UNIVERSITÀ DEGLI STUDI DI BERGAMO

and

VSB – TECHNICAL UNIVERSITY OF OSTRAVA

Doctoral Thesis

Financial Applications of the Conditional Expectation

Author:
Noureddine Kouaissah

Supervisor:
Prof. Sergio Ortobelli Lozza
Co-supervisor:
doc. Ing. Tomas Tichý, Ph.D

A thesis submitted in fullment of the requirements
for the degree of Doctor of Philosophy in
Ph.D. School in Analytics for Economics and Business
and
Ph.D. School in Economic Policy and Administration - Finance

December 2016

Declaration of Authorship

I, Noureddine Kouaissah, declare that this thesis titled, "Financial Applications of the Conditional Expectation" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at these Universities.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at these Universities or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: *Noureddine Kouaissah*

Place and date: Bergamo 10 December 2016

"*An investment in knowledge pays the best interest."* - *Benjamin Franklin*

Abstract

This dissertation examines different financial applications of some conditional expectation estimators. In the first application, we provide some theoretical motivations behind the use of the moving average rule as one of the most popular trading tools among practitioners. In particular, we examine the conditional probability of the price increments and we study how this probability changes over time. In the second application, we present different approaches to evaluate the presence of the arbitrage opportunities in the option market. In particular, we investigate empirically the well-known put-call parity no-arbitrage relation and the state price density. We first measure the violation of the put-call parity as the difference in implied volatilities between call and put options. Furthermore, we propose alternative approaches to estimate the state price density under the classical hypothesis of the Black and Scholes model. In the third application, we investigate the implications for portfolio theory of using conditional expectation estimators. First, we focus on the approximation of the conditional expectation within large-scale portfolio selection problems. In this context, we propose a new consistent multivariate kernel estimator to approximate the conditional expectation. We show how the new estimator can be used for the return approximation of large-scale portfolio problems. Moreover, the proposed estimator optimizes the bandwidth selection of kernel type estimators, solving the classical selection problem. Second, we propose new performance measures based on the conditional expectation that takes into account the heavy tails of the return distributions. Third, we deal with the portfolio selection problem from the point of view of different non-satiable investors, namely risk-averse and risk-seeking investors. In particular, using a well-known ordering classification, we first identify different definitions of returns based on the investors' preferences. The new definitions of returns are based on the conditional expected value between the random wealth assessed at different times. Finally, for each problem, we propose an empirical application of several admissible portfolio optimization problems using the US stock market.

Keywords: Moving Average, conditional probability, systemic risk, arbitrage opportunities, state price density, conditional expectation estimators, and large-scale portfolio selection problems.

Acknowledgements

First and foremost I would like to thank my supervisors Sergio Ortobelli Lozza and Tomas Tichý for giving me the opportunity to carry out this project and for all their invaluable advice, support and encouragement.

I have been exceedingly lucky in both the support, knowledge and skills training I have received from those around me during the course of this thesis, and in addition to Professor Ortobelli there are a number of people who deserve specific mention. I would like to express my gratitude to the directors of the Ph.D. Schools, Prof. Marida Bertocchi (R.I.P) for her constant motivation and encouragement and Prof. Dr. Ing. Zdenek Zmeskal for the support and suggestion received. I extend my gratitude to Professor Milos Kopa for his insightful suggestions and expert guidance. For this dissertation I would like to thank my reading committee members and all my Professors. I also wish to express my gratitude to Professor Armin Schwienbacher for welcoming me at the SKEMA Business School, for four months.

I would like to thank all my family for all their love and continuous encouragement. For my parents who raised me with a love of science and supported me in all my pursuits. I am very grateful to my loving, supportive, encouraging, and patient wife Sara whose faithful support during the stages of this Ph.D. is so appreciated.

Finally, I would like to thank my friends and colleagues in the University of Bergamo, Technical University of Ostrava and SKEMA Business School for their friendship and intellectual camaraderie over the past few years—particularly Serena, Federica, Valentina, Davide, Muhammed, Ferdinand and Matteo. Thank you.

Contents

Dedicated to Sara, Ahmed, Saadia, Mostafa, Fatima and all my family

1. Introduction

In this thesis, we study some financial applications of the conditional expectation and apply this work to three distinguished problems in finance. The main component that connects these three issues is the notion of conditioning, an important concept in probability and statistics, which turn out to be extremely useful in financial modeling. The conditional expectation $E(Y|X)$, represents the best estimate of the random variable Y given the available information about X. Given the importance that conditional expectation plays in modern finance and in several pricing and risk management problems.

First, we extensively use the conditional expectation to provide some theoretical motivations behind the use of the moving average rule as one of the most popular trading tools among practitioners. In particular, we examine the conditional probability of the price increments and we study how this probability changes over time. Then, we compare the ex-post wealth obtained using these trading rules and other portfolio strategies. The ex-post analysis confirms that it is better using these rules to predict the market trends. In this context, we suggest a methodology that incorporates moving average rules as *alarm rules* to predict potential fails of the market.

Second, we present different approaches to evaluate the presence of the arbitrage opportunities in the market. In this context, we propose alternative approaches to estimate the state price density using the conditional expectation estimators. In particular, we use two different methodologies to evaluate the conditional expectation of a random variable X given a random variable Y, namely the kernel method and the OLP method recently proposed by Ortobelli et al. (2015). The kernel nonparametric regression method allows estimating the regression function, which is a realization of the conditional expectation $E(Y|X)$, while the second approach estimates the conditional expectation (intended as a random variable), based on an appropriate approximation of the σ-algebra generated by X.

Third, we examine the use of the conditional expectation, either to reduce the dimensionality of large-scale portfolio problem or to propose alternative risk-reward performance measures. In particular, we focus on three different financial uses. In the first use, we discuss and examine some correlation measures (based on the conditional expectation) used to approximate properly the returns in large-scale portfolio problems. Then, we compare the impact of alternative return approximation methodologies on the ex-post wealth of a classic portfolio strategy. In this context, we show that correlation measures that use properly the conditional expectation perform better than the classical ones. Moreover, the correlation measure typically used for returns in the domain of attraction of a stable law works better than the classical Pearson correlation does. In the second usage, we propose new performance measures based on the conditional expectation that takes into account the heavy tails of the return distributions. Then, we examine portfolio strategies based on the optimization of the proposed performance measures. In particular, we compare the ex-post wealth obtained applying portfolio strategies, which use alternative performance measures based on the conditional expectation. Finally, we deal with the portfolio selection problem from the point of view of different non-satiable investors: namely, risk-averse and risk-seeking. Doing so, we propose alternative use of the conditional expectation in different portfolio problems. Let us proceed to give a more detailed introduction of the content of the three chapters.

1.1 On the use of moving average rule

Technical trading rules are based on past prices and volume information, which help to generate discrete (buy or sell signal) trading recommendation. However, until 1980s the academic society was skeptical towards the use of technical analysis. Accordingly, we can divide technical analysis literature into two periods. The first period supported the impracticability of applying technical analysis for prediction of the future (see Alexander, 1964; Fama and Blume, 1966; Fama, 1970; and the references therein). There are maybe two reasons for such conclusion. The first reason is that earlier studies often assume a random walk model for the stock price, which rule out any profitability of technical trading. The second reason is that no adequate theoretical support for such strategies was provided. This thesis attempts to overcome this gap and provides theoretical foundations for the most popular rule among practitioners, the *moving average rule*. The second period can be considered as a rebirth of technical analysis, where a significant amount of theoretical and empirical works has been developed to support its validity and efficiency (see Brock et al, 1992; Levich and Thomas, 1993; Lo et al., 2000; Chiarella et al., 2006; Moskowitz et al., 2012; and the references therein).

According to many researchers the seminal work of Brock et al. (1992) seems to be the first major study that provides convincing evidence on the profitability of technical analysis. They test two of the simplest and most popular trading rules, the moving average and the trading range break rules. Their overall results, using the bootstrap methodology, provide a strong support to technical strategies against four popular null models: the random walk, the AR (1), GARCHM and the EGARCH models. They find that buy signals generate higher returns than sell signals and the return following buy signals are less volatile than returns following sell signals. Recently, Neely et al. (2013) find that technical indicators, primarily the moving averages, have the forecasting power of the stock market matching or exceeding that of macroeconomic variables. To sum up, there is sufficient evidence in the literature to support the technical analysis as a profitable strategy.

In the first chapter, we discuss and provide some theoretical motivations for the use of the moving average as one of the most popular technical trading rules. In contrast to the vast studies that use the moving average as an indicator function that indicates merely an up or down state of the market, this study sets the theoretical foundation by demonstrating its validity from a statistical point of view under some particular hypothesis. Thus, we prove that when the moving average rule applies, we could have some implications on the up and down trend probabilities. For this reason, we believe that this rule can be better used to predict the probability of the market fails, such as during periods of systemic risk as suggested by Tichý et al. (2015) and Giacometti et al. (2015). Now, we provide a more detailed introduction for the second application of the conditional expectation.

1.2 Alternative methods to evaluate the arbitrage opportunities

The option-pricing theory has had a central role in modern finance ever since the pioneering work of Black and Scholes (1973) (hereinafter BS). The main idea behind the BS option pricing model is that the price of an option is defined as the least amount of initial capital that permits the construction of a trading strategy whose terminal value equals the payout of the option. BS model has a great importance for improving research on the option pricing techniques. Unfortunately, widespread empirical analyses point out that a set of assumptions under which BS model built, particularly normally distributed returns and constant volatility, result in poor pricing and hedging performance. However, different generalizations of the BS model have been proposed in literature – see, e.g., Merton (1976), Heston (1993) and Bates (1996) for more details. Generally, most models that have been proposed so far mainly relax some assumptions of BS model and then trying to be justified via general fundamental theorem of asset pricing-FTAP, Harrison and Kreps (1979). This theorem provides many challenges in asset pricing theory. In particular, it asserts that the absence of arbitrage in a frictionless financial markets if and only if there exist an equivalent martingale measure under which the price process is a martingale.

One fundamental entity in asset pricing theory is the so called State Price Density (hereinafter SPD). Among no-arbitrage models, the SPD is frequently called risk-neutral density, which is the density of the equivalent martingale measure with respect to the Lebesgue measure. The existence of the equivalent martingale measure follows from the absence of arbitrage opportunities, while its uniqueness demands complete markets. Breeden and Litzenberger (1979) proposed an excellent framework to fully recover the SPD in an easy way. In this method, the SPD is simply equal to the second derivative of a European call option with respect to the strike price, see among others Brunner and Hafner (2003), Yatchew and Härdle (2006) for other estimation technique. Furthermore, it is well known that option prices carry important information about market conditions and about the risk preferences of market participants. In this context, the SPD function derived from observed standard option prices has gained considerable attention in the last decades. Indeed, an estimate of the SPD implicit in option prices can be useful in different contexts, see among others Ait-Sahalia and Lo (1998). The most significant application of the SPD is that it allows us to compute the no-arbitrage price of complex or illiquid option simply by integration techniques.

The first fundamental contribution is to evaluate the presence of arbitrage opportunities in the market. To do so, we focus on the violation of the put-call parity no-arbitrage relation and then the nonnegativity of the SPD. Firstly, we measure the violation of put-call parity as the difference in implied volatility between call and put options that have the same strike price, the same expiration date and the same underlying asset. Secondly, we discuss the violation of the nonnegativity of the SPD. This is important, because negative values of the SPD immediately correspond to the possibility of free-lunch in the market.

The second crucial contribution is to propose different approaches to estimate the SPD. We deviate from previous studies in that we estimate SPD directly from the underlying asset under the hypothesis of the BS model. To this end we follow two distinguished approaches to recover the SPD, the first one based on nonparametric estimation techniques “kernel” which are natural candidates (see among others Ait-Sahalia and Duarte, 2003; Benko et al. 2007, for an application to options), then a new method based on conditional expectation estimator proposed by Ortobelli et al. (2015). Firstly, we examine the so called real mean return function using local polynomial smoothing technique. Then, we estimate the conditional expectation under real probability density. According to the hypothesis of BS model, we are able to derive a closed formula for approximating the conditional expectation under risk neutral probability. The main goal of this contribution is to examine and compare the conditional expectation method and the nonparametric technique. These methods allow us extrapolating arbitrage opportunities and relevant information from different markets (futures and options) consistently with the analysis of the underlying. Finally, we give a more detailed introduction for the third application of the conditional expectation.

1.3 On the impact of conditional expectation estimators in portfolio theory

It is well known that the mean-variance model (see Markowitz 1952) is theoretically justified even by utility theory when returns are normally distributed.[1] Unfortunately, the Gaussian distributional assumption of financial return series is mostly rejected, as was already proved by e.g. Mandelbrot (1963) and Fama (1965).[2] Over the years, a significant number of studies have been published related to the topic of portfolio selection problems. Most of these studies have proposed different portfolio selection formulations based on operational research models that try to overcome the mean-variance shortcomings.[3]

According to many researchers, see among others Papp et al. (2005) and Kondor et al. (2007), the portfolio selection problem is extremely related to the estimation of inputs, statistical parameters, which describe the dependence structure of the returns. In particular, the problem of parameter estimation increases with the number of assets. Several approaches have been proposed in the literature to mitigate this problem, among which are naïve diversification, shrinkage estimators, resampling methods, and imposing constraints on the portfolio weights (see DeMiguel et al. 2009; Ledoit and Wolf 2003 and the reference therein). In this context, statistical nonparametric techniques have received significant interest from academics and the investment management community (see, e.g. Ait-Sahalia and Lo 1998, and Scott 2015).

In this chapter, we assess the impact of nonparametric techniques based on the use of conditional expectation estimators in the portfolio theory. In particular, we discuss the use of the conditional expectation for three financial applications: a) approximation problems within large-scale portfolio selection problems, b) performance valuation considering the heavy tails of returns and, c) optimal portfolio choices consistent with different investor preferences.

The first contribution of this chapter is to investigate the impact of alternative return approximation methods depending by k-factors in large-scale portfolio problem (such as in the k-fund separation model of Ross (1979)). In particular, we examine and compare the classical return approximation with a nonparametric approximation of the returns depending on few factors obtained by a principal components analysis (PCA). Furthermore, according to Ortobelli

[1] See, among others, Tobin (1958); and Levy and Markowitz (1979).

[2] For deeper discussion see among others Samorodnitsky and Taqqu 1994; Rachev and Mittnik 2000 and the references therein.

[3] For a survey of recent contributions from operation research and finance to the theory of portfolio selection see Fabozzi et al. (2010).

and Tichý (2015), we determine the principal components (of PCA) either using a correlation matrix suitable for heavy-tailed distribution (called stable linear correlation) or using the classical Pearson correlation matrix (which summarizes the joint dispersion behavior of Gaussian vectors).

The most commonly used approach to estimate the relationship between returns and k factors is the linear approximation based on the ordinary least squares (OLS) estimator (see Ross (1979)). This approximation appears good enough when the returns are normally distributed. Admitting small departures from normality of the returns do not affect the regression coefficients greatly, however errors with a heavier tailed distribution, which is more suitable for modeling asset returns, can significantly affect the estimated OLS regression coefficients, (see Nolan et al. (2013)). Moreover, we believe that there exists substantial evidence of nonlinearity in the financial dataset used to estimate the returns (see among others Rachev et al. 2008). For this reason, according to Ruppert and Wand (1994), we propose a nonparametric regression analysis to approximate the returns. This approach relaxes the assumptions of linearity and it suitable even for non-Gaussian distributions. In this context, we prove that the variability of errors of the return approximation decreases as the number of factor increases even when elliptically distributed returns present heavy tails. In addition, using convex dominance testing, we find that the nonparametric regression outperforms much better than its counterpart parametric (OLS) does. This empirical analysis is provided using portfolios of the components of S&P 500 index.

The second contribution of this chapter is to deal with a proper evaluation of portfolio choices that account the distributional tails of portfolios. In particular, the primary purpose of this contribution is to present theoretically sound portfolio performance measures considering a more realistic behavior of the returns (i.e. heavy-tailed distributions). Using a new alternative conditional expectation estimator proposed by Ortobelli et al. (2015), we are able to forecast the conditional expected portfolio returns with respect to a given sigma algebra of events (either generated by possible profits or generated by possible losses). More specifically, the first suggested performance measure is based on the conditional expectation with respect to two different σ-algebras (the σ-algebra generated by the portfolio losses, and the σ-algebra generated by the portfolio profits). While the second performance measure considers σ-algebras generated by the joint losses and by joint gains in the market. Moreover, we illustrate how the new performance measures can mitigate the shortcoming of the classical Sharpe ratio (see Sharpe (1994)) showing with an ex-post empirical analysis their tested higher capacity to produce wealth in the US market.

Finally, we deal with the portfolio selection problem from the point of view of different non-satiable investors, namely, risk-seeking and risk-averse, see Ortobelli et al. (2015). In particular, using a well-known ordering classification, we first identify different definitions of returns according to the investor's preferences. The new definitions of returns are based on the conditional expected value between the random wealth assessed at different times. Using conditional expectation estimator, we are able to forecast the investors' behavior by comparing the wealth sample path obtained by considering their different preferences.

1.4 The aim of the thesis

Conditional expectation is an important concept in probability and statistics which turn out to be extremely useful in financial modeling. It plays a crucial role in portfolio theory and in several pricing and risk management problems. The aim of this dissertation is to assess the impact of the conditional expectation on different financial applications, e.g. arbitrage opportunities, state price density estimation and large-scale portfolio selection problems etc. Given uncertainty in the input model and parameters, the goal of the study often becomes the estimation of a conditional expectation among different financial variables. The conditional expectation is expected performance conditioned on the selected model and parameters. The distribution of this conditional expectation describes precisely, and concisely, the impact of input uncertainty on performance prediction. Conceptually, from probability theory perspective, the conditional expectation is well studied and its properties are mainly proved.

Given the importance of technical analysis, we extensively use the conditional expectation to provide theoretical foundations for the most popular rule among practitioners, the moving average rule. This contribution attempts to overcome a significance gap in literature which is that no adequate theoretical support for such strategies exists. Moreover, to contribute to the literature on option pricing theory, we present different approaches to evaluate the presence of the arbitrage opportunities in the option market. Then, we propose alternative methods to estimate the SPD. To achieve this aim, we estimate the density of a conditional expectation using two different approaches, namely the classical kernel estimator and a new method recently proposed by Ortobelli et al. (2015). Finally, the last aim of the thesis is to examine and investigate the implications for portfolio theory of using conditional expectation estimators.

The rest of this dissertation is organized as follows. Chapter 2, contains detailed discussion of the conditional expectation and summaries the financial theory needed for the development of the thesis. Chapter 3, provides theoretical and practical motivation behind the use of moving average rules. Chapter 4, presents some methods to evaluate the arbitrage opportunities and proposes alternative methods to estimate the SPD. Chapter 5, examines and discusses the impact of conditional expectation estimators in the portfolio theory. Finally, chapter 6 concludes the thesis.

Chapter 2

Conditional expectation and principles of finance

In this chapter we briefly introduce some of the most important concepts from the probability theory and financial mathematics that are useful in the financial applications of the conditional expectation. The main goal is not to give a comprehensive study or a complete overview of the literature related to this matter. The purpose is to define all concepts that we deemed necessary for the reminder of the thesis and present them in compact way, for deeper discussion we refer to Shreve (2004) and Musiela and Rutkowski (1997), Rachev et al. (2008) among others.

2.1 Probability Theory Preliminaries

In this section we briefly give an overview of those aspects and concepts of the probability theory that will be used throughout the thesis. There is, of course, a plenty of excellent books introducing the probability theory. A few among them are Billingsley (1995), Shiryaev (1996) and Chung (2001).

We start with a recall that a probability space is a triple $(\Omega, \Im, P)$ where

- Ω is a non-empty set of all possible outcome that we are interested in, called sample space.
- $\Im$ is a sigma algebra on Ω, i.e. a collection of sub-sets of Ω closed under all countable set operations. This collection is a σ-*field* and its elements are called events.
- $P : \Im \to [0,1]$ is the *probability measure* such that:

— P is countably additive i.e. if $A_i \subseteq \Im,\ i = 1,2, \ldots$, is a countable collection of pairwise disjoint sets, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$,

— $P(\Omega) = 1$.

A **random variable** X is simply a $\mathfrak{I}$-measurable function mapping Ω to the real numbers i.e. $X: \Omega \to \mathbb{R}$, such that $X^{-1}\big((\infty, x]\big) \in \mathfrak{I}$, for any $x \in \mathbb{R}$, where $X^{-1}\big((\infty, x]\big) = \{\omega \in \Omega | X(\omega) \leq x\}$. In other words, $\{\omega: X(\omega) \leq x\}$ belongs to $\mathfrak{I}$ for all real x.

Next we define the cumulative distribution function (cdf) any function $F_X(\cdot)$ with domain the real line and counter domain the interval [0,1] which satisfies:

$$F_X(x) = P(X \leq x), \quad \forall\, x \in \mathbb{R}, \tag{2.1}$$

where, right-hand side represents the probability that the random variable X takes on a value less than or equal to x. Generally, every cumulative distribution function F_X is a non-decreasing and right continuous function. Furthermore, it has the following two properties:

- $\lim\limits_{x \to -\infty} F_X(x) = 0$.
- $\lim\limits_{x \to \infty} F_X(x) = 1$.

Furthermore, if the derivative of the distribution function exists then we say that its cumulative distribution function is absolutely continuous. In this case we denote $f_X(x) = F_X'(x)$, and $f_X(x)$ is called the probability density function (pdf).

We are often interested in the inverse of the distribution function F which we define as:

$$F_X^{-1}(p) = \inf_{x \in \mathbb{R}} \{F_X(x) \geq p\}, \quad 0 < p < 1. \tag{2.2}$$

$F_X^{-1}(p)$ is correspondingly called the quantile function ($Q_p(X)$) and it can be used to translate results obtained for the uniform distribution to other distributions.

Conceptually, when we deal simultaneously with more than one random variable the joint cumulative distribution function can also be generalized. Indeed, for $\boldsymbol{t} = (t_1, t_2, \dots, t_n)$ and $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ we define $F_{\boldsymbol{t}}(\mathbf{x})$ the multivariate distribution function of the random vector $\big(X_{t_1}, X_{t_2}, \dots, X_{t_n}\big)$ as:

$$F_{\boldsymbol{t}}(\mathbf{x}) = P\big(X_{t_1} \leq x_1, X_{t_2} \leq x_2, \dots, X_{t_n} \leq x_n\big). \tag{2.3}$$

Given $\big(X_{t_1}, X_{t_2}, \dots, X_{t_n}\big)$ an absolutely continuous random vector, then we denote its density in case it exists, by $f_{\boldsymbol{t}}(\mathbf{x})$. The same is true when $\big(X_{t_1}, X_{t_2}, \dots, X_{t_n}\big)$ is discrete, here we define $f_{\boldsymbol{t}}(\mathbf{x})$ as

$$f_{\boldsymbol{t}}(\mathbf{x}) = P\big(X_{t_1} = x_1, X_{t_2} = x_2, \dots, X_{t_n} = x_n\big). \tag{2.4}$$

Finally, we discuss the concept of **conditioning** and briefly introduce some notions of the conditional probability and conditional expectation. To gain some intuition about these concepts, consider two continuous random variables X and Y with joint probability density function $f_{X,Y}(x,y)$, marginal probability functions $f_X(x), f_Y(y)$ and the necessary condition that $f_Y(y) > 0$. The conditional probability density function of *X* given *Y* can be defined as:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}\,. \tag{2.5}$$

With that introduction, we start our study of the conditional expectation. Since both *X* and *Y* are continuous random variables then the conditional expectation of *X* given *Y* can be easily defined as:

$$E(X|Y = y) = \int x f_{X|Y}(x|y)dx. \tag{2.6}$$

The more formal definition of the conditional expectation is presented below. We first define the measurability concept and then next theorem leads to a definition of conditional expectation given a sigma-field.

Definition 2.1: *Given a probability space* $(\Omega, \mathfrak{I}, P)$*, let* $\mathcal{G} \subseteq \mathfrak{I}$ *be sub-σ-field of events. A random variable* $X: \Omega \to \mathbb{R}$ *is measurable with respect to* $\mathcal{G}$ *(or briefly,* $\mathcal{G}$*-measurabble) if and only if the event* $\{X \le x\}$ *is an element of* $\mathcal{G}$ *for all* $x \in \mathbb{R}$.

We know that the conditional expectation of an integrable random variable *X* given a non-null event *G* means:

$$E(X|G) = \frac{E(X\,\mathbb{1}_G)}{P(G)}$$

where, $\mathbb{1}_G$is indicator function and $E(X\,\mathbb{1}_G) = \int_G XdP$ for all $G \in \mathcal{G}$.

Theorem 2.1 (Existence and uniqueness of conditional expectations): *Let X be an integrable random variable, i.e.* $E(X) < \infty$*, defined on the probability space* $(\Omega, \mathfrak{I}, P)$*, and let* $\mathcal{G} \subset \mathfrak{I}$ *be sub-σ-field of* $\mathfrak{I}$*. Then there exists an integrable* $\mathcal{G}$*-measurabe random variable Y such that*

$$E(\mathbb{1}_G Y) = E(\mathbb{1}_G X)$$

for all $G \in \mathcal{G}$*. Furthermore, if there exists another* $\mathcal{G}$*-measurabe random variable* Y' *such that* $E(\mathbb{1}_G Y') = E(\mathbb{1}_G X)$ *for all* $G \in \mathcal{G}$*, then* $Y = Y'$ *a.s.*

Two main general concepts here are: i) we condition with respect to a sub-σ-field and ii) we view the conditional expectation itself as a random variable. Before illustrating the most important properties of conditional expectation, we give a formal definition of the conditional expectation.

Definition 2.2 (Conditional expectation): *Let X be an integrable random variable on the probability space* $(\Omega, \mathfrak{I}, P)$ *and let* $\mathcal{G}$ *be sub-σ-field contained in* $\mathfrak{I}$*. Then there exists an almost surely unique random variable* $E(X|\mathcal{G})$*, called the conditional expectation of X given* $\mathcal{G}$*, which satisfies the following conditions:*

- $E(X|\mathcal{G})$ *is* $\mathcal{G}$*-measurable*
- $E(X|\mathcal{G})$ *satisfies:*

$$\int_G E(X|\mathcal{G})dP = \int_G XdP, \forall G \in \mathcal{G}. \quad (2.7)$$

Now, we present the most important properties of the conditional expectation. Let all random variables appearing below be such that the correspondent conditional expectations are defined, and let $\mathcal{G}$ be sub-σ-field contained in $\Im$. Then, the conditional expectation has the following properties:

- Linearity: $E(aX + bY + c|\mathcal{G}) = aE(X|\mathcal{G}) + bE(Y|\mathcal{G}) + c$ a.s. for constants a, b and c.
- Monotonicity: $X \leq Y$ a.s. $\Rightarrow E(X|\mathcal{G}) \leq E(Y|\mathcal{G})$ a.s.
- Monotone convergence theorem: $X_n \geq 0,\ X_n \nearrow X$ a.s. $\Rightarrow\ E(X_n|\mathcal{G}) \nearrow\ E(X|\mathcal{G})$ a.s.
- Fatou's lemma: $X_n \geq 0 \Rightarrow E(\lim_{n\to\infty} \inf X_n\, |\mathcal{G}) \leq \lim_{n\to\infty} \inf E\,(X_n|\mathcal{G})$ a.s.
- Dominated convergence theorem: If $X_n \to X$ a.s. and $|X_n| \leq Y$ for some integrable random variable $Y \Rightarrow E(X_n|\mathcal{G}) \to\ E(X|\mathcal{G})$ a.s.
- Jensen's inequality: f convex $\Rightarrow\ E(f(X)|\mathcal{G}) \geq f(E(X|\mathcal{G}))$ a.s.
- If X is independent of $\mathcal{G}$ then $E(X|\mathcal{G}) = E(X)$. In particular, $E(X|\mathcal{G}) = E(X)$ if $\mathcal{G}$ is trivial.
- If X is $\mathcal{G}$-measurable, then $E(X|\mathcal{G}) = X$.
- Tower property or law of iterated expectations: if $\mathcal{H} \subseteq \mathcal{G}$ then

$$E(E(X|\mathcal{G})|\mathcal{H}) = E(E(X|\mathcal{H})|\mathcal{G}) = E(X|\mathcal{H}).$$

2.2 Stochastic processes preliminaries

Stochastic processes have a central role in asset pricing theory, for this reason, we will briefly outline some important facts useful in the financial applications of the conditional expectation. Standard references are Karatzas and Shreve (1991), Bjork (2004) and Shreve (2004) among others.

Definition 2.3 (Stochastic process): *A stochastic process indexed by* $t \in \mathbb{R}_+$*, taking its values in* $(\mathbb{R}, \mathcal{B})$*, is a family of measurable mappings* $(X_t)_{t\in\mathbb{R}_+}$*, from a probability space* $(\Omega, \Im, P)$ *into* $(\mathbb{R}, \mathcal{B})$*. The measurable space* $(\mathbb{R}, \mathcal{B})$ *is called the state space and* $\mathcal{B}$ *is the Borel σ-field* [4] *on* $\mathbb{R}$.

A stochastic process $X = (X_t)_{t\geq 0}$ is stationary if its joint probability distribution does not change when shifted in time. In other words, for any integer $k \geq 1$ and real numbers

[4] The Borel σ-field $\mathcal{B}$ on $\mathbb{R}$ is the smallest sigma–field containing every open set.

$0 \le t_1 \le t_2 \le \cdots \le t_k < \infty$ the distribution of the random vector $(X_{t_1+t}, X_{t_2+t}, ..., X_{t_k+t})$ does not depend on *t*. Let us now define an increment of the process $(X_t)_{t\ge 0}$ between time *s* and *t*, $t > s$, as the difference $X_t - X_s$. A stochastic process $(X_t)_{t\ge 0}$ has stationary increments when the probability distribution of any increment $X_t - X_s$ depends only on the length $t-s$ of the time interval, thus it is invariant by translation in time. In particular, increments on equally long time intervals are identically distributed.

Definition 2.4: *A filtration $\mathbb{F} = (\Im_t)_{t\ge 0}$ on the probability space $(\Omega, \Im, P)$ is a non-decreasing family of sub-σ-fields such that:*

$$\Im_s \subset \Im_t \subset \Im \text{ for } 0 \le s < t. \quad (2.8)$$

The filtration is often used to represent the information structure that completely specifies the evolution of information over time. Thus, $\Im_t$ stands for the information presents at time *t*. A stochastic process $X = (X_t)_{t\ge 0}$ is called $(\Im_t)_{t\ge 0}$- adapted if and only if the random variable X_t is $\Im_t$ -measurable for all $t \ge 0$. It simply means that X_t is known at time *t*. Now we are able to define the natural filtration $(\Im_t^X)_{t\ge 0}$ of a continuous stochastic process $(X_t)_{t\ge 0}$ which is the smallest filtration such that the process is adapted.

Definition 2.5: *A stochastic process $X = (X_t)_{t\ge 1}$ is predictable, with respect to a filtration, if X_t is $\Im_{t-1}$- measurable for all $t \ge 1$.*

Now we come to one of the most important concept in financial mathematics, the martingale. A martingale is simply an adapted stochastic process that is constant on average in the following sense:

Definition 2.6 (Martingale): *A stochastic process $X = (X_t)_{t\ge 0}$ is a martingale with respect to a filtration $(\Im_t)_{t\ge 0}$ if*

- *X is $(\Im_t)_{t\ge 0}$ - adapted*
- $E(|X_t|) < \infty$ *for all* $t \ge 0$.
- $E(X_t \mid \Im_s) = X_s \;\; P-a.s.$, *for every pair s, t such that* $0 \le s \le t$.

The third property is of great practical importance and states that the conditional expected value of future observation, given all the past observation, is equal to the last observation. More generally, Martingales turn out to be extremely useful in finance theory since if we can convert any financial assets into martingales, then we consider them as riskless assets under the equivalent martingale measure. On one hand, the process $X = (X_t)_{t\ge 0}$ is called a

supermartingale relative to a filtration $(\Im_t)_{t\geq 0}$ if the third martingale property becomes $E(X_t \mid \Im_s) \leq X_s\ P-a.s.$, for all $0 \leq s \leq t$, here the process has a negative drift. On the other hand, a process $X = (X_t)_{t\geq 0}$ is called **submartingale** whenever we replace the last martingale property with $E(X_t \mid \Im_s) \geq X_s\ P-a.s.$, for all $0 \leq s \leq t$, thus the process has a positive drift.

Let us now introduce one of the most fundamental continuous-time stochastic processes, Wiener process, which has stationary, independent increments, named in honor of Norbert Wiener and first used by Bachelier (1900) in a thesis submitted to the Academy of Paris.

Definition 2.7 (Wiener process): *We define the standard Wiener process $B = (B_t)_{t\geq 0}$ as an $(\Im_t)_{t\geq 0}$-adapted process with Gaussian stationary independent increments and continuous sample paths for which*

$$B_0 = 0,\ E(B_t) = 0,\quad Var(B_t - B_s) = t - s, \text{for all } t \geq 0 \text{ and } s \in [0, t].$$

The Wiener process is also known as **Brownian motion**. The most commonly used process in continuous stock price behavior is the Geometric Brownian Motion (GBM), also known as exponential Brownian motion, which satisfies the following stochastic differential equation (SDE):

$$dS_t = S_t(\mu dt + \sigma dB_t) \tag{2.9}$$

and the solution is given by

$$S_t = S_0 exp(\mu t - 0.5\sigma^2 t + \sigma B_t)\ , \tag{2.10}$$

where, B_t is a Brownian motion and μ and $\sigma > 0$ are drift and volatility constants.

The main underlying assumption of the GBM model is that log-returns are normally distributed. This turn out to play a significant role in many financial theories, including the pioneering Markowitz optimal portfolio model (Markowitz (1952)), Capital Asset Pricing Model (CAPM) (Sharpe 1964; Treynor 1961; Linter 1965) and seminal BS model (1973). However, empirical studies show that the return distributions sensibly diverge from the normal one. Indeed, the profit/loss distributions tend to be asymmetric and present fat tails (Mandelbrot 1963; Fama 1965; Rachev and Mittnik 2000). As a matter of fact, the normality of the assets returns combined with the continuity of the trajectories exhibited by the geometric Brownian model is very often inappropriate since it ignores the fat tails and the fact that real assets prices exhibits jumps.

2.3 Asset pricing

In this section we provide a brief description of the asset pricing methods. Before starting to evaluate the financial strategies we have to define the underlying economy in which we are going to work. Financial markets are places in which people trade financial securities, commodities and other contingent claims at low transaction costs and at prices that reflect supply and demand. Securities include stocks and bonds, and commodities include precious metals or agricultural products. Besides primary assets there are also the secondary instruments that become largely traded during the last decades. In particular, instruments whose payoffs contingent on some primary underlying asset or other factors, the so called the derivatives markets.

The purpose of this section is to provide the main formal ideas that enable us pricing a given financial instrument. Definitely, a pricing model has to be based on an appropriate model of the financial market. First, we consider a continuous-time financial market model where trades can take place continuously during some trading period $t \in [0,T]$, $T > 0$. Second, we impose some rules for the risky asset. We will not allow market friction; there is no default risk, agents are rational and there is no arbitrage. In more concrete terms, no transaction costs (e.g. broker's commission), no bid/ask spread, no taxes, no restrictions on short sales, infinitesimally divisible assets and, if not written otherwise, "no dividends". Clearly, over time we will be concerned only with the price value, therefore only this must be modeled. Since the value of an asset in the future is not known in advance, it is uncertain, so we can model it as a real valued stochastic process.

The concept of risk and uncertainty has a long tradition in modern finance. In this context, the well-developed mathematical theory of probability gives the opportunity to depend on probabilities as a representation of uncertainty. Hence, the uncertain world of a financial market can be described through a probability space $(\Omega, \Im, P)$, where Ω is the sample space, $\Im$ is a sigma algebra on Ω representing the information structure on the market and *P* is a probability measure. In addition, we equip our probability space $(\Omega, \Im, P)$ with a filtration $(\Im_t, 0 \le t \le T)$ of sub-σ-algebra $\Im$ such that $\Im_s \subset \Im_t \subset \Im_T \subset \Im$ for all $0 \le s \le t \le T$. In this context, it is clear that in real life investors can use only information available up to the current time *t*, so $\Im_t$ represents the set of information available to the investor at time *t*, and then $(\Im_t, 0 \le t \le T)$ represents the information flow evolving with time.

Assume that $\Im_t$ is right continuous (i.e. $\Im_t = \cap_{s>t} \Im_s$) and due to technical reasons $\Im_0$ contains all *P*-null sets of Ω, this intuitively means that we know which events are possible and which are not. Without loss of generality assume $\Im = \Im_T$. Now we consider a market with $n+1$ assets, and we denote their price processes by $S_t = (S_{t,0}, S_{t,1}, ..., S_{t,\mathrm{n}})^T$, for all $t \in [0,T]$.

Definitely, price-processes $S_{t,i}, i=1,.....,n$ should be at least $\Im_t$ − adapted which mean that $S_{t,i}, i=1,.....,n$ is $\Im_t$ - measurable for each *t*, thus $S_{t,i}, i=1,.....,n$ is known at time *t* i.e. $\sigma(S_{s,i}, s \le t) \subset \Im_t$ for $0 \le t \le T$. Moreover, we assume the existence of a money market account. If we invest in this account, we have a high degree of certainty about what the return will be over a period of time. Let us call this asset numéraire. In other words, it is a security with strictly positive price at all time. Assume that the 0-th asset is the numéraire; therefore assume that $S_0 = \{S_{t,0}, t \in [0,T]\}$ *P*-a.s. positive. It is often convenient to express the price of a security in units of a chosen numéraire. Therefore, the so called discounted price processes can be defined as $S_t^* = \left(S_{t,i}/S_{t,0}, i=0,....,n, t \in [0,T]\right)$.

The last crucial assumption that needs to be clarified is the no-arbitrage assumption. An arbitrage is a way of trading "strategy" so that one investor starts with zero capital and at some later time say *T* is sure not have lost money and furthermore has a positive probability of having made money "makes money from nothing". The essence of arbitrage-free is that with no initial capital it should not be possible to make a profit without exposure to risk. To be more precise, a definition of strategy is needed and the content of "making money from nothing" needs to be interpreted.

A trading strategy over some time interval $0 \le t \le T$ is a $\Im_t - \textit{predictable}$ process $\theta = \{\theta_{t,i} i=0,...,n, t \in [0,T]\}$ such that $\int_0^t \theta_s dS_s^* < \infty$ and $\int_0^t \theta_s dS_s < \infty$, where predictable means $\theta_{i,t} \in \Im_{t-1}$ for each $0 \le t \le T$, we interpret $\theta_{i,t}$ as the quantity of security *i* (like shares) holds by the investor between time $t-1$ and *t*. By requiring that θ be predictable, we are allowing the investor to select his time *t* portfolio after the prices S_{t-1} are observed. The value of the trading strategy θ can be defined as $V_t(\theta) = \theta^T S_t$. Most compelling is a strategy which requires initial investment $V_0(\theta)$ and thereafter is self-financing. Roughly speaking, a portfolio is called self-financing if and only if an initial investment is made and any reallocation of the portfolio is made without infusion or withdrawal of money, so it is done in a budget neutral way. To be precise, we adopt the following definition. A trading strategy with value process $V_t(\theta)$ is called self-financing if $dV_t(\theta) = \sum_{i=0}^{n} \int_0^t \theta_{s,i} dS_{s,i}$ for $0 \le t \le T$. For some technical constraints on the strategy (i.e. lower bound) we refer to Musiela and Rutkowski (1997).

An appealing property of a financial market is that it is free of arbitrage, meaning that it is impossible to make money out of nothing. In particular, we call a portfolio an arbitrage opportunity if it is self-financing, its value $V_0(\theta)$ at time zero is equal to zero and its value at T is always nonnegative, whereas $V_T(\theta)$ strictly positive is possible. So, with an arbitrage portfolio it is impossible to lose money, whereas making a profit is a possibility. Formally, we say that a self-financing strategy is an arbitrage opportunity if and only if $P(V_0 = 0) = 1$, $P(V_T \geq 0) = 1$ and $P(V_T > 0) > 0$. A market is arbitrage free if no arbitrage possibilities exist. In the sequel the financial market is considered to be arbitrage-free – without any arbitrage-opportunity.

We present now one important relationship between the no-arbitrage assumption and equivalent martingale measures. This is important because this relationship is so relevant for the pricing theory and it is known as Fundamental Theorems of Asset Pricing (FTAP). In this context, following Shreve (2004) we develop and illustrate two fundamental theorems of asset pricing, namely Girsanov and Martingale representation theorems, then we provide some of the main concepts of derivative pricing in continuous time setting. Let $B(t) = (B_1(t), ...B_d(t))$ be a multidimensional Brownian motion on a probability space $(\Omega, \Im, P)$, P is interpreted as the actual probability measure, the one that would be observed from empirical studies of price data.

Theorem 2.2 (Girsanov Theorem): *Let T be a fixed positive time, and let $\Theta(t) = (\Theta_1(t), ..., \Theta_d(t))$ be a d-dimensional adapted process, define:*

$$Z(\mathrm{t}) = \exp\left\{-\int_0^t \Theta(u) \cdot \mathrm{d}\,B(u) - \frac{1}{2}\int_0^t \|\Theta(u)\|^2\, du\right\} \text{ and } \tilde{B}(t) = B(t) + \int_0^t \Theta(u)du \quad (2.11)$$

where, $\int_0^t \Theta dB(u) = \sum_{j=1}^{d} \int_0^t \Theta_j dB_j(u)$ *, and* $\|\Theta(u)\|^2 = \left(\sum_{j=1}^{d} \Theta_j^2(u)\right)^{1/2}$

And assume that: $E\int_0^T \|\Theta(u)\|^2 Z^2(u)du < \infty$

Set $Z = Z(\mathrm{T})$ *.Then* $E(Z) = 1$*, and under the probability measure* Q *given by*

$$Q(A) = \int_A Z(w)dP(w) \text{ for all } A \in \Im, \quad (2.12)$$

The process $\tilde{B}(t)$ *is a d-dimensional Brownian motion* ■

For a complete proof we refer the reader to Shreve (2004) chapter 5. In the following, we consider a stock price process defined with the differential equation:

$$dS_t = \mu_t S_t dt + \sigma_t S_t dB(t)\,,\ t \in [0,T] \tag{2.13}$$

where, the mean rate of return μ_t and the volatility σ_t are allowed to be adapted process, and $\forall t \in [0,T]$, σ_t a.s. not zero. The stock price process is a generalized geometric Brownian motion and an equivalent way of writing it is:

$$S_t = S_0 \exp\left\{\int_0^t \sigma_s dB(s) + \int_0^t (\mu_s - 0.5\sigma_s^2)ds\right\}. \tag{2.14}$$

Moreover, we define a discounted process as: $D_t = \exp\left\{-\int_0^t r_s ds\right\}$ where r_t (money market account) an adapted interest process. It can be shown that $d(D_t S_t) = \sigma_t D_t S_t d\tilde{B}_t$, from which we have:

$$D_t S_t = S_0 + \int_0^t \sigma_u D_u S_u d\tilde{B}(u). \tag{2.15}$$

Using the Girsanov theorem, particularly the relation $dB(t) = -\Theta(t)dt + d\tilde{B}(t)$, where $\Theta(t)$ is the market price of risk (i.e. $\Theta(t) = \frac{\mu_t - r_t}{\sigma_t}$), one can obtain the following formula:

$$S_t = S_0 exp\left\{\int_0^t \sigma_s d\tilde{B}(s) + \int_0^t (r_s - 0.5\sigma_s^2)ds\right\}. \tag{2.16}$$

In continuous time setting, the change from the actual measure *P* to the risk-neutral measure *Q* changes the mean rate of return of the stock but not the volatility. In practice, overall return is reduced from μ_t to the riskless interest rate r_t.

Theorem 2.3 (Martingale representation Theorem): *Let T be a fixed positive time, and assume that* $\Im_t, 0 \le t \le T$, *the filtration generated by the d-dimensional Brownian motion* $B(t), 0 \le t \le T$. *Let* $M_t, 0 \le t \le T$, *be a martingale with respect to this filtration under P. Then there is an adapted, d-dimensional process* $\Gamma(u) = (\Gamma_1(u),....,\Gamma_d(u)), 0 \le u \le T$, *such that*

$$M_t = M_0 + \int_t^t \Gamma(u)dB(u),\ 0 \le t \le T \tag{2.17}$$

If in addition, we assume the notation and assumption of the Girsanov theorem and if $\tilde{M}$, $0 \le t \le T$, *is* $Q-martingale$, *then there is an adapted, d-dimensional process* $\tilde{\Gamma}(u) = (\tilde{\Gamma}_1(u), \ldots, \tilde{\Gamma}_d(u))$ *such that* $\tilde{M}_t = \tilde{M}_0 + \int_t^t \tilde{\Gamma}(u)d\tilde{B}(u),\ 0 \le t \le T$ ■

We call *Q* the measure defined in Girnasov's Theorem, the risk-neutral measure, because it is equivalent to the original measure *P* and it reduces the discounted stock price $D_t S_t$ into a martingale.

Definition 2.8: *A probability measure Q is said to be risk-neutral measure if:*

(i) *Q and P are equivalent (i.e. for every* $A \in \Im$, $P(A) = 0$ *if and only if* $Q(A) = 0$ *) and*

(ii) Under Q, the discounted stock prices $D_t S_{t,i}$ *is a martingale for every* $i = 1,....,n$.

In other words, a market is arbitrage-free if and only if there is a measure Q defined on $(\Omega, \Im)$ equivalent to P (i.e. with same null-sets) such that the discounted price process is a Q-martingale, i.e. $E^Q(S_t \mid \Im_s) = S_s$ for $0 \le s \le t$. In this context, we present two fundamental theorems that are always true whenever the price process follows the classical GBM.

Theorem 2.4 (First FTAP Theorem): *If a market model has a risk-neutral probability measure, then it does not admit arbitrage.* ■

From financial theory perspective one knows that should never offer prices derived from a model that permits arbitrage, hence the assumption of no-arbitrage follows. Indeed, First FTAP provides us a simple condition to apply in order check whether the model adequate or not. Before dealing with the Second FTAP theorem, we provide the following definition:

Definition 2.9: *A market model is complete if every derivative security can be hedged.*

Theorem 2.5 (Second FTAP Theorem): *Consider a market model that has a risk-neutral probability measure. The model is complete if and only if the risk-neutral probability measure is unique* ■

Nicely readable proofs for both theorems can be found in several books, see among others Shreve (2004).

Generally, by uniqueness we mean to find a unique solution of the following system: $\mu_i - r_i = \sum_{j=1}^{d} \sigma_{ij} \Theta_j(t)$, $i = 1,....,m$ which called market price of risk equations. Clearly there are m equations in d unknown market price processes $\Theta_1(t),.....,\Theta_d(t)$.Unfortunately, it already documented in the literature that the equivalent martingale measure Q is not unique. Hence, we denote by Q the set of equivalent martingale measures. Until now, only the primary assets have been considered. The next step towards pricing financial derivatives is simply to define the derivatives. A derivative (or contingent claim) H with expiry date T is some nonnegative ($\Im_t$-measurable) random variable such that $E^Q(H / S_{T,0}) < \infty$ for all $Q \in \boldsymbol{Q}$.

The random variable H models the payoff of the derivative, typically this payoff is determined as a function of the underlying asset (or assets) – the payoff function $H(S_t), t = [0,T]$. A good example could be the plain vanilla call and put options, which they have the following payoff functions $\max(S_t - K; 0)$ and $\max(K - S_t, 0)$ respectively.

The main idea of the pricing based on the no-arbitrage consideration is to find an appropriate (self-financing) strategy that provides the same payoff. Thus, the value of the strategy and the price of such derivative correspond under no-arbitrage conditions. Sadly, this is not always possible. The set of derivatives for which this is fairly possible is called attainable. In this context, a trading strategy θ is called admissible if it is self-financing and $V(\theta) \geq 0$ or if there is $Q^* \in \boldsymbol{Q}$ such that $V^*(\theta)$ is Q^* martingale. Let Φ denote the set of all admissible strategy, a derivative with expiration time T is said to be attainable if there exists some $\theta \in \Phi$ such that $V_T(\theta) = H$ almost surely (see Harrison and Pliska (1981).

Definitely, the price of the claim denoted by $\prod_t(H)$ and the value process $V_t(\theta)$, under no-arbitrage assumption, must coincide for all $t \in [0,T]$. Moreover, it can be demonstrated, see Hafner (2004), that:

$$\prod_t(H) = E^Q(S_{T,0}^{-1}H \mid \Im_t) \text{ for all } t \in [0,T]. \quad (2.18)$$

The last formula is of great practical importance and it is well known as *risk neutral pricing formula*, see Cox and Ross (1976). The price process of the claim is invariant with respect to the choice of the equivalent martingale measure. At this point, it is clear that risk neutral pricing formula (arbitrage free) allows us pricing attainable derivatives. Conceptually, two fundamental markets are well-defined in the literature. The security market model is called complete if every contingent claim is attainable. Otherwise the market is said to be incomplete. For a deep discussion on the completeness of the market we refer to Harrison and Pliska (1981) and Delbaen and Schachermayer (2006). Clearly, the completeness of a market is equivalent to the uniqueness of the risk-neutral measure *Q*, see among others Föllmer and Schied (2002). In the following section a well-known complete market, the BS model, will be introduced.

2.4 Black Scholes market

The pioneering work of BS has a central role in modern finance and a great importance for improving research on the option pricing techniques. The main idea behind BS option pricing model is that the price of an option is defined as the least amount of initial capital that permits the construction of a trading strategy whose terminal value equals the payout of the option. In other words, if options are correctly priced in the financial market, it should not be possible for investors to set up a riskless arbitrage position and earn more than the risk free rate of return. Interestingly, it can be shown that the BS market consisting of a single asset and a risk-free security is a complete market. This means that every contingent claim can be replicated by a dynamic trading strategy, and then the risk-neutral measure is unique. Indeed, the completeness of the market and its assumptions yield the popularity of the BS model.

Since BS market model is complete, any derivatives can be easily priced. Thus, the fundamental result is the pricing formula of European calls and puts options or simply BS pricing formula. Let C_t^{BS} denotes the price of a European call option with expiration date *T* and strike price *K*, see Black and Scholes (1973). Then

$$C_t^{BS}(S_t, K, \tau, r, \sigma) = S_t \Phi(d_1) - Ke^{-r\tau}\Phi(d_2) \tag{2.19}$$

where, $d_1 = \dfrac{\ln(S_t / K) + (r + 0.5\sigma^2)\tau}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$, $\tau = T - t$ is time to maturity, *r* is a riskless interest rate and $\Phi(\cdot)$ is the standard normal distribution function. The power of BS model relies on the construction of a replicating portfolio containing both the option and its underlying asset, which under risk-neutral measure has a return equal to the risk free rate of return *r*. The corresponding price of a European Put option P_t can be obtained from the put-call parity:

$$P_t(S_t, K, \tau, r, \sigma) = C_t(S_t, K, \tau, r, \sigma) - S_t + e^{-\tau r}K. \tag{2.20}$$

For the proof of (2.20) see among others Shreve (2004).

Furthermore, it is possible to consider dividend case. Simply, the BS formula for the price $C(K,T)$ at time zero of European call option on the stock that yields a continuous dividend δ is given by the following formula:

$$C_t^{BS}(S_t, K, \tau, r, \sigma, \delta) = e^{-\delta\tau} S_t \Phi(d_1) - ke^{-r\tau}\Phi(d_2), \tag{2.21}$$

where, $d_1 = \dfrac{\ln(S_t / K) + (r - \delta + 0.5\sigma^2)\tau}{\sigma\sqrt{\tau}}$, $d_2 = d_1 - \sigma\sqrt{\tau}$, and $\tau = T - t$ is the time to maturity and the put-call parity (2.20) becomes:

$$P_t(S_t, K, \tau, r, \sigma) = C_t(S_t, K, \tau, r, \sigma) - e^{-\delta\tau} S_t + e^{-\tau r}K. \tag{2.22}$$

The next measures that are related to financial market model (BS) are the Greeks letters. In particular, for the aim of controlling the risk of their positions, practitioners in the options market varied sensitivities of the BS formula to some variables. Indeed, the sensitivity of a financial instrument (or even a portfolio) with respect to parameters such as spot price (S), volatility (σ), interest rate (r) are denoted by different Greek letters: delta, vega, rho etc. In this context, the sensitivities are measured in terms of derivatives with respect to the parameters, e.g. the Greeks of the call and put options whose price is denoted by $C_t = C_t^{BS}$ and $P_t = P_t^{BS}$ respectively, are summarized in the following Table.

Table 2.1: *Greeks letters, Delta, Vega and Rho*

	Call option	Put option
Delta	$\frac{\partial C_t^{BS}}{\partial S} = \Phi(d_1)$	$\frac{\partial P_t^{BS}}{\partial S} = \Phi(d_1) - 1$
Vega	$\frac{\partial C_t^{BS}}{\partial \sigma} = \frac{\phi(d_1)}{S_t \sigma \sqrt{\tau}}$	$\frac{\partial P_t^{BS}}{\partial \sigma} = \frac{\phi(d_1)}{S_t \sigma \sqrt{\tau}}$
Rho	$\frac{\partial C_t^{BS}}{\partial r} = e^{-r\tau} \tau K \Phi(d_2)$	$\frac{\partial P_t^{BS}}{\partial r} = e^{-r\tau} \tau K (\Phi(d_2) - 1)$

Of course, there are other Greeks that are commonly treated in financial practice, see among others Hull 2015, but they are not considered in this thesis therefore are omitted at this place. The Greeks are very important for hedging – in practice they control the risk of a portfolio position. The main idea is to make a portfolio robust, i.e. insensitive with respect to the parameters changes. Delta, for example, represents the number of shares that must be held at each time in order to perform a perfect dynamic hedge of the options.

2.4.1 Generalizations of the Black Scholes market

The pioneering work of BS model played a crucial role in modern finance and a great importance for improving research on the option pricing methods. Unfortunately, widespread empirical analyses point out that a set of assumptions under which BS model established, particularly normally distributed returns and constant volatility, result in poor pricing and hedging performance. Indeed, the presence of skewness and kurtosis in the market complicates the situation significantly. However, using BS principle different generalizations have been proposed. In this chapter we introduce briefly two of the most common generalizations:

— **Merton Model**, following BS pricing formula, Merton (1976) derived an option pricing formula for the more general case, arguing that the price process might be affected by a sudden shock, e.g. important information. In this case the underlying stock returns are generated as a mixture of both continuous and jump process. The stochastic differential equation governing the dynamics of S_t in this model is given by:

$$dS_t = S_t \mu dt + \sigma S_t dB_t + S_t dZ_t \tag{2.23}$$

where, Z_t is a compounded Poisson process that has logarithmic normal distributed jumps. The Poisson process Z_t models the jump time and it is independent of B_t. The rest of parameters are as in the BS model. Merton model is also known as jump diffusion model.

— **Heston Model** attacks the constant volatility of BS model. In particular, Heston (1993) derived a closed-form solution for pricing European call option on an asset that has stochastic volatility. The Heston stochastic volatility model is given by the following two stochastic differential equations:

$$dS_t = S_t \mu dt + S_t \sqrt{\sigma_t} B_t^{(1)} \quad (2.24)$$

$$d\sigma_t = \kappa(\theta - \sigma_t)dt + \sigma^2 \sqrt{\sigma_t} dB_t^{(2)} \quad (2.25)$$

where, $B_t^{(1)}$ and $B_t^{(2)}$ are two Brownian motions (possibly correlated with ρ). κ models the mean reversion speed of the variance, parameter θ is the long term variance and σ^2 stands for the volatility of the variance. Heston model is also known as stochastic volatility model.

Clearly, advanced generalizations are possible, for an extension of Heston model see for instance Bates (1996).

Given the strong assumptions under which BS model built, in particular normal distribution of returns and constant volatility. Nowadays BS model is still a useful tool among practitioners. A common practice is when BS formula inverted on the market's option. Thus, the so called implied volatility obtained. Indeed, the implied volatility shows that asset prices are more complicated than geometric Brownian motion, so BS parameter σ must be dynamic. Let us briefly introduce the implied volatility concepts in the following section.

2.4.2 Black Scholes Implied Volatility

The Implied volatility (IV), first introduced by Latané and Rendelman (1976), is the parameter estimate obtained by inverting the BS model on market data. In particular, the implied volatility $\tilde{\sigma}(K)$ is defined as the volatility, under which the BS price C_t^{BS} equals the price of call option $\tilde{C}_t$ observed on the market.

For the sake of simplicity, we assume no-dividend (i.e. $\delta = 0$), most variables which are needed to specify the BS model are all directly observable except the volatility of the underlying stock. For example, the risk-free rate can be approximated by government bonds or through the inter-bank offered rates (i.e. Euribor, T-bill). Therefore, observing the market price of an option the implied volatility can be calculated, i.e. a number that assures that BS formula provides the right price. Clearly, due to the non-linearity of BS formula the implied volatility has to be determined by some numerical iterative procedure, typically the Newton-Raphson method is used. Please note that using the same argument as above one can calculate the implied volatility from put option prices. Definitively, under arbitrage free assumption the implied volatility obtained from either put options or call options yields to the some result, see among other Brunner and Hafner (2004). In this context, a related concept is the so called implied volatility surface, which is simply the implied volatility implemented as a function of both strike price and time to maturity, i.e. $\tilde{\sigma}_t(K,\tau)$.

2.5 State Price Density

SPDs derived from cross-sections of observed standard option prices have gained considerable attention during last decades. Since given an estimate of SPD, one can immediately price any path independent derivative. Clearly, the well-known arbitrage free pricing formula is of vital practical importance. In this approach, the option price is given as the expected value of its future payoff with respect to the *risk-neutral measure Q* discounted back to the present time *t*. Formally, the price $\Pi_t(H)$ at time *t* of a derivative with expiration date T and payoff –function $H(\mathrm{S}_T)$ is given by:

$$\Pi_t(H) = e^{-r(T-t)} E^Q\left[H \mid \Im_t\right] = e^{-r(T-t)} \int_0^\infty H(\mathrm{s}) q_{S_T}(\mathrm{s}) ds \quad \text{for all } t \in [0,T], \qquad (2.26)$$

where, $q_{S_T}(\mathrm{s})$ denotes the SPD of S_T conditional on the information $\Im_t$, and $E^Q\left[. \mid \Im_t\right]$ is the conditional expectation operator with respect to the risk-neutral measure. Constant risk-free saving account is assumed. Consider a standard European call option with strike price *K* and maturity T on an underlying asset with price process S_t. Setting $H(\mathrm{S}_T)(S_T) = \max\{S_T - K, 0\}$ the pricing formula (2.18) yields at current time *t* to the price $C_t(K,T)$:

$$C_t(K,T) = e^{-r(T-t)} E^Q(H \mid \Im_t) = e^{-r(T-t)} \int_0^\infty H(\mathrm{S}_T)(s) q_{t,S_T}(s) ds \qquad (2.27)$$

$$= e^{-r(T-t)} \int_0^\infty \max\{s-K, 0\} q_{t,S_T}(s) ds \quad \text{for all } t \in [0,T]$$

where, q_{t,S_T} denotes the SPD of S_T at the current time *t*.

Within the no-arbitrage models, the SPD is frequently called the risk-neutral density, based on the analysis of Ross (1976) and Cox and Ross (1976) see formula (2.18) and thereafter for further discussion.

Breeden and Litzenberger (1978) derived an elegant formula for obtaining an explicit expression for the SPD from option prices. In fact, they observed that the second derivative of the call price function $C_t(K,T)$ with respect to the strike price K is proportional to the SPD. Formally:

$$q_{t,S_T}(x,\tau) = e^{r(T-t)} \left. \frac{\partial^2 C_t(K,T)}{\partial K^2} \right|_{K=x}. \tag{2.28}$$

The last formula is of great practical importance. Since for any fixed time T, the relation between SPD and IV can be obtained simply by a successive application of (2.21) and (2.28). After some algebra, applying chain rule for derivatives one get:

$$q_{t,S_T}(x,\tau) = e^{r\tau} S_t \sqrt{\tau}\phi(\mathrm{d}_1(x,\tau)) \left\{ \begin{array}{l} \dfrac{1}{x^2\sigma(K,\tau)\tau} + \dfrac{2d_1(x,\tau)}{x\sigma(x,\tau)\sqrt{\tau}} \left. \dfrac{\partial\sigma(K,\tau)}{\partial K} \right|_{K=x} + \\ \dfrac{d_1(x,\tau)d_2(x,\tau)}{\sigma(x,\tau)} \left(\left. \dfrac{\partial\sigma}{\partial K} \right|_{K=x} \right)^2 + \left. \dfrac{\partial^2\sigma}{\partial K^2} \right|_{K=x} \end{array} \right\}, \tag{2.29}$$

where, $d_1(x) = \dfrac{\ln\left(S_t/x\right) + \left(r + 0.5\sigma^2(x,\tau)\right)\tau}{\sigma(x,\tau)\sqrt{\tau}}$, $d_2 = d_1(x,\tau) - \sigma(x,\tau)\sqrt{\tau}$ and $\phi(\bullet)$ is the pdf of a standard normal random variable, we refer the reader to Benko et al (2007) and Brunner and Hafner (2003) for further details.

Please note that the equation (2.29) in order to be perfectly defined, the implied volatility function has to be twice-differentiable function with respect to strike price *K*. Moreover, the SPD has a great practical importance. Indeed, an estimate of the SPD implicit in option prices can be useful in different contexts, see among others Ait-Sahalia and Lo (1998). The most significant application of the SPD is that it allows us computing the no-arbitrage price of complex or illiquid option simply by integration techniques.

2.5.1 No-arbitrage conditions implied by the SPD

In this section, we summarize a set of properties that the SPD demands to satisfy in order to be consistent with no arbitrage argument. Therefore, any violation of these properties, particularly non-negativity, implies the existence of the arbitrage opportunities in the market. Since the SPD is a probability density function, then it must satisfy the nonnegativity condition and integrability to one. Moreover, under no arbitrage condition SPD should reprices all calls, hence the so called martingale property holds. Formally:

- Nonnegativity property: the SPD is nonnegative, i.e.:

$$q_{t,S_T}(x) \geq 0, \quad x \in [0,\infty) \tag{2.30}$$

- Integrability property: the SPD integrate to one, i.e.:

$$\int_0^\infty q_{t,S_T}(x)dx = 1 \qquad (2.31)$$

- Martingale property: the SPD reprices all calls, i.e.:

$$\int_0^\infty \max\{x-K,0\} q_{t,S_T}(x)dx = e^{r(T-t)}C_t(K,T), \quad K \geq 0. \qquad (2.32)$$

The first two properties ensure that the SPD is indeed a probability density. Furthermore, if q_{t,S_T} satisfies the three properties, it is a well-defined SPD and the market is free of arbitrage opportunities with respect to maturity *T*. Following Carr (2001), see also Brunner and Hafner 2003, we are able to express the conditions (2.30), (2.31) and (2.32) in terms of call option price function and implied volatility .Indeed, after a set of equivalent equations it can be shown that:

- The value of a call option is bounded from below and above by its intrinsic value and the underlying stock price respectively, i.e.:

$$S_t \geq C_t(K,T) \geq \max\left\{S_t - Ke^{-r(T-t)};0\right\}, \quad K \geq 0 \qquad (2.33)$$

- Setting $K = 0$ the value of call option approaches the value of the stock, while setting $K = \infty$ the call option value vanishes, i.e.:

$$C_t(0,T) = S_t, \quad \lim_{K\to\infty} C_t(K,T) = 0 \qquad (2.34)$$

- The positivity of $e^{r(T-t)}$ and nonnegativity condition of the SPD imply that call option values are convex in strike prices, i.e.:

$$\frac{\partial^2 C_t(K,T)}{\partial K^2} \geq 0, \quad K \geq 0 \qquad (2.35)$$

In this section, we stated the bounds via European call option, deriving the analogous bounds for a put option is straightforward. To conclude this section Bruner and Hafner (2003) point out that the existence of arbitrage opportunities may hinge on options with different maturities (calendar arbitrage) even there exists a time *t* a SPD that satisfies the above three properties for all maturities $t \in (0,T]$. For more details on SPD and arbitrage considerations we refer to Brunner and Hafner (2003) and literature therein.

2.6 Portfolio selection problems

In finance, a portfolio is a collection of investments held by an individual, an institution or a fund. In this context, a fundamental theory of asset choice under uncertainty is expected utility. There are two different approaches to the problem of portfolio selection under uncertainty stemming from the utility theory. One of them is the stochastic dominance approach, while the second is reward–risk analysis, according to which, the portfolio choice is made with respect to two criteria – the expected portfolio reward and the portfolio risk. In particular, a portfolio is preferred to another one if it has higher expected return and lower risk. Markowitz (1952) introduced the first rigorous approximating model to the portfolio selection problem, where the return and risk are modeled in terms of portfolio mean and variance. Markowitz's main idea was to propose variance as a risk measure and he introduced it in a computational model by measuring the risk of a portfolio via the covariance matrix associated with individual asset returns. This leads to a quadratic programming formulation and it was far from being the final answer to the problem of portfolio selection. Generally, the mean-variance approach works well with Gaussian distribution, which is a very restrictive assumption. Indeed, the Gaussian distributional assumption of financial return series is mostly rejected, see for instance Rachev and Mittnik (2000) and the references therein. It follows that several alternative approaches to portfolio selection has been proposed, see among others Rachev et al. (2008), Farinelli et al. (2008) and the references therein.

Several alternative models have been proposed over the last sixty years, see for instance (Konno and Yamazaki, 1991; Sharpe, 1994; Young, 1998; Rockafellar et al., 2006; Rachev et al., 2008; Farinelli et al., 2008, Ortobelli and Tichý 2015 and the literature therein). For example the Mean-Absolute Deviation (MAD) model, proposed by Konno and Yamazaki (1991) and pioneered by Yitzhaki (1982), introduced and analyzed the mean risk model using the Gini's mean difference as a risk measure. While Markowitz model assumes normality of stock returns, the MAD model does not make this assumption. The MAD model also minimizes a measure of risk, where the measure is the mean absolute deviation (Kim et al., 2005; Konno, 2011). This new measure of risk and its formulation has been broadly applied in the financial field (Zenios and Kang, 1993; Simaan, 1997; Ogryczak and Ruszczynski, 1999).

In the following, we suppose that a portfolio contains n assets, we have a frictionless market in which no short selling is allowed and all investors act as price takers. Thus, the general portfolio selection problem among n assets in the reward-risk model consists of minimizing a given risk measure ρ provided that the expected reward v is constrained by some minimal value m, (see Biglova et al. 2004), that is

$$\min_{x} \rho\left(x'z - z_b\right)$$

$$\text{s.t.} \quad \nu\left(x'z - z_b\right) \geq m, \; \sum_{i=1}^{n} x_i = 1, \; x_i \geq 0, \tag{2.36}$$

where, z_b represents the returns of a given benchmark, and $x'z = \sum_{i=1}^{n} x_i z_i$ denotes the returns of a portfolio with weights $x' = (x_1, \dots, x_n)'$. The portfolio that maximizes the expected reward

v per unit of risk ρ is known as the market portfolio and we obtain it solving the problem (2.36) for one value m among all admissible portfolios. In particular, it is obtained by maximizing the ratio between the reward and risk when both are positive measures, (see Ortobelli et al. 2009), i.e.:

$$\max_{x} \frac{v(x'z - z_b)}{\rho(x'z - z_b)}$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_i = 1, \quad x_i \geq 0 . \tag{2.37}$$

Of course, in the literature, we can find many possible performance ratios. For example, Rachev ratio is the ratio between the CVaR of the opposite of the excess return at a given confidence level and the CVaR of the excess return at another confidence level. Let us briefly formalize the two portfolio performance measures (Sharpe ratio and Rachev ratio) that are used in the empirical analysis.

Sharpe ratio (1994). The Sharpe ratio is used to characterize how well the return of an asset compensates the investor for the risk taken. The Sharpe ratio computes the price for unity of risk, by subtracting the risk-free rate from the rate of return of the portfolio and then dividing the result by the standard deviation of the portfolio returns. Formally:

$$SR(x'\mathrm{z}) = \frac{E(x'\mathrm{z}) - z_0}{\sigma_{x'\mathrm{z}}}, \tag{2.38}$$

where, $E(x'z)$ is the portfolio expected returns, z_0 is the risk-free return and $\sigma_{x'z}$ is the portfolio standard deviation.

Rachev ratio. The Rachev ratio, see Biglova et al. (2004), is the ratio between the average of largest earnings and the mean of largest losses. i.e.:

$$RR(x'z, \alpha, \beta) = \frac{CVaR_{\beta}(z_b - x'z)}{CVaR_{\alpha}(x'z - z_b)}, \tag{2.39}$$

where, Conditional Value-at-Risk (CVaR) is a coherent risk measure (see Rockafellar and Uryasev (2002) and Artzner et al. (1999)) defined as

$$CVaR_{\alpha}(X) = \frac{1}{\alpha} \int_0^{\alpha} VaR_q(X)\, dq,$$

and

$$VaR_q(X) = -F_X^{-1}(q) = -\inf\{x \setminus P(X \leq x) > q\},$$

is the Value-at-Risk (VaR) of the random return X. If we assume a continuous distribution for the probability law of X, then $CVaR_{\alpha}(X) = -E[X|X \leq VaR_{\alpha}(X)]$, therefore CVaR can be interpreted as the average loss beyond VaR. Typically, we use historical observations to estimate the portfolio return and risk measures. A consistent estimator of $CVaR_\alpha(X)$ is given by

$$CVaR_\alpha(X) = \frac{-1}{[\alpha M]}\sum_{i=1}^{[\alpha M]} X_{i:M} \quad (2.40)$$

where M is the number of historical observations of X, $[\alpha M]$is the integer part of αM, and $X_{i:M}$ is the ith observation of X ordered by increasing values. Similarly, an approximation of $VaR_q(X)$ is simply given by $-X_{[\alpha M]:M}$. Once we approximate the portfolio return and risk measures, we apply portfolio selection optimization problems to the approximated portfolio returns.

Therefore, when no short sales are allowed $(x_i \geq 0)$ and it is not possible to invest more than a fixed percentage θ in any asset $(x_i \leq \theta)$, we assume that investors will choose the market portfolio solution to the following optimization problem:

$$max_x G(x'r) \quad (2.41)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} x_i = 1$$

$$x_i \geq 0; x_i \leq \theta; i = 1, \dots, n$$

where $G(x'r)$, for example, could be either the Sharpe Ratio or the Rachev Ratio. Generally, many performance measures have been proposed in the literature, for an overview see among others Farinelli et al. (2008) and the references therein. Furthermore, one important aspect in portfolio optimization is the computational complexity. Some recent studies (see Rachev et al., 2008; Stoyanov et al. 2007) classify the computational complexity of reward-risk portfolio selection problems. In particular, Stoyanov et al. (2007) have shown that we can distinguish four various cases of reward and risk that admit a unique optimum in myopic strategies. Thus, in order to optimize some portfolio selection problems in an acceptable computational time, we use a heuristic algorithm for overall optimization such as the one proposed in Angelelli and Ortobelli (2009) for overall portfolio optimization.

Chapter 3

Theoretical and practical motivations for the use of the Moving average rule

In this chapter, we provide some theoretical motivations behind the use of the moving average rule as one of the most popular trading tools among practitioners. In particular, we examine the conditional probability of the price increments and we study how this probability changes over time. We find that under some assumptions the probability of up-trend is greater than the probability of down trend. Finally, we compare the ex-post wealth obtained using these trading rules and other portfolio strategies. The ex-post analysis confirms that it is useful to use these rules to predict the market trends. In this context, we suggest a methodology that incorporate moving average rules as *alarm rules* to predict potential fails of the market.

3.1 Introduction

Technical trading rules are based on past prices and volume information, which help to generate discrete (buy or sell signal) trading recommendation. However, until 1980s the academic society was sceptic towards the usage of technical analysis. Accordingly we can divide technical analysis literature into two periods. The first period supported the impracticability of applying technical analysis for prediction of the future (see Alexander, 1964; Fama and Blume, 1966; Fama, 1970; and the references therein). There are maybe two reasons for such conclusion. The first reason is that earlier studies often assume random walk model for the stock price, which rule out any profitability from technical trading. The second reason is that no adequate theoretical support for such strategies was provided. This paper attempts to overcome this gap and provides theoretical foundations for the most popular rule among practitioners, the *moving average rule*. The second period can be considered as a rebirth of technical analysis, where a significant amount of theoretical and empirical works has been developed to support its validity and efficiency (see Brock et al, 1992; Levich and Thomas, 1993; Lo et al., 2000; Chiarella et al., 2006; Moskowitz et al., 2012; and the references therein).

According to many researchers the seminal work of Brock et al. (1992) seems to be the first major study that provides convincing evidence on the profitability of technical analysis. They test two of the simplest and most popular trading rules, the moving average and the trading range break rules. Overall their results, using the bootstrap methodology, provide a strong support to technical strategies against four popular null models: the random walk, the AR (1), GARCHM and the EGARCH models. They find that buy signals generate higher return than sell signals and the return following buy signals are less volatile than returns following sell signals. Recently, Neely et al. (2013) find that technical indicators, primarily the moving averages, have forecasting power of the stock market matching or exceeding that of macroeconomic variables. To sum up, there is sufficient evidence in literature to support the technical analysis as a profitable strategy.

In this chapter, we discuss and evaluate the use and the impact of the very popular (among practitioners and academics) moving average rule. This rule generates buy and sell signals based on past data, by calculating the differences between a long run and a short run moving averages. The method attempts to predict the direction of the future price without searching to forecast its level. Mainly it is used to detect major upturns or downturns of the financial market. The common rule is to trade with the trend. The trader initiates a position early in the trend and maintains that position as long as the trend continues. Here we distinguish between trend-following, the most used form, and reverse trend-following. The main difficulty of this method is that a rule has to be chosen from an infinite number of alternatives.

The first central contribution of this chapter is to provide some theoretical motivations for the use of the moving average as one of the most popular technical trading rules. In contrast to the vast studies that use the moving average as an indicator function that indicates merely an up or down state of the market, this study sets the theoretical foundation by demonstrating its validity from a statistical point of view under some particular hypothesis. Thus, we prove that when the moving average rule applies, we could have some implications on the up and down trend probabilities. For this reason, we believe that this rule can be better used to predict the probability of market fails, such as during periods of systemic risk as suggested by Tichý et al. (2015) and Giacometti et al. (2015).

The second fundamental contribution of the chapter is to implement some popular moving averages rules using daily data of S&P 500 components. Thus, we introduce an alarm rule that predicts the presence of market systemic risk. The alarm is a simple rule that counts the assets whose average returns on the last *n* days is lower than the mean on the last *N* trading days ($n < N$). If the number of these assets reach a benchmark we deduce that systemic risk is probably present on the market and thus we should not invest in any asset. Finally, we evaluate the usefulness of the moving average rules by comparing their ex-post wealth with those obtained from other portfolio strategies. In particular, we compare the ex-post wealth obtained maximizing a stochastic timing performance proposed by Ortobelli et al 2016, and the wealth obtained maximizing the Sharpe ratio when in both cases we use a moving average rule as alarm of systemic risk.

The rest of the chapter is organized as follows. Section 3.2 describes the theoretical motivations to the use of moving average rules in relation with conditional expectations from a statistical point of view. Section 3.3 examines different applications of moving average rules. Section 3.4 concludes the chapter.

3.2 Theoretical motivations for the use of the moving average rule

In this section we explore, the theoretical foundations behind using, one of the simplest and most popular technical rules: moving average rule. In this method, buy and sell signals are generated by two moving averages, a long period, and a short period:

$$MA_{n,T}\left(x\right)=\frac{\sum_{i=0}^{n-1} x_{T-i}}{n} \text{ and } MA_{N,T}\left(x\right)=\frac{\sum_{i=0}^{N-1} x_{T-i}}{N},$$

where x_T is the price at time *T,* while *n* and *N* are the lengths of the short and long time periods, respectively. The buy and sell signals of this rule are generated as:

$$\text{If } MA_{n,T}\left(x\right) > MA_{N,T}\left(x\right) \text{ and } MA_{n,T-1}\left(x\right) \leq MA_{N,T-1}\left(x\right) \text{ buy at time T} \quad (3.1)$$

$$\text{If } MA_{n,\Gamma}\left(x\right) < MA_{N,\Gamma}\left(x\right) \text{ and } MA_{n,\Gamma-1}\left(x\right) \geq MA_{N,\Gamma-1}\left(x\right) \text{ sell at time } \Gamma \quad (3.2)$$

In other words, buy (sell) signals are generated when the short moving average crosses the long moving average from below (above). Intuitively, the moving average rule detects changes in stock price trend, as the short moving average is more sensitive to recent price movement than longer one. There are many possible combinations of moving average that can be used based on the choice of *n* and *N*, although trial and errors is usually the best way to find an appropriate length. Some popular MA on daily basis rule are (*n, N*) = [(1, 50), (5, 200), (2, 200), (1, 150)] (see Gencay and Stengos, 1998). The usefulness of the moving average is conditional to fact that trends in prices tend to persist for certain time and can be detected. For more details we refer to Orlandini (2008). Clearly we believe that relationships (3.1) and (3.2) will happen with a given probability belonging to (0, 1). However, we first analyze the case in which we are sure that at a given time inequalities (3.1) and (3.2) apply. We demonstrate that the condition (3.1) is indeed an up-trend and consequently condition (3.2) is a down-trend.

Theorem 3.1: *Let* $x = \{x_s\}_{s\in\mathbb{N}}$ *be a stochastic process adapted to a filtered space* $(\Omega, \Im, \{\Im_s\}_{s\in\mathbb{N}}, P)$*. Suppose there exists a sub-sigma algebra* $G \subseteq \Im$ *such that the conditional expected value of one step increments* $\delta_s = x_s - x_{s-1}$ *assumes only two values* $\mu_+ > 0$ *or* $\mu_- < 0$ *for any s, that is* $E(\delta_s \mid G) = \begin{cases} \mu_+ & \text{with probability } p_s \\ \mu_- & \text{with probability } 1-p_s \end{cases}$ *for any s. Suppose there exist two integers n and N (with* $n < N$*) and* $t = s$ *such that:*

$$MA_{n,t}(x) > MA_{N,t}(x) \text{ and } MA_{n,t-1}(x) \le MA_{N,t-1}(x) \text{ a.s.} \tag{3.3}$$

Then $p_t > 1/2$ *if the following inequality is verified:*

$$2nE(k_A) - 2(N-n)E(k_B) \ge N-n, \tag{3.4}$$

where $k_A = \#\{i \mid \mu_i = \mu_+; i = t-N+1,\ldots,t-n\} = \sum_{i=t-N+1}^{t-n} I_{[\mu_i=\mu]}$ *measures how many times* μ_i *assumes the value of* μ_+ *for* $i = t-N+1,\ldots,t-n$ *(*$I_{[\mu_i=\mu]} = 1$ *if* $\mu_i = \mu_+$ *and* $I_{[\mu_i=\mu]} = 0$ *if* $\mu_i = \mu_-$*) , while* $k_B = \#\{i \mid \mu_i = \mu_+; i = t-n+1,\ldots,t-1\} = \sum_{i=t-n+1}^{t-1} I_{[\mu_i=\mu]}$ *measures how many times* μ_i *assumes the value of* μ_+ *for* $i = t-n+1,\ldots,t-1$. ■

Proof of Theorem 3.1: Starting from the hypothesis assumption of the Theorem:

$$\begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1} x_{t-i} > \dfrac{1}{N}\sum_{i=0}^{N-1} x_{t-i} \\ \dfrac{1}{n}\displaystyle\sum_{i=0}^{n-1} x_{t-1-i} \le \dfrac{1}{N}\sum_{i=0}^{N-1} x_{t-1-i}. \end{cases}$$

The system can be rewritten considering that $\sum_{i=0}^{N-1} x_{t-i} = \sum_{i=0}^{n-1} x_{t-i} + \sum_{i=n}^{N-1} x_{t-i}$ and $\sum_{i=0}^{N-1} x_{t-1-i} = \sum_{i=0}^{n-1} x_{t-1-i} + \sum_{i=n}^{N-1} x_{t-1-i}$ as:

$$\begin{cases} \left(\dfrac{1}{n} - \dfrac{1}{N}\right)\displaystyle\sum_{i=0}^{n-1} x_{t-i} > \dfrac{1}{N}\sum_{i=n}^{N-1} x_{t-i} \\ \left(\dfrac{1}{n} - \dfrac{1}{N}\right)\displaystyle\sum_{i=0}^{n-1} x_{t-1-i} \le \dfrac{1}{N}\sum_{i=n}^{N-1} x_{t-1-i}. \end{cases}, \tag{3.5}$$

For notational simplicity, we change the starting point. Let x_0 be the first observation such that $x_t \equiv x_N$ Then, we define the observations as follow:

$$\underbrace{x_0, \overbrace{x_1, x_2, \ldots, x_{N-n-1}, x_{N-n}}^{N\text{-}n}}_{N\text{-}n}, \underbrace{\overbrace{x_{N-n+1}, \ldots, x_{N-1}, x_N}^{n}}_{n}$$

It follows that the system (3.5) can be formulated as:

$$\begin{cases} \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=N-n+1}^{N} x_i > \frac{1}{N}\sum_{i=1}^{N-n} x_i \\ \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=N-n}^{N-1} x_i \le \frac{1}{N}\sum_{i=0}^{N-n-1} x_i \end{cases},$$

then the first inequality of the system can be developed as:

$$\left(\frac{1}{n}-\frac{1}{N}\right)\left(\sum_{i=N-n}^{N-1} x_i - x_{N-n} + x_N\right) > \frac{1}{N}\left(\sum_{i=0}^{N-n-1} x_i - x_0 + x_{N-n}\right),$$

from which it follows

$$\begin{cases} \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=N-n}^{N-1} x_i - \frac{1}{N}\sum_{i=0}^{N-n-1} x_i > \frac{x_{N-n}-x_0}{N} + \left(\frac{1}{n}-\frac{1}{N}\right)\left(x_{N-n}-x_N\right) \\ \left(\frac{1}{n}-\frac{1}{N}\right)\sum_{i=N-n}^{N-1} x_i - \frac{1}{N}\sum_{i=0}^{N-n-1} x_i \le 0 \end{cases}$$

and consequently,

$$\frac{x_{N-n}-x_0}{N} + \left(\frac{1}{n}-\frac{1}{N}\right)\left(x_{N-n}-x_N\right) < 0.$$

Simplifying last inequality, we obtain that (3.3) implies:

$$x_N > \frac{N}{N-n}x_{N-n} - \frac{n}{N-n}x_0, \tag{3.6}$$

At this point the assumption system has been traced as a unique inequality, for which the system is indeed a sufficient condition. To simplify further such inequality we apply the definition of δ_t recursively, so that we obtain $\forall i > 0$: $x_i = x_0 + \sum_{j=1}^{i}\delta_j$ and then the inequality (3.6) becomes $x_0 + \sum_{j=1}^{N}\delta_j > \frac{N}{N-n}\left(x_0 + \sum_{j=1}^{N-n}\delta_j\right) - \frac{n}{N-n}x_0$, which can be reduced after few steps to:

$$\delta_N > \frac{n}{N-n}\sum_{j=1}^{N-n}\delta_j - \sum_{j=N-n+1}^{N-1}\delta_j\ . \tag{3.7}$$

Now we recall the definition of μ_t and by applying $\mu_t = E(\delta_t \mid G)$ we get:

$$\mu_N > \frac{n}{N-n}\sum_{j=1}^{N-n}\mu_j - \sum_{j=N-n+1}^{N-1}\mu_j.$$

Given that μ_t assumes only two values, either $\mu_+ > 0$ or $\mu_- < 0$, for any *t*, and using the fact that the random variables k_A and k_B are the number of times when μ_i assumes the value μ_+ for *i* varying from $t-N+1$ to $t-n$ and from $t-n+1$ to $t-1$, respectively, the inequality becomes

$$\mu_N > \frac{n}{N-n}\left(k_A\mu_+ + (N-n-k_A)\mu_-\right) - \left(k_B\mu_+ + (n-1-k_B)\mu_-\right),$$

which gives,

$$\mu_N > \mu_- + \frac{n(\mu_+ - \mu_-)}{N-n}k_A - (\mu_+ - \mu_-)k_B.$$

Then applying the expected value operator to last inequality we get:

$$E(\mu_N) > E\left(\mu_- + \frac{n(\mu_+ - \mu_-)}{N-n}k_A - (\mu_+ - \mu_-)k_B\right) = \mu_- + (\mu_+ - \mu_-)\left(\frac{n}{N-n}E(k_A) - E(k_B)\right).$$

Therefore system (3.3) is indeed a sufficient condition for the inequality:

$$E(\mu_N) > \mu_- + (\mu_+ - \mu_-)\left(\frac{n}{N-n}E(k_A) - E(k_B)\right).$$

At this point, after simple manipulation, condition (3.4) can be written as:

$$\mu_- + (\mu_+ - \mu_-)\left(\frac{n}{N-n}E(k_A) - E(k_B)\right) \geq \frac{\mu_+ + \mu_-}{2}.$$

Now, considering jointly the two last inequalities we get:

$$E(\mu_N) > \frac{\mu_+ + \mu_-}{2}. \tag{3.8}$$

Given that $E(\mu_N) = \mu_+ p_N + \mu_-(1-p_N)$, then (3.8) is equivalent to:

$$\mu_+ p_N + \mu_-(1-p_N) > \frac{\mu_+ + \mu_-}{2},$$

from which, after few steps, such that $\mu_+ > 0$ and $\mu_- < 0$, we obtain $p_N > \frac{1}{2}$. Q.E.D.

In Theorem 3.1 we do not point out that the process *x* is the price process. Thus, the theorem could be applied to any stochastic process that satisfies the above conditions. Anyway, Theorem 3.1 suggests that the probability to be in up-trend is greater than the probability to be in down trend when the moving average (3.3) and condition (3.4) apply. Clearly, requiring that the moving average rule holds almost surely, is a strong assumption as proved by the following corollary.

Corollary 3.1: *Condition (3.3) cannot be applied almost surely for independent stationary processes.* ■

Proof of Corollary 3.1: Since the process $\{\delta_i\}_{i=1,..,T}$ is stationary and independent, then $E(\delta_i) = E(\delta_j) = \mu \quad \forall\ i, j$. From equation (3.7) derived by condition (3.3) we should have that:

$$\mu = E(\delta_N) > n\frac{\sum_{j=1}^{N-n} E(\delta_j)}{N-n} - (n-1)\frac{\sum_{j=N-n+1}^{N-1} E(\delta_j)}{n-1} = \mu .$$

Thus, condition (3.3) cannot be true. Q.E.D.

These results essentially prove that the moving average rule has some implications in terms of probability of the future price. Moreover, when we consider stationary and independent increments of log prices (typically Lévy processes), the moving average rules holds for log prices $x = \{x_s\}_{s\in\mathbb{N}}$ only with probability lower than 1. Therefore, under the assumption the process of log prices $x = \{x_s\}_{s\geq 0}$ is a Lévy process, we have that:

$$P\left(MA_{n,T}(x) > MA_{N,T}(x), MA_{n,T-1}(x) \leq MA_{N,T-1}(x)\right) < 1$$

and

$$P\left(MA_{n,T}(x) < MA_{N,T}(x), MA_{n,T-1}(x) \geq MA_{N,T-1}(x)\right) < 1 .$$

On the one hand, Theorem 3.1 cannot be applied to log price processes which are generally considered with stationary and independent increments in financial literature, on the other hand, moving average rules are applied to the price processes which generally do not present stationary and independent increments. In addition, we could obtain a result similar to Theorem 3.1 when we consider non-stationary independent increments. In this case, we can better specify the conditional probability and sub-sigma algebra *G* used in Theorem 3.1.

Proposition 3.1: *Let $x = \{x_s\}_{s\leq T}$ be a stochastic process adapted to a filtered space $(\Omega, \Im, \{\Im_s\}_{s\leq T}, P)$ and suppose the one step increments $\delta_s = x_s - x_{s-1}$ are independent. Consider the σ-algebra G=<$\bigcup_{s=1}^{T} G_s$>, where the σ-algebra G_s is given by $G_s = \{\Omega, \emptyset, \{\delta_s > 0\}, \{\delta_s \leq 0\}\}$, then $E(\delta_s \mid G) = \begin{cases} \mu_{+,s} & \text{with probability } p_s \\ \mu_{-,s} & \text{with probability } 1-p_s \end{cases}$, where $E(\delta_s \mid \delta_s > 0) = u_{+,s}$* and *$E(\delta_s \mid \delta_s \leq 0) = u_{-,s}$ for any s. If we assume that $\mu_{+,s}, \mu_{-,s}$ are constant over time (i.e. $\mu_{+,s} = \mu_+$ and $\mu_{-,s} = \mu_-$), $t = s$ and there exist two integers n and N (with $n < N$) such that:*

$$MA_{n,t}(x) > MA_{N,t}(x) \text{ and } MA_{n,t-1}(x) \leq MA_{N,t-1}(x) \text{ a.s.} \tag{3.9}$$

Then $p_t > \frac{1}{2}$ *if the following inequality is verified:*

$$2n\sum_{i=t-N+1}^{t-n} p_i - 2(N-n)\sum_{i=t-n+1}^{t-1} p_i \geq N-n \qquad (3.10)$$

where, $p_i = P(\delta_i > 0)$ ■

Proof of Proposition 3.1: To prove this result we need the following lemma:

Lemma 3.1: *Let Y be a real random variable defined on a probability space* $(\Omega, \Im, P)$. *Let* $\Im_1$ *be a finite sub-σ-algebra contained in the σ-algebra generated by Y namely* $\sigma < Y >$. *Given a finite sequence of finite sub- σ-algebras* $\Im_i \subseteq \Im$ $i = 2,....,n$ *independent on* $\Im_1$, *then:*

$$E(Y| < \cup_{i=1}^{n} \Im_i >) = E(Y|\Im_1).$$

Proof of Lemma 3.1: First we prove the result for $n = 2$. Since σ-algebra $< \Im_1 \cup \Im_2 >$ is finite, then any event $A \in < \Im_1 \cup \Im_2 >$ can be seen as the intersection of events belonging to $\Im_1$ and $\Im_2$ (i.e. $A = C \cap D$ where $C \in \Im_1$ and $D \in \Im_2$).

Thus $\forall A \in < \Im_1 \cup \Im_2 >$, $A = C \cap D$, $C \in \Im_1$, $D \in \Im_2$

$$\int_A YdP = \int_{C\cap D} E(Y| < \Im_1 \cup \Im_2 >)dP = \int_C YI_D dP = P(D)P(C)\int_C YdP$$
$$= P(D)P(C)\int_C E(Y|\Im_1)dP = \int_A E(Y|\Im_1)dP,$$

where the 1st and 4th equalities are derived from conditional expected value definition, while the 3rd and 5th equalities are derived from the independence between $\Im_1$ and $\Im_2$ (that implies the independence between $E(Y | \Im_1)$ and $\Im_2$). Thus from the unicity we get $E(Y | \Im_1) = E(Y |< \Im_1 \cup \Im_2 >)$ P.a.s. Furthermore, the σ-algebra generated by $\cup_{i=2}^{n} \Im_i$ is still finite and independent on $\Im_1$, thus, considering that $< \Im_1 \cup < \cup_{i=2}^{n} \Im_i > > = < \cup_{i=1}^{n} \Im_i >$. Then applying the result obtained for $n = 2$, the equality $E(Y| < \cup_{i=1}^{n} \Im_i >) = E(Y|\Im_1)$. P.a.s. holds.

Q.E.D.

Now we can prove Proposition 3.1. Observe that any σ-algebra G_s is a sub- σ-algebra of the σ-algebra generated by δ_s. Thus, we can apply Lemma 3.1 to any increment δ_s and we get that:

$$E(\delta_s| < \cup_{q=1}^{T} G_q >) = E(\delta_s|G_s). P. \text{ a. s.}$$

Since G_s is a finite σ-algebra if we apply the conditional expected value definition we get:

$$E(\delta_s | G_s) = E(\delta_s | \delta_s > 0)I_{[\delta_s > 0]} + E(\delta_s | \delta_s \leq 0)I_{[\delta_s \leq 0]}, \quad P. \text{a.s.}$$

where, $E(\delta_s | \delta_s > 0) = u_{+,s}$ and $E(\delta_s | \delta_s \leq 0) = u_{-,s}$. Thus, if we apply Theorem 3.1 we get the thesis.

Q.E.D.

So considering non-stationary independent increments, Proposition 3.1 states that the probability to be in up-trend is greater than the probability to be in down-trend if the moving average (3.9) and condition (3.10) apply.

3.3 Ex-post empirical analysis

In this section, we examine two possible uses of the moving average rules. First, we apply the moving average rules to predict possible losses when we use the components of the S&P 500 index. Secondly, we use the moving average rules as proper portfolio strategies. Finally, we compare the effectiveness of the moving average rule with other portfolio strategies. In all cases we consider daily observations of the S&P 500 components from January 1, 2000 to January 10, 2015 - using a collection of 15 years of daily prices. From the beginning, the list included large well-known and actively traded stocks. In recent years the S&P 500 represent about 70 percent of the American equity market by capitalization. The data set is taken from Thomson-Reuters DataStream. Simultaneously, in neither case the short sales are allowed. We use $x = [x_1, \dots, x_{500}]'$ to denote the vector of percentages invested in each asset and $R = [R_1, \dots, R_{500}]'$ for the vector of returns.

According to several studies, one of the main difficulties in evaluating the profitability using moving average rules is that performance of trading rules depends on accurate choice of rule parameters. In essence, these rules crucially depend on the choice of the short and the long period moving average, n and N respectively. Therefore, the possibility that various combinations of the moving average rules are suitable cannot be dismissed. Although a complete remedy for this issue does not exist, trial and errors remains the best way to find an appropriate length, we mitigate this problem by reporting results from different choices that cover a wide range of possible parameter values, see Table 3.1. In addition, as suggest by several papers, see for instance Brock et al. (1992), these wide range of combinations are quite substantial to give a full picture about the performance of the moving average rules. The rules differ by the length of the short and long period. For example (1,200) indicates that the short period is one day and the long period is 200 days.

In the first empirical analysis, we compare two possible uses of the moving average rules. Firstly, we propose a methodology to predict the periods of systemic risk as suggested by Giacometti et al. (2015). In particular, we consider the assets whose mean on the last n days is lower than the mean on the last N trading days $(n < N)$. If the number of the assets satisfying this rule is higher than 75% of the all assets, we deduce that the probability of systemic losses in the market is high because 3/4 of the assets[5] during last n days are losing their value. Therefore, in presence of systemic risk we do not invest in the market for any asset. Conversely, we invest in the uniform portfolio (0.2% in each components), when we do not observe periods of jointly losses among the assets (for more details on the uniform investment strategy see DeMiguel et al. (2009) and Pflug et al. (2012)). Secondly, since we want to value the impact of the moving average rule, we do not consider (n, N) alarm rule for the second type of strategies. In particular, the second strategy suggests to invest on all assets whose average over the last n days is greater than the average over the last N trading days. In this analysis, starting from January 1, 2000 we calibrate the portfolio every 15 trading days.

Table 3.1 reports summary statistics (mean, standard deviation, skewness, kurtosis, VaR 5%, CVaR 5%, final wealth) of the ex-post returns of a wide range of strategies with and without alarms. In addition, in Table 3.2, we compute the Sharpe Ratio and the performance measure $STARR_\alpha$ defined by Martin et al. (2003)

$$STARR_\alpha(X) = \frac{E(X)}{CVaR_\alpha(X)}, \tag{3.11}$$

with a confidence level $1 - \alpha = 95\%$. STARR allows us to overcome the drawbacks of the standard deviation as a risk measure (Artzner et al. (1999)) and focuses on the downsides risk.[6]

[5] Obviously, it is possible to apply several barriers for detecting systemic risk. However as suggested by several papers (see Biglova et al. (2014) and the references therein) the ¾ of the assets distressed in the market appears sufficient to guarantee contagion and the presence of systemic risk in the market.

[6] STARR ratio is not a symmetric measure of risk when returns present heavy-tailed distributions, see Martin et al. (2003).

Table 3.1: Average of some statistics of the ex-post returns and the final wealth obtained by different moving average rules combinations

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5%	Final W
1 – 50 with alarm	0.046%	1.083%	-0.1362	7.8801	1.690%	3.063%	4.6002
1 – 150 with alarm	0.052%	1.034%	-0.1297	6.5956	1.643%	2.903%	5.7265
2 – 200 with alarm	0.044%	1.081%	-0.3198	7.5438	1.699%	3.067%	4.1908
5 – 200 with alarm	0.048%	1.070%	-0.3230	7.6047	1.667%	3.032%	4.8585
5 – 100 with alarm	0.050%	1.072%	-0.2737	8.6372	1.643%	3.081%	5.3458
5 – 150 with alarm	0.050%	1.066%	-0.3077	7.7075	1.655%	3.011%	5.2298
10 – 100 with alarm	**0.054%**	1.090%	-0.2499	8.3118	1.656%	3.045%	**6.1948**
10 – 150 with alarm	0.050%	1.066%	-0.2893	7.7514	1.648%	3.003%	5.2609
15 – 100 with alarm	0.053%	1.093%	-0.2046	8.4293	1.656%	3.050%	5.8536
15 – 150 with alarm	0.051%	1.061%	-0.2732	7.7636	1.644%	2.976%	5.5344
20 – 100 with alarm	0.052%	1.078%	-0.2263	8.7559	1.642%	3.030%	5.6249
20 – 150 with alarm	0.052%	1.078%	-0.2690	7.8337	1.643%	2.965%	5.7556
25 – 100 with alarm	0.054%	1.083%	-0.2148	8.6427	1.643%	3.033%	6.0565
25 – 150 with alarm	0.053%	1.075%	-0.2640	7.6521	1.655%	3.002%	5.9502
1 – 50 (no alarm)	0.021%	1.239%	-0.3778	10.459	1.943%	3.509%	1.6678
1 – 150 (no alarm)	0.036%	1.151%	-0.2696	8.0333	1.830%	3.243%	3.0084
2 – 200 (no alarm)	0.034%	1.146%	-0.2892	7.9222	1.820%	3.242%	2.8575
5 – 200 (no alarm)	0.034%	1.152%	-0.2996	8.0004	1.824%	3.270%	2.8138
5 – 100 (no alarm)	0.038%	1.187%	-0.3186	8.4779	1.844%	3.331%	3.1872
5 – 150 (no alarm)	0.041%	1.156%	-0.3106	7.6341	1.821%	3.250%	3.6248
10 – 100 (no alarm)	0.037%	1.218%	-0.3914	9.2551	1.880%	3.419%	3.0101
10 – 150 (no alarm)	0.041%	1.170%	-0.3070	8.0923	1.820%	3.280%	3.6720
15 – 100 (no alarm)	0.037%	1.225%	-0.4697	10.074	1.888%	3.439%	2.9868
15 – 150 (no alarm)	0.041%	1.184%	-0.3654	8.8566	1.840%	3.320%	3.6034
20 – 100 (no alarm)	0.038%	1.218%	-0.3705	9.7320	1.883%	3.409%	3.1943
20 – 150 (no alarm)	0.043%	1.183%	-0.3157	8.7960	1.816%	3.311%	3.9271
25 – 100 (no alarm)	0.040%	1.223%	-0.4209	11.558	1.866%	3.403%	3.4130
25 – 150 (no alarm)	0.043%	1.184%	-0.2993	8.8991	1.812%	3.311%	3.9178
S&P 500	0.017%	1.283%	0.0104	11.191	1.977%	3.561%	1.3917

Table 3.2: Sharpe and STARR ratios of the ex-post returns obtained by different moving average strategies with and without alarm rule

	(5,100) alarm	(5,150) Alarm	(10,150) Alarm	(15,100) alarm	(15,150) alarm	(20,100) alarm	(20,150) alarm	(25,100) alarm	(25,150) alarm
Sharpe	4.682%	4.649%	4.665%	4.832%	4.807%	4.788%	4.917%	4.952%	4.935%
STARR	1.663%	1.646%	1.655%	1.732%	1.714%	1.703%	1.754%	1.768%	1.768%
	(5,100) No alarm	**(5,150) No Alarm**	**(10,150) No Alarm**	**(15,100) No alarm**	**(15,150) No alarm**	**(20,100) No alarm**	**(20,150) No alarm**	**(25,100) No alarm**	**(25,150) No alarm**
Sharpe	3.187%	3.534%	3.536%	2.985%	3.466%	3.142%	3.661%	3.277%	3.652%
STARR	1.135%	1.257%	1.261%	1.063%	1.236%	1.122%	1.308%	1.178%	1.306%
	(10,100) alarm	**(1,150) Alarm**	**(2,200) Alarm**	**(5,200) alarm**	**(10,100) No alarm**	**(1,150) No alarm**	**(2,200) No alarm**	**(5,200) No alarm**	**S&P 500**
Sharpe	4.980%	4.991%	4.055%	4.452%	2.991%	3.116%	3.006%	2.960%	1.324%
STARR	1.782%	1.777%	1.429%	1.571%	1.260%	1.106%	1.063%	1.043%	0.477%

From Tables 3.1 and 3.2 we observe that:

1. The moving average rules used as portfolio strategies without alarms present the ex-post lowest return mean, Sharpe ratio (mean/St. dev.) and STARR performance, but also the highest risk (standard deviation, VaR 5%, CVaR 5%), compared the strategies that use moving average rules as alarms.
2. The strategies that use moving average rules as alarms achieve the greatest average, final wealth, Sharpe ratio, STARR performance, and also the lowest risk (standard deviation, VaR 5%, CVaR 5%) compared to the strategies without alarms and S&P 500 benchmark.
3. The moving average strategies with and without alarms are performing much better than S&P 500 benchmark, which presents the worst results in terms of mean return, final wealth and risk measures (standard deviation, VaR 5%, CVaR 5%).
4. The ex-post returns are strongly leptokurtic for all strategies presented in Table 1. In addition, all strategies expect S&P 500 show some signs of skewness.
5. Overall, we observe that some strategies with moving average rules as alarms are performing much better than others. For example, the rule (10,100) with alarm presents the highest mean return, STARR performance and final wealth, while the strategy (1,150) shows the greatest Sharpe ratio and the lowest CVaR.

Interestingly, these preliminary results give us a general overview about the profitability and usefulness of the use of the moving average rules either as proper portfolio strategies or as systemic risk alarms. However, a further analysis is necessary to compare the effects of the different moving average rules. Therefore, we evaluate and test the observed ex-post dominances between the proposed portfolio strategies. In particular, we examine the ex-post log-returns obtained with different strategies and we check whether there exist stochastic dominance relations between the ex-post log-returns of the optimal portfolios. We test for first-order (FSD), second-order (SSD), Third-order (TSD) and increasing-convex-order (ICX) that accounts for the choice of non-satiable risk-seeking investors (for formal definition and deeper discussion on stochastic dominance see Appendix B). In this thesis, we consider the week form of the stochastic dominance (see, among others, Muller and Stoyan 2002; and Davidson and Jean-Yves 2000). Thus, we use 3779 daily observations (returns realizations), from January 1, 2000 to January 10, 2015, for these stochastic dominance comparisons.

In Table 3.3, we examine whether there are dominance orderings between the optimal portfolios obtained when the moving average rules are used as systemic risk alarms.

According to the stochastic dominance tests of Table 3.3, applied to all strategies with alarms, the optimal portfolio (10,100) as a systemic risk alarm rule dominates most portfolio strategies in the ICX sense. Additionally, we observe that the strategy (1,150) as alarm rule, whereas it is dominated by (10,100) in terms of ICX, dominates most remaining strategies in the SSD sense (and thus also for the TSD). Generally, from Table 3.3, we conclude that some moving average rules are performing much better than others in terms of stochastic dominance test. These results confirm that the choice of the moving average lengths remains very crucial even used as a systemic risk alarm. Furthermore, all optimal portfolio strategies with alarms dominate the S&P 500 benchmark in terms of SSD (and thus also for the TSD). However, the most interesting analysis is to test whether there are dominance orderings between the optimal portfolios obtained by the moving average strategies with and without alarm rules. Table 3.4 contains our results.

Table 3.3: Dominance relations between optimal portfolios obtained applying different strategies with alarms

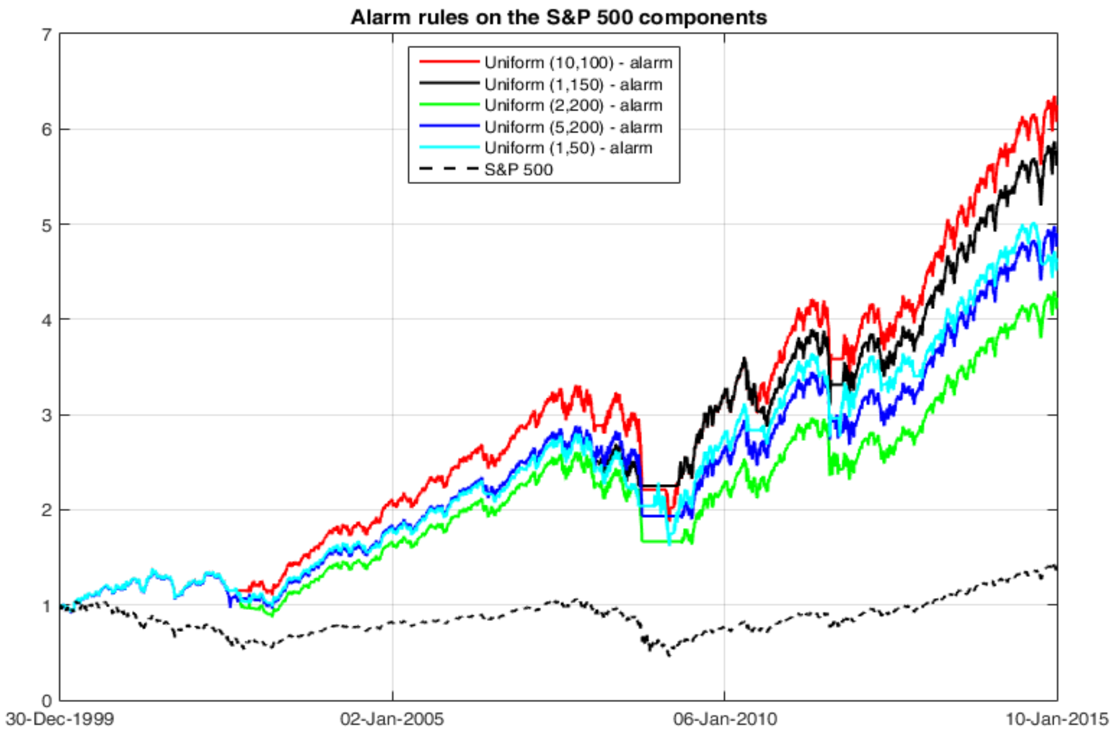
	(1,50) alarm	**(1,150) alarm**	**(2,200) alarm**	**(5,200) alarm**	**(5,100) alarm**	**(5,150) alarm**	**(10,100) alarm**	**(10,150) alarm**	**(15,100) alarm**	**(15,150) alarm**	**(20,100) alarm**	**(20,150) alarm**	**(25,100) alarm**	**(25,150) alarm**	**S&P 500**
(1,50) alarm	**n. c**	≺ SSD	n. c	n. c	n. c	n. c	n. c	n. c	n. c	n. c	n. c	n. c	n. c	n. c	≻ SSD
(1,150) alarm	≻ SSD	**n. c**	≻ SSD	≻ SSD	≻ SSD	≻ SSD	≺ ICX	≻ SSD	≻ SSD ≺ ICX	n. c	≻ SSD	n. c	≺ ICX	n. c	≻ SSD
(2,200) alarm	n. c	≺ SSD	**n. c**	n. c	n. c	n. c	≺ ICX	n. c	≺ ICX	n. c	n. c	n. c	n. c	n. c	≻ SSD
(5,200) alarm	n. c	≺ SSD	n. c	**n. c**	n. c	n. c	≺ ICX	n. c	≺ ICX	n. c	n. c	n. c	n. c	n. c	≻ SSD
(5,100) alarm	n. c	≺ SSD	n. c	n. c.	**n. c.**	n. c	≺ ICX	n. c	≺ ICX	n. c	≺ ICX	≺ SSD	n. c	n. c	≻ SSD
(5,150) alarm	n. c	≺ SSD	n. c	n. c.	n. c	**n. c**	≺ ICX	n. c	≺ ICX	n. c	n. c	n. c	≺ ICX	n. c	≻ SSD
(10,100) Alarm	n. c	≻ ICX	≻ ICX	≻ ICX	≻ ICX	≻ ICX	**n. c**	≻ ICX	n. c	≻ ICX	n. c	≻ ICX	n. c	≻ ICX	≻ SSD
(10,150) Alarm	n. c	≺ SSD	n. c	n. c.	n. c	n. c	≺ ICX	**n. c**	≺ ICX	n. c	n. c	n. c	≺ ICX	n. c	≻ SSD
(15,100) alarm	n. c	≺ SSD ≻ ICX	≻ ICX	≻ ICX	≻ ICX	≻ ICX	n. c	≻ ICX	**n. c**	≻ ICX	n. c	≻ ICX	n. c	≺ SSD	≻ SSD
(15,150) alarm	n. c	n. c	n. c	n. c.	n. c	n. c	≺ ICX	n. c	≺ ICX	**n. c**	n. c	n. c	≺ ICX	n. c	≻ SSD
(20,100) alarm	n. c	≺ SSD	n. c	n. c.	≻ ICX	n. c	n. c	n. c	n. c	n. c	**n. c**	≺ SSD	n. c	n. c	≻ SSD
(20,150) alarm	n. c	n. c	n. c	n. c.	≻ SSD	n. c	≺ ICX	n. c	≺ ICX	n. c	≻ SSD	**n. c**	≺ ICX	n. c	≻ SSD
(25,100) alarm	n. c	≻ ICX	n. c	n. c.	n. c	≻ ICX	n. c	≻ ICX	n. c	≻ ICX	n. c	≻ ICX	**n. c**	n. c	≻ SSD
(25,150) alarm	n. c	n. c	n. c	n. c.	n. c	n. c	≺ ICX	n. c	≻ SSD	n. c	n. c	n. c	n. c	**n. c**	≻ SSD
S&P 500	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	≺ SSD	**n. c**

Table 3.4: Dominance relations between optimal portfolios obtained applying different strategies with and without alarms

	(1,50) No alarm	**(1,150) No alarm**	**(2,200) No alarm**	**(5,200) No alarm**	**(5,100) No alarm**	**(5,150) No alarm**	**(10,100) No alarm**	**(10,150) No alarm**	**(15,100) No alarm**	**(15,150) No alarm**	**(20,100) No alarm**	**(20,150) No alarm**	**(25,100) No alarm**	**(25,150) No alarm**	**S&P 500**
(1,50) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(1,150) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(2,200) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(5,200) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(5,100) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(5,150) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(10,100) Alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(10,150) Alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(15,100) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(15,150) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(20,100) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(20,150) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(25,100) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
(25,150) alarm	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD	$\succ$ SSD
S&P 500	n. c	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	$\prec$ SSD	n. c	$\prec$ SSD	n. c

According to the stochastic dominance tests of Table 3.4, as one could expect, all optimal portfolio strategies with alarms dominate all other portfolio strategies, as far as alarm rule is not used, in the SSD sense. Generally, our results provide strong support for the use of the moving average rules as alarms to detect the presence of systemic risk. To stress this point, Figure 3.1 reports the ex post wealth obtained with the application of different alarm systemic rules considering the following possible combination: (n, N) = [(1, 50), (5, 200), (2, 200), (1, 150), (10, 100)], as suggested by Gencay and Stengos (1998) and Brock et al. (1992).

Figure 3.1: Ex-post wealth obtained with different moving average rules with alarms



Observe that the alarm inserted to detect the presence of systemic risk works well enough since it is able to identify and forecast the largest period of systemic risk of the recent crises, sub-prime crisis 2007-2009 and the European credit risk crisis (fall-winter 2011). Moreover, Figure 3.1 shows that the rule based on (10, 100) presents the highest final wealth, and appears to be the most appropriate strategy since it increases more during the last period of crisis, while the rule (2, 200) was the worst among all. On the one hand, the rule based on (1, 150) has been relatively less affected by the sub-prime crisis 2007-2009 than other trading rules, on the other hand, the rule (10, 100) was comparatively less affected by European credit risk crisis 2010-2011. Furthermore, as expected, all strategies are performing much better than the S&P 500 benchmark.

In order to evaluate the effectiveness of the moving average we go further with the following empirical analysis, where we compare four portfolio strategies. In the first strategy we consider the uniform portfolio with (10, 100) systemic risk alarm rule. The second strategy suggests to invest on all the assets whose average over the last 10 days is greater than the average over the last 100 trading days. Since we want to value the impact of this moving average rule, we do not consider (10, 100) alarm rule for this strategy. In the third strategy we maximize the Sharpe ratio considering the (10, 100) systemic risk alarm rule. The last strategy optimizes the ratio between two expected first passage times − as suggested by Ortobelli et al. (2016) − the expected first time we loss more than 2% and the expected first time we earn more than 20%. Even for this portfolio strategy we consider the (10, 100) systemic risk alarm rule. The *Sharpe* and *Timing* type portfolio strategies can be formalized and described as follows.

***Sharpe ratio (1994)*:** The Sharpe ratio is used to characterize how well the return of an asset compensates the investor for the risk taken. The Sharpe ratio computes the price for unity of risk and is calculated by subtracting the risk-free rate from the rate of return on a portfolio and by dividing the result by the standard deviation of the portfolio returns. Formally:

$$SR(x'R) = \frac{E(x'R) - r_f}{\sigma_{x'R}}, \tag{3.12}$$

where $E(x'R)$ is the portfolio expected return, r_f is the risk-free return and $\sigma_{x'R}$ is the portfolio standard deviation. In Sharpe strategy an *alarm* is inserted to predict the systemic risk.

Timing strategy: Ortobelli et al. (2016) suggests to optimize the average of two first passage times under different distributional assumptions of the wealth Markov process. In this paper, we apply the same algorithm under the assumptions the wealth process follows a nonparametric Markov process approximated by a Markov chain according to Angelelli and Ortobelli (2009). On Markov chain see Ortobelli et al. (2006) and (2007). In this case, it is still possible to compute the distributions of the following stopping times:

$$\tau_d(x) = \inf\{k \in \mathrm{N} \mid W_k(x) \le 0.98\} \wedge T,$$

$$\tau_u(x) = \inf\{k \in \mathrm{N} \mid W_k(x) \ge 1.2\} \wedge T,$$

where $W_k(x)$ is the future wealth at any time $k = 1, 2, ..., T$ (see also Angelelli and Ortobelli, 2009). We consider τ_d the first time the future wealth loses 2% and τ_u the first time the future wealth increases by 20%. Then, at any recalibration time we maximize the following timing portfolio performance:

$$f(\tau_d,\tau_u,T)=\frac{E(\tau_d I_{[\tau_d<T]}+1000I_{[\tau_d\geq T]})}{E(\tau_u I_{[\tau_u<T]}+1000I_{[\tau_u\geq T]})}, \tag{3.13}$$

where $I_{[\omega\in B]}=\begin{cases}1 \text{ if } \omega\in B\\ 0 \text{ otherwise}\end{cases}$ as suggested by Ortobelli et al. (2015). In performance measure (3.13) we penalize 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 (i.e. $\tau_d \geq T$). This timing strategy is implemented with alarm (10, 100) rule to detect systemic risk. As for the first empirical analysis we recalibrate the portfolio every 15 trading days. Thus, at the *k*-th recalibration time, the following steps are performed for Sharpe and Timing strategies:

Step 1 (alarm rule) Verify if the percentage of assets whose mean over the last 10 days is lower than the mean on the last 100 days is higher than the benchmark barrier 75%. If the alarm is not verified, proceed to Step 2, otherwise to Step 3.

Step 2 Compute the optimal portfolio solution $x_M^{(k+1)}$ of the optimization problem:

$$\max_x \rho(x'R)$$
$$s.t.$$
$$\sum_{i=1}^{500} x_i = 1; \quad x_i \geq 0; i=1,...,500,$$

where $\rho(x'R)$ is one of the performance measure (3.12) or (3.13) associated to the portfolio $x'R$. When we use the Sharpe optimization problem

Step 3 Calculate the ex-post final wealth as follows:

$$W_{k+1}=\begin{cases} W_k \text{ } if \text{ } alarm \text{ } applies \\ W_k\left(\left(x_M^{(k+1)}\right)'(1+R_{k+1})\right) otherwise, \end{cases}$$

where R_{k+1} is the ex-post vector of the returns between the *k*-th time and *k*+1-th time.

We apply the algorithm until the observations are available. The results of this analysis are reported in Figure 3.2.

Figure 3.2: Ex-post wealth obtained with Uniform, Sharpe, S&P 500 and Timing strategies

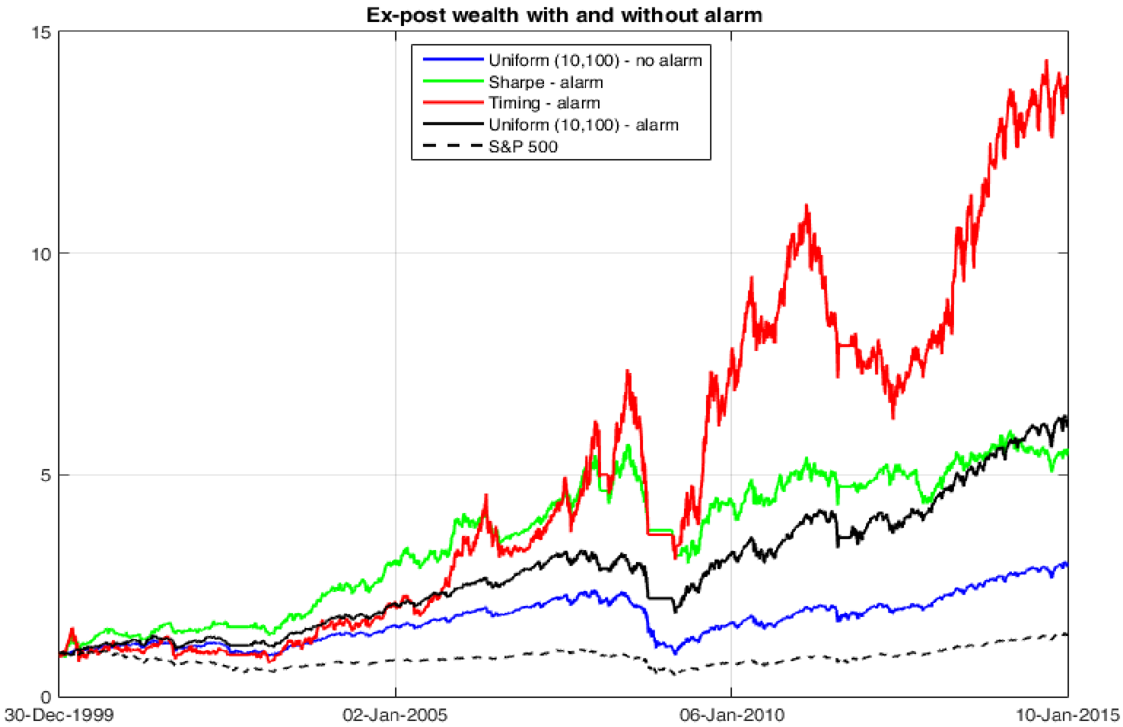


Figure 3.2 reports the ex-post wealth evolution obtained with the four portfolio strategies and the S&P 500 benchmark. Clearly, the Timing portfolio strategy outperforms the strategies based on Sharpe ratio and moving average rule (10, 100). Another remarkable observation is that the Sharpe strategy is also better than the moving average rule strategy applied without alarm. Furthermore, Figure 3.2 shows that during sub-prime crisis 2007-2009 the moving average rule (10, 100) without alarm has a significant loss of 60%, while the moving average rule (10,100) with alarm, Sharpe and timing strategies present losses of 38%, 48% and 57% respectively. However, the same effect does not applies for the European credit risk crisis where we observe a big loss only for the very aggressive Timing strategy, while the other strategies do not present significant losses.

Table 3.5 reports the basic statistics (mean, standard deviation, skewness, kurtosis, VaR 5%, CVaR 5%, final wealth) of the ex-post returns of all strategies of Figures 3.1 and 3.2. Instead, Table 3.6 contains Sharpe ratio and STARR performance of the ex-post returns of all strategies depicted in two Figures.

Table 3.5: Average of some statistics of the ex-post returns plus the final wealth obtained by different strategies

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5%	Final Wealth
2 – 200 rule	0.044%	1.081%	-0.3198	7.8801	1.699%	3.067%	4.1908
5 – 200 rule	0.048%	1.070%	-0.3230	6.5956	1.667%	3.032%	4.8585
1 – 150 rule	0.052%	1.034%	-0.1297	7.5438	1.643%	2.903%	5.7265
1 – 50 rule	0.046%	1.083%	-0.1362	7.6047	1.690%	3.063%	4.6002
10 – 100 rule	0.054%	1.090%	-0.2500	8.6372	1.656%	3.045%	6.1948
Sharpe (with alarm)	0.054%	1.321%	-0.3561	7.7075	2.064%	3.087%	5.5182
Timing (with alarm)	0.092%	2.131%	-0.2696	8.3118	3.319%	5.021%	13.817
10 – 100 (no alarm)	0.036%	1.218%	-0.3909	9.2544	1.879%	2.892%	2.9826
S&P 500	0.017%	1.283%	0.0104	11.191	1.977%	3.561%	1.3917

Table 3.6: Sharpe and STARR ratios of the ex-post returns obtained by different strategies with and without alarm rules

	(10,100) Alarm	(1,150) Alarm	(5,200) Alarm	(1,50) alarm	(2,200) alarm	Timing strategy Alarm	(10,100) No alarm	Sharpe strategy Alarm	S&P 500
Sharpe	4.980%	4.991%	4.452%	4.275%	4.055%	4.335%	2.991%	4.091%	1.324%
STARR	1.782%	1.777%	1.571%	1.512%	1.429%	1.839%	1.260%	1.750%	0.477%

From Tables 3.5 and 3.6 we observe that:

1. S&P 500 benchmark presents the ex-post lowest return mean, final wealth, Sharpe ratio and STARR performance.
2. The moving average rule (10,100) used as a portfolio strategy without alarm presents the ex-post lowest return mean, Sharpe ratio and STARR performance compared to other portfolio strategies with alarms.
3. The strategy, that invest 0.2% in each asset once that the (1, 150) alarm rule applies, presents the lowest risk (standard deviation, VaR 5% and CVaR 5%) and the highest Sharpe ratio.
4. Timing strategy achieves the greatest average and STARR performance, but also the highest risk (standard deviation, VaR 5%, CVaR 5%).

Conceptually, Timing strategy depends crucially on the parameter of stopping time. In this study, we follow Ortobelli et al. (2016), however, it would be interesting to examine the results for some others values of theses parameters. Thus, we consider three different symmetrical cases, the results of this analysis are reported in Table 3.7

Table 3.7: Average of some statistics of the ex-post returns plus the final wealth obtained by timing strategy with different parameters of stopping times

Timing	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5 %	Final W	Sharpe	STARR
(0.9-1.1)	0.102%	2.218%	-0.3204	3.4918	3.570%	5.230%	18.112	4.591%	1.947%
(0.95-1.05)	0.103%	2.057%	-0.3486	3.5866	3.230%	4.848%	21.772	5.018%	2.129%
(0.98-1.02)	0.077%	1.805%	-0.3052	3.2702	2.930%	4.308%	9.8804	4.292%	1.798%

Overall, from Table 3.7, we observe that Timing strategy presents the greatest average finale wealth, Sharpe and STARR performance, but also the highest risk (standard deviation, VaR 5%, CVaR 5%). In particular, the best performing strategy is the one that considers the first time the future wealth loses 5% and the first time the future wealth increases by 5%.

In summary, these results give us a nice overview about the profitability and usefulness of the use of the moving average method as a systemic risk alarm rule. For further confirmation, as before, we examine whether there are dominance orderings between the optimal portfolios obtained with the moving average rules and the ones obtained with the Sharpe and Timing strategies with systemic risk alarm rule. Table 3.8 summarizes our tests.

Table 3.8: Dominance relations between optimal portfolios obtained applying different strategies with and without alarms

Optimal portfolios	Timing strategy Alarm	(10,100) Alarm	Sharpe strategy Alarm	(10,100) No alarm	S&P 500
Timing strategy Alarm	**n. c.**	≻ICX	≻ICX	≻ICX	≻ICX
(10,100) Alarm	≺ICX	**n. c.**	≻SSD	≻SSD	≻SSD
Sharpe strategy Alarm	≺ICX	≺SSD	**n. c.**	n. c.	n. c.
(10,100) No alarm	≺ICX	≺SSD	n. c.	**n. c.**	≻SSD
S&P 500	≺ICX	≺SSD	n. c.	≺ SSD	**n. c.**

The Timing strategy dominates all strategies presented in Figure 3.2 in the ICX sense. As we could expect, the moving average rule (10,100) is dominated by the rest of the strategies as far as the moving average method is not used as a systemic risk alarm rule. However, the strategy becomes comparable in the stochastic ordering sense as long as moving average is used as a systemic risk alarm rule. Indeed, according to the stochastic dominance tests of Table 3.8, we observe that the rule (10,100) with alarm dominate Sharpe strategy with alarm, the (10,100) without alarm and S&P 500 benchmark in terms of SSD. Thus, Tables 3.2, 3.3 and 3.8 confirm with the proper tests the dominance observed in Figures 3.1 and 3.2. This fact strengthens the hypothesis that the moving average rule cannot be used as a profitable strategy, but rather than as a useful tool to detect the presence of systemic risk.

Overall results, we deduce that the moving average rules are much more effective when used as alarms to detect the presence of systemic risk.

Further research could involve theoretical and empirical studies. On the one hand, investors may employ complex versions of the moving average rules. On the other hand, the impact of calendar periods such as the weekend effect, the turn-of-the-month effect, the holiday effect and the January effect. Future research will investigate this aspects. Another promising direction for future research is to consider other technical indicators, which may be easier to detect algorithmically, to examine whether or not such indicators are able to predict the presence of systemic risk.

3.4 Conclusions

In this chapter, we provide some theoretical motivations behind the use of the moving average rule as trading strategy. In particular, we demonstrate that under some technical assumptions the probability to be in up-trend is greater than the probability to be in down-trend. For this reason, we propose to use moving average rules to predict periods of systemic risk. Thus, we examine the impact of the moving average rules on the U.S. stock market. Firstly, a comparison among different moving average trading rules with and without alarms of losses is performed. Secondly, we compare the ex post wealth obtained with the best performing systemic risk rule used as trading strategy with the wealth obtained maximizing two different portfolio performances. From the comparison among different strategies and stochastic dominance tests, we deduce that the best use of the moving average rules is obtained to predict periods of market distress. These empirical analyses suggest that the moving average rules are much more effective and performing when used to detect the presence of systemic risk.

Chapter 4

On the valuation of the arbitrage opportunities and the SPD estimation

In this chapter, we present different approaches to evaluate the presence of the arbitrage opportunities in the market. In particular, we investigate empirically the well-known put-call parity no-arbitrage relation and the SPD. First, we measure the violation of the put-call parity as the difference in implied volatilities between call and put options. Then, we discuss the usefulness of the nonnegativity of the SPD. We evaluate the effectiveness of the proposed approaches by an empirical analysis on S&P 500 index options data. Moreover, we propose alternative approaches to estimate the SPD under the classical hypothesis of the BS model. To this end, we use the classical nonparametric estimator based on kernel and a recent alternative the so called OLP estimator that uses a different approach to evaluate the conditional expectation consistently.

The remainder of this chapter is structured as follows. The first part describes some theoretical properties of two approaches. In particular, section 4.1.1 focuses on local polynomial estimator, while section 4.1.2 introduces the conditional expectation estimator (OLP). The second part presents alternative methods to evaluate the arbitrage opportunities and describes the procedure followed towards estimating the SPD. The rest of this part is organized as follows. Section 4.2 presents some methods to evaluate the arbitrage opportunities. Section 4.2.1 illustrates the first empirical analysis. Section 4.3 proposes alternative methods to estimate the SPD. Section 4.3.1 includes the second empirical analysis. Concluding remarks are contained in Section 4.4

4.1 Nonparametric estimation

Regression analysis is surely one of the most suitable and widely used statistical techniques. In general, it explores the dependency of the so-called dependent variable on one (or more) explanatory or independent variables.

$$Y = E(Y \mid X = x) + \varepsilon = g(x) + \varepsilon \tag{4.1}$$

It is well known that, if we know the form of the function $g(x) = E(Y \mid X = x)$, (e.g. polynomial, exponential, etc.), then we can estimate the unknown parameters of $g(x)$ with several methods (e.g. least squares). In particular, if we do not know the general form of $g(x)$, except that it is a continuous and smooth function, then we can approximate it with a nonparametric method, as proposed by E. A. Nadaraya (1964) and G. S. Watson (1964). The aim of nonparametric technique is to relax assumptions on the form of regression function, and allows data search for an appropriate function that represent well the available data, without assuming any specific form of the function. Thus, $g(x)$ can be estimated by:

$$\hat{g}_n(x) = \frac{\sum_{i=1}^{n} y_i k\left(\frac{x - x_i}{h(n)}\right)}{\sum_{i=1}^{n} k\left(\frac{x - x_i}{h(n)}\right)}, \tag{4.2}$$

where, $k(\cdot)$ is a density function such that: i) $k(x) < \mathrm{C} < \infty$, ii) $\lim_{x \to \pm\infty} |xk(x)| = 0$, iii) $h(n) \to 0$ when $n \to \infty$. *h* is a bandwidth, also called a smoothing parameter, which controls the size of the local averaging. The function $k(x)$ is denoted by the *kernel;* observe that kernel functions are generally used for estimating probability densities nonparametrically (see for instance V. A. Epanechnikov (1965)).

It was proved in E. A. Nadaraya (1964) that if *Y* is quadratically integrable then $\hat{g}_n(x)$ is a consistent estimator for $g(x)$. In particular, observe that, if we denote by $f(x, y)$ the joint density of (X, Y), the denominator of (4.2) converges to the marginal density of *X* $\int f(x, y)dy$, while the numerator converges to the function $\int yf(xy)dy = \int_{-\infty}^{\infty} \int_{\{X=x\}} yP(dx, dy)$. Please note that, if *X* is continuous, the function $\int_{\{X=x\}} yP(dx, \mathrm{dy}) / \int_{-\infty}^{\infty} \int_{\{X=x\}} P(dx, dy)$ has to be intended as a regular conditional probability. An overview of nonparametric regression or smoothing techniques may be found, e.g., in Härdle (1990); Simonoff (1996); Fan and Gijbels (1996); Härdle et al. (2004). In the next section, one poplar type of nonparametric estimation techniques is presented, the so called local polynomial estimation.

4.1.1 Local polynomial regression

For the sake of clarity, consider a random sample $\left(t_i, Y_i\right)$ $i = 1,....,n$. A regression model can be presented as follows:

$$Y_i = X(t_i) + \varepsilon_i \text{ for } i = 1,....,n \tag{4.3}$$

where, $\boldsymbol{Y} = (Y_1,....,Y_n)$ are dependent variables, $\boldsymbol{t} = (t_1,....,t_n)$ are explanatory variables, X is the regression function, and ε_i are i.i.d random variable with zero mean and $\text{var}(\varepsilon_i) = \sigma(t_i)$, $i = 1,....,n$. In local polynomial estimation, a lower-order Weighted Least Square (WLS) is fit at each point of interest t using some data from its neighborhoods. Assume that function X has $(p+1)^{th}$ derivatives, and then it is possible to use Taylor expansion to approximate the regression function X at point t_i as:

$$X(t_i) \approx X(t) + X^{(1)}(t)(t - t_i) + ... + X^{(P)}(t)(t - t_i)^p \frac{1}{p!} \tag{4.4}$$

where, $X^{(j)}$ stands for j^{th} derivative of X and $(t - t_i)^{p+1}$ is order of approximation error. In practice it possible to estimate these terms using WLS by solving for $\alpha_l = \dfrac{X^{(l)}(t)}{l!}, l = 0,..., p$. Hence the estimate of regression function X at the point t is given by minimization of the following criterion:

$$\hat{\alpha}(t,h,p,k) = \arg\min_{\alpha} \sum_{i=1}^{n} \left[Y_i - \sum_{J=0}^{p} \alpha_j (t_i - t)^j \right]^2 k\left(\frac{t_i - t}{h}\right), \tag{4.5}$$

where, the weighting function $k(\cdot)$ is *kernel* function and h is the bandwidth controlling the size of the local averaging. According to (4.5) there are three parameters that may have direct impact on the quality of the fit. Mainly, the bandwidth h, the local polynomial order p, and the kernel function k (see below). It clear form WLS estimation (4.5) and Taylor expansion (4.4) that $\alpha_0 = \hat{X}(t)$, where $\hat{X}(t)$ is an estimate of the regression function at the point t, furthermore:

$$\hat{X}^{(v)}(t) = v!\hat{\alpha}_v(t,h,p,k), \tag{4.6}$$

where, $\hat{\alpha}_v$ is an estimate of the *vth* derivative of the function X. Thus, the local polynomial estimator provides us not only an estimate of the function X but also its derivatives up to the order p. A special case of this estimator is when $p = 0$. Indeed in this case we obtain (4.2), which is also known as Nadaraya-Watson (kernel) estimator. For a complete discussion of the theoretical properties of this estimator we refer to Fan and Gijbels (1996).

Clearly, it is usually simpler to work with matrix notation. In this context, (4.5) becomes:

$$\hat{\alpha} = \arg\min_{\alpha} (\boldsymbol{Y} - \boldsymbol{t}_{t,n,p}\alpha)^{T} \mathbf{W}_{t,n,p,h,k} (\boldsymbol{Y} - \boldsymbol{t}_{t,n,p}\alpha), \tag{4.7}$$

where, $\boldsymbol{Y} = (Y_1, ..., Y_n)^T$, $\alpha = (\alpha_0, ..., \alpha_p)^T$ while $\boldsymbol{t}_{t,n,p}$ and $\mathbf{W}_{t,n,p,h,k}$ are as follow:

$$\boldsymbol{t}_{t,n,p} = \begin{pmatrix} 1 & (t_1 - t) & (t_1 - t)^2 \cdots & \\ \vdots & \vdots & & \ddots \\ 1 & (t_n - t) & (t_n - t)^2 \cdots & \end{pmatrix} \tag{4.8}$$

$$\mathbf{W}_{t,n,p,h,k} = \text{diag}\left\{ k\left(\frac{t_1 - t}{h}\right), ..., k\left(\frac{t_n - t}{h}\right) \right\}. \tag{4.9}$$

Definitively, from (4.5) it is clear that higher order polynomials yield higher computational costs and typically the choice $p = v + 1$ is preferred, for detailed discussion see Fan and Gijbels (1996).

Concerning the parameter that may have some effect on the quality of polynomial regression, it is documented that large h increase the bias and decrease the variance, while small h has the opposite effect. In general, the choices of the bandwidth-selection methods are based on the balancing of the bias and variance. Since bandwidth plays an important role in the practical usage of the local polynomials, several automated selection rules have been proposed. Mainly there are two approaches in literature. The first method is the so called global bandwidth choice, a bandwidth valid for all points $t \in [0,1]$, typically we choose the bandwidth by minimizing MISE of the estimate, for an overview see Härdle et al. (2004). A more sophisticated approach is known as local bandwidth choice, essentially it is chosen by minimizing the MSE individually for each t where the estimate is constructed; see Härdle (1990), Spokoiny (2006). For the bandwidth selection within no arbitrage argument and option implied volatility, see Kopa and Tichý (2014).

Another question concerns the choice of the kernel function k. It has been shown that the choice of kernel function has not a crucial role in the practice, see among others Härdle et al. (2004). Normally a probability density function is used, some common choices are:

- Uniform kernel $k(u) = \frac{1}{2}\mathbf{1}(|\boldsymbol{u}| \le 1)$,
- Epanechnikov kernel $k(u) = \frac{3}{4}(1 - u^2)\mathbf{1}(|\boldsymbol{u}| \le 1)$,
- Quadratic kernel $k(u) = \frac{15}{16}(1 - u^2)^{\mathbf{2}}\mathbf{1}(|\boldsymbol{u}| \le 1)$,
- Gaussian Kernel $k(u) = (2\pi)^{-1/2} \exp\left(\frac{-u^2}{2}\right)$.

The third issue in local polynomial regression is the choice of the order of the local polynomial. In practice for a given bandwidth h, a large value of p would cause a large variance and a considerable computational cost, but it would reduce the modeling bias. Since the bandwidth is used to control the size (complexity), it is recommended to use the lowest odd order, i.e. $p = v+1$, or sometimes $p = v+3$.

To conclude this section we note that local polynomial estimators belong to an interesting class of smoothing methods. It well known that smoothing methods can be written as weighted local average of responses variables:

$$\hat{X}(t) = \sum_{i=1}^{n} w_i(t) Y_i. \tag{4.10}$$

A clear example could be the Nadaraya-Watson -smoothers (4.2) where the weights $w_i(t)$ can be written as:

$$w_i(t) = \frac{k\left(\frac{t-t_i}{h}\right)}{\sum_{i=1}^{n} k\left(\frac{t-t_i}{h}\right)}. \tag{4.11}$$

In this context, the local polynomial estimator $\hat{\alpha}_v$ can be given as:

$$\hat{\alpha}_v = \sum_{i=1}^{n} w_v\left(\frac{t_i - t}{h}\right) Y_i, \tag{4.12}$$

where, $w_v^T(u) = e_{v+1}^T \left(\boldsymbol{t_{t,n,P}}^T \boldsymbol{W_{t,n,p,h,k}} \boldsymbol{t_{t,n,P}}\right)^{-1} \{1, uh, \dots, (uh)^p\}^T \, k(u)/h,$ for $v = 1, \dots, p$. Please note that local polynomial estimator can be expressed in the form of the so-called equivalents kernel. In particular formula (4.12) shows that the estimator $\hat{\alpha}_v$ is very much like a standard kernel estimator except that the 'kernel' w_v^T depends on the locations and design points, for more details see Fan and Gijbels (1996), section 3.2.2.

The idea of the local polynomial estimation can be extended to multi-dimensional regression problems in straightforward way. Many of their properties have been rigorously investigated and are well understood, see, for example, Fan and Gijbels (1996), Gyorfi et al. (2002) and Tsybakov (2009).

4.1.2 Conditional expectation estimators

The kernel nonparametric regression method allows estimating the regression function $g(x)$, which is a realization of the conditional expectation $E(Y \mid X)$. A recent alternative approach consists in estimating the conditional expectation (intended as a random variable), based on an appropriate approximation of the σ-algebra generated by *X*. In this section, we present the procedure of estimating the distribution of the conditional expectation based on the kernel method, so that it is possible to compare the two approaches by verifying which one better estimates the true distribution of $E(Y \mid X)$. In particular, if we assume that the two-dimensional variable $E(Y \mid X)$ is normally distributed, then the true distribution of $E(Y \mid X)$ can be computed quite easily, and the comparison can be performed in terms of goodness-of-fit tests.

We now describe an alternative nonparametric approach, see Ortobelli et al (2015), for approximating the conditional expectation; the method is denoted by "OLP", which is an acronym of the authors' names.

Define by $\mathfrak{I}_X$ the σ-algebra generated by X (i.e.$\mathfrak{I}_X = \sigma(X) = X^{-1}(\mathcal{B}) = \{X^{-1}(B): B \epsilon \mathcal{B}\}$, where $\mathcal{B}$ is the Borel σ-algebra on ℝ). Observe that the regression function is just a "pointwise" realization of the random variable $E(Y \mid \mathfrak{I}_X)$, which can equivalently be denoted by $E(Y \mid X)$. The following methodology is aimed at estimating $E(Y \mid X)$ rather than function $g(x)$. For this reason, we propose the following consistent estimator of the random variable $E(Y \mid X)$.

Let $X: \Omega \to \mathbb{R}$ and $Y: \Omega \to \mathbb{R}$ be integrable random variables in the probability space $(\Omega, \mathfrak{I}, P)$ and define by $\mathfrak{I}_X$ the σ-algebra generated by *X*. Notice that: $E(Y \mid X)$ is equivalent to $E(Y \mid \mathfrak{I}_X)$. We can approximate $\mathfrak{I}_X$ with a σ-algebra generated by a suitable partition of Ω. In particular, for any $k \in \mathbb{N}$, we consider the partition $\{A_j\}_{J=1}^{b^k} = \{A_1, \dots, A_{b^k}\}$ of Ω in b^k subsets, where *b* is an integer number greater than 1 and:

- $A_1 = \{\omega: X(\omega) \le F_X^{-1}\left(\frac{1}{b^k}\right)\}$,
- $A_h = \{\omega: F_X^{-1}\left(\frac{h-1}{b^k}\right) < X(\omega) \le F_X^{-1}\left(\frac{h}{b^k}\right)\}$, for $h = 2, \dots, b^k - 1$,
- $A_{b^k} = \Omega - \bigcup_{J=1}^{b^k-1} A_j = \{\omega: X(\omega) > F_X^{-1}\left(\frac{b^k-1}{b^k}\right)\}$.

Thus, starting with the trivial σ-algebra $\mathfrak{I}_0 = \{\emptyset, \Omega\}$, we can generate a sequence of σ-algebras generated by these partitions obtained by varying k ($k = 1, \dots, m, \dots$). Thus, $\mathfrak{I}_1 = \sigma\{\emptyset, \Omega, A_1, \dots, A_b\}$ is the σ-algebra generated by $A_1 = \{\omega: X(\omega) \le F_X^{-1}(1/b)\}$, $A_s = \left\{\omega: F_X^{-1}\left(\frac{s-1}{b}\right) < X(\omega) \le F_X^{-1}\left(\frac{s}{b}\right)\right\}$, $s = 1, \dots, b-1$ and $A_b = \{\omega: X(\omega) > F_X^{-1}((b-1)/b)\}$, moreover:

$$\mathfrak{I}_k = \sigma\left(\{A_j\}_{J=1}^{b^k}\right), k \in \mathbb{N} \tag{4.13}$$

Proposition 4.1 *Given the sequence of σ-algebras* $\{\mathfrak{I}_k\}_{k \in \mathbb{N}}$ *defined above*:

$$E(Y|X) = \lim_{k\to\infty} E(Y|\Im_k), \quad \text{a.s.} \tag{4.14}$$

where , $E(Y|\Im_k)(\omega) = \sum_{j=1}^{b^k} E(Y|A_j)1_{A_j}(\omega)$ a.s. *and* $1_{A_j}(\omega) = \begin{cases} 1 & \omega \in A_j \\ 0 & \omega \notin A_j \end{cases}$.

Proof:

Observe that the increasing sequence of simple functions (i.e. $s_k \leq s_{k+1}$):

$$s_k(w) = \sum_{j=1}^{b^k} F_X^{-1}\left(\frac{j-1}{b^k}\right) 1_{A_j}(\omega) \tag{4.15}$$

converges to X almost surely, i.e. $X = \lim_{k\to\infty} s_k$ a.s.. Moreover, the sequence of σ-*algebras* $\{\Im_k\}_{k\in\mathbb{N}}$ is a filtration and $\Im_X = \sigma(\bigcup_{k\in\mathbb{N}} \Im_k)$, because the σ-algebra generated by s_k is $\Im_k$ and $\Im_k \subset \Im_{k+1}$ for $k \in \mathbb{N}$. According to Ortobelli et al. (2015), the equality $E(Y|X) = \lim_{k\to\infty} E(Y|\Im_k)$ holds, since the family of random variables $E(X|\Im_k)$ is uniformly integrable. Therefore, using the definition of conditional expectation, we can easily verify that $E(Y|\Im_k)$ is defined by

$$E(Y \mid \Im_k)(\omega) = \sum_{j=1}^{b^k} \frac{1_{A_j}(\omega)}{P(\mathrm{A}_j)} \int_{A_j} YdP,$$

because $E(Y|\Im_k)$ is the unique $\Im_k$-measurable function such that for any set $A \in \Im_k$ (that can be seen as a union of disjoint sets, in particular $A = \bigcup_{A_j \subseteq A} A_j$) we obtain the equality

$$\int_A E(Y \mid \Im_k) dP = \sum_{j=1}^{b^k} \frac{\int_{A_j} YdP}{P(\mathrm{A}_j)} \int_A 1_{A_j}(\omega) dP(\omega) =$$

$$= \sum_{A_j \subseteq A} \int_{A_j} YdP = \int_A Y(\omega) dP(\omega) \tag{4.16}$$

Since $E(Y \mid X) = \frac{1}{P(\mathrm{A}_j)} \int_{A_j} YdP$, we obtain $E(Y \mid \Im_k)(\omega) = \sum_{j=1}^{b^k} E(Y \mid A_j) 1_{A_j}(\omega)$ Q.E.D.■

When b is large enough, even $E(Y|\Im_1)$ can be a good approximation of the conditional expected value $E(Y|X)$ because from equation (4.15) we get $X = \lim_{b\to\infty} s_1$ a.s. On the one side, given *N* i.i.d. observations of *Y*, we get that $\frac{1}{n_{A_j}} \sum_{y\in A_j} y$ (where n_{A_j} is the number of elements of A_j) is a consistent estimator of $E(Y|A_j)$. On the other side, if we know that the probability p_i is the probability of the *i*-th outcome y_i of random variable *Y*, we get $E(Y|A_j) = \sum_{y_i\in A_j} y_i p_i / P(A_j)$, Otherwise, we can give uniform weight to each observation, which yields the following consistent estimator of $E(Y|A_j) = \frac{1}{n_{A_j}} \sum_{y_i\in A_j} y_i$, where n_{A_j} is the number of elements of A_j. Therefore, we are able to estimate $E(Y|\Im_k)$, that is a consistent estimator of the conditional expected value $E(Y|X)$ as a consequence of Proposition 4.1.

4.1.2.1 Comparison in case of normality

If we assume that X and Y are jointly normally distributed, i.e. $(X,Y)\sim N(\mu_X,\mu_Y,\sigma_X,\sigma_Y,\rho)$, we can obtain the distribution of the random variable $E(Y|X)$ quite easily. Indeed, we know that:

$$g(x) = E(Y|X = x) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(x - \mu_X), \tag{4.17}$$

therefore, as $X\sim N(\mu_X,\sigma_X)$, we obtain that:

$$E(Y|X) = \mu_Y + \rho \frac{\sigma_Y}{\sigma_X}(X - \mu_X)\sim N(\mu_Y,|\rho|\sigma_Y). \tag{4.18}$$

Of course, if we simulate data from (X,Y) and approximate $E(Y|X)$ with the estimator $E(Y|\Im_k)$ defined in (4.14), we can finally compare the true and the theoretical (estimated) distribution by performing a goodness-of-fit test. Differently, the kernel nonparametric regression method does not allow to estimate $E(Y|X)$, but only yields a consistent estimator of $g(x)$. However, assume that the random variable X' is independent from X and moreover $X =_d X'$ (that is, $X'\sim N(\mu_X,\sigma_X)$ and $\rho(X,X') = 0$): in this case we can estimate $E(Y|X')$ with

$$g_n(X') = \frac{\sum_{i=1}^{n} y_i k\left(\frac{X'-x_i}{h(n)}\right)}{\sum_{i=1}^{n} k\left(\frac{X'-x_i}{h(n)}\right)}, \tag{4.19}$$

and thereby we can also estimate the distribution of $E(Y|X)$, because $E(Y|X') =_d E(Y|X)$. Obviously, the estimate depends on the choice of the kernel function k. It is proved that $E(Y|\Im_k)$ converges almost surely to $E(Y|X)$ i.e. ($E(Y|\Im_k) \to_{a.s.} E(Y|X)$). Moreover, note that also $g_n(X')$ satisfies a weaker convergence property (convergence in distribution). Indeed, we have that

$$g_n(X') \to_{a.s.} E(Y|X') =_d E(Y|X), \tag{4.20}$$

thus we obtain that $g_n(X') \to_d E(Y|X)$.

Finally, it is possible to compare the two methods by verifying which one better estimates the distribution of $E(Y|X)$. Without significant loss of generality, the mathematical notation changes in the next section (the distinction of the variables will always be clear from context). Interpret Y as S_T, while X as S_t.

4.2 Methods to evaluate the arbitrage opportunities

4.2.1 Black and Scholes methodology

Fisher Black and Myron Scholes (1973) achieved a major breakthrough in European option pricing. In this model we assume that the price process follows a standard geometric Brownian motion defined on filtered probability space $(\Omega, \Im, P, \{\Im_t\}_{t\geq 0})$, where $\{\Im_t\}_{t\geq 0}$ is the natural filtration of the process completed by the null sets. Under these assumptions we know that $\mathrm{E}(S_T \mid \Im_t) = \mathrm{E}(S_T \mid S_t)$ as consequence of Markovian property. The model of stock price behavior used is defined as:

$$dS = \mu S dt + \sigma S dB, \tag{4.21}$$

where, μ is the expected rate of return, σ is the volatility of stock return and B denotes a standard Brownian motion. Under this hypothesis we know that the log price is normally distributed:

$$\ln S_T \sim \phi\left[\ln S_0 + \left(\mu - 0.5\sigma^2\right)T, \sigma^2 T\right], \tag{4.22}$$

where, S_T is the stock price at future time T, S_0 is the stock price at time 0 and ϕ denotes a normal distribution. Please note that μ in equation (4.21) represents the expected rate of return in real world, while in BS model (risk neutral world) it becomes risk-free rate r.[7]

4.2.2 Put-call parity

We recall an important relationship between the prices of European put and call options that have the same strike price and the same time to maturity. This relationship is known as put-call parity, see Stoll (1969). In particular, it shows that the value of a European call option with a certain strike price and expiration date can be deduced from the value of a European put option with the same strike price and expiration date, and vice versa. Formally, in perfect markets, the following equality must hold for European options on non-dividend-paying stocks:

$$C - P = S_0 - Ke^{-rt}, \tag{4.23}$$

where, S_0 is the current stock price, C and P are the call and put prices, respectively, that have the same strike price K, the same expiration date and the same underlying asset.

To illustrate the arbitrage opportunities when equation (4.23) does not hold, we measure the violation of put-call parity as the difference in implied volatility between call and put options that have the same strike price, the same expiration date and the same underlying asset. In this context, it is well known that the BS model satisfies put-call parity for any assumed value of the volatility parameter σ. Hence,

---

[7] For more details about BS assumptions we refer to Hull (2015)

$$C^{BS}(\sigma) - P^{BS}(\sigma) = S_0 - Ke^{-rt}, \qquad \forall \sigma > 0, \tag{4.24}$$

where, $C^{BS}(\sigma)$ and $P^{BS}(\sigma)$ denotes BS call and put prices, respectively, as a function of the volatility parameter σ . At this point, from equation (4.23) and (4.24) we can deduce that:

$$C^{BS}(\sigma) - P^{BS}(\sigma) = C - P \qquad \forall \sigma > 0, \tag{4.25}$$

By definition, the implied volatility (IV) of a call option (IV^{call}) is that value of the volatility of the underlying asset, which matches the BS price with the price actually observed on the market. In formal way:

$$C^{BS}(IV^{call}) = C, \tag{4.26}$$

Now, it is straightforward form equation (4.25) that:

$$P^{BS}(IV^{call}) = P, \tag{4.27}$$

this in turn implies that:

$$IV^{call} = IV^{put}. \tag{4.28}$$

Put-call parity holds only for European options. Thus, for this type of options, put-call parity is equivalent to the statement that the BS implied volatilities of pairs of call and put options must be equal. Therefore, any violation of put-call parity may contain useful information about the presence of tradable arbitrage opportunities. No attempt will be made to formulate the case of American option, which beyond the scope of this study. However, it possible to derive some results for American options price, where put-call parity takes the form of an inequality.

In this study, we will carry the analysis on the European options style. Since put-call parity is one of the best known no-arbitrage relations, we use the difference in implied volatility between pairs of call and put options in the spirit of equation (4.28) in order to detect the presence of arbitrage opportunities in the market. Intuitively, lower call implied volatilities relative to put implied volatilities means that calls are less expensive than puts, and lower put implied volatilities with respect to call implied volatilities suggest the opposite.

We compute the difference in implied volatilities between call and put options that have the same strike price, the same maturity and are written on the same underlying asset. Hence, we refer to such difference as volatility spread (VS) which may represent a valid indicator of the presence of arbitrage opportunities in the market, especially close to at-the-money options. Formally, given call and put options with the same strike price and expiration date, we compute the VS as:

$$VS = \max | IV^{call} - IV^{put} | \tag{4.29}$$

Of course, higher volatility spread is a significant indicator of arbitrage opportunities since put-call parity is a fundamental relation of no-arbitrage. A simple example illustrates intuitively this result.

Example: Consider a put option on S&P 500 index with strike price $K = 2200$ and has 6 months to maturity. The current underlying asset price is $S_0 = 2100$ and the 6-month risk free rate of return is $r = 0.08\%$. Let us assume that the price of this put option is $P = 160$ and the price of the call option on the S&P 500 index with the same strike price and the same maturity is $C = 120$. It is very simple to verify that the put-call parity does not hold and that the volatility of call option is greater than the volatility of the put option. Indeed,

$$P + S_0 = 2240 < C + Ke^{-rT} = 2299.1,$$

$$IV^{call} = 0.2723,\ IV^{put} = 0.1699 \text{ and } VS = 0.1024$$

Arbitrage position: Buy the put option at $P = 160$ and the stock at $S_0 = 2100$, then sell the call option at $C = 120$. To finance this position, borrow:

$$D = P + S_0 - C = 160 + 2100 - 120 = 2140 \text{ at } r = 0.08\%.$$

Payoff to this arbitrage position:

- If $S_T < 2100$, the trader exercises the put option and the payoff is:

$$(K - S_T) + S_T - De^{rT} = 59.14$$

- If $S_T > 2100$, the short call option exercised and the payoff is:

$$S_T - (S_T - K) - De^{rT} = 59.14$$

In both cases, the trader ends up with a payoff of 59.14 and selling the stock at $K = 2200$.

This example illustrates the situation when the call implied volatility is greater than the put implied volatility, such that the option has the same strike price, the same maturity and is written on the same underlying asset. On the opposite, lower call implied volatility relative to put implied volatility means that call option is less expensive than put option. Therefore, one may follow the simplest strategy that involves buying the call option and shorting both the put option and the stock.

The efficacy of this theoretical arbitrage mechanism in maintaining put and call price parity will be examined empirically. However, several papers argue that violations of the put-call parity can be justified via the short sale constraint, data-related issues or even the payment of dividend streams, see among others Ofek et al (2004). To overcome these issues and to have a valid confirmation of this approach we could combine the IV smoothing with SPD estimation, which requires some properties in order to be consistent with no-arbitrage argument. In particular, the nonnegativity property of SPD since its negative values immediately corresponds to the possibility of the arbitrage opportunities in the market. For complete treatment of this method we refer the reader to a relatively conservative approach adopted by Benko et al. (2007).

4.2.3 First empirical analysis

In this section, we report numerical experiments obtained using the methods introduced to detect the presence of arbitrage opportunities in the market. To evaluate the empirical importance of these techniques and the corresponding SPD estimate, we present some applications to the S&P 500 index using daily data obtained from DataStream for the sample period December 26, 2012 to May 13, 2015. Of course, S&P 500 Index options are among the most actively traded financial derivatives in the world.

In the first empirical application to S&P 500 index options we present the analysis concerning the estimation of IVs. For this purpose we use as dataset all options listed on May 13, 2015. The options are European style and the average daily volume during the sample day was 82.65 and 179.01 contracts for call and put respectively. Strike price is at 130 percent and barrier at 70 percent of the underlying spot price at 2098.48, while strike price intervals are 5 points. During sample period, the mean and standard deviation of continuously compounded daily returns of the S&P index are 1.078 percent and 11.268 percent, respectively. Throughout this period short-term interest rates exhibit a very low level. They range from 0.01 percent monthly to 0.89 percent in almost three years. The options in our sample vary significantly in price and terms, for example the time-to-maturity varies from 2 days to 934 days.

The row data present some challenges that must be addressed. Clearly, in-the-money (ITM) options are rarely traded relative to at-the -money (ATM) and out-the-money (OTM) options. For example, the average daily volume for puts that are 25 points OTM is 2553 contracts, in contracts, the volume for puts that are 25 points ITM is 2. This can be justified by the strong demand of portfolio managers for protective puts.

In this analysis, we consider different kernel functions and different choices of bandwidth selection. It was shown that the violation of arbitrage-free conditions heavily depends on all these settings. Theoretically, there has been major progress in recent years in data-based bandwidth selection for kernel density estimation (for a more complete treatment, from a historical viewpoint, with complete references, and detailed discussion of variations that have been suggested, see among others Jones et al. 1996 and Scott 2015). Some methods, including plug-in and smoothed bootstrap techniques, have been developed that are far superior to well-known earlier methods, such as rules of thumb, least squares cross-validation, and biased cross-validation. Several researchers, see among others Jones et al. (1996) and Scott (2015), recommend a plug-in bandwidth selector as being most reliable in terms of overall performance. Thus, for bandwidth selection, we consider three different methods Bowman and Azzalini rule, Scott's rule and Freedman-Diaconis rule, for more details about these rules see among others Scott (2015). Conceptually, we have two different kinds of bandwidths, i.e. moneyness bandwidth h_m and maturity (calendar) bandwidth h_τ. The following Table reports the optimal bandwidths obtained from the three reference rules for Gaussian kernel function.

Table 4.1: Optimal bandwidths obtained from three reference rules

Type of bandwidth	Freedman-Diaconis rule	Bowman and Azzalini rule	Scott's rule
Optimal band. h_m for Gaussian	0.0885	0.0548	0.0551
Optimal band. h_τ for Gaussian	0.1449	0.1123	0.1344

From Table 4.1, we observe that Scott's and Bowman and Azzalini rules have closer moneyness bandwidth, while Freedman-Diaconis rule gives higher bandwidth (for deeper discussion about the differences between the three rules see Scott's 2015). Empirically, in order to consider different kernel functions, it will be convenient to fix one of these bandwidth selection methods. In particular, in what follows we start showing the results obtained by applying the Scott's rule, because it provided better approximations in these analyses. We recall that the optimal bandwidth, according to the Scott's rule, is given by 3.5 $\sigma_X n^{-1/3}$.

For a variety of reasons, there is no single kernel that can be recommended for all circumstances. One potential candidate is the normal kernel; however, it is relatively inefficient and has infinite support. The optimal Epanechnikov kernel is not continuously differentiable and cannot be used to estimate derivatives. In practice, the ability to switch between different kernels without having to reconsider the calibration problem at every turn is convenient. This task is easy to accomplish, but only for kernels of the same order. As Scott (2015) noted, if h_1 and h_2 are smoothing parameters to be used with kernels K_1 and K_2, respectively, then Table 4.2 gives a summary of factors for equivalent smoothing bandwidths among popular kernels.

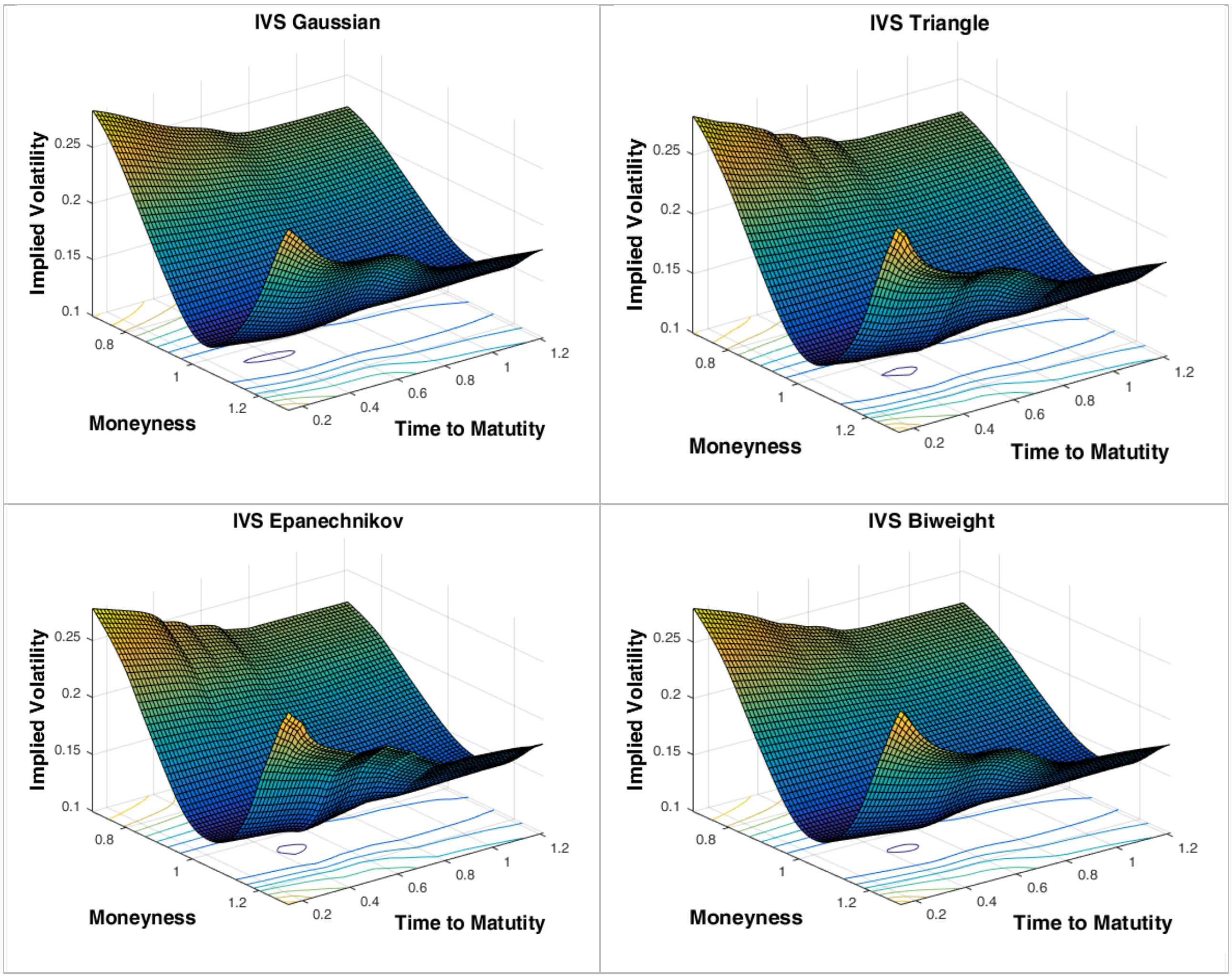
Table 4.2: Factors for equivalent smoothing among popular kernels

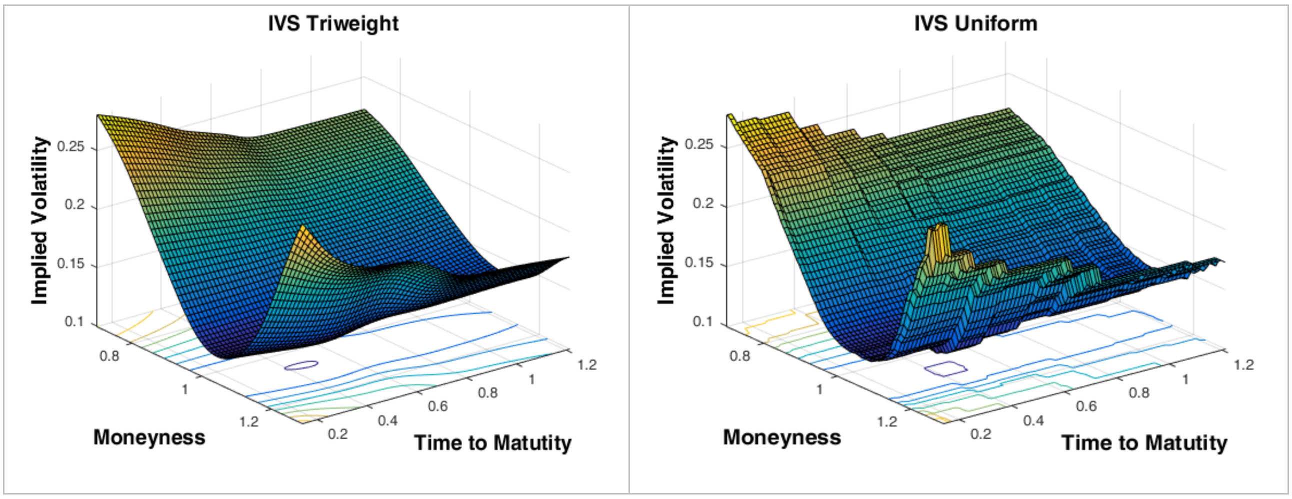
From \ To	Gaussian	Uniform	Epanech.	Triangle	Biweight	Triweight
Gaussian	1	1.740	2.214	2.423	2.623	2.978
Uniform	0.575	1	1.272	1.398	1.507	1.711
Epanech.	0.452	0.786	1	1.099	1.185	1.345
Triangle	0.411	0.715	0.910	1	1.078	1.225
Biweight	0.3881	0.663	0.844	0.927	1	1.136
Triweight	0.336	0.584	0.743	0.817	0.881	1

To investigate the impact of kernel functions on the proposed measure of arbitrage we proceed as follows. We start using Scott's rule for bandwidth selection with different kernel functions. Therefore, with each kernel function we use the correspondent factor for equivalent smoothing, summarized in Table 4.2, to determine the optimal bandwidth.

Figure 4.1 shows the IV surface estimated using put options for the daily data on May 13, 2015, Scott's rule and all kernel functions presented in Table 4.2. The IV smile is very clear for small maturities and still evident as time to maturity increases for all examined kernel functions.

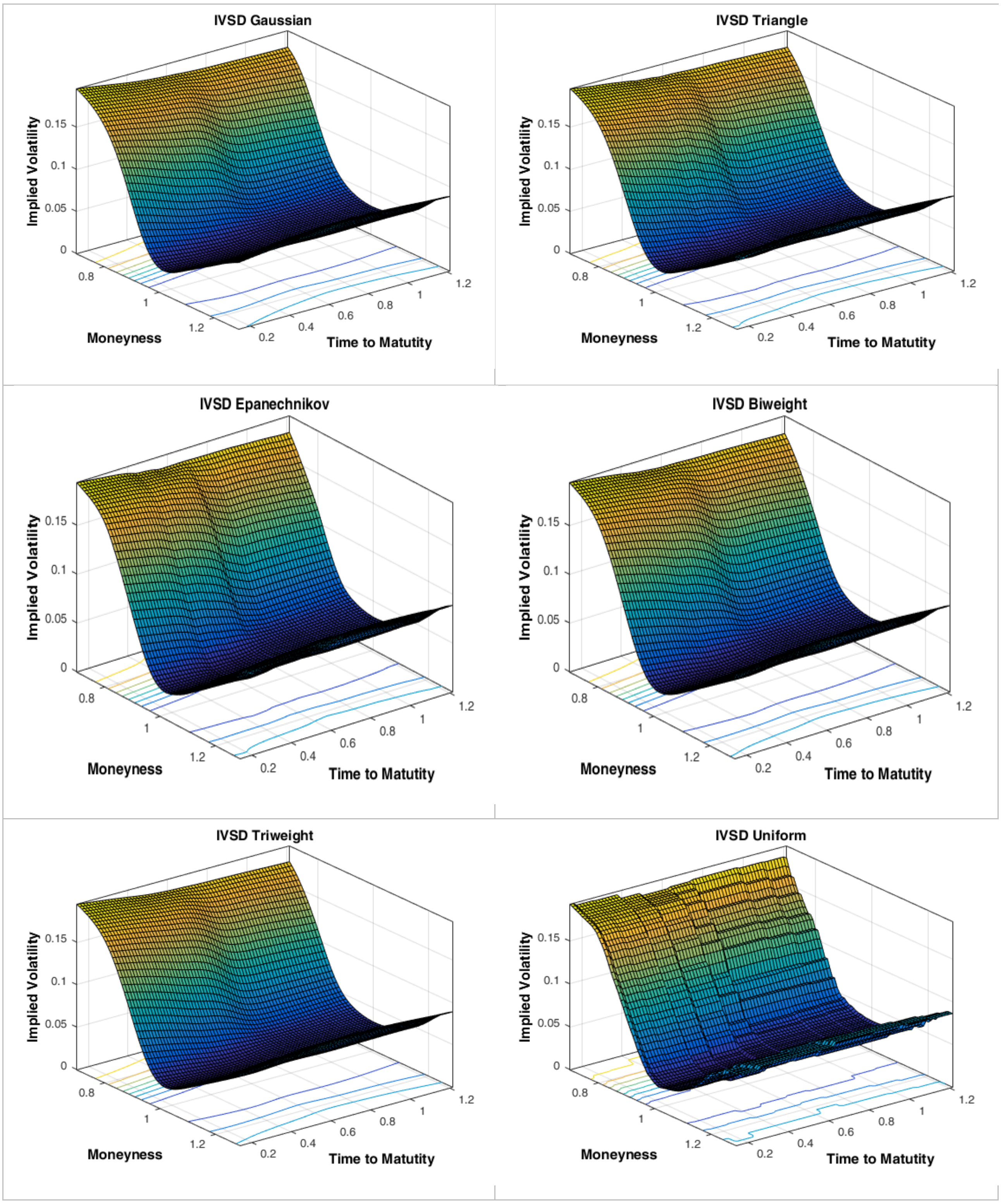
Figure 4.1: Implied Volatility Surface (IVS) of S&P 500 put options obtained with Scott's rule and different kernel functions (i.e. Gaussian, Triangle, Epanechnikov, Biweight, Triweight and Uniform)





Clearly, from Figure 4.1, we observe almost the same surface for all kernel functions. Overall, these results confirms that the choice of kernel is not critical, while the performance of the smoothed IVS is more a question of bandwidth choice. To evaluate the presence of arbitrage opportunities, we compute the difference in implied volatilities between call and put options that have the same strike price, the same maturity and are written on the same underlying asset. In particular, we consider the differences that are greater than 80 percent of the maximum absolute value of the differences between call and put implied volatilities. In this way, we rule out some differences due to the noisy data or transaction costs. Figure 4.2 shows the differences in implied volatilities between call and put options using Scott's rule and different kernel functions (i.e. Gaussian, Epanechnikov, Triangle, Uniform, Biweight and Triweight).

Figure 4.2: Implied volatility surface differences (IVSD) with Scott's rule and different kernel functions (i.e. Gaussian, Triangle, Epanechnikov, Biweight, Triweight and Uniform)



So far as IV surface, we obtain almost the same surface hence the type of kernel function does not matter. Interestingly, in Figure 4.2, it is clear that the differences are significant at lower moneyness which corresponds to OTM put options and ITM call options. However, since the market increases and it is well known that OTM put options and ITM call options are not reliable data to evaluate arbitrage opportunities, we focus on at ATM options. From figure 4.2, we observe even at ATM option there are small differences for all considered kernel functions, which may represent arbitrage opportunities. In particular, the differences increase as the maturities increase.

Generally, it is well known that bandwidth selection in nonparametric smoothing problems is very crucial step, for this reason, we consider other selection methods. To simplify the exposition of our results, we show first the impact of bandwidths choices for Gaussian kernel function considering different normal reference rules on total VS (which is the sum of VS (4.29) for all considered maturities). In practice, we try to change the bandwidth h_m, fix the optimal h_τ and observe what happens. In particular, we increase the optimal bandwidth h_m for each normal reference rule with step length 0.01.

Table 4.3: Comparison of the total VS with different bandwidth choices using Gaussian kernel function

	Freedman-Diaconis rule	**Bowman and Azzalini rule**	**Scott's rule**
Optimal bandwidth (h_m))	0.0885	0.0548	0.0551
Optimal bandwidth (h_τ)	0.1449	0.1123	0.1344
Total VS	361.44	380.31	383.46
h_m	0.0985	0.0648	0.0651
Total VS	355.99	373.62	376.51
h_m	0.1085	0.0748	0.0751
Total VS	351.62	366.88	369.48
h_m	0.1185	0.0848	0.0851
Total VS	348.33	360.59	362.91
h_m	0.1285	0.0948	0.0951
Total VS	346.08	355.05	357.16

From Table 4.3, we observe that Scott's rule presents the highest total VS and an optimal bandwidth between the two normal reference rules (i.e. Bowman and Azzalini rule and Freedman-Diaconis rule). Generally, for the three normal reference rules, the total VS decreases as the bandwidth increases. This result confirms the importance of bandwidth selection methods.

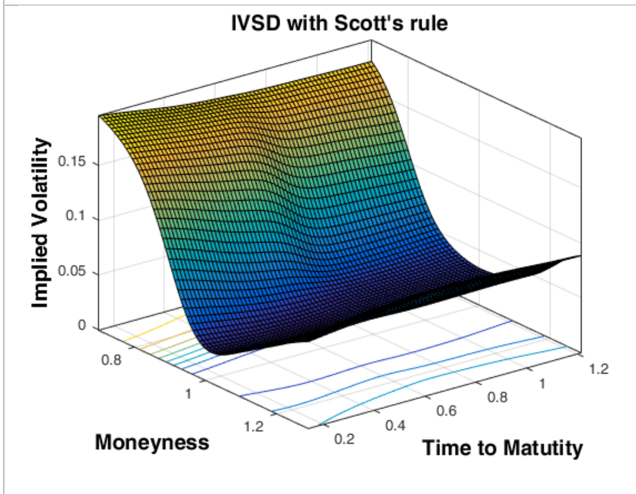
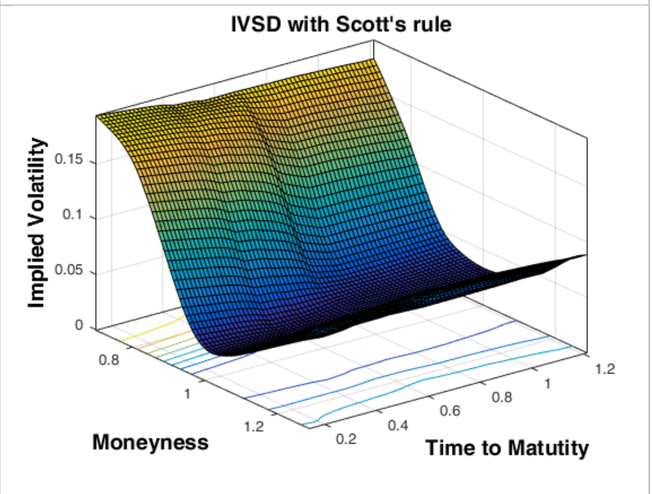
Next, we compare the optimal bandwidth choices and the total VS from three different normal reference rules either using Gaussian kernel function or Epanechnikov kernel function. In this context, we use the factor for equivalent smoothing bandwidths between Gaussian and Epanechnikov of 2.214, see Table 4.2. The result of this analysis is reported in Table 4.4.

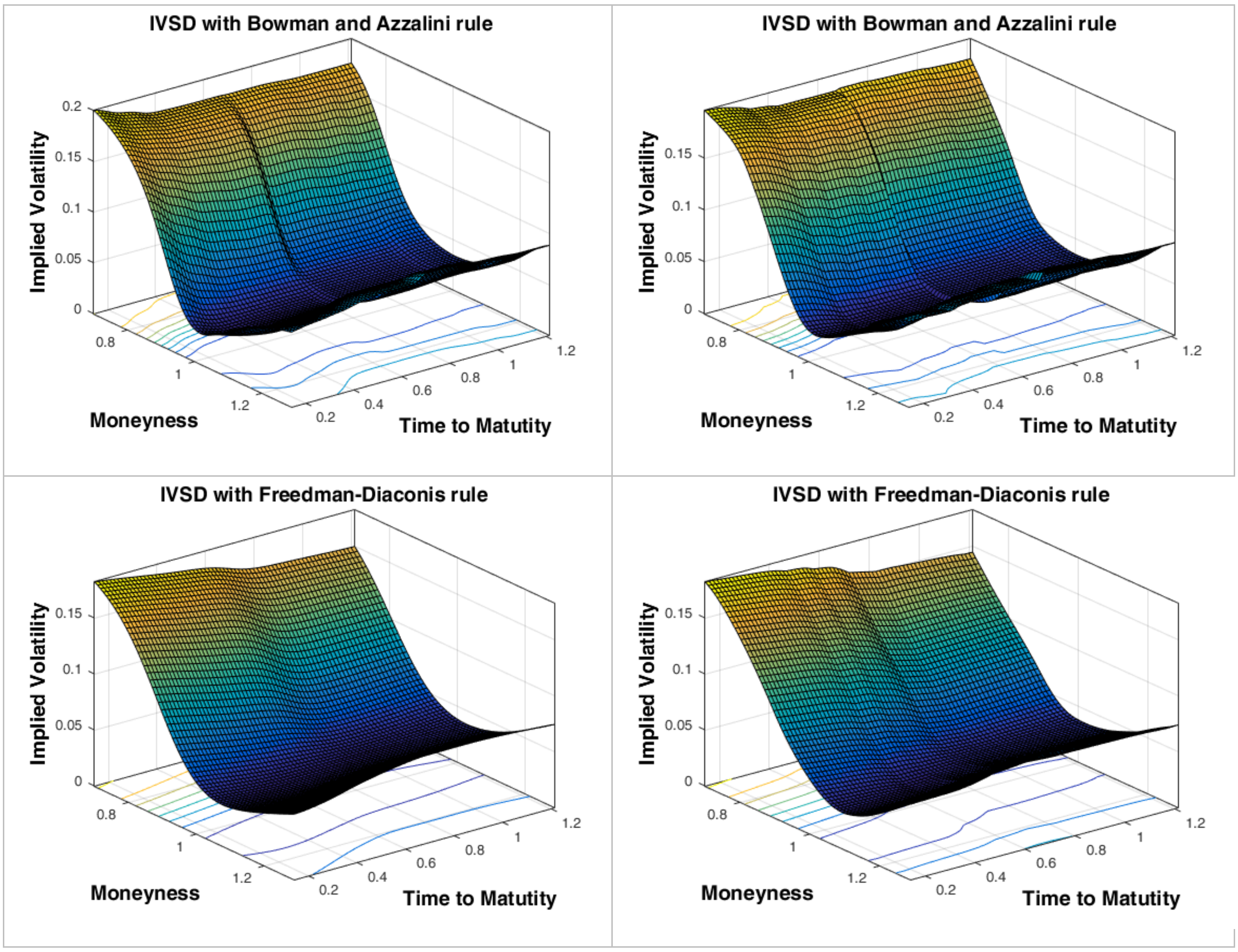
Table 4.4: Comparison of the optimal bandwidth choices and total VS from three normal reference rules either with Gaussian kernel or with Epanechnikov function

	Type of bandwidth	**Freedman-Diaconis rule**	**Bowman and Azzalini rule**	**Scott's rule**
Optimal band. Gaussian	h_m	0.0885	0.0548	0.0551
Optimal band. Gaussian	h_τ	0.1959	0.1213	0.1219
Total VS with Gaussian		361.44	380.31	383.46
Optimal band. Epanech.	h_m	0.1959	0.1213	0.1219
Optimal band. Epanech	h_τ	0.0885	0.0548	0.0551
Total VS with Epanech.		358.71	380.98	383.52

Using Epanechnikov kernel function, which requires a higher optimal bandwidth (2.214 factor) than the Gaussian kernel function, we obtain total VS less than normal case for Freedman-Diaconis rule, while the opposite holds for the others reference rules. Yet the Scott rule achieves the biggest total VS either using Gaussian or Epanechnikov kernel function. Overall, Tables 4.3 and 4.4 numerically show the importance of the bandwidth selection. Visually, Figure 4.3 reports the IVSD using three reference rules either with Gaussian or Epanechnikov kernel function.

Figure 4.3: comparison of the implied volatility surface differences (IVSD) obtained with three different normal reference rules and either with Gaussian kernel function (column 1) or with Epanechnikov one (column 2)

Gaussian kernel	**Epanechnikov kernel**
	



From figure 4.3, we observe some differences either between the three normal reference rules or between the two types of kernel function (that use different optimal bandwidths). For example, among the three normal references rules, observe the difference between Scott's rule and Bowman and Azzalini rule. While for the difference among kernel functions see IVSD based on Bowman and Azzalini rule either with Gaussian or Epanechnikov kernel function. These results confirm the impact of bandwidth selection methods on smoothed surface and the marginal contribution of kernel functions. Generally, from all Figures, we observe at ATM option there are small differences, which may represent arbitrage opportunities. In particular, the differences increase as the maturities increase for all kernel functions and bandwidths selection methods. Hence, we argue that the proposed method measure of arbitrage, even it is sensitive for the examined settings, remains theoretically and empirically valid method to detect the presence of arbitrage opportunities in the option market.

To evaluate the size of the arbitrage opportunities, one also could combine the IV smoothing with SPD estimation. This is important, because the SPD requires some properties in order to be consistent with no-arbitrage argument. In particular, the nonnegativity property of SPD since its negative values immediately corresponds to the possibility of the arbitrage opportunities in the market. For complete treatment of this method see a relatively conservative approach adopted by Benko et al. (2007). The last approach is of great practical importance and generally confirms the result obtained via the violation of the put call parity relation.

The second contribution of this study is to propose different methods to estimate SPD under the classical hypothesis of BS model. In particular, we use two different methodologies to evaluate the conditional expectation. Namely, the nonparametric estimator based on kernel estimator and a new alternative technique the so called OLP estimator proposed by Ortobelli et al. (2015). Differently from previous studies we estimate SPD directly from the underlying asset under the hypothesis of the BS model. To do so, firstly we examine the real mean return function using local polynomial smoothing technique. Then, we estimate the conditional expectation under real probability density. According the hypothesis of the BS model, we are able to derive a closed formula for approximating the conditional expectation under risk neutral probability. Now, we describe in details our alternative approach towards estimating the SPD.

4.3 Alternative methods to estimate the SPD

For the sake of clarity, denote S^{RW} for a real world price and S^{RN} for the risk neutral price. Under the hypothesis of the BS model it is straightforward to write:

$$S_T^{RN} = S_T^{RW} e^{-(\mu - r)T}, \tag{4.30}$$

Since $S_t = e^{-r(\mathrm{T-t})} E(S_T^{RN} \mid \Im_t)$, we can write $S_t = e^{-r(\mathrm{T-t})} E(S_T^{RW} e^{-(\mu - r)T} \mid \Im_t)$ from which we obtain:

$$E(S_T^{RN} \mid \Im_t) = e^{-\mu T + rt} E(S^{RW} \mid \Im_t), \tag{4.31}$$

If we assume μ changes over time in model (4.21), then equation (4.31) becomes

$$e^{-\int_0^T (\mu(\tau) - r) d\tau} E(S_T \mid \Im_t) = E^Q(S_T \mid \Im_t), \tag{4.32}$$

where, $E^Q(S_T \mid \Im_t)$ denotes expectation under risk neutral world and $E(\mathrm{S}_T \mid \Im_t)$ the conditional expected price under real world. Moreover, (4.32) is equivalent to:

$$e^{-\int_0^T \mu(\tau) d\tau} \int_0^\infty s q_{RW}(\mathrm{s}) ds = e^{-rT} \int_0^\infty s q_{RN}(\mathrm{s})\, \mathrm{ds}, \tag{4.33}$$

where, $\mathrm{q}_{RW}(\mathrm{s})$ and $q_{RN}(s)$ denotes SPDs under real and risk neutral world respectively. Please note that under the BS hypothesis S_{T-t} has the same distribution as $S_T e^{-\mu t}$.

4.3.1 Local polynomial smoothing technique

The first step in this approach is to propose a direct method of estimating the real mean return function. Therefore, we use a local estimator that automatically provides an estimate of the real mean function and its derivatives. The input data are daily prices. Denoting the intrinsic value by $\tilde{\mu}_i$ and the true function by $\mu(t_i)$, $i = 1, ..., n,$ we assume the following regression model:

$$\tilde{\mu}_i = \mu(t_i) + \varepsilon_i, \tag{4.34}$$

where, ε_i models the noise, *n* denotes the number of data considered. The local quadratic estimator $\tilde{\mu}(t)$ of the regression function $\mu(t)$ in the point *t* is defined by the solution of the following local least squares criterion:

$$\min_{\alpha_0, \alpha_1, \alpha_2} \sum_{i=1}^{n} \{\tilde{\mu}_i - \alpha_0 - \alpha_1 (t_i - t) - \alpha_2 (t_i - t)^2\} \, k_h(t - t_i) \tag{4.35}$$

where, $k_h(t - t_i) = \frac{1}{h} k\left(\frac{t - t_i}{h}\right)$ is *kernel* function, see Fan and Gijbels (1996) for more details.

Comparing the last equation with the Taylor expansion of μ yields:

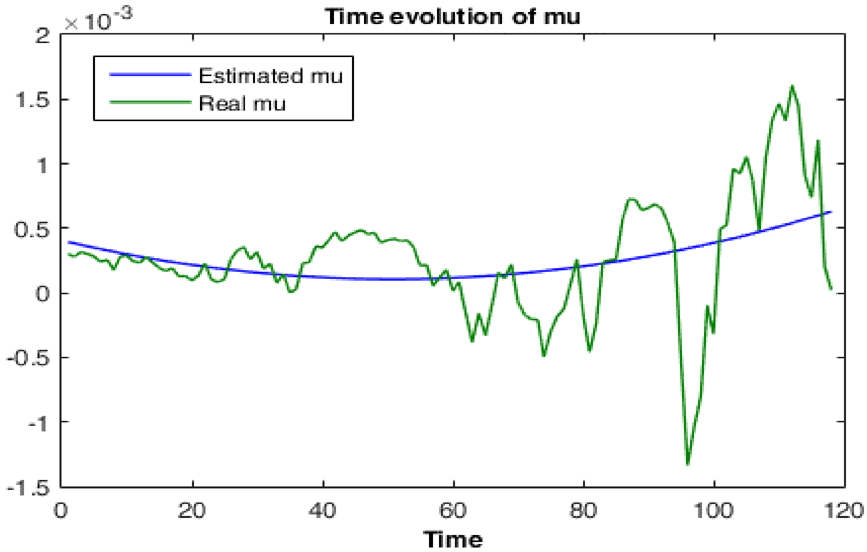
$$\alpha_0 = \hat{\mu}(t_i), \; \alpha_1 = \hat{\mu}'(t_i), \; 2\alpha_2 = \hat{\mu}''(t_i), \tag{4.36}$$

which make the estimation of the regression function and its two derivatives possible. The second step towards estimating the SPD is to use two methodologies, namely OLP estimator and kernel estimator, to estimate the quantity $E(S_T \mid S_t)$.

4.3.2 Second empirical analysis

In the second empirical analysis, we present an application to the S&P 500 index using daily data for the sample period April 28, 2014 to April 28, 2015. In this context, we use Treasury Bond 3 months as a riskless interest rate for a period matching our selecting data. Firstly, we examine the real mean return function using local polynomial smoothing technique (4.35). The results of this analysis are reported in Figure 4.4.

Figure 4.4: Real mean return function estimation over time (mu $\equiv \mu$)



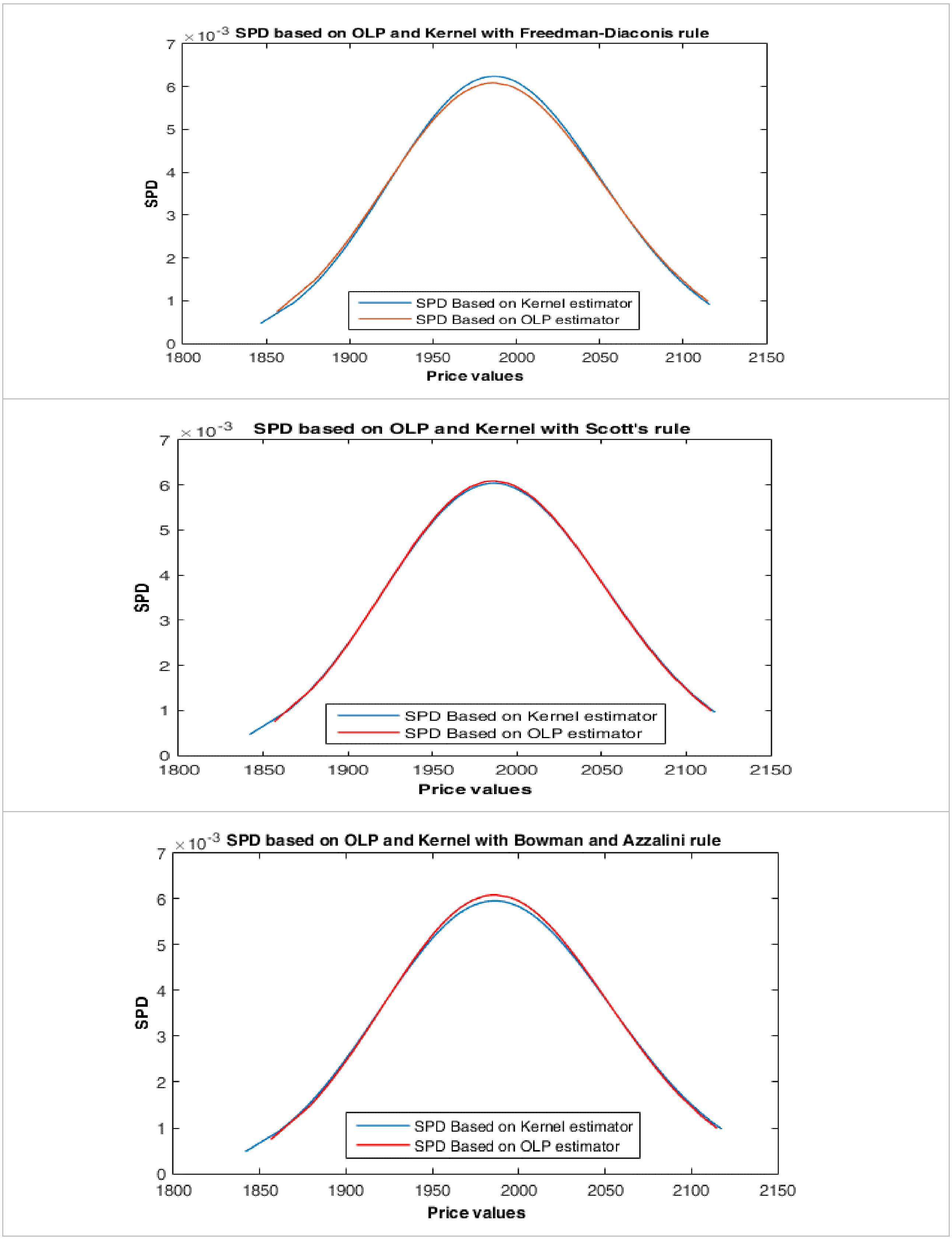
Secondly, we evaluate the conditional expected price using both estimators, namely kernel estimator and OLP, to estimate $E(S_T \mid S_t)$ as described above. In particular, for kernel method, it is well known that bandwidth choice is very crucial step, for this reason, as previous analysis we consider three selection methods (i.e. Scott's rule, Bowman and Azzalini rule and Freedman-Diaconis rule). Following Table reports the optimal bandwidth for each normal reference rule:

Table 4.5: Comparison of the optimal bandwidth choices from three normal reference rules

	Freedman-Diaconis rule	**Bowman and Azzalini rule**	**Scott's rule**
Optimal bandwidth	0.6586	0.3637	0.4310

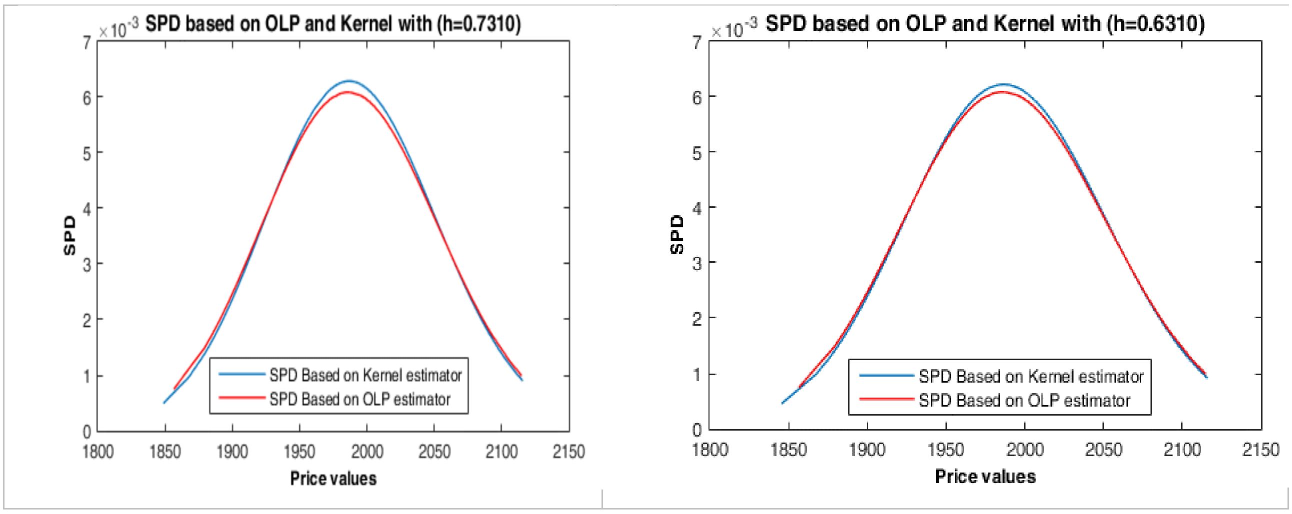
So far as previous empirical analysis, Scott's rule gives an optimal bandwidth between the two other normal reference rules (i.e. Bowman and Azzalini rule and Freedman-Diaconis rule). Furthermore, in what follows we choose Gaussian as a kernel function (motivated by assumptions of the BS model). Finally, we use the relationships (4.32) and (4.33) in order to recover the SPD. The results of this analysis are reported in Figure 4.5

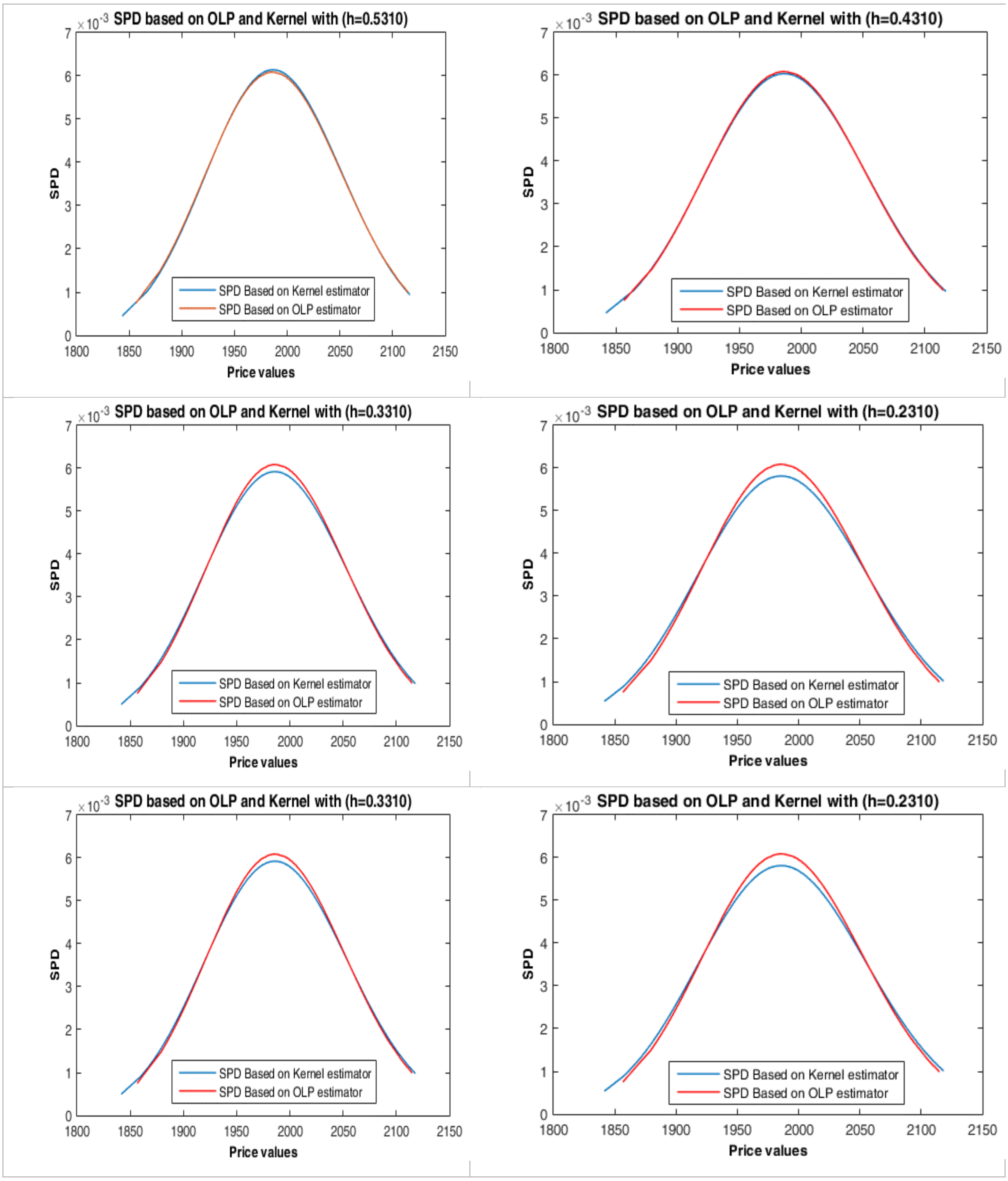
Figure 4.5: State Price Densities obtained with Kernel and OLP estimators



From Figure 4.5, we observe that Scott's rule gives the best result (SPDs from both estimators are almost identical). While for the other rules (Bowman and Azzalini rule and Freedman-Diaconis rule) we note a slight difference in the result obtained from both estimators. This result can be explained by the nature of the two methodologies. In particular, the OLP method proposed by Ortobelli et al (2015) yields a consistent estimator of the random variable $E(Y|X)$, while the generalized kernel method proposed in equation (4.2) yields a consistent estimator of the distribution function of $E(Y|X)$.Thus, OLP method that yields consistent estimators of random variables $E(Y|X)$ can be used to evaluate the SPD. Overall, the small differences between the two methods principally could be explained by the impact of bandwidth choice and then by other parameters (such as polynomial degree, type of kernel function etc.). Therefore, especially for kernel method, we examine the impact of bandwidth selection in a different way. In particular, we use Scott's rule then we try to increase or decrease the optimal bandwidth with step length 0.1 to observe what happens. The results of this analysis are reported in Figure 4.6.

Figure 4.6: State Price Densities obtained with OLP and Kernel estimator with different bandwidth choices

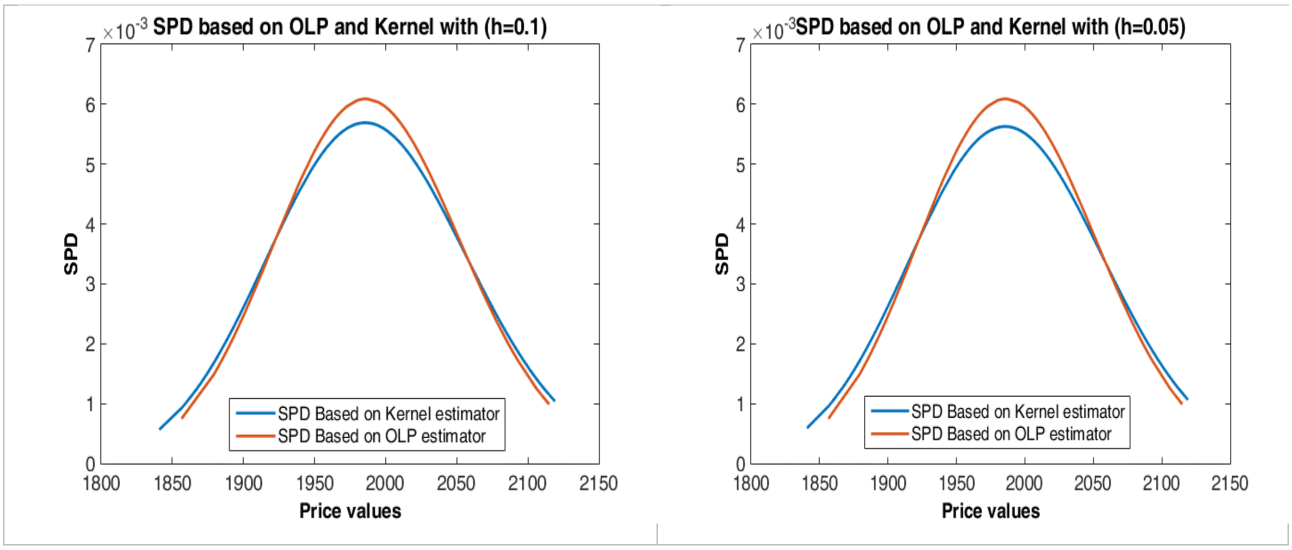




From Figure 4.6, with Scott's optimal bandwidth ($h = 0.4310$) we obtain the best result (SPDs from both estimators are almost identical). While increasing or decreasing the optimal bandwidth with step length 0.1, we observe increasing differences between the two estimators. In particular, when we increase the optimal bandwidth the kernel method surpasses the OLP estimator. Whereas the opposite holds when we decrease the optimal bandwidth. Furthermore, it would be of interest to consider even smaller bandwidth (e.g. when minimizing the bias).

Thus, we examine the shape of the SPD using two smaller bandwidths ($h = 0.1$ and $h = 0.05$). The results of this analysis are reported in Figure 4.7.

Figure 4.7: State Price Densities obtained with OLP and Kernel estimator with small bandwidth choices



From Figure 4.7, we clearly observe that the SPD based on kernel method, with smaller bandwidths, allocates under the SPD based on OLP estimator. The gap between the two SPDs increases as the bandwidth decreases. This analysis confirms that the bandwidth selection methods have direct impact on the arbitrage free condition. In particular, it was shown that as the bandwidth parameter increases the degree of no-arbitrage violation decreases (for deeper discussion see Kopa and Tichý 2014). On the one hand, these analyses show the importance of bandwidth selection methods. On the other hand, observe that OLP estimator also could be improved depending on the parameter b^k that we choose (in this analysis we use the rule of thumb proposed by Ortobelli et al. (2015)).

Furthermore, we examine the third polynomial degree (In practice for a given bandwidth h, a large value of polynomial would cause a large variance and a considerable computational cost, but it would reduce the modeling bias. Since the bandwidth is used to control the size (complexity), it is recommended to use the lowest odd order) and other kernel functions. Overall, these analysis confirm the previous conclusion (either polynomial degree or kernel function types contribute marginally in comparison with the impact of bandwidth selection methods).

Finally, we use the same procedure to extend our analysis to the tridimensional case towards examining the evolution of SPDs over time. In particular, with kernel method we use Scott's rule and Gaussian as a kernel function. Figures 4.8 and 4.9 show SPDs estimated via kernel and OLP estimator respectively, from which we observe as well a slight difference between the two estimators.

Figure 4.8: State Price Densities obtained with Kernel estimator

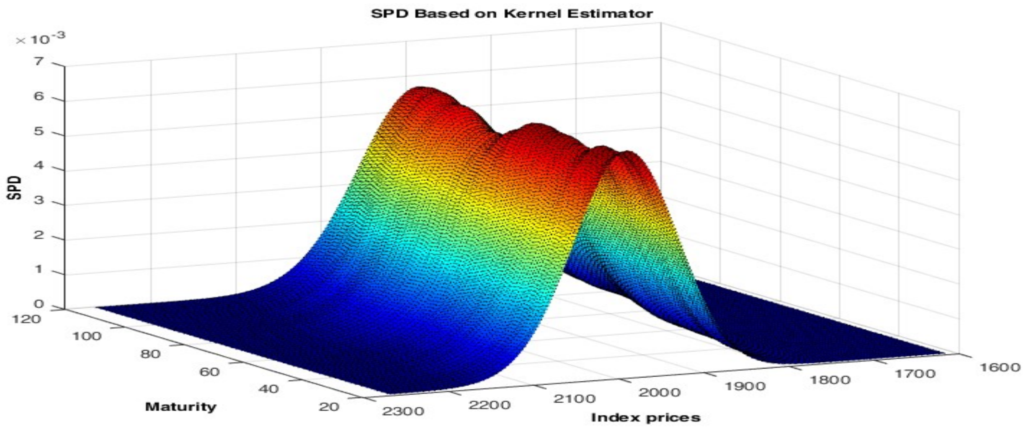
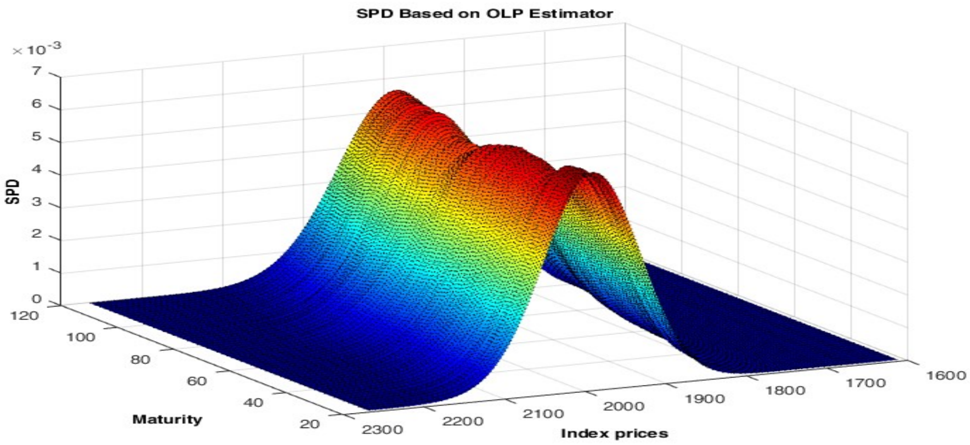


Figure 4.9: State Price Densities obtained with OLP estimator



4.4 Concluding remarks

In this chapter, we present alternative methods to evaluate the presence of arbitrage opportunities in the market. In particular, we examine the violation of the well-known put-call parity no-arbitrage relation and discuss the nonnegativity of the SPD. Then, we propose different methods to estimate SPD. Particularly, we use two distinct methodologies for estimating the conditional expectation, namely the kernel method and the OLP method recently proposed by Ortobelli et al. (2015). We deviate from previous studies in that we estimate SPD directly from the underlying asset under the hypothesis of BS model. To this end, firstly we examine the real mean return function using local polynomial smoothing technique. Then, we estimate the conditional expectation under real probability density. Under the hypothesis of BS model, we are able to derive a closed formula for approximating the conditional expectation under risk neutral probability. This analysis allows us extrapolating arbitrage opportunities and relevant information from different markets (futures and options) consistently with the analysis of the underlying.

Chapter 5

On the impact of conditional expectation estimators in portfolio theory

In this chapter, we investigate the implications for portfolio theory using conditional expectation estimators. In particular, we focus on three financial applications: i) approximation of the conditional expectation within large-scale portfolio selection problems, ii) performance valuation considering the heavy tails of returns, and iii) the optimal portfolio choices for different investors' preferences.

In the first application, we discuss and examine some correlation measures (based on the conditional expectation) used to approximate properly the returns in large-scale portfolio problems. Then, we compare the impact of alternative return approximation methodologies on the ex-post wealth of a classic portfolio strategy. In this context, we show that a correlation measures that use properly the conditional expectation perform better than the classical ones. Moreover, the correlation measure typically used for returns in the domain of attraction of a stable law works better than the classical Pearson correlation does.

In the second application, we propose new performance measures based on the conditional expectation that takes into account the heavy tails of the return distributions. Then, we examine portfolio strategies based on the optimization of the proposed performance measures. In particular, we compare the ex-post wealth obtained applying portfolio strategies, which use alternative performance measures based on the conditional expectation.

In the third application, we first propose a new consistent multivariate kernel estimator to approximate the conditional expectation. We show how the new estimator can be used for the return approximation of large-scale portfolio problems. Moreover, the proposed estimator optimizes the bandwidth selection of kernel-type estimators, solving the classical selection problem. Second, we deal with the portfolio selection problem from the point of view of different non-satiable investors, namely risk-averse and risk-seeking investors. In particular, using a well-known ordering classification, we first identify different definitions of returns based on the investors' preferences. The new definitions of returns are based on the conditional expected value between the random wealth assessed at different times.

Finally, for each problem, we propose an empirical application of several admissible portfolio optimization problems applied to the US stock market. The proposed empirical analysis allows us to evaluate the impact of the conditional expectation estimators in portfolio theory.

5.1 Introduction

The mean-variance model has had a central role in portfolio theory ever since the pioneering work of Markowitz (1952). The main idea behind this model is that the return and the risk are modelled in terms of the portfolio mean and variance. This may be particularly suitable for small rational investors whose investments cannot influence market prices and who prefer smaller risks to larger ones and higher yields to lower ones. Unfortunately, widespread research has pointed out that the set of assumptions under which the classical mean-variance framework is established is not consistent with all investors' preferences. For this reason, several alternative approaches to portfolio selection have been proposed: see, among others, Biglova et al. (2004), Rachev et al. (2008), and Ortobelli et al. (2009). Despite the fact that portfolio optimization based on several advanced measures has been introduced in the literature (Konno and Yamazaki, 1991; Young, 1998; Rockafellar et al., 2006; Rachev et al., 2008; Ortobelli and Tichý, 2015 and the literature therein) and has been utilized in practice, the literature on the impact of the conditional expectation estimation in portfolio theory is limited. Therefore, the general aim of this chapter is to assess the impact of conditional expectation estimators on different financial applications within portfolio theory.

The first contribution of this chapter is to investigate the impact of alternative return approximation methods depending by k-factors in large-scale portfolio problem (such as in the k-fund separation model of Ross (1979)). In particular, we examine and compare the classical return approximation with a nonparametric approximation of the returns depending on few factors obtained by a principal components analysis (PCA). Furthermore, according to Ortobelli and Tichý (2015), we determine the principal components (of PCA) either using a correlation matrix suitable for heavy tailed distribution (called stable linear correlation), or using the classical Pearson correlation matrix (which summarizes the joint dispersion behavior of Gaussian vectors). The most commonly used approach to estimate the relationship between returns and k factors is the linear approximation based on the ordinary least squares (OLS) estimator (see Ross (1979)). This approximation appears good enough when the returns are normally distributed. Admitting small departures from normality of the returns do not affect the regression coefficients greatly, however errors with a heavier tailed distribution, which is more suitable for modeling asset returns, can significantly affect the estimated OLS regression coefficients, (see Nolan et al. (2013)). Moreover, we believe that there exists substantial evidence of nonlinearity in the financial dataset used to estimate the returns (see among others Rachev et al 2008). For this reason, according to Ruppert and Wand (1994), we propose a nonparametric regression analysis to approximate the returns. This approach relaxes the assumptions of linearity and it suitable even for non-Gaussian distributions. In this context, we prove that the variability of errors of the return approximation decreases as the number of factor increases even when elliptically distributed returns present heavy tails. In addition, using concave dominance testing, we find that the nonparametric regression outperforms much better than its counterpart parametric (OLS) does. This empirical analysis is provided using portfolios of the components of S&P 500 index.

The second contribution of this chapter is to deal with a proper evaluation of portfolio choices that account the distributional tails of portfolios. In particular, the main purpose of this contribution is to present theoretically sound portfolio performance measures considering a more realistic behavior of the returns (i.e. heavy tailed distributions). Using a recent alternative conditional expectation estimator proposed by Ortobelli et al. (2015), we are able to forecast the conditional expected portfolio returns with respect to a given sigma algebra of events (either generated by possible profits or generated by possible losses). More specifically, the first suggested performance measure is based on the conditional expectation with respect to two different σ-algebras (the σ-algebra generated by the portfolio losses, and the σ-algebra generated by the portfolio profits). While the second performance measure considers σ-algebras generated by the joint losses and by joint gains in the market. Moreover, we illustrate how the new performance measures can mitigate the shortcoming of the classical Sharpe ratio (see Sharpe (1994)) showing with an ex-post empirical analysis their tested higher capacity to produce wealth in the US market.

The third contribution of this chapter proposes a new consistent multivariate kernel estimator to approximate the conditional expectation, and we stochastically compare the errors of the return approximation. We show that the approximation error is reduced with the new estimator, and that we can optimize the bandwidth parameters (whose approximation is always considered a problem for kernel-type estimators).Then, we deal with the portfolio selection problem from the point of view of different non-satiable investors, namely risk-seeking and risk-averse investors (see Ortobelli et al. (2015)). In particular, by using the conditional expected value properly, we first identify different definitions of returns based on the investors' preferences. The new definitions of returns are based on the conditional expected value between the random wealth assessed at different times. Finally, we compare the ex-post wealth obtained by maximizing some well-known performance ratios applied to the different returns definitions. In doing this, we can examine the impacts of the choices of investors with different risk aversion attitudes.

The rest of the chapter is organized as follows. In section 5.2, we discuss and examine the impact of the approximation methods/correlation measures within large-scale portfolio selection problems. In section 5.3, we introduce the optimal portfolio choices using consistent estimation of the conditional expectation. In section 5.4, we deal with the optimal portfolio choices for different investors' preferences. In this context, we propose a new methodology to solve bandwidth selection problem by introducing a new multivariate kernel estimator. Each section is empirically self-contained. Finally, our conclusions are summarized in section 5.5.

5.2 Practical and theoretical aspects of return approximation in large-scale portfolio selection problems

In this section, we focus on the approximation methods within large-scale portfolio selection problems. In particular, we theoretically and empirically compare different methodologies based on the approximation of the dependence between returns and factors obtained by a PCA applied to different return correlation matrices.

In this chapter, we consider n risky assets defined on a probability space $(\Omega, \Im, P)$. We point out the portfolios gross returns, $x'z$, where $x \in S = \{y \in R_+^n | \sum_{i=1}^n x_i = 1; x_i \geq 0; \forall\, i = 1, \dots, n\}$ is the vector of non-negative allocations (i.e., no short sale are allowed $x_i \geq 0$) among n risky limited liability investments with gross returns[8] $z = [z_i, \dots, z_n]'$. Moreover, we generally assume that portfolio of gross returns belong to $L^p(\Omega, \Im, P) = \{X | E(|X|^p) < \infty\}$ for some $p \geq 1$.

According to many researchers, see among others Rachev et al. (2005), the portfolio dimensional problem is strongly related to the estimation of statistical parameters input, which describes the dependence structure of the returns. In this setting, aiming to get a good

[8] We define the i-th gross return between time t and time $t+1$ as $z_{i,t} = \frac{P_{t+1,i} + d_{[t,t+1],i}}{P_{t,i}}$, where $P_{t,i}$ is the price of the i-th asset at time t and $d_{[t,t+1],i}$ is the total amount of cash generated by the instrument between t and $t + 1$.

approximation of the portfolio risk-reward measures, Papp et al. (2005) and Kondor et al. (2007) (among others) have shown that the number of observations should increase with the number of assets. Therefore, in order to get a sound approximation of portfolio input measures, it is important to find the right tradeoff between the number of historical data and a statistical approximation of the historical observations that depend only on a few parameters. Many studies illustrate that the problem of parameter uncertainty increases with the number of assets (see Kan and Zhou 2007).

In practice, there are many different ways to reduce the dimensionality dependence of a large-scale portfolio selection problem. In this section, we compare parametric and nonparametric regression models used to reduce the dimensionality of the large-scale portfolio problem. In particular, we reduce the dimensionality of the portfolio problem and we approximate the return series through a multifactor model that depends on a proper number (not too large) of factors. Toward this end, we perform a principal component analysis (PCA) of the correlation matrix of returns in order to identify the main portfolio factors (principal components), whose variability is significantly different from zero. According to Ortobelli and Tichý (2015), we can determine the principal components applying the PCA to a proper linear correlation matrix of returns that is different from the Pearson one when the returns are in the domain of attraction of a stable non-Gaussian law. Therefore, in the following subsections we define the crucial aspects useful for the proposed dimensionality reduction, which are:

1) the description of different correlation measures to represent the dependence structure between random variables that are used to identify the main factors which account all return variability (through a PCA);
2) the description of alternative (parametric and nonparametric) methodologies to approximate the dependence between returns and factors;
3) An ex-post analysis discusses the impact of the use of different correlation measures and approximation methods.

5.2.1 Correlation measures and principal component analysis

The measurement of dependency among random variables plays a central role in several financial decision-making problems. Therefore, various measures of dependence between random variables have been proposed and extensively studied; see among others Scarsini (1984), Cherubini et al. (2004) and the literature therein. Since empirical evidence shows that Pearson correlation ρ_P does not well approximate the dependence structure of returns (see Ortobelli and Tichý (2015)), we consider an alternative correlation measure that is suitable for heavy tail random variables belonging to $L^p(\Omega, \Im, P)$ with $p \in (0,1)$. In particular, for any

α stable sub-Gaussian vector (with $\alpha > 1$) we can define the following linear correlation measure (called stable correlation measure):

$$\rho_s(X,Y) = \frac{E((X-E(X))sign(Y-E(Y)))}{2E(|X-E(X)|)} + \frac{E((Y-E(Y))sign(X-E(X)))}{2E(|Y-E(Y)|)}. \quad (5.1)$$

The stable correlation measure and the Pearson linear correlation represent alternative measures used to evaluate the returns dependence structure. Moreover, according to Ortobelli and Tichý (2015), for any couple of random variables X_1, X_2 we can apply Pearson or stable correlation measures (simply indicated as ρ) to all the random variables $Z_i = (X_i - E(X_i|\mathfrak{I}_1))$ $(i = 1,2)$ orthogonal to $L^2(\Omega, \mathfrak{I}_1, P)$ (when we use the scalar product $(U,V) \longrightarrow E(UV)$) where $\mathfrak{I}_1$ is a proper sub-σ-algebra of $\mathfrak{I}$ (i.e. $\mathfrak{I}_1 \subset \mathfrak{I}$) and the random variables X_i $(i = 1,2)$ are not $\mathfrak{I}_1$-measurable. Doing so, we obtain the following alternative correlation measures (commonly indicated with the same symbol ρ that could be either the Pearson linear correlation or the stable one):

$$O_{\rho,\mathfrak{I}_1}(Z_1, Z_2) = \rho(Z_1, Z_2), \quad (5.2)$$

that extend the Pearson correlation measure and the stable one when $\mathfrak{I}_1 = \{\emptyset; \Omega\}$. Recall that the orthogonal projection $E(X|\mathfrak{I}_1)$ minimizes the expected squared differences $E((X-Y)^2)$ among all $\mathfrak{I}_1$-measurable random variables Y. In other words, it is the best predictor of X based on the information in $\mathfrak{I}_1$. Besides, we know that $E((X - E(X|\mathfrak{I}_1))Y) = 0$ for all $\mathfrak{I}_1$-measurable random variables Y.[9] Generally, we consider linear correlation measures (as (5.2)) when we want to reduce the dimensionality of large-scale portfolio selection problems. In this case, the sigma algebra $\mathfrak{I}_1$ is the sigma algebra generated by a market index, and the measure (5.2) identifies the variability of the part of random variables $X - E(X|\mathfrak{I}_1)$, which are "uncorrelated" with the market index.

According to Ortobelli and Tichý (2015), we consider the σ-algebra $\mathfrak{I}_1$ generated by the so-called upper stochastic bound $max_i z_i$. The upper stochastic bound $max_i z_i$ satisfies the relation $max_i z_i \geq x'z$ for all portfolio of weights $x \in S$. Essentially, this choice is based on both unambiguous financial evidence and compelling theoretical argument from the probability theory. To give a meaningful financial interpretation, we start with this simple consideration; every investor wishes to determine 'today' the best daily optimal asset that will be the nearest to the future maximum, i.e. the upper stochastic bound of the market. Considering correlation measure (5.2) with the sigma algebra $\mathfrak{I}_1$, generated by the upper stochastic bound investors take

[9] This result theoretically supported by a well-known Theorem from probability theory, see among others Chung (2001).

into account the variability of the part of their portfolio uncorrelated with the upper stochastic bound. To estimate $E(Y|\Im_1) = E(Y|max_i z_i)$ we can use the Nadaraya-Watson kernel estimator:

$$E(Y|max_k z_k = x) = \frac{\sum_{i=1}^{n} y_i K\left(\frac{x-x_i}{h(n)}\right)}{\sum_{i=1}^{n} K\left(\frac{x-x_i}{h(n)}\right)},$$

where the *kernel function* $K(x)$ is a density function such that i) $K(x) < C < \infty$; ii) $\lim_{x\to\pm\infty} |xK(x)| = 0$; and iii) the positive bandwidth function $h(n) \to 0$ when $n \to \infty$. Several studies confirm that the choice of the kernel function is not critical, while the performance of the nonparametric regression is more a question of bandwidth choice.[10] In the proposed analysis, we use the Gaussian kernel univariate estimator with bandwidth $h(n) = 3.5n^{-1/3}\sigma_{max_i z_i}$ as suggested by Scott (2015).

Once we identify the main factors (obtained by a PCA on the proper correlation matrix) that summarize the variability of market returns, we have to approximate the relationship between the returns and these factors. The return approximation can be described using parametric or nonparametric regression models.

5.2.2 Approximation with parametric and nonparametric regression models

With parametric regression models, we can replace the original *n* correlated time series $\{z_i\}_{i=1}^{n}$ with the *n* uncorrelated time series $\{R_i\}_{i=1}^{n}$ (obtained by the PCA) assuming that each z_i is a linear combination of R_i. In particular, the dimensionality reduction is obtained by choosing only the first factors that sufficiently summarize a large part of the variability. In this setting, we call portfolios factors f_i the first *s* principal components R_i with a significant variability, while the remaining $n - s$ principal components with smaller variability are summarized by an error ε. Typically, the OLS estimator is widely used to approximate the returns through the following linear relation.

$$z_i = b_{i,0} + \sum_{j=1}^{s} b_{i,j} f_j + \sum_{j=s+1}^{n} b_{i,j} R_j = b_{i,0} + \sum_{j=1}^{s} b_{i,j} f_j + \varepsilon_i, \quad i = 1,...,n \tag{5.3}$$

where z_i is the gross return for asset *i*, $b_{i,0}$ is the fixed intercept for asset i, $b_{i,j}$ is the coefficient for the factor f_j, *s* is the number of factors, ε_i is the error term for asset *i* and *n* is the number of assets.

Generally, the OLS estimator is a well-established and very useful procedure for solving

[10] For a more complete treatment, from a historical viewpoint, with complete references, and detailed discussion of variations that have been suggested, see Scott (2015).

regression problems when the returns are normally distributed. However, the returns are generally characterized by a heavier tailed distribution (see among others Rachev and Mittnik (2000)) and therefore we cannot generally assume that the dependence between returns and the principal components is linear. For this reason, we suggest to use the nonparametric regression analysis as an alternative to the classical parametric approach (5.3). Typically, in several financial models (APT, CAPM, etc.) the returns are assumed to be elliptically distributed, and the large-scale portfolio problem is solved approximating the returns with a regression model on some uncorrelated market factors. Even in this chapter we reduce the complexity of the large portfolio model using a nonparametric regression model where *s* factors are determined applying a PCA to a linear correlation measure and the $(s+1)$-th factor M_{s+1} is a market index i.e.

$$z_i = E(z_i | f_1, \dots, f_s, M_{s+1}) + \varepsilon_i^{(s)}. \quad (5.4)$$

Clearly, when we regress the returns to the uncorrelated factors obtained by a PCA using an OLS estimator we obtain that the variability of the errors decreases increasing the number of factors. This aspect is also true for nonparametric estimators as suggested by the following theorem.

Theorem 5.1: *Let $\boldsymbol{W} = (z, f_1, \dots, f_s)$ be an s+1-dimensional elliptically distributed vector $Ell(\boldsymbol{\mu}, \Sigma)$, where $\boldsymbol{\mu}$ and Σ are respectively the mean vector and the dispersion matrix. Moreover, $f_1, \dots, f_s$ are uncorrelated one-dimensional factors. Assume that $\boldsymbol{W} = \boldsymbol{\mu} + A\boldsymbol{G}$, where A is a continuous positive random variable, which is independent from the Gaussian vector $\boldsymbol{G}$ that has null mean and variance covariance matrix Σ. Then, it follows that*

$$E(z | f_1, \dots, f_s) = \mu_z + \sum_{i=1}^{S} (E(z|f_i) - \mu_z) \sim Ell\left(\mu_z, \sum_{i=1}^{s} \frac{\sigma_{zf_i}^2}{\sigma_{f_i}^2}\right). \quad (5.5)$$

where σ_{zf_i} is the covariation between z and f_i (component $(1,i)$ o f matrix Σ) and $\sigma_{f_i}^2$ is dispersion of factor f_i (component (i,i) of matrix Σ).

Proof of Theorem 5.1

To proof Theorem 5.1 we need to prove the following more general Proposition.

Proposition 5.1: *Let $\boldsymbol{W} = (\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{Z})$ be an n-dimensional elliptically distributed $Ell(\boldsymbol{\mu}, \Sigma)$ vector, where **X**, **Y** and **Z** are respectively p-dimensional, q-dimensional and n-p-q dimensional vectors $(n > p > q \geq 1)$, with dispersion matrix*[11] $\Sigma =$

[11] With a little abuse of notation, in this chapter we write the dispersion matrix as

$$\Sigma = \begin{bmatrix} \Sigma_X & \Sigma_{XY} & \Sigma_{XZ} \\ \Sigma_{YX} & \Sigma_Y & \Sigma_{YZ} \\ \Sigma_{ZX} & \Sigma_{ZY} & \Sigma_Z \end{bmatrix} = \left((\Sigma_X, \Sigma_{XY}, \Sigma_{XZ}), (\Sigma_{YX}, \Sigma_Y, \Sigma_{YZ}), (\Sigma_{ZY}, \Sigma_{XY}, \Sigma_Z)\right)$$

$\big((\Sigma_X, \Sigma_{XY}, \Sigma_{XZ}), (\Sigma_{YX}, \Sigma_Y, \Sigma_{YZ}), (\Sigma_{ZY}, \Sigma_{XY}, \Sigma_Z)\big)$ *and mean* $\boldsymbol{\mu} = (\boldsymbol{\mu_X}, \boldsymbol{\mu_Y}, \boldsymbol{\mu_Z})$ *such that* $\boldsymbol{Y}$ *and* $\boldsymbol{Z}$ *are uncorrelated. Moreover, assume that* $\boldsymbol{W} = \boldsymbol{\mu} + A\boldsymbol{G}$, *where A is a continuous positive random variable, which is independent from the Gaussian vector* $\boldsymbol{G}$ *that has null mean and variance covariance matrix* Σ *. Then,*

$$E(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{Z}) = E(\boldsymbol{X}|\boldsymbol{Y}) + E(\boldsymbol{X}|\boldsymbol{Z}) - \boldsymbol{\mu_X} \sim Ell\big(\boldsymbol{\mu_X}, \Sigma_{XY}\Sigma_{\boldsymbol{K}}^{-1}\Sigma_{YX} + \Sigma_{XZ}\Sigma_{\boldsymbol{K}}^{-1}\Sigma_{ZX}\big). \quad (5.6)$$

Proof: Let be $\boldsymbol{K} = (\boldsymbol{Y}, \boldsymbol{Z})$, *then* $\Sigma_{\boldsymbol{K}} = \big((\Sigma_Y, \boldsymbol{0}), (\boldsymbol{0}, \Sigma_Z)\big)$, because $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are uncorrelated (i.e. $\Sigma_{YZ} = \boldsymbol{0}$). This implies that:

$$\Sigma = \begin{bmatrix} \Sigma_X & \Sigma_{XK} \\ \Sigma_{KX} & \Sigma_K \end{bmatrix} \text{ and } \Sigma_{\boldsymbol{K}}^{-1} = \begin{bmatrix} \Sigma_Y^{-1} & \boldsymbol{0} \\ \boldsymbol{0} & \Sigma_Z^{-1} \end{bmatrix},$$

where $\Sigma_{XK} = [\Sigma_{XY}, \Sigma_{XZ}]$. According to the Corollary 1 proposed by Ortobelli and Lando (2015), we know that:

$$E(\boldsymbol{X}|\boldsymbol{K}) = \boldsymbol{\mu_X} + \Sigma_{XK}\Sigma_{\boldsymbol{K}}^{-1}(\boldsymbol{K} - \boldsymbol{\mu_K}) \sim Ell\big(\boldsymbol{\mu_X}, \Sigma_{XK}\Sigma_{\boldsymbol{K}}^{-1}\Sigma_{KX}\big),$$

from which we get

$$E(\boldsymbol{X}|\boldsymbol{Y},\boldsymbol{Z}) = \boldsymbol{\mu_X} + \Sigma_{XY}\Sigma_Y^{-1}(\boldsymbol{Y} - \boldsymbol{\mu_Y}) + \Sigma_{XZ}\Sigma_Z^{-1}(\boldsymbol{Z} - \boldsymbol{\mu_Z}) = E(\boldsymbol{X}|\boldsymbol{Y}) + E(\boldsymbol{X}|\boldsymbol{Z}) - \boldsymbol{\mu_X}.$$

Thus, the thesis holds. Q.E.D

The proof of the Theorem 5.1 is a logical consequence of the above Proposition. As a matter of fact, we get the proof with two factors, if we assume *Y* and Z of the previous theorem two uncorrelated one-dimensional factors. The general proof follows by induction. Q.E.D

Some examples of jointly elliptical distributions that admit the decomposition $\boldsymbol{W} = \boldsymbol{\mu} + A\boldsymbol{G}$ include α stable sub-Gaussian distributions, multivariate t-Student distributions, and all distributions used for symmetric Lévy processes where *A* is the subordinator such as Normal Inverse Gaussian (NIG) symmetric vectors, Variance Gamma (VG) symmetric vectors (see among others Shoutens (2003) and Samorodnitsky and Taqqu (1994)).

From Theorem 5.1 we deduce that the dispersion of the random errors $\sigma^2_{\varepsilon(s)}$ of formula (5.4) decreases as the number of factors *s* increases when returns are elliptical distributed. Thus, Theorem 5.1 suggests that nonparametric regression analysis (5.4) maintains important intuitive properties of classic parametric regression even for elliptical vectors that do not admit finite variance (such as t-Student distribution with freedom of degrees less than 2 or stable sub-Gaussian vectors). Thus, the nonparametric approximation (5.4) is consistent even with the assumption of stable distributed returns.

5.2.2.1 Multivariate kernel methodology

Regression analysis is surely one of the most commonly used statistical techniques in economics and finance, as well as in physical sciences. Generally, it explores the relationship between a dependent variable and one (or more) explanatory or independent variables.

$$z = E(z \mid F = \boldsymbol{f}) + \varepsilon = m(\boldsymbol{f}) + \varepsilon \,. \tag{5.7}$$

where $\boldsymbol{f} = (f_1, \dots, f_s)$ is the vector of uncorrelated factors. Observe that, we can estimate the unknown parameters of $m(\boldsymbol{f})$ with several methods (e.g. least squares), if the form of the function $m(\boldsymbol{f}) = E(z \mid F = \boldsymbol{f})$ is known (e.g. polynomial, exponential, etc.). However, when the general form of $m(\boldsymbol{f})$ is not known, then it can be approximated with nonparametric techniques, as proposed by E. A. Nadaraya (1964) and G. S. Watson (1964). In this case, we relax the assumptions on the form of regression function that can be any continuous and smooth non-linear function.

According to Stanton (1997), one potentially serious problem with any parametric model is misspecification, particularly when we have no economic reason to prefer one functional form to another. The same conclusion reported by Backus et al. (1998) who show that misspecification of interest rate models can lead to serious pricing and hedging errors. However, in order to avoid the misspecification problems Ait-Sahalia and Lo (1998) show that the use of nonparametric techniques often solves the issue.

A commonly used estimator for $m(\cdot)$ of (5.7) is the multivariate version of the Nadaraya-Watson kernel estimator, which is given by:

$$\hat{m}(\boldsymbol{f}, h) = \frac{n^{-1}\sum_{i=1}^{n} K_h(\boldsymbol{f} - \boldsymbol{f}_i) z_i}{n^{-1}\sum_{j=1}^{n} K_h(\boldsymbol{f} - \boldsymbol{f}_j)}, \tag{5.8}$$

where $h = (h_1, h_2, \dots, h_s)'$ is a vector of positive real numbers commonly called bandwidths, and

$$K_h(f) = \frac{1}{h_1 h_2 \dots h_s} K\left(\frac{f_1}{h_1}, \frac{f_2}{h_2}, \dots, \frac{f_s}{h_s}\right),$$

where $K(\cdot)$ is a suitable multivariate kernel function. For further discussion of nonparametric regression analysis, see Härdle (1990) or Härdle and Müller (2000). However, the Nadaraya-Watson estimator has certain disadvantages. In particular, it corresponds to the local constant fit. To overcome these drawbacks, a general class of nonparametric regression estimator based on locally weighted least squares has been proposed (see among others Ruppert and Wand (1994)). In this context, an estimate of regression function $m(\boldsymbol{f})$ is obtained punctually by the optimal $\hat{a}$, which minimize the following criterion:

$$min_{a,b} \sum_{i=1}^{n} \{z_i - a - b^T(\boldsymbol{f}_{(i)} - \boldsymbol{f})\}^2 K_{\boldsymbol{H}}(\boldsymbol{f}_{(i)} - \boldsymbol{f}), \quad (5.9)$$

where $\boldsymbol{H}$ is $s \times s$ symmetric positive definite matrix depending on the number of observations n, $\boldsymbol{f}_{(i)}$ is the i-th observation of vector $\boldsymbol{f}$. For deeper discussion about the properties of multivariate locally weighted least squares regression, we refer the readers to Ruppert and Wand (1994). These authors derive the leading bias and variance terms for general multivariate kernel weights using weighted least squares matrix theory. This estimator has nice properties and serves as a comparison to the classical parametric estimator (OLS) in the empirical analysis. In particular, it is straightforward weighted least squares and has a solution for a and b of (5.9) given by:

$$\begin{bmatrix} a \\ b \end{bmatrix} = \left(F_f^T W_f F_f\right)^{-1} F_f^T W_f Z \text{, where } \quad F_f = \begin{pmatrix} 1 & \left(\boldsymbol{f}_{(1)} - \boldsymbol{f}\right)^T \\ \vdots & \vdots \\ 1 & \left(\boldsymbol{f}_{(n)} - \boldsymbol{f}\right)^T \end{pmatrix},$$

where $W_f = diag\left\{K_{\boldsymbol{H}}\left(\boldsymbol{f}_{(1)} - \boldsymbol{f}\right), \dots, K_{\boldsymbol{H}}\left(\boldsymbol{f}_{(n)} - \boldsymbol{f}\right)\right\}$ and $Z = [z_1, \dots, z_n]^T$. In this context, the local least squares regression has a closed solution for $m(\boldsymbol{f})$ as:

$$\hat{m}(\boldsymbol{f}) = \hat{a} = e_1^T \left(F_f^T W_f F_f\right)^{-1} F_f^T W_f Z, \quad (5.10)$$

where e_1^T is a $(s+1)$ −dimensional vector with one in the first entry and zeros in all other entries.

Several researchers have shown that the choice of kernel is not critical, while the performance of the smoothed regression function is more a question of bandwidth choice. Fan and Gijbels (1996) give a survey on bandwidth selection for univariate local polynomial smoothing technique, which contains the Nadaraya-Watson estimator as a special case. However, there is little guidance in literature on bandwidth selection for multivariate kernel density estimation, which certainly remains an important issue in empirical studies. The most widely used bandwidth selection methods are the rule-of-thumb and the plug-in bandwidth selection. In particular, the former is the normal reference rule for kernel density estimation presented in Bowman and Azzalini (1997). For general multivariate kernel estimators $K_{\boldsymbol{H}}$, where $\boldsymbol{H} = diag(h_1, \dots, h_s)$, Scott (2015) suggests the following bandwidth selectors:

$$\text{Scott's rule in } R^s\text{:} \quad \hat{h}_i = \hat{\sigma}_i \, n^{-1/(s+4)}, \quad i = 1, \dots, s \quad (5.11)$$

where $\hat{\sigma}_i$ is the usual estimate of the standard deviation of each variable f_i, and n is the sample size. This method of bandwidth selection has the property that it minimizes the mean integrated squared error (MISE) of the estimate. In the following empirical analysis, we use a normal kernel function $K_{\boldsymbol{H}}$ and we employ the bandwidth selection suggested by Scott (2015).

5.2.3 Empirical comparison between parametric and nonparametric estimators

In this section, we first compare the ex-ante errors we get considering OLS or Ruppert-Wand (hereinafter RW) regression return approximation. Secondly, we compare the ex-post wealth obtained maximizing the Sharpe ratio of differently approximated return portfolios.

In the empirical analysis, we use all active stocks on S&P 500 index from January 1, 2008 to June 10, 2016. The dataset is obtained from Thomson-Reuters DataStream, and it consists of 2126 daily observations. The choice of this period is not arbitrary; it is characterized by a higher variability due to both subprime crisis and European credit risk crisis.

5.2.3.1 Ex-ante impact of regression errors

We first examine the impact of two approximation methods considering the portfolio dimensionality reduction problems. In particular, starting from January 1, 2008 every month (20 trading days) we evaluate the errors obtained using different regression methodologies. Thus, as suggested by Ortobelli and Tichý (2015), we first perform a PCA on some correlation matrices of the stock returns in order to identify few factors with the highest return variability. Then, we compare the errors obtained regressing the returns on these principal components.

According to Section 5.2.1, we apply the PCA on four different correlation matrices here summarized:

P1: Pearson linear correlation of the gross returns z_i.

P2: Pearson linear correlation applied to the random variables $z_i - E(z_i| \Im_1)$ orthogonal to $(\Omega, \Im_1, P)$, where $\Im_1$ is the sigma algebra generated by the upper stochastic bound $max_i z_i$. The estimator used to approximate $E(z_i| \Im_1)$ is given in Section 5.2.1. For simplicity, we refer to this correlation matrix as Pearson conditional.

S1: Stable linear correlation (given by the formula (5.1)) of the gross returns z_i.

S2: Stable linear correlation applied to the random variables $z_i - E(z_i| \Im_1)$ orthogonal to $(\Omega, \Im_1, P)$. For simplicity, we refer to this correlation matrix as Stable conditional.

From the PCA we select the 20 principal components (about 4% of all active assets) with the highest return variability (for the four alternative concepts of variability). Then we investigate which approximation method (i.e. OLS and RW regression models) gives the lowest variability of the errors for every group of factors obtained by the PCA. Thus, we have to compare eight (4x2) error approximations 107 times for all asset returns of the S&P500 components. Thus, every month (20 trading days) we use a moving average window of one year (252 trading days) for the PCA computation and regression models.

Observe that, the residuals of OLS and RW regression models in (5.3) and (5.4) present null mean. Therefore, we test concave dominance among the estimated errors, in order to evaluate which residual is "greater".[12] This test allows us to determine those errors with the lowest variability.

This study essentially enables us to make three important comparisons: first between parametric estimator (i.e. OLS) and nonparametric estimator (i.e. RW), then between stable correlation matrix and the classical linear Pearson correlation, and finally between the correlation with conditional expectation (i.e. P2 and S2) and unconditional expectation cases (i.e. P1 and S1). In particular, Table 5.1 reports the average (over the time) of the average percentage of regression errors with higher variability (in the sense of concave dominance) of the criterion on the raw with respect to the criterion on the column. We point out with $n.c$ (not comparable) the elements on the diagonal of Table 5.1.

Table 5.1: Average percentage of regression errors with higher variability considering two regression models (i.e. OLS and RW) applied to few factors obtained by the PCA of four alternative correlation matrices (P1, P2, S1, S2).

	OLS Pearson	**OLS Pearson conditional**	**OLS Stable**	**OLS Stable conditional**	**RW Pearson**	**RW Pearson conditional**	**RW Stable**	**RW Stable conditional**
OLS Pearson	$n.c$	0.0058	0.01	0.0097	0.4939	0.5032	0.5041	0.513
OLS Pearson conditional	0.0045	$n.c$	0.0096	0.0098	0.4839	0.4977	0.5037	0.507
OLS stable	0,0468	0,0474	$n.c$	0,0045	0.4639	0.4938	0.5014	0.503
OLS Stable conditional	0.0467	0.0473	0.0041	$n.c$	0.4589	0.4837	0.4937	0.505
RW Pearson	0	0	0	0	$n.c$	0.337	0.4628	0.4627
RW Pearson conditional	0	0	0	0	0.0689	$n.c$	0.4621	0.4625
RW stable	0	0	0	0	0.0201	0.0205	$n.c$	0.2245
RW Stable conditional	0	0	0	0	0.0203	0.0208	0.2023	$n.c$

From Table 5.1 we observe that:

- The errors we get with nonparametric regression analysis (i.e. RW) never present higher variability than those obtained with the classical parametric regression analysis (i.e.

[12] For more details on stochastic dominance orders and tests see among others Muller and Stoyan (2002), Davidson and Jean-Yves (2000).

OLS). Moreover, we test that about the 50% of OLS error approximation present higher variability (in the sense of concave order) than RW error approximation.

- When we compare the RW regression errors for different concepts of variability, we observe that the same number of the principal components (20) obtained with a PCA on a stable type correlation matrix are often able to explain a higher variability than the principal components obtained with a PCA on a Pearson type correlation matrix. Indeed, we test that about the 45% of RW error approximation based on "Pearson type factors" present higher variability of RW regression errors for "Stable type factors". While we do not observe a significant supremacy in terms of higher variability using different correlation matrices, when we compare the OLS regression errors using the same number of the principal components (20) for different concepts of variability.
- Finally, comparing the regression errors obtained regressing the returns on conditional and unconditional correlation matrices, we observe that the same number of the principal components obtained with a PCA on conditional-type correlation matrices are often able to explain a higher variability than the principal components obtained with a PCA on unconditional-type correlation matrices. From this surprising result, we deduce the importance of identifying the right factors representing the explainable variability.

Overall, the best results are obtained when we approximate the return regressing them on few factors obtained by a PCA applied to a stable conditional correlation matrix (i.e. S2) using the nonparametric RW regression model.

5.2.3.2 On the ex-post wealth obtained with different return approximating methodologies

In order to evaluate the impact of different approximation methodologies on large-scale portfolio problems, we suggest to compare the ex-post wealth obtained with a classical strategy when we approximate in different ways the returns. We use the same database introduced in section 5.2.3. In particular, we compare the ex-post wealth obtained maximizing the Sharpe ratio,

a) Either on returns approximated regressing them on few factors, selected using the Random Matrix Theory criterion,[13] obtained by a PCA applied to a stable conditional correlation matrix (i.e. S2) and using the nonparametric RW regression model.
b) Or on returns approximated regressing them on few factors obtained by a PCA (i.e. P1)

[13] We use as factors the portfolios whose weights are orthonormal eigenvectors of the correlation matrix with eigenvalue (i.e. variance of factors) greater than $\left(1+\sqrt{\frac{number\ of\ assets}{number\ of\ observations}}\right)^2$ see, among others, Coqueret and Milhau 2014, Daly et al. (2008) and references therein.

and using the OLS regression model (classical strategy used by practitioners).

The ex-post wealth obtained with the two strategies is also compared with the wealth we get with a "take and hold" strategy on the market index S&P 500 (that is the classic benchmark of the market.). Let us briefly formalize the portfolio performance measure (Sharpe ratio) used in this empirical analysis.

Sharpe ratio (1994). The Sharpe ratio is used to characterize how well the return of an asset compensates the investor for the risk taken. The Sharpe ratio computes the price for unity of risk, by subtracting the risk-free rate from the rate of return of the portfolio and then dividing the result by the standard deviation of the portfolio returns. Formally:

$$SR(x'z) = \frac{E(x'z) - z_0}{\sigma_{x'z}}, \tag{5.12}$$

where, $E(x'z)$ is the portfolio expected returns, z_0 is the risk-free gross return and $\sigma_{x'z}$ is the portfolio standard deviation.

In this analysis the riskless is not allowed, i.e. $z_0 = 1$. Moreover, we recalibrate the portfolio every three months (60 trading day) using a moving average window of one year (252 trading days) for approximating the returns and computing the optimal portfolios. Starting with an initial wealth $W_0 = 1$ that we invest on January 1, 2008, we evaluate the ex-post wealth sample paths obtained from the two compared cases. Thus, at the *k*-th recalibration time, the following steps are performed for Sharpe strategies:

Step 1. On the one hand, apply PCA to Pearson correlation matrix (i.e. P1) and then approximate the gross returns by OLS estimator. On the other hand, apply PCA to the stable conditional correlation matrix (i.e. S2) and then approximate the gross returns by RW estimator. In both cases, we select the number of principal components factors using the Random Matrix Theory criterion (see footnote 13).

Step 2. Determine the market portfolio $x_M^{(k)}$ that maximizes the performance ratio $\rho(x'z)$ applied to the approximated returns:

$$\begin{aligned} &max_x \; \rho(x'z) \\ &\text{s.t. } \textstyle\sum_{i=1}^{n} x_i^{(k)} = 1, \\ &x_i^{(k)} \geq 0 \quad i = 1, \dots, n \end{aligned} \tag{5.13}$$

Here the performance measure $\rho(x'z)$ is the Sharpe ratio (5.12). The maximization of the Sharpe ratio can be solved as a quadratic-type problem (see Stoyanov et al. 2007).

Step 3. Since we recalibrate the portfolio every 60 trading days, we calculate the ex-post final wealth as follows:

$$W_{t_{k+1}} = \left(W_{t_k} - t.c._{t_k}\right)\left(x_M^{(k)}\right)' z_{t_{k+1}}^{(expost)}, \tag{5.14}$$

where $t.c._{t_k}$ are the proportional transaction costs of 20 basis points and $z_{t_{k+1}}^{(expost)}$ is the vector of observed gross returns in the period between t_k and t_{k+1}, such that $t_{k+1} = t_k + 60$.

We apply the algorithm until the observations are available. The results of this analysis are reported in Figure 5.1.

Figure 5.1: Ex-post wealth obtained by Sharpe ratio either with Pearson correlation matrix and OLS regression model (simply SR (OLS)) or with stable conditional correlation and RW regression model (simply SR (RW)), compared to the S&P 500 benchmark.

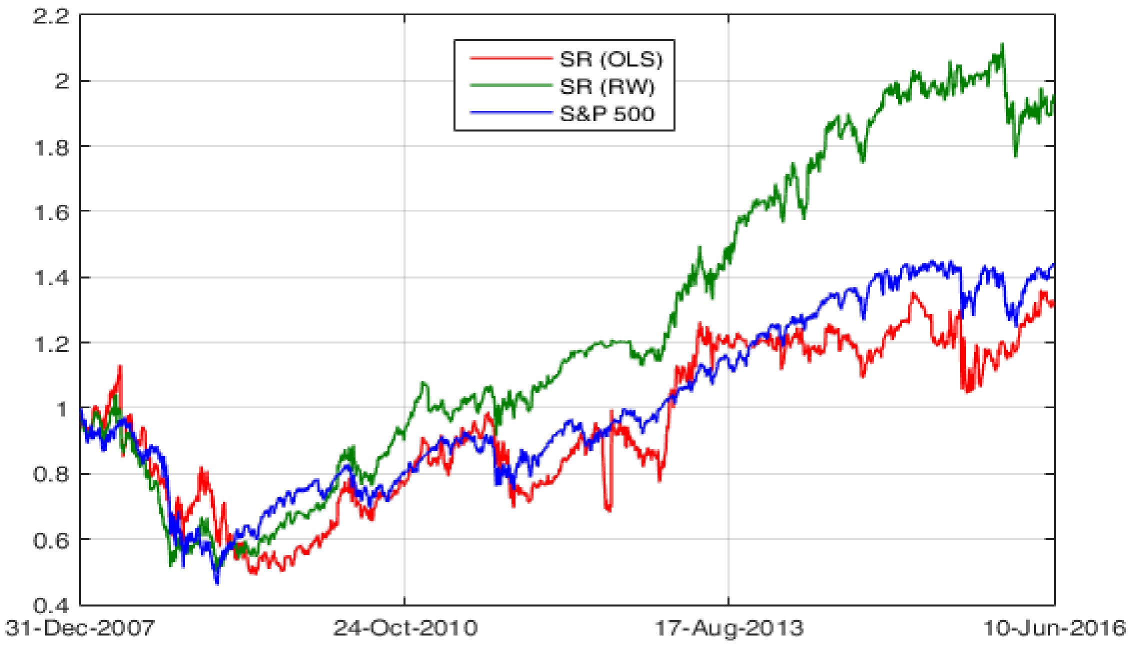


Figure 5.1 reports the ex-post wealth evolution obtained by the two strategies (i.e. SR (RW) and SR (OLS)) and the S&P 500 benchmark. Clearly, the SR (RW) outperforms both the strategy based on SR (OLS) and the S&P 500 benchmark. Furthermore, we observe even S&P 500 benchmark is slightly better than SR (OLS). For further confirmation, we examine the ex-post log-returns obtained with the three strategies, and we check whether there exist stochastic dominance relationships between them. In particular, we test whether a portfolio strategy is preferable (with preference relation $\succ$) to another one from the point of view of different classes of investors. In particular, we evaluate the performances for all non-satiable investors (first-order dominance - FSD), all non-satiable risk-averse investors (second-order dominance - SSD), all non-satiable risk-averse investors with preference of positive skewness (third-order dominance - TSD) (see, among others, Muller and Stoyan (2002) Davidson and Jean-Yves (2000)). The results of this analysis are reported in the following Table.

Table 5.2: Dominance relationships among optimal portfolios obtained applying SR (OLS), SR (RW) and S&P 500 benchmark.

	SR (OLS)	**SR (RW)**	**S&P 500**
SR (OLS)	$\boldsymbol{n.c}$	≺SSD	≺SSD
SR (RW)	≻SSD	$\boldsymbol{n.c}$	≻SSD
S&P 500	≻SSD	≺SSD	$\boldsymbol{n.c}$

We do not find First-order stochastic dominance relationships. However, we find that SR (RW) dominates both the SR (OLS) and the S&P 500 benchmark in terms of second-order stochastic dominance (SSD) (and thus also for the TSD). Moreover, both the SR (RW) and the S&P 500 benchmark dominate the SR (OLS) in the SSD sense. Therefore, these results provide strong support for the use of the RW regression model in the context of the portfolio theory.

In Table 5.3, we report eight different statistics of the ex-post log returns obtained with the three strategies. In particular, we compute the mean, standard deviation, skewness, kurtosis, VaR_α, $CVaR_\alpha$, Sharpe ratio, and the performance measure $STARR_\alpha$ defined by Martin et al. (2003) as follows

$$STARR_\alpha(x'z) = \frac{E(x'z)}{CVaR_\alpha(x'z)}, \tag{5.15}$$

with a confidence level $1 - \alpha = 95\%$. STARR ratio allows us to overcome the drawbacks of the standard deviation as a risk measure (Artzner et al. (1999)) and focuses on the downside risk.[14]

Table 5.3: Average of some statistics of the ex-post returns obtained maximizing the Sharpe ratio (SR) with two different correlation matrices (Pearson and Stable conditional) and two different approximation methods (i.e. OLS and RW).

	Mean	**St dev**	**Skewness**	**Kurtosis**	**VaR 5 %**	**CVaR 5 %**	**Final W**	**Sharpe**	**STARR**
SR (OLS)	0.033%	2.015%	3.6586	8.8133	2.655%	4.491%	1.3226	1.628%	0.731%
SR (RW)	0.041%	1.351%	0.0809	7.3628	2.037%	3.265%	1.9580	3.022%	1.251%
S&P 500	0.027%	1.380%	-0.06366	9.7897	2.087%	3.399%	1.4407	1.936%	0.786%

From Table 5.3, we observe that:

- The SR (OLS) strategy presents the highest risk (standard deviation, VaR 5%, CVaR 5%) and the lowest finale wealth, Sharpe (mean/St dev.) and STARR performance.

[14] STARR ratio is not a symmetric measure of risk when returns present heavy-tailed distributions; see Martin et al. (2003).

- The SR (RW) strategy is performing much better, than the other strategies, presenting the highest results in terms of the mean, final wealth, Sharpe (mean/St dev.) and the STARR performance, and more interestingly the lowest risk (standard deviation, VaR 5%, CVaR 5%).

Generally, the worst result is obtained with the SR (OLS) strategy. This is not surprising, since it confirms that a better approximation (observed in Section 5.2.3.1) implies better choices. In practice, recalibration time period and moving average window, used to compute the optimal portfolios, are very important parameter choices that should be accounted for by the portfolio manager. For this reason, we enrich our analysis considering some other choices of these parameters. The results of this analysis are reported in Table 5.4. For example OLS (126-10) stands for the SR (OLS) strategy when: "we use a moving average window of 126 trading days for the computation of each optimal portfolio and we recalibrate the portfolio every 10 trading days".

Table 5.4: Average of some statistics of the ex-post returns obtained maximizing the SR (OLS) and the SR (RW) with different parameter choices of moving average window and recalibration time.

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5 %	Final W	Sharpe	STARR
OLS (126-10)	-0.006%	1.927%	0.1933	14.579	2.728%	4.819%	0.5873	-0.331%	0.132%
OLS (126-20)	-0.026%	1.967%	0.2853	14.534	2.662%	4.943%	0.3813	-1.317%	-0.524%
OLS (126-60)	0.023%	1.785%	-0.6701	7.7633	2.818%	4.598%	1.1525	1.280%	0.497%
OLS (252-10)	-0.001%	1.715%	-0.7689	12.168	2.522%	4.379%	0.7051	-0.087%	-0.034%
OLS (252-20)	0.015%	1.467%	-0.2461	11.309	2.157%	3.555%	1.0853	1.005%	0.415%
OLS (252-60)	0.033%	2.015%	3.6586	8.1338	2.655%	4.491%	1.3226	1.628%	0.731%
OLS (500-60)	0.031%	1.835%	0.2717	6.1445	2.712%	4.326%	1.3501	1.690%	0.717%
RW (126-10)	0.018%	1.360%	-0.0569	10.000	1.942%	3.263%	1.2077	1.341%	0.559%
RW (126-20)	0.039%	1.386%	-0.0100	10.928	1.876%	3.317%	1.8489	2.787%	1.165%
RW (126-60)	0.034%	1.362%	-0.2035	5.8101	2.114%	3.364%	1.6723	2.466%	0.998%
RW (252-10)	0.013%	1.314%	-0.2190	11.664	1.976%	3.180%	1.0916	0.980%	0.405%
RW (252-20)	0.023%	1.306%	-0.3989	7.4001	2.030%	3.284%	1.3567	1.762%	0.701%
RW (252-60)	0.041%	1.351%	0.0809	7.3628	2.037%	3.265%	1.9580	3.022%	1.251%
RW (500-60)	0.048%	1.351%	-0.0432	4.9902	2.123%	3.307%	2.2814	3.556%	1.453%

From Table 5.4, we clearly observe the impact of these parameter choices (i.e. moving average window and portfolio recalibration time) on the ex-post wealth obtained by the examined strategies. Most importantly, the SR (RW) presents better results than the SR (OLS) for the same parameter combination. In particular, the SR (RW) achieves the highest mean, final wealth, Sharpe ration, STARR performance, and also the lowest risk (standard deviation, VaR 5%, CVaR 5%).

Overall, from Figure 5.1 and Tables 5.1, 5.2, 5.3 and 5.4, we deduce that it is better using: a) the stable conditional correlation matrix with respect to the classical Pearson linear correlation for determining the main factors, and; b) nonparametric regression approximation of the returns rather than the OLS parametric regression.

5.3 Optimal portfolio choices accounting the heavy tails

Although the Sharpe ratio is fully compatible with normally distributed returns, it will lead to incorrect investment decision when returns present heavy-tailed distribution, see among others Biglova et al. (2004). Several alternatives to the Sharpe ratio for optimal portfolio selection have been proposed over the years, for an overview see among others Farinelli et al. (2008) and Rachev et al. (2008). The common aspect of these alternative reward-risk approaches consists in taking into account the expected losses and the expected gains. However, these approaches do not consider wholly the losses and the gains and their probability of realization.

In this context, we propose new performance measures based on the conditional expectation that takes into account the portfolio distributional behavior on the tails. More specifically, the first suggested performance measure is based on two different σ-algebras (the σ-algebra generated by the portfolio losses, and the σ-algebra generated by the portfolio profits). While the second performance measure considers σ-algebras generated by the joint losses and by joint gains of all assets in the market. These sigma algebras are approximated using sigma algebras generated by proper partitions (of losses or of gains).

Let us consider the first performance measure based on losses and gains of the return portfolio. Formally, for approximating the σ-algebra generated by losses we consider the partition $\{A_j\}_{J=1}^{d} = \{A_1, \dots, A_d\}$, that accounts the portfolio losses less than $(d/100)$-percentile of portfolio (where d is an integer number greater than 1), in d subsets as follows:

- $A_1 = \{x'z \le F_{x'z}^{-1}(0.01)\}$,
- $A_h = \{F_{x'z}^{-1}\left(\frac{h-1}{100}\right) < x'z \le F_{x'z}^{-1}\left(\frac{h}{100}\right)\}$, for $h = 2, \dots, d-1$ (5.16)
- $A_d = \{x'z > F_{x'z}^{-1}\left(\frac{d-1}{100}\right)\}$.

Clearly, the sets A_i depend on the portfolio weights x and thus the σ-algebra generated by these sets depends on portfolio weights. We point out the sigma algebra with $\Im_L(x) = \sigma(A_1, \dots, A_d)$ that is our approximation of the σ-algebra generated by all the portfolio losses. While for approximating the σ-algebra generated by the portfolio profits, we consider the following partition $\{B_j\}_{J=1}^{u} = \{B_1, \dots, B_u\}$, which accounts the possible portfolio profits greater than $((100-u)/100)$-percentile of portfolio (where u is an integer number greater than 1), in subsets as follows:

- $B_1 = \{x'z \geq F_{x'z}^{-1}(0.99)\}$,
- $B_h = \{F_{x'z}^{-1}\left(\frac{100-h}{100}\right) \leq x'z < F_{x'z}^{-1}\left(\frac{101-h}{100}\right)\}$, for $h = 2, \dots, u-1$ (5.17)
- $B_u = \{x'z < F_{x'z}^{-1}\left(\frac{101-u}{100}\right)\}$.

Thus, we point out with $\Im_P(x) = \sigma(B_1, \dots, B_u)$ the σ-algebra generated by the portfolio profits. These two σ-algebras contain the information of portfolio losses and gains that must be considered entirely in portfolio performance measure. To consider these losses, we calculate the portfolio net returns conditional on σ-algebra $\Im_L(x)$ and conditional on σ-algebra $\Im_P(x)$. In other words, we approximate the portfolio of net returns (considering the σ-algebras generated by profits and losses) as follows $y_P = E(x'z|\Im_P(x)) - 1$ and $y_L = E(x'z|\Im_L(x)) - 1$. These random variables give us information of periodic (say daily) portfolio losses and gains. So given T historical observations, we can consider the wealth generated by the gains given by $W_{P,T} = \prod_{t=1}^{T}(1+y_P)_t$ and the discount we apply if we consider the losses given by $W_{L,T} = \prod_{t=1}^{T}(1-y_L)_t$. Generally, each investor wants to maximize $W_{P,T}$ and minimize $W_{L,T}$. Therefore, we suggest to maximize the following performance ratio:

$$TOK(x'z) = \frac{W_{P,T}}{W_{L,T}}. \quad (5.18)$$

The proposed TOK ratio is a very flexible performance measure that considers the expected portfolio returns given the σ-algebra $\Im_P(x)$ generated by the portfolio profits and the σ-algebra $\Im_L(x)$ generated by the portfolio losses. The main advantage of the TOK ratio with respect to other performance measures is that we can account of all the losses and gains obtained in a given period $(0, T)$ maintaining the flexibility in choosing the partition that generated the σ-algebras $\Im_L$ and $\Im_P$. For example, we could consider σ-algebras $\Im_P$ and $\Im_L$ generated by a limited number of sets. In the empirical analysis, we consider this performance ratio with different parameter choices (e.g. $d = u = 5, 7, 10$ in formulas (5.16) and (5.17)).

Moreover, in view of this approach, we propose a second performance measure that considers one important feature of the market, which is the joint losses and gains among the assets. Let $L_i^{(q1)} = \{z_i \leq F_{z_i}^{-1}(q_1)\}$ be a set of the losses bigger than Value at Risk ($VaR_{q_1}(z_i) = -F_{z_i}^{-1}(q_1)$) of i-th asset and consider the function that counts the total number of assets which are jointly losing more than their $VaR_{q_1}(z_i)$, i.e., $g(\omega) = \sum_{i=1}^{\#assets} I_{L_i^{(q_1)}}(\omega)$ (in this study we use $q_1 = 0.01$). Thus, we know that for larger values of function g we are considering the joint risk of loss existing in the market. For this reason, we consider the "bear market σ-algebra" $\Im_{Loss}$ generated by the following *m* sets (where *m* is an integer number greater than 1):

- $A_1 = \{\omega: g(\omega) \geq F_g^{-1}(0.99)\}$,
- $A_i = \left\{\omega: F_g^{-1}\left(\frac{100-i}{100}\right) \leq g(\omega) < F_g^{-1}\left(\frac{101-i}{100}\right)\right\}$, for $i = 2, \dots, m-1$, (5.19)
- $A_m = \left\{\omega: g(\omega) < F_g^{-1}\left(\frac{101-m}{100}\right)\right\}$.

Observe that in this case the sets A_i do not depend on the portfolio weights. Moreover, the σ-algebra $\Im_{Loss} = \sigma(A_1, \dots, A_m)$ generated by the joint losses take into account the period in which the market appears in a ''bear" period. To get some insight, the term "bear market" describes the downward trend of stock index or negative stock index returns over a period.

Analogously, let $G_i^{(q_2)} = \{z_i \geq F_{z_i}^{-1}(q_2)\}$ be the set of the gains bigger than q_2-th percentile of the *i*-th asset. Then, the function that counts the total number of assets whose returns are bigger than their q_2-th percentile is given by: $f(\omega) = \sum_{i=1}^{\#assets} I_{G_i^{(q_2)}}(\omega)$ (in this study, we use $q_2 = 0.99$). Thus large values of the function f summarize the idea of a ''bull market'' period. The term bull market describes the upward trend of stock index or positive stock index returns over a period. For this reason, we consider the bull market σ-algebra $\Im_G$ generated by the following *s* sets:

- $B_1 = \{\omega: f(\omega) \geq F_f^{-1}(0.99)\}$,
- $B_i = \left\{\omega: F_f^{-1}\left(\frac{100-i}{100}\right) \leq f(\omega) < F_f^{-1}\left(\frac{101-i}{100}\right)\right\}$, for $i = 2, \dots, s-1$, (5.20)
- $B_s = \left\{\omega: g(\omega) \leq F_f^{-1}\left(\frac{101-s}{100}\right)\right\}$.

Thus, the σ-algebra $\Im_G = \sigma(B_1, \dots, B_s)$ generated by the joint profits takes into account the period in which the market appears in a ''bull' period. In this way, we approximate the portfolio net returns (considering σ-algebras $\Im_{Loss}$ and $\Im_G$) as follows $y_{Loss} = E(x'z|\Im_{Loss}) - 1$ and $y_G = E(x'z|\Im_G) - 1$. These random variables give us information of periodic (say daily) market joint losses and gains. Thus, given *T* historical observations, we can consider the wealth generated by the joint gains given by $W_{G,T} = \prod_{t=1}^{T}(1 + y_G)_t$ and the discount we apply if we consider the joint losses given by $W_{Loss,T} = \prod_{t=1}^{T}(1 - y_{Loss})_t$. In this case, each investor aims to maximize $W_{G,T}$ and minimize $W_{Loss,T}$. Therefore, we maximize the following performance ratio:

$$Joint\_TOK(x'r) = \frac{W_{G,T}}{W_{Loss,T}}. \quad (5.21)$$

This performance measure is, theoretically appealing approach to portfolio selection problems, based on the conditional expectation that takes into account the heavy tails of all return jointly. In particular, it considers an important feature of the market that should be accounted by the portfolio managers (i.e. the joint losses among the assets).

The earliest definitions of returns (i.e. y_P , y_G, y_L and y_{Loss}) are based on the conditional expected value with respect to a given σ-algebra. To evaluate this conditional expected value we use an alternative nonparametric approach, called "OLP", described in Ortobelli et al. (2015). This method consider a σ-algebra $\Im_k = \sigma(A_1, \dots, A_k)$ generated by a partition $A_1, \dots, A_k$. The conditional expected value of a random variable Y with respect σ-algebra $\Im_k$ is simply given by $E(Y|\Im_k)(\omega) = \sum_{j=1}^{\mathrm{k}} E(Y|A_j)1_{A_j}(\omega)$ a.s. and $1_{A_j}(\omega) = \begin{cases} 1 & \omega \in A_j \\ 0 & \omega \notin A_j \end{cases}$.

Therefore, on the one side, given *N* i.i.d. observations of *Y*, we get that $\frac{1}{n_{A_j}} \sum_{y \in A_j} y$ (where n_{A_j} is the number of elements of A_j) is a consistent estimator of $E(Y|A_j)$. On the other side, if we know that the probability p_i is the probability of the *i*-th outcome y_i of random variable *Y*, we get $E(Y|A_j) = \sum_{y_i \in A_j} y_i p_i / Pr(A_j)$, otherwise, we can give a uniform weight to each observation, which yields the following consistent estimator of $E(Y|A_j)$=$\frac{\mathbf{1}}{\boldsymbol{n_{A_j}}} \sum_{\boldsymbol{y_i \in A_j}} \boldsymbol{y_i}$, where n_{A_j} is the number of elements of A_j. Therefore, we have a consistent estimator of $E(Y|\Im_k)$.

5.3.1 An ex-post empirical analysis according to conditional expectation estimation

In this section, we compare the optimal portfolio approaches solved according to the new performance measures (i.e. TOK and Joint_TOK ratios) with different parameter choices (e.g. $d = u = m = s = 5, 7, 10$ in formulas (5.16), (5.17), (5.19) and (5.20)). The proposed empirical analysis allows us to evaluate the ex-post wealth of the optimal portfolios when we consider the new performance measures, stable conditional correlation matrix, and different approximation methods (i.e. RW and OLP). We use the same database introduced in section 5.2.3. In particular, we employ the proposed techniques to reduce the dimensionality of large-scale portfolio problems, see section 5.2. Specifically, we perform PCA on stable conditional correlation matrix (i.e. S2) of the returns; then we approximate the portfolio returns using RW regression model. Moreover, in order to evaluate the conditional expected value with respect to a given σ-algebra, we use the OLP estimator as suggested in Section 5.3.1.

We recalibrate the portfolio monthly (i.e. every 20 trading days). No short selling is allowed (i.e., $x_i \geq 0$) and, in order to guarantee enough diversification for all portfolio problems we assume that we cannot invest in more than 20% in a single stock (i.e., $x_i \leq 0.2$).

We also consider proportional transaction cost of 20 basis points.

We use a moving average window of two years (500 trading days) for the computation of each optimal portfolio, and we recalibrate the portfolio every month (20 trading days). Starting with an initial wealth $W_0 = 1$ that we invest on January 1, 2008, we evaluate the ex-post wealth sample paths obtained maximizing the new performance measures. Therefore, at the k-th optimization, three steps are performed to compute the ex-post final wealth:

Step 1. Apply the PCA to the stable conditional correlation matrix (i.e. S2). Then, approximate the portfolio returns as suggested in sections 5.2 (via RW estimator).

Step 2. Determine the market portfolio $x_M^{(k)}$ that maximizes the performance ratio $\rho(x)$ applied to the approximated returns:

$$\begin{aligned} &max_x \, \rho(x'z) \\ &\text{s.t. } \textstyle\sum_{i=1}^{n} x_i^{(k)} = 1, \\ &0 \le x_i^{(k)} \le 0.2 \quad i = 1, \dots, n \end{aligned} \tag{5.22}$$

Here the performance measure $\rho(x'z)$ is either TOK performance measure (5.18) or Joint_TOK ratio (5.21). These optimization problems may have several local optima. In order to overcome this limitation, we use as a starting point the optimal solution obtained with Angelelli and Ortobelli heuristic algorithm (see Angelelli & Ortobelli (2009)). We then improve this solution by applying the heuristic function pattern search implemented in Matlab 2015 to solve global optimization problems.

Step 3. Compute the ex-post final wealth as follows:

$$W_{t_{k+1}} = \left(W_{t_k} - t.c._{t_k}\right)\left(x_M^{(k)}\right)' z_{t_{k+1}}^{(expost)}, \tag{5.23}$$

where $t.c._{t_k}$ are the proportional transaction costs of 20 basis points and $z_{t_{k+1}}^{(expost)}$ is the vector of observed gross returns in the period between t_k and t_{k+1}, such that $t_{k+1} = t_k + 20$.

We apply the algorithm until the observations are available. The results of these analyses are reported in Figures 5.2 and 5.3.

Figure 5.2: Ex-post wealth obtained with TOK performance measure with different parameter choices ($u = d = 5, 7, 10$ for simplicity, we write TOK (5), etc.) compared with S&P 500 benchmark.

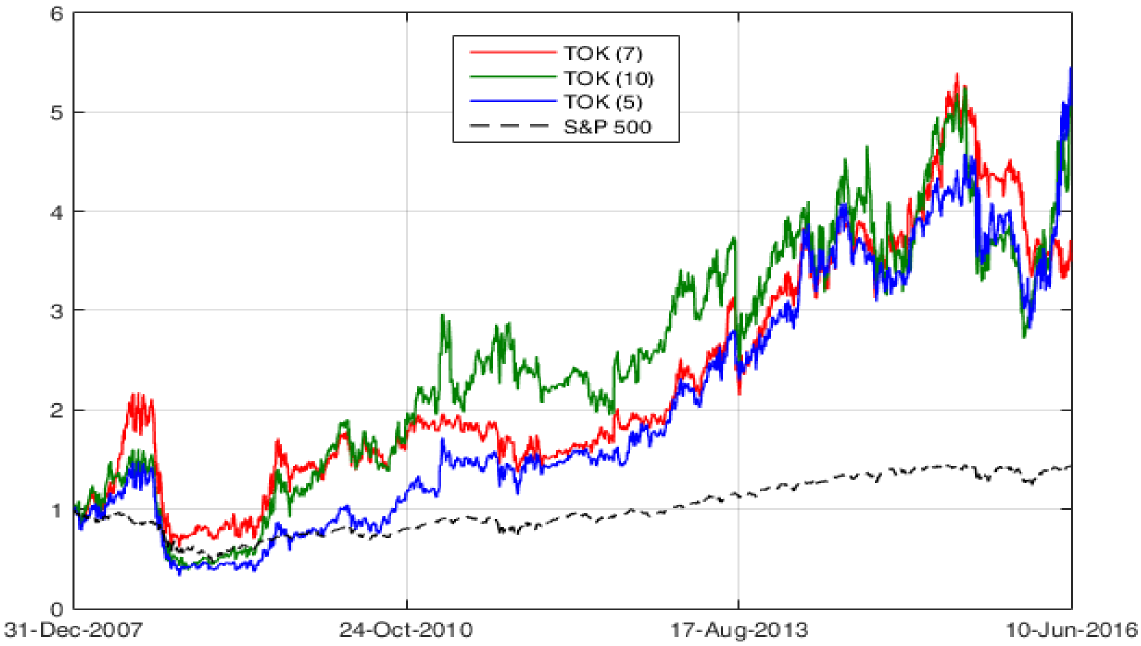


Figure 5.2 reports the sample paths of the ex-post wealth obtained maximizing TOK performance measure with three different parameter choices (i.e. ($u = d = 5, 7, 10$ in formulas (5.16) and (5.17)) and the S&P 500 benchmark. Clearly, we observe that all strategies based on TOK performance are performing much better than the S&P 500 benchmark. On the one hand, the TOK performance with $u = d = 10$ (simply TOK (10)) seems the best-performing strategy among other strategies. While the strategies TOK (5) and TOK (7) allocate in the second position. On the other hand, from Figure 5.2, we clearly observe that TOK performance measure is performing much better than the classical Sharpe ratio observed in Figure 1 of section 5. 2.3.

In Table 5.5, we present summary statistics (mean, standard deviation, skewness, kurtosis, VaR 5%, CVaR 5%, final wealth) for the ex-post log returns obtained maximizing the TOK ratio with stable conditional correlation matrix, and using two estimators of the conditional expectation (i.e. OLP and RW). Furthermore, we compute the Sharpe ratio and the performance measure $STARR_{\alpha}$ (with $\alpha = 5\%$).

Table 5.5: Average of some statistics of the ex-post returns obtained by maximizing the TOK ratio for different parameters

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5 %	Final W	Sharpe	STARR
TOK (5)	0.125%	3.108%	-0.1068	5.6820	4.537%	7.368%	5.0636	4.021%	1.696%
TOK (7)	0.105%	3.004%	-0.2494	7.3197	4.382%	7.222%	3.4974	3.479%	1.417%
TOK (10)	0.145%	3.706%	0.2328	9.4630	5.394%	8.744%	5.0957	3.925%	1.663%
S&P 500	0.027%	1.380%	-0.0636	9.7897	2.087%	3.399%	1.4407	1.936%	0.786%

From Table 5.5, we observe that:

- The TOK (10) strategy presents the greatest average and final wealth, while the TOK (5) strategy achieves the highest Sharpe ratio and STARR performance. However, both strategies show the highest risk (standard deviation, VaR 5%, CVaR 5%).
- The strategies based on TOK performance measure are significantly greater than the S&P 500 benchmark in terms of the mean, final wealth, Sharpe ratio, and STARR performance. Nevertheless, compensating, the S&P 500 index achieves the lowest risk (standard deviation, VaR 5%, CVaR 5%) among all strategies.
- The ex-post returns are strongly leptokurtic for all strategies presented in Table 5.5. In addition, all strategies (expect S&P 500 benchmark) show a slight asymmetry, since the skewness is different from zero.

Interestingly, these preliminary results give us a general overview about the TOK performance measure using different parameter choices (i.e. $u = d = 5, 7, 10$). Moreover, we also test if the observed ex-post results can be ordered from the point of view of some classes of investors. In particular, we examine the ex-post log-returns obtained with different strategies, and we check whether there exist a stochastic dominance relationship between the ex-post log-returns of the optimal portfolio strategies. Thus, we evaluate and test the dominance for all non-satiable investors (first-order dominance - FSD), all non-satiable risk-averse investors (second-order dominance - SSD), all non-satiable risk seeker investors (increasing-convex-order - ICX) (see, among others, Muller and Stoyan (2002) Davidson and Jean-Yves (2000)). We test for first-order (FSD), second-order (SSD) and increasing-convex-order (ICX) (see, among others, Davidson and Jean-Yves (2000)). The ICX ordering (generally less used in financial decision problems) accounts for the choice of non-satiable risk-seeking investors according to Muller and Stoyan (2002) studies.

Table 5.6: Dominance relationships between optimal portfolios obtained applying TOK performance measure with three different parameter choices (i.e., $\boldsymbol{u = d = 5, 7, 10}$) and the S&P 500 index.

	TOK (5)	TOK (7)	TOK (10)	S&P 500
TOK (5)	$\boldsymbol{n.c}$	$n.c$	<ICX	>ICX
TOK (7)	$n.c$	$\boldsymbol{n.c}$	<ICX	>ICX
TOK (10)	>ICX	>ICX	$\boldsymbol{n.c}$	>ICX
S&P 500	<ICX	<ICX	<ICX	$\boldsymbol{n.c}$

In this analysis, we never observe a preference for some strategy by all non-satiable risk-averse investors (i.e. SSD). We always observe that all non-satiable risk seeker investors will choose the TOK (10) strategy. Moreover, as one could expect, all strategies with TOK performance dominate the S&P 500 benchmark from the point of view of non-satiable risk seeker investors. On the one hand, from Figure 5.2 and Tables 5.5 and 5.6, we conclude that TOK performance strategies (especially TOK (10) and TOK (5)) are performing much better than S&P 500 benchmark. On the other hand, according to the empirical analysis of section 5.2.3, we deduce that TOK performance gives a better result than the classical Sharpe ratio. Overall, these results confirm and provide strong support for the TOK performance measure and to the proposed methodology to reduce the dimensionality of the problem.

Another important feature of the market, which should be accounted by asset managers, is the joint risk of loss that is the basic of the systemic risk. When we consider this important aspect using Joint_TOK ratio, as suggested in section 5.3, we report the results in Figure 5.3 and Table 5.7.

Figure 5.3: Ex-post wealth obtained with Joint_TOK performance measure with different parameter choices ($m = s = 3, 5, 10$, simply we write Joint_TOK (5), etc.) compared with S&P 500 benchmark.

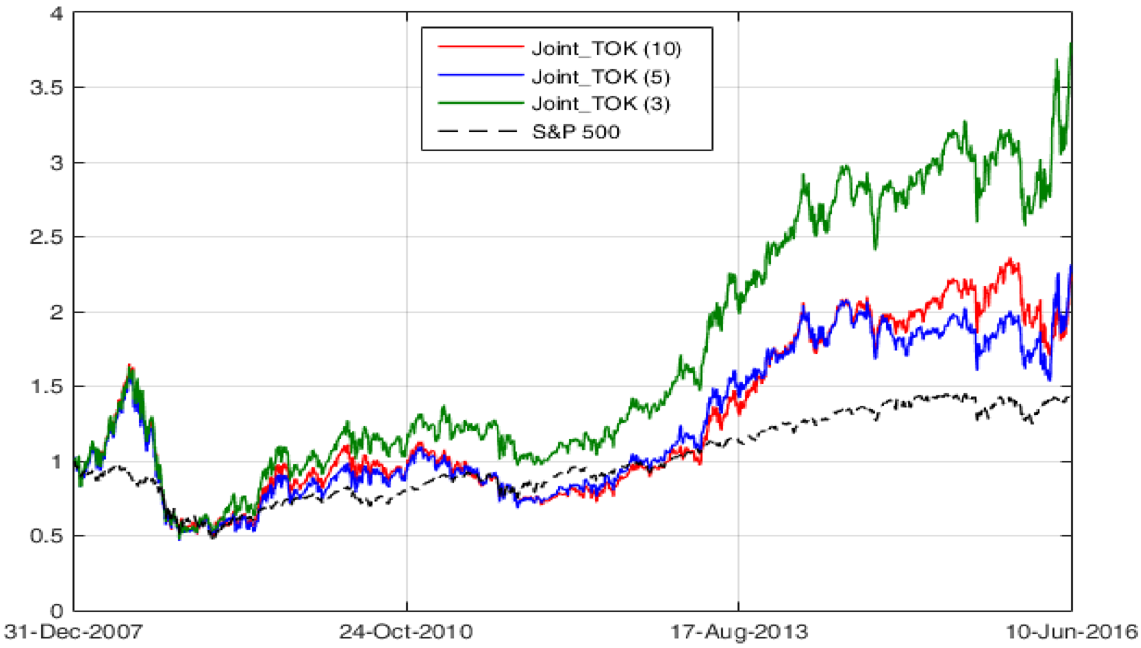


Figure 5.3 reports the ex-post wealth evolution obtained maximizing Joint_TOK performance measure with three different parameter choices (e.g. $m = s = 3, 5, 10$ in formulas (5.19) and (5.20)) and the S&P 500 benchmark. On the one hand, we clearly observe that the Joint_TOK (3) is the best performing strategy among the other strategies. On the other hand, all strategies based on the Joint_TOK performance measure are performing better than the S&P 500 benchmark. Overall, the Joint_TOK performance measure gives promising results while considering all possible joint losses and joint gains present in the market.

Table 5.7 shows summary statistics for the ex-post wealth obtained maximizing Joint_TOK ratio with a Stable conditional correlation matrix. In particular, we consider mean, standard deviation, skewness, kurtosis, VaR 5%, CVaR 5%, final wealth, Sharpe ratio and $STARR_\alpha$ performance (with $\alpha = 5\%$).

Table 5.7: Average of some statistics of the ex-post returns obtained by maximizing the Joint_TOK ratio with stable conditional correlation matrix and two approximation methods (i.e. OLP and RW).

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5 %	Final W	Sharpe	STARR
Joint_TOK (3)	0.092%	2.448%	-0.3735	3.5074	4.061%	6.101%	3.7225	3.763%	1.510%
Joint_TOK (5)	0.068%	2.437%	-0.3520	3.7471	3.915%	6.138%	2.2647	2.809%	1.115%
Joint_TOK (10)	0.067%	2.415%	-0.3083	3.6887	3.879%	6.084%	2.2071	2.671%	1.096%
S&P 500	0.027%	1.380%	-0.06366	9.7897	2.087%	3.399%	1.4407	1.936%	0.786%

In Table 5.8, we examine whether there are dominance orderings between the optimal portfolios presented in Figures 5.1, 5.2 and 5.3 and the S&P 500 benchmark.

Table 5.8: Dominance relations between optimal portfolios of all strategies depicted in Figures 1, 2 and 3 and the S&P 500 index.

	SR (OLS)	SR (RW)	S&P 500	TOK (5)	TOK (7)	TOK (10)	Joint_TOK (3)	Joint_TOK (5)	Joint_TOK (10)
SR (OLS)	$\boldsymbol{n.c}$	≺SSD	≺SSD	$n.c$	$n.c$	≺ICX	$n.c$	$n.c$	$n.c$
SR (RW)	≻SSD	$\boldsymbol{n.c}$	≻SSD	≺ICX	≺ICX	≺ICX	≺ICX	≺ICX	≺ICX
S&P 500	≻SSD	≺SSD	$\boldsymbol{n.c}$	≺ICX	≺ICX	≺ICX	≺ICX	≺ICX	≺ICX
TOK (5)	$n.c$	≻ICX	≻ICX	$\boldsymbol{n.c}$	$n.c$	≺ICX	≻ICX	≻ICX	≻ICX
TOK (7)	$n.c$	≻ICX	≻ICX	$n.c$	$\boldsymbol{n.c}$	≺ICX	≻ICX	≻ICX	≻ICX
TOK (10)	≻ICX	≻ICX	≻ICX	≻ICX	≻ICX	$\boldsymbol{n.c}$	≻ICX	≻ICX	≻ICX
Joint_TOK (3)	$n.c$	≻ICX	≻ICX	≺ICX	≺ICX	≺ICX	$\boldsymbol{n.c}$	$n.c$	$n.c$
Joint_TOK (5)	$n.c$	≻ICX	≻ICX	≺ICX	≺ICX	≺ICX	$n.c$	$\boldsymbol{n.c}$	$n.c$
Joint_TOK (10)	$n.c$	≻ICX	≻ICX	≺ICX	≺ICX	≺ICX	$n.c$	$n.c$	$\boldsymbol{n.c}$

From these two tables, we deduce the following points.

- The Joint_TOK (3) strategy presents the greatest average, final wealth, Sharpe ratio and STARR performance, but also the highest risk in terms of the standard deviation and VaR 5%.
- The strategies based on Joint_TOK performance are performing much better than S&P 500 benchmark, which shows the worst results in terms of mean, final wealth, Sharpe ratio, and STARR performance. However, in compensation, it achieves the lowest risk (standard deviation, VaR 5%, CVaR 5%) among all strategies.
- The ex-post returns are strongly leptokurtic for all strategies presented in Table 5.7. Moreover, all strategies (expect S&P 500 benchmark) show some signs of skewness.
- All strategies based on TOK and Joint_TOK performance measures dominate the S&P 500 benchmark and SR (RW) in the ICX sense. While TOK strategies dominate the Joint_TOK ones in terms of the ICX sense.

Comparing Joint_TOK results with the ones obtained with the TOK performance measure, we generally observe that the Joint_TOK ratio gives the lowest reward (mean, final wealth, Sharpe ratio, and STRR performance), but in compensation, it presents a lower risk. This result is not surprising since it accounts for the joint losses among the assets. Therefore, considering the joint losses among the assets has important implications for portfolio choices.

Interestingly, from these empirical analyses, we deduce that both performance measures (i.e. TOK and Joint_TOK) give better results than the classical Sharpe ratio and the S&P 500 benchmark. Overall, these results confirm and provide strong support for the use of the proposed new performance measures.

5.4 Optimal portfolio choices for different investors preferences

Portfolio selection problems can be characterized and classified based on the motivations and intentions of investors (see Ortobelli et al. (2013)). Thus, it is important to classify the optimal choices for any admissible ordering of preferences. Commonly, the consistency of a probability function within a given preferences ordering is used to describe and characterize the investor's optimal choice coherently with his/her preferences. For this reason, several classifications of the reward{risk probability function consistent with different orderings have been proposed in the financial literature (see, among others, Szegö (2004), Stoyanov et al. (2007) and Ortobelli et al. (2015)). In this context, we use a completely different approach to the portfolio selection problem and we show the impact of this new method on the possible choices for the US stock market during the period 2003-2015.

Let us start with this simple consideration: the investor wants his/her wealth at a given time to dominate his/her wealth at a previous time with respect to his/her preferences. In other words, the investor wishes to determine today the wealth that will be dominant (with respect to his/her attitude and preferences) in the near future. Therefore, we suggest a different return definition, one that is characterized by a well-known ordering classification used in portfolio theory, namely concave ordering, increasing concave ordering, and increasing convex ordering (see, among others, Strassen (1965), Shaked and Shanthikumar (1994), and Müller and Stoyan (2002)). First, we identify different definitions of returns with respect to the behaviour of non-satiable investors who could be risk-averse or risk-seeking. We then discuss some portfolio strategies that use the new returns definitions. To this end, we must recall the following classic results of ordering theory.

Lemma 5.1 *Given two random variables X and Y, the following equivalence relationships hold.*
1) (Martingale property): Every risk-averse investor (i.e. with concave utility function) prefers X to Y if and only if there exist two random variables X', Y' defined on the same probability space $(\Omega, \Im, P)$ *that have the same distribution of X and Y such that:*

$$E(Y'|X') = X' \ a.s..$$

2) (Super-martingale property): Every non-satiable risk-averse investor (i.e. with increasing concave utility function) prefers X to Y if and only if there exist two random variables X', Y' *defined on the same probability space* $(\Omega, \Im, P)$ *that have the same distribution of X and Y such that:*

$$E(Y'|X') \leq X' a.s..$$

3) (Sub-martingale property): Every non-satiable risk-seeking investor (i.e. with increasing convex utility function) prefers X to Y if and only if there exist two random variables X', Y' defined on the same probability space $(\Omega, \Im, P)$ *that have the same distribution of X and Y such that:*

$$E(Y'|X') \geq X' a.s..$$

The proof of this lemma based on the analysis proposed by Strassen 1965 and is a well-known result of ordering theory (see also Shaked and Shanthikumar, 1994, Müller and Stoyan 2002, and the reference therein).

Following Ortobelli et al. (2015), let $W = \{W_t\}_{t\in\mathbb{N}}$ be a discrete time wealth process and assume that the investor's temporal horizon is T. the main goal of any portfolio investor with a temporal horizon T is to determine the portfolio that, starting with an initial wealth W_{s-T} at time $s - T$, optimizes the future random wealth W_s. In particular, any non-satiable investors prefers W_s to W_{s-T} if $W_s \geq W_{s-T}$ and for this reason all the non-satiable investors optimizes the gross returns W_s/W_{s-T} that mean the increments of wealth during the period $[s - T, s]$. From Lemma 1, we are able to identify what is preferable for particular classes of investors: any non-satiable risk-averse (risk-seeking) investor prefers W_s to W_{s-T} if $W_s \geq E(W_{s-T}|W_s)$ ($E(W_s|W_{s-T}) \geq W_{s-T}$). For this reason, it makes sense to define the wealth increments from the point of view of a given class of investors, as $W_s/E(W_{s-T}|W_s)$ and $E(W_s|W_{s-T})/W_{s-T}$ (for non-satiable risk-averse investors and non-satiable risk-seeking investors, respectively). According to Rachev et al. (2008), we normally use the average return (for one unit of time of the process W)

$$r_s = \left(\frac{W_s}{W_{s-T}}\right)^{1/T} - 1,$$

in order to consider the aggregate risk of the period $[s - T, s]$. Clearly, by virtue of Lemma 1, we can distinguish different possible returns definitions with respect to the investors' preferences, the investors' temporal horizon and the portfolio wealth strategy.

Definition 5.1 *Let us assume that the wealth process of a given portfolio strategy follows a discrete stochastic process* $W = \{W_t\}_{t\in\mathbb{N}}$ *defined on a filtered probability space. Assume the investor has a temporal horizon equal to T (unities of time of the wealth process). Then, for any* $s > T$*,* *1) we define the risk-averse returns of the wealth process W for investors with a temporal horizon T as:*

$$r_S^{RA} = \left(\frac{W_S}{E(W_{S-T}|W_S)}\right)^{\frac{1}{T}} - 1; \tag{5.24}$$

2) we define the risk-seeking returns of the wealth process W for investors with temporal horizon T as:

$$r_S^{RS} = \left(\frac{E(W_S|W_{S-T})}{W_{S-T}}\right)^{\frac{1}{T}} - 1; \tag{5.25}$$

Using the new definitions of returns, the increment of the future wealth is defined from the point of view of different investors and thus we confirm that these return definitions are consistent (coherent) with certain investors' preferences. Furthermore, any risk measure that is consistent with the choice of non-satiable investors (see among others Rachev et al., 2008) according to the new definitions of risk-averse and risk-seeking investors' returns is a measure of risk for the respective investors' preferences. Therefore, these new definitions allows us to identify and classify the optimal choices of the investors with respect to their attitude towards risk in a better way. In particular, when we use the classical reward-risk portfolio selection models, applying risk-averse (risk-seeking) returns, we identify the choices of non-satiable risk-averse (risk-seeking) investors.

The portfolio selection problem is typically examined in a reward-risk framework, according to which the portfolio choice is made with respect to two criteria – the expected portfolio return and the portfolio risk. In particular, one portfolio is preferred to another if it has a higher expected return and a lower risk. Markowitz (1952) introduced the first rigorous approximating model for the portfolio selection problem, in which the return and risk are modelled in terms of portfolio mean and variance. However, a different generalization has been proposed in the literature (see, among others, Biglova et al. (2004), Rachev et al. (2008) and the references therein). In the following, we suppose that a portfolio contains n assets that we have a frictionless market in which no short selling is allowed and that all investors act as price takers. Let us briefly formalize another portfolio performance measure (Rachev ratio) that is used in the next empirical analysis.

Rachev ratio. The Rachev ratio (see Biglova *et al.* (2004)) is the ratio between the average of earnings and the mean of losses. i.e.:

$$RR(x'z,\alpha,\beta) = \frac{CVaR_\beta(z_b - x'z)}{CVaR_\alpha(x'z - z_b)} \tag{5.26}$$

where the Conditional Value-at-Risk (CVaR) is a coherent risk measure (see Rockafellar and Uryasev (2002) and Artzner et al. (1999)) that is defined as:

$$CVaR_{\alpha}\left(X\right)=\frac{1}{\alpha}\int_{0}^{\alpha}VaR_{q}\left(X\right)dq,$$

and

$$VaR_{q}\left(X\right)=-F_{X}^{-1}\left(q\right)=-inf\left\{x\setminus P\left(X\leq x\right)>q\right\},$$

is the Value-at-Risk (VaR) of the random return X. If we assume a continuous distribution for the probability law of X, then $CVaR_{\alpha}\left(X\right)=-E\left(X\mid X\leq -VaR_{\alpha}\left(X\right)\right)$, and therefore CVaR can be interpreted as the average loss beyond VaR. Again, many performance measures have been proposed and studied in the literature; for an overview see, among others, Farinelli et al. (2008) and the references therein. To conclude this section, one important aspect in portfolio optimization is the computational complexity. Some recent studies (see Rachev et al. (2008) and Stoyanov et al. (2007)) classify the computational complexity of reward-risk portfolio selection problems. In particular, Stoyanov et al. (2007) have shown that we can distinguish four different cases of reward and risk that admit a unique optimum in myopic strategies. Thus, in order to optimize these portfolio selection problems in an acceptable computational time, we use a heuristic algorithm for overall optimization, such as the one proposed by Angelelli and Ortobelli (2009) for overall portfolio optimization.

The earlier definitions of returns are based on the conditional expected value of the wealth process at a given time with respect to the wealth process at another time. To evaluate the conditional expected value we use a nonparametric estimator that allows us to estimate the random variable $E(Y|X)$. To contribute to the literature in this context, we propose a new multivariate kernel estimator that allows us to find, locally, the optimal bandwidths. We show that the new estimator is substantially more precise. To this end, we first present the bandwidth selection problem in the multivariate setting, and then we propose the new estimator in more detail.

5.4.1 Bandwidth selection problem

The bandwidth selection in the nonparametric technique is a very crucial step. Several methods have been proposed to select an optimal bandwidth (for a more complete treatment from a historical viewpoint, with complete references and a detailed discussion of the variations that have been suggested, see, among others, Jones et al. (1996) and Scott (2015)). Most of the proposed methods are mainly based on a simple but important rule of balancing bias and variance. This bias-variance trade-off works well for many densities, especially in the univariate or bivariate setting. However, as the dimensionality increases, the so-called curse of dimensionality becomes important. In particular, the increase in dimensionality has the effect of slowing the convergence rate of the MISE, and alters the trade-off between bias and variance. Therefore, as the dimensionality increases, larger and larger bandwidths are required to control the increased variability and especially the contributions from the tails. In this setting, it is well-known that the normal reference rules and plug-in approaches come close to the asymptotically optimal bandwidth. In particular, the optimal bandwidths $h=\left(h_1,h_2,....,h_d\right)'$ are proportional to the convergence rate $n^{-1/(d+4)}$, for a deeper discussion see Scott (2015).

In this chapter, we propose an alternative approach to bandwidth selection that uses a locally optimal bandwidth choice. To achieve this aim, we first optimize a criterion that allows us to find, locally, the optimal convergence rate. Then, using this optimal rate of convergence, we are able to determine, locally, the optimal bandwidth. In essence, using this procedure, we improve the rate of convergence and we drastically simplify the choice of the bandwidths.

5.4.2 New alternative multivariate kernel estimator

In this section, we propose a new kernel-type nonparametric regression estimator in the multivariate setting. The estimator we consider uses two well-known methodologies in a refined way; these are the multivariate locally weighted least squared regression and the classical multivariate version of Nadaraya-Watson estimator. In particular, this estimator fits locally without assuming any form of the function. Since choosing suitable bandwidths is a critical step in the nonparametric regression, much of our attention it will be devoted to propose a consistent estimator that allows us to find, locally, the optimal bandwidths. This method is direct and intuitively appealing. Let $(y_i, \boldsymbol{x}_i)$, for $i=1,2,....,n$ and $\boldsymbol{x}_i \in \mathbb{R}^d$ be a sample of observations drawn independently and identically from the distribution with density function $f_{X,Y}$. The solution for the optimal bandwidths and b coefficients is obtained by minimizing the following criterion:

$$\text{Minimize } \sum_{i=1}^{n} \{\mathrm{y_i} - \widehat{m}(\boldsymbol{x}, h) - b^T(\boldsymbol{x}_i - \boldsymbol{x})\}^2 K_{\boldsymbol{H}}(\boldsymbol{x}_i - \boldsymbol{x}), \quad (5.27)$$

where $\widehat{m}(\boldsymbol{x}, h)$ is the classical kernel estimation of $E(Y|X = \boldsymbol{x})$ and h represents a vector of positive optimal bandwidth, which controls the size of the local averaging and satisfies $h(n) \to 0$ when $n \to \infty$. The rest of the parameters are the same as locally linear weighted least squares given in (5.9). This method allows us to estimate the optimal local bandwidths consistently; therefore, we are able to use a multivariate local kernel estimator precisely. Indeed, this procedure relaxes one of the greatest drawbacks of the classical Nadaraya -Watson kernel estimator, which is the constant fit. Hence, instead of using local constant approximations or even locally linear functions, we fit locally without assuming any form of the function.

The proposed estimator is new contribution of this chapter and serves as a comparison to the classical estimators in the empirical analysis. In particular, it contains as a special case the classical locally linear weighted least squares. In fact, if we assume $m(\boldsymbol{x}, h) = \alpha$ we obtain (5.9). For more details about the classical locally weighted least squares see Ruppert and Wand (1994) and references therein. Higher order of polynomials, such as local cubic fit, can also be investigated similarly, but the number of parameters to be estimated increases rapidly, which demands a higher computational cost in practice. To conclude this section, we should note that the proposed new estimator is a consistent estimator since it is based on two consistent estimators.

5.4.3 Ex-post empirical analysis according to different investors' preferences

In this section, we first describe the data set and the methodology used to compare the different estimators and models. Then, we propose an ex-post comparison among portfolio models based on performance measures presented in (5.13) and (5.24). We use all active stocks on S&P 500 index from March 17, 2003 to February 24, 2015 using the previous 3000 daily observations. The data set is taken from Thomson-Reuters DataStream.

We examine the ex-post impact of different estimators, considering two portfolio problems: portfolio dimensionality reduction problems and portfolio performance problems. In particular, starting from September 12, 2003 we preselect the 100 stocks with the highest Rachev performance ratio (5.24). We preselect using Rachev ratio because the portfolios that maximize this reward-risk performance measure generally present higher earnings, a positive skewness, and lower losses (see Biglova et al. (2004)). Then, using the preselected stock, we reduce the dimensionality of the portfolio problem. Thus, we perform a PCA on the return of the selected stocks in order to identify the few factors that represent the highest return variability. We apply PCA on the Pearson correlation matrix of returns and then we regress the return series on these factors so that we are able to approximate the returns z (using either the factor model (5.3) or the nonparametric model (5.4)).

Since our aim in this section is to investigate the impact of regression analysis on large-scale portfolio selection problems, we first compare the classical OLS estimator, the Ruppert and Wand (hereinafter RW) estimator presented in (5.9) and the new proposed multivariate kernel estimator (hereinafter the KOT estimator). In particular, we compare the time evolution of the sum of the mean square error (Sum-MSE) for the different return approximations obtained by the nonparametric methods (namely the RW and KOT estimators) and the OLS estimator. On the preselected gross returns, we maximize both the Sharpe and the Rachev ratios such that the vector of weights x belongs to the simplex:

$$S = \left\{ x \in R^n \setminus \sum_{i=1}^{n} x_i = 1; x_i \geq 0; x_i \leq 0.1 \right\}. \tag{5.28}$$

This means that short sales are not allowed and that we invest no more than 10% in each asset. In this empirical analysis, we ignore the transaction costs, we assume a risk-free rate of zero, and α and β in the Rachev ratio are set to 5%.

We use a moving average window of 125 trading days for the computation of each optimal portfolio and we recalibrate the portfolio every month (20 trading days). First, we compare the Sum-MSE calculated from all methodologies at each k-th recalibration time. The results of this analysis are reported in Figure 5.4.

Figure 5.4: Sum of the MSE at each k-th recalibration by the OLS, RW and KOT estimators

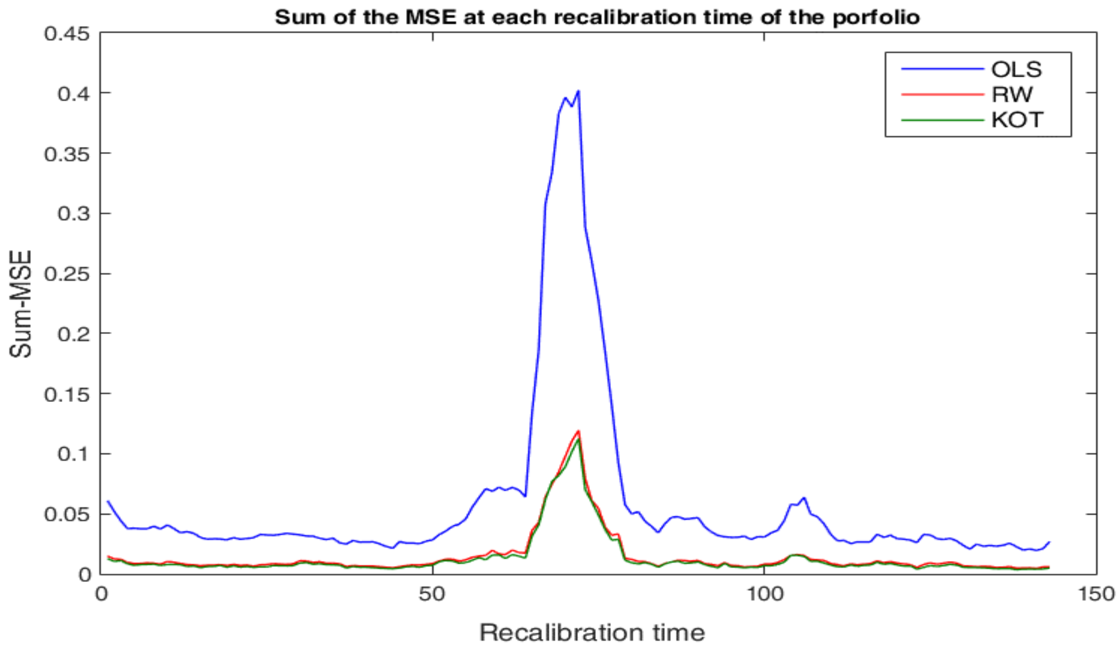


Figure 5.4 reports the sum-MSE obtained with three methods, namely the classical OLS and the nonparametric techniques based on RW and KOT estimators. On the one hand, we observe that both KOT and RW estimators are much better than the classical OLS. This means that the nonparametric estimators perform much better than the parametric estimators. On the other hand, the KOT estimator outperforms the classical OLS and the RW estimators in terms of the Sum-MSE. In the following table, we summarize average of some statistics of the Sum-MSE calculated from each of the methods.

Table 5.9: Average of some statistics of the Sum-MSE obtained by different estimators

	Mean(Sum-MSE)	St dev.(Sum-MSE)	Total(Sum-MSE)
KOT	0.0131	0.0184	1.8755
RW	0.0148	0.0194	2.1162
OLS	0.0583	0.0769	8.3358

Table 5.9 confirms the previous observations on the difference between the three estimators. For further confirmation, we compare the ex-ante errors we get considering OLS or KOT regression return approximation (see section 5.2.1). In particular, Table 5.10 reports the average (over the time) of the percentage of regression errors with higher variability (in the sense of concave dominance) of the criterion on the raw with respect to the criterion on the column.

Table 5.10: Average percentage of regression errors with higher variability considering two regression models (i.e. OLS and KOT) applied to few factors obtained by the PCA of four alternative correlation matrices (P1, P2, S1, S2)

	OLS Pearson	OLS Pearson conditional	OLS Stable	OLS Stable conditional	KOT Pearson	KOT Pearson conditional	KOT Stable	KOT Stable conditional
OLS Pearson	$n.c.$	0.0024	0.0129	0.0143	0.4493	0.4527	0.4978	0.4878
OLS Pearson conditional	0.002	$n.c.$	0.0149	0.0143	0.4473	0.4513	0.4967	0.4864
OLS stable	0.0415	0.0409	$n.c.$	0.0048	0.4464	0.4509	0.4951	0.4853
OLS Stable conditional	0.039	0.0374	0.0034	$n.c.$	0.4439	0.4504	0.4945	0.4847

From Table 5.10, the errors we get with nonparametric regression analysis (i.e. KOT) never present higher variability than those obtained with the classical parametric regression analysis (i.e. OLS). Moreover, we test that about the 50% of OLS error approximation present higher variability (in the sense of concave order) than KOT error approximation. According to this analysis, it makes sense to consider the new approximation method (i.e. KOT) in the portfolio selection problems. Thus, we propose the following ex-post empirical analysis.

We compare the optimal portfolio approaches solved for different investor's preferences. The proposed empirical analysis allows us to evaluate the ex-post impact of the optimal choices when we consider the new definitions of returns. For each strategy and for each returns definition, we compute the optimal portfolio composition every month (20 trading days). Therefore, at the k-th optimization, three steps are performed to compute the ex-post final wealth.

Step 1. Preselect the first 100 stocks with the highest Rachev ratio. Apply the PCA component to the correlation matrix of the preselected stocks. Then, approximate the portfolio returns as suggested in sections 5.2 and 5.4 (via KOT estimator).

Step 2. Determine the market portfolio $x_M^{(k)}$ that maximizes the performance ratio $\rho(x)$ applied to the approximated returns and constrained to the new definitions of returns:

$$max_x \; \rho(x)$$
$$\text{s.t. } \sum_{i=1}^{n} x_i^{(k)} = 1,$$
$$0 \leq x_i^{(k)} \leq 0.1 \quad i = 1, \dots, n \tag{5.29}$$

Here the performance measure $\rho(x)$ is either the Sharpe ratio (5.12) or the Rachev performance measure (5.26), constrained by different return definitions (5.24) and (5.25). Accordingly, we distinguish Sharpe risk-averse (SRA), Sharpe risk-seeking (SRS), Rachev risk-averse (RRA) and Rachev risk-seeking (RRS). These optimization problems may have several local optima. In order to overcome this limitation, we use as a starting point the optimal solution obtained with Angelelli and Ortobelli heuristic algorithm (see Angelelli & Ortobelli (2009)). We then improve this solution by applying the heuristic function pattern search implemented in Matlab 2015 to solve global optimization problems.

Step 3. Compute the ex-post final wealth as follows:

$$W_{t_{k+1}} = W_{t_k} \left(\left(x_M^{(k)} \right)' z_{(t_{k+1})}^{(expost)} \right), \tag{5.30}$$

where $z_{(t_{k+1})}^{(expost)}$ is the vector of observed gross returns in the period between t_k and t_{k+1}, such that $t_{k+1} = t_k + 20$.

We apply the three steps until the observations are available. The results of this analysis are reported in Figure 5.5.

Figure 5.5: Ex-post wealth obtained with SRA, SRS, RRA and RRS type strategies compared with S&P 500 benchmark

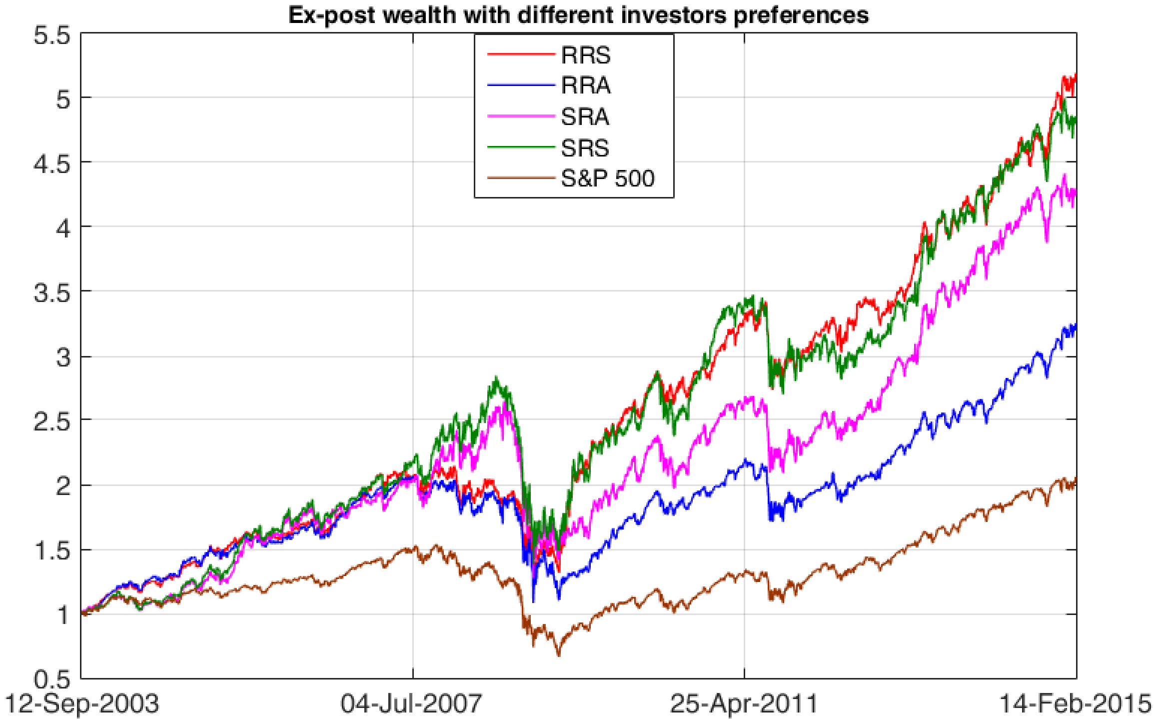


Figure 5.5 shows the comparison among the ex-post final wealth processes that match different investor's preferences. The strategies of non-satiable risk-seeking investors of RRS and SRS give a larger ex-post final wealth than the other strategies. Furthermore, all proposed strategies perform much better than the S&P 500 stock index. Table 5.11 reports the basic statistics (mean, standard deviation, skewness, kurtosis, VaR 5%, CVaR 5%, final wealth) of the ex-post returns of all strategies of Figure 5.5.

Table 5.11: Average of some statistics of the ex-post returns plus the final wealth obtained by different strategies

	Mean	St dev	Skewness	Kurtosis	VaR 5 %	CVaR 5%	Final W
SRA	0.060%	1.386%	-0.1472	8.5657	2.066%	3.333%	4.2926
SRS	0.065%	1.398%	-01337	9.2183	2.029%	3.385%	4.8566
RRA	0.048%	1.166%	-0.3338	15.467	1.628%	2.803%	3.2529
RRS	0.064%	1.143%	0.2358	16.236	1.505%	2.640%	5.1926
S&P 500	0.033%	1.232%	-0.0861	11.950	1.810%	3.032%	2.0586

From Table 5.11 and Figure 5.5 we deduce that:

- RRS presents the best Sharpe (mean/St. dev.) and STARR performance (mean/CVaR), and has the lowest risk (standard deviation, VaR 5%, CVaR 5%).

- SRS presents the highest mean but also the greatest risk (standard deviation, VaR 5%, CVaR 5%).
- All the performance strategies always perform better than the S&P500 stock index, in addition, they present a slight skewness and a pronounced kurtosis.

Overall, RRS gives the best result (RR is consistent with the preferences of a non-satiable investor who is neither risk seeking nor risk averse see Biglova et al. 2004). A more general empirical analysis with further studies and comparisons of the proposed model should be an object of future researches. In particular, we examine these new approaches using different performance measures that are account for particular investors' preferences.

To conclude this section, the experimental analysis suggests that the incorporation of new definitions of returns into an optimal portfolio framework leads to remarkable stabilization of the optimal portfolio strategy that is consistent with different investor's preferences. Moreover, solving the optimization problems constrained by the new returns definitions leads to better understanding of how the wealth evolves over time according to investors' preferences, namely non-satiable risk-averse and non-satiable risk-seeking.

5.5 Conclusion

Portfolio selection problems often involve unknown parameters that have to be properly approximated from the data. Therefore, in this chapter, we consider the implications of the conditional expectation estimators on portfolio theory. In particular, we focus on three financial applications, e.g. approximation problems within large-scale portfolio selection problems and optimal portfolio choices with consistent estimation of the expected returns. In the first application, we discuss and examine the impact of the correlation matrices and approximation methods in the portfolio theory. In this context, we suggest to approximate the returns using nonparametric regression analysis rather than the classical parametric approach. Using convex dominance testing, we find that the nonparametric regression outperforms its parametric counterpart. Moreover, we show that the dependence measure used to evaluate the joint behavior of returns (stable correlation matrix vs. Pearson correlation matrix) plays a crucial role in the dimensionality reduction of large-scale portfolio problems. For this reason, we propose to use the stable conditional correlation matrix to determine the few factors on which regress the return series and as a regression model the nonparametric ones.

In the second application, we suggest new performance measures that account for the heavy-tailed distribution of the returns. In this context, using stable conditional correlation matrix and the nonparametric techniques, we find that the new suggested methods typically yield the best performance, as measured by the Sharpe ratio and new performance measures. In the third application, we propose a new consistent multivariate kernel estimator of the conditional expectation and we show that the mean square error in the return approximation is generally lower than the mean square error for other estimators used in literature. Moreover, we deal with the portfolio selection problem from the point of view of different non-satiable investors: namely, risk-averse and risk-seeking. In particular, using a recent returns definition based on the conditional expectation we are able to compare the choices of different investor categories (according to their risk aversion attitude). Therefore, we propose an empirical comparison in which we optimize some classical performances on the returns (according to their new definitions). Thus, even this proposed empirical analysis allows us to evaluate the optimal choices for different categories of investors by using a conditional expectation estimator.

Chapter 6

Conclusion and Future Research

Conditional expectation is an important concept in probability and statistics which turn out to be extremely useful in financial modeling. It plays a crucial role in portfolio theory and in several pricing and risk management problems. We already stressed in the introductory chapters the importance of the topic: option market, technical analysis and portfolio theory. The aim of this work is to propose theoretical and methodological approaches to cover different portfolio managers' goals. This thesis contributes to such rich and challenging environment in at least four ways.

In chapter 2 we briefly introduce some of the most important concepts from the probability theory and financial mathematics that are useful in the financial applications of the conditional expectation: we hope that it could represent a useful map for researchers navigating through this vast and growing corpus of resources.

In chapter 3, we provide some theoretical motivations behind the use of the moving average rule as one of the most popular trading tools among practitioners. In particular, we examine the conditional probability of the price increments and we study how this probability changes over time. We find that under some assumptions the probability of up-trend is greater than the probability of down trend. For this reason, we propose to use moving average rules to predict periods of systemic risk. In this context, we suggest a methodology that incorporate moving average rules as alarm rules to predict potential fails of the market. Thus, we examine the impact of the moving average rules on the U.S. stock market. Firstly, a comparison among different moving average trading rules with and without alarms of losses is performed. Secondly, we compare the ex post wealth obtained with the best performing systemic risk rule used as trading strategy with the wealth obtained maximizing two different portfolio performances. From the comparison among different strategies and stochastic dominance tests, we deduce that the best use of the moving average rules is obtained to predict periods of market distress. These empirical analyses suggest that the moving average rules are much more effective and performing when used to detect the presence of systemic risk.

Further research could involve theoretical and empirical studies. On the one hand, investors may employ complex versions of the moving average rules. On the other hand, the impact of calendar periods such as the weekend effect, the turn-of-the-month effect, the holiday effect and the January effect. Future research will investigate this aspects. Another promising direction for future research is to consider other technical indicators, which may be easier to detect algorithmically, to examine whether or not such indicators are able to predict the presence of systemic risk.

In chapter 4, we present alternative approaches to evaluate the presence of the arbitrage opportunities in the option market. In particular, we empirically investigate the well-known put-call parity no-arbitrage relation and the SPD. First, we measure the violation of the put-call parity as the difference in implied volatilities between call and put options that have the same strike price, the same expiration date and the same underlying asset. Then, we discuss the usefulness of the nonnegativity of the SPD. We evaluate the effectiveness of the proposed approaches by an empirical analysis on S&P 500 index options data. Moreover, we propose alternative approaches to estimate the SPD under the classical hypothesis of the BS model. To this end, we first examine the real mean return function using local polynomial smoothing technique. Then, we estimate the conditional expectation under real probability density. Under the hypothesis of BS model, we are able to derive a closed formula for approximating the conditional expectation under risk neutral probability.

We use the classical nonparametric estimator based on kernel and a recent alternative the so called OLP estimator that uses a different approach to evaluate the conditional expectation consistently. This analysis allows us extrapolating arbitrage opportunities and relevant information from different markets (futures and options) consistently with the analysis of the underlying. Future research will focus on the extension of those concepts analyzing other possible development and uses of the conditional expectation estimators.

In chapter 5, we examine the use of the conditional expectation in portfolio theory. In particular, we propose three alternative financial applications based on the conditional expectation and a new conditional expectation estimator.

Firstly, we discuss and examine the impact of the correlation matrices and approximation methods in the portfolio theory. In this context, we suggest to approximate the returns using nonparametric regression analysis rather than the classical parametric approach. Using convex dominance testing, we find that the nonparametric regression outperforms its parametric counterpart. Moreover, we show that the dependence measure used to evaluate the joint behavior of returns (stable correlation matrix vs. Pearson correlation matrix) plays a crucial role in the dimensionality reduction of large-scale portfolio problems. For this reason, we propose to use the stable conditional correlation matrix to determine the few factors on which regress the return series and as a regression model the nonparametric ones.

Secondly, we suggest new performance measures that account for the heavy-tailed distribution of the returns. In this context, using stable conditional correlation matrix and the nonparametric techniques, we find that the new suggested methods typically yield the best performance, as measured by the Sharpe ratio and new performance measures.

Finally, we propose a new consistent multivariate kernel estimator of the conditional expectation and we show that the mean square error in the return approximation is generally lower than the mean square error for other estimators used in literature. Moreover, we deal with the portfolio selection problem from the point of view of different non-satiable investors: namely, risk-averse and risk-seeking. In particular, using a recent returns definition based on the conditional expectation we are able to compare the choices of different investor categories (according to their risk aversion attitude). Therefore, we propose an empirical comparison in which we optimize some classical performances on the returns (according to their new definitions). Thus, even this proposed empirical analysis allows us to evaluate the optimal choices for different categories of investors by using a conditional expectation estimator.

Overall, the thesis contributes the literature in several ways and achieves the general aim. In particular, it allows us to assess the impact and usefulness of the conditional expectation on different financial applications, e.g. arbitrage opportunities, large-scale portfolio selection problems, and optimal portfolio choices.

6. Literature

1. Ait-Sahalia, Y., Duarte, J. (2003). Nonparametric option pricing under shape restrictions. *Journal of Econometrics*, **116**(1-2), 9-47.
2. Ait-Sahalia, Y., Lo, A. W. (1998). Nonparametric estimation of state-price densities implicit in financial asset prices. *Journal of Finance* **53**(2), 499–547.
3. Alexander, S. (1961). Price movements in speculative markets: trends or random walks. *Industrial Management Review*, **2**(5), 7−26.
4. Angelelli, E., Ortobelli, S. (2009). American and European portfolio selection strategies: the Markovian approach. Chapter 5 in *Financial Hedging*, P. N. Catlere (ed.), pp. 119−152. New York: Nova Science Publishers.
5. Artzner, P., Delbaen, F., Eber, J.M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, **9**(5), 203-228.
6. Bachelier, L. (1900). Theorie de la speculation. *Annales Scientifiques de l'Ecole Normale Suprieure*, **3**(17), 21-86.
7. Backus, D., Foresi, S., Zin, S. (1998). Arbitrage opportunities in arbitrage-free models of bond pricing. *Journal of Business and Economic Statistics*, **16**, 13-24.
8. Bates, D. S. (1996). Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Review of Financial Studies*, **9**, 69–107.
9. Bawa, V. S. (1975). Optimal rules for ordering uncertain prospects. *Journal of Financial Economics*, **2**(1), 95-121.
10. Bawa, V. S. (1978). Safety-first, stochastic dominance, and optimal portfolio choice. *Journal of Financial and Quantitative Analysis*, **13**(02), 255-271.
11. Benko, M., Fengler, M., Härdle, W., Kopa, M (2007). On Extracting Information Implied in Options. *Computational statistics*, **22**, 543–553.
12. Bjork, T. (2004). Arbitrage Theory in Continuous Time. Oxford University Press.
13. Biglova, A., Ortobelli, S., Fabozzi, F. (2014). Portfolio selection in the presence of systemic risk. *The Journal of Asset Management*, **15**(5), 285−299.
14. Billingsley, P. (1995). Probability and Measure. Wiley, New York.
15. Bingham, N.H., Kiesel, R. (1998). Risk-Neutral Valuation. Springer-Verlag, Berlin Heidelberg, New York.
16. Black, F., Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, **81**, 637-659.

17. Bliss, R.R., Panigirtzoglou, N. (2002). Testing the stability of implied probability density functions. *Journal of Banking and Finance*, **26**, 381–422.
18. Breeden, D. T., Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of Business* **51**(4), 621–51.
19. Bowman, A.W., Azzalini, A. (1997). Applied Smoothing Techniques for Data Analysis. Oxford University Press, London.
20. Brock, W., Lakonishok, J., LeBaron, B**.** (1992). Simple technical rules and the stochastic properties of stock returns. *Journal of Finance*, **47**(5), 1731−1764.
21. Brunner, B., Hafner, R. (2003). Arbitrage-free estimation of the risk-neutral density from the implied volatility smile. *Journal of Computational Finance*, **7**(1), 75–106.
22. Chen, S. X., Xu, Z. (2014). On implied volatility for options: Some reasons to smile and more to correct. *Journal of Econometrics*, **179**(1), 1-15.
23. Cherubini, U., Luciano, E., Vecchiato, W. (2004). Copula Methods in Finance. John Wiley & Sons, Hoboken.
24. Chiarella, C., He, X., Hommes, C. (2006). A dynamic analysis of moving average rules. *Journal of Economic Dynamics and Control*, **30**(9-10), 1729−1753.
25. Chung, K.L. (2001). A course in probability theory. Academic Press, San Diego.
26. Coqueret, G., Milhau, V. (2014). Estimating covariance matrices for portfolio optimization. *ERI Scientific Beta White Paper*.
27. Cont, R., Da Fonseca, J. (2002). The dynamics of implied volatility surfaces. *Quantitative Finance*, **2**(1), 45–60.
28. Cox, J. E., Ross, S. A. (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics*, **76**, 145–166.
29. Daly, J., Crane, M., Ruskin, H. (2008). Random matrix theory filters in portfolio optimization: a stability and risk assessment. *Physica A: Statistical Mechanics and its Applications*, **387**(16-17), 4248-4260.
30. Davidson, R., Jean-Yves D. (2000). Statistical inference for stochastic dominance and for the measurement of poverty and inequality. *Econometrica*, **68**(6), 1435-1464.
31. De Giorgi, E. (2005). Reward-risk portfolio selection and stochastic dominance. *Journal of Banking & Finance*, **29**(4), 895-926.
32. De Giorgi, E., Post, T. (2008). Second-order stochastic dominance, reward-risk portfolio selection, and the capm. *Journal of Financial and Quantitative Analysis*, **43**(02), 525-546.
33. Delbaen, F., Schachermayer, W. (2006). The mathematics of arbitrage. Springer-Verlag Berlin Heidelberg, New York.

34. DeMiguel, V., Garlappi, L., Uppal, R. (2009). "Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?" *Review of Financial Studies*, **22**, 1915-1953.
35. Denuit, M., Dhaene, J., Goovaerts, M.J., Kaas, R. (2005). Actuarial Theory for Dependent Risks. John Wiley&Sons.
36. Dentcheva, D., Ruszczynski, A. (2006). Portfolio optimization with stochastic dominance constraints. *Journal of Banking & Finance*, **30**(2), 433-451.
37. Duffie, D. (2001). Dynamic Asset Pricing Theory, third edition. Princeton University Press, Princeton
38. Dupire, B. (1994). Pricing with a smile. *RISK*, **7**(1),18–20.
39. Epanechnikov, V. A. (1965). Nonparametric estimation of a multivariate probability density. *Theory of Probability and its Applications,* **14**(1), 153-158.
40. Fabozzi, F. J., Dashan, H., & Guofu, Z. (2010). Robust portfolios: Contributions from operations research and finance. *Annals of Operations Research*, **176**(1), 191–220.
41. Fama, E., (1965). The behavior of stock market prices. *Journal of Business* **38**, 34-105.
42. Fama, E.F. (1970). Efficient capital markets: a review of theory and empirical work. *Journal of Finance*, **25**(2), 383−417.
43. Fama, E.F., Blume, M.E. (1966). Filter rules and stock market trading. Security prices: a supplement. *Journal of Business*, **39**(1), 226−241.
44. Fan, J., Gijbels, I. (1996). Local polynomial modelling and its applications. Chapman and Hall, London.
45. Farinelli, S., Ferreira, M., Rossello, D. (2008). Beyond Sharpe ratio: Optimal asset allocation using different performance ratios. *Journal of Banking & Finance*, **32**(10), 2057-2063.
46. Fengler, M.R., Härdle, W. and Mammen, E. (2007). A semiparametric factor model for implied volatility surface dynamics. *Journal Financial Economics*, **5**, 189–218.
47. Föllmer, H., Schied, A. (2002). Stochastic Finance: An Introduction in Discrete Time. Wiley Series in Probability and Mathematical Statistics. Walter de Gruyter, Berlin, New York.
48. Fong, W. M. . Wong, W. K., Lean, H. H. (2005). International momentum strategies: a stochastic dominance approach. *Journal of Financial Markets*, **8**(1), 89-109.
49. Giacometti, R., Ortobelli, S., Tichy, T. (2015). Portfolio selection with uncertainty measures consistent with additive shifts. *Prague Economic Papers*, **24**(1), 3−16.
50. Gencay, R., Stengos, T. (1998). Moving average rules, volume and the predictability of security returns with feedforward networks. *Journal of Forecasting*, **17**(5-9), 401–414.

51. Gyorfi, L., Kohler, M. (2002). Nonparametric regression estimation. in *Principles of Nonparametric Learning*, Gyorfi, L (Ed.), Springer-Verlag, Wien, 59-118.
52. Hadar, J., Russell, W.R. (1969). Rules for Ordering Uncertain Prospects, *American Economic Review* **59**, 2-34.
53. Hafner, R. (2004). Stochastic Implied Volatility. Springer, Berlin.
54. Härdle, W. (1990). Applied Nonparametric Regression. Cambridge University Press, Cambridge.
55. Härdle, W., Müller, M. (2000). Multivariate and semiparametric kernel regression. In: Schimek, M.G. (Ed.), Smoothing and Regression: Approaches, Computation, and Application. John Wiley & Sons, New York, 357-392.
56. Härdle, W., Müller, M., Sperlich, S. and Werwatz, A. (2004). Nonparametric and Semiparametric Models. Springer-Verlag, Berlin, Heidelberg, New York.
57. Härdle, W., Hlavka, Z. (2009). Dynamics of state price densities. *Journal of Econometrics*, **150**(1), 1-15.
58. Harrison J., Kreps M. (1981). Martingales and arbitrage in multiperiod securities markets. *Journal Economic Theory* **20**,381–408.
59. Harrison, J., Pliska, S., (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications,* **11**, 215–260.
60. Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**, 327–343.
61. Hull, J. (2015). Options, Futures, and Other Derivatives. 9th edition. Prentice Hall, New Jersey.
62. Jackwerth, J. C. (2000). Recovering Risk Aversion from Option Prices and Realized Returns. *Review of Financial Studies*, **13**(2).
63. Jones, M. C., Marron, J. S., Sheather, J. S., (1996). A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91(433), 401-407.
64. Kahalé, N. (2004). An arbitrage-free interpolation of volatilities. *RISK*, **17**(5), 102–106.
65. Karatzas, I., Shreve, S.E. (1991). Brownian motion and stochastic calculus, Graduate Texts in Mathematics, 113, Springer-Verlag, New York.
66. Kan, R., Zhou, G. (2007). Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, **42**(3), 621-656.
67. Kim, J. S., Kim, Y. C., Shin, K. Y. (2005). An algorithm for portfolio optimization problem. *Informatica*, 16(1) 93-106.

68. Kondor, I., Pafka, S., Nagy, G. (2007). Noise sensitivity of portfolio selection under various risk measures. *Journal of Banking and Finance*, **31**, 1545-1573.
69. Konno, H. (2011). Mean absolute deviation model. *Stochastic Programming*, pages 239-255.
70. Konno, H., Yamazaki, H. (1991). Mean absolute deviation portfolio optimization model and its applications to Tokyo stock market. *Management Science*, **37**(5), 519-531.
71. Kopa, M., Post, T. (2009). A portfolio optimality test based on the first-order stochastic dominance criterion. *Journal of Financial and Quantitative Analysis*, **44**(05), 1103-1124.
72. Kopa, M., Tichy, T. (2014). No arbitrage condition of implied volatility and bandwidth selection. *Anthropologist* **17** (3), 751-755.
73. Kouaissah, N., Orlandini, D. Ortobelli, S., Tichý, T (2016). Theoretical and practical motivations for the use of the Moving average rule. *OR Spectrum* (Submitted).
74. Kouaissah, N., Ortobelli, S. (2015). Alternative Methods to Evaluate the Arbitrage Opportunities. *WSEAS Transactions on Business and Economics*, E-ISSN: 2224-2899, **12**, 416-425.
75. Kouaissah, N., Ortobelli, S. On the estimation of the state price density. In Proceedings of the *33rd International Conference on Mathematical Methods in Economics*. 09-11 September 2015, Plzeň: UWB, Czech Republic. ISBN: 978-80-261-0539-8, pp. 389-394.
76. Latané, H. A., Rendelman, J. (1976). Standard deviations of stock price ratios implied in option prices. *Journal of Finance*, **31**, 369–381.
77. Ledoit, O., Wolf. M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance*, **10**, 603- 621.
78. Levich, R. M., Thomas, L. R. (1993). The significance of technical trading-rule profits in the foreign exchange market: a bootstrap approach. *Journal of International Money and Finance*, **12**(5), 451−474.
79. Levy, H. (1992). Stochastic dominance and expected utility: survey and analysis. *Management Science*, **38**(4), 555-593.
80. Levy, H., Markowitz, H., M. (1979). Approximating expected utility by a function of mean and variance. *American Economic Review* **69**, 308-317.
81. Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics, **47**, 13-37.
82. Lo, A., Mamaysky, H., Wang, J. (2000). Foundations of technical analysis: computational algorithms, statistical inference, and empirical implementation. *Journal of Finance*, **55**(4), 1705–1770.

83. Mandelbrot, B. (1963). The variation of certain speculative prices. The Journal of Business **36**, 394-419.
84. Markowitz, H. (1952). Portfolio Selection. The Journal of Finance **7**(1), 77-91.
85. Martin, D., Rachev, S., Siboulet, F. (2003). Phi-alpha optimal portfolios and Extreme Risk Management. *Wilmott Magazine of Finance*, November, 70−83.
86. Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science*, **4** (Spring), 141–183
87. Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, **3**, 125–144.
88. Moskowitz, T. J, Ooi, Y. H., Pedersen, L. H. (2012). Time series momentum. *Journal of Financial Economics*, **104**(2), 228–250.
89. Müller, A., Stoyan, D. (2002). Comparison methods for stochastic models and risks. New York: John Wiley & Sons.
90. Musiela, M., Rutkowski, M. (1997). Martingale Methods for Financial Modelling. Springer-Verlag, Berlin, Heidelberg, New York.
91. Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability and its Applications*, **9**(1), 141-142.
92. Neely, C. J., Rapach, D. E., Tu, J., Zhou, G. (2014). Forecasting the equity risk premium: the role of technical indicators. *Management Science*, **60**(7), 1772−1791.
93. Nolan, J.P., Ojeda-Revah, D. (2013). Linear and nonlinear regression with stable errors. *Journal of Econometrics*, **172**(2), 186–194.
94. Ogryczak, W., Ruszczynski, A. (1999). From stochastic dominance to mean-risk models: Semideviations as risk measures. *European Journal of Operational Research*, **116**(1), 33-50.
95. Orlandini, D. (2008). *L'utilizzo delle trading rules nell'ottimazzazione di portafolio*. PhD Thesis. Cap. 3, Bergamo University.
96. Ortobelli, S., Kouaissah, N., Tichý, T (2016). On the impact of the conditional expectation estimators in the portfolio theory. *Computational management science* (submitted).
97. Ortobelli, S., Kouaissah, N., Tichý, T (2016). On the use of the conditional expectation in portfolio selection problems. *Annals of operation research* (submitted).
98. Ortobelli, S. (2001). The classification of parametric choices under uncertainty: analysis of the portfolio choice problem. *Theory and Decision*, **51**(2-4), 297-328.

99. Ortobelli, S., Angelelli, E., Ndoci, A. (2016), Timing portfolio with Markov processes: the example of exponential Lévy processes. To appear in *European Journal of Operational Research*.
100. Ortobelli S., Leccadito A., Russo E., Iaquinta G. (2006). Financial Risk Modeling with Markov Chains, Springer-Verlag Lecture Notes in Computer Science **4284**, 1275-1282.
101. Ortobelli S. Leccadito A., Russo E.(2007). Portfolio Selection, VaR and CVaR models with Markov Chain. *International Journal of Computer Science & Network Security* **7**(6), 115-123.
102. Ortobelli, S., Kouaissah, N. (2015). Alternative methods to estimate the State Price Density. In MSCI 2015, Sliema - Malta. DOI 10,1109/MSCI-2015.19 on Conference Publishing Services (CPS) of IEEE Computer Society, 244-246.
103. Ortobelli, S., Lando, T. (2015). Independence tests based on the conditional expectation. *WSEAS Transactions on Mathematics*, **14**, 335-344.
104. Ortobelli, S., Rachev, S., Shalit, H., Fabozzi, F. (2009). Orderings and probability functionals consistent with preferences. *Applied Mathematical Finance*, **16**(1), 81-102.
105. Ortobelli, S., Petronio, F., Lando, T. (2015). A portfolio return definition coherent with the investors preferences. *IMA Journal of Management Mathematics*, 1-13.
106. Ortobelli, S., Shalit, H., Fabozzi, F. (2013). Portfolio Selection Problems Consistent with Given Preference Orderings. *International Journal of Theoretical and Applied Finance*, **16**(4), 1-34.
107. Ortobelli, S., Tichý, T. (2015). On the impact of semidefinite positive correlation measures in portfolio theory. *Annals of Operations Research*, **235**(1), 625-652.
108. Papp, G., Pafka, S., Nowak, M.A., Kondor, I. (2005). Random matrix filtering in portfolio optimization. *ACTA Physica Polonica B*, **36**, 2757-2765.
109. Pflug, G., Pichler, A., Wozabal, D. (2012), "The 1/N investment strategy is optimal under high model ambiguity". *Journal of Banking and Finance*, **36**(2), 410-417.
110. Post, T. (2003). Empirical tests for stochastic dominance efficiency. *Journal of Finance*, **58**, 1905-1931.
111. Post, T., Kopa, M. (2016). Portfolio Choice based on Third-degree Stochastic Dominance. *Forthcoming in Management Science*.
112. Post, T., Kopa, M. (2013). General linear formulations of stochastic dominance criteria. *European Journal of Operational Research*, **230**(2):321-332.

113.Post, T., Levy, H. (2005). Does risk seeking drive stock prices? a stochastic dominance analysis of aggregate investor preferences and beliefs. *Review of Financial Studies*, 18 (3), 925-953.
114.Quirk, J. P., Saposnik, R (1962). Admissibility and measurable utility functions. *Review of Economic Studies* **29**, 140-146.
115.Rachev, S.T., Menn, C., Fabozzi, F.J. (2005). Fat-tailed and Skewed Asset Return Distributions: Implications for Risk Management, Portfolio Selection, and Option Pricing. John Wiley & Sons, New York.
116.Rachev, S., Mittnik, S. (2000) .Stable paretian models in finance. Wiley, Chichester.
117.Rachev, S., Ortobelli, S., Stoyanov, S., Fabozzi, F. Biglova, A. (2008). Desirable properties of an ideal risk measure in portfolio theory. *International Journal of Theoretical and Applied Finance*, **11**(1), 19-54.
118.Rockafellar, R. T., Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance,* **26**(7), 1443–1471.
119.Rockafellar, R. T., Uryasev, S., Zabarankin, M. (2006). Generalized Deviations in Risk Analysis. *Finance and Stochastics*, **10**, 51-74.
120.Ross, S. (1978). Mutual fund separation in financial theory-the separating distributions. *Journal of Economic Theory*, **17**, 254-286.
121.Ruppert, D., Wand, M.P. (1994). Multivariate locally weighted last squares regression. *The annals of Statistics*, **22**(3), 1346-1370.
122.Samorodnitsky, G., Taqqu, M. S. (1994). Stable non-Gaussian random processes: stochastic models with infinite variance, volume 1. CRC Press.
123.Samuelson, P. A. (Spring 1965). Rational Theory of Warrant Pricing. *Industrial management Review*, **6**, 13-31.
124.Scarsini, M. (1984). On measures of concordance. *Stochastica*, **8**, 201-218.
125.Schoutens, W. (2003). Levy Processes in Finance. John Wiley & Sons, Ltd.
126.Scott, D.W. (2015). Multivariate Density Estimation: Theory, Practice, and Visualization. John Wiley & Sons, New York.
127.Shaked, M., Shanthikumar, G. (1994). Stochastic orders and their applications. New York: Academic Press Inc. Harcourt Brace & Company, New York.
128.Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. Journal of Finance, **19**(3), 425-442.
129.Sharpe, W. F. (1994). The Sharpe ratio. *The Journal of Portfolio Management*, **21**(1), 49-58.

130.Shimko, D. (1993). Bounds on probability. *RISK*, **6**(4), 33–374.

131.Shiryaev, A. (1996). Probability. Springer, New York.

132.Shreve, S. E. (2004). Stochastic Calculus for Finance II, Continuous-Time Models. Springer-Verlag, Berlin, Heidelberg, New York.

133.Simaan, Y. (1997). Estimation risk in portfolio selection: the mean variance model versus the mean absolute deviation model. *Management Science*, **43**(10), 1437-1446.

134. Simonoff, J. S. (1996). Smoothing Methods in Statistics, Springer-Verlag, New York.

135. Stanton, R. (1997). A nonparametric model of term structure dynamics and the market price of interest rate risk. *The Journal of Finance*, **52**(5), 1973-2002.

136. Stoyanov, S., Rachev, S., Fabozzi, F. (2007). Optimal financial portfolios. *Applied Mathematical Finance*, **14**(5), 401-436.

137. Strassen, V. (1965). The existence of probability measures with given marginals. *Annals of Mathematical Statist*ics, **36**(2), 423-439.

138. Szegö, G. (2004). Risk measures for the 21st century. Chichester: John Wiley & Son.

139. Taylor, S. J. (2011). Asset price dynamics, volatility, and prediction. Princeton university press. Princeton.

140. Tichý, T. (2011). Lévy Processes in Finance: Selected applications with theoretical background. *SAEI*, vol. 9. Ostrava: VŠB-TU Ostrava.

141. Tichý, T. (2012). Some results on pricing of selected exotic options via subordinated Lévy models. In *Managing and Modelling of Financial Risks*, Ostrava, 610-617.

142. Tichý, T., Kouaissah, N., Ortobelli, S., Orlandini, D. (2015). On the use of the moving average rule. *12th International Conference on Service Systems and Service Management (ICSSSM),* 5p.

143.Tobin, J. (1958). Liquidity preference as behavior towards risk. *Review of Economic studies*, **25**, 65-86.

144. Treynor, J. (1961). Towards a theory of market value of risky assets. Unpublished manuscript.

145. Tsybakov, A.B. (2009). Introduction to Nonparametric Estimation. Springer, New York

146. Watson, G. S. (1964). Smooth regression analysis. Sankhya, Series A, 26(4), 359-372.

147. Yatchew, A., Härdle W. (2006). Nonparametric state price density estimation using constrained least squares and the bootstrap. *Journal of Econometrics*, **133**(2), 579-599.

148. Yitzhaki (1982). Stochastic dominance, mean variance, and Gini's mean difference. *The American Economic Review*, 178-185.

149. Young, M. R. (1998). A minimax portfolio selection rule with linear programming solution. *Management Science*, **44**(5), 673-683.
150. Zenios, S. A., Kang, P. (1993). Mean-absolute deviation portfolio optimization for mortgage backed securities. *Annals of Operations Research*, **45**(1), 433-450.
151. Whitmore, G., A. (1970). Third-degree Stochastic Dominance. *American Economic Review*, **60**, 457-59.

Appendix A

Gaussian and Alpha Stable Distributions

In this Appendix, we review two main principal distributions that are used throughout this dissertation.

Gaussian distribution

The class of normal distributions, or Gaussian distributions, is certainly one of the most important probability distributions in statistics and, due to some of its appealing properties, also the class that is used in most applications in finance (e.g. BS model, mean-variance framework, CAPM etc.). Here we introduce some of its basic properties.

The random variable X is said to be normally distributed with parameters μ and σ, simply abbreviated by $X \sim N(\mu, \sigma)$, if density function of the random variable is given by the following formula:

$$f(X) = \frac{1}{\sqrt{2\pi\sigma^2}} exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}. \quad \text{(A.1)}$$

The parameter μ is called a location parameter because the middle of the distribution equals μ and σ is called a shape parameter or a scale parameter. If $\mu = 0$ and $\sigma = 1$, then X is said to have a standard normal distribution. For small values of σ, the density function becomes more narrow and peaked whereas for larger values of σ the shape of the density widens. These observations lead to the name shape parameter or scale parameter for σ. An important property of the normal distribution is the location-scale invariance of the normal distribution. Another interesting and important property of normal distributions is their summation stability.

The last important property that is often misinterpreted to justify the nearly exclusive use of normal distributions in financial modeling is the fact that the normal distribution possesses a domain of attraction. A mathematical result called the central limit theorem states that under certain technical conditions the distribution of a large sum of random variables behaves necessarily like a normal distribution. In the eyes of many, the normal distribution is the unique class of probability distributions having this property. This is wrong and actually it is the class of stable distributions (containing the normal distributions) that is unique in the sense that a large sum of random variables can only converge to a stable distribution. In the sequel, we discuss the stable distribution.

Alpha Stable Distribution

The class of the stable distributions is defined by means of their characteristic functions. With very few exception, no closed-form expressions are known for their densities and distribution functions. Research on stable distributions in the field of finance has a long history (for a more complete treatment, from a historical viewpoint, with complete references, and detailed discussion see among others Samorodnitsky and Taqqu, 1994 and Rachev and Mittnik, 2000). In 1963, the mathematician Benoit Mandelbrot first used the stable distribution to model empirical distributions that have skewness and fat tails.

To distinguish between Gaussian and non-Gaussian stable distributions, the latter are generally referred to as stable Paretian, Levy stable, or α-stable distributions. Stable Paretian tails decay more slowly than the tails of the normal distribution and therefore better describe the extreme events present in the data (Rachev et al., 2011). Like the Student's t-distribution, stable Paretian distributions have a parameter responsible for the tail behavior, called tail index or index of stability.

It is possible to define the stable Paretian distribution in two ways. The first one establishes the stable distribution as having a domain of attraction. That is, (properly normalized) sums of IID random variables are distributed with the α-stable distribution as the number of summands n goes to infinity. Formally, let $Y_1, Y_2, \ldots, Y_n$ be an IID random variable and $a_n$and $b_n$be sequences of real and positive numbers, respectively. A variable X is said to have the stable Paretian distribution if

$$\frac{\sum_{i=1}^{n} Y_i - a_n}{b_n} \overset{d}{\rightarrow} X \qquad \text{(A.2)}$$

The density function of the stable Paretian distribution is not available in a closed form expression in the general case. Therefore, the distribution of a stable random variable X is alternatively defined through its characteristic function. The density function can be obtained through a numerical method. The characteristic function of the α-stable distribution is given by

$$\varphi_X(t\backslash\alpha,\sigma,\beta,\mu) = E\left[e^{itX}\right] = \begin{cases} \exp\left(i\mu t - |\sigma t|^{\alpha}\left(1 - \beta(sign\ t)tan(\pi\alpha/2)\right)\right), & \alpha \neq 1 \\ \exp\left(i\mu t - \sigma|t|(1 - \beta(2/\pi)(sign\ t)ln|t|)\right), & \alpha = 1 \end{cases}$$

(A.3)

where, $$sign\ t = \begin{cases} 1 & t > 0 \\ 0 & t = 0 \\ -1 & t < 0 \end{cases} \tag{A.4}$$

The four parameters appearing in equation (A.3) are the following:

- $\alpha \in (0,2)$ is the index of stability or the tail exponent
- $\beta \in [-1, +1]$ is a skewness parameter.
- $\sigma \in (0, +\infty)$ is a scale parameter.
- $\mu \in (-\infty, +\infty)$ is a location parameter.

The parameter α determines how heavy the tails of the distribution are. That is why the α-stable distribution is highly flexible and suitable for modeling non-symmetric, highly kurtosis, and heavy-tailed data. Since stable are uniquely determined by the four parameters, the common notation is $S_\alpha(\sigma, \beta, \mu)$. The three special cases where there is a closed-form solution for the densities are the Gaussian case ($\alpha = 2$), the Cauchy case ($\alpha = 1, \beta = 0$) and the Lévy case $\left(\alpha = \frac{1}{2}, \beta \pm 1\right)$.

Apart from the appealing feature that the probabilistic properties of only the stable distributions are close to probabilistic properties of sums of i.i.d. random variables, there is another important characteristic which is the stability property. According to stability property, appropriately centered and normalized sums of i.i.d. α-stable random variables is again α-stable. This property is unique to the class of stable law.

Appendix B

Stochastic Dominance and increasing convex order

The stochastic dominance is one of the fundamental concepts of the decision theory; see among others Levy (1992). It introduces a partial order in the space of real random variables. The first-order dominance (Quirk and Saposnik 1962) assumes non-satiation (nondecreasing utility functions), the second-order dominance (Hadar and Russel 1969) assumes also risk aversion (concave nondecreasing utility functions), and the third-order dominance (Whitmore 1970) adopts as well decreasing risk aversion (concave nondecreasing utility functions that have a non-negative third derivative). In portfolio theory, stochastic dominance rules have been widely used to justify the reward risk approaches (see Ortobelli et al., 2009 and the reference therein). In this context, several theoretical formulations and empirical application have been proposed in the last decades (Post and Kopa, 2013; Ortobelli et al., 2009; Rachev et al., 2008; Dentcheva and Ruszczynski, 2006; De Giorgi, 2005; Fong et al., 2005; Post and Levy, 2005; Post, 2003).

We consider the space $L^1(\Omega, \Im, P)$ of integrable random variables defined in a probability space $(\Omega, \Im, P)$. The right-continuous cumulative distribution function $F_X(\eta)$ of X is defined as follows:

$$F_X(\eta) = P(X \leq \eta) \quad for \;\; \eta \in \mathbb{R} \tag{B.1}$$

Definition B.1. A random return X is said to be stochastically dominate another random return Y in the first order stochastic dominance sense, denoted $X \succeq_{FSD} Y$, if

$$F_X(\eta) \leq F_Y(\eta) \quad for \;\; \eta \in \mathbb{R} \tag{B.2}$$

The usual first-order definition of stochastic dominance (FSD) gives a partial order in the space of real random variables (Kopa and Post, 2009; Levy, 1992; Bawa, 1978). More important from the portfolio point of view is the notion of the second-order dominance (SSD), which is also defined as a partial order. It is one of the most debated topics in the financial portfolio selection, due to its connection to the theory of risk-averse investor behavior and tail minimization (De Giorgi and Post, 2008; Ortobelli, 2001; Bawa, 1975). The integrated distribution function $F_2(Z;\eta)$ is defined as follows:

$$F_X^{(2)}(\eta) = \int_{-\infty}^{\eta} F_X(t)dt \quad for\ \eta \in \mathbb{R}. \tag{B.3}$$

Accordingly, the weak relation of the SSD is defined as:

Definition B.2. A random return X stochastically dominates another random return Y in the second order stochastic dominance sense, denoted $X \succeq_{SSD} Y$, if

$$F_X^{(2)}(\eta) \leq F_Y^{(2)}(\eta) \quad for\ \eta \in \mathbb{R} \tag{B.4}$$

It is equivalent to this statement: a random variable X dominates the random variable Y if $E[u(X)] \geq E[u(Y)]$ for all non-decreasing concave functions $u(\cdot)$ for which these expected values are finite. Thus, no risk-averse decision maker will prefer a portfolio with return rate Y over a portfolio with return rate X (Ortobelli et al. 2013). Changing the order of integration, the ordering $X \succeq_{SSD} Y$ is equivalent to the expected shortfall (Ortobelli et al., 2013; Ogryczak and Ruszczynski, 1999):

$$F_X^{(2)}(\eta) = E[(\eta - X)_+] \quad for\ \eta \in \mathbb{R} \tag{B.5}$$

where $(\eta - X)_+ = max(\eta - X, 0)$. In this case, the function $F_X^{(2)}(t)$ is continuous, convex, nonnegative and non-decreasing. It is well defined for all random variables X with finite expected value.

Similarly, a condition which called third-order stochastic dominance (TSD) is considered; see Whitmore (1970). TSD is a less restrictive form than SSD since it considers preference ordering only for those risk-averse investors who exhibit decreasing risk aversion, see among others Post and Kopa (2016). The economic meaning of the TSD can be explained as follows. Denote $\mathcal{U}_3$ the set of all utility functions that are non-decreasing, concave and have a non-negative third derivative. Thus, $\mathcal{U}_3$ represents the class of non-satiable, risk-averse investors (with decreasing risk aversion) who prefer positive to negative skewness.

Definition B.3. We say that a random return X dominates another random return Y in the TSD sense, denoted $X \succeq_{TSD} Y$, if

$$F_X^{(3)}(\eta) \leq F_Y^{(3)}(\eta) \quad for\ \eta \in \mathbb{R}, \tag{B.6}$$

where $F_X^{(3)}(\eta) = \int_{-\infty}^{\eta} F_X^{(2)}(t)dt \quad for\ \eta \in \mathbb{R}$.

Equivalently, $X \succeq_{TSD} Y$, if $Eu(X) \geq Eu(Y), \forall\ u \in \mathcal{U}_3$.

The set of utility function $\mathcal{U}_3$ is contained in the set of nondecreasing concave utilities, $\mathcal{U}_3 \subset \mathcal{U}_2$, therefore, the condition (B.4) for SSD is only sufficient in the case of TSD

$$X \succeq_{SSD} Y \Rightarrow X \succeq_{TSD} Y \tag{B.7}$$

The condition that characterizes the TSD can be derived as follows,

$$X \succeq_{TSD} Y \Leftrightarrow E[(X-\eta)_+^2] \leq E[(Y-\eta)_+^2] \quad for\, \eta \in \mathbb{R} \tag{B.8}$$

where $(x-\eta)_+^2$ notation means the maximum between $x-\eta$ and zero raised to the second power, $(x-\eta)_+^2 = \left(max(x-\eta, 0)\right)^2$. The quantity $(X-\eta)_+^2$ is known as the second lower partial moment of the random variable *X*. It gauges the variability of *X* below a target payoff level η. Assume that *X* and *Y* have the same mean and variance but different skewness. If *X* has a positive skewness and *Y* has a negative skewness, then the variability of *X* below any target payoff level η will be smaller than the variability of *Y* below the same target payoff level, for deeper discussion see, among others, Rachev et al. (2008) and Whitmore (1970).

This method can be generalized to the *n*-th order stochastic dominance. Denote by $\mathcal{U}_n$ the set of all utility functions, the derivatives of which satisfy the inequalities $(-1)^{k+1}u^{(k)}(x) \geq 0$, $k = 1,2,\dots,n$ where $u^{(k)}(x)$ denotes the *k*-th derivatives of $u(x)$. Assume that the absolute moments $E|X|^k$ and $E|Y|^k$, $k = 1,2,\dots,n$ of the random variables X and Y are finite. We say that a random variable *X* dominates a random variable *Y* with respect to the *n*-th order stochastic dominance, $n = 1, 2, \dots$, denoted $X \succeq_n Y$, if

$$X \succeq_n Y \text{ if } Eu(X) \geq Eu(Y), \forall\, u \in \mathcal{U}_n. \tag{B.9}$$

Therefore, the first-order, second-order, and third-order stochastic dominance can be seen as special cases from the *n*-th order stochastic dominance when $n = 1, 2, 3$ respectively.

There exists an equivalent way to define the *n*-th order stochastic dominance in terms of the CDFs. The condition is as follows

$$X \succeq_n Y \quad \Leftrightarrow \quad F_X^{(n)}(\eta) \leq F_Y^{(n)}(\eta) \quad for\ \eta \in \mathbb{R} \tag{B.10}$$

where $F_X^{(n)}(\eta)$ denotes the *n*-th integral of the CDF of *X*, which can be defined recursively as

$$F_X^{(n)}(\eta) = \int_{-\infty}^{\eta} F_X^{(n-1)}(t)dt \quad for\ \eta \in \mathbb{R} \tag{B.11}$$

Equivalently a condition that characterizes *n*-th stochastic dominance can be derived as follows

$$X \succeq_n Y \quad \Leftrightarrow E[(\eta - X)_+^{n-1}] \leq E[(\eta - Y)_+^{n-1}] \quad for\ \eta \in \mathbb{R} \tag{B.12}$$

$(\eta - x)_+^{n-1} = \left(max(\eta - x, 0)\right)^{n-1}$. This condition clarifies why it is necessary to assume that all absolute moments until order *n* are finite. Since, in the *n*-th order stochastic dominance, we provide the conditions on the utility function as *n* increases, the following relation holds,

$$X \succeq_1 Y \quad \Rightarrow \quad X \succeq_2 Y \quad \Rightarrow \dots \Rightarrow \quad X \succeq_n Y \tag{B.13}$$

which generalizes the relationship between FSD, SSD and TSD.

To conclude, it is possible to extend the *n*-th order stochastic dominance to the α order stochastic dominance in which $\alpha \geq 1$ is a real number and instead of the ordinary integrals of the CDFs, fractional integrals are involved. Generally speaking, X dominates Y with respect to the α stochastic dominance order $X \succeq_\alpha Y$ (with $\alpha \geq 1$) if and only if $E[u(X)] \geq E[u(Y)]$ for all u belonging to a given class $\mathcal{U}_\alpha$ of utility functions (Ortobelli et al., 2009).

Increasing convex order

One of the main problems of mathematical finance is the comparison of risks; see Muller and Stoyan (2002). In this context, stochastic order is too restrictive and it does not capture the important notion that risk should also depend on variability. For this reason alternative notions of partial order for distributions have been investigated extensively, and the increasing convex order (ICX) has turned out to play a major role in this application area.

Definition B.4. Given two random variables *X* and *Y* with respective distribution functions F and G, *X* is said to be smaller than *Y* in the increasing convex order, denoted $Y \succeq_{icx} X$, if

$$E[v(Y)] \geq E[v(X)],$$

for all increasing convex function v, provided the expectations exists. Equivalently, if

$$E[(X-t)_+] \leq E[(Y-t)_+] \qquad \text{for all } t \geq 0,$$

see, for example, Theorem 1.5.7 in Muller and Stoyan (2002).

Many properties and applications of this order can be found in the books by Muller and Stoyan (2002), Denuit et al. (2005), Shaked and Shanthikumar (2007). For example, the ICX is characterized in terms of the tail value-at –risk as follows:

$$Y \succeq_{icx} X \text{ if and only if } TVaR_p(X) \leq TVaR_p(Y), \text{ for all } p \in (0,1)$$

where $TVaR_p(X) = \frac{1}{1-p}\int_p^1 F^{-1}(t)dt, p \in (0,1)$. Therefore, for continuous random variables, we have

$$Y \succeq_{icx} X \text{ if and only if } CVaR_p(X) \leq CVaR_p(Y), \text{ for all } p \in (0,1).$$

From definitions B.2 and B.4, we clearly observe that the SSD relation ICX relation are connected as follows

$$Y \succeq_{icx} X \quad \Leftrightarrow \quad -X \succeq_{SSD} -Y$$

The ICX ordering (generally less used in financial decision problems) accounts for the choice of non-satiable risk-seeking investors according to Muller and Stoyan (2002) studies.

Abbreviation of the terms used in this dissertation

Terms	
a.s.	almost surely
BS	Black and Scholes
cdf	cumulative distribution function
CVAR	Conditional Value-at-Risk
pdf	probability density function
FTAP	Fundamental Theorem of asset pricing
FSD	First-order stochastic dominance
GBM	Geometric Brownian motion
ICX	Increasing convex dominance
IV	Implied volatility
IVS	Implied volatility surface
IVSD	Implied volatility surface differences
MA	Moving average
MSE	mean square error
NYSE	New York Stock Exchange
OLP	Ortobelli Lando and Petronio
OLS	Ordinary least squares
OTC	Over-the-counter
KOT	Kouaissah Ortobelli and Tichý
MAD	Mean-Absolute Deviation
PCA	principal components analysis
TOK	Tichý Ortobelli and Kouaissah
TSD	Third-order stochastic dominance
SPD	State Price Density
SSD	Second-order stochastic dominance
RR	Rachev ratio
RW	Ruppert and Wand
SR	Sharpe ratio
S&P 500	Standard & Poor's 500
VS	Volatility spread
VAR	Value-at-Risk

List of Figures and Tables

Figures

Tables