MSc. Thesis

Financial modeling and risk measurement on an emerging market using Lévy processes

Author:
Charlene T. Chipoyera

Supervisor:
Prof. Diane L. Wilcox
Co-Supervisor:
Dr Virginie K. Socgnia

Programme in Advanced Mathematics of Finance
School of Computer Science and Applied Mathematics
University of the Witwatersrand
Private Bag 3, Wits-2050, Johannesburg
South Africa

August 7, 2018

Declaration

I declare that this dissertation is my own, unaided work. It is being submitted for the Degree of Master of Science at the University of the Witwatersrand. It has not been submitted before for any degree or examination in any other University.

August 7, 2018

Abstract

Efficient financial risk management is fundamental to good business decision making. Risk management heavily relies on the use of mathematical techniques to measure risk, hence it is important to use accurate models when measuring risk. The movement from using models based on geometric Brownian motion is due to its inability to capture many stylized facts of asset returns. Some of the well known shortcomings of using Brownian motion when modeling asset returns that can be addressed by using Lévy processes include jumps, skewness and heavy tails. This thesis focuses on generalized hyperbolic Lévy processes to model asset returns. The two representations of the generalized hyperbolic distribution (GHD) considered in this thesis are the normal mean variance mixture introduced by McNeil *et al.* (2005) and the subordinated Brownian motion representation. The results presented in this thesis argue the case for using GHD models to model intraday data to using the normal distribution. The goodness of fit tests performed showed that there were no significant differences between the performance of the two representations of the generalized hyperbolic distribution. Risk measures based on the GHD and normal distribution are defined and evaluated. The results show that the GHD risk measures perform remarkably better than the Gaussian risk measures.

Acknowledgments

Firstly I would like to thank my co-supervisor Dr Virginie Konlack Socgnia for giving me this research topic and for her guidance. I am very grateful to my supervisor Professor Diane Wilcox for her consistent guidance and invaluable suggestions. I would like to express my sincere appreciation to the QueriLab team for giving me access to the data that I used in this project and for all their suggestions that helped make this work better. Furthermore, I would like to acknowledge the NRF[1] for funding projects with grant numbers 87830, 74223, 70643 that laid foundations for the data that was used in this project. I am forever thankful to the Head of School, Professor Joel Moitsheki for his encouragement and the administrative work that he put in order for this project to be marked. To my parents, relatives and friends I cannot thank you enough for your support and encouragement throughout my journey. Finally, I would like to thank God for giving me the strength and guidance to produce this work.

Trust in the Lord with all your heart; do not depend on your own understanding. Seek His will in all you do, and He will show you which paths to take. - Proverbs 3 vs 5-6

[1] The opinions expressed in this document are not necessarily those of the National Research Foundation.

Acronym	Meaning
ACF	Autocorrelation Function
AD	Anderson Darling
ADF	Augmented Dickey Fuller
ARMAD	Average Relative Mean Absolute Deviation
BIC	Bayesian Information Criterion
BIC_0	Bayesian Information Criterion per observation
EM	Expectation Maximization
ES	Expected Shortfall
Eqn.	Equation
GBM	Geometric Brownian Motion
GHD	Generalized Hyperbolic Distribution
GIG	Generalized Inverse Gaussian
GMM	Generalized Method of Moments
HYP	Hyperbolic Distribution
$i.i.d.$	independent and identically distributed
JSE	Johannesburg Stock Exchange
KPSS	Kwiatkowski Phillips Schmidt Shin
KS	Kolmogorov Smirnov
LR	Likelihood Ratio
MASE	Mean Absolute Scaled Error
MLE	Maximum Likelihood Estimation
MME	Method of Moment Estimators
NIG	Normal Inverse Gaussian
NMM	Numerical Method of Moments
PDF	Probability Density Function
POF	Proportion of Failures
PSD	Positive Definite Symmetric
RIC	Reuters Instrument Code
SKT	Skew students t
UC	Unconditional Coverage
VG	Variance Gamma
VaR	Value at Risk
VR	Variance Ratio

Tab. 0.1: List of accronyms and their meanings

Contents

List of Figures

List of Tables

Chapter 1

Introduction

A common practice throughout financial modeling literature is using quantitative techniques to model asset prices through price movements. The idea of using a stochastic process to model asset prices was introduced by Bachelier (1900). In his PhD thesis, Bachelier (1900) evaluated option prices using arithmetic Brownian motion, this theory did not receive significant attention due to the fact that it accommodated negative asset prices. To correct Bachelier's mistake, Samuelson (1965) used geometric Brownian motion (GBM) to model asset prices and his ideas were later used by Black and Scholes (1973) to produce the famous Black-Scholes formula for pricing European options.

GBM is one of the most common models used to model the behavior of the price of a non-dividend paying stock. GBM paths are continuous, so only small price movements are permitted in a short period of time, however asset prices may have big jumps in a short period of time. GBM increments are independent and stationary and follow the normal distribution. The drawback of using GBM to model asset prices is that it will not be able to capture some of the known features of asset returns such as skewness, jumps and semi-heavy to heavy tails .

Lévy processes are a better model for asset returns compared to GBM in that they weaken the condition of path-wise continuity to continuity in probability, allowing them to model jumps. Depending on the type of the *Lévy measure*[1], it is possible to have a distribution that is asymmetric with semi-heavy to heavy tails. Lévy processes are named after Paul Lévy (1886-1971) a French mathematician. Mandelbrot (1963) was the first to use Lévy processes to model asset returns.

[1] The expected number per unit time of jumps in any given time interval.

Lévy processes can be divided into two categories, namely infinite activity processes and jump diffusion processes. In jump diffusion models, price trajectories are driven by a diffusion process that is punctuated by jumps at random times and jumps represent rare events such as crashes. Price trajectories of infinite activity models are characterized by an infinite number of jumps and small jumps are more frequent than large jumps.

Whilst jump diffusion models are easy to simulate and the distribution of jump sizes is known, statistical estimation of these models is difficult since their closed form densities are unknown. Some process in the infinite activity category can be represented as Brownian motion that is time changed by a Lévy process that is increasing almost everywhere $(a.e)$ making them analytical tractability. Two such processes that can be represented as time changed Brownian motion that were considered in this thesis are the normal inverse Gaussian (NIG) and the variance gamma (VG).

Geman (2002) provides the following properties of Lévy processes that strengthen the argument for using them in financial modeling :

1. Representation of asset returns using Lévy process are consistent with the no arbitrage assumption.

2. Price changes can be expressed as changes resulting from large amount of shocks in the economy since Lévy processes are related to infinitively divisible distributions.

3. Discontinuous Lévy processes can be introduced if the data deviates significantly from normality.

Infinitely divisible distributions are an important class of statistical distributions in the theory of Lévy processes. An infinitely divisible random variable can be thought of as a sum of a large number of independent and identically distributed $(i.i.d.)$ increments. According to Proposition 3.1 in Cont and Tankov (2004), for every infinitely divisible distribution there is an associated Lévy process and every Lévy process has an infinitely divisible distribution.

The generalized hyperbolic distribution (GHD) is a parametric family of infinitely divisible distributions, a property that gives it a relationship with

Lévy processes. The GHD was introduced by Barndorff-Nielsen (1977), the birth and study of the GHD was due to the discovery of irregularities in the distribution of wind blown sands according to size by Bagnold (1941). Eberlein and Keller (1995) were the first to apply the GHD and its subclasses to finance. In mathematical finance the GHD has been used in portfolio optimization, pricing portfolio credit risk and risk management see Hu (2005) for example.

The GHD has several different parameterizations, however only two parameterizations were used in this thesis. The parametrization of the GHD introduced by McNeil *et al.* (2005) is discussed in chapter 4 and the representation of the GHD by Barndorff-Nielsen (1977) is given in chapter 5. McNeil *et al.* (2005) represented the GHD as a normal variance-mean mixture using the generalized inverse Gaussian (GIG) distribution as the mixing distribution. Section 5.1 gives a brief discussion on the construction generalized hyperbolic Lévy process using the representation in Barndorff-Nielsen (1977).

As the dimension of the data increases, it gets more difficult to fit data to the GHD using maximum likelihood estimation (MLE). McNeil *et al.* (2005) representation allows us to compute maximum likelihood estimators of the GHD using the Expectation-Maximization (EM) algorithm. The EM algorithm is an iterative procedure that is used to estimate parameters using two stages. The advantage of the time subordinated Brownian motion representation discussed in chapter 5 is that the parameters have a physical interpretation.

The computer program 'hyp', developed by Blæsild and Sørensen (1992) can estimate the parameters of the hyperbolic distribution (a subclass of the GHD) up to 3 dimensions. Blæsild and Sørensen (1992) reported that if the dimension of the data was four or greater, it would be impossible to calibrate parameters of the GHD. Prause (1999) managed to fit the skewed GHD to data of dimension three and reports that the estimation was slow and unstable for higher dimensions. However, Prause (1999) managed to fit the symmetric GHD to data of dimension greater than three.

Protassov (2003) was able to fit five dimension data the NIG. The algorithm to fit the GHD to high dimension data, developed by McNeil *et al.* (2005)

generalizes Protassov (2003) algorithm. Hu (2005) reports that McNeil *et al.* (2005) algorithm might be unstable in some cases. Konlack Socgnia and Wilcox (2014) fitted daily returns of seven stocks to the normal variance-mean mixture of the GHD by McNeil *et al.* (2005). The work in this thesis extends the work by Konlack Socgnia and Wilcox (2014) to intraday data and to the subordinated Brownian representation of the NIG and the VG.

To ensure that the EM algorithm of the GHD was stable the MLE algorithm of the GIG developed by Jørgensen (1982) and the algorithm developed by Luo and Zhao (2014) for the boundary cases of the GIG were incorporated in the EM algorithm. The discussion of the GIG and its boundary cases is presented in section 4.1. The parameters of the subordinated Brownian motion representation were estimated using three methods, namely the MLE, the generalized method of moments and the numerical method of moments. The EM algorithm is implemented only for the representation by McNeil *et al.* (2005) because the estimation involves first estimating some of the parameters of the GHD in the expectation step and then using a combination of the previous parameters and those obtained from the expectation step to compute parameters of the GIG in the maximization step.

The drawback of the GHD is that they are not closed under convolution except for the NIG and the VG. It is therefore difficult to calibrate the GHD when the data is not on the same time scale for example it is difficult to fit the GHD to a price sheet with options that have different maturities. It is easier to calibrate the GHD to data that is from the same time scale, hence the inhomogeneous time series obtained from the tick data was first resampled into homogeneous time series before calibration.

Long term distributions of asset returns based on monthly or even weekly data points are close to the normal distribution. However intra-day data or even daily data are generally not normally distributed so one can use Lévy process to model returns distributions of intraday data. However, returns may not even be independent on small time scales so one needs to be careful about which time scales to fit using Lévy processes.

Before fitting any Lévy model one has to check whether the asset prices at that sampling frequency have independent and stationary increments. If the increments in consideration are not independent, there are techniques

one can apply before calibration for example Hu (2005) applied GARCH filtering before fitting the returns.

Lévy processes can be modeled using the following three approaches

1. **Specifying the Lévy measure:** The distribution of the process is known since the process is modeled directly from the jump structure.
2. **Time change Brownian motion using an increasing Lévy process:** The density is usually known in closed form although the Lévy measure may not have an explicit formula.
3. **Specify the density of the increments using an infinitely divisible density:** This can be done for any sampling frequency, the process of the generalized hyperbolic processes can be simulated using this approach.

The smooth functioning of financial institutions is imperative to modern society. If risk is not properly managed, the collapse of major financial institutions can result in cascading failures that may bring about financial instability. One such example is the 2008 financial crisis also known as the global financial crisis.

The global financial crisis resulted in the collapse of many financial institutions and massive bailouts of major banks by the United States federal government. The crisis caused stock markets to drop world wide and triggered a severe recession. The lack of adequate capital reserves by financial institutions to fulfill their financial obligations in times of extreme portfolio losses is cited as one of the reasons of the cause of the crisis.

Since the global financial crisis many countries adopted Basel III capital and liquidity standards. Financial institutions are authorized by regulators to use in house models to calculate risk and assess capital requirements based on these risk measures. From the crisis we learn the importance of using risk measures based on models that adequately capture the properties of asset returns.

In order to ensure that funds are allocated efficiently one must use models that correctly capture the riskiness of an asset. Generally there is more probability in tails of asset returns in comparison to that of the normal dis-

tribution. Using a risk measure that is based on the assumption of normality of asset returns will result in a risk measure that generally undermines the risk of a position.

The GHD allows greater probability mass in tails in comparison to the normal distribution and can therefore be used to reduce the modeling error. The GHD will also be able to capture asymmetry in the distribution and thus provide better risk measures in comparison to those based on the normal distribution.

1.1 Thesis Outline

Chapter 2
This chapter gives a brief discussion on how to resample an irregularly spaced time series to a regularly spaced time series. Stylized facts of asset returns are presented as well as a comparison of the properties of asset returns from the JSE to stylized facts of asset returns. Tests to determine whether asset returns from the JSE are independent and stationary are conducted.

Chapter 3
The theory presented in this chapter discusses the mathematical concepts used in construction of Lévy properties. Properties of Lévy processes are also discussed.

Chapter 4
This chapter discusses the properties of the GIG as well as the GHD and its subclasses. A discussion of the EM algorithm is given for the particular case of the GHD that is represented using the a variance-mean mixture with the GIG as a mixing distribution. The results from the implementation of the EM algorithm are presented.

Chapter 5
This chapter discusses two specific Lévy process, namely NIG and the VG. These processes can be represented as Brownian motion time changed by a Lévy process that is increasing almost everywhere. The parameters are estimated using three methods, in particular the MLE, the generalized method of moments (GMM) and the numerical method of moments (NMM). The

NMM that was implemented was developed by Figueroa-López *et al.* (2011) and it is only specific to the VG and NIG represented as time changed Brownian motion with drift.

Chapter 6
In this chapter focus is given to two common risk measures, in particular the value at risk (VaR) and the expected shortfall (ES). It is shown the VaR and ES that are based on the assumption of normality undermine the risk in comparison to the VaR and ES based on the GHD models.

Chapter 7
Concluding remarks and a scope for further research is presented.

Chapter 2

Empirical analysis of intraday returns from the JSE

Since it is difficult to calibrate the GHD to data that is from different time scales, the data was first converted to regularly spaced time intervals before calibration. The time series obtained from tick data was irregularly spaced in time, this is because market ticks arrive at random times in the market. In order to obtain a time series of regularly spaced returns, interpolation methods were used.

In this chapter a brief discussion of the methods that can be used to obtain a regularly spaced time series from a raw time series is given in section 2.1. Section 2.2 discusses stylized facts of asset returns. Since Lévy processes assume independent and stationary increments, section 2.3 and 2.4 presents the results obtained from the tests for independence and stationarity respectively.

2.1 Resampling a financial time series

Homogeneous (regularly spaced) time series for intraday data are usually constructed from high frequency data, which is inhomogeneous (irregularly spaced). One has to apply an interpolation method to obtain a homogeneous time series. The two main interpolation techniques used to re-sample a financial time series are linear interpolation and previous tick interpolation.

Let x_k represent the kth log price of an irregularly spaced time series and Δt denote the time interval between consecutive observations. We would like

to construct an evenly spaced time series of log prices $(x_{t_0+i\Delta t})_{i\in\mathbb{N}}$, starting at time t_0. Let

$$k' = \max(k|t_k \leq t_0 + i\Delta t), \quad t_{k'} \leq t_0 + i\Delta t \leq t_{k'+1}. \tag{2.1}$$

Using previous tick interpolation the log price is

$$x_{t_0+i\Delta t} = x_{k'} \quad \text{and} \tag{2.2}$$

using the linear interpolation method the log price is

$$x_{t_0+i\Delta t} = x_{k'} + \frac{(t_0 + i\Delta t) - t_{k'}}{t_{k'+1} - t_{k'}} \left(x_{k'+1} - x_{k'}\right). \tag{2.3}$$

Linear interpolation makes use of future information from time $t_0 + i\Delta t$, whilst previous tick interpolation only uses exclusive information at time $t_0 + i\Delta t$. The drawback of previous tick interpolation is that in the case of an illiquid stock, extreme jumps in the price process may be experienced which may affect the statistical analysis of extreme returns. It is advisable to use linear interpolation when dealing with illiquid stocks.

The stocks analyzed in this project are amongst the most liquid stocks on the JSE in terms of the number of trades, hence only previous tick interpolation was used to obtain the price process for each day. Note that only trades data were used to construct the time series used to conduct analysis in this project. The log returns were computed at different sampling frequencies for each day, none of the returns were computed using data from different days. The term daily returns is used interchangeably with 8 hour returns. The time span between consecutive observations were 2, 5, 10, 20, 30 minutes and 1 , 2, 4 and 8 hours.

The billing model change was implemented on the Johannesburg Stock Exchange (JSE) on the $30th$ of September 2013. The fee restructuring of trades on the JSE removed the floor price on trades. The initial aim of the project was to calibrate the GHD models before and after the fee restructuring using a year's worth of data for each period. However at the time when the GHD models were calibrated there was only three months worth of data available after the fee restructuring. The GHD models were calibrated using data from 01 October 2012 to 30 September 2013 which is one year before the fee restructuring.

Table A.1 shows summary of the number of trades for top 42 stocks on the JSE in January 2013. The Reuters Instrument Code (RIC) of the stocks that had the highest number of trades in January 2013 are MTN, SHP, GFI, IMP, WHL, NPN and AGL. The names of these stocks can be found in Table A.1. SHP is in the Food and Drug Retailers sector, MTN in the Mobile Telecommunications sector, NPN in the Media sector, WHL in the General Retailers sector and AGL, GFI and IMP are in the Mining sector. AGL, SHP, MTN and NPN log returns did not reject the hypothesis of random walk hypothesis for returns that were sampled at 20 minutes or greater (see Section 2.3). Log returns of AGL, SHP, MTN and NPN passed the stationarity tests (see section 2.4). The GHD models were fitted to log returns of AGL, MTN and NPN.

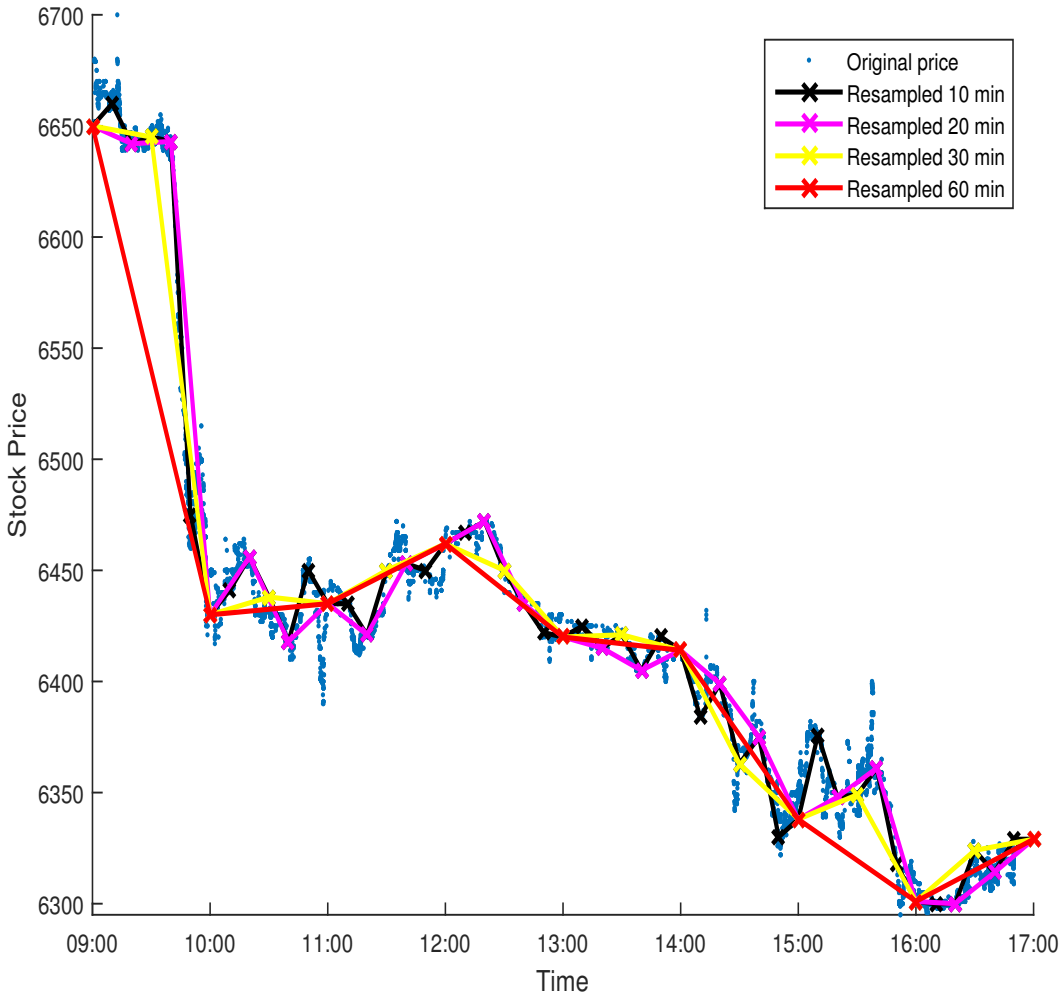


Fig. 2.1: Resampled time series using previous tick interpolation for WHL trade prices on 15 January 2013.

2.2 Stylized empirical facts of financial asset returns

Stylized facts of asset returns are common statistical properties that are expected to present themselves when one analyses any time series of asset returns. The four main categories of stylized facts are distributional issues, scaling properties, autocorrelation of returns and seasonality. Stylized facts help to design models that can adequately capture the properties of return series which then result in more accurate risk measures.

The discussion of stylized facts of asset returns is included to motivate the

use of Lévy process to model asset returns. The discussion is based on the discussion of asset returns presented in Dacorogna *et al.* (2001) and in Cont and Tankov (2004). A summary of the observed properties of asset returns based on four stocks from the JSE is included at the end of this section. The observed properties are in conjunction with the stylized facts of asset returns.

1. **Skewness:** Large negative returns are more common than corresponding large positive returns which in turn results in a distribution that has a larger left tail .

2. **Heavy tails:** The heavy tailed and peaked of asset returns is most probably the most prevalent stylized fact. Fama (1965) observed the leptokurtic nature of assets returns using 30 stocks that were at the time included in the Dow Jones Industrial Average returns using daily data from 1957 to 1962. Leptokurtosis reportedly becomes stronger as the sampling frequency increases.

3. **Aggregational Normality:** The shape of the distribution of asset returns tends to normal distribution as the sampling time span increases. Dacorogna *et al.* (2001) reports that as the sampling time span decreases, the distribution of asset returns become fat-tailed with the second moment probably finite but the fourth moment tends to diverge.

4. **Absence of autocorrelations:** Usually linear autocorrelations are often insignificant for sampling time spans greater than 20 minutes. For sampling frequencies less than 20 minutes microstructure effects are assumed into come in play.

5. **Intermittency:** For any time scale asset returns display high variability.

6. **Volatility clustering:** Return series seem uncorrelated whilst the autocorrelation of squared returns are significant, positive and slowly decaying. This quantifies the fact that large price movements are clustered in time, small price movements also seem to follow small price movements.

7. **Conditional heavy tails:** Residuals from a filtered return process using GARCH or somesuch models still exhibit heavy tails even though the tails are less heavy compared to those of the original return series.

8. **Leverage effects:** When the equity value of a company increases its leverage decreases resulting in lower volatility. However, a decrease in a company's equity value results in higher leverage which results in higher volatility. This phenomenon that volatility measures and returns are negatively correlated was first identified by Black (1976).

9. **Slow decay of autocorrelation in absolute returns:** The autocorrelation of absolute returns slowly decays as a function of time lag.

10. **Volume/volatility correlation:** All measures of volatility seem to be positively correlated with trading volume.

11. **Asymmetry of time scales:** Historical coarse-grained measures of volatility predicts better realized measures of volatility than the other way round.

Whilst Cont and Tankov (2004) state that returns are not correlated, Dacorogna *et al.* (2001) reports that returns are often negatively correlated at the first lag for highest sampling frequencies. When the price formation process is over the correlations are expected to disappear. MTN and AGL log returns do not have significant negative autocorrelations at lag 1 whilst NPN and SHP log returns have significant autocorrelations at lag 1 for both 5 and 10 minute log returns reaching as low as -0.42 for SHP returns see figure 2.3.

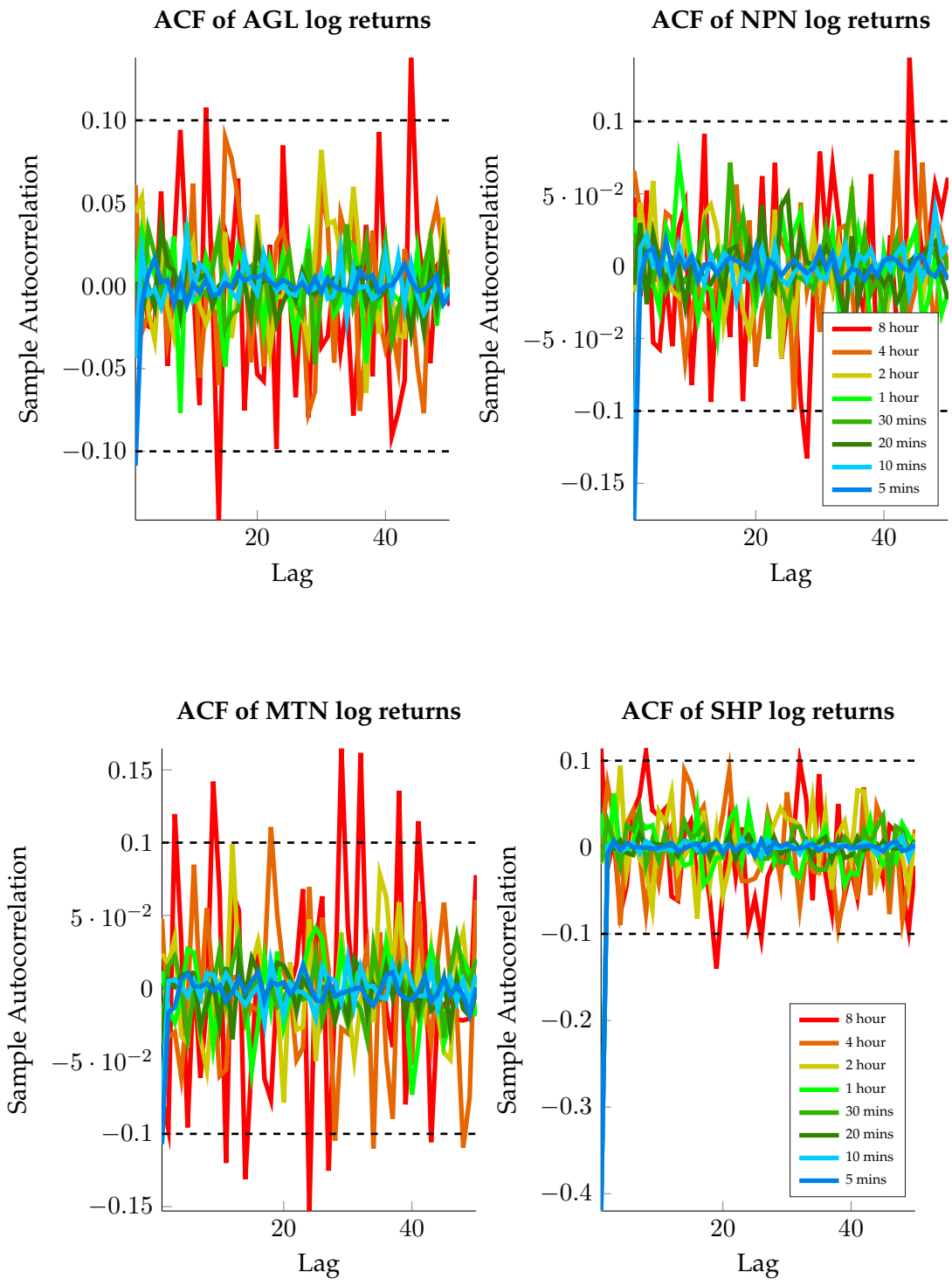


Fig. 2.3: Plots of the ACF of AGL, NPN, MTN and SHP log returns from 01 October 2012 to 30 September 2013.

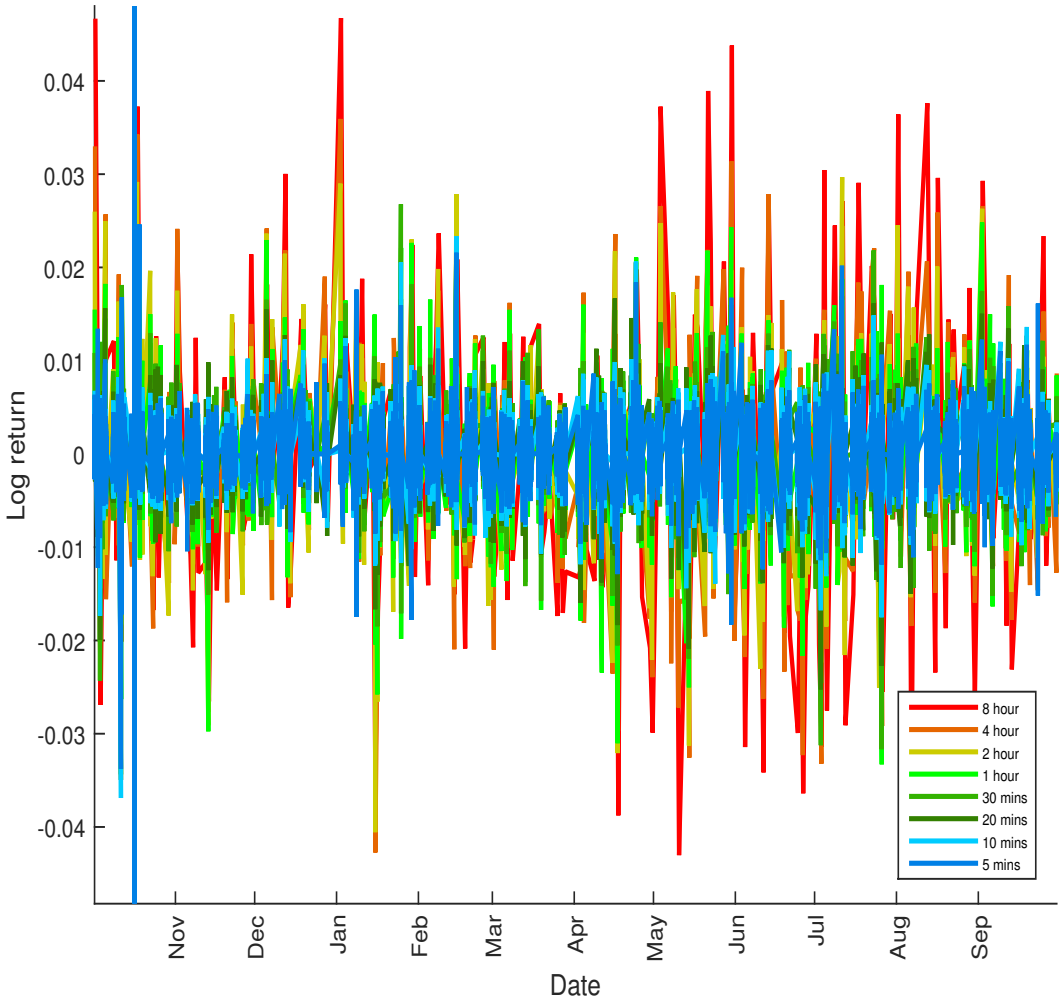


Fig. 2.4: Plots of the log returns of AGL from 01 October 2012 to 30 September 2013.

Figure 2.4 shows that for all sampling frequencies, large price movements are clustered in time. Table 2.1 and Table A.2 - A.4 shows that as the sampling time span decreases the distribution of log returns increasingly deviate from the normal distribution. Figure 2.5 shows that the cumulative frequencies of the log returns of AGL and MTN are S-shaped for all sampling frequencies as fat tailed distributions.

Interval	Sample size	Min	Max	Mean	Variance	Skewness	Kurtosis
8 hours	249	-0.0430	0.0467	-0.0000696	0.0161	0.1867	3.2827
4 hours	498	-0.0427	0.0359	-0.0000348	0.0111	-0.0480	3.8261
2 hours	996	-0.0406	0.0297	-0.0000174	0.0077	-0.0596	5.7063
1 hour	1992	-0.0333	0.02481	-0.0000087	0.0056	-0.2473	6.6447
30 minutes	3984	-0.0344	0.0268	-0.0000044	0.0041	-0.3011	9.1770
20 minutes	5976	-0.0350	0.0230	-0.0000029	0.0033	-0.2636	10.0204
10 minutes	11952	-0.0369	0.0234	-0.0000015	0.0024	-0.1381	13.7521
5 minutes	23904	-0.0482	0.0480	-0.0000007	0.0018	-0.0679	57.4069
2 minutes	59760	-0.0362	0.0479	-0.0000003	0.0012	0.5839	74.8507

Tab. 2.1: Summary statistics of AGL log returns from 01 October 2012 to 30 September 2013.

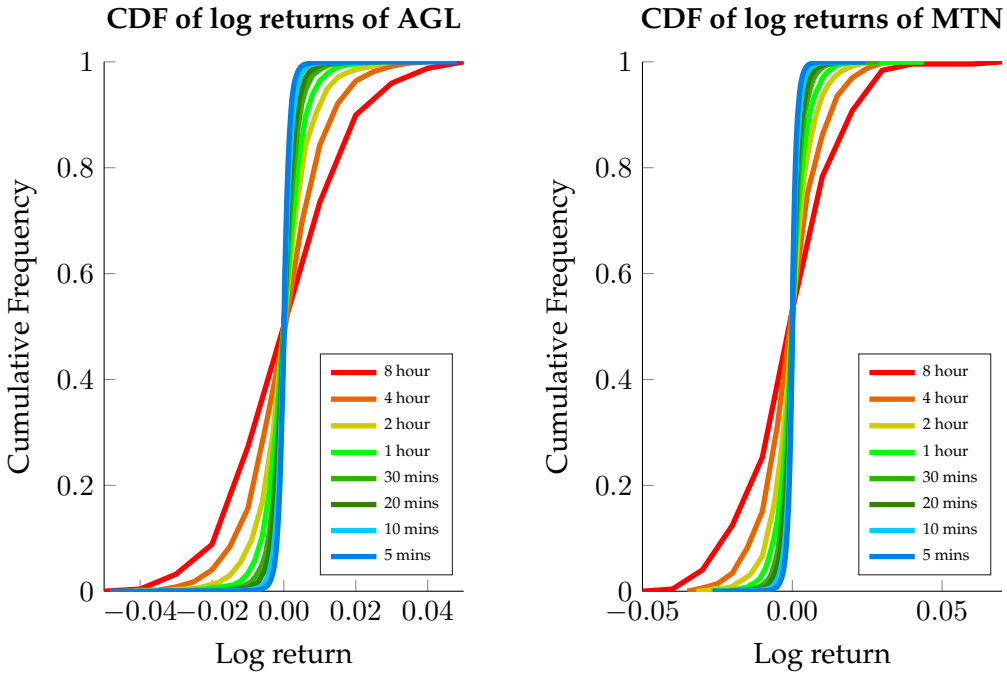


Fig. 2.5: Plots of the cumulative frequency of the AGL and MTN log returns from 01 October 2012 to 30 September 2013.

Log returns	Lévy processes
Absence of autocorrelations	True: Independent increments → no autocorrelations for all Lévy processes.
Heavy or semi-heavy tails	Possible, can be reproduced by choosing a Lévy measure with heavy/semi-heavy tails.
Price trajectories can have jumps	True: continuity only in probability → trajectories can have jumps.
Finite variance	True for Lévy processes with finite second moments.
Aggregational normality for log prices	True for Lévy processes with finite second moments.
Skewed distribution	Possible by choosing an asymmetric distribution/measure.
Volatility clustering (clustering of large events)	Not true (large events occur at random independent times)
Positive autocorrelation in absolute returns	Not true: Independent of increments imply no autocorrelation for returns or any power of returns.
"Leverage" effect: $\mathrm{Cov}(r_t^2 r_{t+\Delta t}) < 0$	Not true: independent of increments

Tab. 2.2: Comparison of the statistical properties of Lévy processes and empirical stylised facts of asset returns

Table 2.2 was adopted from Cont and Tankov (2004) and the table shows a comparison of stylized facts of log returns with the properties of Lévy properties. Although Lévy processes can model asset returns better than models based on the normal distribution, the drawback of using Lévy processes is that the assumption of independent increments will not allow them to reproduce volatility clustering, leverage effects as well as long memory.

Numerous studies have been conducted to investigate the potential presence of long range dependence (or long memory) asset returns. Many of these studies reported little evidence of long range dependence. Studies that reported evidence of long memory for asset returns include Christodoulou-Volos and Siokis (2006) and Heyde and Liu (2001).

For our purposes we shall say a stationary process is said to exhibit long

range dependence if the sum of the autocovariance process diverges i.e. $\sum_{k=1}^{\infty} \gamma_k = \infty$, for integers $k \geq 1$ with positive autocovariances $\{\gamma_k\}$. The long memory property can be difficult to quantify as it is depended upon the behavior of autocorrelation function at large lags. Models that exhibit long memory are usually formulated in terms of self similarity because of the quantification problem. The long time behavior of a self-similar process can be extrapolated from the short time behavior allowing one to quantify long memory using readily observable variables.

The connection between long memory in a stationary time series and self similar processes was given by Mandelbrot and Van Ness (1968). A typical example of a self similar process whose increments exhibit long memory is fractional Brownian motion. Note that self-similarity does not imply long memory nor does long memory imply self-similarity.

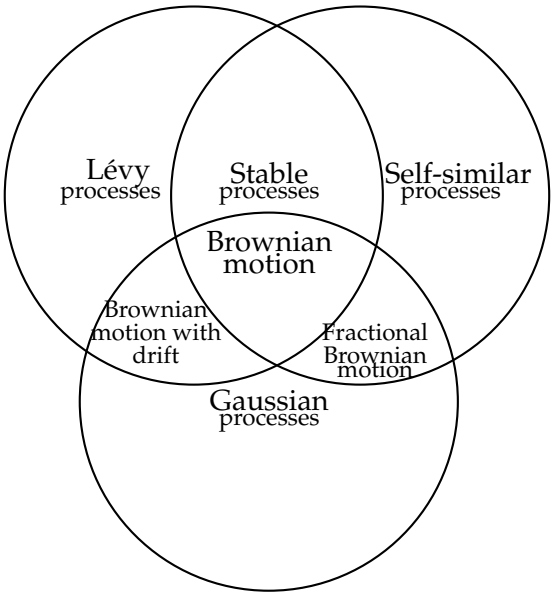


Fig. 2.6: Lévy processes and their relation to self-similar processes and Gaussian processes.

In the absence of high variability, self similarity can arise from strong dependence between increments (e.g. fractional Brownian motion), this effect is referred to as the "Joseph effect" . Self similarity can also arise from high variability in situations when the increments are heavy tailed and independent (e.g stable Lévy process), this is referred to as the "Noah effect". Cont and Tankov (2004) states that the two effects can be mixed to construct pro-

cess that exhibit both long memory and heavy tails.

2.3 Test for independent increments

Since Lévy processes assume that the increments are independent, the variance ratio test (VR) was used to check for independent increments of the log prices. Plots of autocorrelation function (ACF) were used to check if there were any significant autocorrelations of the log return process, see figure 2.3 .

2.3.1 The Variance Ratio test

The variance ratio test uses a comparison of variances of time series at different sampling time intervals. The variance of the increments of a random walk are directly proportional to the length of the time interval, for example the variance of hourly returns is expected to be six times the variance of the corresponding 10 minutes returns. Let X_t denote a stochastic process that satisfies the recursive relationship,

$$X_t = X_{t-1} + \mu + \varepsilon_t, \quad \mathbb{E}[\varepsilon_t] = 0 \quad \forall t, \tag{2.4}$$

where μ is an arbitrary drift parameter. For a random walk, the residuals ε_t should not be predictable from past values and should not be correlated.

Let q (an integer greater than 1) denote the number of overlapping time periods and n be an integer greater than 1, satisfying the equation $T = nq + 1$. Given a time series $(X_0, \ldots, X_T)$ of log prices of a stock, denote the return at time t to be $r_t = X_t - X_{t-1}$. The variance ratio test statistic is defined to be

$$VR(q) = \frac{\bar{\sigma}_b^2}{\bar{\sigma}_a^2}, \tag{2.5}$$

where

$$\hat{\mu} = (X_{nq} - X_0), \qquad \bar{\sigma}_a^2 = \frac{1}{nq-1}\sum_{k=1}^{nq}(r_k - \hat{\mu})^2, \tag{2.6a}$$

$$m = q(nq - q + 1)\left(1 - \frac{n}{nq},\right), \quad \hat{\sigma}_b^2 = \frac{1}{m}\sum_{k=q}^{nq}(X_k - X_{k-q} - q\hat{\mu})^2. \tag{2.6b}$$

The variance ratios are expected to be asymptotically close to one for a random walk, significantly greater than 1 if the time series is mean-averting (explosive) and less than 1 if the series is mean-reverting. To test whether the variance ratios significantly deviate from one enough to reject the random walk hypothesis, Lo and MacKinlay (1988) derived a test static that is asymptotically normal, the test statistic is given by the following equation

$$z(q) = \frac{\sqrt{nq}(\mathrm{VR}(q) - 1)}{\sqrt{\sum_{j=1}^{q} \left(\frac{2(q-j)}{q}\right)^2 \hat{\delta}(j)}}, \quad \text{where} \tag{2.7}$$

$$\hat{\delta}(j) = \frac{nq \sum_{k=j+1}^{nq} (X_k - X_{k-1} - \hat{\mu})^2 (X_{k-j} - X_{k-j-1} - \hat{\mu})^2}{\left(\sum_{k=1}^{nq} (X_k - X_{k-1} - \hat{\mu})^2\right)^2}.$$

Wilcox and Gebbie (2008) used the variance ratio test to test the random walk hypothesis of daily returns from January 1993 to December 2002 for stocks listed on the JSE. Wilcox and Gebbie (2008) grouped the data into six subsets and for each subset they computed correlation matrix of the data. For each correlation matrix, Wilcox and Gebbie (2008) ordered the eigenvalues from largest to smallest and derived the corresponding time series

$$x^{(k)}(t) = \sum_{i=1}^{N} u_i^{(k)} r_i(t), \tag{2.8}$$

where $u^{(k)}$ denotes the corresponding k-th eigenvalue and $u_i^{(k)}$ are its components, these are referred to as eigenmodes. Wilcox and Gebbie (2008) grouped the data into two cases, in particular case A did not have any missing returns and case B excluded repeated returns when there were missing returns or no trading took place. For $q = 2$, Wilcox and Gebbie (2008) report that for case B only the eigenmode that corresponds to the largest eigenvalue exhibits serial autocorrelation whilst the 4 leading eigenmodes exhibits positive serial autocorrelation and the remaining eigenmodes exhibit slight negative autocorrelations. Wilcox and Gebbie (2008) rejected the random walk hypothesis for the leading and second eigenmodes for case A and for the leading eigenmode for case B.

Sampling Frequency	q	2	4	8	16	32	64
Daily	VR(q)	0.9326	0.9617	0.9729	1.1189	1.0385	1.3184
	z(q)	1.1589	-0.3356	-0.1452	0.4146	0.0889	0.5046
	p value	0.2465	0.7372	0.8846	0.6784	0.9292	0.6139
	Decision	NR	NR	NR	NR	NR	NR
4 hours	VR(q)	1.0416	1.0355	1.0418	1.0322	1.2096	1.1029
	z(q)	0.9593	0.4301	0.3185	0.1637	0.7132	0.2365
	p value	0.3374	0.6671	0.7501	0.8699	0.4757	0.8130
	Decision	NR	NR	NR	NR	NR	NR
2 hours	VR(q)	1.0310	1.0975	1.0913	1.0995	1.0798	1.2493
	z(q)	1.1980	1.7648	0.9809	0.7152	0.3965	0.8517
	p value	0.2309	0.0776	0.3266	0.4745	0.6918	0.3944
	Decison	NR	NR	NR	NR	NR	NR
1 hour	VR(q)	0.9962	1.0383	1.0760	1.0555	1.0590	1.0295
	z(q)	-0.1443	0.8172	1.0866	0.5422	0.4027	0.1423
	p value	0.8852	0.4138	0.2772	0.5877	0.6872	0.8868
	Decison	NR	NR	NR	NR	NR	NR
30 mins	VR(q)	0.9769	0.9624	0.9974	1.0302	1.0153	1.0182
	z(q)	-1.1896	-1.0353	-0.0473	0.3950	0.1415	0.1191
	p value	0.2342	0.3005	0.9623	0.6928	0.8875	0.9052
	Decison	NR	NR	NR	NR	NR	NR
20 mins	VR(q)	1.0018	0.9853	0.9986	1.0346	1.0455	1.0327
	z(q)	0.0970	-0.4468	-0.0285	0.5165	0.5003	0.2577
	p value	0.9227	0.6550	0.9773	0.6055	0.6169	0.7966
	Decison	NR	NR	NR	NR	NR	NR
10 mins	VR(q)	0.9558	0.9514	0.9329	0.9456	0.9799	0.9904
	z(q)	-3.6002	-2.1710	-1.9497	-1.1053	-0.3014	-0.1052
	p value	0.0003	0.0299	0.0512	0.2690	0.7631	0.9162
	Decison	R	R	NR	NR	NR	NR
5 mins	VR(q)	0.8909	0.8301	0.8199	0.7985	0.8058	0.8339
	z(q)	-3.5005	-3.5638	-3.1200	-3.0652	-2.5858	-1.8986
	p value	0.0005	0.0004	0.0018	0.0022	0.0097	0.0576
	Decison	R	R	R	R	R	NR
2 mins	VR(q)	0.8886	0.8072	0.7575	0.7450	0.7350	0.7303
	z(q)	-8.1525	-8.1049	-7.4495	-6.4480	-5.6891	-4.8239
	p value	$\ll 0$	$\ll 0$	$\ll 0$	$\ll 0$	0.0013	0.1408
	Decison	R	R	R	R	R	R

The decision tells whether the random walk hypothesis is rejected(R) or failed to reject (NR) at 5% significance .

Tab. 2.3: The results of variance ratio test for AGL log returns from 01 October 2012 to 30 September 2013.

2.4 Test for stationary increments

It is crucial to understand or at least think abut the dynamical structure of the data before building a model to describe it. Lévy process assume that increments of their trajectories are stationary. The Augmented Dickey-Fuller (ADF) test and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test were used to test stationarity of log returns. Unit root tests such as the ADF test, test for the presence of a unit root whilst stationarity tests such as the KPSS test, test for trend stationarity.

The difference between the hypothesis of unit root tests and the stationarity tests are as follows

ADF test

- H_0 : Series has a unit root.
- H_1 : Series does not have a unit root implying that the series is either stationary or trend stationary.

KPSS test

- H_0 : Series is trend stationary.
- H_1 : Series has a unit root.

ADF decision	KPSS decision	Implication
Fail to reject H_0	Reject H_0	Series has a unit root
Reject H_0	Fail to reject H_0	Series is stationary
Fail to reject H_0	Fail to reject H_0	Number of observations not enough to make a conclusion
Reject H_0	Reject H_0	No conclusion can be made (series may be heteroskedastic or contain structural breaks)

Tab. 2.4: Summary showing how the ADF and KPSS tests complement each other

2.4.1 The Augmented Dickey-Fuller test

The ADF test was developed by Said and Dickey (1984) to address the shortcomings of the Dickey-Fuller test.

The Dickey-Fuller (DF) test

Consider an autoregressive variable, that follows the AR(1) model

$$x_t = \alpha x_{t-1} + \varepsilon_t, \quad \text{where} \quad \varepsilon_t \sim N(0, \sigma^2).$$

If $\alpha \geq 1$, the time series will be non-stationary and a value $|\alpha| < 1$, implies that the time series is stationary. The hypothesis of DF test are

$$H_0 : \alpha = 1,$$
$$H_1 : |\alpha| < 1.$$

The DF test is one sided, denote the least squares estimate by $\hat{\alpha}$ and the standard error by SE($\hat{\alpha}$), the DF statistic is a t-statistic that is given by

$$t_{\alpha=1} = \frac{\hat{\alpha} - 1}{\text{SE}(\hat{\alpha})}.$$

The dynamical structure of a financial time series are complicated and cannot be represented by an AR(1) model. If the residuals are serially autocorrelated then the results of the DF test will be biased, this reason led to the development of the ADF test. The ADF test uses a general ARMA(p,q) model of unknown orders unlike the DF test which uses an AR(1) model. The hypothesis and test statistic of the ADF test are the same to those of the DF test, the only exception is the form autoregressive variable. The form of the autoregressive variable for the ADF test is given by

$$x_t = \mathbf{\Psi}' \mathbf{Y}_t + \alpha y_{t-1} + \sum_{k=1}^{p} \beta_k \Delta x_{t-k} + \varepsilon_t, \tag{2.9}$$

where Y_t is a vector of deterministic terms and p is set such that the error term (ε_t) is serially uncorrelated.

2.4.2 The KPSS test

The KPSS test was introduced by Kwiatkowski *et al.* (1992), the test assumes that the time series $(x_t, t = 1, \ldots, T)$ can be written as a sum of a random walk, deterministic trend and a stationarity error,

$$x_t = r_t + \mu t + \varepsilon_t, \text{where} \tag{2.10}$$
$$r_t = r_{t-1} + u_t, \quad u_t \sim i.i.d(0, \sigma_u^2). \tag{2.11}$$

The initial value (r_0) is treated as a constant, since the variance of a random walk is expected to be 0, and ε_t is assumed to be stationary then x_t is trend stationary under the null hypothesis.

Interval	ADF test		KPSS test		Stationary
	Statistic	Decision	Statistic	Decision	Increments ?
2 minutes	-273.3574	Reject H_0	0.0564	Fail to reject H_0	Yes
5 minutes	-172.4640	Reject H_0	0.0617	Fail to reject H_0	Yes
10 minutes	-114.2578	Reject H_0	0.0725	Fail to reject H_0	Yes
20 minutes	-77.1107	Reject H_0	0.0768	Fail to reject H_0	Yes
30 minutes	-64.4991	Reject H_0	0.0753	Fail to reject H_0	Yes
1 hour	-44.5992	Reject H_0	0.0787	Fail to reject H_0	Yes
2 hours	-30.3357	Reject H_0	0.0829	Fail to reject H_0	Yes
4 hours	-21.1266	Reject H_0	0.0796	Fail to reject H_0	Yes
8 hours	-16.6298	Reject H_0	0.0755	Fail to reject H_0	Yes

Tab. 2.5: ADF and KPSS test for AGL log returns from 01 October 2012 to 30 September 2013.

Interval	ADF test		KPSS test		Stationary
	Statistic	Decision	Statistic	Decision	Increments ?
2 minutes	-415.1031	Reject H_0	0.0024	Fail to reject H_0	Yes
5 minutes	-172.1401	Reject H_0	0.0707	Fail to reject H_0	Yes
10 minutes	-118.6459	Reject H_0	0.0791	Fail to reject H_0	Yes
20 minutes	-80.3545	Reject H_0	0.0865	Fail to reject H_0	Yes
30 minutes	-64.6226	Reject H_0	0.0902	Fail to reject H_0	Yes
1 hour	-44.0528	Reject H_0	0.0922	Fail to reject H_0	Yes
2 hours	-30.7373	Reject H_0	0.0981	Fail to reject H_0	Yes
4 hours	-21.1676	Reject H_0	0.0949	Fail to reject H_0	Yes
8 hours	-16.8326	Reject H_0	0.0816	Fail to reject H_0	Yes

Tab. 2.6: ADF and KPSS test for MTN log returns from 01 October 2012 to 30 September 2013.

Chapter 3

Properties of Lévy Processes

Lévy processes are named after the French mathematician Paul Lévy (1886-1971) who characterized their distribution (Lévy-Khintchine formula) and described their structure (Lévy-Itô decomposition). The Lévy-Khintchine formula and the Lévy-Itô decomposition are theorems and will be discussed in detail in section 3.4. Lévy increments are independent and stationary. Lévy processes assume continuity in probability allowing them to model processes that exhibit jumps for example stock prices. Brownian motion is the only Lévy process that has continuous trajectories whilst compound Poisson process is the only Lévy process that is piecewise constant.

This chapter first discusses the mathematical concepts that will be used in the construction of Lévy process in section 3.1, followed by a discussion of Poisson processes in section 3.2. A discussion of square integrable martingales is given in section 3.4.1 and a discussion of Brownian motion follows in section 3.3. Compound Poisson processes, Brownian motion and square integrable martingales are all building blocks used in the construction of general Lévy processes. The general structure of Lévy process is given in section 3.5. The aim of this chapter is to give a basic introduction to Lévy processes, hence most statements will not be proved. For in depth discussions on Lévy process the reader is referred to Cont and Tankov (2004) and Sato (1999).

3.1 Mathematical tools

An understanding of probability spaces is crucial to the discussion of Lévy process given in this project, hence a brief discussion of probability spaces is given below. The concept of a probability space was introduced in the

1930s by Andrew Kolmogorov, a Russian mathematician.

Definition 3.1. σ-algebra
Let Ω be a nonempty set, and let $\mathscr{F}$ be a collection of subsets of Ω. We say that $\mathscr{F}$ is a σ-algebra provided that:

i the empty set $\emptyset$ to $\mathscr{F}$,

ii whenever a set A belongs to $\mathscr{F}$, its complement A^c also belongs to $\mathscr{F}$, and

iii whenever a sequence of sets A_1, A_2, ... belongs to $\mathscr{F}$, their union $\cup_{\infty}^{n=1} A_n$ also belongs to $\mathscr{F}$.

Definition 3.2. Probability Space
Let Ω be a nonepty set, and let $\mathscr{F}$ be a σ-algebra of subsets of Ω. A probability measure $\mathbb{P}$ is a function that, to every set $A \in \mathscr{F}$, assigns a number in $[0, 1]$, called the probability of A and written $\mathbb{P}(A)$. We require

i $\mathbb{P}(\Omega) = 1$, and

ii (countable additivity) whenever $A_1, A_2, ...$ is a sequence of disjoint sets in $\mathscr{F}$, then

$$\mathbb{P}\left(\cup_{\infty}^{n=1} A_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(A_n). \tag{3.1}$$

The triple (Ω, $\mathscr{F}$, $\mathbb{P}$) is called a probabiliy space.

Single execution of a model must result only in one outcome. Let E be the set of all numerical values that can be taken by the random variable X. Some random variables have a one to one map with their numerical values, for example when one rolls a die the set of possible outcomes is $\Omega = \{1, 2, 3, 4, 5, 6\}$ and $X(\omega) = \omega$. However, stock price models are complicated. Knowing only the stock price at time T cannot help one to pin down the exact events that occurred to give the result.

Definition 3.3. Lévy Process
A **cadlag**[1] stochastic process $(\mathbf{X_t})_{t\geq 0}$ on $(\Omega, \mathbb{F}, \mathbb{P})$ with values in $\mathbb{R}^d$, $d \geq 1$ is called a Lévy process if it exhibits the following properties

[1] A function $f : A \to B$ is a **cadlag** function if $\forall x \in A$

- $f(x_-) := lim_{x\uparrow} f(x)$ i.e. the left limit exists and,
- $f(x_+) := lim_{x\downarrow} f(x)$ exists and equals to f(x)

i **Zero initial value:** $\mathbf{X_0} = 0$

ii **Independent increments:** $\mathbf{X_{t_j}} - \mathbf{X_{t_{j-1}}}, j = 1, \ldots, n$ are independent for all $0 \leq t_0 < \ldots t_n$

iii **Stationary increments:** the distribution of $\mathbf{X_{t+s}} - \mathbf{X_t}$ does not depend on t

iv **Stochastic continuity:** $\forall \epsilon > 0, \lim_{h \to 0} \mathbb{P}(|\mathbf{X_{t+h}} - \mathbf{X_t}| \geq \epsilon) = 0$

Definition 3.4. Infinite divisibility
The probability distribution of a random variable X is *infinitely divisible* if $\forall\, n \in \mathbb{N}$ there exists n *i.i.d.* random variables $X_1, \ldots, X_n$ such that $X = X_1 + \ldots + X_n$.

Examples of infinitely divisible distributions on $\mathbb{R}^d$ include the Gaussian and Cauchy distributions. The exponential, Poisson and geometric distributions are infinitely divisible on $\mathbb{R}$, whilst the binomial and uniform distributions are not infinitely divisible.

Example 3.5. Let X_t be a Lévy process on $\mathbb{R}^d$, for any $n = 1, 2, \ldots$, X_t can be represented as

$$X_t = X_{\frac{t}{n}} + \left(X_{\frac{2t}{n}} - X_{\frac{t}{n}}\right) + \ldots + \left(X_t - X_{\frac{(n-1)t}{n}}\right),$$

showing that for every t, the distribution of X_t is infinitely divisible.

Proposition 3.6. *Let $(X_t)_{t \geq 0}$ be a Lévy process. Then for every t, X_t has an infinitely divisible distribution. Conversely, if F is an infinitely divisible distribution then there exists a Lévy process (X_t) such that the distribution of X_1 is given by F.*

The proof of the direct statement is given by example 3.5 and see Sato (1999), Corollary 11.6 for proof of the converse statement. The direct implication of proposition 3.6 is that every Lévy process has an associated infinitely divisible distribution and every infinitely divisible distribution has an associated Lévy process.

Definition 3.7. Lévy measure
Let $(X_t)_{t \geq 0}$ be a Lévy process on $\mathbb{R}^d$, $d \geq 1$. For any measurable set, $A \subseteq \mathbb{R}^d$, the measure ν defined on $\mathbb{R}^d$ defined by:

$$\nu(A) = \mathbb{E}[\#\{t \in [0,1] : \Delta X_t \neq 0,\ \Delta X_t \in A\}], \quad A \in \mathscr{B}(\mathbb{R}^d) \tag{3.2}$$

is called the Lévy measure of X, $\nu(A)$ is the expected number per unit time of jumps whose size belong to A.

The jump behavior of a discontinuous Lévy process is determined by the Lévy measure.

3.2 Poisson and Compound Poisson Processes

A Poisson process is basically a process that counts the number of events that occur in a given time period. Poisson processes are used to model arrival times of an object in a system and can also be used to model departure times from the system. Poisson processes have discontinuous trajectories and are often used as building blocks for more complicated process. A Compound Poisson process has arbitrary jumps unlike a Poisson process whose jump size is always one. Jump size in financial market models need to be of an arbitrary size, Compound Poisson processes, a generalisation of Poisson processes are used as building blocks to construct general Lévy processes.

Definition 3.8. Poisson random variable
A discrete random variable X is said to have Poisson distribution with parameter $\lambda > 0$ if the probability mass function of X is given by

$$f(k;\lambda) = \mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad \text{for } k = 0, 1, 2, \ldots. \tag{3.3}$$

Definition 3.9. Poisson process
A Poisson process is a stochastic process N_t that satisfies the following

i $N_0 = 0$ almost surely.

ii $N_t - N_s$ is Poisson distributed with parameter $(t-s)\lambda$ for $s < t$.

iii $N_{t_j} - N_{t_{j-1}}$ are independent for all $0 \leq t_0 < \ldots < t_n$.

iv N_t has left limits and almost surely right continuous.

Definition 3.10. Compound Poisson process
Let $(X_i)_{i\geq 1}$ be a sequence of $i.i.d.$ random variables independent to the Poisson process N_t, we define the compound Poisson process as

$$Q_t = \sum_{i=1}^{N_t} X_i, \;\; t \geq 0. \tag{3.4}$$

The jumps of Q_t occur at the same times as the jumps of N_t but the jumps of Q_t are of random size unlike the jumps of N_t that are always of size 1.

Proposition 3.11. *The characteristic function of a compound Poisson process Q_t on $\mathbb{R}^d$ has the following representation*

$$\mathbb{E}[exp^{iu.Q_t}] = exp\{t\lambda \int_{\mathbb{R}^d} (e^{iu.q} - 1) f(dx)\}, \quad \forall\, u \in \mathbb{R}^d, \tag{3.5}$$

where f denotes the jump size distribution and λ the jump intensity.

Proof. Denote the characteristic function of f by $\hat{f}$ and using iterated conditioning on Q_t we have

$$\begin{aligned}
\mathbb{E}[e^{iu.Q_t}] &= \mathbb{E}[\mathbb{E}[e^{iu.Q_t}]|Q_t] \\
&= \mathbb{E}[(\hat{f}(u))^{Q_t}] \\
&= \sum_{q=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^q (\hat{f}(u))^q}{q!} \\
&= e^{\lambda t(\hat{f}(u)-1)} \\
&= \exp\{t\lambda \int_{\mathbb{R}^d} (e^{iu.q} - 1) f(dx)\}.
\end{aligned} \tag{3.6}$$

□

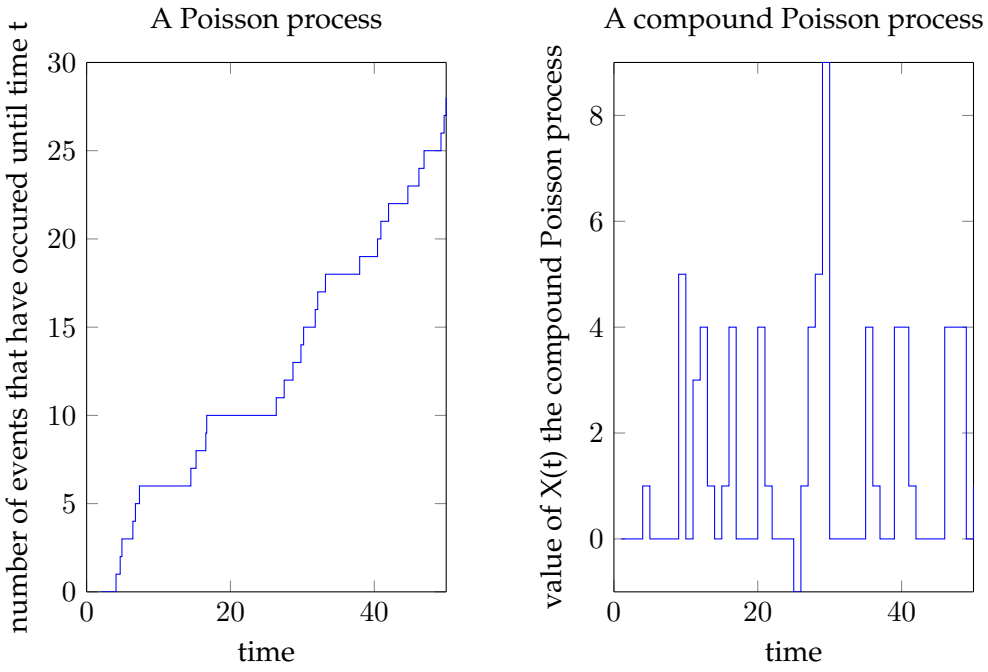


Fig. 3.1: Sample plots of Poisson process and Compound Poisson process where the rate of occurrence of an event $\lambda = 0.7$.

3.3 Brownian Motion

Brownian motion, also known as Wiener process are named after the botanist Robert Brown who noted the phenomenon whilst looking through a microscope in 1827 and Nobert Wiener who mathematically formalized the process. Brownian motion trajectories are continuous everywhere but their paths are nowhere differentiable.

Definition 3.12. Brownian Motion
A stochastic process $W = (W_t)_{t\geq 0}$, on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ is called a standard *Brownian motion* or a *Wiener process* if

1. $W(0) = 0$, almost surely.
2. $W(t)$ has independent increments i.e. $W(t_j) - W(t_{j-1})$, $j = 1, \ldots, n$ are independent for all $0 \leq t_0 < t_1 < \ldots t_n$.
3. W has continuous paths i.e. $\forall\, \omega \in \Omega$, $W(t, \omega)$ is continuous .
4. W has Gaussian increments i.e. for all $0 \leq t_1 < t_2$, $W(t_2) - W(t_1) \sim N(0, t_2 - t_1)$.

Corollary 3.13. *Characteristic function of Brownian motion*
For every $\omega \in \Omega$, the path of a Brownian motion $W_t \sim N(\mu t, \sigma^2 t)$ has the characteristic function

$$\mathbb{E}[e^{i\gamma W_t}] = e^{t(i\mu\gamma - \frac{1}{2}\sigma^2\gamma^2)}. \tag{3.7}$$

Proof. The characteristic function of a Gaussian random variable $V \sim N(\mu, \sigma^2)$ is given by

$$\mathbb{E}[e^{i\gamma V}] = e^{i\gamma\mu - \frac{\sigma^2\gamma^2}{2}}. \tag{3.8}$$

Using Eqn. 3.8 we get the characteristic function of $W_t \sim N(\mu t, \sigma^2 t)$ to be

$$\mathbb{E}[e^{i\gamma W_t}] = e^{t(i\mu\gamma - \frac{1}{2}\sigma^2\gamma^2)}. \tag{3.9}$$

□

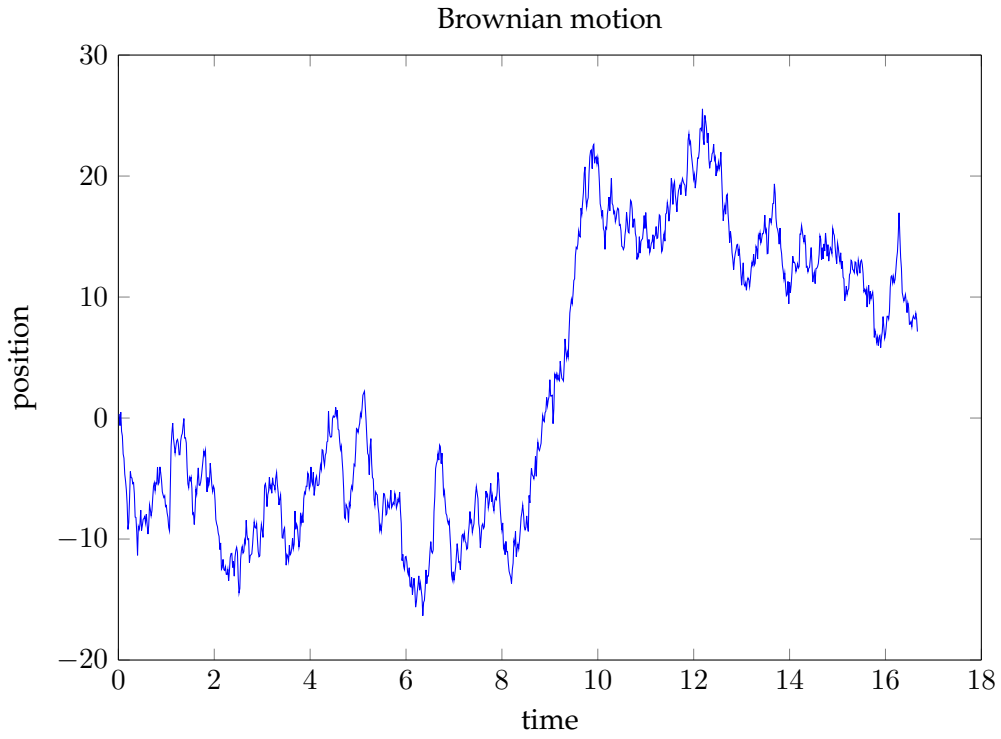


Fig. 3.2: Graph of sampled Brownian motion.

3.4 The general properties of Lévy process

The Lévy-Khintchine representation also known as the Lévy-Khintchine formula provides a useful representation of the characteristic exponent of all infinitely divisible distributions. Bruno de Finetti is considered to be the pioneer of infinitely divisible distributions, in the period 1929-1931 de Finetti published problems of finding the general form of characteristic functions for infinitely divisible distributions. Kolmogorov (1932) presented a special case of the representation with second moments. Lévy (1934) presented the general representation and Khintchine (1937) gave a much simpler proof than the one given by Lévy (1934).

Theorem 3.14. *Lévy-Khintchine formula*
Let $(X_t)_{t\geq 0}$ *be a Lévy process on* $\mathbb{R}^d$*, then*

$$\mathbb{E}[e^{i.zX_t}] = e^{t\Psi(z)},\ z \in \mathbb{R}^d \tag{3.10}$$

with

$$\Psi(z) = -\frac{1}{2}z'Az + i\gamma.z + \int_{\mathbb{R}^d}(e^{iz.x} - 1 - iz.x1_{|x|\leq 1})\nu(dx), \tag{3.11}$$

where $\gamma \in \mathbb{R}^d$, $A \in \mathbb{R}^{d\times d}$ is positive definite and $\nu(dx)$ is a positive random measure on $\mathbb{R} \setminus \{0\}$ that satisfies the conditions $\int_{\mathbb{R}}(1 \wedge x^2)\nu(dx) < \infty$.

Proof: See Theorem 8.1, Sato (1999).

The Lévy-Khintchine representation entails that every Lévy process can be characterized by a positive definite matrix A, a vector γ and ν a positive measure that uniquely determines distribution. The triplet, (A, ν, γ) is called the Lévy triplet or the characteristic triplet. It is important to note that from the Lévy-Khintchine formula the measure $\nu(dx)$ must be integrable near zero. Decomposing the Lévy-Khintchine formula, we notice that the formula can be seen as having three parts:

1. A linear deterministic part **(drift)** term controlled by the drift coefficient γ.
2. A Brownian motion **(diffusion)** part controlled by the matrix A.
3. A **pure jump** part controlled by the Lévy measure $\nu(dx)$, which dictates how the jumps occur.

Theorem 3.15. *Lévy-Itô Decomposition*
Given any $\sigma \geq 0$, $\gamma \in \mathbb{R}$ and a concentrated measure $\nu(dx)$ on $\mathbb{R}\setminus\{0\}$ satisfying

$$\int_{\mathbb{R}}(1 \wedge x^2)\nu(dx) < \infty, \tag{3.12}$$

then there exists a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ on which three independent Lévy processes $X^{(1)}$, $X^{(2)}$ and $X^{(3)}$ exist where

1. *$X^{(1)}$ is a Brownian with drift.*
2. *$X^{(2)}$ is a compound Poisson process.*
3. *$X^{(3)}$ is a square integrable pure jump martingale that almost surely has a countable number of jump on each finite interval.*

Taking $X = X^{(1)} + X^{(2)} + X^{(3)}$ we see that a probability space exists on which a Lévy process has a characteristic exponent

$$\Psi(z) = -\frac{1}{2}\sigma^2 z^2 + i\gamma z + \int_{\mathbb{R}}(e^{izx} - 1 - izx1_{|x|<1})\nu(dx), \tag{3.13}$$

for $z \in \mathbb{R}$.

The Lévy-Itô decomposition entails that the structure of a general Lévy process is described by three independent supplementary Lévy processes, each with different types of behavior. To gain an intuitive understanding of the Lévy-Itô decomposition we split the Eqn. 3.13 into three

$$\begin{aligned}\Psi(z) &= \Psi^{(1)}(z) + \Psi^{(2)}(z) + \Psi^{(3)}(z) \\ &= \left\{ -\frac{1}{2}\sigma^2 z^2 + i\gamma z \right\} \\ &+ \left\{ \nu(\mathbb{R}\backslash(-1,1)) \int_{|x|\geq 1} (e^{izx} - 1) \frac{\nu(dx)}{\nu(\mathbb{R}\backslash(-1,1))} \right\} \\ &+ \left\{ \int_{0<|x|<1} (e^{izx} - 1 - izx)\nu(dx) \right\}. \end{aligned} \tag{3.14}$$

The properties of compound Poisson processes and Brownian motion were discussed in section 3.2 and section 3.3 respectively. It is evident that $\Psi^{(1)}(z)$ corresponds to the characteristic function of Brownian motion with drift. The function $\Psi^{(2)}(z)$ corresponds to the characteristic function of a compound Poisson process whose random variates are concentrated on $\{x\colon |x| \geq 1\}$ with distribution $\frac{\nu(dx)}{\nu(\mathbb{R}\backslash(-1,1))}$. The process $X^{(2)}$ is identically zero when $\nu(\mathbb{R}\backslash(-1,1)) = 0$.

To prove that characteristic exponent of a general Lévy has the characteristic form given by Eqn. 3.13, we have to show that there exists a Lévy process $X^{(3)}$ with characteristic function $\Psi^{(3)}(z)$. By rewriting the characteristic function $\Psi^{(3)}(z)$ into the form given by the following equation it is established that the process $X^{(3)}$ is a superposition of (at most) countable compound Poisson processes with different arrival rates and additional linear drift.

$$\begin{aligned}&\int_{0<|x|<1} (e^{izx} - 1 - izx)\nu(dx) \\ &\qquad = \sum_{n\geq 0} \Big\{ \lambda_n \int_{2^{-(n+1)}\leq|x|<2^{-n}} (e^{izx} - 1) F_n(dx) \\ &\qquad\qquad - iz\lambda_n \left(\int_{2^{-(n+1)}\leq|x|<2^{-n}} x F_n(dx) \right) \Big\},\end{aligned}$$

where $\lambda_n = \nu\left(\left\{x\colon 2^{-(n+1)} \leq |x| < 2^{-n}\right\}\right)$ and

$$F_n(dx) = \lambda_n^{-1}\nu(dx)|_{\left\{x\colon 2^{-(n+1)}\leq|x|<2^{-n}\right\}}.$$

Note that if $\lambda_n = 0$, then the n^{th} integral is absent. To establish a rigorous mathematical understanding of the above superposition and consequently of the general structure of Lévy process, a brief discussion of square integrable martingales is provided below.

3.4.1 Square integrable martingales

For a fixed $T > 0$, let $\mathscr{F}_t^*\colon t \geq 0$ be a right continuous filtration in the sense that $\mathscr{F}_t^* = \cap_{s>t}\mathscr{F}_s^*$ is complete with respect to the null sets of $\mathbb{P}$. Assume that $(\Omega, \mathscr{F}, \{\mathscr{F}_t^*\colon t \in [0,T]\}, \mathbb{P})$ is a filtered probability space. A martingale M_t defined on this space is said to be square integrable if for every $t \geq 0$, $\mathbb{E}\left[M_t^2\right] < +\infty$.

Definition 3.16. Fix $T > 0$, for a finite interval $[0,T]$ define
$M_T^2 = M_T^2(\Omega, \mathscr{F}, \{\mathscr{F}_t^*\colon t \in [0,T]\}, \mathbb{P})$ to be the space of zero mean right continuous, square integrable $\mathbb{P}$ martingales that are real valued.

For each $n = 1, 2, 3, ...$, suppose that $N^{(n)} = \left\{N_t^{(n)}\colon t \geq 0\right\}$ is a Poisson process with rate $\lambda_n \geq 0$. If $\lambda_n = 0$ then the process $N_t^{(n)}$ is identically zero (i.e. $N_t^{(n)} = 0\,\forall t \geq 0$). Throughout this section it will be assumed that the processes $N^{(n)}$ are mutually independent.

For each $n = 1, 2, 3, ...$, we introduce the $i.i.d.$ sequences $\left\{\xi_i^{(n)}\colon i = 1, 2, 3, ...\right\}$ with common distribution F_n which satsfies $\int_{\mathbb{R}} x^2 F_n(dx) < \infty$ and which does not assign mass to the origin.

Lemma 3.17. *Suppose that* $\int_{\mathbb{R}} |x| F(dx) < \infty$.

1. *The process* $M = \{M_t\colon t \geq 0\}$ *where*

$$M_t :- \sum_{i=1}^{N_t} \xi_i - \lambda t \int_{\mathbb{R}} xF(dx) \tag{3.15}$$

   is a martingale with respect to its natural filtration.

2. *If moreover* $\int_{\mathbb{R}} x^2 F(dx) < \infty$ *then M is a square integrable martingale such that*

$$\mathbb{E}(M_t^2) = \lambda t \int_{\mathbb{R}} x^2 F(dx) \tag{3.16}$$

Proof. See Lemma 4.1 Kyprianou (2007).

From the discussion of the intuitive reasoning of Lévy-Itô decomposition one can see that it is imperative that the countably infinite number of such processes are superimposed together in such a way that the resulting sum converges in an appropriate sense. The following theorem precisely addresses this point.

Theorem 3.18. *If*

$$\sum_{n=1}^{\infty} \lambda_n \int_{\mathbb{R}} x^2 F_n(dx) < \infty \tag{3.17}$$

then there exists a Lévy process $X = \{X_t : t \geq 0\}$ that is a square integrable martingale on the same probability space as $\{M^{(n)} : n \geq 1\}$ which has a characteristic exponent of the form

$$\Psi(\gamma) = \int_{\mathbb{R}} (e^{i\gamma x} - 1 - i\gamma x) \sum_{n=1}^{\infty} \lambda_n F_n(dx). \tag{3.18}$$

Moreover for each $\gamma \in \mathbb{R}$ such that for each fixed $T > 0$ we have,

$$\lim_{k \uparrow \infty} \mathbb{E}[\sup_{t \leq T} (X_t - \sum_{n=1}^{k} M_t^{(n)})^2] = 0. \tag{3.19}$$

See Theorem 4.1, Kyprianou (2007) for proof. Note that the characteristic exponent of a Lévy processes in Kyprianou (2007) is the negative of the one presented in this chapter.

Proof. (Lévy-Itô Decomposition)
Re-organizing the form of the characteristic exponent of any infinitely divisible distribution described by the Lévy-Khintchine formula we get

$$\begin{aligned} \Psi(z) = \{-\frac{1}{2}\sigma^2 z^2 + i\gamma z\} + \int_{|x| \geq 1} (e^{izx} - 1)\nu(dx) \\ + \int_{0<|x|<1} (e^{izx} - 1 - izx1_{|x|<1})\nu(dx). \end{aligned} \tag{3.20}$$

The first two parts of the sum are the characteristic exponent of Brownian motion with drift and compound Poisson process on $\mathbb{R}\backslash 0$ respectively. To verify the existence of a square integrable pure jump martingale, let

$$\lambda_n = \nu(\{x : 2^{-(n+1)} \leq |x| < 2^{-n}\}) \quad \text{and}$$

$$F_n(dx) = \lambda_n^{-1} \nu(dx)|_{\{x : 2-(n+1) \leq |x| < 2^{-n}\}}.$$

Note that in particular

$$\sum_{n=1}^{\infty} \lambda_n F_n(dx) = \int_{(-1,1)\setminus\{0\}} x^2 \nu(dx) < \infty, \tag{3.21}$$

hence the assumptions of theorem 3.14 hold and over each finite time interval $X^{(3)}$ has a countable number of jumps that are less than unity in magnitude. □

3.5 Stochastic time change

The existence of jumps as well as variations in time spans between price changes in financial markets can be modeled using a time changed Lévy process. On a more active business day a trader's decision rate increases implying a faster business clock and equivalently, the decision rate of a trader is usually slow in times of market instability. This randomness in market activity generates randomness in the volatility. By time changing Brownian motion with an almost surely increasing Lévy process the randomness in business activity can be replicated. Krichene (2005) suggests that in-order to capture leverage effects, innovations in the Lévy process must be correlated with the innovations in the random clock on which it is run.

Proposition 4.1 Let $(S_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}$. The following conditions are equivalent:

i $S_t \geq 0$ almost surely for some $t > 0$.

ii $S_t \geq 0$ almost surely for every $t > 0$.

iii Sample paths of (S_t) are almost surely nondecreasing: $t \geq s \implies S_t \geq S_s$ almost surely.

iv The characteristic triplet of (S_t) satisfies $A = 0, b \geq 0,\ \rho((-\infty, 0)) = 0$ and $\int_0^\infty (x \wedge 1)\rho(dx) < \infty$, that is, S_t has positive drift, positive jumps of finite variation and no diffusion component.

See Cont and Tankov (2004), Proposition 3.10 for proof.

Definition 3.19. Subordinator
A Lévy process $(S_t)_{t\geq 0} \in \mathbb{R}$ is a subordinator[2] if and only if it satisfies one of the conditions of Proposition 4.1.

S_t can be described using Laplace transforms since it is a positive random variable. Let $(0, \rho, b)$ be the characteristic triplet of S, then the moment generating function of S_t is

$$\mathbb{E}(e^{uS_t}) = e^{tl(u)} \; \forall u \leq 0, \text{ where } l(u) = ub + \int_0^\infty (e^{ux} - 1)\rho(dx). \tag{3.22}$$

Theorem 3.20. *Subordination of Lévy process*
Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$. Let $(S_t)_{t\geq 0}$ be a subordinator with characteristic triplet $(0, \rho, b)$ and Laplace exponent l(u). Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with characteristic exponent $\Psi(u)$ and Lévy triplet (A, ν, γ). Suppose that (S_t) and (X_t) are independent, define

$$Y_t(\omega) = X_{S_t}(\omega), \quad \omega \in \Omega. \tag{3.23}$$

Then the process $(Y_t)_{t\geq 0}$ is a Lévy process on $\mathbb{R}^d$ with characteristic function

$$\mathbb{E}(e^{iuY_t}) = e^{tl(\Psi(u))}. \tag{3.24}$$

The Lévy triplet (A^Y, ν^Y, γ^Y) of $(Y_t)_{t\geq 0}$ is

$$A^Y = bA, \tag{3.25a}$$

$$\nu^Y(B) = b\nu(B) + \int_0^\infty p_s^X(B)\rho(ds), \; \forall B \in \mathscr{B}(\mathbb{R}^d), \tag{3.25b}$$

$$\gamma^Y = b\gamma + \int_0^\infty \rho(ds) \int_{|x|\leq 1} x p_s^X(dx). \tag{3.25c}$$

where p_t^X is the probability distribution of X_t.

See Cont and Tankov (2004), Theorem 4.2 for proof.

Theorem 3.20 entails that by taking the composition of the Laplace exponent of the subordinator S_t with the characteristic exponent of X_t one can obtain the characteristic function of the subordinated process Y_t.

Brownian motion has continuous paths hence its Lévy measure is zero also Brownian motion increments follow the lognormal distribution. The VG and NIG representations discussed in chapter 5 are those obtained by time changing Brownian motion with drift using the gamma and inverse Gaussian processes as subordinators respectively.

[2] A Lévy process that is increasing almost surely.

Definition 3.21. Multivariate normal density function
The multivariate normal density function of $\mathbf{x} \in \mathbb{R}^d$ that is normally distributed with mean $\boldsymbol{\gamma}$ and covariance matrix $\boldsymbol{\Sigma}$ is given by

$$f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\gamma}, \boldsymbol{\Sigma}) = \frac{\exp(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\gamma})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\gamma}))}{\sqrt{(2\pi)^k |\boldsymbol{\Sigma}|}}, \tag{3.26}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of $\boldsymbol{\Sigma}$.

Theorem 3.22. *Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ and let $(S_t(\omega))_{t\geq 0} \in \mathbb{R}$ be subordinator with Lévy triplet $(0, \rho, b)$. Let $X_t(\omega) = \sigma W_t(\omega) + \gamma t$ be a Brownian motion with drift on $\mathbb{R}^d$ that is independent to S_t. Then the process $(Y_t(\omega))_{t\geq 0}$ defined by*

$$Y_t(\omega) = X_{S_t}(\omega) = \sigma W_{S_t}(\omega) + \theta S_t(\omega), \tag{3.27}$$

is a Lévy process with Lévy triplet $(\Sigma^Y, \nu^Y, \theta^Y)$, where

$$\Sigma^Y = b\Sigma, \tag{3.28a}$$

$$\nu^Y(x) = \int_0^\infty f_{\boldsymbol{X}}(\boldsymbol{x}; s\gamma, s\Sigma)\rho(s)ds, \tag{3.28b}$$

$$\gamma^Y = b\gamma + \int_0^\infty \int_{|x|\leq 1} x f_{\boldsymbol{X}}(\boldsymbol{x}; s\gamma, s\Sigma)\rho(s)dxds. \tag{3.28c}$$

where $f_{\boldsymbol{X}}(\boldsymbol{x}; \theta, \Sigma)$is multivariate normal density. For all t, the characteristic function of Y_t, is given by

$$\mathbb{E}[e^{iuY_t}] = e^{tl(iu\gamma - \frac{\sigma^2 u^2}{2})}. \tag{3.29}$$

Theorem 3.22 is a direct result of theorem 3.20. From theorem 3.22 we can see that the assumption of normality of log returns is affected by time changing Brownian motion. The representation of Lévy processes using Brownian subordination adds tractability and makes the model easier to understand, however it imposes limitations on the form of the Lévy measure. From Eqn. 3.28, if the subordinator S_t is independent to the Brownian motion with drift X_t, then the two integrals can be evaluated separately. By introducing correlation between S_t and X_t one may introduce asymmetry in the distribution of Y_t.

Theorem 3.23. *Let $\mu \in \mathbb{R}$ and $\nu(dx)$ be a Lévy measure on $\mathbb{R}$. There exists a Lévy process $(X_t)_{t\geq 0}$ with measure $\nu(dx)$ such that*

$$X_t = W_{S_t} + \mu S_t,$$

for some subordinator $(S_t)_{t\geq 0}$ and some Brownian motion $(W_t)_{t\geq 0}$ independent from S if and only if the following conditions are satisfied

1. *ν is absolutely continuous with density $\nu(x)$.*
2. *$\nu(x)e^{-\mu x} = \nu(-x)e^{\mu x}$.*
3. *$\nu(\sqrt{\mu})e^{-\mu\sqrt{\mu}}$ is a completely monotonic function on $(0, \infty)$.*

See Theorem 4.3, Cont and Tankov (2004) for proof. Lévy processes that are represented as Brownian motion subordinated by an almost surely increasing Lévy process are discussed in detail in chapter 5.

Chapter 4

The multivariate generalized hyperbolic distribution

The generalized hyperbolic distribution (GHD) has several different parameterizations, this chapter focuses on the parametrization introduced by McNeil *et al.* (2005). The representation of the GHD by McNeil *et al.* (2005) is a normal variance-mean mixture with the generalized inverse Gaussian (GIG) as a mixing distribution. The parametrization introduced by McNeil *et al.* (2005) allows one to compute maximum likelihood estimates using the Expectation Maximization (EM) algorithm. The EM algorithm is a two step iterative procedure and the EM algorithm is more stable than the maximum likelihood estimation (MLE). Although the literature discussed in this chapter is that of the multivariate GHD, the fitting was only done to single stocks.

Section 4.1 discusses the GIG, the MLE of the GIG is also presented in this section. Boundary cases of the GIG are discussed in section 4.2, a stable algorithm to estimate the parameters of the boundary cases of the GIG developed by Luo and Zhao (2014) is also presented. The representation of the GHD by McNeil *et al.* (2005) is given in section 4.3 and a brief discussion of the subclasses of the GHD are also presented. Section 4.4 mentions the limiting cases of the GHD and a brief discussion of the tail behavior of the univariate GHD is presented in section 4.5. The EM algorithm of the GHD is discussed in depth in section 4.6. Section 4.7 mentions the measures used to evaluate goodness of fit in this project. The results obtained from fitting the GHD models to financial data from the JSE are presented in section 4.8.

4.1 The generalized inverse Gaussian distribution

The generalized inverse Gaussian (GIG) also known as the Sichel distribution after the statistician Herbert Sichel ($1915 - 1995$) who made great advances in the area of the GIG distribution. To construct Poisson distributions as a mixture of normals, Sichel (1974) used the GIG as the mixing distribution. Barndorff-Nielsen (1977) obtained the GHD as a mixture of normals using the GIG as the mixing distribution. Barndorff-Nielsen and Halgreen (1977) proved that GIG is an infinitely divisible distribution. Jørgensen (1982) contains an in depth discussion of the statistical properties of the GIG.

The probability density function GIG distribution is given by

$$f_{\text{GIG}}(w;\lambda,\chi,\psi) = \frac{\left(\sqrt{\psi/\chi}\right)^{\lambda}}{2K_{\lambda}(\sqrt{\chi\psi})} w^{\lambda-1} e^{-\frac{1}{2}(\chi w^{-1}+\psi w)},\ w>0, \tag{4.1}$$

where $K_{\lambda}(\sqrt{\chi\psi})$ given by Eqn. B.5, an equation is closely related to modified Bessel equation of the third kind given by Eqn. B.2.

Throughout the chapter, the symbol $N^{\vdash}(\lambda,\chi,\psi)$ is used for the GIG distribution. Let $(\lambda,\chi,\psi) \in \mathbb{R}^3$ and denote the parameter space of the GIG by Θ_{GIG}, the cases of the parameter space of the GIG distribution are given by

$$\Theta_{\text{GIG}}(\lambda,\chi,\psi) = \begin{cases} \textbf{Normal case:} & \lambda \in \mathbb{R}, \chi > 0, \psi > 0, \\ \textbf{Boundary case 1:} & \lambda > 0, \chi = 0, \psi > 0, \\ \textbf{Boundary case 2:} & \lambda < 0, \chi > 0, \psi = 0. \end{cases}$$

The GIG distribution is a generalization of many distributions that are used as mixing distributions in the construction of subclasses of the GHD. Below is a table with an overview of the parameter space for the subclasses of the GIG, for detailed discussion of the subclasses see Paolella (2007).

4.1.1 Moments of a GIG random variable

If $W \sim N^{\vdash}(\lambda,\chi,\psi)$, then the r-th raw moment of W is given by

$$\begin{aligned} \mathbb{E}[W^r] &= \frac{\chi^{-\lambda}(\chi\psi)^{\frac{\lambda}{2}}}{2K_{\lambda}(\sqrt{\chi\psi})} \int_0^{\infty} w^{\lambda+r-1} e^{-\frac{1}{2}(\chi w^{-1}+\psi w)}\, \mathrm{d}w \\ &= \left(\frac{\chi}{\psi}\right)^{\frac{r}{2}} \frac{K_{\lambda+r}(\sqrt{\chi\psi})}{K_{\lambda}(\sqrt{\chi\psi})}. \end{aligned} \tag{4.2}$$

Distribution name	Parameter Range		
Normal case	$\lambda \in \mathbb{R}$	$\chi > 0$	$\psi > 0$
Inverse Gamma	$\lambda < 0$	$\chi > 0$	$\psi = 0$
Gamma	$\lambda > 0$	$\chi = 0$	$\psi > 0$
Lévy	$\lambda = -1/2$	$\chi > 0$	$\psi = 0$
Inverse Gaussian	$\lambda = -1/2$	$\chi > 0$	$\psi > 0$
Exponential	$\lambda = 1$	$\chi = 0$	$\psi > 0$
Positive hyperbolic	$\lambda = 1$	$\chi > 0$	$\psi > 0$

Tab. 4.1: Subclasses of the GIG

Using Eqn. 4.2, the expectation of W is given by the following expression

$$\mathbb{E}[W] = (\frac{\chi}{\psi})^{\frac{1}{2}} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}, \tag{4.3}$$

the variance of W is given by

$$\begin{aligned} \text{Var(W)} &= \mathbb{E}[W^2] - (\mathbb{E}[W])^2 \\ &= \left(\frac{\chi}{\psi}\right) \frac{K_{\lambda+2}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} - \left((\frac{\chi}{\psi})^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}\right)^2. \end{aligned} \tag{4.4}$$

The expectation of the natural logarithm of W is derived in Appendix B and is given by the following equation

$$\mathbb{E}[\ln W] = \frac{1}{2}\ln\left(\frac{\chi}{\psi}\right) + \frac{\frac{K_\lambda(\sqrt{\chi\psi})}{\mathrm{d}\lambda}}{K_\lambda(\sqrt{\chi\psi})}. \tag{4.5}$$

4.1.2 MLE of the normal case

Jørgensen (1982) discussed in depth how to compute the maximum likelihood estimate of the GIG given n independent random variables. The algorithm presented in this section is based on theory in chapter 4 of Jørgensen (1982). The MLE for the GIG when either $\chi = 0$ or $\psi = 0$ is given under the discussion of the boundary cases of the GIG in section 4.2.

Suppose one is given $w_1, \ldots w_n$ independent random variables, let

$$\bar{w} = \frac{1}{n}\sum_{i=1}^{n} w_i, \quad \bar{w}_{-1} = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{w_i}, \quad \text{and} \quad \bar{w}_{\sim} = \frac{1}{n}\sum_{i=1}^{n} \log w_i. \tag{4.6}$$

The well known mathematical relationship between the arithmetic mean, geometric mean and harmonic mean is given by the following inequality

$$\bar{w} \geq \exp \bar{w}_{\sim} \geq \bar{w}_{-1}^{-1}, \tag{4.7}$$

hence

$$\bar{w} \cdot \bar{w}_{-1} \geq 1. \tag{4.8}$$

The equality in equations 4.7 and 4.8 occurs if and only if $w_1 = w_2 = \ldots = w_n$. Since $\chi > 0$ and $\psi > 0$ for the normal case of the GIG, define

$$\omega = \sqrt{\chi\psi} \quad \text{and} \quad \eta = \sqrt{\chi/\psi}. \tag{4.9}$$

Substituting η and ω into Eqn. 4.1 the probability density function of the GIG becomes

$$f_{\text{GIG}}(w; \lambda, \omega, \eta) = \frac{\eta}{2K_\lambda(\omega)} \left(\frac{w}{\eta}\right)^{\lambda-1} \exp\left(-\frac{\omega}{2}\left(\frac{w}{\eta} + \frac{\eta}{w}\right)\right), \quad w > 0, \tag{4.10}$$

where $K_\lambda(\omega)$ is given by Eqn. B.5. It follows from Eqn. 4.10 that the log likelihood of the GIG is

$$\begin{aligned}\mathscr{L}_{\text{GIG}}(\lambda, \eta, \omega) = n \times [&-\log(2K_\lambda(\omega)) + (\lambda - 1)\log(w) - \lambda\log(\eta)] \\ &- \frac{\omega n}{2}\left(\bar{w}^{-1}\eta + \frac{\bar{w}}{\eta}\right).\end{aligned} \tag{4.11}$$

To obtain the maximum likelihood estimators $(\hat{\lambda}, \hat{\chi}, \hat{\psi})$ of the GIG, first estimate $\hat{\chi}$ and $\hat{\psi}$ using a fixed λ, then estimate $\hat{\lambda}$ using $\hat{\chi}$ and $\hat{\psi}$.

Estimation of $(\hat{\chi}, \hat{\psi})$ when λ is fixed

Define the equations

$$D_\lambda(\omega) = \frac{K_{\lambda+1}(\omega)K_{\lambda-1}(\omega)}{K_\lambda^2(\omega)}, \quad R_\lambda = \frac{K_{\lambda+1}(\omega)}{K_\lambda(\omega)}, \tag{4.12}$$

where $K_\lambda(\omega)$ is given by Eqn. B.5. The relation between $R_\lambda(\omega)$ and $D_\lambda(\omega)$ is given by

$$D_\lambda(\omega) = R_\lambda(\omega)R_{-\lambda}(\omega). \tag{4.13}$$

The equations used in the maximum likelihood estimation are of the form

$$\bar{w} = R_\lambda(\omega)\eta \quad \text{and} \quad \bar{w}_{-1} = R_{-\lambda}(\omega)\eta^{-1}. \tag{4.14}$$

From the product and ratio of equation 4.14, we get the following alternative equations for the maximum likelihood

$$D_\lambda(\omega) = \bar{w} \cdot \bar{w}_{-1}, \tag{4.15}$$

$$\eta = \sqrt{\frac{\bar{w}}{\bar{w}_{-1}}}\sqrt{\frac{R_{-\lambda}(\omega)}{R_\lambda(\omega)}}. \tag{4.16}$$

If a solution exists for the likelihood equations, it may be found by first solving for ω in Eqn. 4.15 and then inserting ω into Eqn. 4.16 obtain η.

Jørgensen (1982) proved that the function $f(\omega) = R_\lambda(\omega) + R_{-\lambda}(\omega)$ is monotonically decreasing on $(0, \infty)$ and maps onto $(2, \infty)$. Hence for a fixed λ and η, the full GIG family with density given by Eqn. 4.10 has a unique maximum likelihood estimate $\hat{\omega}_{\lambda,\eta}$ which is the solution of the equation

$$R_\lambda(\omega) + R_{-\lambda}(\omega) = \eta\bar{w}_{-1} + \eta^{-1}\bar{w}. \tag{4.17}$$

Algorithm 1 Estimation of $(\hat{\chi}, \hat{\psi})$ for the GIG when λ is fixed

if $|\lambda| = \frac{1}{2}$ **then**
  $\hat{\omega} = \frac{1}{\bar{w}\cdot\bar{w}_{-1}-1}$,
  plug $\hat{\omega}$ into Eqn. 4.16 to get $\hat{\eta}$
  insert $\hat{\omega}$ and $\hat{\eta}$ into Eqn. 4.9 to obtain $(\hat{\chi}, \hat{\psi})$
else if $|\lambda| \leq 1$ **then** ▷ The family $N_\lambda^{\vdash}$ is steep
  solve Eqn. 4.17 to get $\hat{\omega}$ and plug $\hat{\omega}$ into Eqn. 4.16 to get $\hat{\eta}$
  insert $\hat{\omega}$ and $\hat{\eta}$ into Eqn. 4.9 to obtain $(\hat{\chi}, \hat{\psi})$
else ▷ The family $N^{\vdash}$ is not steep when $|\lambda| > 1$
  if $\bar{w}\bar{w}_{-1} < \frac{|\lambda|}{|\lambda|-1}$ **then**
    solve Eqn. 4.17 to get $\hat{\omega}$ and plug $\hat{\omega}$ into Eqn. 4.16 to get $\hat{\eta}$
    insert $\hat{\omega}$ and $\hat{\eta}$ into Eqn. 4.9 to obtain $(\hat{\chi}, \hat{\psi})$
  else ▷ $\bar{w}\bar{w}_{-1} \geq \frac{|\lambda|}{|\lambda|-1}$
    if $\lambda > 1$ **then** ▷ Corresponds to boundary case 1 of the GIG
      $\hat{\chi} = 0$; $\hat{\psi} = \frac{2\lambda}{\bar{w}}$;
    else ▷ Corresponds to boundary case 2 of the GIG
      $\hat{\chi} = \frac{-2\lambda}{\bar{w}_{-1}}$; $\hat{\psi} = 0$;
    end if
  end if
end if

Eqn. 4.15 can only be solved numerically but when $\lambda + \frac{1}{2}$ is an integer the equation has an exact solution. For the cases when $\lambda = \frac{1}{2}$ or $\lambda = -\frac{1}{2}$ we have

$$R_{1/2}(\omega) = 1 + \frac{1}{\omega}, \quad R_{-1/2} = 1, \tag{4.18}$$

therefore when $|\lambda| = \frac{1}{2}$

$$\hat{\omega}_{\pm\frac{1}{2}} = \frac{1}{\bar{w}\cdot\bar{w}_{-1} - 1}. \tag{4.19}$$

If $\bar{w} \cdot \bar{w}_{-1} > 1$, the maximum likelihood estimate of the GIG exists and is unique. When $\bar{w} \cdot \bar{w}_{-1} = 1$, the likelihood function of the GIG will not attain its supremum. Algorithm 1 is based on Theorem 4.1 Jørgensen (1982) and when using the algorithm the assumption is that $\bar{w} \cdot \bar{w}_1 > 1$.

Estimation of $\hat{\lambda}$

Using $(\hat{\chi}, \hat{\psi})$ obtained by algorithm 1, the maximum likelihood estimator $\hat{\lambda}$ of the GIG is the value of λ that maximizes the log-likelihood of the GIG given by Eqn. 4.11. Define

$$u = \frac{\bar{w} \cdot \bar{w}_{-1}}{\bar{w} \cdot \bar{w}_{-1} - 1}. \tag{4.20}$$

It was proved by Jørgensen (1982) that $\mathscr{L}_{\text{GIG}}(\lambda, \eta, \omega)$ is strictly convex on $(-u, u)$. Jørgensen (1982) mentioned using the tabulation method to find $\hat{\lambda}$, however the tabulation method is very time consuming so instead the minimization algorithm by Brent (1973) was used to find $\lambda \in (-u, u)$ that maximizes Eqn. 4.11.

Brent (1973) minimization algorithm uses a combination of successive parabolic interpolation and golden search method. The golden search ensures reliability while the quadratic interpolation optimizes convergence. The algorithm is guaranteed to converge to a local minimum with convergence rate faster than that of the bisection method.

The negative of Eqn. 4.11 was minimized using Brent (1973) algorithm, in order to find the value $\hat{\lambda}$. Since the GIG is unimodal Brent (1973) algorithm will converge to a global maximum.

If x is the minimum (or a maximum) of a parabola that goes through the three points $(a, f(a))$, $(b, f(b))$ and $(c, f(c))$ then

$$x = b - \frac{1}{2} \frac{[f(b) - f(c)](b-a)^2 - [f(b) - f(a)](b-c)^2}{[f(b) - f(c)](b-a) - [f(b) - f(a)](b-c)}. \tag{4.21}$$

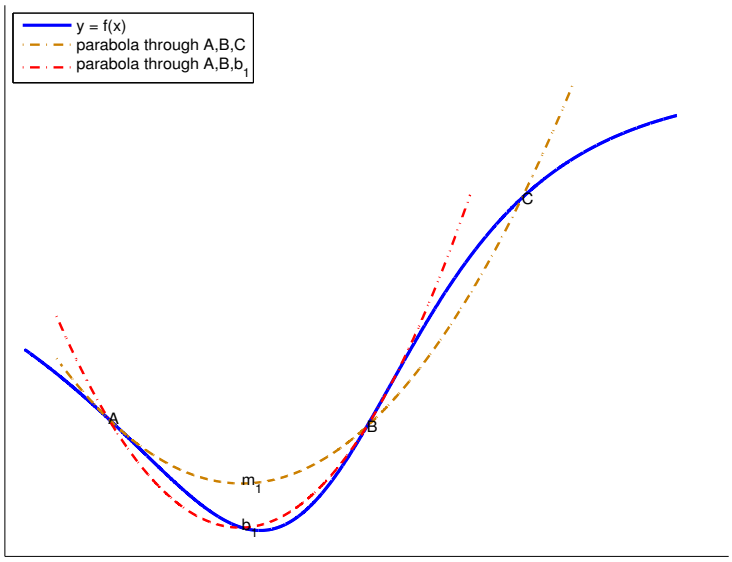


Fig. 4.1: Convergence to a minimum by inverse parabolic interpolation. A parabola through three original points A, B and C on the curve f is drawn. The function, f is evaluated on the minimum of the parabola (b_1) which replaces C. A new parabola is drawn through the points A, b_1 and B and the minimum of this parabola is close to minimum of the curve.

If the three points a, b and c are collinear then the denominator of equation 4.21 is zero, and parabolic interpolation will fail. In-order to guarantee convergence to a minimum Brent (1973) uses golden search when the function is un-corporative and switches to using parabolic interpolation when the function allows.

At any particular stage Brent (1973) method keeps track of six function points a, b, u, v, w and x that are not necessarily distinct. The minimum is bracketed between a and b. The point of the most recent function evaluation is u, while x is the point of the least function value evaluated so far and in the case of a tie x is the most recent function value. The point of the second least function value thus far is w and v is the previous value of w.

Brent (1973) first attempts parabolic interpolation fitting through x, v and w. The parabolic step is accepted if the parabolic step falls within the bounding interval (a, b) and if the movement implied by the parabolic step is less half the movement of the previous step. The second condition simply

ensures that the algorithm does not bounce around in a cycle but actually converges to something. Brent (1973) algorithm is given in Appendix C.

4.2 The boundary cases of the GIG

As $\chi \downarrow 0$, the GIG turns into the gamma distribution with parameters $(\lambda, 2/\psi)$ and as $\psi \downarrow 0$, the GIG turns into the inverse gamma distribution with parameters $(-\lambda, \chi/2)$. The probability density function the gamma distribution is given by

$$f_\Gamma(w; \alpha, \beta) = \frac{w^{\alpha-1}}{\beta^\alpha \Gamma(\alpha)} e^{\frac{-w}{\beta}}, \quad w > 0 \,\text{and}\, \alpha,\ \beta > 0. \tag{4.22}$$

where $\Gamma(\alpha)$ is the Gamma function. The probability of the inverse gamma distribution is given by

$$f_{I\Gamma}(w; \alpha, \beta) = \frac{\beta^\alpha w^{-\alpha-1}}{\Gamma(\alpha)} e^{-\frac{\beta}{w}}, \quad w > 0 > \,\text{and}\, \alpha,\ \beta > 0. \tag{4.23}$$

The limits of the probability density function of boundary case 1 and 2 are given by

$$\lim_{\chi \downarrow 0} f_{\text{GIG}}(w; \lambda, \chi, \psi) = f_\Gamma\left(w; -\lambda, \frac{2}{\psi}\right) \quad \text{and} \lim_{\psi \downarrow 0} f_{\text{GIG}}(w; \lambda, \chi, \psi = f_{I\Gamma}\left(w; -\lambda, \frac{\chi}{2}\right). \tag{4.24}$$

respectively.

4.2.1 The generalized gamma distribution

The gamma and inverse gamma distributions correspond to the boundary cases of the GIG. The boundary cases of the GIG are also special cases of the generalized gamma distribution (GGD). The GGD was introduced by Stacy (1962) with a positive power parameter ζ. The parametrization by Stacy and Mihram (1965) allows for negative power parameters whose absolute value is greater than 0. The probability density function of the GGD is given by

$$f_{\text{GGD}}(w; \zeta, \alpha, \beta) = \frac{|\zeta|}{\beta\Gamma(\alpha)} \left(\frac{w}{\beta}\right)^{\zeta\alpha-1} e^{-\left(\frac{w}{\beta},\right)^\zeta}, \quad \zeta \neq 0,\ w \geq 0 \,\text{and}\, \alpha, \beta > 0. \tag{4.25}$$

The gamma distribution and the inverse gamma distribution corresponds to GGD with power parameter $\zeta = 1$ and $\zeta = -1$, respectively. Given n

independent random variables $w_1, \ldots w_n$, the log likelihood function of the GGD is given by

$$\ln L(w;\zeta,\alpha,\beta) = n \times \left(\ln|\zeta| - \ln\Gamma(\alpha) - \zeta\alpha\ln\beta\right) - (\zeta\alpha - 1)\sum_{i=1}^{n}\ln w_i - \frac{1}{\beta^{\zeta}}\sum_{i=1}^{n} w_i^{\zeta}. \tag{4.26}$$

Fix the power parameter ζ and denote the stationary points of the log likelihood of the GGD by $(\hat{\alpha},\hat{\beta})$. The stationary points of the GGD can be obtained by solving the following system of equations

$$0 = \frac{\zeta}{(\hat{\beta})^{\zeta+1}}\sum_{i=1}^{n} w_i^{\zeta} - \frac{n\hat{\alpha}\zeta}{\hat{\beta}}, \tag{4.27}$$

$$0 = \zeta\sum_{i=1}^{n}\ln w_i - n\zeta\ln\hat{\beta} - n\psi(\hat{\alpha}), \tag{4.28}$$

where $\psi(\alpha)$ is the digamma function. Equation 4.27 yields the following expression of β in terms of α

$$\hat{\beta} = \left(\frac{\sum_{i=1}^{n} w_i^{\zeta}}{n\hat{\alpha}}\right)^{1/p}. \tag{4.29}$$

Substituting Eqn. 4.29 into Eqn. 4.28 we get the following equation in terms of α only

$$0 = n(\ln\hat{\alpha} - \psi(\hat{\alpha})) + \zeta\sum_{i=1}^{n}\ln w_i - n\sum_{i=1}^{n} w_i^{\zeta} + n\ln n, \tag{4.30}$$

where $\psi(x)$ is the digamma function.

Definition 4.1. The digamma function
The logarithmic derivative of the gamma function is given by

$$\psi(z) = \frac{\mathrm{d}}{\mathrm{d}z}\ln\Gamma(z) = \frac{\Gamma'(z)}{\Gamma(z)}, \tag{4.31}$$

where $z \in \mathbb{C}$ and $x \neq 0, -1, -2, \ldots$.

Three of the useful properties from Abramowitz and Stegun (1964) that are useful in the analysis are

1. **Asympototical expansion as $z \to \infty$ in $|arg|\, z < \pi$**

$$\psi(z) = \ln z - \frac{1}{2z} - \sum_{m=1}^{\infty}\frac{B_{2m}}{2mz^{2m}}, \tag{4.32}$$

where B_m is the mth Bernoulli number.

2. **Taylor expansion at** $z_0 = 1$

$$\psi(z+1) = -\gamma_0 - \sum_{m=1}^{\infty}(-1)^{m+1}\zeta(m+1)z^m, \tag{4.33}$$

where $\zeta_0(m)$ denotes the Riemann zeta function and $\gamma_0 = \psi(1) \simeq 0.5772156649$

3. **Recurrence formula**

$$\psi(z+1) = \psi(z) + \frac{1}{z}, \quad z \neq 0, -1, -2, \ldots \tag{4.34}$$

Solving the above system of non-linear equations analytically or numerically is difficult. In the presence of multiple roots, many iterative techniques usually converge to wrong values and do not guarantee convergence. One must be very careful when using iterative methods to solve for the stationary points.

To solve the likelihood equations of the GGD, Wingo (1987) developed an algorithm that makes use of Jones *et al.* (1978) heuristic root isolation method. However the drawback of Wingo (1987) algorithm is that it will work only when the parameters meet some specified conditions. Luo and Zhao (2014) developed an algorithm to estimate the parameters of the GGD by making use of the properties of certain functions and the Intermediate value theorem. Luo and Zhao (2014) algorithm uses the hyperbolic interpolation procedure and guarantees super-linear convergence to a global optimum.

Definition 4.2. Completely monotonic (Van Haeringen (1993))
A function f, is said to be completely monotonic on the interval I, if the derivatives of all orders of $f(x)$ exist and for $n \geq 0$, $0 \leq (-1)^n f^n(x) < \infty$.

Denote the set of completely monotonic functions on the interval I by $\mathscr{C}[I]$.

Theorem 4.3. *Let* $\alpha \in \mathbb{R}$. *Then*

$$f(x) = \psi(x) - lnx + \frac{\alpha}{x} \in \mathscr{C}(0,\infty) \iff \alpha \geq 1, \tag{4.35}$$

$$g(x) = lnx - \psi(x) - \frac{\alpha}{x} \in \mathscr{C}(0,\infty) \iff \alpha \leq 2 \tag{4.36}$$

For proof see Qi (2007), Theorem 2 . Following 4.3 we have the corollary,

Corollary 4.4. *The following inequality holds on* $(0, \infty)$

$$\frac{1}{2x} < lnx - \psi(x) < \frac{1}{x}. \tag{4.37}$$

From Qi and Guo (2010) we get the following corollary,

Corollary 4.5. *The following inequality holds*

$$\frac{(k-1)!}{x^k} + \frac{k!}{2x^{k+1}} < (-1)^{k+1}\psi^{(k)}(x) < \frac{(k-1)!}{x^k} + \frac{k!}{x^{k+1}} \tag{4.38}$$

for $k \in \mathbb{N}^+$ *on* $(0, \infty)$.

Define the right hand side of Eqn. 4.30 as a function

$$f(\alpha) := Q + n(\ln\alpha - \psi(\alpha)), \tag{4.39}$$

where

$$Q = \zeta \sum_{i=1}^{n} \ln w_i - n \sum_{i=1}^{n} w_i^{\zeta} + n\ln n \tag{4.40}$$

$$= n \left(\ln \sqrt[n]{\prod_{i=1}^{n} w_i^{\zeta}} - \ln \frac{\sum_{i=1}^{n} w_i^{\zeta}}{n} \right) < n, \tag{4.41}$$

according to Jensen's inequality. Using Eqn. 4.32 we get the following approximation

$$f(\alpha) \simeq Q + \mathcal{O}(\frac{1}{\alpha}) = Q < 0, \quad \text{as } \alpha \to \infty. \tag{4.42}$$

Exploiting equation 4.33, we obtain the following approximation

$$f(\alpha) \simeq Q + n \left(\ln\alpha + \gamma_0 + \frac{1}{\alpha} + \mathcal{O}(\alpha) \right) \to \infty, \quad \text{as } \alpha^+ \to 0. \tag{4.43}$$

Due to Corollary 4.5, with $k = 1$ one can deduce that

$$f'(\alpha) = n \left(\frac{1}{\alpha} - \psi'(\alpha) \right) < -\frac{n}{\alpha^2} < 0, \tag{4.44}$$

inferring that $f(\alpha)$ decreases monotonically. The Intermediate value theorem guarantees the existence of a root of $f(\alpha)$ on $(0, \infty)$, since the function is decreasing monotonically and $f(\alpha)$ posses different signs as $\alpha^+ \to 0$ and $\alpha \to \infty$.

Proposition 4.6. $f(\alpha)$ *has a unique root* $\hat{\alpha}$.

Since we are guaranteed of a solution $(\hat{\alpha}, \hat{\beta})$ for a fixed ζ, all that is left to do is show that the point is a maximum. Define

$$\begin{aligned} A(\hat{\alpha}, \hat{\beta}) &= \frac{\partial^2}{\partial \alpha^2} \ln L(w; \zeta, \alpha, \beta, \zeta) \\ &= -n\psi'(\hat{\alpha}), \end{aligned} \tag{4.45}$$

$$\begin{aligned} B(\hat{\alpha}, \hat{\beta}) &= \frac{\partial^2}{\partial \alpha \partial \beta} \ln L(w; \zeta, \alpha, \beta, \zeta) \\ &= -\frac{n\zeta}{\hat{\beta}}, \end{aligned} \tag{4.46}$$

$$\begin{aligned} C(\hat{\alpha}, \hat{\beta}) &= \frac{\partial^2}{\partial \beta^2} \ln L(w; \zeta, \alpha, \beta, \zeta) \\ &= -\frac{\zeta(\zeta+1)}{\hat{\beta}^{\zeta+2}} \sum_{i=1}^{n} w_i^{\zeta} + \frac{n\hat{\alpha}\zeta}{\hat{\beta}^2} \\ &= \frac{-n\hat{\alpha}\zeta^2}{\hat{\beta}}. \end{aligned} \tag{4.47}$$

From Corollary 4.5 we have that

$$\alpha\psi'(\hat{\alpha}) > \left(1 + \frac{1}{2\hat{\alpha}}\right) > 1. \tag{4.48}$$

Therefore we have

$$A(\hat{\alpha}, \hat{\beta})C(\hat{\alpha}, \hat{\beta}) - B(\hat{\alpha}, \hat{\beta})^2 = \left(\frac{n\zeta}{\beta}\right)^2 (\hat{\alpha} - 1) > 0, \tag{4.49}$$

thus the stationary point $(\hat{\alpha}, \hat{\beta})$ satisfies the sufficient condition of a maximum.

Proposition 4.7. *Given a fixed ζ, the stationary point $(\hat{\alpha}, \hat{\beta})$ is the maximum of ln $L(w; \zeta, \alpha, \beta, \zeta)$.*

From Corollary 4.4, it can be deduced that

$$Q + \frac{n}{2\alpha} < f(\alpha) < Q + \frac{n}{\alpha}. \tag{4.50}$$

The starting interval in the root search of $\hat{\alpha}$ can be obtained by setting the left and right hand side of Eqn. 4.50

$$(a_0, b_0) = (-\frac{n}{2Q}, -\frac{n}{Q}), \tag{4.51}$$

and is clear that

$$f(a_0) > 0 \quad \text{and} \quad f(b_0) < 0.$$

Since $f(\alpha)$ is monotonically decreasing on $(0, \infty)$ and $f(a_0)$ and $f(b_0)$, possess opposite signs, the Intermediate value theorem guarantees the existence of a unique root in the interval (a_0, b_0). Define

$$\eta_a(\alpha) := f(a) - \frac{n}{2}\left(\frac{1}{a} - \frac{1}{\alpha}\right), \quad \eta_b(\alpha) := f(b) - \frac{n}{2}\left(\frac{1}{b} - \frac{1}{\alpha}\right),$$
$$\xi_a(\alpha) := f(a) - n\left(\frac{1}{a} - \frac{1}{\alpha}\right), \quad \xi_b(\alpha) := f(b) - n\left(\frac{1}{b} - \frac{1}{\alpha}\right).$$

Observe that the functions $\eta_a(\alpha)$ and $\xi_a(\alpha)$ pass through the point $(a, f(a))$ while the functions $\eta_b(\alpha)$ and $\xi_b(\alpha)$ pass through $(b, f(b))$. Using corollary 4.4 for $\alpha \in (a, b)$ we get the following result

$$\begin{aligned}\eta_a(\alpha) - f(\alpha) &= n\left(\ln a - \psi(a) - \frac{1}{2a}\right) - n\left(\ln\alpha - \psi(\alpha) - \frac{1}{2\alpha}\right)\\ &> n(\frac{1}{2a} - \frac{1}{2a}) - n(\frac{1}{2\alpha} - \frac{1}{2\alpha})\\ &= 0.\end{aligned}$$

Using similar arguments, it can be deduced that

$$f(\alpha) - \eta_b(\alpha) > 0, \quad f(\alpha) - \xi_a(\alpha) > 0,$$
$$\xi_b(\alpha) - f(\alpha) > 0.$$

This implies that the root is contained in the curved quadrilateral ABCD that is surrounded by the curves $y = \xi_a(\alpha)$, $y = \eta_a(\alpha)$, $y = \xi_b(\alpha)$ and $y = \eta_b(\alpha)$, the quadrilateral is illustrated in figure 4.2 as ABCD.

Denote the roots of $\xi_a(\alpha)$, $\xi_b(\alpha)$, $\eta_a(\alpha)$ and $\eta_b(\alpha)$ by t_a, t_b, s_a, s_b respectively. It follows that

$$t_a = \frac{na}{n - af(a)}, \quad t_b = \frac{nb}{n - bf(b)}, \tag{4.53a}$$
$$s_a = \frac{na}{n - 2af(a)}, \quad s_b = \frac{nb}{n - 2bf(b)}. \tag{4.53b}$$

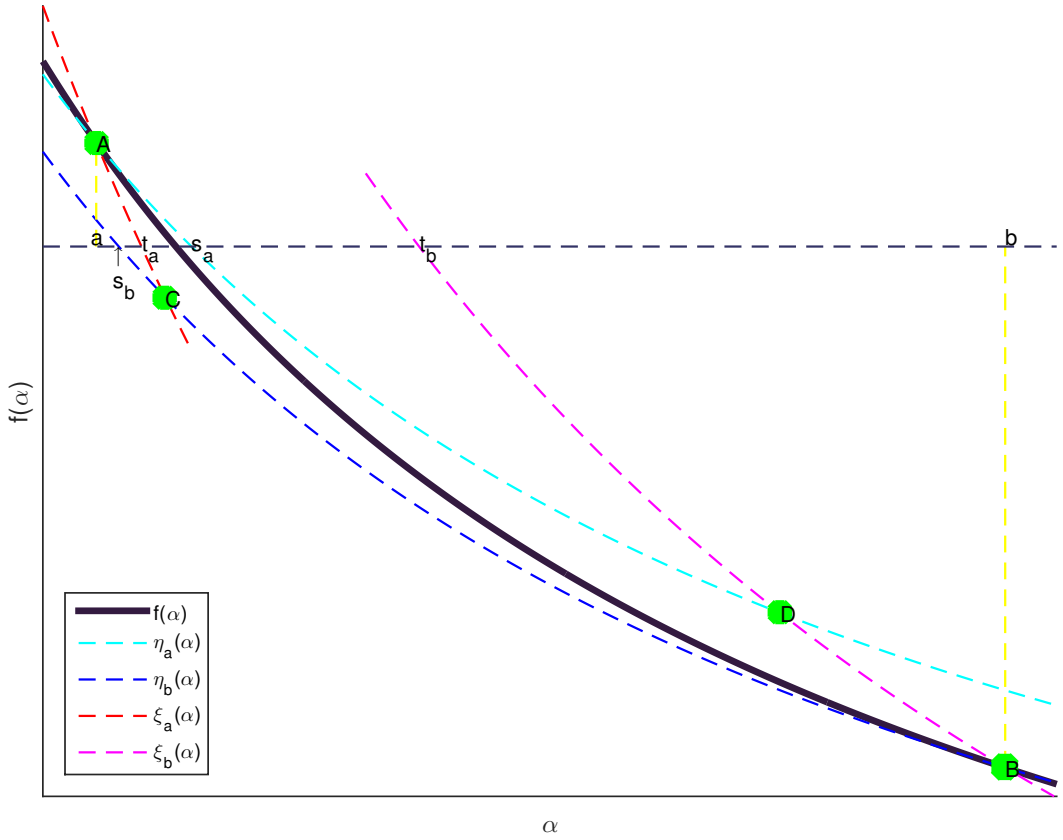


Fig. 4.2: The curved quadilateral ABCD

The curved quadrilateral intersects with the axis on two points. The left intersection is given by $\min\{s_a, t_a\}$ whilst the right intersections is given by $\max\{s_b, t_b\}$. Hence one can be sure that the root truly lies in the subinterval $(\min\{s_a, t_a\}, \max\{s_b, t_b\})$ and can successfully update the initial interval with one of its subintervals.

Luo and Zhao (2014) method iteratively squeezes the search interval by utilizing rational interpolation.The lower and upper bounds in Eqn. 4.50 have a hyperbolic form, so it is reasonable to use the form

$$g(\alpha) = c_0 + \frac{c_1}{\alpha} = 0. \tag{4.54}$$

If $f(\alpha) = g(\alpha)$ at the points $(a, f(a))$ and $((b, f(b))$ then

$$c_0 = \frac{af(a) - bf(b)}{a - b}, \quad c_1 = \frac{ab}{a - b}(f(a) - f(b)). \tag{4.55}$$

If we solve for α using Eqn. 4.54 we obtain

$$\alpha = \frac{ab(f(a) - f(b))}{af(a) - bf(b)}. \tag{4.56}$$

4.3 The generalized hyperbolic distribution

McNeil *et al.* (2005) used the normal mean-variance representation with the GIG distribution as a mixing distribution to represent the multivariate GHD.

Definition 4.8. Normal mean-variance representation
The random vector X is said to have a normal mean-variance distribution if

$$X \overset{d}{=} m(W) + \sqrt{W}AZ, \tag{4.57}$$

where

i $Z \sim N_k(\mathbf{0}, I_k)$

ii $A \in \mathbb{R}^{d \times k}$

iii $m : [0, \infty) \to \mathbb{R}^d$ is a measurable function

iv $W \geq 0$ is a positive, scalar random variable independent of Z.

The specification of the function $m(w)$ in McNeil *et al.* (2005) is

$$m(w) = \boldsymbol{\mu} + W\boldsymbol{\gamma}, \quad \boldsymbol{\mu}, \boldsymbol{\gamma} \in \mathbb{R}^d. \tag{4.58}$$

Using the specification of $m(w)$ by McNeil *et al.* (2005), from Eqn. 4.57, one can see that

$$\mathbf{X}|W \sim N(\boldsymbol{\mu} + W\boldsymbol{\gamma}, W\Sigma), \quad \Sigma = AA', \tag{4.59}$$

Eqn. 4.59 is the reason why the GHD distributions with this representation are known as normal mean-variance mixture models. When the mixing variable in Eqn. 4.57 is $W \sim N^{\vdash}(\lambda, \chi, \psi)$, the random variable, X is said to have a generalized hyperbolic distribution. Following the representation in McNeil *et al.* (2005) , the expectation and variance of the multivariate GHD is given by

$$\begin{aligned} \mathbb{E}[X] &= \boldsymbol{\mu} + \mathbb{E}[W]\boldsymbol{\gamma} \\ &= \boldsymbol{\mu} + \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}\boldsymbol{\gamma}, \end{aligned} \tag{4.60}$$

$$\begin{aligned} \text{Cov}[X] &= \mathbb{E}[W]\Sigma + \text{Var}[W]\boldsymbol{\gamma}\boldsymbol{\gamma}' \\ &= \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}\Sigma \\ &+ \left(\frac{\chi}{\psi}\frac{K_{\lambda+2}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} - (\frac{\chi}{\psi})^{1/2}\left(\frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})}\right)^2\right)\boldsymbol{\gamma}\boldsymbol{\gamma}'. \end{aligned} \tag{4.61}$$

The expectation and variance of the mixing variable W are given by equations 4.3 and 4.4.

4.3.1 The density and parameters of the GHD

Theorem 4.9. *The density function of the GHD*
When Σ is nonsingular and the mixing variable is $W \sim N^{\vdash}(\lambda, \chi, \psi)$, the probability density of GHD random variable $x \in \mathbb{R}^d$ is given by

$$\begin{aligned} f_{GHD}(x; \lambda, \chi, \psi, \mu, \Sigma, \gamma) = cK_{\lambda-\frac{d}{2}}\left(\sqrt{\big(\chi + \rho(x)\big)\big(\psi + \gamma'\Sigma^{-1}\gamma\big)}\right) \\ \times e^{(x-\mu)'\Sigma^{-1}\gamma} \\ \times \big(\big(\chi + \rho(x)\big)\big(\psi + \gamma'\Sigma^{-1}\gamma\big)\big)^{\frac{\lambda}{2}-\frac{d}{4}}, \end{aligned} \tag{4.62}$$

with the normalizing constant

$$c = \frac{(\psi\chi)^{-\frac{\lambda}{2}}\psi^{\lambda}(\psi + \gamma'\Sigma'\gamma)^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}K_{\lambda}(\sqrt{\chi\psi})}, \tag{4.63}$$

where $|.|$ denotes the determinant and $\rho(x) = (x-\mu)'\Sigma^{-1}(x-\mu)$.

For illustrative purposes the univariate GHD (i.e. the GHD when $d = 1$) was used to show how the shape of the distribution changes when one parameter is changed whilst all the other parameters are kept constant.

- $\lambda \in \mathbb{R}$ defines the subclass of the GHD and it is also related to steepness of the density. The density becomes increasingly steep as λ decreases. This has implications to the behavior of the tails in the sense that a steeper density has lighter tails.

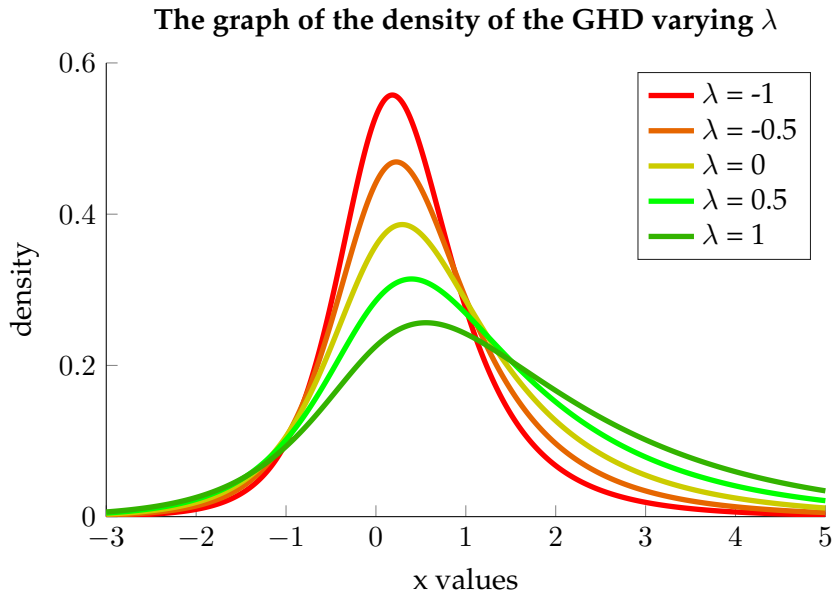


Fig. 4.3: Graph showing how the density function of the GHD changes as λ changes.

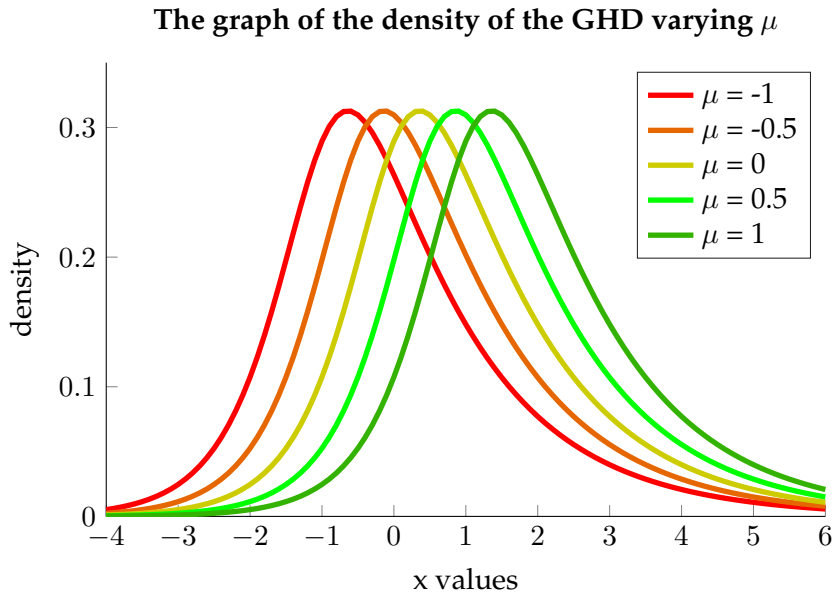


Fig. 4.4: Graph showing how the density function of the GHD changes as μ changes.

- $\mu \in \mathbb{R}^d$ is the parameter that determines where the distribution will be centered.

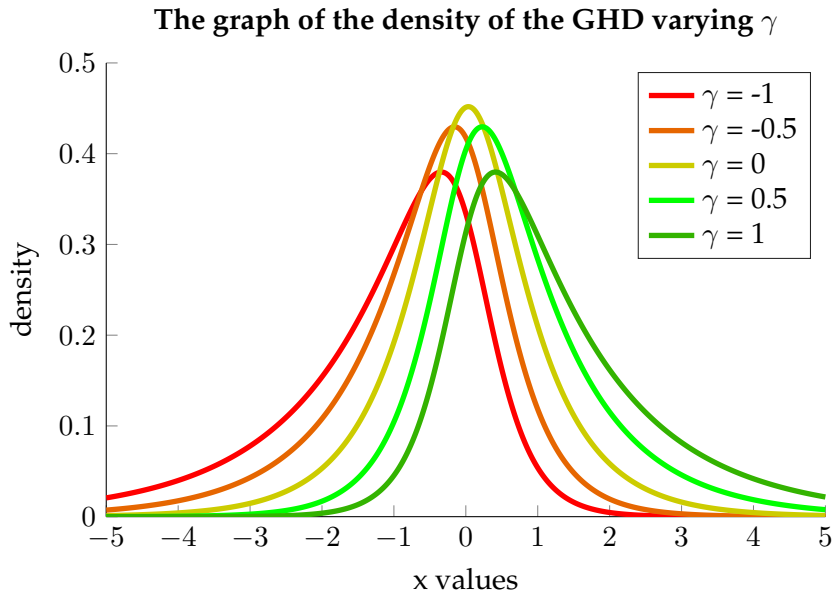


Fig. 4.5: Graph showing how the density function of the GHD changes as γ changes.

- $\gamma \in \mathbb{R}^d$ is the skewness parameter.

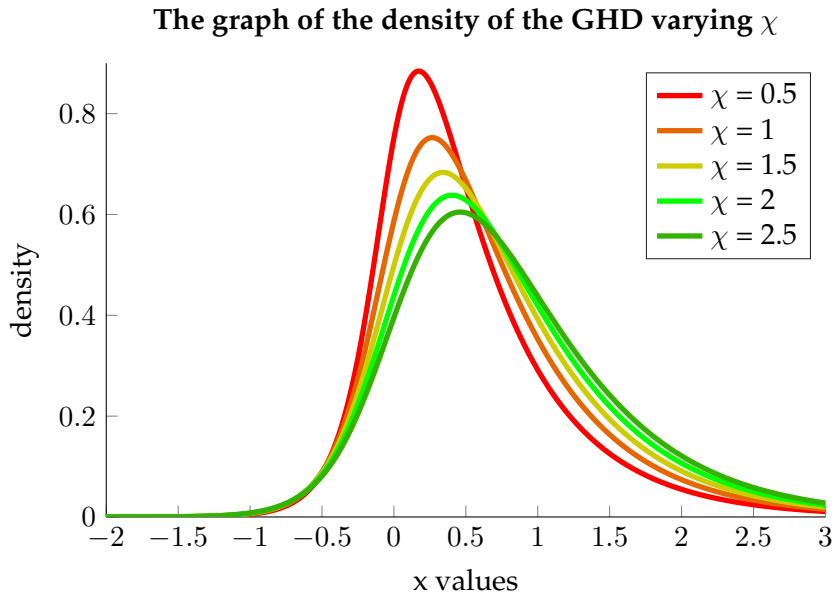


Fig. 4.6: Graph showing how the density function of the GHD changes as χ changes.

- $\chi > 0$ and $\psi > 0$ are shape parameters and they describe the tail

flatness of the GHD distribution.

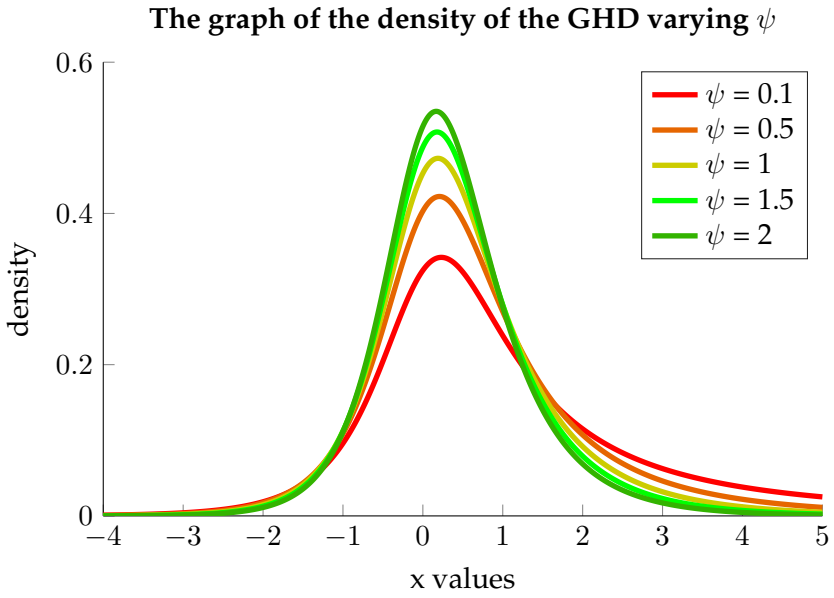


Fig. 4.7: Graph showing how the density function of the GHD changes as ψ changes.

- Σ is a dispersion parameter, the larger the Σ the more dispersed the distribution.

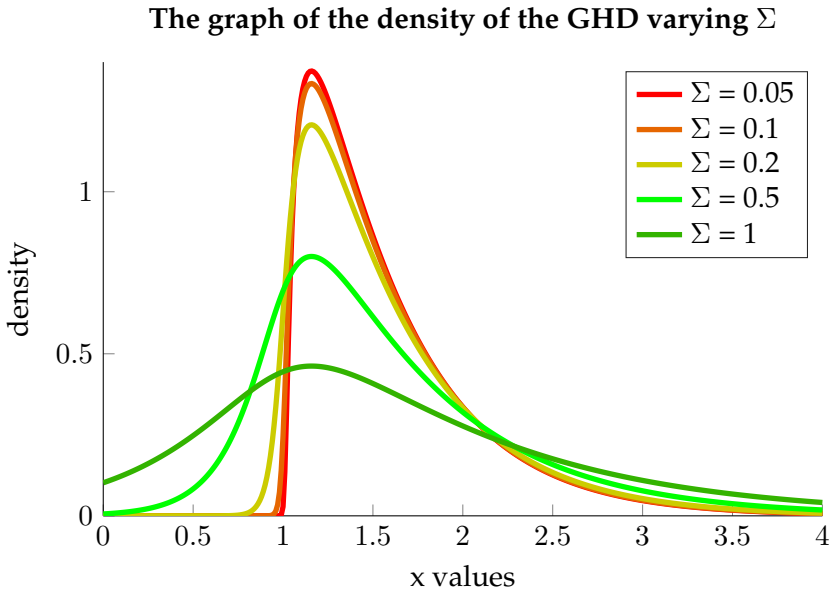


Fig. 4.8: Graph showing how the density function of the GHD changes as Σ changes.

4.3.2 Subclasses of the GHD

The GHD has many subclasses known with special names as well as limiting cases. Table 4.2 shows the connections between subclasses of the GIG with subclasses of the GHD. It is worth mentioning that only two subclasses of the GHD, namely the normal inverse Gaussian and the variance gamma are closed under convolutions. In this section only the subclasses of the GHD that were used for parameter estimation are discussed in detail.

GHD subclass	Mixing distribution	Parameter Range		
Proper GHD	Normal case GIG	$\lambda \in \mathbb{R}$	$\chi > 0$	$\psi > 0$
Hyperbolic asymmetric t	Inverse Gamma	$\lambda < 0$	$\chi > 0$	$\psi = 0$
Student's t	Inverse Gamma ($\Sigma^{-1}\gamma = \mathbf{0}$)	$\lambda < 0$	$\chi > 0$	$\psi = 0$
Variance gamma	Gamma	$\lambda > 0$	$\chi = 0$	$\psi > 0$
Asymmetric Cauchy	Lévy	$\lambda = -1/2$	$\chi > 0$	$\psi = 0$
Cauchy	Lévy ($\Sigma^{-1}\gamma = \mathbf{0}$)	$\lambda = -1/2$	$\chi > 0$	$\psi = 0$
Normal inverse Gaussian	Inverse Gaussian	$\lambda = -1/2$	$\chi > 0$	$\psi > 0$
Asymmetric Laplace	Exponential	$\lambda = 1$	$\chi = 0$	$\psi > 0$
Laplace	Exponential ($\Sigma^{-1}\gamma = \mathbf{0}$)	$\lambda = 1$	$\chi = 0$	$\psi > 0$
Hyperbolic	Positive hyperbolic	$\lambda = d/2$	$\chi > 0$	$\psi > 0$

For all the subclasses of the GHD, $\mu \in \mathbb{R}^d$, $\gamma \in \mathbb{R}^d$ and $A \in \mathbb{R}^{d \times d}$

Tab. 4.2: Subclasses of the GHD

4.3.3 The Hyperbolic distribution

The hyperbolic (HYP) distribution was introduced by Barndorff-Nielsen (1977), when he introduced the GHD. The logarithm of the hyperbolic distribution is a hyperbola whereas the logarithm of the density of the normal distribution is a parabola. This observation resulted in the name hyperbolic distribution. The definition of the distribution was inspired by Bagnold (1941) who noted that logarithm of the density of wind blown sands look more like a hyperbola than a parabola.

- When $\lambda = \frac{(d+1)}{2}$, the word generalized is dropped and the distribution becomes multivariate hyperbolic distribution. The marginal distributions will no longer be hyperbolic distributions but will have $\lambda = \frac{1}{2}$.
- When $\lambda = 1$, one obtains the multivariate GHD that has univariate marginal distributions that are hyperbolic.

The expressions of the expectation and variance of the hyperbolic distribution can be obtained by substituting λ by 1 in equations 4.60 and 4.61 respectively.

4.3.4 The Normal Inverse Gaussian distribution

The normal inverse Gaussian (NIG) was introduced by Barndorff-Nielsen (1997) using the normal mean-variance representation with the inverse Gaussian as the mixing distribution. For the NIG, $\lambda = -\frac{1}{2}$ and the NIG has slightly heavier tails than the GHD. Barndorff-Nielsen (1997) also provides an alternate representation of the NIG via time changing Brownian motion using the inverse Gaussian distribution as a stochastic clock. Barndorff-Nielsen (1998) discusses in depth, the alternative representation of the NIG and the feasibility of likelihood inference using the model. The aim of Barndorff-Nielsen (1998) was to model key stylized facts of financial time series and turbulence of stochastic process with NIG marginals. The expressions of the expectation and variance of the NIG can be obtained by substituting λ by $-\frac{1}{2}$ in equations 4.60 and 4.61 respectively.

4.3.5 The Variance Gamma distribution

The variance gamma (VG) as known as the generalized Laplace distribution is obtained when $\lambda > 0$ and $\chi = 0$. The VG was introduced by Madan and Seneta (1990) as a symmetric distribution to model share market returns. The asymmetric distribution was introduced by Madan *et al.* (1998). Madan and Seneta (1990) and Madan *et al.* (1998) represented the VG process as Brownian motion time changed by an independent gamma process. Lévy processes with increments that follow the VG are known as Laplace motion.

From Abramowitz and Stegun (1964) we obtain the following property of the Bessel function,

$$K_\lambda(x) \sim \Gamma(\lambda)2^{\lambda-1}x^{-\lambda}, \quad x \to 0, \quad \lambda > 0, \tag{4.64}$$

and of the Gamma function

$$\Gamma(x+1) = x\Gamma(x). \tag{4.65}$$

We compute the expressions

$$\begin{aligned}\frac{K_{\lambda+1}(x)}{K_\lambda(x)} &\simeq \frac{\Gamma(\lambda+1)2^{(\lambda-1)+1}x^{-\lambda-1}}{\Gamma(\lambda)2^{\lambda-1}x^{-\lambda}} \\ &= \frac{2\lambda}{x},\end{aligned} \tag{4.66}$$

and

$$\begin{aligned}\frac{K_{\lambda+2}(x)}{K_\lambda(x)} &\simeq \frac{\Gamma(\lambda+2)2^{(\lambda-1)+2}x^{-\lambda-2}}{\Gamma(\lambda)2^{\lambda-1}x^{-\lambda}} \\ &= \frac{4\lambda(\lambda+1)}{x^2}.\end{aligned} \tag{4.67}$$

Using equations 4.66 and 4.67 the following expressions for the variance and expectation of the VG were obtained

$$\mathbb{E}[X] = \boldsymbol{\mu} + \frac{2\lambda}{\psi}\boldsymbol{\gamma}, \tag{4.68}$$

$$\mathrm{Cov}[X] = \left(\frac{2\lambda}{\psi}\right)\Sigma + \left(\frac{4\lambda}{\psi^2}\right)\boldsymbol{\gamma}\boldsymbol{\gamma}'. \tag{4.69}$$

To obtain the probability density function of the VG, replace the term $\frac{(\chi\psi)^{-\lambda/2}}{K_\lambda(\sqrt{\chi\psi})}$ by $\frac{1}{\Gamma(\lambda)2^{\lambda-1}}$ and χ by 0 in Eqn. 4.62. When $\rho(x) \to 0$, a limiting case arises and for $\lambda - \frac{d}{2} > 0$ the probability density function of the VG reduces to

$$f_{\mathrm{VG}}(x) = \frac{\psi^\lambda(\psi + \gamma'\Sigma\gamma)^{d/2-\lambda}\Gamma(\lambda - \frac{d}{2})}{2^d\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\lambda)}. \tag{4.70}$$

When $\lambda - \frac{d}{2} < 0$, the density diverges.

4.3.6 The skewed Student's t distribution

The skewed Student's t (SKT) will also be referred to as the Student's t throughout the chapter. The idea to use a skew extension of the of the Student's t distribution to model asset returns was first proposed by Hansen (1994). The Student's t distribution is a subclass of the GHD that is obtained when $\lambda < 0$ and $\psi = 0$. Define $\nu = -2\lambda$, using the symmetric property of the Bessel function ($K_\lambda(x) = K_{-\lambda}(x)$) and the property of the Bessel function given by Eqn. 4.64, we compute the following expressions

$$\begin{aligned}\frac{K_{\lambda+1}(x)}{K_\lambda(x)} &= \frac{K_{-\lambda-1}(x)}{K_{-\lambda}(x)} \\ &\simeq \frac{\Gamma(-\lambda-1)2^{-\lambda-2}x^{\lambda+1}}{\Gamma(-\lambda)2^{-\lambda-1}x^{\lambda}} \\ &= \frac{x}{2(\lambda+1)},\end{aligned} \tag{4.71}$$

$$\begin{aligned}\frac{K_{\lambda+2}(x)}{K_{\lambda}(x)} &= \frac{K_{-\lambda-2}(x)}{K_{-\lambda}(x)} \\ &\simeq \frac{\Gamma(-\lambda-2)2^{-\lambda-3}x^{\lambda+2}}{\Gamma(-\lambda)2^{-\lambda-1}x^{\lambda}} \\ &= \frac{x^2}{4(\lambda+1)(\lambda+2)}.\end{aligned} \tag{4.72}$$

Therefore, using equations 4.71 and 4.71 expressions for the variance and expectation of the skew Student's t distribution are given by

$$\mathbb{E}[X] = \boldsymbol{\mu} + \frac{\chi}{\nu-2}\boldsymbol{\gamma}, \tag{4.73}$$

$$\text{Cov}[X] = \frac{\chi}{\nu-2}\Sigma + \frac{\chi^2}{(\nu-2)^2(\nu-4)}\boldsymbol{\gamma}\boldsymbol{\gamma}'. \tag{4.74}$$

To ensure that the covariance always exists one can constrain $\nu > 4$ which implies that $\lambda < -2$. To obtain the density function of the SKT, replace $\frac{(\chi\psi)^{-\lambda/2}\psi^{\lambda}}{K_{\lambda}(\sqrt{\chi\psi})}$ with $\frac{\chi^{-\lambda}}{\Gamma(-\lambda)2^{-\lambda-1}}$ and ψ with 0 in Eqn. 4.62. When $\gamma \to 0$, a limiting case arises and the probability density function of the SKT becomes

$$f_{SKT}(x) = \frac{\chi^{-\frac{\nu}{2}}\Gamma(\frac{\nu+d}{2})(\rho(x)+\chi)^{\frac{\nu+d}{2}}}{\pi^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}\Gamma(\frac{\nu}{2})}. \tag{4.75}$$

4.4 Limiting cases of the GHD

It is worth mentioning that the GHD has two limiting cases namely the normal distribution and the GIG. Given the parameters $\alpha, \delta \geq 0$ with $\beta \in [-\alpha, \alpha]$, set $\chi = \delta^2$ and $\psi = \alpha^2 - \beta^2$. Let $\sigma^2 > 0$ and $\mu \in \mathbb{R}$, Eberlein and von Hammerstein (2004) showed that for $\lambda \in \mathbb{R}$

$$\text{GHD}(\lambda, \alpha, \beta, \delta, \mu, \sigma^2) \to N(\mu + \beta\sigma^2, \sigma^2),$$

as $\alpha \to \infty$ and $\delta \to \infty$ with $\frac{\delta}{\alpha} \to \sigma_0^2$. From the result one can infer that as $\chi \to \infty$ and $\psi \to \infty$ the GHD becomes more and more like the normal distribution.

Let $(\lambda_0, \chi_0, \psi_0)$ be in the parameter space of the GIG. Eberlein and von Hammerstein (2004) showed that as $\alpha \to \infty$, $\beta \to \infty$ and $\delta \to 0$ such that $\alpha\delta^2 \to \chi_0$ and $\alpha - \beta \to \psi_0/2$,

$$\text{GHD}(\lambda_0, \alpha, \beta, \delta, 0, \sigma^2) \to \text{GIG}(\lambda_0, \chi_0, \psi_0).$$

Note that α and β converge to ∞ at the same rate because of the condition $\alpha - \beta \to \psi_0/2$. As the parameter values increase the GHD becomes more and more skewed to right until there is no mass on the negative half line which results in the positive GIG distribution.

4.5 Tail behavior of univariate GHD

Tails of the GHD

Let $\alpha = \sqrt{\frac{\psi+\gamma^2/\sigma^2}{\sigma^2}}$ and $\beta = \frac{\gamma}{\sigma^2}$, the asymptotic behavior of the density function of the GHD is given by

$$f(x;\lambda,\chi,\psi,\mu,\sigma,\gamma) \sim c|x|^{\lambda-1}e^{-\alpha|x|+\beta x}, \quad \text{as} \quad x \to \pm\infty, \tag{4.76}$$

where c is a constant. The above relationship shows that the tails of GHD are a product of an exponential and power function therefore the GHD has semi-heavy tails. The different decay rates in the left and right tail are due to the asymmetry parameter γ.

Note that $\alpha \geq |\beta|$ and since $\psi > 0$ by definition, the term $e^{-\alpha|x|+\beta x}$ does not grow in the tails. Decay is ensured by $\lambda < 0$, when $\alpha = |\beta|$.

- $\gamma < 0$, results in heavier left tail whose decay is given by

$$f(x;\lambda,\chi,\psi,\mu,\sigma,\gamma) \sim cx^{\lambda-1}e^{\alpha x+\beta x}, \quad \text{as} \quad x \to -\infty. \tag{4.77}$$

- $\gamma > 0$, results in heavier right tail whose decay is given by

$$f(x;\lambda,\chi,\psi,\mu,\sigma,\gamma) \sim cx^{\lambda-1}e^{-\alpha x+\beta x}, \quad \text{as} \quad |x| \to +\infty. \tag{4.78}$$

Tails of the SKT

Let $\alpha = \frac{|\gamma|}{\sigma^2}$ and $\beta = \frac{\gamma}{\sigma^2}$, if X follows the SKT then

$$f(x;\nu,\mu,\sigma,\gamma) \sim c|x|^{-\frac{\nu}{2}-1}e^{-\alpha|x|+\beta x} \quad \text{as} \quad x \to \pm\infty, \tag{4.79}$$

where c is a constant.

- $\gamma < 0$ results in heavier left tail that decays as

$$f(x;\nu,\mu,\sigma,\gamma) \sim c|x|^{-\frac{\nu}{2}-1} \quad \text{as} \quad x \to -\infty. \tag{4.80}$$

- $\gamma > 0$ results in a heavier right tail that decays as

$$f(x;\nu,\mu,\sigma,\gamma) \sim c|x|^{-\frac{\nu}{2}-1} \quad \text{as} \quad x \to +\infty. \tag{4.81}$$

The SKT decays as a power function and hence has tails heavier than the GHD.

4.6 The Expectation Maximization algorithm

The computation of maximum likelihood estimate of the GHD assumes the knowledge of GIG mixing variables. The problem is that the mixing variables are unknown and have to be inferred from the given time series. In order to facilitate the computation of the MLE when data is incomplete or being treated as incomplete an iterative technique called the Expectation Maximization (EM) algorithm can be used.

Dempster *et al.* (1977) discussed the general properties of the EM algorithm and proved that the technique can be used to find maximum likelihood estimates. Liu and Rubin (1995) discussed the EM algorithm and its extensions and used the techniques to calibrate the student's t distribution. The EM algorithm is an iterative procedure that consists of two steps at each iteration, namely the Expectation step (E step) and Maximization step (M step).

Let $\Pi = (\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ summarize the parameters of the GHD. Given a time series $x_1, \ldots, x_n$ of n independent random variables. The maximum likelihood estimate, is the value of Π that maximizes the logarithm of equation 4.62. The problem can be defined as

$$\textbf{maximize } \ln \mathscr{L}(\Pi; x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f_X(x_i; \Pi) \tag{4.82}$$

where $f_X(x; \Pi)$ denotes the probability density function of the GHD given by Eqn. 4.62.

Definition 4.10. Suppose that X and W are continuous random variables with joint probability density function $f_{X,W}(x, w)$ and marginal probability density functions $f_X(x)$ and $f_W(w)$ respectively. The conditional probability density function of X given $W = w$ is defined as

$$f_{X|W}(x|w) = \frac{f_{X,W}(x, w)}{f_W(w)}. \tag{4.83}$$

The problem defined by Eqn. 4.82 is not easy, particularly because of the number of parameters to be estimated. However, if the latent GIG mixing variables $w_1, \ldots, w_n$ were readily available, the problem could be recon-

structed using the concept of conditional probability as follows

$$\begin{aligned}\ln\mathscr{L}(\Pi; x_1, \ldots, x_n) &= \sum_{i=1}^{n} \ln f_{W_i}(w_i; \lambda, \chi, \psi) + \sum_{i=1}^{n} \ln f_{X_i|W_i}(x_i|w_i; \mu, \Sigma, \boldsymbol{\gamma}) \\ &= \mathscr{L}_1(\lambda, \chi, \psi; w_1, \ldots w_n) \\ &+ \mathscr{L}_2(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}; x_1, \ldots, x_n|w_1, \ldots, w_n), \end{aligned} \tag{4.84}$$

where $f_W(w)$ is the probability density function of the GIG given by Eqn. 4.1. Following Eqn. 4.59 we have that

$$f_{\mathbf{X}|W}(x|w) = \frac{e^{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma} - \frac{w}{2}\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}w^{\frac{d}{2}}} \times e^{-\frac{1}{2w}(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}, \tag{4.85}$$

where $|.|$ denotes the determinant.

The maximum likelihood estimates of the GHD can be obtained by maximizing $\mathscr{L}_1$ and $\mathscr{L}_2$ separately. From Eqn. 4.1 we get that

$$\begin{aligned}\mathscr{L}_1(\lambda, \chi, \psi; w_1, \ldots, w_n) = n \times \left(\frac{\lambda}{2}\ln\psi - \frac{\lambda}{2}\ln\chi - \ln(2K_\lambda(\sqrt{\chi\psi}))\right) \\ + (\lambda - 1)\sum_{i=1}^{n}\ln w_i - \frac{\chi}{2}\sum_{i=1}^{n}w_i^{-1} - \frac{\psi}{2}\sum_{i=1}^{n}w_i. \end{aligned} \tag{4.86}$$

Following Eqn. 4.85 we have that

$$\begin{aligned}\mathscr{L}_2 = \sum_{i=1}^{n}\ln f_{X_i|W_i}(x_i|w_i, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) = -\frac{n \times d}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{d}{2}\sum_{i=1}^{n}\ln w_i \\ + \sum_{i=1}^{n}(x_i - \boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma} - \frac{1}{2}\sum_{i=1}^{n}\frac{1}{w_i}Q(x_i) - \frac{1}{2}\boldsymbol{\gamma}'\Sigma\boldsymbol{\gamma}\sum_{i=1}^{n}w_i, \end{aligned} \tag{4.87}$$

where $Q(\mathbf{x}) = (\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)$.

After k cycles of the algorithm, let $\Pi^{(k)}$ denote the current estimate of Π.

E step

In this step one has to estimate the complete data sufficient statistics. In the case of the GHD this means one has to infer the values of the mixing variables w, $\frac{1}{w}$ and $\ln w$ using conditional expectations. The inferred values are then used obtain the values of $\boldsymbol{\mu}$, $\boldsymbol{\gamma}$ and Σ that maximizes $\mathscr{L}_2$. To maximize $\mathscr{L}_2$, compute the partial derivatives of $\mathscr{L}_2$ with respect to Σ, γ and μ and set the derivatives to be equal to zero.Then we get that

$$\gamma^{(k+1)} = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i}(\bar{x} - x_i)}{\frac{1}{n^2}\left(\sum_{i=1}^{n} w_i\right)\left(\sum_{i=1}^{n}\frac{1}{w_i}\right) - 1}, \tag{4.88}$$

$$\mu^{(k+1)} = \frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i}x_i - \gamma^{(k+1)}}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i}}, \tag{4.89}$$

$$\Sigma^{(k+1)} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{w_i}(x_i - \mu^{(k+1)})(x_i - \mu^{(k+1)})' - \frac{1}{n}\sum_{i=1}^{n} w_i \gamma^{(k+1)}\gamma^{(k+1)'}. \tag{4.90}$$

M step

Using the latent mixing variables, maximize $\mathscr{L}_1$ to get $\lambda^{(k+1)}$, $\chi^{(k+1)}$ and $\psi^{(k+1)}$.

The maximization of $\mathscr{L}_1$ corresponds to the maximum likelihood of the GIG which is explained in detail in section 4.1. Since λ is fixed in the case of the HYP and NIG, use algorithm 1 to compute $\chi^{(k+1)}$ and $\psi^{(k+1)}$. In the case of the GHD, first use algorithm 1 to compute $\chi^{(k+1)}$ and $\psi^{(k+1)}$ and then use Brent (1973) algorithm to compute $\lambda^{(k+1)}$.

In the case of the VG, $\lambda > 0$ and $\chi = 0$ and the mixing distribution (GIG) turns into the gamma distribution. The modified Bessel equation tends to the relation given by Eqn. 4.64. To obtain $\lambda^{(k+1)}$ and $\psi^{(k+1)}$ for the VG, compute the partial derivatives of $\mathscr{L}_1$ with respect to λ and ψ and set them equal to zero. After algebraic computations we obtain

$$\begin{aligned} \frac{\partial \mathscr{L}_1}{\partial \psi} = 0 &= -\frac{1}{2}\sum_{i=1}^{n} w_i + \frac{n\lambda}{\psi} \\ \implies \psi &= \frac{2n\lambda}{\sum_{i=1}^{n} w_i}. \end{aligned} \tag{4.91}$$

and

$$\frac{\partial \mathscr{L}_1}{\partial \lambda} = 0 = \sum_{i=1}^{n} \ln w_i + n\ln\psi - n\ln 2 - n\psi_0(\lambda), \tag{4.92}$$

where ψ_0 is the digamma function. After plugging in Eqn. 4.91 into Eqn. 4.92, ψ is cancelled and the remaining equation is only in terms of λ and is given by

$$0 = \sum_{i=1}^{n} \ln w_i - n\ln\sum_{i=1}^{n} w_i + n\ln n + n(\ln(\lambda) - \psi_0(\lambda)). \tag{4.93}$$

Solving for λ in Eqn. 4.93 is the same as solving for α in Eqn. 4.30. To obtain $\lambda^{(k+1)}$ for the VG, use the latent mixing variables $w_1, \ldots, w_n$ in algorithm 8 with $\zeta = 1$. The parameter $\psi^{(k+1)}$ is given by plugging the value of $\lambda^{(k+1)}$ in place of λ in Eqn. 4.91.

In the case of the SKT, $\lambda < 0$ and $\psi = 0$ and the mixing distribution (GIG) turns into the inverse gamma distribution. Using similar concepts to the ones used for the VG we have that

$$\chi = \frac{-2n\lambda}{\sum_{i=1}^{n} w_i^{-1}}, \tag{4.94}$$

and

$$0 = \sum_{i=1}^{n} w_i + n\ln \sum_{i=1}^{n} w_i^{-1} - n\ln n - n(\ln(-\lambda) + \psi_0(-\lambda)). \tag{4.95}$$

Since $\nu = 2\lambda$, if we set $\zeta = -1$ and multiply Eqn. 4.95 by -1 we get

$$0 = -\sum_{i=1}^{n} w_i - n\ln \sum_{i=1}^{n} w_i^{-1} + n\ln n + n(\ln(\frac{\nu}{2}) + \psi_0(\frac{\nu}{2})). \tag{4.96}$$

If we define

$$f(\alpha) = Q + n(\ln(\frac{\alpha}{2}) - \psi_0(\frac{\alpha}{2})), \tag{4.97}$$

then solving for $\frac{\nu}{2}$ is similar to solving for α using algorithm 8 with the output in the subroutine algorithm 9 computed using Eqn. 4.97. After algebraic computations, using corollary 4.4 the initial interval (a, b) in algorithm 8 becomes $(-\frac{n}{Q}, -\frac{-2n}{Q})$.

Solve for α using algorithm 8 as described above and then set $\lambda^{(k+1)} = -\alpha$. To obtain $\chi^{(k+1)}$, substitute the value of $\lambda^{(k+1)}$ in place of λ in Eqn. 4.94.

After computation of the E step and the M step, the $(k+1)th$ maximum likelihood estimate of the GHD will be

$$\Pi^{(k+1)} = (\lambda^{(k+1)}, \chi^{(k+1)}, \psi^{(k+1)}, \boldsymbol{\mu}^{(k+1)}, \Sigma^{(k+1)}, \boldsymbol{\gamma}^{(k+1)}).$$

4.6.1 Conditional density of the mixing variable

Since the latent mixing variables are not readily observable one must use the inferred values δ_i, η_i and ε_i in place of w_i^{-1}, w_i and $\ln w_i$ respectively. This section discusses how one can compute the values of δ_i, η_i and ε_i.
To perform the E step of the EM algorithm one must calculate the values of the latent mixing variables w_i, w_i^{-1} and $\ln w_i$, conditional on x_i. The following conditional density functions of the GHD models are derived.

GHD, NIG and HYP

$$\begin{aligned} f_{W|X}(w) &= \frac{f_{X,W}(x,w)}{f_X(x)} = \frac{f_{X,W}(x,w)f_{\text{GIG}}(w)}{\int_0^\infty f_{X|W}(x)f_{\text{GIG}}(w)\,\mathrm{d}w} \\ &= \left(\frac{\psi+\gamma'\Sigma^{-1}\gamma}{\chi+\rho(x)}\right)^{\frac{1}{2}(\lambda-\frac{d}{2})} w^{\lambda-\frac{d}{2}-1} \\ &\times \frac{\exp\{-\frac{1}{2}\left(w(\psi+\gamma'\Sigma^{-1}\gamma)+\frac{\chi+\rho(x)}{w}\right)\}}{2K_{\lambda-\frac{d}{2}}\sqrt{(\chi+\rho(x))(\psi+\gamma'\Sigma^{-1}\gamma)}} \end{aligned} \tag{4.98}$$

where $f_{X,W}(x,w)$ is given by Eqn. 4.62, $f_{\text{GIG}}(w)$ is given by Eqn. 4.1 and following equation 4.59, $f_{X|W}(x)$ is probability density function of the normal distribution when $X \sim N(\mu+W\gamma, W\Sigma)$ and

$$\rho(x) = (x-\mu)'\Sigma^{-1}(x-\mu). \tag{4.99}$$

From 4.98 we can see that $W|X \sim N^{\vdash}(\lambda-\frac{d}{2}, \chi+\rho(x), \psi+\gamma'\Sigma^{-1}\gamma)$. For convenience the notation in McNeil *et al.* (2005) is adopted, using equations 4.3 and 4.5 we obtain

$$\begin{aligned} \delta_i^{[k]} &= E[w_i^{-1}|x_i, \Pi^{[k]}] \\ &= \left(\frac{\chi^{[k]}+\rho(x_i)^{[k]}}{\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]}}\right)^{-\frac{1}{2}} \frac{K_{\lambda^{[k]}-\frac{d}{2}-1}\left((\chi^{[k]}+\rho(x_i)^{[k]})(\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]})\right)}{K_{\lambda^{[k]}-\frac{1}{2}}\left((\chi^{[k]}+\rho(x_i)^{[k]})(\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]})\right)}, \end{aligned} \tag{4.100}$$

$$\begin{aligned} \eta_i^{[k]} &= E[w_i|x_i, \Pi^{[k]}] \\ &= \left(\frac{\chi^{[k]}+\rho(x_i)^{[k]}}{\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]}}\right)^{\frac{1}{2}} \frac{K_{\lambda^{[k]}-\frac{d}{2}+1}\left((\chi^{[k]}+\rho(x_i)^{[k]})(\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]})\right)}{K_{\lambda^{[k]}-\frac{1}{2}}\left((\chi^{[k]}+\rho(x_i)^{[k]})(\psi^{[k}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]})\right)}, \end{aligned} \tag{4.101}$$

and

$$\begin{aligned} \xi_i^{[k]} &= E[\ln w_i|x_i, \Pi^{[k]}] \\ &= \frac{1}{2}\ln\left(\frac{\chi^{[k]}+\rho(x_i)^{[k]}}{\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]}}\right) \frac{\frac{\partial K_{(\lambda^{[k]}-\frac{1}{2})+\alpha}\left((\chi^{[k]}+\rho(x_i)^{[k]})(\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]})\right)}{\partial\alpha}|_{\alpha=0}}{K_{\lambda^{[k]}-\frac{1}{2}}\left((\chi^{[k]}+\rho(x_i)^{[k]})(\psi^{[k]}+\gamma^{[k]\prime}\Sigma^{[k]^{-1}}\gamma^{[k]})\right)}. \end{aligned} \tag{4.102}$$

When λ is fixed there is no need to calibrate ξ_i and in order to calculate the derivatives around zero in ξ_i, numerical methods are used.

VG

Following the calculations of the conditional distribution for the GHD, for the VG we have that

$$\begin{aligned} f_{W|X}(w) &= \frac{f_{X,W}(x,w)}{f_X(x)} = \frac{f_{X,W}(x,w)f_\Gamma(w)}{\int_0^\infty f_{X|W}(x)f_\Gamma(w)\,\mathrm{d}w} \\ &= \left(\frac{\psi+\gamma'\Sigma^{-1}\gamma}{\rho(x)}\right)^{\frac{1}{2}(\lambda-\frac{d}{2})} w^{\lambda-\frac{d}{2}-1} \\ &\times \frac{\exp\{-\frac{1}{2}\left(w(\psi+\gamma'\Sigma^{-1}\gamma)+\frac{\rho(x)}{w}\right)\}}{2K_{\lambda-\frac{d}{2}}\sqrt{\rho(x)(\psi+\gamma'\Sigma^{-1}\gamma)}} \end{aligned} \tag{4.103}$$

where $f_\Gamma(w)$ is given by Eqn. 4.22. From Eqn. 4.103 we can see that for the $W|X \sim N^{\vdash}(\lambda-\frac{d}{2}, \rho(x), \psi+\gamma'\Sigma^{-1}\gamma)$, therefore substitute χ by 0 in equations 4.100, 4.101 and 4.102 to calculate the latent mixing variables.

When $\rho(x_i) \to 0$, the latent mixing variables are computed by the following formulas

$$\delta_i^{[k]} = \frac{(\psi^{[k]}+\gamma^{[k]\prime}{\Sigma^{[k]}}^{-1}\gamma^{[k]})}{2(\lambda^{[k]}-\frac{d}{2})}, \tag{4.104}$$

$$\eta_i^{[k]} = \frac{2(\lambda^{[k]}-\frac{d}{2})}{(\psi^{[k]}+\gamma^{[k]\prime}{\Sigma^{[k]}}^{-1}\gamma^{[k]})}, \tag{4.105}$$

$$\xi_i^{[k]} = \ln\left(\frac{\psi^{[k]}+\gamma^{[k]\prime}}{\gamma^{[k]\prime}{\Sigma^{[k]}}^{-1}\gamma^{[k]}}2\right) + \psi_0\left(\lambda^{[k]}-\frac{d}{2}\right). \tag{4.106}$$

where $\psi_0(\cdot)$ is the digamma function.

SKT

Following the calculations of the conditional distribution for the GHD, for the SKT we have that

$$\begin{aligned} f_{W|X}(w) &= \frac{f_{X,W}(x,w)}{f_X(x)} = \frac{f_{X,W}(x,w)f_{I\Gamma}(w)}{\int_0^\infty f_{X|W}(x)f_{I\Gamma}(w)\,\mathrm{d}w} \\ &= \left(\frac{\gamma'\Sigma^{-1}\gamma}{\chi+\rho(x)}\right)^{\frac{1}{2}(\lambda-\frac{d}{2})} w^{\lambda-\frac{d}{2}-1} \\ &\times \frac{\exp\{-\frac{1}{2}\left(w(\gamma'\Sigma^{-1}\gamma)+\frac{\chi+\rho(x)}{w}\right)\}}{2K_{\lambda-\frac{d}{2}}\sqrt{(\chi+\rho(x))(\gamma'\Sigma^{-1}\gamma)}} \end{aligned} \tag{4.107}$$

where $f_{I\Gamma}(w)$ is given by Eqn. 4.23. From Eqn. 4.103 we can see that for the $W|X \sim N^{\vdash}(-\frac{\nu}{2}-\frac{d}{2}, \chi+\rho(x), \gamma'\Sigma^{-1}\gamma)$, therefore substitute ψ by 0 and

λ by $-\frac{\nu}{2}$ in equations 4.100, 4.101 and 4.102 to calculate the latent mixing variables.
When $\gamma \to 0$, the latent mixing variables are computed by the following formulas

$$\delta_i^{[k]} = \frac{\nu^{[k]} + d}{\chi^{[k]} + \rho(x_i)^{[k]}}, \tag{4.108}$$

$$\eta_i^{[k]} = \frac{\rho(x_i)^{[k]} + \chi^{[k]}}{\nu^{[k]} + d - 2}, \tag{4.109}$$

$$\xi_i^{[k]} = -\ln\left(\frac{\rho(x_i)^{[k]} + \chi^{[k]}}{2}\right) + \psi_0\left(\frac{\nu^{[k]} + d}{2}\right). \tag{4.110}$$

where $\psi_0(\cdot)$ is the digamma function.

4.6.2 Procedure when calibrating the GHD

The problem with the representation by McNeil *et al.* (2005) is that for any $k > 0$, the GHD with parameters $(\lambda, \frac{\chi}{k}, k\psi, \mu, k\Sigma, k\gamma)$ have the same distribution with the GHD with parameters $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$. To address this *identifiablity problem* one can restrict the determinant of Σ to be equal to one, however we constrained the expectation of the GIG mixing variable to 1. Set

$$\bar{\omega} = \sqrt{\chi\psi}, \tag{4.111}$$

following Eqn. 4.3, define

$$\mathbb{E}[W] = \left(\frac{\chi}{\psi}\right)^{\frac{1}{2}} \frac{K_{\lambda+1}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} = 1. \tag{4.112}$$

It follows from equations 4.111 and 4.112 that

$$\psi = \bar{\omega}\frac{K_{\lambda+1}(\bar{\omega})}{K_\lambda(\bar{\omega})} \quad \text{and} \quad \chi = \bar{\omega}\frac{K_\lambda(\bar{\omega})}{K_{\lambda+1}(\bar{\omega})} \tag{4.113}$$

Since the Bessel function is symmetric (i.e. $K_\lambda(x) = K_{-\lambda}(x)$), we have $\psi = \chi = \bar{\omega}$ for the NIG.

For the VG, $\chi = 0$ by definition. Using the relation given by Eqn. 4.64, ψ transforms to

$$\psi = \bar{\omega}\frac{K_{\lambda+1}(\bar{\omega})}{K_\lambda(\omega)} \xrightarrow{\bar{\omega}\to 0} 2\lambda. \tag{4.114}$$

For the SKT, $\psi = 0$ by definition. Using Eqn. 4.64 we obtain that

$$\chi = \bar{\omega}\frac{K_\lambda(\bar{\omega})}{K_{\lambda+1}(\omega)} \xrightarrow{\bar{\omega}\to 0} \nu - 2. \tag{4.115}$$

Therefore in the M step of the EM algorithm instead of computing $\psi^{[k+1]}$ for the VG as described before, simply set $\psi^{[k+1]} = 2\lambda^{[k+1]}$. Similarly for the SKT set $\chi^{[k+1]} = \nu^{[k+1]} - 2$.

4.7 The benchmark and goodness of fit

The normal distribution is one of the most commonly used models in financial modeling. Due to this reason the normal distribution was set as a benchmark in this project. In this study we compare the GHD and its subclasses to the normal distribution. The GHD models were calibrated using the Expectation Maximization algorithm.

The goodness of fit measures that were used are the Bayesian information criterion per observation (BIC_0), the Kolmogorov Smirnov (KS) test and the Anderson Darling (AD). For the three criteria, the lower the value the better the fit. The extensive simulations of different sample sizes conducted by Laio *et al.* (2009) based on numerous distributions concluded that the AD generally performs poorly in comparison the Bayesian information criterion (BIC) and eventually the BIC_0.

The AD test is more sensitive to the tails of the distribution in comparison to the KS test, for this reason the AD is considered the more powerful test of the two. Let F_{est} be the estimated cumulative distribution and F_{emp} the empirical cumulative distribution of a random variable x. The KS and AD distances for a continuous distribution are given by

$$\text{KS} = \max_{x\in\mathbb{R}}\{|F_{est}(x) - F_{emp}(x)|\}, \tag{4.116}$$

$$\text{AD} = \max_{x\in\mathbb{R}}\left\{\frac{|F_{emp}(x) - F_{est}(x)|}{\sqrt{F_{est}(x)(1-F_{est}(x))}}\right\}. \tag{4.117}$$

Let n be the number of estimated parameters and $\mathscr{L}$ denote the log likelihood, then the famous Akaike information criterion (AIC) is given by

$$\text{AIC} = -2\mathscr{L} + 2n, \tag{4.118}$$

whilst the BIC is given by

$$\text{BIC} = -2\mathscr{L} + n\ln T. \tag{4.119}$$

where T is the sample size. The AIC and BIC penalizes models according to the number of estimated parameters. From equations 4.118 and 4.119,

the number of parameters (n) is multiplied by a factor of $\frac{1}{2}\ln T$ with respect to the AIC. Due to this fact, when the sample size is at least eight the BIC leans more to lower dimensional models than the AIC. To avoid reporting very large numbers, the BIC_0 was reported instead of the BIC. The BIC_0 is simply the BIC divided by the number of observations and is given by

$$\text{BIC}_0 = -2\frac{\mathscr{L}}{T} + n\frac{\ln T}{T}. \tag{4.120}$$

4.8 Fitting to financial data

Log returns of three stocks from the JSE were fitted to the GHD and normal distribution. The motivation and summary statistics of the data used is given in chapter 2. The three stocks used in the analysis are Anglo American PLC (AGL), MTN Group Ltd (MTN) and Naspers Ltd (NPN). The sampling time spans that were calibrated are daily (8 hours), 4 hours, 2 hours, 1 hour, 30 minutes and 20 minutes. Sampling frequencies below 20 minutes were not considered because the time series failed the independent increments test.

Table 4.3, 4.4 and 4.5 show the parameters obtained for the different stocks. Note that for all stocks the variability parameter Σ generally decreases as the sampling time span decreases.

Sampling Time Span	Class	λ	χ	ψ	μ	Σ	γ
Daily	GHD	5.9543	0.0829386	12.0635	-2.629e-03	2.593e-04	2.566e-03
	HYP	1	0.383763	2.71853	-1.787e-03	2.815e-04	1.715e-03
	NIG	-0.5	1.12698	1.12698	-1.717e-03	2.960e-04	1.659e-03
	SKT	-3.31654	4.63309	0	-2.319e-03	2.595e-04	2.251e-03
	VG	5.8908	0	11.7816	-1.559e-03	2.603e-04	1.867e-03
4 hours	GHD	2.50924	0.189004	5.29278	3.434e-04	1.239e-04	-3.777e-04
	HYP	1	0.376465	2.70811	3.290e-04	1.295e-04	-3.633e-04
	NIG	-0.5	1.07536	1.07536	8.798e-05	1.356e-04	-1.232e-04
	SKT	-1.46925	0.9385	0	6.471e-05	1.244e-04	-9.956e-05
	VG	5.8296	0	11.6592	3.655e-04	1.217e-04	-4.009e-04
2 hours	GHD	-0.0566198	0.677339	1.3557	-3.565e-05	5.770e-05	1.852e-05
	HYP	1	0.362188	2.68761	-3.659e-05	5.557e-05	1.922e-05
	NIG	-0.5	0.958482	0.958482	-3.490e-05	5.722e-05	1.743e-05
	SKT	-2.02715	2.0543	0	3.017e-05	5.941e-05	-5.963e-05
	VG	1.11225	0	2.22451	-1.893e-05	5.986e-05	1.516e-06
1 hour	GHD	-1.04082	1.4404	0.681224	-7.481e-07	3.074e-05	-8.037e-06
	HYP	1	0.369523	2.69817	-1.259e-05	2.971e-05	3.902e-06
	NIG	-0.5	0.993006	0.993006	-8.371e-06	3.073e-05	-3.344e-07
	SKT	-1.62341	1.24683	0	1.160e-04	3.221e-05	-1.250e-04
	VG	1.79893	0	3.59786	-2.444e-05	3.064e-05	1.586e-05
30 minutes	GHD	-0.591757	1.00526	0.938405	-3.955e-05	1.571e-05	3.614Ee-05
	HYP	1	0.36652	2.69385	-4.467e-05	1.495e-05	4.050e-05
	NIG	-0.5	0.965995	0.965995	-3.914e-05	1.541e-05	3.481e-05
	SKT	-1.7365	1.473	0	1.193e-04	1.632e-05	-1.245e-04
	VG	1.26552	0	2.53104	-1.496e-05	1.607e-05	1.069e-05
20 minutes	GHD	-0.599161	1.0101	0.938772	-6.346e-06	1.027e-05	3.545e-06
	HYP	1	0.367327	2.69501	-3.096e-06	9.760e-06	1.957e-07
	NIG	-0.5	0.96899	0.96899	-6.275e-06	1.007e-05	3.378e-06
	SKT	-1.79525	1.59049	0	1.136e-04	1.066e-05	-1.175e-04
	VG	1.22003	0	2.44006	3.251e-07	1.051e-05	-3.249e-06

Tab. 4.3: Estimated parameters of the GHD for AGL log returns from 01 October 2012 to 30 September 2013.

Sampling Time Span	Class	λ	χ	ψ	μ	Σ	γ
Daily	GHD	4.9996	0.0988085	10.1211	-1.311e-03	2.408e-04	9.225e-05
	HYP	1	0.377111	2.70904	-1.343e-03	2.581e-04	1.247e-04
	NIG	-0.5	1.10252	1.10252	-1.285e-03	2.717e-04	6.677e-05
	SKT	-1.49351	0.98702	0	-1.371e-03	2.405e-04	1.526e-04
	VG	5.1637	0	10.3274	-1.284e-03	2.413e-04	8.223e-05
4 hours	GHD	1.53057	0.281942	3.52177	-6.772e-04	1.050e-04	6.755e-05
	HYP	1	0.367469	2.69522	-6.735e-04	1.070e-04	6.419e-05
	NIG	-0.5	1.04077	1.04077	-7.120e-04	1.120e-04	1.027e-04
	SKT	-1.34282	0.685633	0	-6.673e-04	1.048e-04	5.812e-05
	VG	1.82821	0	3.65643	-3.692e-04	1.069e-04	-2.378e-04
2 hours	GHD	-0.0200158	0.671977	1.37621	-2.375E-04	4.906e-05	-6.746e-05
	HYP	1	0.361028	2.68594	-2.287e-04	4.754e-05	-7.594e-05
	NIG	-0.5	0.965259	0.965259	-2.423e-04	4.909e-05	-6.209e-05
	SKT	-1.29394	0.587882	0	-2.574e-04	5.171e-05	-4.724e-05
	VG	1.09856	0	2.19713	2.660e-05	5.067e-05	-3.297e-04
1 hour	GHD	-0.0538801	0.668363	1.34398	-1.909e-04	2.578e-05	3.923e-05
	HYP	1	0.359721	2.68405	-1.912e-04	2.483e-05	3.898e-05
	NIG	-0.5	0.951518	0.951518	-1.935e-04	2.559e-05	4.107e-05
	SKT	-1.62477	1.24954	0	-2.694e-04	2.714e-05	1.173e-04
	VG	1.08637	0	2.17274	-1.489e-04	2.670e-05	-3.376e-06
30 minutes	GHD	-0.0930063	0.663691	1.30774	-5.223e-05	1.300e-05	-2.454e-05
	HYP	1	0.358184	2.68183	-5.103e-05	1.246e-05	-2.519e-05
	NIG	-0.5	0.932504	0.932504	-5.387e-05	1.278e-05	-2.216e-05
	SKT	-1.98575	1.9715	0	-6.131e-05	1.368e-05	-1.860e-05
	VG	1.08387	0	2.16775	-5.959e-05	1.353e-05	-1.665e-05
30 minutes	GHD	-0.586311	0.967996	0.895562	-8.642e-06	8.839e-06	-4.312e-05
	HYP	1	0.35746	2.68078	-3.295e-06	8.461e-06	-4.766e-05
	NIG	-0.5	0.91797	0.91797	-8.958e-06	8.655e-06	-4.161e-05
	SKT	-1.41483	0.829657	0	-1.697e-05	9.234e-06	-3.381e-05
	VG	1.08628	0	2.17256	-6.378e-06	9.286e-06	-4.496e-05

Tab. 4.4: Estimated parameters of the GHD for MTN log returns from 01 October 2012 to 30 September 2013.

Sampling Time Span	Class	λ	χ	ψ	μ	Σ	γ
Daily	GHD	2.49209	0.189442	5.27948	4.640e-04	1.868e-04	6.790e-04
	HYP	1	0.374804	2.70574	6.239e-04	1.945e-04	5.176e-04
	NIG	-0.5	1.07256	1.07256	6.039e-04	2.038e-04	5.399e-04
	SKT	-3.07932	4.15865	0	-6.897e-04	1.871e-04	1.839e-03
	VG	6.00185	0	12.0037	2.231e-04	1.832e-04	9.203e-04
4 hours	GHD	1.4628	0.286765	3.46907	-1.391e-04	8.525e-05	7.153e-04
	HYP	1	0.369008	2.69743	-1.270e-04	8.604e-05	6.977e-04
	NIG	-0.5	1.03697	1.03697	-5.038e-05	8.999e-05	6.222e-04
	SKT	-1.73681	1.47361	0	2.271e-04	1.123e-04	3.813e-04
	VG	1.88104	0	3.76209	-1.422e-04	8.657e-05	7.118e-04
2 hours	GHD	-0.14925	0.641865	1.2344	-1.977e-04	4.196e-05	5.036e-04
	HYP	1	0.352135	2.67307	-2.124E-04	3.996E-05	4.996E-04
	NIG	-0.5	0.889786	0.889786	-1.937e-04	4.065e-05	4.753e-04
	SKT	-1.93127	1.86253	0	-6.449e-05	4.368e-05	3.506e-04
	VG	1.06112	0	2.12224	6.268e-05	4.397e-05	2.265e-04
1 hour	GHD	-0.567684	0.969964	0.912349	1.211e-05	2.107e-05	1.330e-04
	HYP	1	0.358977	2.68298	4.434e-06	2.022e-05	1.388e-04
	NIG	-0.5	0.929726	0.929726	1.279e-05	2.072e-05	1.294e-04
	SKT	-1.50152	1.00304	0	2.892e-05	2.804e-05	1.269e-04
	VG	1.08885	0	2.17771	4.441e-05	2.208e-05	9.910e-05
30 minutes	GHD	-0.639847	0.986981	0.878257	-1.205e-04	1.011e-05	1.997e-04
	HYP	1	0.359705	2.68403	-1.336e-04	9.580e-06	2.062e-04
	NIG	-0.5	0.917256	0.917256	-1.187e-04	9.795e-06	1.897e-04
	SKT	-1.72288	1.44575	0	-5.713e-05	1.112e-05	1.294e-04
	VG	1.07291	0	2.14581	5.348e-05	1.063e-05	1.825e-05
20 minutes	GHD	-0.657783	0.998567	0.879107	2.250e-05	6.933e-06	2.628e-05
	HYP	1	0.3618	2.68705	1.890e-05	6.540e-06	2.888e-05
	NIG	-0.5	0.924289	0.924289	2.274e-05	6.697e-06	2.484e-05
	SKT	-0.5	0.924289	0.924289	2.274e-05	6.697e-06	2.484e-05
	VG	1.08111	0	2.16221	3.910e-05	7.282e-06	8.674e-06

Tab. 4.5: Estimated parameters of the GHD for NPN log returns from 01 October 2012 to 30 September 2013.

Sampling	RIC	AGL			MTN			NPN		
Time Span	Class	KS	AD	BIC_0	KS	AD	BIC_0	KS	AD	BIC_0
Daily	GHD	0.035461	0.10095	-5.2945	0.026789	0.09315	-5.3709	0.029003	0.070963	-5.643
	HYP	0.046032	0.098151	-5.2975	0.031662	0.12014	-5.3797	0.036907	0.091577	-5.6607
	NIG	0.044725	0.10721	-5.2926	0.03151	0.12734	-5.3718	0.038331	0.09536	-5.6552
	SKT	0.041423	0.090399	-5.3292	0.081396	0.2657	-5.2753	0.030149	0.076053	-5.6835
	VG	0.045629	0.096328	-5.3385	0.027347	0.093548	-5.4153	0.033437	0.070452	-5.6823
	Gaussian	0.039114	0.47189	-5.3756	0.045447	0.58277	-5.7059	0.04166	0.35057	-5.4515
4 hours	GHD	0.019061	0.050183	-6.1098	0.034601	0.071492	-6.2865	0.041458	0.083054	-6.5052
	HYP	0.026422	0.063102	-6.1175	0.032034	0.074607	-6.2978	0.040086	0.080319	-6.5175
	NIG	0.025715	0.074627	-6.1129	0.030741	0.081045	-6.2895	0.039796	0.084865	-6.5125
	SKT	0.075404	0.20456	-6.0217	0.090557	0.29142	-6.1515	0.043601	0.098047	-6.5145
	VG	0.02293	0.084394	-6.1302	0.034163	0.072212	-6.3149	0.039473	0.079096	-6.5335
	Gaussian	0.035367	0.99777	-6.1352	0.053927	2.018	-6.503	0.054533	1.8012	-6.3092
2 hours	GHD	0.018575	0.05254	-6.9789	0.027444	0.056092	-7.1303	0.035762	0.092257	-7.309
	HYP	0.022661	0.088339	-6.9793	0.030716	0.082822	-7.1332	0.044318	0.11673	-7.2986
	NIG	0.019452	0.050831	-6.9853	0.028108	0.057434	-7.1357	0.03708	0.08677	-7.314
	SKT	0.02231	0.065642	-6.9862	0.072599	0.23134	-7.0432	0.041512	0.10648	-7.3076
	VG	0.013966	0.047244	-6.9961	0.022275	0.04466	-7.1573	0.034627	0.069267	-7.3409
	Gaussian	0.063024	9.0436	-6.8803	0.091496	16.417	-7.1482	0.064109	8.2839	-7.0500

Tab. 4.6: Goodness of fit for AGL, MTN and NPN log returns from October 2012 to September 2013.

Sampling	RIC	AGL			NPN			MTN		
Time Span	Class	KS	AD	BIC_0	KS	AD	BIC_0	KS	AD	BIC_0
60 minutes	GHD	0.013461	0.07724	-7.6329	0.026751	0.055419	-7.8001	0.023071	0.059977	-8.0105
	HYP	0.013162	0.13997	-7.6298	0.031595	0.065281	-7.7978	0.026637	0.107	-8.0037
	NIG	0.013247	0.078163	-7.6361	0.027621	0.057193	-7.8023	0.022868	0.058801	-8.0148
	SKT	0.03013	0.072941	-7.6274	0.029191	0.1012	-7.7919	0.019859	0.043212	-8.0172
	VG	0.017293	0.15537	-7.6299	0.01606	0.038026	-7.8236	0.017256	0.055744	-8.0262
	Gaussian	0.054081	14.024	-7.5239	0.072389	23.723	-7.861	0.064284	Inf	-7.6828
30 minutes	GHD	0.013018	0.19411	-8.3316	0.023292	0.049999	-8.5033	0.025	0.083316	-8.7753
	HYP	0.015549	0.66093	-8.3236	0.030008	0.084215	-8.4953	0.028417	0.20653	-8.7608
	NIG	0.013054	0.19427	-8.3337	0.024667	0.051787	-8.5043	0.025113	0.081956	-8.778
	SKT	0.021434	0.057727	-8.3345	0.02773	0.058123	-8.4985	0.022115	0.04753	-8.7831
	VG	0.015246	0.38612	-8.3287	0.014155	0.041158	-8.519	0.024455	0.083502	-8.7868
	Gaussian	0.061693	38.042	-8.1757	0.07733	Inf	-8.5699	0.072343	45.456	-8.3576
20 minutes	GHD	0.0092743	0.1694	-8.763	0.025851	0.054272	-8.8968	0.018307	0.23117	-9.1658
	HYP	0.015469	0.3228	-8.754	0.031545	0.13163	-8.8856	1	1855.1	-9.1503
	NIG	0.010175	0.17236	-8.7643	0.02557	0.053692	-8.8991	0.01857	0.23779	-9.1677
	SKT	0.015342	0.040366	-8.7669	0.040376	0.13546	-8.8756	0.015783	0.069082	-9.1726
	VG	0.075709	0.1889	-8.7591	0.017686	0.067159	-8.9112	0.017821	0.62218	-9.1719
	Gaussian	0.057863	55.949	-8.6008	0.073725	Inf	-8.9463	0.076359	78.88	-8.7207

Tab. 4.7: Goodness of fit for AGL, MTN and NPN log returns from October 2012 to September 2013.

From tables 4.6 and 4.7 one can see that the model fit for the GHD become better as the sampling time span decreases, whilst the performance of the performance of the normal distribution diminishes. For all the three stocks fitted, the normal distribution gives the best fit for daily returns.

For mid range sampling time spans (1 and 2 hours), the VG has the best fit for all three stocks except for AGL at 1 hour where the best fit is given by NIG. The SKT gives the best fit for AGL and NPN for 20 minute returns whilst the best fit for MTN is given by the VG.

For visual comparisons of models, density plots of the calibrated models were plotted over a histogram of the log returns of NPN. Figure 4.9 - 4.12 shows the density plots of the GHD as well as the density plot of the normal distribution on the histogram of the data. In each case the curve with the lowest peak the normal curve. The distance of the peak of the normal curve to the peak of the GHD curves become more pronounced as the sampling time span decreases. The peak of the VG becomes sharper as the sampling time span decreases.

To assess the temporal stability of the parameters thirty nine, 12 months windows were considered. The first window started on the 3rd of January 2012 and ended on the 28th December 2013 and the last period started on the 25th of September 2012 and ended on the 23rd of September 2013. The beginning of each period was the first business day of each week. To ensure that each window had the same number of observations, observations from the day of the period began and observations from the 249 business days that followed it were used.

Table 4.8 shows that the variability of the parameters obtained for the GHD from rolling time windows generally decreases with increasing sampling frequency. The hyperbolic distribution has smallest variability followed by the NIG for the parameters from the GIG. The highest variability of the GIG parameters is observed from the VG.

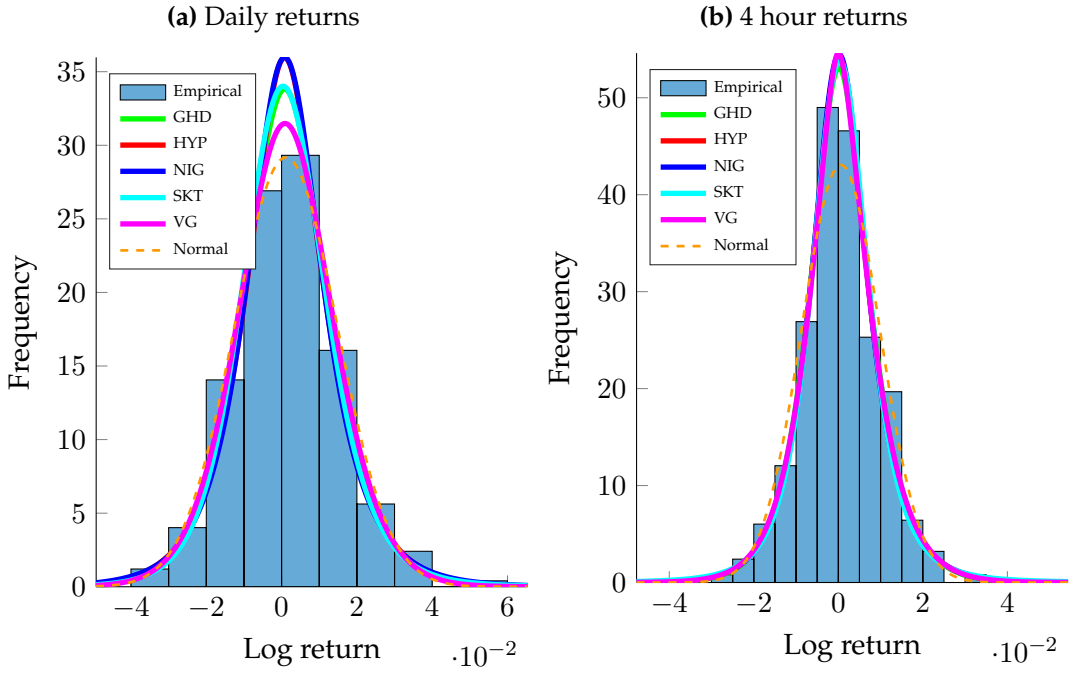


Fig. 4.9: Density plots of the fitted GHD for NPN log returns from Oct 2012 to Sep 2013.

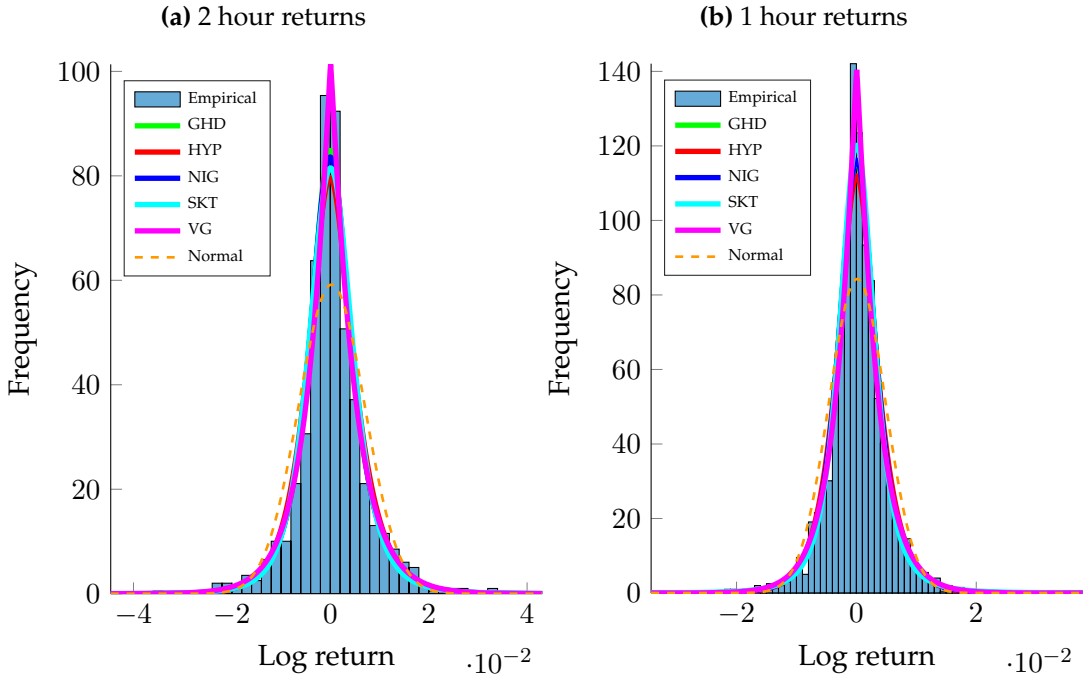


Fig. 4.10: Density plots of the fitted GHD for NPN log returns from Oct 2012 to Sep 2013.

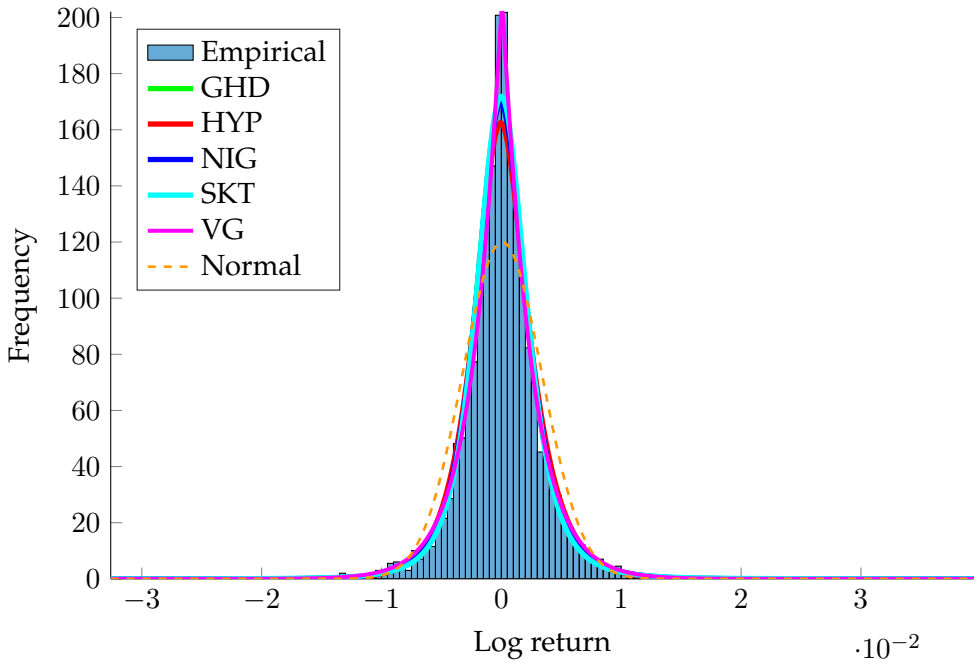


Fig. 4.11: Density plots of the fitted GHD for 30 minute NPN log returns from Oct 2012 to Sep 2013.

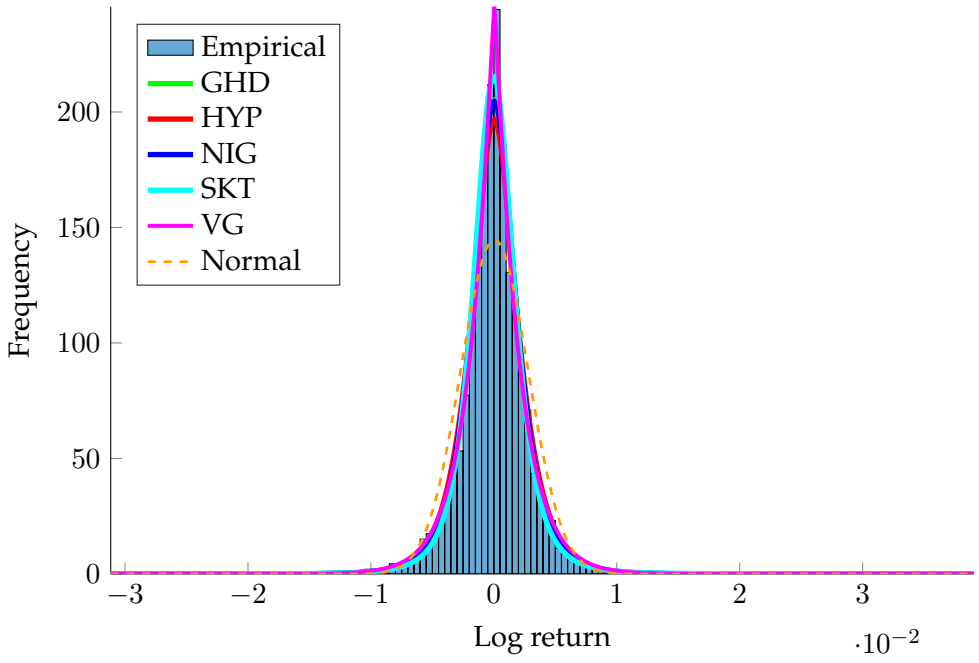


Fig. 4.12: Density plots of the fitted GHD for 20 minute NPN log returns from Oct 2012 to Sep 2013.

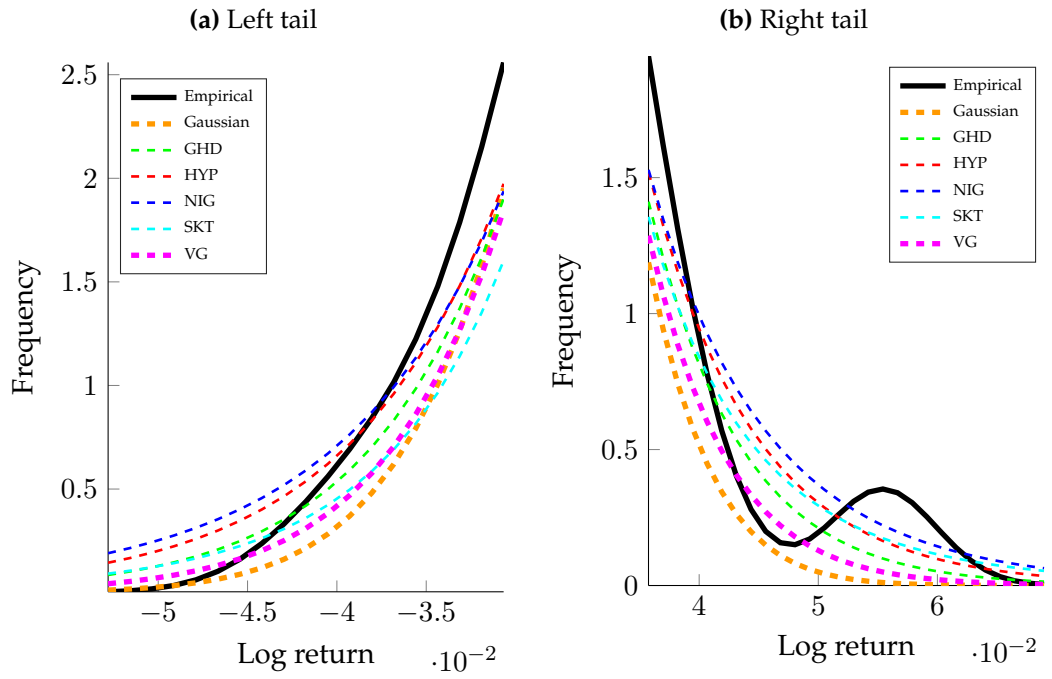


Fig. 4.13: Density plots of the tails for daily NPN log returns from Oct 2012 to Sep 2013.

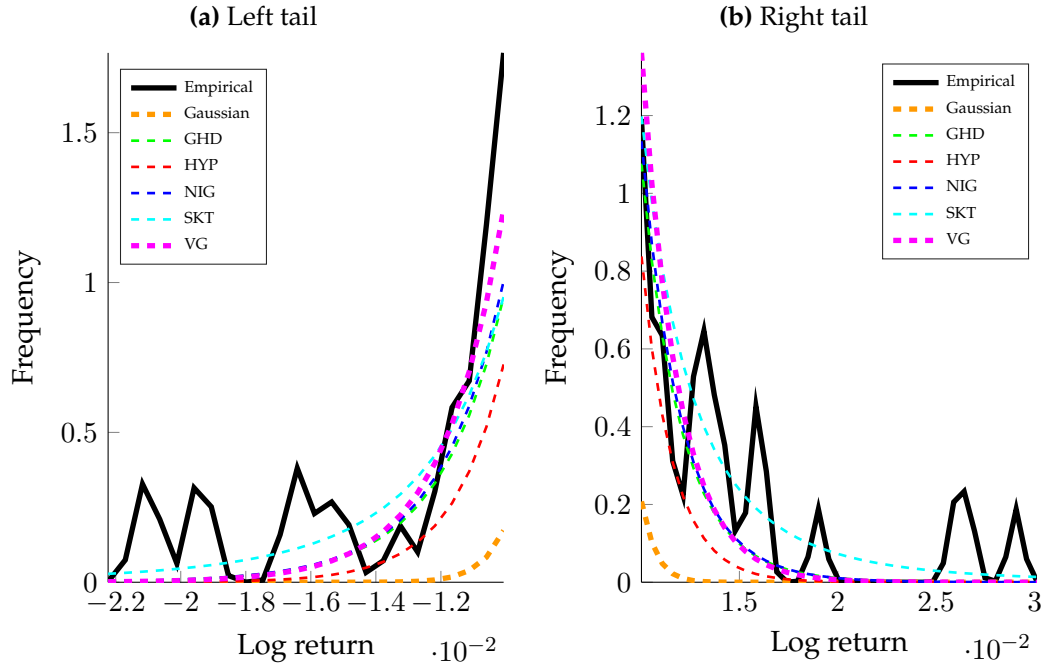


Fig. 4.14: Density plots of the tails for 20 minute NPN log returns from Oct 2012 to Sep 2013.

Class	Sampling Time Span	λ		χ		ψ	
		Mean	Std	Mean	Std	Mean	Std
GHD	Daily	0.0959457	1.93426	1.3121	1.24226	2.38952	2.07543
	4 hours	1.67539	0.242439	0.264403	0.0278524	3.81619	0.425685
	2 hours	-0.0723973	0.0463053	0.662807	0.00933614	1.3234	0.048029
	1 hour	-0.124547	0.185144	0.72506	0.105947	1.30069	0.161814
	30 minutes	-0.666789	0.200146	1.01738	0.0798745	0.892925	0.0616666
	20 minutes	-0.588555	0.0359125	0.987317	0.00530395	0.917269	0.0209195
HYP	Daily	1	0 0.374231	0.00134987	2.70492	0.0019321	
	4 hours	1	0	0.370451	0.00143348	2.6995	0.00205673
	2 hours	1	0	0.358423	0.00286256	2.68217	0.00413755
	1 hours	1	0	0.362704	0.00199525	2.68835	0.00287678
	30 minutes	1	0	0.365094	0.0084701	2.69177	0.0121357
	20 minutes	1	0	0.361908	0.000412649	2.68721	0.000595117
NIG	Daily	-0.5	0	1.04521	0.019434	1.04521	0.019434
	4 hours	-0.5	0	1.04911	0.00721838	1.04911	0.00721838
	2 hours	-0.5	0	0.93812	0.0242618	0.93812	0.0242618
	1 hour	-0.5	0	0.962416	0.0176338	0.962416	0.0176338
	30 minutes	-0.5	0	0.940752	0.0106999	0.940752	0.0106999
	20 minutes	-0.5	0	0.94246	0.0105366	0.94246	0.0105366
SKT	Daily	-2.21444	0.503127	2.42888	1.00625	0	0
	4 hours	-2.26455	0.727089	2.52911	1.45418	0	0
	2 hours	-1.85807	0.207107	1.71613	0.414214	0	0
	1 hour	-1.60011	0.273583	1.20021	0.547166	0	0
	30 minutes	-1.68837	0.194918	1.37673	0.389836	0	0
	20 minutes	-1.76395	0.25882	1.5279	0.51764	0	0
VG	3.36606	1.36969	0	0	6.73212	2.73939	
	4 hours	2.37211	0.532116	0	0	4.74421	1.06423
	2 hours	1.09322	0.0299914	0	0	2.18644	0.0599829
	1 hour	1.34699	0.325219	0	0	2.69398	0.650438
	30 minutes	1.10298	0.0244101	0	0	2.20597	0.0488201
	20 minutes	1.09275	0.00896842	0	0	2.1855	0.0179368

Tab. 4.8: The mean and standard deviation of the mixing (GIG) parameters of the GHD obtained from rolling time windows.

Class	Sampling Time Span	μ		Σ		γ	
		Mean	Std	Mean	Std	Mean	Std
GHD	Daily	-5.045e-04	2.352e-03	1.822e-04	5.910e-05	1.752e-03	2.678e-03
	4 hours	-2.947e-04	5.010e-04	8.447e-05	1.951e-05	9.227e-04	6.513e-04
	2 hours	-2.729e-04	1.829e-e4	4.198e-05	7.317e-06	6.093e-04	1.729e-04
	1 hour	3.151e-05	6.125e-05	2.125e-05	3.134e-06	2.125e-05	7.064e-05
	30 minutes	-2.855e-05	8.377e-05	1.009e-05	1.439e-06	1.105e-04	9.633e-05
	20 minutes	4.322e-05	2.217e-05	6.978e-06	8.640e-07	9.730e-06	3.735e-05
HYP	Daily	2.326e-04	2.060e-03	1.903e-04	6.543e-05	9.881e-04	2.410e-03
	4 hours	-3.540e-04	3.770e-04	8.537e-05	1.974e-05	9.727e-04	5.314e-04
	2 hours	-2.919e-04	1.825e-04	4.002e-05	6.831e-06	6.059e-04	1.644e-04
	1 hour	2.293e-05	6.114e-05	2.038e-05	2.908e-06	2.038e-05	7.340e-05
	30 minutes	-3.637e-05	1.045e-04	9.583e-06	1.339e-06	1.149e-04	1.175e-04
	20 minutes	4.006e-05	2.149e-05	6.605e-06	8.169e-07	1.248e-05	3.558e-05
NIG	Daily	2.192e-04	1.925e-03	2.000e-04	6.942e-05	1.010e-03	2.281e-03
	4 hours	-1.747e-04	4.375e-04	8.943e-05	2.083e-05	7.945e-04	4.978e-04
	2 hours	-2.689e-04	1.810e-04	4.080e-05	6.832e-06	5.768e-04	1.611e-04
	1 hour	3.470e-05	6.241e-05	2.095e-05	2.877e-06	2.095e-05	6.973e-05
	30 minutes	-2.769e-05	8.312e-05	9.830e-06	1.332e-06	1.054e-04	9.184e-05
	20 minutes	4.330e-05	2.205e-05	6.773e-06	8.173e-07	9.113e-06	3.580e-05
SKT	Daily	-8.467e-04	2.663e-03	1.931e-04	5.929e-05	2.069e-03	2.981e-03
	4 hours	-1.373e-04	6.033e-04	9.872e-05	4.322e-05	7.730e-04	6.514e-04
	2 hours	-1.375e-04	1.730e-04	4.429e-05	7.282e-06	4.525e-04	1.318e-04
	1 hour	6.051e-05	9.286e-05	2.421e-05	4.739e-06	2.421e-05	1.032e-04
	30 minutes	-1.253e-05	9.297e-05	1.086e-05	1.477e-06	9.061e-05	1.053e-04
	20 minutes	-2.376e-05	4.083e-05	7.452e-06	1.346e-06	7.627e-05	4.537e-05
VG	Daily	-6.825e-05	2.391e-03	1.816e-04	6.041e-05	1.417e-03	2.716e-03
	4 hours	-1.688e-04	3.779e-04	8.545e-05	1.924e-05	7.883e-04	6.212e-04
	2 hours	2.249e-05	1.714e-04	4.404e-05	7.773e-06	2.945e-04	9.446e-05
	1 hour	3.160e-05	7.430e-05	2.187e-05	3.615e-06	2.187e-05	7.421e-05
	30 minutes	2.807e-05	7.604e-05	1.050e-05	1.623e-06	5.046e-05	6.242e-05
	20 minutes	4.403e-05	2.777e-05	7.310e-06	9.433e-07	8.561e-06	7.859e-06

Tab. 4.9: The mean and standard deviation of the parameters of the GHD obtained from rolling time windows.

Chapter 5

Lévy processes based on time subordinated Brownian motion

Prices changes of an asset are influenced by the arrival of information in the market. For identical time intervals, the price process evolves differently this is because information arrives to traders at different rates. Prices change faster if there are a lot of transactions, suggesting stock prices are not driven by time itself. The concept of a subordinated process was introduced by Bochner (1955). Clark (1973) considered a model with normally distributed asset returns run on an independent stochastic clock that measured the evolution of time and had log-normal increments.

Ané and Geman (2000) extended the work by Clark (1973) and obtained normality of stock returns by using the number of trades as measure of time. Ling (2016) uses the occurrence of a certain amount of trading volume and the occurrence of a certain number of transactions to measure time. To estimate parameters of the normal distribution, Ling (2016) uses the generalized method of moments and the kernel method. Ling (2016) shows that the returns are normally distributed when time is measured using event clocks especially transaction clock.

As mentioned in chapter 4, the GHD is a parametric family of infinitely divisible distributions. Proposition 3.6 informs us that for every generalized hyperbolic distribution there exists a homogeneous Lévy process X_t such that the distribution of X_t at a fixed point time is that particular generalized hyperbolic distribution.

For generalized hyperbolic Lévy processes the distribution of a value X_s at a time point s is usually different to the distribution of the value X_t that at

time t, hence the process X_t will not be generalized hyperbolic. However in the cases of the VG and the NIG, the paths are closed under convolution. Therefore for all time points the value of a Lévy process that is VG or NIG will be VG or NIG distributed respectively, this makes the simulation of the VG and NIG more natural than all the other GHD processes.

Lévy processes can be simulated based on a compound Poisson approximation. However the compound Poisson approach is involved, generalized hyperbolic Lévy processes can simulated by time changing Brownian motion. VG and NIG processes can be simulated by Brownian motion whose time is measured by a stochastic process. The VG processes can be constructed by setting the time of the Brownian motion to be a gamma process, likewise NIG process can be constructed using an inverse Gaussian process as a random clock.

This chapter is organized as follows, section 5.1 gives a brief discussion of generalized hyperbolic Lévy processes. Subordinators with special focus on the gamma process as well as the inverse Gaussian process are discussed in section 5.2. Sections 5.3 and 5.4 discuss the representation of the NIG and VG using subordinated Brownian motion respectively. The discussion in section 5.5 focuses on the possible simulation techniques for the VG and NIG processes. An in-depth discussion is given on three parametric estimation methods namely the Maximum Likelihood Estimation (MLE), Numerical Method of Moments (NMM) and Generalized Method of Moments (GMM). The results from fitting the VG and NIG to financial data from the JSE is presented in section 5.7. A comparison of the performance of the estimation methods from the subordinated Brownian motion representation and the representation by McNeil *et al.* (2005) is given in section 5.8.

5.1 Generalized Hyperbolic Lévy process

The representation of the GHD used by Barndorff-Nielsen (1977) to introduce the GHD has five parameters. The probability density function (PDF) of a random variable X that follows the GHD is given by

$$f_{\text{GHD}}(x;\lambda,\alpha,\beta,\delta,\mu) = \frac{(\gamma/\delta)^{\lambda}}{\sqrt{2\pi}K_{\lambda}(\delta\gamma)} \cdot \frac{K_{\lambda-1/2}(\alpha\sqrt{\delta^2+(x-\mu)^2})}{(\sqrt{\delta^2+(x-\mu)^2}/\alpha)^{1/2-\lambda}} \cdot e^{\beta(x-\mu)}, \quad x \in \mathbb{R}, \tag{5.1}$$

where K_λ is the modified Bessel equation of the third kind and $\gamma^2 = \alpha^2 - \beta^2$.The domain of variation of the parameters is

$$\begin{aligned}
&\text{if } \lambda < 0, \quad \delta > 0, \quad \alpha \geq 0, \quad \alpha^2 \geq \beta^2, \\
&\text{if } \lambda = 0, \quad \delta > 0, \quad \alpha > 0, \quad \alpha^2 > \beta^2, \\
&\text{if } \lambda > 0, \quad \delta \geq 0, \quad \alpha > 0, \quad \alpha^2 > \beta^2,
\end{aligned}$$

with $\mu \in \mathbb{R}$ for all three cases. The GHD is symmetric when $\beta = 0$. We will write $X \sim \text{GHD}(\lambda, \alpha, \beta, \delta, \mu)$ when the random variable X follows the GHD. The $(\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ was calibrated in chapter 4, note that only the univariate case calibrated in chapter 4 hence Σ will have dimension 1. To switch between the two parameterizations $(\lambda, \alpha, \beta, \delta, \mu) \rightarrow (\lambda, \chi, \psi, \mu, \Sigma, \gamma)$ one can use the following formulas

$$\delta = \Sigma, \; \chi = \delta^2, \; \psi = \alpha^2 - \delta\beta^2, \; \gamma = \delta\beta. \tag{5.3}$$

The Laplace transform of the GHD is given by

$$L(z) = e^{\mu z} \cdot \frac{\gamma^\lambda \cdot K_\lambda(\delta\gamma_z)}{\gamma_z^\lambda \cdot K_\lambda(\delta\gamma)}, \quad \alpha > |\beta + z|, \tag{5.4}$$

where $\gamma_z^2 = \alpha^2 - (\beta + z^2)$. When $\lambda = 1$, the hyperbolic distribution is obtained and when $\gamma = 0$ or equivalently when $\alpha = |\beta|$ one obtains the asymmetric scaled t distribution. The NIG is obtained when $\lambda = -\frac{1}{2}$ and VG obtained when $\delta = 0$ and $\mu = 0$.

When a random variable X follows the NIG we will write $X \sim \text{NIG}(\alpha, \beta, \delta, \mu)$ and similarly write $X \sim \text{VG}(\lambda, \alpha, \beta, \mu)$ when X follows the VG. The Laplace transform of the VG is given by

$$L(z) = e^{\mu z} \left(\frac{\gamma}{\gamma_z} \right)^{2\lambda}, \quad \alpha > |\beta + z|, \tag{5.5}$$

where $\gamma_z^2 = \alpha^2 - (\beta + z^2)$. The Laplace transform of the NIG is given by

$$L(z) = e^{\mu z + \delta(\gamma - \gamma_z)}, \quad \alpha > |\beta + z|. \tag{5.6}$$

It follows from Eqn. 5.5 that if X_1 and X_2 are independent such that $X_i \sim \text{VG}(\lambda_i, \alpha, \beta, \mu_i)$, $i = 1, 2$, then

$$X_1 + X_2 \sim \text{VG}(\lambda_1 + \lambda_2, \alpha, \beta, \mu_1 + \mu_2). \tag{5.7}$$

Therefore the VG is closed under convolutions. It follows from Eqn. 5.6 that if X_1 and X_2 are independent such that $X_i \sim \text{NIG}(\alpha, \beta, \delta_i, \mu_i)$, $i = 1, 2$, then

$$X_1 + X_2 \sim \text{NIG}(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2). \tag{5.8}$$

The VG and NIG are the only two subclasses of the GHD that are closed under convolution.

The expression of the Lévy measure of the GHD is complicated as it involves integrals of Bessel function. The Lévy measure of the GHD is given by

$$q(x) = \begin{cases} \frac{e^{\beta x}}{|x|}\left(\lambda e^{-\alpha|x|} + \int_0^\infty \frac{\exp(-|x|\sqrt{\alpha^2+2y})}{\pi^2 y(J_\lambda^2(\delta\sqrt{2y})+Y_\lambda^2(\delta\sqrt{2y}))}\,\mathrm{d}y\right) & \text{if } \lambda \geq 0, \\ \frac{e^{\beta x}}{|x|}\left(\int_0^\infty \frac{\exp(-|x|\sqrt{\alpha^2+2y})}{\pi^2 y(J_\lambda^2(\delta\sqrt{2y})+Y_\lambda^2(\delta\sqrt{2y}))}\,\mathrm{d}y\right) & \text{if } \lambda < 0. \end{cases}$$

where J_λ given by Eqn. B.13 is the Bessel function of the first kind and Y_λ is the Bessel function of the second and is given by Eqn. B.15.

The behavior of the Lévy measure in the neighborhood of zero is of particular importance especially when one wants to construct another Lévy process by changing the measure of another process. The behavior of the Lévy measure of the GHD around zero is given by

$$x^2 q(x) = \frac{\delta}{\pi} + \frac{\lambda + \frac{1}{2}}{2}|x| + \frac{\delta\beta}{\pi}x + \mathcal{O}(|x|), \quad \text{as } x \to 0. \tag{5.9}$$

From Eqn. 5.9 we can see that the Lévy measure of the GHD has infinite mass around zero. The Lévy measure of the NIG is given by

$$q(x) = \frac{1}{\pi}\delta\alpha\frac{1}{|x|}K_1(\alpha|x|)e^{\beta x}, \tag{5.10}$$

where $K_1(x)$ is the modified Bessel function of the third kind. A generalized hyperbolic Lévy process can be written in the form

$$X_t = Z_t + \lambda t, \tag{5.11}$$

where $q(x)$ is the Lévy measure of the GHD and Z_t is a pure jump martingale with infinitely many jumps in every finite interval. The process Z_t is given by

$$Z_t = \int_0^t \int_{\mathbb{R}\setminus\{0\}} x\left(\mu^X(du, dx) - q(x)dudx\right), \tag{5.12}$$

where μ^X is an integer valued random measure defined by

$$\mu^X(dt, dx) = \sum_{s>0} 1_{\{\Delta X_s \neq 0\}}\varepsilon_{(s,\Delta X_s)}(dt, dx). \tag{5.13}$$

where ε_a is the Dirac measure at a and $\Delta X_s = X_s - X_{s^-}$ is the jump process at time s. Note that μ^X is a Poisson random measure with intensity

$q(x)dxdt$.

The jump behavior of a discontinuous Lévy process is determined by its Lévy measure. It is possible to simulate a Lévy process by approximating it using compound Poisson process. The Lévy measure that is truncated in the neighborhood of zero is the jump distribution of the normalized process. Rydberg (1997) took this approach to simulate NIG Lévy motions.

It is difficult to analytically examine the density of the Lévy measure of the GHD, moreover it is cumbersome to numerically evaluate the expression of the Lévy measure. Fortunately the generalized hyperbolic Lévy processes can be represented as Brownian motion that is time subordinated by processes that follow the GIG.

Let $\tau(t)$ be a Lévy processes whose increments ($\tau(1)$) follow the GIG and W be a standard Brownian motion, then the process

$$X(t) = W(\tau(t)) + \beta\tau(t) + \mu t, \tag{5.14}$$

is a generalized hyperbolic Lévy process. Since the increments of any GIG distributed random variables are positive, the process $\tau(t)$ is increasing and thus $\tau(t)$ can be thought of as a stochastic clock.

Almost surely increasing Lévy process are known as subordinators and the construction in Eqn. 5.14 is known as time subordination. Stochastic time change of Lévy processes is described in section 3.5. The advantage of the time change representation is that it allows us to interpret Lévy processes in calendar time as Brownian motions in business time. One of the consequences of time subordination is the emergence of non-Gaussianity in calendar time.

5.2 Subordinators

The increments of the subordinators of generalized hyperbolic Lévy process follow the GIG. Table 4.2 shows the relationship between subclasses of the GHD with those of the GIG. The VG and NIG are the only subclasses of the GHD that are closed under convolution, hence the construction of VG and NIG processes are more natural than the construction of other generalized hyperbolic Lévy processes. Only the construction of the VG and the

NIG processes will be discussed in this chapter.

The VG process is constructed by setting the stochastic clock to be a gamma process whose increments follow the gamma distribution. Similarly the NIG is constructed by setting the stochastic clock to be an inverse Gaussian process whose increments follow the inverse Gaussian distribution. Note that both the gamma distribution and inverse Gaussian distribution are subclasses of the GIG.

5.2.1 The inverse Gamma process

Definition 4.2 Inverse Gaussian Process
An inverse Gaussian process is a Lévy process $(S_t)_{t\geq 0} = S_t(\mu, \lambda)$ with increments distributed according to inverse Gaussian distribution

$$f_{S_t}(x) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\frac{\lambda(x-\mu)^2}{2\mu^2}}, \quad x \geq 0,\ \mu, \lambda > 0. \tag{5.15}$$

The inverse Gaussian distribution describes the distribution of the time taken by the Brownian motion with positive drift to reach a fixed positive level. The inverse Gaussian distribution belongs to the special class of infinitely divisible distribution and has the following mean and central moments

$$\text{Mean}[S_t] = \mu, \qquad \text{Variance}[S_t] \quad = \frac{\mu^3}{\lambda}, \tag{5.16a}$$

$$\text{Skewness}[S_t] = 3\sqrt{\frac{\mu}{\lambda}}, \text{ Excess Kurtosis} = 15\frac{\mu}{\lambda}. \tag{5.16b}$$

From Eqn. 5.16 one can tell that as λ increases the variance, skewness and excess kurtosis of the inverse Gaussian random variable decreases. The skewness and kurtosis are always positive and the distribution is skewed to the right with a single peak. It is possible to observe extremely large values when λ is small. Inverse Gaussian random variables can be simulated using algorithm 2.

The left plot of figure 5.1 shows the inverse Gaussian increments obtained from a simulation of an inverse Gaussian process for $T = 1$ with ($\Delta t = 0.01$) using parameters $\mu = 1$ and $\lambda = 1/(0.0774\Delta t)$. The right plot is the inverse Gaussian process which is the cumulative sum of the inverse Gaussian increments. The bottom plot shows the frequency of the inverse Gaussian increments.

Algorithm 2 Algorithm to generate inverse Gaussian random variables, adopted from Cont and Tankov (2004)

1: Generate $N \sim normal(0,1)$
2: Set $Y = N^2$ and $X = \mu + \frac{\mu^2 Y}{2\lambda} - \frac{\mu}{2\lambda}\sqrt{\mu^2 Y^2 + 4\mu\lambda Y}$
3: Generate $U \sim uniform(0,1)$
4: **if** $U \leq \frac{\mu}{X+\mu}$ **then**
5: Return X.
6: **else**
7: Return $\frac{\mu^2}{X}$.
8: **end if**

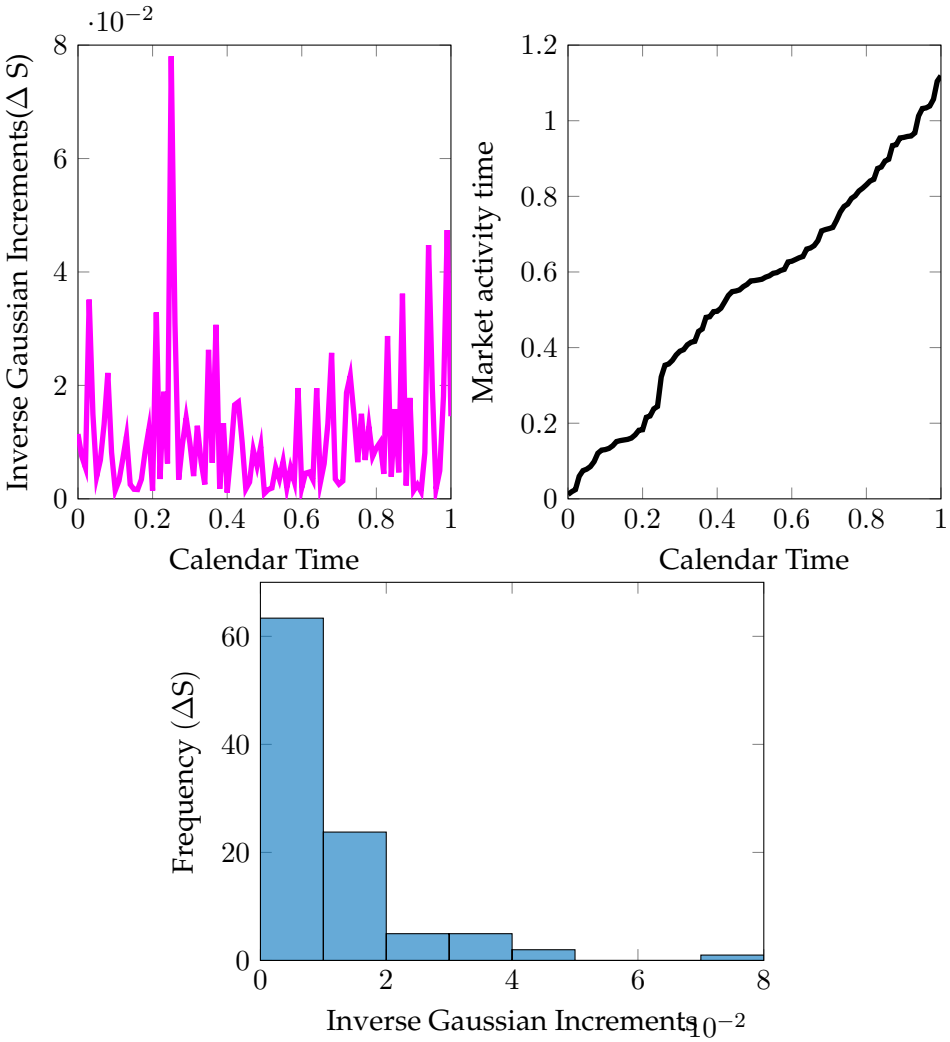


Fig. 5.1: Plots of simulated Inverse Gaussian process

5.2.2 The gamma process

Definition 4.3 Gamma Process
A gamma process is a Lévy process $(S_t)_{t\geq 0} = S_t(\alpha, \beta)$ with increments distributed according to gamma distribution

$$f_{S_t}(x) = \frac{\beta^{\alpha t} e^{-\beta x} x^{\alpha t - 1}}{\Gamma(\alpha t)},\ x \geq 0,\ \alpha, \beta > 0. \tag{5.17}$$

The gamma process will is used as a stochastic clock in the construction of the VG process. In-order to synchronize the gamma process with calendar time choose the parameters of a gamma process such that $\mathbb{E}[S_t] = t$. The gamma process has a scaling property, $\lambda S_t(\alpha, \beta) = S_t(\alpha, \frac{\beta}{\lambda})$ which makes it sufficient if one is able to simulate gamma random variable with the PDF given by

$$p(x) = \frac{x^{a-1}}{\Gamma(a)} e^{-x}. \tag{5.18}$$

The mean and variance of the marginal distribution:

$$\mathbb{E}[S_t] = \frac{\alpha t}{\beta} \qquad , Var[S_t] = \frac{\alpha t}{\beta^2}\ . \tag{5.19}$$

To obtain the desired expectation choose $\alpha = \beta$ and define $\kappa = \alpha^{-1}$, the variance of the VG processes becomes

$$Var[S_t] = \kappa t, \tag{5.20}$$

under the new parametrization. Since the gamma process is a positive process one can derive its Lévy measure using its moment generating function. Using the moment generating function of the gamma process$\left(\left(\frac{\beta}{\beta - u}\right)^{\alpha} \right)$ and integration techniques, Sato (1999) derived the Lévy measure of the gamma process to be

$$\rho(x) = \frac{\alpha e^{-\beta x}}{x} = \frac{e^{-x/\kappa}}{\kappa x}, \quad \forall x > 0. \tag{5.21}$$

Although $\int_0^\infty \rho(dx) = \infty$ implies that jumps arrive infinitely, often for a gamma process, $\rho(x)$ satisfies the conditions of Proposition 3.1. The gamma random variable with the PDF given by Eqn. 5.18 can be simulated by one of the following two algorithms, depending on the value of α.

Algorithm 3 Johnk's generator of gamma variables, $\alpha \leq 1$

1: Set $X, Y = 1$
2: Generate two $i.i.d.$ random variables U and V that are $uniform(0, 1)$
3: **while** $X + Y > 1$ **do**
4: $\quad X = U^{\frac{1}{\alpha}}$, $Y = V^{\frac{1}{1-\alpha}}$
5: **end while**
6: Generate an exponential random variable E.
7: Return $\frac{XE}{X+Y}$

Algorithm 4 Best's generator of gamma variables, $\alpha \geq 1$

1: Set $b = \alpha - 1$, $c = 3\alpha - \frac{3}{4}$
2: Generate *i.i.d.* random variables U and V that are $uniform(0, 1)$.
3: Set $W = U(1 - U)$, $Y = \sqrt{\frac{c}{W}}(U - \frac{1}{2})$ and $X = b + Y$.
4: **if** $X < 0$ **then**
5: $\quad$ Return to line 2
6: **else**
7: $\quad$ Set $Z = 64W^3V^3$.
8: $\quad$ **if** $\log(Z) \leq 2(b\log(X/b) - Y)$ **then**
9: $\quad\quad$ Go to line 14
10: $\quad$ **else**
11: $\quad\quad$ Return to line 2
12: $\quad$ **end if**
13: **end if**
14: Return X

The left plot of figure 5.2 shows the gamma increments obtained from a simulation of a gamma process for $T = 1$ with ($\Delta t = 0.01$) using shape parameter $\beta = 0.0774$ and scale parameter $\alpha = \Delta t/\beta$. The right plot is the gamma process which is the cumulative sum of the gamma increments. The bottom plot shows the frequency of the gamma increments.

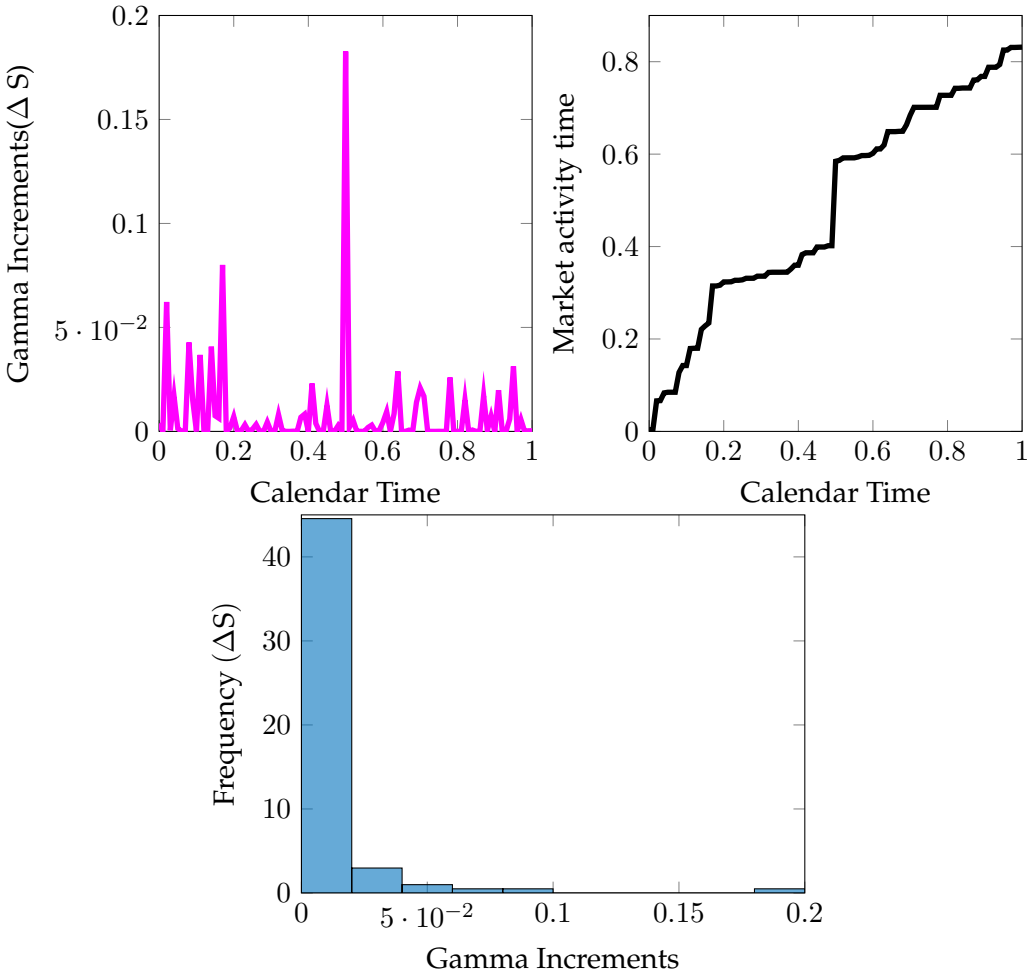


Fig. 5.2: Plots of simulated Gamma process.

5.3 The Normal Inverse Gaussian model

For convenience the notation used in Figueroa-López *et al.* (2011) was adopted. Let $\tau(t)$ be an inverse Gaussian random variable with shape parameter $\lambda = 1/\kappa t$ and mean $\mu = 1$. The NIG process $X(t)$ can be represented as follows

$$X(t) = \sigma W(\tau(t)) + \theta\tau(t) + bt, \quad \sigma > 0,\ b, \theta \in \mathbb{R}, \tag{5.22}$$

where W is a standard Brownian and t is the time interval between consecutive observations.

The parameter θ is the drift component in business time and b is the drift component in calendar time. The PDF of the NIG is symmetric when $\theta = 0$,

positively skewed when $\theta > 0$ and negatively skewed when $\theta < 0$. The parameter κ is the volatility of time subordinator ($\tau(t)$) and κ describes the tail heaviness. The volatility of log returns are described by σ. The probability density function of the NIG is given by

$$p_t(x) = \frac{te^{\frac{t}{\kappa}+\frac{\theta(x-bt)}{\sigma^2}}}{\pi}\left(\frac{(x-bt)^2+\frac{t^2\sigma^2}{\kappa}}{\frac{\theta^2}{\kappa\sigma^2}+\frac{1}{\kappa^2}}\right)^{-1/2} K_1\left(\frac{\sqrt{\theta^2+\frac{\sigma^2}{\kappa}}\sqrt{(x-bt)^2+\frac{t^2\sigma^2}{\kappa}}}{\sigma^2}\right), \tag{5.23}$$

where K_1 is the modified Bessel equation of the third kind with index parameter 1.

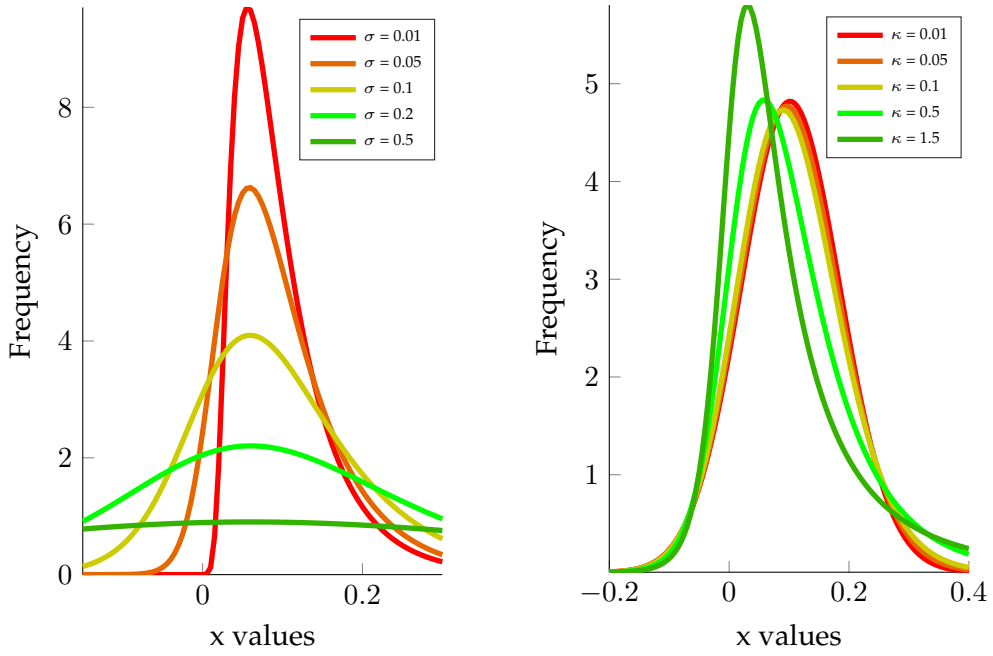


Fig. 5.3: Graphs showing how the PDF of the NIG changes as σ and κ change.

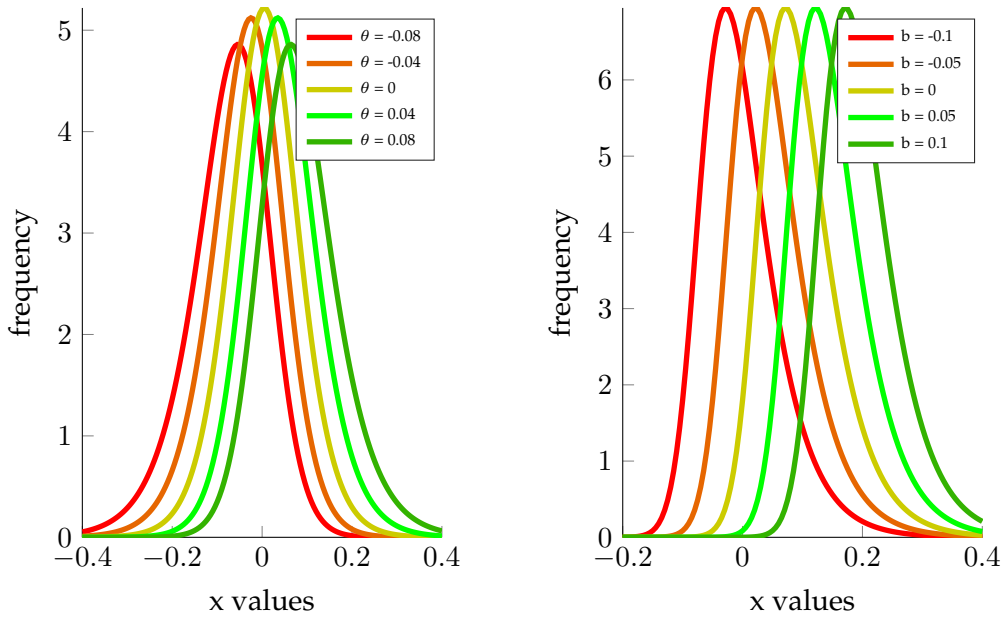


Fig. 5.4: Graphs showing how the PDF of the NIG changes as θ and b change.

The mean and the first three moments of the NIG are

$$M_1(X_t) := \mathbb{E}(X_t) = (\theta + b)t, \tag{5.24a}$$

$$M_2(X_t) := \mathbb{E}(X_t - \mathbb{E}X_t)^2 = (\sigma^2 + \theta^2\kappa)t, \tag{5.24b}$$

$$M_3(X_t) := \mathbb{E}(X_t - \mathbb{E}X_t)^3 = (3\sigma^2\theta\kappa + 3\theta^3\kappa^2)t, \tag{5.24c}$$

$$M_4(X_t) := \mathbb{E}(X_t - \mathbb{E}X_t)^4 = (3\sigma^4\kappa + 18\sigma^2\theta^2\kappa^2 + 15\theta^4\kappa^3)t + 3M_2(X_t)^2. \tag{5.24d}$$

Algorithm 5 Algorithm to simulate a NIG process adapted from Cont and Tankov (2004)

1: Simulate n independent inverse Gaussian random variables $\Delta S_1, \ldots \Delta S_n$ with mean $\mu = 1$ and shape parameter $\lambda = \frac{1}{\kappa t}$ using algorithm 2.
2: Simulate n standard normal random variables $N_1, \ldots N_n$.
3: Set $\Delta X_i = \theta \Delta S_i + \sigma\sqrt{\Delta S_i} N_i + bt, \quad \forall i = 1, \ldots, n.$
4: The discretized trajectory of the NIG processes is $X(t_k) = \sum_{i=1}^{k} \Delta X_i$.

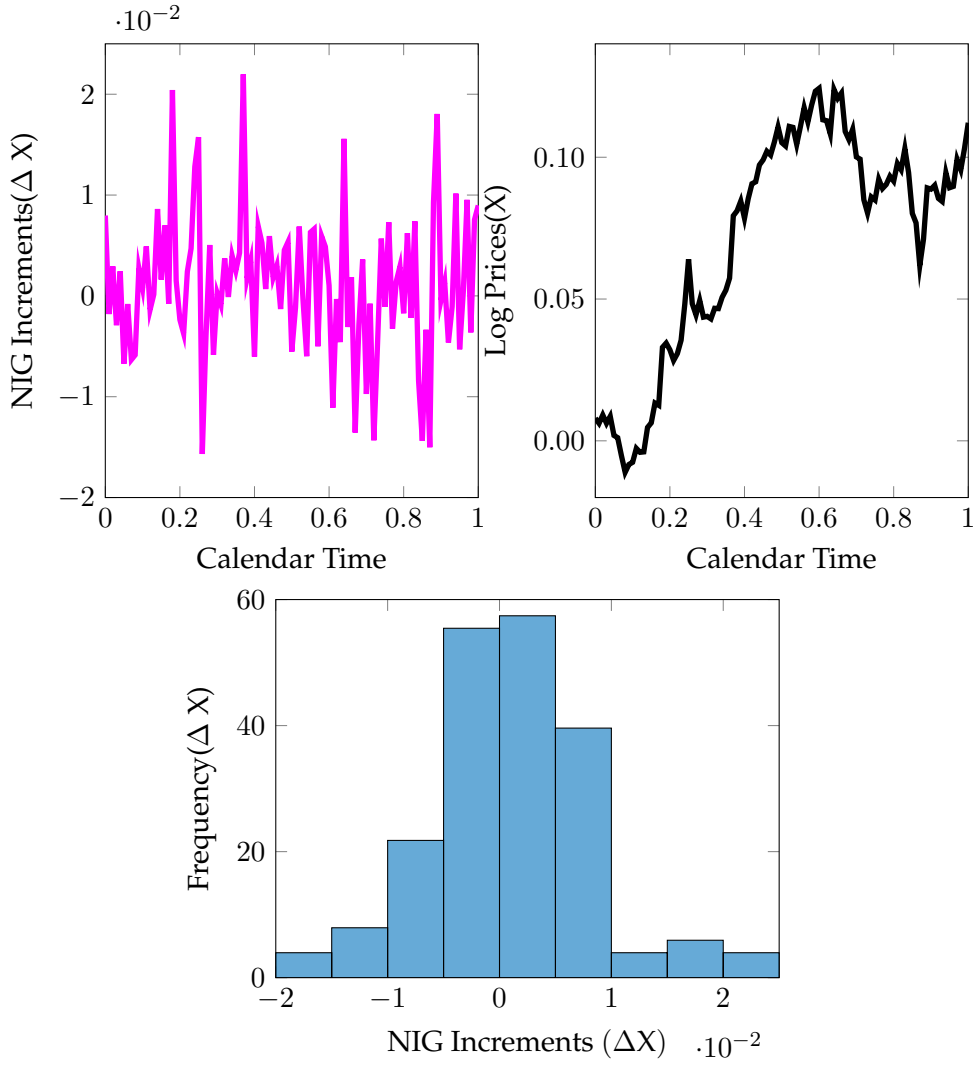


Fig. 5.5: Plots of simulated Normal inverse Gaussian process

The left plot of figure 5.5 shows the NIG increments obtained from a simulation of a NIG process for $T = 1$ with $\Delta t = 0.01$ using the parameters $b = -0.0015$, $\sigma = 0.0133$, $\theta = 0.0043$ and $\kappa = 0.0774$. The right plot is NIG process which is the cumulative sum of the NIG increments. The bottom plot shows the frequency of the NIG increments.

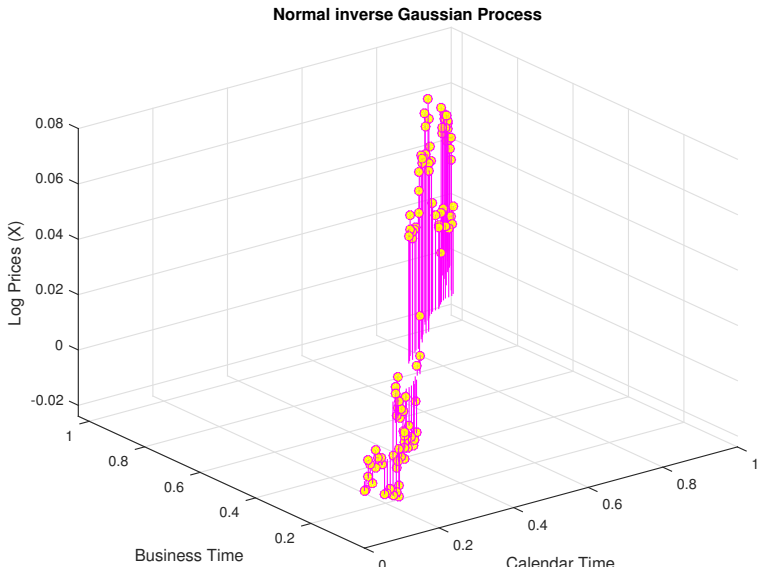


Fig. 5.6: 3D plot of simulated NIG process

Figure 5.6 shows a simulation of a NIG process (Z axis) using parameters $b = -0.0015$, $\sigma = 0.0133$, $\theta = 0.0043$ and $\kappa = 0.0774$, driven by an inverse Gaussian process (business time) as a function of real time (calendar time).

5.4 The Variance Gamma model

The VG can be represented as a difference of two independent gamma processes or as Brownian motion with drift that is time subordinated by a gamma process. The time subordinated Brownian motion representation will be discussed in this section. Let $\tau(t)$ be a gamma process with shape parameter $1/\kappa$ and scale parameter κ, the VG process can be represented as follows

$$X(t) = \sigma W(\tau(t)) + \theta\tau(t) + bt, \quad \sigma > 0, \ b, \theta \in \mathbb{R}. \tag{5.25}$$

where W is a standard Brownian and t is the time interval between consecutive observations.

The parameters of the VG have the same interpretation as the parameters of the NIG, see the interpretation in section 5.3. In the VG model a large κ is will cause fatter tails and result in frequent jumps. The probability density

function of the VG is

$$p_t(x) = \frac{\sqrt{2}e^{\frac{\theta(x-bt)}{\sigma^2}}}{\sigma\kappa^{\frac{t}{\kappa}}\sqrt{\pi}\Gamma(\frac{t}{\kappa})}\left(\frac{|x-bt|}{\sqrt{\frac{2\sigma^2}{\kappa}+\theta^2}}\right)^{\frac{t}{\kappa}-\frac{1}{2}} K_{\frac{t}{\kappa}-\frac{1}{2}}\left(\frac{|x-bt|\sqrt{\frac{2\sigma^2}{\kappa}+\theta^2}}{\sigma^2}\right), \tag{5.26}$$

where K_λ is the modified Bessel equation of the third kind.

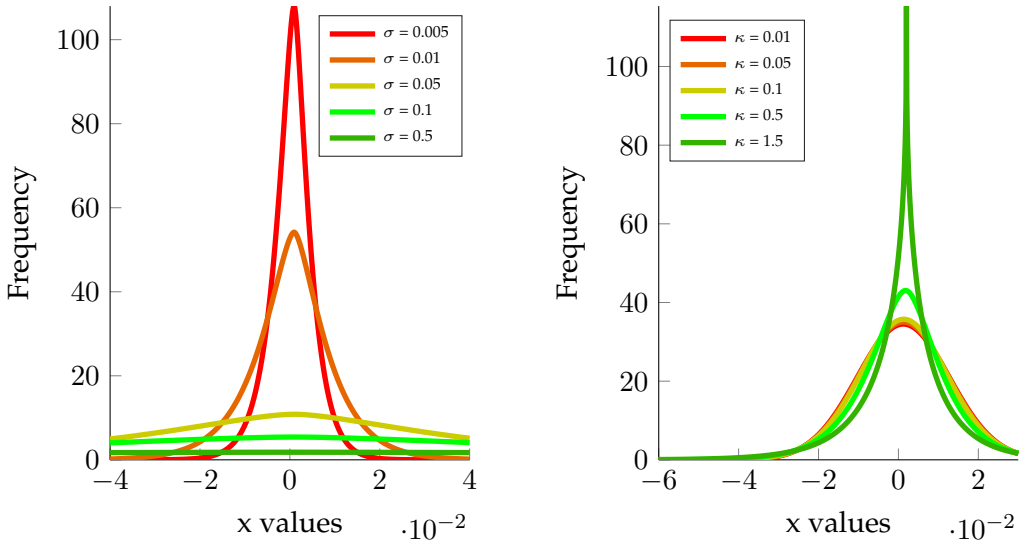


Fig. 5.7: Graphs showing how the PDF of the VG changes as σ and κ change.

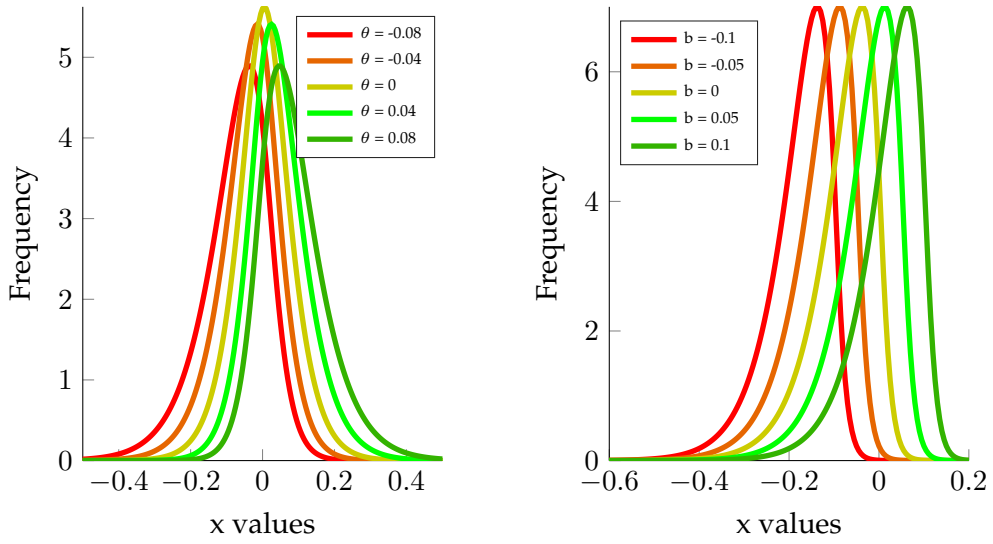


Fig. 5.8: Graphs showing how the PDF of the VG changes as θ and b change.

The mean and the first three moments of the VG are

$$M_1(X_t) := \mathbb{E}(X_t) = (\theta + b)t, \tag{5.27a}$$
$$M_2(X_t) := \mathbb{E}(X_t - \mathbb{E}X_t)^2 = (\sigma^2 + \theta^2\kappa)t, \tag{5.27b}$$
$$M_3(X_t) := \mathbb{E}(X_t - \mathbb{E}X_t)^3 = (3\sigma^2\theta\kappa + 2\theta^3\kappa^2)t, \tag{5.27c}$$
$$M_4(X_t) := \mathbb{E}(X_t - \mathbb{E}X_t)^4 = (3\sigma^4\kappa + 12\sigma^2\theta^2\kappa^2 + 6\theta^4\kappa^3)t + 3M_2(X_t)^2. \tag{5.27d}$$

The physical interpretation of the parameters of the VG and NIG was presented in section 5.3. From figures 5.3 and 5.7, one can interpret that σ is the variability parameter whilst κ is a shape parameter. From figures 5.4 and 5.8, the theoretical interpretation is that b is a location parameter and θ an asymmetry parameter.

Algorithm 6 Algorithm to simulate a VG process adapted from Cont and Tankov (2004)

1: Simulate n independent gamma variables $\Delta S_1, \dots \Delta S_n$ with shape parameter $\alpha = \frac{t}{\kappa}$ and scale parameter $\beta = \kappa$, using either algorithm 3 or 4 depending on the value of α.
2: Simulate n standard normal random variables $N_1, \dots N_n$.
3: Set $\Delta X_i = \theta \Delta S_i + \sigma\sqrt{\Delta S_i} N_i + bt, \quad \forall i$.
4: The discretized trajectory of the VG processes is $X(t_k) = \sum_{i=1}^{k} \Delta X_i$.

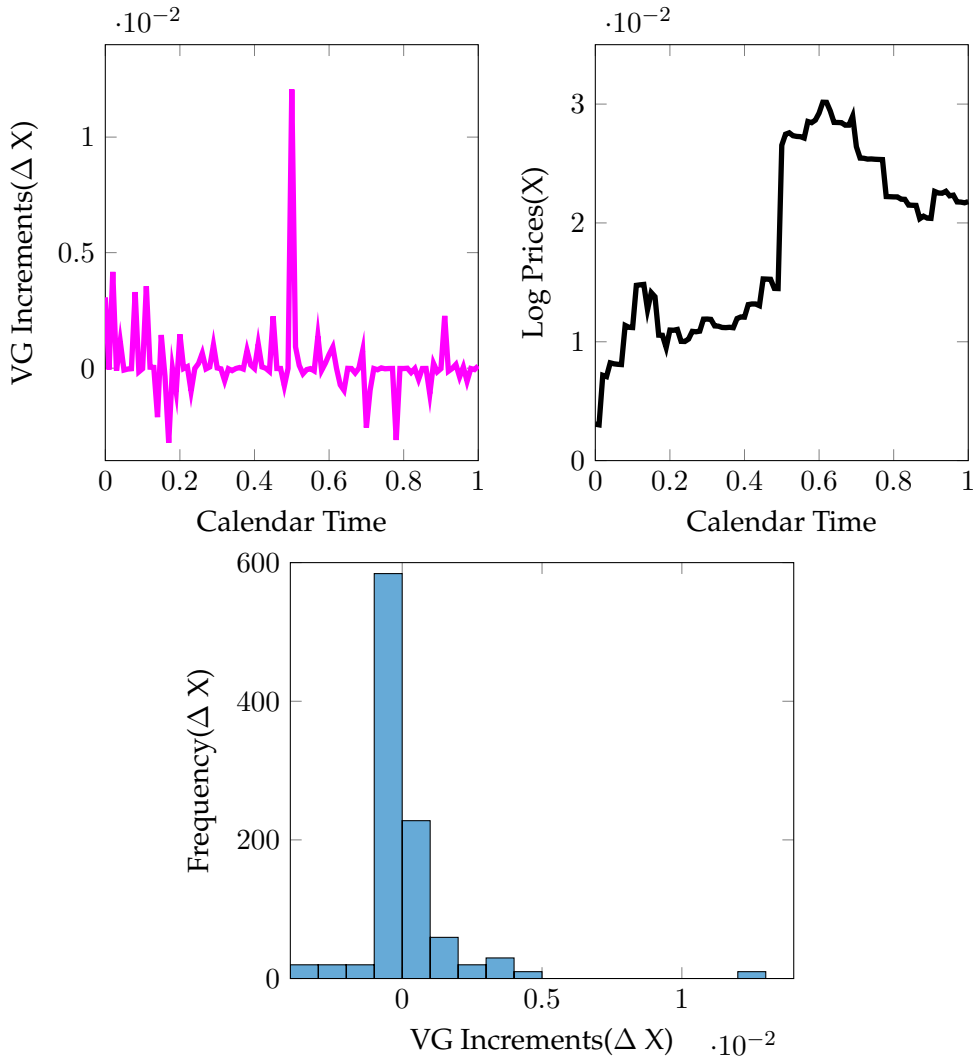


Fig. 5.9: Plot of simulated Variance Gamma process

The left plot of figure 5.9 shows the VG increments obtained from a simulation of a VG process for $T = 1$ with $(\Delta t = 0.01)$ using parameters $b = -0.0015$, $\sigma = 0.0133$, $\theta = 0.0043$ and $\kappa = 0.0774$. The right plot is the VG process which is the cumulative sum of the VG increments. The bottom plot shows the frequency of the VG increments.

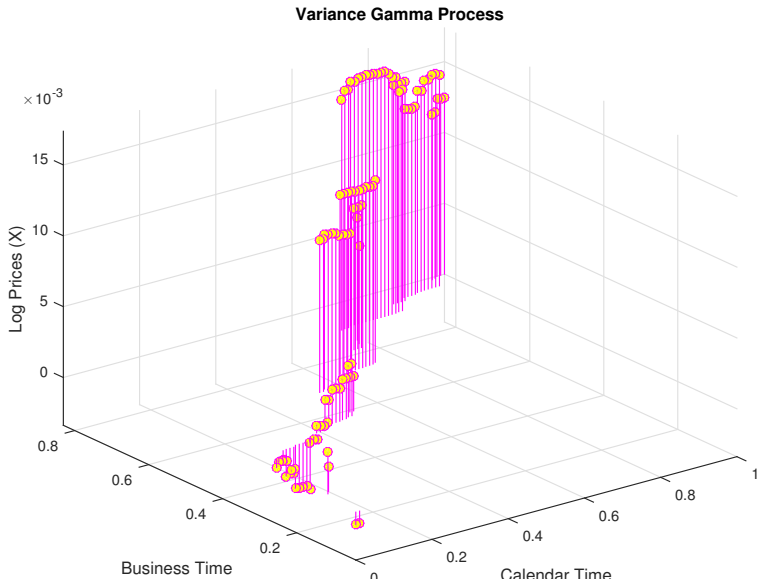


Fig. 5.10: 3D simulation of the VG process

Figure 5.10 shows a simulation of a VG process (Z axis) using parameters $b = -0.0015$, $\sigma = 0.0133$, $\theta = 0.0043$ and $\kappa = 0.0774$, driven by gamma process as a function of real time

5.5 Methods of estimation

5.5.1 The generalized method of moments

The generalized method of moments (GMM) is a generalization of the method of moments (MME) that was developed by Hansen (1982). The GMM is based only on the condition of sample moments and can therefore be constructed without fully specifying the data generation process. Unlike the MME estimators which are only consistent under weak assumptions, GMM estimators are consistent, asymptotically normal and efficient. Lars Peter Hansen shared the Nobel Prize of Economics in 2013 for developing the GMM in part of his work for empirical analysis of stock prices.

In-order to obtain the GMM estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \ldots, \hat{\beta}_l)$ of the model in question, minimize the quadratic form,

$$\hat{\boldsymbol{\beta}} = \text{argmin}_{\boldsymbol{\beta}} \mathbf{g}_n(X, \beta)' \mathbf{W}_n \mathbf{g}_n(X, \boldsymbol{\beta}), \tag{5.28}$$

where $X = (X_1, \ldots X_n)$ is a given time series with $i.i.d.$ observations, W is a positive definite symmetric (PSD) matrix and $\mathbf{g}$ are the average mo-

ment conditions derived from the given time series defined by the following equation

$$\mathbf{g}_n = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(X_i, \boldsymbol{\beta}). \tag{5.29}$$

The PSD matrix will ensure that the estimator converges to a unique point. In order to obtain efficient estimators, the heteroskedastic and autocorrelation (HAC) matrix W proposed by Newey and West (1986) was used in the implementation. The iterated GMM is implemented instead of the two step GMM developed by Hansen (1982). The estimators obtained after computation of the two step GMM are used as starting parameters in the next step of the iterated GMM. The iterations are computed until a certain level of tolerance is reached.

Given a time series $X_1, \dots X_n$ with $i.i.d.$ observations, where the time interval between consecutive observations is t. The procedure for obtaining GMM estimates for VG and NIG is as follows:

Step 1: Assume that $|\theta| \ll 0$ and θ^2, θ^3 and θ^4 are negligible. Compute the starting values $\boldsymbol{\beta}_0$ using the following equations

$$\sigma = \sqrt{V/t}, \quad \kappa = \frac{Kt}{3} - t, \tag{5.30a}$$

$$\theta = \frac{S\sigma\sqrt{t}}{3\kappa}, \quad b = X/t - \theta. \tag{5.30b}$$

where X is the mean, V is the variance, S is the skewness and K is the kurtosis.

Step 2: Let $W = 4 \times 4$ identity matrix, minimize the quadratic from given by Eqn. 5.28 to get β_1.

Step 3: Compute the HAC matrix W, using the method that was introduced by Newey and West (1986). The procedure is as follows:

i Compute the optimum bandwidth using the current moment conditions (in the implementation the Bartlett Kernel was used).

ii Estimate the kernel D of the data, using the optimum bandwidth.

iii Compute the long run auto-covariance matrix S, using the estimated kernel D.

iv Compute V the inverse of the covariance matrix (in the implementation Gaussian elimination was used to improve efficiency.)

v Compute the Jacobian G, using the average current moment conditions.

vi The PSD HAC matrix W is given by

$$W = (GVG)^{-1}G^{'}WSW(GWG)^{-1}. \tag{5.31}$$

Step 4: Compute $\boldsymbol{\beta}_2$ by minimizing Eqn. 5.28.
Step 5: If $\|\boldsymbol{\beta}_2 - \boldsymbol{\beta}_1\|<$ tolerance then stop, else go to step 2.
The GMM estimator is defined to be β_2.

5.5.2 The numerical method of moments

It is usually difficult to solve a system of non linear equations for the MME. Figueroa-López *et al.* (2011) developed a method of estimate parameters of the NIG and VG using sample moments that does not require solving a system of equations and we will refer to this method as the numerical method of moments (NMM). The NMM involves rewriting the moments of the NIG or VG in terms of $\varepsilon = \frac{\sigma\theta^2}{\kappa}$ and solving numerically for an ε that satisfies a specific condition. Note that the NMM is specific to the representation of the VG and NIG given in Figueroa-López *et al.* (2011). Given a time series $X_0, X_1, \ldots X_n$ whose observations are $i.i.d.$ where the time interval between consecutive observations is t, the NMM estimates can be obtained by computing the following steps
Step 1: Compute

$$\Delta_i^n X = \Delta_i^n = X_i - X_{i-1}, \quad i = 1, \ldots, n \text{ and } \bar{\Delta}^n = \frac{1}{n}\sum_{i=1}^{n}\Delta_i^n \tag{5.32}$$

Step 2: Estimate the central moments using

$$\hat{M}_{k,n} = \frac{1}{n}\sum_{i=1}^{n}(\Delta^n - \overline{\Delta}^n), \quad k = 2, 3, 4. \tag{5.33}$$

Step 3: Using the estimated moments in step 2, compute the sample variance, skewness and kurtosis

$$\widehat{\text{Var}_n} = \frac{1}{n-1}\sum_{i=1}^{n}(\Delta_i^n - \bar{\Delta}^n), \quad \widehat{Skw_n} = \frac{\hat{M}_{3.n}}{\hat{M}_{2,n}^{3/2}}, \quad \widehat{Krt_n} = \frac{\hat{M}_{4,n}}{\hat{M}_{2,n}^2}. \tag{5.34}$$

Step 4: For the VG, solve numerically for ε^* that satisfies the following equation

$$\frac{\varepsilon(3+\varepsilon)^2}{(1+\varepsilon)(1+4\varepsilon+2\varepsilon^2)} = \frac{3\widehat{Skw}_n^2}{\widehat{Krt}_n}, \tag{5.35}$$

For the NIG, solve numerically for ε^* that satisfies the following equation

$$\frac{9\varepsilon}{5\varepsilon+1} = \frac{3\widehat{Skw}_n^2}{\widehat{Krt}_n}. \tag{5.36}$$

Step 5: Obtain the numerical MME estimators for the VG using the following equation

$$\hat{\kappa} = \frac{t}{3}\widehat{Krt}_n\left(\frac{(1+\varepsilon^*)^2}{(1+4\varepsilon^*+2\varepsilon^{*2})}\right), \quad \hat{\sigma}^2 = \frac{\widehat{Var}_n}{t}\left(\frac{1}{(1+\varepsilon^*)}\right), \tag{5.37a}$$

$$\hat{\theta} = \frac{M_{3,n}}{t\hat{\sigma^2}\hat{\kappa}}\left(\frac{1}{3+2\varepsilon^*}\right), \qquad \hat{b} = \frac{1}{t}\bar{\Delta}^n - \hat{\theta}. \tag{5.37b}$$

The numerical estimator for the NIG is given by the following formulas

$$\hat{\kappa} = \frac{t}{3}\widehat{Krt}_n\left(\frac{(1+\varepsilon^*)}{(1+5\varepsilon^*)}\right), \quad \hat{\sigma}^2 = \frac{\widehat{Var}_n}{t}\left(\frac{1}{(1+\varepsilon^*)}\right), \tag{5.38a}$$

$$\hat{\theta} = \frac{M_{3,n}}{t\hat{\sigma^2}\hat{\kappa}}\left(\frac{1}{3+3\varepsilon^*}\right), \qquad \hat{b} = \frac{1}{t}\bar{\Delta}^n - \hat{\theta}. \tag{5.38b}$$

5.5.3 Maximum likelihood estimation

When the full shape of the data is unknown the MLE may be inapplicable and the GMM is considered better estimator. The closed form of the PDF of the VG and NIG models makes the MLE applicable. The MLE is often a better estimator since it uses more information in comparison to the GMM which only considers moment conditions. The parameters obtained from the GMM will be used as starting values for the MLE.

Since Lévy processes assumes independent increments we assume that the time series $X = (X_1, \dots X_n)$ has $i.i.d.$ observations. The joint PDF is the product of the marginal PDF under the assumption of independent observations. In-order to obtain the MLE estimate $\hat{\boldsymbol{\beta}}$ for the VG or NIG model, maximize the natural logarithm of the joint PDF

$$\hat{\boldsymbol{\beta}} = \text{argmax}_{\boldsymbol{\beta}} \sum_{i=1}^{n} \log\left(p_t(\boldsymbol{\beta}; X_i)\right), \tag{5.39}$$

where p_t is the PDF of the VG model given by Eqn. 5.26 or the PDF of the NIG model given by Eqn. 5.23 and t is the time interval between consecutive observations.

5.6 Estimated results for the simulated data

Denote the time span between consecutive observations by Δt. The values of Δt used to simulate paths of the VG and NIG are $\frac{1}{32}$, $\frac{1}{24}$, $\frac{1}{16}$, $\frac{1}{8}$, $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$ and 1. For each Δt, 200 paths of the NIG process were simulated using algorithm 5 and 200 paths of the VG process were simulated using algorithm 6. The period used for the simulations is $T = 252$, which is usually used as the number of business days in a year.

The GMM, NMM and MLE were used to estimate parameters of either the VG or NIG model for each path that was simulated. The sample mean and sample standard deviation of the estimated parameters were evaluated. Since the estimates obtained from the GMM were used as starting values for the MLE, figures 5.11 - 5.14 only shows the mean and deviation of the parameters from the MLE and NMM.

The Mean Absolute Scaled Error (MASE) and the Average Relative Mean Absolute Deviation (ARMAD) were evaluated for all sampling frequencies for both the VG and NIG. MATLAB's sequential quadratic programming (sqp) algorithm was used to obtain the MLE estimators for the VG and NIG.

The deviation of b and σ is small when $\Delta t \leq 0.5$ for both the NMM and MLE of the NIG, however the deviation significantly increases when the sampling time span is 1. MLE estimators for θ of the VG have high deviation when $\Delta t \leq \frac{1}{24}$, suggesting that the MLE of the VG is not stable for very small Δt. The deviation of b from the VG increase as the time span increases.

The mean of κ estimators for the MLE are closer to the true value than those given by the NMM for the NIG. The deviation of κ and that of σ increase as Δt increases NIG for the NIG. When $\Delta t \leq \frac{1}{24}$ the mean of σ from the VG varies significantly from the true value.

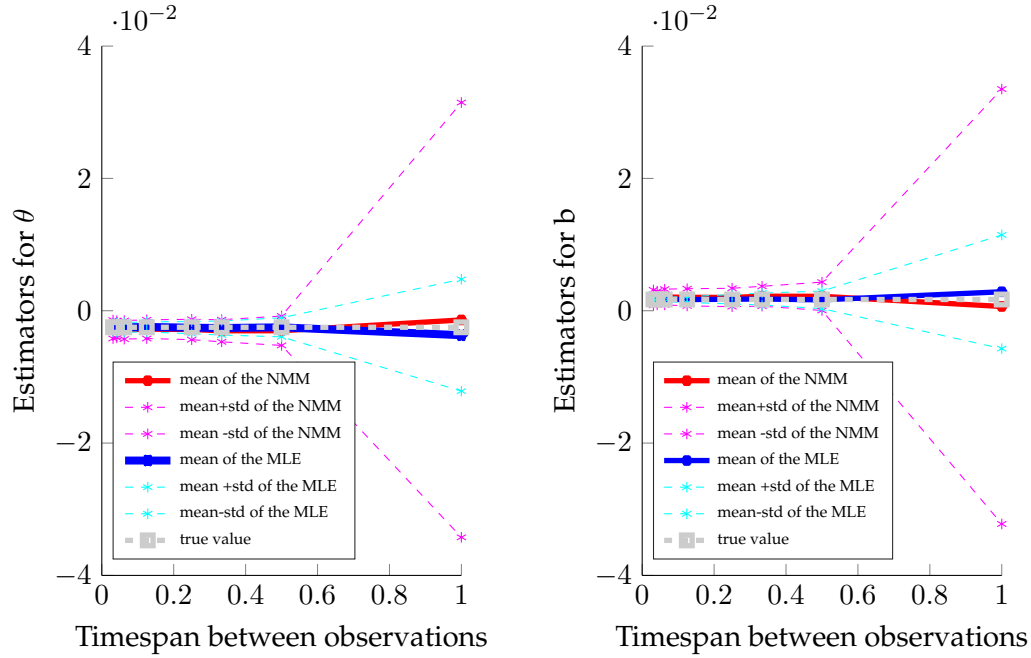


Fig. 5.11: Sampling mean and sampling deviation of the NMM and MLE based on 200 simulations of the NIG using parameters

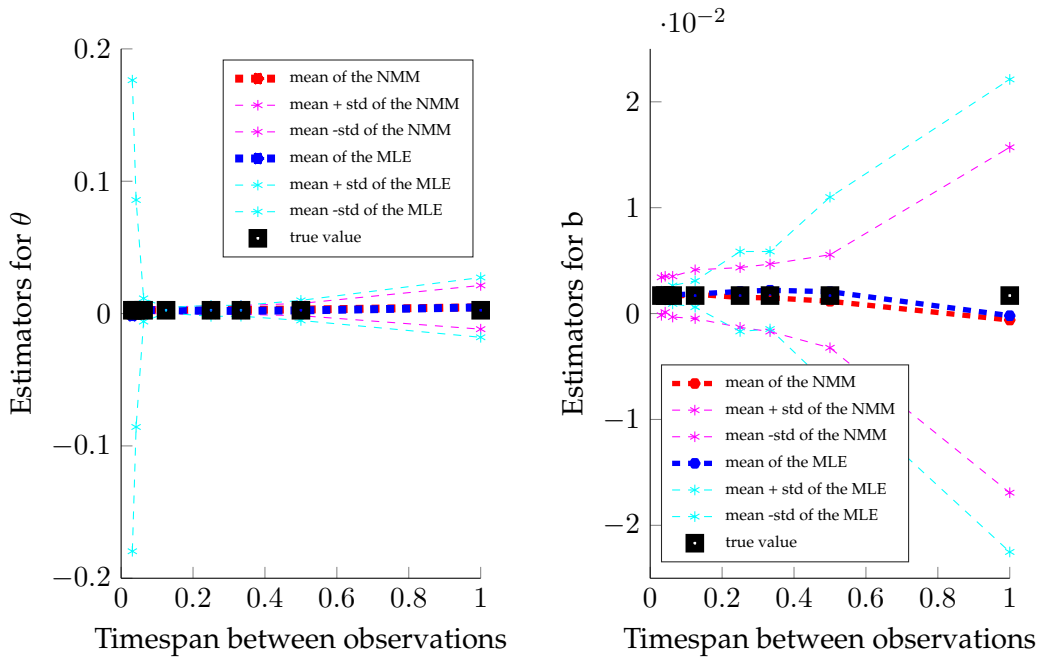


Fig. 5.12: Sampling mean and sampling deviation of the NMM and MLE based on 200 simulations of the VG

Figures 5.11 and 5.13 show the sampling mean and sampling deviation of the NMM and MLE based on 200 simulations of the NIG using parameters $b = 0.0017$, $\sigma = 0.0093$, $\theta = -0.0025$ and $\kappa = 0.4774$ with $T = 252$.

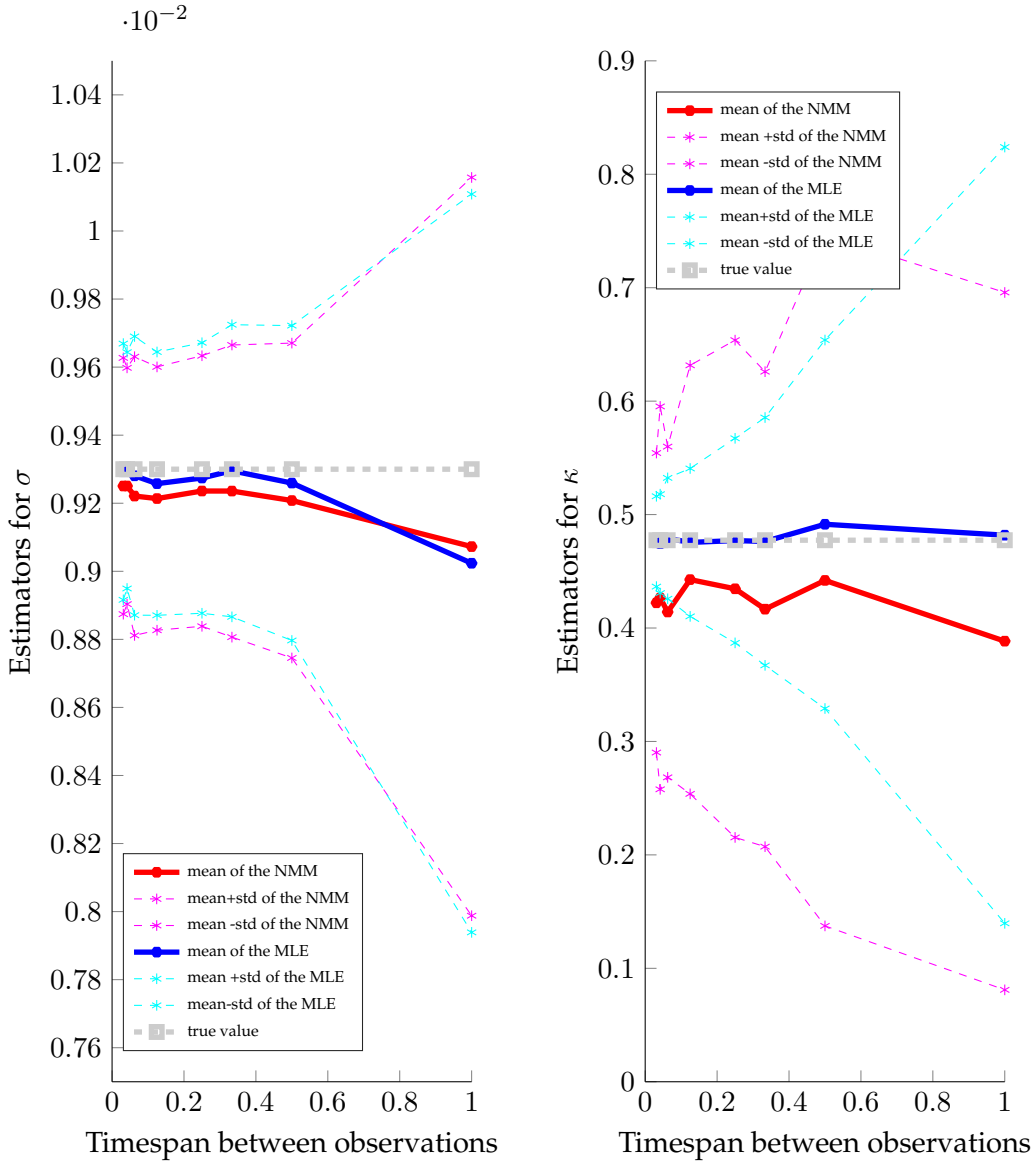


Fig. 5.13: Sampling mean and sampling deviation of the NMM and MLE based on 200 simulations of the NIG

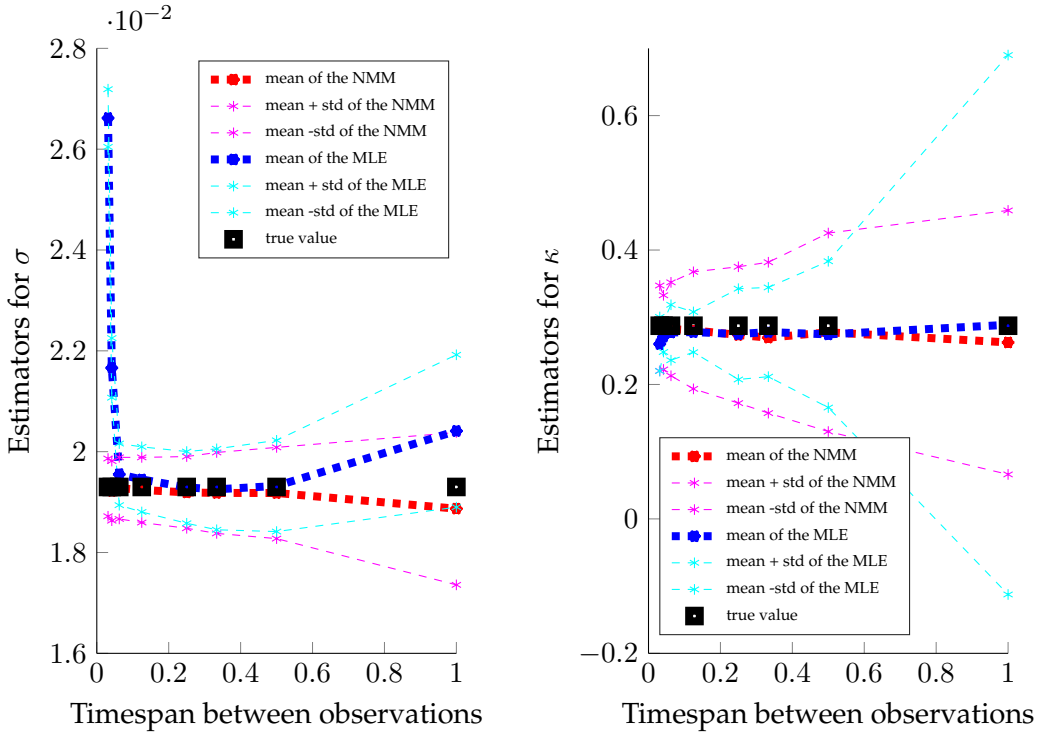


Fig. 5.14: Sampling mean and sampling deviation of the NMM and MLE based on 200 simulations of the VG

Figures 5.12 and 5.14 show the sampling mean and sampling deviation of the NMM and MLE based on 200 simulations of the VG using parameters $b = 0.0017$, $\sigma =$, $\theta = 0.0025$ and $\kappa = 0.28740.0193$ with $T = 252$.

The MASE was introduced in Hyndman and Koehler (2006) to measure forecast accuracy. It is advisable to use the MASE because it is scale invariant, easy to interpret, penalizes negative and positive errors equally and it does not rely on division by the parameter which can be a problem if the value of the parameter is almost or equal to zero.

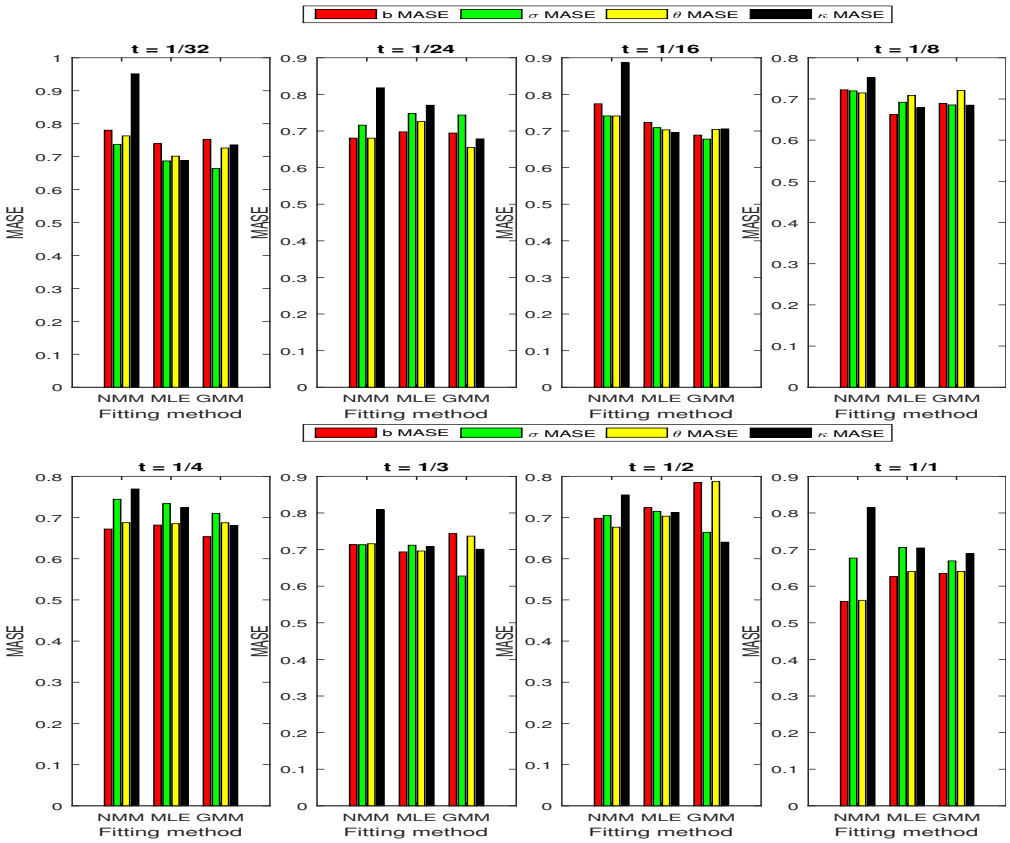


Fig. 5.15: A comparison of the MASE betweeen fitting method for different sampling frequencies of 200 simulated NIG processes.

For N simulations of process of either the VG or NIG models, estimate the parameters of each of the process using either NMM, GMM or MLE. Denote the parameters of the kth path by $\boldsymbol{\beta}_k = (b_k, \sigma_k, \theta_k, \kappa_k)$. The MASE is given by

$$\text{MASE} = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{|e_n|}{\frac{1}{N-1}\sum_{n=2}^{N}|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n-1}|}\right), \tag{5.40}$$

where e_n is the difference between the nth estimate of the parameter and the true value. If the MASE gives a value less than one then the in sample estimates are good.

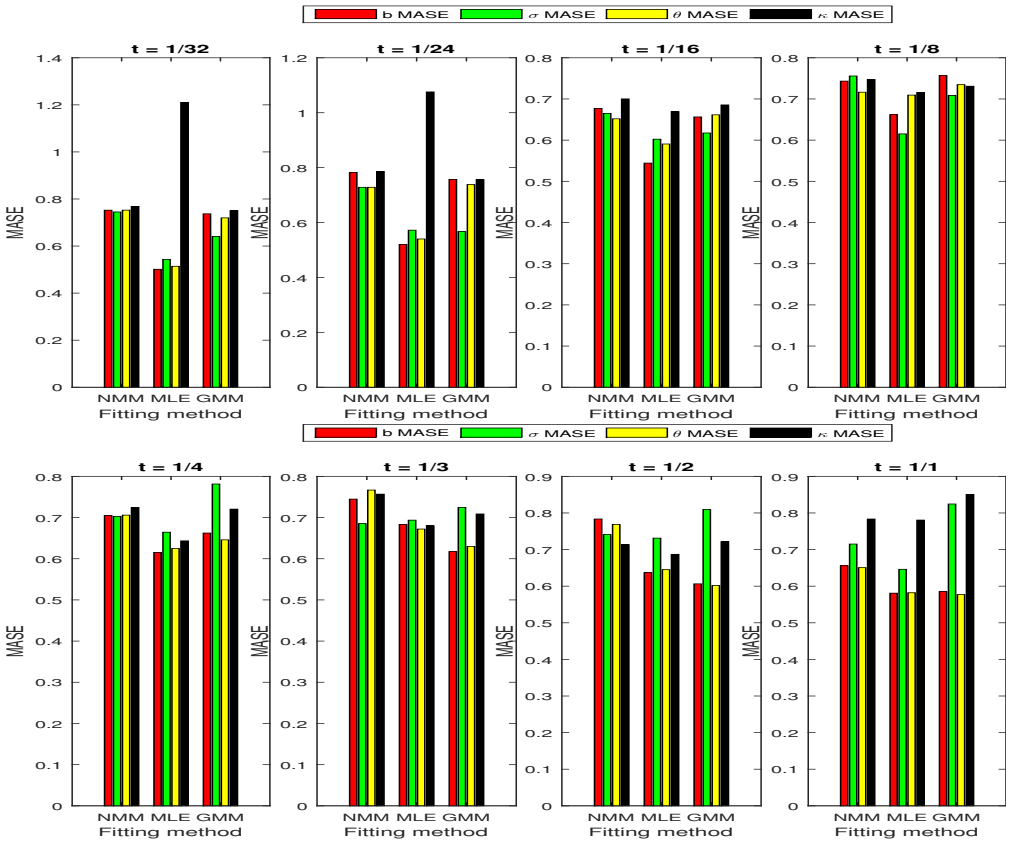


Fig. 5.16: A comparison of the MASE betweeen fitting method for different sampling frequencies of 200 simulated VG processes.

The ARMAD is included for ease of comparison between the methods over the computed 200 simulations for the VG and NIG. The ARMAD is simply the mean absolute deviation of the estimated parameters. The ARMAD for the VG or NIG is given by

$$\frac{1}{200}\sum_{i=1}^{200}\left\{\left|\frac{\hat{b}_i}{b}-1\right|+\left|\frac{\hat{\sigma}_i}{\sigma}-1\right|+\left|\frac{\hat{\theta}_i}{\theta}-1\right|+\left|\frac{\hat{\kappa}_i}{\kappa}-1\right|\right\}. \tag{5.41}$$

Tab. 5.1: Partial horizontal line

Δt	NIG			VG		
	GMM	MLE	NMM	GMM	MLE	NMM
1	0.7601	0.8417	1.8578	2.1123	2.4607	2.2753
$\frac{1}{2}$	0.4768	0.3380	0.5357	1.0073	1.0251	0.9795
$\frac{1}{3}$	0.3446	0.2429	0.4048	0.6343	0.4962	0.7276
$\frac{1}{4}$	0.3085	0.1987	0.3760	0.6535	0.4402	0.6496
$\frac{1}{8}$	0.2823	0.1500	0.3503	0.4476	0.2823	0.5352
$\frac{1}{16}$	0.2489	0.1225	0.3348	0.4273	0.3738	0.4566
$\frac{1}{24}$	0.2393	0.1080	0.3041	0.3677	1.4751	0.4084
$\frac{1}{32}$	0.2479	0.0992	0.3289	0.4144	3.1970	0.4351

Tab. 5.2: The ARMAD from 200 simulations for NIG process and 200 simulations of the VG

The NIG seems to perform well at all sampling time spans in comparison with the VG. For all three estimation methods considered the ARMAD deviation increases as the time span between consecutive observations (Δt) increases for the NIG. However for the VG small values of the ARMAD are realized when $\Delta t < 0.5$ for the GMM and NMM. The MLE gives an ARMAD of less than one for mid-range sampling frequencies only that is when $\Delta t >= 1/2$ or $\Delta t <= 1/24$. The GMM, MLE and NMM ARMAD is greater than one when $\Delta t = 1$ for the NIG. From table 5.2, the NIG performs better in comparison to the VG.

5.7 Fitting to financial data

Results from the calibration of the NIG and VG are presented in this section. The normal distribution is used as a benchmark. The motivation and discussion of the measures for goodness of fit used is given in section 4.7. The GMM, NMM and MLE were used to estimate the parameters. The GMM estimates were used as starting values for the MLE.

Summary statistics and the motivation for the period of of the data used is provided in chapter 2. The sampling time spans of the log returns used for calibration ranges from 20 minutes to 8 hours. The 8 hour log returns are also referred to as daily returns.

Fig. 5.17: 20 minute returns

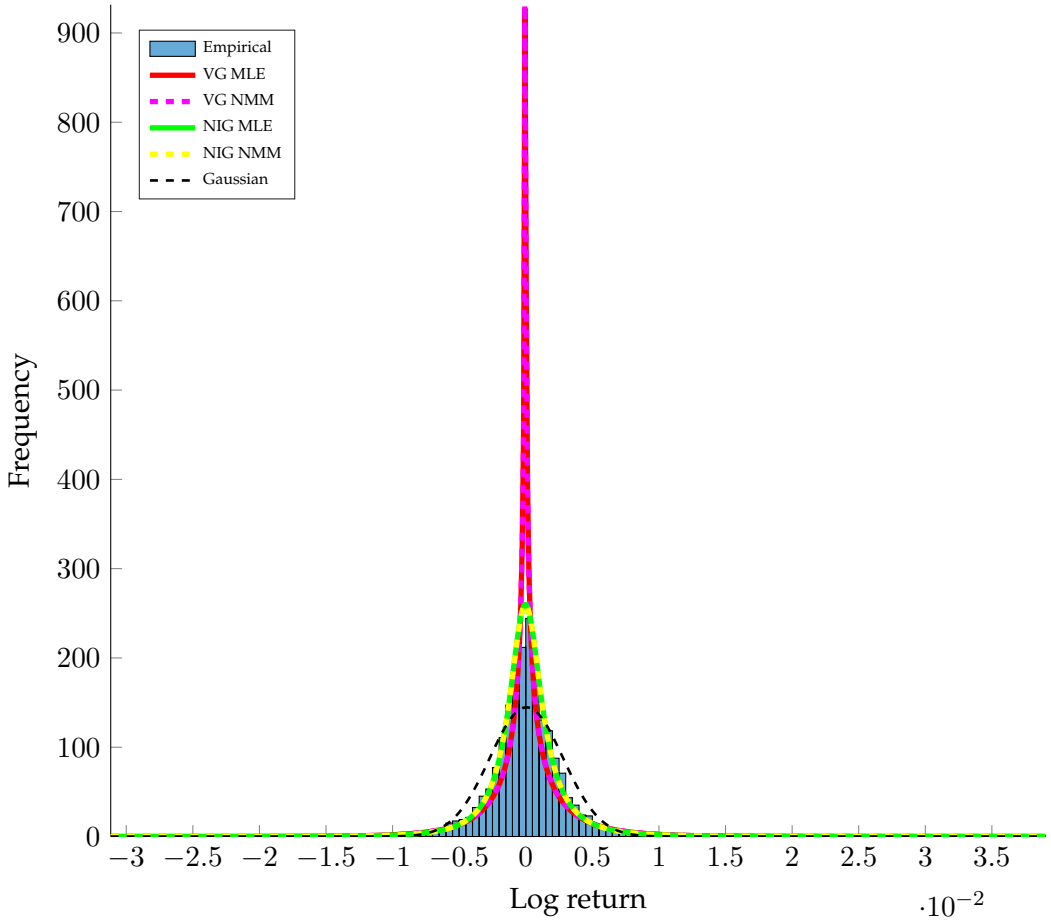


Fig. 5.18: Density plots for the NMM and MLE of 20 minute NPN log returns based on a 12 month window.

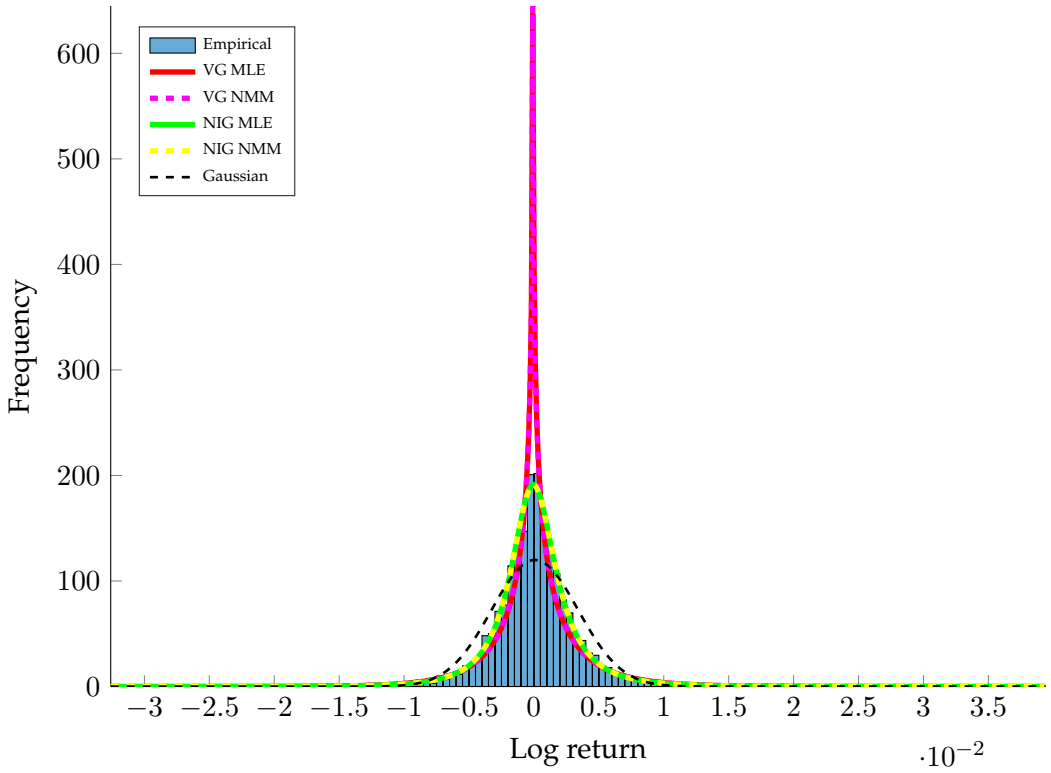


Fig. 5.19: Density plots for the NMM and MLE of 30 minute NPN log returns based on a 12 month window.

The VG has a tendency to overestimate the peak of the distribution and this tendency becomes more pronounced as Δt becomes smaller. The NIG and the normal distribution underestimate the peak of the empirical distribution. The inability of the normal distribution to accurately capture the peak of the empirical distribution increases as Δt decreases. There are no significant differences between the curves of the MLE and NMM for both the VG and NIG models.

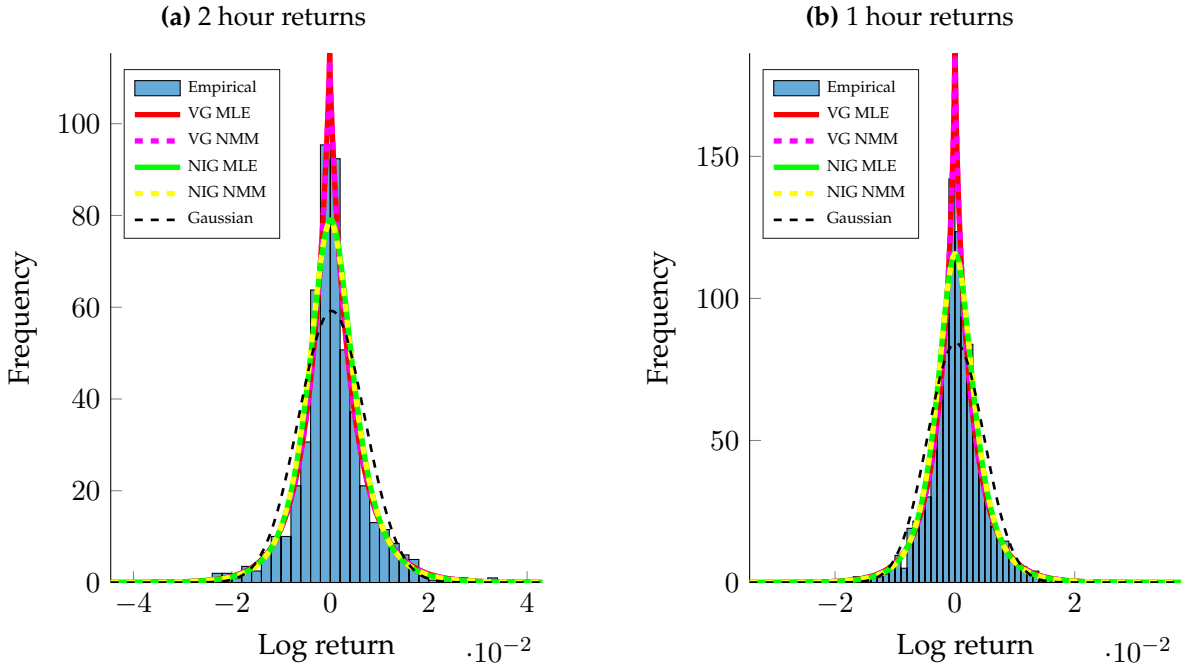


Fig. 5.20: Density plots for the NMM and MLE based for NPN log returns based on a 12 month window.

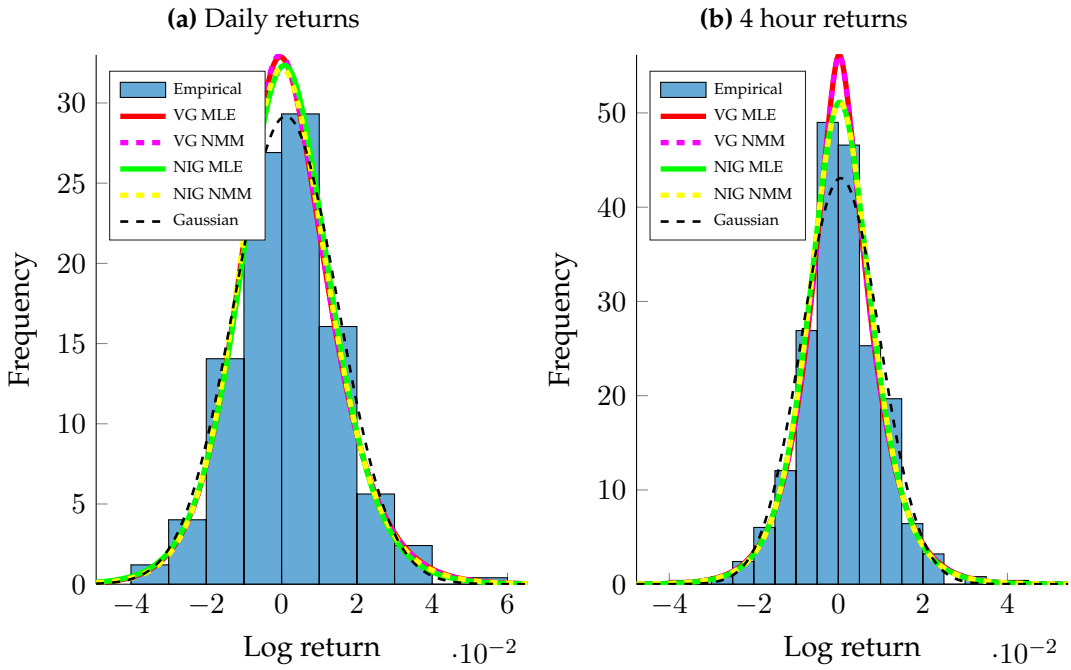


Fig. 5.21: Density plots for the NMM and MLE for NPN log returns based on a 12 month window.

Sampling Frequency	Class	Method	AGL			MTN			NPN		
			KS	AD	BIC_0	KS	AD	BIC_0	KS	AD	BIC_0
Daily	NIG	MLE	0.033581	0.090162	675.79	0.028088	0.090766	685.18	0.02652	0.06873	718.96
		NMM	0.033696	0.090455	675.75	0.031188	0.09809	685.11	0.025454	0.079835	718.94
	VG	MLE	0.036583	0.095795	675.67	0.031408	0.098993	685.11	0.025687	0.072129	718.99
		NMM	0.033819	0.089113	675.77	0.031114	0.098166	685.15	0.025109	0.079405	718.99
4 hours	NIG	MLE	0.020994	0.059204	1539.5	0.039912	0.13651	1578.5	0.045259	0.090631	1637.7
		NMM	0.020915	0.058994	1539.5	0.046003	0.095366	1581.4	0.046128	0.092364	1637.7
	VG	MLE	0.020979	0.059164	1539.8	0.046004	0.095368	1582.1	0.046118	0.092345	1639.4
		NMM	0.020894	0.058938	1539.8	0.046003	0.095366	1582.1	0.04612	0.092347	1639.3
2 hours	NIG	MLE	0.025599	0.05573	3494.2	0.033488	0.069466	3567.3	0.047304	0.097784	3656.5
		NMM	0.027503	0.056237	3494.0	0.033955	0.069213	3567.3	0.045024	0.096226	3656.5
	VG	MLE	0.027496	0.056222	3497.5	0.033948	0.069199	3575.8	0.044909	0.095912	3674.0
		NMM	0.027487	0.056204	3497.5	0.033946	0.069194	3575.8	0.044835	0.095557	3673.9
1 hour	NIG	MLE	0.025676	0.051366	7622.3	0.027977	0.056201	7792.9	0.023866	0.049239	8004.4
		NMM	0.02572	0.051455	7622.4	0.028022	0.05629	7792.9	0.023813	0.049129	8004.3
	VG	MLE	0.02564	0.051293	7582.9	0.02791	0.056067	7791.4	0.023768	0.049038	8006.7
		NMM	0.025715	0.051444	7583.9	0.027936	0.056119	7791.8	0.023762	0.049027	8006.8
30 minutes	NIG	MLE	0.028948	0.060213	16606	0.025476	0.053719	16966	0.014082	0.030377	17528
		NMM	0.028636	0.059293	16607	0.025553	0.053869	16966	0.014085	0.029573	17528
	VG	MLE	0.028937	0.06019	16344	0.025551	0.053865	16985	0.014091	0.0304	17365
		NMM	0.028819	0.060276	16352	0.025551	0.053864	16985	0.014026	0.029677	17369
20 minutes	NIG	MLE	0.032723	0.075455	26174	0.023046	0.048474	26632	0.03381	0.083931	27404
		NMM	0.032429	0.074748	26176	0.023071	0.048527	26632	0.033482	0.083067	27405
	VG	MLE	0.032723	0.075453	25719	0.023061	0.048507	26617	0.0338	0.083905	26685
		NMM	0.032729	0.075475	25726	0.023061	0.048505	26617	0.033784	0.083862	26693

Tab. 5.3: Goodness of fit for AGL, MTN and NPN log returns based on a 12 month horizon window.

Sampling Frequency	Class	Method	AGL			MTN			NPN		
			KS	AD	BIC_0	KS	AD	BIC_0	KS	AD	BIC_0
Daily	NIG	MLE	0.086045	0.25193	353.18	0.057483	0.17156	349.77	0.083301	0.22193	378.83
		NMM	0.087235	0.25447	352.99	0.057249	0.17749	349.45	0.084297	0.22584	378.66
	VG	MLE	0.091566	0.26362	352.29	0.057497	0.17823	349.47	0.085221	0.22798	378.38
		NMM	0.08683	0.25361	353.03	0.057165	0.17752	349.50	0.083792	0.22485	378.65
4 hours	NIG	MLE	0.063818	0.15777	791.43	0.083229	0.17813	802.67	0.064408	0.13193	840.83
		NMM	0.063723	0.15756	791.40	0.094632	0.19883	800.85	0.065278	0.13256	840.84
	VG	MLE	0.063778	0.15768	791.68	0.094633	0.19883	801.37	0.065268	0.13278	842.10
		NMM	0.063697	0.15751	791.70	0.094632	0.19883	801.37	0.06527	0.13249	842.11
2 hours	NIG	MLE	0.061748	0.13946	1768.1	0.0699	0.14181	1800.6	0.071232	0.14485	1854.5
		NMM	0.06028	0.13877	1767.9	0.069435	0.1409	1800.6	0.071232	0.14485	1854.4
	VG	MLE	0.060266	0.13877	1774.2	0.069434	0.1409	1808.6	0.068242	0.14451	1866.6
		NMM	0.06026	0.13871	1774.2	0.069427	0.14088	1808.6	0.068185	0.14417	1866.5
1 hour	NIG	MLE	0.029934	0.10519	3851.9	0.037076	0.090568	3968.7	0.036079	0.076813	4035.0
		NMM	0.030367	0.10648	3851.6	0.03718	0.090186	3968.7	0.036207	0.076845	4035.0
	VG	MLE	0.030449	0.10672	3836.3	0.037143	0.090604	3971.4	0.036181	0.07685	4039.7
		NMM	0.030196	0.10597	3836.8	0.037111	0.090277	3971.7	0.036175	0.07672	4039.8
30 minutes	NIG	MLE	0.035723	0.1111	8417.4	0.040216	0.098659	8577.7	0.019954	0.067152	8813.1
		NMM	0.035807	0.1110	8417.5	0.040162	0.098538	8577.7	0.019889	0.066824	8813.3
	VG	MLE	0.035713	0.11113	8295.9	0.040161	0.098535	8594.1	0.019963	0.067125	8732.8
		NMM	0.035796	0.11061	8300.9	0.04016	0.098533	8594.2	0.019889	0.066447	8734.5
20 minutes	NIG	MLE	0.021297	0.077182	13287	0.048795	0.11923	13471	0.025868	0.064215	13758
		NMM	0.021406	0.079082	13287	0.048792	0.11922	13471	0.025539	0.063362	13759
	VG	MLE	0.021304	0.077148	13095	0.048783	0.1192	13486	0.025858	0.064189	13419
		NMM	0.021351	0.078361	13099	0.04878	0.1192	13486	0.025841	0.064147	13423

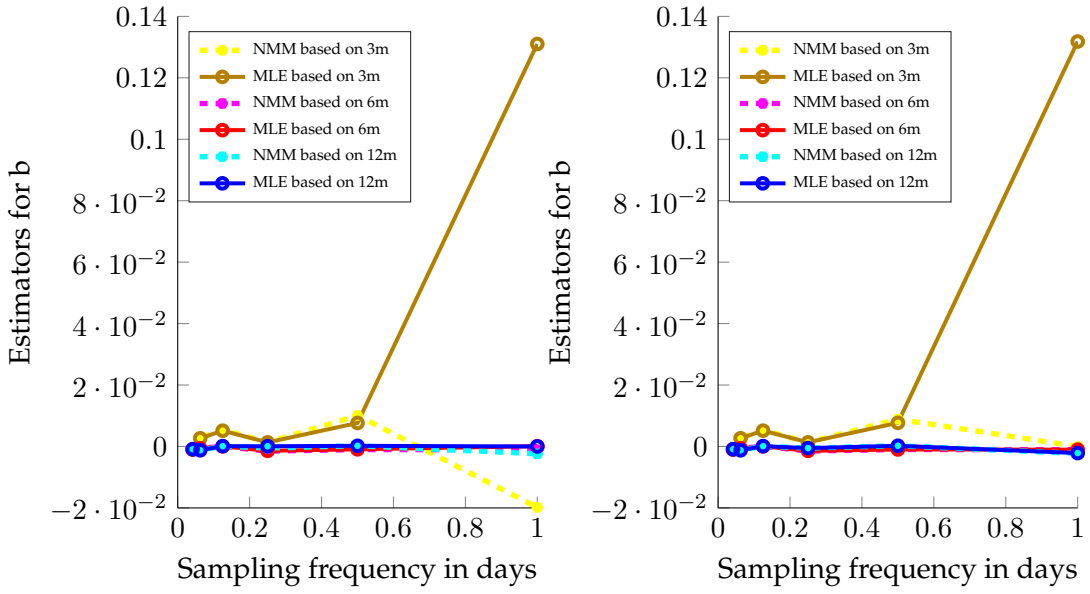
Tab. 5.4: Goodness of fit for AGL, MTN and NPN log returns based on a 6 month horizon window.

Sampling Frequency	Class	Method	AGL			MTN			NPN		
			KS	AD	BIC_0	KS	AD	BIC_0	KS	AD	BIC_0
Daily	NIG	MLE	0.086664	0.22648	176.70	0.12229	0.26693	181.13	0.12448	0.34433	194.76
		NMM	0.08809	0.2292	176.61	0.12529	0.27271	180.94	0.12547	0.34927	194.54
	VG	MLE	0.092067	0.23853	176.30	0.12552	0.27315	180.96	0.1264	0.35124	194.30
		NMM	0.087624	0.22828	176.64	0.12522	0.27256	180.99	0.12497	0.34855	194.48
4 hours	NIG	MLE	0.06543	0.15752	394.73	0.080485	0.17734	417.04	0.096592	0.2111	424.45
		NMM	0.065326	0.15731	394.71	0.099129	0.21731	414.41	0.09746	0.21134	424.45
	VG	MLE	0.065376	0.15743	395.01	0.099129	0.21731	414.65	0.09745	0.21155	425.24
		NMM	0.065299	0.15726	395.01	0.099129	0.21731	414.65	0.097449	0.21134	425.25
2 hours	NIG	MLE	0.048871	0.12457	885.79	0.10871	0.22783	936.30	0.0774	0.15592	926.71
		NMM	0.048981	0.12487	885.51	0.10916	0.22869	936.29	0.074325	0.14984	926.64
	VG	MLE	0.048972	0.12486	887.72	0.10915	0.22868	943.54	0.074288	0.14976	931.68
		NMM	0.048953	0.12481	887.72	0.10915	0.22867	943.54	0.074235	0.14966	931.67
1 hour	NIG	MLE	0.039043	0.1490	1955.1	0.077373	0.16711	2048.8	0.032621	0.10926	2014.8
		NMM	0.039823	0.14997	1954.9	0.077537	0.16744	2048.8	0.032721	0.10914	2014.8
	VG	MLE	0.03969	1948.7	0.15034	0.077312	0.16699	2057.3	0.032726	0.10929	2015.7
		NMM	0.039481	0.14961	1948.9	0.077298	0.16696	2057.4	0.032724	0.10916	2015.7
30 minutes	NIG	MLE	0.044107	0.13152	4243.6	0.061129	0.13831	4409.5	0.017326	0.064517	4375.9
		NMM	0.044191	0.13161	4243.7	0.061206	0.13844	4409.5	0.017229	0.063837	4376.1
	VG	MLE	0.044097	0.13155	4182.4	0.061204	0.13844	4426.0	0.017335	0.064534	4322.9
		NMM	0.04418	0.13111	4185.8	0.061204	0.13844	4426.0	0.017256	0.063963	4323.7
20 minutes	NIG	MLE	0.034902	0.15722	6685.4	0.073547	0.17393	6948.5	0.046257	0.11484	6848.1
		NMM	0.034984	0.1609	6685.6	0.073572	0.17393	6948.5	0.045929	0.11396	6848.5
	VG	MLE	0.03491	0.15715	6584.7	0.073564	0.17391	6986.3	0.046247	0.11482	6673.3
		NMM	0.03496	0.1595	6587.1	0.073563	0.1739	6986.4	0.046231	0.11477	6676.5

Tab. 5.5: Goodness of fit for AGL, MTN and NPN log returns based on a 3 month horizon window.

Table 5.3 - 5.5 report the goodness of fit of the NIG and VG for AGL, MTN and NPN log returns. The higher the log likelihood the better the model is to fitting the returns and conversely the lower the KS statistic or the AD statistic the better the model. For both models there are no significant difference between the log likelihood, KS statistic as well the AD statistic between the MLE and the NMM. However as Δt decreases the difference between the values of the log likelihood for the VG and NIG increases. The NIG performs better than the VG for AGL and NPN log returns when $\Delta t \leq 1$ hour and the opposite for MTN returns when $\Delta \leq 1$ hour. The value of the log likelihood increases significantly as Δt decreases for all three time windows, suggesting that more data will result in a better fit.

For all sampling time spans the parameters from the GMM and the MLE of the NIG were identical see table 5.6. Note that the GMM estimates were used as starting values for the MLE. There were slight differences between the estimates from the GMM and MLE of the VG.

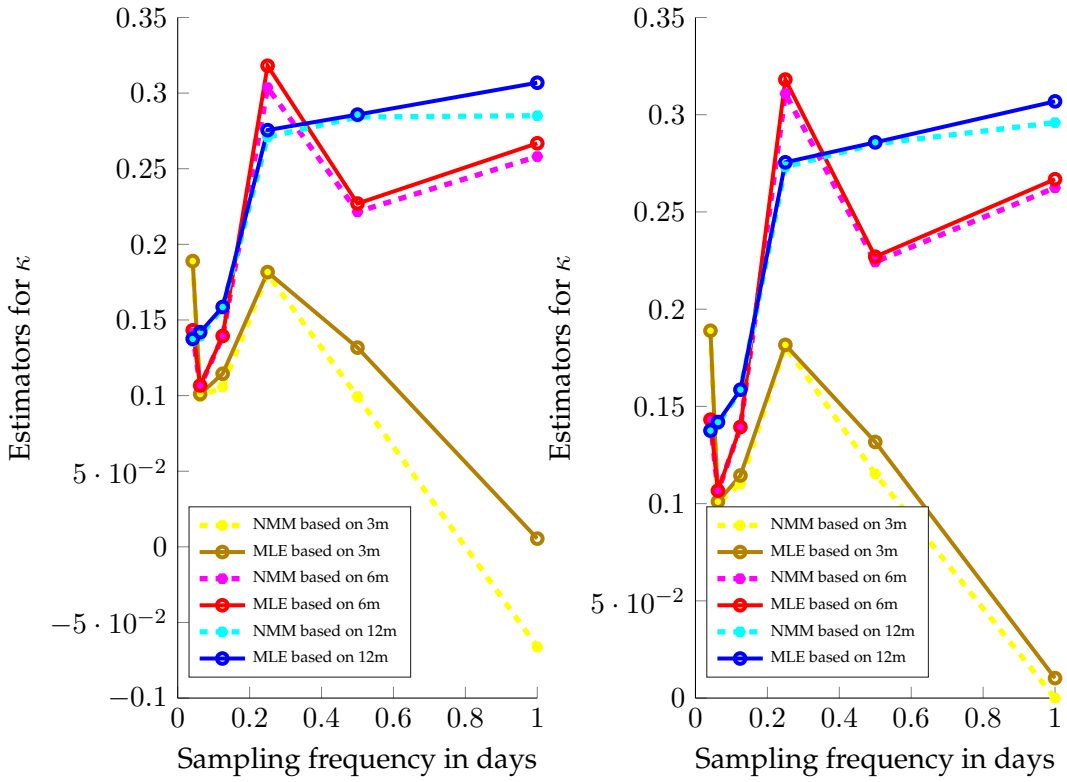


(a) NIG model **(b)** VG model

Fig. 5.22: Signature plots for the NMM AND MLE of NPN log returns based on different time windows.

Sampling Frequency	Class	Method	b	σ	θ	κ	$\mathscr{L}$
Daily	NIG	GMM	3.3961e-06	0.0010836	0.013635	0.30686	718.96
		MLE	3.3961e-06	0.0010836	0.013635	0.30686	718.96
		NMM	-0.0023783	0.0035205	0.013546	0.28506	718.94
	VG	GMM	5.7151e-06	0.0010812	0.013635	0.30686	719.12
		MLE	-0.0021481	0.0032902	0.013677	0.30686	718.99
		NMM	-0.0022703	0.0034125	0.013549	0.29593	718.99
4 hours	NIG	GMM	0.00021094	0.00089364	0.013098	0.28578	1637.7
		MLE	0.00021094	0.00089364	0.013098	0.28578	1637.7
		NMM	0.00022731	0.00091487	0.013089	0.2842	1637.7
	VG	GMM	0.00021092	0.0008937	0.013098	0.28578	1639.4
		MLE	0.00022964	0.00091254	0.013098	0.28578	1639.4
		NMM	0.00022943	0.00091275	0.013089	0.28499	1639.3
2 hours	NIG	GMM	7.7267e-05	0.0010473	0.01344	0.27552	3656.5
		MLE	7.7267e-05	0.0010473	0.01344	0.27552	3656.5
		NMM	-0.00054401	0.0016862	0.013455	0.27091	3656.5
	VG	GMM	7.3855e-05	0.0010508	0.01344	0.27552	3673.7
		MLE	-0.00051833	0.0016605	0.013483	0.27552	3674.0
		NMM	-0.00053217	0.0016744	0.013455	0.27322	3673.9
1 hour	NIG	GMM	0.00010576	0.0010277	0.01339	0.15858	8004.4
		MLE	0.00010576	0.0010277	0.01339	0.15858	8004.4
		NMM	0.00010702	0.0010352	0.013383	0.15798	8004.3
	VG	GMM	0.00010576	0.0010277	0.01339	0.15858	8006.7
		MLE	0.00011014	0.001032	0.01339	0.15858	8006.7
		NMM	0.00010864	0.0010335	0.013383	0.15828	8006.8
30 minutes	NIG	GMM	-0.0012333	0.002373	0.013306	0.14195	17528
		MLE	-0.0012333	0.002373	0.013306	0.14195	17528
		NMM	-0.0012774	0.0024196	0.013275	0.13938	17528
	VG	GMM	-0.0012333	0.002373	0.013306	0.14195	17365
		MLE	-0.0012345	0.0023767	0.013306	0.14195	17365
		NMM	-0.001259	0.0024012	0.013275	0.14066	17369
20 minutes	NIG	GMM	-0.00087653	0.0020158	0.013508	0.13754	27404
		MLE	-0.00087653	0.0020158	0.013508	0.13754	27404
		NMM	-0.00090201	0.0020442	0.013487	0.13585	27405
	VG	GMM	-0.00087653	0.0020158	0.013508	0.13754	26685
		MLE	-0.00087738	0.0020196	0.013508	0.13754	26685
		NMM	-0.00089147	0.0020337	0.013487	0.1367	26693

Tab. 5.6: Estimated parameters from NPN log returns using a 12 month time window.

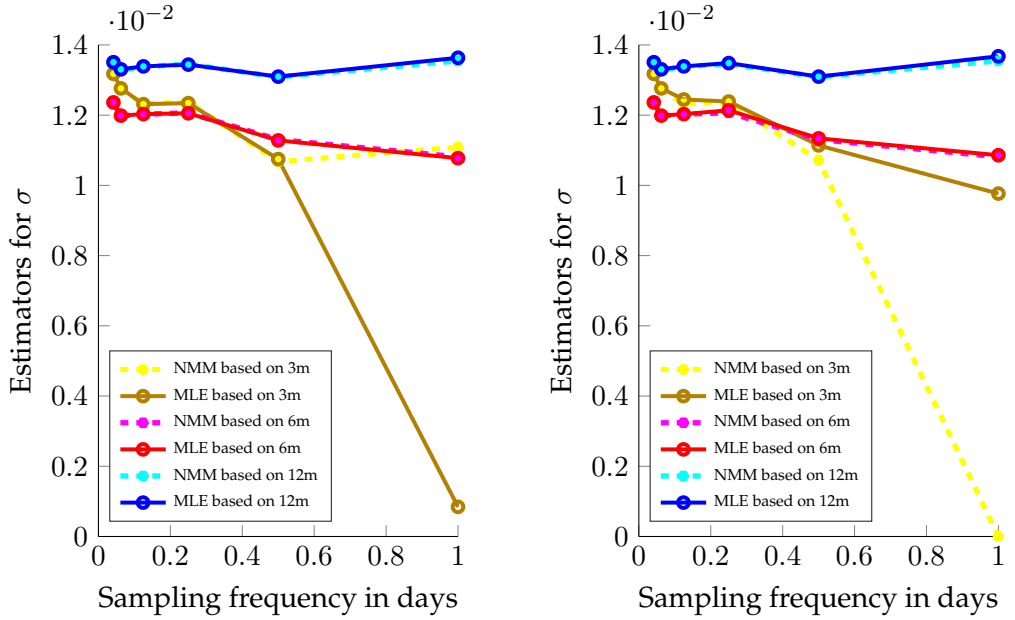


(a) NIG model **(b)** VG model

Fig. 5.23: Signature plots for the NMM AND MLE of NPN log returns based on different time windows.

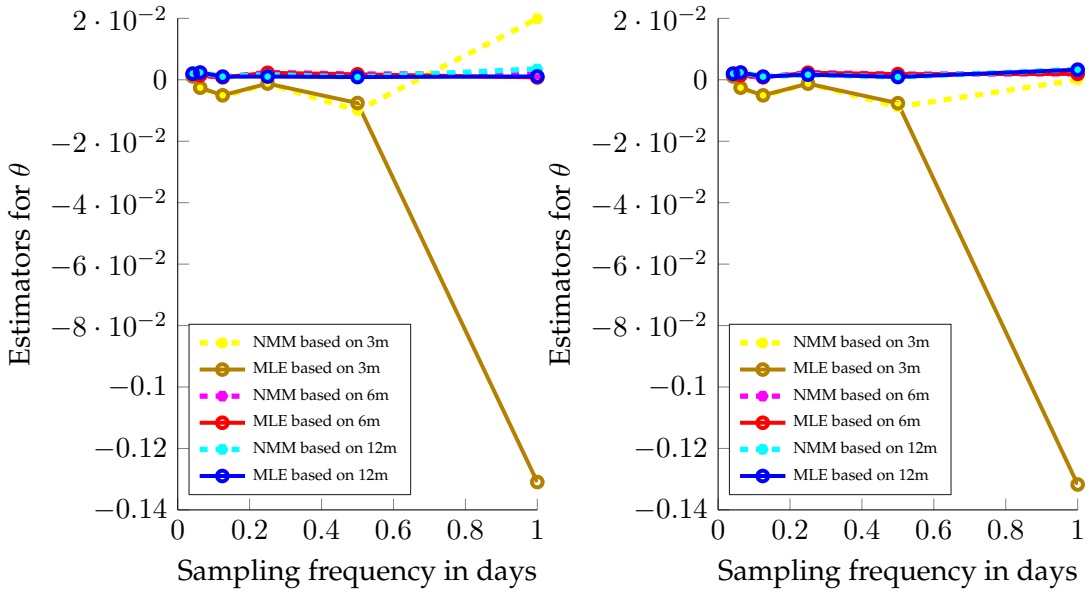
Figueroa-López *et al.* (2011) states that κ is the excess percentage kurtosis multiplied by the time span when $\theta = 0$. The drawback of the NMM is that the estimate of κ can be negative if $\hat{\varepsilon} < -1/5$ for the NIG. In figure 5.23, κ obtained from the NMM of the NIG model is negative for daily log returns for the three month window.

The values of σ for NPN returns do not vary that much except for daily sampling frequency for the three month window. The values of σ and θ obtained from the MLE and NMM of NPN returns are close to zero for all sampling except for daily returns from the three month window for both the VG and NIG.



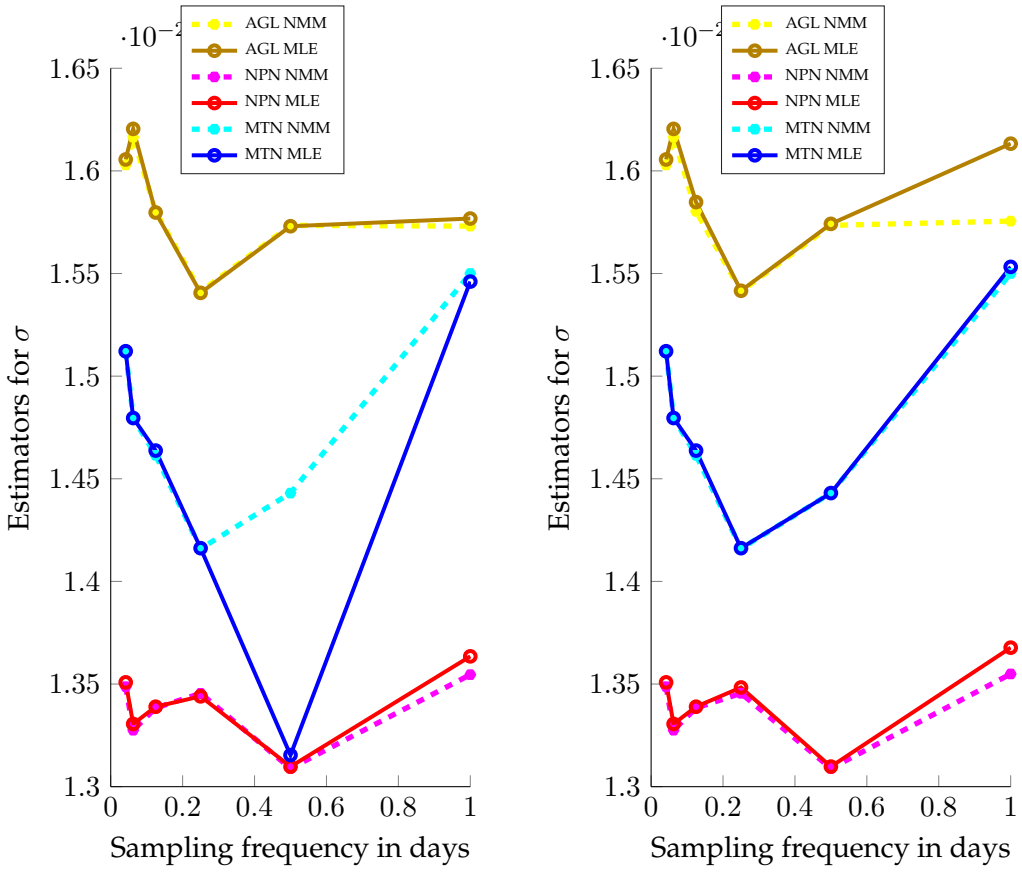
(a) NIG model **(b)** VG model

Fig. 5.24: Signature plots for the NMM AND MLE bsed on different time windows.



(a) Estimators for b using the NIG model. **(b)** Estimators for θ using the NIG.

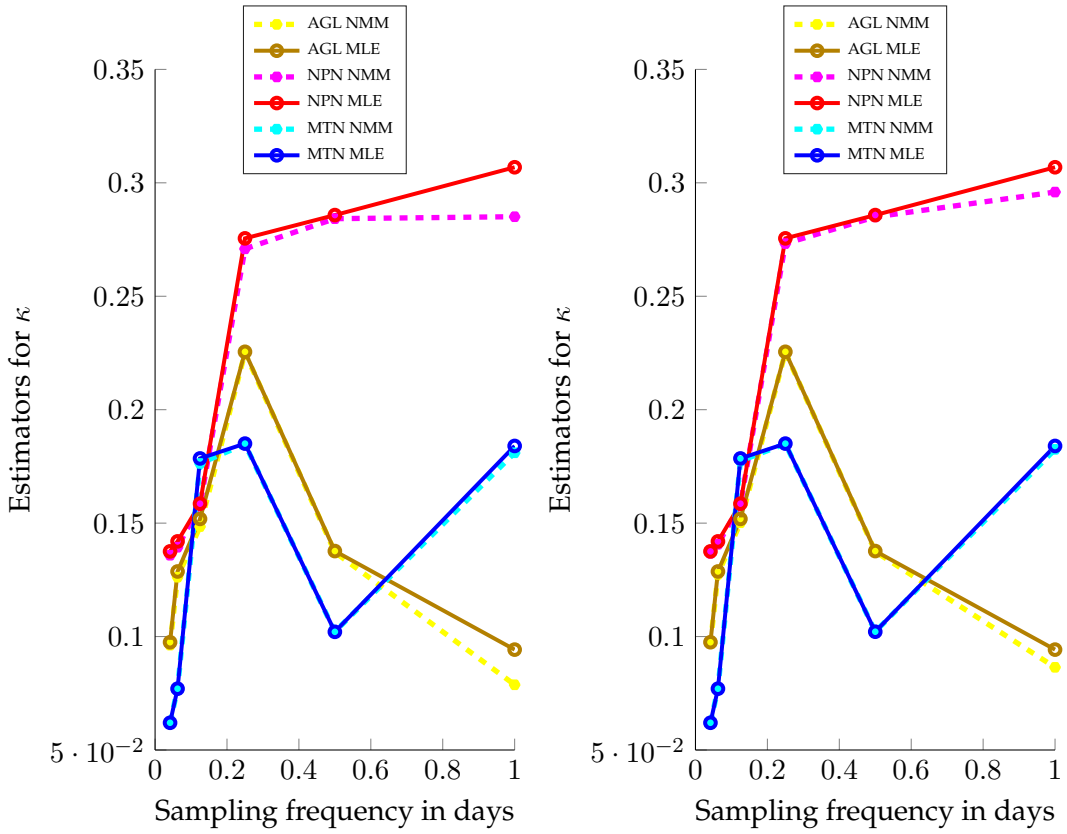
Fig. 5.25: Signature plots for the NMM AND MLE of NPN log returns based on different time windows.



(a) VG estimators **(b)** NIG estimators

Fig. 5.26: Signature plots for the NMM AND MLE using log returns of AGL, MTN based on a 12 month window.

The assumption of independent of increments poses a limitation when using Lévy processes to model high frequency data. Due to this limitation, Figueroa-López *et al.* (2011) suggested using signature plots of point estimators to assess the suitability of Lévy models. The variability parameter (σ) is stable in comparison to the kurtosis estimator (κ) even for sampling time spans as small as 20 minutes, see figures 5.24 and 5.23. Figures 5.26 and 5.27 show signature plots of three different stocks and from the figures the same conclusion about the variability and kurtosis estimators can be drawn.



(a) VG estimators

(b) NIG estimators

Fig. 5.27: Signature plots for the NMM AND MLE using log returns of AGL, MTN based on a 12 month window.

To assess the temporal stability of the parameters thirty four, 12 months periods were considered. The first window started on the 6th of February 2012 and ended on 4 February 2013 and the last period started on the 25th of September 2012 and ended on the 23rd of September 2013. The beginning of each period was the first business day of each week. To ensure that each window had the same number of observations, observations from the day of the period began and observations from the 249 business days that followed it were used.

Figure 5.28 shows that σ has an outlier for 30 minute sampling observation. The values of κ are stable as one moves through time for 20 minute, 30 minute and 1 hour returns. Daily returns have the greatest variability for κ. The drift parameter θ seems fairly well behaved for mid range frequencies. The parameter b has a distinguished outlier for 30 minute returns otherwise

the values are fairly stable.

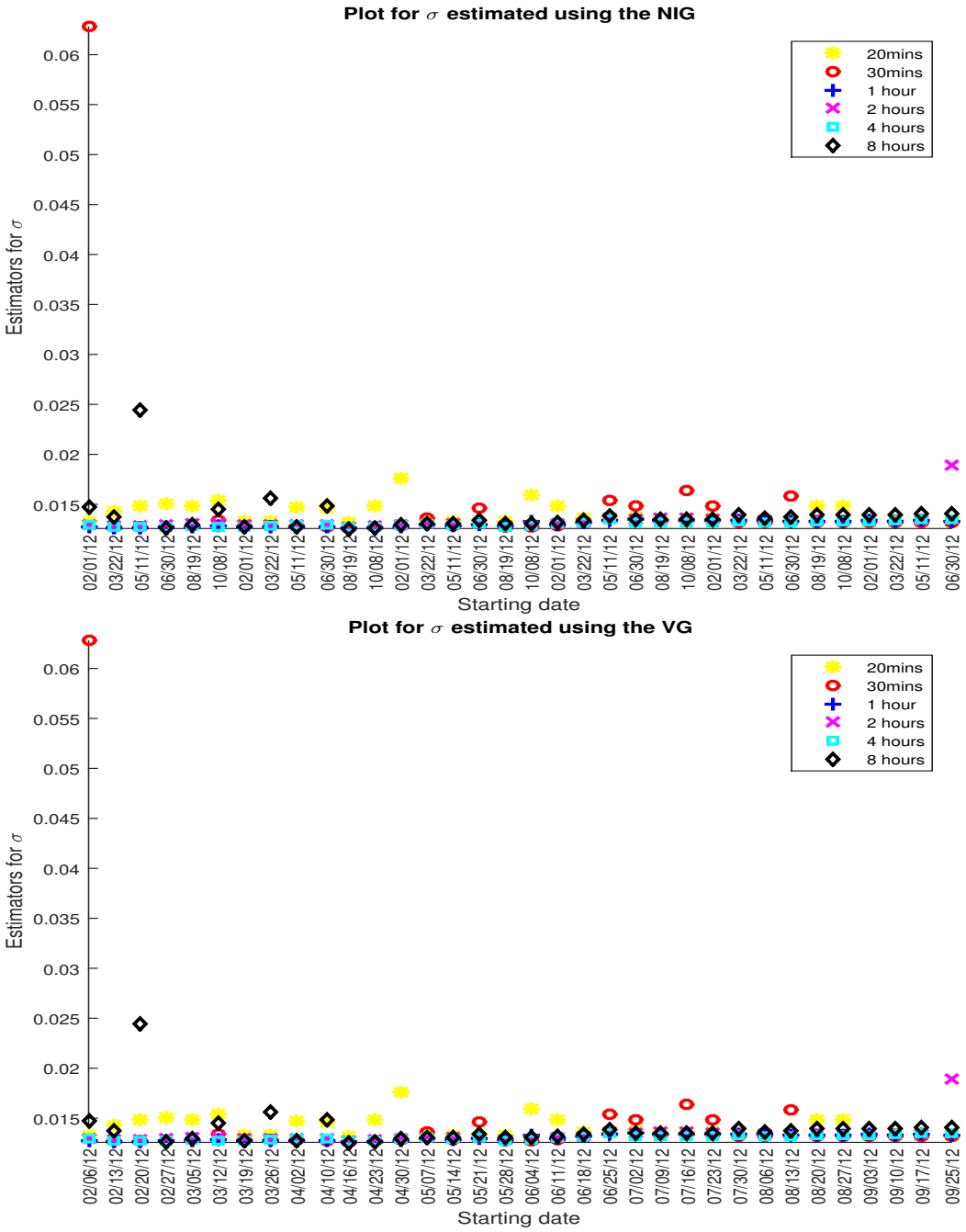


Fig. 5.28: Plot showing how σ varies when rolling 12 month windows for NPN returns.

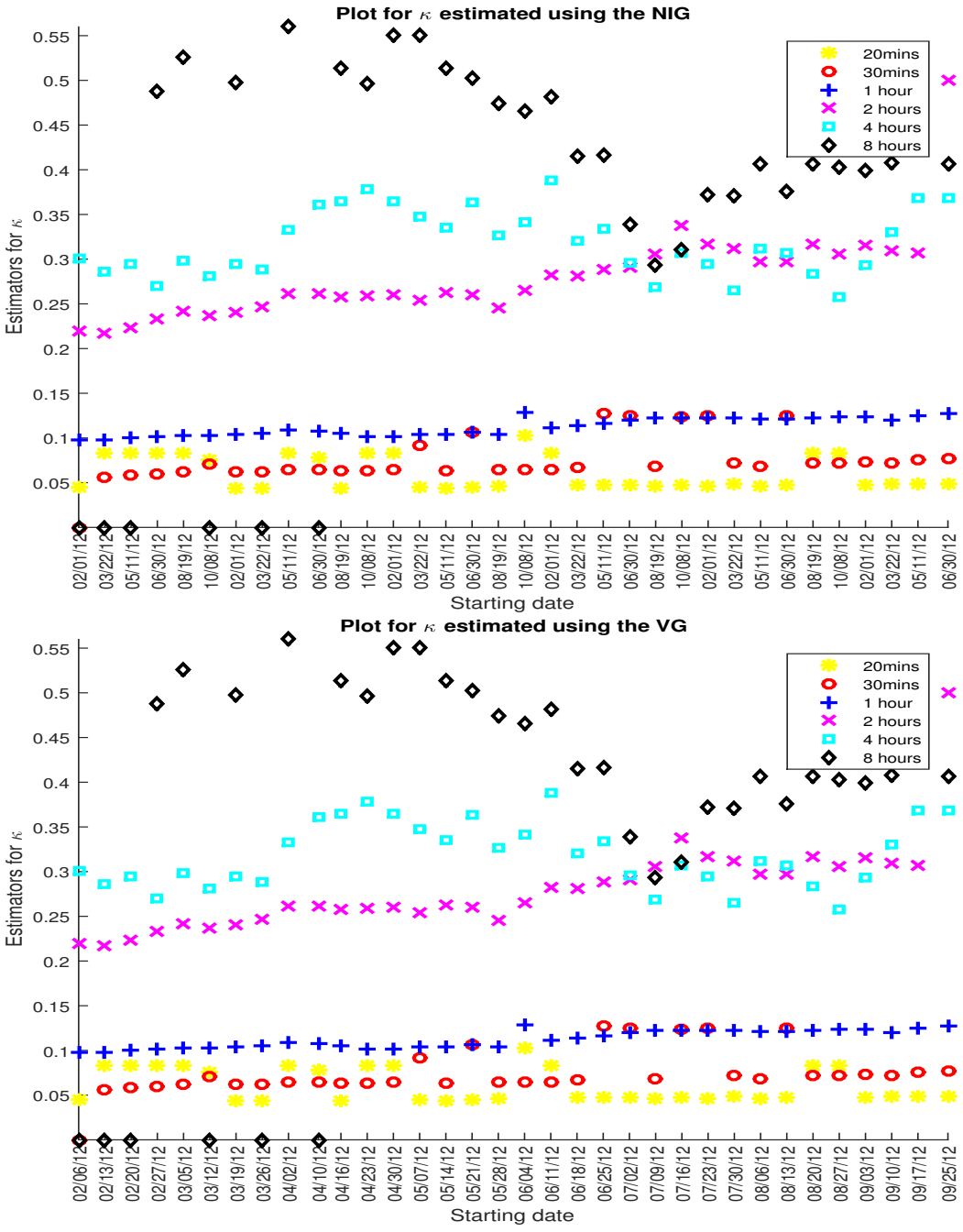


Fig. 5.29: Plot showing how σ varies when rolling 12 month windows for NPN returns using the MLE.

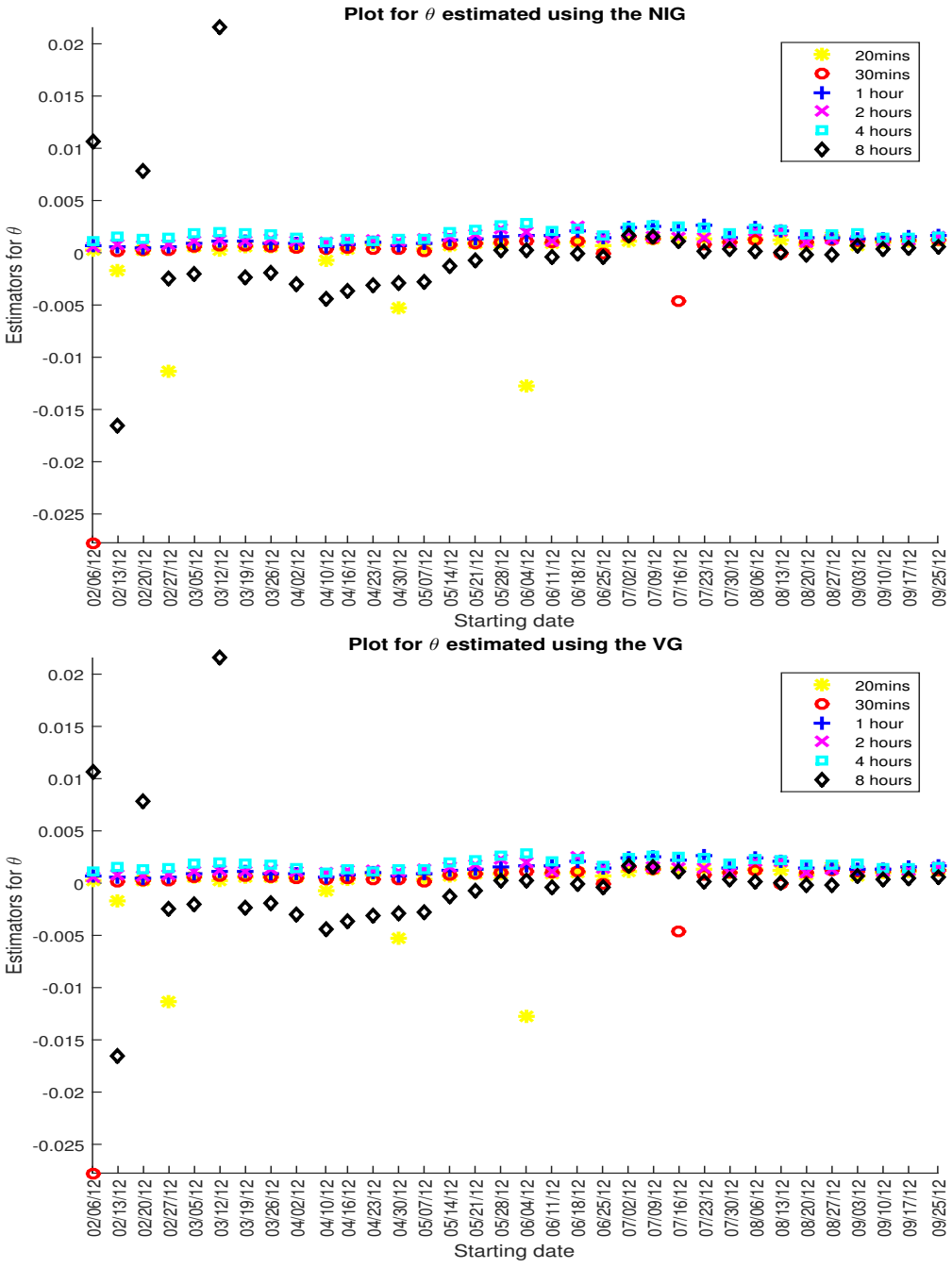


Fig. 5.30: Plot showing how σ varies when rolling 12 month windows for NPN returns using the MLE.

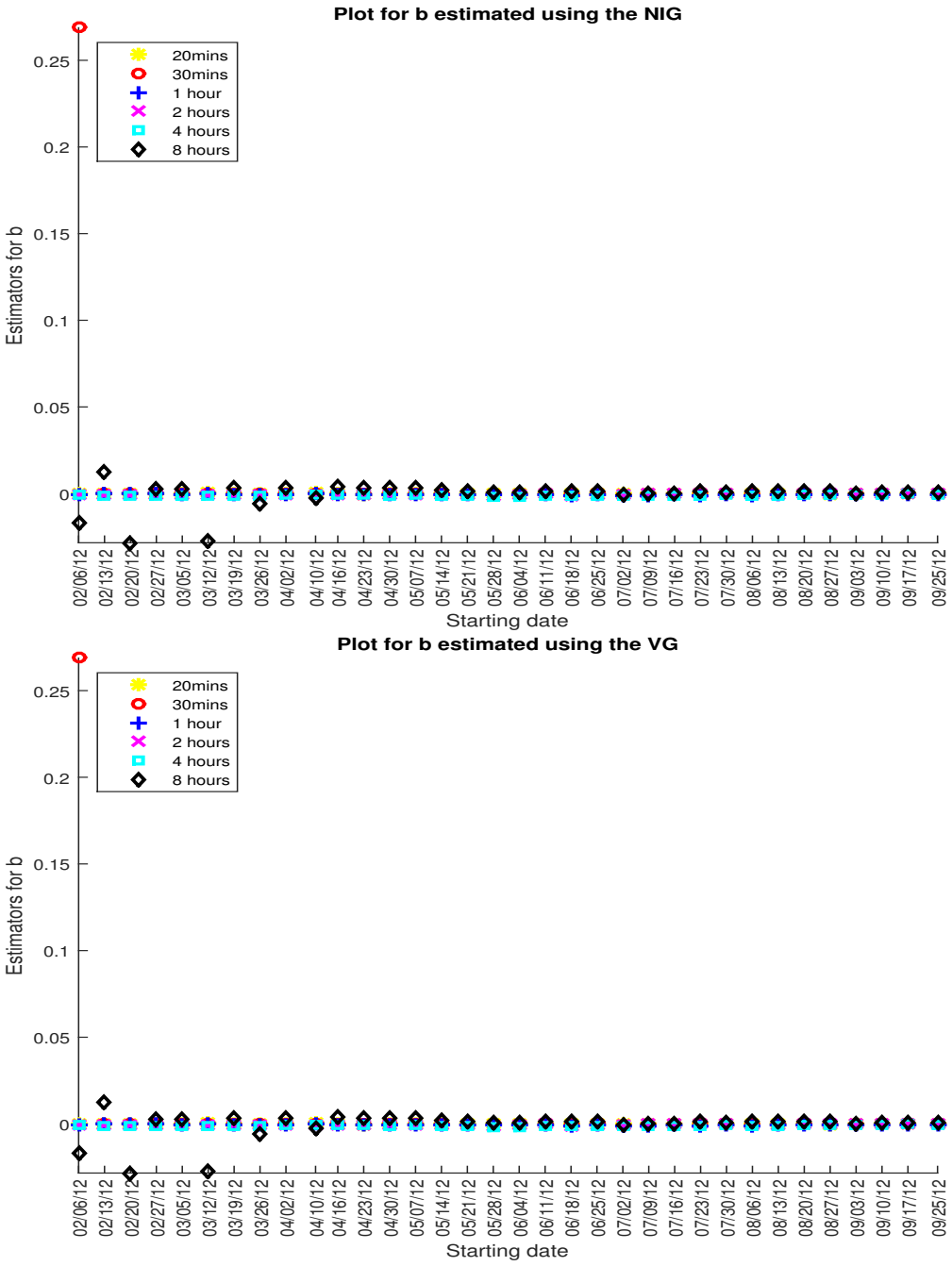


Fig. 5.31: Plot showing how b varies when rolling 12 month windows for NPN returns using the MLE.

5.8 Comparison of the representations

Tables 5.7 and 5.8 show a comparison of the goodness of fit of the estimation methods of the subordinated representation with the normal variance mean mixture for the NIG and VG. The subordinated Brownian motion representation of the VG and the NIG was discussed in this chapter. The NMM, MLE and GMM were used to calibrate the parameters of the subordinated representation. The goodness of fit of the GMM was not included in the tables because the parameters estimated from the GMM were used as starting values for the MLE. The normal mean variance representation was discussed in chapter 4, and the EM algorithm was used to estimate the parameters.

The normal distribution produces higher KS static values than the MLE, NMM and EM for most cases of the NIG, see table 5.7. The difference between the values of the AD test static from the NIG estimation methods becomes more pronounced as Δt decreases, with the normal distribution producing larger values. Using the BIC_0 values in table 5.7, the MLE performs slightly better than the NMM and EM for most cases, followed by the EM. For all sampling frequencies the values of the BIC_0 do not vary significantly for all three estimation methods of the NIG. Although there are slight differences between the BIC_0 values from the NIG methods and the normal distribution, the high values of the AD test statistic highlights the failure of the normal distribution in comparison with the NIG.

The high values of the AD test statistic from the normal distribution in table 5.8 makes a stronger case for using the VG to model intra-day data. Using the BIC_0 values in table 5.8, one can conclude that the EM performs better for most cases followed by the NMM and then the MLE. There are no significant differences between the values of the BIC_0 from the VG estimation methods.

From the visual comparisons of figures 4.9-4.12 and 5.18 -5.21 one can conclude that the VG from the normal mean variance representation gives a better fit of the log returns in comparison to that of the subordinated representation. The VG from the subordinated representation has a tendency to overestimate the peak of the distribution. Whilst for both representations the NIG slightly underestimates the peak of the distribution.

Sampling Frequency	Class	Method	AGL			MTN			NPN		
			KS	AD	BIC_0	KS	AD	BIC_0	KS	AD	BIC_0
Daily	NIG	MLE	0.033581	0.090162	-5.3394	0.028088	0.090766	-5.4148	0.02652	0.06873	-5.6862
		NMM	0.033696	0.090455	-5.3391	0.031188	0.09809	-5.4143	0.025454	0.079835	-5.6860
		EM	0.044725	0.10721	-5.2926	0.03151	0.12734	-5.3718	0.038331	0.09536	-5.6552
	Gaussian	MLE	0.039114	0.47189	-5.3756	0.045447	0.58277	-5.7059	0.04166	0.35057	-5.4515
4 hours	NIG	MLE	0.020994	0.059204	-6.1328	0.039912	0.13651	-6.2894	0.045259	0.090631	-6.5273
		NMM	0.020915	0.058994	-6.1328	0.046003	0.095366	-6.3009	0.046128	0.092364	-6.5273
		EM	0.025715	0.074627	-6.1129	0.030741	0.081045	-6.2895	0.039796	0.084865	-6.5125
	Gaussian	MLE	0.035367	0.99777	-6.1352	0.053927	2.018	-6.503	0.054533	1.8012	-6.3092
2 hours	NIG	MLE	0.025599	0.05573	-6.9887	0.033488	0.069466	-7.1356	0.047304	0.097784	-7.3147
		NMM	0.027503	0.056237	-6.9884	0.033955	0.069213	-7.1356	0.045024	0.096226	-7.3146
		EM	0.019452	0.050831	-6.9853	0.028108	0.057434	-7.1357	0.03708	0.08677	-7.314
	Gaussian	MLE	0.063024	9.0436	-6.8803	0.091496	16.417	-7.1482	0.064109	8.2839	-7.0500
1 hour	NIG	MLE	0.025676	0.051366	-7.6376	0.027977	0.056201	-7.809	0.023866	0.049239	-8.0213
		NMM	0.02572	0.051455	-7.6378	0.028022	0.05629	-7.8089	0.023813	0.049129	-8.0212
		EM	0.013247	0.078163	-7.6361	0.027621	0.057193	-7.8023	0.022868	0.058801	-8.0148
	Gaussian	MLE	0.054081	14.024	-7.5239	0.072389	23.723	-7.861	0.064284	Inf	-7.6828
30 minutes	NIG	MLE	0.028948	0.060213	-8.3281	0.025476	0.053719	-8.5086	0.014082	0.030377	-8.7907
		NMM	0.028636	0.059293	-8.3285	0.025553	0.053869	-8.5086	0.014085	0.029573	-8.7908
		EM	0.013054	0.19427	-8.3337	0.024667	0.051787	-8.5043	0.025113	0.081956	-8.778
	Gaussian	MLE	0.061693	38.042	-8.1757	0.07733	Inf	-8.5699	0.072343	45.456	-8.3576
20 minutes	NIG	MLE	0.032723	0.075455	-8.7540	0.023046	0.048474	-8.9072	0.03381	0.083931	-9.1655
		NMM	0.032429	0.074748	-8.7545	0.023071	0.048527	-8.9072	0.033482	0.083067	-9.1659
		EM	0.010175	0.17236	-8.7643	0.02557	0.053692	-8.8991	0.01857	0.23779	-9.1677
	Gaussian	MLE	0.057863	55.949	-8.6008	0.073725	Inf	-8.9463	0.076359	78.88	-8.7207

Tab. 5.7: Goodness of fit for AGL, MTN and NPN log returns based on a 12 month horizon window for the NIG.

Sampling Frequency	Class	Method	AGL			MTN			NPN		
			KS	AD	BIC_0	KS	AD	BIC_0	KS	AD	BIC_0
Daily	VG	MLE	0.036583	0.095795	-5.3385	0.031408	0.098993	-5.4146	0.025687	0.072129	-5.6864
		NMM	0.033819	0.089113	-5.3392	0.031114	0.098166	-5.4146	0.025109	0.079405	-5.6864
		EM	0.045629	0.096328	-5.3385	0.027347	0.093548	-5.4153	0.033437	0.070452	-5.6823
	Gaussian	MLE	0.039114	0.47189	-5.3756	0.045447	0.58277	-5.7059	0.04166	0.35057	-5.4515
4 hours	VG	MLE	0.020979	0.059164	-6.1340	0.046004	0.095368	-6.3038	0.046118	0.092345	-6.5339
		NMM	0.020894	0.058938	-6.1340	0.046003	0.095366	-6.3038	0.04612	0.092347	-6.5339
		EM	0.02293	0.084394	-6.1302	0.034163	0.072212	-6.3149	0.039473	0.079096	-6.5335
	Gaussian	MLE	0.035367	0.99777	-6.1352	0.053927	2.018	-6.503	0.054533	1.8012	-6.3092
2 hours	VG	MLE	0.027496	0.056222	-6.9954	0.033948	0.069199	-7.1526	0.044909	0.095912	-7.3498
		NMM	0.027487	0.056204	-6.9955	0.033946	0.069194	-7.1525	0.044835	0.095557	-7.3496
		EM	0.013966	0.047244	-6.9961	0.022275	0.04466	-7.1573	0.034627	0.069267	-7.3409
	Gaussian	MLE	0.063024	9.0436	-6.8803	0.091496	16.417	-7.1482	0.064109	8.2839	-7.0500
1 hour	VG	MLE	0.02564	0.051293	-7.5981	0.02791	0.056067	-7.8075	0.023768	0.049038	-8.0236
		NMM	0.025715	0.051444	-7.5991	0.027936	0.056119	-7.8079	0.023762	0.049027	-8.0237
		EM	0.017293	0.15537	-7.6299	0.01606	0.038026	-7.8236	0.017256	0.055744	-8.0262
	Gaussian	MLE	0.054081	14.024	-7.5239	0.072389	23.723	-7.861	0.064284	Inf	-7.6828
30 minutes	VG	MLE	0.028937	0.06019	-8.1966	0.025551	0.053865	-8.5182	0.014091	0.0304	-8.7089
		NMM	0.028819	0.060276	-8.2006	0.025551	0.053864	-8.5182	0.014026	0.029677	-8.7108
		EM	0.015246	0.38612	-8.3287	0.014155	0.041158	-8.519	0.024455	0.083502	-8.7868
	Gaussian	MLE	0.061693	38.042	-8.1757	0.07733	Inf	-8.5699	0.072343	45.456	-8.3576
20 minutes	VG	MLE	0.032723	0.075453	-8.6015	0.023061	0.048507	-8.9022	0.0338	0.083905	-8.9248
		NMM	0.032729	0.075475	-8.6040	0.023061	0.048505	-8.9022	0.033784	0.083862	-8.9275
		EM	0.075709	0.1889	-8.7591	0.017686	0.067159	-8.9112	0.017821	0.62218	-9.1719
	Gaussian	MLE	0.057863	55.949	-8.6008	0.073725	Inf	-8.9463	0.076359	78.88	-8.7207

Tab. 5.8: Goodness of fit for AGL, MTN and NPN log returns based on a 12 month horizon window for the VG.

Chapter 6

Value at Risk and Expected Shortfall

The underestimation of risk can result in the collapse of a financial institution as seen in the global financial crisis whilst the overestimation of risk would jeopardize potential investment opportunities. The motivation for using the GHD models to measure risk was given in chapter 1. The representation of the GHD by McNeil *et al.* (2005) was used to measure risk. There were no significant differences in the performance of the VG and NIG from the representation by McNeil *et al.* (2005) and the time subordinated Brownian motion representation, see section 5.8. Given the insignificant differences in the performances of the GHD and due to time constraints only the representation by McNeil *et al.* (2005) was used to measure risk.

Section 6.1 discusses the concept of *Value at risk* (VaR) and some of the methods used to backtest VaR. The results presented in section 6.1 shows that the Gaussian VaR given is more likely to be exceeded by losses in comparison to VaR from the GHD models. This evidence supports the fact that Gaussian VaR generally underestimates risk. Section 6.2 discusses the concept of *expected shortfall* (ES) and coherent risk measures. An explanation of why ES is considered to be a better risk measure than VaR is also given in section 6.2.

6.1 Value at risk

The concept of VaR was introduced by JP Morgan in their first publication of *RiskMetrics* and the metric was based on the assumption that returns are normally distributed. For a given probability $\alpha \in (0, 1)$, VaR is the smallest

value l such that the probability that a loss L greater than l is not greater than $1-\alpha$. Let F denote the cumulative distribution function of losses L, then mathematically VaR can be expressed as

$$\text{VaR}_\alpha = \inf\{l \in \mathbb{R} \colon \mathbf{P}[L > l] \leq 1-\alpha\} = \inf\{l \in \mathbb{R} \colon F_L(l) \geq \alpha\}. \tag{6.1}$$

If the distribution of the losses is strictly increasing then

$$\text{VaR}_\alpha = F^{-1}(\alpha). \tag{6.2}$$

Under the assumption that returns are normally distributed, VaR is given by

$$\text{VaR}_\alpha = \mu + \sigma\Phi^{-1}(\alpha) \tag{6.3}$$

where $\Phi^{-1}(\alpha)$ is the inverse cumulative density of the standard normal distribution. The VaR_α of the GHD models is obtained by finding q such that $\int_{-\infty}^{q} f(x)dx = \alpha$.

Losses (the negative of the return series) were used to evaluate VaR. The sumarry statistics of the return series are presented in chapter 2. For density estimation, the maximum likelihood estimates were computed using the EM algorithm of the GHD. After the calibration of the density VaR_α was evaluated and the following results were obtained.

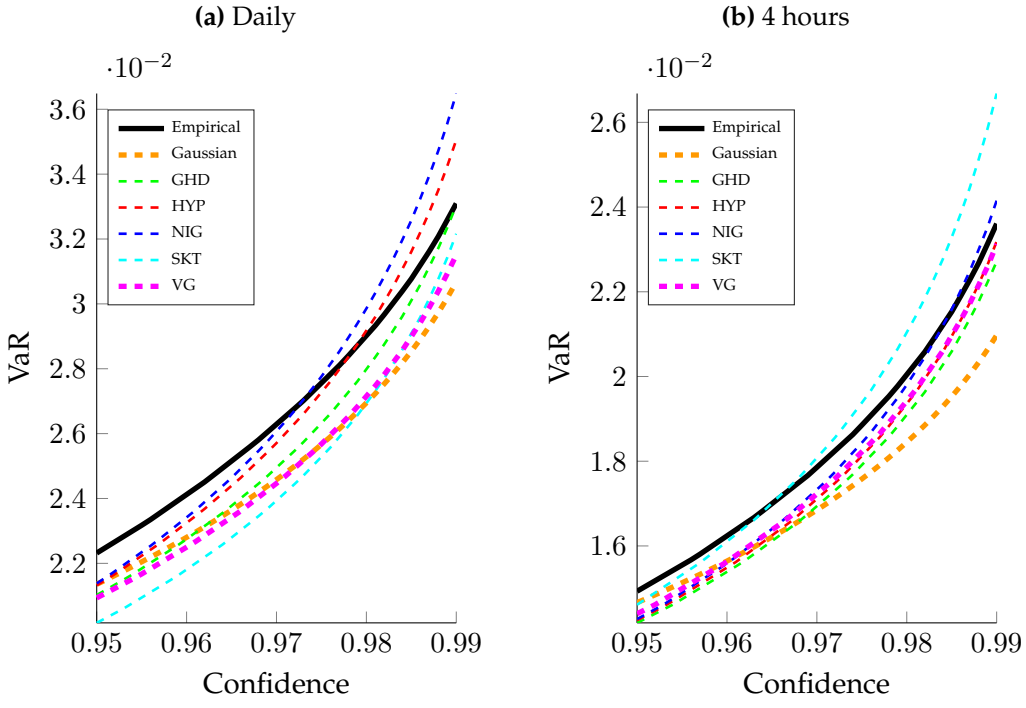


Fig. 6.1: VaR trajectories based upon the GHD and normal distribution.

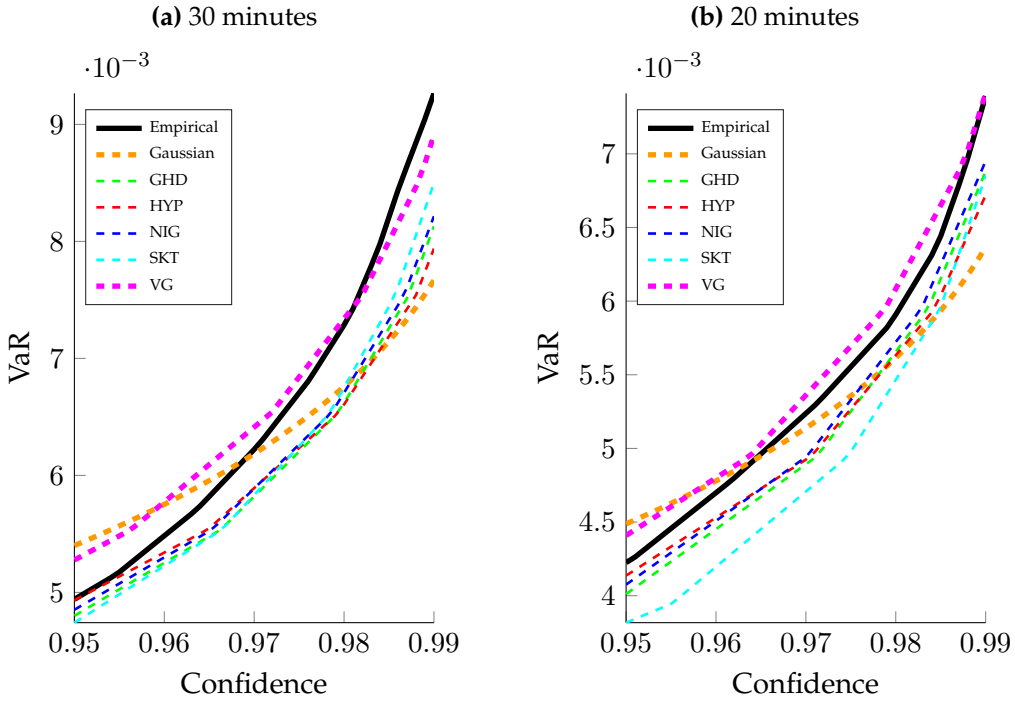


Fig. 6.2: VaR trajectories based upon the GHD and normal distribution.

Notice that the difference between the empirical VaR and the Gaussian VaR becomes more and more distinguished as the sampling time span of the returns decreases. The empirical VaR generally exceeds VaR from most GHD models and Gaussian VaR at high sampling frequencies, this maybe a result of increased in volatility that is caused by intraday trading. Note that intraday VaR can be used by traders who rebalance their portfolios on intraday time scales.

Sampling Frequency	RIC	Sample	Gaussian	GHD	HYP	NIG	SKT	VG
Daily	AGL	0.039653	0.037600	0.038400	0.042259	0.042259	0.039152	0.038405
	MTN	0.039404	0.037352	0.039215	0.043366	0.044793	0.041405	0.039225
	NPN	0.033095	0.030676	0.033043	0.035055	0.036486	0.032161	0.031511
4 hours	AGL	0.029872	0.02593	0.028521	0.030363	0.031503	0.029454	0.027055
	MTN	0.027807	0.024347	0.027238	0.027782	0.028973	0.026282	0.02787
	NPN	0.023608	0.020974	0.022714	0.023189	0.024154	0.026681	0.023124
2 hours	AGL	0.021473	0.017949	0.020354	0.019651	0.020513	0.020876	0.0212
	MTN	0.02006	0.016777	0.019316	0.018559	0.019406	0.018209	0.020289
	NPN	0.019146	0.015398	0.016259	0.01589	0.016415	0.016589	0.01765
1 hour	AGL	0.015403	0.013043	0.014836	0.014371	0.015014	0.015414	0.01452
	MTN	0.014341	0.012191	0.013766	0.013261	0.013848	0.013678	0.014343
	NPN	0.013401	0.01087	0.012017	0.011602	0.012054	0.013392	0.012646
30 minutes	AGL	0.011187	0.0094287	0.01054	0.010156	0.01063	0.011084	0.010867
	MTN	0.0105	0.0086813	0.0097847	0.0094105	0.0098714	0.0098946	0.010228
	NPN	0.0092657	0.007667	0.0081301	0.0079388	0.0082148	0.0084996	0.0089093
20 minutes	AGL	0.009083	0.0076271	0.0085726	0.0082741	0.0086456	0.0089777	0.0088312
	MTN	0.0085085	0.0072314	0.0081402	0.0077704	0.0081784	0.0080428	0.0085159
	NPN	0.0073924	0.006367	0.0068813	0.0067192	0.0069513	0.0068497	0.007411

Tab. 6.1: $VaR_{0.99}$ of AGL, MTN and NPN from 01 Oct 2012 to 30 September 2013

6.1.1 Back-testing VaR

Since financial institutions are allowed by regulatory bodies to use in-house risk models to assess capital adequacy, it is therefore important to have a rigorous approach to backtesting. By backtesting we mean the assessment of the performance of a risk model in comparison to the actual realized returns.

Daily VaR is usually used in capital allocation, hence this section focuses on backtesting VaR computed using daily returns for the GHD and normal distribution. To test daily performance of the risk model, The Basel Committee on Banking Supervision recommends using the most recent 250 daily observations, however because of data restrictions only 187 observations were used to make comparisons.

If the VaR is not able to cover the daily return then we will say a violation occurred. Denote the event that the loss at time t exceeds the $VaR_\alpha(t)$ by the hit function

$$I_\alpha(t) = \begin{cases} 0 & \text{if} \quad x_{t,t+1} \geq -\text{VaR}_\alpha(t) \\ 1 & \text{if} \quad x_{t,t+1} < -\text{VaR}_\alpha(t) \end{cases} \tag{6.4}$$

Let λ_t be the probability of having a violation at time t, then I_t is a Bernoulli random variable that is

$$I_t \sim B(\lambda_t). \tag{6.5}$$

Kupiec (1995) proposed a *proportion of failures* (POF) test statistic to assess the performance of VaR. For a given *significance level*[1] the sequence $\{I_\alpha(t)\}_{t=1}^{t=T}$ is identical to $\{I_\beta(t)\}_{t=1}^{t=T}$. The POF test is based on the fact that for a given *significance level* β, the ratio of violations in the sequence $\{I_\beta(t)\}_{t=1}^{t=T}$ must be exactly equal to the probability of $I_\beta(t) = 1$ if the VaR model is accurate. Here

$$\hat{\beta} = \mathbb{P}[I_\beta(t) = 1] = \frac{\sum_{t=1}^{T} I_\beta(t)}{T}. \tag{6.6}$$

The accuracy of the VaR model becomes questionable when the number of violations differ significantly from $\beta \times 100\%$. The POF suggested by Kupiec

[1] $\beta = 1 - \alpha$

(1995) is defined as

$$\text{POF} = 2\log\left(\left(\frac{1-\hat{\beta}}{1-\beta}\right)^{T-\sum_{t=1}^{T} I_\alpha(t)} \cdot \left(\frac{\hat{\beta}}{\beta}\right)^{\sum_{t=1}^{T} I_\alpha(t)}\right) \tag{6.7}$$

Note that the POF test statistic will be zero when the proportion of VaR violations $\hat{\beta} \times 100\%$ is exactly equal to $\beta \times 100\%$. The bigger the proportion of VaR violations to $\hat{\beta} \times 100\%$ the bigger the POF test statistic, indicating the inaccuracy of the VaR model.

According to Christoffersen (1998), the problem of determining the performance of a VaR model can be reduced to testing whether the hit sequence $\{I_\alpha(t)\}_{t=1}^{t=T}$ satisfies two properties, namely the independence property and the unconditional coverage property. The drawback of Kupiec (1995) POF test is that the POF only tests for the unconditional coverage property.

The unconditional coverage property places a restriction on the number of VaR violations that are allowed to occur (i.e. $\mathbb{P}[I_\beta(t) = 1] = \beta$). If the number of cases where the actual losses exceed the reported VaR are less than $\beta \times 100\%$ then the risk model overestimates the risk of the portfolio. Similarly more cases where the VaR is violated would signal that the model underestimates the risk of the portfolio.

If the VaR model is accurate then the violations should be independent of each other. The independence property therefore places a restriction on how the VaR violations should occur. Specifically two elements in the set $\{I_\beta(t)\}_{t=1}^{t=T}$ must be independent of each other. If VaR violations are clustered in time then there might be a possibility that the capital reserve may not be enough to cover the losses in that period, hence the model will need to be revised.

Christoffersen (1998) proposed a likelihood ratio framework that can be used to test for both independence and unconditional coverage of a hit sequence simultaneously. Christoffersen (1998) test specifies the likelihood ratio (LR) for independence and the LR for unconditional coverage. To completely test for both independence and condition coverage Christoffersen (1998) combined the likelihood ratios of independence and unconditional coverage. The test for independence and unconditional coverage is referred to as the conditional coverage test.

6.1.2 The likelihood ratio test for unconditional coverage

This test assumes that the sequence $\{I_\beta(t)\}_{t=1}^{t=T}$ is independent, the meaning of independent is defined in the following subsection. The null and alternative hypothesis for the unconditional coverage property are as follows

- $H_0 : \mathbb{E}[I_\beta(t)] = \beta$.
- $H_1 : \mathbb{E}[I_\beta(t)] \neq \beta$.

Let n_0 denote the number of elements in $\{I_\beta(t)\}_{t=1}^{t=T}$ whose value is zero and similarly n_1 the number whose value is 1. Under the null hypothesis the likelihood is

$$L(\beta; I_\beta(1), \ldots, I_\beta(T)) = (1-\beta)^{n_0}\beta^{n_1}, \tag{6.8}$$

and under the alternative hypothesis

$$L(\pi; I_\beta(1), \ldots, I_\beta(T)) = (1-\pi)^{n_0}\pi^{n_1}. \tag{6.9}$$

Christoffersen (1998) formulated the unconditional coverage (UC) test as a standard likelihood ratio test,

$$\text{LR}_{uc} = -2\log\left(\frac{L(\beta; I_\beta(1), \ldots, I_\beta(T))}{L(\hat{\pi}; I_\beta(1), \ldots, I_\beta(T))}\right) \sim \chi^2(1), \tag{6.10}$$

where $\hat{\pi} = \frac{n_1}{n_0+n_1}$ is the maximum likelihood estimate of π.

6.1.3 The likelihood ratio test for independence

If a VaR model is accurate then the risk of violating tomorrow's VaR should not depend on whether VaR was violated today. Given a hit sequence $\{I_\beta(t)\}_{t=1}^{t=T}$, first create a 2×2 contingency table that records the violations of VaR on adjacent days

	$I_\beta(t-1)=0$	$I_\beta(t-1)=1$	
$I_\beta(t)=0$	n_{00}	n_{10}	$n_{00}+n_{10}$
$I_\beta(t)=1$	n_{01}	n_{11}	$n_{01}+n_{11}$
	$n_{00}+n_{01}$	$n_{10}+n_{11}$	$n_{00}+n_{01}+n_{10}+n_{11}=T$

If the VaR model is accurate then the proportion of VaR violations given that the previous day was a violation should be equal to the proportion

of VaR violations given that VaR was not violated the previous day, i.e. $\frac{n_{00}}{n_{00}+n_{01}} = \frac{n_{10}}{n_{10}+n_{11}}$. Denote the transition probability matrix of the first order Markov chain $I_\beta(t)$ by

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

where $\pi_{ij} = \mathbb{P}[I_\beta(t) = j | I_\beta(t) = i]$, the likelihood function of this process is approximated by

$$L(\Pi_1; I_\beta(1), \ldots, I_\beta(T)) = (1-\pi_{01})^{n_{00}} \pi_{01}^{n_{01}} \times (1-\pi_{11})^{n_{10}} \pi_{11}^{n_{11}}. \tag{6.11}$$

Note that independence corresponds to

$$\Pi_2 = \begin{bmatrix} 1-\pi_2 & \pi_2 \\ 1-\pi_2 & \pi_2, \end{bmatrix}$$

hence the likelihood under the null hypothesis becomes

$$L(\Pi_2; I_\beta(1), \ldots, I_\beta(T)) = (1-\pi_{01})^{n_{00}+n_{10}} \pi_{01}^{n_{01}+n_{11}}. \tag{6.12}$$

The maximum likelihood estimate is

$$\hat{\Pi}_2 = \hat{\pi}_2 = \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}}. \tag{6.13}$$

Christoffersen (1998) formulated the likelihood ratio test for independence as

$$LR_{\text{ind}} = -2\log\left(\frac{L(\hat{\Pi}_2; I_\beta(1), \ldots, I_\beta(T))}{L(\hat{\Pi}_1; I_\beta(1), \ldots, I_\beta(T))}\right) \sim \chi^2(1), \tag{6.14}$$

where

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{n_{00}}{n_{00}+n_{01}} & \frac{n_{01}}{n_{00}+n_{01}} \\ \frac{n_{10}}{n_{10}+n_{11}} & \frac{n_{11}}{n_{10}+n_{11}} \end{bmatrix}.$$

6.1.4 The test for conditional coverage

Christoffersen (1998) combined the test for independence with the the test for unconditional coverage to form a test for conditional coverage. In the conditional coverage test the null hypothesis of the unconditional coverage test is against the alternative of the independence test, this resulted in the following test

$$L_{\text{cc}} = -2\log\left(\frac{L(\beta; I_\beta(1), \ldots, I_\beta(T))}{L(\hat{\Pi}_1; I_\beta(1), \ldots, I_\beta(T))}\right) \sim \chi^2(2). \tag{6.15}$$

The numerically relationship between the above three tests is given by

$$\mathrm{LR}_{cc} = \mathrm{LR}_{uc} + \mathrm{LR}_{ind}. \tag{6.16}$$

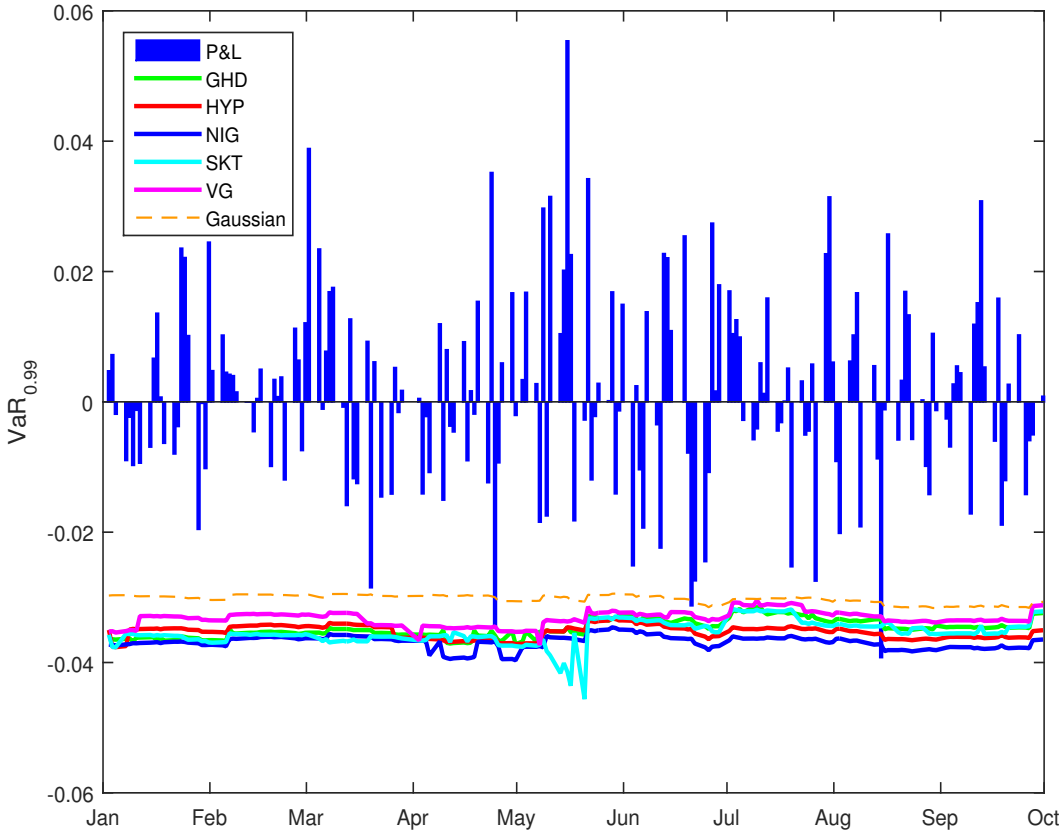


Fig. 6.3: Comparison of the daily returns of NPN with VaR computed by rolling 12 month windows.

The windows of data used were rolled on a daily basis. The first window used to compute VaR in Figure 6.3 started on 3 January 2012 and ended 28 December 2012 and the last period started on 28 September 2012 and ended 27 September 2013. The negative of the VaR calculated using data from the first window was compared with the return on the following day in the data, i.e. 2 January 2013. Similar calculations for all the 187 periods were computed and the results are presented in figure 6.3.

Notice that in the figure 6.3 the normal distribution gives smaller VaR than the GHD at all times. For the rolling windows considered, the expected number of VaR violations at 99% significance is 1.87. The normal distribution violates VaR three times whereas the GHD violates VaR only once,

note that VaR is violated when the return is strictly smaller than the reported VaR. The dates used in figure 6.3 are the dates of the returns used to compare the reported VaR.

Model	LR_{uc}		LR_{ind}		LR_{cc}	
	statistic	p value	statistic	p value	statistic	p value
GHD	0.4829	0.4871	0.0108	0.9172	0.4937	0.4937
HYP	0.4829	0.4871	0.0108	0.9172	0.4937	0.4937
NIG	0.4829	0.4871	0.0108	0.9172	0.4937	0.4937
SKT	0.4829	0.4871	0.0108	0.9172	0.4937	0.4937
VG	0.4829	0.4871	0.0108	0.9172	0.4937	0.4937
Gaussian	0.5953	0.4404	0.0984	0.7538	0.6937	0.7069

Tab. 6.2: Test for indpendence and coverage for VaR violations of NPN log returns.

6.2 Expected shortfall

Suppose there were 95 days when either a profit or a reasonable loss was made and 5 days of excessive loss. If one calculates VaR at 95% confidence to be 0.5 dollars, then an investor would infer that at a 95% confidence level that the maximum they can lose 0.5 dollars if they hold the portfolio. However, an investor is generally interested in knowing the expected loss of the 5 days when there was massive loss. One of the criticisms of VaR is that it is not conclusive about the expected size of the loss when a certain level is exceeded. VaR simply tells one the number of times the loss is expected to exceed a given level, however, ES tells one the expected average loss if the loss exceeds a certain level.

By writing axioms that a "good financial risk measure" should satisfy, Artzner *et al.* (1999) crystalized using precise statements the intrinsic nature of the concept. Fix a probability space $(\Omega, \mathscr{F}, \mathbb{P})$, consider the set V of random variables $X : V \to \mathbb{R}$, which represents the possible returns from holding a portfolio of assets over a single period of time. For each X we like to define a risk measure $\rho : V \to \mathbb{R}$ that represents the risk of holding this portfolio of assets over a single period of time. Note that there is risk associated with the portfolio if and only if $\rho(X) \geq 0$.

Since risk was measured in this project to show that the GHD models provide better risk estimates for the amount of capital an investor could reserve when holding a single stock over a finite time horizon, in the context of VaR and ES. It therefore makes sense to assume that the expectation of the random variable X is finite. However, it is worth mentioning that Delbaen (2002) extended the definition of coherent risk measure by Artzner *et al.* (1999) to arbitrary or general probability spaces. Delbaen (2002) concluded that there is no coherent risk measures that only take finite values on an arbitrary probability space. Also, Delbaen (2002) mentioned that a risk of $+\infty$ or $-\infty$ is not acceptable in any reasonable model. Hence the work in this project will solely use the definition presented by Artzner *et al.* (1999).

Definition 6.1. Coherent risk measure Artzner *et al.* (1999)
A real valued function ρ is a coherent risk measure if it satisfies the following four properties,

1. **Monotonicity**. For all $X, Y \in V$ with $X \leq Y$, $\rho(X) \leq \rho(Y)$.
2. **Positive homogeneity**. For all $\lambda \geq 0$ and all $X \in V$, $\rho(\lambda X) = \lambda\rho(X)$.
3. **Translation invariance**.For all $X \in V$ and any $a \in \mathbb{R}$, $\rho(a + X) = \rho(X) - a$.
4. **Subadditivity**. For all $X, Y \in V$, $\rho(X + Y) \leq \rho(X) + \rho(Y)$.

The monotonicity axiom entails that if the payoff of a financial instrument X is less than that of Y, then the risk of holding the instrument X should not exceed the risk of holding the instrument Y. The monotonicity axiom also implies that if the portfolio has positive value then there is no risk. Positive homogeneity signifies that a measure has same scalability (dimension) as the random variable X for example double the size of the portfolio would result in the doubling of risk. The translation invariance axiom states that if a guaranteed profit added to a portfolio will reduce the risk by the same amount. Subadditivity encapsulates the idea that portfolio diversification should reduce risk i.e. combining two portfolios cannot increase the total individual risk.

As mentioned previously, VaR has been widely criticized for not being conclusive about the expected size of loss should loss exceed a specific confidence level. Another criticism of VaR is that it violates the subadditivity

property when the loss distribution is not elliptical[2], see McNeil *et al.* (2005).

A common practice in risk management is measuring the risk of subportfolios or even single stocks and then aggregating the risks to get the risk of a firm. For example, if risk manager wants to know the risk of a firm then they can calculate the risk of the bond trading desk and the commodity trading desks and all the other risk taking portfolios and then combine the risks of the individual portfolios the bank has undertaken. Since VaR violates the subadditivity property when the distribution is not elliptical, one cannot simply combine the VaR of individual stocks to get the VaR of a portfolio of stocks.

It is important to note that if a regulator imposes capital requirements based on a risk measure that does not satisfy the subadditivity axiom then it may provide incentive for a firm or bank to split into separate independent firms or banks.

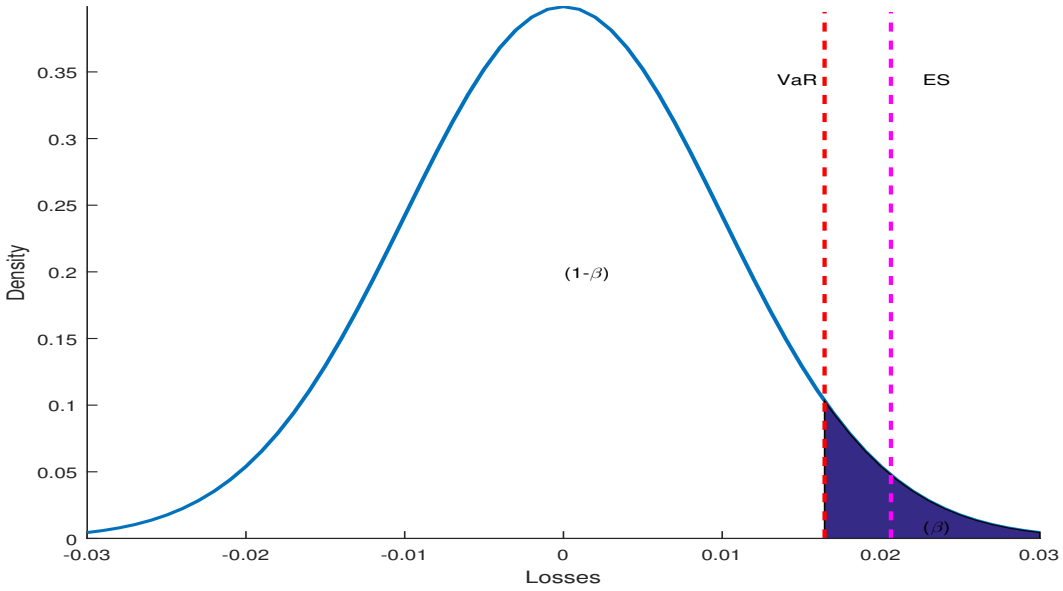


Fig. 6.4: Density of loses with VaR and ES.

Before discussing expected shortfall, it is worthwhile to mention that there

[2] A $(n \times 1)$ random vector X is said to have an elliptical distribution with parameters $\mu(n \times 1)$ and $\Sigma(n \times n)$ if its characteristic function can be expressed as

$$\mathbb{E}[\exp(it^T X)] = \exp(it^T \mu) \cdot \phi(t^T \Sigma t), \tag{6.17}$$

for some scalar function ϕ.

exists a number of alterative coherent risk measures that consider losses beyond VaR. The reader is referred to the following literature for mathematical descriptions and properties for some of the well-known coherent risk measures:

- Worst Conditional Expectation: introduced by Artzner *et al.* (1999), the measure is not law invariant, hence for two portfolios with the same loss distribution using the risk measure can result in different risk values.
- Tail Conditional Expectation: introduced by Artzner *et al.* (1999) the measure violates the subadditivity property for general distributions, the measure is coherent for continuous distributions only.
- Expected Shortfall: introduced by Artzner *et al.* (1999) to overcome some of the theoretical shortfalls of VaR, Acerbi and Tasche (2002) proved that the measure is coherent. Most importantly it was proved by Inui and Kijima (2005) that Expected Shortfall gives the minimum value amongst a class of plausible coherent risk measures.
- Conditional Value at Risk: Also known as Expected Shortfall was introduced by Rockafellar and Uryasev (2002).

Expected shortfall is derived from VaR but is coherent and is also known as the average value at risk and tail conditional and expected tail loss. For a given confidence level α, the expected shortfall is given by

$$\mathrm{ES}_\alpha = \mathbb{E}[L|L \geq \mathrm{VaR}_\alpha] = \frac{1}{1-\alpha}\int_\alpha^1 q_u(F_L(u))\,\mathrm{d}u, \tag{6.18}$$

where F_L is the loss distribution and $q_u(\cdot)$ is the quantile function. The expression of ES in terms of VaR is

$$\mathrm{ES}_\alpha = \frac{1}{1-\alpha}\int_\alpha^1 \mathrm{VaR}_u(L)\,\mathrm{d}u, \tag{6.19}$$

hence the ES can be thought of as the average value at risk in the interval $(\alpha, 1)$.

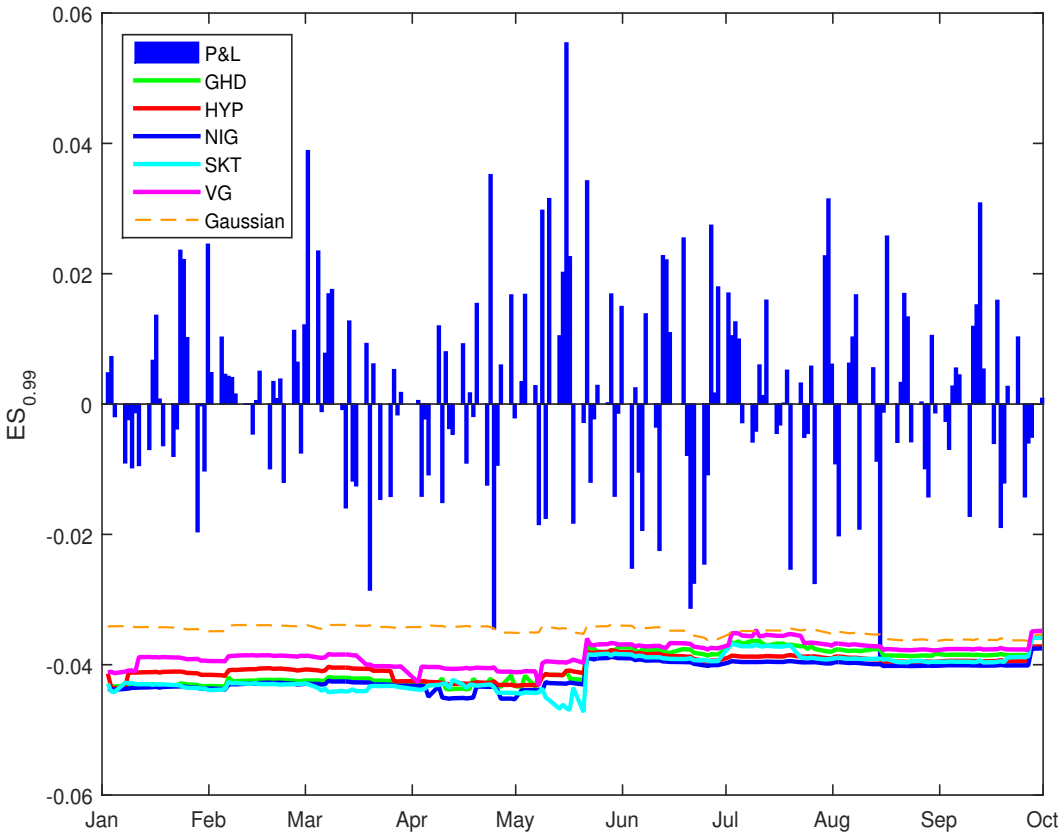


Fig. 6.5: Comparison of the daily returns of NPN with ES computed by rolling 12 month windows.

The 12-month window periods used to compute the ES reported in figure 6.5 is the same data that was used to compute the reported VaR in section 6.1. The Gaussian VaR exceeds the losses three times whereas the corresponding ES exceeds losses twice. The ES is always equal to or slightly more than the VaR hence it is less likely to exceed the losses than VaR.

Sampling Frequency	RIC	Sample	Gaussian	GHD	HYP	NIG	SKT	VG
Daily	AGL	0.043282	0.043067	0.040472	0.041175	0.041175	0.040845	0.040484
	MTN	0.04327	0.042615	0.04221	0.043607	0.043644	0.043423	0.042214
	NPN	0.037169	0.035311	0.036168	0.037193	0.03756	0.035853	0.035024
4 hours	AGL	0.034637	0.029702	0.033503	0.035625	0.036939	0.037001	0.03143
	MTN	0.031153	0.027804	0.030972	0.031464	0.032374	0.031702	0.03145
	NPN	0.028836	0.024113	0.027371	0.028197	0.029517	0.032725	0.027811
2 hours	AGL	0.027908	0.020561	0.025461	0.024015	0.025883	0.027784	0.02614
	MTN	0.024613	0.019177	0.023444	0.022458	0.023654	0.02437	0.024234
	NPN	0.023411	0.017682	0.020434	0.019496	0.020772	0.02218	0.021862
1 hour	AGL	0.022308	0.014942	0.018901	0.01767	0.019009	0.021787	0.017696
	MTN	0.017855	0.013945	0.017124	0.016138	0.017364	0.018319	0.017629
	NPN	0.017135	0.012474	0.015206	0.014197	0.015247	0.018075	0.015619
30 minutes	AGL	0.015815	0.010801	0.013628	0.012774	0.013716	0.016107	0.013652
	MTN	0.013491	0.0099347	0.012307	0.011602	0.012516	0.013398	0.01268
	NPN	0.012387	0.0087942	0.010398	0.0098198	0.010484	0.011842	0.011142
20 minutes	AGL	0.01302	0.0087377	0.011224	0.010405	0.011299	0.013138	0.011216
	MTN	0.011726	0.0082774	0.01038	0.0096375	0.010415	0.012	0.010576
	NPN	0.010375	0.0073014	0.0089286	0.0083794	0.0090225	0.0096318	0.0092988

Tab. 6.3: $ES_{0.99}$ of AGL, MTN and NPN from 01 Oct 2012 to 30 September 2013

Chapter 7

Conclusion and recommendations

7.1 Concluding remarks

The aim of this thesis was to model the behavior of stock market returns using Lévy process in particular generalized hyperbolic Lévy processes. Stock prices exhibit jumps so it seemed more natural to use processes that incorporate jumps to model their behavior. In chapter 4 asset returns from the JSE were fitted to the normal distribution and the normal mean variance mixture representation of the GHD introduced by McNeil *et al.* (2005) GHD. The GHD gave better fit in comparison the normal distribution for intraday data.

Simulations of general Lévy processes requires knowledge of numerous simulation techniques and is quite involved. However, the NIG ad VG are closed under convolution and can be simulated more naturally than general Lévy process. In chapter 5 simulation algorithms of the NIG and VG paths represented as subordinated Brownian motion that is time changed by an increasing Lévy process are presented. The were no significant differences in the performance of the VG and NIG from both representations.

The VaR and ES evaluated from the normal variance-mean representation of the GHD and the normal distribution were presented in chapter 6. The results show that the GHD markedly improves modeling risk even at daily sampling frequencies.

The results presented in this dissertation provides evidence that Lévy processes are able to capture more of the empirical stylized facts of asset re-

turns than the normal distribution for intraday data. Although the research ideas undertaken in this thesis only focused on univariate estimation they can be extended to multivariate setting.

7.2 Recommendations for further research

The simulation of the VG and the NIG in chapter 5 focused on paths with independent increments. Finlay and Seneta (2008) provides a method of simulating the VG and t distributed increments that exhibit long memory. Finlay and Seneta (2008) reported that their models allow returns to be lightly correlated or uncorrelated returns with highly correlated squared returns, a phenomenon that is observed in actual data. The drawback of the simulation method provided by Finlay and Seneta (2008) is that $\frac{2}{\kappa}$ must be an integer.

In-order to construct NIG process with dependence, Kumar and Vellaisamy (2012) subordinated fractional Brownian motion to a NIG process. The first order increments of the process constructed by Kumar and Vellaisamy (2012) are stationary and exhibit long range dependence.

For simplicity when VaR was backtested in chapter 6, the risk management horizon considered was $\Delta = 1$ day, however the Basel Committee recommends a $\Delta = 10$ days for market risk and the use of VaR at 99% confidence level. Denote the 10-day VaR at 99% confidence level calculated on day j by $\text{VaR}_{0.99}^{j,10}$. The risk capital for banks that use internal models (IM) for the market risk (MR) approach is given by

$$\text{RC}_{\text{IM}}^{t}(\text{MR}) = \max\{\text{VaR}_{0.99}^{t,10}, \frac{k}{60}\sum_{i=1}^{60}\text{VaR}_{0.99}^{t-i+1,10}\} + C_{\text{SR}}, \tag{7.1}$$

where $3 \geq k \leq 4$ is a stress factor that is determined as a function of the overall variability of the bank's internal model. A specific risk (SR) component has to be added to all VaR numbers, hence the component C_{SR} in Eqn. 7.1 see Crouchy *et al.* (2000) for examples. The work in this project can be extended to calculation of risk capital.

Bibliography

Abramowitz, M. and Stegun, I. (1964). *Handbook of mathematical functions: with formulas, graphs, and mathematical tables*, Vol. 55, Courier Corporation.

Acerbi, C. and Tasche, D. (2002). On the coherence of expected shortfall, *Journal of Banking & Finance* **26**(7): 1487–1503.

Ané, T. and Geman, H. (2000). Order flow, transaction clock, and normality of asset returns, *The Journal of Finance* **55**(5): 2259–2284.

Artzner, P., Delbaen, F., Eber, J. and Heath, D. (1999). Coherent measures of risk, *Mathematical finance* **9**(3): 203–228.

Bachelier, L. (1900). *Théorie de la spéculation*, PhD thesis, Gauthier-Villars.

Bagnold, R. (1941). *The physics of blown sand and desert dunes*, London, Methuen.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **353**(1674): 401–419.

Barndorff-Nielsen, O. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling, *Scandinavian Journal of statistics* **24**(1): 1–13.

Barndorff-Nielsen, O. (1998). Processes of normal inverse Gaussian type, *Finance and stochastics* **2**(1): 41–68.

Barndorff-Nielsen, O. and Halgreen, C. (1977). Infinite divisibility of the hyperbolic and generalized inverse Gaussian distributions, *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **38**(4): 309–311.

Black, F. (1976). Studies of stock price volatility changes, *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, American Statistical Association, Washington, D.C., pp. 177–181.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities, *The journal of political economy* pp. 637–654.

Blæsild, P. and Sørensen, M. (1992). "Hyp": A computer program for analyzing data by means of the hyperbolic distribution, *Department of Theoretical Statistics, University of Aarhus* .

Bochner, S. (1955). Harmonic analysis and the theory of probability, *Berkeley University of California Press* .

Brent, R. (1973). Algorithms for minimization without derivatives, *Dover Publications, New York* .

Christodoulou-Volos, C. and Siokis, F. (2006). Long range dependence in stock market returns, *Applied Financial Economics* **16**(18): 1331–1338.

Christoffersen, P. (1998). Evaluating interval forecasts, *International economic review* pp. 841–862.

Clark, P. (1973). A subordinated stochastic process model with finite variance for speculative prices, *Econometrica: Journal of the Econometric Society* **41**: 135–155.

Cont, R. and Tankov, P. (2004). *Financial modelling with jump processes*, CRC press.

Crouchy, M., Galai, D. and Mark, R. (2000). A comparative analysis of current credit risk models, *Journal of Banking & Finance* **24**(1): 59–117.

Dacorogna, M., Gencay, R., Muller, U., Olsen, R. and Pictet, O. (2001). *An introduction to high frequency finance*, Academic Press: San Diego, CA.

Delbaen, F. (2002). Coherent risk measures on general probability spaces, *Advances in Finance and Stochastics: Essays in Honour of Dieter Sondermann*, Springer Berlin Heidelberg, pp. 1–37.

Dempster, A., Laird, N. and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm, *Journal of the royal statistical society. Series B (methodological)* **39**(1): 1–38.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance, *Bernoulli* pp. 281–299.

Eberlein, E. and von Hammerstein, E. (2004). Generalized hyperbolic and inverse Gaussian distributions: limiting cases and approximation of processes, *Seminar on stochastic analysis, random fields and applications IV*, Vol. 58, Birkhäuser Verlag, Boston, pp. 221–264.

Fama, E. (1965). The behavior of stock-market prices, *The journal of Business* **38**(1): 34–105.

Figueroa-López, J., Lancette, S., Lee, K. and Mi, Y. (2011). Estimation of NIG and VG models for high frequency financial data, *Handbook of Modeling High-Frequency Data in Finance* .

Finlay, R. and Seneta, E. (2008). Stationary-increment variance-gamma and t model: Simulation and parameter estmation, **76**(2): 167–186.

Geman, H. (2002). Pure jump Lévy processes for asset price modelling, *Journal of Banking & Finance* **26**(7): 1297–1316.

Hansen, B. (1994). Autoregressive conditional density estimation, *International Economic Review* **35**(3).

Hansen, L. (1982). Large sample properties of generalized method of moments estimators, *Econometrica: Journal of the Econometric Society* pp. 1029–1054.

Heyde, C. and Liu, S. (2001). Empirical realities for a minimal description risky asset model. the need for fractal features, *Journal of Korean Mathematical Society* **38**(5): 1047–1059.

Hu, W. (2005). *Calibration of multivariate generalized hyperbolic distributions using the EM algorithm, with applications in risk management, portfolio optimization and portfolio credit risk*, PhD thesis, The Florida state university.

Hyndman, R. and Koehler, A. (2006). Another look at measures of forecast accuracy, *International Journal of Forecasting* **22**(4): 679–688.

Inui, K. and Kijima, M. (2005). On the significance of expected shortfall as a coherent risk measure, *Journal of Banking & Finance* **29**(4): 853–864.

Jones, B., Waller, W. and Feldman, A. (1978). Root isolation using function values, *BIT Numerical Mathematics* **18**(3): 311–319.

Jørgensen, B. (1982). *Statistical properties of the generalized inverse Gaussian distribution*, Vol. 9, Springer - Verlag New York Inc.

Khintchine, A. (1937). A new derivation of one formula by Levy P, *Bull. Moscow State Univ., I* (1): 1–5.

Kolmogorov, A. (1932). Sulla forma generale di un processo stocastico omogeneo, *Atti Accad. Naz. Lincei* **15**: 805–808.

Konlack Socgnia, V. and Wilcox, D. (2014). A comparison of generalized hyperbolic distribution models for equity returns, *Journal of Applied Mathematics* **2014**.

Krichene, N. (2005). Subordinated Lévy processes and applications to crude oil options, *IMF Working Paper* .

Kumar, A. and Vellaisamy, P. (2012). Fractional normal inverse Gaussian process, *Methodology and Computing in Applied Probability* **14**(2): 263–283.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models, *The journal of Derivatives* **3**(2): 73–84.

Kwiatkowski, D., Phillips, P., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of econometrics* **54**(1-3): 159–178.

Kyprianou, A. (2007). An introduction to the theory of Lévy processes. Satellite Summer school on Levy Processes : Theory and Application, Sønderborg Denmark.

Laio, F., Di Baldassarre, G. and Montanari, A. (2009). Model selection techniques for the frequency analysis of hydrological extremes, *Water Resources Research* **45**(7).

Lévy, P. (1934). Sur les intégrales dont les éléments sont des variables aléatoires indépendantes, *Annali della Scuola Normale Superiore di Pisa-Classe di Scienze* **3**(3-4): 337–366.

Ling, X. (2016). *Changing the way we measure time to more accurately estimate the probability of informed trading*, PhD thesis, UQ Business School, The University of Queensland.

Liu, C. and Rubin, D. (1995). ML estimation of the t distribution using EM and its extensions, ECM and ECME, *Statistica Sinica* **5**: 19–39.

Lo, A. and MacKinlay, A. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test, *Review of financial studies* **1**(1): 41–66.

Luo, R. and Zhao, W. (2014). A robust iterative algorithm for parameter estimation of the generalized gamma distribution. Proceedings: 2014 IEEE 17th International Conference on Computational Science and Engineering.

Madan, D., Carr, P. and Chang, E. (1998). The variance gamma (V.G.) model for share market returns, *European finance review* **2**(1): 79–105.

Madan, D. and Seneta, E. (1990). The variance gamma (V.G.) model for share market returns, *The Journal of business* **63**(4): 511–524.

Mandelbrot, B. (1963). The variation of certain speculative prices, *The Journal of Business* **36**(4): 394–419.

Mandelbrot, B. and Van Ness, J. (1968). Fractional Brownian motions, fractional noises and applications, *SIAM review* **10**(4): 422–437.

McNeil, A., Frey, R. and Embrechts, P. (2005). *Quantitative risk management:Concepts, techniques and tools*, Vol. 10, Princeton university press.

Newey, W. and West, K. (1986). A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix, *National Bureau of Economic Research Cambridge, Mass., USA* .

Paolella, M. (2007). *Intermediate probability: A computational approach*, John Wiley & Sons.

Prause, K. (1999). *The generalized hyperbolic model:Estimation, financial derivatives, and risk measures*, PhD thesis, University of Freiburg.

Protassov, R. (2003). EM-based maximum likelihood parameter estimation for multivariate generalized hyperbolic distributions with fixed λ, *Statistics and Computing* **14**(1): 67–77.

Qi, F. (2007). Three classes of logarithmically completely monotonic functions involving gamma and psi functions, *Integral Transforms and Special Functions* **18**(7): 503–509.

Qi, F. and Guo, B. (2010). Two new proofs of the complete monotonicity of a function involving the psi function, *Bull. Korean Math. Soc.* **47**(1): 103–111.

Rockafellar, R. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions, *Journal of Banking & Finance* **26**(7): 1443–1471.

Rydberg, T. (1997). The normal inverse Gaussian Lévy process: simulation and approximation, *Communications in statistics. Stochastic models* **13**(4): 887–910.

Said, S. and Dickey, D. (1984). Testing for unit roots in autoregressive-moving average models of unknown order, *Biometrika* **71**(3): 599–607.

Samuelson, P. (1965). Rational theory of warrant pricing, *IMR; Industrial Management Review (pre-1986)* **6**(2).

Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press.

Sichel, H. (1974). On a distribution representing sentence-length in written prose, *Journal of the Royal Statistical Society. Series A (General)* pp. 25–34.

Stacy, E. (1962). A generalization of the gamma distribution, *The Annals of mathematical statistics* pp. 1187–1192.

Stacy, E. and Mihram, G. (1965). Parameter estimation for a generalized gamma distribution, *Technometrics* **7**(3): 349–358.

Van Haeringen, H. (1993). Completely monotonic and related functions, Report 93-108, Faculty of Technical Mathematics and Informatics, *Delft University of Technology, Delft, The Netherlands* .

Wilcox, D. and Gebbie, T. (2008). Serial correlation, periodicity and scaling of eigenmodes in an emerging market, *International Journal of Theoretical and Applied Finance* **11**(7): 739–760.

Wingo, D. (1987). Computing maximum-likelihood parameter estimates of the generalized gamma distribution by numerical root isolation, *IEEE Transactions on Reliability* **36**(5): 586–590.

Appendix A

Summary statistics of intraday data from the JSE

Interval	Sample size	Min	Max	Mean	Variance	Skewness	Kurtosis
8 hours	249	-0.0456	0.0602	-0.00122	0.0155	0.0822	3.5521
4 hours	498	-0.0348	0.0323	-0.00061	0.0102	-0.0029	3.6128
2 hours	996	-0.0306	0.0346	-0.00030	0.0071	-0.0448	5.2206
1 hour	1992	-0.0260	0.0436	-0.00015	0.0052	0.1996	7.2843
30 minutes	3984	-0.0229	0.0285	-0.00008	0.0037	-0.0186	6.6970
20 minutes	5976	-0.0205	0.0203	-0.00005	0.0031	-0.046	7.4641
10 minutes	11952	-0.0189	0.0218	-0.00003	0.0023	-0.1061	9.1656
5 minutes	23904	-0.0266	0.0242	-0.00001	0.0017	-0.2403	13.439
2 minutes	59760	-1.0057	1.0051	-0.0000051	0.0059	-0.1581	27710

Tab. A.2: Summary statistics for MTN from 01 October 2012 to 30 September 2013.

Interval	Sample size	Min	Max	Mean	Variance	Skewness	Kurtosis
8 hours	249	-0.0650	0.0574	-0.0022	0.0170	-0.0333	3.8479
4 hours	498	-0.0491	0.0411	-0.0011	0.0123	-0.2412	4.4951
2 hours	996	-0.0503	0.0350	-0.0005	0.0088	-0.7104	7.7699
1 hour	1992	-0.0390	0.0223	-0.0002736	0.0059	-0.6051	7.1707
30 minutes	3984	-0.0402	0.0209	-0.0001368	0.0042	-0.6928	9.2943
20 minutes	5976	-0.4108	0.4104	-0.0000912	0.0083	-0.0711	2026
10 minutes	11952	-0.4105	0.4086	-0.0000456	0.0059	-0.4166	3925
5 minutes	23904	-0.4080	0.4083	-0.0000228	0.0042	0.0701	7410
2 minutes	59760	-0.4077	0.4072	-0.0000091	0.0027	-0.2559	17151

Tab. A.3: Summary statistics for SHP from 01 October 2012 to 30 September 2013.

Interval	**Sample size**	**Min**	**Max**	**Mean**	**Variance**	**Skewness**	**Kurtosis**
8 hours	249	-0.0392	0.0554	0.00114218	0.0137	0.2215	3.9206
4 hours	498	-0.0377	0.0444	0.0005711	0.0093	0.0845	4.7147
2 hours	996	-0.0346	0.0332	0.0002855	0.00674	0.2036	6.3062
1 hour	1992	-0.0243	0.0278	0.0001428	0.0047	0.1037	6.8059
30 minutes	3984	-0.0226	0.0294	0.0000714	0.0033	0.3043	9.8136
20 minutes	5976	-0.0212	0.0291	0.0000476	0.0028	0.3022	12.9031
10 minutes	11952	-0.0498	0.0477	0.0000238	0.0022	-0.2006	54.6148
5 minutes	23904	-0.0508	0.0470	0.0000119	0.0016	-0.3284	80.2991
2 minutes	59760	-0.0509	0.0239	0.0000048	0.0011	-1.3681	83.6099

Tab. A.4: Summary statistics for NPN from 01 October 2012 to 30 September 2013.

Interval	ADF test		KPSS test		Stationary
	Statistic	Decision	Statistic	Decision	Increments ?
2 minutes	-377.8247	Reject H_0	0.0095	Fail to reject H_0	Yes
5 minutes	-241.5188	Reject H_0	0.0099	Fail to reject H_0	Yes
10 minutes	-171.0369	Reject H_0	0.0101	Fail to reject H_0	Yes
20 minutes	-120.1070	Reject H_0	0.0102	Fail to reject H_0	Yes
30 minutes	-64.1155	Reject H_0	0.0593	Fail to reject H_0	Yes
1 hour	-42.82362	Reject H_0	0.0595	Fail to reject H_0	Yes
2 hours	-32.0151	Reject H_0	0.0540	Fail to reject H_0	Yes
4 hours	-22.0313	Reject H_0	0.0555	Fail to reject H_0	Yes
8 hours	-13.8061	Reject H_0	0.0585	Fail to reject H_0	Yes

Tab. A.5: ADF and KPSS test for SHP log retruns from 01 October 2012 to 30 September 2013.

Interval	ADF test		KPSS test		Stationary
	Statistic	Decision	Statistic	Decision	Increments ?
2 minutes	-290.1195	Reject H_0	0.0287	Fail to reject H_0	Yes
5 minutes	-184.5635	Reject H_0	0.0345	Fail to reject H_0	Yes
10 minutes	-124.9706	Reject H_0	0.0393	Fail to reject H_0	Yes
20 minutes	-78.2907	Reject H_0	0.0488	Fail to reject H_0	Yes
30 minutes	-61.6966	Reject H_0	0.0503	Fail to reject H_0	Yes
1 hour	-43.1026	Reject H_0	0.0497	Fail to reject H_0	Yes
2 hours	-32.0779	Reject H_0	0.0492	Fail to reject H_0	Yes
4 hours	-20.7790	Reject H_0	0.0520	Fail to reject H_0	Yes
8 hours	-14.9254	Reject H_0	0.0480	Fail to reject H_0	Yes

Tab. A.6: ADF and KPSS test for NPN log retruns from 01 October 2012 to 30 September 2013.

Sampling Frequency	q	2	4	8	16	32	64
Daily	VR(q)	0.9340	0.8566	0.8443	0.9234	0.8303	1.1206
	z(q)	-0.9749	-1.1236	-0.7615	-0.2512	-0.3782	0.1830
	p value	0.3296	0.2612	0.4464	0.8016	0.7053	0.8548
	Decision	NR	NR	NR	NR	NR	NR
4 hours	VR(q)	1.0508	1.0319	0.9761	0.9686	1.0649	0.9484
	z(q)	1.0617	0.3586	-0.1693	-0.1501	0.2119	-0.1151
	p value	0.2884	0.7199	0.8656	0.8807	0.8322	0.9084
	Decision	NR	NR	NR	NR	NR	NR
2 hours	VR(q)	1.0252	1.0754	1.0569	1.0034	1.0039	1.1031
	z(q)	0.7608	1.1826	0.5594	0.0224	0.0176	0.3230
	p value	0.4468	0.2370	0.5759	0.9822	0.9860	0.7467
	Decision	NR	NR	NR	NR	NR	NR
1 hour	VR(q)	1.0126	0.9951	1.0015	0.9870	0.9383	0.9356
	z(q)	0.4152	-0.0919	0.0195	-0.1135	-0.3755	-0.2773
	p value	0.6780	0.9268	0.9844	0.9097	0.7073	0.7816
	Decision	NR	NR	NR	NR	NR	NR
30 minutes	VR(q)	0.9762	0.9911	0.9811	0.9754	0.9632	0.9184
	z(q)	-1.1087	-0.2308	-0.3271	-0.3042	-0.3221	-0.5053
	p value	0.2676	0.8175	0.7436	0.7610	0.7474	0.6134
	Decision	NR	NR	NR	NR	NR	NR
20 minutes	VR(q)	0.9611	0.9489	0.9590	0.9278	0.9420	0.8915
	z(q)	-2.0433	-1.5097	-0.8141	-1.0402	-0.6098	-0.8162
	p value	0.0410	0.1311	0.4156	0.2982	0.5420	0.4144
	Decision	R	NR	NR	NR	NR	NR
10 minutes	VR(q)	0.9177	0.8852	0.8706	0.8796	0.8522	0.8641
	z(q)	-5.8025	-4.5723	-3.4319	-2.2755	-2.0670	-1.4043
	p value	$\ll 0$	$\ll 0$	0.0006	0.0229	0.0387	0.1602
	Decision	R	R	R	R	R	NR
5 minutes	VR(q)	0.8929	0.8162	0.7862	0.7740	0.7835	0.7605
	z(q)	-7.8619	-8.1021	-6.8162	-5.4147	-3.9331	-3.3144
	p value	$\ll 0$	$\ll 0$	$\ll 0$	0.0001	0.0839	0.9185
	Decision	R	R	R	R	R	R
2 minutes	VR(q)	0.5150	0.2718	0.1474	0.0859	0.0555	0.0406
	z(q)	-1.0072	-1.0081	-1.0118	-1.0124	-1.0122	-1.0118
	p value	0.3138	0.3134	0.3116	0.3113	0.3114	0.3116
	Decision	NR	NR	NR	NR	NR	NR

The decision tells whether the random walk hypothesis is rejected(R) or failed to reject (NR) at 5% significance .

Tab. A.7: The results of variance ratio test for MTN log returns from 01 October 2012 to 30 September 2013.

Sampling Frequency	q	2	4	8	16	32	64
Daily	VR(q)	1.0478	1.0865	1.0445	1.0276	0.9020	0.7146
	z(q)	0.6278	0.6132	0.1995	0.0861	-0.2116	-0.4069
	p value	0.5301	0.5397	0.8419	0.9314	0.8324	0.6841
	Decision	NR	NR	NR	NR	NR	NR
4 hours	VR(q)	1.0685	1.1387	1.1894	1.1396	1.1200	0.9808
	z(q)	1.4680	1.4760	1.2863	0.6443	0.3840	-0.0425
	p value	0.1421	0.1400	0.1983	0.5194	0.7010	0.9661
	Decision	NR	NR	NR	NR	NR	NR
2 hours	VR(q)	0.9827	1.0254	1.0867	1.1329	1.0873	1.0625
	z(q)	-0.5362	0.4068	0.8378	0.8727	0.4002	0.1998
	p value	0.5918	0.6842	0.4021	0.3828	0.6890	0.8417
	Decision	NR	NR	NR	NR	NR	NR
1 hour	VR(q)	1.0335	1.0308	1.0074	1.0747	1.1251	1.0894
	z(q)	1.2064	0.5915	0.0948	0.6598	0.7836	0.4012
	p value	0.2277	0.5542	0.9245	0.5094	0.4333	0.6883
	Decision	NR	NR	NR	NR	NR	NR
30 minutes	VR(q)	1.0224	1.0308	1.0282	1.0060	1.0788	1.1345
	z(q)	1.1310	0.8378	0.4875	0.0726	0.6778	0.8264
	p value	0.2581	0.4021	0.6259	0.9421	0.4979	0.4086
	Decision	NR	NR	NR	NR	NR	NR
20 minutes	VR(q)	0.9870	0.9977	1.0099	0.9843	1.0290	1.0679
	z(q)	-0.6875	-0.0701	0.1975	-0.2215	0.2989	0.5075
	p value	0.4917	0.9441	0.8434	0.8247	0.7650	0.6118
	Decision	NR	NR	NR	NR	NR	NR
10 minutes	VR(q)	0.8667	0.8236	0.8131	0.8193	0.7950	0.8270
	z(q)	-2.9562	-2.5644	-2.2501	-1.8954	-1.8767	-1.3574
	p value	0.0031	0.0103	0.0244	0.0580	0.0606	0.1746
	Decision	R	R	R	NR	NR	NR
5 minutes	VR(q)	0.8246	0.7342	0.7111	0.7101	0.7167	0.6968
	z(q)	-4.4658	-4.4443	-4.0253	-3.5947	-3.1350	-2.9401
	p value	$\ll 0$	$\ll 0$	0.0001	0.0003	0.0017	0.0033
	Decision	R	R	R	R	R	R
2 minutes	VR(q)	0.8304	0.7000	0.6149	0.5878	0.5890	0.5935
	z(q)	-12.2700	-12.0676	-11.1078	-9.7450	-8.2392	-6.8144
	p value	$\ll 0$	$\ll 0$	$\ll 0$	$\ll 0$	$\ll 0$	0.9463
	Decision	R	R	R	R	R	R

The decision tells whether the random walk hypothesis is rejected(R) or failed to reject (NR) at 5% significance .

Tab. A.8: The results of variance ratio test for NPN log returns from 01 October 2012 to 30 September 2013.

Sampling Frequency	q	2	4	8	16	32	64
Daily	VR(q)	1.1213	1.1846	1.1635	1.4590	1.2290	1.0056
	z(q)	1.7505	1.4783	0.8441	1.5979	0.5335	0.0091
	p value	0.0800	0.1393	0.3986	0.1101	0.5937	0.9928
	Decision	NR	NR	NR	NR	NR	NR
4 hours	VR(q)	1.0067	1.1060	1.1329	1.1020	1.3919	1.1490
	z(q)	0.1399	1.1265	0.9266	0.4981	1.3260	0.3469
	p value	0.8888	0.2600	0.3542	0.6184	0.1848	0.7287
	Decision	NR	NR	NR	NR	NR	NR
2 hours	VR(q)	0.9826	1.0006	1.0920	1.1166	1.0768	1.3614
	z(q)	-0.6603	0.0112	0.9314	0.7925 0.3645	1.1816	
	p value	0.5091	0.9911	0.3516	0.4281	0.7155	0.2374
	Decision	NR	NR	NR	NR	NR	NR
1 hour	VR(q)	1.0398	1.0694	1.1146	1.1957	1.2236	1.1809
	z(q)	1.4395	1.3864	1.4745	1.7195	1.4040	0.8206
	p value	0.1500	0.1656	0.1403	0.0855	0.1603	0.4119
	Decision	NR	NR	NR	NR	NR	NR
30 minutes	VR(q)	0.9835 1.0093	1.0451	1.1009	1.1872	1.2184	
	z(q)	-0.8088	0.2558	0.8165	1.2674	1.6279	1.3555
	p value	0.4186	0.7981	0.4142	0.2050	0.1035	0.1753
	Decision	NR	NR	NR	NR	NR	NR
20 minutes	VR(q)	0.5857	0.3742	0.2703	0.2318	0.2219	0.2174
	z(q)	-1.0064	-1.0134	-1.0125	-0.9943	-0.9736	-0.9618
	p value	0.3142	0.3109	0.3113	0.3201	0.3303	0.3361
	Decision	NR	NR	NR	NR	NR	NR
10 minutes	VR(q)	0.5801	0.3739	0.2689	0.2191	0.2053	0.2072
	z(q)	-1.0363	-1.0301	-1.0309	-1.0275	-1.0115	-0.9921
	p value	0.3000	0.3030	0.3026	0.3042	0.3118	0.3211
	Decision	NR	NR	NR	NR	NR	NR
5 minutes	VR(q)	0.5813	0.3678	0.2656	0.2149	0.1914	0.1898
	z(q)	-1.0634	-1.0705	-1.0658	-1.0633	-1.0595	-1.0444
	p value	0.2876	0.2844	0.2865	0.2877	0.2894	0.2963
	Decision	NR	NR	NR	NR	NR	NR
2 minutes	VR(q)	0.5902	0.3744	0.2611	0.2081	0.1840	0.1732
	z(q)	-1.0819	-1.1010	-1.1145	-1.1148	-1.1116	-1.1081
	p value	0.2793	0.2709	0.2651	0.2650	0.2663	0.2678
	Decision	NR	NR	NR	NR	NR	NR

The decision tells whether the random walk hypothesis is rejected(R) or failed to reject (NR) at 5% significance .

Tab. A.9: The results of variance ratio test for SHP log returns from 01 October 2012 to 30 September 2013.

RIC	Stock name	Total number of trades	RIC	Stock name	Total number of trades
AGL	Anglo American PLC	69582	AMS	Aglo American Platinum Ltd	49399
ANG	AngloGold Ashanti Ltd	64179	APN	Aspen Pharmacare Holdings Limited	38332
ARI	African Rainbow Minerals Ltd	26810	ASA	Barclays Africa Group Ltd	46646
ASR	Assore Litd	13805	BIL	BHP Billiton PLC	47294
BTI	British America Tobacco	50005	BVT	Bidvest Group Ltd	49455
CFR	Compagnie Financiere Richemont SA	43535	DSY	Discovery Limited	17804
EXX	Exxara Resources Ltd	46969	FSR	First Rand Ltd	52574
GFI	Gold Fields Ltd	81790	GRT	Growthpoint Properties Ltd	16511
IMP	Impala Platinum Holdings Ltd	76172	INL	Investec Ltd	31984
INP	Investec PLC	30465	IPL	Imperial Holdings Ltd	49996
ITU	Intu Properties PLC	-	KIO	Kumba Iron Ore Ltd	48097
MDC	Mediclinic International Ltd	22799	MND	Mondi Ltd	23244
MNP	Mondi PLC	26323	MSM	Massmart Holdings Ltd	22226
MTN	MTN Group Ltd	90374	NED	Nedbank Group Ltd	24038
NPN	Naspers Ltd	70660	OML	Old Mutual Plc	24029
REM	Remgro Ltd	30442	RMH	RMB Holdings Ltd	22009
SAB	SABMiller Ltd	57869	SBK	Standard Bank Group Ltd	51097
SHF	Steinhoff International Holdings	41783	SHP	Shoprite Holdings Ltd	87598
SLM	Sanlam Ltd	35304	SOL	Sasol Ltd	69503
TBS	Tiger Brands Ltd	48020	TRU	Truworths International Ltd	42523
VOD	Vodacom Group Ltd	50312	WHL	Woolworths Holdings Ltd	72904

Tab. A.1: Summary of the top 42 stocks on the JSE data from the 2nd to the 28th of January 2013

Appendix B

Bessel Functions

In this section, properties of the modified Bessel equation of the third kind are discussed, since the equation is related to the probability density function of the generalized inverse Gaussian distribution that is discussed in chapter 5. A good reference for properties of Bessel functions is Abramowitz and Stegun (1964).

The modified Bessel differential equation is written as follows

$$x^2y'' + xy' + (x^2 - n^2)y = 0, \tag{B.1}$$

where $n \in \mathbb{N}$. In the integral form, the modified Bessel equation of the third kind can be represented as

$$K_\nu(x) = \frac{1}{2}\int_0^\infty t^{\nu-1}e^{-\frac{1}{2}x(t+t^{-1})}dt,\ x > 0. \tag{B.2}$$

Define an integral that is similar to the Bessel function

$$k_\lambda(\chi,\psi) = \int_0^\infty x^{\lambda-1}e^{-\frac{1}{2}(\psi x+\chi x^{-1})}dx. \tag{B.3}$$

Let $\eta = \sqrt{\chi/\psi}$ and $\omega = \sqrt{\chi\psi}$ and substitute $x = \eta y$, to get

$$\begin{aligned}
k_\lambda(\chi,\psi) &= \int_0^\infty x^{\lambda-1}e^{-\frac{1}{2}(\psi x+\chi x^{-1})}dx \\
&= \int_0^\infty x^{\lambda-1}e^{-\frac{1}{2}\omega(\frac{x}{\eta}+\frac{\eta}{x})}dx \\
&= \int_0^\infty (\eta y)^{\lambda-1}e^{-\frac{1}{2}\omega(y+y^{-1})} \times \eta dy \\
&= 2\eta^\lambda \times \frac{1}{2}\int_0^\infty y^{\lambda-1}e^{-\frac{1}{2}\omega(y+y^{-1})}dy \\
&= 2\eta^\lambda K_\lambda(\omega).
\end{aligned} \tag{B.4}$$

From equations B.2 and B.4, we have that

$$K_\lambda(\sqrt{\chi\psi}) = \frac{1}{2}k_\lambda(\chi,\psi)\left(\frac{\psi}{\chi}\right)^{(\lambda/2)}. \tag{B.5}$$

To compute $\mathbb{E}[\ln W]$ when $W \sim N^{\vdash}(\lambda, \chi, \psi)$ we use the pdf of the GIG given by equation 4.1. Let

$$I = \int_0^\infty \frac{\left(\frac{\psi}{\chi}\right)^\lambda}{2K_\lambda(\sqrt{\chi\psi})} w^{\lambda-1} e^{-\frac{1}{2}(\chi x^{-1}+\psi x)} dw = 1. \tag{B.6}$$

Note that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\chi}{\psi}\right)^{-\lambda/2} = \frac{\mathrm{d}}{\mathrm{d}\lambda} e^{\frac{-\lambda}{2}\ln(\frac{\chi}{\psi})} = -\frac{1}{2}\ln\frac{\psi}{\chi}, \tag{B.7}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}(w^{\lambda-1}) = \frac{\mathrm{d}}{\mathrm{d}\lambda} e^{(\lambda-1)\ln w} = \ln w \times w^{\lambda-1}. \tag{B.8}$$

Using the chain rule we have that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} = -\frac{\frac{K_\lambda(\sqrt{\chi\psi})}{\mathrm{d}\lambda}}{\left(K_\lambda(\sqrt{\chi\psi})\right)^2} \tag{B.9}$$

The only terms that have a λ term in I are of the form

$$q = \frac{\left(\frac{\chi}{\psi}\right)^{-\lambda/2}}{K_\lambda(\sqrt{(\chi\psi)})} w^{\lambda-1}, \tag{B.10}$$

using the product rule and chain rule of differentiation, we get the following

$$\begin{aligned}\frac{\mathrm{d}I}{\mathrm{d}\lambda} = 0 = -\frac{1}{2}\ln\left(\frac{\chi}{\psi}\right) I - \frac{\frac{K_\lambda(\sqrt{\chi\psi})}{\mathrm{d}\lambda}}{K_\lambda(\sqrt{\chi\psi})} \times I \\ + \int_0^\infty \ln(w) \frac{\left(\frac{\psi}{\chi}\right)^\lambda}{2K_\lambda(\sqrt{\chi\psi})} w^{\lambda-1} e^{-\frac{1}{2}(\chi x^{-1}+\psi x)} dw\end{aligned} \tag{B.11}$$

$$\begin{aligned}\implies 0 = -\frac{1}{2}\ln\left(\frac{\chi}{\psi}\right) - \frac{\frac{K_\lambda(\sqrt{\chi\psi})}{\mathrm{d}\lambda}}{K_\lambda(\sqrt{\chi\psi})} + \mathbb{E}[\ln W], \\ \therefore \mathbb{E}[\ln W] = \frac{1}{2}\ln\left(\frac{\chi}{\psi}\right) + \frac{\frac{K_\lambda(\sqrt{\chi\psi})}{\mathrm{d}\lambda}}{K_\lambda(\sqrt{\chi\psi})}.\end{aligned} \tag{B.12}$$

For $\lambda \in \mathbb{R}$ and $x > 0$, the Bessel function of the first kind is defined as follows

$$J_\lambda(x) = \frac{1}{\pi}\int_0^\pi \cos(x\sin u - \lambda u)\,\mathrm{d}u - \frac{sin(\lambda\pi)}{\pi}\int_0^\infty e^{-x\sinh(u)-\lambda u}\,\mathrm{d}u, \quad x > 0. \tag{B.13}$$

When $\lambda > -\frac{1}{2}$, the Bessel function of the first kind is given by

$$J_\lambda(x) = \frac{2(\frac{1}{2}x)^\lambda}{\sqrt{\pi}\Gamma(\lambda+\frac{1}{2})}\int_0^1 (1-u^2)^{\lambda-\frac{1}{2}}\cos(xu)\,\mathrm{d}u, \quad x \in \mathbb{R}, \tag{B.14}$$

where Γ denotes the Gamma function. When $\lambda \in \mathbb{R}$ and $x > 0$, the Bessel function of the second kind is defined as follows

$$Y_\lambda(x) = \frac{1}{\pi}\int_0^\pi \sin(x\sin(u) - \lambda u)\,\mathrm{d}u - \frac{1}{\pi}\left(e^{\lambda u} + e^{-\lambda u}\cos(\lambda\pi)\right)e^{-x\sinh(u)}\,\mathrm{d}u. \tag{B.15}$$

Appendix C

Algorithms

Algorithm 7 Brent (1973) minimization algorithm

inputs a,b, f(x), eps
(a,b) the interval in consideration and tolerance eps which is not than the square root of the machine precision
procedure
$c = (3 - \sqrt{5})/2; e = 0$
$x = a + c * (b - a);$ ▷ always the last evaluated point of f
$v = x; w = x;$
$funx = f(x); funw = funx\ ; funv = funx;$
Main loop ▷ Iterate until convergence
$m = 0.5 * (a + b);$
$tol = eps1 * |x| + eps/3; t2 = 2 * tol;$
if $|x - m| \leq (t2 - 0.5 * (b - a))$ **then** ▷ convergence has been reached
    return (x,funx) ; ▷ the minimum point in (a,b)
if $|x - m| > t2 - 0.5 * (b - a)$ **then** ▷ check stopping criterion
    $p = 0; q = 0; r = 0;$
    if $|e| > tol$ **then** ▷ Fit hyperbola
        $r = (x - w) * (funx - funv);$
        $q = (x - v) * (funx - funw);$
        $p = ((x - v) * q) - (x - w) * r;$
        $q = 2 * (q - r);$
        if $q > 0$ **then**
            $p = -p;$
        else
            $q = -q;$
        $r = e; e = d1;$
    if $|p| < abs|0.5 * q * r| \wedge p < q * (a - x) \wedge p < q * (b - x)$ **then**
        ▷ A parabolic interpolation step
        $d1 = (p/q); u = x + d1;$

Algorithm 7 Brent (1973) minimization algorithm continued

28: **if** $(u - a) < t2 \vee (b - u) < t2$ **then**;
29: ▷ f must not be evaluated too close to a or b
30: **if** $x < m$ **then**
31: $d1 = tol$;
32: **else**
33: $d1 = -tol$;
34: **else** ▷ A golden search step
35: **if** $x < m$ **then**
36: $e = b - x$;
37: **else**
38: $e = a - x$;
39: $d1 = c * e$;
40: **if** $|d1| \geq tol$ **then** ▷ f must not be evaluated too close to to x
41: $u = x + d1$;
42: **else if** $d1 > 0$ **then**
43: $u = x + tol$;
44: **else**
45: $u = x - tol$;
46: $funu = f(u)$;
47: **if** $funu < funx$ **then** ▷ update a,b,v,w and x
48: **if** $u < x$ **then**
49: $b = x$;
50: **else**
51: $a = x$;
52: $v = w$; $funv = funw$;
53: $w = x$; $funw = funx$;
54: $x = u$; $funx = funu$;
55: **else**
56: **if** $u < x$ **then**
57: $a = u$;
58: **else**
59: $b = u$;
60: **if** $funu < funw \vee w = x$ **then**
61: $v = w$; $funv = funw$;
62: $w = u$; $funw = funu$;
63: **else if** $funu < funw \vee v = x \vee v = x$ **then**
64: $v = u$; $funv = funu$;
65: **End Main Loop**
66: **return** (x, funx) ▷ the minimum point in (a,b)

Algorithm 8 Luo and Zhao (2014) root finding algorithm for the GGD

inputs I_ζ the interval of ζ, W sample set, n size , δ tolerance, procedure $compF$
outputs $\hat{\alpha}$, $\hat{\beta}$, $\hat{\zeta}$
$y \leftarrow -\infty$
for all $\zeta \in I_\zeta$ **do**
    raise all the points $w \in W$ to the power of ζ, w^ζ
    calculate Q according to equation 4.40
    set $a = \frac{n}{2Q}$; $b = \frac{n}{Q}$
    $f_a = compF(a, n, Q)$; $f_b = compF(b, n, Q)$
    if $|f_a| > |f_b|$ **then**
        $k_0 = a$; $k_1 = b$;
    else
        $k_0 = b$; $k_1 = a$;
    end if
    while $(|a - b| \geq \delta | f_a \neq 0 | f_b \neq 0)$ **do**
        calculate t_a, t_b, s_a, s_b according to equation 4.53;
        set $\sigma = \max\{t_a, s_b\}$; $\tau = \min\{s_a, t_b\}$;
        if $k_r \neq k_{r-1}$ **then**
            $w = \frac{k_r k_{r-1}(f_{k_r} - f_{k_{r-1}})}{k_r f_{k_{r-1}} - k_{r-1} f_{k_{r-1}}}$; (derived from equation 4.56)
        else
            $w = \frac{\sigma+\tau}{2}$;
        if $w < \sigma$ **then**;
            $k_{r+1} = \sigma$
        else if $w > \tau$ **then**
            $k_{r+1} = w$;
        else
            $k_{r+1} = w$;
        end if
        $f_{k_{r+1}=compF(k_{r+1},n,Q)}$;
        if $f_{k_{r+1}} > 0$ **then**
            $a = k_{r+1}$; $b = \tau$;
            $f_a = f_{k_{r+1}}$; $f_b = compF(b, n, Q)$
        else
            $a = \sigma$; $b = k_{r+1}$;
            $f_a = compF(a, n, Q)$; $f_b = f_{k_{r+1}}$;
        end if
        $r = r + 1$;
    end while
    if $|f_a| < |f_b|$ **then**
        $\alpha^* = a$;
    else
        $\alpha^* = b$;
    calculate β^* and $\ln L(w; \zeta, \alpha, \beta)$ by Eqn. 4.29 and 4.26.
    if $\ln L(w; \zeta, \alpha, \beta) > y^*$ **then**
        $\hat{\alpha} = \alpha^*$; $\hat{\beta} = \beta^*$; $\hat{\zeta} = \zeta$; $y^* = \ln L(w; \zeta, \alpha, \beta)$
    end if
end for
return $\hat{\alpha}$, $\hat{\beta}$, $\hat{\zeta}$

Algorithm 9 The subroutine $compF(x, n, Q)$ of algorithm 8

input x; sample size n; offset Q
output f_x
function $compF(x)$
$\psi_x \leftarrow compPsi(x)$
$f_x \leftarrow n(\ln x - \psi_x)$
return f_x
end function

Algorithm 10 The subroutine $compPsi(x)$ of algorithm 9

function $compPsi(x)$
if $x \leq 10^{-7}$ **then**
    $res \leftarrow -\gamma - \frac{1}{x} + \frac{\pi^2 x}{6}$ ▷ γ is the Euler's constant
else if $x \geq 8.5$ **then**
    $res \leftarrow \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6}$
else
    $res \leftarrow -1/x + compPsi(x + 1)$
end if
return res
end function