Pricing interest rate derivatives in an illiquid market

GL Grobler

20068549

Thesis submitted for the degree *Philosophiae Doctor* in Business Mathematics and Informatics at the Potchefstroom Campus of the North-West University

Promoter: Prof F Lombard

Graduation October 2017

http://www.nwu.ac.za/

Acknowledgements

Firstly, I want to thank my supervisor, Prof. Freek Lombard, for his patience throughout. He provided me with expert guidance and helped me grow as a researcher. He accomplished this by encouraging me to take ownership of the problem and gave me room to develop my own ideas. I believe this will help me immeasurably in my future development as a researcher.

Besides my supervisor, my colleagues at the university also played a crucial role in my development. I want to especially thank Dr. Antoinetta Venter and Prof. Louis Labuschagne for their words of encouragement from time to time.

I also want to thank my parents for their support. They helped me in whichever capacity they could.

Lastly, but most importantly, I want to express my sincere gratitude to my wife Karin, two-year old toddler Katrien and my baby boy Jaco who helped me keep things in perspective.

Declaration

I declare that, apart from the assistance acknowledged, the research presented in this thesis is my own unaided work. It is being submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Business Mathematics and Informatics at the Potchefstroom campus of the North-West University. It has not been submitted before for any degree or examination to any other University.

Nobody, but myself is responsible for the final version of this thesis.

Signature.....

Date.....

Abstract

Globally, one-factor diffusion processes have been popular models for the short rate by virtue of their analytically tractable features. However, due to shortcomings of these models in certain markets a number of models, such as two-factor diffusion and jump diffusion models, have been developed over time. Interest rate models for the South African market have not been researched thoroughly. As a consequence, one-factor diffusion models remain the popular choice in South African interest rate markets. We will investigate, by empirical means, whether one-factor diffusion models are suitable for the modelling of domestic short dated low risk interest rate data.

We will show evidence that the South African short rate should be modelled by a pure jump process. The evidence is found through empirically analysing and applying hypothesis tests for jumps on historical 3-month Johannesburg Interbank Agreed Rate (JIBAR) data. We fit a nonstationary compound Poisson process with stably distributed jumps and rate dependent intensities to the 3-month JIBAR. As a result we use a slightly altered model to price options on the 3-month forward JIBAR. We find potentially large changes of these option prices compared to prices derived from a nonparametric one-factor diffusion short rate model.

In order to fit a distribution from the family of stable distributions we show how to estimate its parameters. We apply two methods and compare the results with each other. To calculate maximum likelihood estimators (MLEs) we develop a method to estimate stable density function values. We compare these estimators to integrated least squared estimators (ILSEs). ILSEs are asymptotically less efficient than MLEs. However, we develop an algorithm to calculate the ILSEs that is quicker to apply than the method used to find MLEs.

Key terms: Pure jump processes, Nonstationary compound Poisson processes, Short rate models, JIBAR, Maximum likelihood estimators, Integrated least squared estimators.

Opsomming

Enkel-faktor diffusieprosesse was tot op hede, wêreldwyd, gewilde modelle vir die oombliklike termynrentekoers as gevolg van hul analities-beheerbare kenmerke. Die modelle het egter tekortkominge met hul toepassings in sekere markte, met die gevolg dat dubbel-faktor diffusiemodelle met die tyd ontwikkel is. Rentekoersmodelle vir die Suid-Afrikaanse mark is nie deeglik nagevors nie. As 'n gevolg is enkel-faktor modelle steeds gewild in die Suid-Afrikaanse markte. Ons stel ondersoek in, deur empiriese metodes toe te pas, of enkel-faktor diffusiemodelle gepas is vir die modellering van lokale korttermyn, lae-risiko, rentoekoers data.

Ons sal aantoon dat die Suid-Afrikaanse oombliklike termynrentekoers gemodelleer moet word met 'n suiwer sprongproses. Bewysstukke is gevind deur empiriese analises en die toepassing van hipotesetoetse vir spronge, toegepas op historiese 3-maand 'Johannesburg Interbank Agreed Rate' (JIBAR) data. Ons pas 'n nie-stasionêre saamgestelde Poisson proses met stabiel verdeelde spronge en rentekoersvlak afhanklike intensiteit op die 3-maand JIBAR data. As 'n gevolg prys ons opsies op die 3-maand JIBAR termynkontrak met 'n effens aangepaste model. Ons vind potensiële groot veranderinge in die opsie pryse in vergelyking met pryse wat verkry is vanaf 'n nie-parametriese enkel-faktor diffusie oombliklike termynrentekoersmodel.

Met die doelwit om 'n verdeling vanuit die familie van stabiele verdelings te pas toon ons aan hoe om die parameters te beraam. Ons pas twee metodes toe en vergelyk die resultate met mekaar. Ons ontwikkel 'n metode om stabiel verdeelde digtheidsfunksiewaardes te bereken met die doelwit om Maksimumaanneemlikheidsberamers (MLEs) te bereken. Ons vergelyk hierdie beramers met Geïntegreerde Kleinstekwadrateberamers (ILSEs). ILSEs is asimptoties minder effektief in vergelyking met MLEs. Ons ontwikkel egter 'n algoritme om ILSEs vinniger te bereken in vergelyking met MLEs.

Key terms: Diffusieproses, Nie-stasionêre saamgestelde Poisson proses, oombliklike termynrentekoersmodelle, JIBAR, Maksimumaanneemlikheidsberamers, Geïntegreerde Kleinstekwadrateberamers.

Basic Notations

$p(t,T)$ - Price at time t of a zero coupon bond maturing at time T.
r_t - Short rate.
$\mathbb{P}$ - Market measure.
$\mathbb{P}^*$ - Risk-neutral martingale measure.
μ - Drift coefficient of model OR location parameter of a stable distribution.
σ - Diffusion coefficient of diffusion process OR scale parameter of a stable distribution.
α - Index parameter of a stable distribution.
β - Skewness parameter of a stable distribution.
$\boldsymbol{\theta} = [\alpha, \beta, \sigma, \mu]$ - Parameter vector associated with a stable distribution.
W_t - Brownian motion (standard).
$N(\mathrm{d}t, \mathrm{d}x)$ - Marked point process differential.
N_t - Poisson process.
Z_n - I.i.d random variables representing jumps.
λ - Constant arrival intensity of a stationary Poisson process.
$\lambda(r_{t-})$ - Short rate dependent arrival intensity of a nonstationary Poisson process.
$\phi_X(t)$ - Characteristic function of a stochastic variable X.
$\hat{\phi}(t)$ - Empirical characteristic function.
$\mathscr{I}(\theta)$ - Fisher information.

Abbreviations

ILSE(s) - Integrated least squared estimation (estimators)
JIBAR - Johannesburg Interbank Agreed Rate
LIBOR - London Interbank Offer Rate
MLE(s) - Maximum likelihood estimation (estimators)
MPC - Monetary Policy Committee
OTC - Over-the-counter
Q-Q - Quantile-Quantile
SARB - South African Reserve Bank
T-Bill - Treasury Bill

Contents

Introduction

Interest rate derivatives have become popular financial instruments since the successful introduction of the Treasury bill (T-bill) futures contract on the Chicago Mercentile Exchange (CME) in 1975. The popularity of interest rate derivatives grew to the extent that the Eurodollar contract, created by the CME in 1982, became the "most actively traded of all futures contracts" (Chance, 1995).

A starting point in the process of pricing interest rate derivatives is to find the price of zero-coupon bonds. The price at time t of a zero-coupon bond maturing at time T, is denoted by $p(t,T)$. The bond price can be calculated as a function of the interest rate dynamics (defined by a stochastic process). More formally, a family of bond price stochastic processes (called the term structure), can be determined by the interest rate dynamics under a risk-neutral martingale measure $\mathbb{P}^*$ (Bjork, 2004). In other words, the price of a zero coupon bond can be calculated by taking the following expectation under $\mathbb{P}^*$

$$p(t,T) = E^* \left[e^{-\int_t^T r_s \mathrm{d}s} \right],$$

where

$$\mathrm{d}r_t = \mu(r_t)\mathrm{d}t + \sigma(r_t)\mathrm{d}W_t. \tag{1}$$

The stochastic process W_t is a Brownian motion under $\mathbb{P}^*$. The functions μ and σ are called the drift and diffusion coefficients respectively.

Various short rate models of the form (1) have been developed, with each model having advantages and disadvantages. Brigo and Mercurio (2006, p.54) list a number of factors to be taken into account to evaluate the appropriateness of a short rate model. For instance, the Vasiček model (Vasicek (1977)) with constant parameters a, b and σ given by

$$\mathrm{d}r_t = a\left(b - r_t\right)\mathrm{d}t + \sigma \mathrm{d}W_t, \tag{2}$$

ensures mean reversion of the interest rates to an interest rate level of b and explicit computation of bond and various standard option prices. However, the model produces a poor fit to the initial term structure of interest rates and allows for negative interest rates. Although negative interest rates have been observed in several countries recently, it was thought of as an impossible event (Leonhardt, 2016). The CIR model developed by

Cox et al. (1985), improved the Vasiček model by setting the diffusion coefficient equal to $\sigma\sqrt{r_t}$, which ensures positive interest rates. Another improvement on the Vasiček model is the more general short rate model by Hull and White (1990), which extends the Vasiček model by allowing for time dependent coefficients. The adjustment by Hull and White (1990) improves the ability of the model to "be fitted to the term structure of interest rates and the term structure of spot or forward rate volatilities" (Brigo and Mercurio, 2006, p.73). Short rate models imply a certain distribution for the short rate. For example, the Vasiček model implies a normal distribution of the short rate. The question is which model implies the best fit to the South African short rate distribution? How much do we sacrifice in terms of analytical tractability, by choosing a model which fits the short rate most accurately?

The models introduced above are all examples of one-factor diffusion processes. Johannes (2004) questions the applicability of a one-factor diffusion model in the interest rate market of the United States of America. He investigates whether a model, of the form (1), can be used to model the American short rate. Historical prices of the United States Treasury bill with 3 months maturity (called the 3-month T-bill) were used as a proxy for the American short rate. Johannes (2004) found that economic events, such as announcements on the federal funds target rate, lead to jumps which should be incorporated in the short rate process. A model incorporating jumps will have the form

$$\mathrm{d}r_t = \mu(r_t)\mathrm{d}t + \sigma(r_t)\mathrm{d}W_t + \int_{-\infty}^{\infty} J(t,x)N(\mathrm{d}t,\mathrm{d}x),$$

where $N(\mathrm{d}t,\mathrm{d}x)$ is a marked point process and J a real valued function on $\mathbb{R}$ (Protter, 2005, p. 26).

Jump diffusion models have been employed to describe the dynamics of various quantities. Originally, Merton (1976) developed the model to describe the dynamics of stock returns, which is given by

$$\frac{\mathrm{d}S_t}{S_{t-}} = (\mu - \lambda\kappa)\,\mathrm{d}t + \sigma\mathrm{d}W_t + \left(e^J - 1\right)N(\mathrm{d}t), \tag{3}$$

where S_t denotes the price process and N_t is a homogeneous Poisson process with constant arrival intensity λ and J represents i.i.d. normally distributed jump sizes. Jumps were added to model stock returns due to fat tailed distributions being observed as well as the inability of some diffusion models to be calibrated to an implied volatility smile (Kou and Wang, 2004). By adding jumps higher implied volatilities of options may be obtained, when pure diffusion models are unable to do so (Brigo and Mercurio, 2006, p.110).

From this brief history of short rate diffusion models (and jump diffusion models in general), the following questions were identified which need to be addressed. Firstly, is a diffusion model adequate to model the short rate in South Africa? If a liquid market, such as the US T-bill market, leads to jump diffusion short rate models, then surely questions

surrounding the appropriateness of diffusion short rate models in less liquid markets need to be answered.

Secondly, if jumps are present in the South African short rate, then, which short rate model should be used? If a pure jump model is adequate, then the distributions of the jump frequency as well as the jump sizes have to be modelled.

Lastly, if a short rate process include jumps, do they have an impact on interest rate derivative prices compared to prices obtained from a diffusion model? Practitioners will only be interested in the results if the impact on prices of interest rate derivatives is significant.

These three questions will be addressed in this thesis in relation to South African interest rates.

The first step is to identify a risk-free interest rate on a short term contract to use as a proxy for the mathematically defined, but not directly observable, short rate r_t. Johannes (2004) uses the 3-month T-bill in the US market. The interest rates derived from T-bills are called Treasury rates, or T-bill rates. Aling and Hassan (2012) use the Treasury rate derived from the 91-day T-bill traded in the South African market "which is commonly used as a market-determined proxy for the domestic short-term risk free rate"(Aling and Hassan, 2012, p.308). Alternatively, the 3-month Johannesburg Interbank Agreed Rate (JIBAR) can also be used as a proxy for the short rate as "derivative traders do not usually use Treasury rates as risk-free rates" (Hull, 2006, p. 76). Globally the JIBAR is similar to the London Interbank Offer Rate (LIBOR) as both rates are derived from a 3-month deposit a bank makes with other banks. The rate we used in our study was influenced by the data available. Although JIBAR data is only available from 1999 (compared to T-bill data being available from 1984), it is calculated each trading day. However, T-bills are weekly auctioned instruments, hence the T-bill rate is calculated once a week. In this thesis we will test for jumps, hence the frequency of observations is an important factor to take into account. We compared the T-bill data to the JIBAR data, and found similar patterns, and therefore decided to use the 3-month JIBAR as a proxy for the South African short rate mainly due our requirement in terms of observation frequency.

The method used in Johannes (2004) will be applied to test whether the diffusion process in formula (1) on page 1 can be used to model the short rate. The method is a hypothesis test, where the kurtoses from simulated paths generated by the process under the null hypothesis are compared to the sample kurtosis. If a nonparametric one-factor diffusion process is assumed under the null hypothesis, then the null hypothesis is rejected which indicated a pure diffusion is not adequate to model the short rate. If under the null hypothesis normally distributed jumps were added to the nonparametric one-factor diffusion model, then the model is not rejected. However, for low levels of the interest rate the volatility in the JIBAR changes is fully described by the volatility generated by

the jumps in the model, with the volatility from the diffusion component in the model being zero. This result corresponds with initial observations made where no diffusion component was evident for the period from 20 September 2010 to 11 April 2014. The conclusions from these tests were that a pure diffusion model is rejected, but whether any diffusion component should be incorporated into the model was questioned.

Another nonparametric test, described by Lee and Mykland (2008), was employed to identify which day-to day movements from the historical rates will not realistically be a result from a diffusion process. A statistical hypothesis test was used to test whether a realised return at a certain time point is within realistic bounds, assuming a sequence of preceding returns are from a diffusion process. In most cases either the JIBAR changes was equal to zero or the changes were rejected to be a realisation of a diffusion process. The results from applying the test from Lee and Mykland (2008) therefore confirmed our doubts whether or not a jump diffusion model should be used. The conclusion was made that a pure jump model should be used to model the short rate in South Africa.

To model the short rate by a pure jump model the interest rate returns, when jumps occur, therefore need to be modelled. Our analysis suggested that the returns do not emanate from either the normal distribution or from a Cauchy distribution, which is a thick tailed symmetric distribution. The normal and Cauchy distributions belong to the family of stable distributions, which also allows for non-symmetric distributions. We therefore fitted the jumps with a distribution from the family of stable distributions, and analysed whether it is a good fit.

We also fitted several distributions to the time elapsed between jumps. We did not find an ideal distribution to fit the data. However, we found an exponential distribution with rate dependent parameter $\lambda(r_{t-})$ provides us with an improved fit compared to an exponential distribution with constant parameter λ. This ensures the volatility in our model to be rate dependent, which is evident in the South African market (Aling and Hassan, 2012). The pure jump model is then a nonhomogeneous marked point process or more specifically a compound Poisson process with nonstationary increments. Therefore, the dynamics of our pure jump South African short rate model is given by

$$\mathrm{d}r_t = \mathrm{d}\left(\sum_{n=1}^{N_t} Z_n\right), \tag{4}$$

where N_t is a Poisson process with arrival intensity $\lambda(r_{t-})$ and Z_n are i.i.d. stably distributed random variables.

The aim in developing the short rate model in (4) is for the implied distribution from the short rate to fit the South African empirical distribution as best as possible. However, a realistic implied distribution of the short rate is only one of the factors to measure a short rate model against (Brigo and Mercurio, 2006, p.54). For example, the model in (4) does not imply mean reversion and bond and option prices are not explicitly computable.

One criterion mentioned is that the model should be suited for Monte Carlo simulation. If we apply our model in (4) to the pricing of interest rate derivatives, numerical methods are our only option as none of them are explicitly computable. Now, we can simulate short rate paths from a nonstationary compound Poisson process since stably distributed random variables can be simulated quickly and efficiently (Weron, 1996). However, stable variates (for index parameter $\alpha < 2$) have infinite second moments, which makes a pricing method using Monte Carlo simulation ineffective. Our main aim is to price interest rate derivatives and not to model the short rate. We therefore decided to use some of the observations made during the development of the model in (4) and define a new short rate model for the purpose of derivative pricing.

Let the dynamics of the short rate under a risk-neutral measure $\mathbb{P}^*$ be given by

$$\mathrm{d}r_t = \mu(r_t)\mathrm{d}t + \mathrm{d}\left(\sum_{n=1}^{N_t} Z_n\right),$$

where Z_n are now stably distributed random variables truncated at levels $\pm L$ with no drift and no skewness, i.e., the location and skewness parameter equals zero. The drift of the short rate is therefore fully determined by the drift coefficient $\mu(r_t)$, which can be chosen to ensure mean reversion (see Example 21 on page 30.) The volatility in the model is determined entirely by the jump process and is also dependent on interest rate levels through the rate dependent intensity $\lambda(r_{t-})$. A fat tailed distribution of the short rate can be implied by the model through the correct choice (or calibration) of the index parameter α. Importantly, Monte Carlo variates from the family of stable distributions are simulated to simulate jump sizes with finite variance through

$$Y = X\mathbb{1}_{\{|X|\leq L\}} + L\mathbb{1}_{\{X>L\}} - L\mathbb{1}_{\{X<-L\}},$$

with X a stably distributed random variable and $L \in \mathbb{R}_+$. Being able to simulate finite variance jumps, enables us to explore Monte Carlo methods to price various interest rate derivatives.

The thesis is structured as follows. Chapter 1 contains a literature survey conducted on the South African interest rate derivative market as well as a survey of models developed specifically for the South African market. Throughout the thesis the calculus for pure jump and jump diffusion processes is used to derive important results and make vital observations. The theory of the calculus is fully covered in existing literature (Cont and Tankov (2004) and Protter (2005)). However, Chapter 2 is used to outline the calculus specific to stationary and nonstationary compound Poisson processes. The objective of the chapter is to present the theory of the calculus in a more approachable manner, with the goal to apply Ito's lemma and Girsanov's theorem. The rest of the thesis is divided into two parts.

Part *I* of the thesis contains the empirical process used to answer the research questions. In Chapter 3 the history of the 3-month JIBAR is analysed. It seems as if occasional

rate changes are abnormally large if the underlying process is a pure diffusion process. In Chapter 4 the methods from Johannes (2004) and Lee and Mykland (2008) are used to test for jumps. Thereafter (Chapter 5), a nonstationary compound Poisson process is fitted to the interest rate changes, by first modelling the jump intensities (Section 5.1) and then the jump size (Section 5.2). In Chapter 6 a nonstationary compound Poisson process for the short rate is adapted to pricing interest rate derivatives. The impact of pure jump processes on a European call option as well as a barrier option on the 3-month forward JIBAR is analysed. The first part of the thesis addresses the main aim of our study, which is to price interest rate derivatives in the South African market. However, it is difficult to know which instruments are traded frequently, due to most instruments being traded over-the-counter (OTC), West (2008) indicates that options on the 3-month forward JIBAR are traded frequently.

To achieve our main objectives of the thesis, a study of the family of stable distributions is outlined in Part II. To fit the model in Chapter 5 a distribution from the family of stable distributions was fitted to the jump sizes. The main aim of Part II is to develop methods to estimate parameters from the family of stable distributions. To achieve this goal (by maximum likelihood estimation) density function values needed to be approximated (Section 8.2). In Chapter 9 parameter estimation methods are developed. The maximum likelihood estimation (MLE) (Section 9.2) and Integrated least squared error (ILSE) methods (Section 9.3) were used to fit the jump sizes to the family of stable distributions.

In Part I we simulated stably distributed variates in some of our applications. The simulation method by Weron (1996) is outlined in Section 8.4.

The rest of Part II is allocated to compare the asymptotic variance of the two parameter estimation methods used. In Chapter 10 the theory to calculate the asymptotic variance of ILSEs and MLEs (estimates from the ILSE and MLE methods) is described. In Chapter 11 we approximate the relative efficiency of the ILSEs, compared to the most efficient MLEs. We found that the relative efficiency can be improved by choosing various weight functions in the integrated least squared error formula (Section 11.3).

Chapter 1

Interest rate derivatives in the South African market

In this Chapter we conduct a literature survey of the interest rate derivative market in South Africa. We specifically survey the market itself as well as literature available on various interest rate models developed for the South African market. Hassan (2013) reviewed South African capital markets including, amongst other, the interest rate derivatives market in South Africa. Derivatives are traded on the Johannesburg Stock Exchange (JSE) or OTC. The total turnover on interest rate derivatives traded OTC far exceeds the turnover on the JSE (OTC daily average of $16 billion in 2013 versus approximately $35 billion on the JSE in 2012 in total). Although interest rate options, which can be difficult to price, are not the main contributor to the OTC interest rate derivative market, the turnover is large compared to interest rate option markets from other countries (larger than the option markets in Australia and Canada in 2010). The only interest rate options traded on the JSE are on Bond future prices. However, they are European style options which are priced by the standard Black formula for pricing options, where the underlying asset is a futures contract (Black, 1976). The JSE also accommodates trading on futures such as Bonds, 3-month JIBAR and Swaps. In our study we will develop a model to price interest rate derivatives where one of the objectives is to price exotic options. These types of options are traded OTC, where contracts are not standard.

Some instruments have the prime rate as underlying rate. According to West (2008) preferential shares have returns that are prime linked and he therefore developed a method to price derivatives with a payoff as a function of a forward prime rate. The value of many assets of financial institutions is dependent on the prime rate (such as home loans), but the exposure to the prime rate cannot directly be hedged by selling those assets. A solution is therefore to trade in derivatives on the 3-month JIBAR rate. Thus, vanilla options in the OTC market are "typically on the 3-month JIBAR" (West, 2008). He also did a cointegration analysis to detect common stochastic trends between the prime rate

and the 3-month JIBAR. As a result of the analysis the prime rate is shown to be a linear dependent function of the 3-month JIBAR. A standard call option, with the forward rate as underlying rate, is a caplet. West (2008) shows how a caplet on the forward prime rate can be written as a caplet on the forward 3-month JIBAR with a changed strike rate. Pricing of such a caplet using Black's formula is standard.

Various models have been developed to price the options traded in the South African market. Aling and Hassan (2012) compared the efficiency of a number of one-factor short rate models in the South African market, all special cases of the model given by

$$\mathrm{d}r_t = (\alpha + \beta r_t)\,\mathrm{d}t + \sigma r_t^{\gamma}\mathrm{d}W_t. \tag{1.1}$$

The model specifications ensure mean reversion (to $-\frac{\alpha}{\beta}$), where β is the speed of reversion to the mean and where the parameter γ measures the sensitivity of the variance to the short rate level. He applied the Bergstrom-Nowman maximum likelihood method to estimate model parameters (Nowman, 1997) using historical 3-month T-bill data from June 1984 to July 2011.

Since 2000 the monetary policy in South Africa adopted an inflation targeting framework. The results found by Aling and Hassan (2012) differ for the period before and after inflation targeting. The most appropriate model for the period before inflation targeting is the constant elasticity of variance (CEV) model (Cox, 1975). The CEV model has the following dynamics for the short rate

$$\mathrm{d}r_t = \beta r_t \mathrm{d}t + \sigma r_t^{\gamma}\mathrm{d}W_t,$$

where the estimate for γ is 0.7186, while the estimate for β is close to zero.

For the period since inflation targeting Aling and Hassan (2012) found the Brennan and Schwartz model (Brennan and Schwartz, 1980) to be the most appropriate model, with the CEV model also not rejected. The Brennan and Schwartz model dynamics is given by

$$\mathrm{d}r_t = (\alpha + \beta r_t)\,\mathrm{d}t + \sigma r_t \mathrm{d}W_t.$$

Both estimates for α and β are close to zero and the parameters are found to be statistically insignificant. The estimate for the γ parameter in the CEV model for this period is 0.9925 and is therefore statistically significant. This implies the variance in the model to be more sensitive to the short rate level in the time since inflation targeting, when compared to the time before. This observation is also indirectly true in the Brennan and Schwartz model as the measure of sensitivity is equal to one.

One-factor short rate models have been popular due to its analytical tractability. However, some disadvantages of these models do exist and this lead to various models being developed. An important aspect is that the model should be calibrated to the yield curve. In South Africa the JSE calculates three yield curves daily, which are plots "depicting the

yields on zero-coupon bonds for a continuum of maturities, in some time interval" (JSE, 2012). The three types of curves differ in the type of instruments used to construct them. The nominal bond curve is constructed by calculating the yields from zero-coupon T-bills with 91, 182, 273 and 365-day maturities and coupon bearing government bonds through the Government Bond (GOVI) index. The nominal swap curve is constructed by the yields on 1-month JIBAR, 3-month JIBAR, FRAs and Swaps, with various maturities in order to calculate a 30 year yield curve. For both curves the South African Futures Exchange (SAFEX) overnight rate is used as the starting point in the curves. This rate represents the average rate that SAFEX receives on its deposits with the banks. A third yield curve is the real bond curve, which "represents the real zero-coupon yields which the South African government can obtain funding" (JSE, 2012). The real bond curve is constructed by using inflation linked government bonds. Efficient one-factor short rate models can be easily calibrated to the starting point of the yield curve, but it assumes all the rates on the curve to be perfectly correlated, which is not true in reality. Two-factor models such as Longstaff and Schwartz (1992) enable rates in the yield curve to have imperfect correlation, and have the potential to provide a better fit to the yield curve.

A second objective of interest rate models is calibration to an implied volatility curve or surface. The procedure to calibrate a model is to calibrate the model to zero-coupon bond prices (or the zero-coupon yield curve) and to liquid options. To easily calibrate models to instruments, the formulas to value these instruments should have explicit forms. For example, an advantage of the Hull and White model (Hull and White, 1990) is the explicitly computable zero-coupon bonds and caps. An implied volatility curve is the relationship between the implied volatility obtained from inverting the Black formula (in the case of where the underlying rate of a European option is a forward rate, such as a cap or swaption) and a range of strike prices. Typically, the shape of the curve is convex (smile) or the volatility is a decreasing function of the strike (skew). In general one-factor short rate models do not produce the volatility curves (or surfaces) which are observed in the market. Therefore, stochastic volatility models - where the diffusion coefficient is replaced by a stochastic process - produces models that can replicate observed stochastic volatility as well as market smiles and skews (Brigo and Mercurio, 2006, p.495), while high implied volatilities can be obtained by adding jumps to a diffusion model. These models have therefore become popular models due to the criteria of accurate calibration to an implied volatility surface.

A one-factor diffusion model incorporating jumps has recently been developed for the South African market in Chapter 5 of the thesis by Malumisa (2015). In his model jumps were added to the model in (1.1). He thus describes the dynamics of the short rate by

$$\mathrm{d}r_t = (\alpha + \beta r_t)\,\mathrm{d}t + \sigma r_t^{\gamma}\mathrm{d}W_t + JN(\mathrm{d}t),$$

where the jumps J are normally distributed and the intensity of the jumps is constant. Malumisa (2015) uses 91 day T-bill data from January 1990 to August 2011 which gives

him 6784 observations. T-bills are auctioned once a week and therefore there are only approximately 1126 observations in his dataset where the rate could possibly have changed. The main results from Malumisa (2015) are that the returns have a high kurtosis, which cannot be replicated by the diffusion component of his model and jumps should be included into the model to obtain higher moments for short rate returns. By applying a likelihood test, a null-hypothesis of no jumps is rejected at a 1% significance. With at least 80% of his sample having values of zero, it is logical that a test for jumps will have this outcome. In our opinion the sample should only include the dates when the T-bill is auctioned, which will result in low frequency data. The assumption that the 91-day T-bill rate (as a proxy for the short rate) does not change between these dates will have an impact on the statistical results obtained. If the T-bill was auctioned on days in between current auction dates the rates would probably have changed more frequently. In our thesis we will revisit the problem addressed by Malumisa in part *I*.

A wide class of interest rate models such as short rate models and forward rate models have an underlying quantity that is not observed in the market. The short rate and instantaneous forward rate are mathematically defined functions, which cannot be calibrated directly to market data. For this reason, market models such as the forward LIBOR market models and forward Swap market models have become popular. In South Africa Gumbo (2012) prices caps and floors within a South African context with the forward JIBAR rate as underlying forward rate. For instance, if Y_t^T is the 3-month forward JIBAR rate from time T observed at time t then Y_t^T is modelled by a geometric Brownian motion:

$$\mathrm{d}Y_t^T = Y_t^T \sigma(t) \mathrm{d}W_t.$$

The SAFEX-JIBAR market models are formed by modelling a range of forward rates with a geometric Brownian motion. These types of models are popular as the international markets for caps and swaptions are two main interest rate option markets (Brigo and Mercurio, 2006, p.195). In South Africa caps and swaptions are only traded OTC, but West (2008) does give us an indication that caps are popular instruments in the South African interest rate options market because the underlying rate of a cap is the forward JIBAR rate.

Given the wide spectrum of models available we decided to start our investigation by questioning whether one-factor diffusion models can be used to price interest rate derivatives efficiently. Aling and Hassan (2012) states that one-factor diffusion models are "consistent with the high frequency of changes in market rates". From observing historical 3-month JIBAR and 3-month T-bill rates, this assumption is questioned. It seems as if jumps, which will lead to higher volatility, are observed in the market and, moreover, market rates also do not change frequently in some periods. This begs the question whether or not short rate diffusion models are appropriate in the South African market?

Chapter 2

Jump diffusion calculus

In this Chapter standard results from the calculus of jump diffusion processes used in pricing of financial derivatives are presented. These results will be used to explain some of the more technical aspects in Part I of this thesis. They are indispensable for the discussion of the following problems that will be considered.

- By applying Ito's lemma we find that e^{r_t} is modelled by a compound Poisson process with jumps $e^{Z_n} - 1$, if r_t is modelled by a compound Poisson process with jumps Z_n. This result will be used in Section 4.
- In the presence of jumps the quadratic variation of a jump diffusion process (Example 4 on page 16) justifies the estimation of instantaneous volatility by realised bipower variation rather than by realised power variation (Section 4.2).
- The definition of truncated stably distributed jumps (Section 6) is justified by the requirement to have jumps of finite variance.

The main result from this chapter, which impacts on the methodology used in Part I, is Girsanov's theorem and its applications. In practise, models for the short rate r_t are given relative to an assumed risk-neutral measure $\mathbb{P}^*$ (Bjork, 2004, p.327). However, the parameters in these models cannot be estimated under $\mathbb{P}^*$, but can be observed in the real world, which is represented by a measure $\mathbb{P}$. Fortunately, as a consequence of the result in Example 17 on page 25 the diffusion coefficient of a jump diffusion process does not change with a change of measure. However, the evidence from Chapters 3 and 4 suggests that a pure jump process is better suited to model the short rate. We will see in Example 19 on page 26 and Example 20 on page 29 that the jump intensity as well as the distribution of the jump sizes of a compound Poisson process changes with a change of measure. These factors have to be implemented in our pricing approach in Chapter 6.

The results in this chapter are available in existing literature, such as in Protter (2005) and Cont and Tankov (2004). However, in many cases an advanced background in pure

Mathematics is needed to read the literature on jump diffusion processes. A second objective of this Chapter is therefore to present the results in a more approachable way, such that a reader with a background in elementary probability theory or actuarial science can use the results to apply in their field of research. To reach this objective some results will be derived heuristically rather than formally.

2.1 Compound Poisson process

In this section we present standard stochastic calculus theorems such as Itô's lemma and Girsanov's theorem for compound Poisson processes, which are defined in terms of counting processes. The counting process $N = (N_t)_{0 \leq t \leq \infty}$ associated with a sequence of random jumping times $(T_n)_{n \geq 1}$ is defined by

$$N_t = \sum_{n \geq 1} \mathbb{1}_{\{t \geq T_n\}},$$

The process N_t therefore counts the number of jumps up to, and including, time t (Protter, 2005, p.12). Consequently, the range (called the state space) of N_t is a subset of natural numbers, including zero, and we have $N_{T_n} = n$. The process N_t therefore has right-continuous paths with left limits as seen in Figure 2.1.

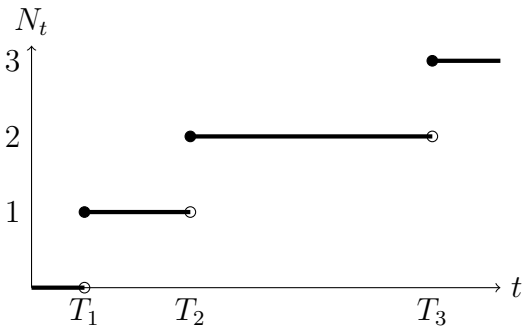

Figure 2.1: A realisation of a counting process N_t.

Let $N_{t-} = \lim_{\mathrm{d}t \to 0^+} N_{t-\mathrm{d}t}$ be a stochastic process with left continuous paths with right limits, associated with a counting process N. The jump (if any) at t is defined as

$$\Delta N_t = N_t - N_{t-}$$

A valuable property of counting processes, is that if we assume that a counting process N has independent and stationary increments, then the increments has the Poisson distribution. This makes it a natural candidate for applications in statistical modelling. Such a process is called a Poisson process.

Therefore, N is a Poisson process if

i N has increments independent of the past; that is, $N_t - N_s$ is independent of N_u, $0 \leq u \leq s < t < \infty$.

ii N has stationary increments; that is, $N_t - N_s$ has the same distribution as $N_v - N_u$, $0 \leq s < t < \infty$, $0 \leq u < v < \infty$, $t - s = v - u$.

The increments $N_t - N_s$ then has the Poisson distribution with parameter $\lambda(t-s)$. Therefore, the density function for $N_t - N_s$ is given by

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

Poisson process jumps are either equal to one or zero, while many applications require a stochastic process to have continuous valued random jump sizes. The compound Poisson process has this required property.

Definition 1. A compound Poisson process, Y_t, is defined by

$$Y_t = \sum_{n=1}^{N_t} Z_n, \tag{2.1}$$

where N_t is a Poisson process with arrival intensity λ and $(Z_n)_{n\geq 1}$ are i.i.d. random variables with probability distribution $\nu(\mathrm{d}y)$, independent of the jumping times $(T_n)_{n\geq 1}$ (Cont and Tankov, 2004, p.70).

From the definition, the jumps of Y_t at time t can be written as

$$\Delta Y_t = Z_{N_t} \Delta N_t,$$

which enables us to form a stochastic integral representation of Y (Privault, 2013, p.452)

$$Y_t = Y_0 + \int_0^t Z_{N_s} N(\mathrm{d}s) = Y_0 + \sum_{s\leq t} Z_{N_s} \Delta N_s.$$

The dynamics of Y can now be written as

$$\mathrm{d}Y_t = Z_{N_t} N(\mathrm{d}t). \tag{2.2}$$

Some properties of compound Poisson processes will be important in the next chapter. Privault (2013) on page 450 provides the proof for the moment generating function of a compound Poisson process as well as its expected value. The moment generating function can also be used to show that the compound Poisson process has independent increments.

Theorem 1. *Let Y_t be a compound Poisson process with constant intensity λ and probability distribution of the jumps $\nu(\mathrm{d}y)$.*

(a) The moment generating function of the increment $Y_T - Y_t$ is given by

$$E\left[e^{u(Y_T - Y_t)}\right] = \exp\left[\lambda(T-t)\int_{-\infty}^{\infty}(e^{yu}-1)\,\nu(\mathrm{d}y)\right], u \in \mathbb{R},$$

for any $t \in [0,T]$.

(b) The expected value and variance of Y_t are given by

$$E[Y_t] = \lambda t E[Z_1],$$

and

$$Var[Y_t] = \lambda t E\left[Z_1^2\right].$$

(c) Y_t has independent increments.

The definitions and notation we used thus far enable us to define some concepts fairly easily. However, variants of both the definition of a compound Poisson process as well as notation will be helpful in some applications. For instance, a compound Poisson process Y_t can be defined in terms of a dirac delta function as follows:

$$\begin{aligned} Y_t &= \int_0^t \int_{-\infty}^{\infty} yN(\mathrm{d}s, \mathrm{d}y) \\ &= \sum_{s \le t, \triangle Y_s \neq 0} \triangle Y_s \delta_{(s, \triangle Y_s)}((0,t], \mathrm{d}y), \end{aligned}$$

where

$$\delta_{(s,\triangle Y_s)}(\mathrm{d}t, \mathrm{d}y) = \begin{cases} 1 & \text{if} \quad s \in \mathrm{d}t \quad \text{and} \quad \triangle Y_s \in \mathrm{d}y \\ 0 & \text{if} \quad s \notin \mathrm{d}t \quad \text{or} \quad \triangle Y_s \notin \mathrm{d}y \end{cases}$$

is the dirac measure and $N(\mathrm{d}t, \mathrm{d}y)$ is a counting measure defined by

$$N(\mathrm{d}t, \mathrm{d}y) = \sum_{s \le t, \triangle Y_s \neq 0} \delta_{(s,\triangle Y_s)}(\mathrm{d}t, \mathrm{d}y) \quad \text{(see Cont and Tankov (2004, p.62))}.$$

The equation above is therefore an alternative way to express the compound Poisson process in (2.1), while the dynamics of Y_t can be written as

$$\mathrm{d}Y_t = \int_{-\infty}^{\infty} yN(\mathrm{d}t, \mathrm{d}y).$$

These alternative forms of a compound Poisson process and its dynamics will be used when we apply Itô's lemma in Section 2.3.

2.2 Quadratic covariation

The quadratic covariation of processes X and Y (denoted by $[X,Y]_t$) can be defined in terms of its dynamics

$$\mathrm{d}[X,Y]_t = \mathrm{d}X_t \mathrm{d}Y_t, \tag{2.3}$$

which implies that

$$[X,Y]_t = \int_0^t \mathrm{d}X_s \mathrm{d}Y_s.$$

If $X = Y$, then the process $[X,X]$ is called the quadratic variation process of X.

In the following two examples, the quadratic variation of two standard processes (a Brownian motion and a Poisson process) will be derived heuristically.

Example 1. (Cont and Tankov, 2004, p.266) Let W_t be a standard Brownian motion, then

$$\mathrm{d}W_t \approx W_{t+\delta t} - W_t,$$

Now, $W_{t+\delta t} - W_t$ is normally distributed with mean zero and variance δt. Therefore,

$$E\left[(W_{t+\delta t} - W_t)^2\right] = \mathrm{Var}\left[W_{t+\delta t} - W_t\right] = \delta t \to \mathrm{d}t$$

and

$$\begin{aligned}\mathrm{Var}\left[(W_{t+\delta t} - W_t)^2\right] &= E\left[(W_{t+\delta t} - W_t)^4\right] - \left\{E\left[(W_{t+\delta t} - W_t)^2\right]\right\}^2 \\ &= 3\left(\delta t\right)^2 - \left(\delta t\right)^2 = 2\left(\delta t\right)^2 \to 2\left(\mathrm{d}t\right)^2 = 0\end{aligned}$$

We therefore have

$$E\left[(\mathrm{d}W_t)^2\right] = \mathrm{d}t$$

and

$$\mathrm{Var}\left[(\mathrm{d}W_t)^2\right] = 0,$$

which implies that $\mathrm{d}W_t = \mathrm{d}t$ or $[W,W]_t = t$.

Example 2. The differential $N(\mathrm{d}t)$ has value either one or zero, which implies that $\{N(\mathrm{d}t)\}^2 = N(\mathrm{d}t)$. Therefore, the quadratic variation process of a Poisson process is given by $[N,N]_t = N_t$.

The results from the two preceding examples are used frequently and summarised in the following Itô multiplication table (Privault, 2013)

Table 2.1: Itô multiplication table

$\cdot$	$\mathrm{d}t$	$\mathrm{d}W_t$	$N(\mathrm{d}t)$
$\mathrm{d}t$	0	0	0
$\mathrm{d}W_t$	0	$\mathrm{d}t$	0
$N(\mathrm{d}t)$	0	0	$N(\mathrm{d}t)$

A few examples will now be shown to show how the Itô multiplication table can be used to derive the quadratic variation for some processes, including the compound Poisson process defined in (2.1) on page 13.

Example 3. (Cont and Tankov, 2004, p.266) Let the dynamics of a compound Poisson process be given by (2.2) on page 13. Therefore,

$$\mathrm{d}[Y,Y]_t = (\mathrm{d}Y_t)^2 = \{Z_{N_t}N(\mathrm{d}t)\}^2 = Z_{N_t}^2 N(\mathrm{d}t).$$

The quadratic variation process $[Y,Y]$ is therefore a compound Poisson process with jumps Z^2 given by

$$[Y,Y]_t = \sum_{n=1}^{N_t} Z_n^2.$$

Example 4. In this example we will derive the quadratic variation of a jump diffusion process, which shows that the quadratic variation can be divided into two components influenced separately by the diffusion coefficient and the jump sizes. Let the dynamics of a jump diffusion process be given by

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}Y_t,$$

where Y is a compound Poisson process.

The dynamics of a stochastic process can be written as the sum of a diffusion and a jump component

$$\mathrm{d}X_t = \mathrm{d}X_t^c + \Delta X_t, \tag{2.4}$$

where X_t^c is the diffusion component and ΔX_t is the jump component.

Now, from (2.4) and the Itô multiplication table (Table 2.1) we can write the dynamics of the quadratic variation of X as

$$\begin{aligned}\mathrm{d}[X,X]_t &= (\mathrm{d}X_t^c)^2 + (\triangle X_t)^2 \\ &= (\mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t)^2 + (\mathrm{d}Y_t)^2 \\ &= \sigma_t^2 \mathrm{d}t + Z_{N_t}^2 N(\mathrm{d}t).\end{aligned}$$

Therefore, the quadratic variation process of X is given by

$$[X,X]_t = \int_0^t \sigma_s^2 \mathrm{d}s + \sum_{n=1}^{N_t} Z_n^2.$$

By applying Example 4 we can find an unbiased estimate of the diffusion coefficient in the presence of jumps. From the result above an estimate of the quadratic variation may be influenced by jumps and will therefore not be an unbiased estimate for the diffusion coefficient. However, the realised bipower variation is not influenced by jumps and provides us with an unbiased estimate of the diffusion coefficient, even if jumps are observed.

Example 5. Let $X_{t_0}, X_{t_1}, ..., X_{t_n}$ be a sample of data from times $t_0, t_1, ..., t_n$. An estimate for the quadratic variation of a process X is given by the realised power variation defined by

$$[\hat{X}]_t = \sum_{i=1}^{n} |X_{t_i} - X_{t_{i-1}}|^2.$$

Therefore, the realised power variation is a discretisation of (2.3) on page 15. Barndorff-Nielsen and Shepard (2004) shows that

$$[\hat{X}]_t \xrightarrow{\mathrm{P}} [X, X]_t,$$

where $\xrightarrow{\mathrm{P}}$ denotes convergence in probability.

In the case of a jump diffusion process X given by

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \mathrm{d}Y_t,$$

the quadratic variation process for X is given by (see Example 4)

$$[X, X]_t = \int_0^t \sigma_s^2 \mathrm{d}s + \sum_{n=1}^{N_t} Z_n^2.$$

The realised power variation is therefore a biased estimate of the instantaneous volatility σ_t if the sample has infrequent jumps.

However, Barndorff-Nielsen and Shepard (2004) defines the realised bipower variation as

$$[\tilde{X}]_t = \sum_{i=1}^{n-1} |X_{t_i} - X_{t_{i-1}}||X_{t_{i+1}} - X_{t_i}|$$

and shows that for the jump diffusion process X defined above

$$[\tilde{X}]_t \xrightarrow{\mathrm{P}} \int_0^t \sigma_s^2 \mathrm{d}s.$$

Therefore, the realised bipower variation can be used to estimate the instantaneous volatility σ_t, and is not influenced by infrequent jumps.

This result will be used in Section 4.2 to estimate the diffusion coefficient of a jump diffusion process.

2.3 Itô's lemma

With the standard Black-Scholes model the dynamics of an asset price X_t is given by

$$\mathrm{d}X_t = \mu X_t \mathrm{d}t + \sigma X_t \mathrm{d}W_t,$$

which is called the geometric Brownian motion. To solve the geometric Brownian motion, the dynamics of the stochastic process $\log X_t$ is written as

$$\mathrm{d}\log X_t = \mu \mathrm{d}t + \sigma \mathrm{d}W_t.$$

By integrating both sides $\log X_t$ can be solved and statistical properties of X_t, such as the mean and variance can be derived.

Itô's lemma enables us to transform the dynamics of a stochastic variable X_t to $f(t, X_t)$, where the function $f(t, x)$ is a function of two variables, with continuous first order partial derivative relative to variable t and infinitely differentiable relative to variable x. Therefore, $f \in \mathcal{C}^{1,\infty}$.

For instance, in Example 11 on page 22 we show that if X_t is a geometric Poisson process, with dynamics

$$\mathrm{d}X_t = X_{t-} \int_{-\infty}^{\infty} y N(\mathrm{d}t, \mathrm{d}y),$$

then the dynamics of $\log X_t$ is given by

$$\mathrm{d}\log X_t = \int_{-\infty}^{\infty} \log(1+y)\, N(\mathrm{d}t, \mathrm{d}y).$$

A form of Itô's lemma which can be used to apply directly to jump diffusion processes can be found in Cont and Tankov (2004) on page 275:

Theorem 2. *Assume that the process X has a stochastic differential given by*

$$\mathrm{d}X_t = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}W_t + \int_{-\infty}^{\infty} J(t, y) N(\mathrm{d}t, \mathrm{d}y),$$

where μ_t and σ_t are continuous adapted processes with

$$E\left[\int_0^T \sigma_t^2 \mathrm{d}t\right] < \infty.$$

Then, any $\mathcal{C}^{1,2}$-function f has the following stochastic differential form

$$\begin{aligned}\mathrm{d}f(t, X_t) = &\left\{\frac{\partial f}{\partial t} + \mu_t \frac{\partial f}{\partial x} + \frac{1}{2}\sigma_t^2 \frac{\partial^2 f}{\partial x^2}\right\} \mathrm{d}t \\ &+ \sigma_t \frac{\partial f}{\partial x} \mathrm{d}W_t + \left[f(t, X_t) - f(t, X_{t-})\right].\end{aligned}$$

Now, we can write $\mathrm{d}f(t, X_t)$ as

$$\mathrm{d}f(t, X_t) = \mathrm{d}f^c(t, X_t) + \triangle f(t, X_t),$$

where

$$\triangle f(t, X_t) = \left[f(t, X_t) - f(t, X_{t-})\right].$$

Before we can apply Itô's lemma to various problems, we need the following result for $\triangle f(t, X_t)$.

Corollary 1. *Assume X is defined as in Theorem 2. Then, if f is a $\mathcal{C}^{1,\infty}$-function the jump part of $f(t, X_t)$ can be written as*

$$\triangle f(t, X_t) = \sum_{k=1}^{\infty} \left(\int_{-\infty}^{\infty} J^k(t, y) N(\mathrm{d}t, \mathrm{d}y) \right) \frac{1}{k!} \frac{\partial^k f(t, X_{t-})}{\partial x^k}$$

Proof. Applying Taylor's theorem to f we have

$$\begin{aligned} f(t, X_t) - f(t, X_{t-}) &= \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k f(t, X_{t-})}{\partial x^k} (\triangle X_t)^k \\ &= \sum_{k=1}^{\infty} \frac{1}{k!} \frac{\partial^k f(t, X_{t-})}{\partial x^k} \left(\int_{-\infty}^{\infty} J(t, y) N(\mathrm{d}t, \mathrm{d}y) \right)^k. \end{aligned}$$

We will show by induction that

$$\left(\int_{-\infty}^{\infty} J(t, y) N(\mathrm{d}t, \mathrm{d}y) \right)^k = \int_{-\infty}^{\infty} J^k(t, y) N(\mathrm{d}t, \mathrm{d}y), \quad \forall k \geq 2.$$

From the definition of a counting measure $N(\mathrm{d}t, \mathrm{d}y)$, and applying the Itô multiplication table (Table 2.1 on page 15) we have for $k = 2$

$$\begin{aligned} &\left(\int_{-\infty}^{\infty} J(t, y) N(\mathrm{d}t, \mathrm{d}y) \right)^2 \\ &= \int_{-\infty}^{\infty} J(t, y_1) N(\mathrm{d}t, \mathrm{d}y_1) \int_{-\infty}^{\infty} J(t, y_2) N(\mathrm{d}t, \mathrm{d}y_2) \\ &= \iint\limits_{\{y_1 = y_2\}} J(t, y_1) J(t, y_2) N(\mathrm{d}t, \mathrm{d}y_1) N(\mathrm{d}t, \mathrm{d}y_2) \\ &\qquad + \iint\limits_{\{y_1 \neq y_2\}} J(t, y_1) J(t, y_2) N(\mathrm{d}t, \mathrm{d}y_1) N(\mathrm{d}t, \mathrm{d}y_2) \\ &= \int_{-\infty}^{\infty} J^2(t, y) \left[N(\mathrm{d}t, \mathrm{d}y) \right]^2 \\ &= \int_{-\infty}^{\infty} J^2(t, y) N(\mathrm{d}t, \mathrm{d}y). \end{aligned}$$

Assume the result is true for $k = n$:

$$\left(\int_{-\infty}^{\infty} J(t, y) N(\mathrm{d}t, \mathrm{d}y) \right)^n = \int_{-\infty}^{\infty} J^n(t, y) N(\mathrm{d}t, \mathrm{d}y).$$

Now, for $k = n + 1$ we have

$$\begin{aligned}
&\left(\int_{-\infty}^{\infty} J(t,y)N(\mathrm{d}t,\mathrm{d}y)\right)^{n+1} \\
&= \left(\int_{-\infty}^{\infty} J(t,y)N(\mathrm{d}t,\mathrm{d}y)\right)^{n} \int_{-\infty}^{\infty} J(t,y_2)N(\mathrm{d}t,\mathrm{d}y_2) \\
&= \int_{-\infty}^{\infty} J^n(t,y)N(\mathrm{d}t,\mathrm{d}y) \int_{-\infty}^{\infty} J(t,y_2)N(\mathrm{d}t,\mathrm{d}y_2) \\
&= \iint\limits_{\{y=y_2\}} J^n(t,y)J(t,y_2)N(\mathrm{d}t,\mathrm{d}y)N(\mathrm{d}t,\mathrm{d}y_2) \\
&\qquad + \iint\limits_{\{y\neq y_2\}} J^n(t,y)J(t,y_2)N(\mathrm{d}t,\mathrm{d}y)N(\mathrm{d}t,\mathrm{d}y_2) \\
&= \int_{-\infty}^{\infty} J^{n+1}(t,y)N(\mathrm{d}t,\mathrm{d}y).
\end{aligned}$$

Therefore, the result is true for $k \geq 2$. □

Some examples applying Theorem 2 will now de discussed. In Example 6 to Example 8 some results from applying Theorem 2 will be compared to results obtained by algebraic manipulation. In Example 9 we apply Theorem 2 to the dynamics of an asset assumed to be a combination of a geometric Brownian motion and a geometric Poisson process. Example 10 implies that if the short rate r_t is modelled by a compound Poisson process with jumps Z_n, then e^{r_t} has jumps equal to $r_{t-}\left(e_n^Z - 1\right)$. This result will be used in Section 4.1.

Example 6. Let N_t be a Poisson process. The increment $\triangle\left(N_t^2\right)$ will be algebraically derived after which a comparison will be done with the result obtained from Theorem 2.

$$\begin{aligned}
\triangle\left(N_t^2\right) &= N_t^2 - N_{t-}^2 \\
&= \triangle N_t(N_t + N_{t-}) \\
&= \triangle N_t(\triangle N_t + 2N_{t-}) \\
&= (\triangle N_t)^2 + 2N_{t-}\triangle N_t \\
&= (1 + 2N_{t-})\triangle N_t.
\end{aligned}$$

Now, applying Theorem 2 we let $f(x) = x^2$ then

$$\triangle f(N_t) = f'(N_{t-})\triangle N_t + \frac{1}{2}f''(N_{t-})(\triangle N_t)^2 + 0$$

as $f^{(n)}(x) = 0$ for $n \geq 3$. Using the property that $(\triangle N_t)^2 = \triangle N_t$ we get the same result from Itô's formula as from first principles.

Example 7. Using the same process as in the previous example $\triangle \log N_t$ will be evaluated by the same method. Therefore, using a series expansion of the logarithmic function and

the same power property of Poisson processes stated in the previous example, we get the following result:

$$\begin{aligned}\triangle \log N_t &= \log \tfrac{N_t}{N_{t-}} \\ &= \log\left(1+\tfrac{\triangle N_t}{N_{t-}}\right) \\ &= \textstyle\sum_{k\geq 1}(-1)^{k+1}\frac{(\triangle N_t)^k}{k(N_{t-})^k} \\ &= \triangle N_t \log\left(1+\tfrac{1}{N_{t-}}\right).\end{aligned}$$

By applying Theorem 2 the following result is obtained when $f(x) = \log(x)$.

$$\begin{aligned}\triangle f(N_t) &= \tfrac{\triangle N_t}{N_{t-}} - \tfrac{1}{2}\left(\tfrac{\triangle N_t}{N_{t-}}\right)^2 + ... \\ &= \triangle N_t\left[\tfrac{1}{N_{t-}} - \tfrac{1}{2}\left(\tfrac{1}{N_{t-}}\right)^2 + ...\right] \\ &= \triangle N_t \log\left(1+\tfrac{1}{N_{t-}}\right).\end{aligned}$$

Example 8. Next we evaluate $\triangle\left(e^{i\lambda N_t}\right)$.

$$\begin{aligned}\triangle\left(e^{i\lambda N_t}\right) &= e^{i\lambda N_t} - e^{i\lambda N_{t-}} \\ &= e^{i\lambda(\triangle N_t + N_{t-})} - e^{i\lambda N_{t-}} \\ &= e^{i\lambda N_{t-}}\left(e^{i\lambda \triangle N_t} - 1\right) \\ &= e^{i\lambda N_{t-}}\textstyle\sum_{j\geq 1}\frac{(i\lambda)^j}{j!}(\triangle N_t)^j \\ &= e^{i\lambda N_{t-}}(\triangle N_t)\left(\textstyle\sum_{j\geq 1}\frac{(i\lambda)^j}{j!}\right) \\ &= e^{i\lambda N_{t-}}\triangle N_t(e^{i\lambda}-1).\end{aligned}$$

By applying Theorem 2 the following result is obtained when $f(x) = e^{i\lambda x}$.

$$\begin{aligned}\triangle f(N_t) &= f^{(1)}(N_{t-})\triangle N_t + \tfrac{1}{2}f^{(2)}(N_{t-})(\triangle N_t)^2 + \tfrac{1}{6}f^{(3)}(N_{t-})(\triangle N_t)^3 + ... \\ &= e^{i\lambda N_{t-}}(i\lambda)(\triangle N_t) + \tfrac{1}{2}e^{i\lambda N_{t-}}(i\lambda)^2(\triangle N_t)^2 + \tfrac{1}{6}e^{i\lambda N_{t-}}(i\lambda)^3(\triangle N_t)^3 + ... \\ &= e^{i\lambda N_{t-}}\triangle N_t(e^{i\lambda}-1).\end{aligned}$$

Example 9. In the paper of Cheang and Chiarella (2011), the return dynamics of the price of an underlying asset, S_t, is assumed to be given by a jump diffusion process given in formula (3) on page 2.

$$\frac{\mathrm{d}S_t}{S_{t-}} = (\mu - \lambda\kappa)\,\mathrm{d}t + \sigma \mathrm{d}W_t + \left(e^J - 1\right)N(\mathrm{d}t).$$

If the underlying asset pays a continuous dividend at the rate q per unit time. The yield process of the asset is defined as $S_t e^{qt}$. The discounted yield process is defined by

$$e^{-rt}S_t e^{qt} = S_t e^{(q-r)t}, \tag{2.5}$$

where r is the constant risk free short rate.

Now, we apply Theorem 2 to (2.5). Let, $f(t,x) = xe^{(q-r)t}$, then $\frac{\partial f}{\partial x} = e^{(q-r)t}$, $\frac{\partial^k f}{\partial x^k} = 0$ and $\frac{\partial f}{\partial t} = xe^{(q-r)t}(q-r)$. Therefore,

$$\begin{aligned}
\mathrm{d}f(t,S_t) &= \mathrm{d}\left(S_t e^{(q-r)t}\right) \\
&= \{S_{t-}e^{(q-r)t}(q-r) + S_{t-}\left(\mu - \lambda\kappa\right)e^{(q-r)t}(q-r)\}\mathrm{d}t \\
&\qquad + S_{t-}\sigma e^{(q-r)t}\mathrm{d}W_t + S_{t-}e^{(q-r)t}\left(e^J - 1\right)N(\mathrm{d}t) \\
&= S_{t-}e^{(q-r)t}\left[\{q - r + \mu - \lambda\kappa\}\mathrm{d}t + \sigma\mathrm{d}W_t + \left(e^J - 1\right)N(\mathrm{d}t)\right].
\end{aligned}$$

Example 10. Let X_t be modelled by a compound Poisson process with jump sizes given by i.i.d. random variables $(Z_n)_{n\geq 0}$. The dynamics of X_t is then given by

$$\mathrm{d}X_t = \int_{-\infty}^{\infty} yN(\mathrm{d}t, \mathrm{d}y).$$

Now, we want to find the dynamics of $Y_t = e^{X_t}$. Applying Theorem 2 to $f(x) = e^x$, we get

$$\begin{aligned}
\mathrm{d}Y_t &= \sum_{k=1}^{\infty}\frac{1}{k!}e^{X_{t-}}\int_{-\infty}^{\infty} y^k N(\mathrm{d}t, \mathrm{d}y) \\
&= \int_{-\infty}^{\infty} Y_{t-}\left(\sum_{k=1}^{\infty}\frac{1}{k!}y^k\right)N(\mathrm{d}t, \mathrm{d}y) \\
&= \int_{-\infty}^{\infty} Y_{t-}\left(e^y - 1\right)N(\mathrm{d}t, \mathrm{d}y).
\end{aligned}$$

The resulting process Y_t therefore has jumps equal to $Y_{t-}\left(e^{Z_n} - 1\right)$.

Example 11. Let the dynamics of X_t be given by

$$\mathrm{d}X_t = X_{t-}\int_{-\infty}^{\infty} yN(\mathrm{d}t, \mathrm{d}y).$$

Now, we want to find the dynamics of $Y_t = \log(X_t)$. Applying Theorem 2 to $f(x) = \log x$, we get

$$\begin{aligned}
\mathrm{d}Y_t &= \sum_{k=1}^{\infty}(-1)^{k+1}\frac{(k-1)!}{k!}\frac{1}{X_{t-}^k}\int_{-\infty}^{\infty}\left(X_{t-}y\right)^k N(\mathrm{d}t, \mathrm{d}y) \\
&= \int_{-\infty}^{\infty}\left(\sum_{k=1}^{\infty}(-1)^{k+1}\frac{y^k}{k}\right)N(\mathrm{d}t, \mathrm{d}y) \\
&= \int_{-\infty}^{\infty}\log\left(1+y\right)N(\mathrm{d}t, \mathrm{d}y).
\end{aligned}$$

The resulting process Y_t therefore has jumps equal to $\log\left(1 + Z_n\right)$, if $|Z_n| < 1$.

2.4 Girsanov's theorem

In this thesis we will apply jump diffusion calculus to the pricing of financial derivatives. In general, if S_t is the price of an asset at time t, then, assuming a constant risk-free rate r, we have

$$S_t = e^{-r(T-t)} E^* \{S_T \mid \mathcal{F}_t\}, \tag{2.6}$$

where the conditional expectation is taken under the measure $\mathbb{P}^*$ and $\mathcal{F}_t$ is a collection of subsets of the sample space, Ω. which represents the history up to time t and $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$. Therefore, under the measure $\mathbb{P}^*$, the asset S_t has the same return (in expectation) as a risk-free asset. The measure $\mathbb{P}^*$ is therefore called a risk-neutral measure.

If the dynamics of S_t is observed under a measure $\mathbb{P}$, the question arises how to transform from $\mathbb{P}$ to $\mathbb{P}^*$ in such a way that the relation in (2.6) still holds. To answer this question martingales, conditional quadratic covariation and Girsanov's theorem need to be discussed.

A stochastic process M is a martingale under a measure $\mathbb{P}$ if

$$E\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\} = 0,$$

where E is the conditional expectation taken under the measure $\mathbb{P}$. In the following examples we discuss a couple of standard martingales.

Example 12. If W is a standard Brownian motion then W is a martingale as it has independent increments

$$E\{\mathrm{d}W_t \mid \mathcal{F}_{t-}\} = E\{\mathrm{d}W_t\} = 0.$$

Example 13. If N is a Poisson process with intensity λ then $M_t = N_t - \lambda t$ is a martingale as

$$E\{N(\mathrm{d}t) \mid \mathcal{F}_{t-}\} = E\{N(\mathrm{d}t)\} = \lambda \mathrm{d}t.$$

The process M is called a compensated Poisson process

Example 14. If Y is a compound Poisson process with intensity λ then $M_t = Y_t - \lambda E(Z_1)t$ is a martingale. From Theorem 1 (b) and (c) on page 14 we have

$$E\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\} = E\{\mathrm{d}Y_t\} = \lambda E(Z_1)\mathrm{d}t.$$

The process M is called a compensated compound Poisson process.

The conditional quadratic covariation of the processes X and Y can be defined in terms of its dynamics

$$\mathrm{d}\langle X, Y\rangle_t = E\{\mathrm{d}[X, Y]_t \mid \mathcal{F}_{t-}\}. \tag{2.7}$$

If $X = Y$ then $\langle X, Y\rangle_t$ is called the conditional quadratic variation. In the next two examples, it will be shown that the conditional quadratic variation of a Brownian motion and that of a compound Poisson process are both equal to their variance.

Example 15. Let W_t be a Brownian motion, then the conditional quadratic variation of W is equal to its quadratic variation:

$$\mathrm{d}\left\langle W, W\right\rangle_t = E\left\{\mathrm{d}[W,W]_t \mid \mathcal{F}_{t-}\right\} = E\left\{\mathrm{d}t \mid \mathcal{F}_{t-}\right\} = \mathrm{d}t.$$

Example 16. Let Y be a compound Poisson process. The quadratic variation of Y is a compound Poisson process with jumps Z_n^2 (See Example 3 on page 16) and therefore according to Theorem 1 has independent increments. Therefore,

$$\mathrm{d}\left\langle Y, Y\right\rangle_t = E\left\{\mathrm{d}[Y,Y]_t \mid \mathcal{F}_{t-}\right\} = E\left\{Z_{N_t}^2 N(\mathrm{d}t)\right\} = \lambda E\left[Z_1^2\right]\mathrm{d}t$$

From Example 15 we know that the conditional quadratic variation of a Brownian motion is equal to its quadratic variation. This is true for all diffusion processes. However, the conditional quadratic variation for a jump process may differ from its quadratic variation.

We will now turn to Girsanov's theorem, which will enable us to derive some important results used in Chapter 6. Let $\mathbb{P}^*$ and $\mathbb{P}$ be equivalent measures (See definition in Protter (2005) on page 133) and let L be a random variable with $\mathrm{d}\mathbb{P}^* = L\mathrm{d}\mathbb{P}$ and $E(L) = 1$. Now, let

$$L_t = E\left\{\frac{d\mathbb{P}^*}{d\mathbb{P}} \mid \mathcal{F}_t\right\}.$$

L is then a martingale under $\mathbb{P}$. The following lemma from Kuo (2005) on page 141 will be used to prove the predictable version of Girsanov's theorem (Protter, 2005, p.135)

Lemma 1. *Let E^* and E denote the expectation under equivalent measures $\mathbb{P}^*$ and $\mathbb{P}$ respectively. Let X be a stochastic process and let L be defined by $\mathrm{d}\mathbb{P}^* = L\mathrm{d}\mathbb{P}$. Then,*

$$E^*\left\{\mathrm{d}X_t \mid \mathcal{F}_{t-}\right\} = \frac{E\left\{\mathrm{d}X_t L_t \mid \mathcal{F}_{t-}\right\}}{E\left\{L_t \mid \mathcal{F}_{t-}\right\}}.$$

Theorem 3. ***(Girsanov Theorem)*** *Let E^* and E denote the expectation under equivalent measures $\mathbb{P}^*$ and $\mathbb{P}$ respectively. Let M be a martingale under the measure $\mathbb{P}$ and let L be defined by $\mathrm{d}\mathbb{P}^* = L\mathrm{d}\mathbb{P}$. Then,*

$$E^*\left\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\right\} = \frac{1}{L_{t-}}\mathrm{d}\langle M, L\rangle_t.$$

Therefore, the dynamics of a martingale M^* under $\mathbb{P}^*$ is given by

$$\mathrm{d}M_t^* = \mathrm{d}M_t - \frac{1}{L_{t-}}\mathrm{d}\langle M, L\rangle_t.$$

Proof. From Lemma 1 we have

$$\begin{aligned} E^*\left\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\right\} &= \frac{E\left\{\mathrm{d}M_t L_t \mid \mathcal{F}_{t-}\right\}}{E\left\{L_t \mid \mathcal{F}_{t-}\right\}} \\ &= \frac{1}{L_{t-}} E\left\{\mathrm{d}M_t \left(L_{t-} + L_t - L_{t-}\right) \mid \mathcal{F}_{t-}\right\} \\ &= E\left\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\right\} + \frac{1}{L_{t-}} E\left\{\mathrm{d}M_t \mathrm{d}L_t \mid \mathcal{F}_{t-}\right\} \\ &= \frac{1}{L_{t-}} \mathrm{d}\langle M, L\rangle_t \end{aligned}$$

□

If L and M are diffusion processes then the Girsanov's result becomes

$$E^*\left\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\right\} = \frac{1}{L_t} \mathrm{d}[M, L]_t$$

as $L_{t-} = L_t$ and $\mathrm{d}\langle M, L\rangle_t = \mathrm{d}[M, L]_t$.

Example 17. In this example we will show that a standard Brownian motion has added drift with a change of measure. Let the dynamics of L_t be given by

$$\mathrm{d}L_t = L_t h_t \mathrm{d}W_t,$$

where W is a $\mathbb{P}$ standard Brownian motion. Using Theorem 3 we can write W^* as a $\mathbb{P}^*$ martingale where

$$\mathrm{d}W_t^* = \mathrm{d}W_t - \frac{1}{L_t} \mathrm{d}[L, W]_t.$$

Now, by applying Itô's multiplication table (Table 2.1 on page 15) we get

$$\mathrm{d}[L, W]_t = \mathrm{d}L_t \mathrm{d}W_t = L_t h_t \left(\mathrm{d}W_t\right)^2 = L_t h_t \mathrm{d}t.$$

Therefore,

$$\mathrm{d}W_t^* = \mathrm{d}W_t - h_t \mathrm{d}t.$$

An important conclusion from the example above is that the drift of W_t changes from zero under measure $\mathbb{P}$ to $h_t \mathrm{d}t$ under measure $\mathbb{P}^*$. However, the variance of W_t stays the same. This has an important consequence in pricing of interest rate derivatives as pricing occurs under a risk-neutral measure $\mathbb{P}^*$, but market observations are made under a measure $\mathbb{P}$. With a change of measure having no influence on the diffusion coefficient, calibration to market data is possible.

Example 18. In this example we will show that the intensity of a Poisson process changes with a change of measure. Let the dynamics of L_t be given by

$$\mathrm{d}L_t = L_{t-} c \, \mathrm{d}M_t$$

where $M_t = N_t - \lambda t$ is a compensated Poisson process under measure $\mathbb{P}$ and $c > -1$ a constant. Using Theorem 3 we can write

$$\begin{aligned} E^*\left\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\right\} &= \frac{1}{L_{t-}}\mathrm{d}\langle M, L\rangle_t \\ &= \frac{1}{L_{t-}}E\left\{\mathrm{d}M_t\mathrm{d}L_t \mid \mathcal{F}_{t-}\right\} \\ &= \frac{1}{L_{t-}}E\left\{L_{t-}c\left(\mathrm{d}M_t\right)^2 \mid \mathcal{F}_{t-}\right\} \\ &= cE\left\{N(\mathrm{d}t) \mid \mathcal{F}_{t-}\right\} \\ &= c\lambda\mathrm{d}t. \end{aligned}$$

Therefore, we have that

$$E^*\left\{N(\mathrm{d}t) \mid \mathcal{F}_{t-}\right\} = \lambda\mathrm{d}t + c\lambda\mathrm{d}t = \lambda(1+c)\mathrm{d}t,$$

which show that the intensity of the Poisson process N under measure $\mathbb{P}^*$ is given by

$$\tilde{\lambda} = \lambda(1+c).$$

Example 19. In this example we will show that the intensity as well as the jump distribution of a compound Poisson process changes with a change of measure. Let Y_t be a compound Poisson process with intensity λ and jump distribution $\nu(\mathrm{d}y)$. Consider another jump distribution $\tilde{\nu}(\mathrm{d}y)$ and let (Privault, 2013, p. 467)

$$x = \frac{\tilde{\lambda}}{\lambda}\frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\nu}(x) - 1.$$

Now, let

$$\mathrm{d}L_t = L_{t-}\mathrm{d}M_t,$$

where

$$M_t = Y_t - \lambda E_\nu\left[Z_1\right]t$$

is a compensated compound Poisson process under $\mathbb{P}$, where

$$\mathrm{d}Y_t = Z_{N_t}N(\mathrm{d}t).$$

Using Theorem 3 we can write

$$\begin{aligned} E^*\left\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\right\} &= \frac{1}{L_{t-}}\mathrm{d}\langle M, L\rangle_t \\ &= \frac{1}{L_{t-}}E\left\{\mathrm{d}M_t\mathrm{d}L_t \mid \mathcal{F}_{t-}\right\} \\ &= \frac{1}{L_{t-}}E\left\{L_{t-}\left(\mathrm{d}M_t\right)^2 \mid \mathcal{F}_{t-}\right\} \\ &= E\left\{\left(\mathrm{d}Y_t\right)^2 \mid \mathcal{F}_{t-}\right\} \\ &= \lambda E_\nu\left[Z_1^2\right]\mathrm{d}t. \end{aligned}$$

Now,

$$E^* \{ \mathrm{d}Y_t \mid \mathcal{F}_{t-} \} = \lambda E_\nu [Z_1] \, \mathrm{d}t + \lambda E_\nu \left[Z_1^2 \right] \mathrm{d}t.$$

It can be shown that

$$E_\nu [Z_1] = \frac{\tilde{\lambda}}{\lambda} - 1 = c$$

and

$$E_\nu \left[Z_1^2 \right] = (1 + c) E_{\tilde{\nu}} [Z_1] - c,$$

where E_ν and $E_{\tilde{\nu}}$ are the expected values taken w.r.t. the jump distributions ν and $\tilde{\nu}$. Therefore,

$$\begin{aligned} E^* \{ \mathrm{d}Y_t \mid \mathcal{F}_{t-} \} &= \lambda c \mathrm{d}t + \lambda \{ (1 + c) E_{\tilde{\nu}} [Z_1] - c \} \, \mathrm{d}t \\ &= \lambda (1 + c) E_{\tilde{\nu}} [Z_1] \\ &= \tilde{\lambda} E_{\tilde{\nu}} [Z_1] , \end{aligned}$$

which implies that the compensated compound Poisson process under measure $\mathbb{P}^*$ is given by

$$M_t^* = Y_t - \tilde{\lambda} E_{\tilde{\nu}} [Z_1] \, t.$$

Example 17 shows that the drift changes with a change of measure when a diffusion process is used, but the volatility remains unchanged. Example 18 shows that with a Poisson process the intensity changes, while Example 19 shows that when a compound Poisson process is used the intensity as well as the jump distribution changes with a change of measure. Therefore, if a compound Poisson process is used to model the short rate, then all of the distribution parameters changes with a change of measure. Calibration to market data is therefore more difficult when the underlying process is a pure jump process rather than a diffusion process (Bjork, 2004, p.327). This important fact will determine our pricing approach in Chapter 6.

2.5 Nonstationary compound Poisson processes

In our application in Part *I* we find that intensity of the jumps depends on the level of the interest rate. This change of intensity from constant to a deterministic function of the stochastic process level makes the compound Poisson process nonstationary. Now, let a marked point process, Y_t, be defined by

$$Y_t = \int_0^t \int_{-\infty}^{\infty} y N(\mathrm{d}s, \mathrm{d}y),$$

where $N(\mathrm{d}s, \mathrm{d}y)$ is the random counting measure associated with Y_t. In this case N_t can be defined as

$$N_t = \int_0^t \int_{-\infty}^{\infty} N(\mathrm{d}s, \mathrm{d}y),$$

where N is a Poisson process with intensity $\lambda(Y_{t-})$. We can then write the dynamics of Y_t, with similar notation used as in previous sections, by

$$\mathrm{d}Y_t = Z_{N_t} N(\mathrm{d}t) = \int_{-\infty}^{\infty} y N(\mathrm{d}t, \mathrm{d}y),$$

where $(Z_n)_{n\geq 1}$ are i.i.d. random variables with probability distribution $\nu(\mathrm{d}y)$.

Now, using similar notation as Giesecke and Zhu (2013), the process given by

$$m_t = \int_0^t \int_{-\infty}^{\infty} [N(\mathrm{d}s, \mathrm{d}y) - A(\mathrm{d}t)\nu(\mathrm{d}y)]$$

is a martingale associated with the counting process N_t, where the nondecreasing process A_t is given by

$$A_t = \int_0^t \lambda(Y_{t-})\mathrm{d}s.$$

A martingale associated with the marked point process Y_t is given by

$$M_t = \int_0^t \int_{-\infty}^{\infty} y \left[N(\mathrm{d}s, \mathrm{d}y) - A(\mathrm{d}s)\nu(\mathrm{d}y)\right].$$

Therefore, the conditional expectation of $\mathrm{d}Y_t$ can be derived by

$$\begin{aligned}
0 = E\left\{\mathrm{d}M_t \mid \mathcal{F}_{t-}\right\} &= E\left\{\int_{-\infty}^{\infty} y\left[N(\mathrm{d}t, \mathrm{d}y) - A(\mathrm{d}t)\nu(\mathrm{d}y)\right] \mid \mathcal{F}_{t-}\right\} \\
&= E\left\{\mathrm{d}Y_t - A(\mathrm{d}t)\int_{-\infty}^{\infty} y\nu(\mathrm{d}y) \mid \mathcal{F}_{t-}\right\} \\
&= E\left\{\mathrm{d}Y_t - A(\mathrm{d}t)E\left[Z_1\right] \mid \mathcal{F}_{t-}\right\} \\
&= E\left\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\right\} - E\left[Z_1\right] E\left\{A(\mathrm{d}t) \mid \mathcal{F}_{t-}\right\},
\end{aligned}$$

which implies that

$$\begin{aligned}
E\left\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\right\} &= E\left[Z_1\right] E\left\{A(\mathrm{d}t) \mid \mathcal{F}_{t-}\right\} \\
&= E\left[Z_1\right] \lambda(Y_{t-})\mathrm{d}t. \qquad (2.8)
\end{aligned}$$

The quadratic variation of Y_t has the same form as in Section 2.2, which implies that the dynamics of the conditional quadratic variation can be written as

$$\mathrm{d}\left\langle Y, Y\right\rangle_t = E\left\{(\mathrm{d}Y_t)^2 \mid \mathcal{F}_{t-}\right\} = E\left[Z_1^2\right]\lambda(Y_{t-})\mathrm{d}t.$$

However, to find the form in which the conditional quadratic variation or conditional expectation of Y_t is written the following result is needed.

Theorem 4. *Let X_t be a stochastic process with $X_0 = 0$ and $E\left\{\mathrm{d}X_t \mid \mathcal{F}_{t-}\right\} = \mathrm{d}A_t$, then for $u < t$*

$$E\left\{X_t \mid \mathcal{F}_u\right\} = Y_u + \int_u^t E\left\{\mathrm{d}A_t \mid \mathcal{F}_u\right\}$$

Proof. Applying the Tower law for conditional expectation we find the following result

$$\begin{aligned} E\left\{X_t \mid \mathcal{F}_u\right\} &= E\left\{\int_0^t \mathrm{d}X_s \mid \mathcal{F}_u\right\} \\ &= E\left\{\int_0^u \mathrm{d}X_t + \int_u^t \mathrm{d}X_s \mid \mathcal{F}_u\right\} \\ &= X_u + E\left\{\int_u^t \mathrm{d}X_s \mid \mathcal{F}_u\right\} \\ &= X_u + \int_u^t E\left\{\mathrm{d}X_s \mid \mathcal{F}_u\right\} \\ &= X_u + \int_u^t E\left\{E\left[\mathrm{d}X_s \mid \mathcal{F}_{t-}\right] \mid \mathcal{F}_u\right\} \\ &= X_u + \int_u^t E\left\{\mathrm{d}A_s| \ \mathcal{F}_u\right\} \end{aligned}$$

□

In the stationary case, where the conditional estimation of the compound Poisson process Y_t is given by

$$E\left\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\right\} = \lambda E\left[Z_1\right]\mathrm{d}t,$$

then we find that

$$E\left\{Y_t \mid \mathcal{F}_u\right\} = Y_u + \lambda E\left[Z_1\right](t-u).$$

However, in the marked point process case where

$$E\left\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\right\} = \lambda(Y_{t-})E\left[Z_1\right]\mathrm{d}t,$$

then we have that

$$E\left\{Y_t \mid \mathcal{F}_u\right\} = Y_u + E\left[Z_1\right]\int_u^t E\left\{\lambda(Y_{s-})| \ \mathcal{F}_u\right\}\mathrm{d}s.$$

Therefore, the conditional quadratic variation of the marked point process Y_t, where $Y_0 = 0$, is given by

$$\langle Y, Y\rangle_t = E\left[Z_1^2\right]\int_0^t E\left\{\lambda(Y_{s-})| \ \mathcal{F}_0\right\}\mathrm{d}s.$$

To apply Girsanov's theorem to nonstationary compound Poisson processesis similar than the applications we did in the previous section.

Example 20. Let Y_t be a compound Poisson process with intensity $\lambda(Y_{t-})$ and jump distribution $\nu(\mathrm{d}y)$. Consider another jump distribution $\tilde{\nu}(\mathrm{d}y)$ and let

$$x = (c+1)\frac{\mathrm{d}\tilde{\nu}}{\mathrm{d}\nu}(x) - 1, \tag{2.9}$$

where

$$\tilde{\lambda}(Y_{t-}) = (c+1)\lambda(Y_{t-})$$

Now, let

$$\mathrm{d}L_t = L_{t-}\mathrm{d}M_t,$$

where

$$M_t = Y_t - E_\nu\left[Z_1\right]\int_0^t \lambda(Y_{s-})\,\mathrm{d}s$$

is a martingale under measure $\mathbb{P}$.

Using Theorem 3 and following the same steps as in Example 19 we find that

$$E^*\left\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\right\} = \left(E_\nu\left[Z_1\right] + E_\nu\left[Z_1^2\right]\right)\lambda(Y_{t-})\mathrm{d}t.$$

Again, we have

$$E_\nu\left[Z_1\right] = c$$

and

$$E_\nu\left[Z_1^2\right] = (1+c)E_{\tilde{\nu}}\left[Z_1\right] - c,$$

where E_ν and $E_{\tilde{\nu}}$ are the expected values taken w.r.t. the jump distributions ν and $\tilde{\nu}$. Therefore,

$$E^*\left\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\right\} = \tilde{\lambda}(Y_{t-})E_{\tilde{\nu}}\left[Z_1\right],$$

which implies that a martingale under the measure $\mathbb{P}^*$ is given by

$$M_t^* = Y_t - E_{\tilde{\nu}}\left[Z_1\right]\int_0^t \tilde{\lambda}(Y_{s-})\,\mathrm{d}s.$$

In general Girsanov's theorem applied to marked point processes can be found in Bjork et al. (1997). The same conclusions we made in Section 2.4 can be made in the nonstationary case where the distribution of the jumps and the jump intensity changes with a change of measure.

In our final example of this chapter we will derive an expression for the drift of a Vasiček type pure jump short rate model. Importantly, the model is mean reverting. If the jumps have a zero mean, then the pure jump model and original Vasiček one-factor short rate diffusion model given in (2) on page 1 have the same mean-reversion parameter.

Example 21. Let the Vasiček pure jump short rate model be defined by

$$\mathrm{d}r_t = a\left(b - r_t\right)\mathrm{d}t + \mathrm{d}Y_t,$$

where Y_t is a nonstationary compound Poisson process.

Multiply both sides of the equation by e^{at} to get

$$\begin{aligned} e^{at}\mathrm{d}r_t &= e^{at}a\left(b - r_t\right)\mathrm{d}t + e^{at}Z_{N_t}N(\mathrm{d}t) \\ e^{at}\mathrm{d}r_t &= \mathrm{d}\left(e^{at}\right)\left(b - r_t\right) + e^{at}Z_{N_t}N(\mathrm{d}t) \\ e^{at}\mathrm{d}r_t + \mathrm{d}\left(e^{at}\right)r_t &= \mathrm{d}\left(e^{at}\right)b + e^{at}Z_{N_t}N(\mathrm{d}t). \end{aligned} \tag{2.10}$$

Now, by applying the multiplication table be have

$$\begin{aligned} \mathrm{d}\left(e^{at}r_t\right) &= e^{at}\mathrm{d}r_t + \mathrm{d}\left(e^{at}\right)r_t + \mathrm{d}\left(e^{at}\right)\mathrm{d}r_t \\ &= e^{at}\mathrm{d}r_t + \mathrm{d}\left(e^{at}\right)r_t + e^{at}a\mathrm{d}t\left\{a\left(b - r_t\right)\mathrm{d}t + Z_{N_t}N(\mathrm{d}t)\right\} \\ &= e^{at}\mathrm{d}r_t + \mathrm{d}\left(e^{at}\right)r_t. \end{aligned} \tag{2.11}$$

Inserting (2.11) into (2.10) and integrating from 0 to t we find

$$e^{at}r_t = r_0 + b\left(e^{at} - 1\right) + \int_0^t e^{as}Z_{N_s}N(\mathrm{d}s),$$

and

$$r_t = e^{-at}r_0 + b\left(1 - e^{-at}\right) + e^{-at}\int_0^t e^{as}Z_{N_s}N(\mathrm{d}s).$$

Now, taking conditional expectations on both sides we find that

$$E\left\{r_t \mid F_0\right\} = e^{-at}r_0 + b\left(1 - e^{-at}\right) + e^{-at}E\left[Z_1\right]\int_0^t e^{as}E\left\{\lambda(r_{t-}) \mid F_0\right\}\mathrm{d}s$$

As $t \to \infty$, we have

$$E\left\{r_t \mid F_0\right\} \to b + E\left[Z_1\right]\frac{E\left\{\lambda(r_{t-}) \mid F_0\right\}}{a} = b \iff E\left[Z_1\right] = 0,$$

since $\lambda(r_{t-}) > 0$.

Part I

Modelling the South African short rate with a pure jump stochastic process.

Part I: Introduction

In this first part of the thesis we give statistical evidence that a pure jump stochastic process should be used to model the short rate in South Africa. The main result is that for interest rate derivatives, especially when interest rates are at a low level, the impact on the price of a European as well as an exotic option on the forward JIBAR can be significant.

In Chapter 3 the 3-month JIBAR rate is analysed, as the rate is used as a proxy for the South African short rate throughout this thesis. We collected data of the 3-month JIBAR from 1 February 1999 to 31 December 2014. There were periods when the JIBAR changed frequently (Figure 3.2 on page 38) and periods where it did not (Figure 3.3 on page 39). Figure 3.3 is of real significance as there were a number of consecutive days from 20 September 2010 to 11 April 2014, where no changes occurred. Therefore, a pure jump stochastic process is more appropriate than a process with a diffusion component to model the interest rate in such a time span. The question is, can we use a pure jump model to fit the data in general? We answer this question by testing for jumps using the data of the 3-month JIBAR from 1 February 1999 to 17 September 2010 (Chapter 4). We find that jumps did occur frequently, which makes a pure jump model appropriate for the short rate in most circumstances. We also find that if the short rate is modelled by a jump diffusion process, then the volatility obtained from the jumps is the main contributor to the total volatility in the model. This is especially evident when interest rates were low (Figure 4.2 on page 54).

In Chapter 5 a pure jump model is fitted to the data. Firstly we fit an exponential distribution with parameter $\lambda(r_{t-})$ to the time between jumps. As a result the pure jump model is a nonstationary compound Poisson process with an intensity that changes as the interest rate level changes. Secondly, a distribution from the family of stable distributions is fitted to JIBAR changes.

Our main objective is to price interest rate derivatives. In standard pricing methodologies the model used is not a model obtained by the statistical approach we use in Chapter 5. The reason for this is that a martingale approach is used in pricing, where the discounted price of an underlying asset is a martingale. In other words, the underlying asset has the same expected return as a risk-free asset. In reality (under a measure

$\mathbb{P}$) assets have various returns. The dynamics of the underlying asset therefore needs to be transformed to the dynamics of an asset which has the same expected return as a risk-free asset. The transformed dynamics is defined under a transformed measure, called the risk-neutral measure $\mathbb{P}^*$. In Chapter 5 we fit the model with reference to a measure $\mathbb{P}$ and the question arises if the model will still be appropriate with reference to a risk-neutral measure $\mathbb{P}^*$.

An approach in interest rate derivative pricing is to define a short rate model under a measure $\mathbb{P}^*$ as a starting point. The models are defined to support real-world evidence, such as mean-reversion or positive interest rates (observations made under a measure $\mathbb{P}$). We follow the same approach, which is justified by Girsanov's theorem (Example 19 on page 26 and Example 20 on page 29). Firstly, although the model changes with a change of measure, a compound Poisson process under $\mathbb{P}$ will change to a compound Poisson process under $\mathbb{P}^*$, with changed parameter values. Secondly, the jump distribution changes by formula (2.9) on page 29.

Unfortunately, one characteristic of the model we fit in Chapter 5 cannot be used in our model under $\mathbb{P}^*$. We truncate the jumps in our model in order to get more realistic interest rate movements as well as to address a theoretical concern. The theoretical concern is that jumps should be modelled by a distribution with finite variance for a Monte Carlo pricing method to be efficient. The fitted compound Poisson model under $\mathbb{P}$ has stably distributed jumps. Stably distributed random variables have infinite variance (where the index parameter $\alpha < 2$), which will lead to infinite variance of the short rate movements in the model under the measure $\mathbb{P}$. We will simulate Monte Carlo short rate paths to estimate prices of interests rate derivatives, where the efficiency of the method is dependent on the variance of the short rate movements. Short rate movements with infinite variance will therefore lead to an inefficient pricing method.

The short rate model defined under the measure $\mathbb{P}^*$ is given by

$$\mathrm{d}r_t = \mu(r_t)\mathrm{d}t + Z_{N_t}N(\mathrm{d}t), \tag{2.12}$$

where N_t is a Poisson process with intensity $\lambda^*(r_{t-})$ and $(Z_n)_{n\geq 1}$ are i.i.d. stable jump sizes, truncated at levels $\pm L = \pm 0.5$ with jump distribution ν^*. Our pure jump model will have some properties that are similar to a diffusion model. The drift coefficient μ will ensure more flexibility. For example, appropriate characteristics such as mean-reversion can be built into the model (See Example 21 on page 30). The location and skewness parameters of the family of stable distributions are set equal to zero. The mean of the jumps is therefore zero. We made this assumption for two reasons. Firstly, we investigate the impact of the index parameter α on interest rate option prices. The assumption of jumps with zero mean means that only two parameters are left in the jump distribution, which makes the task less complex. Another reason is that the conditional drift and jump

components are separated by

$$E^* \{ \mathrm{d}r_t \mid \mathcal{F}_{t-} \} = \mu(r_t)\mathrm{d}t$$

and

$$E^* \left\{ (\mathrm{d}r_t)^2 \mid \mathcal{F}_{t-} \right\} = \lambda^*(r_{t-}) E_{\nu^*} \left[Z_1^2 \right] \mathrm{d}t, \tag{2.13}$$

where expectation is taken under $\mathbb{P}^*$. This is similar to the separation of a one-factor diffusion model into a drift and a diffusion component. We investigate the impact of pure jumps against diffusion. By setting up our model this way we can determine the impact more efficiently.

The main question is, compared to a one-factor diffusion model, whether our pure jump short rate model influences interest rate derivative prices or not? To answer this question we assume under $\mathbb{P}^*$ a nonparametric one-factor diffusion model given by (1) on page 1. We apply the method by Johannes (2004) to estimate the diffusion coefficient. These parameters under $\mathbb{P}^*$ can be estimated using either historical or market data since the diffusion coefficient does not change with a change of measure. The question is then how to calibrate our pure jump model because these parameter values do change with a change of measure? If we assume no market price of risk we can estimate the parameters using the following equation

$$\lambda^*(r_t) E_{\nu^*}(Z_1^2)\mathrm{d}t = E \left\{ (\mathrm{d}r_t)^2 \mid \mathcal{F}_{t-} \right\}. \tag{2.14}$$

In Chapter 6 we price both a vanilla option and a barrier option on the 3-month forward JIBAR, where the latter can be calculated as the average yield of a zero-coupon bond. We find the pure jump model to have an impact on interest rate derivative prices, especially when interest rate levels are low.

Due to a number of reasons, we are unable to find a unique market determined price in an incomplete interest rate market. We will show that in the case where market price of risk does exist, it will influence the calibration of option prices to a volatility structure. Research into measuring the market price of risk in the South African market is therefore necessary. Even if the market price of risk is zero, the calibration process demands explicitly computable bond and option prices from our model. We will therefore in future research projects need to possibly simplify our model in search of analytical tractability. To add to these problems, interest rate options are mostly traded OTC in South Africa. Therefore, data is sparse, which will negatively impact accurate calibration to an option volatility structure.

However, we do find an impact of jumps on option prices assuming a zero market price of risk. We can therefore assume the impact will be even greater where a market price of risk does exist. Our results can therefore be used as a benchmark model to compare pricing results from more analytically tractable models.

Chapter 3

The JIBAR

Short term interest rates used in calculating the nominal swap yield curve in South Africa are the range of JIBAR rates. In the 1990s the South African Futures Exchange Bank Bill rate was used for this purpose. This rate, which represented a risk free rate, was based on bankers' acceptances (BAs). BAs are "tradable instruments issued by a corporate but accepted by a bank"(SARB., 2012). As the interest rate derivative market in South Africa grew in the 1990s, the need for a 3-month maturity reference rate emerged. The BAs were replaced by negotiable certificates of deposit (NCDs), which are instruments banks had been trading between themselves. The average rate derived from the NCDs is called the JIBAR. The JIBAR is similar to the LIBOR, specifically in the sense that it can be used as a proxy for a risk free interest rate. The rate is available in 1-month, 3-month, 6-month and 12-month maturities.

The 3-month JIBAR is also "accepted as a short-dated interest rate benchmark"(SARB., 2012). There are other rates that could have been considered as a proxy for a short term risk free interest rate. The interest rate derived from the 3-month Treasury Bill (T-bill) issued by the South African Reserve Bank (SARB) is one alternative. The 3-month T-bill rate is only calculated once a week when the instrument is traded on an auction. The accuracy of some of our tests (specifically the test for jumps in Section 4.2) depends on the frequency of data available. The 3-month JIBAR changes daily, which suits our needs better than the rate derived from the 3-month T-bill.

The daily 3-month JIBAR from 1 February 1999 to 31 December 2014, as obtained from Safex. (2014) is shown in Figure 3.1. The Repo rate is also shown in this figure. The Repo rate is the interest rate determined by the SARB through its Monetary Policy Committee (MPC), obtained from SARB. (2015).

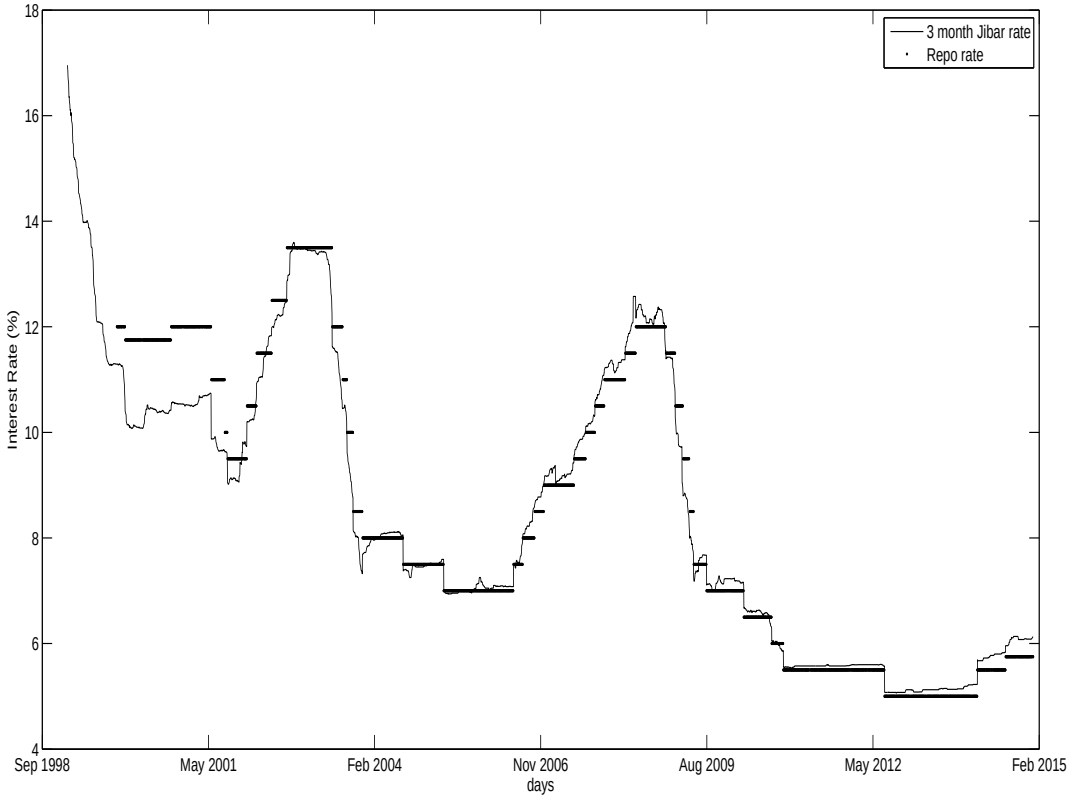

Figure 3.1: The historical daily 3-month JIBAR compared to the Repo rate. The Repo rate remains constant for longer periods, while the JIBAR may have daily changes.

Figure 3.2 shows a part of the time series in Figure 3.1 where interest rates changed frequently. The monetary policy of SARB aims to keep inflation between 3% and 6%. One way to accomplish this is to control interest rates through changes in the Repo rate, which the SARB controls. From August 2007 to September 2008 the inflation rate kept rising from an index in the region of 6% to above 11% (Anon, 2015). This forced the SARB to increase the Repo rate. However, after the global economic downturn in 2008 reached its peak, inflation started to fall, reaching the 6% mark in September 2009. The Repo rate also declined during this period as the SARB wanted to help the economy to recover from the recession.

Generally, the JIBAR changes in anticipation of changes in the Repo rate. On the other hand, there are instances where the market reacted to decisions made by the MPC in terms of the Repo rate. There are examples of this that can be seen in our time series. For instance, from November 2006 to June 2007 (Figure 3.2), the JIBAR kept rising, as the market anticipated a rise in the Repo rate. However, the Repo rate was held at the

same level. The constant Repo rate resulted in a substantial decline in the JIBAR. The JIBAR therefore exhibits both small and large changes and these seem to occur mostly near times when the MPC discusses a possible change in the Repo rate.

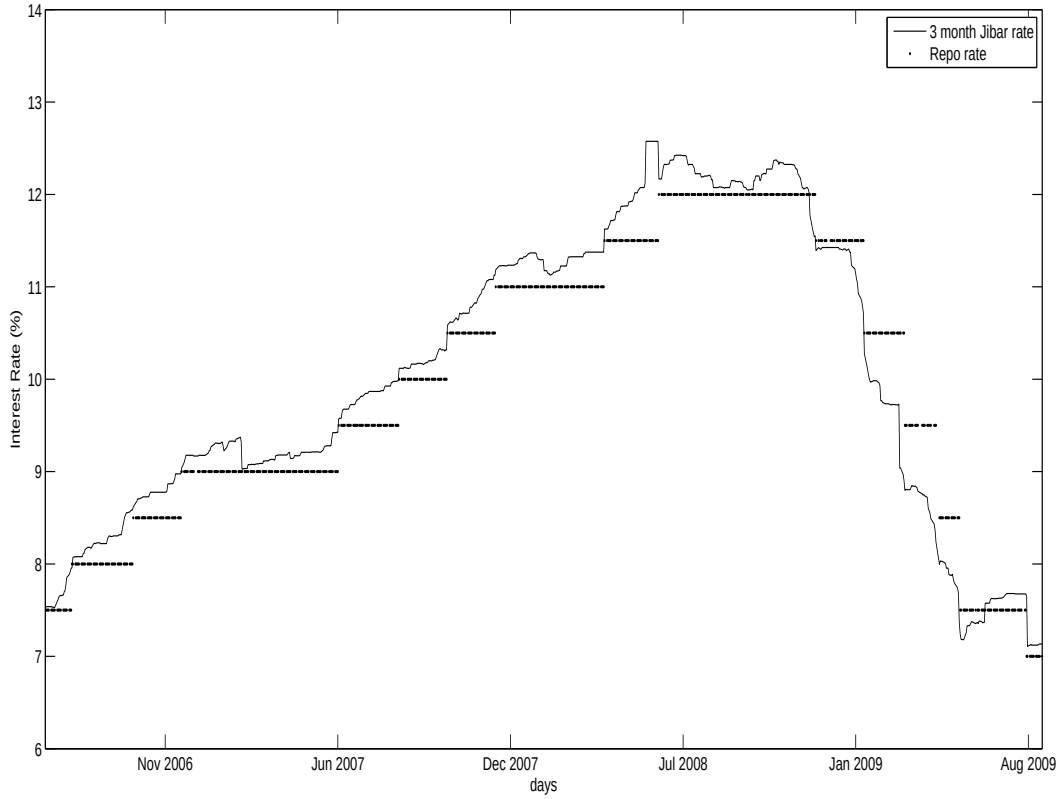

Figure 3.2: A period from Figure 3.1 where the 3-month JIBAR changed frequently.

In Figure 3.3 a period is shown where interest rates hardly changed. In 2009, as the South African economy struggled to recover from the global economic recession, interest rates were kept at very low levels to encourage consumer spending. The JIBAR therefore did not change frequently in this time and the changes in interest rates were small. The exception to these small changes were near dates when the MPC changed the Repo rate.

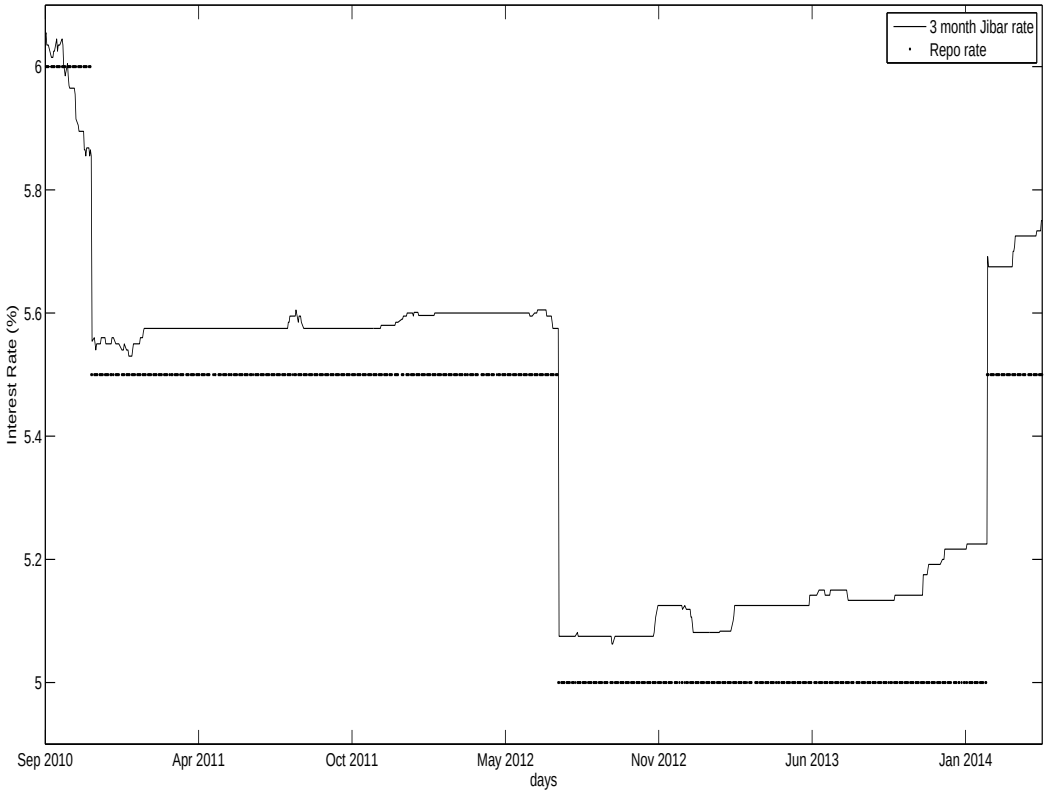

Figure 3.3: A period from Figure 3.1 where interest rates remained fairly constant.

The data shown in Figure 3.3 has significant consequences in the modelling of the short rate. The standard approach in short rate modelling is to assume a diffusion short rate model and to add jumps to the model if jumps in the market occur. However, no diffusion component is evident in Figure 3.3. An alternative to the standard approach is therefore to assume a pure jump model as all movements in data can be seen as jumps. This will enable us to statistically fit a distribution to all interest rate movements. The question then is whether a pure jump model is appropriate to be fitted to the observed data? To answer this question we will test for jumps using data in the period prior to the period shown in Figure 3.3 (Chapter 4).

It is interesting to know whether data from other markets have the same characteristics. In Figure 3.4 we compare the 3-month JIBAR to the 3-month LIBOR. Note that the LIBOR had more rate changes in the period from 2009 until 2014 (shown in the bottom graph). This was in a period when rates were relatively low in both markets. This could be due to the LIBOR being a rate derived from a more liquid financial instrument compared to the derived JIBAR rate. The pure jump characteristic in the Johannesburg

based market therefore seems to be unique compared to a more liquid London based market. Our data could therefore necessitate a unique and different approach to interest rate derivative pricing given the unique characteristics observed.

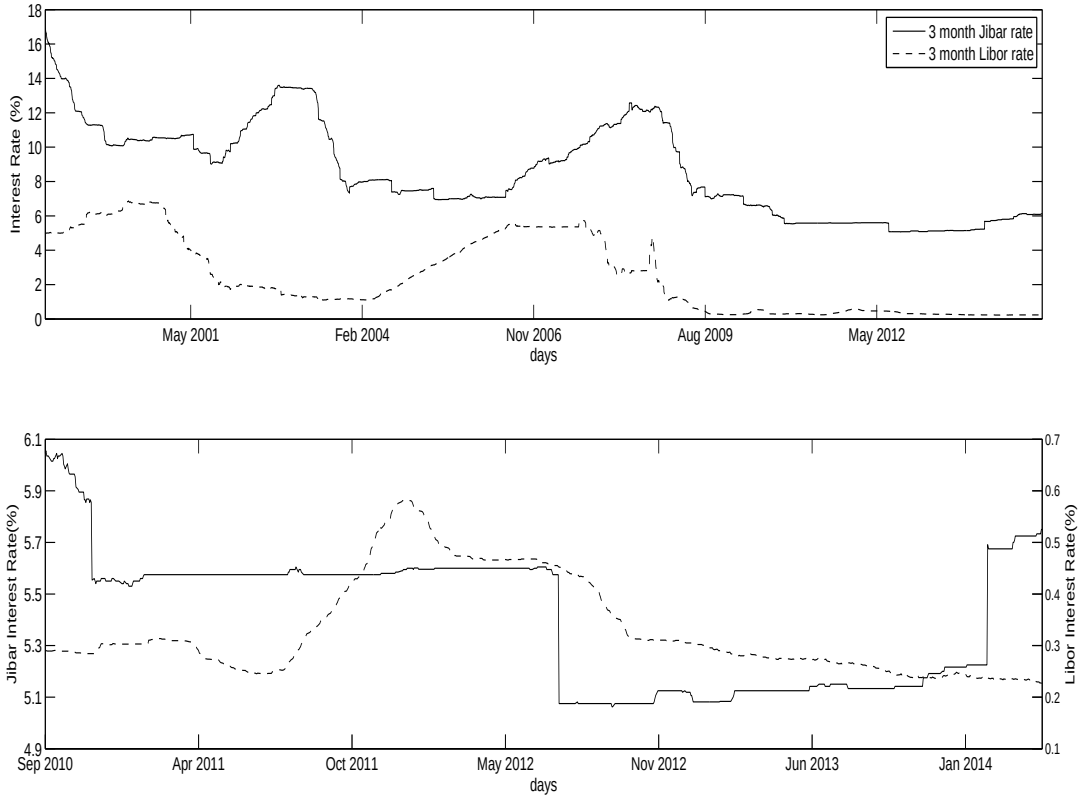

Figure 3.4: The JIBAR rate compared to the LIBOR rate for the period 1 February 1999 to 31 December 2014 (top graph). The bottom graph highlights a period where interest rate levels were low in both markets. A pure jump characteristic in this period is more evident from the JIBAR data than from the LIBOR data.

An alternative to using the JIBAR rate is to use the 91 day T-bill rate. In South Africa T-bills are traded on a weekly auction, but is open to the public, whereas the JIBAR rate is a rate calculated from assets traded between South African banks. The T-bill is a liquid instrument, but the T-bill and JIBAR rate has some comparable characteristics, which is shown in Figure 3.5. In the top graph we show historical rates from 1 February 1999 to 31 December 2014, where it is clear that the two rates are highly correlated. In the bottom graph we show the rates from a period where interest rates were low. The T-bill stayed constant for long periods, even though the T-bill is traded on an auction. From 8 July 2011 to 8 December 2011 (22 consecutive weeks) the T-bill rate was quoted

at a constant rate of 5.49%.

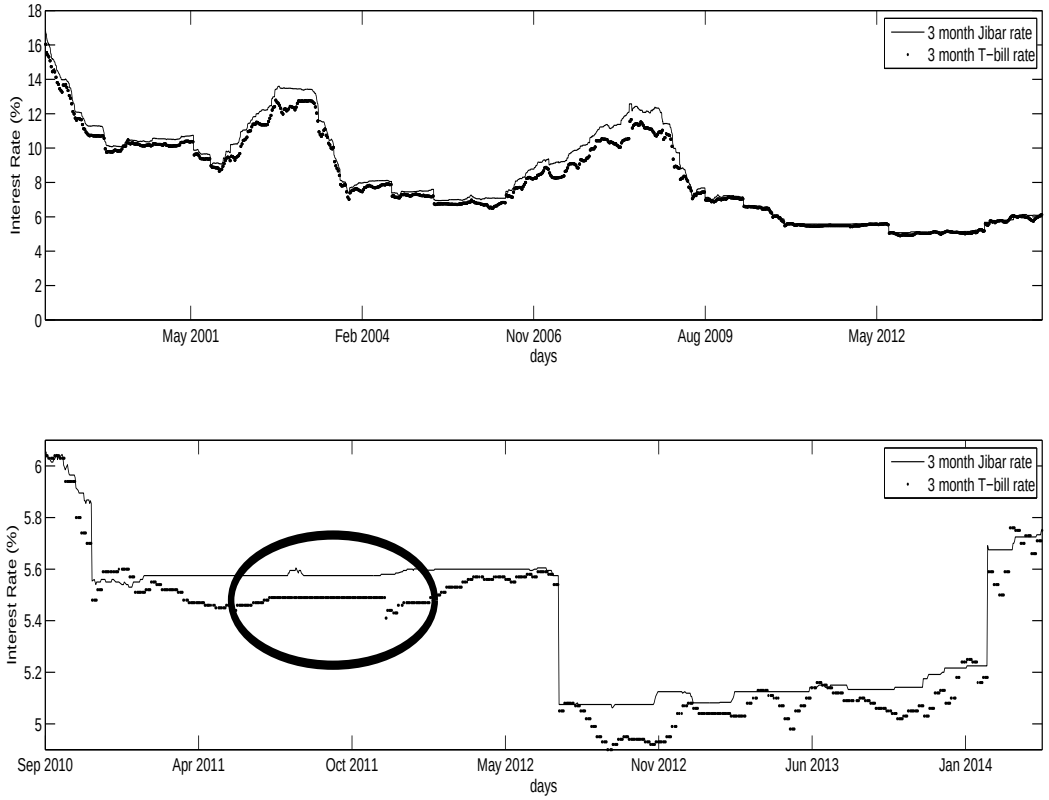

Figure 3.5: The 3-month JIBAR rate compared to the 91-day T-bill rate for the period 1 February 1999 to 31 December 2014 (top graph). The bottom graph highlights a period where interest rate levels were low in both markets. Although the T-bill rate has more changes, it stays constant for some long periods (highlighted in the bottom graph).

In Table 3.1 we compare some basic statistics of the 91-day T-bill rate and the 3-month JIBAR. In order to make a fair comparison we only take the dates on which the T-bill rate are quoted (once a week) and compare the week-on-week changes of the two rates. The changes are measured as a change in the number of basis points where one percentage point of the rate is equal to 100 basis points. The statistics show similar trends for the two rates. What is evident, is the high kurtosis for both samples as well as the high percentage of observations where no rate change occurred.

Table 3.1: Basic statistics of the week-on-week JIBAR and T-bill rate changes (measured in basis points)

Description	JIBAR	T-bill
# observations	805	805
mean	-1.316729	-1.236025
standard deviation	12.925293	13.616603
skewness	-2.608868	-1.292453
kurtosis	15.505196	12.131436
maximum	50	100
minimum	-100	-83
# ZRC	237	140
ZRC as % of total	29.44%	17.39%

ZRC = Zero rate change observations

The efficiency of the statistical tests we use in Chapter 4 to test for jumps, depends on the frequency of data. The T-bill rate changes weekly, while the JIBAR changes daily. We therefore rather use the JIBAR rate than the T-bill rate in our work. The basic statistics of the daily changes in the JIBAR are shown in Table 3.2. The observations we made from Figure 3.3 on page 39 are of real significance. We therefore show the statistics from our entire dataset as well as for the period before 17 September 2010 and after.

Table 3.2: Basic statistics of the daily 3-month JIBAR rate changes (measured in basis points)

Description	JIBAR(1999-2014)	JIBAR(1999-2010)	JIBAR(2010-2014)
# observations	4035	2966	1069
mean	-0.268426	-0.368206	0.008419
standard deviation	4.497886	5.046124	2.370504
skewness	-6.077207	-5.682355	-3.432186
kurtosis	118.107362	94.468961	350.103329
maximum	46.8	46.8	46.7
minimum	-86	-86	-50
# ZRC	2324	1388	936
ZRC as % of total	57.60%	46.80%	87.56%

ZRC = Zero rate change observations

Two statistics correspond with the statistics shown in Table 3.1. The one is the high kurtosis (118) and the other is the high percentage of observations where no rate change occurred (57.6%). The objective of Chapter 4 is firstly to test whether sample paths from

a one-factor diffusion process can obtain such a high kurtosis. If not, should jumps be added? Secondly, is it realistic for a jump diffusion process to produce paths with such a high frequency of zero rate movements? The statistics in Table 3.2 calculated from the rate changes for the period from 2010 to 2014 shows a clear lack of a diffusion component.

The focus will now shift to showing graphs of the changes in the 3-month JIBAR. The changes are measured as a change in the number of basis points. Figure 3.6 shows the daily interest rate changes over time. The interest rate changes seem to be skew, as the large downward movements in general exceed the large upward movements. This phenomenon may be due to the way markets react to certain economic events. What is again evident in this graph is not only the infrequent large jumps, but the absence of interest rate movements for long periods.

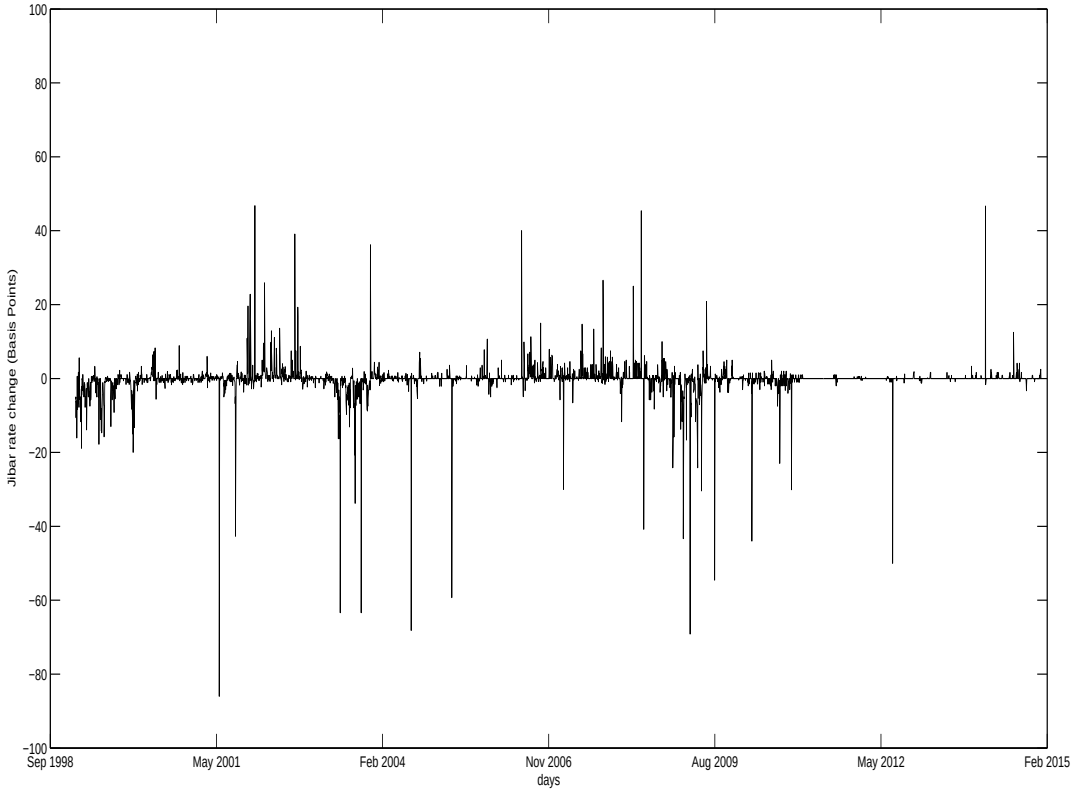

Figure 3.6: Daily three-month JIBAR changes.

Our approach in Chapter 5 will be to fit a model to both the JIBAR movements as well as the time between movements. Figure 3.7 shows a histogram of the JIBAR changes. Most of the changes are near zero, with some extreme changes. The extreme interest rate changes justify investigation into thick tailed distributions, such as the family of stable

distributions (Part II).

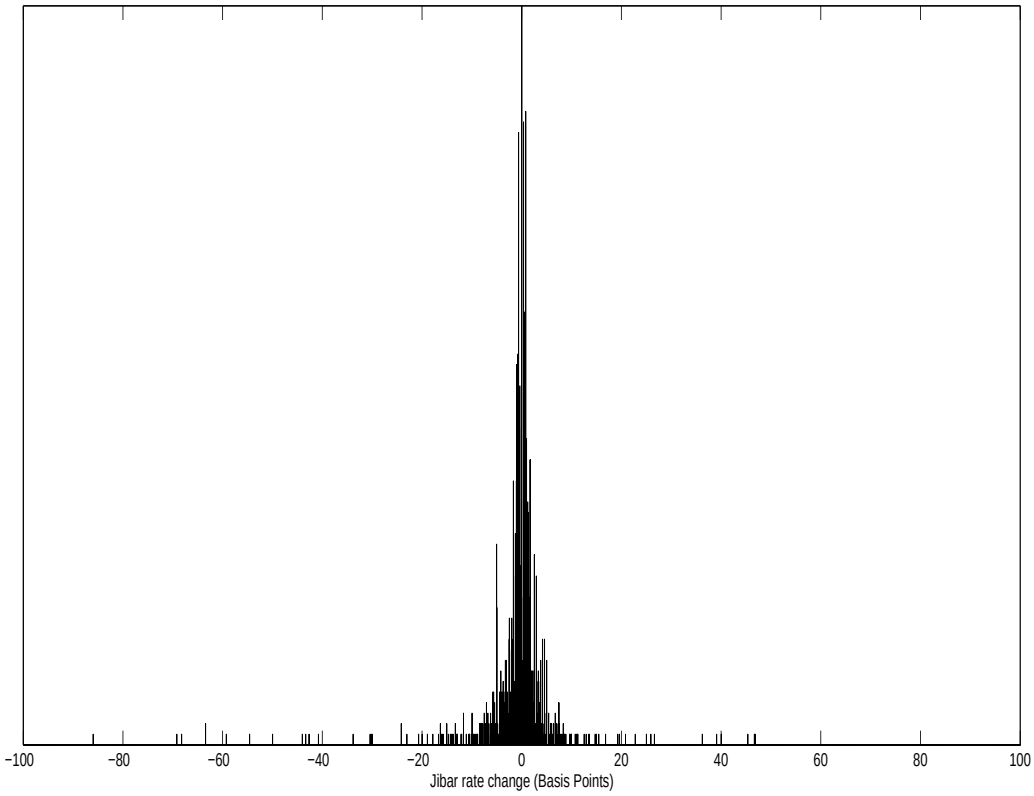

Figure 3.7: A histogram of the three-month JIBAR changes.

Figure 3.8 shows the frequency distribution of the number of days between changes in the JIBAR. There were four separate periods where the number of days between jumps exceeded 60 days. This graph confirms our observation from preceding figures and tables in this section concerning the infrequency of interest rate movements.

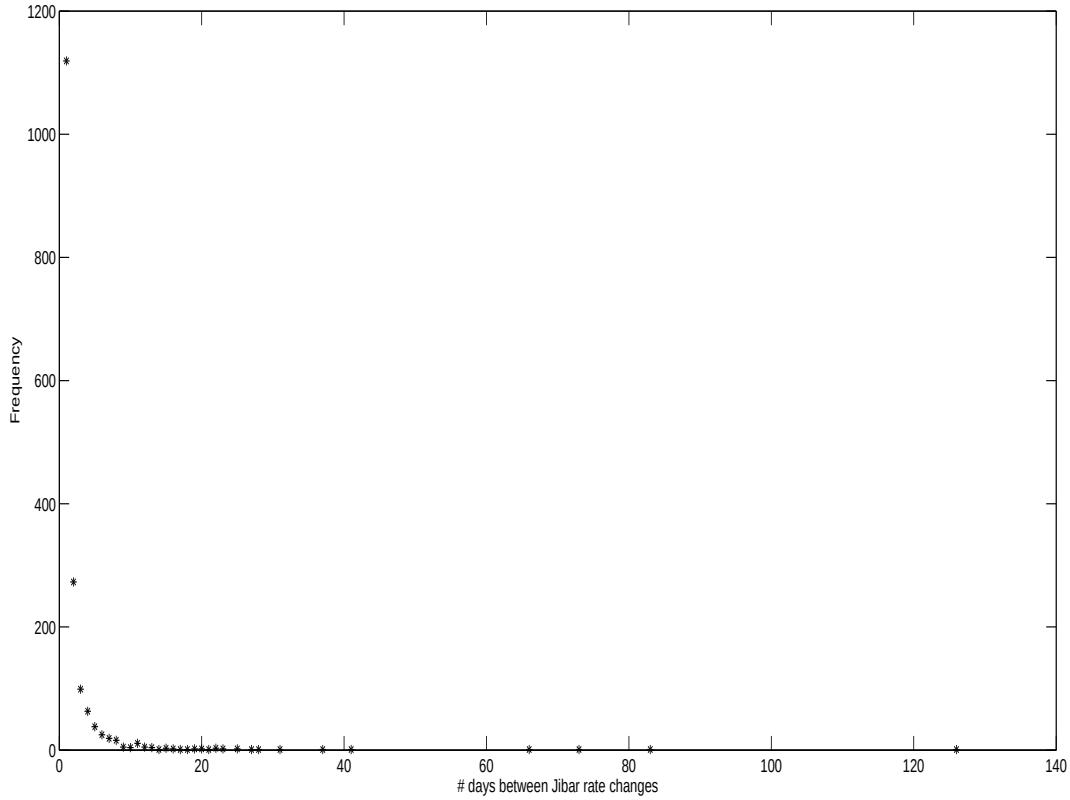

Figure 3.8: The frequency of the number of days between JIBAR changes.

Two characteristics of South African short term interest rates are clearly evident from our empirical study in this chapter, and needs to be addressed by our model. These characteristics are the high kurtosis of interest rate movements as well as the high frequency of zero rate changes, which are not characteristics of pure diffusion models. These observations were made for both the rate changes in the JIBAR and T-bill rate. Jump diffusion models may result in interest rate movements with a high kurtosis. However, interest rate paths with a high frequency of zero rate changes cannot be reproduced by a model with a diffusion component. From this initial analysis it seems that a jump diffusion model is, in general, inappropriate as a model for the short rate and that a pure jump model may be more efficient.

Chapter 4

Testing for jumps

In this chapter we investigate which model is the most suitable for the South African short rate: a one-factor diffusion process, a jump diffusion process or a pure jump process. Several tests exist, with some of them summarised by Hong and Zou (2015). A test for jumps was first introduced by Ait-Sahalia (2002), after which many similar tests were developed. Carr and Wu (2003) developed a test in which the dynamics of an asset can be identified as either a diffusion-, jump diffusion- or pure jump model. This is done by analysing the behaviour of short-maturity options. However, due to the nature of the South African interest rate option market, data is not freely available to apply this test. High-frequency data of the underlying asset is a requirement for a number of tests. For example, the nonparametric test to identify isolated jumps by Lee and Mykland (2008) is an asymptotic test, which implies that high-frequency data is necessary. This is also true for the nonparametric test for jumps by Ait-Sahalia and Jacod (2009) and its extension by Ait-Sahalia, Jacod and Li (2012). Unfortunately, high frequency data is not available for short term South African interest rates. The frequency of the 3-month JIBAR is daily, while the 91-day T-bill rate is calculated once a week. Another test, developed by Johannes (2004), is a Monte Carlo based hypothesis test. Although not ideal, we will apply two tests for jumps using daily data. These are nonparametric tests developed by Johannes (2004) and by Lee and Mykland (2008).

In the preceding chapter we empirically analysed the JIBAR. For the period from 20 September 2010 to 31 December 2014 the JIBAR is characterised by infrequent movements. A strong case can be made that a pure jump model as underlying process for the short rate during this period is the best choice. However, for the period prior to 20 September it is not clearly evident which type of model is most appropriate. We therefore used formal tests for jumps in this period.

Our first main conclusion in this chapter is that a diffusion process is not suitable for modelling of the short rate. By applying the test from Johannes (2004) we show that

a diffusion process cannot replicate the high sample kurtosis. The result from Lee and Mykland (2008) is that, of all JIBAR movements, 11% were identified as jumps. Secondly, a jump diffusion process does replicate sample paths with a similar high kurtosis, compared to the sample kurtosis. However, we overestimated the conditional variance in the jump diffusion model. This is due to estimation errors of the conditional variance emanating from the diffusion model component. We find that the main contributor to the total estimated conditional variance can be ascribed to the jump component. Interestingly, at low interest rate levels, the only contributor to the total variance emanates from the jump component. In addition we find that more than half of the observed day-to-day rate changes are zero jumps. This provides further proof that a pure jump model should be used to model the short rate, especially when interest rate levels are relatively low.

The details of the two tests applied, as well as the results, will be described in the next two sections.

4.1 Monte Carlo hypothesis testing

We start our discussion by using the test developed by Johannes (2004) to test whether the 3-month JIBAR data can be modelled by a one-factor diffusion model under a market measure $\mathbb{P}$, defined by

$$\mathrm{d}r_t = \mu(r_t)\mathrm{d}t + \sigma(r_t)\mathrm{d}W_t, \tag{4.1}$$

where W is a Brownian motion under $\mathbb{P}$. The sample kurtosis of Monte Carlo simulated sample path increments, generated by the diffusion model, will be compared to the sample kurtosis of the 3-month Jibar rate increments. Note that the model in (4.1) is defined under a measure $\mathbb{P}$, while the model in (1) on page 1 is defined under a risk-neutral measure $\mathbb{P}^*$. The parameters of the model in (4.1) will therefore be estimated using market data, in this case historical data.

Now, the conditional expectation of the dynamics of the short rate can be written as

$$E\left\{\mathrm{d}r_t \mid \mathcal{F}_{t-}\right\} = \mu(r_t)\mathrm{d}t + \sigma(r_t)E\left\{\mathrm{d}W_t \mid \mathcal{F}_{t-}\right\} = \mu(r_t)\mathrm{d}t, \tag{4.2}$$

since the Brownian motion W_t is a martingale (see Example 12 on page 23 for the result and the notation).

We also have that the conditional expectation of the quadratic variation of the short rate can be written as (see Example 4 on page 16)

$$E\left\{(\mathrm{d}r_t)^2 \mid \mathcal{F}_{t-}\right\} = \sigma^2(r_t)\mathrm{d}t. \tag{4.3}$$

Now, by discretisation of (4.2) and (4.3) we find that

$$\mu(r_t) \approx \frac{1}{\triangle} E\left\{r_{t+\triangle} - r_t | r_t = r\right\}$$

and

$$\sigma^2(r_t) \approx \frac{1}{\triangle} E\left\{(r_{t+\triangle} - r_t)^2 | r_t = r\right\},$$

where $\triangle$ is the increment length.

Therefore, to estimate μ and σ, the first two conditional moments of short rate increments need to be estimated.

For the set of interest rates $S = \{r_{1\triangle}, r_{2\triangle}, ..., r_{k\triangle}, ..., r_{T\triangle}\}$, fix k and set $r = r_{k\triangle}$. The natural estimator of $E\left[(r_{t+\triangle} - r_t)^j | r_t = r\right]$ is given by

$$\begin{aligned}
&\frac{1}{\#(r_{i\triangle} = r)} \sum_{r_{i\triangle}=r} \left(r_{(i+1)\triangle} - r\right)^j \\
&\approx \frac{\sum_{t=1}^{T-1} \left(r_{(t+1)\triangle} - r\right)^j \mathbb{1}(r_{t\triangle} = r)}{\sum_{t=1}^{T-1} \mathbb{1}(r_{t\triangle} = r)} \\
&= \frac{\sum_{t=1}^{T-1} \left(r_{(t+1)\triangle} - r_t\right)^j \mathbb{1}(r_{t\triangle} = r)}{\sum_{t=1}^{T-1} \mathbb{1}(r_{t\triangle} = r)},
\end{aligned}$$

where $\#(r_{i\triangle} = r)$ denotes the number of times $r_{i\triangle} = r$.

To smooth this estimator (in case we want to fix r at real numbers not equal to a rate in S), notice that if ϕ is the standard normal density, then

$$\phi\left(\frac{r_{t\triangle} - r}{h_j}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{r_{t\triangle} - r}{h_j}\right)^2},$$

where h_j are bandwidths, is a typical Kernel used to estimate a density. As $h_j \to 0$ we have that

$$\phi\left(\frac{r_{t\triangle} - r}{h_j}\right) \to \begin{cases} \frac{1}{\sqrt{2\pi}}, & \text{if} \quad r_{t\triangle} = r \\ 0, & \text{if} \quad r_{t\triangle} \neq r \end{cases} = \frac{1}{\sqrt{2\pi}} \mathbb{1}_{(r_{t\triangle}=r)}.$$

Hence, smoothed estimators for $\frac{1}{\triangle} E\left[(r_{t+\triangle} - r_t)^j | r_t = r\right]$ can be defined by

$$m_j(r_t) = \frac{\frac{1}{\triangle} \sum_{t=1}^{T-1} \left(r_{(t+1)\triangle} - r_t\right)^j \phi\left(\frac{r_{t\triangle} - r}{h_j}\right)}{\sum_{t=1}^{T-1} \phi\left(\frac{r_{t\triangle} - r}{h_j}\right)}. \tag{4.4}$$

The bandwidths h_j used are the same as those suggested by Johannes (2004).

Therefore, the drift and diffusion in (1) on page 1 can be estimated by

$$\hat{\mu}(r_t) = m_1(r_t), \quad \hat{\sigma}(r_t) = \sqrt{m_2(r_t)}.$$

Using these estimates, interest rate paths $\{r_{i,t},\ i = 1, ..., M,\ t = 0, \triangle, ..., T\triangle\}$ are simulated. Starting at $r_{i,0}$, we define

$$r_{i,t+\triangle} = r_{i,t} + \hat{\mu}(r_{i,t})\triangle + \hat{\sigma}(r_{i,t})\sqrt{\triangle} Z_{i,t},$$

where $Z_{i,t}$ are i.i.d. standard normally distributed random variables.

To test for jumps, we compare the sample kurtosis of the JIBAR rate increments to a distribution of sample kurtosis obtained from these simulated interest rate paths.

Implementing this method, we selected $r_{i,0} = 10\%$, $\triangle = 1$, $M = 10,000$ and $T = 2968$, where T is the number of observations we have from 1 February 1999 to 17 September 2010. The estimated drift and variance, $\hat{\mu}(r_t)$ and $\hat{\sigma}^2(r_t)$, from the sample should lie within a confidence band of estimates obtained from the simulated paths. In the top two graphs in Figure 4.1 we show that this is the case.

We have shown that a nonparametric diffusion model can generate sample paths with first- and second order conditional moments similar to those of the observed data. However, the sample kurtosis is very high (Table 3.2 on page 42). The question is whether the model can generate sample paths with similar higher order conditional moments? To investigate we will use conditional skewness and kurtosis estimates calculated by

$$\hat{\Sigma}(r_t) = \frac{m_3(r_t)}{m_2(r_t)^{3/2}}, \quad \hat{\kappa}(r_t) = \frac{m_4(r_t)}{m_2(r_t)}.$$

Confidence bands for the estimated conditional skewness and kurtosis are shown in the two bottom figures in Figure 4.1. The conditional skewness and kurtosis estimated from the sample does not lie within these bands. Informally, we can therefore conclude that a diffusion model is not efficient in modelling the short rate.

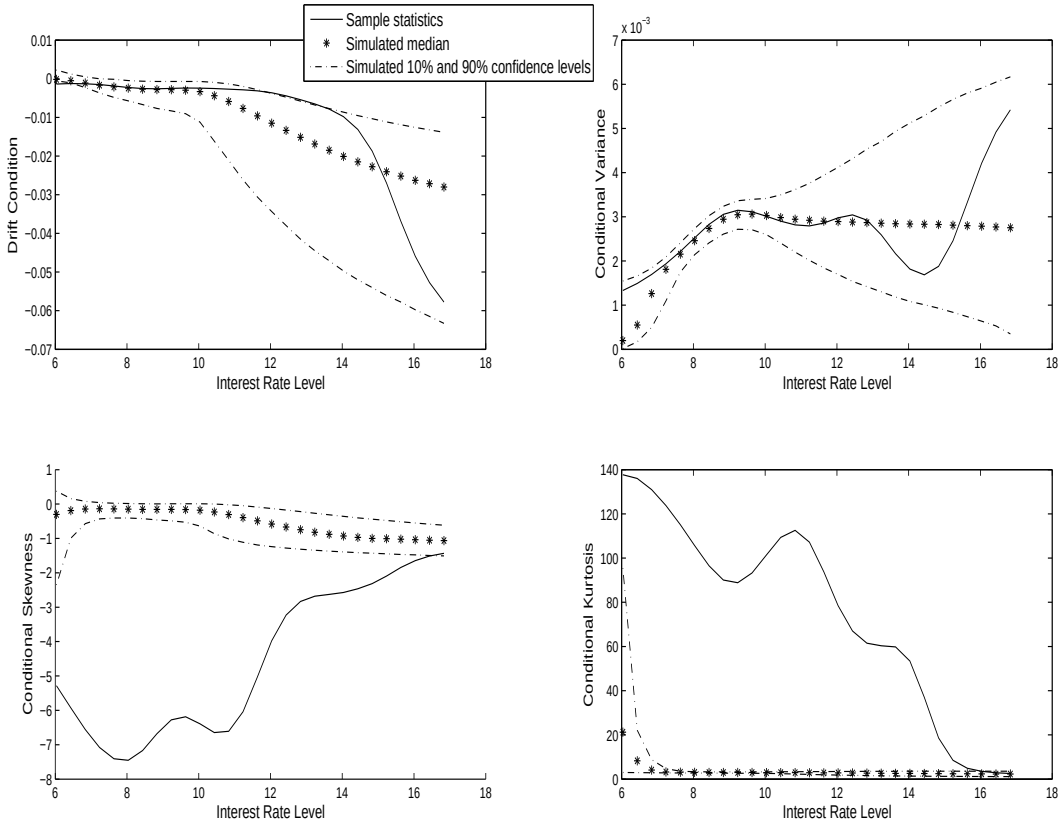

Figure 4.1: The estimated functions $\hat{\mu}$, $\hat{\sigma}$, $\hat{\Sigma}$ and $\hat{\kappa}$ for our dataset as well as Monte Carlo confidence bands assuming a single-factor diffusion model.

To apply a more formal test, we will use a hypothesis test. The unconditional sample excess kurtosis is given by (see Weisstein (2017))

$$\hat{k} = \frac{\kappa_4}{\kappa_2^2}, \tag{4.5}$$

where

$$\kappa_2 = \frac{n}{n-1} m_2$$

and

$$\kappa_4 = \frac{n^2((n+1)m_4 - 3(n-1)m_2^2)}{(n-1)(n-2)(n-3)},$$

with m_2 the sample variance, m_4 the 4^{th} sample central moment and n the sample size.

We apply the test as follows:

1. Estimate the statistic from a sample of interest rate changes. Therefore $\hat{k}_s$ is calculated for the sample of 3-month JIBAR changes of size T.

2. From the stochastic process assumed in the null hypothesis simulate a large number M of sample paths $\{r_{i,1}, ..., r_{i,T}\}_{i=1}^{M}$.

3. For each sample path, compute the kurtosis statistic, $\hat{k}_i$ which produces a set of test statistics of size M.

4. Compare the quantiles from the simulated distribution of kurtosis statistics to the sample kurtosis statistic.

The sample kurtosis, $\hat{k}_s$ from formula (4.5), is used as a test statistic in the hypothesis test. The set of sample kurtoses $\hat{k}_i,\ i = 1, 2, ..., M$, from the Monte Carlo paths can be calculated and ordered to produce the p^{th} percentile $\hat{k}^p$. Then, at a confidence level of $1 - \alpha$, the null hypothesis is rejected if

$$\hat{k}_s < \hat{k}^{\alpha/2} \text{ or } \hat{k}_s > \hat{k}^{1-\alpha/2}.$$

Table 4.1 summarises these Monte Carlo percentiles. It shows that the null hypothesis that a single-factor diffusion process is adequate to model short rate increments is rejected at a significance level of $\alpha = 2\%$ as $\hat{k}_s = 94.30 > 13.42 = \hat{k}^{99\%}$.

Table 4.1: Kurtosis percentiles for a single factor diffusion model

Description	Percentile	Kurtosis
sample kurtosis		94.30
simulated kurtosis	50	2.37
	75	4.49
	90	7.10
	95	9.05
	99	13.42

The second model that we test using the method of Johannes (2004) is a nonparametric jump diffusion model defined by

$$\mathrm{d}r_t = \mu(r_t)\mathrm{d}t + \sigma(r_t)\mathrm{d}W_t + \mathrm{d}\left(\sum_{n=1}^{N_t} r_{\tau_n -}\left\{\exp\left(Z_n\right) - 1\right\}\right) \tag{4.6}$$

where W_t is a Brownian Motion, N_t is a doubly stochastic point process with stochastic intensity $\lambda(r_t)$ and $Z_n \sim N(0, \sigma_z^2)$ are the marks of the point process that arrive at time τ_n.

The pure jump part of the process in (4.6) is equivalent to the dynamics of $\log r_t$ with jumps Z_n. Applying Itô's formula (Example 10 on page 22) to $f(x) = e^x$, where

$$\mathrm{d}\log r_t = \mathrm{d}\left(\sum_{n=1}^{N_t} Z_n\right) = \int_{-\infty}^{\infty} yN(\mathrm{d}t, \mathrm{d}y),$$

we get

$$\mathrm{d}e^{\log r_t} = \mathrm{d}r_t = \int_{-\infty}^{\infty} r_{t-} \left(e^y - 1\right) N(\mathrm{d}t, \mathrm{d}y) = \mathrm{d}\left(\sum_{n=1}^{N_t} r_{\tau_n -} \left\{\exp\left(Z_n\right) - 1\right\}\right).$$

Therefore, to fit the pure jump part in (4.6), a compound Poisson process Y_t, with jumps Z_n, is fitted to the natural logarithm of the interest rates. This leads to the following model

$$\mathrm{d}\log r_t = \mu(r_t)\mathrm{d}t + \sigma(r_t)\mathrm{d}W_t + \mathrm{d}Y_t, \tag{4.7}$$

which is equivalent to (4.6). The coefficients, $\mu(r_t)$ and $\sigma(r_t)$, are redefined as the drift and diffusion of the natural logarithm of the short rate.

The nonparametric functions μ, σ, λ and σ_z can be estimated using the moment estimates from formula (4.4) on page 48.

Now, from formula (2.8) on page 28 we have that

$$E\left\{\mathrm{d}\log r_t \mid \mathcal{F}_{t-}\right\} = \mu(r_t)\mathrm{d}t + E\left\{\mathrm{d}Y_t \mid \mathcal{F}_{t-}\right\} = \mu(r_t)\mathrm{d}t + \lambda(r_{t-})E\left[Z_1\right]\mathrm{d}t. \tag{4.8}$$

The conditional expectation of the quadratic variation of $\log r_t$ can be written as (see Example 4 on page 16, which can be directly extended for a nonstationary compound Poisson process)

$$E\left\{\left(\mathrm{d}\log r_t\right)^2 \mid \mathcal{F}_{t-}\right\} = \sigma^2(r_t)\mathrm{d}t + E\left\{\left(\mathrm{d}Y_t\right)^2 \mid \mathcal{F}_{t-}\right\} = \sigma^2(r_t)\mathrm{d}t + \lambda(r_{t-})E\left[Z_1^2\right]\mathrm{d}t. \tag{4.9}$$

For the natural numbers $j \geq 3$, Y_t^j are compound Poisson processes with jump sizes Z_n^j.

$$E\left\{\left(\mathrm{d}\log r_t\right)^j \mid \mathcal{F}_{t-}\right\} = E\left\{\left(\mathrm{d}Y_t\right)^j \mid \mathcal{F}_{t-}\right\} = \lambda(r_{t-})E\left[Z_1^j\right]\mathrm{d}t. \tag{4.10}$$

Now, by discretisation of (4.8) to (4.10) we will be able to find the parameters μ, σ, λ and σ_z^2 from the following equations

$$\begin{aligned}
\frac{1}{\triangle} E\left[\log(r_{t+\triangle}/r_t) | r_t = r\right] &\approx \mu(r_t), \\
\frac{1}{\triangle} E\left[\log(r_{t+\triangle}/r_t)^2 | r_t = r\right] &\approx \sigma^2(r_t) + \lambda(r_{t-})\sigma_z^2, \\
\frac{1}{\triangle} E\left[\log(r_{t+\triangle}/r_t)^4 | r_t = r\right] &\approx 3\lambda(r_{t-})(\sigma_z^2(r_t))^2,
\end{aligned}$$

and

$$\frac{1}{\triangle} E\left[\log(r_{t+\triangle}/r_t)^6 | r_t = r\right] \approx 15\lambda(r_{t-})(\sigma_z^2(r_t))^3,$$

where higher moment properties of the normal distribution were used to find $E\left[Z_1^4\right]$ and $E\left[Z_1^6\right]$.

Using the moment estimates $m_j,\ j = 1, 2, 4, 6$, defined in (4.4) on page 48, the estimate for $\sigma_z^2(r_t)$ is equal to

$$\hat{\sigma}_z^2(r_t) = \frac{3}{15}\frac{m_6(r_t)}{m_4(r_t)},$$

and the estimate for $\lambda(r_{t-})$ is given by

$$\hat{\lambda}(r_{t-}) = \frac{m_4}{3\left[\hat{\sigma}_z^2(r_t)\right]^2}.$$

Therefore, the estimate for $\sigma^2(r_t)$ is given by

$$\hat{\sigma}^2(r_t) = m_2(r_t) - \hat{\lambda}(r_{t-})\hat{\sigma}_z^2(r_t).$$

The estimate $\hat{\sigma}^2(r_t)$ should be positive, therefore the following inequality should be

$$m_2(r_t) \geq \hat{\lambda}(r_{t-})\hat{\sigma}_z^2(r_t).$$

A disadvantage of the moment estimation technique as proposed by Johannes (2004) is that the inequality above may be violated due to small numerical errors.

Applying these estimation formulas to our data we find negative estimates for $\hat{\sigma}^2(r_t)$ at low interest rate levels, as is shown in Figure 4.2. This will result in complex valued diffusion coefficients. A remedy is to redefine the estimate for the diffusion coefficient as

$$\hat{\sigma}(r_t) = \sqrt{\max\left(0, m_2(r_t) - \hat{\lambda}(r_{t-})\hat{\sigma}_z^2(r_t)\right)}.$$

Unfortunately, the total estimated conditional variance from our simulations will then overestimate the true conditional variance.

We also find that the biggest contributing factor to the total conditional variance comes from jumps, where the conditional variance from the jumps is given by $\lambda(r_{t-})\sigma_z^2$. This conclusion follows from Figure 4.2. The estimated conditional variance of the diffusion component, $\hat{\sigma}(r_t)$, is less of a contributing factor to the total estimated conditional variance for all interest rate levels. This result confirms our earlier doubts whether a diffusion component should be included in our model. At least at low interest rate levels no diffusion component is evident.

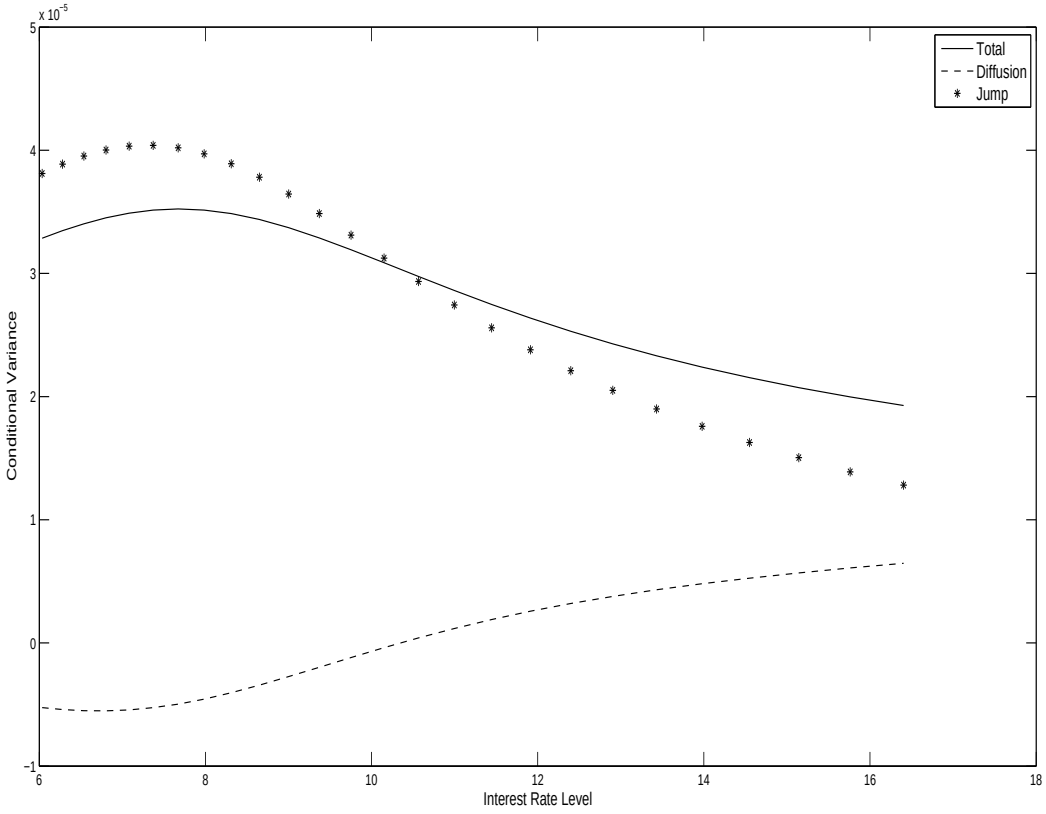

Figure 4.2: The total estimated conditional variance, $m_2(r_t)$, compared to the contributing estimated conditional variance from jumps, $\hat{\lambda}(r_{t-})\hat{\sigma}_z^2$, and from the diffusion component, $\hat{\sigma}^2(r_t)$.

To test for jumps, Monte Carlo paths need to be simulated. The first step is to simulate the number of jumps occurring from one time step to another. If Y_t is a stationary compound Poisson process, then

$$Y_t - Y_{t-\triangle} = \sum_{n=1}^{N_t - N_{t-\triangle}} Z_n,$$

where $N_t - N_{t-\triangle}$ is Poisson distributed with parameter $\lambda\triangle$, and time independent (Section 2.1). However, we have a nonstationary compound Poisson process, where increments $N_t - N_{t-\triangle}$ do not have the same distribution. To simulate paths, we assume that if the distance $\triangle$ is small enough, then $N_t - N_{t-\triangle}$ is approximately Poisson distributed with parameter $\lambda\triangle$.

Therefore, using formula (4.7) on page 52 we can implement the method by Cont and Tankov (2004, p.175) to simulate paths

$\{r_{i,t},\ i = 1, ..., M,\ t = 0, \triangle, ..., T\triangle\}$, starting at $r_{i,0}$, with

$$\log r_{i,t+1} = \log r_{i,t} + \hat{\mu}(r_{i,t})\triangle + \sqrt{\triangle}\hat{\sigma}(r_{i,t})Z_{i,t}^{(1)} + \hat{\sigma}_z \sum_{j=1}^{N_{i,t}} Z_{i,t,j}^{(2)},$$

where $Z_{i,t}^{(1)}$ and $Z_{i,t,j}^{(2)}$ are i.i.d. standard normally distributed random variables and $N_{i,t}$ are Poisson distributed random variables with parameter $\hat{\lambda}(r_{i,t})\triangle$.

We use the same methods to test whether a jump diffusion model can be used to model the short rate, that we used to test whether a single-factor diffusion model is adequate. The graphs shown in Figure 4.3 can be compared to the graphs from Figure 4.1 on page 50. As with the single-factor diffusion model, the jump diffusion model produces interest rate paths with similar conditional drift and conditional variance as the estimated drift and variance from the data. However, from the bottom right graphs, we conclude that the estimated conditional kurtosis from the observed data lies within the 10% and 90% bounds of the estimated conditional kurtosis calculated from the simulated paths. This is a much improved result compared to those obtained from the single-factor diffusion model.

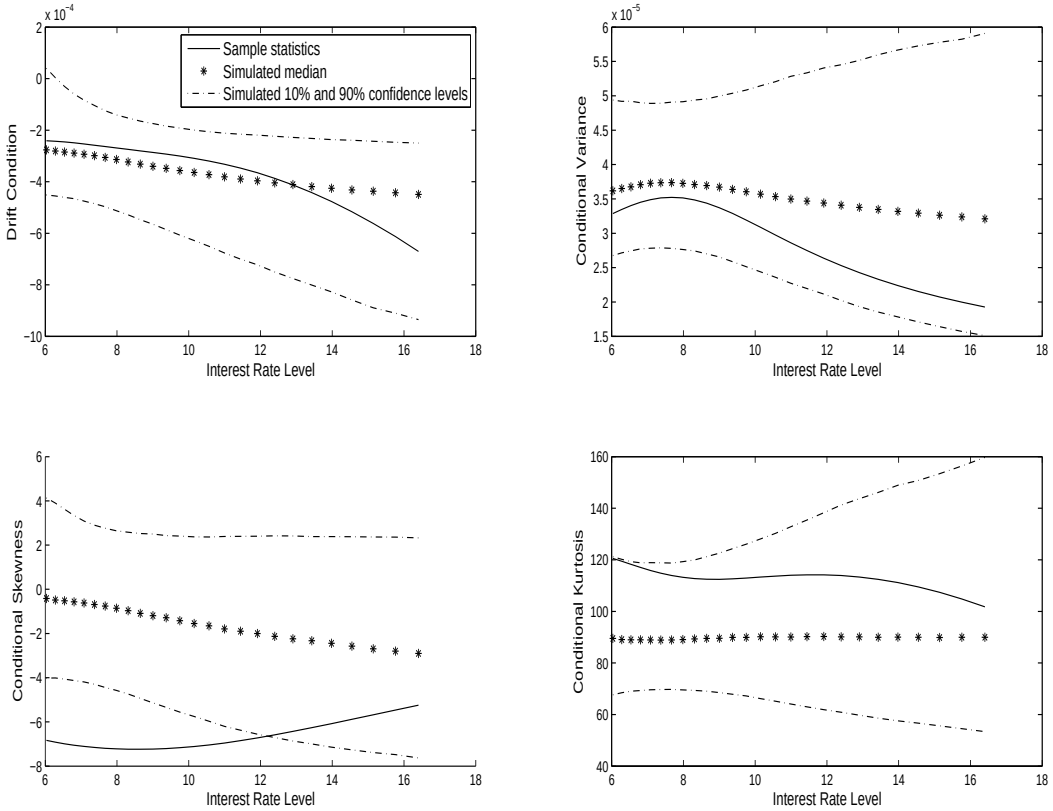

Figure 4.3: The estimated functions $\hat{\mu}$, $\hat{\sigma}$, $\hat{\Sigma}$ and $\hat{\kappa}$ for our dataset and Monte Carlo confidence bands assuming a jump diffusion model.

The formal hypothesis test results are shown in Table 4.2, where we summarise the simulated percentiles for the kurtosis statistic. The sample kurtosis ($\hat{k} = 113.75$) lies within the interval $[\hat{k}^{40\%}, \hat{k}^{60\%}] = [101.14, 116.23]$. Therefore, applying the hypothesis test from Johannes (2004), the null hypothesis that the short rate can be modelled by a jump diffusion model is not rejected at a significance level of $\alpha = 80\%$. Therefore, jumps should be incorporated into the short rate model.

Table 4.2: Kurtosis percentiles for jump diffusion model

Description	Percentile	Kurtosis
sample kurtosis		113.75
simulated kurtosis	1	66.73
	5	75.07
	10	80.59
	25	91.59
	40	101.14
	50	108.10
	60	116.23
	75	132.39
	90	166.38
	95	195.20
	99	272.94

From the results of this section we conclude that a nonparametric single-factor model cannot be used to model the short rate in South Africa. Either jumps should be incorporated (which will result in a jump diffusion model) or a pure jump short rate model should be used. We found that the conditional variance in a jump diffusion model is mostly due to the conditional variance from the jumps (as seen from Figure 4.2 on page 54). We also found difficulties in using moment estimation to estimate the diffusion coefficient. The main conclusion from this is that a pure jump model should be used to model the short rate with the data we used from 1 February 1999 to 17 September 2010. We argued in the preceding chapter that only a pure jump model can be used as a short rate model, given the data from 20 September 2010 to 31 December 2014. We will therefore fit a pure jump model to the JIBAR in Chapter 5 and will customise the fitted model to price interest rate derivatives in Chapter 6.

In the next section the test developed by Lee and Mykland (2008) will be applied to identify isolated jumps.

4.2 Identifying isolated jumps: nonparametric test

In this section a nonparametric test is used to identify isolated jumps from our 3-month JIBAR data. A hypothesis test is used in this method, where the statement that the short rate is given by a nonparametric diffusion process is tested. In the preceding section we asked the same question and used a Monte Carlo based hypothesis test to conclude that a nonparametric diffusion process is not adequate. We define a stochastic variable, $\mathcal{L}$, as

the realised return of the 3-month JIBAR divided by its realised instantaneous volatility. The method of Lee and Mykland (2008) is based on whether $\mathcal{L}$ is within expected bounds if we assume the underlying process is a diffusion process.

In this section a short rate diffusion process is defined by

$$\mathrm{d}r_t = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W_t, \tag{4.11}$$

while a jump diffusion process is given by

$$\mathrm{d}r_t = \mu(t)\mathrm{d}t + \sigma(t)\mathrm{d}W_t + Y(t)N(\mathrm{d}t), \tag{4.12}$$

where N is a point process and Y is the jump size. It differs slightly from the models defined in the previous section as the drift, diffusion and intensity λ are functions of t and not r_t. We apply the method to test whether r_t have jumps, but we get similar results to testing whether $\log r_t$ have jumps. We will therefore only show the results for the models defined in (4.11) and (4.12).

A statistic formed to test for a jump from time t_{i-1} to t_i is defined by

$$\mathcal{L}(i) = \frac{R(t_i) - R(t_{i-1})}{\hat{\sigma}(t_i)},$$

where $R(t_i)$ is the JIBAR at time t_i and

$$\hat{\sigma}^2(t_i) = \frac{1}{K-2}\sum_{j=i-K+2}^{i-1} |R(t_j) - R(t_{j-1})||R(t_{j-1}) - R(t_{j-2})|.$$

The rationale behind this method, as described by Lee and Mykland (2008), is to compare the realised returns of the JIBAR with the local variation coming from the diffusion part of the process. The diffusion is defined in terms of $\sigma(t)$, which is called the instantaneous volatility. The statistic $\hat{\sigma}^2$ is used to calculate the realised bipower variation (discussed in Example 5 on page 17). The statistic is a consistent estimator for σ^2. Importantly, jumps do not affect the consistency of this estimate. Therefore, with the statistic $\mathcal{L}(i)$ we should be able to distinguish whether the realised return is greater (or less) than what one would expect if the underlying process is a diffusion process.

Under the null hypothesis we assume the process in (4.11), while under the alternative hypothesis we assume the process in (4.12). Under the alternative hypothesis, $\mathcal{L}$ can be approximated by

$$\hat{\mathcal{L}}(i) = \frac{Z_i}{c} + \frac{\mathcal{Y}(\tau_i)}{c\sigma\sqrt{\triangle t}},$$

where Z_i is a standard normal variable, $\mathcal{Y}(\tau_i)$ is the actual jump size at actual jump time $\tau_i \in (t_{i-1}, t_i]$, σ is the instantaneous volatility in (4.12), c is given by

$$c = \sqrt{\frac{2}{\pi}},$$

and $\triangle t$ is the length of the sampling interval. Under the null hypothesis $\mathcal{L}$ is approximately distributed as a normal variate with mean zero and variance $\pi/2$, because $\mathcal{Y}(\tau_i)$ equals zero.

The most important result for our purposes from Lee and Mykland (2008) is given by the following Lemma

Lemma 2. *As $\triangle t \to 0$ and under certain conditions for $\mathcal{L}(i)$, K, c, and $\bar{A}_n$, it follows that*

$$\frac{\max_{i \in \bar{A}_n} |\mathcal{L}(i)| - C_n}{S_n} \to \psi, \tag{4.13}$$

where convergence is in distribution, ψ has a cumulative distribution function given by

$$P(\psi \leq x) = \exp\left(-e^{-x}\right),$$

$$C_n = \frac{(2\log n)^{1/2}}{c} - \frac{\log \pi + \log(\log n)}{2c(2\log n)^{1/2}},$$

$$S_n = \frac{1}{c(2\log n)^{1/2}}$$

and where n is the number of observations.

Define

$$T(i) = \frac{|\mathcal{L}(i)| - C_n}{S_n}.$$

If

$$T(i) > -\log\left[-\log(1-\alpha)\right],$$

we conclude that a jump did occur in the interval $(t_{i-1}, t_i]$. The presence of the maximum over $\mathcal{L}(i)$ in formula (4.13) ensures that the overall type I error probability of this procedure is approximately α.

Figure 4.4 shows which of the JIBAR changes in our dataset have been identified as jumps, with $K = 25$ and at a significance level of $\alpha = 5\%$. In total 1555 JIBAR rate changes have been observed from 8 March 1999 (the 26th trading day after 1 February 1999) to 17 September 2010, which is 53% of the total number of trading days in this period. A total of 165 movements were identified as jumps, which is 11% of the total number of movements. Importantly, for more than half of the trading days either a jump or no movement occurred.

In Figure 4.4 the original JIBAR rate realisation over time (top graph) is shown as well as the JIBAR rate changes (bottom graph). From the bottom graph we can see most large JIBAR rate changes were identified as jumps, but in times where the volatility is low small changes were also identified as jumps. The method is therefore effective to distinguish between movements due to diffusion in the market and movements that can be classified as jumps.

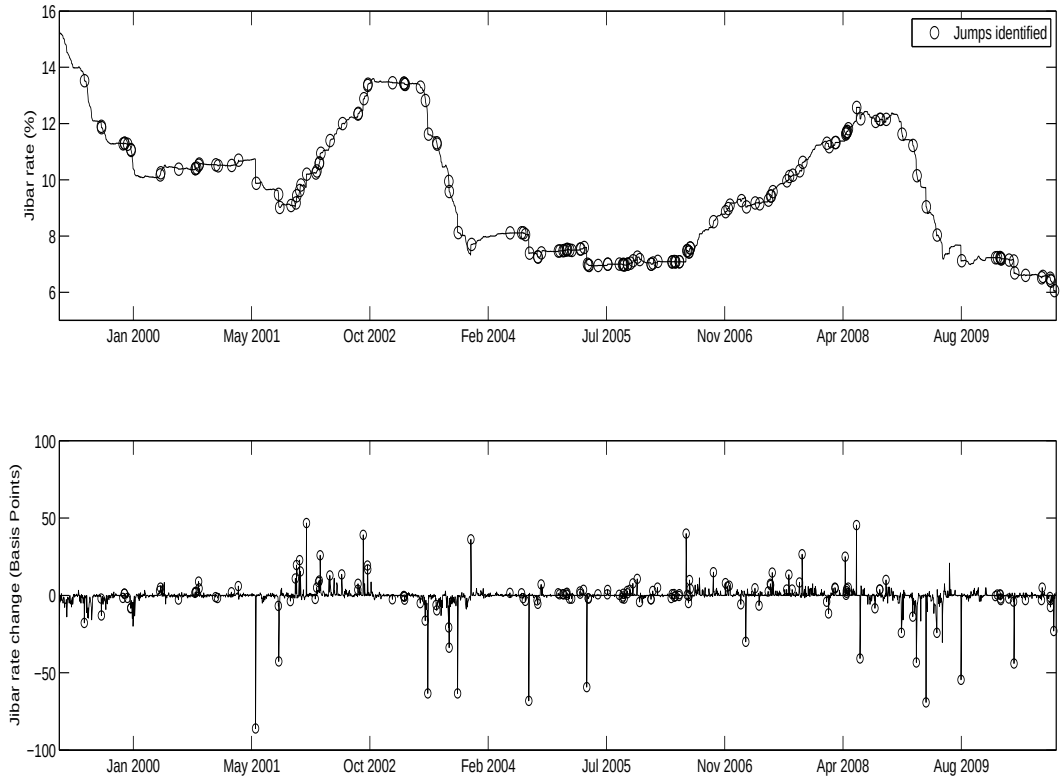

Figure 4.4: The historical JIBAR (top graph) and the first differences(bottom graph), with isolated jumps identified with the nonparametric test from Lee and Mykland (2008).

The conclusion can therefore be made that jumps should be incorporated to accurately model the 3-month JIBAR. The question is whether a pure jump model or a jump diffusion model should be used.

In Figure 4.5 we separate the movements identified as jumps (top graph), from the movements not identified as jumps (bottom graph). We also show the Repo rate changes in the top graph. From the top graph we can conclude that the Repo rate is especially well correlated with large changes in the JIBAR (as expected). However, we cannot say with certainty that all changes identified as jumps are directly due to changes in the Repo rate. Furthermore, there are periods where no clear diffusion component is observed.

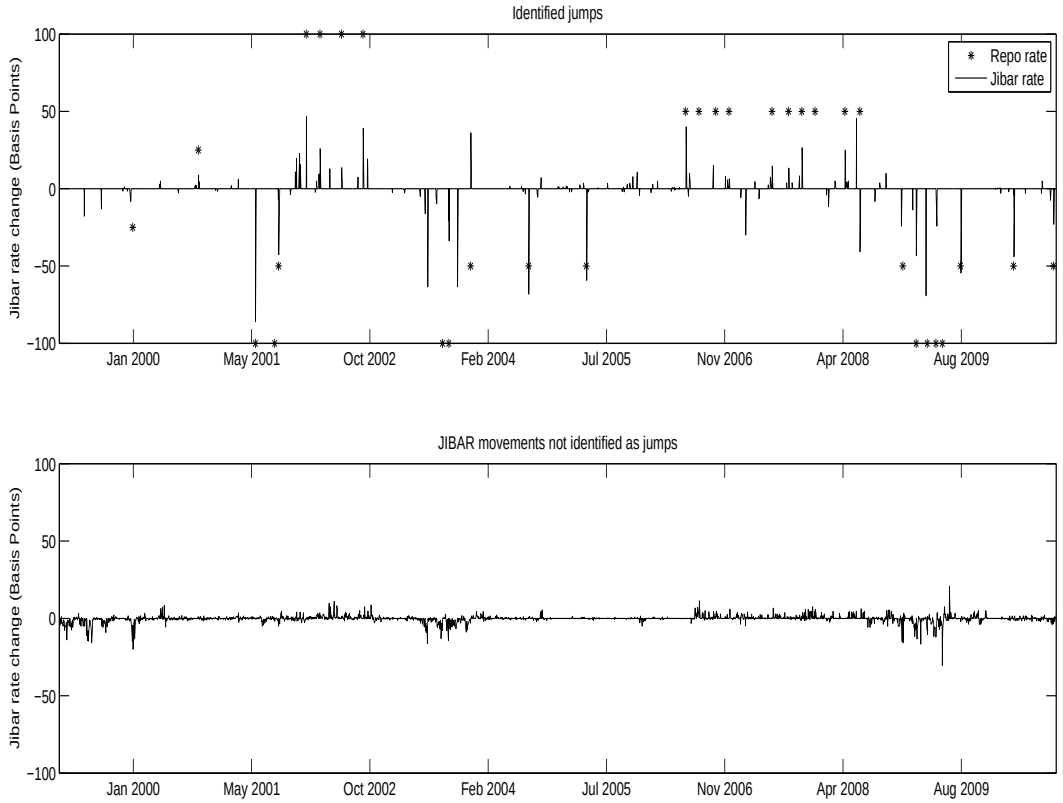

Figure 4.5: Changes in the JIBAR, compared to changes in the Repo rate. The top graph shows the movements identified as jumps by the method from Lee and Mykland (2008). The bottom graph shows the remaining rate changes.

From February 2004 up to June 2006 interest rates were at a relatively low level, where the Repo rate only changed twice. However, in this period a high percentage of movements were identified as jumps and for long consecutive days no rate changes occurred as can be seen from Figure 4.6. A pure jump model should therefore be used for this period.

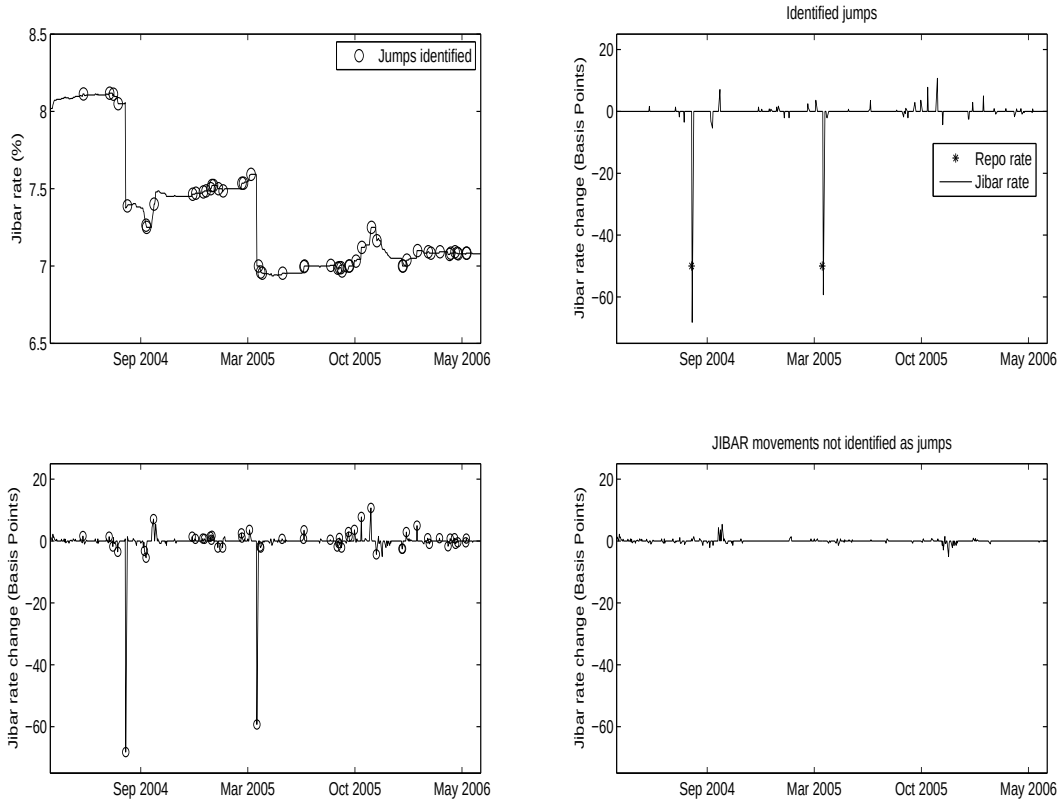

Figure 4.6: The two preceding figures combined for the period from February 2004 to June 2006. These graphs indicate that a pure jump model should be used.

The question is can a pure jump model be used in times where the volatility is high and where a jump diffusion model is possibly more appropriate? We use the period from June 2006 to November 2008, shown in Figure 4.7 that even in a period where interest rates changed frequently, a high number of jumps did occur.

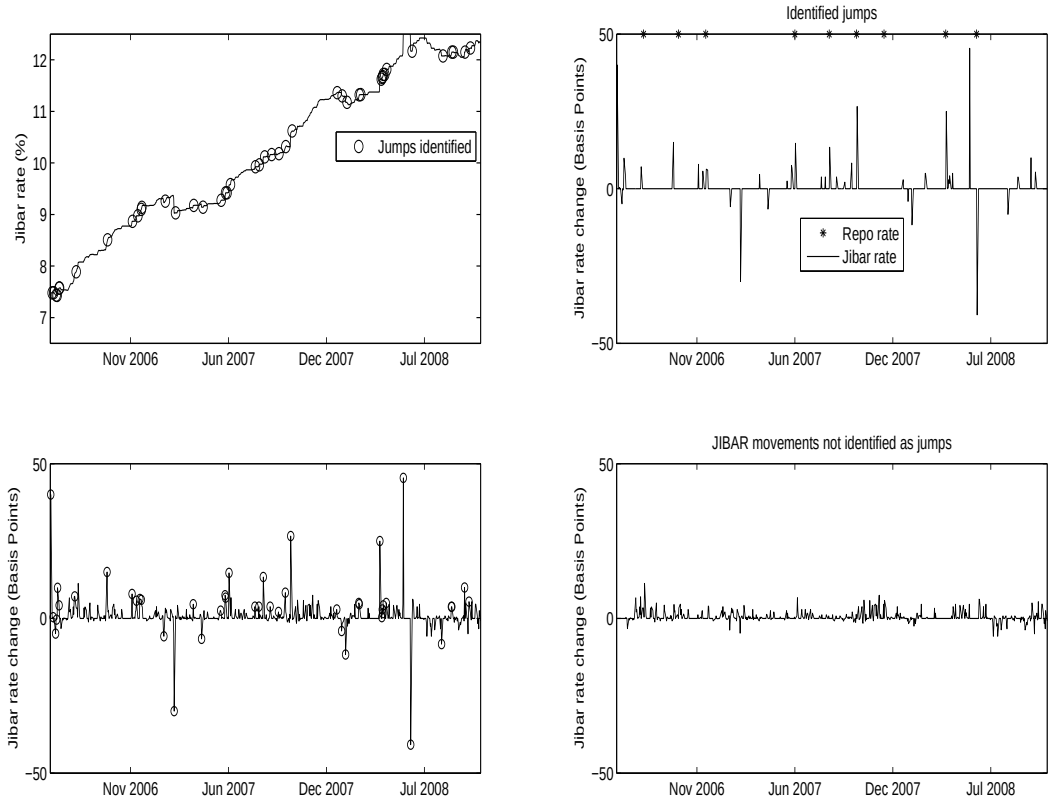

Figure 4.7: The two preceding figures combined for the period from February 2006 to June 2006. These graphs indicate that a pure jump model should be used.

From this section as well as the preceding section we found that when interest rates are at a low level, a pure jump model should be used to model the 3-month JIBAR. For some periods, where interest rates were at a relatively high level, we found a jump diffusion model to be more appropriate. However, our goal is to find a model for all circumstances. If interest rate movements as well as the intensity of the movements are fitted accurately, then a pure jump model can be used in general. We will show in Chapter 5 that these goals can be attained by fitting a compound Poisson model to the JIBAR.

Chapter 5

A pure jump interest rate model

Within various time intervals of the 3-month JIBAR dataset (Figure 3.1 on page 37), interest rate movements can be seen as the realisation of a pure jump process. For example, from Figure 3.3 on page 39 no diffusion component seems to exist. In Section 4.1 we fitted a nonparametric jump diffusion model, where the estimated conditional variance from the jumps is the main driver behind the total estimated conditional variance. The estimated conditional variance obtained from the diffusion component is relatively small, and negligibly small at low interest rate levels. Estimation errors occurred at low interest rate levels. Infrequent jumps at these interest rate levels lead to negative estimated conditional variance emanating from the diffusion component. We also found in Section 4.2 that on more than half of all the trading days from March 1999 to 17 September 2010 either no movements occurred or a jump occurred. Our conclusion from the empirical evidence is that a pure jump model can be used to model the short rate in South Africa. In this section we fit a nonstationary compound Poisson process to the JIBAR rate observed from 1 February 1999 to 31 December 2014.

To fit a pure jump process, two steps are necessary.

(a) We need to fit a distribution to the time between jumps, or in other words, we need to model the jump intensity. If we want to stay within the framework of compound Poisson processes, then the time between jumps should follow an exponential distribution with either a constant parameter λ for a stationary process or $\lambda(r_{t-})$ for a nonstationary process. A variety of distributions are used in Section 5.1 to model the jump intensity. In particular we will use the exponential distribution with a constant parameter and with a rate dependent parameter.

(b) We need to fit a distribution to the jump sizes at all the jumping times. In Section 5.2 we fit a distribution from the family of stable distributions which offers a good fit to the 3-month JIBAR movements.

5.1 Modelling the jump intensity

Let the observed number of days elapsed between successive jumps be given by

$$K_i = T_i - T_{i-1},\ i = 1, ..., N,$$

where $(T_i)_{i\geq 1}$ are the jumping times and N the total number of jumps.

The observed frequencies of days elapsed between jumping times are given by

$$f_k = \sum_{i=1}^{N} \mathbb{1}_{\{K_i = k\}},\ k \in \mathbb{N}. \tag{5.1}$$

Let the time between jumps be defined by the stochastic variable K. If the short rate is modelled by a compound Poisson process, then K is exponentially distributed. This distribution is continuous, whereas the number of days between jumps is discrete.

Now, for $k \in \mathbb{N}$, we define the function

$$f(k;\lambda) = N\mathbb{P}\left(k-1 < K \leq k\right) = N\left[F(k;\lambda) - F(k-1;\lambda)\right], \tag{5.2}$$

where F is the exponential distribution function given by

$$F(k;\lambda) = 1 - e^{-\lambda k}.$$

The observed frequencies in (5.1) can be compared to the expected frequencies, calculated by $f(k;\hat{\lambda})$. The estimate of the parameter λ is calculated by

$$\hat{\lambda} = \frac{1}{\bar{K}},$$

with $\bar{K}$ the sample mean given by

$$\bar{K} = \frac{1}{N}\sum_{i=1}^{N} K_i.$$

The observed frequencies as well as the expected frequencies obtained from the Poisson process are shown in Figure 5.1. Both graphs have a similar shape, but the fit is not good. Note that the maximum number of days is capped at $n = 20$ days, although some outlying observations are greater than 100 days.

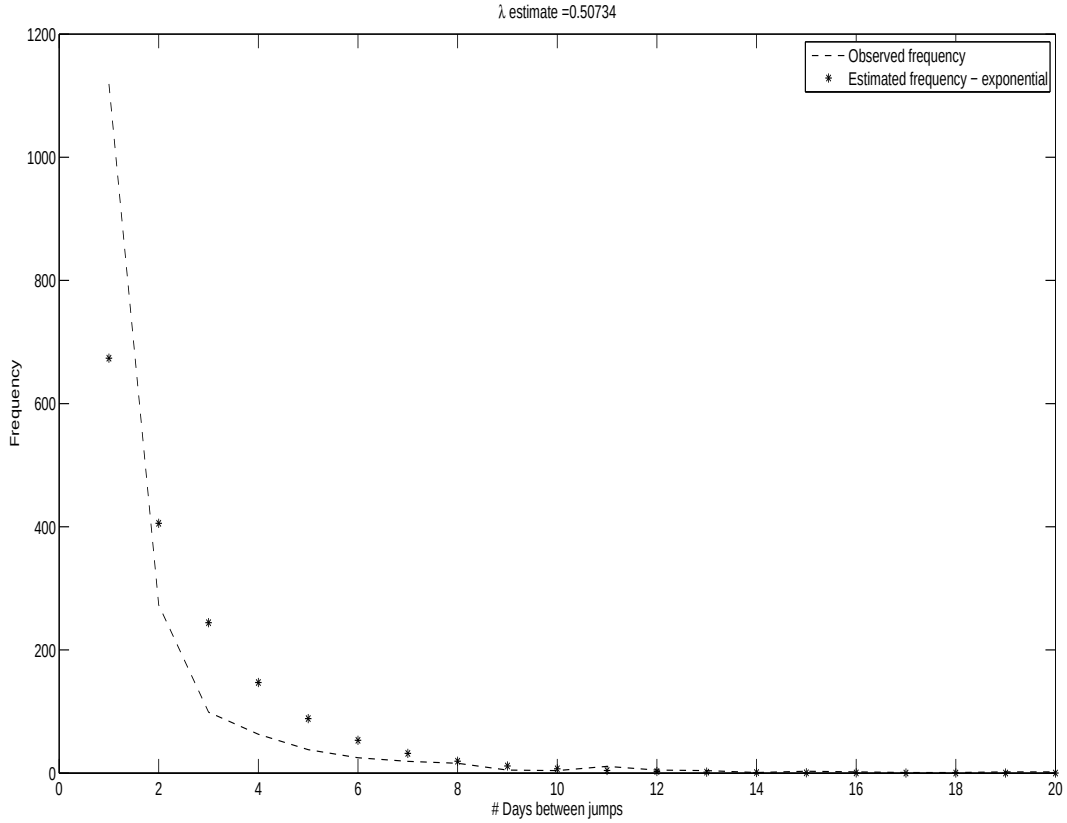

Figure 5.1: Fitted exponential distribution to the days elapsed between 3-month JIBAR movements.

A discrete distribution can be used to model the number of days between jumps. The geometric distribution is defined by the discrete density

$$P(K = k) = p(k) = (1 - \pi)^{k-1}\pi,$$

for $k \in \mathbb{N}$ and $0 \leq \pi \leq 1$. The parameter π is estimated by

$$\hat{\pi} = \frac{1}{\bar{K}}.$$

Another discrete distribution is derived from the geometric distribution by emphasising the weight on the probability of the occurrence of a single day between jumps (i.e., $k = 1$). We define the mixed variable geometric distribution with discrete density as

$$P(X = k) = p(k) = \alpha(1 - \rho)^{k-1}\rho + (1 - \alpha)\mathbb{I}_{\{k=1\}},$$

for $k \in \mathbb{N}$ and $0 \leq \rho \leq 1\,,\ 0 \leq \alpha \leq 1$. The estimates for parameters ρ and α are given by

$$\hat{\rho} = \frac{\sum_{k=2}^{n} f_k}{\sum_{k=2}^{n}(k-1)f_k}$$

and

$$\hat{\alpha} = \frac{\sum_{k=2}^{n} f_k}{(1-\rho)\sum_{k=1}^{n} f_k}.$$

These estimators are obtained through the maximum likelihood estimation method. Notice that

$$\begin{aligned} p(1) &= \alpha\rho + (1-\alpha) = 1 - \alpha(1-\rho), \\ p(2) &= \alpha(1-\rho)\rho, \\ p(3) &= \alpha(1-\rho)^2\rho. \end{aligned}$$

If L is the maximum likelihood function, then $l = \log L$ equals

$$\begin{aligned} l &= f_1 \log(1-\alpha(1-\rho)) + f_2 \log(\alpha(1-\rho)\rho) + ... + f_n \log(\alpha(1-\rho)^{n-1}\rho) \\ &= f_1 \log(1-\alpha(1-\rho)) + \log(\alpha\rho)\sum_{k=2}^{n} f_k + \log(1-\rho)\sum_{k=2}^{n}(k-1)f_k. \end{aligned}$$

Therefore, solving the following two equations for α and p, we get analytic estimates of the parameters.

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= f_1 \frac{\rho - 1}{1 + \alpha(\rho - 1)} + \frac{1}{\alpha}\sum_{k=2}^{n} f_k = 0, \\ \frac{\partial l}{\partial \rho} &= f_1 \frac{\alpha}{1 + \alpha(\rho - 1)} + \frac{1}{\rho}\sum_{k=2}^{n} f_k + \frac{1}{\rho - 1}\sum_{k=2}^{n}(k-1)f_k = 0. \end{aligned} \tag{5.3}$$

From (5.3) we get

$$f_1(\rho - 1)\alpha + (1 + \alpha(\rho - 1))\sum_{k=2}^{n} f_k = 0,$$

i.e.,

$$\alpha(\rho - 1)\left[f_1 + \sum_{k=2}^{n} f_k\right] = -\sum_{k=2}^{n} f_k.$$

Therefore,

$$\alpha = \frac{-\sum_{k=2}^{n} f_k}{(\rho - 1)\sum_{k=1}^{n} f_k}.$$

To solve the equation in (5.3) in terms of ρ we first get an expression for

$$\begin{aligned} f_1 \frac{\alpha}{1 + \alpha(\rho - 1)} &= f_1 \frac{-\sum_{k=2}^{n} f_k}{(\rho - 1)\sum_{k=1}^{n} f_k}\left[1 - \frac{\sum_{k=2}^{n} f_k}{(\rho - 1)\sum_{k=1}^{n} f_k}(\rho - 1)\right]^{-1} \\ &= f_1 \frac{-\sum_{k=2}^{n} f_k}{(\rho - 1)\sum_{k=1}^{n} f_k} \frac{\sum_{k=1}^{n} f_k}{(\sum_{k=1}^{n} f_k - \sum_{k=2}^{n} f_k)} \\ &= \frac{-\sum_{k=2}^{n} f_k}{(p - 1)}. \end{aligned} \tag{5.4}$$

Substituting (5.4) into (5.3) we get

$$\sum_{k=2}^{n} f_k = \rho(-\sum_{k=2}^{n} f_k + \sum_{k=2}^{n} f_k + \sum_{k=2}^{n}(k-1)f_k).$$

Therefore, solving the last equation in terms of p we get the maximum likelihood estimate, $\hat{\rho}$, given by

$$\hat{\rho} = \frac{\sum_{k=2}^{n} f_k}{\sum_{k=2}^{n}(k-1)f_k}.$$

Table 5.1: Parameter estimates

Distribution	parameter	estimate
Exponential	λ	0.507342
Geometric	π	0.507342
Mixed Geometric	ρ	0.349148
Mixed Geometric	α	0.520922

Figure 5.2 shows the goodness of fit for the geometric distribution and the two-factor mixed geometric distribution, compared to the exponential distribution fitted previously. Clearly, the fit provided by the geometric distribution is not a great improvement on the fit obtained from the exponential distribution. The goodness of fit for the mixed geometric distribution seems to be much better.

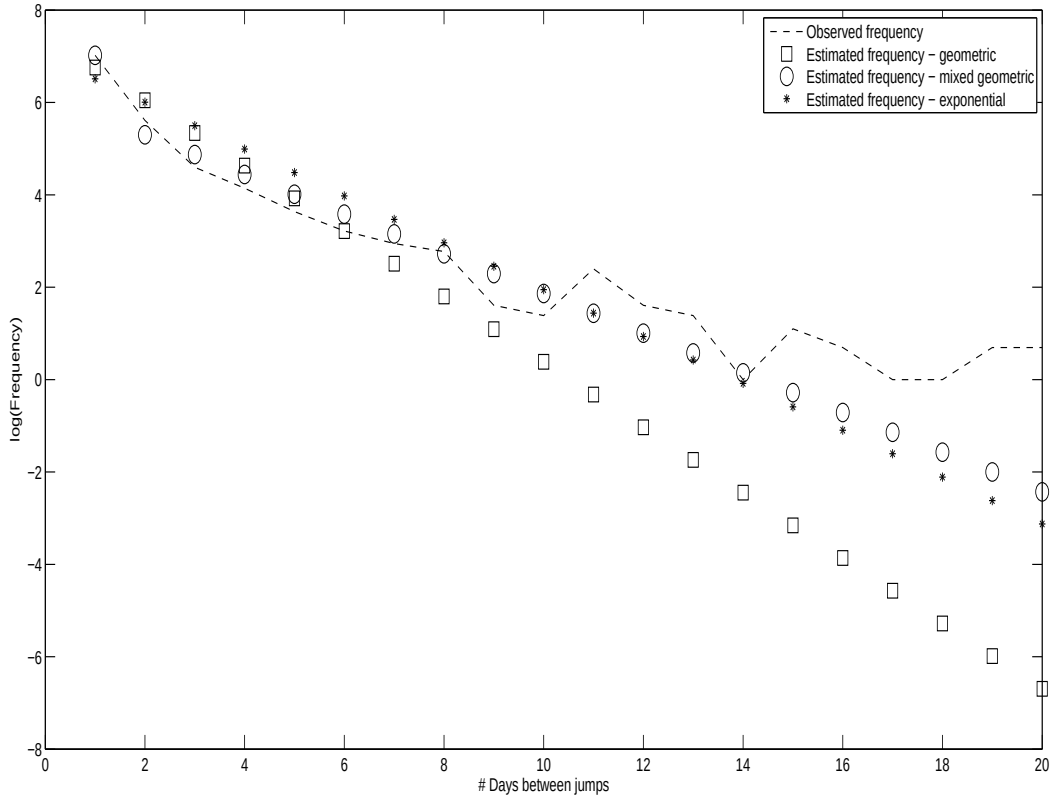

Figure 5.2: The fitted geometric and two-factor mixed geometric distributions compared to the fitted Poisson process. Frequencies are plotted on the log-scale.

A goodness of fit statistic can be defined by

$$X^2 = \sum_{k=1}^{n} \frac{(f_k - f(k))^2}{f(k)},$$

where $f(k)$, $(1 \leq k \leq n)$ are the expected frequencies from either the exponential distribution (5.2), the geometric distribution or the mixed geometric distribution.

The statistic X^2 can be compared to the chi-squared distribution with degrees of freedom $\mathrm{d}f$ to form the Pearson chi-squared test. The number of degrees of freedom is the difference between the number of cells in the frequency table and the number of parameters estimated. The results are shown in Table 5.2. For all distributions, the goodness of fit test fails at a confidence level of 90% if $X^2 > \chi^2_{90\%}$

The number of observed frequency values for $k \geq 12$ are all smaller than 5, which does affect the applicability of the test negatively. However, the difference between the observed and expected frequencies for $k = 1$ are too large for both the exponential and

geometric distributions. Although the mixed geometric distribution fits the observed frequency exactly at $k = 1$ (by definition), the difference between the observed and expected frequencies for $k = 2$ is too large to obtain a good fit.

Table 5.2: Pearson chi-squared test for goodness of fit

Distribution	X^2	df	$\chi^2_{90\%}$
Exponential	525.71	9	14.684
Geometric	438.09	7	12.017
Mixed geometric	51.55	8	13.362

The discretised exponential distribution with a constant parameter λ has been fitted to the time elapsed between JIBAR movements. However, we found other distributions to fit the time between jumps better. The question arises whether we can find a distribution with the advantage of staying within the framework of compound Poisson processes? An exponential distribution with a rate dependent parameter may lead to a better fit, which will lead to a compound Poisson process with nonstationary increments. We will now fit an exponential distribution with rate dependent parameter $\lambda(r_{t-})$ to the time elapsed between jumps and compare the results obtained above.

In Section 4.1 the intensity of a jump diffusion process was estimated by using estimated conditional moments. We found estimation errors of the conditional variance, which might have been due to unreliable higher moment estimates. To avoid possible estimation errors we estimate intensity in this section by estimating the conditional expectation of the short rate intensity, given certain levels of the short rate. To estimate the intensity, the reciprocal of the observed days between JIBAR movements are used as our sample.

Importantly, our estimate for the intensity is a function of the short rate level. By incorporating this property into our model, the short rate will have a rate dependent variance, which is observed in the market. Current literature supports the notion that the volatility in the short rate is high for high levels of the short rate and low for low levels of the short rate (see Chapters 1 and 3).

Let L_i be defined as the reciprocal of the observed days between jumps K_i. An estimate of $\lambda(r_{t-})$ is calculated by

$$\hat{\lambda}(r_{t-}) = \hat{E}\left[\lambda(r_{t-})|r_t = r\right] = \frac{\sum_{i=1}^{N} \phi\left(\frac{r_i - r}{h}\right) L_i}{\sum_{i=1}^{N} \phi\left(\frac{r_i - r}{h}\right)}, \tag{5.5}$$

where r_i is the interest rate level at jump time T_{i-1} for $i = 1, ..., N$. The kernel density ϕ and bandwidth h are defined as in formula (4.4) on page 48.

Figure 5.3 shows how the 3-month JIBAR changes more frequently when interest rates are high, while the intensity is lower when interest rates are low.

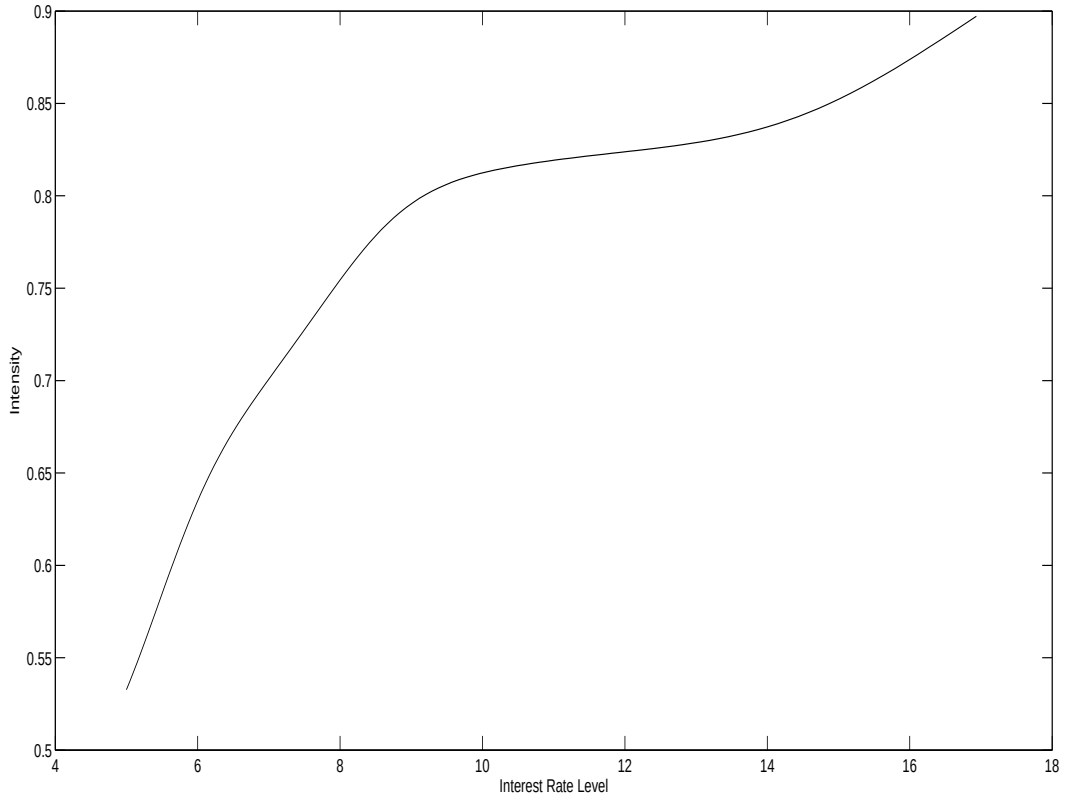

Figure 5.3: The intensity estimate $\hat{\lambda}$ as a function of the interest rate level.

We will now show that the exponential distribution with estimated rate dependent parameter $\hat{\lambda}(r_{t-})$ fits the time between 3-month JIBAR movements data better than by using an exponential distribution with estimated constant parameter $\hat{\lambda}$.

To measure and compare the goodness of fit of the intensity estimate $\hat{\lambda}(r_{t-})$, we use a Monte Carlo simulation method to produce a graph similar to the graphs shown in Figure 5.2 on page 69. To simulate interest paths, we assume that the interest rate levels of the simulated paths are equal to the interest rate levels of our 3-month JIBAR data ($r_i,\ i = 0, ..., N$). This enables us to simulate days elapsed from r_{i-1} to r_i for all $i = 1, ..., N$, using the intensity estimate in (5.5). This removes the possible errors obtained by simultaneously estimating the various interest rate levels and the number of days between interest rate movements.

Let k, $(1 \leq k \leq n)$ be the index of the number of days between jumps. Then, the mean simulated frequency $\bar{f}(k)$ for each k can be calculated by the following algorithm:

(i) Calculate the simulated number of days between jumps $K_{i,j},\ i = 1, ..., N,\ j =$

$1, ..., M$, by

$$K_{i,j} = \lceil -\frac{\log(U_{i,j})}{\hat{\lambda}(r_i)} \rceil,$$

for interest rate level r_i and independent standard uniformly distributed variates $U_{i,j}$.

(ii) Calculate M frequencies for each k by

$$f_j(k) = \sum_{i=1}^{N} \mathbb{1}_{\{K_{i,j}=k\}}.$$

(iii) Estimate the expected frequency by taking the mean of the calculated frequencies $f_j(k)$, i.e.,

$$\bar{f}(k) = \frac{1}{M} \sum_{j=1}^{M} f_j(k).$$

We compare the Monte Carlo simulated frequencies to the frequencies from the geometric, two-factor mixed geometric and Poisson process with constant intensity $\hat{\lambda}$. The results are shown in Figure 5.4. Our simulation parameters are $M = 1,000$ and $N = 1711$, with the latter the number of observed jumps in the JIBAR. The fit seems to be better than at least the Poisson process with constant intensity $\hat{\lambda}$.

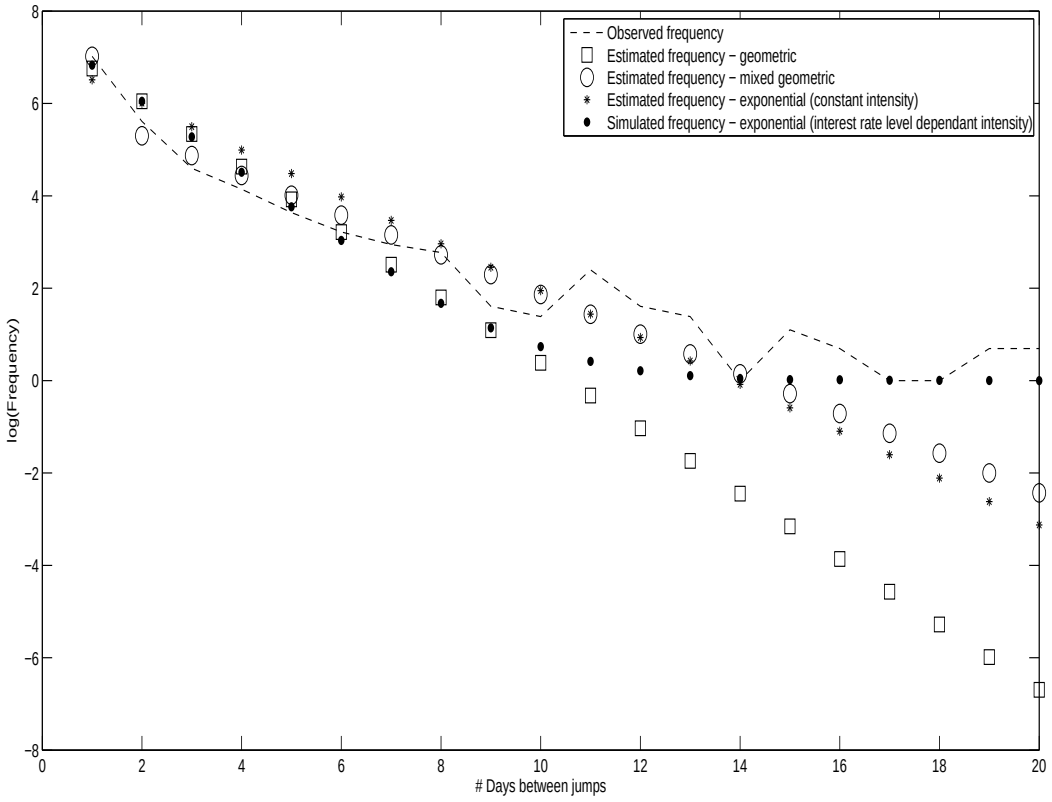

Figure 5.4: The fitted geometric and two-factor mixed geometric distributions compared to the fitted Poisson process with constant intensity $\hat{\lambda}$ and rate dependent intensities $\hat{\lambda}(r_{t-})$. Frequencies are plotted on the log-scale.

In this section, we have shown how well various distributions fit the number of days between jumps. We found the two-factor mixed geometric distribution to be a marginally better fit than an exponential distribution with a rate dependent parameter. However, if we decide to apply the two-factor mixed geometric distribution, then the resulting stochastic process falls outside the scope of compound Poisson processes. This may lead to more involved pricing models. To price interest rate derivatives (Chapter 6), we assume the short rate model to be a nonstationary compound Poisson process with rate dependent intensity. To fit a compound Poisson process to the JIBAR, we first need to fit a distribution to the jumps of the JIBAR.

5.2 Modelling the jump size

In this section several distributions will be fitted to our sample, $X_1, ..., X_n$ ($n = 1711$), of nonzero 3-month JIBAR changes. Q-Q plots as well as a graphical test from Davison and Hinkley (1997, p.150) will be used to determine which distribution gives the best fit.

Firstly, we compare the shape of the relative frequency histogram of the nonzero 3-month JIBAR changes to the shape of the fitted normal density function. The normal density seems to be a poor fit. This is due to the peaks of the histogram from the sample around zero being too high and the tails of the normal distribution not being heavy enough to fit the large movements of the JIBAR. The Cauchy distribution is a distribution with heavy tails and a high peak around its mean. After fitting the Cauchy distribution to the nonzero 3-month JIBAR changes, we found the peak of the Cauchy distribution around the mean to be too high. Details of parameter estimation can be found on page 76 in this section. The tails of the Cauchy distribution are also too heavy. The results are shown in Figure 5.5, where we also show the estimated kernel densities. The kernel densities are calculated by using a routine developed by Dynare (2008), where we choose to use a Gaussian kernel function.

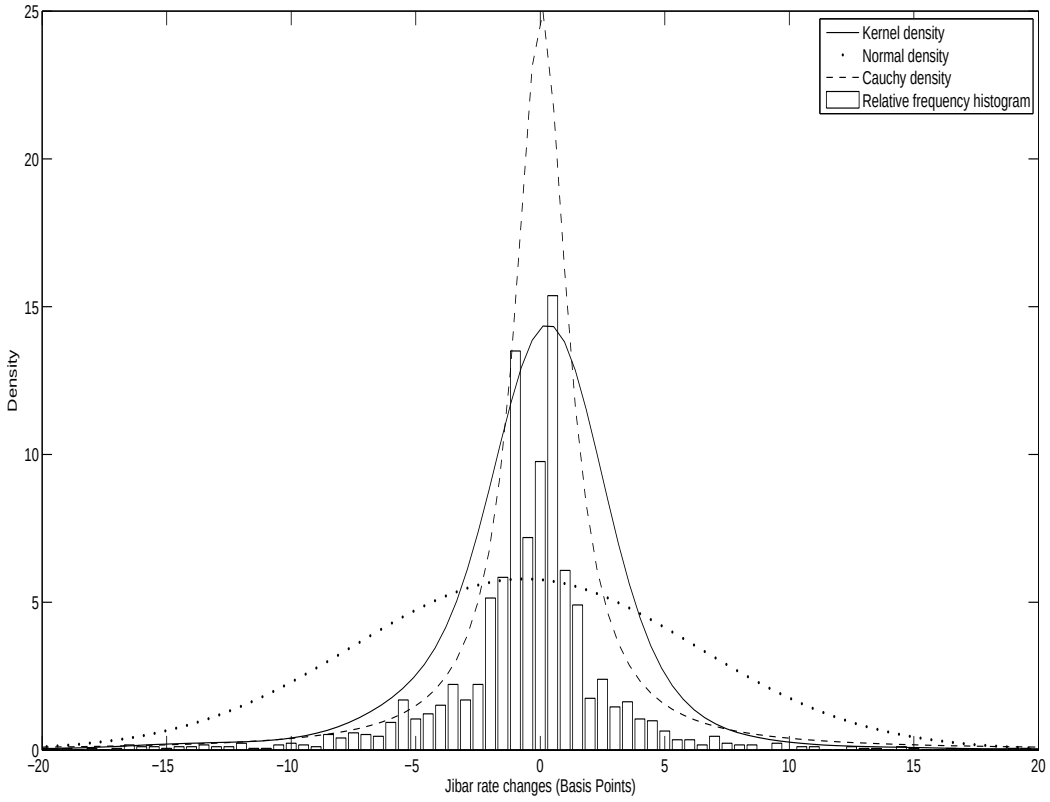

Figure 5.5: The fitted normal and Cauchy densities compared to the relative frequency histogram and kernel density estimate of the nonzero 3-month JIBAR changes. The parameter estimates for the fitted normal and Cauchy densities are shown in Table 5.4 on page 78.

The preliminary results indicate that

(a) a distribution with heavier tails than the normal distribution, but not as heavy as the Cauchy distribution, will provide a good fit.

(b) we are seeking a distribution with a peak around its mean, lower than the Cauchy distribution, but higher than the normal distribution.

The Cauchy as well as the normal distributions belong to the family of stable distributions. Therefore, a distribution from the family of stable distributions will be fitted in this section.

The goodness of fit of the fitted stable distribution will be compared to the goodness of fit of the fitted normal and Cauchy distributions. To determine the goodness of fit

we firstly use Q-Q plots, where the quantiles of the distributions we are investigating need to be calculated. Denote the order statistics of $X_1, X_2, ..., X_n$ by $X_{(1)} < ... < X_{(n)}$. These order statistics will be compared to the $\frac{i}{n+1}$ quantiles of the fitted normal, Cauchy and stable distributions. To distinguish the various distribution the corresponding order statistics are denoted by $X^N_{(i)}$, $X^C_{(i)}$ and $X^S_{(i)}$ respectively. The parameters μ and σ in the normal and Cauchy distributions will be denoted by μ_N, μ_C and σ_N, σ_C respectively.

The density function of the normal distribution is defined by

$$f_N(x; \mu_N, \sigma_N) = \frac{1}{\sigma_N \sqrt{2\pi}} e^{-\frac{1}{2\sigma_N^2}(x-\mu_N)^2}, \ -\infty < x < \infty.$$

Unbiased estimates for the parameters μ_N and σ_N^2 are given by

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

and

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2.$$

If the normal distribution function is given by $F_N(x; \mu_N, \sigma_N)$, then the $\frac{i}{n+1}$-quantile of the fitted normal distribution is given by

$$X^N_{(i)} = F_N^{-1}(\frac{i}{n+1}; \bar{X}, s),$$

where F_N^{-1} is the inverse normal distribution function.

The density function of the Cauchy distribution is defined by

$$f_C(x; \mu_C, \sigma_C) = \frac{1}{\sigma_C \pi} \left[\frac{\sigma_C^2}{\sigma_C^2 + (x-\mu_C)^2}\right], \ -\infty < x < \infty.$$

The maximum likelihood estimates (MLEs) for parameters μ_C and σ_C can be approximated numerically. The log-likelihood function for the Cauchy distribution is given by

$$l(\mu_C, \sigma_C) = -n \log \pi + n \log \sigma_C - \sum_{i=1}^{n} \log\{(X_i - \mu_C)^2 + \sigma_C^2\}.$$

The partial derivatives of l to μ_C and σ_C are given by

$$\frac{\partial l}{\partial \mu_C} = -2 \sum_{i=1}^{n} \frac{X_i - \mu_C}{(X_i - \mu_C)^2 + \sigma_C^2}$$

and

$$\frac{\partial l}{\partial \sigma_C} = \frac{n}{\sigma_C} - 2 \sum_{i=1}^{n} \frac{\sigma_C}{(X_i - \mu_C)^2 + \sigma_C^2}.$$

Therefore, the MLEs, $\hat{\mu}_C$ and $\hat{\sigma}_C$, are the solutions of a system of two nonlinear equations, given by

$$\sum_{i=1}^{n} \frac{X_i - \mu_C}{(X_i - \mu_C)^2 + \sigma_C^2} = 0$$
$$\sum_{i=1}^{n} \frac{\sigma_C^2}{(X_i - \mu_C)^2 + \sigma_C^2} - \frac{n}{2} = 0.$$

We implemented a routine by Axensten (2006) to compute the MLEs of the Cauchy distribution.

The Cauchy distribution function is given by

$$F_C(x; \mu_C, \sigma_C) = \frac{1}{\pi} \arctan\left(\frac{x - \mu_C}{\sigma_C}\right) + \frac{1}{2}.$$

The $\frac{i}{n+1}$-quantile of the fitted Cauchy distribution is calculated by

$$X_{(i)}^C = F_C^{-1}(\frac{i}{n+1}; \hat{\mu}_C, \hat{\sigma}_C),$$

where F_C^{-1} is the inverse Cauchy distribution function given by

$$F_C^{-1}(p, \mu_C, \sigma_C) = \sigma_C \tan\left(\{p - \frac{1}{2}\}\pi\right) + \mu_C. \tag{5.6}$$

Density function values for the family of stable distributions can be calculated by (see Section 8.2)

$$f_S(x; \alpha, \beta, \sigma, \mu) = \frac{1}{\pi} \int_0^\infty e^{-\sigma^\alpha t^\alpha} \cos\left[(\mu - x)t + \sigma^\alpha t^\alpha \beta w(t, \alpha)\right] \mathrm{d}t, \ -\infty < x < \infty.$$

In Section 8.2 we develop a numerical integration method to approximate stable density function values. Subsequently, stable distribution function values can be approximated by

$$F_S(x; \alpha, \beta, \sigma, \mu) \approx \int_{-\sigma c + \mu}^{x} f_S(s; \alpha, \beta, \sigma, \mu)\, \mathrm{d}s,$$

where c, as a function of α, is calculated by (formula (11.4) on page 157)

$$c(\alpha) = a e^{-b(\alpha - 1)}, \ a = 1750, \ b = 4.33.$$

Two methods to estimate parameters from the family of stable distributions are applied in Chapter 9, namely the MLE and ILSE methods. The estimates for the parameters are shown in Table 5.3. The estimates obtained from applying the MLE method are, asymptotically speaking, the most efficient estimates. However, in estimating parameters from the family of stable distributions the running time of the algorithm (or algorithmic

complexity) to calculate the ILSEs is less than the running time of the algorithm to calculate the MLEs (See Section 9.5). This is due to the absence, in general, of a closed form for the density function of a stably distributed random variable. In our application the algoithm's running time is not important and we will therefore use MLEs in the rest of this section.

Table 5.3: Paremeter estimates from family of stable distributions

parameter	MLE	ILSE
α	1.198663	1.198607
β	-0.322786	-0.322872
σ	0.011846	0.007231
μ	-0.010020	-0.008816

The MLEs $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\mu}$ for the family of stable distributions are used to calculate the stable quantiles by

$$X_{(i)}^S = F_S^{-1}(\frac{i}{n+1}, \hat{\alpha}, \hat{\beta}, \hat{\sigma}, \hat{\mu}),$$

where the inverse stable distribution function values were approximated by inverse interpolation.

The values of the parameter estimates for the various fitted distributions are given in Table 5.4.

Table 5.4: Parameter estimates

distribution	parameter	estimate
Normal	μ_N	-0.006330
	σ_N	0.068917
Cauchy	μ_C	0.000416
	σ_C	0.012708
Stable	α	1.198663
	β	-0.322786
	σ	0.011846
	μ	-0.010020

Figure 5.6 shows the Q-Q plots constructed by plotting the points $(X_{(i)}^C, X_{(i)})$, $(X_{(i)}^N, X_{(i)})$ and $(X_{(i)}^S, X_{(i)})$ in the top left, top right and bottom left graphs respectively. In the bottom right graph the three Q-Q plots are plotted on one axis to compare the various fitted distributions. We can conclude that out of the three fitted distributions, the fitted distribution from the family of stable distributions fits the data best.

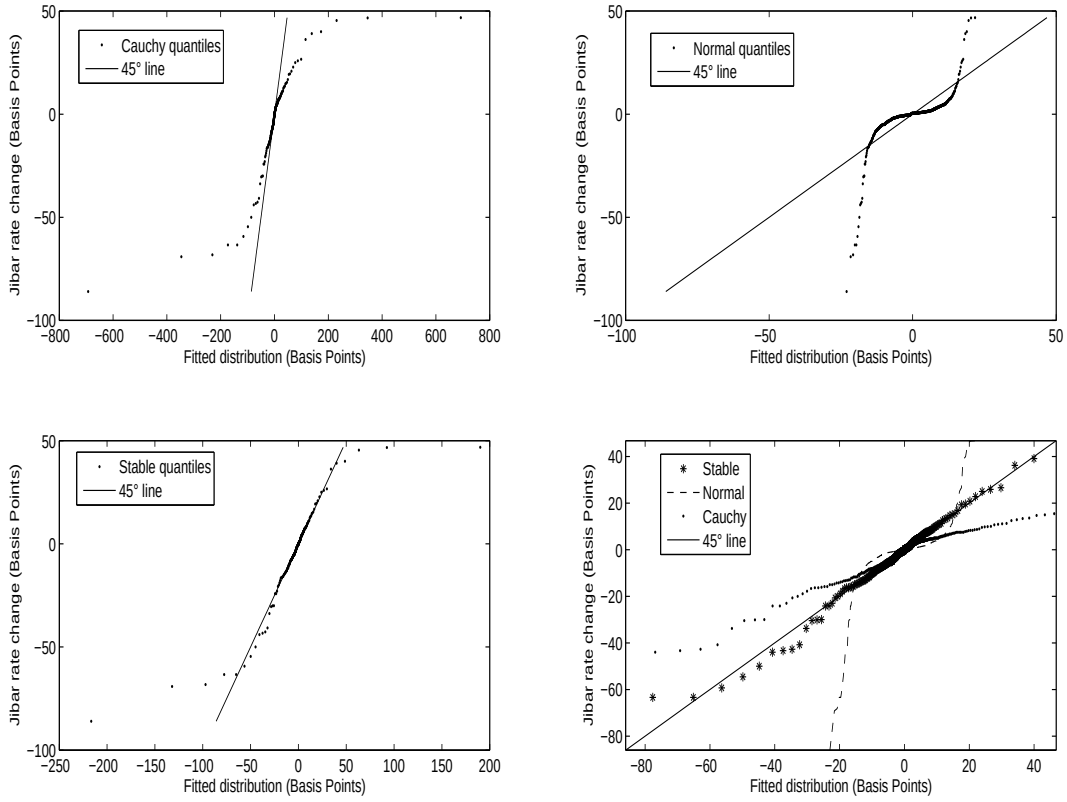

Figure 5.6: Q-Q plots for the quantiles of the 3-month JIBAR changes, compared to quantiles of the fitted Cauchy, normal and stable distributions.

We now use a more formal graphical test from Davison and Hinkley (1997, p.153) to test whether the fitted stable distribution fits the data well. With this test we will compare the sample order statistics with simulated quantiles from the fitted stable distribution. The graphical test compares all points simultaneously, and is therefore a multiple test. To apply the test an overall error rate needs to be approximated. This rate is the chance of any point falling outside a confidence band of quantiles from the fitted stable distribution. An algorithm to approximate the overall error rate, using Monte Carlo simulations, is given by

1. Simulate a sample $X_1^*, ..., X_n^*$ from the family of stable distributions with parameters $\hat{\alpha}$, $\hat{\beta}$, $\hat{\sigma}$ and $\hat{\mu}$.

2. Sort $X_1^*, ..., X_n^*$ to obtain the order statistics $X_{(1)}^* < ... < X_{(n)}^*$.

3. Repeat first two steps m times. The i^{th} order statistic in the j^{th} iteration will be denoted by $X_{(i),j}^*$.

4. For each $i = 1, ..., n$ compute the rank of the m sampled variables $X^*_{(i),j}$, by

$$R_{i,j} = \text{rank}\left\{j : X^*_{(i),j}\right\}.$$

5. Let α be the significance level of a two-sided pointwise test, described in Davison and Hinkley (1997, p.150), and $k = \frac{\alpha}{2}(m+1)$. The overall error rate, p, is approximated by

$$p \approx \frac{1}{m}\sum_{j=1}^{m} I_j,$$

where I_j are indicator function values described by

$$I_j = \begin{cases} 1 & \text{if } R_{i,j} <= k \text{ or } R_{i,j} >= m+1-k \text{ for any } i = 1, ..., n \\ 0 & \text{otherwise.} \end{cases}$$

To implement the algorithm, we use the parameter values: $m = 4999$, $\alpha = 0.1\%$ and n the size of our sample. A significance level of the pointwise test at $\alpha = 0.1\%$ leads to an overall error rate of $p \approx 10\%$. Figure 5.7 shows that our sample order statistics $X_{(i)}$ falls within the 90% confidence bands for all $i = 1, 2, ..., n$. The fitted stable distribution is therefore not rejected as a good fit at a 90% confidence level.

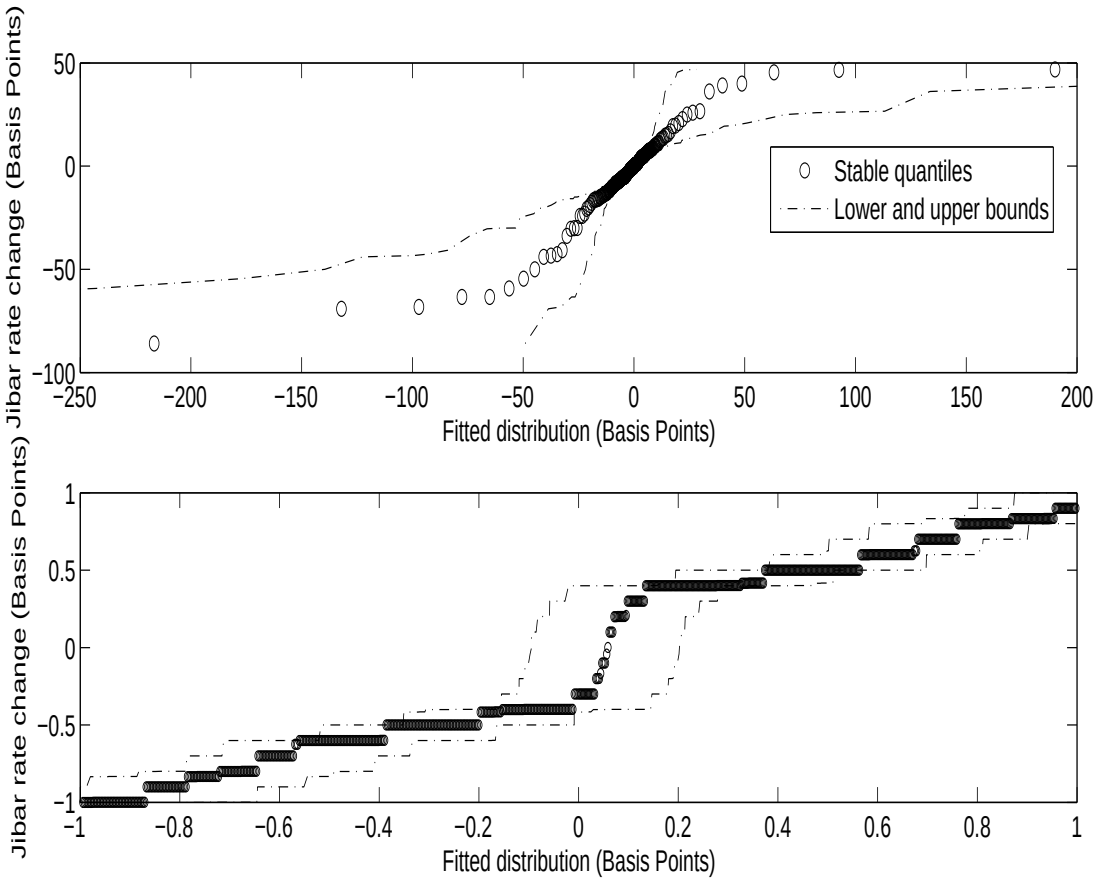

Figure 5.7: Multiple graphical goodness of fit test for the family of stable distributions. The sample order statistics falls within the 90% confidence bands of the quantiles, simulated from the fitted stable distribution. The top graph shows the result over the entire range of sampled basis point values, while the bottom graph shows the results for the range $[-1, 1]$.

In this Chapter we fitted a nonstationary compound Poisson process to the 3-month JIBAR data. We found that a distribution from the family of stable distributions fits the nonzero JIBAR movements well. We also found that the exponential distribution with a rate dependent intensity fits the time between jumps better than an exponential distribution with a constant parameter. Our aim in this chapter was to fit a pure jump stochastic process to the JIBAR as a proxy for the South African short rate. However, our main aim in our study is to apply the model to price interest rate derivatives. In the next chapter we will discuss why we need to adjust the short rate model fitted in this chapter in order to price interest rate derivatives.

Chapter 6

Pricing interest rate derivatives

In this chapter we will address our main objective, which is, pricing of interest rate derivatives. Analytical solutions of interest rate derivative prices do not exist if we apply the model developed in Chapter 5. As a result, we will price various financial instruments numerically. Numerical methods to price derivatives can be divided into three general methods. They are lattice methods, methods to solve partial differential equations numerically and Monte Carlo simulation. An algorithm to simulate stably distributed random variables, with a relatively small cost in terms of computational time, exists (see Section 8.4). This makes the Monte Carlo methodology relatively easy to implement.

Unfortunately, we cannot use the model developed in Chapter 5 directly as a pricing model. The reason for this is, as we will now show, that using stably distributed jumps are not practical. We therefore need to modify the model.

The price of an option is the expected discounted payoff under a risk-neutral measure $\mathbb{P}^*$. Therefore, if $\phi(S, r_S)$ is the payoff at time S, then the price of the option at time t is given by

$$\Phi(t, r_t) = E^* \left[e^{-\int_t^S r_s \mathrm{d}s} \phi(S, r_S) | \mathcal{F}_t \right],$$

where the expectation is taken under $\mathbb{P}^*$ (Brigo and Mercurio, 2006, p.39). Here ϕ does NOT denote a standard normal density function.

To apply the method of Monte Carlo simulation to price the option one therefore needs to simulate interest rate paths r_s for $t \leq s \leq S$ and calculate a discounted payoff

$$X = e^{-\int_t^S r_s \mathrm{d}s} \phi(S, r_S)$$

for each simulated path. Let $X_1, ..., X_M$ the discounted payoffs calculate on M simulated paths. Then

$$\bar{X}_M = \frac{1}{M} \sum_{i=1}^{M} X_i$$

is an estimate of $\Phi(t, r_t)$.

If $\text{Var}(X_i) < \infty$, then the absolute estimation error is bounded at a specific confidence level. From the Central Limit Theorem (Theorem **??** on page **??**) with a probability of $(1-\alpha)\%$ the absolute Monte Carlo estimation error,

$$|\bar{X}_M - \Phi(t, r_t)|,$$

is bounded by $z_{1-\frac{\alpha}{2}}\sqrt{\frac{\text{Var}(X_i)}{n}}$.

However, if the short rate r_t is modelled by a compound Poisson process and each random jump has an infinite second moment, then the variance of the short rate is infinite. Although the discount factor is bounded for positive interest rates the discounted payoff may not be bounded which may lead to an infinite variance of the discounted payoff. This makes the model developed in Chapter 5 not applicable since stably distributed random variables have an infinite second moment for $\alpha < 2$ (Fama and Roll, 1968). Our solution to the problem is to define jumps that are truncated.

Let X be a stably distributed random variable and $L \in \mathbb{R}_+$. A truncated stably distributed random variable can be defined as

$$Y = X\mathbb{1}_{\{|X|\leq L\}} + L\mathbb{1}_{\{X>L\}} - L\mathbb{1}_{\{X<-L\}}.$$

The density function of Y is given by

$$f_Y(y) = f_X(y)\mathbb{1}_{\{|y|<L\}} + \mathbb{P}(X > L)\mathbb{1}_{\{y=L\}} + \mathbb{P}(X < -L)\mathbb{1}_{\{y=-L\}},\ -L \leq y \leq L,$$

where f_X is the density function of X. Importantly, the second moment (and all finite moments) of the truncated stable distribution is, as expected, finite since $|Y| \leq L$.

The value of the parameter L is not estimated using statistical methods, but is an exogenous variable. In our applications the value of L is chosen as 50 basis points. The reason is that history shows the SARB mostly changes the repo rate with either 50 or 25 basis points. We take the larger of the two values as our maximum allowed jump size.

The first problem in applying the model developed in Chapter 5 stems from the infinite variance of stably distributed jumps. This is solved by defining truncated stably distributed jumps. The second problem with the model is that estimates were obtained from historical data under a market measure $\mathbb{P}$. However, pricing occurs under a risk-neutral measure $\mathbb{P}^*$. The problem is that the jump intensity as well as all of the parameters of the jump distribution changes with a change of measure. This was shown in Example 19 on page 26 and Example 20 on page 29 as an application of Girsanov's theorem. The market price of risk gives us an indication of how the parameters in a model change with a change of measure. We will find a formula for the market price of risk in both pure jump and one-factor diffusion models in Section 6.1.

A valid question is why did we model the dynamics of the short rate under $\mathbb{P}$, if all the estimates will change? Chapter 5 gave us evidence of thick-tailed stably distributed

jumps and a jump intensity dependent on r_t. We will incorporate these results by defining a model under a measure $\mathbb{P}^*$ in Section 6.2, which can be used to price interest rate derivatives. Therefore, the short rate model used for pricing interest rate derivatives will both have characteristics observed from fitting a model to historical rates as well as some of the characteristics of a proper pricing model.

The impact of a pure jump model on a European call option on the forward JIBAR rate will be analysed in Section 6.3. We will also investigate the impact on a more exotic option than a European call option. We choose a barrier option with the forward JIBAR as underlying rate, which will be priced in Section 6.4.

6.1 Market price of risk

In this section we will show how the dynamics of both a one-factor diffusion and pure jump model change with a change of measure. Both models have a market price of risk, which influences model calibration. An asset under the risk-neutral measure $\mathbb{P}^*$ has a risk-free return in a no-arbitrage model, while the actual return of an asset under a market measure $\mathbb{P}$ may differ. This is due to the risk associated with the asset. The market price of risk is therefore used to measure the excess return in the market for each unit of risk.

A unit of risk can be measured by the instantaneous volatility of a model. In the general one-factor diffusion model under a measure $\mathbb{P}$, defined in (4.1) on page 47, the instantaneous volatility can be calculated in terms of the quadratic variation or conditional quadratic variation of a process (Chapter 2). Therefore, the instantaneous volatility in a one-factor diffusion model can be calculated by

$$\sigma(r_t) = \sqrt{\frac{\mathrm{d}\langle r, r\rangle_t}{\mathrm{d}t}} = \frac{1}{\sqrt{\mathrm{d}t}}\sqrt{E\left\{(\mathrm{d}r_t)^2 \mid F_{t-}\right\}},$$

where $\langle r, r\rangle$ is the conditional variation of the r process defined in formula (2.7) on page 23. The market price of risk in a general short rate model can thus be calculated by

$$\eta(r_t) = \frac{\frac{1}{\mathrm{d}t}E\left\{\mathrm{d}r_t \mid F_{t-}\right\} - \frac{1}{\mathrm{d}t}E^*\left\{\mathrm{d}r_t \mid F_{t-}\right\}}{\frac{1}{\sqrt{\mathrm{d}t}}\sqrt{E\left\{(\mathrm{d}r_t)^2 \mid F_{t-}\right\}}} = \frac{1}{\sqrt{\mathrm{d}t}}\frac{E\left\{\mathrm{d}r_t \mid F_{t-}\right\} - E^*\left\{\mathrm{d}r_t \mid F_{t-}\right\}}{\sqrt{E\left\{(\mathrm{d}r_t)^2 \mid F_{t-}\right\}}}. \quad (6.1)$$

The excess return is calculated in terms of the conditional expectations of the short rate dynamics under $\mathbb{P}$ and $\mathbb{P}^*$, while a unit of risk is measured in terms of the conditional quadratic variation. Our formula to calculate the market price of risk is similar to the informal interpretation of the Sharpe ratio from Bjork (2004, p. 203). The Sharpe ratio is equivalent to the market price of risk for one-dimensional processes.

The market price of risk for a diffusion model can be calculated by applying Girsanov's theorem to the one-factor diffusion model defined in (4.1) on page 47. From Example 17

on page 25 we have the result that if

$$\mathrm{d}L_t = L_t h(r_t)\mathrm{d}W_t,$$

where L_t then

$$\mathrm{d}W_t^* = \mathrm{d}W_t - h(r_t)\mathrm{d}t,$$

where W^* is a Brownian motion under $\mathbb{P}^*$. The first conditional moment under $\mathbb{P}^*$ of short rate changes can be written as

$$E^*\left\{\mathrm{d}r_t \,|\, F_{t-}\right\} = \mu(r_t)\mathrm{d}t + \sigma(r_t)E^*\left\{\mathrm{d}W_t \,|\, F_{t-}\right\} = \left[\mu(r_t) + \sigma(r_t)h(r_t)\right]\mathrm{d}t.$$

If the drift coefficient under $\mathbb{P}^*$ is given by $\mu^*(r_t)$, then

$$\mu^*(r_t) = \mu(r_t) + \sigma(r_t)h(r_t),$$

which implies that the market price of risk $\eta(r_t)$ is given by

$$\eta(r_t) = \frac{\mu(r_t) - \mu^*(r_t)}{\sigma(r_t)} = -h(r_t). \tag{6.2}$$

Also take note of the fact that the market price of risk does not influence the second conditional moment of short rate changes. This follows from the formula

$$E^*\left\{(\mathrm{d}r_t)^2 \,|\, F_{t-}\right\} = \sigma^2(r_t)\mathrm{d}t,$$

which implies that the diffusion coefficient under $\mathbb{P}^*$ is equal to the diffusion coefficient under $\mathbb{P}$.

If a pure jump model under $\mathbb{P}$ is defined by

$$\mathrm{d}r_t = \mu(r_t)\mathrm{d}t + Z_{N_t}N(\mathrm{d}t), \tag{6.3}$$

where the Z_n, $n \in \mathbb{N}$, are i.i.d. jumps under $\mathbb{P}$ with jump distribution ν and N is a nonstationary Poisson process under $\mathbb{P}$. Let

$$\mathrm{d}L_t = L_{t-}h(r_{t-})\mathrm{d}X_t,$$

where X_t is a martingale with dynamics given by

$$\mathrm{d}X_t = Z_{N_t}N(\mathrm{d}t) - \lambda(r_{t-})E_\nu[Z_1]\mathrm{d}t,$$

where E_ν is expectation taken under jump distribution ν. Similar to Example 19 on page 26 and Example 20 on page 29 we find that the conditional expectation of $\mathrm{d}X_t$ under $\mathbb{P}^*$ is given by

$$E^*\left\{\mathrm{d}X_t \,|\, F_{t-}\right\} = h(r_{t-})\lambda(r_{t-})E_\nu[Z_1^2]\mathrm{d}t.$$

Therefore, the conditional expectation of $\mathrm{d}r_t$ under $\mathbb{P}^*$ is given by

$$E^*\left\{\mathrm{d}r_t \,|\, F_{t-}\right\} = \left\{\mu(r_t) + \lambda(r_{t-})E_\nu[Z_1] + h(r_{t-})\lambda(r_{t-})E_\nu[Z_1^2]\right\}\mathrm{d}t. \tag{6.4}$$

Applying (6.1) on page 84, the market price of risk in a pure jump model $\eta(r_t)$ is given by

$$\eta(r_t) = \frac{-h(r_{t-})\lambda(r_{t-})E_\nu[Z_1^2]}{\sqrt{\lambda(r_{t-})E_\nu[Z_1^2]}} = -h(r_{t-})\sqrt{\lambda(r_{t-})E_\nu[Z_1^2]}. \tag{6.5}$$

We will also show that the conditional second moment of the short rate with a pure jump model is influenced by the market price of risk. Let the dynamics of a martingale D_t be given by

$$\mathrm{d}D_t = (\mathrm{d}X_t)^2 - \lambda(r_{t-})E_\nu[Z_1^2]\mathrm{d}t.$$

We can apply Girsanov's theorem on page 24 to find that the conditional expectation of $\mathrm{d}D_t$ under $\mathbb{P}^*$ is given by

$$\begin{aligned} E^*\left\{\mathrm{d}D_t \,|\, F_{t-}\right\} &= \frac{1}{L_{t-}}E\left\{\mathrm{d}L_t\mathrm{d}D_t \,|\, F_{t-}\right\} \\ &= E\left\{h(r_{t-})\mathrm{d}X_t\mathrm{d}D_t \,|\, F_{t-}\right\} \\ &= h(r_{t-})E\left\{(\mathrm{d}X_t) \mid F_{t-}\right\} \\ &= h(r_{t-})\lambda(r_{t-})E_\nu[Z_1^3]\mathrm{d}t. \end{aligned}$$

Therefore, the conditional second moment of $\mathrm{d}r_t$ under $\mathbb{P}^*$ is given by

$$E^*\left\{(\mathrm{d}r_t)^2 \mid F_{t-}\right\} = E^*\left\{(\mathrm{d}Y_t)^2 \mid F_{t-}\right\} = \left\{h(r_{t-})\lambda(r_{t-})E_\nu[Z_1^3] + \lambda(r_{t-})E_\nu[Z_1^2]\right\}\mathrm{d}t, \tag{6.6}$$

which is influenced by the market price of risk for pure jump models. This result differs from the corresponding result for one-factor diffusion models.

In the following sections we will price interest rate derivatives by first assuming the market price of risk is zero. This implies model calibration to market volatilities are done similarly to diffusion models, where the conditional second moment of short rate changes do not change with a change of measure. Importantly, if we find that there is an impact on interest rate derivative prices when no market price of risk is evident in the market, then the impact will possibly be even greater in the presence of a non-zero market price of risk.

6.2 The $\mathbb{P}^*$-dynamics of the short rate

In this section a pure jump short rate model under a risk-neutral measure $\mathbb{P}^*$ is defined. The pure jump model in Chapter 5 was developed using only historical data (under a measure $\mathbb{P}$). However, pricing of short rate models occur under $\mathbb{P}^*$. In this section a

model will be defined which will incorporate some important characteristics of the model developed in Chapter 5, while being adapted for pricing purposes.

The most important features of the nonstationary compound Poisson model developed in Chapter 5 are heavy-tailed jumps as well as rate dependent intensities of the jumps. We will define a model in this section with the same properties. However, some practical features such as pricing derivatives with Monte Carlo simulation as well as the possibility of mean reversion of the short rate will be incorporated.

The general short rate model we will use to price interest rate derivatives is defined in (2.12) on page 34. Some of the reasons for choosing this model are:

i The finite second moments of the jumps ensure efficiency of the Monte Carlo pricing method. As a results of the model fitted in the previous chapter, the jumps will be stably distributed with truncated jumps. Although this will influence the heavy tailed feature, the jumps will be truncated at levels to prevent unrealistic short rate movements.

ii Model flexibility through choosing an appropriate drift coefficient μ. Some features such as mean reversion of the short rate can be enabled by a certain form of μ (see Example 21 on page 30). The skewness and location parameters of the stably distributed jumps are set equal to zero. This will ensure the first moment of each random jump to be zero. We showed in Example 21 on page 30 that mean reversion of the short rate in a nonstationary compound Poisson model is only possible if $E_{\nu^*}[Z_1] = 0$.

iii The requirement that $E_{\nu^*}[Z_1] = 0$ ensures that the jumps will only impact the stochastic part of the model, and not the deterministic drift in the model. One-factor diffusion models are also split into two parts, a deterministic drift part and a diffusion part. Therefore, this requirement will enable us to analyse the impact on interest rate derivative prices from having a model with jumps instead of a diffusion component.

We now address the question of calibrating our model to the market. The market is represented by a measure $\mathbb{P}$? From previous discussions, we know that the diffusion coefficient in a one-factor pure diffusion model does not change with a change of measure. The assumption is therefore made that the market's prediction of volatility is correct. By assuming a zero market price of risk, we will assume the same in our model. Therefore, from (6.6) on page 86 and (2.13) on page 35 we have that

$$E\left\{(\mathrm{d}r_t)^2 \mid F_{t-}\right\} = E^*\left\{(\mathrm{d}r_t)^2 \mid F_{t-}\right\} = \lambda^*(r_{t-})E_{\nu^*}\left[Z_1^2\right].$$

In the next couple of sections we will price interest rate derivatives to investigate the possible impact of a compound Poisson model with truncated stably distributed jumps,

compared to a pure diffusion model given by

$$\mathrm{d}r_t^c = \mu(r_t)\mathrm{d}t + \sigma(r_t)\mathrm{d}W_t. \tag{6.7}$$

A few assumptions are made which will assist us in analysing how jumps influence the price of interest rate derivatives.

Assumption 1. *The jumps of the short rate under a risk-neutral measure $\mathbb{P}^*$ are truncated stably distributed with no change of parameter values with a change of measure. The parameter estimates for α and σ are therefore given in Table* 5.4*, where $\alpha = 1.198663$ and $\sigma = 0.011846$.*

Assumption 2. *The dynamics of the short rate has no drift under $\mathbb{P}^*$. To ensure this we will set the drift parameter $\mu(r_t)$ in both models equal to zero.*

Assumption 1 may have an impact on our conclusions concerning the question whether we obtain different derivative prices when comparing our pure jump model with a pure diffusion model. We will therefore test Assumption 1 in the next section as it is the most naive assumption that we make.

In order to estimate prices of interest rate derivatives numerically, we will use a Monte Carlo method to simulate interest rate paths (the same method described in Section 4.1).

Therefore, interest rate paths $\{r_{i,t},\ i = 1, ..., M,\ t = 0, \triangle, ..., T\triangle\}$ are simulated, starting at $r_{i,0}$ by

$$r_{i,t+\triangle} = r_{i,t} + \hat{\mu}(r_{i,t})\triangle + \sum_{j=1}^{N_{i,t}} Z_{i,t,j},$$

where the $Z_{i,t,j}$ are i.i.d. truncated stably distributed random variables and the $N_{i,t}$ are Poisson distributed random variables with parameter $\hat{\lambda}(r_{i,t})\triangle$. We can estimate $E\left[Z_1^2\right]$ by m Monte Carlo simulations Z_k, $k = 1, 2, ..., m$. Therefore, $\hat{\lambda}(r_{i,t})$ is given by

$$\hat{\lambda} = \frac{m_2(r_t)}{\frac{1}{m}\sum_{k=1}^{m} Z_k^2},$$

where $m_2(r_t)$ is calculated by (4.4) on page 48.

We expect the intensity estimates from our current model to differ from those of the model developed in Section 5.1. This is due to the different estimation method used in the current model. By using this method we ensure that the conditional second moments in both the pure jump and one-factor diffusion models are equal. Figure 6.1 shows that the intensity estimates do differ somewhat. However, both exhibit an overall increasing tendency as the interest rate level increases.

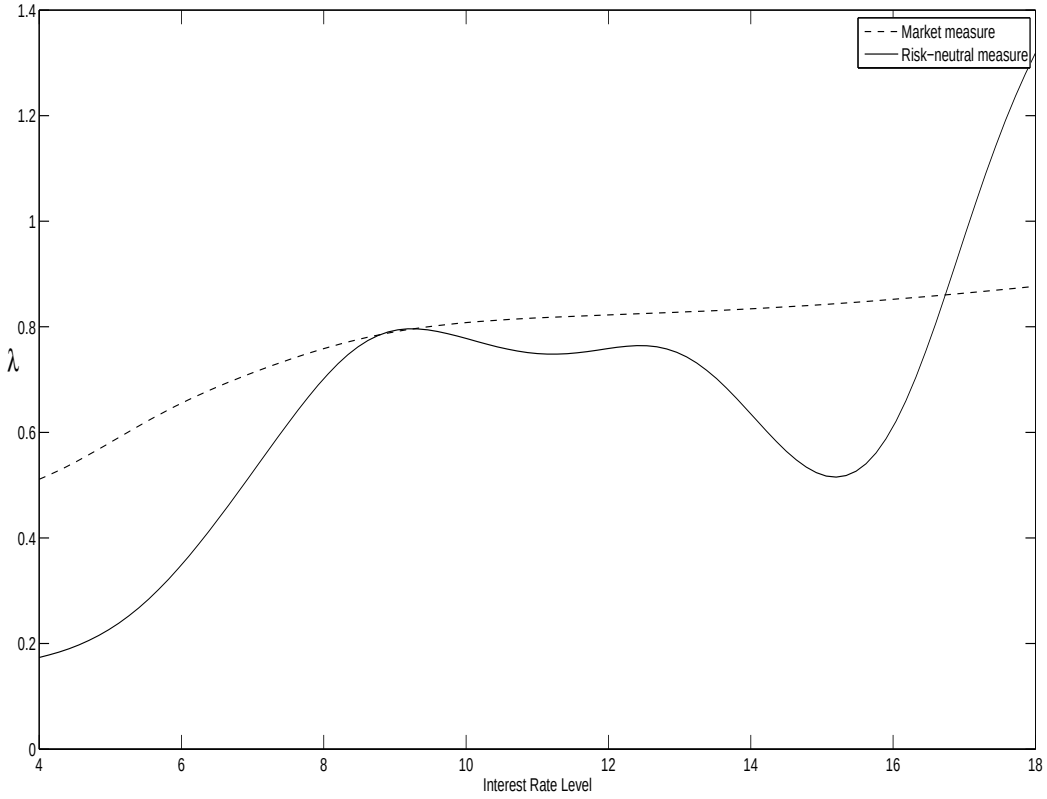

Figure 6.1: Intensity estimates for $\lambda(r_{t-})$ under the risk-neutral measure $\mathbb{P}^*$ as well us under market measure $\mathbb{P}$.

We will now test our simulations by analysing whether we can price a zero-coupon bond by Monte Carlo in both the diffusion model and in our pure jump model. To distinguish quantities pertinent to both models we add a superscript c (for continuous) to those referring to the diffusion model. Interest rate paths from the diffusion model, denoted by $r^c_{i,t}$, will also be simulated by the method described in Section 4.1.

The price of a zero coupon bond with price $p(S,S) = 1$ at maturity can be calculated by

$$p(t,S) = E^* \left[e^{-\int_t^S r_s \mathrm{d}s} | \mathcal{F}_t \right], \tag{6.8}$$

where the expectation is taken under a measure $\mathbb{P}^*$ and the time to maturity S is fixed.

Let $r^c_{i,t}$ and $r_{i,t}$ be the simulated interest rates for $i = 1, 2, ..., M$ and $t = 0, 1, ..., T$, where $T = 360S + 90$ is the number of days until maturity S years plus 3 months. The first step to price our interest rate derivative is to estimate the values of zero coupon bonds $p(0, \frac{j}{360})$, where $j = 1, 2, ..., T$, with (6.8). We first calculate the prices

in terms of Riemann-sum approximations of the integral in (6.8) by

$$\tilde{p}_i^c(0, \frac{j}{360}) = e^{-\frac{1}{360}\sum_{t=1}^{j} r_{i,t}^c}$$

and

$$\tilde{p}_i(0, \frac{j}{360}) = e^{-\frac{1}{360}\sum_{t=1}^{j} r_{i,t}}.$$

The price of the zero coupon bonds $p(0, \frac{j}{360})$ is then estimated by

$$\hat{p}^c(0, \frac{j}{360}) = \frac{1}{M}\sum_{i=1}^{M} \tilde{p}_i^c(0, \frac{j}{360}) \quad \text{and} \quad \hat{p}(0, \frac{j}{360}) = \frac{1}{M}\sum_{i=1}^{M} \tilde{p}_i(0, \frac{j}{360})$$

To test our simulations we price the zero coupon bond $p(0, S)$, where $S = \frac{3}{12}$, $r_0 = 10\%$, $\triangle = 1$ and $M = 100,000$. The results are given in Table 6.1. The exact value can be calculated by

$$p(0, S) = e^{-r_0 S}.$$

Table 6.1: Zero coupon bond price.

Description	Price
$p(0, S)$	0.975310
$\hat{p}^c(0, S)$	0.975311
$\hat{p}(0, S)$	0.975308

These results show that both Monte Carlo methods simulate sets of interest rate paths which can be compared to each other in our further analysis. To analyse the results further we will analyse the average yield at time S on a bond with 3 months to maturity, which is given by

$$Y_S = -4 \log p(S, S + \frac{1}{4}) = -4 \log \frac{p(0, S + \frac{1}{4})}{p(0, S)}. \tag{6.9}$$

The rate Y_S is equivalent to the 3-month forward JIBAR rate, which will be the underlying rate of our options priced in the next section.

From our interest rate paths, we can simulate M random variables given by $\hat{Y}^c(i, S)$ and $\hat{Y}(i, S)$. Now, basic statistics of these random variables are shown in Table 6.2.

Table 6.2: Yield statistics

statistic	Y_T^c	Y_T
mean	9.997155%	9.998003%
standard deviation	0.596099	0.601715
kurtosis	-0.052246	0.291956
minimum	7.661736%	7.38523%
maximum	12.41939%	12.846104%

The shape of the tails of the distributions of $\hat{Y}^c(i,S)$ and $\hat{Y}(i,S)$ should differ, as the sample excess kurtosis of $\hat{Y}^c(i,S)$ and $\hat{Y}(i,S)$ differ. This is confirmed in Figure 6.2.

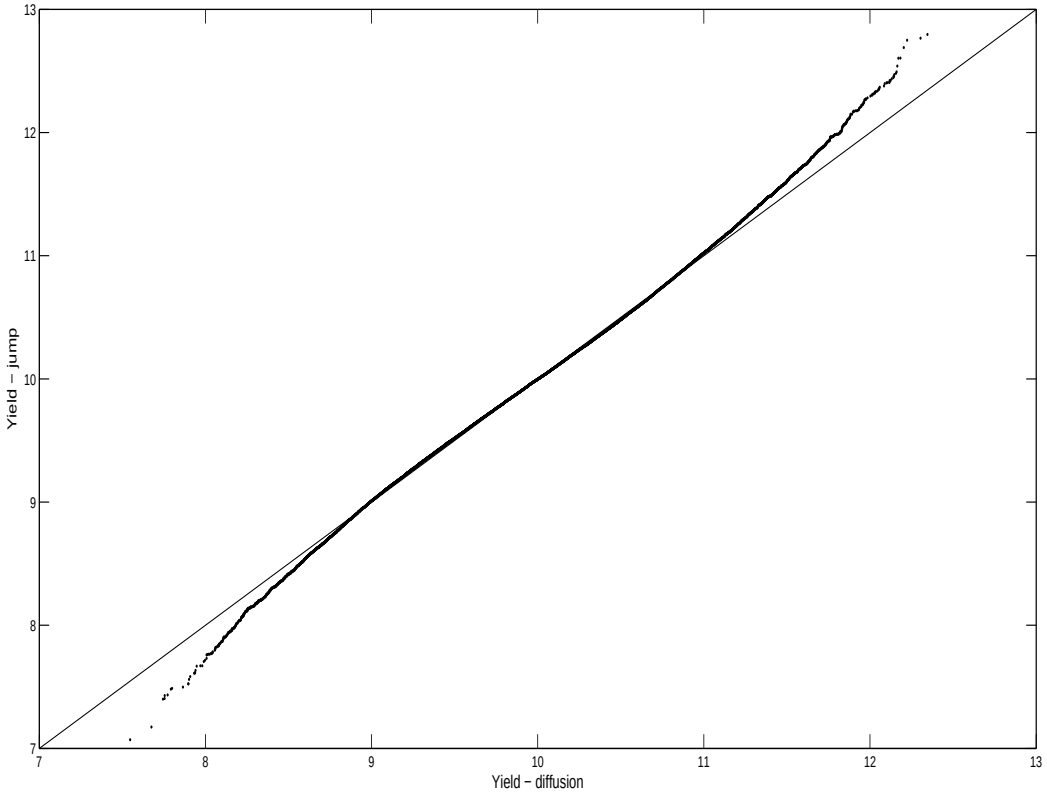

Figure 6.2: This graph is a plot of the estimated quantiles of the average yield obtained from the diffusion process against the average yield obtained from the pure jump process. The quantiles were estimated with a Monte Carlo method.

We will analyse the prices of a European style option (Section 6.3) and a barrier option (Section 6.4) on the 3-month forward JIBAR. We calculate the 3-month JIBAR as the

average yield of a zero-coupon bond with 3 months to maturity. These options are not traded on exchanges, but indications are that similar options are frequently traded in the South African OTC interest rate derivative market (Chapter 1). The results from our analyses will provide us with insight into the possible effect of pricing in a pure jump model.

6.3 Pricing of a European call option on the 3-month forward JIBAR

As in the previous section let Y_t be the average yield of a zero-coupon bond valued at time t with 3-months maturity.

In this section we will analyse the impact of a pure jump model on the price of a European call option on Y_t with strike yield K and maturity date S, where the payoff is given by

$$\Phi(S, K) = \max(Y_S - K, 0). \tag{6.10}$$

A call option on the yield differs with a call option on an asset (such as a stock or a bond) as no asset will be traded at maturity. These assets are cash-settled, which will be illustrated by an example:

Example 22. Let the average yield at time $t = 0$ and the strike yield be equal to 10%, therefore $Y_0 = K = 10\%$. Let us assume that the contract is stipulated in terms of basis points, where one basis point unit is equal to one South African Rand (ZAR). If the option is in-the-money (ITM) at maturity, for example $Y_S = 10.5\%$, then the payoff of the option is equal to 0.5% which equates to 50 basis points. The writer of the option will in this case pay the option buyer 50 ZAR at maturity. Let us assume the price of the option is equal to 20 basis points, then the profit for the buyer equals 30 ZAR. If the option is out-the-money (OTM) at maturity, for example $Y_S = 9.5\%$, then the option will not be exercised. The option buyer then makes a loss equal to 20 ZAR, which is the amount exchanged at $t = 0$.

The question is how would we apply a martingale approach to compute a fair price for such an option? We have for $0 \leq t \leq S$ the discounted value of a bond given by

$$X_t = e^{-\int_0^t r_s \mathrm{d}s} p(t, S),$$

where $p(t, S)$ is the price of the bond at time t. Therefore, with expectation taken under the risk-neutral measure $\mathbb{P}^*$, we have

$$
\begin{aligned}
E^*\left[X_S|\mathcal{F}_t\right] &= E^*\left[e^{-\int_0^S r_s \mathrm{d}s} p(S,S)|\mathcal{F}_t\right] \\
&= E^*\left[e^{-\int_0^S r_s \mathrm{d}s}|\mathcal{F}_t\right] \\
&= E^*\left[e^{-\int_0^t r_s \mathrm{d}s} e^{-\int_t^S r_s \mathrm{d}s}|\mathcal{F}_t\right] \\
&= e^{-\int_0^t r_s \mathrm{d}s} E^*\left[e^{-\int_t^S r_s \mathrm{d}s}|\mathcal{F}_t\right] \\
&= e^{-\int_0^t r_s \mathrm{d}s} p(t,S) \\
&= X_t.
\end{aligned}
$$

Therefore, the discounted value of $p(t,S)$, is a $\mathbb{P}^*$ martingale (Privault, 2013, p. 314). Therefore, $\mathbb{P}^*$ is a risk-neutral martingale measure (Bjork, 2004, p.92). If the payoff is given by $\phi(p(S,R))$, where R is the term of the bond with price $p(S,S+R)$ at time S, then the price at time $t<S$ of a derivative can be calculated by

$$
\Phi(t,S,R,r_t) = E^*\left[e^{-\int_t^S r_s \mathrm{d}s}\phi(p(S,S+R))|\mathcal{F}_t\right]. \tag{6.11}
$$

In our case, the payoff is given by (6.10), where Y_S is a function of $p(S,S+\frac{1}{4})$ by formula (6.9) on page (90). The price at time $t=0$ of a European call option on the average yield is therefore given by

$$
\Phi(S,K,r_t) = E^*\left[e^{-\int_0^S r_s \mathrm{d}s}\max\left(Y_S-K,0\right)|\mathcal{F}_0\right], \tag{6.12}
$$

were r_t is the short rate for $0\leq t\leq S$.

The main objective of this chapter is to compare the prices of interest rate derivatives yielded by our pure jump model with those yielded by one-factor diffusion models. We apply the same assumptions as listed in the previous section.

Since the characteristics of the models differ at various levels of the interest rate, we will estimate the price of a European call option on the average yield for two values of the interest rate Y_0, namely, for $Y_0=10\%$ and $Y_0=6\%$. We expect the impact to be greater at lower levels of the interest rate as no diffusion component is evident from historical data at these low levels.

For $Y_0=10\%$, we take various strike yields ($K=8.5\%,9\%,...,11.5\%$) and we estimate the prices by Monte Carlo simulation. Interest rate paths are simulated to time $S+\frac{1}{4}$ to estimate a value for Y_S for each path. The payoff in (6.10) on page 92 is then discounted using each interest rate path. We then estimate the price in (6.12) by calculating the mean of each simulated discounted payoff.

In our application we take $S=\frac{3}{12}$ (3 months) and the number of simulations as $100,000$ ensuring the estimates to be fairly accurate. The results are shown in Table 6.3, where

we observe relatively larger price differences for larger values of the strike yield. However, when the option is ATM at $t = 0$, the difference is relatively small.

Table 6.3: Call Option prices ($Y_0 = 10\%$)

K	$\Phi(\frac{3}{12}, K, r_t)^c$ ('ZAR)	$\Phi(\frac{3}{12}, K, r_t)$ ('ZAR)	relative change
8.5%	146.10	146.26	0.1109%
9%	98.39	98.66	0.2802%
9.5%	55.14	55.3	0.2894%
10%	23.08	22.96	-0.5137%
10.5%	6.39	6.52	2.0716%
11%	1.04	1.26	21.2805%
11.5%	0.09	0.17	94.757%

Φ^c is the price obtained from one-factor diffusion model.

If we apply the same method for $Y_0 = 6\%$ and various strike yields ($K = 5\%, 5.5\%, ..., 7\%$) then large relative differences are observed for ATM and OTM options ($K \geq 6\%$).

Table 6.4: Call Option prices ($Y_0 = 6\%$)

K	$\Phi(\frac{3}{12}, K, r_t)^c$ ('ZAR)	$\Phi(\frac{3}{12}, K, r_t)$ ('ZAR)	relative change
5%	98.58	98.37	-0.2132%
5.5%	50.05	50.11	0.1233%
6%	13.05	12.15	-6.9326%
6.5%	1.21	1.40	15.3222%
7%	0.05	0.13	187.4501%

Our conclusion therefore is that European call options on the 3-month forward JIBAR are priced inaccurately by a one-factor diffusion model, especially in times where interest rate levels are low.

We now turn to the question of the validity of our assumptions. We assume the parameters of the family of the truncated stable distribution are fixed at certain values. We know from Girsanov's theorem (Example 19 on page 26) that the distribution of jump sizes changes with a change of measure. We will now examine whether our conclusions made in this section will change if the index parameter (α) changes. We therefore take a range of α parameter values and price our European call option on the average yield applying the same methodology used thus far. We have used the following parameter values in our analysis ($Y_0 = 6\%$, $K = 5\%, 6\%, 7\%$, $m = 100,000$, $\alpha = 1.05, 1.1, ..., 1.95, 1.96, ..., 1.99$, $S = \frac{3}{12}$).

The results are shown in Figure 6.3, where the relative price changes for the ITM option does not depend on the value of α, while the ATM and OTM relative changes do. A conclusion can be made that as $\alpha \to 2$, the relative change is minimal, and as $\alpha \to 1$ we could possibly have large relative changes.

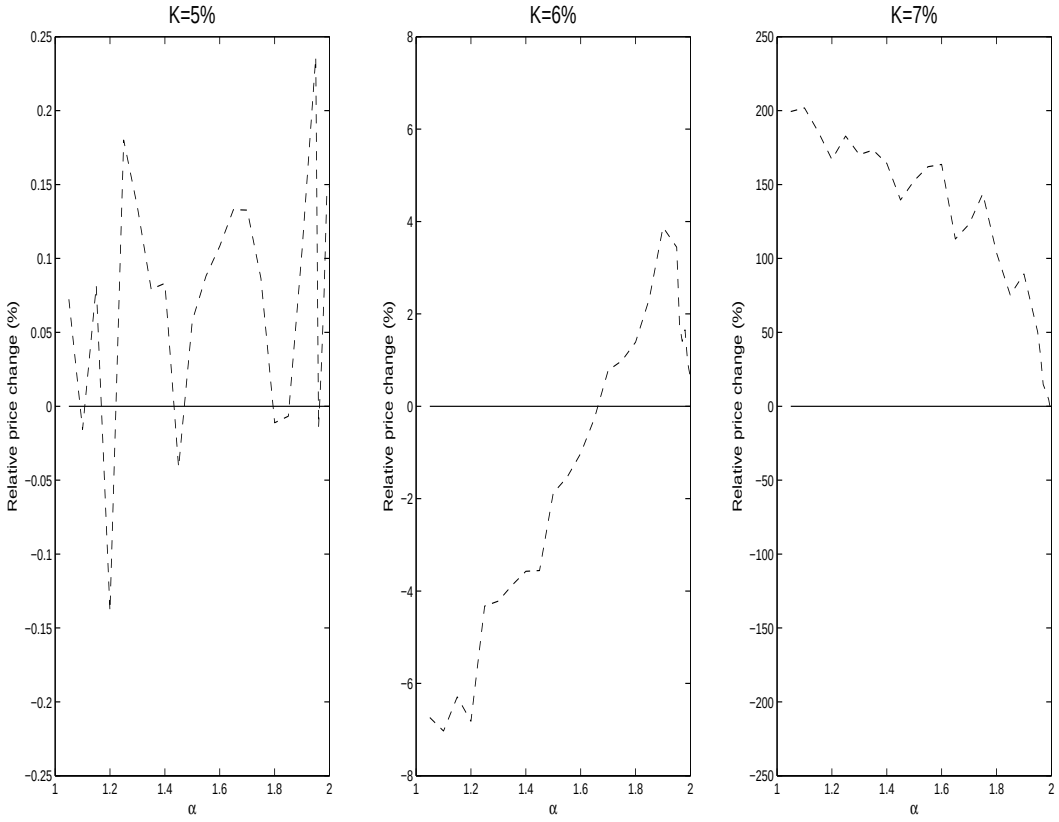

Figure 6.3: Relative price change of the prices obtained from pure jump processes compared with diffusion processes for various values of α. Analyses were done for an ITM call option ($K = 5\% < Y_0 = 6\%$), an ATM call option $K = 6\% = Y_0 = 6\%$ and an OTM call option $K = 7\% > Y_0 = 6\%$. Relatively large changes are observed for the ATM and the OTM option, except if α is close to 2.

Similarly to testing whether the relative price change depends on the value of α, we test whether a choice of σ changes our conclusions. In this case we use values of

$$\sigma = 0.01 \times 2^{-3}, 0.01 \times 2^{-2}, ..., 0.01 \times 2^{9} \quad \text{and} \quad \alpha = 1.2,$$

with all other parameter values kept the same as in the previous test. The results shown in Figure 6.4 clearly depicts a relationship between the relative price change with σ. As $\sigma \to 0$ the price change is minimal and as σ becomes larger the price change is larger.

This result can be explained by the fact that a smaller σ will lead to a larger intensity λ, which means more frequent jumps and vice versa. Therefore, as the pure jump process converges to a diffusion process, no relative price changes are observed, whereas, if the pure jump process has infrequent jumps, larger price changes are observed.

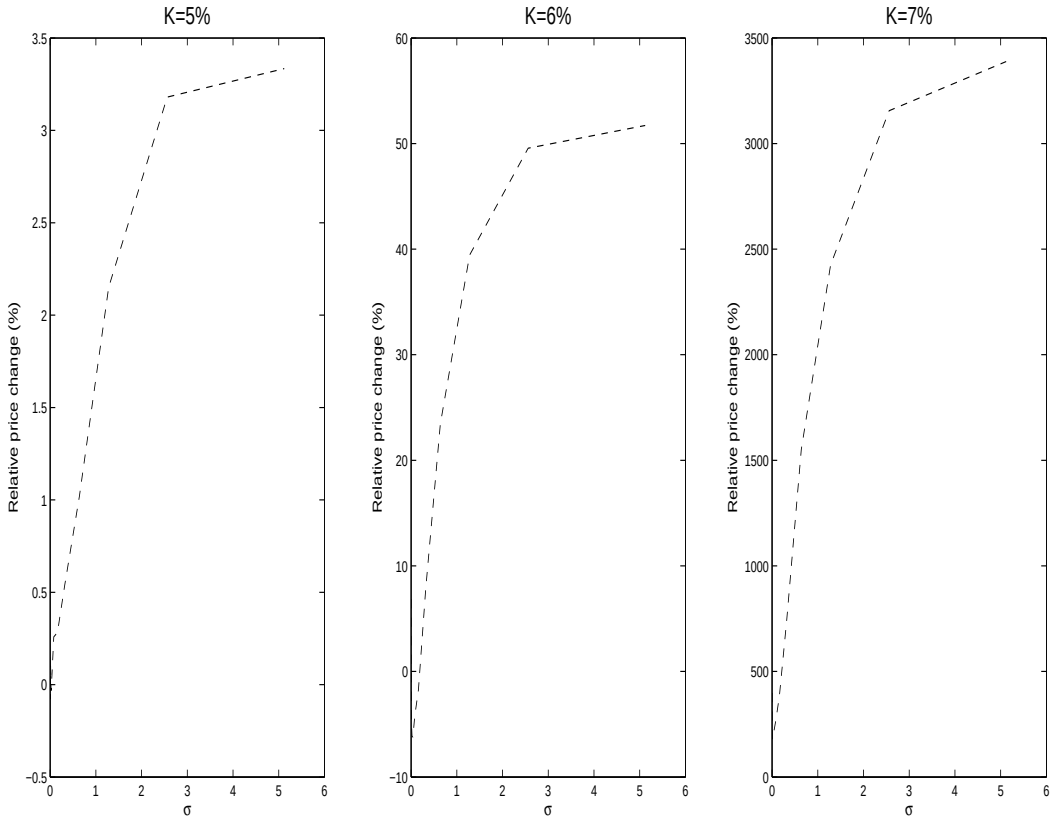

Figure 6.4: Relative price change of the prices obtained from pure jump processes compared with diffusion processes for various values of σ. Analyses were done for an ITM call option ($K = 5\% < Y_0 = 6\%$), an ATM call option $K = 6\% = Y_0 = 6\%$ and an OTM call option ($K = 7\% > Y_0 = 6\%$). Larger values of σ lead to larger relative changes.

The conclusion from this section is that the prices of interest rate derivatives approximated with a pure jump model differ from those approximated using a diffusion model. This result is not only due to the shape of the jump distribution, but also due to the intensity of the jumps.

In the next section we will price a barrier option on the 3-month forward JIBAR.

6.4 Pricing of a barrier option on the 3-month forward JIBAR

Writers of vanilla options, such as European call options take on more risk than buyers of the option. The maximum loss for a buyer is the price paid for the option, which is known at time $t = 0$. The possible loss for the writer is unbounded and not known at time $t = 0$. One way to hedge the writer from this risk is to write an up-and-out call (UOC) option on the average yield rather than a vanilla call option.

Let $Y_B > Y_0$ and $Y_B > K$, then the payoff of an UOC option with maturity date S and strike yield K is zero if $Y_t > Y_B$ for any $t \in (0, S]$ and equal to the payoff of a European call otherwise. The payoff can therefore be written as

$$\Phi(S, K) = \max(Y_S - K, 0)\, I_{\{\tau > S\}}, \tag{6.13}$$

where

$$\tau = \inf\{t \mid Y_t > Y_B\}.$$

An UOC option is an example of a barrier option. A barrier option is path dependent, since the payoff depends not only on the value of the underlying asset at maturity date S, but also on dates between 0 and S.. In practice, models are calibrated to the prices of vanilla options and the results are then used to price exotic options, such as barrier options.

In this section we will take an example of an option where the difference between the price obtained from the diffusion model and the price obtained from our pure jump model is relatively small. We will then analyse whether differences in prices exist for various UOC options. In the previous section the ATM European call option on the average yield, where $Y_0 = K = 10\%$, had a relative price change of -0.51%, which is relatively small. We will now take a range of barrier yields $Y_B = 10.5\%, 10.55\%, ..., 13\%$, and estimate the price of the UOC option for each barrier yield. The price of the option can be calculated by (6.11) on page 93, where the payoff is given by (6.13). We estimate the prices by applying Monte Carlo simulations as in the previous section.

In Figure 6.5 the estimated prices of the UOC option give by applying the pure jump model and one-factor diffusion model are shown in relation to the various levels of the barrier yields Y_B in the top graph. In the bottom graph we compare the two sets of prices by calculating the relative change in prices from the two models in relation to Y_B. We observe that the absolute relative change can be large for lower values of the barrier yield. The absolute relative change can be as large as 4% when Y_B is in the region of 11% to 11.5%.

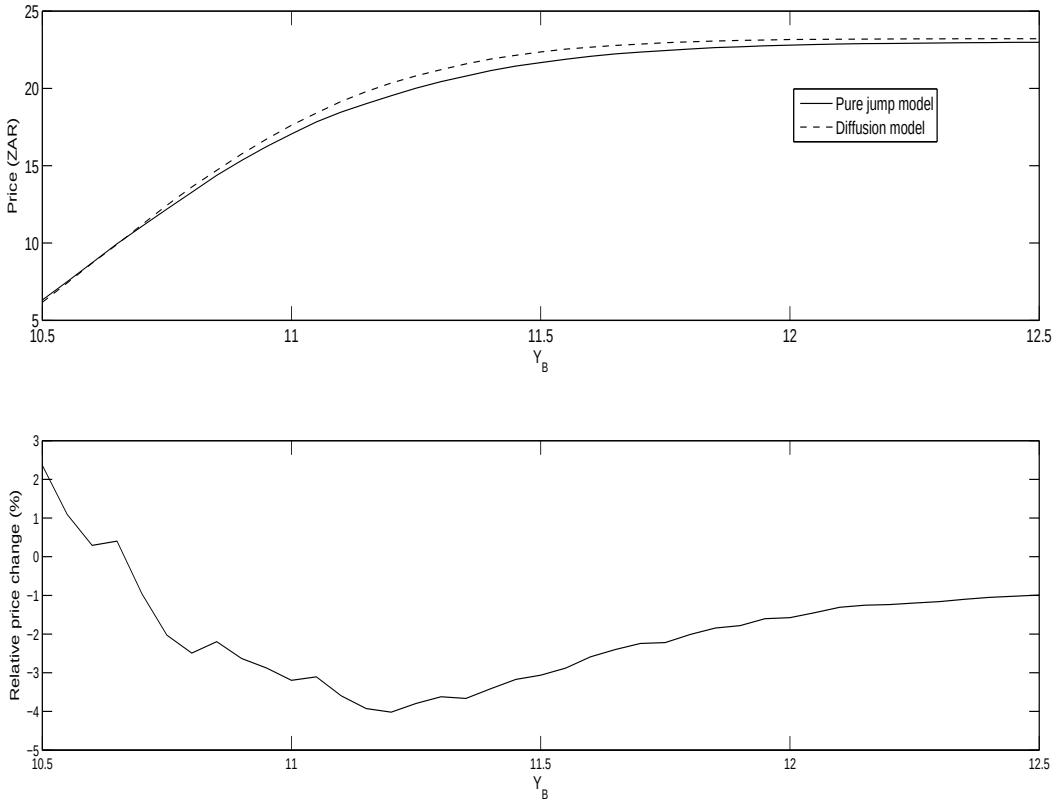

Figure 6.5: In this figure the prices of an up-and-out call (UOC) option on the 3-month forward JIBAR from a pure jump model are compared to the option prices from a one-factor diffusion model. The top graph shows the prices in relation to various levels of the barrier yield Y_B, while the bottom graph shows the change in price obtained from the pure jump model relative to the prices obtained from the diffusion model.

The reliability of the prices shown in Figure 6.5 may be influenced by the number of paths breaching the barrier yield at low levels of Y_B. In the bottom graph of Figure 6.6 we show the percentage of paths where the barrier yield Y_B have been breached for our UOC option with 3 months to maturity. We replicated the analysis for both 12 and 24 months time to maturity. The results are also shown in the top graph of Figure 6.6, where the change in the prices from the pure jump model is calculated relative to the prices from the one-factor diffusion model. We observe potentially high relative price changes at low levels of Y_B with the 24-month maturity option, even though the relative change for the European call option is minimal. From the bottom graph we notice that close to 60% of the sample paths breach the barrier yield of $Y_B = 10.5\%$. However, with $100,000$ sample paths being simulated, more than $40,000$ sample paths do not breach the barrier yield. This should equate to a relatively high accuracy of the option prices estimated

using Monte Carlo simulations.

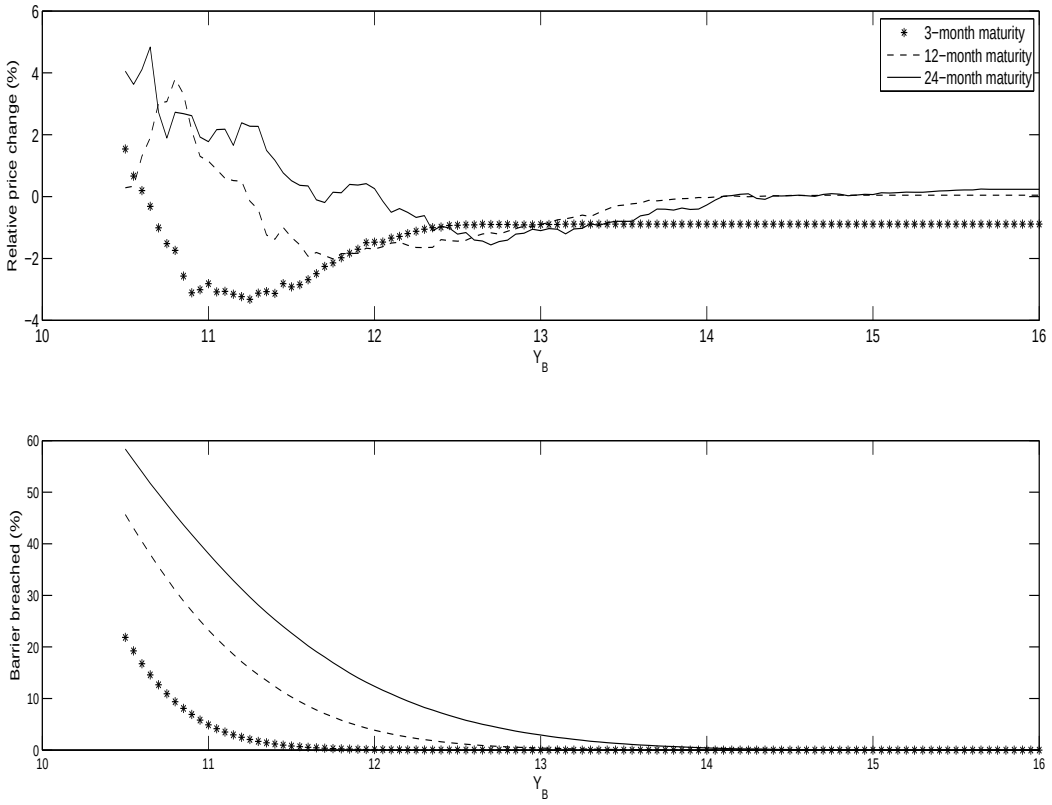

Figure 6.6: In this figure the results of an up-and-out call (UOC) are shown for various option maturity dates. The top graph shows the change in option prices in relation to the barrier yield. The bottom graph shows the percentage of paths where the barrier yield is breached for various levels of the barrier yield.

Our conclusion from the analysis in this section is that a pure jump model has a marked influence on the price of a UOC barrier option. We only looked at cases where the corresponding vanilla option is not influenced by whether the model is a pure jump model or a one-factor diffusion model. We can assume that the potential impact will be even greater in those cases where the corresponding vanilla option price do differ.

We have provided evidence in this chapter to suggest that a pure jump model should be considered when pricing interest rate derivatives on the 3-month forward JIBAR as underlying rate. Not only does the choice of model have an impact on European options on the forward JIBAR, but also on more exotic options such as barrier options.

There are some considerations to take into account when using a model in practice.

We did not calibrate our model to option prices or to a volatility surface. There are two reasons for this. The one is that with our goal to find a model which fits historical rates as perfectly as possible, analytical tractability was not considered. Models used in industry are calibrated daily to yield curves as well as to volatility surfaces. This is a requirement due to the incompleteness of short rate models in general (Bjork, 2004, p.327). Therefore, a model which produces analytical solutions for bond prices as well as vanilla options is preferred in industry. The second reason is that option price data in the South African market is sparse. We mentioned in Chapter 1 that only vanilla options on bonds are traded on the JSE, while other interest rate derivatives are traded OTC. However, the market for OTC interest rate options, especially on the 3-month JIBAR, is relatively large and therefore important to investigate.

Chapter 7

Part I: Conclusion

In this chapter we will summarise our main findings. We will also list possible further research projects that originate from our current research.

The two short term interest rates in South Africa are the 3-month JIBAR and 91-day T-bill rate. We empirically found from historical data that these two rates have similar characteristics. Both rates have infrequent large interest rate movements, with long periods where no interest rate changes occur. These properties also vary at different interest rate levels. A high frequency of jumps occurs at higher interest rates levels, while the period between movements are longer at low interest rate levels. These interest rates can be used as proxies for the South African short rate, which originally would have been modelled by a one-factor diffusion process. Existing literature on interest rate models in the South African context is scant. However, from the literature available it is clear that the volatility in the South African short rate varies for different levels of the short rate. This corresponds with the results of our initial analysis.

As a result of our initial investigation we decided to test for jumps more formally, using statistical methods. We found that jumps occur frequently by applying two tests. Firstly, by applying a Monte Carlo hypothesis test for jumps we found that jumps should be included in the model. Secondly, a hypothesis test to identify isolated jumps from historical data found jumps to occur frequently. The tests for jumps demand high frequency data. We therefore used daily historical data from the 3-month JIBAR, rather than weekly 91-day T-bill data. An important secondary result from the tests for jumps emerged. We found evidence that jumps contribute more to the conditional volatility of the short rate than the diffusion component in a jump diffusion model. This result is valid for all interest rate levels. A compelling result is that no diffusion component is evident at lower interest rate levels. We therefore decided to fit a pure jump short rate model to historical 3-month JIBAR data and to investigate whether the fitted model can be used to price interest rate derivatives.

We followed a classic statistical methodology to fit a pure jump model to historical data. We fitted a distribution to the time between interest rate movements. An exponential distribution with rate dependent intensity was found to be a good fit. We also fitted a distribution to the interest rate movements, where a distribution from the family of stable distributions was found to fit the jumps. The resulting South African short rate model we initially considered was a nonstationary compound Poisson process. The fitted short rate model is one of our main results.

In the field of Quantitative Risk Management, interest rates are standard risk drivers for certain models. Therefore, in many divisions of the banking sector predicting interest rates accurately is an important goal. If calibrated accurately, the mean predicted rate from our model will compare well to that of a fitted diffusion model. However, in certain applications, such as stress testing, the tails of the interest rate distribution are of more importance. Our pure jump model leads to realistic heavy tailed interest rate distributions and would therefore forecast future interest rates better in a stressed environment than a diffusion model would.

Our main objective in our study was to build a model to price interest rate derivatives. We had to adjust our fitted pure jump model in order to achieve this goal. The short rate model we defined to price interest rate derivatives is a pure jump model with symmetric truncated stably distributed jumps around a zero mean, a rate dependent intensity and a rate dependent drift coefficient. To evaluate our model we will measure the adequacy of our model against the criteria listed in Brigo and Mercurio (2006, p.54):

- *Does the dynamics imply positive rates?* No, the model may result in negative interest rates. However, the likelihood of this occurring is low. Interest rates in South Africa are relatively high compared to interest rates in some first world countries and jumps are truncated at realistic values. This criterion is possibly old fashioned as negative rates were historically believed to be an impossible event, but have occurred in several countries recently.

- *What distribution does the dynamics imply?* The implied distribution from our pure jump model is a fat tailed distribution with a higher kurtosis than the implied distribution from one-factor diffusion models. Our aim was to build a model which fits short dated interest rate returns close to perfect. Our altered pricing model will slightly impact the goodness of fit. However, the implied distribution of our model remains its most powerful aspect. This is due to the results found in pricing several options on the 3-month forward JIBAR. We found some considerable changes in option prices in comparison with one-factor diffusion models. This is mainly due to the implied distribution of the 3-month forward JIBAR rate being accurate when applying our pure jump model. We obtained these results by assuming a nonzero market price of risk. However, it is safe to assume that in the presence of a nonzero

market price of risk the implied distribution will change more. This will probably result in larger differences in our pricing results. However, the estimation of the market price of risk and its impact on derivative prices for South African short rate models are topics for future research. We did, however, lay the foundation for this research in terms of pure jump models.

- *Are bond and interest rate option prices explicitly computable?* Our model neither leads to analytical solutions of zero-coupon bond prices nor to interest rate option prices, such as caps or floors. This is the weakest property of our model as it will be difficult to calibrate the model accurately to either market yield curves or implied volatility surfaces. Incomplete market models, such as short rate models, need to be calibrated to market data to obtain unique derivative prices.

- *Is the model mean reverting?* With the general form of our model, mean reversion is possible through the correct choice of model parameters. The short rate model

$$\mathrm{d}r_t = a\,(b - r_{t-})\,\mathrm{d}t + Z_{N_t} N(\mathrm{d}t), \tag{7.1}$$

  will have the mean reversion effect as

$$E\left\{r_t \mid F_0\right\} \to b$$

  The only criterion to enable mean reversion is for the jumps to have a zero mean, which we do assume in our model.

- *What do the volatility structures implied by the model look like?* An important property of our model is a rate dependent volatility due to the intensity of the jumps being rate dependent. However, further research needs to be done to investigate for example the term structure of caplet volatilities and compare it to the resulting structures from other short rate models.

- *Does the model allow for explicit short-rate dynamics under the forward measures?* The calculus in applying Girsanov's theorem for nonstationary compound Poisson processes has been outlined in our study. However, we have not researched how the dynamics of a compound Poisson process will change from a risk-neutral measure to a forward measure. This may be done as part of a future research project.

- *How suited is the model for the numerical pricing of derivatives?* The pricing of derivatives with Monte Carlo simulation methods was fairly straightforward. We did not employ variance reduction techniques. Employing these methods becomes a necessity if Monte Carlo simulations need to be used where computational time is of importance. Therefore, the investigation of variance reduction techniques as well as other pricing methods such as lattice methods or solving partial differential equations numerically can form part of a future research project.

- *Does the chosen dynamics allow for historical estimation techniques to be used for parameter estimation purposes?* We used historical data to fit our original short rate model. Methods to estimate the parameters when the model has a specific form, such as in (7.1) need to be investigated in future.

The advantages and disadvantages of our pure jump short rate model in a pricing context has been summarised above. We also gave an indication of possible future research projects which can be done due to the research we did. The major drawback of our model is the inability to find analytical solutions of bond and interest rate option prices. Our main aim for future research will therefore be to find a model which has similar characteristics as the model built, while being analytically tractable.

An affine term structure of bond prices generally makes option pricing easier. Such a term structure of bond prices does exist if we restrict the drift and intensity in our model to be a first order linear function of the short rate under certain conditions (Bjork et al., 1997, p.237). In addition we possibly need to choose a heavy tailed jump distribution which is analytically uncomplicated to work with.

We feel that the South African interest rate market is unique, as the data is sparse and derived from illiquid instruments. However, the interest rate option market is unstructured, but relatively large. Research in the South African interest rate derivative market is therefore necessary. The findings from our project showcases the possibility of finding unique research results by exploring this often uncharted research field.

Part II

Stable distributions

Part II: Introduction

The empirical evidence from Chapter 3 justifies the research into the properties of thick tailed distributions. We fitted a distribution to the jump sizes in Section 5.2 in order to fit a nonstationary compound Poisson process to interest rate movements. Typically, a normal distribution for the jumps will be assumed, but in Section 5.2 it is shown that a normal distribution does not fit the 3-month JIBAR movements well. A histogram of the data resembles a normal distribution fairly well (in terms of general shape and symmetry around the mean). However, the extreme jumps that occur justify the use of a distribution with thicker tails than those of the normal distribution. One example of such a distribution is the Cauchy distribution. The tails of the Cauchy distribution, however, are too thick to fit the data (Section 5.2). These two observations regarding the normal and Cauchy distributions lead us to consider the family of stable distributions. The family of stable distributions has four parameters: an index parameter α, a skewness parameter β, a scale parameter σ and a location parameter μ. Two examples of distributions from the family of stable distributions are the normal distribution ($\alpha = 2$, $\beta = 0$) and the Cauchy distribution ($\alpha = 1$, $\beta = 0$).

A stably distributed random variable X is defined in terms of its characteristic function. Let X be a random variable with density function f. The characteristic function of X is given by

$$\phi(t) = E\left(e^{itX}\right) = \int_{\mathbb{R}} e^{itx} f(x)\mathrm{d}x, \tag{7.2}$$

where $i = \sqrt{-1}$.

If ϕ is integrable, $\int_{\mathbb{R}} |\phi(t)|\mathrm{d}t < \infty$, then f can be expressed in terms of ϕ by the inversion formula

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi(t)\mathrm{d}t. \tag{7.3}$$

Liu et al. (2014) provides us with various characterisations of the characteristic function, of which the Lévy-Khintchine representation is one.

Definition 2. (Lévy-Khintchine representation) The characteristic function of a stable random variable X has the following form

$$\phi(t; \boldsymbol{\theta}) = \exp\left(i\mu t - |\sigma t|^{\alpha}\left[1 - i\beta\,\mathrm{sgn}(t) w(t, \alpha)\right]\right), -\infty < t < \infty, \tag{7.4}$$

where

$$w(t,\alpha) = \begin{cases} \tan\left(\frac{\pi\alpha}{2}\right) & \text{if} \quad \alpha \neq 1 \\ -\frac{2}{\pi}\log|t| & \text{if} \quad \alpha = 1, \end{cases}$$

and the parameters satisfy the constraints $\alpha \in (0, 2]$, $\beta \in [-1, 1]$, $\sigma \in (0, \infty)$ and $\mu \in \mathbb{R}$. The vector $\boldsymbol{\theta}$ is defined as

$$\boldsymbol{\theta} = [\alpha \ \beta \ \sigma \ \mu]^T.$$

Density function values for the family of stable distributions can then in principle be calculated by applying (7.3). Typically, however, f does not have a closed form.

In Part II various parameter estimation techniques will be explored. The strength of the methods will be measured against the asymptotic variance of consistent parameter estimates as well as against their algorithmic complexity. As a measure for the algorithmic complexity, we take the running time to perform the algorithm. Some of the existing parameter estimation methods are discussed by Fama and Roll (1968), McCulloch (1986), Koutrouvelis (1980) and Paulson et al. (1975).

In Chapter 8 we provide the reader with background about the family of stable distributions. We illustrate in Section 8.2 the difficulties in calculating density function values. Chapter 8 will provide us with the tools needed to implement numerical methods in Chapters 10 and 11.

A method to approximate density function values as well as its partial derivatives (to the parameters of the stable distribution) is derived in Sections 8.2 and 8.3. The method to simulate stably distributed random variables is described in Section 8.4 (Weron, 1996).

In Chapter 9 two parameter estimation techniques will be investigated. Generally, maximum likelihood estimation is the most popular method to estimate the parameters of a number of distributions. Asymptotically maximum likelihood estimates (MLEs) are the most efficient estimates that exist. In other words, if a parameter estimate is computed from a sample, then the estimate will lie within the smallest confidence interval around the true population parameter value. Dumouchel (1975) calculated the MLEs for the stable distribution parameters, which we will use to compare our results.

Although the maximum likelihood estimates are the most efficient estimation techniques, the technique to calculate them is based on the involved calculation of a density function and this can be time consuming. We therefore want to investigate whether other estimation methods exist that use less calculation time while being almost as efficient as the MLEs. This will be done using the parametrisation defined in formula (7.4) on page 106.

A parameter estimation method which we will compare to the MLE method was developed by Paulson et al. (1975) and is called the integrated least squared estimation (ILSE) method. This method is based on the characteristic function of the stable distribution

and will lead to a less time consuming estimation method than the MLE method.

The integrated least squared estimates (ILSEs), can be obtained by finding the values of the parameters which minimises the following function

$$I(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} |\hat{\phi}(t) - \phi(t, \boldsymbol{\theta})|^2 e^{-t^2} \mathrm{d}t, \tag{7.5}$$

where $\hat{\phi}(t)$ is the empirical characteristic function for a sample $X_1, ..., X_n$ and is given by

$$\hat{\phi}(t) = \frac{1}{n} \sum_{i=1}^{n} \exp(itX_i).$$

The characteristic function $\phi(t, \boldsymbol{\theta})$ is defined in formula (7.4) on page 106.

Other parametrisations of the stable distribution exist. Nolan (1999) uses Zolotarev's (M) parametrisation (details in Section 8.2.8). With this parametrisation the characteristic function of a stable random variable M, is given by

$$\phi(t; \alpha, \beta, \lambda, \gamma) = \exp\left\{\lambda\left(it\gamma - |t|^{\alpha} + itw_M(t, \alpha, \beta)\right)\right\}, \tag{7.6}$$

where

$$w_M(t, \alpha, \beta) = \begin{cases} \left(|t|^{\alpha-1} - 1\right)\beta \tan\left(\frac{\pi\alpha}{2}\right), & \text{if} \quad \alpha \neq 1 \\ -\beta\frac{2}{\pi}\log|t|, & \text{if} \quad \alpha = 1, \end{cases}$$

with $\lambda > 0$ and $\gamma \in \mathbb{R}$ (Zolotarev, 1986, p.11). Nolan uses this parametrisation in his research because of the continuity of (7.6) for all values of α. This is in contrast to the characteristic function in (7.4) on page 106 that is discontinuous at $\alpha = 1$. We show in Section 8.2.8 that the approach to approximate density function values is based on the same methodology, whether we use Zolotarev's (M) parametrisation or the Lévy-Khintchine representation. The discontinuity of the characteristic function in the Lévy-Khintchine representation therefore does not have an effect on the accuracy of density function value calculations. The reason Nolan (1999) uses an alternative parametrisation has to do with the estimation of parameter values when α is close to one. However, we limit our research to values of α in the interval $(1, 2)$. The discontinuity of (7.4) at $\alpha = 1$ is therefore of no importance. Computer programs to find maximum likelihood estimates are available at Nolan (2017b) in the form of DOS executables. However, the underlying code is not available which implies that these programs cannot be integrated into a Matlab simulation program. One of our goals is to investigate the efficiency of the ILSE method, compared to the MLE method, which we do partially by Monte Carlo simulation. Therefore, we have develop our own methods to accomplish this. In particular our optimisation routines are written to preclude the occurrence of estimates $\hat{\alpha}$ are less than one.

The MLE method will be discussed in detail in Section 9.2. In Section 9.3 the ILSE method will be described after its brief introduction above. The results we get from

implementing these two methods will be analysed in Section 9.5, before we move on to a more theoretical approach in determining the efficiency of these two methods.

Within the framework of formal asymptotic statistics, the relative efficiency of the ILSE method will be computed. Section 10.1 will outline the basic theory used in deriving the consistency and asymptotic normality results for the ILSE method from Heathcote (1997).

Our first step in applying the theory of asymptotic normality to compute the relative efficiency of the ILSEs, starts by isolating the index parameter, α. We then calculate the asymptotic normality properties of its estimate, while the other parameters are held constant. The asymptotic variance of the MLEs first needed to be computed by calculating the Fisher information (Section 11.1.2). In Section 11.3 the relative efficiency of the estimates for this isolated case is computed.

The methods, applied to the single parameter case, are adjusted for the general case. The Fisher information is computed and the asymptotic covariance matrices for the MLEs and ILSEs are calculated (Sections 11.1.3 and 11.2.2). This enables us to compute the relative efficiency of the ILSEs compared to the MLEs (Section 11.3). The most significant results of this section are provided in the last section, where the relative efficiency of the ILSEs was improved by incorporating various weight functions (Section 11.3).

Chapter 8

Definitions and properties

A stably distributed random variable has the property that the sum of independent stably distributed random variables is again stably distributed (Dumouchel, 1983). We state the definitions in Nolan (2017a), who recently researched the field of the family of stable distributions and their applications extensively. Let X be a random variable, with X_1 and X_2 independent copies of X, then

$$aX_1 + bX_2 \stackrel{d}{=} cX + d. \tag{8.1}$$

(The notation $\stackrel{d}{=}$ means equality in distribution.)

Definition 3. A random variable X is stable (or stable in the broad sense) if for X_1 and X_2 independent copies of X and any positive constants a and b, formula (8.1) holds for some positive c and some $d \in \mathbb{R}$. The random variable is strictly stable (or stable in the narrow sense) if formula (8.1) holds with $d = 0$ for all choices of a and b.

The definition can be expanded to include the summation of independent copies X_1, X_2,..., X_n of the random variable X, where

$$\sum_{i=1}^{n} a_i X_i \stackrel{d}{=} cX + d.$$

It is clear from this definition that a normally distributed random variable is also a stably distributed random variable, as the mean of i.i.d. normally distributed random variables is also normally distributed.

8.1 Characteristic function

The characteristic function of the family of stable distributions was defined in Definition 2 on page 106 in terms of the Lévy-Khintchine representation. The stable distributions

have various parametrisations for their characteristic functions which lead to different interpretations of the stable parameters (Liu et al., 2014).

To get a number of our numerical results we occasionally suppress parameters μ, σ and β of the general characteristic function in (7.6) on page 108 and use the following form

$$\phi(t;\alpha) = \exp\left(-|t|^{\alpha}\right), -\infty < t < \infty. \tag{8.2}$$

When referring to the stable distributions with the above form of the characteristic function we will refer to the "single parameter case of the stable distributions."

In general it is helpful to analyse the characteristic function by rewriting the function in the form

$$\phi(t;\boldsymbol{\theta}) = u(t;\boldsymbol{\theta}) + iv(t;\boldsymbol{\theta}), -\infty < t < \infty, \tag{8.3}$$

with

$$u(t;\boldsymbol{\theta}) = e^{-|\sigma t|^{\alpha}} \cos\left[\mu t + |\sigma t|^{\alpha}\beta \operatorname{sgn}(t) w(t,\alpha)\right],$$

and

$$v(t;\boldsymbol{\theta}) = e^{-|\sigma t|^{\alpha}} \sin\left[\mu t + |\sigma t|^{\alpha}\beta \operatorname{sgn}(t) w(t,\alpha)\right].$$

The characteristic function can therefore be analysed graphically, by drawing graphs of the functions $u(t;\boldsymbol{\theta})$ and $v(t;\boldsymbol{\theta})$ on two separate axes. Three important aspects will be looked at in these graphs. The one is the speed at which the functions approach zero, as t tends to $\pm\infty$. The other factor is the shape of the functions near $t = 0$, while the third factor is the oscillating tendency of the functions.

In Figure 8.1 the parameter values of β, μ and σ are fixed and the graphs for three separate index parameter α values are shown. The one set of graphs that stands out is when $\alpha = 1$. For this specific case, the complex part of the characteristic function has a very steep angle around $t = 0$ and the real part is not differentiable at $t = 0$, contrary to the other two cases of the α parameter.

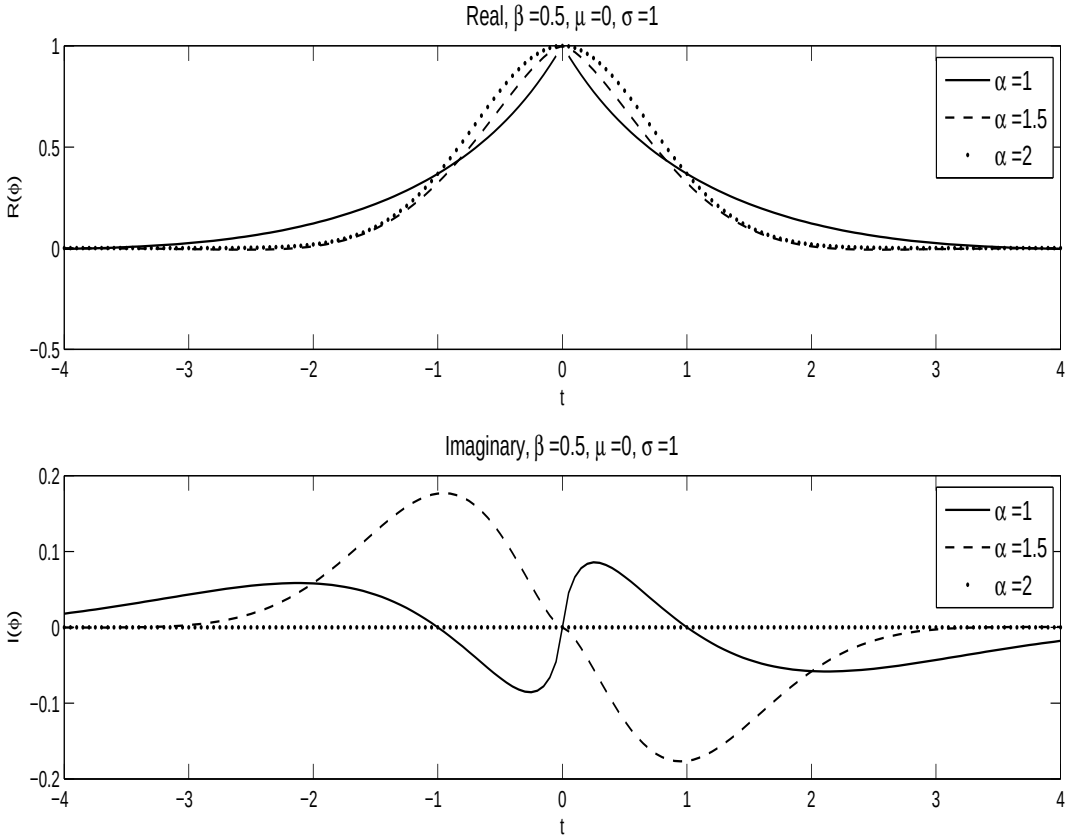

Figure 8.1: Graphs of the characteristic function for a range of α parameter values.

In Figure 8.2 the α, σ and μ parameters are all set at fixed values, while various values of the skewness parameter β are chosen. These graphs are all smooth. Note that the mirror image of the imaginary part of the characteristic function around the t-axis can be obtained by changing the sign of the β parameter.

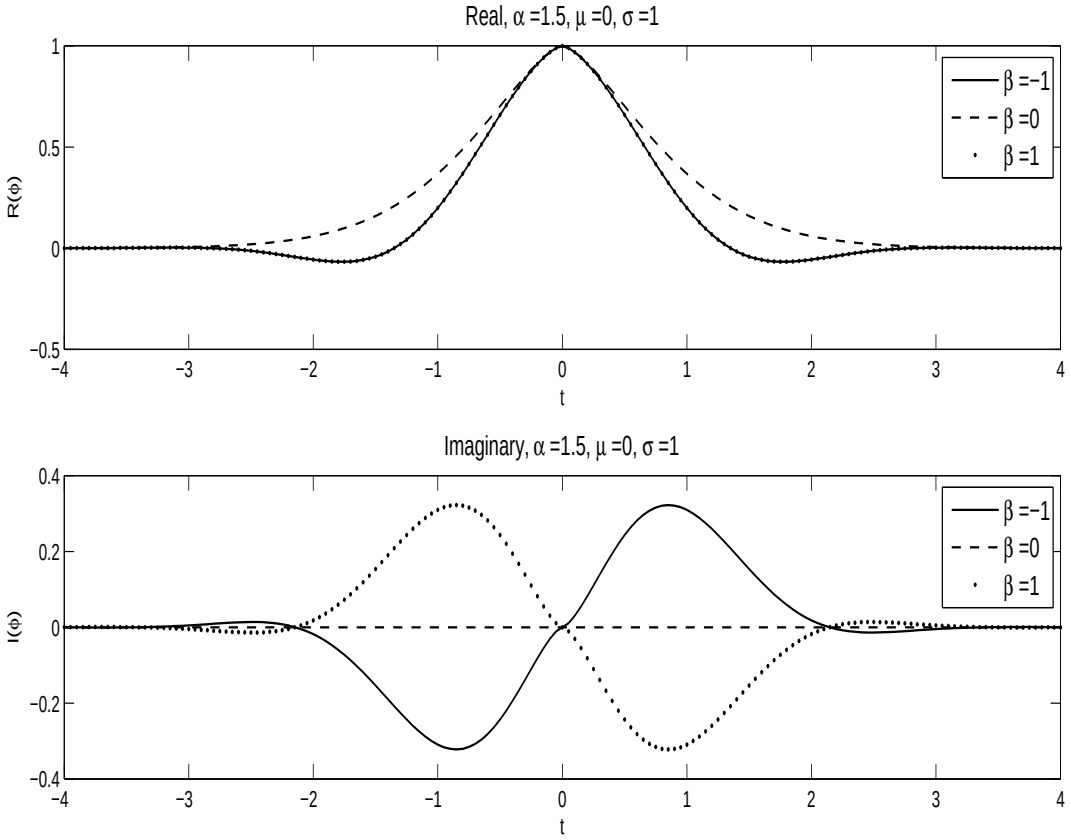

Figure 8.2: Graphs of the characteristic function for various values of the parameter β.

In Figure 8.3 the α, β and σ parameters are all set at fixed values, while various values of the location parameter μ are chosen. The functions shown in these graphs (Figures 8.1 and 8.2) are all oscillating functions. In Figure 8.3 it is shown that the parameter μ can have a significant influence on the oscillating tendencies of these functions. The influence of μ therefore needs to be kept in mind when using the characteristic function in numerical work.

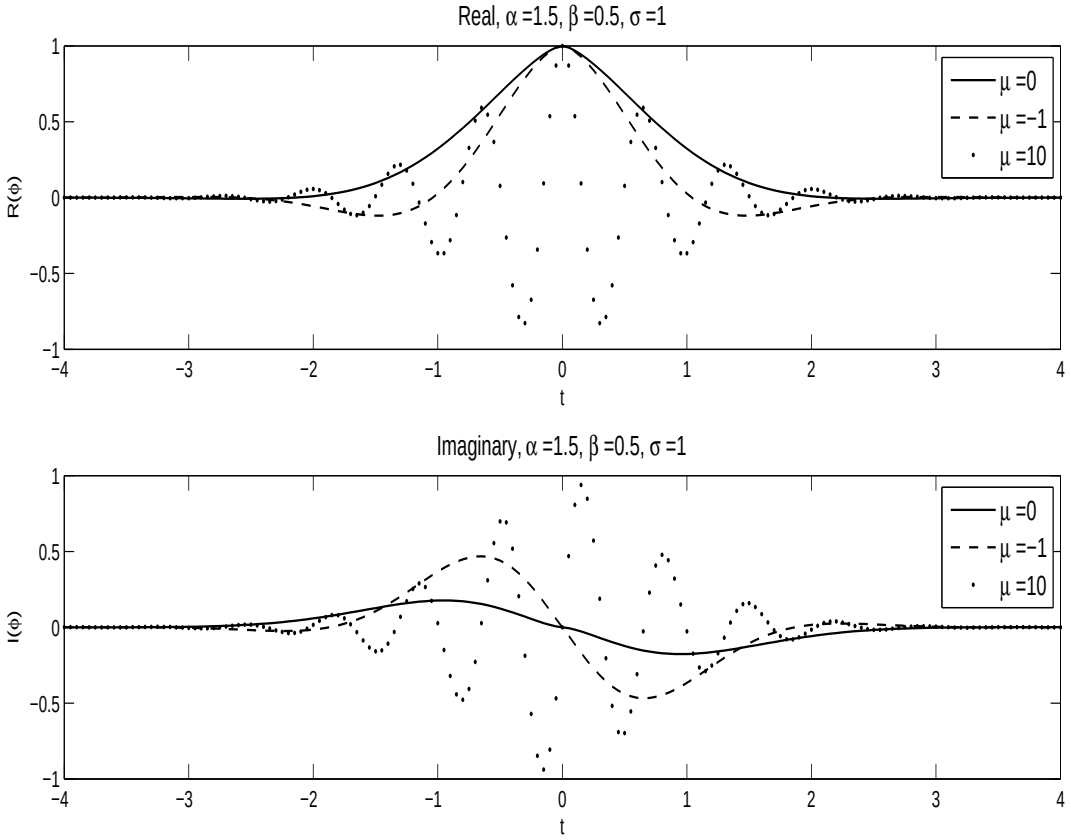

Figure 8.3: Graphs of the characteristic function for various values of the parameter μ.

In Figure 8.4 the α, β and μ parameters are all set at fixed values, while various values of the scale parameter σ are chosen. The parameter σ can have a significant effect on the decay of the real and imaginary graphs of the characteristic function. This can be seen from Figure 8.4.

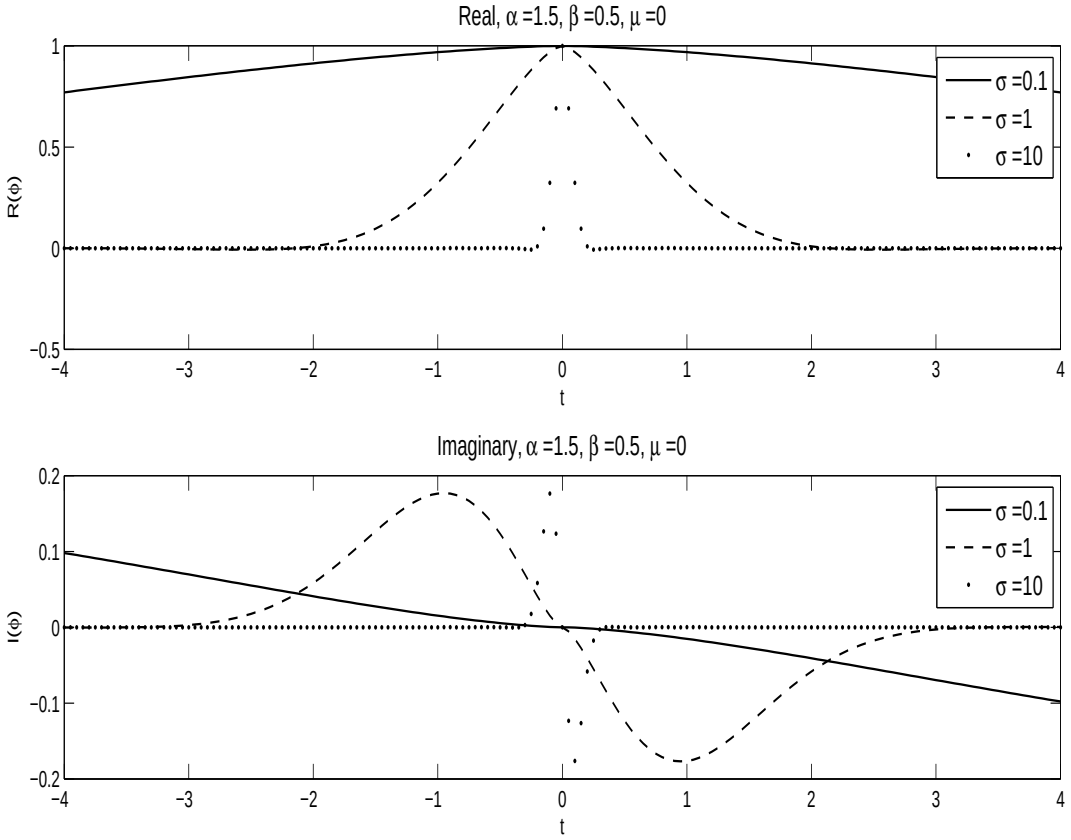

Figure 8.4: Graphs of the characteristic function for various values of the parameter σ.

8.2 Density function

The difficulty in working with the family of stable distributions is the non-existence of closed form density functions in general. To calculate density function values we apply (7.3) on page 106:

$$
\begin{aligned}
f(x;\boldsymbol{\theta}) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}\phi(t;\boldsymbol{\theta})\mathrm{d}t \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itx}e^{i\mu t-|\sigma t|^{\alpha}[1-i\beta\,\mathrm{sgn}(t)w(t,\alpha)]}\mathrm{d}t \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{i[(\mu-x)t+\sigma^{\alpha}|t|^{\alpha}\beta\,\mathrm{sgn}(t)w(t,\alpha)]}e^{\sigma^{\alpha}|t|^{\alpha}}\mathrm{d}t \\
&= \frac{1}{2\pi}\left[\int_{0}^{\infty} e^{i[(\mu-x)t+\sigma^{\alpha}t^{\alpha}\beta w(t,\alpha)]}e^{\sigma^{\alpha}t^{\alpha}}\mathrm{d}t\right. \\
&\qquad\qquad \left.+\int_{-\infty}^{0} e^{i[(\mu-x)t-\sigma^{\alpha}(-t)^{\alpha}\beta w(t,\alpha)]}e^{\sigma^{\alpha}(-t)^{\alpha}}\mathrm{d}t\right] \\
&= \frac{1}{2\pi}\left[\int_{0}^{\infty} e^{i[(\mu-x)t+\sigma^{\alpha}t^{\alpha}\beta w(t,\alpha)]}e^{\sigma^{\alpha}t^{\alpha}}\mathrm{d}t\right. \\
&\qquad\qquad \left.+\int_{0}^{\infty} e^{-i[(\mu-x)t+\sigma^{\alpha}t^{\alpha}\beta w(t,\alpha)]}e^{\sigma^{\alpha}t^{\alpha}}\mathrm{d}t\right] \\
&= \frac{1}{2\pi}\int_{0}^{\infty} e^{-\sigma^{\alpha}t^{\alpha}}\left\{e^{i[(\mu-x)t+\sigma^{\alpha}t^{\alpha}\beta w(t,\alpha)]}+e^{-i[(\mu-x)t+\sigma^{\alpha}t^{\alpha}\beta w(t,\alpha)]}\right\}\mathrm{d}t \\
&= \frac{1}{\pi}\int_{0}^{\infty} e^{-\sigma^{\alpha}t^{\alpha}}\cos\left[(\mu-x)t+\sigma^{\alpha}t^{\alpha}\beta w(t,\alpha)\right]\mathrm{d}t. \qquad (8.4)
\end{aligned}
$$

For the special case, with a single parameter α, we have the density function

$$
f(x;\alpha) = \frac{1}{\pi}\int_{0}^{\infty} e^{-t^{\alpha}}\cos\left(tx\right)\mathrm{d}t. \qquad (8.5)
$$

In this section the approximation of density function values of stably distributed random variables will be addressed. Initially, the density function in formula (8.5) for the single parameter case will be approximated. In general, density function values need to be approximated by evaluating an integral numerically. With the single parameter case two examples exist where the density functions can be evaluated analytically. One example, is the standard Cauchy density (Section 8.2.1) when $\alpha = 1$ and the other is the normal density with mean zero and variance equal to two (Section 8.2.2) when $\alpha = 2$. These special cases will be used to test various numerical methods applied to approximate single parameter stable density functions. The methods formed and tested in the first number of sections will be extrapolated to the computation of density functions for the general case (Section 8.2.7). Note that the results in this thesis will be restricted to $\alpha \in (1, 2)$. This is due to the objective to fit data to a distribution with the property that its tails are heavier than the normal distribution, but not as heavy as the tails of the Cauchy distribution.

8.2.1 The Cauchy distribution

The Cauchy distribution corresponds to the special case in (8.2) on page 111, with $\alpha = 1$. From formula (8.5) on page 116 the density function

$$f(x) = \frac{1}{\pi} \int_0^\infty e^{-t} \cos(tx) \mathrm{d}t, \tag{8.6}$$

for this special case can be computed, which equals the closed form

$$f(x) = \frac{1}{\pi(1+x^2)}; -\infty < x < \infty. \tag{8.7}$$

The density function in (8.6) will be approximated using various numerical integration techniques and tested against (8.7).

8.2.2 The normal distribution

The normal distribution with mean zero and variance equal to two corresponds to the special case in (8.2) on page 111, with $\alpha = 2$. From formula (8.5) on page 116 the density function

$$f(x) = \frac{1}{\pi} \int_0^\infty e^{-t^2} \cos(tx) \mathrm{d}t,$$

provides us with a formula to calculate the density function for the normal distribution, which equals the closed form

$$f(x) = \frac{1}{2\sqrt{\pi}} e^{-\frac{1}{4}x^2}; -\infty < x < \infty. \tag{8.8}$$

For b sufficiently large, we will approximate the integral $\int_0^\infty g(t)\mathrm{d}t$ by

$$\int_0^b g(t)\mathrm{d}t. \tag{8.9}$$

The following paragraphs will explain methods to numerically approximate an integral of the form (8.9).

8.2.3 Riemann-sums

In this section the Riemann-sum method will be described to approximate the integral

$$\int_a^b g(t)\mathrm{d}t$$

numerically. We will apply the method of Riemann-sums to not only approximate integrals of the form (8.9), and is therefore described more generally. Three different cases are defined using left-, right- and midpoint evaluation.

The interval $[a, b]$ is divided into n sub-intervals of equal length

$$\Delta t = t_i - t_{i-1},\ i = 1, 2, ..., n,$$

where $\{t_0, t_1, ..., t_n\}$ is a partition of $[a, b]$. The evaluation points, $t^*_{i,r} \in [t_{i-1}, t_i]$, are defined by

$$t^*_{i,r} = a + (r + i - 1)\Delta t\,, i = 1, ..., n,$$

where

$$r = \begin{cases} 0, & \text{for leftpoint evaluation} \\ \frac{1}{2}, & \text{for midpoint evaluation} \\ 1, & \text{for rightpoint evaluation.} \end{cases}$$

Then

$$\int_a^b g(t)\mathrm{d}t \approx \sum_{i=1}^{n} g(t^*_{i,r})\Delta t.$$

We will, for notation purposes, define the right hand side of the above equation by

$$R_r \int_a^b g(t)\mathrm{d}t = \sum_{i=1}^{n} g(t^*_{i,r})\Delta t.$$

8.2.4 Simpson's rule

The integral $\int_0^b g(t)\mathrm{d}t$ can be approximated using the Simpson rule. This numerical integration method is derived from fitting a second order polynomial $P_i(t)$ to the end points and midpoint of each interval $\{[t_{i-1}, t_i]\,, i = 1, 2, ..., n\}$ and calculating the sum of the integrals

$$\sum_{i=1}^{n} \int_{t_{i-1}}^{t_i} P_i(t)\mathrm{d}t,$$

where $\{t_0, t_1, ..., t_n\}$ is a partition of $[0, b]$.

The following formula can then be derived (Cheney and Kincaid, 1991, p.448)

$$\begin{aligned} &\int_0^b g(t)\mathrm{d}t \\ &\approx \frac{\Delta t}{3}\left[g(t_0) + 4g(t_1) + 2g(t_2) + 4g(t_3) + 2g(t_4) + ... + 4g(t_{n-1}) + f(t_n)\right] \\ &= \frac{\Delta t}{3}\left[g(t_0) + 4\sum_{i=1}^{n/2} g(t_{2i-1}) + 2\sum_{i=2}^{n/2} g(t_{2i-2}) + g(t_n)\right]. \end{aligned} \tag{8.10}$$

Again, for notational purposes the approximation is formally defined by

$$S\int_0^b g(t)\mathrm{d}t = \frac{\Delta t}{3}\left[g(t_0) + 4\sum_{i=1}^{n/2} g(t_{2i-1}) + 2\sum_{i=2}^{n/2} g(t_{2i-2}) + g(t_n)\right].$$

8.2.5 Application of Simpson's rule

The density function calculated by the integral in formula (8.6) on page 117 has an integrand of the form $g(t,x) = h(t)\cos(tx)$, where $h(t)$ is a decreasing, positive function of t (therefore h does not have a root). For fixed x the period of the function $\cos(tx)$ is equal to $\frac{2\pi}{|x|}$. Therefore, the period of $\cos(tx)$ is a function of x. This justifies the usage of a Simpson method application, adapted for this special case. In this application x will determine the number of evaluation points which are directly proportional to the number of periods of the function $\cos(tx)$ in the integration interval $[0,b]$. This method will be named the "adapted Simpson method", as the method will automatically take into account some properties of the integrand, due to a change in the value of x. With other adaptive methods, such as in (Cheney and Kincaid, 1991, p.471), the method automatically changes the length of the subintervals, until an acceptable precision is reached. These adaptive methods will not be used, as the computational time to approximate density functions should be kept to a minimum.

Let $p(x)$ be the number of full cycles in an interval $[0,b]$ and n_1 the number of chosen intervals per full cycle. A new upper bound for the interval can then be defined by

$$b_1(x) = \max\left(1, p(x)\left[\lfloor\frac{b}{p(x)}\rfloor + 1\right]\right).$$

The total number of intervals is $n(x) = n_1[p(x)+1]$ and the interval size is $\Delta t(x) = \frac{b_1(x)}{n(x)}$. The Simpson rule in formula (8.10) can be applied by forming a partition $\{t_0, t_1, ..., t_n\}$ of the interval $[0, b_1(x)]$, where $\Delta t(x)$ is the length between two consecutive points in the partition.

Therefore, when the standard Simpson approximation is used, given by $S\int_0^b g(t,x)\mathrm{d}t$, the values of Δt and n are independent of x. However, when using the adapted Simpson approximation Δt and n are dependent on x. We will use the notation $AS\int_0^b g(t,x)\mathrm{d}t$ for the adapted Simpson approximation to distinguish between the two Simpson approximations used in this thesis.

8.2.6 Results

We will use various tests, of which the first is based on approximating density function values for various values of x in formula (8.6) on page 117. We know that

$$1 = \int_{\mathbb{R}} f(x)\mathrm{d}x = \mathbb{P}\left(|X| > c\right) + \int_{-c}^{c} f(x)\mathrm{d}x,$$

and f approaches zero as x tend to $-\infty$ and ∞. To choose a value of c, such that

$$\left|1 - \int_{-c}^{c} f(x)\mathrm{d}x\right| = \mathbb{P}\left(|X| > c\right) < 0.0001,$$

we approximate the integral by a midpoint Riemann-sum

$$\int_{-c}^{c} f(x)\mathrm{d}x \approx R_{0.5} \int_{-c}^{c} f(x)\mathrm{d}x, \tag{8.11}$$

where x_i is chosen in the interval $[-c, c]$ according to the midpoint rule (with $\Delta x = 0.1$). Approximating the above integral by Riemann-sums should provide accurate results as a density function is smooth.

In the first case, the function values of $f(x_i)$ are calculated by assuming the density function of a Cauchy distribution (formula (8.7) on page 117). Therefore, approximation errors are made from the approximation of the integral in (8.11), and not from the calculations of $f(x_i)$.

The approximation of $\mathbb{P}(|X| > c)$ as a function of c is shown in Figure 8.5. From the figure we can derive that a value of $c \approx 6366.2$ can be used to ensure that

$$\mathbb{P}(|X| > c) < 0.0001.$$

Analytically, we can calculate the above probability if X is a random variable from the Cauchy distribution:

$$\mathbb{P}(|X| > c) = \frac{2}{\pi} \arctan(c).$$

Therefore,

$$c > \tan\left(\frac{\pi}{2}\left[1 - 0.0001\right]\right) = 6366.198$$

ensures that

$$\mathbb{P}(|X| > c) < 0.0001.$$

The result above illustrates the accuracy of a standard Riemann-sum with the approximation in (8.11).

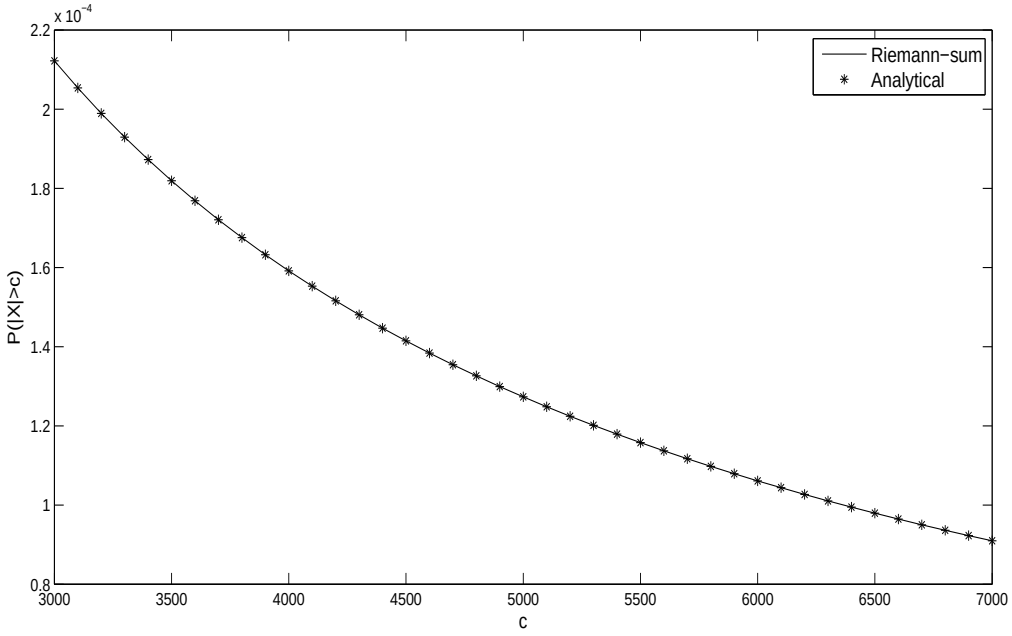

Figure 8.5: The probability that the absolute value of a Cauchy distributed random variable X is larger than a number c, as c becomes larger. The probability is approximated by a Riemann-sum with midpoint evaluation. The probability is also calculated analytically to show the accuracy of the approximation.

Now, for the second case we consider whether the probability

$$\mathbb{P}\left(|X| > c\right) \tag{8.12}$$

can be approximated accurately if X is a normally distributed random variable with mean zero and variance two (i.e. standard stable with $\alpha = 2$). The same method will be used as for the Cauchy distribution, where the closed form of the density function for the normal distribution with mean zero and variance two is given by formula (8.8) on page 117. We then approximate the probability in formula (8.12) by

$$\mathbb{P}\left(|X| > c\right) = 1 - \int_{-c}^{c} f(x)\mathrm{d}x \approx 1 - R_{0.5} \int_{-c}^{c} f(x)\mathrm{d}x,$$

where $\Delta x = 0.01$ and $[-c, c]$ are partitioned according the midpoint rule of a Riemann-sum.

The approximation of $\mathbb{P}\left(|X| > c\right)$ as a function of c is shown in Figure 8.6. From the figure we can derive that a value of $c \approx 5.5050$ can be used to ensure that

$$\mathbb{P}\left(|X| > c\right) < 0.0001.$$

With Matlab's "normcdf" function, we can approximate (Matlab's function also computes an approximate probability) the above probability if X is a random variable from the normal distribution. Therefore, for $c = 5.5050$, the following is obtained

$$\mathbb{P}\left(|X| < c\right) \approx 0.9996.$$

The graph in Figure 8.6 shows that a Riemann-sum provides us with good approximations.

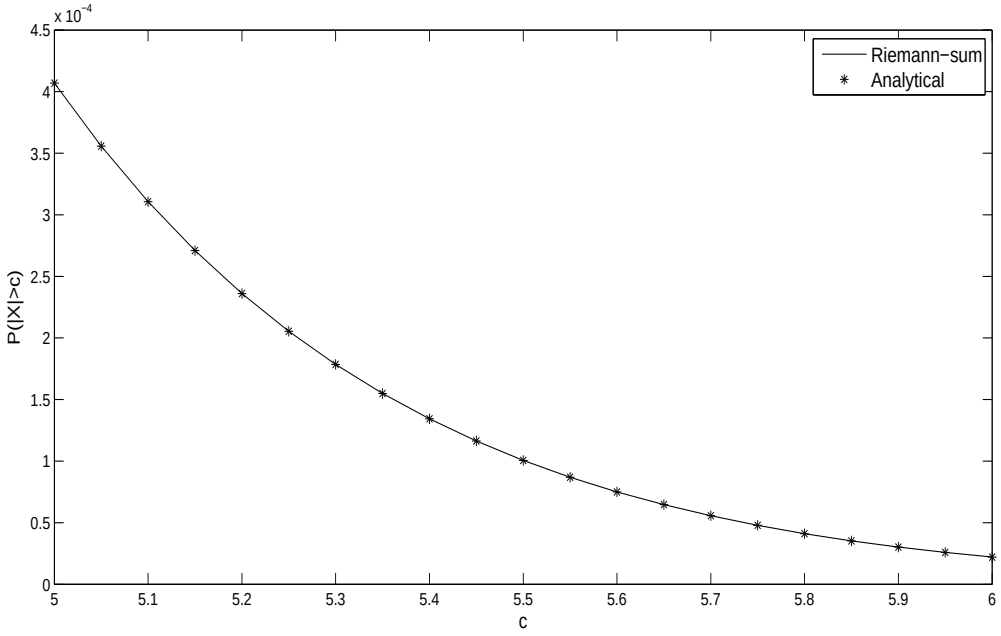

Figure 8.6: The probability that the absolute value of a normally distributed random variable X is larger than a number c, as c becomes larger. The probability is approximated by a Riemann-sum with midpoint evaluation. The accuracy of the approximation is tested against the results from Matlab's "normcdf" function.

The foundation has now been laid to test the accuracy of various numerical methods when approximating the integral

$$\frac{1}{\pi}\int_0^b e^{-t^\alpha}\cos\left(tx\right)\mathrm{d}t \tag{8.13}$$

in order to approximate density function values from formula (8.5) on page 116. To choose the value of b which ensures that

$$\int_b^\infty e^{-t^\alpha}\cos\left(tx\right)\mathrm{d}t \approx 0,$$

we note that

$$\left|e^{-t^\alpha}\cos\left(tx\right)\right| \leq e^{-t^\alpha} < 0.0001 \iff t > \left[\ln(10000)\right]^{1/\alpha}.$$

Figure 8.7 shows the graph of the function $b = [\ln(10000)]^{1/\alpha}$. Therefore, to approximate density function values a larger integration interval needs to be used when α is close to one, which has an impact on the computational time of our numerical integration methods.

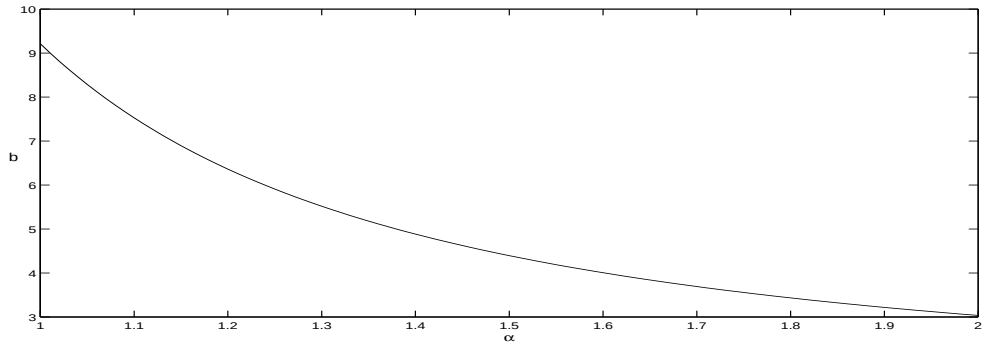

Figure 8.7: The logarithmic function that will be used to calculate the upper bound of the integration interval in formula (8.13) on page 122 (As a function of the parameter α).

We have now shown that

$$1 \approx \int_{-c}^{c} f(x)\mathrm{d}x = \int_{-c}^{c} f(x)\mathrm{d}x \approx R_{0.5}\int_{-c}^{c} f(x)\mathrm{d}x,$$

for c large enough, where $\alpha = 1$ and $\alpha = 2$. With these two cases the exact value of f is known. We assume

$$1 \approx R_{0.5}\int_{-c}^{c} f(x)\mathrm{d}x,$$

for all values $\alpha \in [1, 2]$. Therefore, for b large enough we have

$$1 \approx \frac{1}{\pi}R_{0.5}\int_{-c}^{c}\left(\int_{0}^{b} e^{-t^\alpha}\cos(tx)\mathrm{d}t\right)\mathrm{d}x.$$

The question is, which numerical method will approximate the integral

$$\int_{0}^{b} e^{-t^\alpha}\cos(tx)\mathrm{d}t, \tag{8.14}$$

such that

$$\frac{1}{\pi}R_{0.5}\int_{-c}^{c}\left(\int_{0}^{b} e^{-t^\alpha}\cos(tx)\mathrm{d}t\right)\mathrm{d}x,$$

is the nearest to 1. We will refer to this as the "total density" criterion.

Applying the "total density" criterion, some errors occur when either the Riemann-sum method or Simpson's method is used to approximate the integral in (8.14). For instance, to approximate the total density of a standard stably distributed random variable, X, for $\alpha = 1$ (i.e. X is Cauchy distributed) we have chosen $c = 6366.2$. Therefore, the density $f(x)$ in formula (8.6) on page 117 should be approximated for extreme values of x. A numerical error occurs for some of these extreme values. This is shown in Figure 8.8, where negative approximate density values occur for some values of x.

These numerical errors in Figure 8.8 seem to be cyclical. An application of the Simpson method (or adapted Simpson method, see Section 8.2.5) is used to automatically account for changing period lengths as x changes. This method seems to eliminate the numerical errors shown in Figure 8.8.

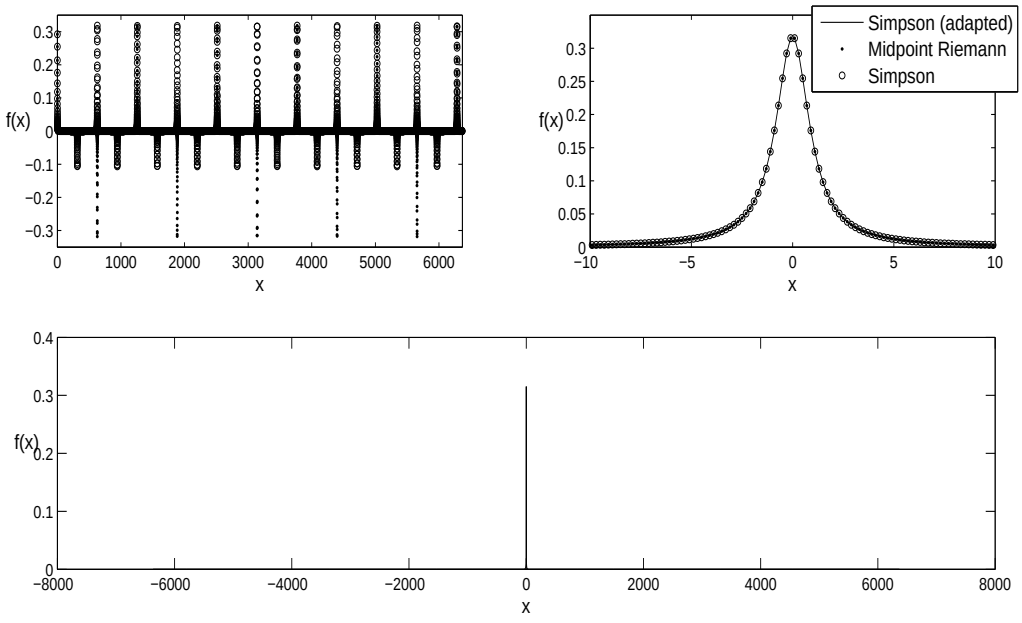

Figure 8.8: The top left figure shows the numerical errors that occur when approximating density values for large values of x, when using a standard Riemann-sum with midpoint evaluation or the Simpson method. The top right graph shows the accuracy of the approximation of density values for values of x close to zero. With the adapted Simpson method, the numerical errors seem to vanish, as can be seen in the top left figure. The density values, for a wide range of values of x, are shown in the bottom figure, if the adapted Simpson method is used.

One problem with the adapted Simpson method, is the number of evaluation points

used in the method for large values of x. The number of evaluation points is given by

$$k(x) = \left\lceil \frac{bx}{2\pi\Delta t} \right\rceil,$$

and for the Riemann or Simpson method by

$$k = \left\lceil \frac{b}{\Delta t} \right\rceil,$$

where $\lceil x \rceil$ denotes the first integer exceeding x.

Now, $k(x) > k \iff x > 2\pi$. This inequality means that for $x > 2\pi$ the adapted Simpson method takes longer to execute than the Riemann or Simpson methods. To reduce computational time, accuracy will be sacrificed. Therefore, for the Cauchy distribution, if we assume

$$c > \tan\left(\frac{\pi}{2}\left[1 - 0.002\right]\right) = 318.3088$$

then

$$\mathbb{P}\left(|X| > c\right) < 0.002. \tag{8.15}$$

Table 8.1 summarises the parameter values used for the various numerical methods. Let $c(\alpha)$, $\Delta x(\alpha)$ and $b(\alpha)$ denote the parameter values as functions of α. The following formulas show how these parameter values are calculated for values of α between 1 and 2:

$$c(\alpha) = \left[c(2) - c(1)\right](\alpha - 1) + c(1),$$

$$\Delta x(\alpha) = \left[\Delta x(2) - \Delta x(1)\right](\alpha - 1) + \Delta x(1),$$

and

$$b(\alpha) = \left[\ln(10000)\right]^{1/\alpha}.$$

The use of linear interpolation in the first two expressions is merely a matter of convenience in order to automate the numerical calculations.

Table 8.1: Parameter values for numerical integration methods.

parameter	$\alpha = 1$	$\alpha = 2$
c	318.31	5.505
Δx	0.1	0.01
Δt	0.01	0.01
b	9.21	3.03

Figure 8.9 shows how close to one the approximations of $P(|X| < c)$ get with the Riemann-sum method (left-,right-and midpoint evaluations), the Simpson rule method and the adapted Simpson method. The leftpoint and rightpoint Riemann-sum methods are clearly not accurate as they over- or underestimate the density values. The Simpson

rule method error shown in Figure 8.8 on page 124 has an influence on the probability value obtained for $\alpha = 1$. From the graphs in Figure 8.9 we conclude that the adapted Simpson method provides the best approximation. Note that the approximate value of

$$P(|X| < c) \approx 0.998$$

corresponds with the result in (8.15) on page 125. The adapted Simpson method therefore passed the "total density"" criterion.

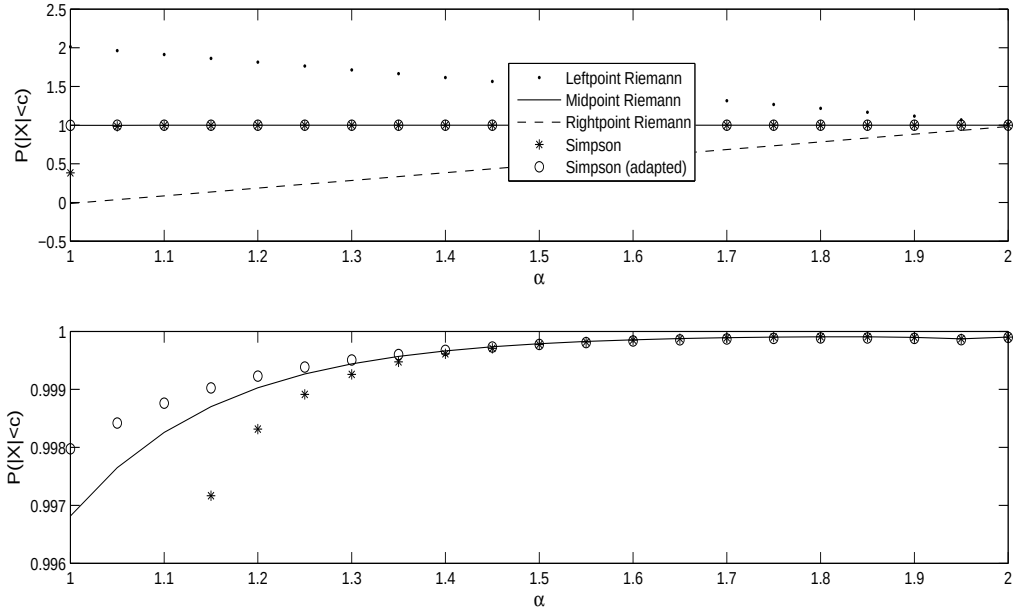

Figure 8.9: In the top and bottom figures the total density is approximated for different values of α, with the only difference being a change in the range on the vertical axis. The top figure shows the large errors when using a left- or rightpoint Riemann-sum, while the bottom figure shows the improved accuracy of the adapted Simpson method.

The second, more straightforward, test for our numerical methods is to calculate the actual error made in approximating density function values. Let

$$\hat{f}_R(x;\alpha) = \frac{1}{\pi} R_{0.5} \int_0^b e^{-t^\alpha} \cos(tx)\, \mathrm{d}t,$$

be the approximated density value using the midpoint Riemann-sum method and

$$\hat{f}_{AS}(x;\alpha) = \frac{1}{\pi} AS \int_0^b e^{-t^\alpha} \cos(tx)\, \mathrm{d}t,$$

be the approximated density value using the adapted Simpson method (These two methods have been chosen as they have shown the best results thus far.) Then the errors,

$df_R(x;\alpha)$ and $df_{AS}(x;\alpha)$ are equal to

$$df_R(x;\alpha) = \hat{f}_R(x;\alpha) - f(x;\alpha)$$

and

$$df_{AS}(x;\alpha) = \hat{f}_{AS}(x;\alpha) - f(x;\alpha).$$

The actual closed form density values, for $\alpha = 1$ and $\alpha = 2$, are obtained from formulas (8.7) and (8.8) on page 117.

The results are shown in Figure 8.10. In both cases the errors from the two methods obtained are almost equal to each other for values of x close to zero. However, the errors from the adapted Simpson method are smaller for larger absolute values of x.

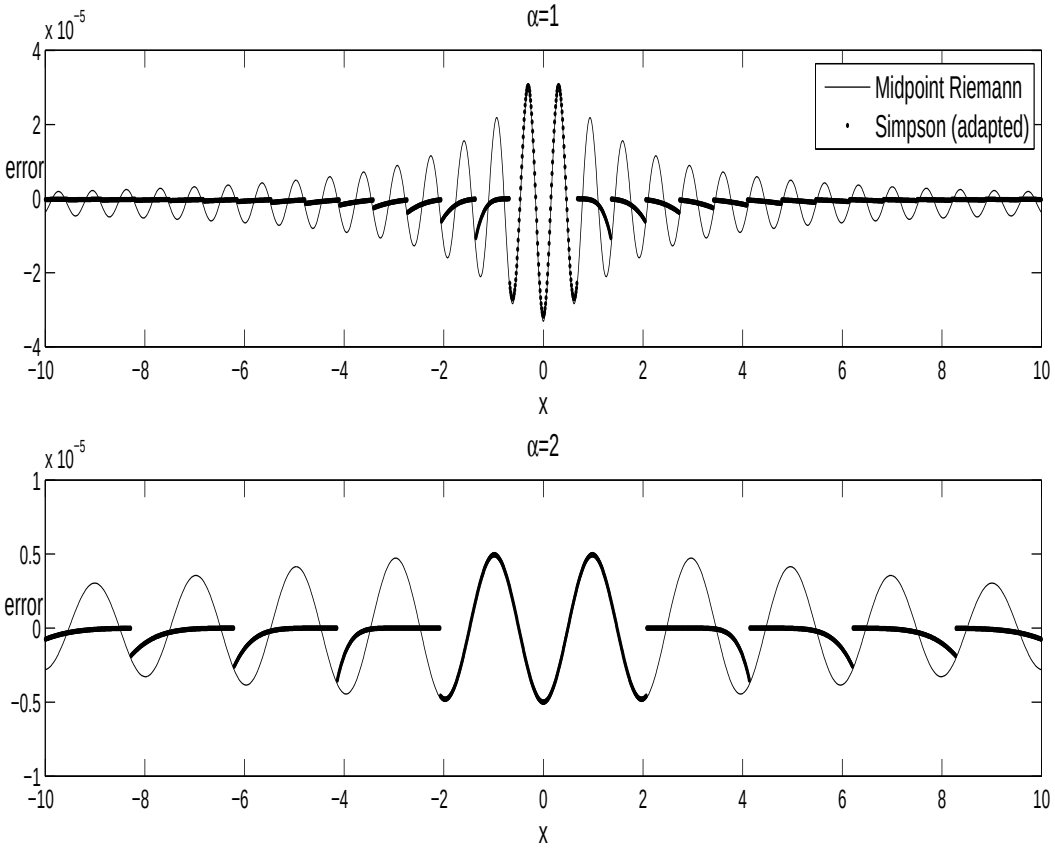

Figure 8.10: The two figures show the actual approximation errors when the Riemann-sum with midpoint evaluation and adapted Simpson methods are used. The top figure shows the error when the Cauchy density values are estimated, while the bottom figure shows the error when the normal density values are approximated.

8.2.7 Stable density functions in general

In this section a method to approximate stable density function values in the general case defined in formula (8.4) on page 116 will be explained.

In the previous two sections (Section 8.2.1 and Section 8.2.2) a numerical integration method was developed to approximate density function values of the form in formula (8.5) on page 116. Two aspects have to be taken into account in the numerical integration method, of which one is to find an interval where the integrand is not close to zero as the variable t tends to $\pm\infty$. Therefore, in the general case finding a value for b such that

$$e^{-(\sigma b)^\alpha} < \epsilon.$$

We will then approximate the integral in formula (8.4) on page 116 by

$$f(x;\boldsymbol{\theta}) \approx \frac{1}{\pi}\int_0^b e^{-\sigma^\alpha t^\alpha}\cos\left[(\mu-x)t+\sigma^\alpha t^\alpha \beta w(t,\alpha)\right]\mathrm{d}t,$$

for b sufficiently large. The value of b, as a function of ϵ, is chosen by using

$$b = \frac{1}{\sigma}\left[\log\left(\frac{1}{\epsilon}\right)\right]^{\frac{1}{\alpha}}.$$

The period of the trigonometric function

$$\cos\left[(\mu-x)t+\sigma^\alpha t^\alpha \beta w(t,\alpha)\right],$$

for the simple case where $\beta = 0$, $\mu = 0$ and $\sigma = 1$ (discussed in the previous section) is constant for fixed x. In this general case, the period is not constant.

Let

$$g(t;\boldsymbol{\theta}) = \begin{cases} At + Bt^\alpha & \text{if} \quad \alpha \neq 1 \\ At + Bt\log(t) & \text{if} \quad \alpha = 1, \end{cases}$$

with

$$A = \mu - x,$$

and

$$B = \begin{cases} \sigma^\alpha\beta\tan\left(\frac{\pi\alpha}{2}\right) & \text{if} \quad \alpha \neq 1 \\ -\frac{2}{\pi}\sigma\beta & \text{if} \quad \alpha = 1, \end{cases}$$

where $t \in [0,\infty)$ (Note that $t\log(t) \to 0$ as $t \to 0$.)

The function g has at most only one critical point on the interval $[0,\infty)$, as the derivative of g to t is given by

$$g'(t;\boldsymbol{\theta}) = \begin{cases} A + B\alpha t^{\alpha-1} & \text{if} \quad \alpha \neq 1 \\ A + B + B\log(t) & \text{if} \quad \alpha = 1, \end{cases}$$

and

$$g'(t_0;\boldsymbol{\theta}) = 0 \iff t_0 = \begin{cases} \left(-\frac{A}{\alpha B}\right)^{\frac{1}{\alpha-1}} & \text{if} \quad \alpha \neq 1 \\ \exp\left(-\frac{A+B}{B}\right) & \text{if} \quad \alpha = 1. \end{cases}$$

The function g can therefore either have one critical point on the interval $[0, \infty)$ or none. An example of each case is shown in Figures 8.11 and 8.12. These figures show that $\cos(g)$ has variable lengths of full cycles. In Figure 8.11 g has a critical point on $[0, \infty)$, which means the length of a full cycle for $\cos(g)$ is much longer near a critical point. Therefore, for larger absolute values of g', the shorter the length of a full cycle for $\cos(g)$.

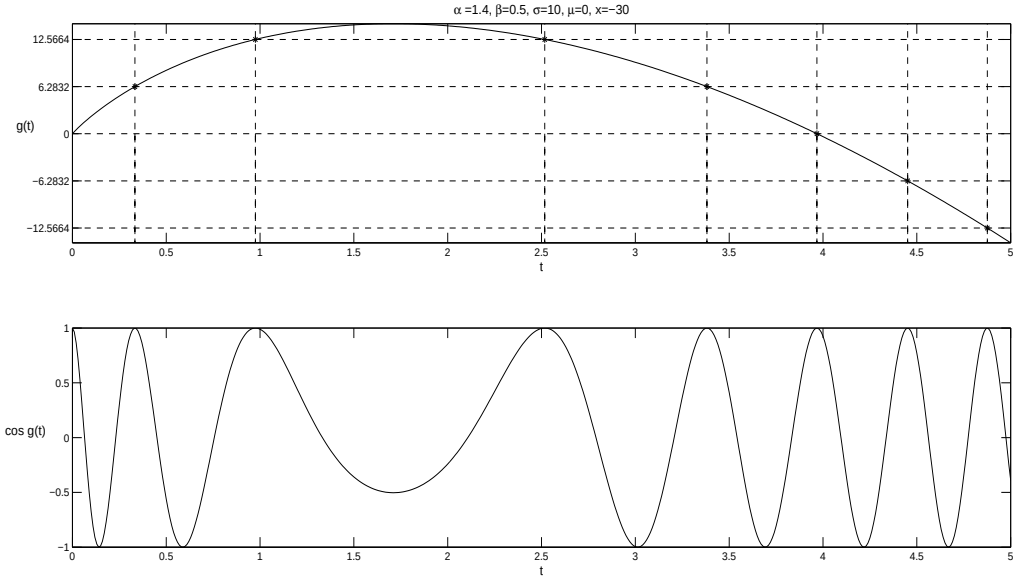

Figure 8.11: The function $g(t)$ (top figure) with a positive critical point, and its influence on the trigonometric function $\cos g(t)$ (bottom figure).

In Figure 8.12 g has no critical points and g becomes steeper as t becomes larger. The leads to shorter full cycle lengths for $\cos(g)$.

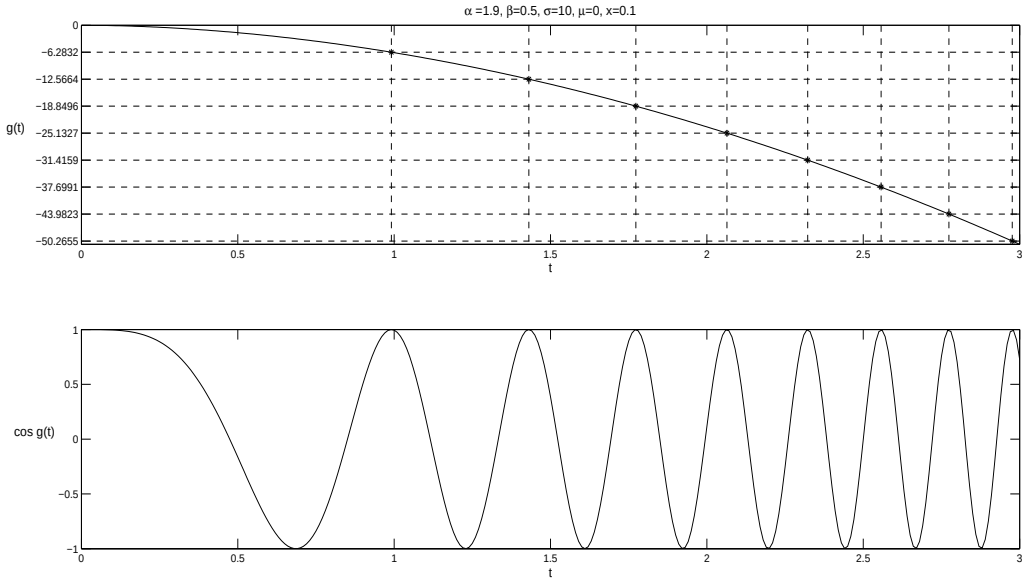

Figure 8.12: The function $g(t)$ (top figure) with a negative critical point, and its influence on the trigonometric function $\cos g(t)$ (bottom figure).

Good results should be obtained if the numerical integration method takes these changing cycle lengths into account. To achieve this the values of t must be obtained where

$$g(t;\boldsymbol{\theta}) = 2\pi k, \quad k \in \mathbb{Z}, \quad t \in [0, b]. \tag{8.16}$$

With this information the Simpson rule method can be used on each full cycle of $\cos\left[g(t;\boldsymbol{\theta})\right]$. Another route will be to count the number of times the function g obtains the values $2\pi k, \quad k \in \mathbb{Z}$. Therefore, let

$$p(x) = \sum_t \mathbb{1}_{(g(t;\boldsymbol{\theta})=k2\pi)}, \quad k \in \mathbb{Z}, \quad t \in [0, b],$$

then if n_1 is the number of chosen intervals per full cycle, the total number of intervals in the adapted Simpson method (Section 8.2.5) is

$$n(x) = n_1[p(x) + 1].$$

The partition $\{t_0, t_1, ..., t_n\}$ of the interval $[0, b]$ will be used in the Simpson rule method, where the length between the points in the partition is $\Delta t(x) = \frac{b}{n(x)}$. This method takes the number of full cycles on $[a, b]$ into account, but is less time consuming than finding the approximate values of t that satisfies (8.16) through a root finding method.

The "total density" criterion formulated in the previous section will be adjusted to apply in the general case. Therefore, approximate the integral

$$\int_{-c}^{c} f(x)\mathrm{d}x \approx R_{0.5} \int_{-c}^{c} f(x)\mathrm{d}x,$$

where f is approximated by

$$f(x) \approx \frac{1}{\pi} AS \int_0^b e^{-\sigma^\alpha t^\alpha} \cos\left[(\mu - x)t + \sigma^\alpha t^\alpha \beta w(t, \alpha)\right] \mathrm{d}t.$$

We use the notation of $AS \int_0^b h(t)\mathrm{d}t$ as the approximation of the integral $\int_0^b h(t)\mathrm{d}t$ by the adapted Simpson method described in this section.

The results from the "total density" criterion are shown in Figure 8.13. Three values of β were chosen to show that a change in β does not influence the accuracy of the adapted Simpson method. We have chosen the same parameter values for c, Δx and Δt given in Table 8.1 on page 125.

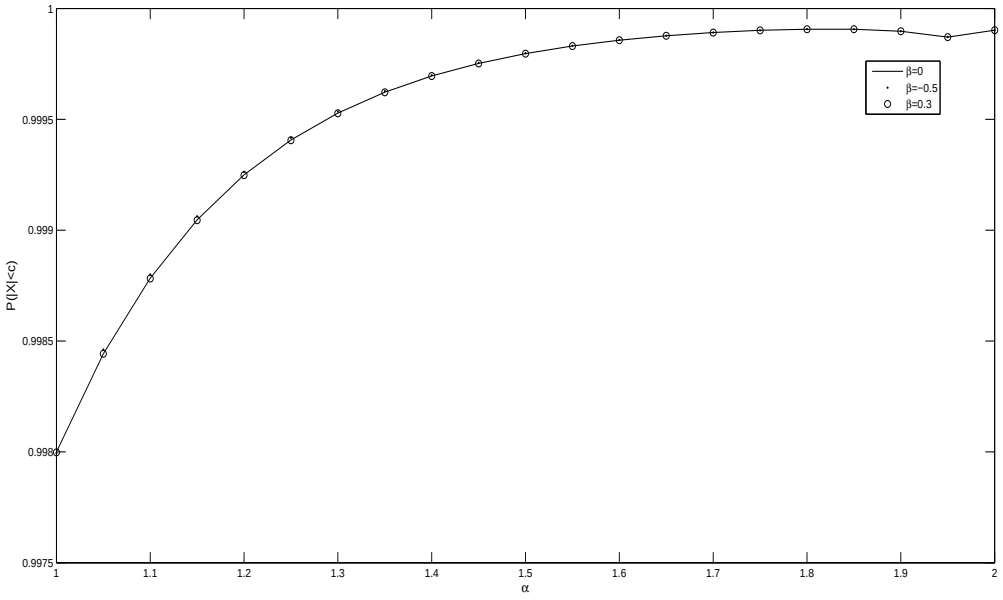

Figure 8.13: The results from the "total density" criterion used in the previous section (see Figure 8.9 on page 126) for varying values of β.

We used the methods described in this section to plot density functions for various values of the stable parameters. Graphs of these density functions are shown in Appendix B.

8.2.8 Zolotarev(M) parametrisation

As mentioned before, more than one parametrisation of stably distributed random variables exist. Nolan (1999) explores the approximation of stable densities using Zolotarev's (M) parametrisation, given in (7.6) on page 108.

The characteristic function of a normalised stable random variable Z ($\gamma = 0$, $\lambda = 1$) is then given by

$$\phi(t;\boldsymbol{\Theta}) = \begin{cases} \exp\left\{-|t|^{\alpha}\left[1 + i\beta\,\mathrm{sgn}(t)\tan\left(\frac{\pi\alpha}{2}\right)\left(|t|^{1-\alpha} - 1\right)\right]\right\}, & \text{if} \quad \alpha \neq 1 \\ \exp\left\{-|t|\left[1 + i\beta\,\mathrm{sgn}(t)\frac{2}{\pi}\log|t|\right]\right\}, & \text{if} \quad \alpha = 1, \end{cases}$$

for $t \in \mathbb{R}$ and $\boldsymbol{\Theta} = [\alpha, \beta]$. Then, by the inversion formula the density function of Z is given by

$$p(z;\boldsymbol{\Theta}) = \frac{1}{\pi}\int_0^{\infty} e^{-t^{\alpha}}\cos\left[h(z,t;\boldsymbol{\Theta}\right]\mathrm{d}t, \tag{8.17}$$

where

$$h(z,t;\boldsymbol{\Theta}) = \begin{cases} xt + \beta\tan\left(\frac{\pi\alpha}{2}\right)(t - t^{\alpha}), & \text{if} \quad \alpha \neq 1 \\ xt + \beta\frac{2}{\pi}t\log(t), & \text{if} \quad \alpha = 1. \end{cases}$$

The function h is a continuous function of the parameters α and β. By L'Hôpital's rule we have

$$\lim_{\alpha\to 1} h(z,t;\boldsymbol{\Theta}) = xt + \beta\left\{\lim_{\alpha\to 1}\frac{t - t^{\alpha}}{\cot\left(\frac{\pi\alpha}{2}\right)}\right\} = xt + \beta\frac{2}{\pi}t\log(t).$$

Therefore h is a continuous function of α at $\alpha = 1$. The continuity of h implies the continuity of the densities p. The density function from formula (8.4) on page 116 is not continuous at $\alpha = 1$, as the limit

$$\lim_{\alpha\to 1}\left\{(\mu - x)t + \sigma^{\alpha}t^{\alpha}\beta\tan\left(\frac{\pi\alpha}{2}\right)\right\},$$

does not exist. Because of this continuity property, Nolan (1999) prefers Zolotarev's (M) parametrisation when applying the maximum likelihood estimation method to estimate stable parameters.

Density function values of a general stable random variable, X, can be calculated from (8.17) by

$$f(x;\boldsymbol{\theta}) = \frac{1}{\sigma}p(z;\boldsymbol{\Theta}), \tag{8.18}$$

where

$$Z \overset{d}{=} \begin{cases} \frac{1}{\sigma}\left\{X - \left(\sigma\beta\tan\left(\frac{\pi\alpha}{2}\right) + \mu\right)\right\}, & \text{if} \quad \alpha \neq 1 \\ \frac{1}{\sigma}\left\{X - \left(\frac{2}{\pi}\beta\sigma\log(\sigma) + \mu\right)\right\}, & \text{if} \quad \alpha = 1. \end{cases}$$

The form of both density functions f and p is the same, which means the same numerical method that was used in the previous sections can be used to approximate densities of Z. Approximating densities of X, by using (8.18) therefore lead to no better results than approximating them directly from (8.4) on page 116.

8.3 The derivative functions

The partial derivatives of the density function to each parameter need to be computed in order to calculate the Fisher information described in Section 11.1. We will also use these derivatives to apply the maximum likelihood technique for parameter estimation (Section 9.2).

We will again distinguish between the single parameter case and the general case. The density function for the single parameter case is given in formula (8.5) on page 116. The partial derivative to α is given by

$$\frac{\partial}{\partial\alpha}f(x;\alpha) = -\frac{1}{\pi}\int_0^\infty e^{-t^\alpha}\cos(tx)\,t^\alpha\log(t)\mathrm{d}t.$$

A formula for the density function from the family of stable distributions in the general case had been derived in Section 8.2. Let the function g be defined by

$$g(t;\boldsymbol{\theta}) = (\mu - x)t + \sigma^\alpha t^\alpha\beta w(t,\alpha).$$

We will use the notation $g_\alpha(t;\boldsymbol{\theta}) = \frac{\partial}{\partial\alpha}g(t;\boldsymbol{\theta})$ for the partial derivative to the parameter α ($\alpha > 1, \beta \neq 0$ as g is not continuous at $\alpha = 1$). Similar formulas for g_β, g_σ and g_μ will be derived.

$$\begin{aligned}
g_\alpha(t;\boldsymbol{\theta}) &= (\sigma t)^\alpha\log(\sigma t)\beta w(t,\alpha) + (\sigma t)^\alpha\beta\sec^2\left(\frac{\pi\alpha}{2}\right)\frac{\pi}{2},\\
g_\beta(t;\boldsymbol{\theta}) &= (\sigma t)^\alpha w(t,\alpha),\\
g_\sigma(t;\boldsymbol{\theta}) &= \alpha\sigma^{\alpha-1}t^\alpha\beta w(t,\alpha),\\
g_\mu(t;\boldsymbol{\theta}) &= t.
\end{aligned}$$

Now, the partial derivatives f_α, f_β, f_σ and f_μ are given by

$$\begin{aligned}
f_\alpha(t;\boldsymbol{\theta}) &= -\frac{1}{\pi}\int_0^\infty e^{-(\sigma t)^\alpha}\left\{(\sigma t)^\alpha\log(\sigma t)\cos\left[g(t;\boldsymbol{\theta})\right]\right.\\
&\qquad\left.+\sin\left[g(t;\boldsymbol{\theta})\right]g_\alpha(t;\boldsymbol{\theta})\right\}\mathrm{d}t, \qquad (8.19)\\
f_\beta(t;\boldsymbol{\theta}) &= -\frac{1}{\pi}\int_0^\infty e^{-(\sigma t)^\alpha}\left\{\sin\left[g(t;\boldsymbol{\theta})\right]g_\beta(t;\boldsymbol{\theta})\right\}\mathrm{d}t, \qquad (8.20)\\
f_\sigma(t;\boldsymbol{\theta}) &= -\frac{1}{\pi}\int_0^\infty e^{-(\sigma t)^\alpha}\left\{\alpha\sigma^{\alpha-1}t^\alpha\cos\left[g(t;\boldsymbol{\theta})\right]\right.\\
&\qquad\left.+\sin\left[g(t;\boldsymbol{\theta}\right]g_\sigma(t;\boldsymbol{\theta})\right\}\mathrm{d}t, \qquad (8.21)\\
f_\mu(t;\boldsymbol{\theta}) &= -\frac{1}{\pi}\int_0^\infty e^{-(\sigma t)^\alpha}\left\{\sin\left[g(t;\boldsymbol{\theta})\right]g_\mu(t;\boldsymbol{\theta})\right\}\mathrm{d}t. \qquad (8.22)
\end{aligned}$$

8.4 Simulation of stably distributed random variables

The paper from Weron (1996) provides us with a method to simulate stable random variables. The method is described by the following algorithm:

1. Generate a random variable V uniformly distributed on $(\frac{-\pi}{2}, \frac{\pi}{2})$ and an independent exponential random variable W with mean 1.

2. For $\alpha \neq 1$ compute

   i
   $$X = S_{\alpha,\beta} \frac{\sin\left[\alpha\left(V + B_{\alpha,\beta}\right)\right]}{\left[\cos(V)\right]^{1/\alpha}} \left\{ \frac{\cos\left[V - \alpha(V + B_{\alpha,\beta})\right]}{W} \right\}^{(1-\alpha)/\alpha}$$
   with
   $$B_{\alpha,\beta} = \frac{\arctan\left[-\beta \tan(\frac{\pi\alpha}{2})\right]}{\alpha}$$
   and
   $$S_{\alpha,\beta} = \left\{ 1 + \beta^2 \left[\tan\left(\frac{\pi\alpha}{2}\right) \right]^2 \right\}^{\frac{1}{2\alpha}}$$

   ii
   $$Y = \sigma X + \mu$$

3. For $\alpha = 1$ calculate

   i
   $$X = \frac{2}{\pi} \left[\left(\frac{\pi}{2} - \beta V \right) \tan V + \beta \log\left(\frac{W \cos V}{\pi/2 - \beta V} \right) \right]$$

   ii
   $$Y = \sigma X - \frac{2}{\pi} \sigma \log \sigma + \mu$$

This method produces a stably distributed random variable Y calculated from the standard ($\mu = 0$, $\sigma = 1$) stably distributed variable X.

Chapter 9

Parameter estimation

In this chapter two parameter estimation methods will be used to estimate the stably distributed parameters. The most efficient estimates are maximum likelihood estimates (MLEs). The question is, is the method to compute these estimates time consuming? If $X_1, ..., X_n$ is a sample of i.i.d. distributed random variables, then the log-likelihood function is defined as

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(X_i, \boldsymbol{\theta}),$$

where $f(X_i, \boldsymbol{\theta})$ is the density function value of X_i for a set of parameter values $\boldsymbol{\theta}$. The parameter estimates are computed by finding the value of the parameter vector $\boldsymbol{\theta}$ which maximises the function L.

A different method to compute parameter estimates is called the integrated least squared estimation (ILSE) method, introduced by Heathcote (1997). The value of the parameter vector $\boldsymbol{\theta}$ needs to be found which minimises the function I, given by

$$I(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} |\hat{\phi}(t) - \phi(t, \boldsymbol{\theta})|^2 e^{-t^2} \mathrm{d}t,$$

where $\hat{\phi}(t)$ is the empirical characteristic function for a sample $X_1, ..., X_n$ and is given by

$$\hat{\phi}(t) = \frac{1}{n} \sum_{i=1}^{n} \exp(itX_i).$$

The ILSE method may be less time consuming than the MLE method. This is due to the closed form of the characteristic functions for the family of stable distributions, compared to density function values that needs to be approximated by a numerical integration method. We have shown how to approximate density function values of f, as well as its derivatives to the parameters α, β, σ and μ. A gradient-based optimisation method in which we approximate the gradient of the function L will be used to find MLEs. The

same method - with the same initial point - will be used to find the ILSEs, this time by approximating the gradient vector of the function I.

9.1 Steepest descent method

Let a function be defined by $G(\boldsymbol{\theta}) : S \mapsto \mathbb{R}$, where $S \subset \mathbb{R}^n$ and $\boldsymbol{\theta} = [\theta_1\theta_2...\theta_n]^T$. The direction of steepest descent (direction in which G has a minimal directional derivative) is given by the unit vector

$$\boldsymbol{s} = \frac{-\nabla G(\boldsymbol{\theta})}{\|\nabla G(\boldsymbol{\theta})\|}.$$

Here $\nabla = \left[\frac{\partial G}{\partial \theta_1}...\frac{\partial G}{\partial \theta_n}\right]$ denotes the gradient vector and $\|\cdot\|$ denotes length. A standard method to find a local minimum value of G is the steepest descent method. A steepest descent optimisation method can be visualised as walking from a starting point $\boldsymbol{\theta}_0$ in the direction of the unit gradient vector

$$\boldsymbol{s}_0 = \frac{-\nabla G(\boldsymbol{\theta}_0)}{\|\nabla G(\boldsymbol{\theta}_0)\|},$$

until a minimum value of G in this direction is found. The point at which the minimum value, $\boldsymbol{\theta}_1$, is found. The value $\boldsymbol{\theta}_1$ is then used to calculate a new direction $\boldsymbol{s}_1$ to walk into until a minimum value of G in the new direction is found. This process can be repeated until no further progress is made. With an unbounded problem an algorithm of the method is given by (Brandimarte, 2006, p.339):

Start at a point $\boldsymbol{\theta}_0$, and calculate for iterations $i = 0, 1, 2, ...$ until $\|\boldsymbol{\theta}_{i+1} - \boldsymbol{\theta}_i\|$ is small enough.

i Calculate

$$\boldsymbol{s}_i = \frac{-\nabla G(\boldsymbol{\theta}_i)}{\|\nabla G(\boldsymbol{\theta}_i)\|}.$$

ii Find the value of $t = t_i$ which minimises the function

$$G(\boldsymbol{r}(t))$$

on the line

$$\boldsymbol{r}(t) = \boldsymbol{\theta}_i + t\boldsymbol{s}_i$$

iii Calculate

$$\boldsymbol{\theta}_{i+1} = \boldsymbol{r}(t_i).$$

To find the minimum value of G on a closed region, the extreme value theorem (see Burden et al. (2016, p.5)) is used.

Theorem 5 (Extreme value theorem)**.** *Let the function $G : [a, b] \in \mathbb{R} \mapsto \mathbb{R}$ be differentiable on (a, b). Then the absolute maximum and minimum values of G are either at points a or b or at any points $c \in (a, b)$ where $G'(c) = 0$.*

To accommodate the extreme value theorem for single variable functions, the second step of the algorithm for the steepest descent method is changed to finding the value of $t = t_i$ which minimises the function

$$G(\boldsymbol{r}(t)), \quad t \in [0, b], \tag{9.1}$$

where b is the point where the line $\boldsymbol{r}(t)$ crosses the boundary of the closed region S.

We will apply the method to functions of the stable parameters. Therefore, the steepest descent method must be applied to the region with boundaries given by

$$S = \left\{(\alpha, \beta, \sigma, \mu) \in \mathbb{R}^4 | 1 \leq \alpha \leq 2, -1 \leq \beta \leq 1, \sigma \geq \delta > 0\right\},$$

where some rules need to be brought into the iteration process (δ is chosen as a small positive number).

Let $\mathbf{i_1}$, $\mathbf{i_2}$, $\mathbf{i_3}$ and $\mathbf{i_4}$ be the unit vectors in the direction of the α, β, σ and μ axes, respectively. The gradient vector can then be written as

$$\nabla G = G_\alpha \mathbf{i_1} + G_\beta \mathbf{i_2} + G_\sigma \mathbf{i_3} + G_\mu \mathbf{i_4},$$

where G_α, G_β, G_σ and G_μ are the partial derivatives relative to the parameters α, β, σ and μ. If

$$\nabla G = G_\sigma \mathbf{i_3} + G_\mu \mathbf{i_4},$$

where $G_\sigma \geq 0$, then the steepest descent method is applied in its original form.

If this is not the case, i.e. $G_\alpha \neq 0$, $G_\beta \neq 0$ or $G_\sigma < 0$, then the steepest descent method, formalised in (9.1) is applied.

9.2 Maximum likelihood estimates

Maximum likelihood estimation is a standard method to estimate parameters of a distribution. An advantage of the method is its theoretical asymptotic properties.

Let the random variables $X_1, ..., X_n$ have a joint density function f with parameter vector $\boldsymbol{\theta}$. The likelihood function is defined in terms of the density function

$$l(\boldsymbol{\theta}) = f(x_1, x_2, ..., x_n | \boldsymbol{\theta}),$$

where $X_i = x_i, \quad i = 1, 2, ..., n$.

If the random variables are i.i.d., then from independence of X_i, the likelihood function can be written as

$$l(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(X_i, \boldsymbol{\theta}),$$

where the density function f can be approximated by formula (8.4) on page 116 if X_i are stably distributed. In some cases the log-likelihood function

$$L(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(X_i, \boldsymbol{\theta}),$$

is more useful. The maximum likelihood estimates are obtained by finding those parameter values that maximises the likelihood function l. The same parameter values will be obtained if we maximise the log-likelihood function L.

The steepest descent method described in the preceding section will be used to maximise the log-likelihood function L. Let $G(\boldsymbol{\theta}) = -L(\boldsymbol{\theta})$. By choosing a set of initial parameter values

$$\boldsymbol{\theta}_0 = [\alpha_0 \quad \beta_0 \quad \sigma_0 \quad \mu_0].$$

We will search for a minimum value of G in the direction of $\nabla G = -\nabla L$, where

$$\nabla L(\boldsymbol{\theta}) = [L_\alpha \; L_\beta \; L_\sigma \; L_\mu] \, .$$

The derivative of L relative to the parameter α is given by

$$L_\alpha = \sum_{i=1}^{n} \frac{f_\alpha(X_i, \boldsymbol{\theta})}{f(X_i, \boldsymbol{\theta})}.$$

Similarly, the partial derivatives L_β, L_σ and L_μ can be approximated using the derivatives of the density function given in formulas (8.19) to (8.22) on page 133.

9.3 The integrated least squared error method

An estimation technique based on the characteristic function was developed by Paulson et al. (1975) and is defined by minimising the objective function defined in formula (7.5) on page 108. The steepest descent method described in Section 9.1 will be used. We therefore need to calculate the gradient vector of I

$$\nabla I = [I_\alpha \; I_\beta \; I_\sigma \; I_\mu] \, .$$

Now, the function I can be rewritten as

$$I(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \left([U_n(t) - u(t;\boldsymbol{\theta})]^2 + [V_n(t) - v(t;\boldsymbol{\theta})]^2 \right) e^{-t^2} \mathrm{d}t,$$

with

$$U_n(t) = \frac{1}{n}\sum_{i=1}^{n}\cos(tX_i),$$
$$V_n(t) = \frac{1}{n}\sum_{i=1}^{n}\sin(tX_i).$$

The functions u and v are defined in formula (8.3) on page 111. The partial derivatives of this function I relative to the stable parameters α, β, σ and μ are given by

$$I'(\boldsymbol{\theta}) = -2\int_{-\infty}^{\infty}\left\{\left[U_n(t) - u(t;\boldsymbol{\theta})\right]u'(t;\boldsymbol{\theta}) + \left[V_n(t) - v(t;\boldsymbol{\theta})\right]v'(t;\boldsymbol{\theta})\right\}e^{-t^2}\mathrm{d}t,$$

where I' denotes the derivative of I relative to any one of the parameters α, β, σ or μ. Therefore, we need formulas for the partial derivatives of u and v to α, β, σ and μ. These formulas will only differ slightly to those that was derived in Section 8.3.

Let the function g be defined by

$$g(t;\boldsymbol{\theta}) = \mu t + |\sigma t|^{\alpha}\operatorname{sgn}(t)\beta w(t,\alpha).$$

Similarly we will derive formulas for g_β, g_σ and g_μ.

$$\begin{aligned}
g_\alpha(t;\boldsymbol{\theta}) &= |\sigma t|^{\alpha}\log|\sigma t|\operatorname{sgn}(t)\beta w(t,\alpha) + |\sigma t|^{\alpha}\operatorname{sgn}(t)\beta\sec^2\left(\frac{\pi\alpha}{2}\right)\frac{\pi}{2},\\
g_\beta(t;\boldsymbol{\theta}) &= |\sigma t|^{\alpha}\operatorname{sgn}(t)w(t,\alpha),\\
g_\sigma(t;\boldsymbol{\theta}) &= \alpha\sigma^{\alpha-1}|t|^{\alpha}\operatorname{sgn}(t)\beta w(t,\alpha),\\
g_\mu(t;\boldsymbol{\theta}) &= t.
\end{aligned}$$

Notice that g_α is not defined at $\alpha = 1$ if $\beta \neq 0$ as $\lim_{\beta\neq 0,\alpha\to 1} g$ does not exist. Now, we can derive the formulas for the partial derivatives u_α, u_β, u_σ and u_μ.

$$\begin{aligned}
u_\alpha(t;\boldsymbol{\theta}) &= -e^{-|\sigma t|^{\alpha}}\left\{|\sigma t|^{\alpha}\log|\sigma t|\cos\left[g(t;\boldsymbol{\theta})\right] + \sin\left[g(t;\boldsymbol{\theta})\right]g_\alpha(t;\boldsymbol{\theta})\right\}\mathrm{d}t\\
u_\beta(t;\boldsymbol{\theta}) &= -e^{-|\sigma t|^{\alpha}}\left\{\sin\left[g(t;\boldsymbol{\theta})\right]g_\beta(t;\boldsymbol{\theta})\right\}\mathrm{d}t\\
u_\sigma(t;\boldsymbol{\theta}) &= -e^{-|\sigma t|^{\alpha}}\left\{\alpha\sigma^{\alpha-1}|t|^{\alpha}\cos\left[g(t;\boldsymbol{\theta})\right] + \sin\left[g(t;\boldsymbol{\theta}\right]g_\sigma(t;\boldsymbol{\theta})\right\}\mathrm{d}t\\
u_\mu(t;\boldsymbol{\theta}) &= -e^{-|\sigma t|^{\alpha}}\left\{\sin\left[g(t;\boldsymbol{\theta})\right]g_\mu(t;\boldsymbol{\theta})\right\}\mathrm{d}t.
\end{aligned} \tag{9.2}$$

Similarly, formulas for v_α, v_β, v_σ and v_μ can be derived.

9.4 Initial estimates

A global extreme value of a function may be difficult to find numerically if the function has multiple critical points. One way to curtail this problem, is by choosing a good starting

point in the optimisation method. Parameter estimates used as starting point estimates do not necessarily need to be the most efficient estimates, but the least time consuming. Therefore, an initial effort is made to choose estimates which can be calculated directly from a sample, such as a sample mean. Fama and Roll (1968) provides us with estimates for the stable parameters μ, σ and α. Various truncated means are used for different values of α, but we decide to use the sample mean for an estimate of μ, knowing that if there are any signs of it being biased it is due to extreme outlier values. Therefore, if $X_1, ..., X_n$ is a sample then the initial estimate for the parameter μ is given by the sample mean

$$\mu_0 = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

The estimate for σ that Fama and Roll (1968) proposes is given by

$$\sigma_0 = \frac{1}{1.654}\left[x_{0.72} - x_{0.28}\right],$$

where x_p is the p^{th} sample percentile. Fama and Roll (1968) proposes an empirical method to estimate α which does not meet our requirements.

To estimate the skewness parameter β, it will be sufficient in the present situation to at least get an estimate with the correct sign for non-symmetrical distributions. A simple method to observe outliers is to identify those values of a sample that lie outside the interval

$$[x_{0.25} - 1,5 \times IQR, x_{0.75} + 1,5 \times IQR], \tag{9.3}$$

where IQR is the interquartile range (Rice, 1995, p.372).

The number of sample points greater than the upper bound of this interval, considered as a fraction of the number of sample points outside this interval, will indicate whether the sample is skewed to the right ($\beta > 0$) or skewed to the left ($\beta < 0$). Therefore, define a statistic x^*

$$x^*(k) = \frac{\#\left(X_i \geq x_{0.75} + k \times IQR\right)}{\#\left(X_i \leq x_{0.25} - k \times IQR\right) + \#\left(X_i \geq x_{0.75} + k \times IQR\right)}. \tag{9.4}$$

The values of x^* will be between 0 and 1, where the value $x^*(k) = 0$ should correspond to $\beta = -1$ and $x^*(k) = 1$ should correspond to $\beta = 1$. Assuming the relationship between β and $x^*(k)$ is linear, we propose the following estimate for β

$$\beta_0 = -1 + 2 \times x^*(k). \tag{9.5}$$

An estimate based on the number of values outside the interval in (9.3) will also be derived for α. A normal distribution will not have many values outside the interval, while a Cauchy distribution will have a greater number of extreme values. Therefore an

initial estimate for α can be given by

$$\alpha_0 = 2 - \frac{\#\left(X_i \leq x_{0.25} - k \times IQR\right) + \#\left(X_i \geq x_{0.75} + k \times IQR\right)}{n\hat{q}(n)}, \tag{9.6}$$

where $\hat{q}(n)$ is an estimate of the number of outliers as a fraction of a sample size for the Cauchy distribution, and n is the size of the sample. The more outliers observed, the closer to one the quotient in formula (9.6) is, as the fraction of outlying values to the sample size will be closer to $\hat{q}$. If there are very few outlying values, then the quotient in formula (9.6) will be close to zero, and the value of α_0 will be close to two, which corresponds to its value for the normal distribution.

9.5 Parameter estimation results

In the first part of this section, the effectiveness of the initial parameter estimate for β is analysed. Formula (9.5) on page 140 is used to get an initial estimate for β. This estimate behaves surprisingly well, and this is shown in Figures 9.1 to $B.6$.

Let $\mathbf{X}(\mu, \beta, \sigma, \mu)$ be a simulated sample from the stable distribution with parameters α, β, σ and μ of size $100,000$. Let $\beta_1 = -0.9$, $\beta_2 = -0.3$, $\beta_3 = 0$ and $\beta_4 = 0.5$ and define the samples

$$\mathbf{X}_j(\alpha) = \mathbf{X}(\alpha, \beta_j, 1, 0), \quad j = 1, 2, 3, 4,$$

$$\mathbf{X}_j(\sigma) = \mathbf{X}(1.5, \beta_j, \sigma, 0), \quad j = 1, 2, 3, 4,$$

and

$$\mathbf{X}_j(\mu) = \mathbf{X}(1.5, \beta_j, 1, \mu), \quad j = 1, 2, 3, 4,$$

for a range of α, σ and μ values. Using formulas (9.4) and (9.5) on page 140 we calculate the following statistic for each of these sets of samples, where $k = 1.75$:

$$x_j^*(\alpha),\ x_j^*(\sigma),\ x_j^*(\mu), \quad j = 1, 2, 3, 4 \tag{9.7}$$

and

$$\beta_{0j}(\alpha),\ \beta_{0j}(\sigma),\ \beta_{0j}(\mu), \quad j = 1, 2, 3, 4. \tag{9.8}$$

Figure 9.1 shows these statistics for a range of α values, Figure $B.5$ for a range of σ values and Figure $B.6$ for a range of μ values. From the top graph of each figure, the conclusion can be made that

$$\beta_{0j}(\alpha) \approx \beta_j, \quad j = 1, 2, 3, 4.$$

except when α is close to 2. In Appendix B we find similar results for $\beta_{0j}(\sigma),\ \ j = 1, 2, 3, 4$, and $\beta_{0j}(\mu),\ \ j = 1, 2, 3, 4$. This shows that the initial estimate β_0 seems to be an unbiased estimator of β.

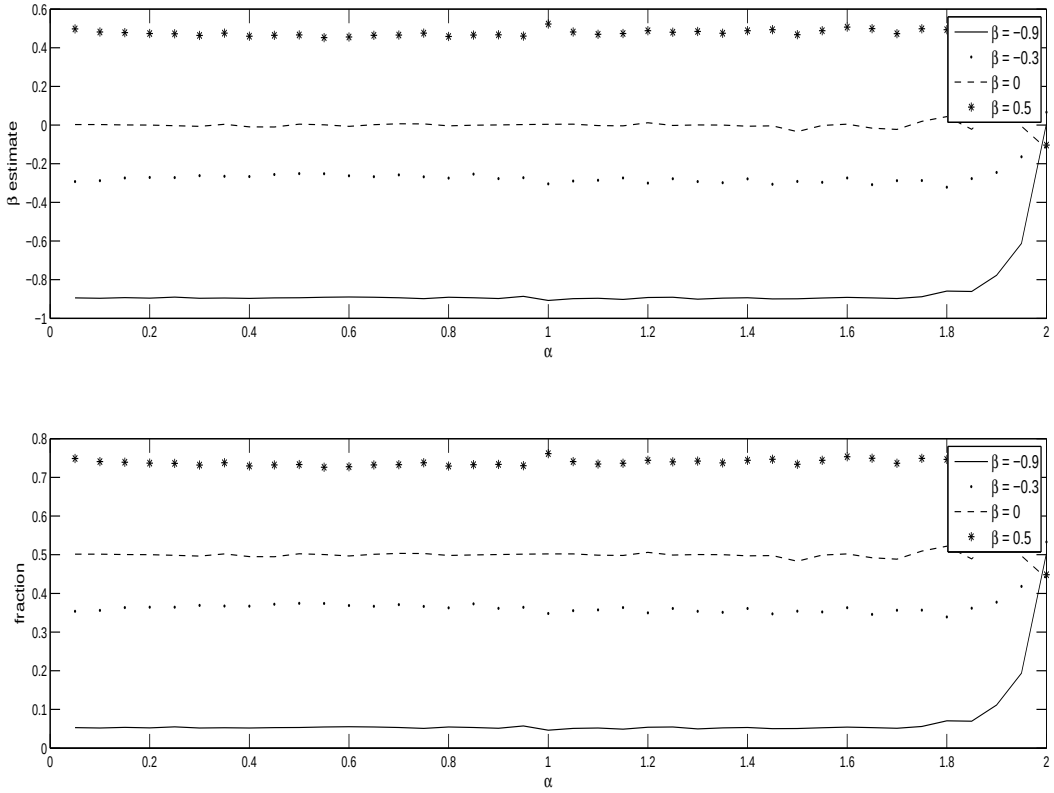

Figure 9.1: The statistic β_0 and x^* for a range of α parameter values.

Figure 9.2 shows the efficiency of the initial estimates (defined in Section 9.4), the MLEs (defined in Section 9.2) and the ILSEs (defined in Section 9.3). To construct these box plots in Figure 9.2, 50 samples of size 1000 were simulated with input parameters (chosen randomly) as follows

$$\alpha = 1.5,\ \beta = 0.4,\ \sigma = 0.4,\ \mu = -0.1.$$

The parameters were then estimated by the respective methods for each sample, to obtain 50 sets of parameter estimates for each method. Statistics could then be calculated from these samples of size 50 and the resulting box plots are shown in Figure 9.2.

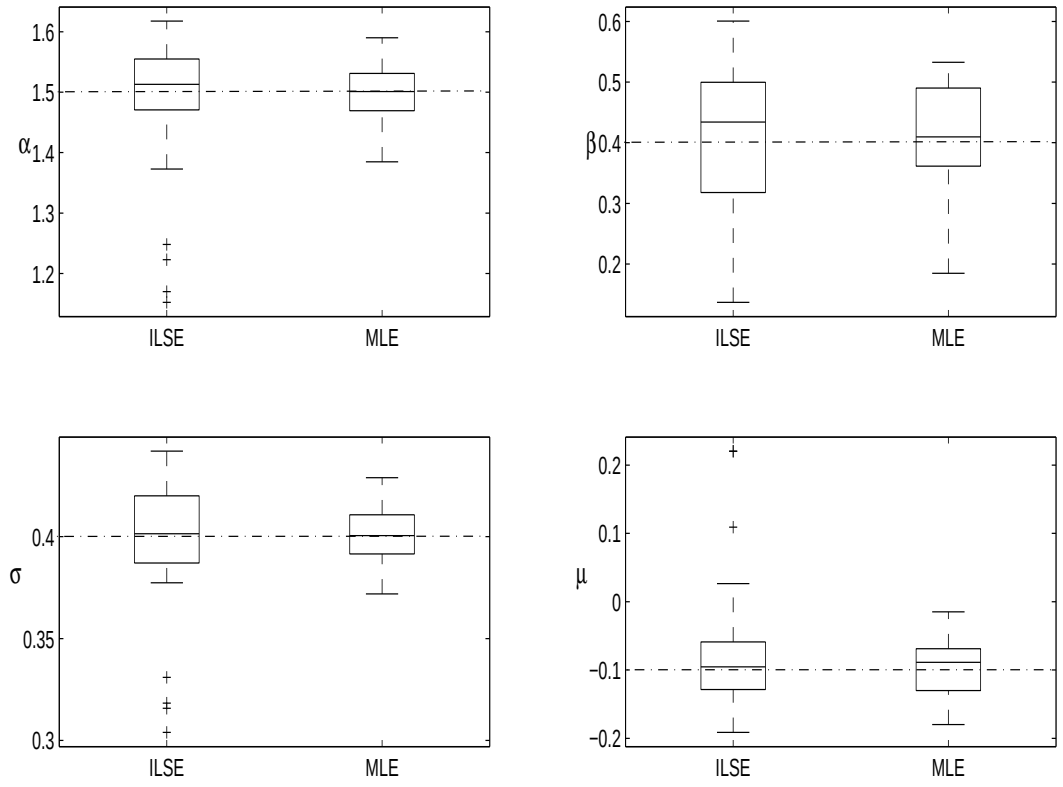

Figure 9.2: Box plots of parameter estimates from both estimation methods and for each parameter.

The box plots show that the MLEs are less median biased and have smaller IQRs than the ILSEs, hence would typically be considered to be preferred. The one statistic not shown in these graphs is the time it took to calculate the estimates. On average the ILSEs were calculated within 50 seconds, while the MLEs required 535 seconds.

The steepest descent method is "not recommended as a serious optimisation procedure" (Bunday and Garside, 1987, p.62), which means improvement can be made in terms of the optimisation method applied. Improvements may also be made applying a different parametrisation. However, the same initial values were used and the same gradient method was applied in finding the MLEs and ILSEs. The conclusion can therefore be made that the ILSE method leads to less time consuming estimates, compared to the MLE method in approximating stable parameters defined by Definition 2 on page 106. This is due to the computational time in approximating stable density function values. The Box plots in Figure 9.2 however suggests the ILSE method to be less efficient in terms of accuracy (which is to be expected). The relative efficiency of the ILSEs, compared to

the MLEs is the subject of research in the following sections.

Chapter 10

Statistical inference in the family of stable distributions

In Chapter 8 we lay the foundation to apply parameter estimation methods. An important result was the method to approximated stable densities. As a result we could approximate stable distribution function values used in Chapter 5. In Chapter 9 we implemented two methods to estimate stably distributed parameters. We implemented the methods to fit a distribution from the family of stable distributions to the changes of the 3-month JIBAR in Chapter 5. The advantage of using the MLE is its favourable asymptotic properties, with the estimation error being asymptotically the smallest of all estimators. However, the numerical approximation error in calculating the MLE is possibly larger than in other estimators. This is due to the involved approximation of the densities and their partial derivatives which are used to find the critical points of a multivariable log-likelihood function.

In this chapter and the next our goal is to determine the asymptotic variance of both the MLEs and the ILSEs. As a consequence we will be able to approximate the relative efficiency of the ILSEs in comparison with the most efficient MLEs. Although none of the results from these two chapters were used directly in Part I, the results give us an indication of how accurate MLEs and ILSEs are.

If an estimate $\tilde{\theta}_n$ of a parameter is consistent, then the accuracy of $\tilde{\theta}_n$ can be analysed by calculating the asymptotic variance of $\tilde{\theta}_n$. In this chapter we will describe the theory to calculate the asymptotic variance of both the MLEs and ILSEs. The asymptotic variance of each MLE is the diagonal elements of the inverse of the Fisher information matrix. We will therefore describe how to calculate the Fisher information matrix in this chapter. Heathcote (1997) provides us with the results to calculate the asymptotic variance of the ILSEs. In this chapter and Appendix A we give a detailed exposition of the results of Heathcote (1997).

The results from this chapter are implemented in Chapter 11 to approximate the relative efficiency of the ILSEs compared to the most efficient MLEs.

10.1 Relative efficiency of integrated least squared estimators

In this section the consistency and asymptotic normality of the ILSEs will be investigated. Some of the derivations will be provided in Appendix A, to complement the derivations and results shown in this section. In Section 9.3 the method was described and implemented to estimate the parameters of the stable distribution. Heathcote (1997) investigated the asymptotic properties of the method in a more general setting. Let θ be a parameter and let the objective function be defined by

$$I_n(\theta) = \int_{-\infty}^{\infty} |\hat{\phi}_n(t) - \phi(t;\theta)|^2 dG(t), \tag{10.1}$$

where

$$\hat{\phi}_n(t) = \frac{1}{n}\sum_{j=1}^{n} e^{itX_j}$$

is the empirical characteristic function of the sample $X_1, ..., X_n$ and $\phi(t;\theta)$ is the characteristic function of a general distribution with one parameter (The multiple parameter case is a straightforward extension of the results). The function $G(t)$ denotes a non-decreasing weight function with a total variation of 1. We can therefore assume that

$$\mathrm{d}G(t) = g(t)\mathrm{d}t, -\infty \leq t \leq \infty,$$

where $g(t)$ is a density function of some distribution. Note that if we do not specify the boundaries of any of the integrals in this section or Appendix A, then we are referring to an integral over $\mathbb{R}$.

The ILSEs $\tilde{\theta}_n$ is defined as the value of θ that minimises the function $I_n(\theta)$. This estimate is a consistent and asymptotically normal estimator of the true parameter value θ_0: See Heathcote (1997). To clarify the derivations of these results we first need to identify all the underlying random sequences involved.

We can construct a sequence of random variables, $\tilde{\theta}_n$, by finding the minimum of $I_n(\theta_n)$ for each n. The true parameter value θ_0 is a constant value, and not random, but the value of $I_n(\theta_0)$ will differ if we draw different samples of size n. Therefore, $I_n(\theta_0)$ is also a sequence of random variables, as is $I_n(\tilde{\theta}_n)$.

In order to derive the asymptotic properties of $\tilde{\theta}_n$, the first and second derivatives of I_n with respect to θ need to be found and their convergence properties investigated. Now,

we can rewrite $\phi(t;\theta)$ and $\hat{\phi}_n(t)$ as

$$\phi(t;\theta) = u(t;\theta) + iv(t;\theta)$$

and

$$\hat{\phi}_n(t) = U_n(t) + iV_n(t).$$

The objective function can then be rewritten as

$$I_n(\theta) = \int \left\{ [U_n(t) - u(t;\theta)]^2 + [V_n(t) - v(t;\theta)]^2 \right\} dG(t),$$

and its derivative with respect to θ is

$$\begin{aligned} I_n'(\theta) &= -2 \int \left\{ [U_n(t) - u(t;\theta)]\, u'(t;\theta) + [V_n(t) - v(t;\theta)]\, v'(t;\theta) \right\} dG(t) \\ &= -\frac{2}{n} \sum \int \left\{ [\cos(tX_j) - u(t;\theta)]\, u'(t;\theta) \right. \\ &\qquad \left. + [\sin(tX_j) - v(t;\theta)]\, v'(t;\theta) \right\} dG(t). \end{aligned}$$

If we define the random variable $K_j(\theta)$, with mean zero, as

$$K_j(\theta) = \int \left\{ [\cos(tX_j) - u(t;\theta)]\, u'(t;\theta) + [\sin(tX_j) - v(t;\theta)]\, v'(t;\theta) \right\} dG(t)$$

then the derivative of I_n can be rewritten as a sum of random variables

$$I_n'(\theta) = -\frac{2}{n} \sum K_j(\theta).$$

Heathcote (1997) showed that $K_j(\theta)$ is bounded which ensures that the central limit theorem as well as the law of large numbers apply to the sequence of independent random variables $\{K_j(\theta)\}$, which are i.i.d. copies of a random variable $K(\theta)$.

Therefore, from the central limit theorem we can conclude that $I_n'(\theta) = -\frac{2}{n} \sum_{i=1}^{n} K_i(\theta)$ has the following asymptotic distribution at θ_0:

$$\sqrt{n} I_n'(\theta_0) \rightsquigarrow N\left[0, 4\mathrm{Var}\left\{K(\theta_0)\right\}\right], \tag{10.2}$$

where $\rightsquigarrow$ denotes convergence in distribution. The asymptotic variance of $K(\theta)$ is

$$\begin{aligned} \mathrm{Var}\left\{K(\theta)\right\} = \iint & \left[\mathrm{Cov}\left\{\cos(tX), \cos(sX)\right\} u'(t;\theta) u'(s;\theta) \right. \\ & + 2\mathrm{Cov}\left\{\cos(tX), \sin(sX)\right\} u'(t;\theta) v'(s;\theta) \\ & \left. + \mathrm{Cov}\left\{\sin(tX), \sin(sX)\right\} v'(t;\theta) v'(s;\theta) \right] dG(t) dG(s), \end{aligned} \tag{10.3}$$

and

$$\begin{aligned} \mathrm{Cov}\left\{\cos(tX), \cos(sX)\right\} &= \frac{1}{2}\left[u(t-s;\theta) + u(t+s;\theta) - 2u(t;\theta)u(s;\theta)\right], \\ \mathrm{Cov}\left\{\cos(tX), \sin(sX)\right\} &= \frac{1}{2}\left[v(t+s;\theta) - v(t-s;\theta) - 2u(t;\theta)v(s;\theta)\right], \\ \mathrm{Cov}\left\{\sin(tX), \sin(sX)\right\} &= \frac{1}{2}\left[u(t-s;\theta) - u(t+s;\theta) - 2v(t;\theta)v(s;\theta)\right]. \end{aligned}$$

We are now ready to state the proof summarised in Heathcote (1997) that a statistic $\tilde{\theta}_n$ exist which is a strongly consistent estimator of θ_0. For any number $\delta > 0$

$$\begin{aligned} & I_n(\theta_0 \pm \delta) - I_n(\theta_0) \\ & = \int \left\{ [u(t;\theta_0 \pm \delta) - u(t;\theta_0)] \, [u(t;\theta_0 \pm \delta) + u(t;\theta_0) - 2U_n(t)] \right\} \, dG(t) \\ & + \int \left\{ [v(t;\theta_0 \pm \delta) - v(t;\theta_0)] \, [v(t;\theta_0 \pm \delta) + v(t;\theta_0) - 2V_n(t)] \right\} \, dG(t). \end{aligned}$$

Now, according to Feuerverger and Mureika (1977) $U_n(t) \xrightarrow{\text{as}} u(t;\theta)$ and $V_n(t) \xrightarrow{\text{as}} v(t;\theta)$, where $\xrightarrow{\text{as}}$ denotes almost sure convergence. Therefore, by the dominated convergence theorem

$$\begin{aligned} I_n(\theta_0 \pm \delta) - I_n(\theta_0) \xrightarrow{\text{as}} \int \big[& (u(t;\theta_0 \pm \delta) - u(t;\theta_0))^2 \\ & + (v(t;\theta_0 \pm \delta) - v(t;\theta_0))^2 \big] \, dG(t) \geq 0, \end{aligned}$$

since $U_n(t), V_n(t), u(t;\theta)$ and $v(t;\theta)$ are all bounded functions.

Therefore, with probability one and for n large enough we have

$$I_n(\theta_0 \pm \delta) \geq I_n(\theta_0).$$

Equality only occurs when $u(t;\theta_0 \pm \delta) = u(t;\theta_0)$ and $v(t;\theta_0 \pm \delta) = v(t;\theta_0)$ for every t which implies $F(x;\theta_0 \pm \delta) = F(x;\theta_0)$ for all x, which is a contradiction, therefore with probability one $I_n(\theta_0 \pm \delta) > I_n(\theta_0)$ for n large enough.

Therefore, with probability one θ_0 is a strict global minimum point for $I_n(\theta)$ if n is large enough. Now, let $\tilde{\theta}_n = \theta_0 + \delta_n$ for some δ_n. For n large enough, δ_n will be equal to zero, as $\tilde{\theta}_n$ and θ_0 are both global minimum points for $I_n(\theta)$. Therefore, with probability one and for any $\epsilon > 0$, the following is true

$$|\tilde{\theta}_n - \theta_0| = \delta_n = 0 < \epsilon,$$

for n large enough. Therefore, $\tilde{\theta}_n \xrightarrow{\text{as}} \theta_0$.

Now, the second derivative of $I_n(\theta)$ is given by

$$\begin{aligned} I_n''(\theta) = 2 \int \Big\{ & [u'(t;\theta)]^2 + [v'(t;\theta)]^2 - \\ & \{U_n(t) - u(t;\theta)\} \, u''(t;\theta) - \{V_n(t) - v(t;\theta)\} \, v''(t;\theta) \} \, dG(t). \end{aligned}$$

To apply the dominated convergence theorem, we need to assume that $u''(t;\theta)$ and $v''(t;\theta)$ are both bounded, then for the same reason as above

$$I_n''(\theta_0) \xrightarrow{\text{as}} 2 \int \left([u'(t;\theta_0)]^2 + [v'(t;\theta_0)]^2 \right) \, dG(t),$$

which implies that

$$I_n''(\theta_0) \rightsquigarrow 2 \int \left([u'(t;\theta_0)]^2 + [v'(t;\theta_0)]^2 \right) dG(t).$$

Now, if we define

$$\lambda(\theta_0) = \int \left([u'(t;\theta_0)]^2 + [v'(t;\theta_0)]^2 \right) dG(t), \tag{10.4}$$

where Heathcote (1997) showed that

$$\lambda(\theta) = \text{Cov}\left\{ K(\theta), \frac{\partial}{\partial\theta} \log f(x,\theta) \right\}.$$

With a Taylor expansion we get

$$I_n'(\tilde{\theta}_n) = I_n'(\theta_0) + (\tilde{\theta}_n - \theta_0) I_n'' \left\{ \theta_0 + \epsilon(\tilde{\theta}_n - \theta_0) \right\},$$

for some $|\epsilon| < 1$.

Let us define

$$\hat{\theta}_n = \theta_0 + \epsilon(\tilde{\theta}_n - \theta_0).$$

We know that $\tilde{\theta}_n \xrightarrow{\text{as}} \theta_0$, which implies $\hat{\theta}_n \xrightarrow{\text{as}} \theta_0$. Therefore, by the continuous mapping theorem (Van der Vaart (1998) Theorem 2.3) we have $I_n''(\hat{\theta}_n) \xrightarrow{\text{as}} I_n''(\theta_0)$. This in turn implies that

$$I_n''(\hat{\theta}_n) \rightsquigarrow 2\lambda(\theta_0).$$

Now, by using Slutsky's theorem (Van der Vaart (1998) Lemma 2.8) and formula (10.2) on page 147 we can conclude that

$$\begin{aligned}
\sqrt{n}(\tilde{\theta}_n - \theta_0) &= -\frac{\sqrt{n} I_n'(\theta_0)}{I_n'' \left(\theta_0 + \epsilon(\tilde{\theta}_n - \theta_0) \right)} \\
&\rightsquigarrow \frac{1}{2\lambda(\theta_0)} N\left[0, 4\text{Var}\left\{K(\theta_0)\right\}\right] \\
&= N\left[0, \frac{\text{Var}\left\{K(\theta_0)\right\}}{\lambda^2(\theta_0)}\right].
\end{aligned}$$

In this section the results were obtained for a sequence of one dimensional random variables. Heathcote (1997) generalised the results for a sequence of multidimensional random vectors, which will be applied in the next section to get the asymptotic normality properties of a random vector, with the ILSEs as vector entries.

10.2 Relative efficiency of the stable ILSEs

The asymptotic variance will be estimated for a family of symmetric stable distributions with only one parameter, α, as well as the asymptotic covariance matrix for the general

case. For the single parameter case we will use the characteristic function defined in (8.2) on page 111 and the density function defined in (8.5) on page 116. The formulas for the general case can be obtained in (8.3) on page 111, (8.4) on page 116 and (8.19) to (8.22) on page 133.

The justification to calculate the relative efficiency of the single parameter case can be found in Heathcote (1997), where the author made an interesting observation in terms of the weight function used. According to Heathcote (1997) the efficiency can be improved by choosing a weight function which is not centered around zero.

To summarise the results from Heathcote (1997) for the single parameter case, we need to calculate the asymptotic variance in

$$\sqrt{n}(\tilde{\alpha}_n - \alpha) \rightsquigarrow N\left[0, \frac{\text{Var}\left\{K(\alpha)\right\}}{\lambda^2(\alpha)}\right],$$

where the asymptotic variance is calculated by (10.3) on page 147 and

$$\lambda(\alpha) = \int \left(\left[u'(t;\alpha)\right]^2 + \left[v'(t;\alpha)\right]^2 \right) dG(t).$$

For the special case we have

$$u(t;\alpha) = \exp\left(-|t|^\alpha\right), -\infty < t < \infty.$$

Therefore,

$$u'(t;\alpha) = \frac{\partial}{\partial\alpha} u(t;\alpha) = -e^{-|t|^\alpha} |t|^\alpha \log|t|.$$

The formula for $\text{Var}\{K(\alpha)\}$ can be simplified to

$$\text{Var}\{K(\alpha)\} = \iint \left[\text{Cov}\left\{\cos(tX), \cos(sX)\right\} u'(t;\alpha) u'(s;\alpha)\right] dG(t) dG(s), \tag{10.5}$$

where

$$\text{Cov}\left\{\cos(tX), \cos(sX)\right\} = \frac{1}{2}\left(e^{-|t-s|^\alpha} + e^{-|t+s|^\alpha} - 2e^{-|t|^\alpha} e^{-|s|^\alpha}\right).$$

The formula for $\lambda(\alpha)$ is also given by

$$\lambda(\alpha) = \int \left[u'(t;\alpha)\right]^2 dG(t) = \int e^{-2|t|^\alpha} |t|^{2\alpha} \left(\log|t|\right)^2 dG(t). \tag{10.6}$$

Therefore, to calculate the asymptotic variance of $\sqrt{n}(\tilde{\alpha}_n - \alpha)$, we need to calculate the integrals in formulas (10.5) and (10.6). These integrals will be approximated numerically, and the description of the methods and results are shown in Section 11.

Heathcote (1997) stated that for the general multi-parameter case, the ILSEs

$$\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} = \left[\tilde{\alpha}_n \quad \tilde{\beta}_n \quad \tilde{\sigma}_n \quad \tilde{\mu}_n\right]^T$$

is a consistent estimator of $\boldsymbol{\theta}$. If the first and second order partial derivatives of $u(t;\boldsymbol{\theta})$ and $v(t;\boldsymbol{\theta})$ with respect to α, β, σ and μ are bounded by functions that are G-integrable, where G is assumed to be a distribution function, then

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}) \rightsquigarrow N\left[0, \Lambda^{-1}\Sigma\Lambda^{-1}\right], \tag{10.7}$$

where Σ is the 4×4 covariance matrix of the random variables

$$K_{\theta_i}(\boldsymbol{\theta}) = \int \left[\{\cos(tX) - u(t;\boldsymbol{\theta})\} \frac{\partial}{\partial \theta_i} u(t;\boldsymbol{\theta}) + \{\sin(tX) - v(t;\boldsymbol{\theta})\} \frac{\partial}{\partial \theta_i} v(t;\boldsymbol{\theta}) \right] dG(t), \quad i = 1,2,3,4 \tag{10.8}$$

and Λ is the 4×4 symmetric matrix with entries

$$\Lambda_{\theta_i\theta_j}(\boldsymbol{\theta}) = \int \left[\frac{\partial}{\partial \theta_i} u(t;\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} u(t;\boldsymbol{\theta}) + \frac{\partial}{\partial \theta_i} v(t;\boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} v(t;\boldsymbol{\theta}) \right] dG(t), \quad i,j = 1,2,3,4. \tag{10.9}$$

Note that $\theta_1 = \alpha$, $\theta_2 = \beta$, $\theta_3 = \sigma$, $\theta_4 = \mu$.

Heathcote (1997) showed that the entries into the covariance matrix of the random variables K_{θ_i} defined in formula (10.8) can be calculated by the following formulas

$$\begin{aligned} \mathrm{Cov}(K_{\theta_i}, K_{\theta_j}) = \iint & [\mathrm{Cov}\{\cos(tX), \cos(sX)\} u_{\theta_i}(t;\boldsymbol{\theta}) u_{\theta_j}(s;\boldsymbol{\theta}) \\ & + \mathrm{Cov}\{\cos(tX), \sin(sX)\} u_{\theta_i}(t;\boldsymbol{\theta}) v_{\theta_j}(s;\boldsymbol{\theta}) \\ & + \mathrm{Cov}\{\sin(tX), \cos(sX)\} v_{\theta_i}(t;\boldsymbol{\theta}) u_{\theta_j}(s;\boldsymbol{\theta}) \\ & + \mathrm{Cov}\{\sin(tX), \sin(sX)\} v_{\theta_i}(t;\boldsymbol{\theta}) v_{\theta_j}(s;\boldsymbol{\theta})] dG(t) dG(s), \end{aligned} \tag{10.10}$$

with

$$\begin{aligned} \mathrm{Cov}\{\cos(tX), \cos(sX)\} &= \frac{1}{2}\left[u(t-s;\boldsymbol{\theta}) + u(t+s;\boldsymbol{\theta}) - 2u(t;\boldsymbol{\theta})u(s;\boldsymbol{\theta})\right], \\ \mathrm{Cov}\{\cos(tX), \sin(sX)\} &= \frac{1}{2}\left[v(t+s;\boldsymbol{\theta}) - v(t-s;\boldsymbol{\theta}) - 2u(t;\boldsymbol{\theta})v(s;\boldsymbol{\theta})\right], \\ \mathrm{Cov}\{\sin(tX), \cos(sX)\} &= \frac{1}{2}\left[v(t+s;\boldsymbol{\theta}) - v(s-t;\boldsymbol{\theta}) - 2v(t;\boldsymbol{\theta})u(s;\boldsymbol{\theta})\right], \\ \mathrm{Cov}\{\sin(tX), \sin(sX)\} &= \frac{1}{2}\left[u(t-s;\boldsymbol{\theta}) - u(t+s;\boldsymbol{\theta}) - 2v(t;\boldsymbol{\theta})v(s;\boldsymbol{\theta})\right]. \end{aligned}$$

The formulas for u, v and its partial derivatives were defined in (8.3) on page 111 and (9.2) on page 139.

Therefore, to calculate the asymptotic covariance matrix in (10.7), we need to calculate the entries into matrices Σ and Λ by calculating the double integrals in formula (10.10)

and the single integrals in (10.9). These integrals will be approximated numerically, and the description of the methods and results are also shown in Section 11.

The Fisher information is needed to calculate the asymptotic variance of $\sqrt{n}\,(\bar{\alpha}_n - \alpha)$, where $\bar{\alpha}_n$ is the MLE of α in the single parameter case.

The Fisher information is defined as

$$\mathscr{I}(\alpha) = E\left[\left(\frac{\partial}{\partial\alpha}\log f(X;\alpha)\right)^2 \middle| \alpha\right], \tag{10.11}$$

where the density function f is defined in formula (8.5) on page 116.

The Fisher information can also be calculated by

$$\mathscr{I}(\alpha) = E\left[\frac{\partial^2}{\partial\alpha^2}\log f(X;\alpha)\middle|\alpha\right] = E\left[\left(\frac{\frac{\partial}{\partial\alpha}f(X;\alpha)}{f(X;\alpha)}\right)^2 \middle|\alpha\right]. \tag{10.12}$$

These expectations will be approximated and the accuracy will be tested using various methods.

To calculate the relative efficiency of $\tilde{\alpha}_n$, compared to $\bar{\alpha}_n$, it should be noted that the asymptotic variance of $\bar{\alpha}_n$ is the minimum asymptotic variance of any unbiased estimator $\tilde{\alpha}_n$. The Cramer-Rao bound states that the variance of an unbiased estimator $\tilde{\alpha}_n$ is bounded by the reciprocal of the Fisher information $I(\alpha)$. Therefore,

$$\mathrm{Var}(\tilde{\alpha}_n) \geq \frac{1}{I(\alpha)}.$$

To evaluate the efficiency of a statistic, the correlation between the efficient score and $K(\alpha)$ is given by

$$\mathrm{e}(\tilde{\alpha}_n) = \frac{\lambda^2(\alpha)}{\mathscr{I}(\alpha)\mathrm{Var}\{K(\alpha)\}}. \tag{10.13}$$

Before we can discuss the relative efficiency of the more general case, asymptotic properties of MLEs for the stable parameters need to be looked at. It is a question whether or not the asymptotic normal properties of the MLEs exist for the more general case. The answer is positive and was provided by Dumouchel (1975). Therefore, if we define a vector of random MLEs

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{n}} = [\bar{\alpha}_n\ \bar{\beta}_n\ \bar{\sigma}_n\ \bar{\mu}_n],$$

which estimates the vector of parameters

$$\boldsymbol{\theta} = [\alpha\ \beta\ \sigma\ \mu],$$

then we know that

$$\bar{\boldsymbol{\theta}}_{\boldsymbol{n}} \xrightarrow{\text{as}} \boldsymbol{\theta},$$

and

$$\sqrt{n}\left(\bar{\boldsymbol{\theta}}_{\boldsymbol{n}} - \boldsymbol{\theta}\right) \rightsquigarrow N\left[0, \mathscr{I}^{-1}(\boldsymbol{\theta})\right],$$

where $\mathscr{I}(\boldsymbol{\theta})$ is the Fisher information matrix for the family of stable distributions.

The elements of the Fisher information matrix are defined by

$$\mathscr{I}(\boldsymbol{\theta})_{i,j} = E\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\boldsymbol{\theta})\right)\middle|\boldsymbol{\theta}\right].$$

Under certain regularity conditions, the entries may also be written as

$$\mathscr{I}(\boldsymbol{\theta})_{i,j} = -E\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\log f(X;\boldsymbol{\theta})\middle|\boldsymbol{\theta}\right].$$

The ILSEs, $\tilde{\boldsymbol{\theta}}_n$, are unbiased estimators of $\boldsymbol{\theta}$, and the Cramer-Rao bound for multiple parameters is given by

$$\Lambda^{-1}\Sigma\Lambda^{-1} \geq \mathscr{I}(\boldsymbol{\theta})^{-1},$$

which means that $\Lambda^{-1}\Sigma\Lambda^{-1} - \mathscr{I}(\boldsymbol{\theta})^{-1}$ is positive semidefinite. To calculate the relative efficiency of the ILSEs, the diagonal elements of $\Lambda^{-1}\Sigma\Lambda^{-1}$, will be compared to the diagonal elements of $\mathscr{I}(\boldsymbol{\theta})^{-1}$. Therefore, the measures we will use to evaluate the relative efficiency of $\tilde{\alpha}_n$, $\tilde{\beta}_n$, $\tilde{\sigma}_n$ and $\tilde{\mu}_n$ are given by

$$\mathrm{e}(\tilde{\alpha}_n) = \frac{\left(\mathscr{I}(\boldsymbol{\theta})^{-1}\right)_{1,1}}{\left(\Lambda^{-1}\Sigma\Lambda^{-1}\right)_{1,1}}, \tag{10.14}$$

$$\mathrm{e}(\tilde{\beta}_n) = \frac{\left(\mathscr{I}(\boldsymbol{\theta})^{-1}\right)_{2,2}}{\left(\Lambda^{-1}\Sigma\Lambda^{-1}\right)_{2,2}}, \tag{10.15}$$

$$\mathrm{e}(\tilde{\sigma}_n) = \frac{\left(\mathscr{I}(\boldsymbol{\theta})^{-1}\right)_{3,3}}{\left(\Lambda^{-1}\Sigma\Lambda^{-1}\right)_{3,3}}, \tag{10.16}$$

and

$$\mathrm{e}(\tilde{\mu}_n) = \frac{\left(\mathscr{I}(\boldsymbol{\theta})^{-1}\right)_{4,4}}{\left(\Lambda^{-1}\Sigma\Lambda^{-1}\right)_{4,4}}. \tag{10.17}$$

Chapter 11

Relative efficiency: Numerical results

In this chapter our goal is to approximate the relative efficiency of the ILSEs compared to the most efficient MLEs. In Chapter 10 we laid the foundation to calculate the asymptotic variance of both the MLEs and ILSEs. We will approximate the Fisher information matrix in Section 11.1, in order to approximate the asymptotic variance of the MLEs. In Section 11.2 we approximate the asymptotic variance of the ILSEs. Combining these results we can approximate the relative efficiency of the ILSEs, compared to the MLEs (Section 11.3). An interesting result is that the relative efficiency can be improved by choosing various weight functions in (10.1) on page 146.

11.1 Fisher Information: Numerical results

11.1.1 Existing tables

Entries into the covariance matrix $\mathscr{I}(\boldsymbol{\theta})^{-1}$ for various combinations of the α and β parameters can be derived from a table in (Dumouchel, 1975), which is shown in Table 11.1 below.

The entries into the table are the correlation between MLEs $\bar{\alpha}_n$, $\bar{\beta}_n$, $\bar{\sigma}_n$ and $\bar{\mu}_n$ as well as the standard deviations of these estimators. Note that the standard deviation of $\bar{\sigma}_n$ and the correlation between $\bar{\sigma}_n$ and any of the other MLEs is given the notation σ_{σ_1} and $\rho_{\cdot\sigma_1}$ to avoid confusion. To convert Table 11.1 to entries of a 4×4 covariance matrix for each combination of the estimators, we use the formula for covariance

$$\text{Cov}_{\theta_i,\theta_j} = \rho_{\theta_i,\theta_j}\sigma_{\theta_i}\sigma_{\theta_j},$$

where ρ_{θ_i,θ_j} is the correlation between parameters estimates θ_i and θ_j and σ_{θ_i} is the

Table 11.1: Asymptotic standard deviations and correlation coefficients from Dumouchel (1975). Missing values are denoted by $*$.

α	β	σ_α	σ_β	σ_{σ_1}	σ_μ	$\rho_{\alpha\beta}$	$\rho_{\alpha\sigma_1}$	$\rho_{\alpha\mu}$	$\rho_{\beta\sigma_1}$	$\rho_{\beta\mu}$	$\rho_{\sigma_1\mu}$
2.00	0.0	0.000	∞	0.707	1.410	*	*	*	*	*	0.000
1.99	0.0	0.478	43.100	0.738	1.430	0.000	0.220	0.000	0.000	0.130	0.000
1.95	0.0	0.885	14.600	0.791	1.490	0.000	0.310	0.000	0.000	0.270	0.000
1.95	0.5	0.861	13.700	0.786	1.490	0.210	0.290	-0.110	0.110	0.210	0.030
1.95	1.0	0.724	0.000	0.763	1.500	*	0.180	-0.270	*	*	0.050
1.90	0.0	1.130	9.060	0.835	1.570	0.000	0.330	0.000	0.000	0.380	0.000
1.90	0.5	1.100	8.480	0.827	1.570	0.220	0.310	-0.160	0.130	0.300	0.060
1.90	1.0	0.908	0.000	0.783	1.590	*	0.170	-0.380	*	*	0.120
1.70	0.0	1.500	4.320	0.965	1.970	0.000	0.340	0.000	0.000	0.650	0.000
1.70	0.5	1.460	3.910	0.947	2.040	0.170	0.320	-0.320	0.120	0.500	0.130
1.70	1.0	1.210	0.000	0.844	2.200	*	0.110	-0.650	*	*	0.310
1.50	0.0	1.540	3.040	1.080	2.670	0.000	0.320	0.000	0.000	0.820	0.000
1.50	0.5	1.500	2.690	1.050	2.980	0.120	0.300	-0.500	0.090	0.590	0.140
1.50	1.0	1.250	0.000	0.901	3.670	*	0.060	-0.820	*	*	0.370
1.30	0.0	1.420	2.360	1.210	4.230	0.000	0.290	0.000	0.000	0.930	0.000
1.30	0.5	1.380	2.070	1.160	5.870	0.090	0.250	-0.720	0.050	0.530	0.080
1.30	1.0	1.120	0.000	0.971	8.460	*	-0.030	-0.930	*	*	0.350
1.10	0.0	1.220	1.880	1.360	11.500	0.000	0.240	0.000	0.000	0.990	0.000
1.10	0.5	0.488	1.690	1.410	18.600	0.230	-0.430	-0.810	-0.110	0.340	0.560
1.10	1.0	0.074	0.000	1.360	12.800	*	-0.610	-0.780	*	*	0.970
1.00	0.0	1.100	1.730	1.450	1.450	0.000	0.210	0.000	0.000	-0.210	0.000
0.80	0.0	0.856	1.440	1.690	4.560	0.000	0.130	0.000	0.000	-0.960	0.000
0.80	0.5	0.853	1.270	1.620	8.450	0.060	0.140	-0.860	-0.050	-0.490	-0.300
0.80	1.0	0.268	0.000	1.370	4.140	*	-0.320	-0.750	*	*	-0.350

standard deviation of the estimate θ_i.

The entries from Dumouchel (1975) for $\alpha = 1.1$ and $\beta = 0.5$ lead to a covariance matrix with a determinant which is negative. This cannot be true as a covariance matrix is positive semi-definite. This is probably due to a small computational error, and will be excluded from our analyses. We will also exclude the cases where $\beta = 1$, $\alpha = 2$ and $\alpha = 0.8$ as they fall either outside the scope of our study or incomplete entries are observed in the table.

These results from Dumouchel (1975) will be used for comparison to our results for the Fisher information. Table 11.2 was constructed by inverting the covariance matrices obtained from Table 11.1.

Table 11.2: Fisher information matrix entries derived from Table 11.1 on page 155

α	β	$\mathscr{I}_\alpha$	$\mathscr{I}_\beta$	$\mathscr{I}_\sigma$	$\mathscr{I}_\mu$	$\mathscr{I}_{\alpha\beta}$	$\mathscr{I}_{\alpha\sigma}$	$\mathscr{I}_{\alpha\mu}$	$\mathscr{I}_{\beta\sigma}$	$\mathscr{I}_{\beta\mu}$	$\mathscr{I}_{\sigma\mu}$
1.99	0.0	4.599	0.001	1.929	0.497	0.000	-0.655	0.000	0.000	-0.002	0.000
1.95	0.0	1.413	0.005	1.768	0.486	0.000	-0.490	0.000	0.000	-0.013	0.000
1.95	0.5	1.572	0.006	1.777	0.485	-0.021	-0.467	0.148	-0.004	-0.013	-0.050
1.90	0.0	0.879	0.014	1.610	0.474	0.000	-0.392	0.000	0.000	-0.031	0.000
1.90	0.5	1.018	0.017	1.641	0.479	-0.034	-0.395	0.182	-0.005	-0.031	-0.089
1.70	0.0	0.503	0.093	1.214	0.446	0.000	-0.266	0.000	0.000	-0.132	0.000
1.70	0.5	0.737	0.105	1.347	0.446	-0.114	-0.392	0.302	0.040	-0.129	-0.210
1.50	0.0	0.470	0.330	0.955	0.428	0.000	-0.214	0.000	0.000	-0.308	0.000
1.50	0.5	1.271	0.370	1.285	0.450	-0.448	-0.677	0.592	0.234	-0.321	-0.359
1.30	0.0	0.541	1.329	0.746	0.414	0.000	-0.184	0.000	0.000	-0.690	0.000
1.30	0.5	6.552	1.565	1.608	0.471	-2.851	-2.374	1.679	1.029	-0.791	-0.620
1.10	0.0	0.713	14.218	0.574	0.380	0.000	-0.153	0.000	0.000	-2.301	0.000
1.00	0.0	0.865	0.350	0.498	0.498	0.000	-0.138	0.000	0.000	0.088	0.000

11.1.2 Single parameter Fisher information

The Fisher information needs to be calculated in order to calculate the relative efficiency of the ILSEs. The Fisher information can be calculated by either the conditional expectation defined in formula (10.11) or formula (10.12) on page 152.

The Fisher information can be approximated using the law of large numbers by simulating a large number of stably distributed random variables $(X_1, X_2, ..., X_n)$ and calculating the following sum

$$\mathscr{I}(\alpha) \approx \frac{1}{n} \sum_{i=1}^{n} \left(\frac{\frac{\partial}{\partial \alpha} f(X_i; \alpha)}{f(X_i; \alpha)} \right)^2. \tag{11.1}$$

Alternatively, the Fisher information can be computed by calculating the following integral:

$$\mathscr{I}(\alpha) = \int_{-\infty}^{\infty} \frac{\left[\frac{\partial}{\partial \alpha} f(x; \alpha)\right]^2}{f(x; \alpha)} \, \mathrm{d}x. \tag{11.2}$$

Dumouchel (1975) uses (11.2) to approximate the Fisher information. He mentions that some of his results may underestimate the value of the Fisher information. We will investigate whether this is the case and whether we can find a more accurate approximation of the Fisher information.

The Fisher information can be approximated by

$$\mathscr{I}(\alpha) \approx \int_{-c}^{c} \frac{\left[\frac{\partial}{\partial \alpha} f(x; \alpha)\right]^2}{f(x; \alpha)} \, \mathrm{d}x, \tag{11.3}$$

for c large enough.

The integral on the right hand side is estimated by a midpoint Riemann-sum approximation:

$$\int_{-c}^{c} \frac{\left[\frac{\partial}{\partial\alpha} f(x;\alpha)\right]^2}{f(x;\alpha)}\,\mathrm{d}x \approx R_{0.5} \int_{-c}^{c} \frac{\left[\frac{\partial}{\partial\alpha} f(x;\alpha)\right]^2}{f(x;\alpha)}\,\mathrm{d}x.$$

We use this numerical integration method as the integrand is smooth. The parameter values for Δx are the same as in preceding sections, where we interpolate linearly between $(\alpha, \Delta x) = (1; 0, 1)$ and $(\alpha, \Delta x) = (2; 0, 01)$.

Figure 11.1 shows which value of c needs to be chosen for $\alpha = 1$ and $\alpha = 1.99$. A sample of estimates for the Fisher information was approximated using the simulation method formulated in (11.1) where $n = 50,000$ for each sample. The 10% and 90% sample quantiles were calculated and are shown in the figure. The estimates from Dumouchel (1975) are calculated by using the value of σ_α in Table 11.1 on page 155 in

$$\mathscr{I}(\alpha) = \frac{1}{\sigma_\alpha^2},$$

as $\mathscr{I}(\alpha)$ is the reciprocal of the asymptotic variance in the single parameter case.

From Figure 11.1 we can conclude that c needs to be larger for smaller values of α, as the tails of the stable distribution are thicker for values of α closer to 1. We will use the value $c = 1750$ when $\alpha = 1$ and $c = 24$ when $\alpha = 1.99$. We find small numerical errors in the final results when linear interpolation is used to find the value of c. After some investigation we found that an exponential graph will be a better fit for the relationship between c and α. Therefore, to find values of c for $\alpha \in [1, 2]$ we use the following formula

$$c = ae^{-b(\alpha-1)},\ a = 1750,\ b \approx 4.33. \tag{11.4}$$

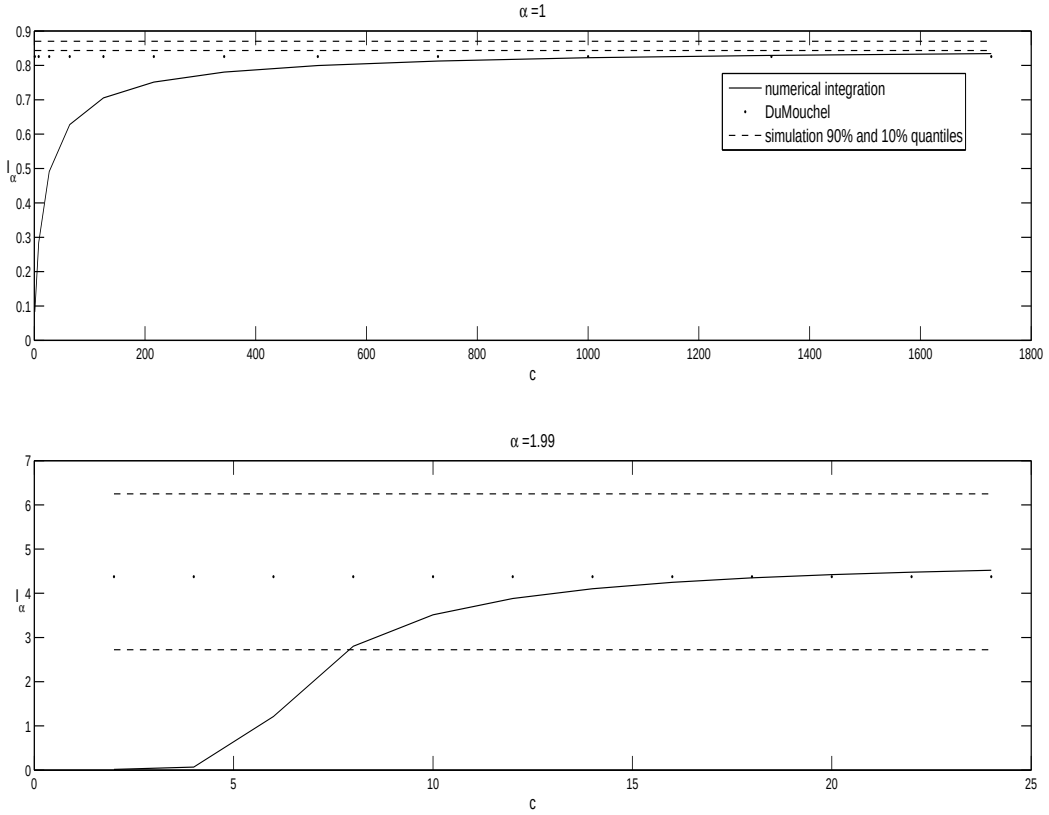

Figure 11.1: This graph shows the necessary width of the integration interval, in order to approximate the Fisher information accurately with a numerical integration method.

Figure 11.2 shows the approximated values for the Fisher information from the numerical integration method and the simulation method for various values of α. The results from Dumouchel (1975) are also shown for comparison. These results agree with the fact that the Fisher information approaches ∞ as α approaches the parameter value $\alpha = 2$ (Dumouchel, 1983).

We simulate $n = 50,000$ random variables in each sample of size 20. The mean as well as the 10% and 90% sample quantiles are shown in Figure 11.2. The mean is greater than the 90% sample quantile for $\alpha = 1.99$. This shows the possible occurrence of outlying values when using simulations to approximate the Fisher information.

The numerical integration method underestimates the value of the Fisher information. If c in (11.3) is large enough the actual error in the approximation becomes negligible. However, for α close to 1, it is difficult to choose c large enough given computational time constraints. The simulation method does not have similar problems.

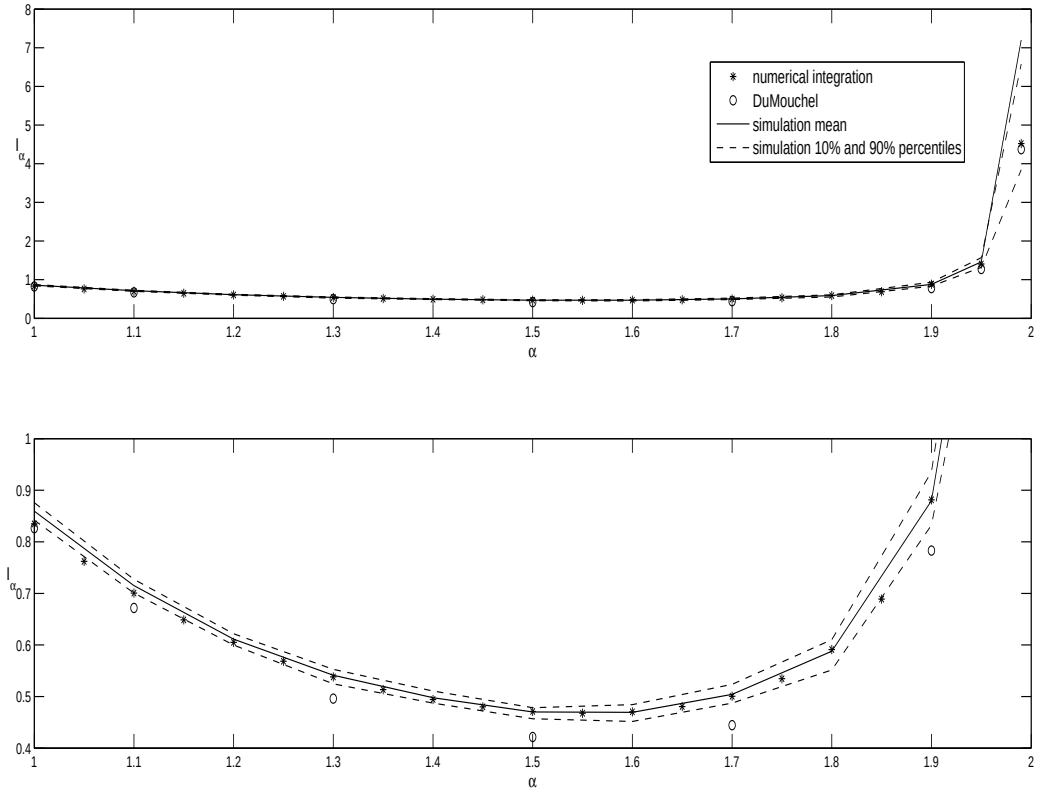

Figure 11.2: The single parameter Fisher information as a function of the parameter α. The two graphs are the same, with only the scale of the vertical axis changing from the top graph to the bottom graph.

Possible numerical errors may occur when approximating the Fisher information. This is more probable to occur using the simulation method. The value of $f(x;\alpha)$ approaches 0 as $x \to \infty$. Therefore, if extreme values of X_i are simulated, then to approximate the Fisher information we divide by a value very close to zero. In our simulations of $1,000,000$ (20 samples times $50,000$ variates) for $\alpha = 1$ an extreme value of $X = -1,527,577$ was simulated. The density value at X is close to $1e-13$. To approximate the Fisher information by (11.3) on page 156 the most extreme values where density function values are calculated are $\pm c$. For example, when $\alpha = 1$ then $c = 1750$ and $f(1750;\alpha = 1) \approx 1e-07$. The risk for numerical errors emanating from division by zero, is therefore manageable when using the numerical integration method, but not when the simulation method is used.

In terms of computational time, the Fisher information, approximated from simulations we ran, were less time consuming than the numerical integration method for values of α

close to 1, but more time consuming for values of α close to 2.

In conclusion, both methods have their advantages and disadvantages. We have decided to use numerical integration methods to approximate the Fisher information in the following sections. However, simulations will be used to test our result where necessary.

11.1.3 Fisher information matrix

The numerical integration method described for the single parameter case in the previous section will be slightly altered to approximate the Fisher information matrix for the general stable distribution. The Fisher information can be calculated by the following integral (In some figures and tables $\mathscr{I}_{\theta_i}$ will be used for $\mathscr{I}_{\theta_i\theta_j}$ if $i = j$ i.e. the diagonal elements of the Fisher information matrix.)

$$\mathscr{I}_{\theta_i\theta_j}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \frac{\frac{\partial}{\partial\theta_i} f(x;\boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} f(x;\boldsymbol{\theta})}{f(x;\boldsymbol{\theta})} \,\mathrm{d}x, \quad i,j = 1,2,3,4, \tag{11.5}$$

in which

$$\theta_1 = \alpha\,, \theta_2 = \beta\,, \theta_3 = \sigma\,, \theta_4 = \mu.$$

The Fisher estimate $\mathscr{I}_{\theta_i\theta_j}(\boldsymbol{\theta})$ will be estimated by a midpoint Riemann-sum. The density function and its partial derivatives are defined in formula (8.4) on page 116 and formulas (8.19) to (8.22) on page 133.

With the numerical integration method, using midpoint Riemann-sums we approximate the Fisher information by

$$\mathscr{I}_{\theta_i\theta_j}(\boldsymbol{\theta}) = R_{0.5} \int_{-\sigma c+\mu}^{\sigma c+\mu} \frac{\frac{\partial}{\partial\theta_i} f(x;\boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} f(x;\boldsymbol{\theta})}{f(x;\boldsymbol{\theta})} \,\mathrm{d}x, \quad i,j = 1,2,3,4. \tag{11.6}$$

The values of c are the same as in the single parameter case, while Δx is equal to σ times the values of Δx used in the single parameter case.

The results from (11.6) will be tested, by comparing them to the results shown in Table 11.2 on page 156 from Dumouchel (1975) and a simulation method.

The Fisher information for the general case can be approximated by

$$\mathscr{I}_{\theta_i\theta_j}(\boldsymbol{\theta}) \approx \frac{1}{n} \sum_{k=1}^{n} \frac{\frac{\partial}{\partial\theta_i} f(X_k;\boldsymbol{\theta}) \frac{\partial}{\partial\theta_j} f(X_k;\boldsymbol{\theta})}{\left[f(X_k;\boldsymbol{\theta})\right]^2}, \quad i,j = 1,2,3,4,$$

where $(X_1, X_2, ..., X_n)$ are simulated stably distributed random variables.

The Fisher information, approximated by the numerical integration method in (11.6), is shown in Table 11.3 and should be compared to the results in Table 11.2 on page 156.

Table 11.3: Fisher information matrix entries calculated for the general stable distribution

α	β	$\mathscr{I}_\alpha$	$\mathscr{I}_\beta$	$\mathscr{I}_\sigma$	$\mathscr{I}_\mu$	$\mathscr{I}_{\alpha\beta}$	$\mathscr{I}_{\alpha\sigma}$	$\mathscr{I}_{\alpha\mu}$	$\mathscr{I}_{\beta\sigma}$	$\mathscr{I}_{\beta\mu}$	$\mathscr{I}_{\sigma\mu}$
1.99	0.0	4.684	0.001	1.946	0.496	0.001	-0.724	0.003	-0.000	-0.002	-0.001
1.95	0.5	1.542	0.006	1.781	0.485	-0.022	-0.468	0.153	-0.004	-0.013	-0.051
1.95	0.0	1.374	0.005	1.765	0.484	0.000	-0.480	-0.000	0.000	-0.013	-0.000
1.90	0.5	1.004	0.016	1.647	0.475	-0.034	-0.398	0.181	-0.004	-0.030	-0.089
1.90	0.0	0.866	0.014	1.614	0.473	0.000	-0.394	-0.000	-0.000	-0.031	-0.000
1.70	0.5	0.723	0.106	1.354	0.451	-0.114	-0.389	0.302	0.040	-0.131	-0.212
1.70	0.0	0.495	0.092	1.220	0.442	0.000	-0.267	-0.000	-0.000	-0.131	-0.000
1.50	0.5	1.238	0.366	1.270	0.443	-0.439	-0.654	0.577	0.228	-0.316	-0.347
1.50	0.0	0.467	0.331	0.957	0.428	0.000	-0.215	-0.000	-0.000	-0.309	0.000
1.30	0.5	6.333	1.498	1.584	0.453	-2.730	-2.307	1.619	0.988	-0.756	-0.599
1.30	0.0	0.534	1.387	0.751	0.431	0.000	-0.181	-0.000	-0.000	-0.722	0.000
1.10	0.0	0.698	17.415	0.578	0.463	-0.000	-0.149	0.000	-0.000	-2.818	0.000
1.00	0.0	0.836	0.349	0.500	0.500	0.000	-0.131	0.000	-0.000	0.085	-0.000

If a number n_1 of independent sets of stably distributed random variables are simulated, each of size n, then a sample of Fisher information matrices of size n_1 can be produced. From these samples, bands are formed by the 10% and 90% percentiles. We will use these bands to investigate whether the results in Table 11.2 on page 156 and Table 11.3 on page 161 are accurate. We use a sample size of $n_1 = 20$ and the number of simulations $n = 10000$ for each sample. There are small differences in Table 11.2 and Table 11.3. Figure 11.3 shows a couple of cases where the results from Table 11.2 do not fall within the 10% and 90% simulation percentiles, where as the results from Table 11.3 do.

In a number of cases, the Fisher information approximations from the simulations are not a good measure of accuracy as they are influenced by outliers. Because of this, we find that the 10% and 90% simulation percentiles, in some cases, are not helpful to analyse the results from the numerical integration method. We therefore use the 20% and 80% simulation percentiles in these cases. Figure 11.4 shows an example where the approximations from Table 11.2 do not fall within the 20% and 80% simulation percentiles, whereas the results from Table 11.3 do.

The numerical integration method seems to provide accurate approximations for the Fisher information and will be used in the following sections, where the relative efficiency of ILSEs are approximated.

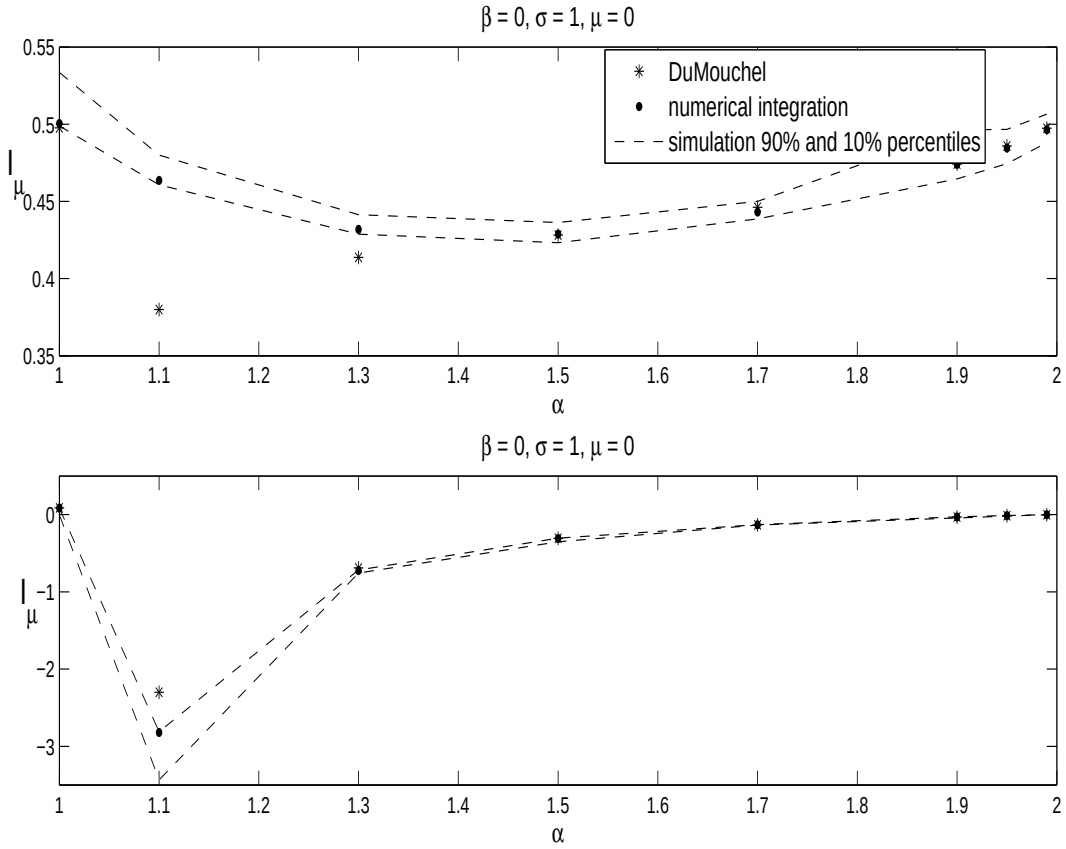

Figure 11.3: These two graphs show two entries of the Fisher information matrix ($\mathscr{I}_\mu$ and $\mathscr{I}_{\beta\mu}$), as functions of α. The results from Table 11.2 on page 156 seems to be inaccurate for a couple of cases.

11.2 Asymptotic normality of the ILSEs: Numerical results

11.2.1 Asymptotic variance

In this sub-section the focus will be on approximating the asymptotic variance

$$\sigma^2(\alpha; G) = \frac{\text{Var}\{K(\alpha)\}}{\lambda^2(\alpha)}, \tag{11.7}$$

in order to approximate the relative efficiency of the ILSEs defined in formula (10.13) on page 152. To approximate $\text{Var}\{K(\alpha)\}$ and $\lambda(\alpha)$, the integrals in formulas (10.5) and (10.6) on page 150 need to be approximated:

$$\text{Var}\{K(\alpha)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\text{Cov}\left\{\cos(tX), \cos(sX)\right\} u'(t;\alpha)u'(s;\alpha)]dG(t)dG(s),$$

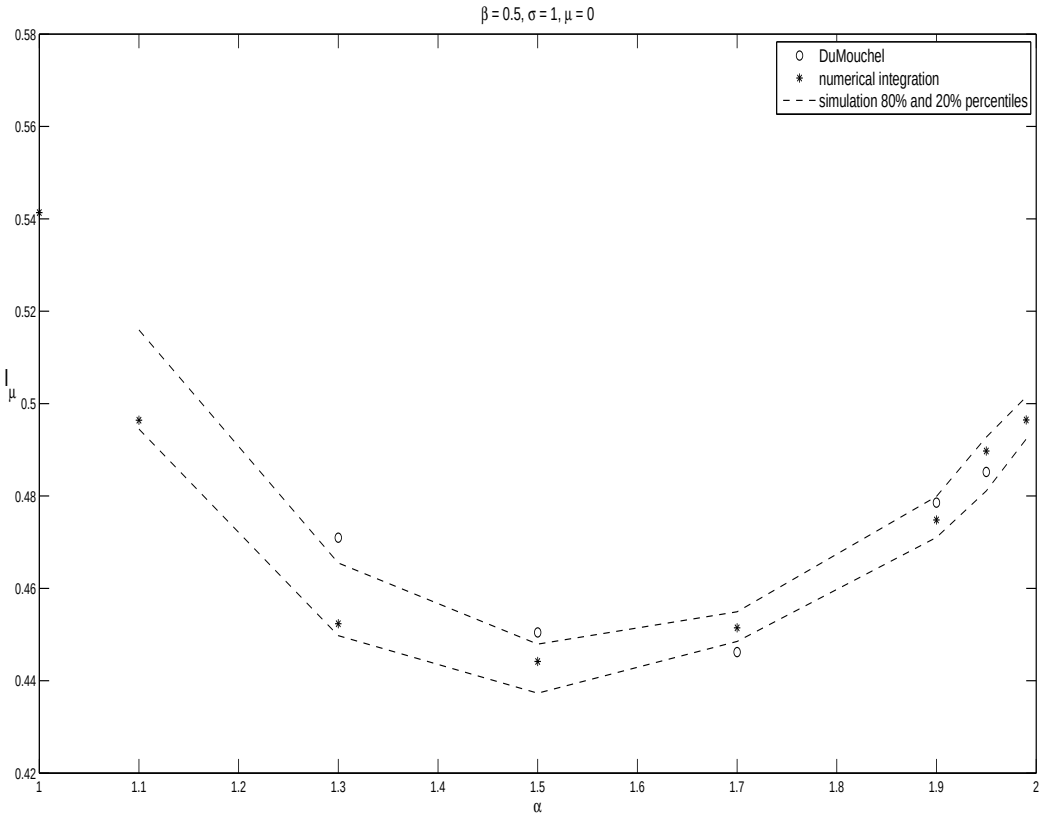

Figure 11.4: This graph is similar to the graphs in Figure 11.3, except that 20% and 80% simulation percentiles were used.

where

$$\mathrm{Cov}\left\{\cos(tX), \cos(sX)\right\} = \frac{1}{2}\left(e^{-|t-s|^\alpha} + e^{-|t+s|^\alpha} - 2e^{-|t|^\alpha}e^{-|s|^\alpha}\right)$$

and

$$u'(t;\alpha) = \frac{\partial}{\partial\alpha}u(t;\alpha) = -e^{-|t|^\alpha}|t|^\alpha \log|t|.$$

$$\lambda(\alpha) = \int \left[u'(t;\alpha)\right]^2 \, dG(t) = \int_{-\infty}^{\infty} e^{-2|t|^\alpha}|t|^{2\alpha}\left(\log|t|\right)^2 \, dG(t).$$

To approximate the integrals accurately, we need a value of c so large that

$$\mathrm{Var}\{K(\alpha)\} \approx \int_{-c}^{c}\int_{-c}^{c}[\mathrm{Cov}\left\{\cos(tX), \cos(sX)\right\} u'(t;\alpha)u'(s;\alpha)]dG(t)dG(s),$$

and

$$\lambda(\alpha) \approx \int_{-c}^{c} e^{-2|t|^\alpha}|t|^{2\alpha}\left(\log|t|\right)^2 \, dG(t),$$

are good approximations. The value of c is calculated by taking the inverse of G, i.e. $c = G^{-1}(1-\epsilon)$ for a small positive number ϵ.

Let us define the weight function g as the density function which corresponds to the distribution function G. Paulson et al. (1975) used the weight function $g(t) = e^{-t^2}$ which is equal to $\sqrt{\pi}$ times the normal density with mean zero and standard deviation $\frac{1}{\sqrt{2}}$. Heathcote (1997) remarked that such a weight function, which puts a lot of weight on the origin, will not lead to high efficiency. The remark will be tested in this section by investigating whether a weight function exists which will lead to a lower asymptotic variance for all values of $\alpha \in [1, 2]$.

More specifically, our objective is to find a density (weight) function which improves the efficiency for all values of $\alpha \in [1, 2]$. To construct Figure 11.5 various density functions were chosen to compare the asymptotic variance, σ^2 (11.7), and relative efficiency, e (10.13). The efficiency of the method can be improved by moving weight away from zero.

Let $g(t, \mu_w, \sigma_w)$ be a normal density function, with mean μ_w and standard deviation σ_w. A symmetric (around zero) density function with density centred at $\pm\mu_w$ can be constructed by

$$g_1(t, \mu_w, \sigma_w) = \frac{1}{2}\left[g(t, -\mu_w, \sigma_w) + g(t, \mu_w, \sigma_w)\right]. \tag{11.8}$$

Figure 11.5 shows the results if we use g_1 as a weight function, with $\mu_w = 0.7$ and $\sigma_w = 1/\sqrt{2}$. An improvement in the relative efficiency corresponds with the remark from Heathcote (1997). However, a contradiction is obtained if we choose $\mu_w = 0$ and $\sigma_w = 0.1$. In this case the relative efficiency for most values of α increases from $\pm 50\%$ to $\pm 75\%$. This is a drastic improvement in the efficiency of the ILSEs.

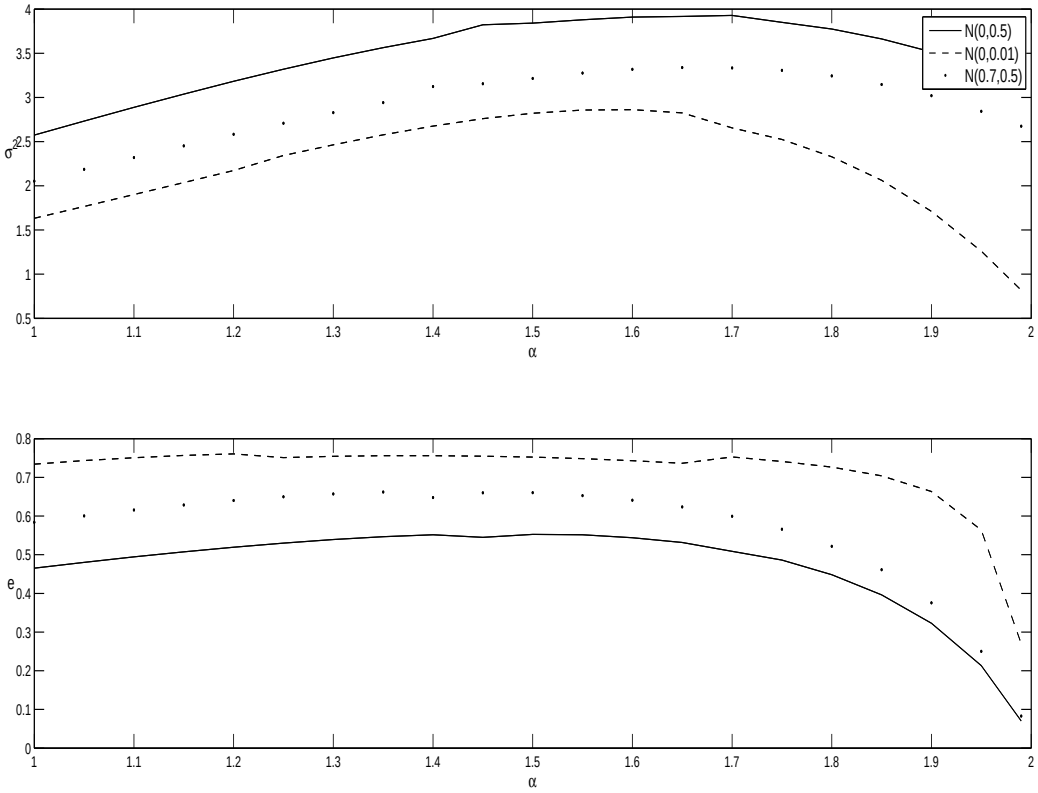

Figure 11.5: Asymptotic variance and relative efficiency using various normally distributed density functions in (11.8).

11.2.2 Asymptotic covariance matrix

11.2.3 The covariance of $K(\boldsymbol{\theta})$ (The Σ matrix).

In this section the numerical method to approximate the double integrals in formula (10.10) on page 151 will be described.

Our focus will now be on calculating one of these double integrals, and then expanding our results to other similar double integrals. Let us start with the integral

$$F(\boldsymbol{\theta}) = \iint_{\mathbb{R}^2} f(t, s, \boldsymbol{\theta}) dG(t) dG(s)$$

with

$$f(t, s, \boldsymbol{\theta}) = u(t - s; \boldsymbol{\theta}) u_\alpha(t; \boldsymbol{\theta}) u_\beta(s; \boldsymbol{\theta}).$$

The functions u, v and their partial derivatives, were defined in formula (8.3) on page 111 and formula (9.2) on page 139.

We will approximate $F(\boldsymbol{\theta})$, by

$$F(\boldsymbol{\theta}) \approx \iint_R f(t,s,\theta)\mathrm{d}G(t)\mathrm{d}G(s), \tag{11.9}$$

where R is a rectangular region

$$R = [-c,c] \times [-c,c].$$

The question is whether the function $|f(t,s;\boldsymbol{\theta})g(t)g(s)|$ is negligibly small on $\mathbb{R}^2 \setminus R$?

Now,

$$\lim_{t\to\infty} u_\alpha(t;\boldsymbol{\theta}) = 0,$$
$$\lim_{s\to\infty} u_\beta(s;\boldsymbol{\theta}) = 0,$$

and

$$\lim_{|t-s|\to\infty} u(t-s;\boldsymbol{\theta}) = 0,$$

except where $t = s$.

On this line in $\mathbb{R}^2$, the function f becomes

$$f(s,s) = u_\alpha(s;\boldsymbol{\theta})u_\beta(s;\boldsymbol{\theta}),$$

which approaches 0 as s tends to ∞. These results can be extended to the case where t, s tends to $-\infty$. Therefore, the function f approaches zero as t, s tends to $\pm\infty$. Therefore,

$$\lim_{t,s\to\pm\infty} f(t,s;\boldsymbol{\theta})g(t)g(s) = 0.$$

We can therefore assume the approximation in (11.9) is accurate by having control over the value of the double integral

$$\iint_{\mathbb{R}^2\setminus R} g(t)g(s)\mathrm{d}t\mathrm{d}s,$$

through our choice of c.

We now have to decide on a numerical integration method to use to approximate the various double integrals over the region R.. With Matlab's *"integral2"* method, the maximum allowed number of evaluations are reached in some cases. The solution to this problem is to divide the function f into disjoint regions where f is differentiable. Some of the functions are not differentiable on the lines $t = s$ and $t = -s$ as well as on various axes. We therefore divide the region R into the union of triangular regions

$$\begin{aligned} R = &\{0 \le s \le c\,, 0 \le t \le s\} \cup \{0 \le s \le c\,, s \le t \le c\} \\ &\cup \{-c \le s \le 0\,, -s \le t \le c\} \cup \{-c \le s \le 0\,, 0 \le t \le -s\} \\ &\cup \{-c \le s \le 0\,, s \le t \le 0\} \cup \{-c \le s \le 0\,, -c \le t \le s\} \\ &\cup \{0 \le s \le c\,, -c \le t \le -s\} \cup \{0 \le s \le c\,, -s \le t \le 0\}\,. \end{aligned}$$

11.2.4 Approximating Σ by Monte Carlo simulation

With the assumption that G is a normal distribution function, it enables us to simulate random variables K_{θ_i} from formula (10.8) on page 151 and calculate the covariance matrix of these simulated random variables. This method will be used to test our results obtained from the numerical integration method.

Let

$$\Psi_i(t, \boldsymbol{\theta}, X) = \{\cos(tX) - u(t; \boldsymbol{\theta})\} \frac{\partial}{\partial \theta_i} u(t; \boldsymbol{\theta}) + \{\sin(tX) - v(t; \boldsymbol{\theta})\} \frac{\partial}{\partial \theta_i} v(t; \boldsymbol{\theta}).$$

Then, from formula (10.8) on page 151, we have

$$\begin{aligned} K_{\theta_i}(\boldsymbol{\theta}) &= \int [\Psi_i(t, \boldsymbol{\theta}, X)] \, dG(t), \\ &= E_G [\Psi_i(t, \boldsymbol{\theta}, X)], \quad i = 1, 2, 3, 4, \end{aligned}$$

where the expectation is taken under the distribution G.

Let $(X_1, X_2, ..., X_n)$ be a set of stably distributed random variables and define for each $i = 1, 2, 3, 4$ a set of normally distributed random variables $(T_{i1}, T_{i2}, ..., T_{im})$, then for each $X_k, \quad k = 1, ..., n$ we can estimate the random variable $K_{\theta_i}^{(k)}$ by the law of large numbers

$$K_{\theta_i}^{(k)}(\boldsymbol{\theta}) = \frac{1}{m} \sum_{l=1}^{m} \Psi_i(T_{il}, \boldsymbol{\theta}, X_k).$$

The expected value of K_{θ_i} is zero, therefore the covariance matrix entries are given by the following formula

$$\hat{\Sigma}_{i,j} = \frac{1}{n-1} \sum_{k=1}^{n} K_{\theta_i}^{(k)}(\boldsymbol{\theta}) K_{\theta_j}^{(k)}(\boldsymbol{\theta}).$$

If we repeat this procedure n_1 times, a sample of covariance matrix entries are approximated, from which we can calculate the 10% and 90% percentiles. This will provide us with a statistic which will be used to test whether the result obtained from the numerical integration method is close to the true value of the covariance matrix.

Figure 11.6 displays an example of a graph which shows that the numerical integration method is accurate for fixed values of β, σ and μ. Only the diagonal elements of Σ are shown in this graph.

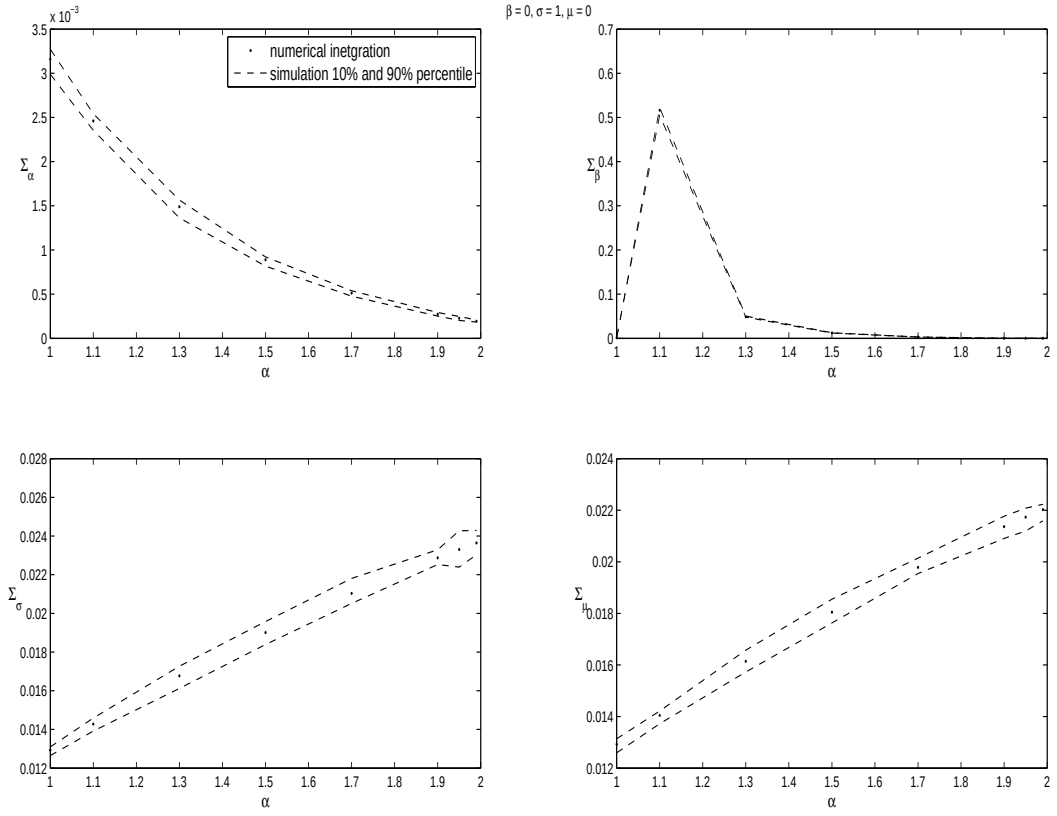

Figure 11.6: These graphs show the accuracy of the numerical integration method to calculate the entries of the covariance matrix, Σ.

11.2.5 The 4×4 symmetric matrix (Λ matrix).

In this section the method to approximate the 4×4 symmetric matrix, denoted by Λ, defined in formula (10.9) on page 151, will be explained and tested.

As mentioned earlier in this section, the entries into Λ are approximated by

$$\lambda(\alpha) \approx \int_{-c}^{c} e^{-2|t|^{\alpha}} |t|^{2\alpha} \left(\log |t|\right)^2 \, dG(t).$$

The value of c is calculated by taking the inverse of G, i.e. $c = G^{-1}(1-\epsilon)$ for a small positive number ϵ.

The integral itself is calculated with the *"quadgk"* function from Matlab.

The entries of this matrix can also be tested by a simulation method. Define for each

$i, j = 1, 2, 3, 4$ a set of normally distributed random variables $(T_{ij1}, T_{ij2}, ..., X_{ijm})$, and let

$$\psi_{i,j}(t, \boldsymbol{\theta}) = \frac{\partial}{\partial \theta_i} u(t; \boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} u(t; \boldsymbol{\theta}) + \frac{\partial}{\partial \theta_i} v(t; \boldsymbol{\theta}) \frac{\partial}{\partial \theta_j} v(t; \boldsymbol{\theta}), \quad i, j = 1, 2, 3, 4$$

then

$$\Lambda_{\theta_i \theta_j}(\boldsymbol{\theta}) = E_G \left[\psi_{i,j}(t, \boldsymbol{\theta}) \right], \quad i, j = 1, 2, 3, 4.$$

Therefore, by the law of large numbers, estimates for these entries are given by

$$\hat{\Lambda}_{i,j}(\boldsymbol{\theta}) = \frac{1}{m} \sum_{l=1}^{m} \psi_{i,j}(T_{ijl}, \boldsymbol{\theta}), \quad i, j = 1, 2, 3, 4.$$

These estimates can be used in the same manner, used in Section 11.2.3, to test whether the matrix entries are accurate.

11.3 Relative efficiency

In this section the relative efficiency of the family of stable distribution parameters is approximated by approximating the measures defined in formulas (10.14) to (10.17) on page 153.

Figure 11.7 shows the approximated relative efficiency if a normally distributed density function with mean zero and standard deviation $\frac{1}{\sqrt{2}}$ is used in (11.8). The efficiency is approximated as a function of the α parameter, when β, σ and μ are given fixed values. Two different values of β were used. The relative efficiency of the ILSEs is quite good, especially when estimating the σ parameter. Notice that the relative efficiency cannot be calculated at $\alpha = 1$, when $\beta \neq 0$, as the characteristic function of the family of stable distributions, used in this study, is not continuous at $\alpha = 1$, when $\beta \neq 0$.

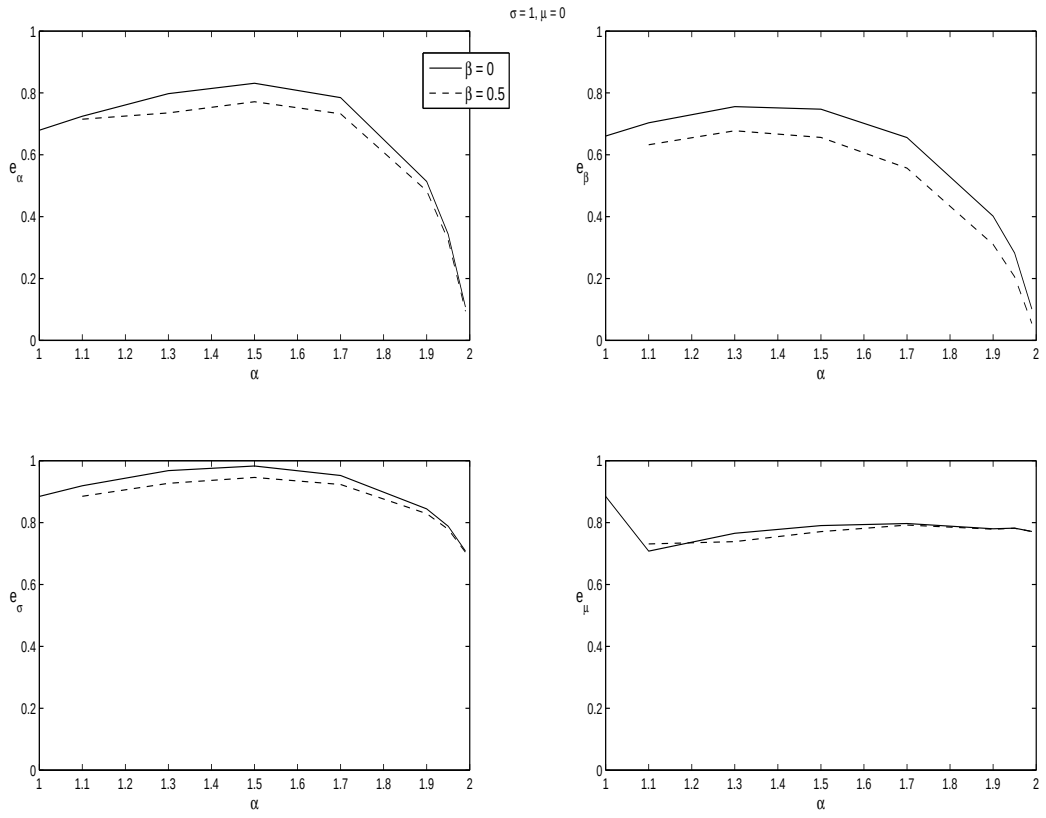

Figure 11.7: The relative efficiency of the four stable parameter estimates when applying the ILSE method.

Various density functions will be used to investigate whether it is possible to improve the efficiency of the ILSEs. In Section 11.2.1 an improvement of the relative efficiency was obtained by using two density functions, $g_1(t, 0, 0.1)$ and $g_1(t, 0.7, \frac{1}{\sqrt{2}})$, where g_1 is defined in (11.8) on page 164.

Both weight functions $g_1(t, 0, 0.1)$ and $g_1(t, 0.7, \frac{1}{\sqrt{2}})$ result in higher relative efficiency for the single parameter case, compared to results using the original normal density function, $g_1(t, 0, \frac{1}{\sqrt{2}}) = g(t, 0, \frac{1}{\sqrt{2}})$, as a weight function. We will investigate whether this is true for the general case.

Figure 11.8 shows us that in the general case, $g_1(t, 0, 0.1)$ and $g_1(t, 0.7, \frac{1}{\sqrt{2}})$ lead to less efficient ILSEs. Notice that the introduction of all parameters into our calculation has increased the efficiency of the ILSE of α when comparing Figure 11.8 to Figure 11.5 on page 165.

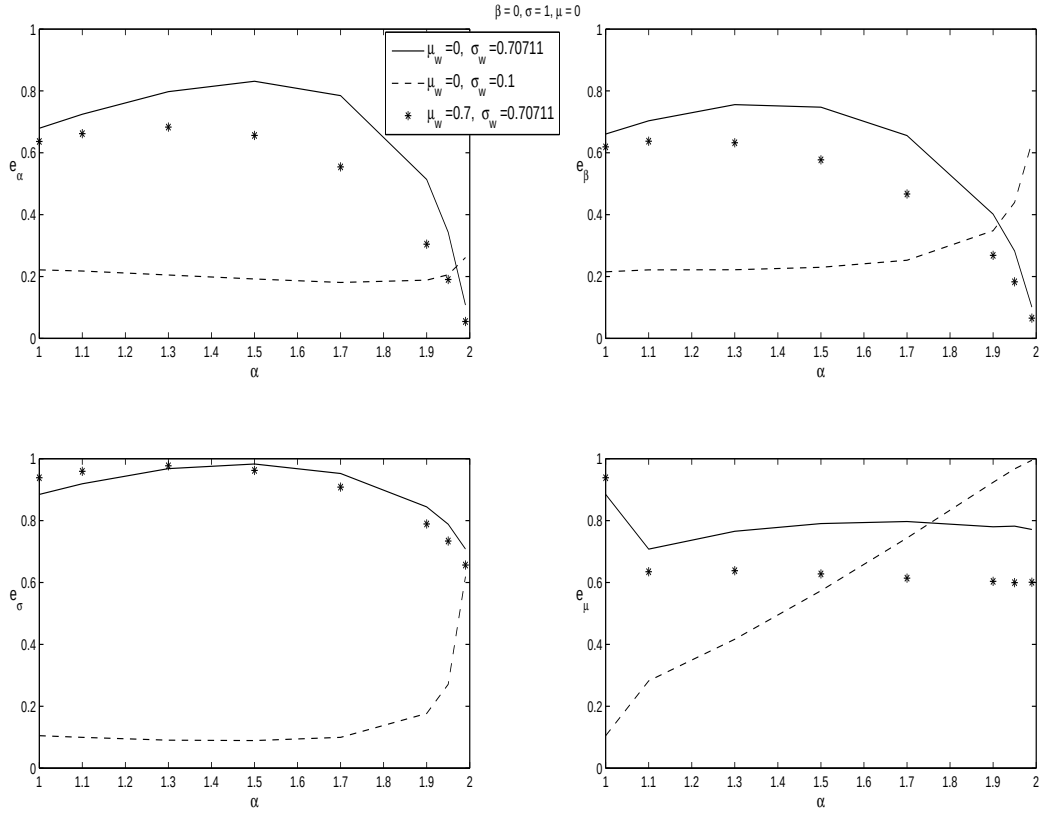

Figure 11.8: Comparing the relative efficiency of various weight functions, that are similar to those in the single parameter case. The stable distribution parameters used are $\beta = 0$, $\sigma = 1$ and $\mu = 0$. The weight function parameters are those that define the weight function g_1 in formula (11.8) on page 164.

Figure 11.9 and Figure 11.10 show a comparison of the relative efficiency, as a function of the α parameter for $\beta = 0$ and $\beta = 0.5$, using various weight functions. We compare the results when using the original weight function $g_1(t, 0, 1./\sqrt{2})$, against the weight functions $g_1(t, 0.1, 0.58)$, $g_1(t, 0.1, 0.5)$ and $g_1(t, 0.1, 0.4)$. From these graphs, we conclude that various weight functions can be used to increase the relative efficiency of ILSEs for higher values of α, while slightly decreasing the efficiency for lower values of α.

Therefore, an improvement in efficiency of the ILSEs can be obtained when fitting a stable distribution, especially if the distribution is fairly close to the normal distribution ($\alpha = 2$).

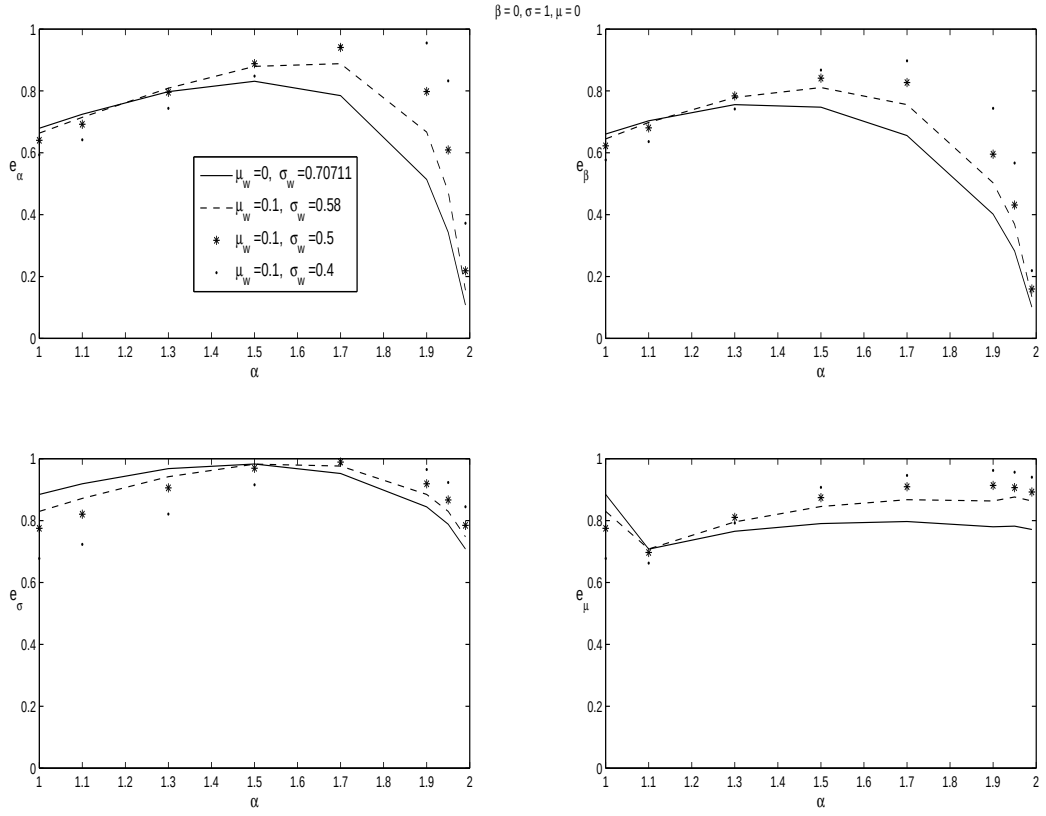

Figure 11.9: Comparing the relative efficiency of various weight functions, which improves the relative efficiency. The stable distribution parameters used are $\beta = 0$, $\sigma = 1$ and $\mu = 0$. The weight function parameters are those that define the weight function g_1 in formula (11.8) on page 164.

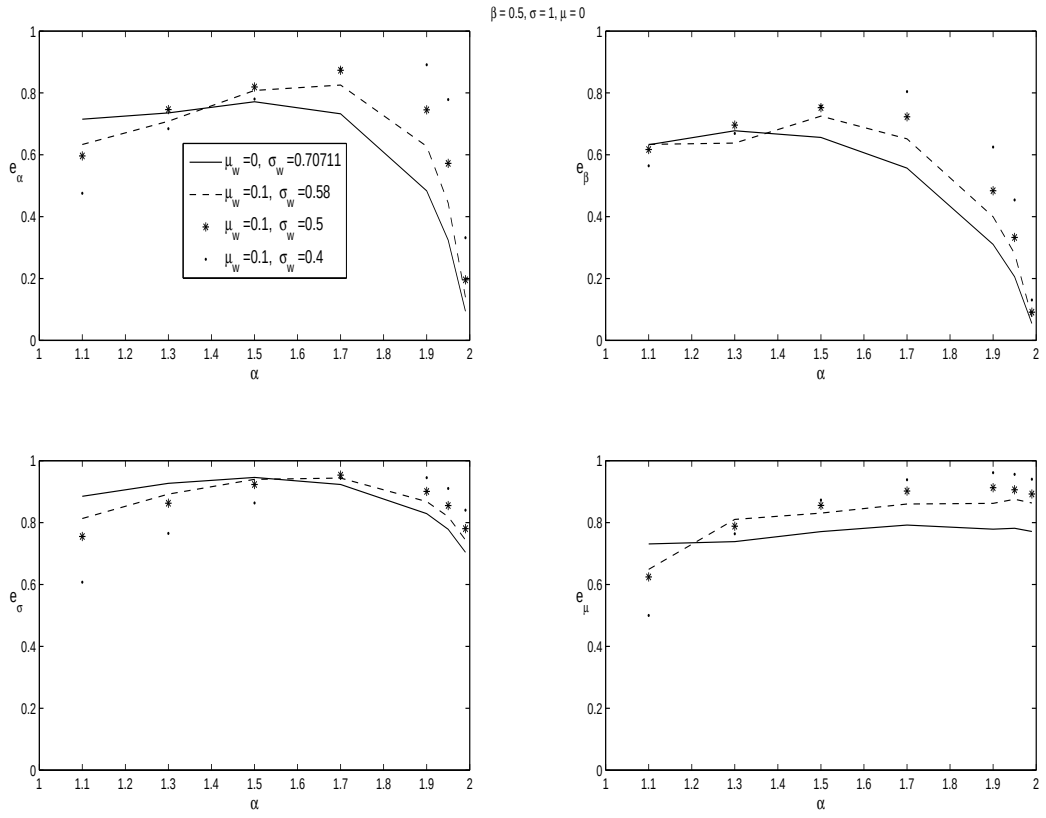

Figure 11.10: Comparing the relative efficiency of various weight functions, which improves the relative efficiency. The stable distribution parameters used are $\beta = 0.5$, $\sigma = 1$ and $\mu = 0$. The weight function parameters are those that define the weight function g_1 in formula (11.8) on page 164.

Appendix A

Consistency and asymptotic normality of ILSEs

The details of the proofs of a number of results from Heathcote (1997) are given in this Appendix. In (10.4) on page 149 it is stated that if we define the function $\lambda(\theta)$ by $\lambda(\theta_0) = \int \left([u'(t;\theta_0)]^2 + [v'(t;\theta_0)]^2 \right) dG(t)$, then

$$\lambda(\theta) = \mathrm{Cov}\left\{ K(\theta), \frac{\partial}{\partial\theta} \log f(x,\theta) \right\}.$$

Expressions for $u'(t;\theta)$ and $v'(t;\theta)$ can be derived from the definitions of u and v. Therefore,

$$u(t;\theta) = E\left[\cos(tX)\right] = \int \cos(tx) f(x,\theta) \mathrm{d}x$$

Now,

$$\begin{aligned}
u'(t;\theta) &= \frac{\partial}{\partial\theta} \int \cos(tx) f(x,\theta) \mathrm{d}x \\
&= \int \cos(tx) \frac{\partial}{\partial\theta} f(x,\theta) \mathrm{d}x \\
&= \int \cos(tx) \frac{\frac{\partial}{\partial\theta} f(x,\theta)}{f(x,\theta)} f(x,\theta) \mathrm{d}x \\
&= \int \cos(tx) \frac{\partial}{\partial\theta} \log f(x,\theta) f(x,\theta) \mathrm{d}x \\
&= E\left[\cos(tx) \frac{\partial}{\partial\theta} \log f(x,\theta) \right] \\
&= \mathrm{Cov}\left\{ \cos(tx), \frac{\partial}{\partial\theta} \log f(x,\theta) \right\},
\end{aligned}$$

where $E\left[\frac{\partial}{\partial\theta} \log f(x,\theta)\right]$ is assumed to be zero and

$$\mathrm{Cov}\left\{X, Y\right\} = E(XY) - E(X)E(Y).$$

Similarly

$$v'(t;\theta) = \mathrm{Cov}\left\{\sin(tx), \frac{\partial}{\partial\theta}\log f(x,\theta)\right\}.$$

Setting

$$K(\theta) = \int \left[\{\cos(tX_j) - u(t;\theta)\}\, u'(t;\theta) + \{\sin(tX_j) - v(t;\theta)\}\, v'(t;\theta)\right] dG(t),$$

then $E\left[K(\theta)\right] = 0$ from the definitions of u and v. The covariance of $K(\theta)$ and $\frac{\partial}{\partial\theta}\log f(x,\theta)$ is then given by

$$\begin{aligned}
&\mathrm{Cov}\left\{K(\theta), \frac{\partial}{\partial\theta}\log f(x,\theta)\right\} \\
&= E\left[K(\theta)\frac{\partial}{\partial\theta}\log f(x,\theta)\right] \\
&= E\left[\left(\int \left[\{\cos(tX_j) - u(t;\theta)\}\, u'(t;\theta) + \right.\right.\right. \\
&\qquad\qquad \left.\left.\left. \{\sin(tX_j) - v(t;\theta)\}\, v'(t;\theta)\right] dG(t)\right) \frac{\partial}{\partial\theta}\log f(x,\theta)\right] \\
&= \int E\left[\cos(tX)\frac{\partial}{\partial\theta}\log f(x,\theta)u'(t;\theta)\right] dG(t) \\
&\qquad - \int E\left[u(t;\theta)\frac{\partial}{\partial\theta}\log f(x,\theta)u'(t;\theta)\right] dG(t) \\
&\qquad\qquad + \int E\left[\sin(tX)\frac{\partial}{\partial\theta}\log f(x,\theta)v'(t;\theta)\right] dG(t) \\
&\qquad\qquad\qquad - \int E\left[v(t;\theta)\frac{\partial}{\partial\theta}\log f(x,\theta)v'(t;\theta)\right] dG(t) \\
&= \int \left[u'(t;\theta)^2 + v'(t;\theta)^2\right] dG(t) \\
&= \int |\phi'(t;\theta)|^2 dG(t) \\
&= \lambda(\theta).
\end{aligned}$$

We used the formula for the variance of $K(\theta)$ in many cases. It is given by

$$\begin{aligned}
\mathrm{Var}\left[K(\theta)\right] &= E\left[K(\theta)^2\right] \\
&= E\left\{\int \left[\left\{\cos(tX) - u(t;\theta)\right\} u'(t;\theta)\right.\right. \\
&\qquad\qquad \left.\left. + \left\{\sin(tX) - v(t;\theta)\right\} v'(t;\theta)\right] dG(t) \\
&\int \left[\left\{\cos(sX) - u(s;\theta)\right\} u'(s;\theta)\right. \\
&\qquad\qquad \left. + \left\{\sin(sX) - v(s;\theta)\right\} v'(s;\theta)\right] dG(s)\Big\} \\
&= \iint [\mathrm{Cov}\left\{\cos(tX), \cos(sX)\right\} u'(t;\theta)u'(s;\theta) \\
&+ 2\mathrm{Cov}\left\{\cos(tX), \sin(sX)\right\} u'(t;\theta)v'(s;\theta) \\
&+ \mathrm{Cov}\left\{\sin(tX), \sin(sX)\right\} v'(t;\theta)v'(s;\theta)]dG(t)dG(s)
\end{aligned}$$

Basic trigonometric identities, for example

$$\cos(A)\cos(B) = \frac{1}{2}(\cos(A-B) + \cos(A+B)),$$

leads to the following properties

$$\begin{aligned}
\mathrm{Cov}\left\{\cos(tX), \cos(sX)\right\} &= \frac{1}{2}\left(u(t-s;\theta) + u(t+s;\theta) - 2u(t;\theta)u(s;\theta)\right) \\
\mathrm{Cov}\left\{\cos(tX), \sin(sX)\right\} &= \frac{1}{2}\left(v(t+s;\theta) - v(t-s;\theta) - 2u(t;\theta)v(s;\theta)\right) \\
\mathrm{Cov}\left\{\sin(tX), \sin(sX)\right\} &= \frac{1}{2}\left(u(t-s;\theta) - u(t+s;\theta) - 2v(t;\theta)v(s;\theta)\right)
\end{aligned}$$

Lastly, to apply the Central Limit Theorem, the random variable $K(\theta)$ needs to be bounded. Observe that (with $|\cos(tX)| \leq 1$ and $|\sin(tX)| \leq 1$)

$$\begin{aligned}
K(\theta)^2 &\leq 4\left(\int \left[(u'(t;\theta))^2 + (v'(t;\theta))^2\right] dG(t)\right) \\
&= 4\int \left(\mathrm{Cov}\left\{\cos(tX), \frac{\partial}{\partial\theta}\log f(x,\theta)\right\}\right)^2 dG(t) \\
&\qquad + 4\int \left(\mathrm{Cov}\left\{\sin(tX)\frac{\partial}{\partial\theta}\log f(x,\theta)\right\}\right)^2 dG(t) \\
&= 4\int \left(E\left[\cos(tX)\frac{\partial}{\partial\theta}\log f(x,\theta)\right]\right)^2 dG(t) \\
&\qquad + 4\int \left(E\left[\sin(tX)\frac{\partial}{\partial\theta}\log f(x,\theta)\right]\right)^2 dG(t) \\
&\leq 4\int E\left[\cos^2(tX)\right] E\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2\right] dG(t) \\
&\qquad + 4\int E\left[\sin^2(tX)\right] E\left[\left(\frac{\partial}{\partial\theta}\log f(x,\theta)\right)^2\right] dG(t) \\
&= 4\mathscr{I}(\theta)\int \left(E\left[\cos^2(tX)\right] + E\left[\sin^2(tX)\right]\right) dG(t) \\
&= 4\mathscr{I}(\theta)\int dG(t) \\
&= 4\mathscr{I}(\theta)
\end{aligned}$$

where the total variation of $G(t)$ is assumed to be one. Therefore, the following two equations are true.

$$|K(\theta)| \leq 2\sqrt{\mathscr{I}(\theta)},$$

and

$$\mathrm{Var}\left[K(\theta)\right] \leq 4\mathscr{I}(\theta).$$

Appendix B

Figures

B.1 Stable density functions

In this section various density functions for the family of stable distributions are shown. The density values were approximated using the method described in Part *II*.

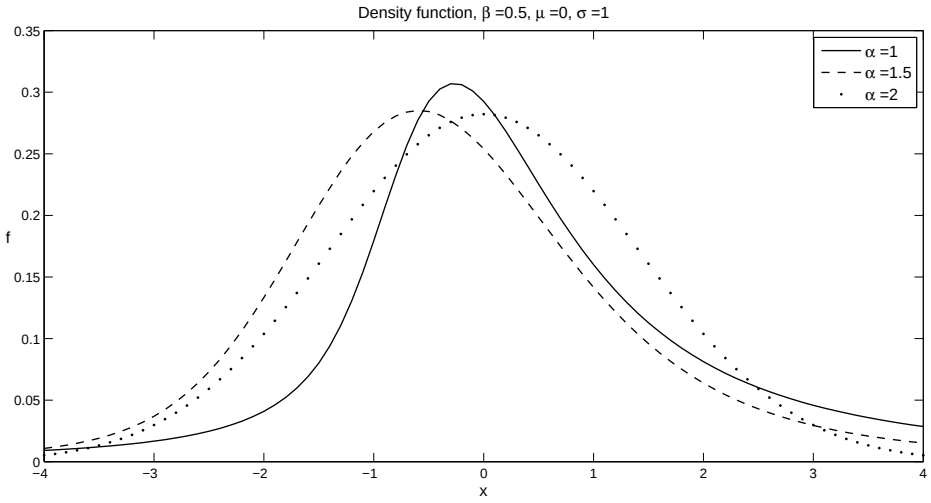

Figure B.1: Figures of the density function for various values of the parameter α.

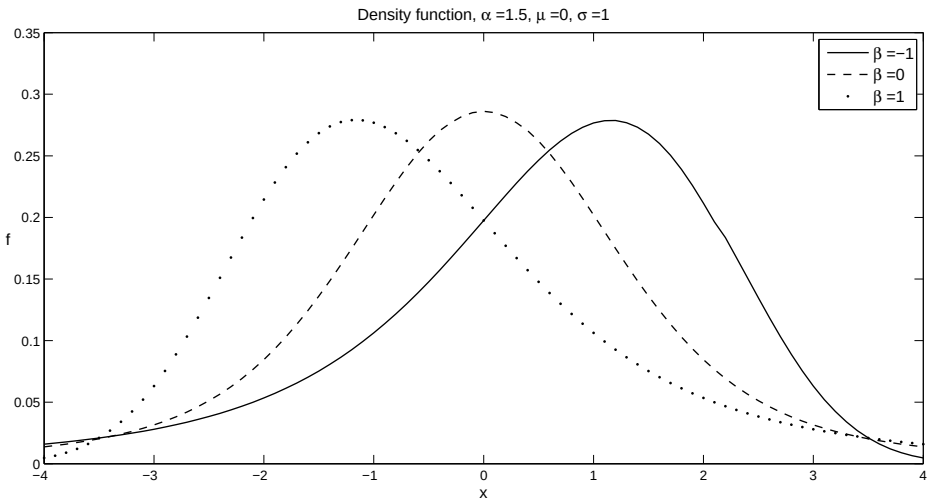

Figure B.2: Figures of the density function for various values of the parameter β.

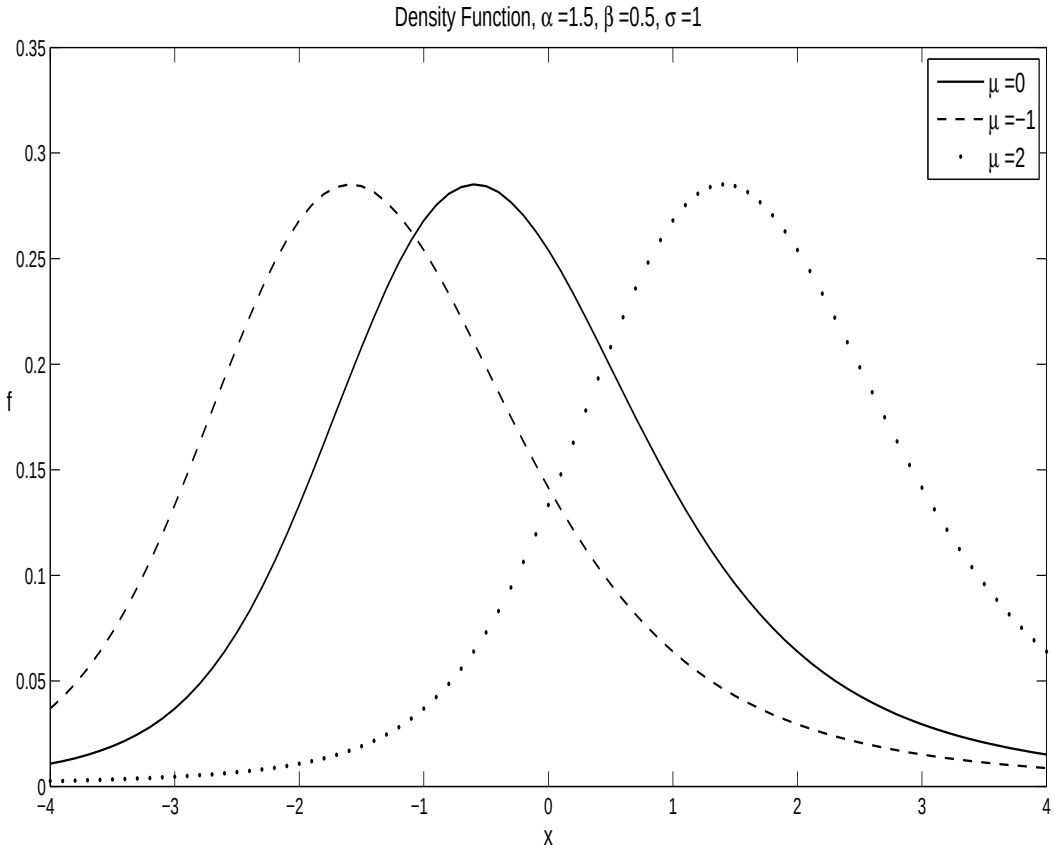

Figure B.3: Figures of the density function for various values of the parameter μ.

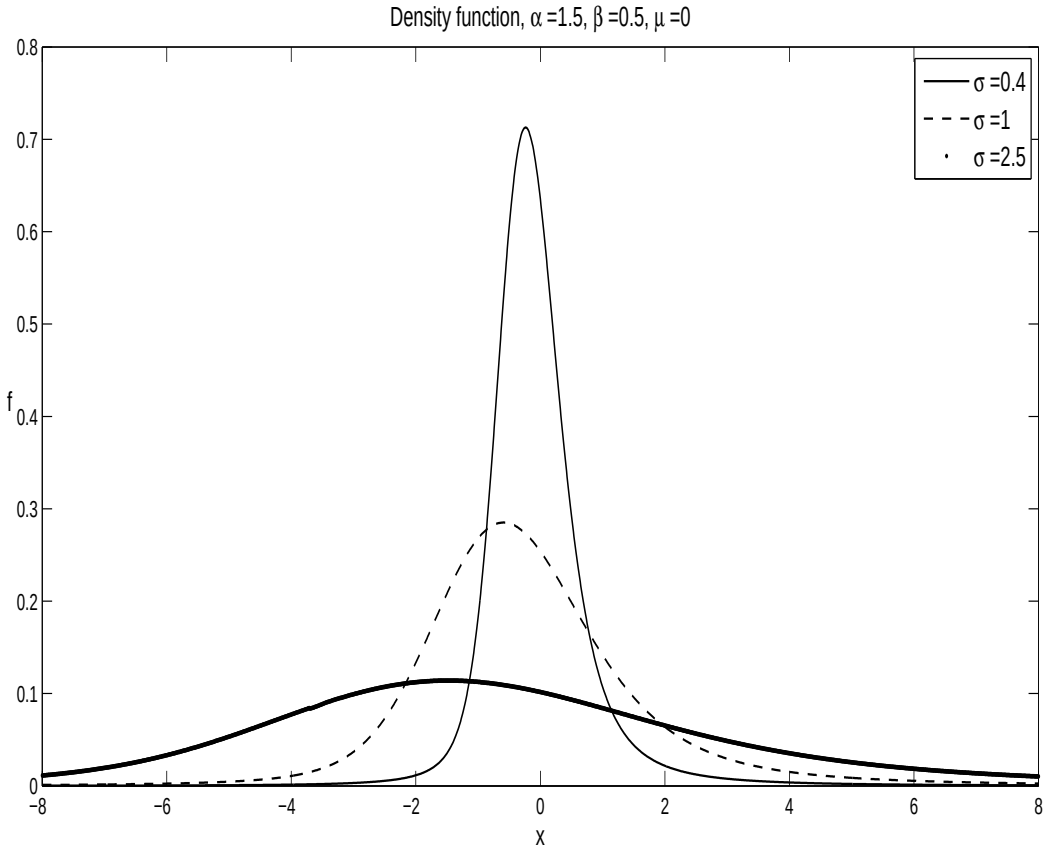

Figure B.4: Figures of the density function for various values of the parameter σ.

B.2 Initial estimator of the skewness parameter in stable distributions.

In Section 9.4 we defined an estimate for the skewness parameter of the family of stable distributions, β. In Section 9.5 we simulated samples of stably distributed random variables for a range of values of parameters α and β. From the simulated samples we estimated β, by β_0, defined in (9.5) on page 140. We showed that β_0 seems to be an unbiased estimator of β, at different values of α. This conclusion was made based on the graphs shown in Figure 9.1 on page 142. In Figure $B.5$ and Figure $B.6$ we show similar graphs for a range of σ and μ values. The conclusion is that the initial estimate β_0, in general, seems to be an unbiased estimator of β.

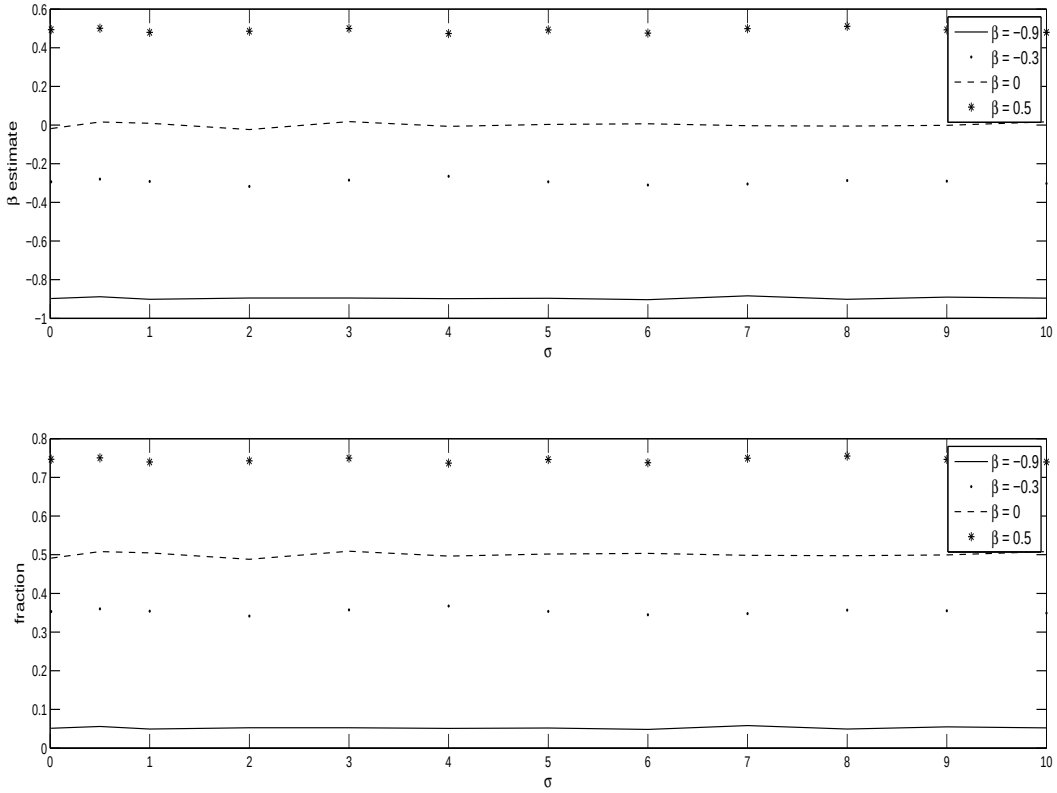

Figure B.5: The statistic β_0 and x^* for a range of σ parameter values.

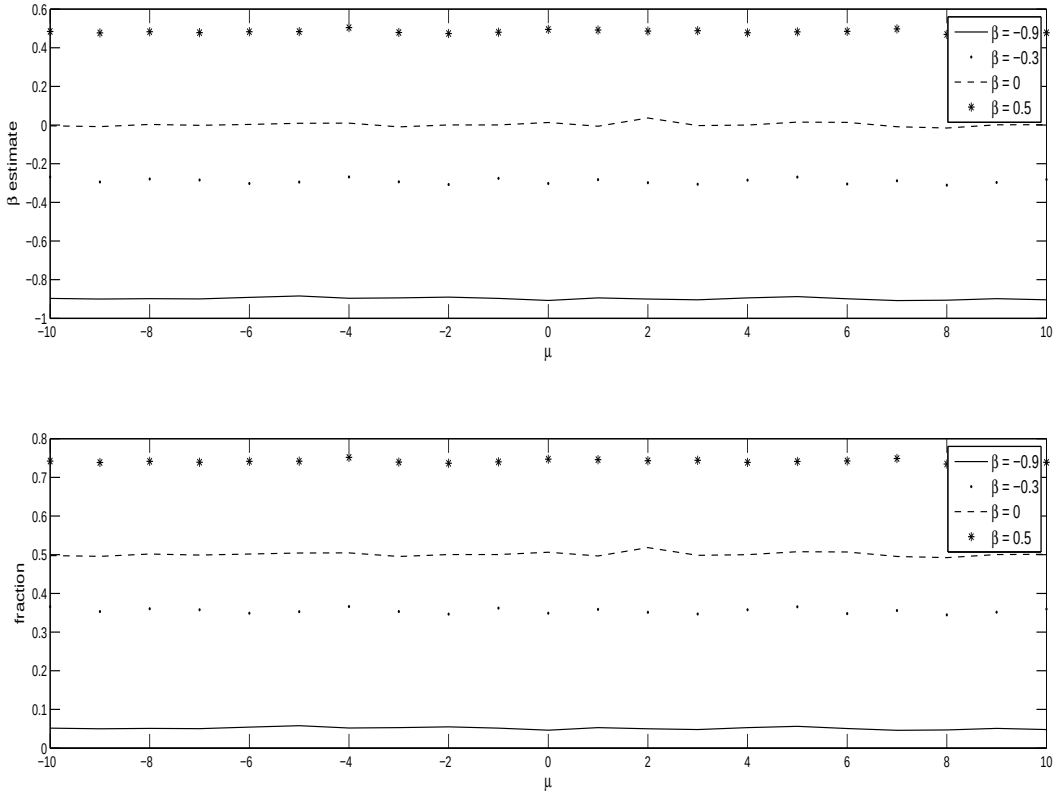

Figure B.6: The statistic β_0 and x^* for a range of μ parameter values.

Bibliography

Anon. 2015. Historic inflation South Africa - CPI inflation. http://www.inflation.eu/inflation-rates/south-africa/historic-inflation/cpi-inflation-south-africa.aspx Date of access: 3 June 2015.

Ait-Sahalia, Y. 2002. Telling from discrete data whether the underlying continuous-time model is a diffusion. *The journal of Finance,* 57(5):2075–2112.

Ait-Sahalia, Y., Jacod, J. & Li, J. 2012. Testing for jumps in noisy high-frequency data. *Journal of Econometrics,* 168:207–222.

Ait-Sahalia, Y. & Jacod, J. 2009. Testing for jumps in a discretely observed process. *Annals of Statistics,* 37:184–222.

Aling, P. & Hassan, S. 2012. No-arbitrage one-factor models of the South African term structure of interest rates. *South African journal of Economics,* 80:301–318.

Axensten, P. 2006. Cauchy cdf, pdf, inverse cdf, parameter fit, and random generator. http://www.mathworks.com/matlabcentral/fileexchange/11749-cauchy/content/paxmle.m Date of access: 10 May 2016.

Barndorff-Nielsen, O. & Shepard, N. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics,* 2:1–48.

Bjork, T. 2004. Arbitrage theory in continuous time. 2nd ed. Oxford: Oxford University Press.

Bjork, T., Kabanov, Y. & Runggaldier, W. 1997. Bond market structure in the presence of marked point processes. *Mathematical Finance,* 7(2):211–223.

Black, F. 1976. The pricing of commodity contracts. *Journal of Financial Economics,* 3:167–179.

Brandimarte, P. 2006. Numerical methods in Finance and Economics. Hoboken, NJ: Wiley.

Brennan, M. & Schwartz, E. 1980. Analysing convertible bonds. *Journal of Financial and Quantitative Analysis,* 15:907–929.

Brigo, D. & Mercurio, F. 2006. Interest rate models - theory and practice: with smile, inflation and credit. 2nd ed. Berlin Heidelberg: Springer-Verlag.

Bunday, B.D. & Garside, G.R. 1987. Optimization methods in Pascal. Baltimore: Edward Arnold.

Burden, R.L., Faires, J.D. & Burden, A.M. 2016. Numerical Analysis. 10th ed. Boston, MA: Cengage Learning.

Carr, P. & Wu, L. 2003. What type of process underlies options? A simple robust test. *Journal of Finance,* 58:2581–2610.

Chance, D. 1995. A chronology of derivatives. *Derivatives quarterly,* 2:53–60.

Cheang, G. & Chiarella, C. 2011. A modern view on Merton's jump diffusion model. *Quantitative Finance Research Centre research paper series,* 287 http://qfrc.uts.edu.au/research/research_papers/rp287.pdf Date of access: 12 May 2017.

Cheney, E. & Kincaid, D. 1991. Numerical Analysis: The Mathematics of Scientific Computing. 1st ed. Belmont, CA: Thomson Brooks/Cole.

Cont, R. & Tankov, P. 2004. Financial modelling with jump processes. 1st ed. London: Chapman and Hall/CRC.

Cox, J. 1975. Notes on option pricing I: constant elasticity of variance diffusions. (Unpublished).

Cox, J., Ingersoll, J. & Ross, S. 1985. A theory of the term structure of interest rates. *Econometrica,* 53(2):385–407.

Davison, A. & Hinkley, D. 1997. Bootstrap methods and their applications. 1st ed. Cambridge: Cambridge University Press.

Dumouchel, W. 1975. Stable distributions in statistical inference: 2. Information from stably distributed samples. *Journal of the American Statistical Association,* 70(350):386–393.

Dumouchel, W. 1983. Estimating the stable index α in order to measure tail thickness: a critique. *Annals of Statistics,* 11(4):1019–1031.

Dynare. 2008. Kernel density estimate. https://github.com/DynareTeam/dynare/blob/master/matlab/kernel_density_estimate.m. Date of Access: 9 Feb. 2017.

Fama, E. & Roll, R. 1968. Some properties of symmetric stable distributions. *Journal of the American Statistical Association,* 63(323):817–836.

Feuerverger, A. & Mureika, R. 1977. The empirical characteristic function and its applications. *Annals of Statistics,* 5(1):88–97.

Giesecke, K. & Zhu, S. 2013. Transform analysis for point processes and applications in credit risk. *Mathematical Finance,* 23:742–762.

Gumbo, V. 2012. The SAFEX-JIBAR market models. *Journal of Mathematical Finance,* 2:321–326.

Hassan, S. 2013. South African capital markets: an overview. *Economic Research Southern Africa,* 391. https://econrsa.org/system/files/publications/working_papers/working_paper_391.pdf Date of access: 12 May 2017.

Heathcote, C. 1997. The integrated squared error estimation of parameters. *Biometrika,* 64(2):255–264.

Hong, L. & Zou, J. 2015. Jump tests for semimartingales. *South African Actuarial journal,* 15:93–108.

Hull, J. 2006. Options, futures, and other derivatives. 6th ed. Upper Saddle River, NJ: Pearson, Prentice Hall.

Hull, J. & White, A. 1990. Pricing interest rate derivative securities. *The review of financial studies* 3(4):573–592.

Johannes, M. 2004. The statistical and economic role of jumps in continuous-time interest rate models. *The journal of Finance,* 59(1):227–260.

JSE. 2012. The JSE zero-coupon yield curves. https://www.jse.co.za/content/JSEIndexCalculationMethodologyandConstructionItems/The%20JSE%20Zero%20Coupon%20Yield%20Curves.pdf Date of access: 12 Jan. 2017.

Kou, S. & Wang, H. 2004. Option pricing under a double exponential jump diffusion model. *Management Science,* 50:1178–1192.

Koutrouvelis, I. 1980. Regression-type estimation of the parameters of stable laws. *Journal of the American Statistical Association,* 75(372):918–928.

Kuo, H. 2005. Introduction to Stochastic Integration. 1st ed. New York, NY: Springer.

Lee, S. & Mykland, P. 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *The review of financial studies,* 21(6):2535–2563.

Leonhardt, D. 2016. Negative 0.5% interest rate: why people are paying to save. *The New York times:* 12 Feb. https://www.nytimes.com/2016/02/13/upshot/negative-interest-rates-are-spreading-across-the-world-heres-what-you-need-to-know.html?_r=0 Date of access: 12 May 2017.

Liu, C., Chen, H., Shao, L., Wang, J. & You, C. 2014. Parameterizations and parameter relations of stable distribution. *Applied Mechanics and Materials,* 543:1721–1727.

Longstaff, F. & Schwartz, E. 1992. Interest rate volatility and the term structure: A two-factor general equilibrium model. *Journal of Finance,* 47:1259–1282.

Malumisa, S. 2015. Stochastic volatility models in financial econometrics: an application to South Africa. Johannesburg: University of the Witwatersrand. (Thesis - PhD).

McCulloch, J. 1986. Simple consistent estimators of stable distribution parameters. *Communications in Statistics - simulation and computation,* 15(4):1109–1136.

Merton, R. 1976. Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics,* 3(1):125–144.

Nolan, J. 1999. An algorithm for evaluating stable densities in Zolotarev's (M) parameterization. *Mathematical and Computer Modelling,* 29:229–233.

Nolan, J. 2017a. Stable distributions - models for heavy tailed data. Boston, MA: Birkhauser. Chapter 1 online at http://fs2.american.edu/jpnolan/www/stable/stable.html Date of access: 12 May 2017.

Nolan, J. 2017b. Information on stable distributions. http://fs2.american.edu/jpnolan/www/stable/stable.html Date of access: 12 May 2017.

Nowman, K. 1997. Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *Journal of Finance,* 52:1695–1706.

Paulson, A., Holcomb, E. & Leitch, R. 1975. The estimation of the parameters of the stable laws. *Biometrika,* 62(1):163–170.

Privault, N. 2013. Stochastic finance: an introduction with market examples. 1st ed. London: Chapman and Hall/CRC.

Protter, P. 2005. Stochastic integration and differential equations. 2nd ed. Berlin Heidelberg: Springer-Verlag.

Rice, J. 1995. Mathematical Statistics and data analysis. 2nd ed. Belmont, CA: Duxbury Press.

Safex. 2014. JIBAR archive. http://www.safex.co.za/PUB/mtmdata/Jibar%20Archive/ Date of access: 2 Mar. 2014.

SARB. 2012. A review of the rate-setting process of the Johannesburg Interbank Agreed Rate (JIBAR) as an interest rate benchmark. https://www.resbank.co.za/Lists/News%20and%20Publications/Attachments/5292/JibarReview.pdf Date of access: 15 Nov. 2012.

SARB. 2015. Selected historical rates. https://www.resbank.co.za/Research/Rates/Pages/SelectedHistoricalExchangeAndInterestRates.aspx Date of access: 3 June 2015.

Van der Vaart, A. 1998. Asymptotic Statistics. Cambridge: Cambridge University Press.

Vasicek, O. 1977. An equilibrium characterisation of the term structure. *Journal of Financial Economics,* 5:177–188.

Weisstein, E.W. 2017. Kurtosis. From MathWorld: a Wolfram web resource. http://mathworld.wolfram.com/Kurtosis.html Date of access: 17 Feb. 2016.

Weron, R. 1996. On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. *Statistics and Probability letters,* 18:165–171.

West, G. 2008. Interest rate derivatives in the South African market based on the prime rate. *Journal for studies in Economics and Econometrics,* 32:77–87.

Zolotarev, V. 1986. One-dimensional stable distributions. *American Mathematical Society,* 65.