and the well-being of children in Guatemala*.
410. S. YERGOU-WORKU, *Marriage markets and fertility in South Africa with comparisons to Britain and Sweden.*
411. J.F. SLIJKERMAN, *Financial stability in the EU*.
412. W.A. VAN DEN BERG, *Private equity acquisitions*.
413. Y. CHENG, *Selected topics on nonparametric conditional quantiles and risk theory*.
414. M. DE POOTER, *Modeling and forecasting stock return volatility and the term structure of interest rates*.
415. F. RAVAZZOLO, *Forecasting financial time series using model averaging.*
416. M.J.E. KABKI, *Transnationalism, local development and social security: the functioning of support networks in rural Ghana.*
417. M. POPLAWSKI RIBEIRO, *Fiscal policy under rules and restrictions.*
418. S.W. BISSESSUR, *Earnings, quality and earnings management: the role of accounting accruals.*
419. L. RATNOVSKI, *A Random Walk Down the Lombard Street: Essays on Banking.*
420. R.P. NICOLAI, *Maintenance models for systems subject to measurable deterioration.*
421. R.K. ANDADARI, *Local clusters in global value chains, a case study of wood furniture clusters in Central Java (Indonesia).*
422. V.KARTSEVA, *Designing Controls for Network Organizations: A Value-Based Approach.*
423. J. ARTS, Essays *on New Product Adoption and Diffusion.*
424. A. BABUS, *Essays on Networks: Theory and Applications.*
425. M. VAN DER VOORT, *Modelling Credit Derivatives***.**
426. G. GARITA, *Financial Market Liberalization and Economic Growth.*
427. E.BEKKERS, Essays *on Firm Heterogeneity and Quality in International Trade.*
428. H.LEAHU, *Measure-Valued Differentiation for Finite Products of Measures: Theory and Applications.*
429. G. BALTUSSEN, *New Insights into Behavioral Finance.*
430. W. VERMEULEN, *Essays on Housing Supply, Land Use Regulation and Regional Labour Markets.*
431. I.S. BUHAI, *Essays on Labour Markets: Worker-Firm Dynamics, Occupational Segregation and Workplace Conditions.*
432. C. ZHOU, *On Extreme Value Statistics.*

433. M. VAN DER WEL, *Riskfree Rate Dynamics: Information, Trading, and State Space Modeling.*
434. S.M.W. PHLIPPEN, *Come Close and Co-Create: Proximities in pharmaceutical innovation networks.*
435. A.V.P.B. MONTEIRO, *The Dynamics of Corporate Credit Risk: An Intensity-based Econometric Analysis.*
436. S.T. TRAUTMANN, *Uncertainty in Individual and Social Decisions: Theory and Experiments.*