Master's Thesis

Using Lévy Processes to Calculate Probability of Default

Author:
Ajla Nurkanović

Mentor:
Prof. Dr. Ralf Korn

Supervisor:
Dr. Robert Knobloch

A thesis in fulfillment of the requirements for the degree of Master of Science

at the

Department of Mathematics

in cooperation with

Walbing Technologies GmbH

TU Kaiserslautern

January 2022

Declaration of Authorship

I, Ajla NURKANOVIĆ, declare that this Master's Thesis titled "Using Lévy Processes to Calculate Probability of Default" is my own work and was written without help or any aids other than those specified therein. I also affirm that all text passages, drawings, sketches, illustrations, and similar that were taken from other works are quoted either literally or correspondingly have been identified as such.

I declare and confirm that this Master's Thesis has not been submitted in this or any other form to any other institution in the context of examination requirements.

Signed: Ajla Nurkanović

Date: 25.1.2022

To my parents Mehmed and Zehra

Acknowledgments

Firstly, I wish to thank my mentor Prof. Dr. Ralf Korn for all his invaluable advice, knowledge and the time he selflessly invested in improving my thesis, but above all, in these not-so-exciting Corona times, I have been most grateful for his human virtues – his great calmness and encouragement – which truly helped me get through this not only sanely but also happily.

I am thankful to Walbing Technologies GmbH for hiring me to do this thesis in cooperation with them – it has been a great pleasure.

I am pleased to have had Dr. Robert Knobloch, who is also a good friend of mine, as the supervisor at Walbing Technologies GmbH. Long-hours discussions with you made it possible to overcome the problems I was facing and made this thesis even more interesting.

I am grateful for my parents who supported my dreams since always, who made it possible for me to come and do this Master's program without worrying about anything else. Thank you for all the fruitful mathematical discussion over the years that built my strong mathematical foundation. But most of all, thank you for believing in me from day one – way more than I did myself. Therefore, it is you whom I devote this thesis to.

I am thankful to my sister, not only for always looking out for me, but also for the things she might not even be aware of. Thanks to her, I got to know about Fraunhofer ITWM where I then did an internship in 2018 and also met Robert – thanks to whom I started my student job at Walbing Technologies GmbH in 2021. Mima, this amazing journey in Germany started thanks to you already in 2018.

I am beyond thankful and thrilled for becoming an aunt in 2020, to my Mia. I have had the cutest and loveliest cheerleader since then.

Finally, my warmest thank you for the last years, especially during Corona times (world's and mine) and this degree, goes to my Goran. You helped me go through this both emotionally and professionally, by comforting me, showing me my strength when I thought I did not have any, by pushing me forward and discussing with me about the topics of my thesis – which have nothing to do with physics – and giving me great ideas or asking insightful questions that led to such. You are an inexhaustible inspiration.

Last but not least, I thank my friends, especially my girls – Vana, Jelena, Amina, Samira, Valjeta and Ayurda, and the rest of my family, as well as my dear Teta, who were there for me over the last few years.

To all of these people, a simple thank you in a few sentences is not enough. Please know this thesis would not be as successful without you. You all had an influence on what I have become and what I am still becoming as a person, and you all have a special place in my heart.

Contents

Chapter 1

Introduction

Banks and other financial institutions have always had interest in the assessment of credit risk, and the probability of default (PD) - an estimate of the likelihood that a borrower will be unable to meet her debt obligations - is one of the essential measures of credit risks used. However, recently this has become a concern not just of the banks and financial institutions, since companies are often obligated to pay a spread over the default-free rate of interest proportional to their default probability to compensate lenders for this uncertainty.

Merton's structural model Merton (1974) is a classical model for estimating the PD, based on an assumption that the company's assets are modeled through a Geometric Brownian Motion (GBM) and by modeling the company's equity as a call option on its assets.

One innovative forecasting model which has been widely applied in practice is a particular application of a modified version of the Merton's model that was developed by Kealhofer, McQuown and Vasicek, refered to as the KMV model (see Crosbie and Bohn (2019) for an exposition). The KMV uses an empirical function, i.e. the cumulative distribution function of the standard normal distribution is replaced by an empirical cumulative distribution function. The KMV Corporation has been tracking the default events of all publicly traded companies in the United States since 1973 – their database includes over 250 000 company-years of data and over 4 700 incidents of default or bankruptcy. From this data a frequency table can be generated which relates the likelihood of default to various levels of distance-to-default, (Crosbie and Bohn, 2019, p. 14). That data provided a comparison of the default probabilities calculated from the log-normal distribution with actual realized default rates, which showed that there are small but important differences between the theoretical and actual default rates, i.e. the resulting empirical distribution of default rates has much wider tails than the normal distribution. For example, under the normality assumption of the Merton's model, any company more than four standard deviations from its default point would have in-fact zero probability of default. However, the empirical default probability for such companies is significantly higher, about 0.5%, (Kealhofer, 2003, p. 32). These small numeric differences in tail probabilities between the normal distribution and the empirical distribution result in economically significant differences in terms of default risk. Therefore, it is of our interest to build a model which will be able to capture a higher probability of default than the one predicted by the Merton model.

In this thesis, we firstly explain the Merton model in Chapter 2, giving the main idea and explaining the problem of unobservable asset values, as well as deriving the formula for the probability of default.

In Chapter 3 we introduce Lévy processes together with some of their main properties, and we also explain the exponential Lévy models together with the induced problem of market incompleteness.

Of our particular interest are subordinated processes such as Variance Gamma (VG) and Normal Inverse Gaussian (NIG) introduced in Chapter 4 and Chapter 5, respectively.

In Chapter 6 we show that the equity market values of German public companies are in most cases not normally distributed - as would be the case when modeling them with GBM. Considering three different distributions: VG, NIG and the normal distribution, most of the German public companies' data are best fitted with a VG, in the second place is NIG and only a very few are best fitted with a normal distribution - which motivates the development and use of NIG and VG Merton based models we introduce and develop in Chapter 6. The NIG model can also be found in Jovan and Ahčan (2017), however our approach to the EM algorithm is slightly different.

In contrast to Brambilla et al. (2015), we explicitly deal with the problem of incomplete markets and an equivalent martingale measure (EMM) for both NIG and VG processes (see Section 4.4 and Section 5.4), which are fundamental for the corresponding option pricing formulae needed for the construction of the models.

The problem of unobservable asset values and unknown underlying process parameters is solved via the EM algorithm, using the log-likelihood formula for equity as introduced in Duan (1994). With the obtained asset and parameter values, together with the equity market and debt values, in Chapter 7 we estimate the NIG and Merton's probabilities of default for some German public companies, whose data is available on Eikon Refinitiv (2020)[1], and compare them between the two models.

[1] The datasets analyzed for this study are owned by the financial data provider Eikon Refinitiv (2020) and restrictions apply to the availability of these data, which were used under license from Refinitiv Eikon, and so are not publicly available. Data are however available from the corresponding author upon reasonable request and permission of Refinitiv Eikon.

Chapter 2

Merton Model

Throughout the chapters we assume that, unless stated otherwise, all random objects are defined on a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

2.1 Introduction

The Merton model is one of the most famous structure credit risk models proposed by economist Robert Merton in 1974, Merton (1974). It is a model for assessing the structural credit risk of a company by modeling the company's equity as a call option on its assets. The model uses the Black-Scholes-Merton option pricing methods and is structural because it provides a relationship between the default risk and the asset (capital) structure of the firm. Analysts and investors utilize the Merton model to understand how capable a company is at meeting financial obligations and servicing its debt, as well as for weighing the general possibility that it will go into credit default.

This model, for starters, assumes the following:

- The company has a single debt liability, equity, and has no other obligations.
- The company makes no other cash payouts (e.g. equity dividends).

Under the assumption that the market value of the company's assets evolves as a log-normal process, Merton showed that this model provides a closed-form solution for the value of the company's debt.

In what follows let $W := (W(t))_{t\geq 0} = (W_t)_{t\geq 0}$ be a Brownian motion on $(\Omega, \mathcal{F}, \mathbb{P})$.

2.2 The main idea

The Merton model takes an overly simple debt structure and assumes that the total value A_t of a firm's assets, at time t, follows a geometric Brownian motion under $\mathbb{P}$, i.e.

$$\mathrm{d}A_t = \mu_A A_t \,\mathrm{d}t + \sigma_A A_t \,\mathrm{d}W_t, \quad A_0 > 0, \tag{2.1}$$

where μ_A is the mean rate of return on the assets, σ_A is the asset volatility and A_0 denotes the initial market value of the firm's assets.

A company balance sheet records book values – the value of a firm's equity E_t, its total assets A_t, and its total liabilities L_t at time t. The relationship between

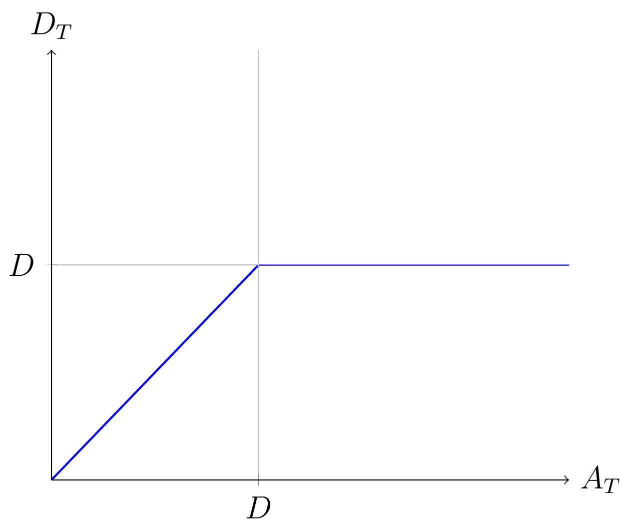


Figure 2.1: Payoff from debt at maturity period T related to value of assets

these values is defined by the following equation, also known as the fundamental identity of accounting,

$$A_t = E_t + L_t. \tag{2.2}$$

Potentially, these book values for E_t, A_t, and L_t are all observable as they are recorded on a firm's balance sheet, if a firm provides a financial balance sheet. However, the book values are reported infrequently and are obtainable only with a significant delay. Therefore, it may be more suitable to only consider the equity market value, which for stock companies is given by the firm's stock market price multiplied by the number of outstanding shares, making it an observable quantity. In this interpretation, the market value of the firm's assets A_t and total liabilities L_t are treated as being unobservable. This problem will be considered in more detail in Section 2.3.

The Merton model is concerned with the situation when a firm has issued two classes of securities: equity and zero-coupon bond, assuming the equity receives no dividends. The bonds represent the firm's debt obligation maturing at maturity time T with principal value D.

As discussed in Zieliński (2013), let us consider A_0, E_0 to be the values of the firm's assets and equity today, respectively, and A_T, E_T the values at date T. If at time T the firm's asset value exceeds the promised payment D, i.e. $A_T \geq D$, the lenders (debt-holders), are paid the full face value D and the equity-holders receive the residual asset value $A_T - D$. If the asset value is lower than the promised payment, the firm defaults, debt-holders receive total asset value, and the equity-holders receive nothing, Hull et al. (2005). Thus, the amount D_T received by the debt-holders on date T can be expressed as (Fig. 2.1)

$$D_T = \begin{cases} D, & \text{if } A_T \geq D \\ A_T, & \text{otherwise} \end{cases} \tag{2.3}$$

This can be rewritten as the payoff from an option position, i.e.

$$D_T = D - \max(D - A_T, 0). \tag{2.4}$$

Hence, option-pricing techniques are used on the problem of pricing risky debt. To better see this, let us consider Eq. (2.4) in more detail.

The first component of Eq. (2.4), i.e. D, is the payoff from investing into a risk-free zero coupon bond maturing at time T with a face value of D.

The second one, $-\max(D - A_T, 0)$, represents the payoff from a short position in a put option on the firm's assets with a strike price of D and a maturity date of T.

This gives us a procedure to calculate the present value of the risky debt in two steps (Fig. 2.2):

1. Identify the present value D of the risk-free debt,
2. subtract the present value $X_{Put}(0) = \max(D - A_T, 0)$ of the put option.

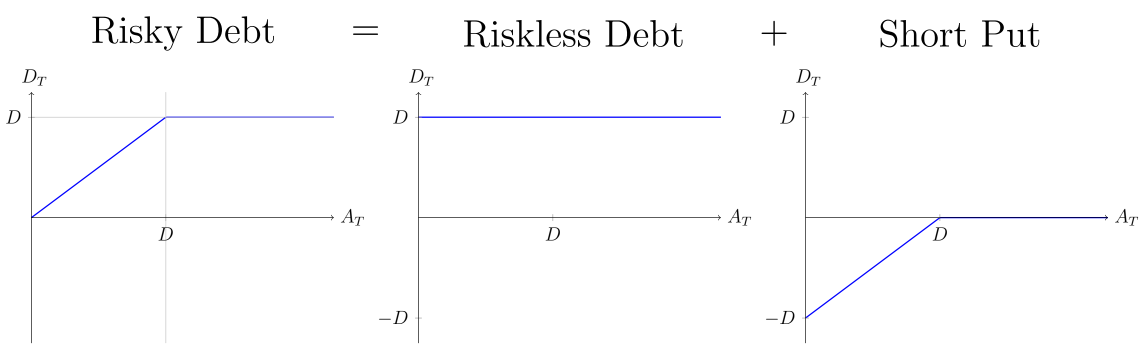


Figure 2.2: Decomposition of debt value at time T

For the first step the formula of continuous compounding of interest is used, and for the second one the Black-Scholes formula. However, firstly, the assumptions of the Black-Scholes model – the firm value A_t following a log-normal distribution with constant volatility σ_A and the constant risk-free rate of interest r – are needed. Now, the value $X_{Put}(t)$ of the put option is obtained via the Black-Scholes pricing formula as follows:

$$X_{Put}(t) = De^{-r(T-t)}\Phi(-d_2(t)) - A_t\Phi(-d_1(t)), \tag{2.5}$$

where

$$d_1(t) := \frac{\ln(A_t/D) + (r + 1/2\sigma_A^2)(T-t)}{\sigma_A\sqrt{T-t}}, \quad d_2(t) := d_1(t) - \sigma_A\sqrt{T-t}, \tag{2.6}$$

and where Φ represents the cumulative distribution function (cdf) of the standard normal distribution function.

2.3 Unobservability of the firm value process

Valuation of the Merton model brings a significant problem – both the firm value A_0 and its volatility σ_A are mostly unobservable. However, not all is doomed! Although they are not directly observable, there is an implicit way to derive these values.

Assume the firm is publicly traded with observable equity prices modelled by the stochastic differential equation

$$\mathrm{d}E_t = \mu_E E_t \,\mathrm{d}t + \sigma_E E_t \,\mathrm{d}W_t, \quad E_0 > 0, \tag{2.7}$$

in the setting of a constant risk-free rate of interest r, where E_0 is the initial value of the firm's equity and σ_E denotes the constant volatility of equity which can be estimated by means of historical data on equity prices. Similarly to the previous section, by expressing the equity via the relationship

$$E_T = \max(0, A_T - D), \tag{2.8}$$

and using the Black-Scholes model, under the assumptions on σ_E and r, the Merton model links the market value of the equity and the market value of the assets as

$$E_t = A_t\Phi(d_1(t)) - De^{-r(T-t)}\Phi(d_2(t)), \tag{2.9}$$

where d_1 and d_2 are given as in Eq. (2.6). As the equity prices E_t are observable, this is an equation with two unknowns – A_t and σ_A. From Eq. (2.9), we see that E_t is both A_t and t dependent, and that we can apply Itô's formula to it which results in

$$\mathrm{d}E_t = \left(\frac{1}{2}\sigma_A^2 A_t^2 \frac{\partial^2 E_t}{\partial A_t^2} + \mu_A A_t \frac{\partial E_t}{\partial A_t} + \frac{\partial E_t}{\partial t}\right)\mathrm{d}t + \sigma_A A_t \frac{\partial E_t}{\partial A_t}\,\mathrm{d}W_t. \tag{2.10}$$

From partial derivation of Eq. (2.9), one can show (see Appendix B.1)

$$\frac{\partial E_t}{\partial A_t} = \Phi(d_1(t)). \tag{2.11}$$

Now, comparing the diffusion terms in Eq. (2.7) and Eq. (2.10), and using Eq. (2.11), we get that the volatility of the equity and the volatility of the asset satisfy the relation

$$\sigma_E = \frac{A_t}{E_t}\Phi\left(d_1(t)\right)\sigma_A. \tag{2.12}$$

By solving the system of nonlinear equations Eq. (2.9) and Eq. (2.12), the unobservable values of A_t and σ_A are determined.

2.4 Distance to default

Based on the main concept and the assumption that a company's value can be modeled by a geometric Brownian motion, it turns out to be possible to compute the company's probability of default.

In (Dar et al., 2019, p. 5), the solution of Eq. (2.1) is derived, i.e. A_t satisfies

$$A_t = A_0 e^{\sigma_A W_t + \left(\mu_A - \frac{1}{2}\sigma_A^2\right)t}. \tag{2.13}$$

As W is a Brownian motion, i.e. W_t is normally distributed, A_t is log-normally distributed:

$$\log A_t = \log A_0 + \left(\mu_A - \frac{1}{2}\sigma_A^2\right)t + \sigma_A \,\mathrm{d}W_t\,. \tag{2.14}$$

From this we can see that the distribution of the stochastic process $\log A$ at time point t is normally distributed as follows

$$\log A_t \sim \mathcal{N}\left(\log A_0 + \left(\mu_A - \frac{1}{2}\sigma_A^2\right)t, \sigma_A^2 t\right). \tag{2.15}$$

Now, we can read off the expectation and variance of $\log A_t$ as

$$\mathbb{E}(\log A_t) = \log A_0 + \left(\mu_A - \frac{\sigma_A^2}{2}\right)t, \quad \mathrm{Var}(\log A_t) = \sigma_A^2 t.$$

This results in the probability of default being

$$\begin{aligned}\mathbb{P}(A_t < B) &= \mathbb{P}(\log A_t < \log D) \\ &= \mathbb{P}\left(\frac{\log A_t - \mathbb{E}(\log A_t)}{\sqrt{\mathrm{Var}(\log A_t)}} < \frac{\log D - \mathbb{E}(\log A_t)}{\sqrt{\mathrm{Var}(\log A_t)}}\right) \\ &= \Phi\left(\frac{\log D - \left(\log A_0 + \left(\mu_A - \frac{\sigma_A^2}{2}\right)t\right)}{\sqrt{\sigma_A^2 t}}\right) = \Phi\left(-DD(t)\right),\end{aligned} \tag{2.16}$$

where DD is given by the following definition.

Definition 1 (Distance to default)**.** *The distance to default in Merton model is defined by*

$$DD(t) := -\frac{\log D - \left(\log A_0 + \left(\mu_A - \frac{\sigma_A^2}{2}\right)t\right)}{\sqrt{\sigma_A^2 t}} = -\frac{\log D - \mathbb{E}(\log A_t)}{\sqrt{\mathrm{Var}(\log A_t)}} \tag{2.17}$$

and indicates how many standard deviations of $\log A_t$ *the expected value of* $\log A_t$ *is away from the failure point* $\log D$.

Graphically the basic concept of the Merton formula can be seen in Fig. 2.3. The horizontal axis represents time, beginning at initial time 0 and terminating at maturity time T. Usually the period under consideration is 1 year ($T = 1$). The vertical axis represents the market value of the company's assets.

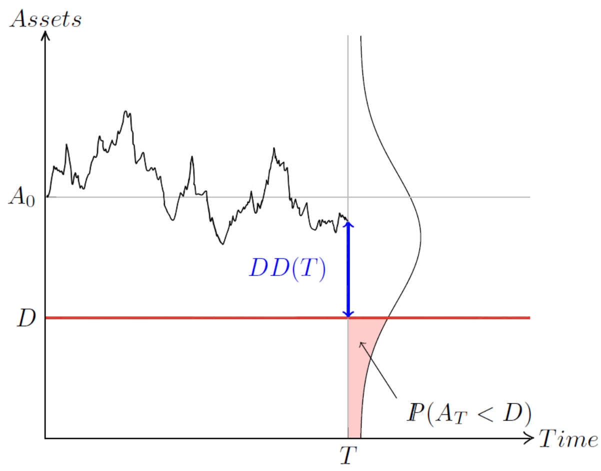


Figure 2.3: Basic concept of Merton's model

At time 0 the assets have a deterministic, observable value A_0, but at maturity time T a range of asset values is possible, and their probability density (vertical bell-curve in Fig. 2.3) describes the likelihood of various asset values at time T. The volatility of the assets conditions the likelihood of extreme outcomes, i.e. the more volatile the asset, the greater the probability of extreme outcomes. The horizontal red line shows the par amount[2] of the liability D due at maturity time T. If the company's asset value A_T at time T is less than the amount of the liability, the company will default. The probability of default is thus given by the shaded area below the default point, which represents the likelihood of the market value of the company's assets in one year being less than what the company owes.

Since a cdf is a non-decreasing function, from Eq. (2.16) we can read that the probability of default increases if the company's market value of assets decreases, if the amount of liabilities increases, or if the volatility of the assets' market value increases. Thus, the three variables – A_t, D and σ_A – are the determinants of the company's default probability.

[2]The amount of money that bond issuers promise to repay bondholders at the maturity date of the bond.

Chapter 3

Lévy Processes

3.1 Motivation

When observing the financial market, one can often notice jumps in the security price processes which motivates exploring stochastic processes with jumps. Lévy processes form a central class of stochastic processes, containing both the Brownian motion and the Poisson process as its examples. They are also examples of Markov processes and semimartingales with jumps occurring at random times. Due to their interesting properties, such as having infinitely divisible distributions, they are very flexible and suitable for usage in financial context and make managing the discontinuities of credit risk possible. We will also later see, when fitting distributions to our data, distributions of Lévy processes with jumps are almost always a better fit than a normal distribution of the BM model.

3.2 Definition and main properties

Definition 2 (Lévy Process)**.** *Let $(X_t)_{t\geq 0}$ be a stochastic process on a probability space $(\Omega, \mathscr{F}, \mathbb{P})$ with values in $\mathbb{R}$ such that $X_0 = 0$. It is called a Lévy process if it has the following properties:*

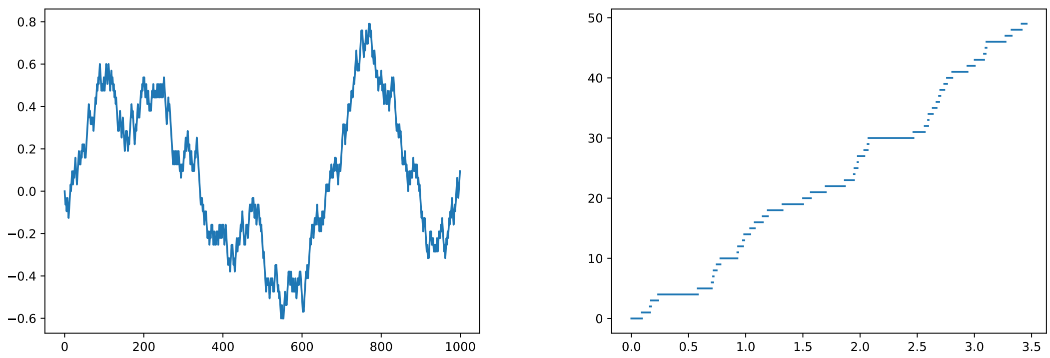
1. *Independent increments: for every increasing sequence of times $t_0, t_1, \dots, t_n$ the random variables $X_{t_0}, X_{t_1} - X_{t_0}, \dots, X_{t_n} - X_{t_{n-1}}$ are independent.*

2. *Stationary increments: the distribution of $X_{t+h} - X_t$ does not depend on t.*

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

One can notice that this definition is not much different than the definition of Brownian motion. The only difference lies in the third property – where the continuity constraint is replaced by a more general condition – stochastic continuity, i.e. continuity in probability. Therefore, Brownian motion itself is an example of a Lévy process, Fig. 3.1a. Some of the other popular examples are the Poisson process (Fig. 3.1b), the compound Poisson process, the α-stable process, the Gamma process, the Merton jump-diffusion process (Fig. 3.2), etc.

An important property of the Lévy process is that it is an infinitely divisible process which we state more clearly in the following Proposition, without proof.



(a) Brownian motion (b) Poisson process

Figure 3.1: Brownian motion and Poisson process as examples of Lévy processes

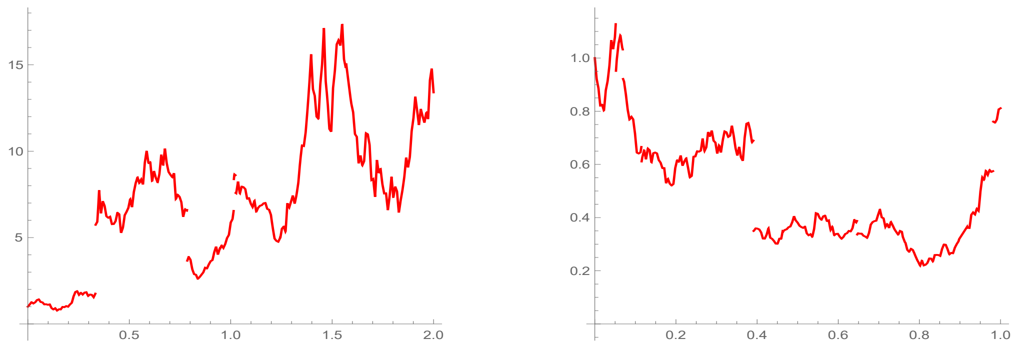


Figure 3.2: Examples of Lévy jump-diffusion processes [3].

Proposition 3 (Infinite divisibility and Lévy processes). *Let $(X_t)_{t\geq 0}$ be a Lévy process. Then for every t, X_t has an infinitely divisible distribution. Whereas, 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 characteristic function of a random variable, denoted with ϕ, is the Fourier transform of its distribution. The characteristic function of a random variable completely characterizes its law: two variables with the same characteristic function are identically distributed. The characteristic functions are an inevitable property of theory on Lévy processes which we will also later state for the Lévy processes used in this thesis. Therefore, we state another Proposition without proof. Interested readers are directed to e.g. Cont and Tankov (2006) or Applebaum (2009).

Proposition 4 (Characteristic function of a Lévy process). *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$. There exists a continuous function $\psi : \mathbb{R} \to \mathbb{R}$ called a characteristic exponent of X, such that*

$$\mathbb{E}[e^{iuX_t}] = e^{t\psi(u)}, \quad u \in \mathbb{R}. \tag{3.1}$$

The function ψ is the cumulant generating function of X_1.

[3]Simulated via "Merton's Jump Diffusion Model" from the Wolfram Demonstrations Project http://demonstrations.wolfram.com/MertonsJumpDiffusionModel.

The law of X_t is determined by the knowledge of the law of X_1, i.e. the only degree of freedom in specifying a Lévy process is in specifying the distribution of X_t for a single time, e.g. $t = 1$, (Cont and Tankov, 2006, p. 70).

For the following results, firstly we need to define a *Lévy measure*, which, as we will see, is of great importance for describing Lévy processes.

Definition 5 (Lévy measure). *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}$. The measure ν on $\mathbb{R}$ defined by:*

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

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

3.2.1 Fundamental results

The first fundamental result of the theory on Lévy processes that we are stating is the so-called *Lévy-Itô decomposition* (see e.g. (Cont and Tankov, 2006, p. 79)).

Proposition 6 (Lévy-Itô decomposition). *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}$ and ν the corresponding Lévy measure, given as in Definition 5. Then the following holds:*

a) ν is a Radon measure on $\mathbb{R} \setminus \{0\}$ and verifies

$$\int_{|x|\leq 1} |x|^2 \nu(\mathrm{d}x) < \infty, \quad \int_{|x|\leq 1} \nu(\mathrm{d}x) < \infty.$$

b) The jump measure of X, denoted by J_X, is a Poisson random measure on $[0,\infty) \times \mathbb{R}^s$ with intensity measure $\nu(\mathrm{d}x)\,\mathrm{d}t$.

c) There exist a $\sigma \in \mathbb{R}_{\geq 0}$ and a Brownian motion $(W_t)_{t\geq 0}$ with variance σ, such that

$$X_t = \gamma t + W_t + X_t^l + \lim_{\epsilon \downarrow 0} \tilde{X}_t^\epsilon, \tag{3.3}$$

where

$$X_t^l = \int_{|x|\geq 1, s\in[0,t]} x J_X(\mathrm{d}s \times \mathrm{d}s)$$

and

$$\tilde{X}_t^\epsilon = \int_{\epsilon\leq|x|<1, s\in[0,t]} x\{J_X(\mathrm{d}s \times \mathrm{d}s) - \nu(\mathrm{d}x)\,\mathrm{d}s\} = \int_{\epsilon\leq|x|<1, s\in[0,t]} x\tilde{J}_X(\mathrm{d}s \times \mathrm{d}x).$$

The terms in Eq. (3.3) are independent and the convergence in the last term is almost sure and uniform in $t \in [0,T]$.

As explained in (Cont and Tankov, 2006, p. 80), the first importance of Proposition 6 lies in the fact that it tells us that for every Lévy process there exist: $\gamma \in \mathbb{R}$ – the drift component, positive σ – the Brownian (motion) component and a positive measure ν, which together uniquely determine its distribution. The triplet (σ, γ, ν) is called the *characteristic triplet* or the *Lévy triplet* of the process X_t. Brownian motion is an example of a Lévy process with a characteristic triplet without the jump component ν, i.e. it's Lévy triplet is $(\sigma, \gamma, 0)$. A pure jump process is an example of a Lévy process without the Brownian component σ, i.e. with the characteristic triplet $(0, \gamma, \nu)$. A diffusion model with a jump term added corresponds to a Lévy process with the characteristic triplet (σ, γ, ν).

The second importance is in the implication that every Lévy process is a combination of a Brownian motion with drift and a (possibly infinite) sum of independent compound Poisson processes, which tells us that every Lévy process can be approximated, arbitrarily precisely, by the sum of a Brownian motion with drift and a compound Poisson process, (Cont and Tankov, 2006, p. 81).

The following famous result on Lévy processes – the *Lévy-Khinchin* representation – expresses the characteristic function of a Lévy process via terms of its characteristic triplet (σ, γ, ν).

Theorem 1 (Lévy-Khinchin representation). *Let $(X_t)_{t\geq 0}$ be a Lévy process on $\mathbb{R}$ with the characteristic triplet (σ, γ, ν). Then*

$$\mathbb{E}[e^{iuX_t}] = e^{t\psi(u)}, \quad u \in \mathbb{R}$$

with

$$\psi(u) = -\frac{1}{2}\sigma u^2 + i\gamma u + \int_{\mathbb{R}} (e^{iux} - 1 - iux\mathbb{1}_{|x|\leq 1})\nu(\mathrm{d}x). \tag{3.4}$$

3.2.2 Activity and variation

Activity and variation are often important properties when talking of Lévy processes. Remember, a Lévy process can be described by its three components: the Brownian motion/diffusion, the drift, and the jump component – which can be either of finite-activity or infinite-activity.

We say the process has finite activity if almost all paths have only a finite number of jumps along any finite time interval. On the other hand, it is said the process has infinite activity, if almost all paths have infinitely many jumps along any finite time interval.

A process is said to have finite (bounded) variation if it has finite total variation, otherwise we say it has infinite variation.

We now state one of the results on activity of Lévy processes, and additionally one for its variation.

Proposition 7. *(Ken-Iti, 1999, Theorem 21.3) Let $X = (X_t)_{t\geq 0}$ be a Lévy process with Lévy measure ν. Then*

a) X has finite activity if $\nu(\mathbb{R}) < \infty$.

b) X has infinite activity if $\nu(\mathbb{R}) = \infty$.

As mentioned in Carr and Wu (2004), VG and NIG are examples of infinite activity processes.

Proposition 8. *(Ken-Iti, 1999, Theorem 21.9) Let $X = (X_t)_{t\geq 0}$ be a Lévy with characteristic triplet (σ, γ, ν). Then*

a) Almost all paths of X have finite variation if $\sigma = 0$ and $\int_{|x|\leq 1} |x|\nu(\mathrm{d}x) < \infty$

b) Almost all paths of X have infinite variation if $\sigma \neq 0$ and $\int_{|x|\leq 1} |x|\nu(\mathrm{d}x) < \infty$.

When it comes to the processes of our consideration, VG is of finite variation, while NIG is an infinite variation process, (Cont and Tankov, 2006, Table 4.5).

3.3 Subordination

One of the ways of constructing new Lévy processes is subordination, where we "time change" a Lévy process with the independent subordinator (another increasing, non-negative Lévy process), i.e. we substitute the physical (calendar) time with the so-called intrinsic (operational) time. As mentioned in (Wang, 2009, p. 11), the idea of time change has strong economic intuitions. The financial market does not evolve identically every day – the trading volume is not uniform during the day and the trading activities vary a lot over time. Intuitively, one could think of the original clock as the calendar time and a random clock as the business time. A more active business day implies a faster business clock. Thus, the concept of business time is used to distinguish from the calendar time and describe the trading activity evolution.

For example, in our case of Brownian subordination, we take Brownian motion $W = (W(t))_{t\geq 0}$ with drift θ and volatility σ to time change it with the independent non-negative increasing stochastic process $T = T(t)_{t\geq 0}$ defined on the same probability space and adapted to the same filtration. Next, we define a new process as $Z = \{Z(t) = W(T(t))\}_{t\geq 0}$ which is said to be subordinated to W by the intrinsic time process T.

Now, let the characteristic triplet of T be $(0, a, \rho)$ and (σ, γ, ν) of the Brownian motion W, then the Lévy triplet of the subordinated process Z is given as $(\sigma^Z, \gamma,^Z, \nu^Z)$ with

$$\sigma^Z = a\sigma, \tag{3.5}$$

$$\gamma^Z = a\gamma + \int_0^\infty \rho(\mathrm{d}s) \int_{|w|\leq 1} p_s^W(\mathrm{d}w), \tag{3.6}$$

$$\nu^Z(B) = a\nu(B) + \int_0^\infty p_s^W(B)\rho(\mathrm{d}s), \forall B \in \mathcal{B}(\mathbb{R}), \tag{3.7}$$

where p_t^W is the probability density of W_t, (Krichene, 2005, p. 10).

Some of the most popular subordinated processes are the Variance Gamma (VG) process and the Normal Inverse Gaussian (NIG) process. Specifically, VG and NIG processes are processes where the Brownian motion is run under a random Gamma and an Inverse Gamma business clock, respectively. Both of these models are pure jump processes, i.e. they have no Brownian motion component. Consequently, the first parameter in their Lévy triplets is 0. These are the processes we will base our models on.

3.4 Exponential Lévy models

The Black-Scholes model

$$\frac{\mathrm{d}S_t}{S_t} = \mu \,\mathrm{d}t + \sigma \,\mathrm{d}W_t \tag{3.8}$$

can also be written in the equivalent exponential form

$$S_t = S_0 e^{(\mu - \sigma^2/2)t + \sigma W_t}. \tag{3.9}$$

As mentioned in Tankov (2011), this motivates two ways of constructing an exponential Lévy model beginning with a (one-dimensional) Lévy process X. First one would be using the stochastic differential equation:

$$\frac{\mathrm{d}S_t^{sde}}{S_{t^-}^{sde}} = r \,\mathrm{d}t + \mathrm{d}X_t \ , \tag{3.10}$$

and second one using the ordinary exponential

$$S_t^{exp} = S_0^{exp} e^{rt + X_t}, \tag{3.11}$$

where we explicitly included the interest rate r (assumed constant) in the formulae, to simplify notation later on. The subscript *sde* stands for stochastic differential equation and *exp* for exponential, in order to emphasize that S^{sde} and S^{exp} are different processes. Often it is convenient to discount the price processes with $B(t,T) = e^{-r(T-t)}$ for a fixed maturity T, i.e. we would then consider $\hat{S}_t := \frac{S_t}{B(t,T)} = e^{r(T-t)} S_t$ which implies an *sde* form of

$$\frac{\mathrm{d}\hat{S}_t}{\hat{S}_{t^-}} = \mathrm{d}X_t \ , \tag{3.12}$$

or an exponential form of

$$\hat{S}_t = \hat{S}_0 e^{X_t}. \tag{3.13}$$

Although *sde* and *exp* forms do not represent the same stochastic processes, Goll and Kallsen (2000) showed that the two approaches are equivalent; if $Z > 0$ is the stochastic exponential of a Lévy process it is then also the ordinary exponential of another Lévy process and vice versa. Let us explain this on an example of the Brownian motion case. Let $X_t = \sigma W_t$, then the ordinary and stochastic exponential are respectively given by

$$Y_t = e^{\sigma W_t}, \ Z_t = e^{\sigma W_t - \frac{\sigma^2 t}{2}}. \tag{3.14}$$

Notice that the stochastic exponential of a Brownian motion is also the ordinary exponential of another Lévy process L, in this case that is $L_t = \sigma W_t - \frac{\sigma^2 t}{2}$. This also holds for Lévy processes with jumps. More on that and on all of the aforementioned can be found in (Cont and Tankov, 2006, p. 286).

3.5 Market incompleteness

It is well-known in financial mathematics that the Black-Scholes model is both arbitrage-free and complete. An arbitrage opportunity means almost surely making money with no initial investment. On the other hand, a complete market is one where every contingent claim can be replicated. Another important property to consider in financial markets is the equivalent martingale measures (EMM). Each EMM gives a possible self-consistent pricing rule for contingent claims. In complete markets the EMM is unique (Second Fundamental Theorem of Asset Pricing, see e.g. (Cont and Tankov, 2006, p. 300), or Korn and Desmettre (2014) for discrete time models), whereas in incomplete markets it exists but is (generally) not unique.

Using a particular change of measure – the Esscher transform, one can show that exponential-Lévy models are arbitrage-free: an equivalent martingale measure always exists (see e.g. Cont and Tankov (2006)). They are also (except for the Brownian motion and the Poisson process cases) incomplete market models: the class of EMMs is infinite. Although at first this may seem as a drawback, we refer you to (Cont and Tankov, 2006, p. 316), where its advantages are discussed.

For us here, it is important to bring attention to market incompleteness in a sense that when one deals with such market models an EMM has to be chosen wisely. In the following, we explain one such way that will be used in the NIG model.

3.5.1 Esscher transform

As discussed in (Cont and Tankov, 2006, p. 310), in the Black-Scholes model, the EMM could be obtained by changing the drift, but that cannot be done in models with jumps without a Gaussian component. However, one can obtain a great variety of EMMs by altering the distribution of jumps, e.g. by using the Esscher transform explained in the following.

For $\theta \in \mathbb{R}$, if $f(x)$ is a probability density, the Esscher transform is defined as

$$f(x;\theta) = \frac{e^{\theta x} f(x)}{\int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x}, \tag{3.15}$$

or more generally if μ is a probability measure, the Esscher transform of μ is a new probability measure $E_\theta(\mu)$ which has density

$$\frac{e^{\theta x}}{\int_{-\infty}^{\infty} e^{\theta x}\,\mathrm{d}\mu(x)}. \tag{3.16}$$

Consider a real number θ and a Lévy process X with the characteristic triplet (σ, γ, ν) s.t. the Lévy measure ν satisfies $\int_{|x|\geq 1} e^{\theta x}\nu(\mathrm{d}x) < \infty$. Using $\phi(x) = \theta x$ for a measure transformation an equivalent probability is obtained[4], under which the process X becomes a Lévy process with the characteristic triplet $(0, \gamma + \int_{-1}^{1} x(e^{\theta x} - 1)\nu(\mathrm{d}x), e^{\theta x}\nu(\mathrm{d}x))$. The measure-change corresponding to the Radon-Nikodym derivative is given by

$$\frac{\mathrm{d}\mathbb{Q}|_{\mathcal{F}_t}}{\mathrm{d}\mathbb{P}|_{\mathcal{F}_t}} = \frac{e^{\theta X_t}}{\mathbb{E}[e^{\theta X_t}]} = \exp(\theta X_t + \gamma(\theta)t), \tag{3.17}$$

[4]By applying (Cont and Tankov, 2006, Proposition 9.8).

where $\gamma(\theta) = -\ln \mathbb{E}[\exp(\theta X_1)]$ represents the log of the moment generating function of X_1 which, up to the change of variable $\theta \leftrightarrow i\theta$, is given by the characteristic exponent of the Lévy process X, (Cont and Tankov, 2006, p. 310).

For further reading and examples also see e.g. Gerber et al. (1993).

Chapter 4

Variance Gamma Process

4.1 Gamma process

As its own name already suggests, Gamma process is the subordinator of Varinace Gamma. Therefore, let us first define the subordinator and name some of its characteristics.

Definition 9. *The density function of the Gamma distribution* $Gamma(a, b)$ *with parameters* $a, b > 0$ *is given as*

$$f_{Gamma}(x; a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} e^{-xb}, \ x > 0, \tag{4.1}$$

where $\Gamma(\cdot)$ *stands for the Gamma function.*[5]

The characteristic function is given by

$$\phi_{Gamma}(u; a, b) = a \ln \left(\frac{1}{1 - iu/b} \right). \tag{4.2}$$

One can also check its infinite divisibility.
Now we define the Gamma process

$$\gamma = (\gamma(t; a, b))_{t \geq 0} \,, \tag{4.3}$$

with parameters $a, b > 0$, as a stochastic process starting at zero and having stationary and independent Gamma distributed increments. In fact, $\gamma(t; a, b)$ follows a $Gamma(at, b)$ distribution. The Lévy triplet of the Gamma process (see e.g. (Schoutens, 2003, p. 52)) is given as

$$(0, a(1 - e^{-b})b^{-1}, a e^{-bx} x^{-1} \mathbb{1}_{x>0} \, \mathrm{d}x), \tag{4.4}$$

which means that it has no Brownian motion component. As Gamma process follows the $Gamma(at, b)$ distribution, using Eq. (4.1) and some well-known properties of Gamma functions, one can straight-forwardly derive mean and variance of the Gamma process, i.e. one gets $\frac{at}{b}$ and $\frac{at}{b^2}$ for the mean and variance, respectively.

[5]The Gamma function is defined as $\Gamma(a) := \int_0^\infty t^{a-1} e^{-t} \, \mathrm{d}t$.

4.2 Variance Gamma process

As already mentioned, the Variance Gamma (VG) process is a Brownian motion (with drift θ and volatility σ) time-changed by a Gamma process.

Let $W = (W(t))_{t\geq 0}$ be a standard Brownian motion, then the process $b(t;\theta,\sigma)$ given as

$$b(t;\theta,\sigma) = \theta t + \sigma W(t) \tag{4.5}$$

is a Brownian motion with drift $\theta \in \mathbb{R}$ and volatility $\sigma > 0$. Now, let $\gamma = \big(\gamma(t; a = 1/\nu, b = 1/\nu)\big)_{t\geq 0}$ be a Gamma process[6], then we define a VG process X^{VG} as follows

$$X^{VG}(t;\sigma,\theta,\nu) := b(\gamma(t;1/\nu,1/\nu);\theta,\sigma), \tag{4.6}$$

or using Eq. (4.5), we can write

$$X^{VG}(t;\sigma,\theta,\nu) = \theta\gamma(t;1/\nu,1/\nu) + \sigma W(\gamma(t;1/\nu,1/\nu)). \tag{4.7}$$

As one can see, VG depends on 3 parameters - drift θ and volatility σ of the Brownian motion, and variance ν of the Gamma process for $t = 1$. We will give more meaning to them when we take a look at the central moments and the Lévy measure representations.

As mentioned in Madan et al. (1998), VG is a finite variation process, and a pure jump process (has no Brownian component) that accounts for high activity by having an infinite number of jumps in any interval of time. Additionally, when modelling the stock behaviour by VG, the level of activity may loosely be measured by the volume or number of transactions, or the associated number of price changes.

As the VG process is of finite variation, it can be introduced in another way as well – as the difference of two increasing processes:

- the first of which accounts for the price increases,
- and the second explaining the price decreases.

In this case, those two processes are Gamma processes. More on that in Madan and Seneta (1990), Madan et al. (1998).

Characteristic function

The random variable of a VG process at unit time follows a 3-parameter VG probability law with characteristic function in the form:

$$\phi_{X^{VG}(t)}(u;\sigma,\theta,\nu) = \frac{1}{\nu}\ln\left(\frac{1}{1 - i\theta\nu u + \frac{1}{2}u^2\sigma^2\nu}\right). \tag{4.8}$$

This distribution is infinitely divisible and the VG process thus has independent and stationary increments for which the increment $X^{VG}_{t+s} - X^{VG}_{s}$ follows a $VG(\sigma\sqrt{t}, \nu/t, t\theta)$ law, (Wang, 2009, p. 11).

[6] whose mean is $\frac{at}{b} = t$ and variance $\frac{at}{b^2} = t\nu$

The central moments

For the sake of simplicity and easier reading, in the following moment formulae we will denote our VG process $X^{VG}(t)$ as $X(t)$. The first 4 central moments of the VG distribution over an interval of length t are given as (Madan et al. (1998)):

$$\mathbb{E}[X(t)] = \theta t, \tag{4.9}$$

$$\mathbb{E}[(X(t) - \mathbb{E}[X(t)])^2] = (\theta^2\nu + \sigma^2)t, \tag{4.10}$$

$$\mathbb{E}[(X(t) - \mathbb{E}[X(t)])^3] = (2\theta^3\nu^2 + 3\sigma^2\theta\nu)t, \tag{4.11}$$

$$\begin{aligned}\mathbb{E}[(X(t) - \mathbb{E}[X(t)])^4] = {} & (3\sigma^4\nu + 12\sigma^2\theta^2\nu^2 + 6\theta^4\nu^3)t \\ & + (3\sigma^4 + 6\sigma^2\theta^2\nu + 3\theta^4\nu^2)t^2.\end{aligned} \tag{4.12}$$

From here, we easily see that the mean of VG is independent of σ and ν. For further analysis, also note that third and fourth moment, when centralized, give skewness (a measure of asymmetry) and kurtosis (a measure of "tailedness"), respectively.

The popularity of the VG process lies in its flexibility of handling the skewness and excess kurtosis exhibited from the historical data of stock prices. While the parameter σ still provides similar behaviour to the volatility parameter as in the Black-Scholes world (Fig. 4.1), the other parameters add quite some flexibility to the distribution. However, the parameters only indirectly reflect the skewness and kurtosis of the return distribution.

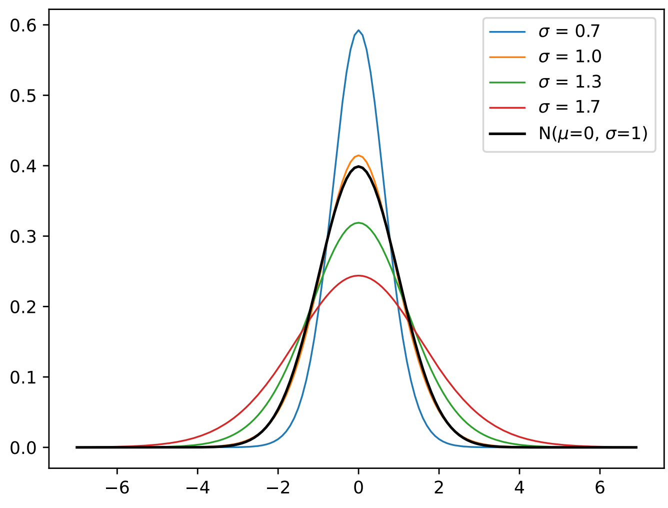
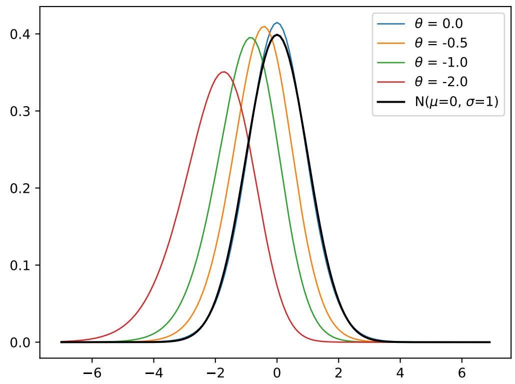
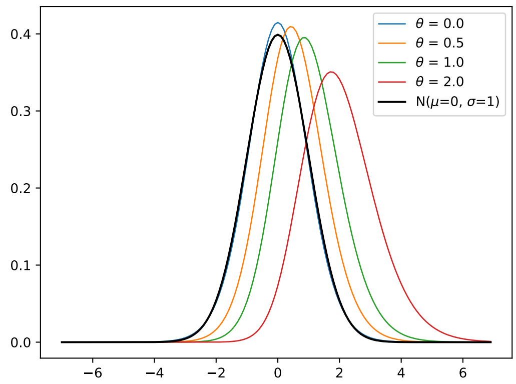


Figure 4.1: Pdfs of VG with different values of σ

From the mean equation we observe that $\theta = 0$ implies there is no skewness, and that the skewness of the distribution has the same sign as θ (see Fig. 4.2). Therefore, a negative skewness, which is generally what can be observed in the market, is the same as requiring a negative drift in the time changed Brownian motion.



(a) Negative values of θ



(b) Positive values of θ

Figure 4.2: Pdfs of VG with different negative (positive) values of θ, i.e. $\theta \in \{0, \pm 0.05, \pm 0.1\}$, compared to the pdf of the standard normal distribution

The kurtosis equation divided by the square of variance equation, for $\theta = 0$, gives

$$3(1+\nu), \tag{4.13}$$

which makes ν the percentage excess kurtosis over 3 – the kurtosis of the normal distribution (see Fig. 4.3). In general, when not considering a single interval, the kurtosis divided by the square of the variance, for $\theta = 0$, yields

$$3\left(1+\frac{\nu}{t}\right). \tag{4.14}$$

Now, when t is increasing, the kurtosis is approaching 3. This VG behaviour is consistent with empirical evidence that fat tails tend to express daily returns while monthly returns tend to be normally distributed. (Fiorani, 2004, p. 33).

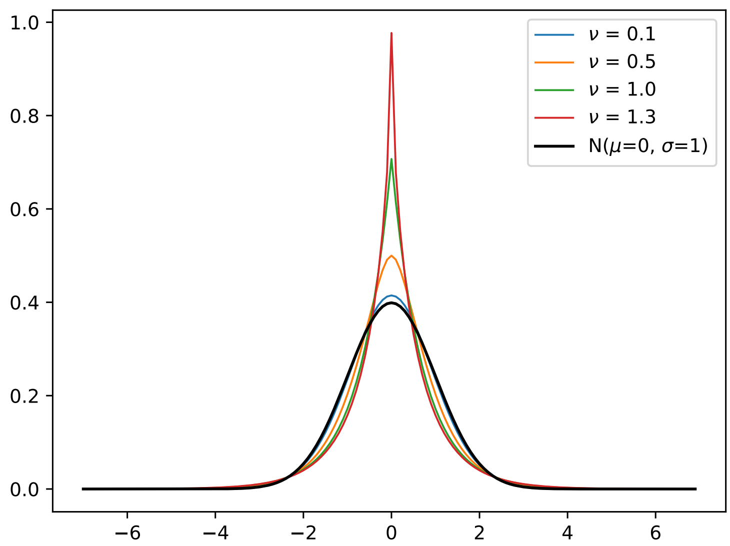


Figure 4.3: Pdfs of VG with different values of ν, $\nu \in \{0.1, 0.5, 1.0, 1.3\}$, compared to the pdf of the standard normal distribution

In Fig. 4.1, Fig. 4.2 and Fig. 4.3 we see probability density functions of the VG process for the following parameters (unless otherwise stated on the plot) $\sigma = 1$, $\theta = 0$, $\nu = 1$, compared to the probability density function of the standard normal distribution.

The Lévy measure

The Lévy measure for the VG process has three representations which we obtain from the representation of VG:

- as the difference of two Gamma processes,
- as the time changed Brownian motion,
- and finally in terms of a symmetric VG process subject to a measure change induced by a constant relative risk aversion utility function as in Madan and Milne (1991) (paper discussed in Section 4.4).

Here, we will introduce only the latter two, as that suffices for our needs in the option pricing. All three representations with its properties can be found in Madan et al. (1998).

In the case of representing the VG process X as a time changed Brownian motion, the Lévy measure has the following form

$$k_X(x)\,\mathrm{d}x = \frac{\exp(\frac{\theta x}{\sigma^2})}{\nu|x|}\exp\left(-\frac{\sqrt{\frac{2}{\nu}+\frac{\theta^2}{\sigma^2}}}{\sigma}|x|\right)\mathrm{d}x\,. \tag{4.15}$$

For $\theta = 0$, the Lévy measure is symmetric around zero. From Eq. (4.15) it follows that for $\theta < 0$, negative values of x receive a higher relative probability than positive ones, which means that a negative value of θ causes a negative skewness. One can also observe that a large value of ν lowers the exponential decay rate of the Lévy measure symmetrically around zero. This increases the probability of large jumps, thereby raising tail probabilities and kurtosis. Thus, the VG process can be expected to flatten the volatility smiles at the low end of the maturity spectrum, Madan et al. (1998). These observations about θ and ν confirm the results obtained by direct computation of the moments of the VG distribution.

Following Madan and Milne (1991), the risk neutral VG process for stock prices can be derived from a Lucas-type general equilibrium economy (Lucas Jr (1978)) where the representative agent has a constant relative risk aversion utility function, with a relative risk aversion ζ and in which the process followed by the logarithm of the stock price is a symmetric VG, i.e. VG with $\theta = 0$ and $\sigma = 1$. In this case the risk neutral Lévy measure is given by[7]

$$k_X(x)\,\mathrm{d}x = \frac{\exp(-\zeta x)}{\nu|x|}\exp\left(-\frac{\sqrt{2}}{s\sqrt{\nu}}|x|\right). \tag{4.16}$$

[7]The measure change function $\exp(-\zeta x)$ is the ratio of marginal utilities for a representative investor holding stock with marginal utility function $S^{-\zeta}$. The ratio of marginal utilities, for a jump in the log price of x, is $(S\exp(x))^{-\zeta}/S^{-\zeta} = \exp(-\zeta x)$, Madan et al. (1998).

When setting the parameters of this equation as

$$\zeta = -\frac{\theta}{\sigma^2}, \tag{4.17}$$

$$s = \frac{\sigma}{\sqrt{1 + (\frac{\theta}{\sigma})^2 \frac{\nu}{2}}}, \tag{4.18}$$

we obtain the previous representation of the Lévy measure as in Eq. (4.15). Therefore, the two representations are in agreement.

4.3 Option pricing

We will firstly explain the idea behind the paper of Madan and Milne (1991) (Section 4.4) which is then further extended in Madan et al. (1998) whose option pricing formula we aim to use (Section 4.5).

Madan and Milne (1991) are using the so-called symmetric VG, i.e. $\theta = 0$, $\sigma = 1$ in Eq. (4.5), whereas Madan et al. (1998) are using the VG exactly as we previously introduced it in Eq. (4.7).

4.4 Symmetric VG and EMM

Using the cost of a hedging strategy is not possible for the pricing of European options for a pure jump driving uncertainty. A self-financing continuous trading strategy in the underlying asset and a riskless bond that replicates the payoff of the option does not exist, as shown in Naik and Lee (1990). They also obtain a necessarily incomplete markets equilibrium value for the option by solving a one-individual equilibrium model, employing a constant relative risk aversion utility function for the individual. This approach is used in the paper to derive the exact equilibrium price interpretation – relation of constant relative risk aversion of a representative agent and the associated measure change density process

$$\lambda(t) = \mathbb{E}_{\mathbb{P}}[d\mathbb{Q}/d\mathbb{P}|\mathcal{F}_t], \tag{4.19}$$

where measure $\mathbb{Q}$ is an equivalent martingale measure, i.e. the following holds

a) $e^{-rt}S(t)$ is a $\mathbb{Q}$ martingale (where r is a constant interest rate), and

b) the price $C(t)$ at time t of a traded call option on the stock with maturity T and strike price K is

$$C(t) = \mathbb{E}_{\mathbb{Q}}[e^{-r(T-t)}(S(T) - K)^+|\mathcal{F}_t]. \tag{4.20}$$

The option pricing formula, derived as a discounted expected value under the identified change of measure, can then be viewed as a first-order approximation of an incomplete markets equilibrium price.

As we discussed in Section 3.5, in complete markets the equivalent martingale measure is unique, whereas in incomplete ones it exists but is (generally) not unique.

Duffie (1988) derives the result in an equilibrium with initial and final consumption for square-integrable semimartingale price processes and a predictable discounting process that pays at maturity and has a price process strictly bounded above and below. Back (1991) allows intermediate consumption and obtains the result for square-integrable semimartingale price processes but uses a continuous discounting process.

Employing these results, the change of measure density process $\lambda(t)$ is first parameterized using results from Jacod and Shiryaev (2013). As most of the jumps for proposed process are near the origin, they construct a one-parameter approximation by a Taylor series expansion to the change in the jump compensator induced by the measure change. The single parameter is then identified using the martingale condition for the discounted asset price process, as it is in the modern versions of the Black-Scholes theory, i.e. the equilibrium martingale condition for discounted asset prices was used to determine the measure change, and the option price was then evaluated by integration. The integration was performed numerically (only for small values of t/ν, for large values the distribution of $N(t)/\sqrt{(t)}$ is approximated by a standard normal variate), where in Madan et al. (1998) for the general VG process they derive a closed-form option pricing formula.

Madan and Milne (1991) observed that symmetric VG option values are typically higher than Black-Scholes values, with the percentage underpricing by Black-Scholes being higher for out of the money options, with long maturity, high mean rates, low variance rates, and high kurtosis.

4.5 VG and option pricing

As we have just discussed in the previous section, Madan and Milne (1991) considered a symmetric VG – a time change of the Brownian motion without drift by a Gamma process. The resulting risk neutral process in the paper of Madan et al. (1998) is similar to the one in Madan and Milne (1991), but with the more general VG process – VG given as in Eq. (4.7), with the drift in the time changed Brownian motion being negative for positive risk aversion. This paper theoretically extends Madan and Milne (1991) by providing closed forms for the return density and the prices of European options on the stock.

4.5.1 Stock price dynamics

In this section the statistical and risk neutral dynamics of the stock price under VG process will be proposed.

Consider a continuous time economy, over the interval $[0, \Upsilon]$, in which a stock, a money market account, and options on the stock are traded for all strikes and maturities $0 < T < \Upsilon$. Suppose a constant continuously compounded interest rate r with money market account value of e^{rt}, stock prices of $S(t)$ and European call option prices of $C(t; K, T)$ for strike K and maturity $T > t$, at time t.

Statistical process

Now we introduce the statistical stock price dynamics under the VG process by replacing the role of Brownian motion in the Black-Scholes GBM model by the VG process.

The statistical process for the stock price is given as

$$S(t) = S(0)\exp(mt + X(t;\sigma_S,\nu_S,\theta_S) + \omega_S t), \tag{4.21}$$

where the subscript S on the VG parameters is indicating the statistical parameters, m is the mean rate of return on the stock under the statistical probability measure, and ω_S is given as[8]

$$\omega_S = \frac{1}{\nu_S}\ln(1 - \theta_S\nu_S - \sigma_S^2\nu_S/2). \tag{4.22}$$

Risk neutral process

Under the risk neutral process, stock prices discounted at the risk free interest rate are martingales and so the expected return on the stock under the risk neutral probability measure is the continuously compounded risk free interest rate r. The risk neutral process is now given as

$$S(t) = S(0)\exp(rt + X(t;\sigma_{RN},\nu_{RN},\theta_{RN}) + \omega_{RN} t), \tag{4.23}$$

where now the subscripts RN stress the risk free parameters, and using the same condition as before, ω_{RN} is given as

$$\omega_{RN} = \frac{1}{\nu_{RN}}\ln(1 - \theta_{RN}\nu_{RN} - \sigma_{RN}^2\nu_{RN}/2). \tag{4.24}$$

4.5.2 Density and option price formulae

In this section we give closed forms for the return density and the prices of European options on the stock. When doing econometric estimation of the statistical and risk neutral processes by maximum likelihood (ML) methods, those closed forms tend to be useful.

The density of the log stock price relative over an interval of length t is, when conditioned on the realization of the Gamma time change, a normal density function. The unconditioned density is obtained by integrating out the Gamma variate and the result is given in terms of the modified Bessel functions of the second kind. The density for the log price relative is given within the following Theorem.

[8] The value of ω_S is determined as a non arbitrage condition, by evaluating the characteristic function for $X(t)$ at $u = 1/i$, so that $\mathbb{E}[S(t)] = S(0)\exp(mt)$, or equivalently $\mathbb{E}[\exp(X(t))] = \exp(-\omega_S t)$.

Theorem 2. *([Madan et al.], [1998], p. 87) The density for the log price relative $z = ln(S(t)/S(0))$ for prices following the VG process dynamics of Eq. (4.21) is given by*

$$f^{VG}(z;\sigma,\theta,\nu) = \frac{2\exp\left(\frac{\theta}{\sigma^2}x\right)}{\nu^{\frac{1}{\nu}}\sqrt{2\pi}\sigma\Gamma\left(\frac{t}{\nu}\right)} \cdot \left(\frac{x^2}{2\frac{\sigma^2}{\nu}+\theta^2}\right)^{\frac{1}{2\nu}-\frac{1}{4}} \cdot K_{\frac{1}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{x^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}\right), \tag{4.25}$$

where K_α is the modified Bessel function of the second kind [9] *of order α* [10]*, and*

$$x = z - mt - \frac{t}{\nu_S}\ln(1-\theta_S\nu_S - \sigma_S^2\nu_S/2).$$

The used VG parameters are the statistical ones.

The price of a European call option $C(S(0), K, T)$, where K is the strike and T is the maturity[11], can be written with the known expression

$$C(S(0), K, T) = e^{-rT}\mathbb{E}[max(S(T) - K, 0)], \tag{4.26}$$

where the expectation is taken under the risk neutral process defined in Eq. (4.23). The first step of the evaluation of the option price in Eq. (4.26) is conditioning on a knowledge of the random time change g that has an independent Gamma distribution. Conditional on g, X(t) is normally distributed and the option value is given by a Black-Scholes type formula. The European option price for VG risk neutral dynamics is then obtained by integrating this conditional Black-Scholes formula over g w.r.t. the Gamma density. This was also the procedure in Madan and Milne (1991). In this paper, Madan et al. (1998), an analytical reduction is obtained – in terms of special mathematical functions. Finally, the following Theorem gives the option pricing formula that we will use.

Theorem 3. *(Madan et al., 1998, Theorem 2) The European call option price on a stock, when the risk neutral dynamics of the stock price is given by the VG process by Eq. (4.23), is (for risk neutral parameters σ, ν, θ)*

$$\begin{aligned} c(S(0); K, T) = S(0) \cdot \Psi\left(d\sqrt{\frac{1-c_1}{\nu}}, (\alpha+s)\sqrt{\frac{\nu}{1-c_1}}, \frac{T}{\nu}\right) \\ -K\exp(-rT) \cdot \Psi\left(d\sqrt{\frac{1-c_2}{\nu}}, \alpha s\sqrt{\frac{\nu}{1-c_2}}, \frac{T}{\nu}\right), \end{aligned} \tag{4.27}$$

where

$$d = \frac{1}{s}\left[\ln\left(\frac{S(0)}{K}\right) + rT + \frac{T}{\nu}\ln\left(\frac{1-c_1}{1-c_2}\right)\right], \tag{4.28}$$

[9]Sometimes also called the Basset function, the modified Bessel function of the third kind, the modified Hankel function or the Macdonald function.

[10]For $\alpha \in \mathbb{R}$, the function K_α solves the differential equation $x^2y'' + xy' - (x^2+\alpha^2)y = 0$.

[11]In Madan et al. (1998) the maturity is denoted with t.

$\alpha = \zeta s$, ζ, s are as defined in Eq. (4.17) and Eq. (4.18),

$$c_1 = \frac{\nu(\alpha+s)^2}{2}, \quad c_2 = \frac{\nu\alpha^2}{2}, \tag{4.29}$$

and the function $\Psi(a,b,\gamma)$ is given as

$$\Psi(a,b,\gamma) = \int_0^{+\infty} \Phi\left(\frac{a}{\sqrt{u}} + b\sqrt{u}\right) \frac{u^{\gamma-1}e^{-u}}{\Gamma(\gamma)} \,\mathrm{d}u, \tag{4.30}$$

or in its closed form in terms of the modified Bessel function of the second kind and the degenerate hypergeometric function of two variables as[12]

$$\begin{aligned}
\Psi(a,b,\gamma) = \; & \frac{c^{\gamma+\frac{1}{2}}\, e^{sign(a)c}(1+u)^{\gamma}}{\sqrt{2\pi}\Gamma(\gamma)\gamma} \cdot \\
& \cdot K_{\gamma+\frac{1}{2}}(c)\, \phi\Big(\gamma, 1-\gamma, 1+\gamma; \frac{1+u}{2}, -sign(a)c(a+u)\Big) \\
& - sign(a)\frac{c^{\gamma+\frac{1}{2}}\, e^{sign(a)c}(1+u)^{1+\gamma}}{\sqrt{2\pi}\Gamma(\gamma)(1+\gamma)} \cdot \\
& \cdot K_{\gamma-\frac{1}{2}}(c)\, \phi\Big(1+\gamma, 1-\gamma, 2+\gamma; \frac{1+u}{2}, -sign(a)c(a+u)\Big) \\
& + sign(a)\frac{c^{\gamma+\frac{1}{2}} \cdot e^{sign(a)c}(1+u)^{\gamma}}{\sqrt{2\pi}\Gamma(\gamma)\gamma} \cdot \\
& \cdot K_{\gamma-\frac{1}{2}}(c)\, \phi\Big(\gamma, 1-\gamma, 1+\gamma; \frac{1+u}{2}, -sign(a)c(a+u)\Big),
\end{aligned} \tag{4.31}$$

where $c = |a|\sqrt{2+b^2}$, $u = \frac{b}{\sqrt{2+b^2}}$, and

$$\phi(\alpha,\beta,\gamma;x,y) = \frac{\Gamma(\gamma)}{\Gamma(\alpha)\Gamma(\gamma-\alpha)} \int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-ux)^{-\beta}\exp(uy)\,\mathrm{d}u.$$

The call price in Eq. (4.27) has a form similar to the Black-Scholes formula, having the stock price multiplied by a probability subtracted by the present value of the strike price multiplied by a second probability element. One can prove that the second probability element is the risk neutral probability that $S(t)$ exceeds K. Now, the first probability element also gives the probability that $S(t) > K$ using the density obtained on normalizing the product of stock price with the risk neutral density of the stock price.

As usual, put prices can be determined by using put-call parity.

The application of the call option pricing formula Eq. (4.27) on real data can be found in Section 5.4.1.

[12]The ϕ in Eq. (4.31) is denoted as Φ in Madan et al. (1998), but we use Φ for a cdf of a standard normal variable. For more details on this closed form and its derivation see Appendix of Madan et al. (1998).

Chapter 5

Normal Inverse Gaussian Process

5.1 Inverse Gaussian process

Let $T(a, b)$ be a random time describing the first time a standard Brownian motion with drift $b > 0$, $(W_t + bt)_{t\geq 0}$, reaches the positive level $a > 0$. Such random time follows the Inverse Gaussian law $IG(a, b)$ which has a characteristic function

$$\phi_{IG}(u; a, b) = e^{-a(\sqrt{-2iu+b^2}-b)}. \tag{5.1}$$

Similarly to the case of the Gamma process, IG distribution is also infinitely divisible. The density function of $IG(a, b)$ is given as

$$f_{IG}(x; a, b) = \frac{a}{\sqrt{2\pi}} e^{ab} x^{-3/2} e^{-\frac{1}{2}(a^2x^{-1}+b^2x)},\ x > 0. \tag{5.2}$$

We define the IG process

$$X^{IG} = (X^{IG(a,b)_t})_{t\geq 0} \tag{5.3}$$

with parameters $a, b > 0$, as the process starting at zero with independent and stationary increments such that

$$\begin{aligned} \mathbb{E}[\exp(iuX_t^{IG})] &= \phi_{IG}(u; at, b) \\ &= e^{-at(\sqrt{-2iu+b^2}-b)}. \end{aligned}$$

Its Lévy triplet is given as

$$\left(0, \frac{a}{b}(2\Phi(b) - 1), (2\pi)^{-1/2} a x^{-3/2} e^{-\frac{1}{2}b^2x}\, \mathbb{1}_{x>0}\ \mathrm{d}x\right), \tag{5.4}$$

where Φ is the normal distribution function, Schoutens (2003). The mean and variance of the IG are $\frac{a}{b}$ and $\frac{a}{b^3}$, respectively.

5.2 Normal Inverse Gaussian process

Let $W = (W(t))_{t\geq 0}$ be a Brownian motion with drift β and diffusion coefficient 1, and let $IG(t) = IG(t; \delta, \gamma)$ be an IG process, independent of W, with parameters δ

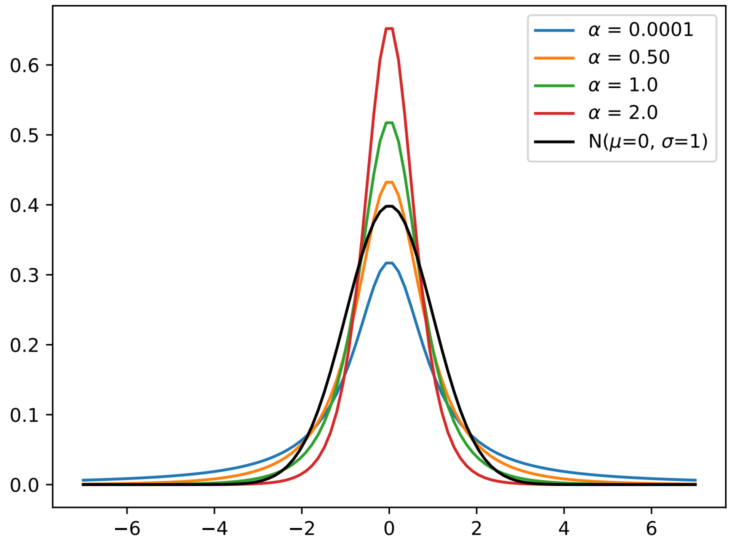


Figure 5.1: Pdfs of NIG with different values of α, i.e. $\alpha \in \{0.0001, 0.5, 1, 2\}$, compared to the pdf of the standard normal distribution

and $\gamma = \sqrt{\alpha^2 - \beta^2}$, then the NIG process can be defined as

$$X^{NIG}(t; \alpha, \beta, \mu, \delta) = W(IG(t; \delta, \gamma)) + \mu t, \tag{5.5}$$

where the IG process has a density function $f_{IG}(x; \delta, \gamma)$ given as in Eq. (5.2) and $\alpha > 0, \delta > 0, 0 < |\beta| < \alpha$. The interpretation of the variate $IG(t)$ can be understood as the first time to reach to the level δt of a Brownian motion with drift $\gamma = \sqrt{\alpha^2 - \beta^2}$ and diffusion coefficient 1, Barndorff-Nielsen (1997). In other words, Eq. (5.5) represents the NIG process as a subordination of the Brownian motion by the IG process.

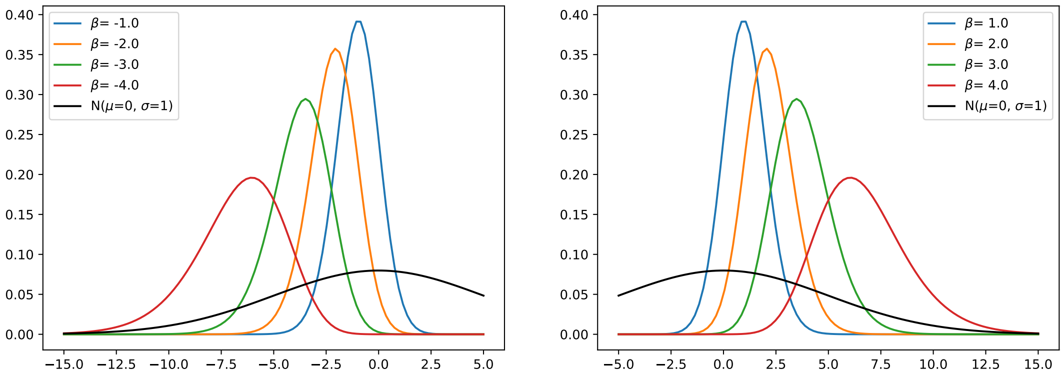
The parameter α is a steepness parameter - it has an effect on the tail heaviness of the distribution, i.e. for a larger value of α we get lighter tails while a smaller value implies heavier tails, see Fig. 5.1. The parameter β is an (a)symmetry parameter - for $\beta > 0$ we get a right-skewed density, $\beta < 0$ implies a left-skewed density, and for $\beta = 0$ we get a symmetric density, as one can observe in Fig. 5.2. On the other hand, δ controls the scale (Fig. 5.3) and μ is a location parameter, Rydberg (1997).

The density function of the NIG process is given as

$$f^{NIG}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} e^{\delta\gamma + \beta(x-\mu)} \frac{K_1(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{\sqrt{\delta^2 + (x-\mu)^2}}, \tag{5.6}$$

where $x \in \mathbb{R}$, $\alpha > 0, \delta > 0, 0 < |\beta| < \alpha$, $\gamma = \sqrt{\alpha^2 - \beta^2}$ and $K_1(x)$ is the modified Bessel function of the third kind with index 1. A distribution with the same name – the $NIG(\alpha, \beta, \gamma, \delta)$ distribution – is defined by such density. The moment generating function of Eq. (5.6) is given by

$$M^{NIG}(u) = \exp(\delta(\sqrt{\alpha^2 - \beta^2}) - \sqrt{\alpha^2 - (\beta + u)^2} + \mu u) \tag{5.7}$$



(a) NIG pdfs with negative values of β (b) NIG pdfs with positive values of β

Figure 5.2: Pdfs of NIG with different negative (positive) values of β, i.e. $\beta \in \{\pm 1, \pm 2, \pm 3, \pm 4\}$, compared to the pdf of the standard normal distribution

from which one can derive the following convolution feature

$$f^{NIG}(\alpha, \beta, \delta_1, \mu_1) * f^{NIG}(\alpha, \beta, \delta_2, \mu_2) = f^{NIG}(\alpha, \beta, \delta_1 + \delta_2, \mu_1 + \mu_2). \tag{5.8}$$

The first 4 central moments of the NIG process (for $t = 1$ or over an interval of length 1) are given as

$$\mathbb{E}[X^{NIG}(1)] = \mu + \delta \frac{\beta}{\sqrt{\alpha^2 - \beta^2}}, \tag{5.9}$$

$$\mathbb{E}[(X^{NIG}(1) - \mathbb{E}[X^{NIG}(1)])^2] = \delta \frac{\beta}{\sqrt{\alpha^2 - \beta^2}^3}, \tag{5.10}$$

$$\mathbb{E}[(X^{NIG}(1) - \mathbb{E}[X^{NIG}(1)])^3] = 3\delta \frac{\beta}{\alpha \sqrt{\delta \sqrt{\alpha^2 - \beta^2}}}, \tag{5.11}$$

$$\mathbb{E}[(X^{NIG}(1) - \mathbb{E}[X^{NIG}(1)])^4] = 3 + 3\left(a + 4\left(\frac{\beta}{\alpha}\right)^2\right) \frac{1}{\delta\sqrt{\alpha^2 - \beta^2}}. \tag{5.12}$$

It is also worth mentioning that the NIG distribution is the special case of the generalized hyperbolic (GH) distrbution (see Barndorff-Nielsen (1978)) for $\lambda = -1/2$ in the parametrization of

$$f^{GH}(x; \lambda, \alpha, \beta, \delta, \mu) = \frac{\zeta^\lambda}{\sqrt{2\pi}\alpha^{2\lambda-1}\delta^{2\lambda} K_\lambda(\zeta)} e^{\beta(x-\mu)} \nu(x)^{\lambda - 1/2} K_{\lambda-1/2}(\nu(x)), \tag{5.13}$$

where $\zeta = \delta\sqrt{\alpha^2 - \beta^2}$ and $\nu(x) = \alpha\sqrt{\delta^2 + (x-\mu)^2}$, Albrecher and Predota (2004).

Based on that, the pdf plots of the NIG distribution (Fig. 5.1, Fig. 5.2, Fig. 5.3) are made with a density function from the PYTHON SCIPY package for a GH random variable[13] with $\lambda = -1/2$, but also $a = \alpha\delta$ and $b = \beta\delta$, as they are using the 4$^{\text{th}}$ parametrization from (Prause et al., 1999, p. 2).

To double-check if this method is valid, we recreated plots of (Forouzanfar et al., 2008, Fig. 1), i.e. we used the same parameters $\alpha \in \{0.0001, 0.5, 1, 2\}$ (see Fig. 5.1)

[13]https://scipy.github.io/devdocs/tutorial/stats/continuous_genhyperbolic.html

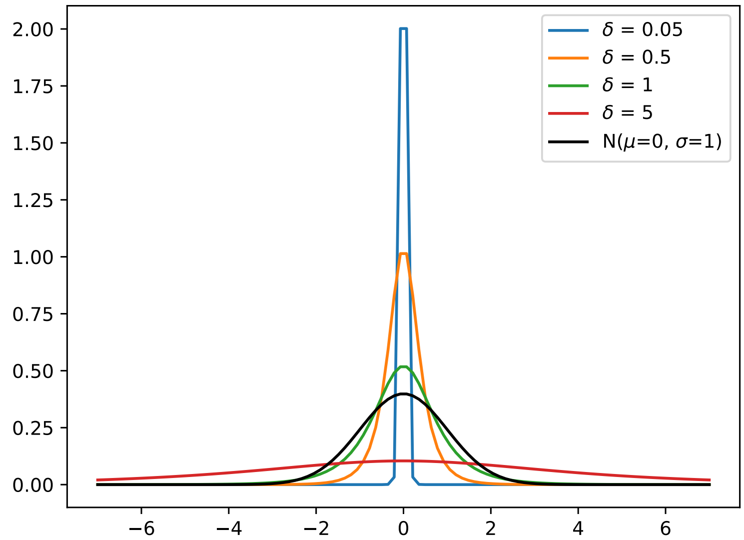


Figure 5.3: Pdfs of NIG with different values of δ, i.e. $\delta \in \{0.05, 0.5, 1, 5\}$, compared to the pdf of the standard normal distribution

and $\beta \in \{1, 2, 3, 4\}$ (see Fig. 5.2b), and they indeed match the ones in (Forouzanfar et al., 2008, Fig. 1).

There is also a package for an NIG random variable[14] which already has $\lambda = -1/2$ included, but the parametrization is still in the same way different to the one we introduced in Eq. (5.6), meaning $a = \alpha\delta$ and $b = \beta\delta$.

5.3 The NIG Lévy asset price model

Let $(S_t)_{t\geq 0}$ be the price of a nondividend-paying stock at time t. As we already mentioned the NIG is the special case of the generalized hyperbolic distribution, thus for the asset price model we will use the idea of Eberlein (2001), as it has been done in Albrecher and Predota (2004) and Jovan and Ahčan (2017).

Using the empirical facts on log-return distributions, the goal is to model asset prices such that log-returns of the model produce exactly a NIG distribution along certain-length time intervals, e.g. one trading day. As the NIG distribution is divisible, it generates a Lévy process $(Z_t)_{t\geq 0}$, in clearer words – a stochastic process with stationary and independent increments, such that $Z_0 = 0$ a.s. and Z_1 is NIG-distributed. From the convolution property in Eq. (5.8) it follows that the increments are NIG-distributed for arbitrary time intervals. Let now the stock price follow the dynamics of the process given as

$$\mathrm{d}S_t = S_t^-(dZ_t + e^{\Delta Z_t} - 1 - \Delta Z_t), \tag{5.14}$$

where $(Z_t)_{t\geq 0}$ denotes the NIG process, Z_t^- the left hand limit of the path at time t, and $\Delta Z_t = Z_t - Z_t^-$ the jump at time t. Then, one can show that the solution of the stochastic differential equation Eq. (5.14), (Eberlein (2001), Albrecher and Predota

[14] https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.norminvgauss.html

(2004)), is given by

$$S_t = S_0 \exp(Z_t), \tag{5.15}$$

from where it follows that the log-returns $ln(S_t/S_{t-1})$ are indeed NIG-distributed, and the $ln(S_t/S_0)$ are also NIG-distributed with the density given by

$$f_{*t}^{NIG}(x) = f^{NIG}(x; \alpha_S, \beta_S, \delta_S t, \mu_S t). \tag{5.16}$$

.

5.4 Esscher EMM and option pricing

As in the VG case, NIG modeling creates an incomplete market model which means that the equivalent martingale measure exists but is not unique. Here we will use the Esscher transform for deriving an EMM (which we introduced in Section 3.5.1), as also proposed in Albrecher and Predota (2004) for the NIG case. This approach is applicable, whenever the stochastic process $(Z_t)_{t\geq 0}$ has stationary and independent increments (see Eberlein and Keller (1995), Gerber et al. (1993)). As we have seen, the density of Z_t is given via $f_{*t}^{NIG}(x)$. Now, for a real number θ we will consider the Esscher transform

$$f_{*t,\theta}^{NIG}(x) = \frac{e^{\theta x} f_{*t}^{NIG}(x)}{\int_{-\infty}^{\infty} f_{*t}^{NIG}(y)\,\mathrm{d}y} = \frac{e^{\theta x}}{M^{NIG}(\theta)^t} f_{*t}^{NIG}(x) \ , \tag{5.17}$$

of the one-dimensional marginal distributions $f_{*t}^{NIG}(x)$ of $(Z_t)_{t\geq 0}$. For any Lévy process $(Z_t)_{t\geq 0}$ (on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P})$) it is now possible to define a locally equivalent probability measure $\mathbb{P}$ as

$$\mathrm{d}\mathbb{P}^\theta = \exp(\theta Z_t - t \log M^{NIG}(\theta))\,\mathrm{d}\mathbb{P} \ , \tag{5.18}$$

where $(Z_t^\theta)_{t\geq 0}$ defined on $(\Omega, \mathcal{F}, \mathbb{P}^\theta)$ is again a Lévy process and the one-dimensional marginal distributions of $(Z_t^\theta)_{t\geq 0}$ are the Esscher transforms of the corresponding marginals of $(Z_t)_{t\geq 0}$. $\mathbb{P}^\theta$ is called the Esscher equivalent measure. The parameter θ can be chosen such that the discounted stock price process $(e^{-rt}S_t)_{t\geq 0}$ is a $\mathbb{P}^\theta$-martingale, i.e. if θ is the (unique) solution of

$$r = \mu + \delta(\sqrt{\alpha^2 - (\beta+\theta)^2} - \sqrt{\alpha^2 - (\beta+\theta+1)^2}) \ , \tag{5.19}$$

where r is the constant daily interest rate. Hubalek and Sgarra (2006) derive a closed-form formula for θ given as

$$\theta = -\beta - \frac{1}{2} - \frac{\mu}{2\delta}\sqrt{\frac{4\alpha^2\delta^2}{\mu^2+\delta^2} - 1} \ . \tag{5.20}$$

The authors in Albrecher and Predota (2004) derive a lemma which states that the Esscher transform of an NIG-distributed random variable is again NIG-distributed, and in particular that

$$f_\theta^{NIG}(x; \alpha, \beta, \delta, \mu, \theta) = f^{NIG}(x; \alpha, \beta+\theta, \delta, \mu)$$

holds. Based on that, they derive the value of a European call option at time t with strike price K and maturity T as

$$
\begin{aligned}
C_{NIG}(t;T,K) = \\
& S_t \int_{\ln \frac{K}{S_t}}^{\infty} f^{NIG}(x;\alpha,\beta+\theta+1,\delta(T-t),\mu(T-t))\,\mathrm{d}x \\
& - e^{-r(T-t)} K \int_{\ln \frac{K}{S_t}}^{\infty} f^{NIG}(x;\alpha,\beta+\theta,\delta(T-t),\mu(T-t))\,\mathrm{d}x\ ,
\end{aligned} \tag{5.21}
$$

which can be computed numerically.

5.4.1 Numerical results

The way the risk neutral parameters of the model are obtained is through the calibration procedure, as they cannot be directly observed in the market. In other words, we attempt to minimize the least-square difference between model prices and market prices. We assume there are N options trading in the market, and that option i has price O_i. For given set of parameters Θ, assume that the model gives option prices O_i^{Θ}. Therefore, the set of risk neutral parameters Θ^* is chosen as

$$
\Theta^* = \mathrm{argmin}_{\Theta} \sum_{i=1}^{N} (O_i - O_i^{\Theta})^2. \tag{5.22}
$$

This can be done in PYTHON with the optimization function *curve_fit*[15].

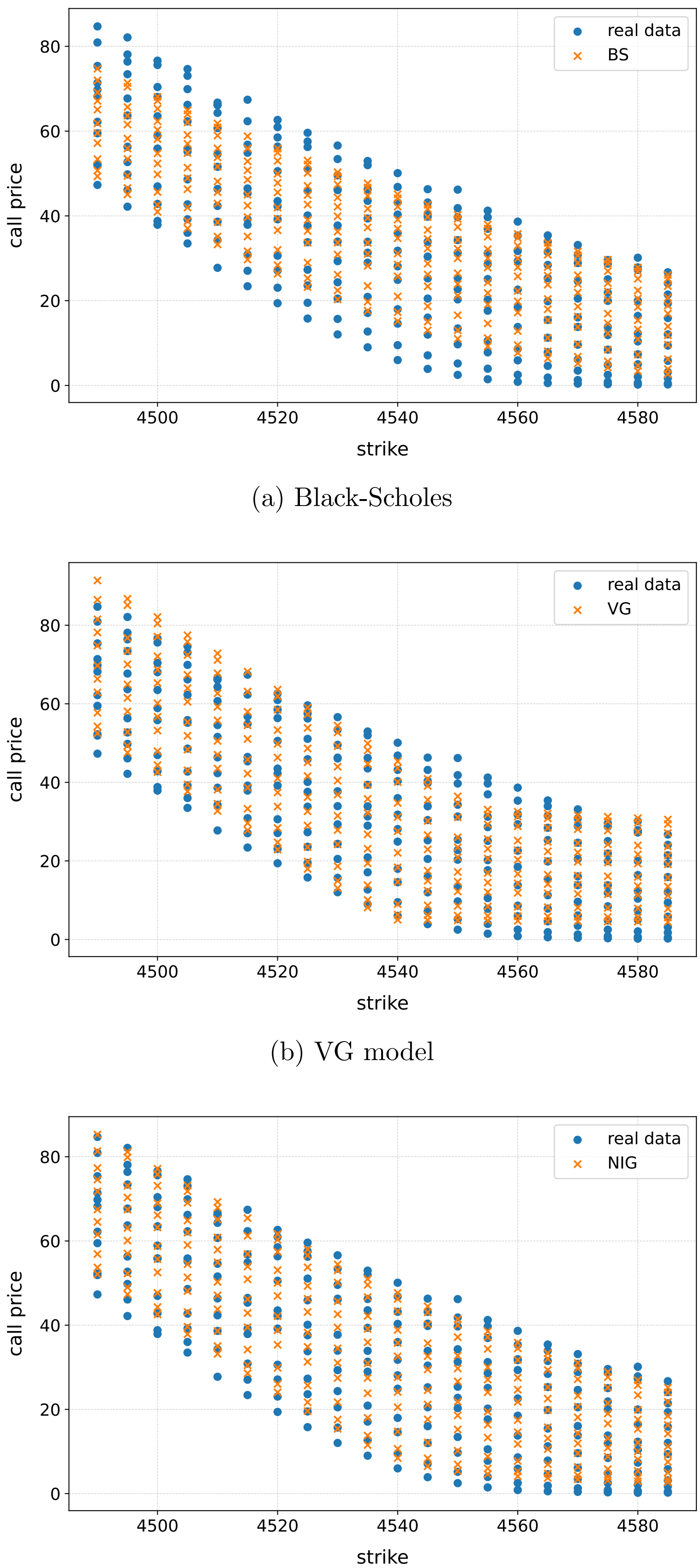
We used the *S&P Index* near-the-money option prices of September 6, 2021 for calibration, where we set $r = 0.021$.

By looking at the plots in Fig. 5.4, we can see that both the VG and NIG models get quite better calibrated than the standard BS model. Reading the R^2 and MSE of the models given in Table 5.1, we can conclude that the best fitted model is the NIG, followed by the VG, and the last one is the BS. Table 5.1 also includes the RN parameters that were obtained from the optimization procedure. We are not exploring the option pricing in more detail as that itself is not the topic of this thesis, for us it is the tool we will use for obtaining the credit risk models.

Model	**Fitted parameters**	**MSE**	**R^2**
BS	$\sigma = 0.089108$	24.979	0.945
VG	$\sigma = 0.490007$, $\nu = 3.287782$, $\theta = 0.155290$	17.634	0.961
NIG	$\alpha = 8.06617343$, $\beta = 75.79304343$, $\delta = 0.13070316$, $\mu = 0.09803445$	6.20	0.986

Table 5.1: Table of the calibrated BS, VG and NIG parameters together with their calibration errors

[15] https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html



(a) Black-Scholes

(b) VG model

(c) NIG model

Figure 5.4: Calibrated option prices of Black-Scholes, VG and NIG model, respectively

Chapter 6

Estimating the Probability of Default

As we mentioned in Section 2.2, the asset values A_t are considered to be unobservable. Thus, the first step in our approach is to estimate them. What one should notice here, is that unlike in the option pricing problem, where the stock prices S_t are the known variables, and the option prices C_t are the unknowns, here the equity – equivalent of option price value – is the observable value, while the assets – stocks equivalent – are not. Additionally, in the option pricing procedure, we could have gotten the risk neutral parameters just by using the calibration technique, but such way is not possible here. When it comes to the Merton model, i.e. when the asset and equity are assumed to be modeled by a geometric Brownian motion, we already explained one way of solving this problem in Section 2.3. However, this approach does not work for the VG nor the NIG model. For that reason, we introduce another way through the ideas of Duan (1994) (and its correction Duan (2000)) applied via the so-called Expectation-Maximization (EM) algorithm, which we will firstly explain for the Merton model and then develop the same idea for the VG and NIG cases.

6.1 Motivation

To motivate the NIG and the VG approach, we fitted NIG, VG and normal distributions to the log-increments of equity data[16] for German companies available on Eikon Refinitiv (2020)[17].

At first, we used all the available data from 2010 to the end of 2021 (with different companies having data available for different time intervals). For the comparison, we used the metric of sum squared errors (SSE). We had in total 611 companies, where 465 of them were best fitted with a VG, 138 of them with an NIG, and only 8 of them with a normal distribution.

Unfortunately, not all of these companies had available data of debts/liabilities, which is crucial for the Merton method, but also for the NIG and VG methods we

[16]Whenever we talk about equity what we actually mean is the equity market value, which for stock companies is given by the firm's stock market price multiplied by the number of outstanding shares.

[17]The equity market value that we used is found under the attribute by the name of "TR.CompanyMarketCapitalization".

Company	SSE			Fitted parameters
	VG	NIG	NORM	NIG
All for One Group SE	52.7	120.3	225.4	$\alpha = 144.63971$, $\beta = -14.64289$, $\delta = 0.02284$, $\mu = 0.00134$
Centrotec SE	63.58	67.63	173.9	$\alpha = 103.70522$, $\beta = -2.94719$, $\delta = 0.01488$, $\mu = -0.00043$
Alstria Office REIT AG	105.6	113.1	138.2	$\alpha = 459.11502$, $\beta = 24.75823$, $\delta = 0.03242$, $\mu = -0.00126$
Bertrand AG	37.38	34.74	74.5	$\alpha = 100.11333$, $\beta = -14.85406$, $\delta = 0.02695$, $\mu = 0.00386$
Jenoptik AG	11.95	18.36	78.91	$\alpha = 64.66791$, $\beta = -1.37641$, $\delta = 0.02246$, $\mu = 0.00131$
CENIT AG	121.1	126.3	197.4	$\alpha = 127.87695$, $\beta = -10.88043$, $\delta = 0.02166$, $\mu = 0.00111$
Datagroup SE	19.19	21.32	36.06	$\alpha = 76.52631$, $\beta = 1.83503$, $\delta = 0.03949$, $\mu = -0.00029$
Mediclin AG	5723.3	2116.5	10057.9	$\alpha = 4.33292$, $\beta = -0.26004$, $\delta = 0.00094$, $\mu = -0.00002$
Masterflex SE	488.8	97.6	447.6	$\alpha = 75.01550$, $\beta = -9.81177$, $\delta = 0.00771$, $\mu = 0.00082$
Nanogate SE	11.36	18.96	118.68	$\alpha = 50.45017$, $\beta = 5.88159$, $\delta = 0.01812$, $\mu = -0.00283$

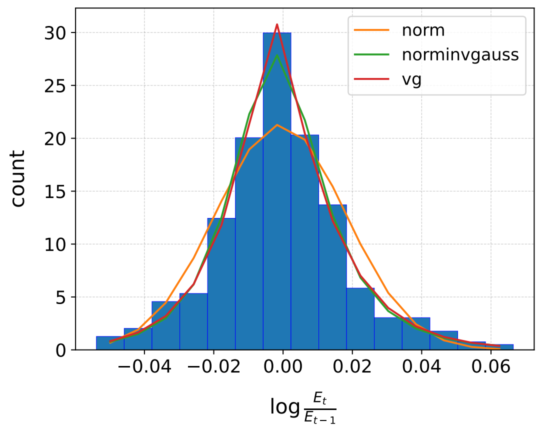
Table 6.1: Table of SSEs for VG, NIG, NORM, and NIG fitted parameters for all of the 10 companies of our consideration.

are about to introduce. Therefore, we restricted our choice to the companies that had debt data available. This resulted in only 268 companies, from which 185 were best fitted with a VG, 81 with an NIG and only 2 of them were best fitted with a normal distribution. As it turns out, those last 2 did not have many available data points – both of them had less than a year of available data, which could be the reason why they were a better fit to a normal distribution rather then to a VG or an NIG. This motivates the development of the NIG and VG models.

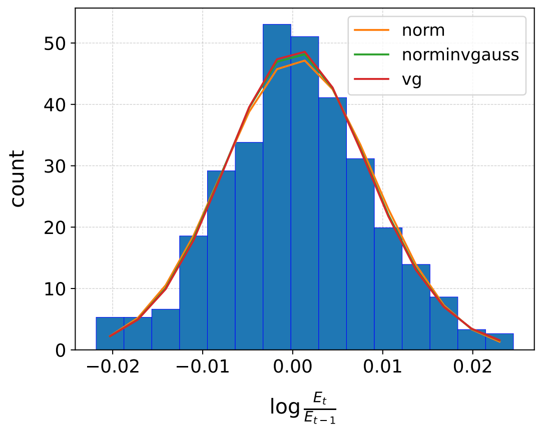
What we noticed is that the difference between SEEs in VG and NIG is usually not significant. In Table 6.1 we can see SSEs for the best fitted distributions of VG, NIG and normal distributions of the 10 companies[18] we chose. The smallest SSE for the company is colored in red – there are 7 companies which showed best performance with VG and 3 of them with NIG. Except for the first company, wherever the SSE is in favor of the VG, we can see that the difference between the VG and the NIG is not significant. In the last column of Table 6.1 we can also find the fitted parameters of the corresponding best NIG distribution, which we will discuss somewhat later. Parameters of the best fitted VG and normal distributions can be found in Table A.1.

As we will see, the VG model is a lot more complicated than the NIG one, and since the difference in their SSEs is not significant, theoretical models are developed both for the NIG and the VG, but in the application we proceed only with the tractable NIG model.

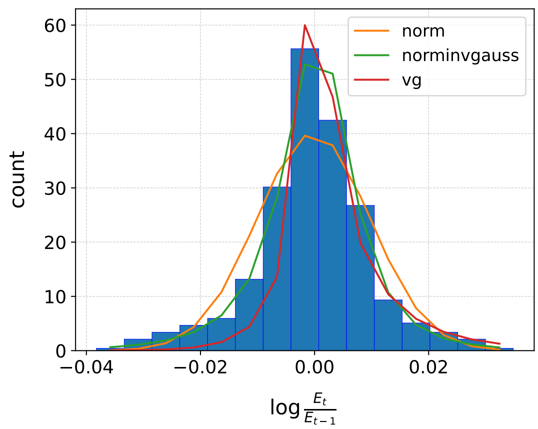
[18]See Chapter 7 for more information on how the 10 companies were chosen.



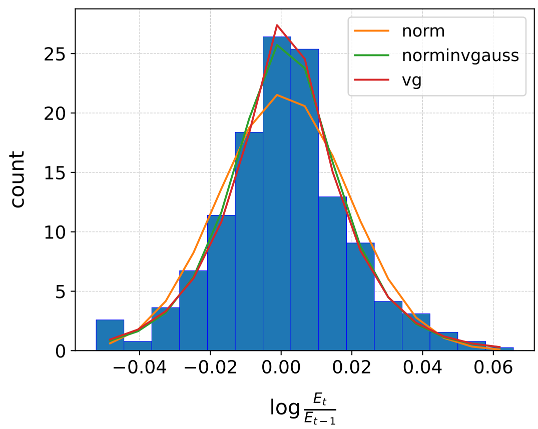
(a) All for One Group SE



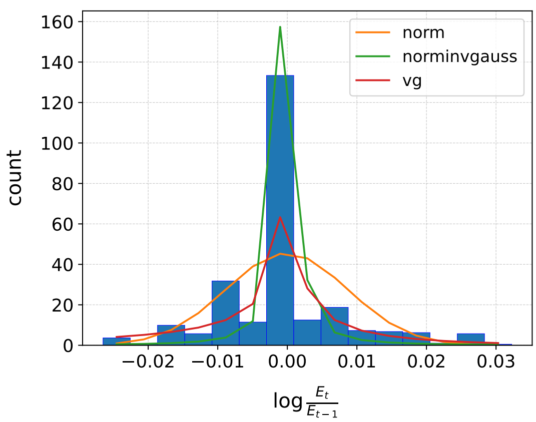
(b) Bertrandt AG



(c) Nanogate SE



(d) CENIT AG



(e) Masterflex SE

Figure 6.1: Fitted VG, NIG and normal distributions to the companies' log-increments of equity data

The fitting and the plots in Fig. 6.1 were done via the PYTHON FITTER package[19]. Unfortunately, that package originally does not include VG distribution functions, so we had to edit the package ourselves and add the VG as one of the available distributions[20]. For the number of bins we used the so-called Rice's rule given as

$$bins = \lceil 2\sqrt[3]{n} \rceil \tag{6.1}$$

where n is the number of data points. We can see that NIG and VG have a much better fit to the companies in Fig. 6.1a, Fig. 6.1c and Fig. 6.1d, where for Fig. 6.1b all 3 of them are pretty close. On the other hand, in Fig. 6.1e we see quite a different situation, where it is difficult to say anything without the error results from Table 6.1.

6.2 Duan's log-likelihood function

Before we develop all the steps for the specific models, we will state the main result from Duan (1994) and afterwards explain the EM algorithm idea, which will be used for all three of the models.

Firstly, the following assumptions are made: X is an n-dimensional vector of unobserved random variates[21] with a distribution in the family $\{F_\theta, \theta \in \Theta\}$ (Θ is an open subset of $\mathbb{R}^k$), the density function $f(x;\theta)$ exists and is continuously twice differentiable in both arguments, and Y is a vector of observed random variates resulting from a data transformation of the unobserved vector X, whose transformation is denoted by $T(\cdot;\theta)$. Additionally, $L(Y;\theta)$ denotes the log-likelihood function of the observed data Y, whereas $L_X(\cdot;\theta)$ is the log-likelihood function of the unobserved random vector X.[22] Using this, we have $Y = T(X;\theta)$ and $X = T^{-1}(Y;\theta)$.

Let us state the result that will be crucial for our estimation procedure, having in mind that in our application $T(X;\theta)$ will represent the option pricing formula of the corresponding model:

Theorem 4. *(Duan, 1994, Theorem 2.2) For an element-by-element basis transformation from X to Y, i.e. $y_i = T_i(x_i;\theta)$ for all i, then*

$$L(Y;\theta) = L_X(\hat{x}_i(\theta), i = 1, ..., n; \theta) - \sum_{i=1}^{n} \ln \left| \frac{\mathrm{d}T_i(\hat{x}_i(\theta);\theta)}{\mathrm{d}x_i} \right|, \tag{6.2}$$

where $\hat{x}_i(\theta) = T_i^{-1}(y_i;\theta)$.

6.3 EM algorithm

To solve the system of equations that we will get using Duan's Theorem 4, we will need to develop an *EM algorithm* for it. In this Section, we briefly explain the idea behind it.

[19] `https://pypi.org/project/fitter/`

[20] We slightly edited the code found at `https://github.com/dlaptev/VarGamma/blob/master/VarGamma.py` and added it to the FITTER package. For more see Appendix C.

[21] A random variate is a particular outcome of a random variable.

[22] Transformation T from $\mathbb{R}^n$ to $\mathbb{R}^n$ is a function of the unknown parameter θ and is one-to-one for every $\theta \in \Theta$. It is also assumed to be continuously differentiable in both arguments.

In 1997, the EM algorithm was named and explained in Dempster et al. (1977), and as they mention themselves, the method has already been proposed many times in special circumstances (Hartley (1958),Carter Jr and Myers (1973), Brown (1974), Chen and Fienberg (1976), Healy and Westmacott (1956)). An EM algorithm is an iterative method for finding (local) maximum likelihood estimates of parameters of a statistical model in cases where the equations cannot be solved directly. Usually, these models involve unobserved latent variables (in our case - asset values) in addition to unknown parameters (both statistical and risk-neutral parameters of BM/VG/NIG) and known data observations (equity values).

The reason behind its name lies in its performance steps. The EM iterates two steps; an *expectation (E) step* and a *maximization (M) step*. The E-step creates a function for the expectation of the log-likelihood evaluated through current values of the parameter-estimates, and the M-step computes parameters that maximize the expected log-likelihood found by the previous E-step. Then, these parameter-estimates are used for determining the distribution of the latent variable in the following E-step.

In such statistical models with latent variables, finding a maximum likelihood solution in a usual way - calculating derivatives of the likelihood function with respect to the unknown values, the latent variables and the parameters, and simultaneously solving the arising system of equations - is typically impossible. What happens is, that the equations in the system are interlocking, and just like in our case, the solution to the values of the latent variables (asset values) requires the parameter values and vice versa, where substitution of a set of equations into another leads to an unsolvable equation.

The EM algorithm starts with the assumption that there is a way of solving those sets of equations numerically. Then, picking arbitrary values for one of the sets of unknowns, estimates the second set, and those new values are then used to find a better estimate of the first set, and so on - it keeps alternating between the two steps until it reaches convergence in both resulting values towards fixed points. The flow chart of an EM algorithm can be seen in Fig. 6.2. Albeit not intuitive, it can be proven that such algorithm works, and that the point is either a local maximum or a saddle point, Wu (1983).

Wu (1983) finds an error in the proof of the convergence result of Dempster et al. (1977), and he studies more broadly two convergence aspects of the EM algorithm and shows that

1. if the unobserved complete-data specification can be described by a curved exponential family with compact parameters space, all the limit points of any EM sequence are stationary points of the likelihood function;

2. if the likelihood function is unimodal and a certain differentiability condition is satisfied, any EM sequence converges to the unique maximum likelihood estimate.

Another important observation by Wu (1983), is that although the EM algorithm has slow numerical convergence, it is a very popular algorithm as the implementation of the E-step and the M-step is easy for many statistical problems thanks to the nice form of the complete-data likelihood function, and in many cases the M-step can be implemented with a standard statistical package, thereby, saving the programming time.

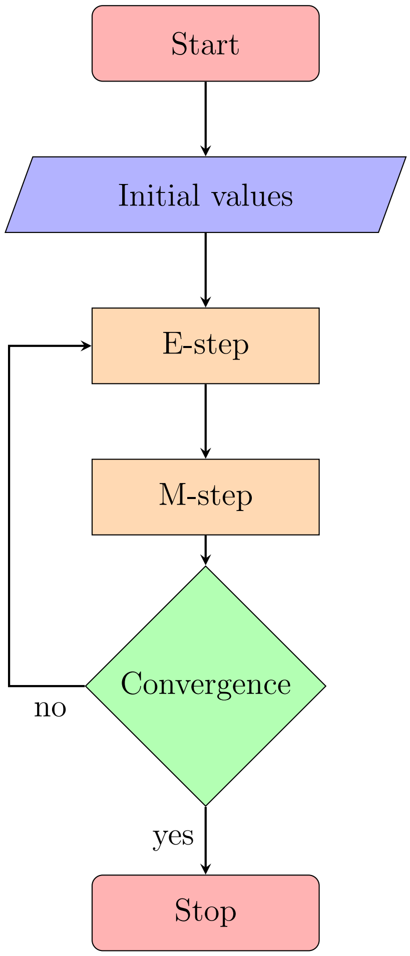


Figure 6.2: Flow chart of the EM algorithm

6.4 Merton model

In the Merton model, our assets A_t follow the geometric Brownian motion Eq. (2.1), therefore we have

$$\ln\left(A_{t+1}/A_t\right) \sim \mathcal{N}\left(\mu_A - \frac{1}{2}\sigma_A^2, \sigma_A^2\right). \tag{6.3}$$

From this, as in Duan (2000), we can write the log-likelihood function for a sample of unobserved values A_t, $t = 1, \ldots, n$[23] as

$$L_A(A_t, t = 1, \ldots, n, \mu_A, \sigma_A) = \sum_{t=2}^{n} \ln\left(f\left(\ln\frac{A_t}{A_{t-1}}\right)\right) - \sum_{t=2}^{n} \ln A_t\ , \tag{6.4}$$

where f is the density of a normal random variable with the distribution given by Eq. (6.3), or when plugged in

$$\begin{aligned} L_A(A_t, t = 1, \ldots, n, \mu_A, \sigma_A) = &- \frac{n-1}{2}\ln(2\pi) - \frac{n-1}{2}\ln\sigma_A^2 - \sum_{t=2}^{n}\ln A_t \\ &- \frac{1}{2\sigma_A^2}\sum_{t=2}^{n}\left[\ln\left(\frac{A_t}{A_{t-1}}\right) - \mu_A + \frac{1}{2}\sigma_A^2\right]^2 . \end{aligned} \tag{6.5}$$

The similar example has been firstly examined in Duan (1994), but there was a mistake that was later corrected in Duan (2000). The term $-\sum_{t=2}^{n}\ln A_t$ was missing, which would be irrelevant if the data of A_t were directly observable, as its derivatives with respect to parameters would then equal zero, where in our case (case of the transformed data setting) this term cannot be ignored, Duan (2000). The exact Merton's example can however be found in Duan et al. (2005). As the equity value formula (the option price formula) Eq. (2.9) is an element-by-element transformation from an unobserved sample of asset values A_t to an observed time series of equity values, and since it is a strictly increasing function in asset value and it depends on the unknown volatility parameter, we can apply Theorem 4 and express the log-likelihood function for the observed sample of equity values E_t as

$$\begin{aligned} L(E_t, t = 1, \ldots, n; \mu_A, \sigma_A) = &- \frac{n-1}{2}\ln(2\pi) - \frac{n-1}{2}\ln\sigma_A^2 - \sum_{t=2}^{n}\ln\hat{A}_t(\sigma_A) \\ &- \frac{1}{2\sigma_A^2}\sum_{t=2}^{n}\left[\ln\left(\frac{\hat{A}_t(\sigma_A)}{\hat{A}_{t-1}(\sigma_A)} - \mu_A + \frac{1}{2}\sigma_A^2\right)\right]^2 \\ &- \sum_{t=2}^{n}\ln\left(\Phi\left(\hat{d}_1(t)\right)\right), \end{aligned} \tag{6.6}$$

[23] As A_t is, by definition, not a stationary process, there exists no proper likelihood value for the first observation. The first observation in the sample merely serves to define the conditional distribution for the observations to follow, Duan (1994).

where $\hat{A}_t(\sigma_A)$[24] is the unique solution to Eq. (2.9) for σ_A, and $\hat{d}_1(t)$ is $d_1(t)$ in Eq. (2.6) with $\hat{A}_t(\sigma_A)$ instead of A_t, and in the last term we used the fact that $\frac{\partial E_t}{\partial A_t} = \Phi(d_1(t))$[25], Duan (2000). As commented and applied in (Duan, 1994, p. 164), using this log-likelihood function Eq. (6.6), an iterative optimization routine can be used for computing the maximum likelihood estimates $\hat{\mu}_A$ and $\hat{\sigma}_A$. In particular, they used the so-called quadratic hill-climbing algorithm of Goldfeld et al. (1966).

This is similar to the procedure that we will use in our VG and NIG models. As the VG distribution and option pricing formulae are much more computationally expensive, we will only state it theoretically, whereas for the NIG and the Merton models we will also do the numerical implementation with real data.

For these reasons let us first explain the NIG model, and afterwards move to the VG model and the applications.

6.5 NIG model

The NIG model can also be found in Jovan and Ahčan (2017), but applied somewhat differently compared to how we will do it using the idea of Duan (1994). Namely, they are not including the second term from Eq. (6.2). Therefore, we need to calculate the partial derivative $\frac{\partial E_t}{\partial A_t}$ for that term, which we give in the Proposition below that one ends up with after some terms get canceled in the calculation which can be found in Appendix B.2.

Proposition 10. *When the statistical dynamics of the asset values is modeled by the NIG process as in Eq. (5.15) and when the equity is given as a call option formula on assets A_t in analogy to Eq. (5.21), i.e.*

$$\begin{aligned} E_{NIG}(t;T,K) = A_t \int_{\ln \frac{D}{A_t}}^{\infty} & f^{NIG}(x;\alpha,\beta+\theta+1,\delta(T-t),\mu(T-t))\,\mathrm{d}x \\ & - e^{-r(T-t)} D \int_{\ln \frac{D}{A_t}}^{\infty} f^{NIG}(x;\alpha,\beta+\theta,\delta(T-t),\mu(T-t))\,\mathrm{d}x, \end{aligned} \tag{6.7}$$

where θ is the solution to Eq. (5.19), the partial derivative $\frac{\partial E_t}{\partial A_t}$ is

$$\frac{\partial E_t}{\partial A_t} = \int_{\ln \frac{D}{A_t}}^{\infty} f^{NIG}(x;\alpha,\beta+\theta+1,\delta(T-t),\mu(T-t))\,\mathrm{d}x \tag{6.8}$$

Remembering the density distribution of log-returns in the NIG model (see Eq. (5.16)), we have

$$\begin{aligned} L_A^{NIG}(A_t, t=1,\dots,n,\alpha,\beta,\delta,\mu) = \\ = \ln\left(\prod_{t=2}^{n} f^{NIG}\left(\ln\frac{A_t}{A_{t-1}};\alpha,\beta,\delta,\mu\right)\right) - \sum_{t=2}^{n} \ln A_t. \end{aligned} \tag{6.9}$$

[24]Inversion does not require μ_A since the equity pricing formula in Eq. (2.9) does not depend on μ_A, as in the risk-neutral evaluation we have the r instead.
[25]see Appendix B.1

Now, applying Theorem 4, and using Proposition 10 we get

$$\begin{aligned} L^{NIG}(E_t, t = 1, \dots, n; \alpha, \beta, \delta, \mu) = \\ \sum_{t=2}^{n} \ln\left(f^{NIG}\left(\ln \frac{\hat{A}_t}{\hat{A}_{t-1}}; \alpha, \beta, \delta(T-t), \mu(T-t) \right) \right) - \sum_{t=2}^{n} \ln \hat{A}_t \\ - \sum_{t=2}^{n} \ln\left(\int_{\ln \frac{D}{\hat{A}_t}}^{\infty} f^{NIG}\left(x; \alpha, \beta+\theta+1, \delta(T-t), \mu(T-t)\right) \mathrm{d}x \right), \end{aligned} \tag{6.10}$$

where $\hat{A}_t$ is the unique solution to Eq. (6.7). Notice that the summands in the last term under the logarithm are actually 1 minus the cumulative distribution function (cdf) of NIG with corresponding parameters, or also known as the *survival function*.

6.6 VG model

Same as for the NIG, for the purpose of applying Theorem 4 we will firstly need to calculate partial derivative of equity by assets $\frac{\partial E_t}{\partial A_t}$, or in the option price and stock notation $\frac{\partial C_t}{\partial S_t}$.

The option price formula Eq. (4.27) from Theorem 3 is given at time 0 for any maturity time T. We obviously cannot use this formula as it is not S_t dependent, but rather S_0 dependent, which is not suitable for our partial derivative. Therefore, first of all we need to write this formula such that it is time-dependent. As the VG process is a Levy process, which by definition has the property of stationary increments, and since Eq. (4.27) holds for any maturity time T, we can rewrite this formula with its value at time t just by replacing every T with $T-t$, and S_0 with S_t, i.e.

$$\begin{aligned} c(S(t); K, T) = S(t) \cdot \Psi\left(d\sqrt{\frac{1-c_1}{\nu}}, (\alpha+s)\sqrt{\frac{\nu}{1-c_1}}, \frac{T-t}{\nu} \right) \\ - Ke^{-r(T-t)} \cdot \Psi\left(d\sqrt{\frac{1-c_2}{\nu}}, \alpha s\sqrt{\frac{\nu}{1-c_2}}, \frac{T-t}{\nu} \right), \end{aligned} \tag{6.11}$$

where

$$d = \frac{1}{s}\left[\ln\left(\frac{S(t)}{K} \right) + r(T-t) + \frac{T-t}{\nu} \ln\left(\frac{1-c_1}{1-c_2} \right) \right]. \tag{6.12}$$

In the following Proposition we state the expression of the partial derivative that one gets after some terms get canceled in the calculation which we give in Appendix B.3.

Proposition 11. *When the risk neutral dynamics of the asset values is given by the VG process as in Eq. (4.23) and when the equity is given as a call option formula on assets A_t in analogy to Eq. (6.11), the partial derivative $\frac{\partial E_t}{\partial A_t}$ is*

$$\frac{\partial E_t}{\partial A_t} = \Psi\left(d\sqrt{\frac{1-c_1}{\nu}}, (\alpha+s)\sqrt{\frac{\nu}{1-c_1}}, \frac{T-t}{\nu} \right), \tag{6.13}$$

where d is now defined as

$$d = \frac{1}{s}\left(\ln\frac{A_t}{D} - (r)(T-t) + \frac{T-t}{\nu}\ln\left(\frac{1-c_1}{1-c_2}\right)\right), \tag{6.14}$$

and $\alpha = \zeta s$, where ζ, s are defined in Eq. (4.17) and Eq. (4.18), respectively, c_1 and c_2 are given by Eq. (4.29), and all of the parameters are the risk neutral ones.

What we also need, is a density function of $z = \ln\frac{A_t}{A_{t-1}}$ for which we will use Theorem 2, but wherever we have t (except in z which will now be $\ln\frac{A_t}{A_{t-1}}$ instead of $\ln\frac{A_t}{A_0}$) we will now have 1, as $t-(t-1) = 1$. To be precise, the density of $z = \ln\frac{A_t}{A_{t-1}}$, for the assets following the VG process dynamics in analogy to Eq. (4.21), is given by

$$f^{VG}(z;\sigma,\theta,\nu) = \frac{2\exp\left(\frac{\theta}{\sigma^2}x\right)}{\nu^{\frac{1}{\nu}}\sqrt{2\pi}\sigma\Gamma\left(\frac{1}{\nu}\right)}\left(\frac{x^2}{2\frac{\sigma^2}{\nu}+\theta^2}\right)^{\frac{1}{2\nu}-\frac{1}{4}} \cdot K_{\frac{1}{\nu}-\frac{1}{2}}\left(\frac{1}{\sigma^2}\sqrt{x^2\left(\frac{2\sigma^2}{\nu}+\theta^2\right)}\right), \tag{6.15}$$

where x is given as $x = z - mt - \frac{1}{\nu_S}\ln(1-\theta_S\nu_S - \sigma_S^2\nu_S/2)$. The VG parameters above are the statistical ones. Now,

$$L_A^{VG}(A_t, t=1,\dots,n,\sigma_S,\theta_S,\nu_S) = \ln\left(\prod_{t=2}^{n} f^{VG}\left(\ln\frac{A_t}{A_{t-1}};\sigma_S,\theta_S,\nu_S\right)\right) - \sum_{t=2}^{n}\ln A_t, \tag{6.16}$$

and using Theorem 4, we have

$$L(E_t, t=1,\dots,n;\sigma_S,\theta_S,\nu_S,\sigma_{RN},\theta_{RN},\nu_{RN}) = L_A(A_t, t=1,\dots,n,\sigma_S,\theta_S,\nu_S) - \sum_{t=2}^{n}\ln\frac{\partial E_t}{\partial A_t}. \tag{6.17}$$

Now, using Proposition 11 for the last term, we get

$$\begin{aligned} L^{VG}(E_t, t=1,\dots,n;\sigma_S,\theta_S,\nu_S,\sigma_{RN},\theta_{RN},\nu_{RN}) = \\ \ln\left(\prod_{t=2}^{n} f^{VG}\left(\ln\frac{\hat{A}_t}{\hat{A}_{t-1}};\sigma_S,\theta_S,\nu_S\right)\right) - \sum_{t=2}^{n}\ln\hat{A}_t \\ - \sum_{t=2}^{n}\ln\Psi\left(d(\hat{A}_t,D)\sqrt{\frac{1-c_1}{\nu_{RN}}}, (\alpha+s)\sqrt{\frac{\nu_{RN}}{1-c_1}}, \frac{T-t}{\nu_{RN}}\right), \end{aligned} \tag{6.18}$$

where $\hat{A}_t$ is the unique solution to Eq. (6.11) in credit risk form (for $c(S(t);K,T)$ being $E_t = E_t(A_t;D,T)$) and $d(\hat{A}_t,D) = d$ is given by Eq. (6.14) with $\hat{A}_t$ instead of A_t. The log-likelihood function given in Eq. (6.18) is obviously way more complicated than the ones in the NIG and the Merton case, as it contains both statistical and risk neutral parameters. Thus, we decide to proceed only with the Merton and the NIG model in our application.

Chapter 7

Numerical Results

7.1 Implementation

We chose 10 companies, with available data of equity and debt for 2017, 2018 and 2019 in Eikon Refinitiv (2020). The equity was again taken from the attribute *Company Market Capitalization* under the code "TR.CompanyMarketCapitalization", and as far as debts are concerned we need those that are due within one year, therefore we used *Current Liabilities - Actual* under "TR.CurrentLiabilitiesActValue".

The EM algorithm was run on the data of 2017 and 2018[26]. When it comes to the E-step of the algorithm, which is calculating the asset values through solving the inversion problem of the option pricing formula (Eq. (2.9) for the Merton, i.e. Eq. (6.7) for the NIG model), we used the *fsolve* function from the package PYTHON SCIPY OPTIMIZE[27], with initial value set to $E_t + D$ (see Eq. (2.2)), where E_t and D are the equity and debt value for the corresponding time point of the asset value.

For the M-step, the maximization of the log-likelihood function was conducted with the help of the *minimize* function from the PYTHON SCIPY OPTIMIZE, which was applied to the negative values of the log-likelihood function $-L$ (i.e. on $-L^{NIG}$ for NIG). The method argument was set to *Nelder-Mead*. The bounds on the parameters in the NIG were set to $(1.0, 500)$ for α, $(-500, 500)$ for β, $(0.0001, 500)$ for δ, and $(-500, 500)$ for μ. Although the theoretical lower bound[28] on α is 0.5, after some numerical examination we concluded that the value of 1.0 was a better choice, as the EM algorithm for α values close to the theoretical one did not converge for some of the examples.

We iterated these two steps until one of the following stopping criteria was reached: either the number of EM iterations became greater than 15 or the difference between each of the parameters from the previous and the current iteration was less than 0.001. However, for all 10 of our companies the number of iterations never exceeded 6.

The interest rate was set to $r = 0$, and the time was counted in years rather than days, i.e. maturity time of 1 year meant $T = 1$, implying that the difference between two days was not 1 anymore, but rather $1/365$. Why are we mentioning this? This

[26] 2017 and 2018 are meant as the "debt year" data and not calendar year data. Debts usually started in, e.g. March 2017, ending in March 2018 – we count this as the "debt year" 2017, and analogously we do for the year 2018.

[27] https://docs.scipy.org/doc/scipy/reference/optimize.html

[28] See Hubalek and Sgarra (2006) for more constraints on parameters of the NIG exponential model.

means that some formulae we stated before must be changed. Log-increments in the Merton model are not distributed as given in Eq. (6.3) anymore, but as

$$\ln\left(A_{(t+1)h}/A_{th}\right) \sim \mathcal{N}\left(\left(\mu_A - \frac{1}{2}\sigma_A^2\right)h, \sigma_A^2 h\right), \tag{7.1}$$

where $h = 1/365$. The same holds for the NIG case, log-increments of asset values are now distributed as $NIG(\alpha, \beta, \delta h, \mu h)$. This means that the L_A and L_A^{NIG} have to be updated accordingly, i.e.

$$L_A(A_{th}, t = 1, \ldots, n, \mu_A, \sigma_A) = \sum_{t=2}^{n} \ln\left(f\left(\ln\frac{A_{th}}{A_{(t-1)h}}\right)\right) - \sum_{t=2}^{n} \ln A_{th}, \tag{7.2}$$

where f is now the density of a normal random variable with a distribution given by Eq. (7.1), and for NIG we obtain

$$\begin{aligned} L_A^{NIG}&(A_{th}, t = 1, \ldots, n, \alpha, \beta, \delta, \mu) = \\ &= \ln\left(\prod_{t=2}^{n} f^{NIG}\left(\ln\frac{A_{th}}{A_{(t-1)h}}; \alpha, \beta, \delta h, \mu h\right)\right) - \sum_{t=2}^{n} \ln A_{th}. \end{aligned} \tag{7.3}$$

As we know how the assets are distributed, or to be precise how $\log A_T/A_0$ are distributed, calculating the probability of default is not difficult. Default happens when the company is not able to pay its debt, i.e. when $A_T < D$. As we calibrated the parameters to the data of years 2017 and 2018, for the first day of 2019 (debt year) we take the E_t and the new D value and calculate the first asset value A_0 of 2019 (denoted as $\mathrm{A}_{\mathrm{new}}$ in Table 7.1, Table 7.2 and Table 7.3) again with the function *fsolve*. Then, for the Merton case the probability of default is calculated as

$$\mathbb{P}(A_T < D) = \Phi_{\hat{\mu}_A, \hat{\sigma}_A}\left(\log\frac{D}{A_0}\right), \tag{7.4}$$

where $\Phi_{\hat{\mu}_A, \hat{\sigma}_A}$ is the cumulative distribution function of the normal variable with parameters $\hat{\mu}_A = \mu_A - 1/2\sigma_A^2$ and $\hat{\sigma}_A^2 = \sigma_A^2$, where μ_A and σ_A are the optimized parameters that we got via the EM algorithm for the Merton case. On the other hand, the probability of default for the NIG case is calculated as

$$\mathbb{P}^{NIG}(A_T < D) = \Phi_{\alpha,\beta,\delta,\mu}^{NIG}\left(\log\frac{D}{A_0}\right), \tag{7.5}$$

where $\Phi_{\alpha,\beta,\delta,\mu}^{NIG}$ is now the cdf of the NIG random variable with the parameters $\alpha, \beta, \delta, \mu$, that we got through the EM algorithm for the NIG model.

Another important thing to mention is that using *norminvgauss* functions from SciPy gets very complicated due to the different parametrization of the NIG density function compared to the one in Eq. (5.6) which we use in our model. Luckily, the package rpy2[29] from Python allows us to call functions from R packages inside Python.

[29] https://rpy2.github.io/doc/latest/html/index.html

The functions we called from the GeneralizedHyperbolic[30] package are *dnig* – the NIG density function, and *pnig* – the NIG cumulative distribution function. These functions are defined in the correct parametrization for our model, i.e. *dnig* is given in the exact form as in Eq. (5.6). The function *dnig* was used for the density function inside L_A^{NIG}, we used $(1 - pnig)$ for the survival function in the option pricing formula and in L^{NIG}, whereas for the calculation of the probability of default we used *pnig*.

7.2 Results

The results of our EM algorithm for the Merton model can be seen in Table 7.1, and for the NIG model in Table 7.2 and Table 7.3. In these tables we can find the values of $\mathrm{A}_{\mathrm{new}}$, of the estimated PDs and the model parameters, as well as the values of the corresponding log-likelihood function.

For the NIG model, we used two different sets of initial values; one was $[1, 0, 1, 0]$[31] – called NIG_1 (Table 7.2), and the other one was $[10, 0, 1, 0]$ – which we will call NIG_{10} (Table 7.3). On the other hand, the initial values for the Merton model were $\sigma = 1$ and $\mu = 0$.

The two NIG cases differ for 5 of the companies – for the cases of All for One Group SE, Centrotec SE, CENIT AG, Mediclin AG and Masterflex SE – where the algorithm reproduced $\alpha = 1$ for NIG_1 (the third row in red in Table 7.2). In those cases, we can see that the $\mathrm{A}_{\mathrm{new}}$ values (the first row in red in Table 7.2) of the NIG_1 are quite different than the ones estimated by the Merton, as well as the ones given by NIG_{10}. The PDs in NIG_1 (the second row in red in Table 7.2) for those 5 companies are so high that they are of an unreasonable order of magnitude. Additionally, the log-likelihood values of the NIG_1 are less than or equal to the ones of the NIG_{10}, for all 10 companies, pointing to the poorer maximization of the log-likelihood function in the NIG_1 case.

For the other 5 companies – Alstria Office REIT AG, Bertrandt AG, Jenoptik AG, Datagroup SE, Nanogate SE – both NIG methods produced almost exactly the same results for the $\mathrm{A}_{\mathrm{new}}$ values, as well as for the PD values and the NIG parameters. For all of these 5 companies parameter α was much greater than 1, in both NIG cases. For the NIG_{10} this actually holds for all of the 10 companies considered, and all of its PDs are of the sensible order of magnitude.

Clearly, the choice of initial values $[10, 0, 1, 0]$ produced more reasonable results overall, as can also be seen from the values of $\mathrm{A}_{\mathrm{new}}$ which are almost exactly the same as in the Merton model. Therefore, one can conclude that the initial set of values $[1, 0, 1, 0]$ is not the best choice, and that the small values of α are causing the troublesome results. In Jovan and Ahčan (2017) we also see the high values of α for all of the examined examples. The same behaviour can be seen in Table 6.1 in blue for the NIG fitted equity data.

All of the aforementioned goes in favor of choosing a greater initial value of α than 1.

Therefore, for the PD values we will only compare the Merton model and the NIG_{10}. The values of the probabilities of default in the Merton case were in most

[30] https://cran.r-project.org/web/packages/GeneralizedHyperbolic/GeneralizedHyperbolic.pdf

[31] The initial values are meant in the form of $[\alpha, \beta, \delta, \mu]$.

cases smaller than the ones in the NIG_{10}[32]. However, for 2 of the companies – Centrotec SE and Datagroup AG – Merton actually gave somewhat higher PDs (0.0039% and 0.0002%, in blue in Table 7.1) than the NIG_{10} (0.0028% and 0.00015%, in blue in Table 7.3). Nevertheless, for all of the 10 companies the log-likelihood values were larger for the NIG_{10} model compared to the Merton one, which inidicates a better maximization of the log-likelihood function in the NIG_{10} case. These results clearly indicate that the NIG model is promising and deserves further, more-detailed examination.

[32]This claim is not explicitly supported by the values in the tables for Alstria Office REIT AG, Jenoptik AG and Masterflex SE, as the values which were smaller than 10^{-7} were rounded off to 0.0. However, if one were to examine the values without rounding them off, it would be clear that the Merton model gave smaller PD values for these companies as well.

Merton [1,0]	1	2	3	4	5	6	7	8	9	10
Company	**All for One Group SE**	**Centrotec SE**	**Alstria Off. REIT AG**	**Bertrandt AG**	**Jenoptik AG**	**CENIT AG**	**Datagroup SE**	**Mediclin AG**	**Masterflex SE**	**Nanogate SE**
A_{new}	305099000	319162636	2448642245	954026423	2111126844	147200405	344101158	353974220	75508498	706479937
PD in %	0.0000027	0.0039504	0.0	0.0000461	0.0	0.0000074	0.0002006	0.0	0.0	0.0000164
mu	-0.284826	-0.21972	0.252765	-0.114422	0.334559	-0.128186	0.406908	0.020408	0.027654	-0.045487
sigma	0.238533	0.204509	0.18442	0.321364	0.371876	0.234445	0.440683	0.167728	0.218735	0.392113
log-likelihood	-4212.69	-8289.21	-9096.22	-9108.94	-9397.94	-8050.96	-8706.57	-8185.00	-7520.28	-9014.74

Table 7.1: The results of A_{new}, PD, parameters and log-likelihood values of the Merton model for the 10 companies

NIG [1,0,1,0]	1	2	3	4	5	6	7	8	9	10
Company	**All for One Group SE**	**Centrotec SE**	**Alstria Off. REIT AG**	**Bertrandt AG**	**Jenoptik AG**	**CENIT AG**	**Datagroup SE**	**Mediclin AG**	**Masterflex SE**	**Nanogate SE**
A_{new}	291798228	288142066	2448642245	954026423	2111126844	136895418	344101284	351671357	73419829	706479937
PD in %	17.0081284	26.4084192	0.0	0.0000701	0.0	19.7274425	0.0001534	1.8914366	8.0054439	0.0000182
alpha	1	1	108.845948	58.116981	38.933677	1	42.145072	1	1	33.95863
beta	-0.032949	-0.044674	10.168514	-10.46751	-0.698807	-0.037464	3.132617	0.033075	-0.07743	4.975503
delta	2.259581	1.872039	3.426173	5.559027	5.465714	2.352158	7.958265	0.372898	1.535509	5.203789
mu	-0.197419	-0.120384	-0.085776	0.85184	0.363505	-0.048531	-0.283353	-0.003538	0.131392	-0.893166
log-likelihood	-4214.43	-8256.55	-9053.51	-9067.88	-9357.06	-8065.02	-8673.19	-8015.10	-7430.35	-8964.08

Table 7.2: The results of A_{new}, PD, parameters and log-likelihood values of the NIG_1 model for the 10 companies

NIG [10,0,1,0]	1	2	3	4	5	6	7	8	9	10
Company	**All for One Group SE**	**Centrotec SE**	**Alstria Off. REIT AG**	**Bertrandt AG**	**Jenoptik AG**	**CENIT AG**	**Datagroup SE**	**Mediclin AG**	**Masterflex SE**	**Nanogate SE**
A_{new}	305099000	319162638	2448642245	954026423	2111126844	147200406	344101284	353971155	75508498	706479937
PD in %	0.0000055	0.0028178	0.0	0.0000701	0.0	0.0000110	0.0001534	0.0093423	0.0	0.0000182
alpha	64.781842	69.154272	108.845974	58.116969	38.933716	75.132328	42.145091	5.289669	41.102025	33.958645
beta	-2.261522	-2.363977	10.168531	-10.46752	-0.698797	-1.539022	3.132635	0.228755	-2.350796	4.975542
delta	3.71638	2.700386	3.426174	5.559023	5.46572	4.142424	7.958264	0.380147	1.978275	5.20379
mu	-0.183428	-0.148226	-0.085776	0.85184	0.363503	-0.070817	-0.283354	-0.010077	0.117071	-0.89317
log-likelihood	-4194.17	-8217.09	-9053.51	-9067.88	-9357.06	-8022.15	-8673.19	-8014.98	-7410.47	-8964.08

Table 7.3: The results of A_{new}, PD, parameters and log-likelihood values of the NIG_{10} model for the 10 companies

Chapter 8

Conclusion

In this section we take a look back on what has been done in this thesis and summarize the most important takeaways.

In Chapter 1, we reported on the problem of the famous Merton model (introduced in Chapter 2) in underestimating the probability of default, which is what we aim to overcome in this thesis with the models we developed.

The most important aspects of the Lévy processes were given in Chapter 3, and an exhaustive descriptions of two cases of subordinated processes – VG and NIG – were given in Chapter 4 and Chapter 5, respectively.

At the beginning of Chapter 6 the motivation for the use of the NIG and VG models can be found, followed by the explanation on the approaches of the three models at hand.

Our approach was based on the log-likelihood formula by Duan (1994) (see Section 6.2) which needed partial derivatives that we calculated from the corresponding option pricing formulae (see Section 4.5 and Section 5.4). The detailed calculation of these derivatives is given in Appendix B.2 and Appendix B.3 for the NIG and the VG, respectively. The problem of unobservable asset values was solved with the use of the EM algorithm (Section 6.3).

As the VG model is computationally quite challenging (Section 6.6), and as the differences in the SSEs of the VG and the NIG equity distributions were not significant (Section 6.1), we chose to proceed only with the NIG model in our numerical implementation.

In Chapter 7, the Python implementation of the Merton and the NIG model was explained, as well as the presentation of the results obtained for 10 German companies.

These results represent positive findings – for most of the companies examined the NIG model estimated higher probabilities of default than the ones estimated via the Merton model, and the log-likelihood function values were higher for the NIG model for all of the 10 companies.

The NIG model clearly seems promising and deserves further examination, as there are still some challenges to overcome, such as choosing appropriate initial values, improving the computation time and testing on a bigger sample.

As George E. P. Box said, *"All models are wrong, but some are useful"*, and with the results of this thesis we definitely showed the usefulness of the introduced models.

Appendix A

Fitted Parameters

In the table below we can see the parameters of the best fitted VG and normal distributions for the 10 companies listed. The parameter c in VG is the shift of location, i.e. it plays the same role as the parameter μ in an NIG or a normal distribution.

Company	VG	NORM
All for One Group SE	$c = 0.00009$, $\sigma = 0.01288$, $\theta = -0.00110$, $\nu = 0.56090$	$\mu = -0.00101$, $\sigma = 0.01271$
Centrotec SE	$c = -0.00053$, $\sigma = 0.01198$, $\theta = -0.00032$, $\nu = 0.55992$	$\mu = -0.00086$, $\sigma = 0.01189$
Alstria Office REIT AG	$c = -0.00094$, $\sigma = 0.00841$, $\theta = 0.00142$, $\nu = 0.08896$	$\mu = 0.00049$, $\sigma = 0.00842$
Bertrand AG	$c = 0.00492$, $\sigma = 0.01632$, $\theta = -0.00508$, $\nu = 0.24865$	$\mu = -0.00016$, $\sigma = 0.01644$
Jenoptik AG	$c = 0.00174$, $\sigma = 0.01870$, $\theta = -0.0009$, $\nu = 0.64868$	$\mu = 0.00083$, $\sigma = 0.01843$
CENIT AG	$c = 0.00087$, $\sigma = 0.01305$, $\theta = -0.00160$, $\nu = 0.35149$	$\mu = -0.00074$, $\sigma = 0.01302$
Datagroup SE	$c = -0.00093$, $\sigma = 0.02258$, $\theta = 0.00162$, $\nu = 0.26553$	$\mu = 0.00069$, $\sigma = 0.02247$
Mediclin AG	$c = 0.0$, $\sigma = 0.01781$, $\theta = -0.00537$, $\nu = 4.74378$	$\mu = -0.00007$, $\sigma = 0.00876$
Masterflex SE	$c = 0.0$, $\sigma = 0.00768$, $\theta = 0.00269$, $\nu = 2.40219$	$\mu = -0.00020$, $\sigma = 0.00995$
Nanogate SE	$c = -0.00280$, $\sigma = 0.01889$, $\theta = 0.00209$, $\nu = 0.78275$	$\mu = -0.00070$, $\sigma = 0.01873$

Table A.1: Parameters of best fitted VG and normal distribution of the 10 companies listed

Appendix B

Derivative Calculations

B.1 Merton model

Here we will derive the partial derivative of Eq. (2.9). We start with a chain rule and using the fact that

$$\frac{\partial \Phi(d(t))}{\partial A_t} = \frac{\partial}{\partial A_t}\int_{-\infty}^{d(t)=g(A_t)} f(x)dx = f(d(t)) \cdot \frac{\partial g(A_t)}{\partial A_t} - 0 = f(d(t)) \cdot \frac{\partial d(t)}{\partial A_t}$$

for $d(t) = d_1(t)$ and $d(t) = d_2(t)$ which are both functions of A_t, and where f is the density function of the standard normal variable, i.e. $f(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}}x^2$. This leads to

$$\begin{aligned}\frac{\partial E_t}{\partial A_t} = \Phi(d_1(t)) + A_t\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(d_1(t))^2}\frac{\partial d_1(t)}{\partial A_t} \\ - De^{-r(T-t)}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(d_2(t))^2}\frac{\partial d_2(t)}{\partial A_t},\end{aligned}$$

and inserting

$$\frac{\partial d_1(t)}{\partial A_t} = \frac{\partial d_2(t)}{\partial A_t} = \frac{1}{A_t\sigma_A\sqrt{T-t}},$$

as well as $d_2(t) = d_1(t) - \sigma_A\sqrt{T-t}$, it becomes

$$\begin{aligned}\frac{\partial E_t}{\partial A_t} = \Phi(d_1(t)) + A_t\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(d_1(t))^2}\frac{1}{\sigma_A\sqrt{T-t}} \\ - \frac{D}{A_t}e^{-r(T-t)}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(d_1(t)-\sigma_A\sqrt{T-t})^2}\frac{1}{\sigma_A\sqrt{T-t}}.\end{aligned} \tag{B.1}$$

Now from

$$d_1(t) := \frac{\ln(A_t/D) + (r + 1/2\sigma_A^2)(T-t)}{\sigma_A\sqrt{T-t}},$$

we can calculate the value $\frac{D}{A_t}$ as

$$\frac{D}{A_t} = e^{-d_1(t)\sigma_A\sqrt{T-t}+(r+\frac{1}{2}\sigma_A^2)(T-t)}.$$

Inserting this into Eq. (B.1) the last summand in Eq. (B.1) becomes equal to the second summand with a difference in the sign, therefore they cancel each other and we finally get

$$\frac{\partial E_t}{\partial A_t} = \Phi(d_1(t)).$$

B.2 NIG model

Let us write the NIG option price formula in the credit risk setting.

$$\begin{aligned} E_t = A_t \int_{\ln \frac{D}{A_t}}^{\infty} f^{NIG}(x; \alpha, \beta + \theta + 1, \delta(T-t), \mu(T-t)) \, \mathrm{d}x \\ - D \cdot e^{-r(T-t)} \int_{\ln \frac{D}{A_t}}^{\infty} f^{NIG}(x; \alpha, \beta + \theta, \delta(T-t), \mu(T-t)) \, \mathrm{d}x. \end{aligned} \tag{B.2}$$

Using the fact that

$$\frac{\mathrm{d}}{\mathrm{d}x} \left(\int_{g_1(x)}^{g_2(x)} f(t) \, \mathrm{d}t \right) = f(g_2(x)) \cdot g_2'(x) - f(g_1(x)) \cdot g_1'(x) \tag{B.3}$$

and by redefining the functions as

$$\begin{aligned} f^{NIG}(x; \alpha, \beta + \theta + 1, \delta(T-t), \mu(T-t)) := f_1(x) \\ f^{NIG}(x; \alpha, \beta + \theta, \delta(T-t), \mu(T-t)) := f_2(x) \,, \end{aligned} \tag{B.4}$$

we can write the partial derivative as

$$\begin{aligned} \frac{\partial E_t}{\partial A_t} &= \int_{\ln \frac{D}{A_t}}^{\infty} f_1(x) \, \mathrm{d}x + A_t \cdot \Big(f_1(\infty) \cdot 0 - f_1(\ln \tfrac{D}{A_t}) \cdot (\ln \tfrac{D}{A_t})'_{A_t} \Big) \\ &\quad - D \cdot e^{-r(T-t)} \Big(f_2(\infty) \cdot 0 - f_2(\ln \tfrac{D}{A_t}) \cdot (\ln \tfrac{D}{A_t})'_{A_t} \Big) \\ &= \int_{\ln \frac{D}{A_t}}^{\infty} f_1(x) \, \mathrm{d}x + \underbrace{f_1(\ln \tfrac{D}{A_t}) - \frac{D}{A_t} e^{-r(T-t)} \cdot f_2(\ln \tfrac{D}{A_t})}_{(*)} \,, \end{aligned} \tag{B.5}$$

Now, using

$$\Big(\ln(\tfrac{D}{A_t}) \Big)'_{A_t} = -\frac{1}{A_t}, \tag{B.6}$$

the NIG density formula

$$f^{NIG}(x; \alpha, \beta, \delta, \mu) = \frac{\alpha\delta}{\pi} e^{\delta\sqrt{\alpha^2+\beta^2}} e^{\beta(x-\mu)} \frac{K_1(\alpha\sqrt{\delta^2 + (x-\mu)^2})}{\sqrt{\delta^2 + (x-\mu)^2}}, \tag{B.7}$$

and Eq. (B.4) we get

$$
\begin{aligned}
(*) &= f_1(\ln \tfrac{D}{A_t}) - \frac{D}{A_t} e^{-r(T-t)} \cdot f_2(\ln \tfrac{D}{A_t}) \\
&= \frac{\alpha\delta(T-t)}{\pi} \, e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\theta+1)^2}} \, e^{(\beta+\theta+1)(\ln \frac{D}{A_t} - \mu(T-t))} \cdot \\
&\quad \cdot \frac{K_1(\alpha\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2}}{\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2}} \\
&\quad - \frac{D}{A_t} e^{-r(T-t)} \, \frac{\alpha\delta(T-t)}{\pi} e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\delta)^2}} \, e^{(\beta+\theta)(\ln \frac{D}{A_t} - \mu(T-t))} \cdot \\
&\quad \cdot \frac{K_1(\alpha\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2})}{\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2}} \\
&= \frac{\alpha\delta(T-t)}{\pi} \, e^{(\beta+\theta)(\ln \frac{D}{A_t} - \mu(T-t))} \cdot \frac{K_1(\alpha\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2})}{\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2}} \cdot \\
&\quad \cdot \left(e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\theta+1)^2}} \, e^{\ln \frac{D}{A_t} - \mu(T-t)} - \frac{D}{A_t} \, e^{-r(T-t)} \, e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\theta)^2}} \right) \\
&= \frac{\alpha\delta(T-t)}{\pi} \, e^{(\beta+\theta)(\ln \frac{D}{A_t} - \mu(T-t))} \cdot \frac{K_1(\alpha\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2})}{\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2}} \cdot \\
&\quad \cdot \left(\frac{D}{A_t} \, e^{-\mu(T-t)} \, e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\theta+1)^2}} - \frac{D}{A_t} \, e^{-r(T-t)} \, e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\theta)^2}} \right) \\
&= \frac{\alpha\delta(T-t)}{\pi} \, e^{(\beta+\theta)\left(\ln \frac{D}{A_t} - \mu(T-t)\right)} \cdot \frac{K_1(\alpha\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2})}{\sqrt{(\delta(T-t))^2 + (\ln \frac{D}{A_t} - \mu(T-t))^2}} \cdot \\
&\quad \cdot \frac{D}{A_t} \left(e^{-\mu(T-t)} \, e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\theta+1)^2}} - e^{r(T-t)} \, e^{\delta(T-t)\sqrt{\alpha^2+(\beta+\theta)^2}} \right) = 0,
\end{aligned}
$$

where for the last equation we used the property of θ from Eq. (5.19). Therefore,

$$
\frac{\partial E_t}{\partial A_t} = \int_{\ln \frac{D}{A_t}}^{\infty} f^{NIG}(x; \alpha, \beta + \theta + 1, \delta(T-t), \mu(T-t)) \, \mathrm{d}x. \tag{B.8}
$$

B.3 VG model

Let us explain how the option price formula is derived in Madan and Milne (1991) for the symmetric VG process $N(t)$. For easier following of that paper we will use their notation, i.e. the maturity will be t and option price will be calculated at time 0. Using a lot of semimartingale theory, they derive a change of measure density process $\lambda(t) = \mathbb{E}_{\mathbb{P}}[d\mathbb{Q}/d\mathbb{P}|\mathcal{F}_t]$ ((Madan and Milne, 1991, Chapter 3)), and then calculate option price value at time 0 as

$$W_0 = E_{\mathbb{Q}}[e^{-rt}(S(t) - K)^+] = E_{\mathbb{P}}[e^{-rt}\lambda(t)(S(t) - K)^+]\,. \tag{B.9}$$

To calculate this, they do the following steps:

1. first condition the Gamma variate $G(t) = G$, conditional on this, as $N(t) = W(G(t)) = W(G)$, we have $N(t) \sim \mathcal{N}(0, G)$,
2. the conditional option $W(G)$ is obtained by standard methods on integrating wrt the normal density,
3. the conditional option is then integrated w.r.t. the gamma density with mean t and variance νt (see Eq. (4.7) and Section 4.2).

In other words, if $W(g)$ is the conditional option price, then the option price W_0 is calculated by integrating the conditional option price $W(g)$ w.r.t. the gamma density with mean t and variance νt (see Eq. (4.7) and Section 4.2), i.e.

$$f(g) = \frac{g^{\frac{t}{\nu}-1}e^{-\frac{g}{\nu}}}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)}, \tag{B.10}$$

therefore

$$W_0 = \int_0^\infty W(g)\frac{g^{\frac{t}{\nu}-1}e^{-\frac{g}{\nu}}}{\nu^{\frac{t}{\nu}}\Gamma\left(\frac{t}{\nu}\right)}\,\mathrm{d}g. \tag{B.11}$$

For the non-symmetric process VG and maturity time T, the option price formula Eq. (4.27) (Theorem 3) is similarly derived, where some steps can be found in the Appendix of Madan et al. (1998). The conditional option value $c(g)$ (what was $W(g)$ for $N(t)$) is in this case given as

$$\begin{aligned} c(g) = S_0\Big(1 - \frac{\nu(\alpha+s)^2}{2}\Big)^{\frac{T-t}{\nu}} e^{\frac{1}{2}(\alpha+1)^2 g} \cdot \Phi\Big(\frac{d}{\sqrt{g}} + (\alpha+1)\sqrt{g}\Big) \\ - Ke^{-r(T-t)}\Big(1 - \frac{\nu\alpha^2}{2}\Big)^{\frac{T-t}{\nu}} e^{\frac{1}{2}\alpha^2 g} \cdot \Phi\Big(\frac{d}{\sqrt{g}} + \alpha\sqrt{g}\Big), \end{aligned} \tag{B.12}$$

and the call option price $C(S_0, K, T)$ is again obtained by integration w.r.t. the gamma density

$$C(S_0, K, T) = \int_0^\infty c(g)\frac{g^{\frac{T}{\nu}-1}e^{-\frac{g}{\nu}}}{\nu^{\frac{T}{\nu}}\Gamma\left(\frac{T}{\nu}\right)}\,\mathrm{d}g. \tag{B.13}$$

Now, at time t the option price $C(S_t, K, T) = C_t$ has the following form

$$\begin{aligned} C_t = S_t \int_0^\infty & \left(1 - \frac{\nu(\alpha+s)^2}{2}\right)^{\frac{T-t}{\nu}} e^{\frac{1}{2}(\alpha+1)^2 g} \cdot \\ & \cdot \Phi\left(\frac{d}{\sqrt{g}} + (\alpha+1)\sqrt{g}\right) \frac{g^{\frac{T-t}{\nu}-1} e^{-\frac{g}{\nu}}}{\nu^{\frac{T-t}{\nu}} \Gamma\left(\frac{T-t}{\nu}\right)} \mathrm{d}g \\ & - K e^{-r(T-t)} \int_0^\infty \left(1 - \frac{\nu\alpha^2}{2}\right)^{\frac{T-t}{\nu}} e^{\frac{1}{2}\alpha^2 g} \cdot \\ & \cdot \Phi\left(\frac{d}{\sqrt{g}} + \alpha\sqrt{g}\right) \frac{g^{\frac{T-t}{\nu}-1} e^{-\frac{g}{\nu}}}{\nu^{\frac{T-t}{\nu}} \Gamma\left(\frac{T-t}{\nu}\right)} \mathrm{d}g. \end{aligned} \tag{B.14}$$

In our case, C_t represents equity E_t, stock price S_t represents asset A_t, and K represents liabilities/debt D. We are interested in $\frac{\partial E_t}{\partial A}$, which would be $\frac{\partial C_t}{\partial S_t}$, therefore let us calculate $\frac{\partial C_t}{\partial S_t}$ and in the end change the notation to the credit risk needs. For this purpose we will use Eq. (B.14) instead of its form via the Ψ function given in Eq. (6.11), as the derivative calculation is much easier this way. For that purpose, notice that for $m \in \{0, s\}$ and with $f(\cdot)$ being the density function of the standard normal variable, we have

$$\begin{aligned} \frac{\partial \Phi\left(\frac{d}{\sqrt{g}} + (\alpha+m)\sqrt{g}\right)}{\partial S_t} &= f\left(\frac{d}{\sqrt{g}} + (\alpha+m)\sqrt{g}\right) \cdot \frac{\partial\left(\frac{d}{\sqrt{g}} + (\alpha+m)\sqrt{g}\right)}{\partial S_t} \\ &= f\left(\frac{d}{\sqrt{g}} + (\alpha+m)\sqrt{g}\right) \cdot \frac{1}{\sqrt{g}} d'_{S_t} \\ &= f\left(\frac{d}{\sqrt{g}} + (\alpha+m)\sqrt{g}\right) \cdot \frac{1}{\sqrt{g}} \frac{1}{s \cdot S_t}, \end{aligned} \tag{B.15}$$

which we now use in the following calculation of $\frac{\partial C_t}{\partial S_t}$

$$\begin{aligned} \frac{\partial C_t}{\partial S_t} = \int_0^\infty & \left(1 - \frac{\nu(\alpha+s)^2}{2}\right)^{\frac{T-t}{\nu}} e^{\frac{1}{2}(\alpha+s)^2 g} \cdot \\ & \cdot \Phi\left(\frac{d}{\sqrt{g}} + (\alpha+s)\sqrt{g}\right) \frac{g^{\frac{T-t}{\nu}-1} e^{-\frac{g}{\nu}}}{\nu^{\frac{T-t}{\nu}} \Gamma(\frac{T-t}{\nu})} \mathrm{d}g \\ & + S_t \frac{\left(1 - \nu(\alpha+s)^2/2\right)^{\frac{T-t}{\nu}}}{\nu^{\frac{T-t}{\nu}} \Gamma(\frac{T-t}{\nu})} \cdot \\ & \cdot \int_0^\infty e^{\frac{1}{2}(\alpha+s)^2 g} f\left(\frac{d}{\sqrt{g}} + (\alpha+1)\sqrt{g}\right) \frac{1}{\sqrt{g}} \frac{1}{s \cdot S_t} g^{\frac{T-t}{\nu}-1} e^{-\frac{g}{\nu}} \mathrm{d}g \\ & - K e^{-r(T-t)} \frac{\left(1 - \nu\alpha^2/2\right)^{\frac{T-t}{\nu}}}{\nu^{\frac{T-t}{\nu}} \Gamma(\frac{T-t}{\nu})} \cdot \\ & \cdot \int_0^\infty e^{\frac{1}{2}\alpha^2 g} \cdot f\left(\frac{d}{\sqrt{g}} + \alpha\sqrt{g}\right) \frac{1}{\sqrt{g}} \frac{1}{s \cdot S_t} g^{\frac{T-t}{\nu}-1} e^{-\frac{g}{\nu}} \mathrm{d}g. \end{aligned} \tag{B.16}$$

Let us now concentrate on the calculation of the last two terms. Omitting the common factor $\left(s\,\nu^{\frac{T-t}{\nu}}\Gamma(\frac{T-t}{\nu})\right)^{-1}$ and denoting them as $(**)$ we have

$$\begin{aligned}(**) =&\Big(1-\frac{\nu(\alpha+s)^2}{2}\Big)^{\frac{T-t}{\nu}}\int_0^\infty e^{\frac{1}{2}(\alpha+s)^2 g}\cdot\\ &\cdot f\bigg(\frac{d}{\sqrt{g}}+(\alpha+s)\sqrt{g}\bigg)g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g\\ &-\frac{K}{S_t}e^{-r(T-t)}\Big(1-\frac{\nu\alpha^2}{2}\Big)^{\frac{T-t}{\nu}}\int_0^\infty e^{\frac{1}{2}\alpha^2 g}\cdot\\ &\cdot f\bigg(\frac{d}{\sqrt{g}}+\alpha\sqrt{g}\bigg)g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g.\end{aligned}\tag{B.17}$$

As $f(\frac{x-\mu}{\sigma})=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2}$, $(**)$ becomes

$$\begin{aligned}(**) =&(1-c_1)^{\frac{T-t}{\nu}}\int_0^\infty e^{\frac{1}{2}(\alpha+s)^2 g}\frac{1}{\sqrt{2\pi g}}e^{-\frac{1}{2}\left(\frac{d}{\sqrt{g}}+(\alpha+s)\sqrt{g}\right)^2}g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g\\ &-\frac{K}{S_t}e^{-r(T-t)}(1-c_2)^{\frac{T-t}{\nu}}\cdot\\ &\cdot\int_0^\infty e^{\frac{1}{2}\alpha^2 g}\frac{1}{\sqrt{2\pi g}}e^{-\frac{1}{2}\left(\frac{d}{\sqrt{g}}+\alpha\sqrt{g}\right)^2}g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g\\ =&(1-c_1)^{\frac{T-t}{\nu}}\cdot\\ &\cdot\int_0^\infty e^{\frac{1}{2}(\alpha+s)^2 g}\frac{1}{\sqrt{2\pi g}}e^{-\frac{1}{2}\left(\frac{d^2}{g}+2d(\alpha+s)+(\alpha+s)^2 g\right)}g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g\\ &-\frac{K}{S_t}e^{-r(T-t)}(1-c_2)^{\frac{T-t}{\nu}}\cdot\\ &\cdot\int_0^\infty e^{\frac{1}{2}\alpha^2 g}\frac{1}{\sqrt{2\pi g}}e^{-\frac{1}{2}\left(\frac{d^2}{g}+2d\alpha+\alpha^2 g\right)}g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g\\ =&(1-c_1)^{\frac{T-t}{\nu}}e^{-ds}\int_0^\infty\frac{1}{\sqrt{2\pi g}}e^{-\frac{1}{2}\left(\frac{d^2}{g}+2d\alpha\right)}g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g\\ &-\frac{K}{S_t}e^{-r(T-t)}(1-c_2)^{\frac{T-t}{\nu}}\int_0^\infty\frac{1}{\sqrt{2\pi g}}e^{-\frac{1}{2}\left(\frac{d^2}{g}+2d\alpha\right)}g^{\frac{T-t}{\nu}-3/2}e^{-\frac{g}{\nu}}\,\mathrm{d}g\end{aligned}\tag{B.18}$$

Now, by using Eq. (6.12) we obtain

$$\begin{aligned}(1-c_1)^{\frac{T-t}{\nu}}e^{-ds} &= (1-c_1)^{\frac{T-t}{\nu}}e^{-\ln\frac{S_t}{K}-r(T-t)-\frac{T-t}{\nu}\ln\left(\frac{1-c_1}{1-c_2}\right)}\\ &=(1-c_1)^{\frac{T-t}{\nu}}\frac{K}{S_t}e^{-r(T-t)}\left(\frac{1-c_1}{1-c_2}\right)^{-\frac{T-t}{\nu}}\\ &=\frac{K}{S_t}e^{-r(T-t)}(1-c_2)^{\frac{T-t}{\nu}}.\end{aligned}\tag{B.19}$$

From this follows that $(**) = 0$, therefore $\frac{\partial E_t}{\partial A_t}$ is

$$\begin{aligned}\frac{\partial E_t}{\partial A_t} = \int_0^\infty \Big(1 - \frac{\nu(\alpha+s)^2}{2}\Big)^{\frac{T-t}{\nu}} e^{\frac{1}{2}(\alpha+s)^2 g} \cdot \\ \cdot \Phi\left(\frac{d}{\sqrt{g}} + (\alpha+s)\sqrt{g}\right) \frac{g^{\frac{T-t}{\nu}-1} e^{-\frac{g}{\nu}}}{\nu^{\frac{T-t}{\nu}}\Gamma(\frac{T-t}{\nu})}\,\mathrm{d}g,\end{aligned} \tag{B.20}$$

or via the Ψ function form

$$\frac{\partial E_t}{\partial A_t} = \Psi\left(d\sqrt{\frac{1-c_1}{\nu}}, (\alpha+s)\sqrt{\frac{\nu}{1-c_1}}, \frac{T-t}{\nu}\right), \tag{B.21}$$

where d is now defined as

$$d = \frac{1}{s}\left(\ln\frac{A_t}{D} - (r)(T-t) + \frac{T-t}{\nu}\ln\left(\frac{1-c_1}{1-c_2}\right)\right). \tag{B.22}$$

Appendix C

Code

C.1 Option pricing

The code and the data used for option pricing in Section 5.4.1 can be found at: https://github.com/nurky17/OptionPricing-BS-VG-NIG.

C.2 Motivation

The code for making plots from Section 6.1 is accessible here: https://github.com/nurky17/ThesisMotivationCode.

C.3 EM algorithms

The code for the EM algorithms and the calculation of probability of default is to be found at: https://github.com/nurky17/Merton-NIG-Probability-of-Default.

Bibliography

Albrecher, H. and Predota, M. (2004). On Asian option pricing for NIG Lévy processes. *Journal of computational and applied mathematics*, 172(1):153–168.

Applebaum, D. (2009). *Lévy processes and stochastic calculus*. Cambridge University Press.

Back, K. (1991). Asset pricing for general processes. *Journal of Mathematical Economics*, 20(4):371–395.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae. *Scandinavian Journal of statistics*, pages 151–157.

Barndorff-Nielsen, O. E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling. *Scandinavian Journal of statistics*, 24(1):1–13.

Brambilla, C., Gurny, M., Lozza, S. O., et al. (2015). Structural credit risk models with Lévy processes: The VG and NIG cases. *Far East Journal of Mathematical Sciences*, 97(1):101–119.

Brown, M. B. (1974). Identification of the sources of significance in two-way contingency tables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 23(3):405–413.

Carr, P. and Wu, L. (2004). Time-changed Lévy processes and option pricing. *Journal of Financial economics*, 71(1):113–141.

Carter Jr, W. H. and Myers, R. H. (1973). Maximum likelihood estimation from linear combinations of discrete probability functions. *Journal of the American Statistical Association*, 68(341):203–206.

Chen, T. and Fienberg, S. E. (1976). The analysis of contingency tables with incompletely classified data. *Biometrics*, pages 133–144.

Cont, R. and Tankov, P. (2006). *Financial Modelling With Jump Processes*, volume 101.

Crosbie, P. and Bohn, J. (2019). Modeling default risk. In *World Scientific Reference on Contingent Claims Analysis in Corporate Finance: Volume 2: Corporate Debt Valuation with CCA*, pages 471–506. World Scientific.

Dar, A. A., Anuradha, N., and Qadir, S. (2019). Estimating probabilities of default of different firms and the statistical tests. *Journal of Global Entrepreneurship Research*, 9.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1):1–22.

Duan, J.-C. (1994). Maximum likelihood estimation using price data of the derivative contract. *Mathematical Finance*, 4(2):155–167.

Duan, J.-C. (2000). Correction: Maximum likelihood estimation using price data of the derivative contract (mathematical finance 1994, 4/2, 155–167). *Mathematical Finance*, 10(4):461–462.

Duan, J.-C., Gauthier, G., and Simonato, J.-G. (2005). On the equivalence of the KMV and maximum likelihood methods for structural credit risk models.

Duffie, D. (1988). Security markets: Stochastic models.

Eberlein, E. (2001). Application of generalized hyperbolic Lévy motions to finance. In *Lévy processes*, pages 319–336. Springer.

Eberlein, E. and Keller, U. (1995). Hyperbolic distributions in finance. *Bernoulli*, pages 281–299.

Eikon Refinitiv (2020). Company Market Capitalization and Current Liabilities, ©2018-2020 Refinitiv.

Fiorani, F. (2004). Option pricing under the variance gamma process. *Available at SSRN 1411741.*

Forouzanfar, M., Moghaddam, H., and Sona, G. (2008). Locally Adaptive Multiscale Bayesian Method for Image Denoising Based on bivariate Normal Inverse Gaussian Distributions. *IJWMIP*, 6:653–664.

Gerber, H. U., Shiu, E. S., et al. (1993). *Option pricing by Esscher transforms.* HEC Ecole des hautes études commerciales.

Goldfeld, S. M., Quandt, R. E., and Trotter, H. F. (1966). Maximization by quadratic hill-climbing. *Econometrica: Journal of the Econometric Society*, pages 541–551.

Goll, T. and Kallsen, J. (2000). Optimal portfolios for logarithmic utility. *Stochastic Processes and their applications*, 89(1):31–48.

Hartley, H. O. (1958). Maximum likelihood estimation from incomplete data. *Biometrics*, 14(2):174–194.

Healy, M. and Westmacott, M. (1956). Missing values in experiments analysed on automatic computers. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 5(3):203–206.

Hubalek, F. and Sgarra, C. (2006). Esscher transforms and the minimal entropy martingale measure for exponential Lévy models. *Quantitative finance*, 6(02):125–145.

Hull, J., Nelken, I., and White, A. (2005). Merton's Model, Credit Risk and Volatility Skews. *Journal of Credit Risk*, 1.

Jacod, J. and Shiryaev, A. (2013). *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media.

Jovan, M. and Ahčan, A. (2017). Default prediction with the Merton-type structural model based on the NIG Lévy process. *Journal of Computational and Applied Mathematics*, 311:414–422.

Kealhofer, S. (2003). Quantifying Credit Risk I: Default Prediction. *Financial Analysts Journal*, 59(1):30–44.

Ken-Iti, S. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge University Press.

Korn, R. and Desmettre, S. (2014). *Moderne Finanzmathematik–Theorie und praktische Anwendung*. Springer.

Krichene, N. (2005). Subordinated Lévy Processes and Applications to Crude Oil Options. *IMF Working Paper*.

Lucas Jr, R. E. (1978). Asset prices in an exchange economy. *Econometrica: journal of the Econometric Society*, pages 1429–1445.

Madan, D. B., Carr, P. P., and Chang, E. C. (1998). The variance gamma process and option pricing. *Review of Finance*, 2(1):79–105.

Madan, D. B. and Milne, F. (1991). Option pricing with V.G. martingale components 1. *Mathematical finance*, 1(4):39–55.

Madan, D. B. and Seneta, E. (1990). The variance gamma (VG) model for share market returns. *Journal of business*, pages 511–524.

Merton, R. C. (1974). On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. *The Journal of Finance*, 29(2):449–470.

Naik, V. and Lee, M. (1990). General equilibrium pricing of options on the market portfolio with discontinuous returns. *The Review of Financial Studies*, 3(4):493–521.

Prause, K. et al. (1999). *The generalized hyperbolic model: Estimation, financial derivatives, and risk measures*. PhD thesis, Citeseer.

Rydberg, T. H. (1997). The normal inverse Gaussian Lévy process: simulation and approximation. *Communications in statistics. Stochastic models*, 13(4):887–910.

Schoutens, W. (2003). *Lévy processes in finance: pricing financial derivatives*.

Tankov, P. (2011). Pricing and hedging in exponential Lévy models: review of recent results. *Paris-Princeton Lectures on Mathematical Finance 2010*, pages 319–359.

Wang, J. (2009). *The multivariate Variance Gamma process and its applications in multi-asset option pricing*. PhD thesis.

Wu, C. J. (1983). On the convergence properties of the EM algorithm. *The Annals of statistics*, pages 95–103.

Zieliński, T. (2013). Merton's and KMV models in credit risk management. *Studia Ekonomiczne*, (127):123–135.