POLITECNICO DI TORINO

Collegio di Ingegneria Gestionale

Corso di Laurea Magistrale in

Engineering **and Management**

Tesi di Laurea Magistrale

Lévy Process in finance and estimation of the Variance Gamma parameters

Supervisor:

Prof. Patrizia Semeraro

Student:

Simone Partenza

Contents

Chapter 1

Introduction

It is well known that there is a strong relationship between finance and mathematics, many people also believe that if anyone is so smart to understand and apply statistics on the prices quoted on the stock market will be able to become rich with no big effort. Sadly, the reality is not in this way, the aim of this thesis is to present in the most clear and easy way the mathematics and statistics underlying the stock prices statistical models, focusing on those that uses the stochastic processes as base.

The thesis is divided in two parts, the first one there will be introduced the mathematical theory behind those models and the second part presents an application of that with a real data sample. In this section one of the most used process to model the assets returns, the Variance Gamma, is estimated starting from a data sample of the MSCI US Investment Market Indices from January 2, 2009 to May 31st, 2013.

In the next chapters the thesis will present one of the most used family of statistical process used for financial modelling purposes, the Lévy process. They have many interesting properties, the most known and visible is about their paths with respect to the time. In fact, their trajectories present many discontinuities that are distributed randomly over the time. This means that in a random time T the trajectories of such processes suddenly jump up or down. This characteristic can be interpreted with the sudden jumps of the asset prices quoted on the stock market due to unforeseen events related or to episodes that interest or the entire market or the single firm related to the financial instrument.

In the specific, in the first chapter there is going to be a presentation of common statistical models, with a particular focus on the most known and famous, the Black Scholes Merton model, presenting how it is structured, its statistical properties and results (the BSM formula for vanilla options). Moreover, in the last section there is going to be a part related to its defects and misalignment with the evidences that come from the real data. In the second chapter the Lévy processes are presented, starting from the simplest one, the Poisson process, and going deeper with the mathematical theory underlying the models that are based on these processes. Specifically, after introducing the theory of the Lévy processes, it is going to be presented two subclasses of this family, the infinite activity process and the subordinator Lévy processes. The former are processes with the property to be composed by many small jumps while the latter are processes that have only positive jumps and can be interpreted as a financial time that changes its unity width basing on what happen in the market. The aim of these sections is to present the difficult mathematical notions in the most clear and effective manner, keeping the formality and the mathematical notation used in the many references used to develop the thesis. The last part is related to the specific Lévy process that is going to be used in the second part of the thesis, the Variance Gamma. After present how it can be build, the section shows all its statistical properties such as mean, variance, Skewness and Kurtosis that will be useful in the following chapters.

In the second part of the thesis we discuss the estimation of the Varianc Gamma process on the real data. The third chapter introduces the estimation methods i.e. the Moments method and the Likelihood method that are the most known and used in this task.
In the last chapter we perform the estimation on data. Specifically, we estimate the Variance Gamma process using the Moments method. Data are elaborated with Matlab.

Chapter 2

The Brownian motion based models

One of the possible application of the stochastic processes is the attempt to represent the behaviour of financial assets quoted in the stock market. For sure the most used, popular and usually building block for many models is the Brownaian motion stochastic process which can be seen as the dynamic representation of the Normal distribution with respect to the time.

Definition Given a probability space $(\Omega; \mathcal{F}; \mathbb{P})$, if a process $(X_t)_{t\geq 0}$ has:

- $X_0 = 0$ almost surely
- X_t has independent increments
- X_t has stationary increments
- $X_{t+s} - X_t$ is a random variable that follows a Normal distribution with mean 0 and variance s

This process has many interesting properties that make it very useful in many fields, finance included. The Brownian motion belongs to the family of processes that sharing the Martingale property:

Definition A process $(X_t)_{t\geq 0}$ defined on a probability space $(\Omega; \mathcal{F}; \mathbb{P})$ is said to have the Martingale property if for each $t > s$:

$$E[X_t|\mathcal{F}_s] = E[X_t|X_s] = X_s \tag{2.1}$$

The interpretation of this property lies behind the interpretation of the filtration $\mathcal{F}$ of the probability space. In a stochastic process $(X_t)_{t\geq 0}$, if the variable t can, as usual, interpreted as the time, the filtration of the probability space can be related to the information on the process past. Given this hypothesis, if a process owns the Martingale property it means that, knowing its history up to time s, the expected value of the process at time t is the value assumed by the variable at time s.

Taking in consideration the path of a Brownian motion is possible to prove that is always continuous with no jumps. The trajectories are, however, very irregular and everywhere, among its domain, the Brownian motion is not differentiable. Moreover, taking the following definition:

Definition Given a function $f(x)$ defined in an interval $[a;b]$ and each partition $[x_0 = a, x_2, ...x_n = b]$ of the interval, the variation of the function is defined as:

$$V_{[a;b]}(f) = \sup_{n\in\mathbb{N}} \sum_{i=1}^{n} |f(x_i) - f(x_{i-1})| \tag{2.2}$$

It is possible to prove that the Brownian motion has infinite variation everywhere ($V_{[a;b]}(f) = \infty$ for each closed interval). This can hamper the definition of the integrals in the Reimann sense respect to the Brownian motion and, thus, require a new definition of stochastic integral. The paths range between all the numbers, both positive and negative, that can be seen using the following property:

$$P(\inf(B_t) = -\infty \quad and \quad \sup(B_t) = +\infty) = 1$$

Another property of the Brownian motion is the selfsimilarity:

Definition A process $(X_t)_{t\geq 0}$ is said to be selfsiliar if:

$$\forall a > 0 \;\; \exists b(a) > 0 \;\; such \;\; that: \quad \frac{X_{at}}{b(a)} \sim X_t$$

In the particular case of the Brownian motion (without drift), it is possible to assert that:

$$\frac{W_{at}}{\sqrt{a}} \sim W_t \tag{2.3}$$

Conversely, if we consider a Brownian motion with drift $S_t = \mu t + W_t$ the property holds only by adding to the equation a translation:

$$\frac{S_{at}}{\sqrt{a}} \sim \mu\sqrt{a}t + W_t \tag{2.4}$$

One of the first model proposed by Louis Bachelier that uses the Brownian motion ($(W_t)_{t\geq 0}$) to describe the stochastic paths of the stock prices was:

$$S_t = S_0 + \sigma W_t$$

Since this model can give rise to negative prices, obviously meaningless in a financial context, it has been improved by the Black-Scholes-Merton one. The model uses the so-called Geometric Brownian Motion process and it states that not the price, S_t, follows a Brownian Motion but the stock returns. Taking into consideration a small time span $[t; t + \Delta t]$ the stock price will change of $S_{t+\Delta t} - S_t$ and dividing this amount by the stock price at time t, it is possible to obtain the percentage return. The Black-Scholes-Merton model assumes that the return of a stock price is composed by two parts, i.e. the systematic and the random. The former is given by a constant μ multiplied by Δt while the latter is given by a Brownian Motion variation, ΔW_t, multiplied by a coefficient, σ, called volatility that can be interpreted by the "degree of randomness" implied in the stock price. Reducing the time span Δt to zero and, with an abuse of notation since the no differential exists for the Brownian motion, the following stochastic differential equation can be written as:

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{2.5}$$

By applying the theory of the stochastic differential equation (that is beyond the purpose of the text) it is possible to prove that it has the following single solution:

$$S_t = S_0 \exp[(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t] \tag{2.6}$$

The distribution of this variable is called *Lognormal*, as the logarithm follows a normal distribution with parameters $((\mu - \frac{1}{2}\sigma^2)t; \sigma^2 t)$. An example of this distribution can be seen in the figure 1.1:

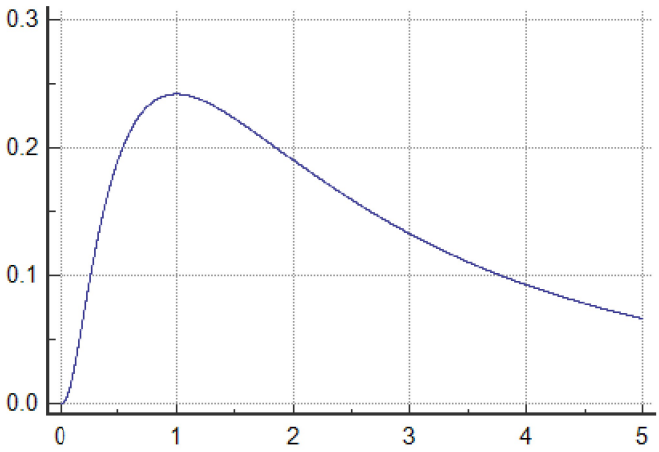

Figure 1.1: The Lognormal distribution

Considering the distributional characteristic of the model the stock price possesses the following statistical properties:

- Expectation: $E(S_t) = S_0 e^{\mu t}$
- Variance: $VAR(S_t) = S_0^2 e^{2\mu t}(e^{\sigma^2 t} - 1)$
- Kurtosis: $e^{4\sigma^2} + 2e^{3\sigma^2} + 3e^{2\sigma^2} - 3$

- The curve is asymmetric with respect its expected value

As we can observe, the Kurtosis is higher than a Gaussian distribution (equal to 3), thus the Black-Scholes-Merton model describes a process that has thin tails that, in financial terms, it means that extreme events are very rare. An example of the trajectory of this process is represented in the picture 1.2:

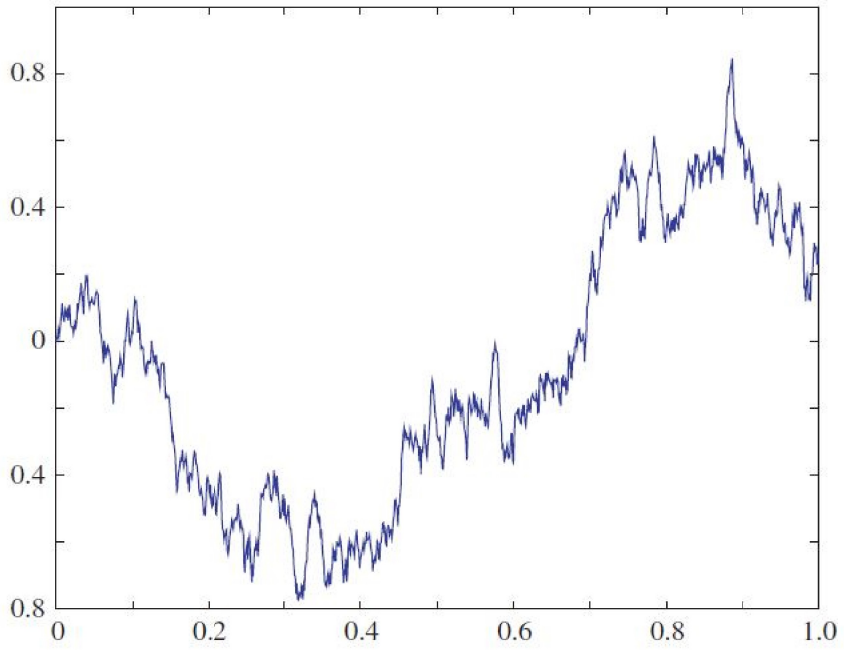

Figure 1.2: The Brownian Model trajectory (ref. Schoutens: Lévy process in finance)

The market described by the Black-Scholes-Merton model is complete, with no frictions and just two quoted assets. The first is of course that according which the model is trying to describe the trajectory while the other is the risk-free asset. Regarding the latter, the model assumes a rate r constant in the short term with a value after t that is $B_t = e^{rt}$.

2.1 The diffusion model limits

The Black-Scholes-Merton model is one of the most famous in the financial world since it allows to find an explicit formula for the price of Vanilla call and put options. Unfortunately the characteristics of this model have many discrepancy to the empirical data observed. The first thing that can be easily inferred by the empirical data is the presence of jumps in the prices' sample paths. While a diffusion model such the Black-Scholes-Merton has continuous trajectories that do not allow for sudden, instantaneous big changes, the price behaviour observed seems to presents them. Of course, since the observation are made at discrete time, it is impossible to certainly assert that the real sample paths effectively presents discontinuity. However, the aim of this a model is maximize the likelihood with respect to the reality at a specific time scale. Another critical point indicating that the Black scholes-Merton model does not represent the reality with acceptable precision, is the size of its tails. This often does not accurately reflect reality as stock prices fluctuate due to external unexpected events. However,it is possible to figure out this problem by remaining in the class of diffusion processes. Instead of presenting a model in which the volatility is a constant over the time, in fact, is possible to make it a random variable with its own distribution as it can be both dependent and independent from the asset's price. Thus, improving the Black-Scholes-Merton model in such a way to affect size tail in order to better reflect reality. This solution, however, has the shortcoming to have highly variable diffusion coefficient in local volatility models or a unrealistic value of "volatility of the volatility" in models in which the volatility is represented by a diffusion-based model.

As pointed out above, the Black-scholes-Merton assumes that the prices follows a Log-normal distribution. This model possesses distributions with very thin tails, indicating that extreme events are very rare to happen. Schoutens proved that modelling the log-price as a normal distribution is going to produce results that are statistically unacceptable. By studying the statistical properties of many indexes of the stock market, evaluated in the period 1997-1999, he demonstrated that none of the indexes log-returns can be described by a Normal distribution. The first evidence is given by the Skewness: in a Normal distribution this value should be 0, however, the value of this parameter has a minimum of -0.2116 for the CAC-40 index and a maximum of -0.5439 for the S&P500.

Concerning the analysis of the Kurtosis of the log-return of the indexes, the results show an even more

clear evidence that the Normal distribution does not represent well enough the indexes behaviour in the real market. For a normal distribution the value of the Kurtosis should be equal to 3, although the collected data present a considerably higher values. The extreme values are again obtained by the S&P500 for the maximum (6.94) and the CAC-40 that present the closest estimation of the Kurtosis to the Normal distribution (4.63). In particular, Shoutens obtained the following result respect to analyses indexes:

Index	Mean	Standard Deviation	Skewness	Kurtosis
S&P500	0.0009	0.0119	−0.4409	6.94
Nasdaq-Composite	0.0015	0.0154	−0.5439	5.78
DAX	0.0012	0.0157	−0.4314	4.65
SMI	0.0009	0.0141	−0.3584	5.35
CAC-40	0.0013	0.0143	−0.2116	4.63

Given the empirical collected data, it is possible to obtain an approximation of the empirical density with the Kernel estimator. This approach approximate the density $f(x)$ of a random variable X using the following formula:

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x_i - x}{h}) \tag{2.7}$$

Where $K(x)$ represents the so-called Kernel function. Basically this method smooths each data collected in the sample into a small density bumps which are then added in the final density approximation $\hat{f}(x)$. The precision of the method depends on the choice of the parameter h, called the *smoothing bandwidth*, that regulates the "degree of smoothness" of the approximation. Using the Silverman's role of thumb, is usually obtained a good result using $h = 1.06\sigma n^{\frac{1}{n}}$, where σ is the standard deviation and n is the sample size. About the Kernel function, there exists many of them that all gives acceptable results and one of the most used is the so called Gaussian kernel:

$$K(x) = \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$$

Taking into account the S&P500 index and plotting both the Kernel density and the Normal using as parameters empirically obtained the mean and the standard deviation the results clearly shows that it will be a very bad fit using the Gaussian distribution for the log-returns. In particular, considering the behaviour in the tails of the log-return is clear that, where the log density of the normal distribution has a quadratic decay, using the Kernel distribution it seems to be more linear as can be seen in picture 1.3.

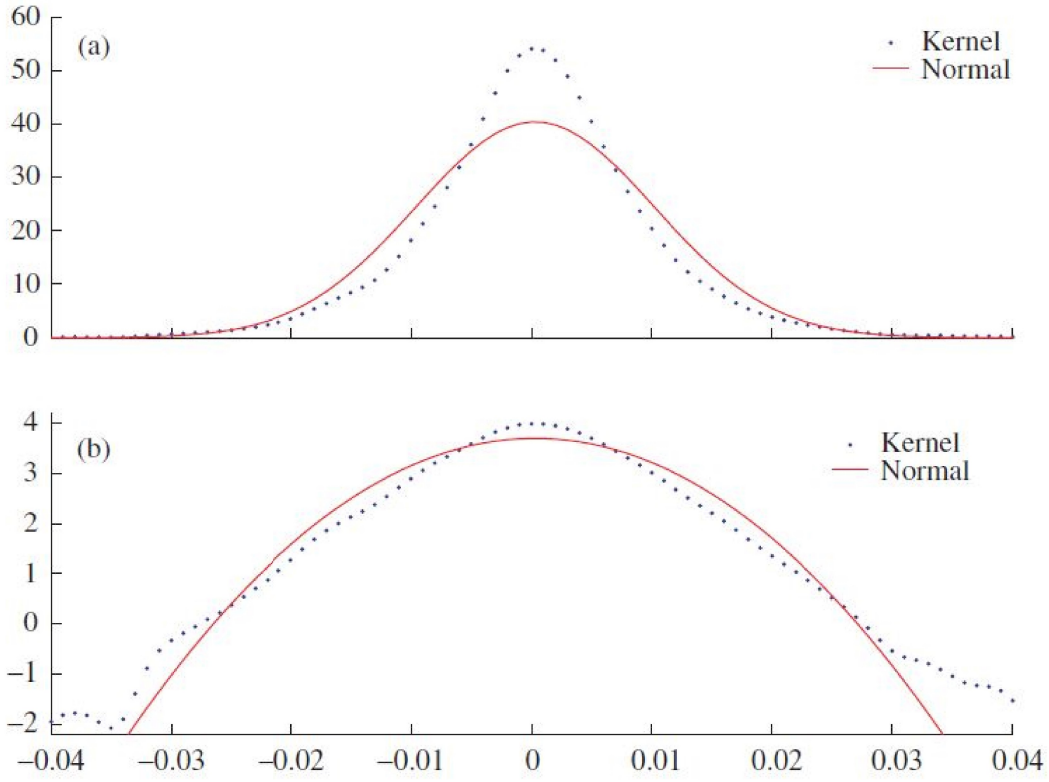

Figure 1: (a)Normal Gaussian Kernel density (b)log-density of the daily log returns of the S&P500 Index(ref. Schoutens: Lévy process in finance)

In order to rigorously demonstrate that the Normal distribution for the log returns does not give a good fit, the hypothesis testing could be the most appropriate. In particular, to test if a density function fairly fits with the data empirically collected, χ^2-test is the most used. This test evaluates the number of points falling in certain intervals and then compare them with those expected. More precisely, given n different (and statistically independent) observation of the random variable, the method divides the support space

(in the case of the log return it is all $\mathbb{R}$) in subset with the same width $(A_1, A_2...A_m)$ and then compares the empirical frequency with the theoretical one. Chosen a particular subset A_k and counting the number of observations that fall inside N_k, in fact, it can be compared with $\pi_k = P(X \in A_k)$ according to the null hypothesis. The comparison can be made by calculating for each subset $\frac{(N_k - \pi_k)^2}{n\pi_k}$ and then summing up all together, thus leading to:

$$\hat{\chi^2} = \sum_{i=1}^{m} \frac{(N_i - \pi_i)^2}{n\pi_i} \tag{2.8}$$

The variable χ^2 follows asymptotically a χ^2 distribution with $m - 1$ degrees of freedom. The *p-value* is calculated as:

$$p = P(\chi^2_{m-1} > \hat{\chi^2})$$

The *p-value* is the probability that values are even more extreme then the specific statistic test and very small values of the *p-value* leads to the rejection of the null hypothesis. Shoutens sets 0.05 as rejection level and accordingly, none of the considered indexes ha a *p-value* that allows to accept the null hypothesis, in particular:

index	*p-value*
S&P500	0.0421
DAX	0.0366
Nasdaq-Composite	0.049
CAC-40	0.0285
SMI	0.0479

Observing the stock price behaviour on the market, it is possible to observe that, for many of them,

the statistical properties are not time invariant. This means that collecting data with a time span Δ and analysing them in order to obtain a model, it is not going to come up with a statistical consistent model with respect to a measure of 2Δ . From a more general point of view it has been observed thatthe more the time scale increases, the more the return distribution is similar to a normal one. This means that by increasing the time scale the model should reduce its heavy tails, that is something that a diffusion model such Black-Scholes-Merton does not allow.

The last point supporting the fact that the Black-Scholes-Merton model is not precise respect to the reality, it is the volatility assessment. Taking the market of vanilla call and put option, being the price of these derivatives freely quoted, it is possible to plug them in the Black-Scholes-Merton formula:

$$C_t^e(t, K, \sigma, S_t) = S_t\Phi(d_1) + Ke^{-r(T-t)}\Phi(d_2) \tag{2.9}$$

Where:

$$d_1 = \frac{ln(\frac{S_t}{K}) + (r + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

$$d_2 = \frac{ln(\frac{S_t}{K}) + (r - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

Given the prices from the market is possible to obtain σ^2, usually called implied volatility. Ranging the prices for different maturities, the hypothesis of the model does not hold: in fact, while the model foresees a constant volatility independent from both strike price and time-to-maturity, the reality shows an implied volatility surface (function of both variables) that is not constant at all.

Given all these empirical facts and criticisms, the need to define a new model that can better fit with the reality. Lèvy processes are in the frame as they allow to have jumps in trajectories as well as have heavy tails. Moreover, at the variance with the Black-Scholes-Merton model, it represent a non complete

market in which complete hedging is impossible and it will always be residual risk.

Chapter 3

Lèvy Processes

3.1 The Poisson process

The Poisson process is the simplest stochastic process with jumps. It can be seen as the building block of the more complex jump process grouped by the family of the Lèvy processes.

The Exponential random variable One of the most important and known random variable is the exponential one. It is well known for its properties that allow to build the Markov processes with jumps. The variable following such density identified by the parameter λ:

$$f(x) = \lambda e^{-\lambda x} 1_{x \geqslant 0} \tag{3.1}$$

Integrating the density in its domain $[0; +\infty]$, it is possible to obtain:

$$F(x) = \int_0^{+\infty} \lambda e^{-\lambda x} dx = 1 - e^{-\lambda x} \tag{3.2}$$

One of the most known and used property of the Poisson random variable is the so called *absence of memory property*:

Theorem 3.1.1 *Given a random variable X with the property: $P\{X > t+s|X > t\} = P\{X > s\}$, the variable X has an exponential distribution.*

Such property of the exponential random variables is used to build a process that has no path dependence. it means that no matter where the process has been in the past to determine the probability with respect of its position in the future. These processes, called Markov processes, are very known and used.

Poisson distribution Going over with the useful distribution in order to define the family of the Lévy processes an other important distribution is the Poisson. A random variable X that follows a Poisson distribution is an integer variable with the below reported probability law identified by the parameter λ :

$$P\{X = n\} = e^{-\lambda}\frac{\lambda^n}{n!} \tag{3.3}$$

while its moment generating function is:

$$M(u) = exp[\lambda(e^u - 1)] \tag{3.4}$$

This distribution is related to the exponential distribution and allows to create what is called the Poisson process. It is in fact possible to proof that:

Theorem 3.1.2 *Given n i.i.d exponential random variables $\{\tau_1, \tau_2 ... \tau_n\}$ the variable defined:*

$$N_t = inf\{n \geq 0, \sum_{i=0}^{n} \tau_i\}$$

Follows a Poisson's distribution with parameter λt

$$P\{N_t = n\} = e^{\lambda t)}\frac{(\lambda t)^n}{n!} \tag{3.5}$$

Moreover, the Poisson distribution has the property to be closed with respect to convolution: this means that the sum of independent Poisson variables leads to a Poisson variable with parameter $\lambda = \lambda_1 + \lambda_2$. From that we can identify a family of random variables called *infinite divisible variables*. Taking the example of the Poisson, each random variables that follows this distribution with parameter λ can be seen as the sum of n i.i.d. Poisson random variables with paramenter $\frac{\lambda}{n}$. Taking the limit is possible to see an infinite divisible random variable as an infinite sum of i.i.d. random variables.

The Poisson process The building block of a process with jumps is the Poisson process that has increasing trajectory piecewise constant with jumps.

Definition Given $(\tau_i)_{i\geq 1}$ a sequence of i.i.d exponential random variables with parameter λ and $T_n = \sum_{i=1}^{n} \tau_i$ (usually called random time) the Poisson process $N(t)$ is defined as:

$$N(t) = \sum_{n=1}^{+\infty} 1_{t\geq T_n} \tag{3.6}$$

Given the definition of the Poisson process is possible to obtain the following interesting properties:

- For any $t \geq 0$, N_t is almost surely finite
- For any possible scenario, N_t is piecewise constant with jump size equal to 1
- The sample path are left-continuous (usually this property is called with the abbreviation *cadlag*)
- N_t is continuous in probability (if $s \longrightarrow t$ $P\{N_s\} \longrightarrow P\{N_t\}$)

- For any t, as pointed out by 2.1.2, N_t follows a Poisson distribution with parameter λt:

$$P\{N_t = n\} = e^{\lambda t}\frac{(\lambda t)^n}{n!} \tag{3.7}$$

  and has an expected value $E(N_t) = \lambda t$

- From the properties of the Poisson distribution, it is possible to obtain the characteristic function:

$$\Phi(u) = E[e^{iuN_t}] = exp\{\lambda t(e^{iu} - 1)\} \tag{3.8}$$

- N_t has independent increments: it means that $\forall t_1 < t_2 < .... < t_n$ the increments $N_{t_{i+1}} - N_{t_i}$ $\forall i$ are independent of each other.

A particular mention is worth to be done for the so-called *cadlag* property. In mathematical terms it means that $N_t = N_{t+}$, and this property requires that the jumps are at most countable and the number of large one must be finite. The choice to have *left-continuous* sample paths comes from the financial interpretation. If we would have a *right-continuous* function, it means that for each t (interpreted as time) the value of the function follows its sample path. This means that in each instant the value of the random variable, given the sample path, is predictable. This, of course, does not happen with the *left-continuous* function, as the jumps are interpreted as sudden, unforeseen events at time t.

The Poisson process can be somehow generalized within the class of the counting processes. A counting process has the same form of a Poisson process, but the distribution of the sum of i.i.d. random variables that compose the random time, can be different from the exponential one. From this point of view, it is possible to see the Poisson process as a particular case of the more general counting processes family, although it can be demonstrate that only the Poisson process is a counting process with independent and stationary increments.

3.2 Random Measure

An other possible point of view of a counting process consists in a count of the number of jumps in a deterministic time span $[0;t]$and, following this idea, given N_t a counting process:

$$N_t = \sharp\{Jumps\} = \{i \geq 1, T_i \in [0;t]\}$$

Given a particular event ω belonging to the probability space Ω and a set A, is possible to set the measure of A as the number of jumps such that $T_i(\omega) \in A$. Such a measure does not depend only from the set A but also on the particular event ω in which the counting process is. For this reason an interesting information of this measure can be the expected value - rather then the particular-value - in a given scenario: Choosing the poisson process, the expected value is easy to calculate:

$$E(M(A)) = \lambda \int_A ds = \lambda |A| \tag{3.9}$$

Where $|A|$ is the Lebesgue measure of the set.

Using the properties of the integral, it is possible to define the random measure as the "derivative" of the Poisson process that is nothing but the sum of Dirac measure located in the jumps.

The idea of the random measure can be generalized in order to define an important class of measure: the Poisson random measure. Taking a Radon measure μ on a set $E \in \mathbb{R}$ provided with a sigma algebra ε, it is possible to define the following measure:

Definition

$$M : \Omega \times \varepsilon \to \mathbb{N}$$

such that:

- $\forall \omega \in \Omega$, $M(\omega, A)$ (with $A \in \varepsilon$) is a Radon measure

- $\forall A$ measurable, $M(.,A)$ is a Poisson random variable with parameter $\mu(A)$:

$$\forall k \in \mathbb{N} \quad P\{M(A) = k\} = e^{-\mu(A)} \frac{(\mu(A))^k}{k!}$$

- For disjoints measurable sets the poisson random measures are independent random variables

The idea behind the Poisson random measure is to create random points using the Poisson variables whose intensity is related to the Radon measure μ used. It should in fact exist a process $\{X_n\}_{n\geq 1}$ according which:

$$M(A) = \sum_{n\geq 1} 1_A(X_i)$$

Using the random measure, it is possible to give a more concise and clear definition of the Poisson process:

$$N_t(\omega) = \int_0^t M(\omega, ds) \tag{3.10}$$

This means that the Poisson random measure create a random series of points and then it counts them. Given this view the Poisson random measure is anything else then a sum of Dirac measure located in these points.

3.3 Definition and properties of the Lévy processes

Using all the informations above, it is possible to define a very broad class of processes: the Léevy processes. The Poisson process and the Brownian motion both belong to this class and, moreover, they can be seen as building block of every Lévy process.

Definition Given a probability space Ω, a sigma-algebra related to it and a probability measure, a *cadlag* stochastic process $(X_t)_{t\geq 0}$ with $X_0 = 0$ is said to belong to the class of Lévy processes if:

- It has independent increments
- It has stationary increments
- Stochastic continuity: $\forall \varepsilon \lim_{h\to 0} P\{X_{h+t} - X_t \geq \varepsilon\} = 0$

A particular mention need to be done for the third property, the stochastic continuity. This property does not mean that there is not any discontinuity points at all (if so there would not be jumps), but that they are isolated. The third property, in fact, just indicates that the probability to have jumps in a specific point is 0, this is true due to the continuity of the distribution of the discontinuity. Because the process X_t is a continue time process, also the distribution of the discontinuity is in continuous time, that implies that the probability of having a discontinuity in a specific point is 0.
If samples of a Lévy process are taken with time interval of Δ, given a specific t, the value of the process in this point, X_t, can ber written as the sum of all the increments from 0 to t with size Δ. Due to the i.i.d increments and the time continuity of the process (that allows for sampling with different δ) it is possible to state that the distribution of X_t, that can be seen as the increments $X_t - X_0$, correspond to each increment. This can be summed up with the infinitely divisible distribution. Since this property must belong to each Lévy process it gives restrictions on the possible increments distribution chosen.
Another property that arises from the first and second property of the Léevy processes provided by the definition, is the exponential form of the characteristic function. Given a Lévy process $(X_t)_{t\geq 0}$, the characteristic function of X_t is:

$$\Phi(z) = E(e^{izX_t})$$

for each $s > t$, the distribution of X_{t+s} can be seen as the sum of the increments $X_t - X_0$ and $X_{t+s} - X_t$. Using both the independence and the stationary properties of the Lévy processes jointly with the prop-

erties of the characteristic function of a sum of independent random variables it is simple to prove that:

$$\Phi_{X_{t+s}}(z) = \Phi_{X_t} \cdot \Phi_{X_s} \tag{3.11}$$

This means that the characteristic function of a Lévy process can be written as an exponential function:

$$\Phi_{X_t}(z) = e^{t\psi(z)} \tag{3.12}$$

Compounded Poisson process One of the most important and used Lévy process is the compounded Poisson process:

Definition Given a Poisson process N_t with intensity $\lambda > 0$ and a jump size distribution f, the compounded Poisson process is defined as:

$$X_t = \sum_{i=1}^{N_t} Y_i \tag{3.13}$$

Where Y_i are i.i.d. random variables with distribution f.

The idea that is behind the compounded Poisson process to create a random process with piecewise constant sample path with random *cadlag* jumps possessing random size (both positive and negative). The Poisson process N_t defines the random time in which the jumps occur while the size is given by the distribution of the Y_i random variables. As pointed out before, since all the not null increments are due to the jump size distribution, it must have the property of infinite divisibility. A very common example is a compounded Poisson process with Gaussian jumps.

The characteristic function of a compounded Poisson process $(X_t)_{t\geq 0}$ defined on $\mathbb{R}$ is:

$$\Phi(z) = E(e^{iz.X_t}) = \exp \lambda t \int_{\mathbb{R}} (e^{iz.x} - 1) f(dx) \tag{3.14}$$

Given the characteristic function, remembering 2.8, it is possible to see the compounded Poisson process as a superposition of independent Poisson processes with different intensity (equal to $\lambda f(dx)$).
To study the behaviour of the jumps in a Lévy process the notion of random measure can be used. Given a set of $\mathbb{R} \times [0; +\infty[$ A is possible to define a random measure associate to a Lévy process $(X_t)_{t\geq 0}$ can be described as follows:

$$J : \mathbb{R} \times [0; +\infty[\to \mathbb{N}$$

$$J_X(A) = \sharp\{(t, X_t - X_{t-}) \in A\} \tag{3.15}$$

This function counts the number of jumps occurring in the process X_t such that their sizes are in A.
J_X can be used to build the jump measure for the compounded Poisson processes:

Definition Given a compounded Poisson process $(X_t)_{t\geq 0}$ with intensity λ and jumps distribution f, J_X is a Poisson random measure with intensity: $\mu(dx \times dt) = \nu(dx)dt = \lambda f(dx)dt$

$\nu(x)$ is called Lévy measure, thus, to define it for all the Lévy processes, is possible to figure out it as the expected number of jumps in the unit time:

Definition Given $(X_t)_{t\geq 0}$ a Lévy process, its Lévy measure $\nu(x)$ is defined as:

$$\nu(A) = E(\sharp\{t \in [0; 1], X_t - X_{t-} = \Delta X_t \neq 0, \Delta X_t \in A\}) \tag{3.16}$$

From this, the following compounded Poisson process using its Lévy measure can be obtained:

$$X_t = \int_{[0;t]\times\mathbb{R}} x J_X(dt \times dx) \tag{3.17}$$

3.4 Infinite activity processes

As previously pointed out, the Lévy measure defines a measure that, due to the *cadlag* property, is almost surely finite. Using the definition 2.16 it can be deduced that it is certainly a finite measure in the set $\mathbb{R} \setminus \{0\}$; getting closer to zero, however, the situation may change. In fact, using the same definition, the Lévy measure $\nu(A)$ counts the expected number of jumps, in the unit time, with the size given by A. Thus, if A tends to zero the Lévy measure counts the number of jumps with a very small size. Is still possible, however, to have a Lévy process in which the behaviour of the Lévy measure presents infinite number of jumps, when the set measured get closer to zero. In this case, the process is called Lévy process with infinite activity, moreover, to figure out how such a process works, it is sufficient to imagine it as something moving with small infinite jumps. Such idea give rise to the famous *Lévy-Ito decomposition*:

Theorem 3.4.1 *Given a Lévy process $(X_t)_{t\geq 0}$ with a Lévy measure ν such that:*

- $$\int_{|x|\leq 1} |x^2|\nu(dx) < \infty, \quad \int_{|x|\geq 1} \nu(dx) < \infty \tag{3.18}$$
- *The Jump measure J_X of the Lévy process is a Poisson random measure with intensity $\nu(dx)dt$*

It Always exist a d dimensional vector γ and a Brownian motion B_t with a covariance matrix A that:

$$X_t = \gamma t + B_t + X_t^l + \lim_{\varepsilon\to 0} X_t^\varepsilon \tag{3.19}$$

Where:

- $$X_t^l = \int_{|x|\geq 0, s\in[0,t]} x J_X(dx \times ds) \tag{3.20}$$

- $$X_t^\varepsilon = \int_{\varepsilon\leq|x|\leq 1, s\in[0,t]} x(J_X(dx \times ds) - \nu(dx)ds) \tag{3.21}$$

The decomposition 2.19 allows to see any Lévy process as a sum of a Brownian motion with drifts, that is a continuous process, and a superposition of compounded Poisson process that are picewise constants with jumps. In order to consider the case of infinite activity process, the process X_t^ε needs to be added. For each ε the process X_t^ε is a compounded Poisson, but it is conceivable to take the limit for ε that tends to 0, since ν has a singularity in 0. Since this limit cannot converge to the jump measure J_X should be substituted with a center one ($E(J_X) = \nu$) in order to ensure it.

Using the Lévy-Ito decomposition it is easy to understand that each Lévy process is uniquely defined by three parameters: γ, A, ν. All together,these three elements, are called Lèvy triplet. Using this decomposition a compact formula also for the characteristic function can be written and this equation is called *Lévy-Khinchin representation*:

Theorem 3.4.2 *Given a Lévy process with Lévy triplet (γ, A, ν), its characteristic function can be written as:*

$$\Phi(z) = E(e^{iz.X_t}) = e^{t\psi(z)}$$

Where:

$$\psi(z) = -\frac{1}{2}z.Az + i\gamma.z + \int_{\mathbb{R}} (e^{iz.x} - 1 - 1_{|x|\leq 1} \cdot iz.x)\nu(dx) \tag{3.22}$$

A Lévy process is said to have a finite variation if its trajectories are functions with finite variation (a function f has finite variation when $\sup \sum_{i=1}^{n-1} f(t_{i+1}) - f(t_i)$ is fintite for each partition of each finite sets). A finite variation Lévy process is endowed with many interesting properties and in particular is possible to said that:

Theorem 3.4.3 *A process $(X_t)_{t\geq 0}$ with finite variation and a characteristic triplet (A, ν, γ) has $A = 0$ and finite activity ($\int_{|x|\leq 1} |x|\nu(dx) < \infty$).*

Given this result and using the Lévy-Ito decomposition, it is easy to prove that a Lévy process with finite variation can be represented as:

$$X_t = bt + \int_{[0;t]\times\mathbb{R}} xJ_X(ds \times dx) \tag{3.23}$$

This means that each process with finite variation can be represented as a linear trend (bt) plus a pure jump process.

3.5 The subordinator Lévy processes

A very interesting class of Lévy processes are those with increasing trajectories. The most simple example is the Poisson one, whose paths are piecewise constant with a sudden jumps of size 1, as previously explained. These kind of processes are important since they can be used as the time variable of other Lévy processes. The financial intuition laying behind this kind of representation is that the volume of asset traded,in different moments, is different. In particular, it is conceivable to note that, even during each day, there are moments in which the assets volume traded in a particular market is lower than others. This empirical fact can be modelled changing the time of the stochastic process used: instead of assuming to have a constant "speed" you can suppose that the time, when the volume of trade is higher,

somehow accelerate. Thus, is possible to have a different "length" of the time unit (that is, of course, stochastic) with a constant volume of trading, instead of having a constant time and a variable volume of trading. It is possible to prove that the following statement, are equivalent definition of a subordinator X_t:

- $X_t \geq 0$ almost surely for some $t > 0$
- $X_t \geq 0$ almost surely for each $t > 0$
- Sample paths of X_t are almost surely increasing
- the characteristic function satisfy: $A = 0$, $\nu([-\infty; 0]) = 0$ and $\int_0^{+\infty} min\{x; 1\}\nu(dx) < \infty$ with $b \geq 0$.

The last point states that a subordinator is also a finite variation Lévy process and, moreover, it has finite activity. Given a subordinator process $(S_t)_{t\geq 0}$ with a characteristic triplet $(0, \rho, b)$, since it is a.s. positive in order to describe it the Laplace transformation can be used instead of the Fourier one. Following this argument is possible to write:

$$E(e^{uS_t}) = e^{tl(u)} \quad \forall u \leq 0 \quad with \quad l(u) = bu + \int_0^{\infty} (e^{ux} - 1)\rho(dx) \tag{3.24}$$

The use of the subordination is a very useful tool, as it allows for financial interpretation of the time and it can be easily simulated and statistically manipulate. It has, however, some drawbacks, since not all the Lévy processes can be seen as a diffusion process with a change of time. In order to assert such property, it is mandatory that the Lévy measure respect the boundaries given by the following theorem:

Theorem 3.5.1 *Let ν be a Lèvy measure on $\mathbb{R}$ and μ a real constant. There exist a Lévy process $(X_t)_{t\geq 0}$ with ν as Lévy measure such that $X_t = B(Z_t) + \mu Z_t$ for some subordinator Z_t, only if ν has:*

- *ν is absolutely continuous with density $\nu(x)$*
- $\nu(x)e^{-\mu x} = \nu(-x)e^{\mu x}$

- $\nu(\sqrt{u})e^{-\mu\sqrt{u}}$ *is a completely monotonic function*

3.6 The Variance Gamma process

In order to understand how the subordinated processes can be used the family of the stable random variable must be defined.

Definition A random variable X is said to be stable if for every $a \geq 0$ exist two functions $b(a)$ and $c(a)$ such that:

$$\Phi_X(z)^a = \Phi_X(zb(a))e^{ic(a).z} \tag{3.25}$$

Moreover, the random variable is said to be *strictly* stable if:

$$\Phi_X(z)^a = \Phi_X(zb(a)) \tag{3.26}$$

It can be proven that for each stable distribution, exists a constant $\alpha \in [0;2]$ such that $b(a) = a^{\frac{1}{\alpha}}$ and for a Lévy process, the Lévy measure can be written as:

$$\nu(x) = \frac{A}{x^{\alpha+1}}1_{x\geq 0} + \frac{B}{|x|^{\alpha+1}}1_{x\leq 0} \quad with \quad \alpha \in [0;2] \tag{3.27}$$

This property, of course, can be applied also to subordinators, and, as only positive jumps are allowed in these processes by definition, the second part of the equation 2.27 can be omitted. Moreover, in order to temperate the size of the big jumps, is possible to multiply the Lévy measure by $e^{-\lambda x}$ leading to this Lévy measure for the processes called *Tempered Stable Subordinator*:

$$\nu(x) = \frac{c}{x^{\alpha+1}} e^{-\lambda x} \tag{3.28}$$

Thus, the Lévy measure of these processes is defined by three parameters that allow to keep under control the size of the jumps of the process. The parameter c gives the relative importance to all the jumps (big and small), and, in other words, it changes the time scale. With the parameter λ is possible to control the size of the big jumps by tempering them in order to avoid unrealistic discontinuity. Finally, with $\alpha \in [0; 1[$ it is possible to control the size of the small jumps. In equation 2.24, for processes that own the above Lévy measure, the Laplace exponent, $l(u)$, is in this form:

$$l(u) = c\Gamma(-\alpha)[(\lambda - u)^{\alpha} - \lambda^{\alpha}] \tag{3.29}$$

Where:

$$\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t} dt \tag{3.30}$$

In this case, the Laplace exponent is true only for $\alpha \neq 0$, when α assumes the value of 0, in fact, the formula changes in:

$$l(u) = -c \log(1 - \frac{u}{\lambda}) \tag{3.31}$$

The case with $\alpha = 0$ is called *Gamma process*, as it will be shown, it plays an important role in the creation of models of stock prices using the Lévy processes. The *Gamma process* has the following statistical properties:

- Lévy measure: $\nu(x) = \frac{ce^{-\lambda x}}{x} 1_{x \geq 0}$
- Laplace transform: $\rho(u) = (1 - \frac{u}{\lambda})^{-ct}$

- probability density: $f(x) = \frac{\lambda^{ct}}{\Gamma(ct)} x^{ct-1} e^{-\lambda x}$

Taking a Brownian Motion with drift $\mu t + \sigma B_t$ and changing the time with the Gamma process, the so-called *Variance Gamma process* can be obtained:

$$VG_t = \mu G_t + \sigma B(G_t) \tag{3.32}$$

Where G_t is a gamma process.

It can be proved that this new process is a Lévy one and it can be described by three parameters: (μ, σ, k). In particular, the statistical properties of the process in closed form can be described:

- Lévy Measure: $\nu(x) = \frac{1}{k|x|} e^{Ax - B|x|}$ Where: $A = \frac{\mu}{\sigma^2}$ and $B = \frac{\sqrt{\mu^2 + 2\sigma^2/k}}{\sigma^2}$
- Characteristic exponent: $\psi(u) = -\frac{1}{k} \log(1 + \frac{u^2\sigma^2 k}{2} - iu\mu k)$
- $E(VG_t) = \mu t$
- $VAR(VG_t) = \sigma^2 t + \mu^2 k t$

The Variance Gamma process can also be obtained following a different reasoning from the subordination of a Brownian motion. In particular, is possible to model the log-return of a stock price as a normal distribution with mean μ and variance $\sigma^2 V$, where V is a random variable that follows a Gamma distribution. The gamma distribution is no more that the underlying distribution of the gamma process. In particular, the Gamma distribution follows this density:

$$g(x) = \frac{\lambda^\gamma x^{\gamma-1} e^{\lambda x}}{\Gamma(\gamma)} \tag{3.33}$$

It is easy to see the similarity with the distribution of the Gamma process with the only difference that in this case γ is not a constant any more, but it depends on time ($\gamma(t) = ct$). From this interpretation of the

Variance Gamma process the density of the log-return price (assuming for seek of simplicity $\mu = 0$)can, thus, be obtained:

$$f(x) = \int_0^{+\infty} \frac{e^{\frac{x^2}{2\sigma^2\nu}}}{\sigma\sqrt{2\pi\nu}} g(\nu)d\nu \tag{3.34}$$

In order to get the characteristic function is enough to conditioning the variance V and for simplicity setting $\gamma = \lambda$, is possible to obtain:

$$\Phi(u) = [1 + \frac{\sigma^2 u^2}{2\lambda}]^{-\lambda} \tag{3.35}$$

Conditioning again the value of the variance V, is possible to calculate the moments of the Variance Gamma. The conditional expectation of the n^{th} moment X^n is:

$$X^n = a_n V^{\frac{n}{2}} \sigma^n \tag{3.36}$$

where a_n is $(n-1)(n-3)...(1)$ if n is even while it is 0 otherwise. From that, the second and fourth moments can be find:

$$E(X^2) = \sigma^2 \quad while \quad E(X^4) = 3\sigma^4(1+\nu)$$

Where $\nu = \gamma/\lambda^2$. From that, defining the Kurtosis as:

$$\kappa = \frac{E(X^4)}{\sigma^4} \tag{3.37}$$

It is easy to see that the Kurtosis of the Variance Gamma is equal to $3(1+\nu)$. Comparing it with the Kurtosis of a normal distribution, it is clear that the Variance Gamma distribution has fatter tails then the normal.

The importance of the Variance Gamma process in modelling the log-return of the stock price is based on its ability to fulfil the properties that a good stock price model should have:

- Tails fatter than the normal distribution in the daily returns, while in the long run it must behave as the Gaussian.
- Finite moments for at least the lower powers.
- The process must be stationary with independent increments no matter the time span taken in consideration.
- Extension to the multivariate processes with elliptical distribution that thereby maintain validity of the CAPM.

The Variance Gamma model fits with all these four properties. Moreover, the process moves in time with many small jumps and the number of discontinuities approaches to infinity, while their sizes is progressively concentrate in the origin. This behaviour fits with the intuition underlying the sample path continuity that is the base of the Brownian motion models. Furthermore, empirically, the Variance Gamma model seems to fits with the evidences that come from the real market. *Madan and Seneta* in 1987 compared the Variance Gamma process with the normal and the Press compound events model. In order to evaluated them they took 19 different stocks quoted in the Sydney Stock Exchange and using to evaluate the models the *chi-squared* method. Accordingly with this method, 12 stocks out of 19 where better represented by the Variance Gamma while five of the remaining seven cases, fitted the Press compound model and the other the normal one.

Part I

Estimation of the Variance Gamma

Chapter 4

Estimation Methods

4.1 Introduction

The aim of this part of the thesis is to present the so many different methods to estimate the parameters of a distribution, that is supposed to be known, of a random variable from a set of empirical observation of the phenomenon. The main, and most known, family of these are the point estimation. They allow to obtain a random variable, which is function of the set of data and that represent the parameter the method is looking for.

4.2 Definitions and Notations

Let X be a random variable, with a distribution function f that is known except for a set of parameters that are grouped in a vector θ for simplicity and to be as much general as possible. Assuming to have m different unknown parameters it is possible to define the subspace Θ of $\mathbb{R}^m$ that is the results of the Cartesian product of the domains of each parameter that compose the vector θ. Given this framework, the density function of X can be written as $f(x, \theta)$. As said above, nothing can be done without a set of data that are the random variable outcomes. Moreover, this set of data must respect some conditions to let the method work in the proper way, in particular, :

Definition Given a set of random variable $\{X_1, X_2, ..., X_n\}$ it is a random data sample if:

- all the variable are independent and identically distributed
- the distribution of these variables is $f(x, \theta)$

One of the most important insights given by this definition, is the need to have independent observations. To have a good approximation of the missing parameters of the distribution each sample collected must be in fact independent, in order to avoid what is called a *bias*, that produces unreliable results. If $X \sim f(x, \theta)$ the distribution of all the random variables is nothing more than:

$$f_{join}(x, \theta) = \prod_{i=1}^{n} f(X_i, \theta) \tag{4.1}$$

An other useful definition is the *statistic of the set*:

Definition A statistic is a function of the data set $g(X_1, X_2, ..., X_n)$. Its distribution is called *sampling distribution*

It is important to emphasize that a statistic is never dependent by the unknown parameters θ, but only its sampling distribution, since the random variable is a function only of the data set, as explained in the definition.
In order to estimate the vector θ, the definitions of characteristic of a sampling and estimator are very useful:

Definition A *characteristic of a sampling* is a function $\kappa(\theta) : \Theta \longrightarrow \mathbb{R}$, where the set Θ is the domain of the vector *theta*. If the function κ is constant over Θ the characteristic is said to be *trivial*

Definition Given a data set $X_1, X_2, ..., X_n$ and a characteristic of the sampling $\kappa(\theta)$, an estimator of the characteristic, based on the data set, is a statistic $T = g(X_1, X_2, ..., X_n)$. The value of the random variable T is said to be an *estimation* of $\kappa(\theta)$

In order to better understand how all these concepts and definitions operate, it is useful to show a simple example that helps also to introduce some of the most common estimators of the characteristics used in statistic. Lets consider a random sample taken from a Gaussian distribution, with unknown *mean* μ and variance σ^2. In this case the vector θ is two dimensional, and it is composed by the two unknown parameters, $(\mu; \sigma^2)$. Of course there are an infinite number of sampling characteristic of the sampling $\kappa(\mu, \sigma^2)$, for example:

- The mean: $\kappa(\mu; \sigma^2) = \mu$
- The variance: $\kappa(\mu; \sigma^2) = \sigma^2$
- The second moment: $\kappa(\mu; \sigma^2) = E(X^2) = \sigma^2 + \mu^2$

If it would be possible to collect a data set, $X_1, X_2, ..., X_n$, that respects all the restrictions explained above, the most known and used statistic for the mean and the variance are the sampling mean $\bar{X}$ and the sampling variance $\bar{S}$ respectively. More precisely, these statistics are calculated as follows:

- $$\bar{X} = \sum_{i=1}^{n} \frac{X_i}{n} \tag{4.2}$$
- $$\bar{S} = \frac{(X_i - \bar{X})^2}{n-1} \tag{4.3}$$

4.3 Mean square error

For a single characteristic $\kappa(\theta)$, there are many different statistics approach that can give an estimation of its value. Since each statistic is a random variable, the choice of the best estimator will be done considering its concentration around κ. Given a set of different statistics of the same characteristic, $T_1, T_2, ..., T_m$, the simplest method to choose is to select a small number ϵ and then the estimator with highest $P(|T_i - \kappa| < \epsilon)$. Another, and more complete way, to chose the best statistic is to obtain an average measure of its proximity to the real value of κ. In order to do this, the best choice is to select the statistic with the lower $E_\theta[(T_i - \kappa(\theta)^2]$. This value is called *Mean Square Error* (or MSE), in order to use it to select the best statistic, which should have finite mean and variance or, alternately, it has a finite second moment. Moreover, it can be proven that the MSE, can be written as follows, if respects the above conditions:

$$E[(T - \kappa(\theta))^2]) = VAR(T) + [E(T) - \kappa(\theta)]^2 \tag{4.4}$$

The quantity $E(T) - \kappa(\theta)$ is called *bias*.

Given this framework, it is possible to make a clear choice between to different estimators, T_1 and T_2.

T_1 will be preferred over T_2 if:

- $E[(T_1 - \kappa(\theta))^2] \leq E[(T_2 - \kappa(\theta))^2]\ \forall\theta \in \Theta$ and
- $E[(T_1 - \kappa(\theta))^2] < E[(T_2 - \kappa(\theta))^2]$ for some $\theta \in \Theta$

Theoretically, the best choice is the estimator T_0 such that:

$$E[(T_0 - \kappa(\theta)^2)] \leq E[(T_i - \kappa(\theta)^2] \quad \forall\theta \in \Theta \ , \ \forall T \tag{4.5}$$

However, this estimator does not exist, in order to prove that, it can be chosen a specific case and prove that, at least for one $\bar{\theta}$ the estimator T_0 is not the best choice.

Lets consider $\kappa(\theta) = \theta$, in this case the estimator $\tilde{T} = K$ for $K \in \Theta$ produces a Mean Square Error equal to 0, when $\theta = K$, that is, of course, the best possible. Thus, in each points, whatever statistic T_0 will be chosen, there will be an other estimator $\tilde{T}$ that behaves, locally, better.

For this reason the choice of the best statistic will be decomposed in two periods: the first will reduce the set of possible estimator by choosing it in the subset of those that have the property of $E(T) - \kappa(\theta) = 0$. This subset is called the *unbiased* statistics and has the property to have the expected value equal to the researched characteristic κ. The second step is finding the estimator with the lowest variance possible for all θ. In this subset are located the sampling mean and the sampling variance for the mean and the variance respectively. The idea of choosing in the subset of the unbiased estimators leads on the basis of not making systematically an error to over- or under-estimate the approximation. In this way, the only uncertainty is given by the variance of the estimator that of course cannot be eliminated. Also in this subset, however, the choice is not always easy, it can happen, in fact, or that the subset is empty (there are not any unbiased statistics) or that it contains many or that that the only one present is meaningless for some reasons.

4.4 UMVUE estimators

Let assume to be in the lucky case in which the subset of unbiased estimators is not empty and its elements have statistical meaning: thus, the *optimal* statistic will be the $\tilde{T}$ such that:

- $\tilde{T}$ is no bias
- $VAR_\theta(\tilde{T}) < VAR_\theta(T_i)$ for each T_i unbiased

If an estimator respects the above two condition, it is said to be an *Uniform Minimum Variance Unbiased Estimator* (UMVUE).

If such estimator exists, it has the following properties:

- It is essentially unique: it means that if exist two UMVUEs T_1 and T_2 it is also true that $P(T_1 = T_2) = 1$.
- Changing the order in which the data sampling $X_1, X_2, ..., X_n$ is collected, the value of the estimator does not change

Moreover, the "good" estimators have some asymptotic properties, that essentially means it would be possible to change the number of the higher observation collected, will be this number, the more precisely the statistics will reflect the characteristic $\kappa(\theta)$. In order to give some mathematical formality to the above statement, the following two definitions can give a quantitative method to select the best estimator:

Definition Given the sequence of i.i.d. random variables $X_1, X_2, ..., X_n$, and T_n, if an estimator of the characteristic $\kappa(\theta)$, has the property:

$$\lim_{n\to\infty} E(T_n) = \kappa(\theta) \tag{4.6}$$

Then estimator is said to be *asymptotically unbiased* .

Definition Given the succession of i.i.d. random variables $X_1, X_2, ..., X_n$, and T_n an estimator of the characteristic $\kappa(\theta)$, if the succession $\{T_n\}_n$ has the property:

$$\lim_{n\to\infty} E[(T_n - \kappa(\theta))^2 = 0 \tag{4.7}$$

The estimator is said to be *mean square consistent.*

It is easy to prove that the property of the mean square consistence embodies the asymptotically unbiased one and it provides, at the same time, the properties of asymptotically null variance.

4.5 Likelihood Method

One of the most known and used concept in the estimation precess is the *Likelihood function.* Lets assume to have a sampling of independent and identically distributed observation of a random variable that follows a known low except for the set of parameters generically identified by the vector θ. In this framework, as explained in the previous paragraphs, the joint distribution is the product of the distribution of each sample and in order to keep it as much general as possible, it will be identified by $f(x, \theta)$.

Definition The Likelihood function of n random variables $X_1, X_2, ..., X_n$ is the joint distribution of them, considered with respect to the parameter θ. Thus, the likelihood function can be written as:

$$\theta \to L_\theta(x_1, x_2, ..., x_n) = \prod_{i=1}^{n} f(x_i, \theta) \tag{4.8}$$

This function can be used to approximate the value of θ. If the likelihood function is interpreted as the likelihood, in fact, to have the obtained sample, by means of the observations, the best choice will be the parameter that maximize such function.

Being a product of the single observation density function, it is usually more easy to treat the *log*-Likelihood function that, with some easy mathematical manipulations can be written as:

$$\log L_\theta = \sum_{i=1}^{n} \log(f(x_i, \theta) \tag{4.9}$$

Maximizing the log function it gives the same value of θ of maximizing the likelihood function.

4.6 Moments method

One of the most known and easy to use method to estimate an unknown parameter from a population, is the Moment method. It is diffused because it does not require a distribution in the closed form, a property that is mandatory with the Likelihood method.

Considering the same sample of i.i.d. random variables $X_1, X_2, ..., X_n$, that represents usually a sampling of observations, the only request to the distribution of each variables is to admit at least the number of moments equal to the dimension of the vector unknown θ. Given this framework, it can be proven that if the moment exists one unbiased estimator for the characteristic $\kappa(\theta) = \nu_k(\theta)$, where $\nu_k(\theta)$ is the j^{th} moment, could be:

$$M_k = \frac{1}{n} \sum_{i=1}^{n} X_i^k \tag{4.10}$$

By building as much estimators of the moments as how many unknown parameters the distribution has a system of independent equation with just one unique solution can be written.

The drawbacks of this method are that the moments must exists and have a closed form. Moreover, it can happen that the method fails if the moment has a particular form. This is the case, for example of the uniform distribution with width 2θ and centred in zero ($U(-\theta;\theta)$). Of course the first moment is zero while generally the moment M_{2r} has the moment in the form:

$$E_\theta(X^{2r}) = \frac{\theta^{2r}}{2r+1} = M_{2r}$$

Since the only unknown variable is the unidimensional value θ, each of this moments can be used for estimate its value leading to complete different results depending by the choice made. Generalizing this concept, whenever a function has more moments depending by θ then the unknown parameters, the choice of which moments to use in the system is completely arbitrary and it will lead to complete different results. Rohatgi and Saleh (1999) sated that the best choice possible is to select the set of moments with the lowest degree possible.

Chapter 5

Variance gamma process estimation

5.1 Introduction

The aim of the last part of the thesis is to apply the Lévy process theory combined with the estimation methods to a real world case. In particular, the Lévy process chosen is the Variance Gamma, due to its statistical properties that are the closest to the actual behaviour of the prices in the real world. In order to estimate the unknown parameters the Moments method has been chosen, as it is the most reasonable choice due to the lack of a closed form for the Variance Gamma density function.

The system that rises from the application of the moments method to the variance gamma density is composed by four non linear equations that cannot be easily solved without the support of a software. Thus, in order to undertake the task, the software Matlab developed by MathWorks Inc. has been chosen. The data sample are the daily log-returns on MSCI US Investment Market Indices from January 2, 2009 to May 31st, 2013. Specifically, consider the first two indices were considered: consumer discretionary (CD), consumer staples (CS), with a total of 1109 observations.

5.2 The data sample

In order to contextualize the estimation in the real world is useful to understand the data sample, and the indexes that have been taken under consideration.

The MSCI US Equity Indexes aims at reflect the investment opportunities across the market capitaliza-

tion, the value and growth investment styles and the industry and sector groups within the US equity market from the perspective of US domestic investors. The MSCI US Equity Indexes includes the following indexes:

- MSCI U.S. Broad Market Index: Captures broad U.S. equity coverage representing about 99% of the U.S. equity universe.
- MSCI U.S. Investable Market 2500 Index: Designed to measure the performance of the large-, mid-, small- cap segment of the U.S. equity market. It is the aggregation of the MSCI U.S. Prime Market 750 (that is, the Large Cap 300 and Mid Cap 450 indexes combined) and Small Cap 1750 indexes.
- MSCI U.S. Large Cap 300 Index: Designed to measure the performance of the large cap segment of the U.S. equity market.
- MSCI U.S. Mid Cap 450 Index: Includes the next largest 450 companies in terms of Market capitalization of the U.S. equity market and designed to measure the performance of the mid cap segment.
- MSCI U.S. Prime Market 750 Index: Represents the aggregation of the Large Cap 300 and Mid Cap 450 Indexes
- MSCI U.S. Small Cap 1750 Index: Includes the remaining smallest 1,750 companies in the U.S.
- The MSCI U.S. Micro Cap Index: Designed to measure the performance of the micro cap segment of the U.S. equity market.

MSCI segregates the eligible US equity universe in three market capitalization segments:

- The investable market segment
- The micro cap segment
- The lower micro cap segment

The investable market segment contains all eligible securities with reasonable size and liquidity whose cost effectively be represented by institutional and pooled retail portfolios of reasonable size. This segment also allows investors to gain exposure to a significant portion of the performance of the US equity universe. The analysis shows that the 2,500 largest companies by full market capitalization, which cover more than 98% of the US equity universe, form an appropriate representation of the investable market segment. The investment performance characteristics of this investable market segment are represented and measured by an Investable Market Index. The micro cap segment will comprise micro cap companies with a market capitalization rank lower than the 2,500 companies in the investable market segment and included in the top 99.5% of the US equity universe ranked by full market capitalization. The micro cap segment is estimated to cover around 1.5% of the market capitalization of the US equity universe.

The investment performance characteristics of this segment of the US equity universe is represented and measured by a Micro Cap Index. The lower micro cap segment covers approximately the bottom 0.5% of the full market capitalization of the US equity universe, and will not be represented by an index. The combination of the Investable Market Index and the Micro Cap Index will form the US Broad Market Index, which thus includes the companies comprised in the top 99.5% of the US equity universe ranked by full market capitalization.

The investable market segment and index is comprised of three market capitalization segments and their corresponding indexes- large cap, mid cap, and small cap. MSCI defines the large cap index as consisting of the 300 largest companies by full market capitalization in the investable market segment, the Mid Cap Index as comprising the next 450 companies, and the Small Cap Index as consisting of the remaining 1,750 companies. The large cap and the mid cap indexes, as defined above, are also combined to create a separate index of the 750 largest companies in the investable market segment ranked by full market capitalization. MSCI uses a fixed number of companies for defining the cut-off levels for the market capitalization segments. Analysis shows that using a fixed number of companies to specify market capitalization cut-off levels leads to a better stability and a lower turnover in the resulting market capitalization indexes over time, when compared to using other factors to define market capitalization segments, such as percentiles of market capitalization or absolute market capitalization levels. In making a determination as to what levels of a fixed number of companies appropriately define the various market capitalization segments within the investable market segment, MSCI considered the behavior of several factors over

time using different levels of fixed number of companies. The factors include the following:

- The absolute market capitalization level of the smallest company.
- The marginal contribution to the relevant index of the smallest company.
- The cumulative proportion of market capitalization covered.
- The liquidity and trading characteristics of companies.
- An analysis of the average size of portfolio holdings of a variety of large, mid cap, and small cap investment managers.

Given this framework the data are collected focusing on two of the many indexes proposed by MSCI, The MSCI UK Consumer Staples and the Consumer Discretionary Index.

The MSCI UK Consumer Staples Index is designed to capture the large and mid cap segments of the UK equity universe. All securities in the index are classified in the Consumer Staples sector as per the Global Industry Classification Standard. On the other hand, The MSCI EMU Consumer Discretionary Index is designed to capture the large and mid cap segments across 10 Developed Markets (DM) countries in the EMU*. All securities in the index are classified in the Consumer Discretionary sector as per the Global Industry Classification Standard.

Both the indexes are based on the MSCI Global Investable Market Indexes (GIMI) Methodology, a comprehensive and consistent approach to index construction that allows for meaningful global views and cross regional comparisons across all market capitalization size, sector and style segments and combinations. This methodology aims to provide exhaustive coverage of the relevant investment opportunity set with a strong emphasis on index liquidity, investability and replicability. The indexes are both reviewed quarterly, in February, May, August and November, with the objective of reflecting change in the underlying equity markets in a timely manner, while limiting undue index turnover. During the May and November semi-annual index reviews, the indexes are both rebalanced and the large and mid capitalization cut-off points are recalculated.

5.3 The MatLab code

In order to apply the Moments Method to the data sample, there is the need to implement it in MatLab code.

The Variance Gamma distribution, according with the previous chapters, has the following statistical properties with respect the three parameters (μ, k, σ):

Statistics	Parametric value
Mean	μ
Variance	$\sigma^2 + k\mu^2$
Skewness	$\frac{\mu k(3\sigma^2+2k\mu^2)}{(\sigma^2+k\mu^2)^{\frac{3}{2}}}$
Kurtosis	$3[1 + 2k - k\sigma^4(\sigma^2 + k\mu^2)^{-2}]$

Moreover, in order to completely consider the variance gamma distribution in the more general case, there is the chance that a drift, namely c, is present in the log-prices path leading to such distribution:

$$LogPrice \sim c + VG(\mu, k, \sigma) \tag{5.1}$$

Given this set of informations and considering the four statistics sample mean, sample variance, sample Skewness and sample Kurtosis, is possible to apply the Moments method. In MatLab, this task has been done creating two *m-files*, a function and a script. The function code is used to take as inputs two vectors that are the log return and the vector with the parameters estimation. More precisely the code is the following:

```
function [error,momentsSample,moments] = getErrorMME(x,logreturns)

    c=x(1);
    theta=x(2);
    sigma=exp(x(3));
    nu=exp(x(4));
    VARIANCE=sigma.^2+nu.*theta.^2;
    moments=[c+theta;VARIANCE;...
        theta*nu*(3*sigma^2+2*nu*theta^2)*VARIANCE^(-3/2);
        3*(1+2*nu-nu*sigma^4*VARIANCE^(-2))];
    momentsSample=[mean(logreturns);var(logreturns);...
        skewness(logreturns);kurtosis(logreturns)];
    error=moments-momentsSample;

end
```

The notation with which the parameters has been defined is quite changed according to the following table:

Notation used in the thesis	Matlab code
σ	σ
μ	θ
k	ν

In order to make more clear the code a brief explanation is needed. After defining the four components of the vector x as the unknown parameters (instead looking directly for σ and ν it has been preferred to look for their exponential due to the positive constraints of this two parameters), following the formulas showed above for the statistics, the function calculate all of them. Subsequently, it calculates the moments using the embedded functions for the sampling Kurtosis and Skewness. Both of them take, as input, the logreturns variable that is an input of the function itself. This variable must be a vector with the log-returns of the sample data collected. After the calculation of the four moments useful in the method, the function calculate the error as the difference between the method defined by the model and the moments sample. The output of this MatLab code are three elements:

- The error between the statistics and the sample moments
- The sample moments set as an array
- The moments as an array

By itself this function is just an half of the moment method, the aim now is to use the power of the MatLab engine to solve the non linear system. To do this, a script has been written. It can be divided in three parts: In the first the log-returns are calculated from the data sample that is just a spot observation of the value of two indexes over the time. In the second part the initial condition are setted, they are the starting point from which the software will start its research of the solution for the non linear system. In the last part the code tries to physically solve the problem. The script is the following:

In the setting of the initial condition it has been assumed that θ is closed to zero, if this statement

```
clear; close all; clc;

load('SampleData')
logreturns=log(prices(2:end,:)./prices(1:end-1,:));

sigma=sqrt(var(logreturns));
nu=kurtosis(logreturns)/3-1;
theta=sigma.*skewness(logreturns)./(3*nu);
c=mean(logreturns)-theta;
x0=[c;theta;log(sigma);log(nu)];

[x,fval,exitflag,output] ...
        =fsolve(@(x) getErrorMME(x,logreturns),x0);

xOpt=[x(1);x(2);exp(x(3));exp(x(4))]
```

holds, the statistics have simpler formulas, specifically:

Statistics	Parametric value
Mean	0
Variance	σ^2
Skewness	0
Kurtosis	$3(1+\nu)$

From this, using the sample variance and sample Kurtosis is easy to obtain the parameters σ and ν. In order to obtain the value of θ, it can be taken in consideration the Skewness and neglecting the power of θ with degree higher than 1. From that is easy to obtain $Skewness = \frac{3\nu\theta}{\sigma}$. The parameter c is calculated considering the formula: $E(logreturns) = c + \theta$.

After defining the initial condition the problem of the non-linear system is solved used the embedded *fsolve function*. The particular instance of this function returns a value exitflag that describes the exit condition of fsolve, and a structure output with information about the optimization process.

Running the code explained the solution obtained is the following:

Statistics	Parametric value
c	0.0017
μ	-0.0008
σ	0.0146
k	0.9940

5.4 Conclusion

We presented the main models used to describe the financial asset returns. We introduced the Black-Scholes-Merton model, underling its points of strength and drawbacks, then we focus on the class of Lévy processes. The description of this family has been using a simple mathematical language (but still rigorous). We then focus on the Variance Gamma process due to its wide use in financial models.

In the second part we recall the principal estimation methods, the Moments and the Likelihood methods. Then, the moments method has been chosen in order to estimate the Variance Gamma parameters from a sample collected in the American stock market. To undertake this task, it has been used the software Matlab.

Acknowledgement: I would like to thank prof. Marina Marena form the University of Torino in the department ESOMAS for providing the Matlab code used in the estimation.

Bibliography

- Chapman and Hall/CRC Financial MathematicsSeries. Financial Modelling With Jump Processes by Rama Cont and Peter Tankov

- Schoutens, W. Lévy Processes in Finance: Pricing Financial Derivatives. 2003.

- Eugene Seneta: Fitting the Variance Gamma Model to Financial Data. Jurnal of applied probability, vol 41, Stochastic Methods and Their Applications (2004).

- $https://en.wikipedia.org/wiki/Lvy_process$

- $https://en.wikipedia.org/wiki/Variance_gamma_process$

- $http://www1.mate.polimi.it/\ ileepi/dispense/0809STAT/stimapuntuale.pdf$

- $http://staff.matapp.unimib.it/\ fcaraven/download/other/dispense-3.3.pdf$

- $https://www.mccormick.northwestern.edu/documents/students/undergraduate/introduction-to-matlab.pdf$

- $https://www.msci.com/us-equity-indexes$

- $https://towardsdatascience.com/the-poisson-distribution-and-poisson-process-explained-4e2cb17d459$

- $http://www.stats.ox.ac.uk/\ winkel/lp1.pdf$

- $https://en.wikipedia.org/wiki/Gamma_distribution$

- $http://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/306439500f660d9fc12574860059ce30/FILE$

Ringraziamenti

Si ringrazia la professoressa Semeraro per il supporto e la comprensione oltre che la competenza che ha mostrato durante il percorso di tesi. Inoltre si ringrazia anche la professoressa Marina Marena dell'Università di Torino, dipartimento ESOMAS, per aver fornito il codice MatLab usato per la stima dei parametri del Variance Gamma Process.

Arrivati alla conclusione di un percorso universitario, nel momento della stesura dei ringraziamenti, non è mai semplice includere tutte quelle persone che hanno fatto sì che ciò si avverasse. Ovviamente il primo pensiero va alla famiglia, Mamma, Babbo e Francesca che ogni giorno ci sono stati per un consiglio, un aiuto o semplicemente una telefonata in questo periodo che ho passato lontano da casa a Torino. A Cristiana, persona senza la quale sarebbe stato tutto più difficile, la ringrazio per la pazienza nel sopportare tutte le difficoltà, facendosi sempre carico dei miei problemi come fossero i suoi e sopportandomi quando ero nervoso o irascibile. La distanza che ci ha separato in questi anni è stata sicuramente difficile da gestire ed anche adesso, all'inizio di un nuovo percorso lavorativo pieno di insidie e difficoltà, la ringrazio per affrontare ogni problema con spensieratezza ed il sorriso. Infine, il pensiero va agli amici, coloro senza i quali tutto sarebbe stato più triste. Ringrazio gli amici di sempre, il Pica per esserci sempre per una telefonata o una discussione accesa sull'ultima partita della Juve, gli amici dell'università di Pisa, Francesco, Daniel e Francesca, che anche se non ci vediamo più assiduamente sono sempre pronti a vederci ridendo e scherzando come se fossimo sempre compagni di Università. Ultimi, ma non per importanza, i ragazzi conosciuti a Torino; ringrazio Federico, Francesco, Irene, Giacomo, Valerio, Fabrizio, Antonio e tutti gli altri che hanno reso questa avventura speciale.