UNIVERSITÀ DEGLI STUDI DI BERGAMO
DIPARTIMENTO DI MATEMATICA, STATISTICA, INFORMATICA
E APPLICAZIONI

Dottorato di ricerca in Metodi computazionali per le Previsioni
e Decisioni Economiche e Finanziarie
XXIV Ciclo

PhD thesis

FIRST PASSAGE TIMES WITH MARKOV PROCESSES IN PORTFOLIO SELECTION PROBLEMS

Alda Ndoci

Supervisor:
Prof. S. Ortobelli

Program Coordinator:
Prof.ssa M. Bertocchi

A mia figlia, Gania

Acknowledgements

A special thank you goes to my supervisor Prof. Sergio Ortobelli. His encouragement and support was unbelievable during this period. He was always available to help me and his advice was immense.
I also thank Prof.ssa Bertocchi and Prof.ssa Giacometti for their support during these years.
I would like to thank Prof.ssa Guegan for her suggestions throughout the period I was in Paris. I really appreciate everything you all did.
A very special thank you goes to my dear parents who gave me the opportunity to study outside Albania. Unfortunately, they live miles away from me however, they never failed to support and put a smile on my face when I stressed.
In addition, I can't find words to thanks my lovely brother Alfred, who had to 'deal' with me during this period. I must say he was very patient.
Another special, thank you goes to my friends Jeta, Elona, Marcelo, Alma, Altin, Filomena, and many more who in different ways contributed toward my thesis.
Last but not least a special thanks goes to my amazing husband, my best friend, Arjan, who always stood by my side all these years. His support and kind words gave me so much strength to continue working very hard on this thesis. He always believed in me and I hope today I make him proud.

Finally, I dedicate this thesis to my daughter Gania! Her existence has been the biggest inspiration for me.

Contents

Introduction **1**

1 An introduction to Markov processes **7**
1.1 Definition of Markov chains and some basic properties . . . 8
1.1.1 Class structure 12
1.1.2 Convergence to equilibrium 19
1.2 The Markov Approximation of Exponential Lévy processes . 21
1.2.1 Definition of Lévy process 21
1.2.2 Lévy processes as Markov processes 30
1.2.3 Exponential Lévy processes and their estimation . . . 34
1.3 Markov regime switching models 47
1.3.1 Description of the Markov Switching framework . . 48
1.3.2 Evaluation of Likelihood Function and EM algorithm 51

2 First passage times distributions with Lévy processes and portfolio application **66**
2.1 Modeling Markov Processes 67
2.1.1 Nonparametric Markov processes 68
2.1.2 Markov Transition matrix associated to a given Lévy process 70
2.2 On the approximation of the wealth distribution 72
2.3 On the approximation of the first passage times 76
2.4 Practical timing portfolio strategies on the US stock market . 80
2.4.1 The Dataset: ex-ante empirical evidence 82
2.4.2 An ex- post comparison among timing portfolio strategies 83

3 Timing portfolio selection with non-Lévy processes 92
3.1 Financial application with alternative models 93
3.1.1 The Markov approximation of Regime Switching model 93
3.1.2 The log-Student-t Model 98
3.2 An empirical comparison among timing portfolio strategies applied to preselected assets 101
3.2.1 Pre-selection criteria 101
3.2.2 An ex- post comparison among timing portfolio strategies applied to preselected assets 107

List of Tables

1.1 Variance Gamma (σ, ν, θ) 39
1.2 Variance Gamma (C,G,M) 40
1.3 The Normal Inverse Gaussian (α, β, δ) 41
1.4 The Meixner (α, β, δ) 42
1.5 Statistical Moments 45
1.6 Parameter Estimation 46

2.1 Statistics on the ex-ante log returns 84
2.2 Descriptive Statistics on the ex-post logreturns 89
2.3 Ex-post wealth of timing portfolio strategies 90

3.1 Descriptive Statistics on the ex-post log-returns (annual basis) 112

List of Figures

2.1 Final probabilities of the nodes at the 2nd step 76
2.2 Ex post final wealth obtained with different portfolio strategies 87
2.3 Turnover and Diversification 88

3.1 Ex post final wealth obtained with non-Lévy processes . . . 111

Contents

Introduction

During the last decade Markov processes have become very popular for modeling the market fluctuations since they describe the time evolution of random phenomena (i.e. the daily price of a risky asset). Originally proposed by the Russian mathematician Markov in 1907, the class of Markov chains have been growing more useful as a method for capturing the stochastic nature of many economic and financial variables. This makes the Markov chains an important example of random processes, that can be also used to approximate other continuous Markov processes.
Applications of Markov chains arise in many different areas. One of the areas where they have been widely used is the option pricing theory. (see among other, Cox et al (1979) [12], D'Amico (2003) [13], D'Amico et al (2009) [14], Iaquinta and Ortobelli (2006) [35].
Among Markovian models we essentially distinguish two categories: parametric models (see among other Duan and Simonato (2001) [20], Duan et al (2003) [18], Staino and Ortobelli (2011)[67]) and non parametric models (see Iaquinta and Ortobelli (2006)[35]). In the first category, the Markovian hypothesis is used for the diffusive models of the underlying returns. The transition matrix depends on the parameters of these Markov processes. In the second category of models, only the historical series are used to estimate the option prices. So, in non parametric Markov models the transition matrix depends on the historical observations. We recall the non-parametric model presented in Iaquinta and Ortobelli (2006) [35], where

the time evolution of the returns is approximated by the Markov chain. With this model, they are able to price American/European and path dependent options in a reasonable computational time. Alternatively, Staino and Ortobelli (2011)[67] proposed to price path dependent options when log-retuns follow a Lévy process. They approximate the Markovian behavior of any Lévy process and using the Markov chain properties they simplify the computation of price path dependent contingent claims. This method has been firstly applied by Duan and Simonato (2001)[20] and Duan et al (2003)[18] to approximate Wiener processes and GARCH processes with Gaussian residuals in order to price American and barrier options.

In addition, several empirical works (see Lamantia et al (2006) [41]) have shown that we cannot reject the Markovian hypothesis of asset returns. Following the methodology proposed by Christoffersen (1998) [10], it is possible to test the null hypothesis that the intervals of the distributional support of a given portfolio are independent against the hypothesis that the intervals follow a Markov chain.

On the one hand a Markov chain should be a good model to describe the evolution of the distributional support of a given portfolio. On the other hand, since its application to predict future wealth presents a high computational complexity, the Markovianity has not always been opportunely used in portfolio theory (see Leccadito et al (2007) [42]).

Angelelli and Ortobelli (2009) [1] proposed some algorithms that reduce the complexity of the portfolio selection problems based on this hypothesis. In particular, they distinguish two different methods to build the transition matrices. With parametric portfolio selection models the transition matrix depends on the parameters of the underlying Markov process and the parameters are functions of the portfolio weights. Instead, with non parametric Markov models the transition matrix depends directly on the portfolio weights. Hence, the transition probabilities are strictly linked to historical observations.

Empirical distributions of asset returns are typically 'fat tailed' with large asset price movements having higher probability than predicted by normality. Alternative models, able to capture tail decays and time varying volatility, have been developed.
Recently, attention has been placed on Markov regime switching models. In such models, the stochastic process remains in one regime for a random amount of time before switching over into a new regime.
Markov regime switching model assumes there are two 'regimes' or states with different mean and volatility levels. This specification can be interpreted as a mixture distribution with dynamics generated by a Markov chain.
In this model the Gaussian distribution is generalized by introducing two regimes with different moments. Leptocurtosis is obtained here because the variance in the two regimes differs.
So, the mixture distribution generates the leptocurtosis and the Markov chain is responsible for the nonlinear dynamics.
The Markov regime switching was first introduced by Hamilton(1989) [30] to modelling changes in time series and business cycle. Engel and Hamilton (1990)[22] are the first to apply Markov regime switching model to financial time series. They study a sample of quarterly returns, finding that Markov regime switching model gives a good fit. They also apply a variety of tests to analyse the performance of their model.
Applications of Markov regime switching model to stock market returns are found in Pagan and Schwert(1990) [56]. There Markov regime switching is compared with GARCH and several models.
Cai (1994)[8], Gray (1996) [27], Klaasen (2002) [39] proposed various specifications for Markov switching ARCH-type models.
Cai(1994)[8] was the first to apply the idea of endogenous regime-switching parameters by Hamilton(1988, 1989) [29] [30] into an ARCH specification to account for possible presence of structural breaks. He uses the ARCH specification instead of GARCH to overcome the problem of infinite path-dependence, i.e to avoid the conditional variance at time t depending on the entire sample path.

Gray(1996)[27] suggests to integrate out the unobserved regime path in the GARCH equation using the conditional expectation of the past variance and his model can be regarded as the first MRS-GARCH.
Klaassen(2002) [39] suggests adopting the conditional expectation of the lagged conditional variance with a broader information than in Gray, so that his model has two main advantages. It allows both a higher flexibility in capturing the persistence volatility shocks, and gives straightforward expressions of multi-step-ahead volatility forecasts that can be calculated recursively as in standard GARCH models.
Duan, Popova and Ritchken (2002)[19] model fills the gap between Black-Scholes(BS) model [5], in this framework can be viewed as a single volatility model with no feedback effects, and the extended GARCH models, which have infinitely many volatility regimes with feedback effects. This include the family of GARCH option pricing models discussed by Duan (1995 [17] and the Markov regime switching models of Hamilton as special cases. His models allow volatility regimes to impact returns but do not allow returns to impact future volatilities. In Duan, Popova and Ritchken (2002) [19] are considered models in which variance updating schemes depend not only on levels of variance and on asset innovations, but also on a second factor that is uncorrelated with asset returns. As a result variance levels are not completely determined by the path of prices. This second factor allows for further flexibility in capturing properties of stock return process.
To price regime switching American and European options, Buffington and Elliot(2002) [7] use partial differential equation with non-smooth boundary conditions, whereas Yao et al (2003)[70] derived the PDE's with smooth boundary conditions. In a recent paper, Fu et al(2012)[25] provided a closed-form formula for price of an European call option, when the Markov process has finite number of states.

This thesis deals with the parametric Markovian approach to model asset returns. We propose and examine innovative timing portfolio selection strategies that use the Markovianity property of some different processes.

First, we discuss how to approximate the distribution of the first passage times of some well known Lévy processes. On the one hand, most of the parametric processes used in portfolio theory are Markov processes such as Lévy processes. On the other hand, forecasting the future wealth of a parametric Markov process could present a high computational complexity in portfolio problems. For this reason in this thesis we apply the method discussed by Duan and Simonato (2001) [20] to approximate the Markovian evolution of the portfolio wealth with a proper Markov chain. Then, using the same logic of Iaquinta and Ortobelli (2006) [35]'s algorithm we are able to propose a new algorithm for computing the distribution of first passage times of the portfolio wealth. This way to estimate first passage time distributions is different and alternative to the approach presented by Mijatović and Pistorius (2013)'s [50] algorithm. Mijatović and Pistorius's approach, as well as, the discreted time Duan's one have been used for pricing barrier options and are not computationally efficient for portfolio selection problems, as suggested by Angelelli and Ortobelli (2009) [1].

While timing portfolio strategies have been treated in financial literature (see Kardaras and Platen (2010) [37]), we are the first's one (by our knowledge) to suggest practical methods to deal with them. In particular, we propose new portfolio selection strategies that optimize proper first passage times of the wealth at some benchmark levels. As a matter of fact, investors want to maximize the first time their wealth decrease and want to minimize the first time their wealth increase. With the methodology proposed we are able to optimize timing portfolio strategies in a reasonable computational time for most of the well known Lévy processes. The new computationally efficient algorithm presented in Chapter 2 is the first contribution of this thesis.

Second, we empirically compare different portfolio strategies based on the optimization of the expected value of proper first passage times of the portfolio wealth.

In this context, we want to evaluate the impact of different distributional assumptions in timing portfolio strategies. In particular, we consider the

following Lévy processes: Normal Inverse Gaussian (NIG) [3] , Variance Gamma (VG) [44], Meixner [63], Brownian motion and α-Stable Lévy process [45] and [34].
A further important contribution of this thesis is the empirical comparison among different timing portfolio selection strategies when the returns are modeled by a non-Lévy process (Markov Regime Switching model and log-Student t model). We assume a Markovian evolution of the portfolio processes and we estimate some proper stopping times. We propose an empirical experiment to show that the techniques previously presented can be applied in practical portfolio problems not only for the Lévy processes but for other processes as well. Finally, we compare the ex post wealth obtained by optimizing (under different distributional assumptions) a performance ratio between the average first passage time of the wealth at a "losing" level and the average first passage time of the wealth at a "winning" level. Doing so, we show how the theoretical discussion can be practically applied to real portfolio problems.

The thesis is organized as follows. Chapter 1 deals with the theory of Markov process. In particular we examine Markov chains, continuous time Markov processes (with more attention to Lévy processes) and Markov regime switching processes.
In chapter 2 we show how to model parametric Markov chains. Then, formalizes a methodology to compute the distribution of bounded stopping times. We compare the ex-post wealth obtained optimizing some parametric timing portfolio selection strategies.
A further financial application considering non-Lévy processes is given in chapter 3. The Markov approximation of the Regime Switching model and the log-Student-t model is presented. Then, we recall a methodology to preselect some assets that optimize some common performance criteria. Finally, we assess the ex-post wealth we obtain maximizing a performance measure based on the average of proper first passage times.

Chapter 1

An introduction to Markov processes

In this chapter we deal with Markov processes and we describe the principal use of the Markov property. In particular we examine Markov chains, continuous time Markov processes (with more attention to Lévy processes) and Markov regime switching processes. Markov chains are the simplest mathematical models for random phenomena evolving in time.

A *Markov process* is a particular type of stochastic process where the future value of a variable depends only on its present value. The characteristic property of this sort of process is that it retains *no memory* of where it has been in the past (*Markov property*). This means that only the current state of the process can influence where it goes next.

The past history of the variable and the way that the present has emerged from the past are irrelevant. The lack of memory property makes possible to predict how a Markov process may behave, and to compute probabilities and expected values which quantify that behavior.

This chapter deals with the theory of Markov process. It is organized as follows. In the section 1.1 discrete-time Markov chains are defined and their behavior is investigated. The calculation of transition probabilities, expected hitting times and invariant distribution are given. Also treated

are recurrence, transience and convergence to the equilibrium.
section 1.2 presents Lévy processes, defining them and stating their characterization by infinitely divisible distributions. Moreover, section 1.2 shows how any Lévy process satisfies Markov property, simplifying its tractability. We focus on the construction of Lévy processes and, in particular, we explain how to define a Lévy process by a subordinator. We describe some exponential Lévy processes and we propose a computationally efficient method to approximate their parameters.
In the section 1.3 is given a description of Markov regime switching framework. The likelihood function is evaluated of and the EM algorithm is described.

1.1 Definition of Markov chains and some basic properties

In this section discrete-time Markov chains are defined and their behavior is investigated. The calculation of transition probabilities, expected hitting times and invariant distribution are given. Also treated are recurrence, transience and convergence to the equilibrium. We follow Norris (1998) [52] and Marcus and Rosen (2006) [47] .

Let us define the Markov chain and its properties.

Definition and basic properties

Let I be a countable set. Each $i \in I$ is called a *state* and I is called the *state-space*. We say that $\lambda = (\lambda_i : i \in I)$ is a distribution if $0 \leq \lambda_i < 1$ for all $i \in I$ and in addition the $\sum_{i \in I} \lambda_i$ equals 1.
Given a probability space $(\Omega, \Im, \mathbb{P})$, recall that a *random variable* X with values in I is a function X: $\Omega \to I$. Suppose we set

$$\lambda_i = \mathbb{P}(X = i) = \mathbb{P}(\{\omega : X(\omega) = i\})$$

Then, λ defines the ditribution of X. We think of X as modelling a random state which takes the value *i* with probability λ_i.
Let define the transition matrix $P = (p_{ij} : i, j \in I)$ where $p_{ij} : i, j \in I$ is the probability to go from *i* to *j*. Next we give the basic definitions of homogeneous Markov chains and its construction.

Definition 1.1. Given a probability space $(\Omega, \Im, \mathbb{P})$, a space of states I and a sequence of random variables $X_n : \Omega \to I$. We say that $(X_n)_{n\geq 0}$ is an homogeneous Markov chain with discrete times if $\forall n \geq 0$ and $\forall i_0, i_1, \ldots, i_{n-1}, i, j \in I$ such that $P(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}, X_n = i) > 0$ then the two properties are verified:

- $P(X_{n+1} = j/X_n = i, X_{n-1} = i_{n-1}, \ldots, X_1 = i_1, X_0 = i_0) = P(X_{n+1} = j/X_n = i)$. This property, known as *Markov property* implies that Markov chains have no memory. In other words, the probability of the system being in any given state at the next time step depends only on the state of the system at the current time step and is independent of the previous states of the system. .
- $P(X_{n+1} = j/X_n = i) = P(X_{k+1} = j/X_k = i) = p(i, j) \quad \forall n, k \in N_0$. This is the case the matrix P does not depend on time and the Markov chain is called homogeneous Markov chains.

Proposition 1.1.1. *Given the space of states I, the transition matrix P and the starting distribution λ associated to X_0(i.e $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$), then it is possible to build a probability space $(\Omega, \Im, \mathbb{P})$ and a Markov chain $(X_n)_{n\geq 0}$ uniquely identified, at less of finite dimensional distribution, by (I, P, λ).*

Since a homogeneous Markov chain is uniquely determined by the initial distribution λ and the transition matrix P, we point out that the Markov chain $(X_n)_{n\geq 0}$ on Markov (λ, P).

In the following we deal only homogeneous Markov chain and with abuse of notation we do not specify that is homogeneous.

Theorem 1.1.2. *A discrete-time random process $(X_n)_{0\leq n\leq N}$ is Markov(λ, P) if and only if for all $i_0, \ldots, i_N \in I$*

$$\mathbb{P}(X_0 = i_0, X_1 = i_1, \ldots, X_N = i_N) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{N-1} i_N}.$$

Proof. See Norris (1998) [52], Theorem 1.1.1 □

An important property of the Markov chains is that they have no memory. This is given by the next result. We write $\delta_i = (\delta_{ij} : j \in I)$ where

$$\delta_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{otherwise} \end{cases} \tag{1.1}$$

Theorem 1.1.3. ***(Markov Property).*** *Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then, conditional on $X_m = i, (X_{m+n})_{n\geq 0}$ is Markov (δ_i, P) and is independent of the random variables $X_0, \ldots, X_m$.*

Proof. We have to show that for any event A determined by $X_0, \ldots, X_m$, we have

$$\begin{aligned} &\mathbb{P}(\{X_m = i_m, \ldots, X_{m+n} = i_{m+n}\} \cap A | X_m = i) \\ &= \delta_{ii_m} p_{i_m i_{m+1}}, \ldots, p_{i_{m+n-1} i_{m+n}} \mathbb{P}(A | X_m = i) \end{aligned}$$

then the result follows by Theorem 1.1.2. First consider the case of elementary events

$$A = \{X_0 = i_0, \ldots, X_m = i_m\}.$$

In that case we have to show that

$$\begin{aligned} &\mathbb{P}(X_0 = i_0, \ldots, X_{m+n} = i_{m+n} \quad \text{and} \quad i = i_m)/\mathbb{P}(X_m = i) \\ &= \delta_{ii} p_{i_m i_{m+1}} \ldots, p_{i_{m+n-1} i_{m+n}} \\ &\times \mathbb{P}(X_0 = i_0, \ldots, X_m = i_m \quad \text{and} \quad i = i_m)/\mathbb{P}(X_m = i) \end{aligned}$$

which is true by 1.1.2. □

During this section we will try to response to the following question: what is the probability that after n steps our Markov chain is in a given state? First, given the transition probability p_{ij} and the initial distribution λ_0 we can calculate the probability $p_i^{(n)}$.

Proposition 1.1.4. *Given the homogeneous Markov chain(λ,P) with transition probability p_{ij} and the initial distribution λ_0 then the following relation holds:*

$$p_i^{(n)} = \mathbb{P}(X_n = i) = \sum_{i_0, i_1, \ldots, i_{N-1}} \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{N-1} i}.$$

We consider distributions and measures λ as row vectors whose components are indexed by I, just as P is a matrix whose entries are indexed by $I \times I$. When I is finite we identify the states 1,2,...,N; then λ will be an N-vector and P an $N \times N$ matrix.
We extend matrix multiplication to the general case in the obvious way, defining a new measure λP and a new matrix P^2 by

$$(\lambda P)_j = \sum_{i \in I} \lambda_i p_{ij}, \qquad (P^2)_{ik} = \sum_{j \in I} p_{ij} p_{jk}.$$

We define P^n similarly for any n given by:

$$(P^n)_{ij} = \sum_{i_1, \ldots, i_{n-1}} p_{ii_1} \cdots p_{i_{n-1}j}$$

Then, the following relation holds:

Proposition 1.1.5. *Given the homogeneous Markov chain(λ,P) with transition probability p_{ij} and the initial distribution λ_0 then:*

$$p_j^{(n)} = \sum_i \lambda_{i_0} . (P^{(n)})_{ij}$$

We agree that P^0, is the identity matrix I, where $(I)_{ij} = \delta_{ij}$. In the context we specify when I refers to the state-space and when to the identity matrix. We write $p_{ij}^n = (P^n)_{ij}$ for the (i,j) entry in P^n.
In the case where $\lambda_i > 0$ we shall write $\mathbb{P}_i(A)$ for the conditional probability $\mathbb{P}(A|X_0 = i)$. By the Markov property at time $m = 0$, under $\mathbb{P}_i$, $(X_n)_{n \geq 0}$ is Markov(δ_i, P). So, the behavior of $(X_n)_{n \geq 0}$ under $\mathbb{P}_i$ does not depend on λ.

Theorem 1.1.6. *Let $(X_n)_{n \geq 0}$ be Markov(λ, P). Then, for all $n, m \geq 0$,*

- $\mathbb{P}(X_n = j) = (\lambda P^n)_j$;

- $\mathbb{P}(X_n = j) = \mathbb{P}(X_{n+m} = j | X_m = i) = p_{ij}^n$

Definition 1.2. We point out with $(p_{ij}^n : i, j \in I) = P_i(X_n = j) = P(x_n = j / X_0 = i)$, and with (I,P) the homogeneous Markov chain with space of states I and transition matrix P.

1.1.1 Class structure

It is sometimes possible to break a Markov chain into smaller pieces, each of which is relatively easy to understand, and which together give an understanding of the whole. This is done by identifying the communicating classes of the chain.
We say that i leads to j and write $i \to j$ if

$$\mathbb{P}_i(X_n = j \quad \text{for some} \quad n \geq 0) > 0.$$

We say i communicates with j and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

Theorem 1.1.7. *For distinct states i and j the following are equivalent:*

- *i)* $i \to j$*;*
- *ii)* $p_{i_0 i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} > 0$ *for some states* $i_0, i_1, \dots, i_n$ *with* $i_0 = i$ *and* $i_n = j$*;*
- *iii)* $p_{ij}^{(n)} > 0$ *for some* $n \geq 0$*.*

It is clear from ii) that $i \to j$ and $j \to k$ imply $i \to k$. So, $\leftrightarrow$ satisfies the conditions for an equivalence relation on I and thus partitions I into *communicating classes*. We say that a class C is *closed* if

$$i \in C, i \to j \quad \text{imply} \quad j \in C$$

Thus a closed class is one from which there is no escape. A state i is *absorbing* if $\{i\}$ is a closed class. The smaller pieces referred to above are these communicating classes. A chain or transition matrix P where I is a single class is called irreducible.

Hitting times and absorption probabilities

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. The *hitting time(or the first passage time)* of a subset A of I is the random variable $H^A : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ given by

$$H^A(\omega) = inf\{n \geq 0 : X_n(\omega) \in A\}$$

where the infimum of the empty set $\emptyset$ is ∞. The probability starting from i that $(X_n)_{n\geq 0}$ ever hits A is then

$$h_i^A = \mathbb{P}_i(H^A < \infty).$$

When A is a closed class, h_i^A is called the *absorption probability*. The mean time taken for $(X_n)_{n\geq 0}$ to reach A is given by

$$k_i^A = \mathbb{E}_i(H^A) = \sum_{n<\infty} n\mathbb{P}(H^A = n) + \infty\mathbb{P}(H^A = \infty).$$

In other words, we write

$$h_i^A = \mathbb{P}_i(\text{hit } A), \quad k_i^A = \mathbb{E}_i(\text{time to hit } A)$$

These quantities can be calculated explicitely by means of certain linear equations associated with the transition matrix P given by the following theorem.

Theorem 1.1.8. *The vector of hitting probabilities $h^A = (h_i^A : i \in I)$ is the minimal non negative solution to the system of linear equations*

$$\begin{cases} h_i^A = 1 & \text{for} \quad i \in A \\ h_i^A = \sum_{j\in I} p_{ij} h_j^A & \text{for} \quad i \notin A \end{cases} \tag{1.2}$$

(Minimality means that if $x = (x_i : i \in I)$ is another solution with $x_i \geq 0$ for all i, then $x_i \geq h_i$ for all i.)

Proof. See Norris (1997) [52], Theorem 1.3.2. □

Now, we give a general result on mean hitting times. Recall that $k_i^A = \mathbb{E}_i(H^A)$, where H^A is the first time $(X_n)_{n\geq 0}$ hits A. We use the notation $\mathbf{1}_B$ for the indicator function of B, so for example, $\mathbf{1}_{X_1=j}$ is the random variable equal to 1 if $X_1 = j$ and equal to 0 otherwise.

Strong Markov Property

Previously, we described the Markov property. This says that for each time m, conditional on $X_m = i$, the process after time m begins afresh from i. Suppose, instead of conditioning on $X_m = i$, we simply waited for the process to hit state i, at some random time H. What can one say about the process after time H? What if we replaced H by a more general random time, for example H-1? In this section we shall identify a class of random times at which a version of the Markov property does hold. This class will include H but non H-1; after all, the process after time H-1 jumps straight to i, so it does not simply begin afresh.
A random variable $T : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ is called a *stopping time* if the event $\{T = n\}$ depends only on $X_0, X_1, \dots, X_n$ for $n = 0, 1, 2, \dots$. Intuitively, by watching the process, you know at the time when T occurs. If asked to stop at T, you know when to stop.

- The *first passage time*

$$T_j = inf\{n \geq 1 : X_n = j\}$$

  is a stopping time because

$$\{T_j = n\} = \{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j\}$$

- The first hitting time H^A above is a stopping time because

$$\{H^A = n\} = \{X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}$$

- The *last exit time*

$$L^A = sup\{n \geq 0 : X_n \in A\}$$

  is not in general a stopping time because the event $\{L^A = n\}$ depends on whether $(X_{n+m})_{m \geq 1}$ visits A or not.

Now, we show that the Markov property holds at stopping times. The crucial point is that, if T is a stopping time and $B \subseteq \Omega$ is determined by $X_0, X_1, \ldots, X_T$, then $B \cap \{T = m\}$ is determined by $X_0, X_1, \ldots, X_m$, for all $m = 0, 1, 2, \ldots$.

Theorem 1.1.9. ***Strong Markov property***. *Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and let T be a stopping time of $(X_n)_{n\geq 0}$. Then, conditional on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n\geq 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \ldots, X_T$.*

Proof. See Norris (1997) [52], Theorem 1.4.2 □

Recurrence and transience

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. We say that a state i is *recurent* if

$$\mathbb{P}_i(X_n = i \quad \text{for infinitely many n}) = 1.$$

We say that i is transient if

$$\mathbb{P}_i(X_n = i \quad \text{for infinitely many n}) = 0.$$

Thus a recurrent state is one to which you keep coming back and a transient state is one which you eventually leave for ever. We shall show that every state is either recurrent or transient.
Recall that the *first passage time* to state i is the random variable T_i defined by

$$T_i(\omega) = inf\{n \geq 1 : X_n(\omega) = i\}$$

where $inf\{\emptyset\} = \infty$. We now define inductively the rth passage time $T_i^{(r)}$ to state i by

$$T_i^{(0)}(\omega) = 0, \quad T_i^{(1)}(\omega) = T_i(\omega)$$

and, for $r = 0, 1, 2, \ldots,$

$$T_i^{(r+1)}(\omega) = inf\{n \geq T_i^{(r)}(\omega) + 1 : X_n(\omega) = i\}$$

The length of the rth excursion to i is then

$$S_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)} & \text{if} \quad T_i^{(r-1)} < \infty \\ 0 \quad \text{otherwise.} \end{cases} \tag{1.3}$$

Our analysis of recurrence and transience will rest on finding the joint distribution of these excursion lengths.

Lemma 1.1.10. *For $r = 2, 3, \ldots$, conditional on $T_i^{(r-1)} < \infty$, $S_i^{(r)}$ is indipendent of $\{X_m : m \leq T_i^{(r-1)}\}$ and*

$$\mathbb{P}(S_i^{(r)} = n | T_i^{(r-1)} < \infty) = \mathbb{P}_i(T_i = n).$$

Recall that the indicator function $\mathbf{1}_{\{X_1=j\}}$ is the random variable equal to 1 if $X_1 = j$ and 0 otherwise. Let us introduce the number of visits V_i to i, which may be written in terms of indicator functions as

$$V_i = \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=i\}}$$

and note that

$$\mathbb{E}_i(V_i) = \mathbb{E}_i \sum_{n=0}^{\infty} \mathbf{1}_{\{X_n=i\}} = \sum_{n=0}^{\infty} \mathbb{E}_i(\mathbf{1}_{\{X_n=i\}}) = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = i) = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

Also, we can compute the distribution of V_i under $\mathbb{P}_i$ in terms of the return probability

$$f_i = \mathbb{P}_i(T_i < \infty)$$

Lemma 1.1.11. *For $r = 0, 1, 2 \ldots$, we have $\mathbb{P}_i(V_i > r) = f_i^r$*

The next theorem gives the definition by which we establish recurrence or transience for a given state. Note that it provides two criteria for this, one in terms of the return probability, the other in terms of the n-th step transition probabilities. Both are useful.

Theorem 1.1.12. *The following dichotomy holds:*

- *if $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent and $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$;*

- *if* $\mathbb{P}_i(T_i < \infty) < 1$, *then i is transient and* $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$;

In particular, every state is either transient or recurrent.

From this theorem we can go on to solve completely the problem of recurrence or transience for Markov chains with finite state-space. We show that recurrence and transience are class properties.

Theorem 1.1.13. *Let C be a communicating class. Then either all states in C are transient or all are recurrent.*

Theorem 1.1.14. *Every recurrent class is closed.*

Theorem 1.1.15. *Every finite closed class is recurrent.*

Remember that irreducibility means that the chain can get from any state to any other, with positive probability. Then:

Theorem 1.1.16. *Suppose P is irreducible and recurrent. Then for all* $j \in I$ *we have* $\mathbb{P}(T_j < \infty) = 1$.

Invariant distributions

Many of the long-time properties of the Markov chains are connected with the notion of an invariant distribution or measure. Remember that a measure λ is any row vector $(\lambda_i : i \in I)$ with non-negative entries. We say λ is invariant if

$$\lambda P = \lambda$$

The terms equilibrium and stationary are also used to mean the same. The first result explains the term stationary.

Theorem 1.1.17. *Let* $(X_n)_{n\geq 0}$ *be Markov* (λ, P) *and suppose that* λ *is invariant for P. Then* $(X_{m+n})_{(n\geq 0)}$ *is also Markov* (λ, P).

Proof. The proof follows from theorem 1.1.6. □

The next result explains the term equilibrium.

Theorem 1.1.18. *Let I be finite. Suppose for $i \in I$ that*

$$p_{ij}^{(n)} \to \pi_j \quad \text{as} \quad n \to \infty \quad \text{for all} \quad j \in I.$$

Then $\pi = (\pi_{ij} : j \in I)$ is an invariant distribution.

Proof. See Norris (1998) [52], Theorem 1.7.2 □

In the next results we show that every irreducible and recurrent stochastic matrix P has an essentially unique positive invariant measure. For a finite state-space I, the existence of an invariant row vector is given by: the row sums of P are all 1, so the column vector of ones is an eigenvector with eigenvalue 1, so P must have a row eigenvector with eigenvalue 1. For a fixed state *k*, consider for each *i* the expected time spent in *i* between visits to *k*:

$$\gamma_i^k = \mathbb{E}_k \sum_{n=0}^{T_k - 1} 1_{\{X_n = i\}}.$$

Here the sum of indicator functions serves to count the number of times n at which $X_n = i$ before the first passage time T_k.

Theorem 1.1.19. *Let P be irreducible and recurrent. Then*

- *i)* $\gamma_k^k = 1$;
- *ii)* $\gamma^k = (\gamma_i^k : i \in I)$ *satisfies* $\gamma^k P = \gamma^k$;
- *iii)* $0 < \gamma_i^k < \infty$ *for all* $i \in I$.

Theorem 1.1.20. *Let P be irreducibile and let λ be an invariant measure for P with $\lambda_k = 1$. Then $\lambda \geq \gamma^k$. If in addition P is recurrent, then $\lambda = \gamma^k$.*

Recall that a state *i* is recurrent if

$$\mathbb{P}_i(X_n = i \quad \text{for infinitely many n}) \quad = 1$$

and we showed in theorem 1.1.12 that this is equivalent to

$$\mathbb{P}_i(T_i < \infty) = 1.$$

If in addition the expected return time

$$m_i = \mathbb{E}_i(T_i)$$

is finite, then we say i is positive recurrent. A recurrent state which fails to have this stronger property is called null recurrent.

Theorem 1.1.21. *Let P be irreducibile. Then the following are equivalent:*

- *i) every state is positive recurrent;*
- *ii) some state i is positive recurrent;*
- *iii) P has an invariant distribution, π say.*
  Moreover, when (iii) holds we have $m_i = 1/\pi_i$ for all i.

1.1.2 Convergence to equilibrium

We analyse the limiting behaviour of the n-step transition probabilities $p_{ij}^{(n)}$ as $n \to \infty$. As we saw in the theorem 1.1.18, if the state space is finite and if for some i the limit exists for all j, then it must be an invariant distribution. But, the limit does not always exists.

Let us call a state i aperiodic if $p_{ii}^{(n)} > 0$ for all sufficiently large n. Note that i is aperiodic if and only if the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$ has no common divisor than 1.

Lemma 1.1.22. *Suppose P is irreducibile and has an aperiodic state i. Then, for all states j and k, $p_{jk}^{(n)} > 0$ for all sufficiently large n. In particular, all states are aperiodic.*

Here is the main result of this section.

Theorem 1.1.23. ***Convergence to the equilibrium**. Let P be irreducible and aperiodic, and suppose that P has an invariant distribution π. Let λ be any distribution. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then*

$$\mathbb{P}(X_n = j) \to \pi_j \quad \textit{as} \quad n \to \infty \quad \textit{for all} j.$$

In particular,

$$p_{ij}^{(n)} \to \pi_j \quad as \quad n \to \infty \quad for\ all i, j.$$

Proof. See Norris (1998) [52], Theorem 1.8.3 □

Theorem 1.1.24. *Let P be irreducible. There is an integer $d \geq 1$ and a partition*

$$I = C_0 \cup C_1 \cup \cdots \cup C_{d-1}$$

such that (setting $C_{nd+r} = C_r$)

- $p_{ij}^{(n)} > 0$ *only if $i \in C_r$ and $j \in C_{r+n}$ for some r;*

- $p_{ij}^{(nd)} > 0$ *for all sufficiently large n, for all $i, j \in C_r$, for all r.*

Proof. See Norris (1998) [52], Theorem 1.8.4 □

We call d the period of P. The theorem just proved shows in particular for all $i \in I$ that d is the greatest common divisor of the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$. This is sometimes useful in identifying d.
Finally, here is a complete description of limiting behavior for irreducible chains. This generalizes theorem 1.1.23 in two respects since we require neither aperiodicity nor the existence of an invariant distribution.

Theorem 1.1.25. *Let P be irreducible of period d and let $C_0, C_1, \ldots, C_{d-1}$ be the partition obtained in theoremm 1.1.24. Let λ be a distribution with $\sum_{i \in C_0} \lambda_i = 1$. Suppose that $(X_n)_{n \geq 0}$ is Markov (λ, P). Then for $r = 0, 1, dots, d-1$ and $j \in C_r$ we have*

$$\mathbb{P}(X_{nd+r} = j) \to d/m_j \quad as \quad n \to \infty$$

where m_j is the expected return time to j. In particular, for $i \in C_0$ and $j \in C_r$ we have

$$p_{ij}^{nd+r} \to d/m_j \quad as \quad n \to \infty.$$

Proof. See Norris (1998) [52], Theorem 1.8.5 □

1.2 The Markov Approximation of Exponential Lévy processes

During the last decade continuous stochastic processes have become very popular for modeling market fluctuation. They describe the time evolution of random phenomena, as the daily price of a risky asset. As random walks are the simplest example of stochastic processes in discrete time, their continuous time relatives processes called Lévy processes in honor to the French mathematician Paul Lévy provide key examples of stochastic processes in continuous time.
There are several empirical investigations (for example Fama(1965) [23] and Mandelbrot(1963) [46] which show how the behavior of the logreturns is more skew with tails fatter than the Normal distribution.
Since Lévy processes are able to capture the skewness and kurtosis of observed asset log returns, their use in finance is becoming very widespread.

1.2.1 Definition of Lévy process

In this section we present the class of stochastic processes, also knows as Lévy processes. At the same time will look to their properties. We refer to general works on Lévy processes given by Cont and Tankov (2004) [11], Sato (1999) [62], and Schoutens (1999) [64].

Definition 1.3. A stochastic process $\{X_t\}_{t\geq 0}$ on $\mathbb{R}^d$ adapted to the filtered space $(\Omega, \Im, P, \Im_{tt\geq 0})$ that satisfies the usual conditions is a Lévy process if:

1. For any choice of $n \geq 1$ and $0 \leq t_0 \leq t_1 \cdots \leq t_n$, random variables $X_{t_0}, X_{t_1} - X_{t_0}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are independent (independent increments property).

2. $X_0 = 0$ a.s.

3. The distribution of $X_{s+t} - X_s$ does not depend on s (temporally homogeneity or stationary increments property).

4. It is stochastically continuous.

5. There is $\Omega_0 \in \Im$ with $P|\Omega_0| = 1$ such that, for every $\omega \in \Omega_0, X_t(\omega)$ is right-continuous in $t \geq 0$ and has left limits in $t > 0$.

A Lévy process on $\mathbb{R}^d$ is called a d-dimensional Lévy process. Any process satisfying (1)-(4) is called a *Lévy process in law*. We define an *additive process* as a stochastic process satisfying the conditions (1), (2), (4) and (5). Dropping condition (5), an *additive process in law* is a stochastic process satisfying (1), (2) and (4).
The conditions (1), (2) and (3) express that the stochastic process $\{X_t\}$ has independent and stationary increments. Under the conditions (2) and (3), the condition (4) can be replaced by:

$$\lim_{s \to t} P[\mid X_s - X_t \mid > \epsilon] = 0 \tag{1.4}$$

for every $t \geq 0$ and $\epsilon > 0$. Equation 1.4 does not imply that the stochastic process $\{X_t\}$ is continuous but it serves to exclude processes with jumps at fixed times. It means that a discontinuity at a fixed time t has probability zero, in other words $\{X_t\}$ is discontinuous at random times. Finally, condition (5) is usually recalled saying that the stochastic process $\{X_t\}$ is cádlág . We can say the process is cádlág, from the French 'continue á droite et limites á gauche'; the term RCLL(right continuous left limit) is sometimes also used.

In order to define the characterization of Lévy process, an important role is that one of the infinitely divisible distributions.

Definition 1.4. A probability distribution μ on $\mathbb{R}^d$ is said to be infinitely divisible if for any integer $n \geq 2$, there exists n i.i.d random variables $Y_1, \ldots, Y_n$ such that $Y_1 + \cdots + Y_n$ has distribution μ.

Since the convolution of two probability measure of μ_1 and μ_2 denoted as $\mu_1 * \mu_2$, is the distribution of the sum of two independent random variables with distribution μ_1 and μ_2, respectively, then the infinite divisibility

of μ implies that, for each n, there are n independent and identically distributed random variables $Y_1, \ldots, Y_n$ such that $Y_1 + \ldots Y_n$ has distribution μ.
Thus, if X is a Lévy process, for any $t > 0$ the distribution of X_t is infinitely divisible. This is a constraint on the possible choice of the distributions for X_t: the distribution of increments of a Lévy process has to be infinitely divisible.
Conversely, given an infinitely divisible distribution μ, for any $n \geq 1$ dividing it into n i.i.d components we can construct a random walk model on a time grid with step size $1/n$ such that the law of the position at $t = 1$ is given by μ. This procedure can be used to construct a continuous time Lévy process $\{X_t\}_{t\geq 0}$ such that the law of X_1 if given by μ:

Proposition 1.2.1. *Let $\{X_t\}_{t\geq 0}$ be a Lévy process. Then, for every t, X_t has an infinitely divisible distribution. Conversely, if μ is an infinitely divisible distribution, then there exists a Lévy process $\{X_t\}$ such that the distribution of X_1 is μ.*

Proof. See Cont and Tankov (2004) [11], Proposition 3.1 □

Given an infinitely divisible distribution μ on $\mathbb{R}^d$, it is possible to prove (see Sato (1999) [62], Lemma 7.6) the existence of an unique continuous function $\psi(z)$ from $\mathbb{R}^d$ into $\mathbb{C}$ such that $\psi(0) = 0$ and

$$\phi_\mu(z) = \int_{\mathbb{R}^d} e^{i\langle z,x\rangle}\mu(dx) = e^{\psi(z)}, \quad z \in \mathbb{R}^d$$

where $\phi_\mu(z)$ is the characteristic function of μ. Further, for every $t \in [0, \infty)$ there exists the infinitely divisible distribution μ^t with characteristic function $\phi_{\mu^t}(z) = e^{t\psi(z)}$ (see Sato (1999) [62], Lemma 7.9). So, if $\{X_t\}_{t\geq 0}$ is a Lévy process and $P_{X_1} = \mu$, then it is possible to prove that $P_{X_t} = \mu^t$ (see Sato (1999)[62] Theorem 7.10). Then, the characteristic function of the Lévy process is given by:

$$\phi_{X_t}(z) = E[e^{i\langle z,X_t\rangle}] = e^{t\psi(z)}, \quad z \in \mathbb{R}^d \tag{1.5}$$

where $\phi_{X_t}(z)$ is the characteristic function of X_t and $\psi(z)$ is called the characteristic exponent of $\{X_t\}$. Further, given the definition of cumulant generating function of a random variable, we see that ψ is the cumulant generating function of X_1:$\psi = \Psi_{X_1}$ and that the cumulant generating function of X_t varies linearly in t: $\Psi_{X_t} = t\Psi_{X_1} = t\psi$. So, from equation 1.5 we obtain the knowledge of the law of X_t is determined by the knowledge of the law of X_1.

Now, we discuss compound Poisson processes, which are the simplest examples of the Lévy processes. They can be considered as Poisson processes with random jump sizes and allow us to introduce two important theoretical tools. First, the Lévy-Khinchin formula that permits us to study distributional properties of Lévy processes and second, the Lévy-Ito decomposition that describes the structure of their sample paths.

Definition 1.5. A compound Poisson process with intensity $\lambda > 0$ and jump size distribution F is a stochastic process $\{X_t\}_{t\geq 0}$ defined as:

$$X_t = \sum_{j=1}^{N_t} Y_j$$

where jump sizes Y_j are independent and identically distributed with distribution F and $\{N_t\}_{t\geq 0}$ is a Poisson process with intensity λ and independent from $\{Y_j : j = 1, 2, \dots\}$.

We deduce from the definition that, the sample paths of X are cadlag piecewise constant functions. Further, the jump times $\{T_i\}_{i\geq 1}$ have the same law as the jump times of the Poisson process N_t. They can be expressed as partial sums of indipendent exponential random variables with parameter λ. Finally, the jump size $\{Y_i\}_{i\geq 1}$ are independent and identically distributed with law F.

Since any cadlag function may be approximated by a piecewise constant function, one can expect that the compound Poisson process gives a good approximation for general Lévy processes. Therefore, by studying

compound Poisson processes one can assume some of the properties of Lévy processes. They allow us to describe the important tool of the Lévy measure.

Proposition 1.2.2. *Let $\{X_t\}_{t\geq 0}$ be a compound Poisson process on $\mathbb{R}^d$. Its characteristic function has the following representation:*

$$E[e^{i\langle z,X_t\rangle}] = exp(t\lambda \int_{\mathbb{R}^d} (e^{i\langle z,x\rangle} - 1)F(dx)), \quad z \in \mathbb{R}^d \tag{1.6}$$

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

Let us introduce a new measure $\nu(A) = \lambda F(A)$, then the formula 1.6 becomes:

$$E[e^{i\langle z,X_t\rangle}] = exp(t \int_{\mathbb{R}^d} (e^{i\langle z,x\rangle} - 1)\nu(dx)), \quad z \in \mathbb{R}^d \tag{1.7}$$

ν is called the Lévy measure of process $\{X_t\}_{t\geq 0}$. ν is positive measure on R but not a probability measure.

The compound Poisson process is the unique Lévy process whose sample paths are piecewise constant functions (see Cont and Tankov (2004) [11] Proposition 3.3).

Let us consider the behaviour of jumps of a compound Poisson process, using the random measure. Since we know that to every compound Poisson process $\{X_t\}$ on $\mathbb{R}^d$ we can assign a random measure on $[0,\infty) \times \mathbb{R}^d$ to describe the jumps of X defined by

$$J_X(B) = \#\{(t, X_t - X_{t-}) \in B\},$$

where B is measurable subset of $[0,\infty) \times \mathbb{R}^d$. So, for every measurable set $A \subset \mathbb{R}^d$, $J_X([t_1,t_2] \times A)$ counts the number of times between t_1 and t_2 such that the size of jumps of $\{X_t\}$ belongs to A and J_X is exactly a Poisson random measure on $\mathbb{R}^d \times [0,\infty)$ with intensity measure $\mu(dx \times dt) = \nu(dx)dt = \lambda F(dx)dt$ (see Cont and Tankov (2004) [11] Proposition 3.5), that is for every measurable set $B \subset \mathbb{R}^d \times [0,\infty)$,

$$P[J_X(B) = k] = e^{-\mu(B)}\frac{\mu(B)^k}{k!}, \quad \forall k \in \mathbb{N} \tag{1.8}$$

From equation 1.8 the Lévy measure of a compound process is given as the average number of jumps per unit of time:

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

where $\Delta X_t = X_t - X_{t-}$. Every compound Poisson process can be represented by

$$X_t = \sum_{s \in [0,t]} \Delta X_s = \int_{[0,t]\times\mathbb{R}^d} x J_X(ds \times dx), \tag{1.9}$$

where J_X is a Poisson random measure with intensity measure $\nu(dx)dt$.

Let $\{X_t^0\}$ be a Lévy process with piecewise constant paths. Since every piecewise constant Lévy process is a compound Poisson process, it can be represented in the form 1.9:

$$X_t^0 = \int_{[0,t]\times\mathbb{R}^d} x J_X(ds \times dx)$$

where J_X is a Poisson random measure with intensity measure $\nu(dx)dt$ and ν is a finite measure defined by

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

Moreover, consider a Brownian motion with drift $\gamma t + W_t$ indipendent of X^0. Then, the sum $X_t = \gamma t + W_t + X_t^0$ is another Lévy process which can be expressed as

$$X_t = \gamma t + W_t + \sum_{s \in [0,t]} X_s^0 = \gamma t + W_t + \int_{[0,t]\times\mathbb{R}^d} x J_X(ds \times dx). \tag{1.10}$$

An expression 1.10 can be proved for every Lévy process. Further, given a Lévy process $\{X_t\}$, we can define its Lévy measure as we have only just done for a compound Poisson process, that is:

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

So, the measure ν is finite ($\nu(A) < \infty$) for any compact set $A \subset \mathbb{R}^d \setminus \{0\}$, otherwise the process would have an infinite number of jumps with size

in A on $[0,T]$. But it contradicts the cadlag property. ν defines a Radon measure on $\mathbb{R}^d \diagdown \{0\}$. But it is not a necessarily a finite measure,indeed the process X could have an infinite number of small jumps on [0,T]. This means that the sum of jumps becomes an infinite series. To guarantee the convergence, we impose some conditions on the measure ν, under which we obtain a decomosition of X as follows:

Proposition 1.2.3. *(Lévy-Ito decomposition). Let $\{X_t\}$ be a Lévy process on $\mathbb{R}^d$ and ν its Lévy measure. Then*

- *ν is a measure on $\mathbb{R}^d \diagdown \{0\}$ and verifies*

$$\int_{|x|\leq 1} \mid x \mid^2 \nu(dx) < \infty \quad \int_{|x|\geq 1} \nu(dx) < \infty.$$

- *The jump measure of X, denoted by J_X, is a Poisson random measure on $[0,\infty) \times \mathbb{R}^d$ with intensity measure $\nu(dx)dt$.*

- *There exist a vector γ and a d-dimensional Brownian motion $\{W_t\}$ with covariance matrix A such that*

$$X_t = \gamma t + W_t + X_t^l + lim_{\epsilon \downarrow 0} \tilde{X}_t^\epsilon \tag{1.11}$$

where

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

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

The terms in 1.11 are indipendent and the convergence in the last term is almost sure and uniform in t on $[0,T]$.

Proof. See Cont and Tankov (2004) [11], Proposition 3.7. □

From the Lévy-Ito decomposition a Lévy process is uniquely determined by a vector γ, a positive definite matrix A and a positive measure ν. The triplet (A, ν, γ) is called characteristic triplet or Lévy triplet of the process $\{X_t\}$.
Let us comment the meaning of the terms in 1.11, which give the importance of this result. The first term, $\gamma t + W_t$ is a continuous Gaussian Lévy process. So, when a Lévy procces is continuous it must be a Brownian motion with drift γ and covariance matrix A.
Then, the two terms $\{X_t^l\}$ and $\{\tilde{X}_t^\epsilon\}$ represent the jumps of $\{X_t\}$ and are described by the Lévy measure ν, where $\{X_t^l\}$ is a compound Poisson process, while $\{\tilde{X}_t^\epsilon\}$ a compensated compound Poisson process (see Cont and Tankov Proposition 2.16). From the condition $\int_{|x|\geq 1} \nu(dx) < \infty$ we deduce that a Lévy process must have a finite number of jumps with absolute value larger than 1.
Therefore, the Lévy-Ito decomposition implies that every Lévy process is a combination of a Brownian motion with drift and a possible infinite sum of independent compound Poisson process. In other words, this means that every Lévy process can be approximated by a jump-diffusion process, which is useful both in theory and in practice.

Finally, using the Lévy-Ito decomposition, we can express the characteristic function of a Lévy process in terms of its characteristic triplet (A, ν, γ).

Theorem 1.2.4. . *Let $\{X_t\}$ be e a Lévy process on $\mathbb{R}^d$ with characteristic triplet (A, ν, γ). Then*

$$E[e^{i\langle z, X_t\rangle}] = e^{t\psi(z)}, \quad z \in \mathbb{R}^d$$

with

$$\psi(z) = -\frac{1}{2}\langle z, Az\rangle + i\langle \gamma, z\rangle + \int_{\mathbb{R}^d} (e^{i\langle z,x\rangle} - 1 - i\langle z, x\rangle 1_{|x|\leq 1})\nu(dx)$$

Proof. See Cont and Tankov (2004) [11], Theorem 3.1. □

The Lévy-Khintchine representation characterizes the characteristic function of any infinitely divisible distribution, because any infinitely divisible

distribution is the distribution of a Lévy process at time t=1. From the characteristic triplet, we deduce some properties of the sample paths of Lévy process. Some paths properties of a Lévy process can be deduced by the characteristic triplet. The next result gives conditions which characterize Lévy processes of finite variation, that is whose trajectories are functions of finite variation with probability 1.

Proposition 1.2.5. *A Lévy process is of finite variation if and only if its characteristic triplet (A, ν, γ) satisfies:*

$$A = 0 \quad \textit{and} \quad \int_{|x|\langle \leq 1|x|} \nu(dx) < \infty$$

Proof. See Cont and Takov (2004) [11], Proposition 3.9 □

In this case we recall that the Lévy-Ito decomposition and Lévy-Khintchine representation can be simplified as follows:

Corollary 1.2.6. *Let $\{X_t\}_{t\geq 0}$ be a Lévy process of finite variation with Lévy triplet given $0, \nu, \gamma$. Then, $\{X_t\}$ can be expressed as the sum of its jumps between 0 and t and a linear drift term:*

$$X_t = bt + \int_{[0,t]\times\mathbb{R}^d} x J_X(ds \times dx) = bt + \sum_{s\in[0,t]}^{\Delta X_s \neq 0} \Delta X_s,$$

and its characteristic function can be expressed as

$$E[e^{i}\langle x, X_t\rangle] = exp(t(i\langle z, b\rangle + \int_{\mathcal{R}^d} (e^{i\langle z,x\rangle} - 1)\nu(dx))), \quad z \in \mathbb{R}^d,$$

where $b = \gamma - \int_{|x|\leq 1} x\nu(dx)$.

See Corollary 3.1 Cont and Takov (2004) [11].

Usually, a pure jump Lévy process (i.e the one with no Brownian component $(\sigma^2 = 0)$) is said of finite activity when $\int_{-1}^{1} \nu(dx) < \infty$ and thus there are finitely many jumps in any finite interval. Instead when $\int_{-1}^{1} \nu(dx) = \infty$ the Lévy process is called of infinite activity and in this case there are infinitely many jumps in any finite interval.

We conclude the section with the notion of completely monotone Lévy density. Given a Lévy measure with density, i.e $\nu(dx) = k(x)dx$, the Lévy density $k(x)$is called completely monotone if it can be written in the form

$$k(x) = \int_0^\infty e^{-ax}\zeta(da)$$

for some positive measure ζ. Thus, a completely monotone Lévy density relates arrival rates of large jump sizes to smaller jump sizes in such a way that large jumps arrive less frequently than small jumps.

1.2.2 Lévy processes as Markov processes

An important property of Lévy processes is the Markov property. This property states that conditionally on X_t, the evolution of the process after time t is independent on its past before this moment. In other words, for every random variable Y depending on the history $\Im_t$[1] of X_t one must have

$$E[Y|\Im_t] = E[Y|X_t]$$

In this section we define Markov processes by using transition functions, and then recall a theorem which characterize Lévy processes with spatially homogeneous transition functions.

Definition 1.6. A mapping $P_{s,t}(x, B)$ of $x \in \mathbb{R}^d$ and $B \in \mathcal{B}(\mathbb{R}^d)$ with $0 \leq s \leq t < \infty$ is called a transition function on $\mathbb{R}^d$ if:

1. it is a probability measure as a mapping of B for any fixed x;

2. it is a measurable in x for any fixed B;

3. $P_{s,s}(x, B) = \delta_x(B)$ for $s \geq 0$;

[1]We assume that filtration is the same generated by the process and satisfies the usual conditions.

4. it satisfies
   $\int_{\mathbb{R}^d} P_{s,t}(x,dy)P_{t,u}(y,B) = P_{s,u}(x,B) \quad \text{for} \quad 0 \leq s \leq t \leq u$
   This property is known as the Chapman-Kolmogorov identity.

   If, in addition,

5. $P_{s+h,t+h}(x,B)$ does not depend on h,
   then it is called a temporally homogeneous transition function and it is given by $P_t(x,B)$ such that
   $P_t(x,B) = P_{s,s+t}(x,B) \quad s \geq 0$

In the case of temporally homogeneous transition function, the property (4) is written as

$$\int_{\mathbb{R}^d} P_s(x,dy)P_t(y,B) = P_{s+t}(x,B) \quad \text{for} \quad s \geq 0 \quad \text{and} \quad t \geq 0. \tag{1.12}$$

(see Sato (1999) [62], Definition 10.1)

In order to define Markov processes, let us recall the celebrated Kolmogorov's extension theorem which is a fundamental for the theory of stochastic processes. Let $\Omega = (\mathbb{R}^d)^{[0,\infty)}$, the collection of all functions $\omega = (\omega(t))_{t\in[0,\infty)}$ into $\mathbb{R}^d$ and define $X_t(\omega) = \omega(t)$. A set

$$C = \{\omega : X_{t_1}(\omega) \in B_1, \ldots, X_{t_n} \in B_n\}$$

for $0 \leq t_1 < \cdots < t_n$ and $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^d)$ is called a cylinder set. Let $\Im$ be the σ-algebra generated by the cilinder sets.

Theorem 1.2.7. *(Kolmogorov's extenstion theorem). Suppose that, for any choice of n and $0 \leq t_1 < \cdots < t_n$, a distribution $\mu_{t_1,\ldots,t_n}$ is given and that, if $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^d)$ and $B_k = \mathbb{R}^d$, then*

$$\mu_{t_1,\ldots,t_n}(B_1 \times \cdots \times B_n) = \mu_{t_1,\ldots,t_{k-1},t_{k+1},\ldots,t_n}(B_1 \times \cdots \times B_{k-1} \times B_{k+1} \times \cdots \times B_n). \tag{1.13}$$

Then, there exists an unique probability measure P on $\Im$ such that

$$P[B_1 \times \cdots \times B_n] = \mu_{t_1,\ldots,t_n}(B_1 \times \cdots \times B_n),$$

for any choice of n, $0 \leq t_0 < \cdots < t_n$, and $B_1, \ldots, B_n \in \mathcal{B}(\mathbb{R}^d)$.

Proof. See Sato (1999) [62], Theorem 1.8 □

If a temporally homogeneous transition function $P_t(x, B)$ on $\mathbb{R}^d$ is given, then, for any $a \in \mathbb{R}^d$ we can construct a stochastic process $\{Y_t\}_{t\geq 0}$ following the theorem given above. Let $\Omega^0 = (\mathbb{R})^{[0,\infty)}$, the collection of all functions ω from $[0, \infty)$ into $\mathbb{R}^d$, $Y_t(\omega) = \omega(t)$ for $t \geq 0$, and $\Im^0$ be the σ-algebra generated by Y_t, $t \geq 0$. Define, for any $0 \leq t_0 < \cdots < t_n$ and $B_0, \ldots, B_n$,

$$\begin{aligned}\mu^a_{t_0,\ldots,t_n}(B_0 \times \cdots \times B_n) \\ &= \int P_{t_0}(a, dx_0)1_{B_0}(x_0) \int P_{t_1-t_0}(x_0, dx_1)1_{B_1}(x_1) \\ &\int P_{t_2-t_1}(x_1, dx_2)1_{B_2}(x_2) \ldots P_{t_n-t_{n-1}}(x_{n-1}, dx_n)1_{B_n}(x_n)\end{aligned}$$

The function $\mu^a_{t_0,\ldots,t_n}$ can be uniquely extended to a probability measure on $(\mathbb{R}^d)^{n+1}$ and, moreover, the family $\{\mu^a_{t_0,\ldots,t_n}\}$ satisfies the condition 1.13 by equation 1.12. Therefore, by theorem 1.2.7 there exists a unique probability measure P^a extending this family.
An important definition is that one of temporally homogeneous Markov process.

Definition 1.7. A stochastic process $\{X_t\}_{t\geq 0}$ defined on a probability space $(\Omega, \Im, P)$ is called a temporally homogeneous Markov process with temporally homogeneous transition function $\{P_t(x, B)\}$ and starting point a, if it is identical in law [2] with the process $\{Y_t\}_{t\geq 0}$ define above on $(\Omega^0, \Im^0, P^a)$. The process $\{Y_t\}$ is the path space representation of the process $\{X_t\}$. If, in addition, the transition function is temporally homogeneous, then $\{Y_t\}$ is called a *temporally homogeneous Markov process*.

[2] Two stochastic processes $\{X_t\}$ and $\{Y_t\}$ are called identical in law if, for any choice of n, $0 \leq t_1 < \cdots < t_n$ and $B_1 \ldots B_n \in \mathcal{B}(\mathbb{R}^d)$,

$$P[X_{t_1} \in B_1, \ldots, X_{t_n} \in B_n] = P[Y_{t_1} \in B_1, \ldots, Y_{t_n} \in B_n]$$

Further, the transition function $P_{s,t}$ on $\mathbb{R}^d$ is said to be spatially homogeneous if

$$P_{s,t}(x, B) = P_{s,t}(0, B - x)$$

for any s,t ,x and B, where $B - x = \{y - x : y \in B\}$.

The next theorem characterizes Lévy processes as temporally homogeneous Markov processes with spatially homogeneous transition functions.

Theorem 1.2.8.

- *Let μ be an infinitely divisible distribution on $\mathbb{R}^d$ and let $\{X_t\}$ be the Lévy process corresponding to μ. Define $P_t(x, B)$ by*

$$p_t(x, B) = \mu^t(B - x)$$

Then $p_t(x, B)$ is a temporally and spatially homogeneous transition function and $\{X_t\}$ is a Markov process with the transition function and starting point 0.

- *Conversely, any stochastically continuous, temporally homogeneous Markov process on $\mathbb{R}^d$ with spatially homogeneous transition function and starting point 0 is a Lévy process.*

Proof. See Sato (1999) [62], Theorem 10.5. □

Markov process satisfy an important property that simplifies their tractability. The so-called Markov property is given by the following:

Proposition 1.2.9. *Consider $\{Y_t : t \geq 0\}$, the path space representation of a temporally homogeneous Markov process with the a transition function $P_t(x, B)$. Let $0 \leq t_0 \leq \cdots < t_n$ and let $f(x_0, \ldots, x_n)$ be a bounded measurable function. Then $E^a[f(Y_{t_0}, \ldots, Y_{t_n})]$ is measurable in a and*

$$E^a[f(Y_{t_0}, \ldots, Y_{t_n})] = \int P_{t_0}(a, dx_0) \int P_{t_1 - t_0}(x_0, dx_1) \int P_{t_2 - t_1}(x_1, dx_2) \ldots P_{t_{n-1} - t_n}(x_{n-1}, dx_n) f(x_0, \ldots, x_n).$$

Moreover, for any $0 \leq s_0 < \cdots < s_m \leq s$ and for any bounded measurable function $g(x_0, \ldots, x_m)$, we have

$$E^a[g(Y_{s_0}, \ldots, Y_{s_m}) f(Y_{s+t_0}, \ldots, Y_{s+t_n})] = E^a[g(Y_{s_0}, \ldots, Y_{s_m}) E^{Y_s}[f(Y_{s+t_0}, \ldots, Y_{s+t_n})]]. \tag{1.14}$$

Proof. See Sato (1999) [62], Proposition 10.6. □

Let us define a filtration a set of σ-algebras $\Im_t$ such that $\Im_s \subseteq \Im_t$, for s ≤ t, and Y_t is measurable with respect to $\Im_t$. Let $\Im_t$ the σ-algebra generated by the random variables Y_s such that $s \leq t$, then the family $\{\Im_t\}$ is the smallest filtration associated to $\{Y_t\}$.
Equation 1.14 is the Markov property and is generally expressed by the filtration of a stochastic process $\{Y_t\}$. It says that if we consider the conditional expectation of $f(Y_{s+t_0}, \ldots, Y_{s+t_n})$ with respect to $\Im_s$, then it is equal to the conditional expectation of $f(Y_{s+t_0}, \ldots, Y_{s+t_n})$ with respect to Y_s:

$$E[f(Y_{s+t_0}, \ldots, Y_{s+t_n})|\Im_s] = E[f(Y_{s+t_0}, \ldots, Y_{s+t_n})|Y_s].$$

More in general, Lévy process satisfy the strong Markov property, where it is present the notion of stopping time. A stopping time T is a mapping from Ω into $[0, \infty]$ such that $\{T \leq t\} \in \Im_t$ for every $t \in [0, \infty)$. From a stopping time T, we could define a σ-algebra $\Im_T$ and a random variable Y_T, and further prove

$$E[f(Y_{T+t_0}, \ldots, Y_{T+t_n})|\Im_T] = E[f(Y_{T+t_0}, \ldots, Y_{T+t_n})|Y_T]. \tag{1.15}$$

Equation 1.15 is called the strong Markov property, and we have the Markov property when T is equal to a constant time t.

1.2.3 Exponential Lévy processes and their estimation

In this section we describe some popular Lévy processes. We start with the increasing Lévy processes, so-called subordinators. This processes can be used as time changes for other Lévy processes. We spend some time looking at density function, characteristic function, Lévy triplets and some of their properties. Then, we compute moments, variance, skewness and kurtosis, if possible. At the same time we look at the semi-heaviness of the tails.

Below is given the definition of the subordinators:

Proposition 1.2.10. *Let $\{X_t\}_{t\geq 0}$ be a Lévy process on $\mathbb{R}$. The following conditions are equivalent:*

- *i) $X_t \geq 0$ almost surely for some $t > 0$.*
- *ii) $X_t \geq 0$ for every $t > 0$.*
- *iii) Sample paths of $\{X_t\}$ are almost surely nondecreasing: $t \geq s \Rightarrow X_t \geq X_s$ almost surely.*
- *iv) The characteristic triplet of $\{X_t\}$ satisfies $A = 0, \nu((-\infty, 0]) = 0, \int_0^\infty (x \wedge 1)\nu(dx) < \infty$ and $b \geq 0$, that is, has no diffusion components, only positive jumps of finite variation and positive drift.*

Proof. See Cont and Tankov (2004) [11], Proposition 3.10. □

Let $\{S_t\}_{t\geq 0}$ be a subordinator on $\mathbb{R}$ with Lévy measure ρ and drift b as defined in the above proposition. For any time t, S_t is a positive random variable and thus we can describe it by its Laplace transform:

$$L_{S_t}(u) = E[e^{uS_t}] = e^{tl(u)}, \quad u \leq 0,$$

where

$$l(u) = bu + \int_0^\infty (e^{ux} - 1)\rho(dx). \tag{1.16}$$

The function l(u) is called Laplace exponent of $\{S_t\}$. The process S_t is increasing and the next theorem justifies the use of a subordinator $\{S_t\}$ as time change of another Lévy process.

Theorem 1.2.11. *Fixed a probability space $(\Omega, \Im, P)$. Let $\{S_t\}_{t\geq 0}$ a subordinator with Lévy measure ρ, drift b, and Laplace exponent $l(u)$, and let $\{X_t\}_{t\geq 0}$ be a Lévy process on $\mathbb{R}^d$ with Lévy triplet (A, ν, γ) and characteristic exponent $\psi(z)$. Then the process $\{Y_t\}_{t\geq 0}$ defined for each $w \in \Omega$ by $Y_t(w) = X_{S_{t(w)}}(w)$ is a Lévy process and its characteristic function is*

$$\phi_{Y_t}(z) = E[e^{i\langle z, Y_t\rangle}] = e^{tl(\psi(z))}, \quad z \in \mathbb{R}^d \tag{1.17}$$

The Lévy triplet (A^*, ν^*, γ^*) of $\{Y_t\}$ is given by

$$A^* = bA,$$

$$\nu^*(B) = b\nu(B) + \int_0^\infty P_{X_s}(B)\rho(ds), \quad \forall B \in \mathcal{B}(\mathbb{R}^d),$$
$$\gamma^* = b\gamma + \int_0^\infty \rho(ds) \int_{|x|\leq 1} xP_{X_s}(dx),$$

where P_{X_t} is the distribution of X_t.

Proof. See Cont and Tankov (2004) [11], Theorem 4.2 □

A way to find a subordinator is to define a Lévy triplet which satisfies the condition (iv) of Proposition 1.2.10. The tempered stable subordinator is defined assuming that its drift b is zero and its Lévy measure is

$$\rho(dx) = \frac{ce^{-\lambda x}}{x^{\alpha+1}} 1_{x>0} dx$$

where c and λ are positive and $0 \leq \alpha < 1$. By equation 1.16 the Laplace exponent of the tempered stable subordinator is given by

$$\begin{cases} l(u) = c\Gamma(-\alpha)\{(\alpha - u)^\alpha - \lambda^\alpha\} & \alpha \neq 0 \\ l(u) = -clog(1 - u/\lambda) & \alpha = 0. \end{cases} \tag{1.18}$$

For financial applications two important subordinators are the Gamma process, $\alpha = 0$, and the Inverse Gaussian process, $\alpha = 1/2$, which have probability density function in explicit form. If $\alpha = 0$ in 1.18, then we have the Gamma (G) process $\{X_t^{(G)}\}_{t\geq 0}$ with parameters $a > 0$ and $b > 0$, where $a = c$ and $b = \lambda$. The Laplace transform and probability density function of X_t^G are, respectively

$$L_{X_t^{(G)}}(u) = (i - u/b)^{-at}, \quad u \leq 0.$$

$$f_{X_t^{(G)}}(x : a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} exp(-xb) 1_{x>0}.$$

The Lévy triplet of the Gamma process $\{X_t^{(G)}\}$ is

$$[\frac{a(1 - exp(-b))}{b}, \quad 0, \quad \frac{a \quad exp(-bx)}{x} 1_{x>0} dx]$$

Instead, if $\alpha = 1/2$ in 1.18, then we have the Inverse Gaussian (IG) process $\{X_t^{(IG)} : t \geq 0\}$ with parameters $a > 0$ and $b > 0$, where $a = c\sqrt{2\pi}$ and

$b = \sqrt{2\lambda}$. The Laplace transform and probability density function of $X_t^{(IG)}$ are, respectively,

$$L_{X_t^{(IG)}}(u) = exp(-at(\sqrt{b^2 - 2u} - b)), \quad u \leq 0.$$

$$f_{X_t^{(IG)}}(x : a, b) = \frac{ta}{x^{3/2}\sqrt{2\pi}} exp(tab) exp(-\frac{1}{2}((ta)^2 x^{-1} + b^2 x)) 1_{x>0}.$$

The Lévy triplet of the Inverse Gamma process $\{X_t^{(IG)}\}$ is

$$[\frac{a}{b}(2N(b) - 1), \quad 0, \quad \frac{a}{x^{3/2}\sqrt{2\pi}} exp(\frac{b^2}{2}x) 1_{x>0} dx]$$

Brownian Motion

The celebrated Black & Scholes model assumes that the underlying asset follows a geometric Brownian motion, that is

$$S_t = S_0 e^{X_t}$$

where $X_t = (\mu - \frac{1}{2}\sigma^2)t + \sigma W_t$ is a Brownian motion with drift. Therefore, the volatility σ, that is the standard deviation of the log return over a time unit, is constant. But, the volatility should depend on the number of transactions occurred during a time unit, that is it should be stochastic. A way to include this other element of randomness is just to model the asset log-return as a subordinated Brownian motion with drift:

$$X_t = \mu Z_t + \sigma W_{Z_t}$$

where $\{Z_t\}$ is a subordinator. In this way we obtain the stochastic volatility $\sigma\sqrt{Z_1}$.
Two important subordinated Lévy processes in finance are the Variance Gamma and Normal Inverse Gaussian processes, that we illustrate in the following.

Variance Gamma process

The Variance Gamma (VG) process can be defined subordinating a Brownian motion with drift $\{X_t = \theta t + \sigma W_t\}$ by a Gamma process $\{Z_t^{(G)}\}$ with parameters $a = 1/\nu > 0$ and $b = 1/\nu > 0$, where $\theta \in \mathbb{R}$ and $\sigma > 0$:

$$X_t^{VG} = \theta Z_t^G + \sigma W_{Z_t^G}$$

By equation 1.17 characteristic function of X_t^{VG} is

$$\phi_{X_t^{VG}}(z; \sigma\sqrt{t}, \nu/t, \theta t) = (1 - iz\theta\nu + \frac{1}{2}\sigma^2\nu z^2)^{-t/\nu}$$

The probability density function of X_t^{VG} is known in explicit form and is given by

$$f_{X_t^{VG}}(z; \sigma\sqrt{t}, \nu/t, \theta t) = \frac{2e^{\frac{\theta x}{\sigma^2}}(\frac{x^2}{2\sigma^2/\nu+\theta^2})^{\frac{t}{2\nu}-\frac{1}{4}}}{\nu^{t/\nu}\sqrt{2\pi}\sigma\Gamma(t/\nu)} \times K_{\frac{t}{2\nu}-\frac{1}{2}}(\frac{1}{\sigma^2}\sqrt{x^2(2\sigma^2/\nu+\theta^2)}),$$

where $K_{\frac{t}{2\nu}-\frac{1}{2}}(x)$ is the modified Bessel function of the third kind with index $\frac{t}{2\nu} - \frac{1}{2}$. It is possible to show that this density is leptokurtic and with semi-heavy tails. The parameters of the model have the following interpretation: σ dictates the overall variability of the log returns of the asset, parameters ν and θ control the kurtosis or tail heaviness and the skewness of the log returns, respectively.
In particular, at time t=1 we have the following characteristics:

mean	θ
variance	$\sigma^2 + \nu\theta^2$
skewness	$\theta\nu(3\sigma^2 + 2\nu\theta^2)/(\sigma^2 + \nu\theta^2)^{3/2}$
kurtosis	$3(1 + 2\nu - \nu\sigma^4(\sigma^2 + \nu\sigma^2)^{-2})$.

We consider an alternative definition of the Variance Gamma process for determining the Lévy triplet. Madan et al (1998) [43] showed that the VG process is also equal to the difference of two independent Gamma process:

$$X_t^{VG} = X_t^{G^1} + X_t^{G^2},$$

where $X_t^{G^1}$ is Gamma process with parameters $a = C$ and $b = M$, and $X_t^{G^2}$ is an independent Gamma process with parameters $a = C$ and $b = G$. The relation between the two definitions is given by the following parametrization:

$$C = 1/\nu > 0$$
$$G = (\sqrt{\frac{1}{4}\theta^2\nu^2 + \frac{1}{2}\sigma^2\nu} - \frac{1}{2}\theta\nu)^{-1} > 0,$$
$$M = (\sqrt{\frac{1}{4}\theta^2\nu^2 + \frac{1}{2}\sigma^2\nu} + \frac{1}{2}\theta\nu)^{-1} > 0.$$

With this second definition the Lévy triplet is immediately given by $[\gamma, 0, \nu_{VG}(dx)]$, where $\nu_{VG}(dx)]$ is given by:

$$\nu_{VG}(dx) = \begin{cases} Cexp(Gx) \mid x \mid^{-1} dx, & x < 0, \\ Cexp(-Mx)x^{-1}dx, & x > 0, \end{cases}$$

$$\gamma = \frac{-C(G(exp(-M) - 1) - M(exp(-G) - 1))}{MG}.$$

With the parametrization in terms of C, G and M, the characteristic function of $X_1^{(VG)}$ reads as follows:

$$\phi_{VG}(u : C, G, M) = (\frac{GM}{GM + (M - G)iu + u^2})^C.$$

In terms of the CGM parameters this read as follows:

Table 1.1: Variance Gamma (σ, ν, θ)

	$\mathbf{VG}(\sigma, \nu, \theta)$
mean	θ
variance	$\sigma^2 + \nu\theta^2$
skewness	$\theta\nu(3\sigma^2 + 2\nu\sigma^2)^{3/2}/(\sigma^2 + \nu\theta^2)$
kurtosis	$3(1 + 2\nu - \nu\sigma^4(\sigma^2 + \nu\theta^2)^{-2})$

For further details we refer to Schoutens (2003) [64] and Madan and Seneta (1990)[44].

The Normal Inverse Gaussian process

The Normal Inverse Gaussian(NIG) process is defined subordinating the Brownian Motion with drift $\{X_t = \beta\delta^2 t + \delta W_t\}$ by an Inverse Gaussian process $\{Z_t^{IG}\}$ with parameters $a = 1$ and $b = \delta\sqrt{\alpha^2 - \beta^2}$, where $\alpha > 0, -\alpha < \beta$ and $\delta > 0$:

$$X_t^{(NIG)} = \beta\delta^2 Z_t^{IG} + \delta W_{Z_t^{IG}}.$$

By equation 1.17 the characteristic function of $X_t^{(NIG)}$ is

$$\phi_{X_t^{NIG}}(z;\alpha,\beta,t\delta) = exp(-t\delta(\sqrt{\alpha^2 - (\beta+iz)^2} - \sqrt{\alpha^2-\beta^2}))$$

The Lévy triplet $[\gamma, o, \nu_{NIG}]$ is computed using Theorem 1.2.11 where,

$$\gamma = \mu + \frac{2\delta\alpha}{\phi}\int_0^1 sinh(\beta x)K_1(\alpha x)dx,$$

$$\nu_{NIG}(dx) = \frac{\delta\alpha}{\phi}\frac{exp(\beta x)K_1(\alpha|x|)}{|x|}dx,$$

where $K_1(x)$ is the modified Bessel function of the third kind with index 1.

Table 1.2: Variance Gamma (C,G,M)

	$\mathbf{VG(C,G,M)}$
mean	$C(G-M)/(MG)$
variance	$C(G^2+M^2)/(MG)^2$
skewness	$2C^{-1/2}(G^3-M^3)/(G^2+M^2)^{3/2}$
kurtosis	$3(1+2C^{-1}(G^4+M^4)/(M^2+G^2)^2))$

Then, the probability density function of $X_t^{(NIG)}$ is known in explicit form and is given by

$$f_{X_t^{NIG}}(z;\alpha\beta,t\delta) = \frac{\alpha t\delta}{\pi} exp(t\delta\sqrt{\alpha^2-\beta^2}+\beta x) \times \frac{K_1(\alpha\sqrt{(t\delta)^2+x^2})}{\sqrt{(t\delta)^2+x^2}}$$

where $K_1(x)$ is the modified Bessel function of the third kind with index 1. Thus, we have a leptokurtic density function with semi-heavy tails. Parameters α and β describe the tail behavior and symmetry of the density function respectively, whereas δ is a scale parameter. The steepness increase monotonically with an increasing α. This has implications for the tail behavior, by the fact that large values of α imply lighter tails while smaller values of α imply heavier tails. At time $t = 1$

mean $\quad \delta\beta/\sqrt{\alpha^2-\beta^2}$

variance $\quad \alpha^2\delta(\alpha^2-\beta^2)^{-3/2}$

skewness $\quad 3\beta\alpha^{-1}\delta^{-1/2}(\alpha^2-\beta^2)^{-1/4}$

kurtosis $\quad 3(1+\frac{\alpha^2+4\beta^2}{\delta\alpha^2\sqrt{\alpha^2-\beta^2}})$

For further details see Schoutens (2003) [64] and Barndorff-Nielsen (1995)[3].

Table 1.3: The Normal Inverse Gaussian (α,β,δ)

	$\mathbf{NIG}(\alpha,\beta,\delta)$
mean	$\delta\beta/\sqrt{\alpha^2-\beta^2}$
variance	$\alpha^2\delta(\alpha^2-\beta^2)^{-3/2}$
skewness	$3\beta\alpha^{-1}\delta^{-1/2}(\alpha^2-\beta^2)^{-1/4}$
kurtosis	$3(1+\frac{\alpha^2+4\beta^2}{\delta\alpha^2\sqrt{\alpha^2-\beta^2}})$

The Meixner Process

The density of the Meixner distribution $Meixner(\alpha,\beta,\delta)$ is given by

$$f_{MXN}(x;\alpha,\beta,\delta) = \frac{(2cos(\beta/2))^{2\delta}}{2\alpha\pi\Gamma(2d)} exp(\frac{bx}{a})|\Gamma(\delta+\frac{ix}{\alpha})|^2,$$

where $\alpha > 0, -\pi < \beta < \pi, \delta > 0$.
The Meixner(α, β, δ) distribution has semi-heavy tails (see Grigelionis 2000 [28]),

$$f_{Meixner}(x : \alpha, \beta, \delta) \sim \begin{cases} C_-|x|_-^{\rho} exp(-\eta_-|x|) & as \quad x \to -\infty \\ C_+|x|_+^{\rho} exp(-\eta_+|x|) & as \quad x \to +\infty \end{cases}$$

where $\rho_- = \rho_+ = 2\delta - 1 \qquad \eta_- = (\pi - \beta)/\alpha, \qquad \eta_+ = (\pi + \beta)/\alpha$ and for some $C_-, C_+ \geq 0$. Moments of all order of this distribution exists and at time $t = 1$

mean	$\alpha\delta tan(\beta/2)$
variance	$\frac{1}{2}\alpha^2\delta(cos^{-2}(\beta/2))$
skewness	$sin(\beta/2)\sqrt{2/\delta}$
kurtosis	$3 + (2 - cos(\beta))/\delta$

Table 1.4: The Meixner (α, β, δ)

	Meixner(α, β, δ)
mean	$\alpha\delta tan(\beta/2)$
variance	$\frac{1}{2}\alpha^2\delta(cos^{-2}(\beta/2))$
skewness	$sin(\beta/2)\sqrt{2/\delta}$
kurtosis	$3 + (2 - cos(\beta))/\delta$

We can clearly see that the kurtosis of the Meixner distribution is always greater than the Normal kurtosis, which always equals 3.

The characteristic function of the Meixner(α, β, δ) distribution is given by

$$\phi_{MXN}(u; \alpha, \beta, \delta) = (\frac{cos(\beta/2)}{cosh((\alpha u - i\beta)/2)})^{2\delta}$$

The Meixner distribution is infinitely divisible:

$$\phi_{MXN}(u;\alpha,\beta,\delta) = (\phi_{Meixner}(u;\alpha,\beta,\delta))^n.$$

We can thus associate with it a Lévy process, which we call the Meixner process. More precisely, a Meixner process

$$X^{(MXN)} = \{X_t^{(MXN)}, t \geq 0\}$$

is a stochastic process which starts at zero, i.e $X_0^{(Meixner)} = 0$, has indipendent and stationary increments, and where the distribution of $X_t^{(Meixner)}$ is given by the Meixner distribution Meixner$(\alpha,\beta,\delta t)$.
We can show (see Grigelionis 1999 [28]) that Meixner process has no Brownian part and a pure jump part governed by the Lévy measure

$$\nu(dx) = \delta\frac{exp(\beta x/\alpha)}{xsinh(\pi x/\alpha)}dx$$

The first parameter in the Lévy triplet equals

$$\gamma = \alpha\delta tan(\beta/2) - 2\delta\int_1^\infty \frac{sinh(\beta x/\alpha)}{sinh(\pi x/\alpha)}dx.$$

Because $\int_{-1}^{+1}|x|\nu(dx) = \infty$, the process is of infinite variation.
The Meixner process was introduced in Schoutens and Teugels(1998) [65]. Schoutens applied this stochastic process to describe the random behavior of asset prices. For more details see Schoutens (2003) [64] and the references therein.

α-Stable Lévy Process

The characteristic function of the α- Stable distribution, $S_\alpha(\sigma,\beta,\mu)$ is:

$$\phi_S(u:\sigma,\beta,\mu) = \begin{cases} exp\{-\sigma^\alpha(|\theta|)^\alpha(1-i\beta(\text{sign}\theta)tan\frac{\pi\alpha}{2}) + i\mu\theta\} & \text{if } \lambda \neq 1 \\ exp\{-\sigma(|\theta|)(1+i\beta\frac{2}{\pi}(\text{sign}\theta)ln|\theta| + i\mu\theta\} & \text{if } \lambda = 1 \end{cases}$$

Here α is the characteristic exponent and lies in the range (0,2] ($\alpha = 2$ represents the Gaussian case), $\sigma \geq 0$ is the scale parameter, β determines the

skewness of the distribution and should lie in range [-1,1] and $\mu \in \mathbb{R}$ is the location parameter. There exist several estimators of the parameters (see Rachev and Mittnick (2000)[59]) and the fast Fourier transform allows us to obtain the density distribution in satisfying computational time. Further details on the properties of stable Paretian distributions can be found in Samorodnitsky and Taqqu (1994) [61].

Estimation

In above section we described some Lévy processes. Now we propose a computationally efficient method to approximate their parameters.

In particular, we use the method of moments for the estimation of the parameters of Brownian Motion, NIG, Variance-Gamma, and Meixner processes,[3] while for the estimation of the stable Paretian parameters we apply the consistent quantile McCulloch's method (see McCulloch (1986) [49]). McCulloch's method requires the knowledge of 5%, 25%, 50%, 75%, 95% quantiles to obtain these estimates in an acceptable computational time for any portfolio. Moreover, these methods are computationally fast and they can be used in proper portfolio optimization models. If Θ is the sample of parameters to estimate, the method of moments (MME) concerned with choosing $\hat{\Theta}$ such that the population moments are equal to the sample moments.
The MME is obtained by solving the system of equations resulting from substituting the central moments of each process. However, solving this system is not straightforward and, in some cases, one will need to rely on a numerical solution of the system. In Tables 1.5 and 1.6 we report respectively the central moments 1.5 and the parameter estimation 1.6 (of some well known Lévy processes that we present here in the following.

[3]For the estimation of NIG and Variance-Gamma see Figueroa and López(2011) [24] and for the estimation of Meixner see Mazzola and Muliere (2011) [48]

Table 1.5: Statistical Moments

Distribution	**Mean**	**Variance**	**Skewness**	**ExcessKurtosis**
Normal(μ, σ^2)	μ	σ^2	0	3
NIG(α, β, δ)	$\frac{\delta\beta}{\sqrt{\alpha^2-\beta^2}}$	$\alpha^2\delta(\alpha^2-\beta^2)^{-3/2}$	$3\beta\alpha^{-1}\delta^{-1/2}(\alpha^2-\beta^2)^{-1/4}$	$3(1+\frac{(\alpha^2+4\beta^2}{\delta\alpha^2\sqrt{\alpha^2-\beta^2}})$
VG(σ, ν, θ)	θ	$(\sigma^2+\nu\theta^2)$	$\theta\nu\frac{(3\sigma^2+2\nu\sigma^2)}{(\sigma^2+\nu\theta^2)}^{3/2}$	$3(1+2\nu-4\sigma^4(\sigma^2+\nu\theta^2)^{-2})$
Meixner(α, β, δ)	$\alpha\delta tan(\frac{\beta}{2})$	$\frac{1}{2}\alpha^2\delta(cos^{-2}(\frac{\beta}{2}))$	$\sqrt{\frac{2}{\delta}}sin(\frac{\beta}{2})$	$3+\frac{2-cos(\beta)}{\delta}$
$\mathbf{S}_\alpha(\sigma, \beta, \mu)$			McCulloch's method[4]	

[4]See McCulloch(1986)[49] provides an efficient technique to derive stable parameters estimation.

Table 1.6: Parameter Estimation

Distribution	**Parameters**		**Constraints**	
NIG$(\hat{\alpha},\hat{\beta},\hat{\delta})$	$\hat{\alpha}=\sqrt{\frac{3m_4-4m_3^2}{m_3^2}\hat{\beta}^2}$	$\hat{\beta}=\frac{3m_3}{\sqrt{m_2}(3m_4-5m_3^2)})$	$\hat{\delta}=m_2\frac{(\sqrt{\hat{\alpha}^2-\hat{\beta}^2})^3}{\hat{\alpha}^2}$	$3m_4-4m_3^2-9>0$
VG$(\hat{\sigma},\hat{\nu},\hat{\theta})$	$\hat{\sigma}^2=m_2(\frac{1}{\hat{\varepsilon}})$	$\hat{\nu}=\frac{m_4}{3}(\frac{(1+\hat{\varepsilon})^2}{1+4\hat{\varepsilon}+2\hat{\varepsilon}^2})$	$\hat{\theta}=\frac{m_1}{\hat{\sigma}^2\hat{\nu}}(\frac{1}{3+2\hat{\varepsilon}})$	$\frac{3m_3^2}{m_4}<2$[5]
Meixner$(\hat{\alpha},\hat{\beta},\hat{\delta})$	$\hat{\alpha}=\sqrt{m_2(2m_4-3m_3^2-6)}$	$\hat{\beta}=sign(m_3)arccos(\frac{m_4-2m_3^2-3}{m_4-m_3^2-3})$	$\hat{\delta}=\frac{1}{m_4-m_3^2-3}$	$2m_4-3m_3^2-6\geq 0$
$\mathbf{S}_\alpha(\hat{\sigma},\hat{\beta},\hat{\mu})$	McCulloch's method			

.

[5] $m_1=\mathbb{E}[X]; m_2=\mathbb{V}[X]; m_3=\mathbb{SKEW}[X]; m_4=\mathbb{KURT}[X]; f(\varepsilon)=\frac{\varepsilon(3+2\varepsilon)^2}{1+4\varepsilon+2\varepsilon^2)(1+\varepsilon)}$.
For the VG we find (numerically) the solution of ε of the equation: $f(\varepsilon)=\dfrac{3\mathbb{SKEW}^2}{\mathbb{KURT}}$

1.3 Markov regime switching models

Recent decades have seen extensive interest in time-varying parameter models of macroeconomic and financial time series. Due to events such as financial crisis (Hamilton 2005)[33] or abrupt changes in government policy Hamilton (1988) [29], many of these economic time series exhibit dramatic breaks in their behavior. For example, the macro economy periodically switches from boom to recession and back again, and the dynamics differ between these two regimes. Financial markets periodically switch from a low-volatility regime to a high-volatility regime and then back again. Model instability is sometimes defined as a occasional jump in a regression equation from one regime to another. It is attractive to model such transitions as a Markov process.

Initially, this process was used for modeling non-stationarities due to abrupt changes of regime in the underlying economy that generate the data. Studies treading this non-stationarity include different approaches. First, based on the assumption that first differences of the observed series follow a linear stationary process: that is, in Beveridge and Nelson (1981)[4], Nelson and Plosser (1982) [51], and Campbell and Mankiw (1987) [9] , optimal forecasts of variables are assumed to be a linear function of their lagged values. Second, a modest alternative to these currently popular approaches is suggested by Hamilton (1989) [30], exploring the consequences of specifying that first differences of the observed series follow a non-linear stationary process rather than a linear stationary process. His model can be viewed as an extension of Goldfeld and Quandt's (1973)[26] model to the important case of structural changes in the parameters of an autoregressive process.

Further, he introduces the Expectation Maximization (EM) algorithm to obtain maximum likelihood estimates of the parameters for time series processes subject to discrete regime shifts.

In the following Kim and Nelson(1999)[38] developed this new framework, casting the Markov-switching model into a state space form which

allows a much broader class of models to be estimated than before. Moreover, his algorithm proves to be much more efficient than the previous one.
In this section, the details on the Markov regime switching models and their estimation are based on the main literature of the subject, Hamilton (1994)[32] and Kim and Nelson(1999)[38].
It proceeds with a brief description of Markov regime switching framework. In the following, the evaluation of likelihood function and the EM algorithm are described. It is given a characterization of the optimal forecasts of the future for the level of a series generated by this model and the Maximum likelihood estimates (MLE) of parameters.

1.3.1 Description of the Markov Switching framework

The following description of the Markov regime switching framework follows closely that of Hamilton(1994) [32].
Consider the following process given by:

$$\begin{aligned} y_t &= \mu_{S_t} + \epsilon_t \\ \epsilon_t &\sim N(0, \sigma^2_{S_t}), \end{aligned} \tag{1.19}$$

where $S_t = 1, 2...k$ and ϵ_t follows a Normal distribution with zero mean and variance given by $\sigma^2_{S_t}$.
This is the simplest case of a model with switching dynamic. For this model, the intercept is switching states given an indicator variable S_t. This means that if there are *k* states, there will be *k* values for μ and σ^2.
If there is only one state of the world $(S_t = 1)$, formula 1 takes the shape of $y_t = \mu_1 + \epsilon_t$ and can be treated as a simple linear regression model under general conditions.
We assume that the model has two states (k=2). An alternative representation is:

$$\begin{aligned} y_t &= \mu_1 + \epsilon_t \qquad \text{for State 1} \\ y_t &= \mu_2 + \epsilon_t \qquad \text{for State 2} \end{aligned} \tag{1.20}$$

where

$$\begin{aligned} \epsilon_t &\sim (0, \sigma_1^2) \qquad \text{for State 1} \\ \epsilon_t &\sim (0, \sigma_2^2) \qquad \text{for State 2} \end{aligned} \tag{1.21}$$

This representation clearly implies two different processes for the dependent variable y_t.
When the state of the world for time *t* is 1(2), then the expectation of the dependent variable is μ_1 (μ_2) and the volatility of the innovations is σ_1^2, (σ_2^2).
Note that we do not identify the states. In general, the S_t variable simply index the states, where the interpretation is given by looking at parameter's values.
Further, we haven't said exactly the switching from one state to the other happens. For instance, how is that one should now which state of the world is for each point in time.
Suppose that we had assumed a deterministic transition of states where state 1 is true for time *t*. This greatly simplifies the model as each state is observable, and, therefore, we can treat the model given before as a regression with dummy variables. This will take the shape of

$$y_t = D_t(\mu_1 + \epsilon_{1,t}) + (1 - D_t)(\mu_2 + \epsilon_{2,t}) \tag{1.22}$$

where D_t is the dummy variable taking value of 1 if $y_t > 0$ and 0 otherwise.

For a markov regime switching model, the transition of states is *stochastic* (and not deterministic). This means that one is never sure whether there will be a switch of state or not. But, the dynamics behind the switching process is known and driven by a *transition matrix*. This matrix will control the probability of making a switch from one state to the other. It

can be represented as:

$$P = \begin{bmatrix} p_{11} & \dots & p_{k1} \\ \vdots & \ddots & \vdots \\ p_{1k} & \dots & p_{kk} \end{bmatrix} \tag{1.23}$$

For the matrix P, the element in the row *j*, column *i* p_{ij} controls the probability of a switch from state *i* to state *j*.
Since we assumed above that the model has two state and the transition probabilities between states are governed by the markov chain[6],we get:

$$\begin{aligned} Pr[S_t = 1|S_{t-1} = 1] &= p_{11} = p \\ Pr[S_t = 2|S_{t-1} = 1] &= p_{12} = 1\text{-}p \\ Pr[S_t = 2|S_{t-1} = 2] &= p_{22} = q \\ Pr[S_t = 1|S_{t-1} = 2] &= p_{21} = 1\text{-}q, \end{aligned} \tag{1.24}$$

and

$$P = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \tag{1.25}$$

In general, the row j, column i element of the P is the transition probability p_{ij}; for example, consider that for some time t the state of the world is 2. This means that the probability of a switch from state 2 to state 1 between t and t+1 will be given by p_{21}. Likewise, a probability of staying in state 2 is determined by p_{22}.

[6]Let s_t be a random variable that can assume only an integer value $\{1, 2, ..., M\}$. Suppose that the probability that s_t equals some particular value j depends on the past only through the most recent value s_{t-1}:

$$P\{s_t = j|s_{t-1} = i, s_{t-2} = k, ...\} = P\{s_t = j|s_{t-1} = i\} = p_{ij} :$$

Such a process is described as an M-*Markov Chain* with transition probabilities $\{p_{ij}\}_{i,j=1,2,...M}$.

These transition probabilities are restricted so that $p_{11}+p_{12} = p_{22}+p_{21} = 1$. This is one of the central points of the structure of a Markov regime switching model, that is, the switching of the states of the world is a stochastic process itself.

The model given by (1.19), in conjunction with the assumptions regarding the transition probabilities (1.25), will be referred as an Markov regime switching (MRS)(1) model.

1.3.2 Evaluation of Likelihood Function and EM algorithm

Optimal Inference About the Regimes

A general MS model can be estimated with two different methods, maximum likelihood or Bayesian inference (Gibbs-Sampling). In this section the models are estimated using maximum likelihood as follows.
Let consider θ as a vector of population parameters that includes $\{\mu_1, \mu_2, \sigma_1, \sigma_2\}$ and various transition probabilities p_{ij}.
One important objective will be to estimate the value of θ based on observation of Y_T. But for the moment let suppose that the value of θ is somehow known with certainty. Even if we know the value of θ, we will not know which regime the process was in at every time of the sample.
Recalling the assumption that the state variable S_t, $t = 1, 2, 3 \ldots, T$, is known a priori, that is, if the dates of the switching or structural are known a priori, the above is nothing more than a dummy variable model, where the dummy variable, S_t, equals 0 in regime 1 and 1 in regime 2.
In this case, the log-likelihood function is given by:

$$lnL = \sum_{t=1}^{T} ln(f(y_t|S_t, \theta)), \tag{1.26}$$

where

$$f(y_t|S_t, \theta) = \frac{1}{\sqrt{2\pi\sigma^2_{S_t}}} exp(-\frac{\{y_t - \mu_{S_t}\}^2}{2\sigma^2_{S_t}}$$

A major problem with the models arises when S_t, $t = 1, 2, \ldots, T$, is not observed. The best we can do is to form a probabilistic inference about the value of S_t.
Let $Pr[S_t = j|Y_t, \theta]$ be the inference about the value of S_t based on data obtained through date t and based on knowledge parameters θ.
This inference takes the form of a conditional probability that we assign to the possibility that the t-th observations was generated by the regime j. We focus on the model (1.19) using two steps to determine the log likelihood function:

STEP 1: First, consider the joint density of y_t and the unobserved S_t variable which is the product of the conditional and marginal densities:

$$f(y_t, S_t|Y_{t-1}; \theta) = f(y_t|S_t, Y_{t-1}, \theta) f(S_t|Y_{t-1}, \theta), \tag{1.27}$$

where Y_{t-1} refers to information up to time t-1 and the conditional density of y_t is given by:

$$\eta_t = \begin{bmatrix} f(y_t|S_t = 1, Y_{t-1}, \theta) \\ f(y_t|S_t = 2, Y_{t-1}, \theta) \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2\pi\sigma_1^2}} exp(-\frac{\{y_t - \mu_1\}^2}{2\sigma_1^2}) \\ \frac{1}{\sqrt{2\pi\sigma_2^2}} exp(-\frac{\{y_t - \mu_2\}^2}{2\sigma_2^2}) \end{bmatrix} \tag{1.28}$$

STEP 2: Then, to obtain the marginal density of y_t, integrate the S_t variable out of the above joint distribution by summing over all possible values of S_t:

$$
\begin{aligned}
f(y_t|Y_{t-1},\theta) &= \sum_{S_t=1}^{1} f(y_t, S_t|Y_{t-1},\theta) \\
&= \sum_{S_t=1}^{1} f(y_t|S_t, Y_{t-1},\theta) f(S_t|Y_{t-1},\theta) \\
&= \frac{1}{\sqrt{2\pi\sigma_1^2}} exp(-\frac{\{y_t-\mu_1\}^2}{2\sigma_1^2}) \times Pr[S_t=1|Y_{t-1},\theta] \\
&+ \frac{1}{\sqrt{2\pi\sigma_2^2}} exp(-\frac{\{y_t-\mu_2\}^2}{2\sigma_2^2}) \times Pr[S_t=2|Y_{t-1},\theta].
\end{aligned}
\tag{1.29}
$$

The log likelihood function then is given by

$$
\ln L = \sum_{t=1}^{T} ln\{\sum_{S_t=1}^{1} f(y_t|S_t, Y_{t-1},\theta) Pr[S_t|Y_{t-1},\theta]\}
$$

The marginal density given above can be interpreted as a weighted average of the conditional densities given $S_t = 1$ and $S_t = 2$, respectively.
To derive the marginal density of y_t ,and thus, the likelihood function, we need to calculate appropriately the weighting factors, $Pr[S_t = 1|Y_{t-1},\theta]$ and $Pr[S_t = 2|Y_{t-1},\theta]$.
Two different cases can be deal with different assumptions about the stochastic behavior of the S_t variable. The first case is when the variable S_t evolves independently of its own past value, and the second one when it follows a Markov chain. We focus on this case.
Consider the simplest case of a two-state, first-order Markov switching process for S_t with the transition probabilities given in 1.25. Under this assumptions, we adopt the following filter for the calculation of the weighting terms:

STEP 1: Given $Pr[S_{t-1} = i|Y_{t-1},\theta],\ i = 1, 2$, at the beginning of time t or the t-th iteration, the weighting terms $Pr[S_t = j|Y_{t-1},\theta],\ j = 1, 2$, are

calculated as

$$\begin{aligned} Pr[S_t = j|Y_{t-1}, \theta] &= \sum_{i=1}^{2} Pr[S_t = j,\ S_{t-1} = i|Y_{t-1}, \theta] \\ &= \sum_{i=1}^{2} Pr[S_t = j|S_{t-1} = i] Pr[S_{t-1} = i|Y_{t-1}, \theta] \\ &= \sum_{i=1}^{2} p_{ij} \times Pr[S_{t-1} = i|Y_{t-1}, \theta], \end{aligned} \tag{1.30}$$

where $Pr[S_t = j|S_{t-1} = i] = p_{ij}\ i = 1, 2,\ j = 1, 2$, are the transition probabilities.

STEP 2: Once y_t is observed at the end of time t, or at the end of the t-th iteration, we can update the probability term in the following way:

$$\begin{aligned} Pr[S_t = j|Y_t, \theta] &= Pr[S_t = j|Y_{t-1}, y_t, \theta] \\ &= \frac{f(S_t = j, y_t|Y_{t-1}, \theta)}{f(y_t|Y_{t-1}, \theta)} \\ &= \frac{f(y_t|S_t = j, Y_{t-1}, \theta) Pr[S_t = j|Y_{t-1}, \theta]}{\sum_{j=1}^{2} f(y_t|S_t = j, Y_{t-1}, \theta) Pr[S_t = j|Y_{t-1}, \theta]} \end{aligned} \tag{1.31}$$

where $Y_t = \{Y_{t-1}, y_t\}$

The above to steps may be iterated to get $Pr[S_t = j|Y_{t-1}],\ t = 1, 2, ..., T$. Following the notation used in Hamilton(1994) [32], we can simplify the optimal inference for each date t in the sample founded precedent equations in:

$$\hat{\xi}_{t|t} = \frac{(\hat{\xi}_{t|t-1} \Theta \eta_t)}{1'(\hat{\xi}_{t|t-1} \Theta \eta_t)} \tag{1.32}$$

where $\hat{\xi}_{t|t} = Pr[S_t = j|Y_t, \theta]$ and $\eta_t =$ represents the conditional density in (1.28), $1'$ is a vector of 1s, and the symbol Θ denotes element-by-element

multiplication.

To start the above filter at time $t = 1$, however, we need $Pr[S_0|Y_0] = \hat{\xi}_{0|0}$.
Several options are available for choosing the starting value.
One approach is by computing the following steady-state or unconditional probabilities of S_t that the process will fall into each regime at an arbitrary date:

$$\pi_j \equiv Pr[S_t = j|Y_0] \text{ for } j = 1, 2. \tag{1.33}$$

These are found by solving the following set of equations:

$$\begin{aligned} &\pi_j = p_{1j}\pi_1 + p_{2j}\pi_2 \; quad\text{for} \quad j = 1, 2. \\ &\pi_1 + \pi_2 = 1 \end{aligned} \tag{1.34}$$

The solution are:

$$\begin{aligned} \pi_1 = Pr[S_0 = 1|\, Y_0] &= \frac{1 - p_{22}}{2 - p_{11} - p_{22}} \\ \pi_2 = Pr[S_0 = 2|Y_0] &= \frac{1 - p_{11}}{2 - p_{11} - p_{22}} \end{aligned} \tag{1.35}$$

Another option is to set $Pr[S_0|Y_0] = \rho$, where ρ is a fixed $(N \times 1)$ vector of nonnegative constants summing to unity, such as $\rho = N^{-1}1$.
Alternatively, ρ could be estimated by maximum likelihood subject to the constraint that $1'\rho = 1$ and $\rho_{ij} \geq 0$ for j=1,2,...N.

Given a starting value $Pr[S_0|Y_0] = \xi_{0|0}$ and assumed value for population parameter θ, one can iterate on (1.31) , for $t = 1, 2, ...T$ to calculate the values of $\psi_{t|t}$ for each date t in the sample.
The log likelihood function $lnL(\theta)$ of the observed data Y_T evaluated at the value of θ that was used to perform the iterations can also be calculated as a by-product of this algorithm from (1.29)

$$lnL(\theta) = \sum_{t=1}^{T} lnf(y_t|Y_{t-1}, \theta),$$

where $f(y_t|Y_{t-1}, \theta)$ is given by (1.29).

In general, the inference problem in a Markov regime switching model consists

- estimating the parameters of the model by maximizing the likelihood function,
- making inferences about S_t, $t = 1, 2, ..., T$.

In the approach we followed, inferences on S_t are usually made conditional on the parameter estimates of the model given known.
Depending on the amount of information used in making inferences on S_t we have filtered probabilities and smoothed probabilities. Filtered probabilities refer to inferences about S_t conditional on information up to t: Y_t. These are obtained by the filter provided in this section. Smoothed probabilities refer to inferences about S_t conditional on all the information in the sample: Y_T.These are obtained by the filter provided in the next step.

Smoothed Inferences for the Regime

Given parameter estimates of the model, we can make inferences on S_t using all information on the sample. This gives us $Pr[S_t = j|Y_T, \theta]$ for $(t = 1, 2, ..., T)$, which is the smoothed probability as opposed to the filtered probability.
Here, it is assumed that we have a model with 2-Markov-Switching and calculate the smoothed inferences using an algorithm developed by Kim and Nelson(1999) [38]. Consider the following derivation of the joint probability that $S_t = j$ and $S_{t+1} = k$ based on the full information:

$$\begin{aligned} Pr[S_t = j, S_{t+1} = k|Y_T] &= Pr[S_{t+1} = k|Y_T] \times Pr[S_t = j|S_{t+1} = k, Y_T] \\ &= Pr[S_{t+1} = k|Y_T] \times Pr[S_t = j|S_{t+1} = k, Y_t] \\ &= \frac{Pr[S_{t+1} = k|Y_T] \times Pr[S_t = j|S_{t+1} = k, Y_t]}{Pr[S_{t+1} = k|Y_t]} \\ &= \frac{Pr[S_{t+1} = k|Y_T] \times Pr[S_t = j|Y_t] \times Pr[S_{t+1} = k|S_t = j]}{Pr[S_{t+1} = k|Y_t]} \end{aligned} \tag{1.36}$$

and

$$Pr[S_t = j|Y_T] = \sum_{k=1}^{2} Pr[S_t = j, S_{t+1} = k|Y_T] \tag{1.37}$$

Again, in vector form, as in Hamilton (1994) [32], this algorithm can be written as

$$\hat{\psi}_{t|T} = \hat{\psi}_{t|t}\Theta\{P' \cdot [\hat{\psi}_{t+1|T}(\div)\hat{\psi}_{t+1|t}]\}, \tag{1.38}$$

The smoothed probabilities $Pr[S_t = j|Y_T]$ are found by iterating on (1.36) backward for $t = T-1, T-2, ...1$. This iteration is started with $Pr = [S_T|Y_T, \theta]$, which is obtained from (1.20) for t=T.

Once the parameter estimates are obtained, we usually compute all the filtering probabilities $Pr[S_t|Y_t]$ and the smoothed probabilities $Pr[S_t|Y_T]$. These probabilities can help us to decide which regime y_t belongs to at each point of time.
We will generally infer that y_t is in state j if $Pr[S_t = j|Y_t] = max_k Pr[S_t = k|Y_t]$ or $Pr[S_t = j|Y_T] = max_k Pr[S_t = k|Y_T]$.
In most applications filtering probabilities and smoothed probabilities would lead to very similar conclusions.
Given that we know in each state we are currently $S_t = j$, it could be interesting to define how long, on average, will the regime last. This is what we are going to discuss in the next section.

Expected Duration of a Regime

Some important information on the expected duration of a state or regime, are given us by the matrix of the transition probabilities.
First of all, we put the transition probabilities for a first-order, k-state Markov-Switching process S_t in equations (1.23), in the following matrix notation.

$$P^* = \begin{bmatrix} p_{11} & p_{21} & \dots & p_{k1} \\ p_{12} & p_{22} & \dots & p_{k2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1k} & p_{2k} & \dots & p_{kk} \end{bmatrix} \tag{1.39}$$

where $i_k^{'} P^* = i_k^{'}$, with
$i_k = \begin{bmatrix} 1 & 1 & \dots & 1 \end{bmatrix}^{'}$. If we let π_t be a vector of $k \times 1$ steady-state probabilities, we have

$$\pi_t = \begin{bmatrix} Pr[S_t = 1] \\ Pr[S_t = 2] \\ \dots \\ Pr[S_t = k] \end{bmatrix} = \begin{bmatrix} \pi_{1t} \\ \pi_{2t} \\ \vdots \\ \pi_{kt} \end{bmatrix} \tag{1.40}$$

$i_M^{'} \pi_t = 1$.
Then according to the definition of steady state probabilities, we have $\pi_{t+1} = P^* \pi_t$ and $\pi_{t+1} = \pi_t$, and thus

$$\pi_t = P^* \pi_t \Rightarrow (I_k - P^*)\pi_t = 0_k$$

where 0_k is an $k \times 1$ matrix of zeros. Combining the above equations we have:

$$\begin{bmatrix} I_k - P^* \\ i_k^{'} \end{bmatrix} \pi_t = \begin{bmatrix} 0_k \\ 1 \end{bmatrix}, \text{ or } A\pi_t = \begin{bmatrix} 0_k \\ 1 \end{bmatrix}$$

Multiply both sides of the above equation by $(A^{'}A)^{-1}A^{'}$. Then,

$$\pi_t = (A^{'}A)^{-1}A^{'}\begin{bmatrix}0_k\\1\end{bmatrix}$$

That is the matrix of steady-state probabilities, π_t, is the last column of the matrix $(A^{'}A)^{-1}A^{'}$. The diagonal elements of the matrix of the transition probabilities contain important information on the expected duration. Note that if we define D as the duration of the state j, we have:

$$\begin{aligned}
&D = 1, if S_t = j, \text{ and } S_{t+1} \neq j;\ Pr[D=1] = (1-p_{jj})\\
&D = 2, if S_t = S_{t+1} = j, \text{ and } S_{t+2} \neq j;\ Pr[D=2] = p_{jj}(1-p_{jj})\\
&D = 3, if S_t = S_{t+1} = S_{t+2} = j, \text{ and } S_{t+3} \neq j;\ Pr[D=3] = p_{jj}^2(1-p_{jj})\\
&D = 4, if S_t = S_{t+1} = S_{t+2} = S_{t+3} = j, \text{ and } S_{t+4} \neq j;\ Pr[D=3] = p_{jj}^3(1-p_{jj})\\
&\vdots
\end{aligned}$$

Then, the expected duration of regime j can be derived as

$$\begin{aligned}
E(D) = \sum_{j=1}^{\infty} jPr[D=j] &= 1 \times Pr[S_{t+1} \neq j | S_t = j]\\
&+ 2 \times Pr[S_{t+1} = j, S_{t+2} \neq j | S_t = j]\\
&+ 3 \times Pr[S_{t+1} = j, S_{t+2} = j, S_{t+3} \neq j | S_t = j]\\
&+ 4 \times Pr[S_{t+1} = j, S_{t+2} = j, S_{t+3} = j, S_{t+4} \neq j | S_t = j]\\
&+ \dots\\
&= 1 \times (1-p_{jj}) + 2 \times p_{jj}(1-p_{jj}) + 3 \times p_{jj}^2(1-p_{jj}) + ...\\
&= \frac{1}{1-p_{jj}}
\end{aligned} \tag{1.41}$$

Forecasts for the regime

In order to complete our analysis, one could also imagine forming forecasts of how likely the process is to be in regime j in period t+1 given observations obtained through date t. Collect these forecasts in a vector $\hat{\xi}_{t+1|t}$,

which is the vector whose j-th element represents $Pr = [S_{t+1} = j|Y_t, \theta]$.
We will work on vector forms and let to the reader to see Hamilton (1989)[30] for more details.
Before that, we need to recall some notation on Markov chain. A useful representation of a Markov chain is given in the form:

$$\xi_{t+1} = P\xi_t + v_{t+1}$$

This expression implies that

$$\xi_{t+m} = vt + m + Pv_{t+m-1} + P^2 v_{t+m-2} + \cdots + P^{m-1} v_{t+1} + P^m \xi_t \quad (1.42)$$

where P^m indicates the transition matrix multiplied by itself m times. It follows from (1.42) that m-period-ahead forecasts for a Markov chain can be calculated from

$$E(\xi_{t+m}|\xi_t, \xi_{t-1}, \dots) = P^m \xi_t \quad (1.43)$$

Since, the jth element of ξ_{t+m} will be unity if $s_{t+m} = j$ and zero otherwise, the jth element of the (Nx1) vector $E(\xi_{t+m}|\xi_t, \xi_{t-1}, \dots)$ indicates the probability that s_{t+m} takes on the value j, conditional on the state of the system at date t.
If the process is in state i at date t, then (1.43) asserts that

$$\begin{bmatrix} Pr[s_{t+m} = 1|s_t = i] \\ Pr[s_{t+m} = 2|s_t = i] \\ \vdots \\ Pr[s_{t+m} = k|s_t = i] \end{bmatrix} = P^m e_i \quad (1.44)$$

where e_i denotes the ith column of I_k.
The above expression indicates that the m-period-ahead transition probabilities for a Markov chain can be calculated by multiplying the matrix P by itself m times. Specifically, the probability that an observation from regime i will be followed m periods later by an observation from regime j, $P[s_{t+m} = j|s_t = i]$, is given by the row j, column i of the matrix P^m.

Now, in order to find the optimal m-period-ahead forecast of ξ_{t+m}, let take the expectations of both sides of (1.42) conditional on information available at date t, Y_t:

$$E(\xi_{t+1}|Y_t) = P^m E(\xi_t|Y_t) \tag{1.45}$$

or

$$\hat{\xi}_{t+m|t} = P^m \hat{\xi}_{t|t}.$$

It works in the same way as optimal inference for the regime. Given a starting value $\hat{\xi}_{0|0}$, one can iterate the above expressions to calculate the optimal forecasts for each date t in the sample, where the $\hat{\xi}_{t|t}$ is given in (1.32).

Forecasts for the Observed Variables

Starting from the conditional density $f(y_t|s_t = j, Y_t, \theta$ it is straightforward to forecast y_{t+1} conditional on knowing Y_t and S_{t+1}. For example, for specification,

$$y_{t+1} = \mu_{S_{t+1}} + \epsilon_{t+1} \tag{1.46}$$

such a forecast is given by:

$$E(y_{t+1}|S_{t+1} = j, Y_t, \theta) = \mu_j + y_t. \tag{1.47}$$

We have consider the case of 2 state Markov switching, so there are 2 different conditional forecasts associated with these possible values for S_{t+1}. It is very interesting the relation between conditional forecasts and the unconditional forecasts based on actual observable variables. Note:

$$
\begin{aligned}
E(y_{t+1}|Y_t,\theta) &= \int y_{t+1}\cdot f(y_{t+1}|Y_t,\theta)\ dy_{t+1} \\
&= \int y_{t+1}\cdot\{\sum_{j=1}^{2} f(y_{t+1},S_{t+1}=j|Y_t,\theta)\}\ dy_{t+1} \\
&= \int y_{t+1}\cdot\{\sum_{j=1}^{2}[f(y_{t+1}|S_{t+1}=j,Y_t,\theta)Pr[S_{t+1}=j|Y_t,\theta]]\}\ dy_{t+1} \\
&= \sum_{j=1}^{2} Pr[S_{t+1}=j|Y_t,\theta]\int y_{t+1}\cdot f(y_{t+1}|S_{t+1}=j,Y_t,\theta)\ dy_{t+1} \\
&= \sum_{j=1}^{2} Pr[S_{t+1}=j|Y_t,\theta]E(y_{t+1}|S_{t+1}=j,Y_t,\theta)
\end{aligned}
\tag{1.48}
$$

Thus, the forecast appropriate for the jth regime, is simply multiplied by the probability that the process will be in the jth regime, and the 2 different products are added together.
If, the j=1,2 forecasts are collected in a vector $h_t^{'}$ then

$$E(y_{t+1}|Y_t,\theta) = h_t^{'}\hat{\xi}_{t+1|t}. \tag{1.49}$$

Note that although the Markov chain itself admits the linear representation, the optimal forecasts of y_{t+1} is a nonlinear function of observables, since the $\hat{\xi}_t|t$ depends nonlineary on Y_t.
Although one may use a linear model to form forecasts within a given regime, if an observation seems unlikely to have been generated by the same regime as preceding observations.
The Markov chain is clearly well suited for forming multi period forecasts as well.

MLE of parameters

In the iteration for the optimal inference and forecasts, the parameter θ was taken to be fixed, a known vector. Once the iteration has been com-

pleted for $t = 1, 2, ...T$ for a given fixed θ, the value of the loglikehood implied by that value of θ is then known. The value of θ that maximizes the log likelihood can be found numerically using the methods described in Hamilton(1994) [32]. If the transition probabilities are restricted only by the conditions that $p_{ij} \geq 0$ and $(p_{i1} + p_{i2} + ... + p_{iN}) = 1$ for all i and j, and if the initial probabilities $\hat{\xi}_{0|0}$ is taken to be a fixed value ρ unrelated to the other parameters, then it is shown in Hamilton (1990) that the maximum likelihood estimates for the transition probabilities satisfy

$$\hat{p}_{ij} = \frac{\sum_{t=2}^{T} Pr[S_t = j, S_{t-1} = i|Y_T; \hat{\theta}]}{\sum_{t=2}^{T} Pr[S_{t-1} = i|Y_T; \hat{\theta}]} \tag{1.50}$$

where $\hat{\theta}$ denotes the full vector of maximum likelihood estimates. Thus, the estimated transition probability $\hat{p}_{ij}$ is essentially the number of times state i seems to have been followed by state j divided by the number of times the process was in state i. These counts are estimated on the basis of the smoothed probabilities.
If the vector of initial probabilities ρ is regarded as a separate vector of parameters constrained only by $1'\rho$ and $\rho \geq 0$, the maximum likelihood estimates of ρ turns out to be the smoothed inference about the initial state:

$$\hat{\rho} = \hat{\xi}_{1|T} \tag{1.51}$$

The maximum likelihood estimate for the vector denoted $\lambda = \{\mu_1, \mu_2, \sigma_1^2, \sigma_2^2\}$ that governs the conditional density of y_t is characterized by

$$\sum_{t=1}^{T} (\frac{\partial log f(y_t|S_t = j, Y_{t-1}, \lambda)}{\partial \lambda'}) \hat{\xi}_{t|T} = 0. \tag{1.52}$$

The second step consists in maximizing the log likelihood with respect to θ.

If the transition probabilities are restricted only by $P_{ij} \geq 0$ and $(p_{i1} + p_{i2}) = 1$ for all i, and if the initial probability $\hat{\xi}_{1|0}$ is assumed to be fixed value ρ, then the maximum likelihood estimates of the transition probabilities are

$$\hat{p_{ij}} = \frac{\sum_{t=2}^{T} P(S_t = j, S_{t-1} = i|Y_T; \hat{\theta})}{\sum_{t=2}^{T} P(S_{t-1} = i|Y_T; \hat{\theta})}$$

where $\hat{\theta}$ is the maximum likelihood estimate of θ. $\hat{p_{ij}}$ is the number of times that state i seemed to be followed by state j divided by the number of times the process was in state i.
The maximum likelihood estimates of the other parameters governing the conditional density above are given by:

$$\sum_{t=1}^{T} (\frac{\partial log \eta_t}{\partial \alpha'})' \hat{\xi_{t|T}} = 0 \tag{1.53}$$

where α is equal to θ without the transition probabilities. Let consider the initial model (1.19). Condition (1.53) yields for

$$\hat{\mu_j} = \frac{\sum_{t=1}^{T} y_t Pr(S_t = j|Y_T, \theta}{\sum_{t=1}^{T} Pr(S_t = j|Y_T, \theta}$$

and for

$$\hat{\sigma^2} = T^{-1} \sum_{t=1}^{T} \sum_{j=1}^{N} (y_t - \hat{\mu_j})^2 Pr(S_t = j|Y_T; \theta)$$

Recalling the Markov switching regression model of the (1.20)

$$y_t = \mu_{S_t} + \epsilon_t \tag{1.54}$$

where ϵ_t is i.i.d $N(0, \sigma^2_{S_t})$. The coefficient vector for this regression is μ_1, σ_1^2 when the process is in regime 1, μ_2 and σ_2^2 when the process is in

regime 2. For this example, the vector η_t would be

$$\eta_t = \begin{bmatrix} \frac{1}{\sqrt{2\pi\sigma}} exp\{\frac{-(y_t-\mu_1)^2}{2\sigma_1^2}\} \\ \frac{1}{\sqrt{2\pi\sigma}} exp\{\frac{-(y_t-\mu_2)^2}{2\sigma_2^2}\} \end{bmatrix} \tag{1.55}$$

and for $\lambda = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$, condition on (1.53) becomes

$$\sum_{t=1}^{T}(y_t - \mu_j) \cdot Pr[s_t = j|Y_T; \hat{\theta}] = 0 \text{ for } j = 1, 2, ...N \tag{1.56}$$

$$\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{N}(y_t - \mu_j)^2 \cdot Pr[s_t = j|Y_T; \hat{\theta}]. \tag{1.57}$$

The estimate of σ^2 in (1.56) is just (1/T) times the combined sum of squared residuals from these N different regressions.
Again, this suggests an appealing algorithm for finding maximum likelihood estimates.
For the case when ρ is fixed a priori, given initial guess for the parameter vector $\theta^{(0)}$ one can evaluate (1.53) and (1.56) to generate a new estimate $\theta^{(1)}$. This estimate of $\theta^{(1)}$ can be used to reevaluate and recalculate the above expressions, producing a new estimate $\theta^{(2)}$.
One continues iterating this fashion until the change between $\theta^{(m+1)}$and$\theta^{(m)}$ is smaller than some specified convergence criterion. This again turns out to be an application of the EM algorithm. Alternatively, if ρ is to be estimated by maximum likelihood, equation $\hat{\rho} = \hat{\xi}_{1|T}$ would be added to the equations that are reevaluated with each iteration. See Hamilton (1990) [31] for details.

Chapter 2

First passage times distributions with Lévy processes and portfolio application

Most of the parametric processes used in portfolio theory are Markov processes such as Lévy processes. Several empirical works (see Lamantia et al (2006) [41]) have shown that we cannot reject the Markovian hypothesis of asset returns.
On the one hand a Markov chain should be a good model do describe the evolution of the distributional support of a given portfolio. On the other hand, since its application to predict future wealth presents a high computational complexity, the Markovianity has not been opportunely used in portfolio theory (see Leccadito et al (2007) [42]).
In this chapter we proposed some algorithms that reduce the complexity of the portfolio selection problems based on this hypothesis. In particular, we distinguish two different methods to build the transition matrixes (see Angelelli and Ortobelli (2009)[1]. With parametric portfolio selection models the transition matrix depends on the parameters of the underlying Markov process and the parameters are functions of the portfolio weights. Instead, with non parametric Markov models the transition matrix depends directly on the portfolio weights. Hence, the transition probabilities

are strictly linked to historical observations.
Since we are interested to value the impact of Markovianity in the portfolio strategies, we apply the method discussed by Duan and Simonato (2001) [20] to approximate the Markovian evolution of the portfolio wealth with a proper Markov chain. Then, using the same logic of Iaquinta and Ortobelli's [35] algorithm we are able to propose a new algorithm for computing the distribution of first passage times of the portfolio wealth.

This chapter is organized as follows: In section 2.1 we show how to model parametric and non parametric Markov chains. The Markovian evolution of the portfolio wealth with a proper Markov chain is given in section 2.2. Section 2.3 formalizes a methodology to compute the distribution of bounded stopping times. In Section 2.4 we compare the ex-post wealth obtained optimizing some parametric timing portfolio selection strategies.

2.1 Modeling Markov Processes

In this section we analyze the time evolution of the process when the portfolios dynamics is described by a homogeneous Markov chain. We distinguish between non parametric Markov processes and parametric Markov processes. In both cases we propose a distributional analysis of the time evolution of the process. We first treat with the non parametric case and then the parametric case.
Since the empirical analysis will focus on the latter kind of processes we will treat the class of Levy processes among all possible parametric Markov processes in the next section.

Consider n risky asset with gross returns [1] $z_{t+1} = [z_{1,t+1}, \ldots, z_{n,t+1}]'$. If we denote by $x = [x_1, \ldots, x_n]'$ the vector of positions taken in the n risky assets, then the portfolio return during the period $[t, t+1]$ is given by:

$$z_{(x),t+1} = x' z_{t+1} = \sum_{i=1}^{n} x_i z_{i,t+1}.$$

2.1.1 Nonparametric Markov processes

In this subsection, we formalize the Markovian nonparametric approach. As in Angelelli and Ortobelli(2009) [1], we describe the behavior of portfolios through a homogeneous Markov chain.

Let us assume that the portfolio of gross returns has support on the interval $(\min_k z_{(x),k}; \max_k z_{(x),k})$, where $z_{(x),k}$ is the k-th past observation of the portfolio $z_{(x)}$. The states are denoted by N gross return $z_{(x)}^{(i)}$ where $i \in \{1, 2, \ldots N\}$. Without loss of generality we assume that $z_{(x)}^{(i)} > z_{(x)}^{(i+1)}$ for $i = 1, \ldots N-1$. Assuming that the initial wealth W_0 is given and equal to 1, the wealth W_t at time $t = 1, \ldots, k$ is a random variable with a number of possible values increasing exponentially with time t.
Let us assume that the initial wealth W_0 at time 0 is equal to 1, while for each possible wealth W_t at time t we have N possible different values $W_{t+1} = W_t \cdot (z_{(x)}^{(i)} (i = 1, \ldots, N)$ at time $t+1$; by indexing as i_t the realized state at time t, the possible values W_t at time t are described by $W_0 \cdot z_{(x)}^{(i_1)} \cdot z_{(x)}^{(i_2)} \cdot \cdots \cdot z_{(x)}^{(i_t)}$. The sequence $\langle i_0, i_1, i_2, \ldots i_t \rangle$ is the path followed by the process up to time t where i_0 represents the initial state of the process.

[1] We define the gross return between time t and t+1 of asset i as

$$z_{i,t+1} = \frac{S_{i,t+1} + d_{i,[t,t+1]}}{S_{i,t}},$$

where $S_{i,t}$ is the price of the i-th asset at time t and $d_{i,[t,t+1]}$ is the total amount of cash dividends paid by the asset between t and t+1. We distinguish the definition of gross return from the definition of return, i.e., $z_{i,t} - 1$ or alternatively, definition of log-retuns $r_{i,t=log z_{i,t}}$.

In order to get the portfolio distribution in a reasonable time, as in the algorithm proposed by Iaquinta and Ortobelli (2006)[35], we first divide the portfolio support $(\min_k z_{(x),k}; \max_k z_{(x),k})$ in N intervals $(a_{(x),i}; a_{(x),i-1}$ where $a_{(x),i}$ is decreasing with index i and it is given by:

$$a_{(x),i} = \left(\frac{\min_k z_{(x),k}}{\max_k z_{(x),k}}\right)^{i/N} \cdot \max_k z_{(x),k}, \qquad i = 0, 1, \ldots, N.$$

Then, we compute the return associated to each state as the geometric average of the extremes of the interval $(a_{(x),i}; a_{(x),i-1})$, that is

$$z_{(x)}^{(i)} := \sqrt{a_{(x),i} a_{(x),i-1}} = \max_k z_{(x),k} \left(\frac{\max z_{(x),k}}{\min z_{(x),k}}\right)^{\left(\frac{1-2i}{2N}\right)}, \quad i = 1, 2, \ldots, N.$$

As a consequence, $z_{(x)}^{(i)} = z_{(x)}^{(1)} u^{1-i}$ where $u = \left(\frac{\max z_{(x),k}}{\min z_{(x),k}}\right)^{1/N} > 1$ and the wealth obtained along a path $\langle i_0, i_1, \ldots, i_t \rangle$ is given by $W_0 \cdot (z_{(x)}^{(1)} u)^t \cdot u^{(i_1+i_2+\cdots+i_t)}$ which can only assume $1 + t(N-1)$ distinct values instead of N^t. In particular note that $(i_1 + i_2 + \cdots + i_t)$ ranges from *t* to *tN* and the final wealth W_t does not depend on the specific path followed by the process, but on the sum of the indices of the states traversed in the first t steps only. We denote such property of a Markov chain as *recombining effect*. Thanks to the recombining effect of the Markov chain on the wealth W, the possible values after k steps of $W_k(z_{(x)}$ are $1 + k(N-1)$. They are given by $w_{(x)}^{(i,k)} = (z_{(x)}^{(1)})^k u^{(1-i)}$ $(i = 1, \ldots, (N-1)k + 1)$, where the i-th node at time k of the Markovian tree corresponds to wealth $w_{(x)}^{(i,k)}$. Moreover, the whole set of possible values of the random variables $W_t (t = 1, \ldots, k)$ can be stored in a matrix with k columns and $1 + k(N-1)$ rows resulting in $k + k^2(N-1) = O(Nk^2)$ memory space requirement. Secondly, we build the transition matrix $P_t = [p_{i,j:t}]_{1 \le i,j \le N}$ valued at time t where the probability $p_{i,j:t}$points out the probability (valued at time t) of the transition process from the state $z_{(x)}^{(i)}$ at time t to state $z_{(x)}^{(j)}$ at time $t+1$. Since we only consider homogeneous Markov chains, the transition matrix does not depend on time and it can be simply denoted by P. The entries $p_{i,j}$ of matrix P are estimated using the maximum likelihood estimates $\widehat{p}_{i,j} = \frac{\pi_{ij}(K)}{\pi_i(K)}$ where $\pi_{ij}(K)$ is the number of observations (out of K observations) that transit

from the i-th state to the j-th state and $\pi_i(K)$ is the number of observations (out of K observations) in the i-th state (see among others D'Amico and Petroni (2012a,b) [15],[16]. Since we are interested to the inter temporal behavior of the return portfolios we need the m-step transition probability matrix $P^m = [p_{i,j}^{(m)}]$ whose element $p_{i,j}^{(m)}$ represent the probability to transit from the i-th state to the j-th state in m-steps. Clearly the estimates of the probabilities $p_{i,j}^{(m)}$ are given by the elements of $\widehat{P}^m = [\widehat{p}_{i,j}^{(m)}]$. We refer to D'Amico (2003) [13] for the statistical properties of these estimators.

2.1.2 Markov Transition matrix associated to a given Lévy process

Many Lévy processes have been used to approximate the log return distributions in financial literature. The most used is the Brownian motion. Since Lévy process are particular Markov processes, we propose a distributional approximation of a parametric Lévy process by using a proper Markov chain. In particular, we use the method discussed by Duan and Simonato (2001) [20] and Staino and Ortobelli (2011) [67] for parametric Lévy processes, to build the transition matrix.

Consider n risky assets with log-returns [2] $r_{k+1} = [r_{1,k+1}, ..., r_{n,k+1}]'$. Then, the portfolio of log-returns during the period $[k, k+1]$ is given by:

$$r_{(x),k+1} = x' r_{k+1} = \sum_{i=1}^{n} x_i r_{i,k+1}, \tag{2.1}$$

where $x = [x_1, ..., x_n]'$ is the vector of the positions taken in the n risky assets at time t. Moreover, we assume that the initial wealth $W_0(x) = \sum_{i=1}^n x_i = 1$, and thus, the weights are subject to the constraint $\sum_{i=1}^n x_i = 1$.

[2] In the paper we adopt the standard definition of log-return between time t and time t+1 of asset i, as $r_{i,t+1} = \ln\left(\frac{S_{i,t+1}+d_{i,[t,t+1]}}{S_{i,t}}\right)$ where $S_{i,t}$ is the price of the i-th asset at time t and $d_{i,[t,t+1]}$ is the total amount of cash dividends paid by the asset between time t and time t+1.

Varying k we generally assume to have m i.i.d. observations of the log-returns (i.e., k=1,...,m). Under this assumption, the log wealth $S_{(x),t} = \sum_{k=1}^{t} r_{(x),k}$ obtained by a given portfolio x follows a process with stationary independent increments. Let us recall the Staino and Ortobelli's algorithm [67] to describe the Lévy behaviour of the log wealth process $\{S_t\}_{0\leq t\leq T}$. The task of this algorithm is to approximate the log-wealth process $\{S_{(x),t}\}_{0\leq t\leq L}$, at discrete times $\{0, \Delta t, 2\Delta t, \dots, s\Delta t = L\}$ by a Markov chain $\{Y_{k\Delta t}, k = 0, 1, 2, \dots, s\}$ with state space $\{p_1, p_2, \dots, p_N\}$ and transition probability matrix $P = [q_{i,j}]_{1\leq i,j\leq N}$, where N is an odd integer. In order to determine the states p_i, $i = 1, .., N$, of the Markov chain, we define the interval $[S_{(x),0} - I(N), S_{(x),0} + I(N)]$ such that

$$\Pr(S_{(x),L} \in [S_{(x),0} - I(N), S_{(x),0} + I(N)]) \approx 1.$$

The quantity $I(N)$ must depend on the number of states N and, precisely, it must satisfy the following conditions:

$$I(N) \to \infty \text{ and } I(N)/N \to 0 \text{ as } N \to \infty.$$

Under these conditions, we can guarantee that the Markov chain $\{Y_{k\Delta t}, k = 0, 1, 2, \dots, s\}$ converges weakly at times $\{0, \Delta t, 2\Delta t, \dots, s\Delta t = L\}$ to the log-wealth process $\{S_{(x),t}\}_{0\leq t\leq L}$ as $N \to \infty$. Given $I(N)$, the N states of the Markov chain are defined as:

$$p_i = S_{(x),0} - \frac{2i - N - 1}{N - 1} I(N), \quad i = 1, \dots, N.$$

Note that the states are decreasing in value such that $p_1 = S_{(x),0} + I(N)$, $p_{(N+1)/2} = S_{(x),0}$, and $p_N = S_{(x),0} - I(N)$. In our numerical applications the quantity $I(N)$ is given by:

$$I(N) = z + \ln(\ln(N)),$$

where $z = max(|z_{0.01}|, |z_{0.99}|)$, and $z_{0.01}$, $z_{0.99}$ are respectively the 1% and 99% quantiles of the $S_{(x),\Delta t}$ distribution. Our construction of the Markov chain goes on defining the cells $(c_{j+1}, c_j]$, $j = 1, \dots, N$, where $c_1 = +\infty$, $c_j = (p_j + p_{j-1})/2$, $j = 2, \dots, N$, and $c_{N+1} = -\infty$. From a computational

point of view, we have to establish a fixed value for ∞, and in our applications we set $c_1 = p_1 + I(N)$ and $c_{m+1} = p_N - I(N)$. Given the cells $(c_{j+1}, c_j]$, $j = 1, \ldots, m$, we define the transition probability q_{ij} between the state i and state j as:

$$\begin{aligned} q_{ij} &= \Pr\left(p_i + S_{(x),\Delta t} \in (c_{j+1}, c_j]\right) = \\ &= \int_{c_{j+1}-p_i}^{c_j - p_i} f_{S_{(x),\Delta t}}(u)du, \end{aligned} \tag{2.2}$$

where $f_{S_{(x),\Delta t}}(\cdot)$ is the density function of $S_{(x),\Delta t}$. Moreover, the initial distribution of the Markov chain that gives the probability to be in each state is simply given by:

$$q_j = \Pr\left(S_{(x),\Delta t} \in (c_{j+1}, c_j]\right) = \int_{c_{j+1}}^{c_j} f_{S_{(x),\Delta t}}(u)du. \tag{2.3}$$

Thus, our construction of the Markov chain characterized by the state space $\{p_1, p_2, \ldots, p_N\}$, the transition matrix $P = [q_{ij}]$ and the initial probability $q = [q_1, ..., q_N]$ is finally complete. Observe that Staino and Ortobelli (2011) [67] suggest a computationally efficient method to approximate integrals (2.3) and (2.2) for those Lévy processes we know either the density $f_{S_{(x),\Delta t}}$ or at least the characteristic function that allows us to determine the density with the fast Fourier transform.

2.2 On the approximation of the wealth distribution

In this section we examine the algorithm to approximate the final wealth distribution.

Assume the increment time $L = 1$ (i.e., $s = 1$, $L = \Delta t = 1$) in the transition matrix of log wealth process. Observe that two adjacent states of the gross returns differ by the multiplicative factor $u = \exp(p_i - p_{i+1}) = \exp(\frac{2}{N-1} I(N))$. Then, we use the algorithm proposed by Iaquinta and Ortobelli (2006)[35] to compute the wealth distribution at a given future time kL. This algorithm (differently from the Mijatović and Pistorius' (2013)

[50] one) allow us to approximate the wealth distribution with a Markovian tree whose number of nodes increase linearly with the time. Thus, we do not generally need to approximate the first step distribution wealth with a big number of states, so reducing the computational time for building the Markov transition matrix. In this context, the wealth process after k periods $W_k(x) = \exp(S_{(x),k})$ assumes $(N-1)k+1$ values (nodes) given by the formula:

$$w_{(x)}^{(i,k)} = (z_{(x)}^{1})^{k}.u^{(1-i)} \qquad \text{for} \quad i = 1, \ldots, (N-1)k+1,$$

where $z_{(x)}^{(j)} = \exp(p_j)$ are the possible states of the wealth. Thus, $w_{(x)}^{(i,k)}$is the wealth at time k in the *i*-th node of the Markovian tree. Note that whereas the distribution function depends on the initial state, the wealth do not depend on the initial state of the process, but only on the portfolio composition x. The procedure to compute the distribution function of the future wealth $W_k(x)$ at any time $k = 1, 2, \ldots, T$ is strictly connected to the recombining feature of the Markovian tree. We can compute the unconditional distribution in a computationally efficient way considering directly a sequence of matrixes $\{Q^{(k)}\}_{k=0,1,\ldots,T}$ where $Q^{(k)} = [q_{i,j}^{k}]_{1\leq i\leq (N-1)k+1, 1\leq j\leq N}$ and $q_{i,j}^{k}$ is the unconditional probability at time k to obtain the wealth $w_{(x)}^{(i,k)}$ and to be in the state $z_{(x)}^{(j)}$. We call the matrix $Q^{(k)}$ *unconditional evolution matrix* of the Markov chain or simply *evolution matrix*. The evolution matrixes $Q^{(k)}$(for $k = 0, 1, \ldots, T$) are defined recursively as follow: $Q^{(0)} = [q_1, \ldots, q_N]$, $p^{(0)} = 1 = Q^{(0)}\mathbf{1}_N$, where $\mathbf{1}_N$ is the N-dimensional column vector with 1 in each component and

$$Q^{(k)} = \text{diagM}(Q^{(k-1)}P)$$

$$p^{(k)} = Q^{(k)}\mathbf{1}_N,$$

where $p^{(k)}$ is the vector of the unconditional probability at time *k* for all nodes. The diagM operator can be defined for any $m, n \in \mathbb{N}$ as diagM:$\mathbb{R}^{mn} \to \mathbb{R}^{(m+n-1)n}$, where to any $m \times n$ matrix $A = [a_{ij}]$ associates the $(m+n-1)\times n$ matrix obtained by simply shifting down the j-th column by $(j-1)$ rows. In particular, we follow two steps:

1. create a new null matrix $D = [d_{ij}]$ of dimension $(m + n - 1) \times n$;

2. copy a_{ij} in $d_{i+(j-1),j}$ for all $1 \leq i \leq m, 1 \leq j \leq n$.

Each zero inserted in the diagonalization process represents the null probability of the respective nodes to be reached. Since diagM: $\mathbb{R}^{mn} \to \mathbb{R}^{(m+n-1)n}$ is a linear operator, we can always find a representative matrix V that for any vector belonging R^{mn} (i.e the $m \times n$ matrix $A = [a_{ij}]$ written as a vector) associate a vector in $\mathbb{R}^{(m+n-1)n}$ corresponding to the new $(m + n - 1) \times n$ matrix $D = [d_{ij}]$.
As proved by Iaquinta and Ortobelli (2006) [35], the algorithm to compute the probabilities of the Markovian tree after *k* steps has a computational complexity of $O(N^3k^2)$. This algorithm is the starting point to determine the distribution of passage time using the tree of the wealth. The following example practically shows three steps of the wealth evolution algorithm.

Example 1: Assume the wealth process evolves following a simple trinomial tree, where the initial wealth is equal to 1 (i.e., $W_0 = 1$) and the Markovian chain is homogenoeous with three wealth states: $\{1.2; 1; 0.9\}$. Let P be the 3×3 transition matrix given by:

$$P = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} = \begin{bmatrix} 0.1\% & 39.9\% & 60\% \\ 25\% & 50\% & 25\% \\ 40\% & 59\% & 1\% \end{bmatrix}$$

Suppose the matrix $Q^{(0)}$ is given by $Q^{(0)} = [q_1; q_2; q_3] = [1\%; 95\%; 4\%]$. Thus, $Q^{(1)} = \texttt{diagM}(Q^{(0)} \cdot P) = \texttt{diagM}(q_1p_{11} + q_2p_{21} + q_3p_{31}; q_1p_{12} + q_2p_{22} + q_3p_{32}; q_1p_{13} + q_2p_{23} + q_3p_{33}) = \texttt{diagM}(0.2535, 0.5026, 0.2439)$ and applying the `diagM` operator we get the (3×3) matrix: $Q^{(1)} = \begin{bmatrix} 0.2535 & 0 & 0 \\ 0 & 0.5026 & 0 \\ 0 & 0 & 0.2439 \end{bmatrix}$.
The probability at each wealth node $\{1.2, 1, 0.9\}$ after one period is given by the vector $p^{(1)} = Q^{(1)} \cdot \mathbf{1}_3 = [0.2535; 0.5026; 0.2439]$.

The (5×3) matrix $Q^{(2)}$ is given by

$$Q^{(2)} = \texttt{diagM}(Q^{(1)} \cdot P) = \texttt{diagM}\left(\begin{bmatrix} 0.00025 & 0.10115 & 0.15210 \\ 0.12565 & 0.25130 & 0.12565 \\ 0.09765 & 0.14390 & 0.00244 \end{bmatrix}\right)$$
$$= \begin{bmatrix} 0.00025 & 0 & 0 \\ 0.12565 & 0.10115 & 0 \\ 0.09765 & 0.25130 & 0.15210 \\ 0 & 0.14390 & 0.12565 \\ 0 & 0 & 0.00244 \end{bmatrix}$$

Therefore, after two periods, the probabilities on the five wealth nodes $\{1.44, 1.2, 1.08, 0.9, 0.81\}$ of the Markovian tree are given by the vector:

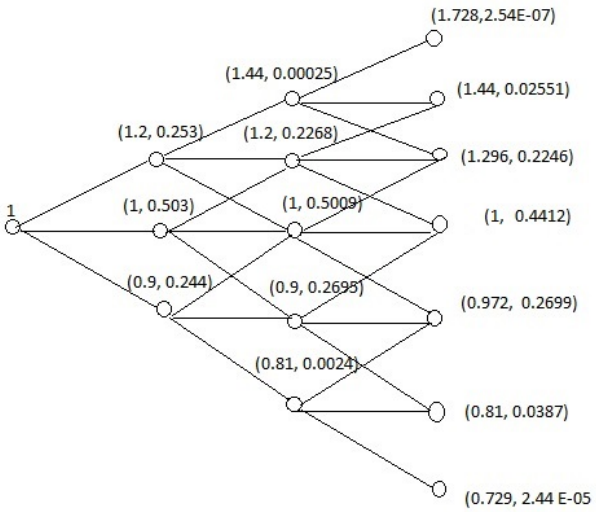
$$p^{(2)} = Q^{(2)} \cdot \mathbf{1}_3 = \begin{bmatrix} 0.00025 \\ 0.22680 \\ 0.50096 \\ 0.26955 \\ 0.00244 \end{bmatrix}.$$

After three periods, the probabilities on the seven wealth nodes are given by:

$$p^{(3)} = Q^{(3)} \cdot \mathbf{1}_3 = \begin{bmatrix} 2.54E-07 \\ 0.02551 \\ 0.22462 \\ 0.44123 \\ 0.26994 \\ 0.03867 \\ 2.44E-05 \end{bmatrix}.$$

Figure 2.1 reports the trinomial tree of Example 1 where the wealth value and the associated probability $(w^{(i,k)}, p_i^{(k)})$ are given in each node.

Figure 2.1: Final probabilities of the nodes at the 2nd step



2.3 On the approximation of the first passage times

We often have to deal with events happening at random times (i.e a positive random variable $T \geq 0$ which represents the time at which some event is going to take place). One can determine whether the event happened ($\tau \leq t$)) or not ($\tau > t$), given the information in the information flow ($\Im_t$). If the answer is yes, the random time τ is called a nonanticipating random time or *stopping time. In other words, T is a nonanticipating random time (($\Im_t$)-stopping time) if*

$$\forall t \geq 0, \qquad \{T \leq t\} \in \Im_t$$

If T_1 and T_2 are stopping times then $T_1 \wedge T_2 = inf\{T_1, T_2\}$ is also a stopping time (for further details, we refer to Karatzas and Shreve (1997) [36] and Cont and Tankov (2004) [11]). The term 'stopping time' seems to imply that something is going to stop at the τ: given a stopping time τ and a nonanticipating process $\{X_t\}$ one can define a new process $X_{\tau \wedge t}$, the process X stopped at τ, by:

$$X_{\tau \wedge t} = X_t \quad \text{if} \quad t < \tau \qquad X_{\tau \wedge t} = X_\tau \quad \text{if} \quad t \geq \tau.$$

An example of stopping times are the *hitting times*: given a nonanticipating cadlag process X, the hitting time of an open set A is defined by the first time when X reaches A:

$$T_A = inf\{t \geq 0, X_t \in A\}.$$

At any given time t, it is enough to know the past positions of X_t to see whether the set A has been reached $(T_A \leq t)$ or not $(T_A > t)$.

$$T_a = inf\{t > 0, X_t > a\}.$$

An example of a random time which is not a stopping time is the first instant $t \in [0, T]$ when X reaches its maximum:

$$T_{max} = inf\{t \in [0, T], X_t = sup_{s\in[0,T]} X_s\}.$$

Obviosly in order to know the value of the maximum one must first wait until T to observe the whole path on [0,T]. Therefore given the information $\Im_t$ at time $t < T$ one cannot decide whether T_{max} has occured or not. Given an information flow $\Im_t$ and a nonanticipating random time τ, the information set $\Im_\tau$ can be defined as the information obtained by observing all nonanticipating (cadlag) processes at τ, i.e., the $\sigma-$algebra generated by these observations: $F_\tau = \sigma(X_\tau, X \text{nonanticipating cadlag process})$. It can be shown that this definition is equivalent to the following one:

$$\Im_\tau = \{A \in \Im, \forall t \in [0, T], A \cap \{t \leq \tau\} \in \Im_t\}.$$

(see Cont and Tankov (2004) [11].

The assumption that the wealth evolves following a Markovian tree allows us to compute the distribution of several bounded stopping times in a recursive way. Considering that the investors with temporal horizon $(0, T]$ want to minimize the time to reach an high level of wealth and want to maximize the time to reach a low level of wealth, then the first time the wealth reaches a given closed set A is a stopping time extremely useful in portfolio problems. Therefore, let us consider the following first passage time of portfolio wealth $W_k(x)$:

$$\tau_A(x) = \inf\{k \geq 0 \mid k \in \mathbb{N}_0; W_k(x) \in A \subseteq \mathbb{R}\} \wedge T,$$

where A is a closed set in $\mathbb{R}$ and T is an integer greater or equal to 1. As for the wealth distribution we can easily evaluate the unconditional distribution of the stopping time[3] τ_A using two sequences of matrixes $\bar{Q}_A^{(k)} = [\bar{q}_{l,s}^{(k)}]_{\substack{1\leq l\leq (N-1)k+1\\ 1\leq s\leq N}}$ and $\tilde{Q}_A^{(k)} = [\tilde{q}_{l,s}^{(k)}]_{\substack{1\leq l\leq (N-1)k+1\\ 1\leq s\leq N}}$ for $k = 0, 1, \dots T$.
Entries $\tilde{q}_{l,s}^{(k)}$ represent the unconditional probability of obtaining, at time k, a wealth $w^{(l,k)}$ being in the state s (i.e $z^{(s)}$) while never obtaining wealths belonging to A in previous periods $t = 0, \dots, k-1$.
Entries $\bar{q}_{l,s}^{(k)}$ represent the unconditional probability of obtaining, at time k, a wealth $w^{(l,k)} \notin A$ being in the state s (i.e $z^{(s)}$) while never obtaining wealths belonging to A in previous periods $t = 0, \dots, k-1$. Accordingly, $\bar{q}_{l,s}^{(k)} = 0$ if $w^{(l,k)} \in A$.
The sequence of matrixes can be computed as follows. By definition, $\tilde{Q}_A^{(0)} = [q_1, \dots, q_N]$ whereas $\bar{Q}_A^{(0)} = [q_1, \dots, q_N]$ if $w^{(1,0)} \notin A$, otherwise $\bar{Q}_A^{(0)} = [0, \dots, 0]$ (q_s is the probability of being in state s at time 0). The other matrixes can be computed using recursion:

$$\tilde{Q}_A^{(k)} = \text{diagM}(\bar{Q}_A^{(k-1)} \cdot P),$$
$$\bar{Q}_A^{(k)} = \text{ZeroA}(\tilde{Q}_A^{(k)});$$

for $k = 1, 2, \dots, T$; where P is the transition matrix, **diagM** is the linear operator introduced above and the **ZeroA** replicates matrix $\tilde{Q}_A^{(k)}$ and then sets to 0 all entries in the rows l corresponding to a wealth $w^{(l,k)} \in A$. The idea is to calculate the probability to reach a node $w^{(l,k)}$ of the wealth-tree as the sum of probabilities of each path reaching it from $w^{(1,0)}$ while excluding from the summation all paths hitting A at a previous time. Doing so, we are able to describe the probability to reach A the first time using the recursive sequence of evolution matrixes. Finally, the distribution of stopping time τ_A can be computed as follow or in a more compact formulation:

$$P(\tau_A = k) = \begin{cases} y_A^{(k)} \cdot \tilde{Q}_A^{(k)} \cdot \mathbf{1}_N & \text{for} \quad k = 1, \dots, T-1 \\ \mathbf{1}'_{(N-1)k+1} \cdot \tilde{Q}_A^{(k)} \cdot \mathbf{1}_N & \text{for} \quad k = T \\ 0 \quad \text{otherwise,} \end{cases}$$

[3] In the following we omit (for simplicity) the portfolio composition x in the notation for the stopping time τ_A, the states $z^{(j)}$, and the wealth nodes $w^{(i,k)}$.

where $y_A^{(k)}$ is a $(N-1)k+1$ row vector whose l-th component is 1 if $w^{(l,k)} \in A$, and 0 otherwise.

This algorithm has the same computational complexity of Iaquinta and Ortobelli's [35] one and it's convergence properties are the same as D'amico (2003) [13] and Mijatović and Pistorius (2013) [50]. Note that, as for the case of wealth, the distribution of stopping time conditioned to the initial state s can be easily computed by simply setting $q_j = 1$ for $j = s$ and $q_j = 0$ for $j \neq s$ in matrix $\tilde{Q}_A^{(0)}$. The following example practically shows three steps of the first passage time algorithm.

Example 2: Assume the wealth process evolves following the same trinomial tree of Example 1. Suppose $T = 3$, $A = [1.2, +\infty)$ and $\tau_A = \inf\{k \geq 0 \mid k \in \mathbb{N}_0; W_k \geq 1.2\} \wedge 3$. If at the starting point the wealth is equal to 1 (i.e, $w^{(1,0)} \notin A$), then $\bar{Q}_A^{(0)} = \tilde{Q}_A^{(0)} = [q_1; q_2; q_3] = [1\%; 95\%; 4\%]$. Thus, the matrix $\tilde{Q}_A^{(1)}$ is given by:

$$\tilde{Q}_A^{(1)} = \text{diagM}(\bar{Q}_A^{(0)} \cdot P) = \begin{bmatrix} 0.2535 & 0 & 0 \\ 0 & 0.5026 & 0 \\ 0 & 0 & 0.2439 \end{bmatrix},$$

and

$$\bar{Q}_A^{(1)} = \text{ZeroA}(\tilde{Q}_A^{(1)}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0.5026 & 0 \\ 0 & 0 & 0.2439 \end{bmatrix}.$$

Similarly, $\tilde{Q}_A^{(2)}$ is given by:

$$\tilde{Q}_A^{(2)} = \text{diagM}(\bar{Q}_A^{(1)} \cdot P) = \begin{bmatrix} 0 & 0 & 0 \\ 0.12565 & 0 & 0 \\ 0.09765 & 0.25130 & 0 \\ 0 & 0.14390 & 0.12565 \\ 0 & 0 & 0.00244 \end{bmatrix}$$

and

$$\bar{Q}_A^{(2)} = \text{ZeroA}(\tilde{Q}_A^{(2)}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.09765 & 0.25130 & 0 \\ 0 & 0.14390 & 0.12565 \\ 0 & 0 & 0.00244 \end{bmatrix}.$$

Finally, in T=3, the matrix $\tilde{Q}_A^{(3)} = \text{diagM}(\bar{Q}_A^{(2)} \cdot P)$ is given by:

$$\tilde{Q}_A^{(3)} = \texttt{diag M} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.0629 & 0.1646 & 0.1214 \\ 0.0862 & 0.1461 & 0.0372 \\ 0.0010 & 0.0014 & 2.44E-5 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0.0629 & 0 & 0 \\ 0.0862 & 0.1646 & 0 \\ 0.0010 & 0.1461 & 0.1214 \\ 0 & 0.0014 & 0.0372 \\ 0 & 0 & 2.44E-5 \end{bmatrix}.$$

Since the wealth nodes after one period are given by $\{1.2; 1; 0.9\}$, then $y_A^{(1)} = [1; 0; 0]$ and the probability the stopping time $\tau_A = 1$ is:

$$P(\tau_{[1.2,+\infty)} = 1) = y_A^{(1)} \cdot \tilde{Q}_A^{(1)} \cdot \mathbf{1}_3 = 0.25351.$$

After two periods the five wealth nodes are $\{1.44; 1.2; 1.08; 0.9; 0.81\}$, then $y_A^{(2)} = [1; 1; 0; 0; 0]$ and the probability the stopping time $\tau_A = 2$ is:

$$P(\tau_{[1.2,+\infty)} = 2) = y_A^{(2)} \cdot \tilde{Q}_A^{(2)} \cdot \mathbf{1}_3 = 0.12565.$$

In the last step, when T=3, the probability the stopping time $\tau_A = 3$ is given by:

$$P(\tau_{[1.2,+\infty)} = 3) = \mathbf{1'}_7 \cdot \tilde{Q}_A^{(3)} \cdot \mathbf{1}_3 = 0.62084.$$

Observe that, as we expected, $P(\tau_{[1.2,+\infty)} = 1)+P(\tau_{[1.2,+\infty)} = 2)+P(\tau_{[1.2,+\infty)} = 3) = 1$.

2.4 Practical timing portfolio strategies on the US stock market

In this section we propose an empirical comparison based on some timing portfolio strategies on the US market. In particular, we compare the ex-post wealth we obtain maximizing either the classic Sharpe ratio or a performance measure obtained using the average of proper first passage times. The main objective of this empirical experiment is to show that the techniques previously presented can be applied in practical portfolio

problems and they could furnish alternative interesting portfolio functionals. Moreover the ex-post comparison among different distributional assumptions could give some indications on the best process to use in timing portfolio strategies.

The classic static portfolio selection problem when no short sales are allowed, can be represented as the maximization of a functional $f : (\Omega, \Im, P) \longmapsto \mathbb{R}$ applied to the random portfolio of gross returns $z_{(x)}$, subject to the portfolio weights belonging to the $(n-1)$-dimensional simplex $S = \{x \in \mathbb{R}^n | \sum_{i=1}^{n} x_i = 1; x_i \geq 0\}$, i.e.,

$$\max_{x \in S} f(z_{(x)}).$$

Typically, the functional $f(.)$ is a performance measure or an utility functional. In both cases the functional $f(.)$ should be isotonic with a particular ordering of preference $\succeq$, that is, if X is preferred to Y ($X \succeq Y$), then $f(X) \geq f(Y)$. Probably the most known performance measure used in financial choices is the Sharpe ratio (see, among others, Sharpe (1994) [66]):

$$f(z_{(x)}) = SR(z_{(x)}) = \frac{E(z_{(x)} - z_f)}{\sigma_{z_{(x)}}}, \tag{2.4}$$

which evaluate the excess return with respect to the riskless gross return z_f for unity of risk where the risk is the standard deviation $\sigma_{z_{(x)}}$. The choice of the functional $f(.)$ plays a crucial role in the portfolio strategy. Isotonic utility functionals with non satiable and risk averse preferences (such as the Sharpe ratio) have been used in several financial applications. However, as suggested in behavioural finance all investors prefer more to less and they could be neither risk averse nor risk lover. For this reason it makes sense to consider functionals that are monotone, even though they are not consistent with an uncertainty/aggressive order (see, among others, Rachev et al. (2008)[60]). In this work we order the choices with respect to their timing. Thus, we use and describe only functionals that consider the forecasted first passage times of the wealth before the investor's temporal horizon T. Since investors want to maximize the first time they lose wealth and minimize the first time their wealth increases enough, we

suggest to optimize the average of two first passage times under different distributional assumptions of the wealth process. In particular, using the algorithm presented in the previous section, we are able to compute the distributions of the following stopping times:

$$\tau_d(x) = \inf\{k \in \mathbb{N} | W_k(x) \leq 0.98\} \wedge T$$

$$\tau_u(x) = \inf\{k \in \mathbb{N} | W_k(x) \geq 1.2\} \wedge T,$$

where we consider the first time the future wealth loses 2% and the first time the future wealth increases 20% . The choice of these percentages (2% and 20%) also depends on the temporal horizon T, that in our experiments is equal to 1 month (20 trading days). We choose these percentages because the index S&P 500 presents positive probability to increase more than 20% and decrease more than 2% in 1 month in all the examined period. Then, under different distributional assumptions, we suggest to maximize the following timing portfolio performance:

$$f(\tau_d, \tau_u, T) = \frac{E(\tau_d I_{[\tau_d<T]} + 1000 I_{[\tau_d \geq T]})}{E(\tau_u I_{[\tau_u<T]} + 1000 I_{[\tau_u \geq T]})}, \tag{2.5}$$

where $I_{[\omega \in B]} = \begin{cases} 1 & \text{if} \quad \omega \in B \\ 0 & \text{otherwise} \end{cases}$. In performance measure (2.5) we penalize[4] the case the first passage time τ_u overcome the temporal horizon T (i.e. $\tau_u \geq T$) while we reward the possibility that the first passage time τ_d overcome the temporal horizon T (i.e. $\tau_d \geq T$).

2.4.1 The Dataset: ex-ante empirical evidence

In this section, we use the historical observations of the components of the S&P500[5] (which were components at time 03/10/2013)) from July

[4]Optimizing the functional (2.5) among the components of the S&P 500, we could observe that the optimal portfolio components do not change using penalty values 1000, 10 000 or 100 000 in formula (2.5). For this reason we use and fix the penalty value 1000.

[5] We download the historical series from Thomson Reuter Datastream.

3, 2002 to October 3, 2013 for a total of 2832 daily observations. Let us describe the statistical characteristics of the used returns. First we test the different distributional assumptions for which we evaluate the impact of timing portfolio strategies (Gaussian, NIG, VG, Meixner and α Stable distributions). For this ex-ante empirical analysis we use the same window of daily observations used in the ex-post analysis. Thus, as for the portfolio problem, every 5 trading days (starting from December 31, 2002) we compute the statistics for each asset based on the previous 125 observations. Then, we compute the average results over the time and among all the log-returns. In particular, Table 2.1 reports the average values for the parameter estimates of each distribution and the average results of the Jarque-Bera, and Kolmogorov-Smirnov tests with a 95% confidence level.

Generally the log returns present heavy tails as suggested by the average of the kurtosis (bigger than 3) and the other heavy tails parameters (α). Moreover, Table 2.1 reports that the log returns are (in average) asymmetric (as suggested by all the skewness parameters which are different from zero).

To test whether log returns follow a normal distribution, we compute the Jarque-Bera statistic. Similarly, we employ Kolmogorov-Smirnov statistic to test whether log returns follow a NIG, VG, Meixner or α-Stable distribution. In Table 2.1 we report the percentage of assets whose distribution is rejected either for the JB test or for the KS tests. It is clear that the Gaussian distribution is rejected for most of the assets while the other distributions present much better results (in average).

2.4.2 An ex- post comparison among timing portfolio strategies

In the out of sample portfolio analysis we determine and re-calibrate weekly the optimal portfolio (every 5 trading days) using a window of six months historical observations (125 trading days). Since the number of observations should increase proportionally with the number of assets (see Papp et al. (2008) [57], Kondor et al.(2007) [40]), it is necessary to

Table 2.1: Statistics on the ex-ante log returns

Average of ex-ante parameters			
Mean	St.deviation	Skewness	Kurtosis
0.0004	0.0205	-0.0412	6.3931
NIG(α, β, δ)			
$\alpha = 55.17$	$\beta = -0.176$	$\delta = 0.0165$	
Variance Gamma(σ, ν, θ)			
$\sigma = 0.020$	$\nu = 1.9161$	$\theta = 1.58e-05$	
Meixner(α, β, δ)			
$\alpha = 1.4838$	$\beta = 0.0088$	$\delta = 1.770$	
$S_\alpha(\sigma, \beta, \mu)$			
$\alpha = 1.729$	$\sigma = 0.0121$	$\beta = 0.0329$	$\mu = 0.0004$

Average of statistic tests on the used assets				
JB test 95%	KS-Stable	KS-NIG	KS-VG	KS -Meixner
0.6375	0.00867	0.03196	0.1719	0.0579

find the right trade-off between a statistical approximation of the historical series depending only on a few parameters and the number of historical observations. Therefore, we suggest to reduce the dimensionality of this large scale portfolio problem preselecting the 10% most relevant stocks (50 assets) according to the highest Sharpe ratio (see, among others, Ortobelli et al (2011) [53]). This procedure is applied at any recalibration time and implies that there is a very high turnover. In all the empirical analysis we assume that:

1. the temporal horizon T=20, in order to have a large number of nodes of the final wealth;

2. the performance (2.5) is optimized for all the Lévy processes of Section 2 (Brownian motion, NIG, VG, Meixner and α Stable processes);

3. the approximating Markov chains have N=9 states (this number of states is the right trade-off between the computational complexity of the algorithm and the proper approximation of the underlying process[6]);

4. the initial wealth W_0 is equal to 1 at the date December 31, 2002;

5. there are proportional transaction costs[7] of 5 basis points [8];

6. we cannot invest more than 20% in a single asset (i.e. $x_i \leq 0.2$) in order to guarantee a proper diversification.

Portfolio optimization leads to different results depending on the adopted distribution for the process. For each strategy, we have to compute the optimal portfolio composition exactly 548 times and at the k-th optimization three main steps are performed to compute the ex-post final wealth:

1. Step 1. Preselect the first 50 assets with the highest Sharpe ratio among all those active in the last six months.

[6]On the computational complexity of Markovian portfolio problems see Angelelli and Ortobelli (2009b) [2].

[7]The portfolio selection problem with transaction costs has been studied in several papers. See, among others, Dumas and Luciano (1991) [21] and the references therein. In this paper we optimize strategies with transaction costs using the same methodology proposed by Ortobelli et al. (2010) [54].

[8]We consider these transaction costs level since in some international trading platforms are payed about 10 basis points with some maximum transaction costs for order (see for example IB platform on the web site: https://www.interactivebrokers.com/en/index.php?f=commission&p=stocks2). Thus (using the limits of IB platform) we could test and find that the level of proportional transaction costs are lower than 5 basis points on average optimizing the Sharpe ratio for any initial wealth bigger than 250,000 USD.

2. Step 2. Determine the optimal portfolio $x_M^{(k)}$ that maximizes a performance ratio $\rho(x^{(k)})$ (either the Sharpe ratio (2.4) or the timing portfolio performance (2.5)) associated to the strategy, i.e. the solution of the following optimization problem:

$$\max_{x^{(k)}} \rho(x^{(k)})$$

$$s.t.$$

$$\sum_{i=1}^{n} x_i^{(k)} = 1; \;\; 0 \leq x_i^{(k)} \leq 0.2; \;\; i = 1, ..., n$$

Clearly, the computational complexity of the problem with portfolio performance (2.5) is much more higher than optimizing the Sharpe ratio, because in the first case the problem present more local optima. Therefore, we suggest to use the function *patternsearch* of Matlab 2013 with starting point the optimal solution of (2.5) obtained with the heuristic (for global optimization) proposed by Angelelli and Ortobelli (2009) [1].

3. Step 3. Compute the ex-post final wealth given by:

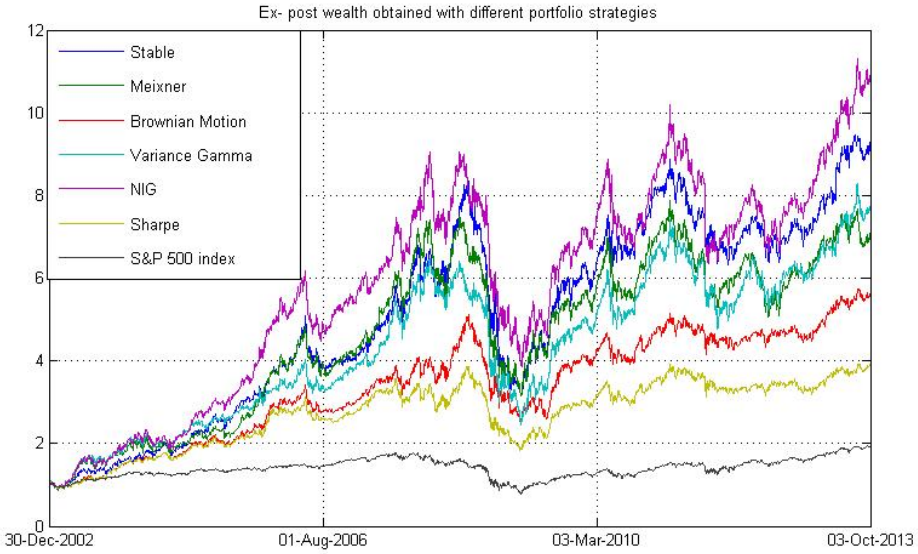
$$W_{t_{k+1}} = (W_{t_k} - tc_{t_k})(x_M^{(k)\prime}) z_{(t_k+1)}^{(\text{ex post})},$$

where tc_{t_k} are the proportional transaction costs we get changing the portfolio, $z_{(t_k+1)}^{(\text{ex post})}$ is the vector of observed gross returns between t_k and t_{k+1}.

Steps 1, 2, and 3 are repeated for the two performance ratios (2.4) and (2.5), and all the distributional assumptions until some observations are available. The results of this ex-post empirical analysis are reported in Figures 2.2, 2.3 and Tables 2.2 and 2.3.

In Figure 2.2 all the timing portfolio strategies outperform the strategy based on the optimization of the Sharpe ratio (that, however, performs better than the S&P 500 index in the same analyzed period). Moreover, we observe that the timing portfolio strategies based on a non-Gaussian

Figure 2.2: Ex post final wealth obtained with different portfolio strategies



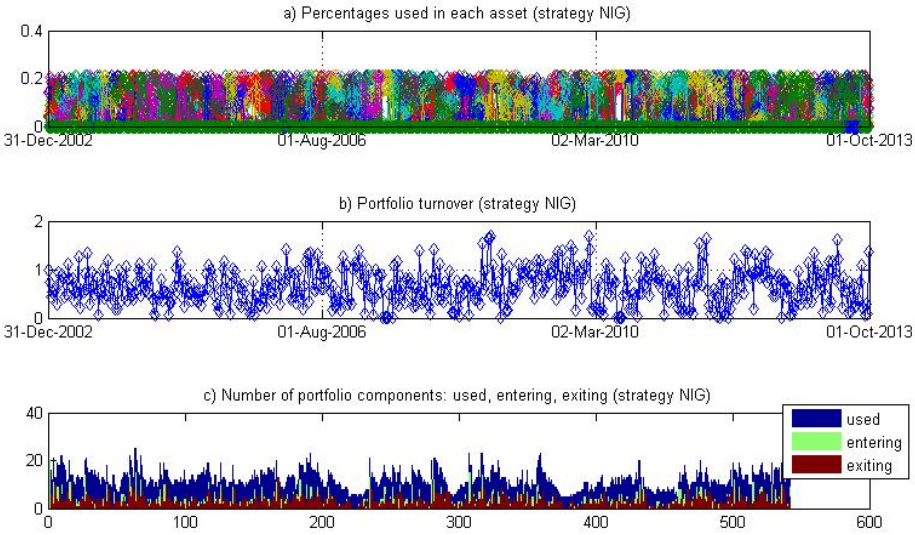
Lévy process outperforms the one based on the Brownian Motion. Graphically the "NIG strategy" seems the most promising in terms of ex-post final wealth.

Figure 2.3, examines the portfolio diversification and turnover for the NIG strategy. Since we obtain similar portfolio diversification and turnover for all the other portfolio strategies, we report the analysis only of the "NIG strategy". In particular, Figure 2.3 a) shows the changes of the portfolio composition of the "NIG" strategy. As we see these percentage proportions are not concentrated on their maximum value 0.2. Therefore, we witness a good diversification for each optimal portfolio. Figure 2.3 b) points out the percentages φ_k (k=1,...,548) of the portfolio that change every 5 trading days by the formula:

$$\varphi_k = \sum_{i=1}^{500} \left| x_{M,i}^{(k)} - x_{M,i}^{(k-1)} \right|, \tag{2.6}$$

where the vector $x_M^{(0)} = [x_{M,1}^{(0)}, ..., x_{M,500}^{(0)}]'$ is a vector with all the components equal to zero. This turnover measure belongs to the interval [0,2], where a value of 0 means that the portfolio composition has never been

Figure 2.3: Turnover and Diversification



changed during the period $[t_{k-1}, t_k]$, while a value of 2 corresponds to the case in which the portfolio would be re-structured completely. Since the values are in average around to 0.87, we can say that the "NIG on preselected" strategy (as the other optimization strategies) presents a significant turnover in the portfolio composition. Clearly, the relatively high turnover adversely impacts portfolio return because of transaction costs. For this reason, we have computed the ex-post wealth and the total return for all the strategies considering proportional transaction costs. The last Figure 2.3 c) shows (for each optimization) the number of: assets with positive proportions; the new assets in the optimal composition; and the assets leaving the optimal composition. This figure confirms the previous results that the NIG strategy almost uses more than 10 assets in its optimization. Table 2.2 reports some descriptive statistics of the ex-post log-returns obtained maximizing the performance ratios under different distributional assumptions. In particular, we report the mean, the standard deviation, the skewness, the kurtosis, the average value at risk $AVaR_{5\%}$

(where $AVaR_\alpha(X) = \frac{-1}{\alpha}\int_0^\alpha F_X^{-1}(u)du$) and the Sharpe ratio of the ex-post log-returns .

Table 2.2: Descriptive Statistics on the ex-post logreturns

Ditribution	Mean	St.Dev	Skewn	Kurtosis	Final Wealth	AVaR 5%
Sharpe	0.00050	0.0141	-0.141	8.189	3.896	0.002016
Brownian Motion	0.00064	0.0143	-0.168	7.815	5.589	0.002192
Meixner	0.00072	0.0175	-0.185	6.082	7.072	0.002625
NIG	0.00088	0.0182	-0.301	5.761	10.902	0.002823
Stable	0.00082	0.0171	-0.334	8.286	9.280	0.002686
VG	0.00075	0.0176	-0.313	7.504	7.728	0.002671
S&P500	0.00024	0.0128	-0.307	13.532	1.925	0.001632

First, we observe that the ordering of the mean and the risk $AVaR_{5\%}$ of the log-returns is the same we have in terms of the ex-post final wealth (that is, from the biggest to the smallest are: NIG, $\alpha-$Stable, VG, Meixner, Brownian Motion, Sharpe, and the index S&P500). Moreover, the $\alpha-$stable Paretian hypothesis shows a lower variability (standard deviation) than the other non-Gaussian Lévy distributional assumptions, probably because it is able to capture the asymptotic behavior of the log-return series. However, from the Sharpe Ratio of the ex-post log-returns we deduce that the Brownian Motion strategy presents higher reward-risk performance of Variance Gamma and Meixner strategies. Secondly, we observe that all the ex-post log-return series are asymmetric and leptokurtic, since they present a small negative skewness and a kurtosis significantly higher than the Gaussian one.

In Table 2.3 we report the ex-post wealth obtained with the used strategies in three different equidistant times: 01-Aug-06, 03-Mar-10, 03-Oct-2013. This table suggest that:

Table 2.3: Ex-post wealth of timing portfolio strategies

Distribution	Wealth 01-Aug-06	Wealth 03-Mar-10	Final Wealth 03-Oct-13
Sharpe	2.58	3.17	3.90
Brownian Motion	2.77	4.25	5.59
Meixner	3.89	5.81	7.07
NIG	4.88	7.45	10.90
Stable	3.85	6.40	9.28
VG	3.37	5.13	7.73
S&P500	1.44	1.27	1.93

- NIG and Meixner strategies present the best performance in the period before the crisis (Jan, 2003- Aug, 2006);
- During the sub-prime crisis (Aug, 2006 - Mar, 2010) the Stable paretian and the Brownian Motion strategies present the best performance;
- During the credit risk crisis (Mar, 2010 - Oct, 2013) all the strategies present similar performance and are not able to outperform the S & P 500 index;
- During the crises (Aug, 2006 - Oct, 2013) the Stable, NIG, VG and Sharpe strategies maintain a constant behavior even if much different from the one before the crisis (Jan, 2003 - Aug, 2006). Brownian Motion strategy outperforms the Meixner one during both crises (Aug, 2006 - Oct, 2013), but present much lower final wealth before the crisis (Jan, 2003- Aug, 2006). This last observation justifies the better ex-post Sharpe ratio of the Brownian Motion strategy with respect to the Meixner one observed in Table 2.3.

Finally, with this ex-post empirical analysis we demonstrate that:

- timing portfolio strategies based on the average of proper first passage times can be practically used in portfolio selection problems;
- timing portfolio strategies present better performance than the U.S. market stock index (S&P 500) and than the classic strategy based on the maximization of the Sharpe ratio;
- the better distributional approximation could have an important impact in the optimal portfolio choices and, in particular, the NIG process and the $\alpha-$Stable process seem to better capture the log-return behaviour in the U.S. stock market of the decade 2003-2013.

In conclusion, in this chapter we described an algorithm to approximate the distributions of first passage times of some Lévy processes and proposes a practical empirical analysis to evaluate, under different distributional hypotheses, the impact of new timing portfolio strategies in portfolio theory. In particular, we assume that investors minimize the time to reach a high level of wealth and maximize the time to reach a low level of wealth. Therefore, we describe how to approximate the distribution of some stopping times using the underlying parametric transition matrix of a given Lévy process. Finally, we propose an ex post empirical comparison among timing portfolio strategies based on different distributional hypotheses.

Chapter 3

Timing portfolio selection with non-Lévy processes

In this chapter we deal with non-Lévy processes. We consider the Markov Regime switching model and the log-Student-t model, which are not based on Lévy processes. In particular, we suppose that the daily log-returns distributions are approximated by these distributions. Then, we assume a Markovian evolution of the portfolio processes and we estimate some proper stopping times.
We extend the application of Chapter 2, proposing an empirical comparison based on some timing portfolio strategies on the US market. The main objective of this empirical experiment is to show that the techniques previously presented can be applied in practical portfolio problems not only for Lévy processes but for other processes as well (when we assume a Markov evolution of the process). Moreover, the ex-post comparison among different distributional assumptions could give some indications on the best process to use in timing portfolio strategies. The chapter is organized as follows: In the first section the Markov approximation of the Regime Switching model and the log-Student-t model is given. Section 2 discusses an ex-post comparison among timing portfolio strategies using these non-Lévy processes.

3.1 Financial application with alternative models

Since the publication of the Black and Scholes (1973) [5] option pricing model, a vast amount of literature has been dedicated to describe the theory and practice of pricing. Important assumptions of the Black-Scholes model are that the underlying asset distribution is log-normal and that the volatility is a fixed constant.
However, empirical evidence suggests that Black-Scholes model fails to react the stochastic variability in the market parameters. In particular the findings of fat tails and volatility clustering are contrary evidence.
Empirical distributions of asset and index returns are typically 'fat tailed' with large asset price movements having higher probability than predicted by normality. Alternative models, based on non-normal, more realistic distributional assumptions, have been developed.
In this section, we will focus in two different classes of models, to better understand asset returns behavior.
For this reason in this chapter, we use the log-Student-t model and the Markov Regime Switching model.

3.1.1 The Markov approximation of Regime Switching model

Consider n risky assets with log-returns $r_{k+1} = [r_{1,k+1}, ..., r_{n,k+1}]'$. We adopt the standard definition of log-return between time t and time t+1 of asset i, as $r_{i,t+1} = \ln\left(\frac{S_{i,t+1}+d_{i,[t,t+1]}}{S_{i,t}}\right)$ where $S_{i,t}$ is the price of the i-th asset at time t and $d_{i,[t,t+1]}$ is the total amount of cash dividends paid by the asset between time t and time t+1.

The empirical evidence suggests that the log-returns exhibits fat tails and skewness, behavior that deviates from normality. Therefore, models that accurately fit return distributions are needed. Recently, attention has been placed on regime switching models. In such models, the stochastic

process remains in one regime for a random amount of time before switching over into a new regime.
Markov Regime Switching model assumes there are two 'regimes' or states with different mean and volatility levels. This specification can be interpreted as a mixture distribution with dynamics generated by a Markov chain.
In this model the Gaussian distribution is generalized by introducing two regimes with different moments. Leptocurtosis is obtained here because the variance in the two regimes differs.
So, the mixture distribution generates the leptocurtosis and the Markov chain is responsible for the nonlinear dynamics. It has been recognized that the dynamics of asset return can be adequately described by the regime switching model.

Suppose, the dynamics of the log-returns are modeled by a Markov Regime Switching process in which prices remain in one regime for a random amount of time before switching over into a new regime.
Let us simplify their tractability by considering a variable y_t which depends on random shocks ϵ_t and some regime process, S_t. Regimes are generally modeled through a discrete variable, $S_t \in \{0, 1, \ldots, k\}$, tracking the particular regime inhabited by the process at a given point in time. Although regimes could affect the entire distribution, they are often limited to affect the intercept, μ_{S_t} and the volatility σ_{S_t}, of the process. Let us recall the MRS model illustrated in section 1.3, chapter 1.

$$y_t = \mu_{S_t} + \sigma_{S_t}\epsilon_t \qquad \epsilon_t \sim \text{iid}(0, 1) \tag{3.1}$$

To complete the model, the process governing the dynamics of the underlying regime, S_t, needs to be specified. We have supposed that the dynamics of the underlying switch between two regimes, but we haven't said how this switching between states happens.
The crucial point in the structure of a Markov Regime Switching model is that,the switching of states is stochastic process itself.
We assume that the transition probabilities between states is governed by

the homogeneous first-order Markov chain $Pr(S_t = j|S_{t-1} = i) = p_{ij}$. For example, in the common case with two regimes, $S_t = 1$ or 2

$$Pr(S_t = 1|S_{t-1} = 1) = p_{11} \quad \text{and} \quad Pr(S_t = 2|S_{t-1} = 2) = p_{22}. \tag{3.2}$$

It is often convenient to collect the transition probabilities in an (2×2) matrix P known as the transition matrix:

$$P = \begin{bmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{bmatrix} \tag{3.3}$$

In general, the row j, column i element of the P is the transition probability p_{ij}; for example, the row 2, column 1 element gives the probability that state 1 will be followed by state 2. These transition probabilities are restricted so that $p_{11} + p_{12} = p_{22} + p_{21} = 1$.

A key issue in regime switching model is whether the same regimes repeat over time, as in the case of repeated recession and expansion periods, or if new regimes always differ from previous ones. If'history repeats' and the underlying regimes do not change, all regimes will occur at some time. With only two regimes this will happen if $p_{ii} < 1, i = 1, 2$. Models with recurring regimes have been used to characterize bull and bear markets, calm versus turbulent markets, and recession and expansion periods.

In practice, if the process is not irreducible and not all states are visited with non-zero probability in the steady state, then the moment analysis can simply be conducted on the subset of states occuring with non-zero stationary probability. The higher is p_{ii} the longer the process is expected to remain in the state i. For this reason we shall refer to p_{ii} as measuring the 'persistence' of the mixing of the underlying state densities. This persistence parameters are very important in determining the higher order moments of the Markov switching process.

Estimation techniques

Different econometric methods can be used to estimate regime switching models. Maximum likelihood and EM algorithms are outlined by Hamilton (1988, 1989)[29], [30] and Gray(1996)[27]. The maximum likelihood algorithm involves a Bayesian updating procedure which infers the probability of being in a regime given all available information up until that time, $Pr(S_t|Y_t)$, where Y_t is the information set at time t.

An alternative to the maximum likelihood estimation is the method of moment. We will concentrate on this.
Let $\pi = (\pi_1, \ldots, \pi_k)'$ be the k-vector of steady state probabilities that solve the system of equations $P'\pi = \pi$. These probabilities can be computed as the eigenvector (scaled so that its elements sum to one) associated with the unit eigenvalue of P′. π is the vector of unconditional probabilities applying to the k states (see equation in chapter 1). The following proposition provides the moments of the basic Markov switching model.

Proposition 3.1.1. *Suppose the stationary Markov Switching process 3.1 e 3.2 started from its steady state characterized by the set of unconditional probabilities* (π)*. Then the centered moments of the process are given by*

$$E[(y_t - \mu)^n] = \sum_{i=1}^{k} \pi_i \sum_{j=0}^{n} {}_nC_j \sigma_i^j E[\epsilon_t^j](\mu_i - \mu)^{n-j} \tag{3.4}$$

where ${}_nC_j = \frac{n!}{(n-j)!j!}$. When ϵ_t follows a normal distribution we have

$$E[(y_t - \mu)^n] = \sum_{i=1}^{k} \pi_i \sum_{j=0}^{n} {}_nC_j \sigma_i^j b_j (\mu_i - \mu)^{n-j}, \tag{3.5}$$

where

$$b_j = \prod_{h=1}^{j/2} (2h-1), \quad \text{provided j is even and} \tag{3.6}$$

$$b_j = 0, \quad \text{otherwise.} \tag{3.7}$$

Proof. See Timmerman (2000). □

An attractive feature of regime switching model is that they capture central statistical features of asset returns. To illustrate this, consider the model given above, where the (unconditional) probability that $S_t = 1$ is π_1 and $S_t = 2$ with probability $1 - \pi_1$.
The mixture of the two normals produces pronounced negative skewness. Timmerman(2000) derives the moments of a general regime-switching process with constant transition probabilities. As a special case, it can be shown that the first four central moments of the process in equation 3.1 are given by:

$E[y_t] = \pi_1\mu_1 + (1-\pi_1)\mu_2$
$\text{Var}[y_t] = \pi_1(1-\pi_1)(\mu_1-\mu_2)^2 + \pi_1\sigma_1^2 + (1-\pi_1)\sigma_2^2$
$\text{Skew}[y_t] = \pi_1(1-\pi_1)(\mu_1-\mu_2)[(1-2\pi_1)(\mu_1-\mu_2)^2 + 3(\sigma_1^2-\sigma_2^2)]$
$\text{Kurt}[y_t] = \pi_1(1-\pi_1)(\mu_1-\mu_2)^2[((1-\pi_1)^3+\pi_1^3)(\mu_1-\mu_2)^2 + 6\pi_1\sigma_2^2 + 6(1-\pi_1)\sigma_1^2]$
$+ 3\pi_1\sigma_1^4 + 3(1-\pi_1)\sigma_2^4.$

Differences in means across regimes, $\mu_1 - \mu_2$, enter the higher moments such as variance, skew and kurtosis. In particular, the variance is not simply the average of the two variances across the two regimes: the difference in means also imparts an effect because the switch to a new regimes contributes to volatility. Intuitively, the possibility of changing to a new regime with a different mean introduce an extra source of risk. Skew only arises in this model if the means differ across two regimes ($\mu_1 \neq \mu_2$).
Importantly, differences in means in addition to differences in variances can generate persistence in levels as well as squared values-akin to volatility persistence observed in many return series.
Again differences in means play an important role in generating autocorrelation in first moments-without such differences, the autocorrelation will be zero. In contrast, volatility persistence can be induced either by differences in means or by differences in variances across regimes. In both cases, the persistence tend to be greater, stronger the combined persistence of the regimes, as measured by $p_{11} + p_{22} - 1$.

We use the method of moments to approximate the Markov Regime Switching model.

3.1.2 The log-Student-t Model

In this section, an alternative approach to model asset prices is analyzed. This approach is based on the subordination principle introduced by Bochner (1955) [**?**]. The option pricing theory based on a subordinated return was developed by Hurst et al (1997) [34].

The general version of the subordinated asset price model is given by:

$$S_{t+s} = S_t exp\{\mu s + \sigma[W(T(t+s)) - W(T(t))]\}, \quad t, s \geq 0, \tag{3.8}$$

where $W(t)$ is a Wiener process, which is subordinated to the intrinsic time process, T(t) (see Rachev and Mittnik (2000) [59], chapter 13). Taking the natural logarithm, the process can be written in the equivalent form:

$$\Delta L_{t,s} = L_{t+s} - L_t = \mu s + \sigma[W(T(t+s)) - W(t))]. \tag{3.9}$$

By using a stochastic time scale, the subordinated model allows for time periods of high and low volatility (measured in the physical time scale), depending on the specification of the directing process $\{T(t), t \geq 0\}$. For example, one could specify the intrinsic time process such that the intrinsic time process is equal to the physical time, on average, i.e $E[\Delta T_{t,s}] = s$. Calm periods of low volatility with a slow information flow and low trading volume in the markets would occur if $\Delta T_{t,s} < s$, i.e, intrinsic time passes more slowly than physical time. On the other hand, periods where unexpected information arrives with a high rate combined with high volume and volatility are compatibile with $\Delta T_{t,s} > s$, i.e, operational time passes more quickly than physical time.

Subordinated processes of the form in 3.8 can therefore interpreted as stochastic volatility processes. We assume that $\{T(t), t > 0\}$ is a process with stationary independent increments.

The probabilistic features of the returns can be studied by combining the

constant volatility parameter σ with the subordinated process $W(T(t))$ to get the new process:

$$Z = \tilde{W}(T(t)) = \sigma W(T(t)), \tag{3.10}$$

where $\tilde{W}(t)$ is a Wiener process having stationary independent increments

$$\Delta\tilde{W}_{t,s} = \tilde{W}(t+s) - \tilde{W}(t) \sim N(0, \sigma^2 s), \quad s, t > 0. \tag{3.11}$$

In this section to describe the log-Student t model we follow Praetz(1972) [58], Blattberg and Gonedes (1974) [6].
The t distribution belongs to a larger class of probability distribution the so-called 'scale mixture of normals'. These distributions are constructed by multiplying a normal random variable Y with zero mean with an independent and positively distributed random variable T. In log-Student t model the unit time increments are governed by an 'inverse' Chi-square distribution, i.e.

$$\Delta T_{t,1} \sim \frac{\nu}{\chi^2_\nu}, \quad \nu > 0,$$

where χ^2_ν denotes a Chi-square distribution with ν degrees of freedom. The probability density function of $\Delta T_{t,1}$ is given by

$$f_{T,1}(x) = \frac{\tilde{\nu}^{\tilde{\nu}}}{\Gamma(\tilde{\nu})} x^{-\tilde{\nu}-1} exp(-\frac{\tilde{\nu}}{x}), \quad \tilde{\nu} = \frac{\nu}{2}, \quad x > 0,$$

where $\Gamma(u) = \int_0^\infty x^{u-1}e^{-x}dx$ denotes the Gamma function. The expected value of a unit time increment is then:

$$\mu_{T,1} = \frac{\nu}{\nu - 2} \quad \text{for} \quad nu > 2$$

If $\nu \leq 2$, the mean does not exist.

The probability density function of the increments $\Delta Z_{t,1}$ of Z has the following form:

$$f_{Z,1} = \frac{1}{\sigma} f_\nu(\frac{x}{\sigma}), \quad x \in \mathbf{R}$$

Here, $f_\nu(u)$ denotes the probability density function of a Student -t distributed random variable with ν degrees of freedom:

$$f_\nu(u) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{(1+\frac{u^2}{\nu})^{-\frac{\nu+1}{2}}}{\sqrt{\nu\pi}}, \quad u \in \mathbf{R} \tag{3.12}$$

Hence, the unit increment $\Delta Z_{t,1}$ follows a scaled t-distribution, i.e.

$$\Delta Z_{t,1} \sim \sigma t_\nu, \quad \nu > 0,$$

where t_ν denotes a t-distributed random variable with ν degrees of freedom. The mean and the variance of the increments of Z are therefore $\mu_{Z,1} = 0$ for $\nu > 1$ and $\sigma^2_{Z,1} = \sigma^2 \frac{\nu}{\nu-2}$ for $\nu > 2$. As the degrees of freedom increase $\nu \to \infty$, the inner time process converges asymptotically to the deterministic physical time process, so that thus model nests the Gaussian model as a special case. In particular, for $\nu \to \infty$ the increments $\Delta Z_{t,1}$ converge to a normal distribution with variance σ^2.

In the empirical analysis, we use the maximum-likelihood method to estimate the degrees of freedom ν at each portfolio.

t based estimator

Suppose variable Z has a t distribution with ν degrees of freedom. The density function is given by 3.12. Let u_i $i = 1, 2, ..., n$ be the observed values in a sample of size n. The log likelihood is

$$L = -\frac{\nu+1}{2} \sum_{i=1}^{n} \log(1+\frac{u_i^2}{\nu}) - n\log(\frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \cdot \frac{1}{\sqrt{\nu\pi}}) \tag{3.13}$$

Since, the degrees of freedom will be estimated for all the itearations of the optimization problem, it is important to have an estimator in reasonable time. Using maximum likelihood estimates of ν degrees of freedom we easly reach the solution with the function $fminunc$ in Matlab.

3.2 An empirical comparison among timing portfolio strategies applied to preselected assets

Following the application in Chapter 2, we propose an empirical comparison based on some timing portfolio strategies on the US market. In this section we consider a portfolio selection procedure that is divided in two phases. In the first phase we preselect some assets that satisfy some performance criteria. Moreover, we define the 'good' assets considering performance ratios based on a Markovian evolution of the assets (see Angelelli and Ortobelli (2009) [1], In the second phase we assess some portfolio selection strategies based on the optimization of functionals that consider the forecasted first passage times of the wealth before the investor's horizon T. In particular, we compare the ex-post wealth we obtain maximizing either the classic Sharpe ratio or a performance measure obtained using the average of proper first passage times.
The main objective of this empirical experiment is to show that the techniques previously presented can be applied in practical portfolio problems not only for the Lévy processes but for other processes as well. Moreover, the ex-post comparison among different distributional assumptions could give some indications on the best process to use in timing portfolio strategies.

3.2.1 Pre-selection criteria

In this section we deal with the pre-selection criteria in the portfolio problems. The preselection has a very important impact on the portfolio choices (see Ortobelli et al (2011) [53]). Further, since the number of observations should increase proportionally with the number of assets (see Papp et al. (2008) [57], Kondor et al.(2007)[40]), it is necessary to find the right trade-off between a statistical approximation of the historical series depending only on a few parameters and the number of historical observations. Therefore, we suggest to reduce the dimensionality of this large scale portfolio problem preselecting the most relevant stocks (150

assets) according to different criterio. (see, among others, Ortobelli et al (2011)[53]). This procedure is applied at any recalibration time and implies that there is a very high turnover.

In particular at each recalibration time we preselected no more than 150 assets among m ones for each optimization problem. The results obtained with different pre-selection criteria are merged in order to identify the assets with common characteristics appealing for investors.
As in Ortobelli et al (2011) [53], we suggest to select those asset that satisfy some optimality criteria, considering several properties of the wealth behaviour, either that Markovian or asymptotic. Another factor we should consider as preselecting criteria is the timing (i.e., we consider that investors want to maximize the time the wealth arrive to a given lower bound or to minimize the time the wealth arrive to a given upper bound.) The association with the market stochastic bounds is important as well (for further details see Ortobelli and Tichy (2009, 2010) [69], [55]. Since the investors want to increase their wealth, their portfolios are concordant as much as possible with the upper stochastic market bound. Similarly, they want to reduce their losses and their portfolio are concordant as much as possible with the lower stochastic market bound.
Therefore we take account of these investors preferences to select a resticted number of assets.

Let us recall that the classic static portfolio selection problem when no short sales are allowed, can be represented as the maximization of a functional $f : (\Omega, \Im, P) \longmapsto \mathbb{R}$ applied to the random portfolio of gross returns $z_{(x)}$, subject to the portfolio weights belonging to the $(n-1)$-dimensional simplex $S = \{x \in \mathbb{R}^n | \sum_{i=1}^n x_i = 1; x_i \geq 0\}$, i.e.,

$$\max_{x \in S} f(z_{(x)}).$$

The approach to the future wealth could be generally considered static or dynamic:

- i) with the static approach we assumes that historical observations are equally distributed. Moreover the investor maximize a functional of the future wealth that is independent from his temporal horizon [0,T], i. e he solves problems of the type:

$$max_{x \in S} f(z_{(x),t+1})$$

- ii)with the dynamic approach we consider that the gross returns follow a Markov chain. Therefore the investor maximizes a functional of the future wealth, that should be determined under this hypothesis, i.e. he solves problems of the type:

$$max_{x \in S} f(W_T(z_{(x)}))$$

where $W_T(z_{(x)})$ is the predicted wealth obtained after T working days investing in the portfolio $z_{(x)} = x'z$.

Starting from the static approach (i) we represent the classic *myopic approach* that does not use the time evolution of the wealth process. In the dynamic context (ii) we rather consider all admissible wealth Markovian processes $W(z_{(x)}) = (W(z_{(x)})_{t \geq 0})$. They are defined on the filtered probability space that depends by an initial portfolio of wights $x \in S$.

In the preselection criteria of our work we consider both the approaches. We select some assets assuming approach (i) and other assets assuming approach (ii). Then the union of these assets is used for the portfolio selection.
In order to select some desirable assets, both the selection approaches are based on following procedure. First, we select a desirable ordering criterion (we will discuss in details some ordering criteria in the following section). Then, the assets are ordered by this ordering criterion. Lastly, we select the best assets satisfying the criteria of the dynamic approach and the best assets satisfying the criteria of the static approach.

In order to deal with the preselection methodology we first argue how to approximate the Markovian evolution of the wealth, then we discuss

three ordering criteria with the static approach and five criteria with the dynamic approach.

Some consequences of the Markovian hypothesis

When we assume that the portfolios follow a Markov chain we can distinguish different types of possible strategies. According to the definition given in Angelelli and Ortobelli (2009)[1] we call *OA expected utility* the above functional $E(u(W_T(z_{(x)})))$ when it is computed under the assumption that the gross return of each portfolio follows a Markov chain with N states. The OA expected utility is given by:

$$E(u(W_T(z_{(x)}))) = u(\hat{W}_T(z_{(x)})) \cdot Q^{(T)} \cdot \mathbf{1}_N = u(\hat{W}(z_{(x)})) \cdot p^{(T)} \tag{3.14}$$

where $\hat{W}_T(z_{(x)}) = [w_{(x)}^{(1,T)}, \dots, w_{(x)}^{((N-1)T+1,T)}]$ is the (N-1)T+1 dimensional vector of the final wealth and $u(\hat{W}_T(z_{(x)})) = [u(w_{(x)}^{(1,T)}), \dots, u(w_{(x)}^{((N-1)T+1,T)})]$ is the utility valued on the final wealth. Formula 3.14 is a logical consequence of the methodology to describe the Markovian tree. As a matter of fact, $p^{(T)} = Q^{(T)} \cdot \mathbf{1}_N$ gives the distribution of the final wealth.

Pre-selection criteria with the static approach

With the preselection criteria of the static approach we only account the consistency with investors' preference, and the association with the market stochastic bounds. In particular, we order the assests considering the following three criteria:

1. the wealth obtained in the last 120 working days, i.e., the assets are ordered with respect to the ration

$$\frac{P}{P_{t-120}} \tag{3.15}$$

where P_t and P_{t-120} are, respectively, the adjusted prices at time t and $t - 120$(where 120 working days are about six month of data)

2. the Sharpe ratio:

$$\frac{E(z_i - z_f)}{St.dev(z_i)} \tag{3.16}$$

where the mean and the standard deviation of the i-th asset are approximated by the empirical mean and the standard deviation of the last 120 working days (we assume that the riskless is null, that is $z_f = 1$):

$$E(z_i) \cong \frac{1}{120} \sum_{t=1}^{120} z_{i,t}$$

and

$$St.dev(z_i) \cong \left(\frac{1}{120} \sum_{t=1}^{120} (z_{i,t} - E(z_i))^2 \right)^{0.5}.$$

3. the Gini performance ratio:

$$\frac{\gamma(z_i, \max_{i \leq m} z_i)}{1.1 + \gamma(z_i, \min_{i \leq m} z_i)} \tag{3.17}$$

This ratio is based on the Gini γ concordance measure; the sample estimation of this measure is given by:

$$\gamma(X, Y) = \frac{1}{\left[\frac{n^2}{2}\right]} \sum_{i=1}^{n} |p_i - q_i - n - 1| - |p_1 - q_1| \tag{3.18}$$

where $n = 120$ is the number of observations, p_i and q_i are, respectively, the ranks of the random variables X and Y.

The first two criteria are consistent with choices of non-satiable investors. While maximizing the Gini performance ratio, we maximize the concordance between the portfolio and the upper stochastic bound and we minimize the concordance between the portfolio and the lower stochastic bound. Using three criteria, we preselect 50 desirable assets for non satiable investors. In particular, the choice includes the first 50 assets, among m, with the best common performance measure (Sharpe ratio, wealth obtained in the last six months, and Gini performance ratio).

Pre-selection criteria with the dynamic approach

With the preselection criteria of the dynamic approach we account the consistency with investors' preferences, the timing of the choices, the association with the market stochastic bounds, the Markovian and asymptotic behavior of the wealth. In particular, we assume that each portfolio of returns follows a Markov chain. Then, we preselect the assets considering these five ordering criteria:

1. The expected power utility

$$E\left(u(W_T(z_i))\right) \tag{3.19}$$

this is the predicted wealth $W_T(z_i)$ obtained after $T = 20$ working days investing in the i-th asset (for any $i = 1, \ldots, n$). In formula 3.19 $u(W) = \frac{W^g}{g}$ with the $g = 0.9$.

2. The Sharpe ratio:

$$\frac{E(W_T(z_i)) - 1}{St.dev(W_T(z_i))} \tag{3.20}$$

where $W_T(z_i)$ is the predicted wealth obtained afte $T = 20$ working days investing in the i-th asset (for any $i = 1, \ldots, n$).

3. The Pearson performance ration, based on the Pearson linear correlation (τ):

$$\frac{\tau\left(W_T(z_i), W_T(max_{i\leq m} z_i)\right)}{1.1 + \tau\left(W_T(z_i), W_T(min_{i\leq m} z_i)\right)} \tag{3.21}$$

4. The timing ratio:

$$\frac{E(\pi_1)}{E(\pi_2)} \tag{3.22}$$

where π_1 and π_2 are two stopping times of the filtration defined as:

$$\pi_1(z_i) = min(T, inf\{k \in [0, T] | W_k(z_i) \leq 0.98\})$$

and

$$\pi_2(z_i) = min(T, inf\{k \in [0, T] | W_k(z_i) \geq 1.2\})$$

These stopping times give the first time (belonging to [0,T] the wealth produced by the i-th asset reaches, respectively, the bounds 0.98 and 1.2.

5. The OA-stable ratio:

$$\frac{\delta_{W_T(z_i)}}{ETL_\beta(W_T(z_i) - E(W_T(z_i)))} \tag{3.23}$$

where $\delta_{W_T(z_i)}$ is the location parameter of the best quantile, stable Paretian approximation of the wealth $W_T(z_i)$, the denominator ETL_β (expected tail loss or average value at risk for $\beta = 0.05$) is computed under the distributional assumption using the Stoyanov et al algorithm (2006) [68].

Therefore we choose the first 100 assets (among $m = 500$) with the highest common performance measures of the predicted wealth: the expected power utility, Sharpe, association type ratio, timing ratio and stable ratio. Together with the 50 assets selected with the static approach, the total of the pre-selected desirable assets is between 100 and 150. As a matter of fact, some assets could be selected with both approaches (static and dynamic).

3.2.2 An ex- post comparison among timing portfolio strategies applied to preselected assets

In our analysis, we use the historical observations of the components of the S&P500[1] (which were components at time 22/07/2014)) from January 1, 2000 to July 22, 2014 for a total of 3662 daily observations.
We considered all the equities active during the last six months (120 working days) of the market. The data source is DATASTREAM. We recalibrate the portfolio every month (T=20 working days) to be coherent with dynamic preselection analysis. At each recalibratin time there are about 500 active assets that can be selected for the portfolio. For the evaluation

[1]The data source is DATASTREAM.

we consider only the observations of the last six months (120 working days), because they are supposed to have the highest impact in the future choices.

Typically, the functional $f(.)$ is a performance measure or an utility functional. In both cases the functional $f(.)$ should be isotonic with a particular ordering of preference $\succeq$, that is, if X is preferred to Y ($X \succeq Y$), then $f(X) \geq f(Y)$. The benchmark performance measure used in financial choices is the Sharpe ratio (see, among others, [66]):

$$f(z_{(x)}) = SR(z_{(x)}) = \frac{E(z_{(x)} - z_f)}{\sigma_{z_{(x)}}}, \tag{3.24}$$

which evaluate the excess return with respect to the riskless gross return z_f for unity of risk where the risk is the standard deviation $\sigma_{z_{(x)}}$. In Chapter 2, we described functionals that consider the forecasted first passage times of the wealth before the investor's temporal horizon T. Since investors want to maximize the first time they lose wealth and minimize the first time their wealth increases enough, we suggest to optimize the average of two first passage times under different distributional assumptions of the wealth process. In this empirical analysis we propose an alternative experiment where we use the historical observations of the components of the S&P 500. In particular, under different distributional assumptions given in Section 3.1 , we suggest to maximize the following timing portfolio performance:

$$f(\tau_d, \tau_u, T) = \frac{E(\tau_d I_{[\tau_d < T]} + 1000 U_{[\tau_d \geq T]})}{E(\tau_u I_{[\tau_u < T]} + 1000 I_{[\tau_u \geq T]})}, \tag{3.25}$$

where $I_{[\omega \in B]} = \begin{cases} 1 & \text{if} \quad \omega \in B \\ 0 & \text{otherwise} \end{cases}$. In performance measure (3.25) we penalize[2] the case the first passage time τ_u overcome the temporal horizon T (i.e., $\tau_u \geq T$) while we reward the possibility that the first passage time τ_d overcome the temporal horizon T (i.e., $\tau_d \geq T$).

[2]Optimizing the functional (3.25) among the components of the S&P 500, we could observe that the optimal portfolio components do not change using penalty values 1000, 10 000 or 100 000 in formula (3.25). For this reason we use and fix the penalty value 1000.

In all the empirical comparison, we optimize each performance measure monthly (using 20 days) using six month of daily historical observations (120 working days) to compute the performance measures we have to optimize. Thus, at any optimization time, every 20 trading days, we use a moving window of 120 working days which are used in the optimization process. The principal parameters used in the optimization process are:

1. the investors have a temporal horizon T=20 trading days,

2. the performance (3.25) is optimized for the processes in Section 2 (Markov regime switching model and log-Student t);

3. Markov chains have N=9 states, so the final wealth W_{20} presents 161 nodes in the Markov tree. On the computational complexity of Markovian portfolio problems see [1].);

4. the initial wealth W_0 is equal to 1 at the date December 31, 1999;

5. two significant constraints were introduced in the optimization function: a) the maximum share of the portfolio that can be invested in a single title is 20% ; b) short sales are not allowed, in other words, the percentage wight of each security in the portfolio can't be negative $(i.ex_i \in [0, 0.2])$.
   The first constraints aims to achieve a well-diversified portfolio and not overly concentrated; while the second constraints excludes the possibility of short selling, since the securities sale not owned directly is a technique not easily implementable by a private investor.

Portfolio optimization leads to different results depending on the adopted distribution for the process. For each strategy, we have to compute the optimal portfolio composition exactly 183 times and at the k-th optimization three main steps are performed to compute the ex-post final wealth:

1. Step 1. Preselect the first 150 assets as discussed in the previos section.

2. Step 2. We implement an optimization function on 150 pre-selected assets. The optimization aim to identify which of these assets, and what percentage , should compose the portfolio that maximizes the performance measure. Therefore, we determine the optimal portfolio $x_M^{(k)}$ that maximizes a performance ratio $\rho(x^{(k)})$ (either the Sharpe ratio (3.24) or the timing portfolio performance (3.25)) associated to the strategy, i.e. the solution of the following optimization problem:

$$\max_{x^{(k)}} \rho(x^{(k)})$$

$$s.t.$$

$$\sum_{i=1}^{n} x_i^{(k)} = 1; \;\; 0 \leq x_i^{(k)} \leq 0.2; \;\; i = 1, ..., n$$

Clearly, the computational complexity of the problem with portfolio performance (3.25) is much more higher than optimizing the Sharpe ratio, because in the first case the problem present more local optima. Therefore, we suggest to use the function *patternsearch* of Matlab 2013 with starting point the optimal solution of (3.25) obtained with the heuristic (for global optimization) proposed by [1].

3. Step 3. Compute the ex-post final wealth given by:

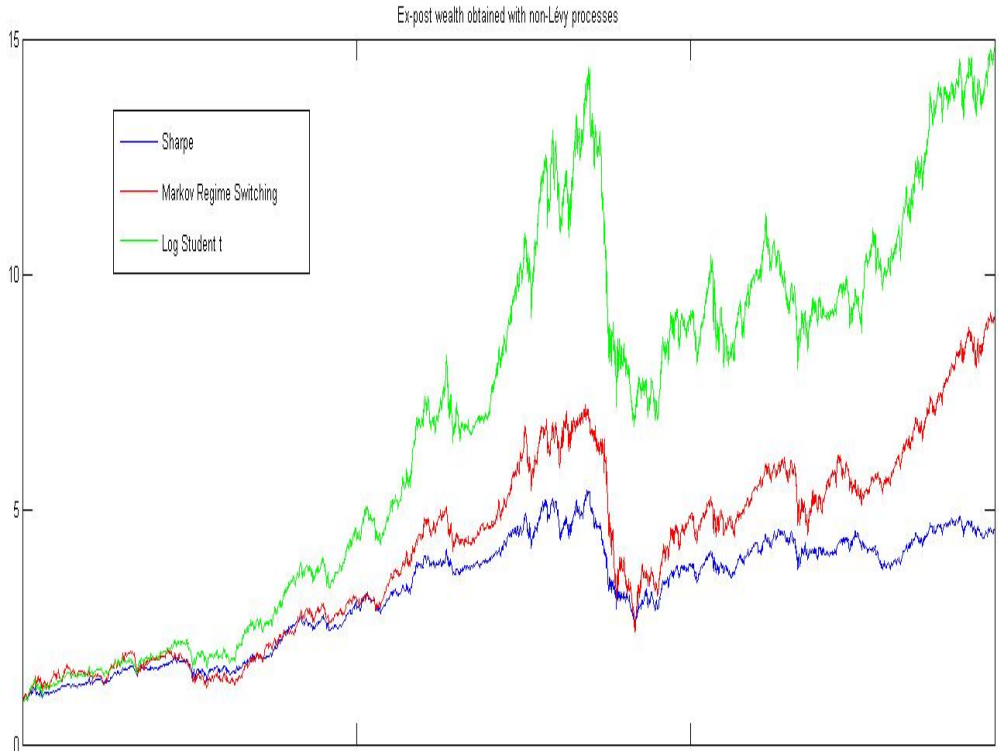
$$W_{t_{k+1}} = (W_{t_k} - tc_{t_k})(x_M^{(k)'}) z_{(t_k+1)}^{(\text{ex post})},$$

where tc_{t_k} are the proportional transaction costs we get changing the portfolio, $z_{(t_k+1)}^{(\text{ex post})}$ is the vector of observed gross returns between t_k and t_{k+1}.

Steps 1, 2, and 3 are repeated for the two performance ratios (3.24) and (3.25), and all the distributional assumptions until some observations are available. The results of this ex-post empirical analysis are reported in Figure 3.1, and Table 3.1

In Figure 3.1 we report the ex-post wealth obtained optimizing the Sharpe ratio and the timing portfolio strategies based on different distributional hypothesis. In particular we observe a growing trend of the wealth

Figure 3.1: Ex post final wealth obtained with non-Lévy processes



that involves the entire period of the analysis.
In Figure 3.1 both the timing portfolio strategies outperform the strategy based on the optimization of the Sharpe ratio. Moreover, we observe that the timing portfolio strategies based on log Student t model outperforms the one based on the Regime switching model. As a matter of fact, the Student t strategy allows to reach a final wealth equal to 15 times the initial one and the MRS strategy 9 times. The Sharpe strategy, instead leads to a final wealth equal to 4.5 times the initial one.
Graphically the "Student-t strategy" seems the most promising in terms of ex-post final wealth.

In Figure 3.1 we observe that all the strategies present a growing trend before the crisis (1999-2007) a decreasing trend during the sub-prime crisis (2008-2009) and a slightly increasing trend during the credit risk cri-

Table 3.1: Descriptive Statistics on the ex-post log-returns (annual basis)

Ditribution	Mean	St.Dev	Skewn	Kurtosis	Final Wealth
Sharpe	10.25%	19.94%	-0.3725	7.1458	4.553
Markov Switching	15%	28.18%	-0.2039	7.6900	9.067
Log-Student t	18.50%	25.44%	-0.3473	6.2008	14.793

sis (2009-2014). In particular, we observe that both investment strategies present similar ex-post wealth until 2004, then there is a high improvement of the Student-t strategy during the crisis.

In Table 3.1, first, we observe that the ordering of the mean of the log-returns is the same we have in terms of the ex-post final wealth (that is, from the biggest to the smallest are: Student t, Markov regime Switching and Sharpe). Moreover, the Student t hypothesis shows a lower variability (standard deviation) than the other distributional assumptions. Secondly, we observe that all the ex-post log-return of optimal strategies are asymmetric and leptokurtic, since they present a small negative skewness and a kurtosis significantly higher than the Gaussian one.

Finally, in this empirical analysis we demonstrate that timing portfolio strategies have a strong impact on the final wealth. Further, a better distributional approximation could have an important impact in the optimal portfolio choices and, in particular, the Student t distribution seems to better capture the log-return behavior in the U.S. stock market of the decade 1999-2014.

The fundamental contribution of this experiment consists in the computational accessible methodology to solve dynamic portfolio strategies without making too strong distributional assumptions. Moreover, the idea developed in the thesis can be extended to several other possible timing

portfolio strategies which can be imported by the option theory.
We believe this is the starting point for future discussions, analyses and empirical comparisons.

Bibliography

[1] E. Angelelli and S. Ortobelli. *Financial Hedging*, chapter American and European portfolio selection strategies: the Markovian approach, pages 119–152. Nova Science Publishers, 2009.

[2] E. Angelelli and S. Ortobelli. Maximum expected utility of markovia predicted wealth. In *Lecture notes in Computer Science*, pages 588–597. G.Allen et al, 2009.

[3] O.E. Barndorff Nielsen. Normal inverse gaussian distributions and the modeling of stock returns. Research report 300, Department of theoretical statistics, Aarhus University, 1995.

[4] S. Beveridge and C.R. Nelson. A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the business cycle. *Journal of Monetary Economics (1981) 151-174.*, 7:151–174, 1981.

[5] F. Black and M. Scholes. The pricing of options and corporate liabilities. *The Journal of Political Economy*, 81:637–654, 1973.

[6] R.C. Blattberg and N. J. Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *Journal of Business*, 47:244–280, 1974.

[7] J. Buffington and R Elliot. American options with regime-switching. *International Journal of Theoretical and Applied Finance*, 5:497–514, 2002.

[8] J. Cai. A markov model of switching-regime arch. *Journal of Business and Economic Statistics*, 12(3):309–316, July 1994.

[9] J. Y. Campbell and M. N. Gregory. Are output fluctuations transitory? *The Quarterly Journal of Economics*, 102:857–880, 1987.

[10] P. Christoffersen. Evaluating interval forecasts. *International Economic Review*, 39(4):841–862, 1998.

[11] R. Cont and R. Tankov. *Financial modelling with jump processes*. Chapman and Hall, 2004.

[12] J.C. Cox, S. A. Ross, and M Rubinstein. Option pricing: a simplified approach. *Journal of Financial Economics*, 7:229–263, 1979.

[13] G. D'Amico. Markov chain european option: Statistical estimation. Technical report, Università di Roma "Sapienza", 2003.

[14] G. D'Amico, J. Janssen, and R. Manca. European and american options:the semi-markov case. *Phys*, 388(15):3181–3194, 2009.

[15] G. D'Amico and F. Petroni. A semi-markov model for price returns. *Physica A: Statistical Mechanics and its Applications*, 391(20):4867–4876, 2012.

[16] G. D'Amico and F. Petroni. Weighted-indexed semi-markov models for modeling financial returns. *Journal of statistical mechanics: theory and experiment*, 07, 2012.

[17] J.C. Duan. The garch option pricing model. *Mathematical Finance*, 5:13–32, 1995.

[18] J.C. Duan, E. Dudley, G. Gauthier, and J.G. Simonato. Pricing discretely monitored barrier options by a markov chain. *Journal of Derivatives*, 10:9–23, 2003.

[19] J.C. Duan, I. Popova, and P. Ritchken. Option pricing under regime switching. *Quantitative Finance*, 2:1–17, 2002.

[20] J.C. Duan and J.G. Simonato. American option pricing under garch by a markov chain approximation. *Journal of Economic Dynamics and Control*, 25:1689–1718, 2001.

[21] B. Dumas and E. Luciano. An exact simulation to a dynamic portfolio choice problem under transiction costs. *Journal of finance*, 46:577–595, 1991.

[22] C. Engle and Hamilton. Long swings in the dollar: Are they in the data and do markets know it? *The American EconomicReview*, 80:689–713, 1990.

[23] H. Fama. The behavior of stock market prices. *Journal of business*, 38:34–105, 1965.

[24] J. E. Figueroa-López, S. R. Lancette, L. Kiseop, and Y. Mi. Estimation of nig and vg models for high frequency financial data. *Handbook of Modeling High-Frequency Data in Finance*, 4, 2011.

[25] C.-D. Fuh, K. W. R. Ho, I. Hu, and R.-H. Wang. Option pricing with markov switching. *Journal of Data Science*, 10:483–509, 2012.

[26] S.M. Goldfeld and R.E. Quandt. A markov model for switching regressions. *Journal of Econometrics*, 1(1):3–15, March 1973.

[27] S. F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Fina*, 42:27–62, 1996.

[28] B. Grigelionis. Process of meixner type. *Lithuanian mathematics journal*, 39(1):33–41, 1999.

[29] J. D. Hamilton. Rational-expectation econometric analysis of changes in regime an investigation of the term structure if interest rates. *Journal of Economic Dynamics and Control*, (12):385–423, 1988.

[30] J. D. Hamilton. A new approach to the economic analysis of nonstationary time series and business cycle. *Econometrica*, 57(2):357–384, March 1989.

[31] J. D. Hamilton. Analysis of time series subject to changes in re. *Journal of Econometrics*, 45:39–70, 1990.

[32] J. D. Hamilton. *Handbook of Econometrics*, chapter 50:State-Space Models, pages 3041–3077. North-Holland, 1994.

[33] J. D. Hamilton. Regime-switching models. In Steven N. Durlauf and Lawrence E. Blume, editors, *The New Palgrave Dictionary of Economics*. Palgrave Macmillan, 2nd edition, 2005.

[34] S.H. Hurst, E. Platen, and S. Rachev. Subordinated market index model: a comparison. *Financial Engineering and the Japanese Markets*, 4:97–124, 1997.

[35] G. Iaquinta and S. Ortobelli. Distributional approximation of asset returns with nonparametric markovian trees. *International Journal of Computer Science and Network Security*, 6:69–74, 2006.

[36] I. Karatzas and S. E. Shreve. *Brownian motion and sS.tochastic calculus*. Springer, 1997.

[37] C. Kardaras and E. Platen. Minimizing the expected market time to reach a certain wealth level. *Society for Industrial and Applied mathematics*, 1:16–29, 2010.

[38] C. J. Kim and C. R Nelson. *State-Space models with regime switching*. MIT press, 1999.

[39] F. Klaassen. Improving garch volatility forecasts with regime-switching garch. *Empirical Economics*, 27:363–394, 2002.

[40] I. Kondor, S. Pafka, and G. Nagy. Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking and Finance*, 31:1545–1573, 2007.

[41] F. Lamantia, S. Ortobelli, and S. Rachev. An empirical comparison among var models and time rules with elliptical and stable distributed returns. *Investment Management and Financial Innovations*, 3:8–29, 2006.

[42] A. Leccadito, S. Ortobelli, and E. Russo. Portfolio selection and risk management with markov chains. *International Journal of Computer Science and Network Security*, 7:115–123, 2007.

[43] D. B. Madan, P. Carr, and E. C. Chang. The variance gamma process and option pricing. *European finance review*, 2(79-105), 1998.

[44] D.B. Madan and E. Seneta. The vg model for share market returns. *Journal of Business*, 63:511–524, 1990.

[45] B. Mandelbrot and M. Taylor. On the distribution of stock price differences. *Operations Research*, 15:1057–1062, 1967.

[46] B. B. Mandelbrot. The variation of certain speculative prices. *Journal of business*, 36(394-419), 1963.

[47] M. B. Marcus and J. Rosen. *Markov processes, Gaussian processes, and local times*. Cambridge University Press, 2006.

[48] E. Mazzola and P. Muliere. Reviewing alternative characterizations of meixner process. *Probability Surveys*, 8:127–154, 2011.

[49] J.H. McCulloch. Simple consistent estimators of stable distribution parameters. *Communications in statistics - simulation and computation*, 15:1109–1136, 1986.

[50] A. Mijatović and M. Pistorius. Continuously monitored barrier options under markov processes. *Mathematical Finance*, 23(1-38), 2013.

[51] C. R. Nelson and C. I. Plosser. Trends and random walks in macroeconomic time series. *Journal of Monetary Economics*, 10:139–162, 1982.

[52] J. R. Norris. *Markov chains*. Cambridge University Press, 1997.

[53] S. Ortobelli, E. Angelelli, and D. Toninelli. Set-portfolio selection with the use of market stochastic bounds. *Emerging Markets Finance and Trade*, 47:5–24, 2011.

[54] S. Ortobelli, S. Rachev, and F. Fabozzi. Risk management and dynamic portfolio selection with stable paretian distributions. *Journal of empirical finance*, 17(2):195–211, 2010.

[55] S. Ortobelli and T. Tichy. On the impact of association measures in the portfolio theory. Technical report, University of Bergamo, 2010.

[56] A. R. Pagan and W. G. Schwert. Alternative models for conditional stock volatility. *Journal of Econometrics*, 45:267–290, 1990.

[57] G. Papp, S. Pafka, M.A. Nowak, and I. Kondor. Random matrix filtering in portfolio optimization. *ACTA Physica Polonica*, B36:2757–2765, 2008.

[58] P. Praetz. The distribution of share price changes. *Journal of Business*, 45:49–55, 1972.

[59] S. Rachev and S. Mittnik. *Stable Paretian model in finance*. Chichester, 2000.

[60] S. Rachev, S. Ortobelli, S. Stoyanov, F. Fabozzi, and Biglova A. Desirable properties of an ideal risk measure in portfolio theory. *International Journal of Theoretical and Applied Finance*, 11:19–54, 2008.

[61] G. Samorodnitsky and M. S. Taqqu. *Stable non-Gaussian random processes:Stochastic models with infinite variance*. Cha, 1994.

[62] K. Sato. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999.

[63] W. Schoutens. The meixner process in finance. Eurandom report 2001-2002., Eurandom, Eindhoven., 2001.

[64] W. Schoutens. *Lévy processes in finance:pricing financial derivatives*. Wiley series in probability and statistics, 2003.

[65] W. Schoutens and J. L. Teugels. Lévy processes, polynomials and martingales. *Communications in statistics: Stochastic models*, 14:335–349, 1998.

[66] W.F Sharpe. The sharpe ratio. *Journal Portfolio Management*, 21:49–58, 1994.

[67] A. Staino and S. Ortobelli. Exotic options with lévy processes: the markovian approach. *Investment Management and Financial Innovations*, 8(1):140–156, 2011.

[68] S. Stoyanov, G. Samorodnitsky, S. T. Rachev, and S. Ortobelli. Computing the portfolio conditional value-at-riskchanges α-stable case. *Probability and Mathematical Statistics*, 26, 2006.

[69] T. Tichy and S. Ortobelli. Concordance measures and portfolio selection problem. *Journal of Economics, Management and Business*, 2009.

[70] D.D. Yao, Q. Zhang, and X.Y. Zhou. A regime switching model for european option pricing. *nternational Series in Operations Research & Management Science*, 94:281–300, 2006.