Simulation based performance measures applied to new technical analysis and portfolio optimization strategies

Thesis

presented to the Faculty of Economics and Social Sciences
at the University of Fribourg (Switzerland)

by

Didier MARTI

from Tschugg BE

in fulfilment of the requirements for the degree of
Doctor of Economics and Social Sciences

Fribourg, 2013

The Faculty of Economics and Social Sciences at the University of Fribourg neither approves nor disapproves the opinions expressed in a doctoral thesis. They are to be considered those of the author. (Decision of the Faculty Council of 23 January 1990)

A mes parents

Preface

J'aimerais tout d'abord remercier mes parents qui m'ont soutenu et encouragé tout le long de mes études. En effet, mon début de parcours scolaire ne laissait pas penser à ce qu'il aboutisse à une thèse universitaire. Je tiens également à relever l'exemple fourni par mon grand frère, Claude-Oliver. En effet, je n'allais quand même pas le laisser être le seul Dr. Marti de la famille!

Ma seconde pensée va à mon directeur de thèse, le Prof. Dr. Dušan Isakov qui m'a guidé à partir de mon travail de Master et m'a activement introduit dans le monde de la recherche académique. Je lui suis particulièrement reconnaissant de sa confiance et de m'avoir laissé une grande liberté de recherche qui a rendu mes quelques années passées dans sa Chaire de Gestion Financière de l'Université de Fribourg très agréables et enrichissantes. J'aimerais également remercier le Prof. Dr. Martin Wallmeier pour son évaluation en tant que deuxième rapporteur.

Finalement, je remercie chaleureusement mes amis et anciens collègues de la faculté des sciences économiques. Je suis particulièrement redevable à Philippe Masset sans oublier Jean-Philippe Weisskopf, Florent Ledentu et David Ardia pour son expertise en économétrie.

A thinker sees his own actions as experiments and questions - as attempts to find out something. Success and failure are for him answers above all.

Friedrich Nietzsche, *The Gay Science.*

Contents

x

List of tables

List of figures

List of abbreviations

AR: Auto regressive

ARMA: Auto regressive moving average

BH: Buy-and-hold strategy

BLL: Brock, Lakonishok and LeBaron (1992)

CAPM: Capital Asset Pricing Model

CBOE: Chicago Board of Options Exchange

DAX: Deutscher Aktien Index

DJIA: Dow Jones Industrial Average

EMH: Efficient markets hypothesis

GARCH: Generalized Autoregressive Conditional Heteroskedasticity

MA: Moving average

RC: Reality Check

RW: Random walk

S&P 500:   Standard and Poor's 500

STA: Superior Predictive Ability

TRB: trading range break-out

USD: United States Dollar

Notation

1	A vector of ones
α:	Strategy or portfolio alpha
B:	Bandwidth
β:	Asset beta computed with respect to the market return
Bl:	Block length used in a block bootstrap simulation
j:	A portfolio optimization specification set
JB:	Jarques-Bera statistic
Ku:	Kurtosis
M:	Markov chains transition probability matrix
$M_{t,S}$:	Short moving average computed over the last S observations
$M_{t,L}$:	Long moving average computed over the last L observations
Nb:	Number of buy signals
Ns:	Number of sell signals
N:	Length of the time series
n:	Number of assets in the portfolio optimization setting
$O_{A,t}$:	Option ask price at time t
$O_{B,t}$:	Option bid price at time t
$O_{C,t}$:	Option closing price at time t
P_t:	Index price
p:	Markov chain state
R_t:	Index return (equivalent to the buy-and-hold return)
R^*_t:	Simulated series return (a strategy or a artificial buy-and-hold series)
$R_{assets,t}$:	Vector of assets returns used in the portfolio optimization setting
$R_{opt,t}$:	Return of an option
$R_{B,t}$:	Borrowing rate
$R_{C,t}$:	Call option return
$R_{DL,t}$:	Return of a debt leveraged strategy
$R_{L,t}$:	Lending rate (or risk-free rate when only one rate is used)
$R_{P,t}$:	Put option return
$R_{Strat,t}$:	Return of an investment strategy (with or without leverage)
S_t:	Trading signal at time t (1 for buy, -1 for sell and 0 for neutral)
S^*_t:	Simulated trading signal at time t (1 for buy, -1 for sell and 0 for neutral)
$S_{BH,t}$:	Dummy variable that takes a 1 (-1) if the market is in a bull (bear) phase
St_{org}:	A statistic computed with the original strategy returns series

St^*:   A statistic computed with a simulated strategy returns series

Sk: Skewness

SB_t: One if the strategy generates a buy signal at time t, zero otherwise

SS_t: One if the strategy generates a sell signal at time t, zero otherwise

σ: Standard deviation

Σ_t: Covariance matrix used as input in the portfolio optimization

Σ_{S_t}:   Sample covariance matrix

TC: One-way transaction cost

UP_t: Option underlying price at time t

μ_t: Expected returns vector used as input in the portfolio optimization

X: Option strike price

x_t: Mean-variance optimal portfolio weights

W^*_t: Simulated portfolio weights

w_t: Shrinkage factor

Thesis introduction

The efficient market hypothesis (EMH) is a pillar of the modern finance theory. It asserts that financial markets integrate all information immediately and accurately. Fama (1970) further define the EMH by proposing three levels of efficiency; the weak, semi-strong and strong form. They differ according to the kind of information considered; the first case states that past prices contains no useful information to predict future prices movements, the second supposes that prices are adapted immediately to all publicly known information, while the last form suggests that all information, including the private and hidden one, is reflected in prices. The EMH implies that prices follow a random walk, and thus, are not predictable. Jensen (1978) further differentiates between the statistical predictability and the economic performance. He argues that a market is efficient with respect to a set of information if it is not possible to generate economic profits, once the returns are adjusted to risk and transaction costs are taken into account. Thus, an investor should expect a normal return on his portfolio, which is defined according to the amount of risk that he accepts to bear.

Despite the importance of this hypothesis in the academic literature, the financial industry is based on the opposite idea; active management is able to outperform its benchmark, either by timing the market or by selecting investment opportunities that are not correctly priced by the market. This is illustrated by the large amount of actively managed funds proposed to investors and the amount of predictions about the market evolution and companies future earnings and returns found in the financial media. A major motivation for the academic literature about active asset management is that investment methods that have been either introduced or improved by academic studies are widely used in practise. For instance, several surveys indicate that professional money managers commonly employ technical analysis methods. In addition, Amenc, Goltz and Lioui (2011) show that standard or a variation of the mean variance optimization, introduced by Markowitz (1952), to construct portfolios is also widely used by asset managers in Europe. As a concrete example, we may cite the various funds belonging to the "QUAM" family

proposed by the asset manager "Edmond de Rothschild PriFund"[1]. They consist in a purely mathematical approach to construct their portfolio, which is largely inspired by the standard mean variance optimization method.

As a consequence, a myriad of studies examine whether assets returns are predicable, whether mutual funds investing is more valuable than a buy-and-hold strategy and whether investment strategies based on past information, such as technical analysis, momentum or value/growth strategies, are profitable. These studies are a mean to test the EMH, and while we do not pretend to provide a comprehensive literature review about these issues, we present some of the major studies briefly.

First, Keim and Stambaugh (1986), Ferson and Harvey (1993), or more recently Ang and Bekaert (2007) find that stocks or bond returns can be predicted, at least to some extent, with lagged variables such as the credit-risk spread, the term structure of interest rates or the dividend yield. Cochrane (2008) argues that as the dividend growth is not predictable, then returns should be predictable in order to have the observed variation in the dividend yield. In addition, Lo and MacKinlay (1988) find that stock returns do not follow a random walk, but they also point out that this result is not a sufficient proof to reject the EMH. Second, in addition to these studies supporting some level of predictability, many investment strategies, which can be either explicit or result from anomalies in stock returns, appear to be able to generate abnormal returns. For example, Brock, Lakonishok and LeBaron (1992) show that simple technical analysis strategies are able to outperform the buy-and-hold performance. Jegadeesh (1990), Jegadeesh and Titman (1993) and Lakonishok, Shleifer and Vishny (1994) document various aspects of prices reversal or momentum according to the temporal horizon. Among the anomalies that are used to develop investment strategies, we may cite the size and the calendar effects and the highest return obtained by value stocks compared with growth stocks.

Nonetheless, despite all these elements, the literature on the performance of actively managed mutual funds inclines toward the validity of the EMH. Indeed, most of these studies find that the vast majority of funds are not able to provide a higher risk-adjusted performance than passively investing in the benchmark. This raises the question about the opportunity to exploit the returns predictability and these anomalies in a real trading setting. Indeed, transaction costs, the data-

---

[1] A prospectus can be freely downloaded on the "Banque Privé Edmond de Rothschild" website. No direct link is provided as each user has to be identified first.

mining issue, an incomplete risk adjustment are among the reasons why these findings may only hold in academic studies. Malkiel (2003) presents a review of these anomalies. While he does not pretend that market prices are always perfectly set, he questions the presence of trading opportunities that would enable to earn a risk-adjusted abnormal return. However, the jury is still out on whether this hypothesis is valid or not.

As empirical studies about investment strategies require a testing procedure in order to determine whether a risk adjusted abnormal return can be obtained, we consider these two issues with a common framework. In a first step, we focus on investment strategies that mainly use equity indices as their investment universe. We analyse some new specifications of well-known investment strategies, i.e. technical analysis and mean variance portfolio optimization. We choose them as they are widely used in practice and they do not involve any subjectivity from the manager in their construction. The second objective is to point out the importance of the testing procedure used to determine if a strategy generates abnormal returns, and to propose a new technique based on the simulation of trading positions. Indeed, the performance measure is primordial both for professional managers and for academics. The former's remuneration and overall success depends on it and for the latter, it may provide evidence in favour or against the EMH.

The overall structure of this thesis is the following: In the first Part, we examine in detail some new technical analysis strategies based on moving-averages. The novelty of our approach resides in exploiting long-term trends, as we use long moving-averages up to four years, while other studies only consider trends up to one year. Indeed and to the best of our knowledge, we are the first to examine commonly used technical trading rules in a long-term setting. As data-snooping is always an issue of this kind of study, we focus on out-of-sample processes that select the parameters used by the investment strategy objectively. We find that our rules produce returns twice as high as the benchmark over a 15 years sample. Then, we propose a new market timing test that focuses on market phases, i.e. bull and bear markets, and we show evidence supporting the fact that the trading strategy invests according to these trends. Consequently, it changes trading position very rarely, and thus, common issues such as transaction cost or market microstructure do not impact its effective implementation. Nonetheless, despite its strong economic performance, Student t-tests are not able to reject the null hypothesis of equal means, between the strategy and the benchmark.

The second Part is dedicated to mean-variance optimal portfolios. Specifically, we investigate the impact of using various lengths of the estimation window. Indeed, this aspect is seldom addressed in the related literature, which usually focuses on developing sophisticated parameters

estimation models. First, we show that using shorter lengths than the standard five years for the estimation window improves the optimization performance dramatically. Moreover, this aspect of the inputs estimation seems to affect the performance to a greater extent than the choice of various popular estimation models. However, as the optimization specification[2] has to be chosen arbitrarily, the data-snooping issue arises again. Hence, we propose to employ selection processes that are similar to the ones used in the technical analysis setting. We show that the resulting portfolios generate large excess returns compared with the equally weighted portfolio. However, the performance strongly depends on the ability to take short positions and the level of transaction costs. In contrast with the results obtained for the technical analysis strategies, the optimal portfolios, despite their large returns, can probably not challenge the EMH.

Finally, the last Part re-examines the conclusions of the first, i.e. the technical analysis strategies can produce large abnormal returns that are not statistically significant. We argue that the standard Student t-test used to determine statistically whether a strategy mean return differs from its benchmark is not a powerful procedure. Indeed, we first show that a strategy should generate a return at least three times as high as the benchmark strategy in order to find a statistically significant difference. Hence, we propose a new test based on the simulation of trading positions. This approach is not new; however, our proposed test is designed to keep the structure of the original trading positions. As the strategy invests according to long-term trends and changes trading positions very rarely, we argue that the artificial positions series should have a similar pattern. We propose to model the trading positions as first-order Markov chains. This has the merit to limit the number of assumptions to only one; the next trading signal depends on the current one only. Thus, we do the performance analysis, presented in the first part, once again with the proposed test, and we find that all p-values associated with various performance measures are lower than the standard 5% confidence level. This means that the large abnormal returns can not be replicated by taking random positions, with a similar structure to the original one, in the market. The difference in conclusions between the two testing procedure in striking: With Student t-tests, we would conclude that the strategies possess some predictive power, but it can not generate statistically significant abnormal returns. This contrasts sharply with our proposed simulation test that concludes that the performance is neither due to luck nor to risk bearing.

---

[2] By specification, we mean the models used to estimate the parameters and the window length over which the estimation is performed.

We also adapt the simulation test to strategies that invest in more than one asset simultaneously, such as the optimal portfolios examined in the second Part. In this case, the Markov chains are not suitable, thus, we propose to use a block bootstrap to keep the dependencies in the portfolio weights at least to some extent. Then, we apply this test to these portfolios and compare the results with standard Student t-test. The findings are in line with those related to technical analysis, even if the difference in conclusions with a 5% confidence level is less marked.

To conclude this last Part, we conduct a comprehensive Monte-Carlo experiment to compare the power of various testing procedures. We focus on the level of abnormal returns required to reject the null hypothesis of equal performance. We find that the proposed simulation test power is more than or at least equal to Student t-tests computed on several performance measures, such as the difference in mean returns, the Sharpe ratio or the Jensen's alpha. Nonetheless, large economical returns are necessary to reject the null, even with our test.

Part I: A long term perspective on technical analysis strategies

1. Introduction

Technical analysis comprises a wide range of methods used to forecast future price movements of stocks, currencies or commodities, based on their past prices and volumes. These methods may be classified into two broad categories, charting and technical trading systems. The first group consists of methods analysing charts to detect price patterns that are supposed to repeat themselves. The second group includes a variety of quantitative rules aimed at detecting trends and generating trading signals accordingly. Among them, the moving average, referred to as MA thereafter, and the filter trading rule are the most popular[3].

Several surveys conducted with professional investment mangers show that the vast majority of them use technical analysis to some extent. Allen and Taylor (1990) in the London foreign exchange market, Lui and Mole (1998) in Hong Kong and Oberlechner (2001) in various European markets document that technical trading is broadly used in order to forecast short-term trends. These surveys also point out that its utilization by professional diminishes as the forecasting horizon increases. In addition, these techniques are not regarded as being in contradiction with fundamental analysis, but they are used in a complementary approach. According to Gehrig and Menkhoff (2006), the use of technical analysis increases during the

---

[3] In this thesis, the trading rule relates to the method according to which trading signals are generated. The trading system consists in using more than one indicator to produce these signals. Strategies reflect the effective positions taken in the market by either combining long and short position or using financial leverage.

nineties. They reach this conclusion by comparing surveys conducted in 1992 and again in 2001 among German and Austrian foreign exchange dealers and funds managers.

On the other hand, academics have been skeptical about the utility of these forecasting methods. Among the various conceivable reasons, we can mention that they lack a sound theoretical basis and the parameters selection is usually not disclosed or justified. In addition, empirical evidence of technical trading profitability is mixed, and strongly depends on the choice of the time interval, the set of methods considered or the underlying asset. Another concern about technical analysis arises as reported evidence of profitability may be biased by data-snooping issues. Indeed, using the same data and trading rules repeatedly may lead to find a few profitable rules by luck.

The objective of this first Part is to examine some new MA strategies based on the out-of-sample approach of Sullivan, Timmermann and White (1999), Skouras (2001) and Fong and Yong (2005). Instead of choosing the parameters arbitrarily, we utilize various selection processes that combine or select trading signals issued by simple MA rules. Furthermore, we extend the literature by considering a much wider range of parameters, for both the short and the long MA. Indeed, all studies related to MA trading rules focus on short-term trends, as they use a long MA up to only 200 or 250 days usually. To the best of our knowledge, nobody examines MA rules in a long-term setting, with parameters longer than 250 days. Indeed, long-term trends may be more easily identified, and they may be less noisy than their short-term counterparts. In addition, there is a possibility that when a short-term trend is identified, it is already too late to exploit it. We find evidence that our trading systems provide much higher returns than usual MA rules. The annual mean returns of our strategies lie between 10.7% and 14.6%, compared with 6.1% for the buy-and-hold. Moreover, the strategies are especially profitable during the most recent period, while the majority of studies find that the technical analysis performance decreases over time. However, Student t-tests fail to reject the null hypothesis of equal means between the strategies and the benchmark. Appendix I-A presents a statistical reminder about these tests. In Part III, we argue that these results are due to the Student t-test low power, and thus, we propose a more powerful test based on the simulation of trading signals. In this first part, we also develop a new market timing test that determines whether a strategy relies on long-term trends related to the business cycle. We find that the strategies trading position coincide with bull and bear market phases to an extent that can not be achieved by luck.

A second objective of this first Part is to examine technical strategies with financial leverage. Indeed, technical trading is widely used by commodity trading advisors (CTA), a class of hedge funds, which may rely on financial leverage to increase the performance. For this purpose, we

consider debt leverage and exchange-traded options. However, we first ensure that the complex rules possess significant forecasting abilities. Indeed, using leverage without these abilities should not add any significant value after the risk adjustment. A second reason for using financial leverage lies in the possibility that even successful forecasting methods may not generate abnormal returns if the market follows a strong upward trend. We find that the superior performance of debt leveraged strategies could not be attributed to leverage itself, but to their predictability. This contrasts with options based strategies, which are not profitable mainly because of the loss of time value. Finally, we provide some insights about the strategies performance analysed with alternative risk measures, which include higher moments of the returns distribution. We show that our strategies are attractive as they may hedge skewness risk without sacrificing returns.

Finally, we perform a simulation analysis to shed light on a puzzling aspect of our results. Indeed, we find that strategies yield economically significant returns in excess of the buy-and-hold. However, we also find that their percentages of correctly predicted trading signals are only slightly higher than the buy-and-hold. Thus, we examine this relationship, and we conclude that only a small increase in these percentages leads to high excess returns.

The structure of this Part is the following: Section 2 presents a comprehensive literature review about related studies. In Section 3, we describe the trading systems. Section 4 includes the description and the empirical implementation of the trading strategies. It is divided in four parts; the first one examines the performance of the simple MA rules to shed light on the impact of using an extended set of parameters, while the three other ones display the performance analysis of the strategies without leverage, with options or with debt. Section 5 displays a risk analysis with alternative measures. In Section 6, we describe and implement our new market timing test based on market phases. In Section 7, we conduct a simulation experiment to explain the relationship between the percentage of right signals and a strategy performance in excess of the buy-and-hold. Finally, Section 8 concludes.

2.   Literature review

In this literature review, we provide a rather comprehensive overview that do not only covers the issues investigated in this thesis, but also other important aspects that decide whether abnormal returns can be obtained in a real trading setting. Specifically, it focuses on the various proposed trading systems, which combine several simple technical indicators. Appendix I-B

briefly presents the most commonly used simple trading rules. Then, we summarize studies related to the performance of technical analysis, and the underlying reasons why technical analysis may be useful as a prediction method. This review focuses on papers that apply trading rules to equity indices or stocks, but we sometimes refer to studies on foreign exchange when they provide some interesting methodological aspects.

2.1   Complex trading systems

Here, we define a complex trading system, or complex trading rules, as a procedure that combines several simple technical trading indicators to generate trading signals. The objective is to use more information than a single indicator, but also to take into account various aspects considered by the simple indicators. Some others trading systems, referred to as complex rules, are based on a selection and an out-of-sample test sample. These methods should mitigate the data-snooping bias.

The most straightforward way to combine several trading signals is simply to compute them independently, and then, take a position only if they reach a consensus. Pruitt and White (1988) propose the CRISMA trading system, which stands for "Cumulative volume, Relative strength, Moving Average" and combines three simple technical indicators. The system takes a position only if the three indicators generate the same signal. Fang and Xu (2003) follow the same method to compute complex trading signals. They combine MA rules signals with time series forecasts generated by different models, such as the auto-regressive model or various kinds of GARCH models. They show that these two forecasting methods capture different aspects of the price predictability, and thus, are complementary. Indeed, the MA rules tend to produce better forecasts for upward trends, while time series forecasts are more accurate for detecting downward trends.

Gençay, Dacorogna, Olsen and Pictet (2003) examine the performance of a commercial real-time trading model, the RTT model, developed by Olsen & Associates. It uses a trend following technique based on an equally-weighted iterative moving average of 20 days. In addition, it also determines the percentage of the capital to be invested according to the strength of the trading signal. This contrasts with most of the technical analysis strategies that either invest the entire capital or take a short position corresponding to the whole capital. The trading system also contains a contrarian indicator to diminish the position exposure to extreme exchange rate movements. Nevertheless, such a position against the current trend may be taken only if the profitability has already reached a yearly return objective level, set at 3%. The authors argue that

analysing this model may shed light on the usefulness of technical analysis, as its performance may proxy very well those of an investor (or a currency dealer) in a real trading setting. Indeed, it considers intraday data, transaction costs with the bid-ask spread, real trading hours and the model trading frequency is in line with those of a real trader.

Bessembinder and Chan (1998) use a portfolio approach to aggregate the set of 26 rules used by Brock, Lakonishok and LeBaron (1992), referred as to BLL thereafter. They allocate an equal fraction of the invested capital according to the signal emitted by each rule. This is also a way to avoid, or at least to minimize, the data-snooping bias resulting from using only the best rules in an in-sample analysis. A slightly different approach is proposed by Chang and Osler (1999). Their strategy takes a position twice as large as usual if both the head-and-shoulders signal and a signal issued by the most profitable MA or momentum rule coincide. Otherwise, the system does not take any position.

Another way to combine signals from simple technical indicators is to use selection algorithms. They consist in selecting, during an evaluation interval, some rules specifications according to a criterion. Then, these specifications are evaluated over an out-of-sample test interval. Hsu and Kuan (2005) implement complex trading rules[4] only based on simple technical indicators, such as MA rules, trading range break-out (referred as to TRB thereafter) and head-and-shoulders among others. They implement three kinds of strategies in order to combine the signals issued by the simple rules. The first is a learning strategy that selects the best-performing specification from a particular class of rules during a predetermined test sample. The second is the voting strategy that gives one "voting right" to each specification, and the system uses the position that has received the largest number of votes. The last one, the fractional position strategy represents an average of the signals produced by all specifications inside a class of rules[5]. The purpose of these complex strategies is to reach a consensus by using more information than it would be possible with a single indicator. Skouras (2001) proposes to use only MA rules with a short window restricted to one day. Nevertheless the length of the long window is not constrained to take a predetermined value, such as 10, 20, 50, 100 or 200 days, which is usually done in other studies. Here, he considers all values from one to 200 days. He shows that the

---

[4] The rules considered by our empirical work are closely related to them. Thus, more details are given thereafter.

[5] For instance, if the 55 MA rules generate 32 buy and 23 sell signals, the position taken by the complex strategy is a long position amounting to 16% of the available capital.

optimal rule chosen by the so called "artificial technical analyst" varies significantly through time and presents pronounced discontinuities. Fong and Yong (2005) consider a similar approach but without restraining the short MA to one day. Lee and Mathur (1996), Maillet and Michel (2000) and Olson (2004) employ an easier procedure: They divide their sample in five years subsamples and apply the best in-sample rule to the next out-of-sample period. Moreover, these strategies are a way to conduct out-of-sample tests, as the parameters are not, or to a lesser extent, chosen arbitrarily at the beginning of the sample. Furthermore, by comparing the performance of a complex strategy with the best rule over the whole sample, we can determine whether the trading system is able to identify the best rule ex-ante. For instance, Cooper (1999) shows that out-of-sample profits are between 25% and 40% lower than those obtained during the in-sample interval used to select the parameters.

Finally, some methods inspired by biology models are adapted to generate trading programs. Allen and Karjalainen (1999) propose a genetic program to combine simple MA and trading range break-out rules to produce "optimal" complex trading rules. They are generated during an estimation and validation period, and then, they are applied to an out-of-sample test period. The idea is to combine simple rules by weighting them according to a fitness criterion, in this case the return of the rule in excess the buy-and-hold. While it is appealing in the sense of minimizing the in-sample selection bias, these complex systems may have a very complex structure, and thus, they are rather difficult to interpret. Neely, Weller and Dittmar (1997) use a similar method with another criterion, as they apply the genetic program to foreign exchange markets. Thus, they do not use the return of the strategies in excess of the buy-and-hold, but they adapt the returns to the interest rate differential. This approach to generate trading rules is also examined by Neely (2003) with an emphasis on risk, Potvin, Soriano and Vallée (2004) with individual Canadian stocks and Neely and Weller (2003) with intraday data. Finally, Brabazon and O'Neill (2004) propose a grammatical evolution model inspired by the biological process of mapping of genes to proteins[6]. They use MA, momentum and TRB indicators as inputs into their evolutionary automatic programming model to generate complex trading rules. They also consider the maximal drawdown as the fitness criterion to account for risk and to avoid extreme losses.

---

[6] Brabazon and O'Neill (2004), page 312.

2.2   Technical analysis profitability

This review focuses on modern studies on the stock market, starting with Brock, Lakonishok and LeBaron (1992). Indeed, they usually include several risk measures and a comprehensive performance analysis. For a comprehensive review of early studies, the interested reader may refer to Park and Irwin (2007).

2.2.1   Studies in favour of technical analysis profitability

Brock, Lakonishok and LeBaron (1992) find that simple technical trading rules (MA and TRB) applied to the DJIA over a century of data produce significant excess returns. In addition, they implement a bootstrap technique to address some limitations of traditional hypothesis tests with Student t-tests, such as the hypothesis of normal returns. They find that the profitability can not be replicated by rules simulated on prices series obtained with various returns generating models, such as the random walk or GARCH models. Their methodology, i.e. the set of rules and the bootstrap test, has influenced a multitude of studies using the same approach for various time intervals and markets. Skouras (2001) shows that using a selection method to find the "optimal" MA rule parameters results in both higher returns and their associated t-statistics. The latter test whether returns are different from the buy-and-hold or zero, and they are much higher than those obtained, on average, by BLL. The difference is also economically significant, as the profits generated by the optimal rule are more than three times higher than those obtained by BLL. Fong and Ho (2001) find that basic trading rules based on MA are profitable when they are tested on US Internet stocks, even after considering transaction costs and a time-varying risk premium. Moreover, they argue that these profits result from the rules predictive power, as they find that the percentage of right buy signals (58% on average) is not due to luck.

Further evidence of profitability is provided by Wong, Manzur and Chew (2003) who consider various kinds of MA rules and the relative strength index on the main Singaporean index between 1974 and 1994. The rules produce, on average, strategies returns that are statistically different from zero with confidence levels ranging from 1% to 10%. Furthermore, even the unprofitable strategies show some predictability, as buy returns are higher than those following sell signals. Isakov and Hollistein (1999) find that simple technical indicators produce significant returns on a Swiss index from 1969 to 1997. Indeed, a simple MA rule yields a yearly return of 24.6% compared with only 6.25% for the buy-and-hold. However, they show that these profits are achievable only with low transaction costs, as that obtained by large institutional investors. Fang and Xu (2003) propose a trading strategy that combines signals from MA rules and various time series forecasting models. The strategies generate positive returns with break-even transaction

costs ranging from 1% to 2% for three US stock indices. They also show that analysing the two forecasting techniques separately results in lower break-even transaction costs.

Gençay (1998) reports that his feedforward network[7] model generates an after transaction costs profitable trading strategy on the DJIA over the period 1963-1988, and also during all six subsamples considered. The percentage of right signals ranges from 57% to 61%. He also provides statistical evidence of market timing abilities according to the Pesaran and Timmermann (1992) and the Henriksson and Merton (1981) tests[8]. Using a similar model, Fernández-Rodríguez, González-Martel and Sosvilla-Rivero (2000) reach the same conclusion with the General Index of the Madrid stock exchange. Nevertheless, the strategy is not profitable during the last subsample, from 1996 to 1997, and they ignore transaction costs. This is not surprising as the index is characterized by an exceptionally strong upward-trend, and thus, beating the buy-and-hold, which is always long, is nearly impossible without leverage.

The CRISMA trading system applied to individual US stocks by Pruitt and White (1988) outperforms the buy-and-hold over the 1976-1985 sample, even after adjusting for transaction costs and risk. They consider a similar methodology to event studies to compute abnormal returns with various models, such as the market model or the mean adjusted returns model. With transaction costs of 1%, excess returns are statistically and economically significant, as they range from 16.4% to 25.4% in an annual frequency. Even with transaction costs as high as 2%, the trading system still provides evidence of profitability. In this case, excess returns lie between 6.1% and 15.13%. Furthermore, they report in a subsequent paper, Pruitt and White (1989), that it could be profitable to use options instead of stocks to benefit from leverage. They are able to identify 171 trading signals that match available in-the-money call options. The average options return is 28.7% (12%) without transactions costs (with maximal retail costs) for an average holding period of 25 days. In addition, they show that more than 70% of the transactions are profitable. This percentage is significantly higher than what could be archived by luck. Pruitt, Tse and White (1992) suggest that the profitability of the CRISMA trading system remains significant during a more recent period, from 1986 to 1990.

---

[7] This is a variant of an artificial neural network model, thus, it is nonparametric.

[8] Market timing consists in increasing (decreasing) the portfolio exposure to the market when it follows an upward (downward) trend. More details about these tests are given in part III.

Cooper (1999) investigates a contrarian strategy based on filter rules and also volume data. It produces statistically and economically significant returns in an out-of-sample analysis. Furthermore, he finds that the technical analysis strategy is more profitable compared with traditional contrarian strategies that divide stocks between looser and winner with either market-adjusted or raw past returns. He suggests that the filter rule helps to distinguish between truly overreacting stocks from noisy signals. He also argues that the profitability is not likely due to either risk or microstructure issues, as only large US stocks are considered. Finally, transaction costs do not eliminate the profits, at least for institutional investors.

Szakmary, Shen and Sharma (2010) document that MA rules are similar to momentum strategies. Indeed, they find a comparable performance when the trading rules are calibrated with the same horizon. They find that most of the rules, applied to commodity futures, are profitable.

Skouras (2001) proposes another method for testing whether technical analysis strategies may challenge the EMH. Instead of testing if the strategies are able to generate returns in excess of the buy-and-hold, he proposes to investigate whether these strategies improve an investor utility. He compares the technical analysis strategies with others that do not use past prices, such as the buy-and-hold. Furthermore, the degree of market efficiency may also be tested by defining for which classes of investors (i.e. determined by a specific behaviour toward risk) these technical strategies improve utility. He finds that the MA strategies would increase an investor utility if he is either risk neutral or govern by a quadratic utility function (i.e. the mean-variance case). In case of risk-averse investors, characterized by a concave utility function, the results are mixed. Those who are particularly averse to extremely low returns would not use technical analysis. Nevertheless, some of them would benefit from taking past prices into account for their investment decision. Thus, market efficiency may be rejected for risk neutral and mean-variance investors, as well as for some of the risk-averse investors. Transaction costs do not invalidate these conclusions, at least for institutional investors. Dewachter and Lyrio (2005) also find that simple MA rules, applied to foreign exchange markets, increase investor utility. They consider the standard utility maximisation problem with constant relative risk aversion (CARRA). They propose to solve Euler equations, which are used to find the optimal weight to allocate in the risky asset. They show that using MA rules trading signal as conditional information in the Euler equation improves the utility significantly.

This Section provides a brief review of studies supporting technical analysis profitability. They are summarized in Appendix I-C. Some studies presented above find that some strategies are able to generate excess returns after transaction costs and risk adjustments. In this case, the

EMH may be rejected. Nevertheless some arguments favouring that these results are only theoretical but not realistic in a real trading context are reviewed in Section 2.5.

2.2.2   Studies that find no technical analysis profitability

The studies that do not find technical analysis excess returns may be classified into two distinct groups. The first one includes studies in which technical indicators possess predictability to some extent; however, the strategies based on them are not profitable. One of the main explanations is that the transaction costs eliminate the profit. The second group consists of studies that conclude that technical analysis has no predictive power.

Hudson, Dempsey and Keasey (1996) consider a similar methodology to BLL to examine the performance trading rules on the main UK stock index, the FT 30, from 1935 to 1994. They find that the profitability decreases over time, and moreover, the strategies are not profitable after transaction costs over the most recent subsamples. Nonetheless, predictability can not be rejected, as conditional buy returns are higher than conditional sell returns. These results are in line with Ratner and Leal (1999) and Taylor (2000) who uses a broader index, the FT all Shares, but also 12 individual UK stocks and the two major US indices. In addition, he can not reject the random walk hypothesis for all these series. The strategies applied to the DJIA are more promising as they produce break-even transaction costs[9] of 1.1%. Further evidence that predictability does not necessarily generate statistically significant profits after transaction costs is presented by Allen and Karjalainen (1999). They use genetic programming trading rules on the S&P 500 over the period 1928-1995. Furthermore, Neely (2003) points out the importance of considering, not only excess returns, but also risk adjustments in order to test market efficiency. He uses a genetic programming approach to generate trading rules in an out-of-sample setting. He finds that they do not outperform the buy-and-hold significantly with risk-adjusted measures, such as the Sharpe ratio, the Jensen's alpha or another measure based on investor utility. He also proposes to use these risk-adjusted measures as the fitness criterion during the in-sample selection process, but the new trading rules are not more successful. Nevertheless, the Cumby and Modest (1987) test of market timing provide significant evidence of predictability. Similar

---

[9]Break-even transaction cost is the level of transaction cost, which includes the bid ask spread and brokers commissions, that makes a strategy mean return equal to zero or to the buy-and-hold return. Thus, if the costs faced by an investor are lower than them, the strategy is profitable. As we use this measure later on, more details about the computation are given in Section 4

results are found by Goodacre and Kohn-Speyer (2001) who show that the excess returns of the CRISMA system are only due to risk bearing during the 1986-1996 period on individual US stocks. Indeed, raw returns are significant and range from 6.2% to 17.6% on an annual basis, depending on the level of transaction costs. However, they conduct a performance analysis inspired by event studies with several returns generating models, and they find that the performance is not significant anymore. In addition, it is even negative with high transaction costs, such as 1% or 2% round trip. These results contrast with Pruitt and White (1988), so, they argue that the trading system performance is highly sensitive to the choice of the parameters.

Predictability occurs when the fraction of right signals, i.e. the percentage of positive (negative) market returns after buy (sell) signals, is higher than 50%[10]. In this case, a trading rule is able to identify market trends. Mills (1997) reaches this conclusion; however, the strategies based upon these signals are not profitable over the last subsample considered (1975-1994). Testing the CRISMA on UK individual stocks, Goodacre, Bosher and Dove (1999) find that the proportion of profitable trades given a buy signal (the strategy takes only long positions) is between 51% and 60% according to various risk-adjusted models. Nonetheless, transaction costs eliminate the profits. They also show that the system performs better for stocks with a high market capitalisation. This is rather counter-intuitive, as the market for these stocks should be more "efficient" than small stocks. However, they provide evidence in favour of technical analysis predictability as these large stocks returns are less likely to be affected by microstructure issues, which could induce spurious autocorrelation in returns. They also examine an option strategy similar to Pruitt and White (1989). They use traded options on the LIFFE, but they are able to identify only 64 signals out of 176 that match a traded at-the-money call option with available data. Nevertheless, the results are promising as they obtained an average return of 27.7% (10.2%) without (with a high level of) transaction costs for a holding period of 40 days in average.

Another issue about the lack of consistent technical analysis profitability resides in its behaviour over various subsamples. Indeed, when the market follows a strong upward trend, yielding higher returns than the buy-and-hold strategy may be nearly impossible. For example,

---

[10] This definition may differ among authors. For instance we argue that a strategy that possesses predictability should have a percentage of right signals higher than those of the buy-and-hold, and not 50%.

Fong and Yong (2005) find that the double-or-out strategy[11] applied to individual US stocks related to Internet during the speculative "bubble" from 1998 to 2002 fails to produce significant profits. However, they find that the strategies are able to outperform the buy-and-hold during the bear subsample. This mainly results from being outside the market most of the time (65% of the days). Further evidence is provided by Potvin, Soriano and Vallée (2004). They apply a genetic programming approach to individual Canadian stocks from 1992 to 2000, and find that the trading rules do not generate profits in excess of the buy-and-hold on average. However they are profitable when the market prices drops, or when they do not follow a trend. The technical trading rules inability to produce significant excess returns during a bull market is also reported by Fernández-Rodríguez, González-Martel and Sosvilla-Rivero (2000) on the Spanish market. Even if the technical trading strategies do not outperform during a strong upward trend, they may be valuable to avoid extreme losses during bear periods.

Some studies find significant profits only for a fraction of the assets on which trading rules are simulated. Among them, Cheng, Cheung and Yung (2003) show that the CRISMA trading system applied to 37 national equity indices does not yield excess return on average. They also point out the arbitrary choice of the system parameters, and show that its performance could be improved by reducing the length of the moving average window. They report that the system generates profits for Hong Kong individual stocks with a high turnover. Cesari and Cremonini (2003) compare the performance of various dynamic asset allocation strategies, such as technical analysis, the constant mix, the constant proportion and option based strategies. They perform historical and Monte-Carlo simulations, and find that the optimal strategy depends on the market state. In addition, the technical strategies are optimal, in term of risk-adjusted performance, only for the Pacific market. For the other markets[12], the constant mix or the constant proportion strategies systematically provide higher risk-adjusted returns than the technical strategies.

---

[11] The double-or-out strategy takes a position that is twice larger than the capital after a buy signal and invests in the risk-free asset otherwise. It never goes short.

[12] They are; North America, Europe and the world market.

Brown, Goetzmann and Kumar (1998) investigate the profitability of the Dow Theory[13], which is based on the Hamilton's editorials published in the Wall Street Journal during 27 years from 1902 to 1929. Contrarily to Cowles (1933), who does not adjust returns for risk, they find that the strategy is profitable with risk-adjusted performance measures. Furthermore, theses findings are confirmed by two bootstrap tests, the first is the standard one based on resampling returns (using the random walk model), and the second consists in bootstrapping in the strategy trading signals space. In addition, they use an event-study framework to assess whether the Dow Theory buy (sell) signals are followed by positive (negative) market moves. They use an 81 days event window[14] and find evidence in favour of predictability. They also infer the Dow Theory predictions with a neural network model. They use the 29 years interval as the training period and the following years (1930-1997) as an out-of-sample test interval. They analyse this strategy with and without a one-day lag between the signal emission and the day the trading position is taken. They find that when including the lag, the strategy outperforms the benchmark only during the bear market that occurred during the 70s. Thus, the theory can not be seen as a threat to the EMH anymore.

In most of the studies presented in this Section, technical indicators possess some forecasting abilities; however, transaction costs or risk adjustments eliminate the profits. A summary of these studies is provided in Appendix I-D.

2.3   Evolution of profits across geographical areas and over time

2.3.1   The market influence

The performance of technical trading rules may vary across different markets and over time as the degree of efficiency of these markets is not constant.

Bessembinder and Chan (1995) compare the technical trading performance in emerging and mature Asian markets. They find that it is profitable only for the informationally inefficient

---

[13] According to them, The Dow Theory has been created by Charles Henry Dow, founder and editor of the Wall Street Journal. However most of what we know come from his successor, William Peter Hamilton. The latter wrote 255 editorials containing forecasts of the US stock market. This theory is not explicit, but it is based on the hypothesis that markets follow trends and examining past fluctuations, in order to detect recurrent patterns, allows identifying them.

[14] 40 days before the editorial publication in the Wall Street Journal, 40 days after and the publication date.

emerging markets. They are characterized by a poor diversification, as they are dominated by a few large companies, high ownership concentration, financial disclosures are less stringent and finally, by a low transaction volume. All these characteristics lead to substantial deviations from the random walk behaviour of stocks prices, and thus, technical analysis can be useful to exploit them to generate abnormal returns. Tian, Wan and Guo (2002) compare the profitability of trading rules on the Chinese market and the DJIA, and they reach similar conclusions. While the set of trading rules (MA and TRB) applied to the DJIA produces, on average, slightly negative break even transaction costs (-0.06%), they amount to 2.22% for the Chinese market (both the Shenzhen and Shanghai exchanges) over the period ranging from 1991 to 2000.

These findings are in line with Fifield, Power and Sinclair (2005) on eleven European stock markets. They compare filter and MA rules applied to seven developed markets (Finland, France, Germany, Ireland, Italy, Spain and the UK) and four emerging markets (Greece, Hungary, Portugal and Turkey) over the period 1990-2000. Only a few strategies simulated on the small, developed markets generate excess returns, while none of them outperformed the buy-and-hold for the three largest developed markets. On the other hand, more than half of the filter rules and some MA rules applied to the emerging markets yield excess returns, even after considering transaction costs. In addition, a few of the filter rules generate statistically significant differences in returns. These results suggest that technical trading strategies are, nowadays, only profitable on less efficient markets.

Hsu and Kuan (2005) argue that this distinction is also valid for different markets in the same country. Indeed, their results indicate that less mature US indices, such as the NASDAQ Composite and the Russell 2000, provide opportunities for technical trading rule to generate profits. Their strategies are successful even after transaction costs and when considering the data-snooping bias with simulation based tests. However, these results in favour of technical analysis are not shared with more mature indices, such as the DJIA and the S&P 500.

These studies support that technical analysis is a useful forecasting method for inefficient markets. They are summarized in Appendix I-E. Another widely addressed issue is the temporal evolution of technical analysis profits.

2.3.2   Profit evolution through time

Following the BLL methodology for the UK market, Mills (1997) shows that the profitability of simple technical rules declines sharply over time. They are even not profitable anymore during the last subsample (1975-1994). This is in line with Wong, Manzur and Chew (2003). They find that profits from MA rules on the main Singaporean index are significant only at the 10% level

over the last subsample (1988-1995), while this level lies between 1% and 5% over earlier periods. Bessembinder and Chan (1998) provide an illustration of decreasing profits over a long DJIA sample, ranging from 1926 to 1991. The BLL set of trading rules yields statistically significant profit over the entire sample, on average, but not during the last subsample under investigation (1976-1991). In addition, break-even transaction costs decline over time, from 0.44% during the first subsample (1926-1943) to 0.11% for the last.

LeBaron (2000) examines a single MA rule applied to the DJIA index over a century, from 1897 to 1999. This sample encompasses the one used by BLL, but it is also ten years longer. This allows him to test whether the trading rules still generate excess returns over this most recent period. He finds a dramatic change in the trading rule performance. Indeed, the difference between conditional buy and sell mean returns is not statistically significant anymore over these 10 years, and in addition, it is even negative. Nevertheless, he can not claim if this is due to a change in the price process during last ten years, or whether the evidence in favour of technical analysis over the entire century results from data-snooping. Arguments in favour of the first hypothesis are provided by Sullivan, Timmermann and White (1999), as they show that BLL results are not due to data-snooping. Olson (2004) examines the evolution of technical trading strategies, based on a process to select the parameters, in the foreign exchanges[15] over the period 1971-2000. He divides this sample in six subsamples of five years. First, he finds that annual excess returns decrease from 3.34% over the period 1976-1980 to -0.12% during the last period (1996-2000). Second, he performs a regression analysis to confirm this finding statistically. He regresses the trading strategies excess returns on a temporal variable and an intercept. For 17 currencies out of 18, the regression coefficient is negative and even significant for 5 currencies. He concludes that foreign exchange market inefficiencies in the seventies and eighties have been corrected, and thus, technical analysis is not profitable anymore. On the contrary, Dueker and Neely (2007) find that complex rules, combining predictions from a Markov switching model and MA rules, are still profitable during the last out-of-sample period considered (from 2002 to 2005). However, standard MA rules alone are not successful during this recent subsample.

Kidd and Brorsen (2004) investigate why the profitability of technical trading strategies decreases over time, especially after 1990. They argue that a structural change in various futures

---

[15] We decided to include this study, and the following one, on foreign exchanges in order to show that this temporal evolution of technical trading profits is similar to stock markets.

markets[16] occurred when the profits began to decrease. They show that volatility of futures prices and the volatility of the close-to-open changes have decreased, and therefore, there are fewer opportunities for technical trading profits. The decline in daily autocorrelations observed between the periods 1975-1990 and 1991-2001 is consistent with lower technical trading profits, as most of them are trend-following techniques. They conclude that the markets are more efficient, and they incorporate new information faster than in the past. Similarly, Day and Wang (2002) argue that the decrease in the daily first-order autocorrelation coefficient of the price-weighted DJIA is due to a significant increase in trading volume during the eighties. Thus, the autocorrelation of stocks indices, which has been exploited by the technical trading rules during earlier periods, was due to the non-synchronous trading bias. For example, the coefficient declines from 0.27 over the 1967-1972 sample to 0.05 from 1982 to 1986.

Some technical analysis strategies are based on the hypothesis that prices patterns repeat themselves over time. Thus, Summers, Griffiths and Hudson (2004) investigate another issue related to the temporal evolution of the profitability. They test if trading rules, based on neural networks, developed at an early period keep their predictive ability during a later interval. They focus on the FT 30 index from 1935 to 1994. First, they confirm that technical profits are not statistically significant anymore over the last subsample (1979-1994). However, when they apply the neural network model generated during the first subsample (1936-1950) to the last one, the profits are economically and statistically significant. They conclude that: "they are unchanging factors over time, but also [support is found] for the presence of factors which confound modelling in more recent periods."[17] They argue that the volatility increases over the most recent period and this results in more noisy time series, and therefore, the detection of prices patterns is more difficult.

Most of these studies indicate that profits of technical analysis have decreased since the nineties to an extent that even institutional traders with low transaction costs can not generate significant excess returns anymore. Appendix I-F presents a summary.

---

[16] They include a wide range of futures in the analysis, such as commodities, currencies, stock market indices or treasury bonds.

[17] Summer, Griffiths and Hudson (2004), page 213.

2.4   Why technical analysis may have forecasting power?

Scepticism about technical analysis is usually based on its lack of theoretical basis. Thus, in this Section, we detail some elements supporting that technical analysis possesses forecasting abilities and why.

2.4.1   Market microstructure and uninformed traders

The first explanation is related to the market microstructure, and specially the ability to recover information from orders flows. Kavajecz and Odders-White (2004) investigate whether stock prices predictability depends on its liquidity and the limit order book. The assumption is that peaks of liquidity in the limit order book on the sell (buy) side is the first important barrier for further price increases (decreases). Thus, they may be related to technical resistance (support) levels. They find evidence supporting this hypothesis, as they report striking similarities between support and resistance levels and various measures of order book liquidity. They confirm this finding with a regression between the limit order book measures and the technical support and resistance levels. They find a positive and significant relationship, even after including various control variables, such as market trends or trading volume. The relation is stronger on the sell side of the order book associated with the resistance levels. Finally, they perform a Granger causality test, and find that is it more likely that order book measures influence support and resistance levels than the opposite. They conclude that technical analysis can discover peaks of liquidity of the limit order book and predict futures prices. Osler (2003) analyses whether the clustering of stop-loss orders and take-profit orders[18] can explain two common predictions of technical analysis. The first is that trends are likely to reverse at levels that can be predicted ex-ante. She shows that take-profits orders cluster more strongly than stop-loss orders at prices ending in 00, and thus, this may induce a trend reversal. The second prediction is that trends accelerate after crossing a support or resistance level. This may be explained by the concentration of large stop-loss buy (sell) orders just above (below) round numbers. She shows that stop-loss sell orders cluster more strongly at these levels than take-profits buy orders (which would induce a trend reversal). Thus, the order flow is dominated by sell orders that accelerate the downward trend and consequently, this provides evidence in favour of using the trading break range rule rationally. The reasons advanced for this order clustering are the information efficiency and the irrational preference for round numbers.

---

[18] The stop-loss orders may be in line with trend following strategies, while take-profit orders imply trend reversals.

Reitz (2006) proposes a model where asset prices are driven by fundamental information, which is revealed to market participants with a lag. As this information is not directly observable, technical analysis may be useful to identify regime shifts. Within a Markov switching model setting, he shows that a MA rule can recover hidden fundamental information, and thus, it is useful for predicting exchange rates when they are subject to regime shifts. He concludes that technical analysis is a "cheap proxy of Bayesian learning"[19].

These papers argue that technical analysis is a suitable method for uninformed traders to obtained information, which would be known only by some informed market participants, such as market makers for instance. They support the fact that markets are not fully efficient.

2.4.2   Nonlinearities in the price process

The link between nonlinearities in the price process and technical trading rules may be more or less obvious according to the type of strategies. Indeed, charting try to identify nonlinear patterns, such as head-and-shoulders, in past prices. On the other hand, this link is less obvious for mathematical rules. The following studies provide evidence supporting the nonlinear explanation of technical analysis forecasting power.

Neftci (1991) argues that technical trading rules may be useful for prediction if and only if prices follow a nonlinear process. Indeed, if the price process is linear, vector auto-regressions should provide the best prediction according to the mean squared error criterion. To determine whether MA rules have forecasting power beyond an autoregressive process, he performs a regression between the prices and their lags and dummy variables representing the positions taken by the rule[20]. As the beta coefficients associated with the dummy variables are jointly significantly different from zero, he concludes that the trading rule can predict the price evolution beyond its linear forecast. Similar results are found by Gençay (1996). Moreover, he also points out the positive impact of using a 1% band on the predictive power of a few MA rules to avoid noisy signals.

Gençay (1999) shows that nonparametric models, such as nearest neighbours and feedforward network regressions[21], provide better exchange rate forecasts than the random walk and the

---

[19] Reitz (2006), page 135.

[20] The position taken by the rules are lagged as well to include long-term trends.

[21] These models take nonlinearities into account.

GARCH (1, 1) models. Kwan, Lam, So and Yu (2000) examine the price trend model proposed by Taylor (1980), which supposes that prices are made up of an unobservable trend and a noise process. As the trend follows a Bernoulli process, forecasts based on an ARMA model may not be appropriate because the price model is nonlinear by construction. They find that trading signals issued from the best linear one-step-ahead forecast are a weighted average of past log prices, and thus, they are very similar to a MA rule with a short window of one day. Clyde and Osler (1997) conduct a simulation analysis to examine whether charting predictability arises from nonlinearities in the price process. They find evidence supporting this hypothesis. Indeed, the profits from head-and-shoulders indicators are significantly higher when they are applied to simulated nonlinear data, compared with random data. They also demonstrate that technical trading rules produce higher returns than random positions when the price follows a nonlinear process. They conclude that "Technical methods may generally be crude but useful methods for doing nonlinear analysis"[22]. Nam, Washer and Chu (2005) argue that the nonlinearity of the returns process is due to an asymmetric reverting property, i.e. positive returns tend to persist longer than negative returns.

Dewachter (2001) further investigates the ability of technical trading rules to exploit nonlinearities in the foreign exchange market. For this purpose, he uses a Markov switching model. It belongs to the class of price-trend models, and thus, past observations contain useful information. To examine whether MA rules exploit nonlinearities, he compares the performance of the Markov model and its linear projection, i.e. an ARMA model. The ARMA model left unexplained between 15% and 71% of the MA rules profits, while this percentage is lower for the Markov model prediction, between 2% and 60%. To investigate further this issue, he conducts a Monte-Carlo analysis. He shows that MA rules applied to the series simulated with the Markov model can replicate the original profit for all currencies at traditional confidence level. On the other hand, this is the case only for two out of four currencies when series are simulated with the linear projection. Finally, he concludes that MA rules take advantages of nonlinearities in the exchange rates to generate better predictions than linear models. The link between a regime-switching model and MA rules is also confirmed by Dueker and Neely (2007). They also shed light on the fact that "Markov rules" rely rather on long-term trends, while MA rules focus on short-term trends. Consequently, trading rules combining both methods generate higher profits than the individual rules. Further evidence is provided by Dewachter and Lyrio (2005), who point

---

[22] Clyde and Osler (1997), page 511.

out a nonlinear relation not only between a MA rule signals and foreign exchange returns, but also with the variance.

These studies show that technical trading rules may generate more accurate forecasts than linear forecasting methods, as they may take advantage from nonlinearities in the price process. These findings should be taken as evidence against the EMH. However, to challenge it, trading strategies using technical rules should provide excess return after accounting for transaction costs and risk adjustments. Nonetheless, these findings support the usefulness of technical analysis as a valuable forecasting method.

2.5   A caveat about markets microstructure and data-snooping

The last part of this literature review focuses on two issues that may question the usefulness of technical analysis strategies when they are implemented in a real context.

2.5.1   Microstructure issues

Beside transaction costs that may reduce or even eliminate the profitability of technical trading strategies, Day and Wang (2002) argue that microstructure is another challenge when implementing them. Indeed, they point out that correcting the index series used by BLL for non-synchronous prices make the difference in returns between the buy-and-hold and the strategies not significant anymore. This is due to the assumption that trading positions can be executed at the closing price used to generate the signal. However, they point out that trading volume on DJIA stocks was low during the time interval under consideration by BLL, and thus, some of the prices adjustments had to occur during the next day opening. As buy (sell) signals are likely to be generated when prices are moving up (down), the non-corrected closing price is understated (overstated) and this creates a positive bias. Among others, Scholes and Williams (1977) and Cohen, Hawawini, Maier, Schwartz and Whitcomb (1980) show that non-synchronous prices introduce spurious auto-correlation in individual securities and in stock indices. Thus, if the simulated trading strategy uses this auto-correlation to generate abnormal profits, they will disappear when the strategy is implemented in a real context. However, we may argue that the increase in trading volume, more efficient electronic markets and the possibility to trade directly in market indices, with futures contracts or trackers, minimizes this bias for more recent studies. For instance, Chordia, Roll and Subrahmanyam (2008) find that an increase in market liquidity leads to lower auto-correlations and more efficient markets.

2.5.2   Data-snooping

Finally, we review the data-snooping bias that arises from a repetitive use of the same data sets and techniques. Indeed, spurious patterns are more likely to occur when a great deal of researches are performed on the same data set. In this case, a data-snooping affected strategy may yield excess returns on past prices series, but not with future prices that will not share the same pattern. This is a serious problem in financial research as stocks or equity index series are among the most analysed series in economics. For example, Lo and MacKinlay (1990) document this bias for empirical tests of asset pricing models.

The risk that the data-snooping bias affects empirical studies on technical analysis is qualified by BLL to be "immense"[23]. Indeed, the rules that appear to be profitable in the past are more likely to be highlighted in future research than less successful trading rules. Thus, technical analysis profitability reported in some studies may result from picking some "lucky" rules out of a universe that has no forecasting power on average. BLL try to minimize this effect by providing results for all tested trading rules, and not only for the most successful, by analysing a very long time series (the DJIA over 90 years) and by performing a robustness analysis with subsamples. Another straightforward solution to control for this bias, at least to some extent, is to use complex trading rules that rely on an out-of-sample process to select the parameters objectively.

Nevertheless, to resolve properly the data-snooping problem, a testing procedure should take into account the dependencies between results from several rules. For this purpose, White (2000) develops the Reality Check (RC) test that determines whether the best model drawn from a wide specifications universe has forecasting abilities. This joint test is based on a bootstrap approach to compute empirical *p*-values. The procedure is the following:

1. The null hypothesis, *H0*, states that there is no trading rule with a superior predictability in the universe of rules under consideration.

2. Choose the benchmark strategy; it is either the buy-and-hold return or zero.

3. Compute the performance in excess of the benchmark of each trading rule.

4. Select the rule with the highest mean excess return and normalize it by multiplying it by the square root of the number of periods. This is the test statistic.

---

[23] Brock, Lakonishok and LeBaron (1992), page 1736.

5. Get the empirical distribution of the test statistic with the following stationary bootstrap procedure: Create artificial price series by resampling the original one, and then apply each trading rule to these new series to generate artificial strategies returns and calculate the excess mean return.

6. Repeat the last step B times. White suggests that 500 times is enough. Each time, select the strategy that generates the largest positive difference between the excess mean return obtained with the simulated prices series and the original one. Compute the difference in excess mean returns and multiply it by the square root of the number of periods. These B statistics represents the empirical distribution under the null hypothesis of the test statistic defined in point 4.

7. Compare the test statistic with its empirical distribution and compute the p-value. For example, if 30 out of 500 statistics obtained in point 6 are higher than the test statistic, the p-value is $30/500 = 6\%$. The lower the p-value, the less likely it is that there is no model with predictive power, as one of them generates higher returns on the original prices series than on a multitude of artificial ones.

8. Determine whether the null hypothesis is rejected at the desired level of confidence. This is the case when this level is higher than the p-value.

He provides an illustration of this test by computing the p-values step-by-step, adding one trading rule after another. If an additional rule added to the universe does not yield a higher performance of the previous best rule, the p-value associated with the null hypothesis that the best rule does not outperform the benchmark increases. This is due to the fact that the best rule is chosen from a larger universe of rules, and therefore, it is more likely that its superior predictive ability is affected by the data-snooping bias. Sullivan, Timmermann and White (1999) investigate if the results obtained by BLL are due to this bias, and whether the rules are still profitable over a more recent subsample. They test an extended set of nearly 8'000 trading strategies, and conduct tests with the RC test either with the entire set of rules or only on the best performing rule, i.e. without considering data-snooping. First, they show that the BLL results are not due to data-snooping, as both p-values are equal to zero. This contrasts with the most recent period from 1987 to 1996. Indeed, not adjusting for data-snooping would lead to misleading conclusions, as the single p-value is 0.004 (0.055) for the whole (BLL) set of rules compared with a RC p-value of 0.341 (0.154). Qi and Wu (2006) and Marshall, Cahan and Cahan (2008) find similar results, respectively on the foreign exchange market and an US exchange-traded fund replicating the Standard & Poor's 500 with intraday data.

Hansen (2005) points out that this test is conservative as the null distribution is based on the least favourable configuration. A related issue is the fact that the test power drops as more unsuccessful forecasting models are included in the universe. Thus, he proposes a more powerful extension of the RC test, the superior predictive ability (SPA) test. Hsu and Kuan (2005) examine a wide universe of 39'832 rules[24] applied to two mature markets, the DJIA and the S&P 500 and to two more recent indices, the NASDAQ Composite and the Russell 2000. Both RC and SPA tests are computed to investigate how the data-snooping bias influences the trading rules profitability. During the in-sample period, from 1990 to 2000, the profits are statistically significant only for the two more recent indices. Moreover, these profits are not due to data-snooping, as their p-values are equal to zero, and the profits are also economically significant even after considering transaction costs. The years 2001 and 2002 are reserved for an out-of-sample analysis, as they test the best rules determined during the in-sample period. The findings are similar, and thus, they conclude that these more recent indices have not reached the weak form of efficiency yet. They also illustrate the fact that the SPA test is more powerful than the RC test, as SPA p-values are consistently smaller.

Finally, Romano and Wolf (2005) and Hsu, Hsu and Kuan (2010) extend the RC and SPA test with a stepwise multiple testing approach. These new tests are able to identify all models that outperform the buy-and-hold, while the RC and SPA test determine only whether at least one specification is successful. Hsu, Hsu and Kuan (2010) apply the new stepwise SPA to US growth oriented and emerging markets indices, but also to the corresponding Exchange Traded Funds (ETF). For each index, they also separate the sample into a pre and post-ETF period. They consider a set of 16'380 trading strategies based on MA and filter rules. Regarding the US indices, they find some successful specifications during the pre-ETF sample, while none of them generate an abnormal performance during the post-ETF period. The findings are similar for the emerging markets. They argue that ETF improve the markets liquidity and efficiency, and thus, there are fewer opportunities for trading rules to generate profits. They also point out that trading rules profitability is not necessary linked with autocorrelations. Indeed, the stepwise SPA test is able to find successful rules on indices that are not serially correlated.

The only definitive conclusion that we can draw from this literature review is that the jury is still out on the usefulness of technical analysis. Nonetheless, the following aspects stand out:

---

[24] They consider commonly used rules, such as MA and filter rules, but also less popular strategies such as graphic rules, contrarian strategies and complex rules.

1. Most of the studies find that technical analysis possesses some forecasting abilities; however it is less likely that they are sufficient to generate profitable investment strategies when transaction costs and risk are included.

2. Its profitability seems to decrease from the nineties, and it may be present on the less efficient markets only.

3. Complex trading rules show a better performance than the simple indicators, which are used to construct the complex systems.

4. The data-snooping bias is an important issue and should not be ignored. Complex trading system and simulation based test can be use to control or mitigate this bias.

3.   The trading system

The first part of this Section focuses on the simple MA rules description, while the second one presents how we combine them to construct the complex trading systems.

3.1   The simple MA rules

Each trading system in this study relies on simple MA rules, which compare two MA computed with a different horizon. Their purpose is to smooth prices series and their comparison is supposed to detect trends in the market. An upward (downward) trend occurs when the short MA arises above (slides below) the long MA. A security band, B, may also be added to avoid non informative or noisy signals when the difference between the two MA is small. It is usually set to 1%. Hence, three parameters, which must be chosen, characterize a MA rule; the length of the short and the long MA, respectively S and L, and the bandwidth, B. They are computed as

$$M_{t,S} = \frac{1}{S}\sum_{j=1}^{S} P_{t-S+j} \quad \text{and} \quad M_{t,L} = \frac{1}{L}\sum_{j=1}^{L} P_{t-L+j}, \tag{1.1}$$

where P is the asset price and their relative difference is

$$D_t = \frac{M_{t,S} - M_{t,L}}{M_{t,L}}. \tag{1.2}$$

A buy signal is generated whether $D_t > B$ and a sell signal if $D_t < -B$. When $B \geq D_t \geq -B$, the two MA are close, resulting in a neutral signal. The trading rule consists in taking a long (short)

position in the market after a buy (sell) signal, and investing in the risk free rate after a neutral signal. In line with the related studies, this implies that there is no lag between the signal computation and the transaction on the market as we use the same closing price. It keeps the same position as long as the trading signal does not change.

The MA rule is one of the most commonly used technical indicator, both by academics and professional. Nonetheless, the novelty of our approach lies in the range of parameters that we consider. In contrast with other studies, we compute long MA with more than one year of data, which is the standard approach. Our universe of simple MA rules includes 1876 combinations of parameters: 23 different lengths[25] for the short MA, with values ranging from one day to 100 days and 48 lengths[26] between five and 990 days for the long MA. We also consider these combinations with and without a security band of 1%. Figure 1 illustrates the MA rule with a long (short) MA of 500 (20) days and the corresponding positions in the market.

Figure 1: MA rule

The long (short) MA is computed with 500 (20) days. The horizontal line at the bottom of the graph indicates whether the rule takes a long (short) position in blue (red).

It suggests that the rule exploits the long-term trends, as the trading positions seem to be in line with bull and bear market phases. However, there is a significant lag between the trend

---

[25] From one to 10 days with an incremental step of 2 days, and then from 15 to 100 days with a step of 5 days.

[26] From 5 to 50 days with an incremental step of 5 days, and then from 65 to 990 days with a step of 25 days.

reversal and the time at which the rule changes the position. Nonetheless, we argue that this lag does not prevent from exploiting the trends, as they last longer.

3.2   The complex trading rules

To determine the profitability of technical analysis and whether it can challenge the EMH, we focus on complex trading rules inspired by those introduced in Hsu and Kuan (2005). In addition to using more information than a simple rule, they are designed to remove the subjectivity in the choice of parameters, and thus, they should be less affected by data mining issues. Indeed, they are based on a selection and a test sample. If the in-sample performance is due to this bias, there is no reason that the selected specification performs well during the test sample. Thus, we argue that if a complex system outperforms the benchmark, this is not due to the data-snooping bias but to the rule predictive power. Our entire time interval spans over 19 years, from January 1990 to December 2008. The four complex rules are described below:

Opt_all. The first rule contains a continuous selection process. Each day, we compute the 1876 simple rules cumulative return over the entire history, and then, we select the best performing rule to generate the effective trading signal. This procedure is repeated every day, which means that the test sample extends over time. Thus, the last selection interval covers the period from 1990 to 2008. We leave four years of data as the first test sample, and the effective strategy starts in 1994. It is identical to the learning strategy in Hsu and Kuan (2005).

The three other complex rules are based on a selection sample and a test sample with constant length. For each of these samples, it is set to four years in order to match the longest MA length. The first four years of data, from January 1990 to December 1993, constitute the first selection sample. Each complex rule uses the information in a different manner, as explained below, and then the rule consists in investing over the next four years (i.e. the test sample). Then, this four years sample, from January 1994 to December 1997, is used as the new selection sample, and then, the trading position are taken over the next four years. This process is repeated until the end of the year 2008 and thus, we have four test periods lasting four years each.

Opt_4. We propose this new rule to address some issues that may affect the first rule. Indeed, the later may suffer from over specification, as a change in the parameters should result from a modification in the trend. Suppose for instance that an event occurs, such as a flash crash, without modifying the current trend. The Opt_all rule will be affected for a certain amount of time. Instead of computing cumulated returns over the entire history, the Opt_4 rule records simple MA rule returns only over the selection period. Moreover, the parameters are not revised

every day, but the selected rule is evaluated over the entire test sample. In other words, the best performing rule is identified during the selection sample, and it is run to produce the effective trading signals over the test sample. These two first rules are similar in the sense that they rank the simple MA rules according to their past performance in order to select a single specification only. In Hsu and Kuan (2005), they only consider selection and evaluation periods up to one year.

The last two complex rules are different in their design, as they combine several rules signals to generate the effective signal. In contrast with Hsu and Kuan (2005), we do not use the whole set of simple rules, but we consider only a subsample. The first step consists in identifying rules that have a return higher than or equal to the market over the selection sample. Then, these rules are used over the next test sample to generate the effective trading signals according to two procedures. The idea is to eliminate the non-performing specifications, and thus, to reduce the impact of the initial universe of rules composition, which has to be set arbitrary.

Voting. The third rule counts, each trading day, the number of buy and sell signals suggested by the rules selected during the previous evaluation sample. The trading signal corresponds to the position that has received the highest number of "votes". For example, suppose that we identify 825 MA rules with higher returns than the buy-and-hold during the selection sample. The first day of the test sample, we generate trading signals with each of these 825 rules. For instance, there are 500 buys and 325 sells, so the rule takes a long position on the next day. This process is repeated each day during the four years test sample with the same 825 rules.

Partial. The last rule averages the signals (-1 for sell, 0 for a neutral and 1 for a buy) emitted by the selected rules, and thus, produces a fractional position. According to the same example as the one presented above, the Partial rule will take a long position in the market corresponding to 21%, i.e. (500-325)/825, of the available capital over the next day. The invested amount varies according to the confidence in the forecast. Indeed, if there are only a few more rules that produce a buy signal than a short one, only a small percentage of the capital is invested. This is consistent with the findings of Blanchet-Scalliet, Diop, Gibson, Talay and Tanré (2007) and Zhu and Zhou (2009).

We highlight the fact that each of these four complex trading rules follows an entirely out-of-sample process, and thus, is not subject to any look ahead bias.

Beside the buy-and-hold strategy, we also compare the profitability of complex rules with two other benchmarks. The first is the random walk strategy that takes a long position at t when the index return is positive at $t-1$ and a short position otherwise. Finally, we also report the best

performing rule, which is named Best, over the entire sample from 1994 to 2008. Obviously, it can not be used to assess whether technical trading has forecasting power, as its performance is in-sample and is strongly affected by the data-snooping bias. However, it indicates whether our complex trading systems are able to approach its performance, or even surpass it.

4.   The investment strategies: Description and empirical results

In this Section, we describe and implement several investment strategies, based on the four trading systems presented above, with three investment approaches; the first one consists in the standard investment strategy that invests (shorts) 100% of the capital after a buy (sell) signal. The following two differ as they use financial leverage to improve the performance either with exchange-traded options or debt. We consider leverage for the two following reasons. First, technical analysis is often used by alternative asset managers, such as hedge funds, who are likely to use leverage. Second, it may help to identify successful strategies when the market follows a strong upward trend. Indeed, these strategies will be long most of the time, and thus, will not be different from the buy-and-hold return without leverage. However, as this chapter marks the beginning of the empirical part of this thesis, we briefly present the notation linked to returns and to the main performance measures. They are similar for the three Parts of this thesis.

4.1   Statistical notations

4.1.1   Returns characteristics

There are two ways to compute the return yielded by holding an asset during a time interval, the simple and the continuously compounded return. The difference lies in the number of times the return is computed. They are calculated as;

$$R_{simple,t} = \frac{P_t - P_{t-1}}{P_{t-1}} \qquad R_{continuous,t} = \ln(P_t) - \ln(P_{t-1}), \tag{1.3}$$

where P_t is the asset price at time t. While the simple return is the difference in prices during the time interval divided by the first price, the continuously compounded return assumes that this interval of time tends toward zero. In order to have a consistent method to compute returns across this thesis, we use simple returns. Indeed, as we use options in the investment strategies as well, we cannot use continuously compounded returns as the option price may be zero and thus,

the return would be minus infinity. We also present the cumulated compounded returns to present an investor point of view. The series are calculated with the simple returns as;

$$R_{compounded,t} = \left(\left(1 + R_{compounded,t-1}\right)\cdot\left(1 + R_{simple,t}\right)\right) - 1, \tag{1.4}$$

where the first compounded return is the first simple return. They represent the cumulative performance of the investment strategy.

The formulas to obtain mean returns are different, as the simple mean return is calculated with the arithmetic mean while the compounded mean return is calculated with the geometric mean;

$$\overline{R}_{simple} = \frac{1}{n}\sum_{t=1}^{n} R_{simple,t} \qquad \overline{R}_{compounded} = \sqrt[n]{\prod_{t=1}^{n}(1 + R_{simple,t})} - 1, \tag{1.5}$$

where n is the number of periods. In order to have a more intuitive understanding of the various strategies' economical performance, we display annualized returns instead of daily returns and they are computed as;

$$\overline{R}_{simple,annual} = 252\cdot\overline{R}_{simple} \qquad \overline{R}_{compounded,annual} = \left(\left(1 + \overline{R}_{compounded}\right)^{252}\right) - 1. \tag{1.6}$$

Note that the simple annualized mean return is only an approximation.

Another characteristic of a return distribution that is usually displayed in investment strategies performance analysis is the volatility, σ. This is the most commonly used risk measure, as it describes the observations dispersion and is computed as the square root of the returns variance;

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(R_{simple,t} - \overline{R}_{simple}\right)^2}. \tag{1.7}$$

We display this measure for the entire strategy and for each of its constituents, i.e. the part after long position and short positions independently. The presented measures are also annualized, by multiplying the result of equation (1.7) by the square root of 252.

4.1.2   Main performance measures

Before turning to the main risk adjusted performance measure, we detail how we calculate the break-even transaction cost, which makes sense as not all investors (retail, institutional, financial institutions) face the same level of transaction costs. It represents the level of transaction cost

that eliminates the profit in excess of a benchmark strategy, such as the buy-and-hold, and an approximation[27] is obtained as;

$$\text{Break-even TC} = \frac{\left(\sum_{t=1}^{N} R_{strategy,t} - \sum_{t=1}^{N} R_{benchmark,t} \right)}{N_{trade}}, \qquad (1.8)$$

where N is the number of periods and N_{trade} is the number of one-way transaction, thus, buying and selling the asset is considered as two transactions. Then, if an investor can trade at a lower cost than this measure, the strategy is profitable. Note that it can be negative if the strategy yields a lower cumulated return than the benchmark, but in this case, it is meaningless.

Two of the most commonly used risk adjusted measures are the Jensen's alpha and the Sharpe ratio. They determine whether the performance is only due to more risk bearing or to genuine predictability. The first one uses the CAPM beta as a risk measure (the market model can also be used) and the second one the volatility. Again, we display the annual values, and they are calculated as;

$$R_{strat,t} - Rf_t = \alpha + \beta(R_{market,t} - Rf_t) + e_t$$
$$\alpha_{annual} = 252\alpha, \qquad (1.9)$$

$$Sharpe = \frac{\overline{R}_{strat,annual} - \overline{R}f}{\sigma_{strat,annual}}. \qquad (1.10)$$

where Rf_t is the risk free rate at time t. The regression in (1.9) is estimated with the ordinary least squares method (OLS).

4.2   The standard investment setting

As mentioned above, we first present the results obtained in the standard investment setting, in which the strategy invests (shorts) the whole capital after a buy (sell) signal without financial leverage. Of course, with the Partial trading system, only the percentage obtained as the trading signal is invested.

---

[27] This is only an approximation as this statistic has to be calculated by numerical optimization as it depends on the time structure of trades.

4.2.1   Data

We apply these four complex trading rules to the S&P 500 price index daily closing prices from January 1990 to December 2008. The first four years of data are kept as the initial estimation period, and consequently, we assess the trading strategies performance over the sample from 1994 to 2008, which contains 3761 daily observations. We choose this relatively short interval as earlier samples have been widely examined, and the recent literature suggests that technical trading profits decrease sharply over time, to an extent that most studies do not find profits since 2000. In addition, this sample corresponds to the period over which we have the exchange-traded options data used in Section 4.3. We use the one month Euro dollar deposit rate as the lending rate. The two series are extracted from *Thomson Reuters Datastream.*

We also simulate the trading rules with weekly observations. As the results are very similar to those in the daily frequency, we present a selection of the findings in Appendix I-G.

4.2.2   The performance of the simple MA rule

Before turning to complex trading rules, Figure 2 presents the annualized mean simple returns of all MA rules considered in this study.

Figure 2: Simple MA rules returns

Note: Each point on this graph represents the mean annualized returns of the simple MA rule with an SMA and an LMA equal to the respective value on the axes over the period 1994–2008. The vertical axis represents the mean annualized return of a strategy. The length of the SMA of the strategy is shown on the short MA axis and the length of the LMA of the strategy is shown on the long MA axis. The horizontal plane is the annualized mean return of the buy-and-hold, i.e. 6.1%
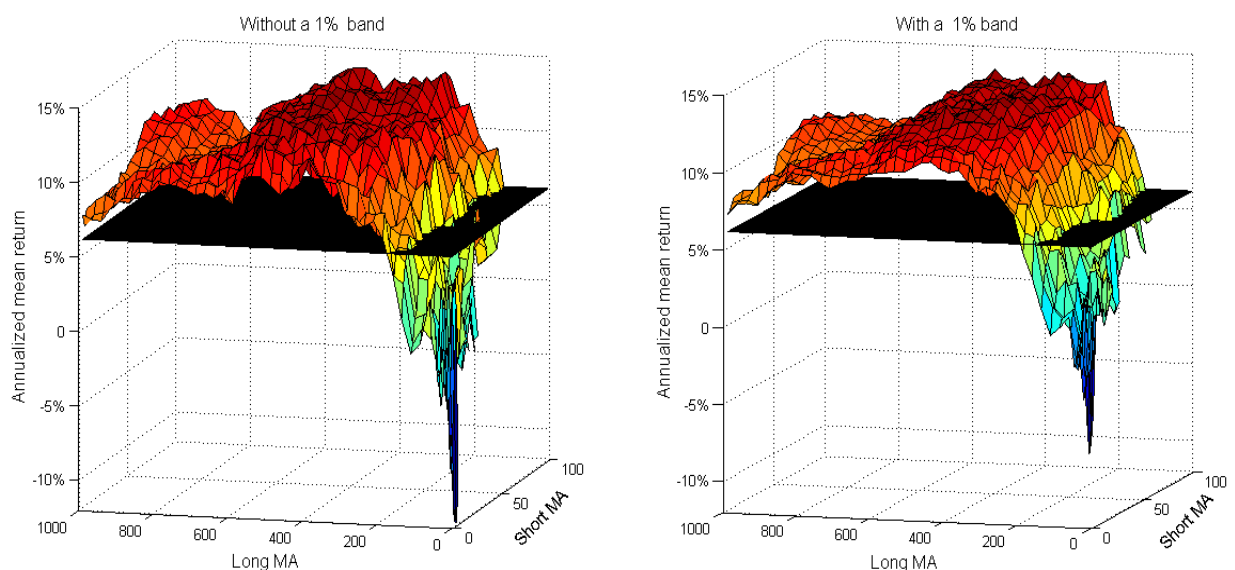
Indeed, the MA rule performance is required to construct the complex rules, and they are computed according to the standard investment setting. It shows that MA rules commonly used in academic studies, which rely on short term trends, perform poorly. Their returns are at best equal to the buy-and-hold and even often negative. On the other hand, the new rules that rely on longer trends yield a return twice as high as the benchmark. This indicates that these trading rules are able to detect and exploit long-term trends. Furthermore, these results ignore transaction costs, which would further reduce the performance of rules based on short-term trends, as they change trading positions more often. It is also worth noting that the trading rules performance is rather insensitive to small variations in parameters. Even if no test can be made, this may indicate that these results are probably not strongly affected by data-snooping issues. Indeed, we are not in a situation where only a very low percentage of rules generate higher returns. The negative returns associated with rules that rely on very short-term trends may indicate that they identify a trend too late, and thus, when they take a position, the trend reverses. The security band seems to affect only these rules, but not the ones computed over an extended number of observations.

Figure 3 displays the performance of the simple trading rules without the security band over four subsamples, including the first one that will be used as the selection sample in the complex trading rules setting. First, the rules performance is consistent in all subsamples, as traditionally used specifications generate a very poor performance. Second, these results contrast with the majority of recent studies that conclude that the technical analysis performance decreases sharply over time. Indeed, we find that the new proposed MA specifications outperform the benchmark during the most recent samples. As we will show in Section 6, the rules computed over long horizons invest according to market phases. As the two first subsamples are characterized by a strong upward trend in the market, it is impossible for these simple rules to outperform it without leverage. The excess profits are generated during bear markets.

Figure 3: Simple MA rules returns on four subsamples

Note: The points on these graphs represent the mean annualized returns of the different simple MA rules without band over different subperiods. The vertical axis represents the mean annualized return of a strategy, the length of the SMA and LMA of the strategy are respectively shown on the short and long MA axis. The horizontal planes represent the annualized mean return of the buy-and-hold strategy. Note that the "Z" axes do not have the same limits, as the mean returns differ considerably.
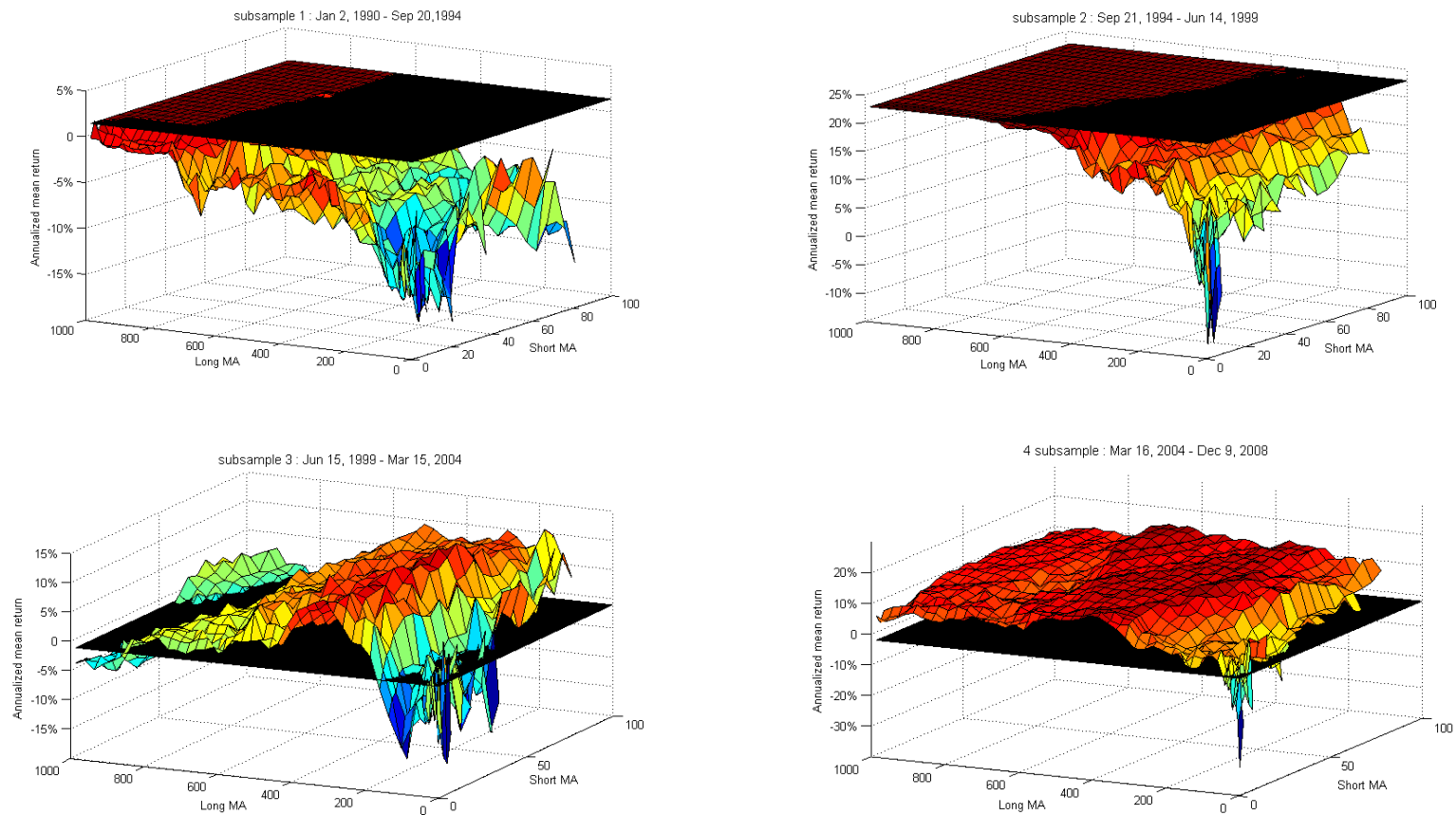
4.2.3   The complex trading system implemented in the standard investment setting

Table 1 presents the performance analysis for the four complex trading rules according to the standard investment approach and the three reference strategies in the first three rows.

First, we examine the percentages of right signals. These measures are not directly related to the performance, but they are useful to shed light on the trading rules forecasting ability. They are defined as

$$Right\_buy_j = \sum_{t=1}^{Nb_j} \boldsymbol{1}\{SB_{j,t}|R_{t+1}>0\ \}\frac{1}{Nb_j} \tag{1.11}$$

$$Right\_sell_j = \sum_{t=1}^{Ns_j} \boldsymbol{1}\{SS_{j,t}|R_{t+1}<0\ \}\frac{1}{Ns_j} \tag{1.12}$$

$$Right\_strat_j = \frac{1}{N}\left[\sum_{t=1}^{Nb_j} \boldsymbol{1}\{SB_{j,t}|R_{t+1}>0\ \}+\sum_{t=1}^{Ns_j} \boldsymbol{1}\{SS_{j,t}|R_{t+1}<0\ \}\right] \tag{1.13}$$

where Nb_j, Ns_j are the jth strategy number of buy and sell signals, $SB_{j,t}$ and $SS_{j,t}$ represent the buy or sell signals at time t, depending on the position taken by the strategy at this time. For the Partial rule, we consider a buy (sell) signal when the proportion invested in the index is higher (smaller) than zero. Thus, these statistics are similar for both the partial and the Voting rules. R_t is the buy-and-hold return at time t and N the total number of observations. $\boldsymbol{1}$ is a function that generates a one if the condition in brackets is filled. The first five lines of Table 1 display these statistics. The results indicate that the four strategies have percentages slightly higher than the buy-and-hold for long positions, as they range from 54.2% to 54.7%, compared with 53.1% for the buy-and-hold. This latter value is higher than 50% because there are more positive than negative returns. The percentages are lower for short positions; however we argue that the forecasting ability is stronger on the short than the long side of the strategy.

Table 1: Complex rules returns

Note: This table reports the results of three benchmark strategies: buy-and-hold (BH), random walk (RW) and best in-sample rule (BEST), as well as the four complex strategies over the entire evaluation period, from January 1994 to December 2008. *Nb Buy* and *Nb Sell* are the number of long and short daily positions generated by the strategies, *% Right* corresponds to the percentage of right positions as described above. For the Partial rule, Buy (Sell) is represented by a % invested higher (lower) than 0. *Buy*, *Sell* and *Strategy* are the annualized mean return of long, short positions and the overall strategy. The *t*-statistics test the null hypothesis that the mean of the specific series is significantly different from the mean return of the buy-and-hold strategy. *Strategy compounded* is the mean annualized return of the strategy in term of compounded returns. *Nb Trades* is the number of trades generated by the strategy. *Break Even TC* is the level of transaction costs that makes a strategy excess return over the buy-and-hold equal to zero. *Volatility Buy, Sell* and *Strategy* are respectively the annualized volatility of the long, short and overall positions. *Beta* and *Alpha* are estimated in the static CAPM framework. The *t*-statistics of the alpha and beta test the null hypothesis that these parameters are equal to 0. *Alphas* and *Sharpe Ratios* are expressed in annual terms.

	Investment strategies						
	BH	RW	Best	Opt_all	Opt_4	Voting	Partial
Nb Buy	3761	1999	2893	2907	2872	2924	2924
% right Buy	53.1%	51.3%	54.6%	54.2%	54.7%	54.4%	54.4%
Nb Sell		1758	868	853	889	837	837
% right Sell		44.6%	51.8%	50.5%	51.7%	51.1%	51.1%
% right Strategy	53.1%	48.2%	54.0%	53.4%	54.0%	53.7%	53.7%
Buy		-0.022	0.133	0.110	0.136	0.117	0.116
t-statistic		*-1.03*	*1.05*	*0.70*	*1.09*	*0.80*	*0.80*
Sell		-0.158	0.177	0.097	0.179	0.131	0.113
t-statistic		*-2.41*	*0.90*	*0.28*	*0.92*	*0.54*	*0.40*
Strategy	0.062	-0.086	0.143	0.107	0.146	0.120	0.115
t-statistic		*-2.10*	*1.16*	*0.65*	*1.20*	*0.83*	*0.78*
Strategy compounded	0.044	-0.099	0.133	0.093	0.136	0.107	0.103
Volatility buy		0.172	0.149	0.151	0.148	0.151	0.145
Volatility sell		0.212	0.293	0.291	0.292	0.293	0.279
Volatility strategy	0.192	0.192	0.192	0.192	0.192	0.192	0.183
Nb trade	1	3895	7	39	11	7	11.54
Break even TC		-0.1%	17.4%	1.7%	11.5%	12.5%	6.9%
Beta	1	-0.15	-0.08	-0.04	-0.09	-0.03	-0.03
t-statistic		*-9.07*	*-4.68*	*-2.75*	*-5.74*	*-2.12*	*-1.92*
Alpha		-0.110	0.116	0.079	0.120	0.092	0.087
t-statistic		*-2.25*	*2.35*	*1.59*	*2.42*	*1.85*	*1.83*
Sharpe	0.168	-0.599	0.594	0.404	0.609	0.472	0.467

Indeed, shorting continuously the market would only result in 46.88% of winning trades, while the percentages associated with the strategies lie between 50.53% and 51.74%. In addition, negative market returns are higher, in absolute value, than positive returns.

Turning to the economic performance, we first observe that each complex strategy yields a higher mean return than the market. They range, in annual terms, from 10.70% for Opt_all to 14.61% for Opt_4, while the buy-and-hold yields 6.16% only. While the Opt_4 performance is the highest among our four complex rules, those of the three others is rather close. This may result from the homogeneity of the simple MA rules returns, as shown in Figure 2 and 3. Thus, the strategies constructed with the various selection processes provide similar results. It is also interesting to note that the Opt_4 strategy generates a slightly higher return than the best in-sample simple MA rule, which yield 14.33%. This may indicate that the changing structure of the complex rule improves the performance, as trends may change as well. The high percentages of right sell signals are reflected in the fact that each of the four complex rules has a rather large positive return in the sell side of the strategy. For two of them (Voting and Opt_4), the mean sell return is even higher than their buy returns. The RW performs poorly, as indicated by a negative mean return. On the other hand, a contrarian strategy, which takes a long (short) position after a negative (positive) market return would yield a higher return than the market, i.e. a simple mean return of 8.54%. Nonetheless, each of our complex trading rules return is higher than this latter. In addition, this contrarian strategy trades very frequently with 3895 transactions[28], and thus, its performance is not achievable in a real trading setting.

Indeed, transaction costs are not included in the strategies performance directly. Nevertheless, they switch trading positions only between three and 19 times[29] over the whole 15 years of the test sample. Thus, the inclusion of realistic transaction costs would only diminish the overall performance marginally. The one-way break even transaction costs range between 1.74% and 12.45%, while 0.1% is not uncommon for large investors nowadays.

Despite the promising economic performance of these complex trading rules shown by high mean returns, their associated *t*-statistics are not large enough to reject the null hypothesis of equal means with the buy-and-hold return. Indeed, they lie between 0.65 and 1.20. This is far from the critical value associated with standard confidence levels. In the last Part of this thesis,

---

[28] This is higher than the number of days, as switching from a long to a short position implies two transactions.

[29] Note that moving from a long to a short position over one day is counted as two trades.

we show that Student *t*-tests are not a powerful procedure to test whether investment strategies mean returns differs from those of a benchmark strategy. Thus, we propose a new simulation based test and we repeat this performance analysis, which gives opposite conclusions regarding the statistical significance of these differences in mean returns.

To further illustrate the contrast between the economic and statistical performance, Figure 4 presents the cumulated compounded returns of the complex strategies, which are relevant from an investor point of view. They range between 274% for Opt_all and 572% for Opt_4, compared with only 90% for the buy-and-hold.

Figure 4: Complex rules compounded returns

This figure presents the total compounded return over the entire test period for two benchmark strategies, the buy-and-hold and the Best strategy and the four complex strategies.
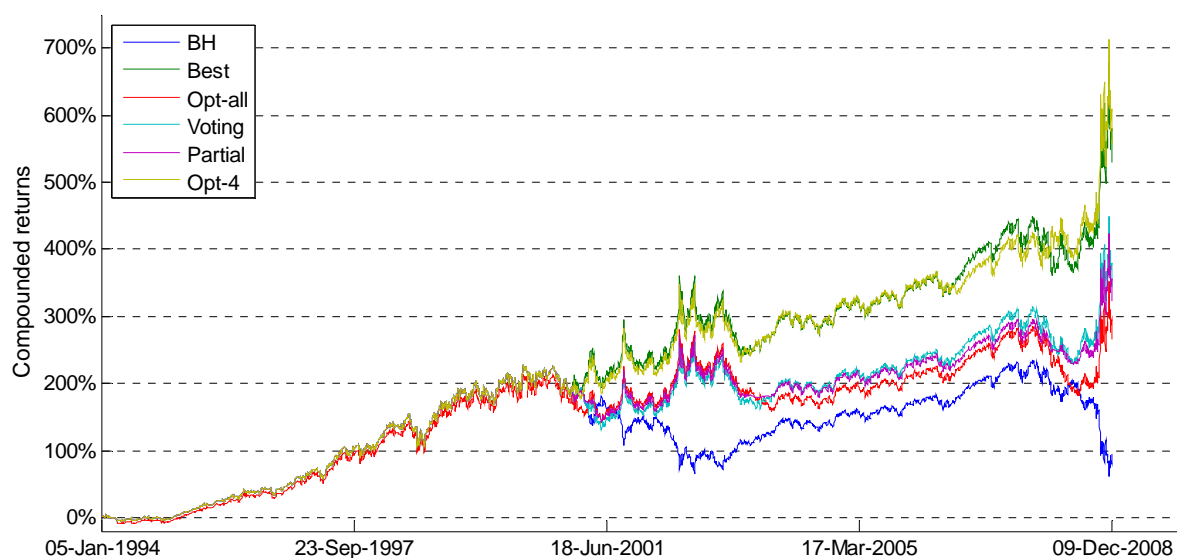
Finally, we examine whether the excess returns may compensate for higher risk bearing. Jensen's alphas provide strong evidence against this hypothesis, as the alphas are even larger than the differences between the strategies and the buy-and-hold returns. For instance, the average return of our four complex rules in excess of the buy-and-hold is 6.05%, while their average alpha is 9.87%. This can be explained by the fact that the strategies betas are very close to zero, and as the buy-and-hold return is higher than the risk-free rate, the alphas are higher than the difference in means. In addition, the alphas *t*-statistics are sharply higher than those associated with the differences in mean returns. The alpha associated with Opt_4 is even statistically different from zero at the 5% confidence level, and those of Voting and Partial are significant at

the 10% level. These findings are confirmed by the Sharpe ratios, which are at least twice as high as those of the buy-and-hold. This is not surprising as the complex strategies volatilities are similar to the market.

4.2.4   The anatomy of the complex trading systems

The results presented above suggest that the complex strategies also rely on long MA values longer than usual. Indeed, the performance is in line with the simple MA rules displayed in Figure 2, and the examination of the compounded returns in Figure 4 suggests that they exploit long-term trends. Indeed, this latter figure shows that MA strategies' higher returns come from shorting the market during when it is subject to downward trends. Thus, in this Section, we detail the behaviour of the complex trading rules, and we assess the performance of the Opt_4 selection process.

First, the Opt_all long MA lengths are highly concentrated around three values: 615, 665 and 940 days[30]. They represent 99% of the distribution. In addition, the shortest long MA is equal to 340 days, which is still sensibly higher than the longest parameter usually used. This is in close agreement with the Opt_4 rule, which takes only four different parameters, as the long MA lengths are 665, 515, 415 and 240 days. Moreover, the best performing rule in-sample has a long (short) MA length of 465 (60) days. This is confirmed with the correlation coefficients between the complex trading systems and the simple MA rules, which are reported in the following figure. They are close to one (zero) with the simple MA specifications that have a long MA rule associated with the highest (lowest) in-sample returns presented in Figure 2.

---

[30] The steadiness in the parameter is also explained by the optimization process. Indeed, if more than one rule produce the highest cumulated return over more than two periods, the optimization process will keep the same parameter.

Figure 5: Correlation between complex and simple trading rules

Note: These figures present the correlation coefficient between the returns of the four complex trading systems returns and those of the 938 simple MA rules with a SMA and LMA as stated on the Long MA and Short MA axes. They use the simple MA rules without the 1% security band. The sample covers the evaluation period from 1994 to 2008.
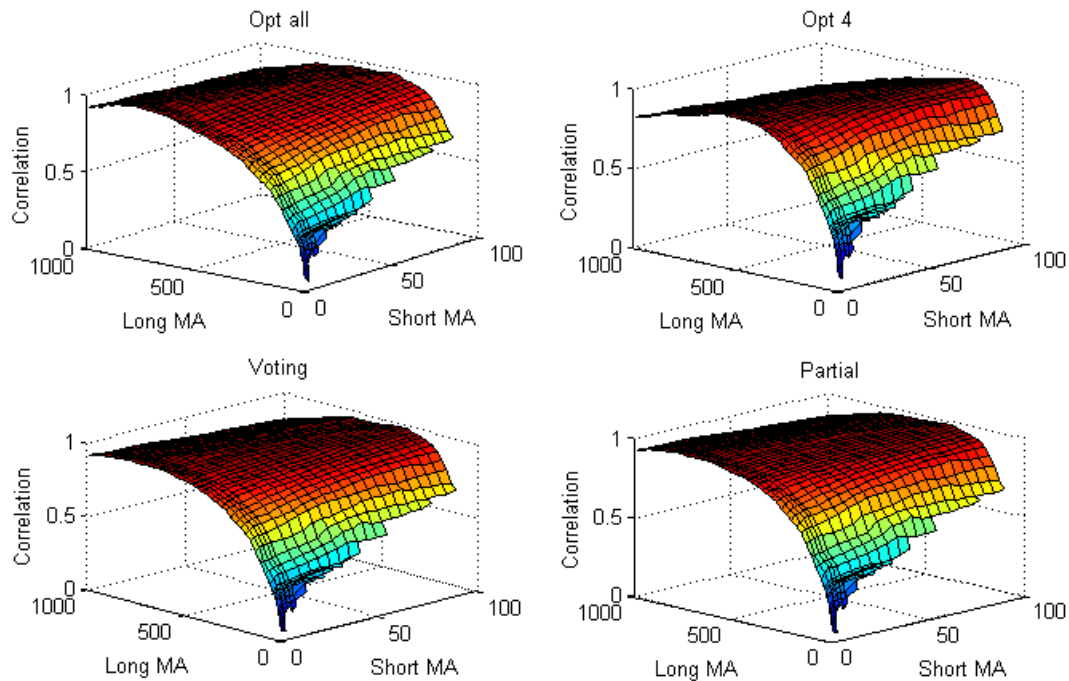
Finally, we provide some insights about the selection processes we use to construct the complex trading rules. Table 2 provides some insights about the performance of the selection process used by the Opt_4 rule, which contains four test samples. Thus, for each of them, we compare the performance of the selected rule with the average of all rules. We compute the percentage of simple MA rules with a higher or lower mean return than the Opt_4 strategy.

These statistics are reported in the third and fourth lines of the table. The selected rule has a higher return than the average of MA rules in the initial universe over each of the four subsamples. We also report the percentage of rules with a return higher or lower than the selected rule. Although the out-of-sample performance of the selected rule is the not the best among all specifications, only a small fraction of rules is able to generate higher returns, while the vast majority of rules have lower returns. The two statistics do not add up to 100% in the first subsample, because the markets followed a strong upward trend, and thus, many rules, like the selected one, consists in long positions only and generate a similar return.

Table 2: Out-of-sample performance of the Optim_4 strategy

Note: *Long MA* and *Short MA* are respectively the length of the long and short MA of the rule selected by the optimization process used by Opt_4. *Selected rule* is the out-of-sample simple mean yearly return of this strategy. *Mean* is the mean yearly return across the whole set of simple MA rules considered. *% high (low)* is the percentage of rules with a higher (lower) return than the selected specification.

	Subsamples			
	01.1994-12.1997	12.1997-01.2002	01.2002-01.2006	01.2006-12.2008
Long MA	665	515	415	240
Short MA	65	35	25	50
Selected rule	0.187	0.129	0.104	0.120
Mean	0.156	0.068	0.051	0.074
% high	0.00	0.09	0.03	0.03
% low	0.56	0.91	0.97	0.97

4.3   Leverage with exchange-traded options

4.3.1   The investment strategy

Options give the opportunity to take positions with leverage, since their premiums represent only a fraction of the underlying price. However it is rather unlikely that an investor would invest her entire capital in traded options because of the possibility of having a return of -100%, and therefore, losing the total value of her investment. This is even more relevant for our strategies that do not change of trading positions frequently. Thus, we propose to invest only a fraction of the available capital in options and the rest in the market, similarly to the standard setting. We evaluate these strategies with three percentages of options; 5%, 10% and 15%. Hence, a buy (sell) signal implies taking a long (short) position in the market equal to 95%, 90% or 85% and the remaining amount is used to buy call (put) options.

4.3.2   Data: Preparation and descriptive statistics

We use the daily observations of exchange-traded S&P 500 options obtained from the *Market Data Express service* of the *Chicago Board of Option Exchange*. Non continuous options, i.e. those with missing data, and those with prices that violate significantly the arbitrage conditions are removed from the sample. It contains 11'464 different call and 11'377 put options corresponding to a total of 1'828'800 daily observations.

Some preliminary work has to be done on the exchange-traded option database, as it does not include returns. We use simple returns since a log-transformation would give a minus infinity return for options that expire worthless. First of all, we compute closing prices with the last bid or last ask prices and the last sale prices according to the methodology given by the *CBOE*[31]. The bid-ask spread is also included in the options returns to have a realistic proxy. Consequently, the option return differs from days to days whether a new position is initiated or not. We calculate these returns as

$$
R_{opt,t+1} = \begin{cases}
\dfrac{O_{C,t+1} - O_{C,t}}{O_{C,t}} & \text{if } S_{t-1} = S_t = S_{t+1} \\[2ex]
\dfrac{O_{C,t+1} - O_{B,t}}{O_{B,t}} & \text{if } S_{t-1} \neq S_t = S_{t+1} \\[2ex]
\dfrac{O_{A,t+1} - O_{C,t}}{O_{C,t}} & \text{if } S_{t-1} = S_t \neq S_{t+1} \\[2ex]
\dfrac{O_{A,t+1} - O_{B,t}}{O_{B,t}} & \text{if } S_{t-1} \neq S_t \neq S_{t+1},
\end{cases}
\tag{1.14}
$$

where $O_{C,t}$, $O_{B,t}$, $O_{A,t}$ are respectively the closing, bid and ask option price at time t, and S_t is the trading signal. The first equation is used when there is no change in the trading position, hence closing prices are used. The second one corresponds to the initiation of a new position that lasts at least two days. The option is bought at the bid price at time t and the closing price is used at time $t+1$ as there is no trade. The next equation enables to compute returns for the last day of a position, the ask price is used as the option is sold. The return of a position which is kept only one day is calculated with the last equation. We apply the same method when the trading signal does not change, but the selected option does. We also have to calculate the moneyness and the options beta:

$$
Money_{call,t} = \frac{UP_t - X}{X} \qquad Money_{put,t} = \frac{X - UP_t}{UP_t},
\tag{1.15}
$$

where UP_t and X are the underlying (i.e. the S&P 500 index) price at time t and the option strike. This calculation ensures that in (out of) the money options have a positive (negative) moneyness.

---

[31] If the last sale is between the last bid and last ask the close is on the last sale. If the last sale is less than or equal to the last bid the option series is closed on the last bid and similarly if the last sale is greater than or equal to the last ask, the close is on the last ask. In the case where there is no last sale for an option series the previous day's close is looked at as if it were the last sale and the same rules are applied.

$$\beta_C = \frac{UP}{C} \Phi \left[\frac{\ln\left(UP/X\right) + \left(r + \sigma^2/2\right)t}{\sigma\sqrt{t}} \right] \beta_s \qquad (1.16)$$

where UP is the underlying current price, C is the call option price according to the Black and Scholes formula, X is its strike, r is the risk free rate, t is the time to maturity, represented as a fraction of a year, β_s is the underlying beta, which is equal to one, σ^2 the volatility estimated by the implied volatility index (VIX) and Φ is the cumulative density function of the standard normal distribution. Figure 6 illustrates the maximum and minimum strike prices, both for the sample of call and put options.

Figure 6: Strike prices of options in our database
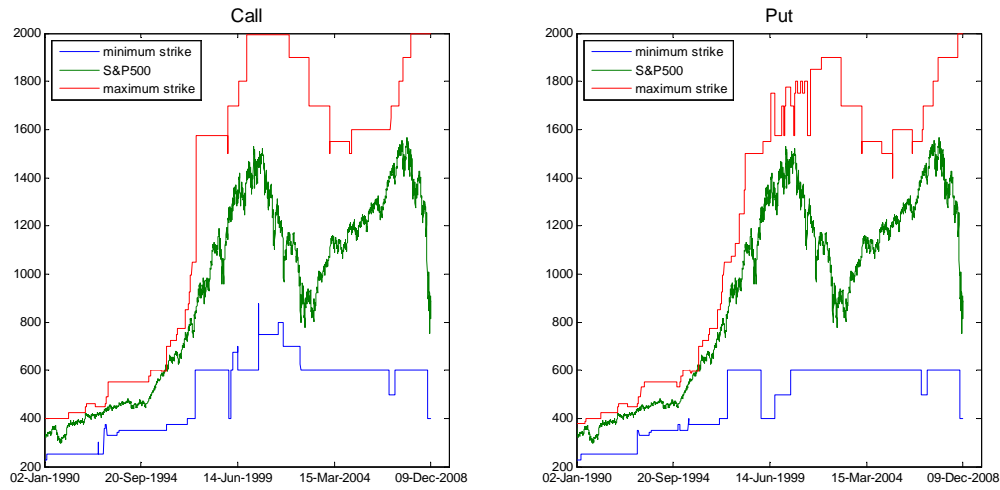
Table 3 presents the option returns descriptive statistics for various groups of options formed according to their maturity and moneyness. Daily returns are calculated with closing prices. The last line of each panel contains the mean option beta obtained under the Black and Scholes assumptions.

Call option returns share similarities with Wilkens (2007) who investigates option returns on the German DAX index in the CAPM framework. First of all, the returns increase from deep-in-the-money call options to at-the-money call options due to a higher leverage; however, returns of the out-of-the-money options are usually negative, especially when the maturity decreases. This is in contradiction with the hypothesis of positive expected call option returns, and it is probably due to a large loss in temporal value which is not compensated by an increase in intrinsic value, even if the underlying has an upward trend. Obviously, this is more important for option with short time to maturity. It is worth noting that each group, including the put options, has a positive skewness, which is consistent with the option asymmetric payoff.

Table 3: options returns descriptive statistics

Note: The statistics are computed from the entire option sample, from January 1990 to December 2008. The Short (long) maturity table contains options with less (more) than 30 (90) days to maturity. Medium maturity comprises options with a maturity between 30 and 90 days. *DOM* stands for deep-out-of-the money (moneyness < -0.15), *OM* for out-of-the-money (-0.15 ≤ moneyness < -0.05), *AM* for at-the-money (-0.05 ≤ moneyness < 0.05), *IM* for in-the-money (0.05 ≤ moneyness < 0.15) and *DIM* for deep-in-the-money (money >= 0.15). Returns are calculated with options closing prices and *Beta* is the average beta obtained under the Black and Scholes assumptions. All statistics are in a daily frequency.

Panel A: Call Options

	Short maturity					Medium maturity				
	Moneyness					Moneyness				
	DOM	OM	AM	IM	DIM	DOM	OM	AM	IM	DIM
Nb option	1291	3083	5352	3688	2936	1272	3288	5505	3615	2841
Nb obs	17079	34822	71749	46301	48326	28465	55248	108380	64042	75692
Mean	-0.020	-0.093	0.019	0.018	0.006	-0.011	-0.035	0.020	0.009	0.004
Median	0.000	0.000	-0.028	0.002	0.001	0.000	0.000	0.000	0.000	0.000
Min	-1	-1	-1	-0.62	-0.39	-0.98	-0.9	-0.8333	-0.46	-0.34
Max	14.70	19.00	72.68	3.81	1.82	19.00	11.00	473.00	1.28	0.93
Var	0.267	0.268	0.912	0.018	0.004	0.212	0.123	2.279	0.008	0.003
Skewness	9.41	6.03	28.78	4.42	3.05	15.27	5.06	290.13	1.42	1.99
Kurtosis	168.05	123.06	1501.21	72.47	60.92	398.89	78.16	89481	13.73	28.39
Beta	221.74	146.30	57.39	11.88	4.75	34.51	29.08	19.87	9.83	4.49

	Long maturity				
	Moneyness				
	DOM	OM	AM	IM	DIM
Nb option	860	1509	2116	1417	1169
Nb obs	57716	65754	86811	64729	108737
Mean	-0.010	-0.004	0.005	0.005	0.005
Median	0.000	0.000	0.000	0.000	0.000
Min	-0.98	-0.9982	-0.63636	-0.37	-0.79
Max	8.00	9.64	1.74	0.82	104.77
Var	0.061	0.026	0.009	0.003	0.167
Skewness	12.11	7.04	1.62	0.88	232.35
Kurtosis	279.52	255.72	19.15	11.30	55429
Beta	14.23	12.80	9.91	6.83	3.80

Table 3 (continued)

Panel B: Put Options

	Short maturity						Medium maturity				
	Moneyness						Moneyness				
	DOM	OM	AM	IM	DIM		DOM	OM	AM	IM	DIM
Nb option	2528	3852	5312	2942	1403		2494	3743	5462	3227	1376
Nb obs	41448	51105	70719	31811	17857		67322	71400	107006	49920	29674
Mean	-0.050	-0.107	-0.010	0.039	0.022		-0.019	-0.017	-0.001	0.021	0.021
Median	0.000	-0.111	-0.072	0.000	0.003		0.000	-0.031	-0.014	0.000	0.001
Min	-1	-1	-1	-0.66	-0.43		-0.974	-0.844	-0.725	-0.51	-0.41
Max	16.00	19.10	152.17	3.75	1.69		10.50	9.27	11.78	2.20	0.88
Var	0.159	0.376	1.181	0.041	0.013		0.097	0.093	0.046	0.019	0.009
Skewness	5.62	9.07	63.69	3.77	2.75		6.70	5.88	5.37	1.95	1.70
Kurtosis	133.40	181.70	6944.85	34.75	22.61		126.40	101.40	164	14.13	11.70
Beta	-69.04	-136.23	-57.76	-11.76	-4.04		-58.96	-40.62	-20.58	-9.65	-3.82

	Long maturity				
	Moneyness				
	DOM	OM	AM	IM	DIM
Nb option	1053	1417	2047	1373	830
Nb obs	101269	72713	81903	52709	48093
Mean	-0.006	-0.006	0.000	0.006	0.009
Median	0.000	0.000	0.000	0.000	0.000
Min	-0.9	-0.6472	-0.51545	-0.43	-0.36
Max	8.00	2.71	1.95	21.46	0.66
Var	0.025	0.014	0.011	0.016	0.004
Skewness	6.70	2.65	1.83	93.22	1.35
Kurtosis	159.10	34.92	18.94	15455	11
Beta	-27.06	-16.22	-10.86	-6.77	-3.17

The 2% mean daily return for at-the-money call options with a medium maturity represents a 500% annualized return, which is much higher than its risk measured by its beta. Nevertheless, the large variance and skewness and a mean return equal to zero indicate that this mean return depends on a few positive extreme observations, as the maximum daily return is 473%. Constantinides, Jackwerth and Perrakis (2009) show that similar options to ours are overpriced relative to their theoretical bounds during the 1997-2006 sample. Furthermore, as this is particularly pronounced for out-of-the-money call options, their results confirm our findings that the out-of-the-money options returns are systematically lower than their at or in-the money counterparts. This is not compatible with the CAPM; nonetheless, this may be in line with the

economic theory. Indeed, out-of-the-money put options are frequently used as insurance against a sharp drop in the index price. Consequently, they have a positive coskewness with the market, and they should have, according to the Kraus and Litzenberger (1976) framework, a higher price, and thus, a lower return.

For instance, investors prefer to hold a portfolio with positive skewness, and they would agree to sacrifice some returns in order to acquire an asset that increases their portfolio skewness. This hypothesis requires that such put options are non redundant securities. These various issues may have a negative influence on the performance of trading strategies with traded options.

Hence, the option selection method tries to lessen them by choosing at-the-money options. Indeed, we have to extract a continuous options time series from this options database, as only one call and one put option can be used every day. In line with studies dedicated to the analysis of options returns, we propose the following method: For every trading day, a single put and call option is selected according to the following criterions: First, we identify options with a maturity between 25 and 90 days and a moneyness level between -5% and +5%. Then, the option with the highest daily open interest is chosen. The use of a liquidity measure in the selection process should limit mispricing, which is more likely to happen when there is no trade over a few days. Finally, this option is kept until its time to maturity reaches 10 days. Table 4 presents the statistics of the selected options, which are effectively used in the strategies.

Table 4: Statistics of selected call and put options

Note*:* This table describes the options selected for our investment strategies. *Mean return* is the simple daily mean return, Open interest is the average number of contracts that are neither closed nor delivered. *Daily volume* is the average number of contract traded during a day, *Relative BA spread* is the relative bid-ask spread, *Moneyness* is the average moneyness as calculated in equation (1.15), *Maturity* is the average time to maturity in days and *BS beta* is the mean beta obtained under the Black and Scholes assumptions.

	Call	Put
Mean return	0.0048	-0.0076
Median return	-0.0069	-0.0292
Open interest	41915.7	45851.3
Daily volume	2806.0	3660.1
Relative BA spread	0.0685	0.0674
Moneyness	-0.0048	-0.0164
Maturity	51.1	51.1
BS beta	20.5	-24.2

On average, the relative bid-ask spread is between 6% and 7%[32]. As changing options every day would result in very high trading costs, we keep the same options as long as no other signal is emitted and its maturity is higher than two weeks. Nevertheless, it is important to include this bid-ask spread in order to obtain realistic results, as ignoring it would overestimate the profitability.

4.3.3   The performance of strategies using options

Before turning to the strategies, Table 5 describes the characteristics of their pure options component. This would correspond to invest the entire capital with options. Thus, returns of - 100% are likely and would eliminate the entire wealth. In this table, the buy-and-hold strategy in Panel A (B) consists in having a continuous position in call (put) options. For complex trading strategies, Panel A (B) displays various statistics describing the returns of call (put) options taken after a buy (sell) signal. Panel C provides the same statistics for the global strategies, or in other word, call (put) options during long (short) positions. First, the buy-and-hold call option series over the entire sample provides some insights about the merits and drawbacks of using options.

Both the mean and the median simple returns are negative, while the mean underlying return is positive. This results from the loss of value due to the passage of time[33]. Another issue arising from options is the systematic difference between mean and median returns. Indeed, the mean is predominantly positive for each strategy and both for the call and put components, while the median is systematically negative (or equal to zero for the call options of the Opt_4 strategy). The positive asymmetry is clearly due to the options payoff, as the maximum daily call return is 600%, and they range between 183% and 517% for puts. Nevertheless, negative medians indicate that most of returns are negative, even tough trading rules possess market timing abilities.

---

[32] The reported spread might be overstated as it is calculated with last bid and last price and does not correspond to an effective trade.

[33] This effect might be magnified in our study as we use options with a relative short time to maturity in order to limit options mispricing.

Table 5: Options returns

Note: This table reports the results of investing only with options according to signals generated by our strategies under consideration. *BH* corresponds in investing continuously with call options. *Mean return* is the simple mean return and *Annual mean* its annualized value. *Median return* is the median of the daily returns distribution. *Min (Max) return* is the minimum (*maximum*) daily return. *Volatility, Skewness* and *Kurtosis* are the daily volatility, skewness and kurtosis of the options returns series. *% neg ret after correct* is the percentage of negative option returns, while index returns are positive. *% pos ret after wrong* is the opposite.

Panel A: Long position - call options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Nb obs	3761	2893	2907	2872	2924	2924
Mean return	-0.0042	0.0030	0.0018	0.0045	0.0019	0.0026
Annual mean	-1.0573	0.7608	0.4599	1.1335	0.4851	0.6521
Median return	-0.0079	-0.0017	-0.0027	0	-0.0027	-0.0022
Min return	-1	-1	-1	-1	-1	-1
Max return	6	6	6	6	6	5.9698
Volatility	0.336	0.330	0.341	0.331	0.342	0.330
Skewness	4.50	4.86	5.20	4.87	5.16	5.15
Kurtosis	56.68	63.28	66.51	62.79	65.66	66.74
% neg ret after correct	0.152	0.152	0.152	0.151	0.153	0.153
% pos ret after wrong	0.047	0.047	0.047	0.048	0.047	0.047

Panel B: Short position - put options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Nb obs	3761	868	853	889	837	837
Mean return	-0.0166	0.0132	-0.0019	0.0136	0.0066	0.0058
Annual mean	-4.1843	3.3183	-0.4758	3.4201	1.6533	1.4715
Median return	-0.0538	-0.0355	-0.0429	-0.0357	-0.0400	-0.0267
Min return	-1	-0.69	-0.69	-0.69	-0.69	-0.68
Max return	9.23	5.17	1.83	5.17	5.17	2.29
Volatility	0.341	0.330	0.279	0.327	0.329	0.263
Skewness	7.15	5.00	1.10	5.06	5.21	1.92
Kurtosis	163.04	71.98	6.42	73.01	75.68	13.59
% neg ret after correct	0.177	0.162	0.174	0.154	0.173	0.173
% pos ret after wrong	0.033	0.005	0.005	0.005	0.005	0.005

Panel C: Strategy - call and put options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Nb obs		3761	3760	3761	3761	3761
Mean return		0.0054	0.0010	0.0066	0.0030	0.0033
Annual mean		1.3510	0.2476	1.6740	0.7451	0.8345
Median return		-0.0091	-0.0110	-0.0083	-0.0101	-0.0088
Min return		-1	-1	-1	-1	-1
Max return		6	6	6	6	5.97
Volatility		0.339	0.330	0.339	0.328	0.330
Skewness		4.89	4.68	4.91	5.17	4.79
Kurtosis		65.27	60.90	65.10	67.69	62.79

This is reflected in the percentage of days with a negative return, while the trading signal was right. They are named as *% neg ret after correct* in Table 5, and they are computed, respectively for calls and puts, as

$$\frac{\sum_{t=1}^{N} \boldsymbol{1}\{R_{C,t+1} < 0 | R_{t+1} > 0 \ \& \ S_t = 1\}}{\sum_{t=1}^{N} \boldsymbol{1}\{R_{t+1} > 0 \ \& \ S_t = 1\}}, \tag{1.17}$$

$$\frac{\sum_{t=1}^{N} \boldsymbol{1}\{R_{P,t+1} < 0 | R_{t+1} < 0 \ \& \ S_t = -1\}}{\sum_{t=1}^{N} \boldsymbol{1}\{R_{t+1} < 0 \ \& \ S_t = -1\}}, \tag{1.18}$$

where N is the number of trading signals, $R_{C,t}$, $R_{P,t}$ and R_t stand for the call, put and index return at time t, and S_t the rule trading signal at time t. The operator $\boldsymbol{1}$ takes a one if the condition between brackets is true and zero otherwise. This level is situated around 15% for call options and slightly above for puts. Mispricing and decreases in volatility may also be an explanation; however the opposite statistic, i.e. the percentage of positive returns after an erroneous signal, is much lower. These results suggest that options should be used only when the trading strategies possess very strong market timing abilities but they may be counterproductive otherwise.

These findings are also reflected in Table 6, which displays the complex strategies performance with a percentage of the capital invested in options. The difference between means of simple and compounded returns is especially noteworthy. The former are sharply higher than the latter, and these difference increases with the percentage invested in options grows. For instance, this difference for the Opt_all strategy is 7.41% in annual term with 5% of options, and it reaches 35.7% with the highest proportion of options. Nevertheless, only compounded returns reflect the effective performance obtained by an investor who follows these strategies.

Options affect the strategies performance very differently. Indeed, the mean compounded return of the Opt_4 strategy increases from 13.5% to 19.2%, in annual terms, when 15% of the capital is invested in options. On the other hand, performance of the Opt_all strategy becomes negative. In addition, the volatility of these leveraged strategies increases massively. While it is around 19% on an annual basis for the unleveraged strategies, it is, on average, 40%, 65% and 90% with the three percentages of the capital invested in options. Taking this into account, the higher return is not worth the increase in volatility, with the possible exception of the best-performing rule, the Opt_4 strategy.

Table 6: Complex rules returns with options

Note: This table presents the performance of the trading strategies with traded options according to three levels of leverage. *Buy*, *Sell* and *Strategy* are respectively the simple mean annual return of the long, short and overall positions. *Strategy compound* is the annual mean compounded return. *Volatility buy, sell* and *strategy* are the annualized volatilities of the two components of the strategies and the global one

Panel A: 5% of options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy		0.2241	0.1873	0.2450	0.1959	0.2017
Sell		0.3689	0.1135	0.3760	0.2448	0.2083
Strategy	0.0737	0.2575	0.1705	0.2760	0.2068	0.2032
Strategy compound	-0.0070	0.1959	0.0964	0.2182	0.1342	0.1431
Volatility buy	0.366	0.368	0.379	0.368	0.380	0.360
Volatility sell	0.245	0.499	0.473	0.496	0.497	0.435
Volatility strategy	0.408	0.402	0.402	0.402	0.409	0.378

Panel B: 10% of options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy		0.3151	0.2649	0.3541	0.2750	0.2877
Sell		0.5607	0.1295	0.5731	0.3583	0.3033
Strategy	0.0858	0.3718	0.2341	0.4059	0.2936	0.2912
Strategy compound	-0.1088	0.1944	0.0415	0.2359	0.0969	0.1286
Volatility buy	0.637	0.615	0.635	0.616	0.636	0.601
Volatility sell	0.366	0.735	0.671	0.729	0.731	0.607
Volatility strategy	0.655	0.644	0.643	0.645	0.658	0.602

Panel C: 15% of options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy		0.406	0.342	0.463	0.354	0.374
Sell		0.752	0.145	0.770	0.472	0.398
Strategy	0.098	0.486	0.298	0.536	0.380	0.379
Strategy compound	-0.242	0.134	-0.059	0.192	0.007	0.066
Volatility buy	0.914	0.866	0.895	0.868	0.897	0.846
Volatility sell	0.494	0.978	0.874	0.970	0.974	0.783
Volatility strategy	0.909	0.893	0.890	0.893	0.914	0.832

We do not use Sharpe ratios to examine this trade-off, as the returns distributions depart clearly from normality. In Section 5 of this first part, we present some risk measures that include the distribution higher moments.

Another puzzling consequence of using options is that the increase in performance is not consistent. For the Opt_4 trading system, the mean annual compounded return increases from 13.6% for the unleveraged strategy to 21.82% with 5% invested in options. Then it increases again up to 23.59% as the percentage in option is 10%. Nonetheless, it decreases to 19.22% with the highest level of options. Figure 7 displays the compounded returns for the four versions of the Opt_4 trading rules. It illustrates the volatility increase induced by the options, and it shows that the strategies returns are much less consistent. For example, the compounded return of the strategy with 15% of options moves below the one without options at the end of the sample before increasing sharply. Thus, the performance is much more dependent on the sample under investigation for strategies with options than for those without leverage.

Another issue is that the selected options have an average maturity of 51 days. This contrasts with the long-term perspective that characterizes the positions taken by the complex trading rules. However, we can not choose a more appropriate set of options, i.e. those with longer maturity in this case, without relying on the look-ahead bias. Indeed, the complex trading rules may suddenly start to exploit short-term trends. In Appendix I-G, we show the same results but with options that have a time-to-maturity of 100 days on average. They are similar to those presented in this Section. This indicates that the issues of using options to increase the profitability are not specific to a particular option time-series.

Figure 7: Opt_4 strategy compounded returns with options
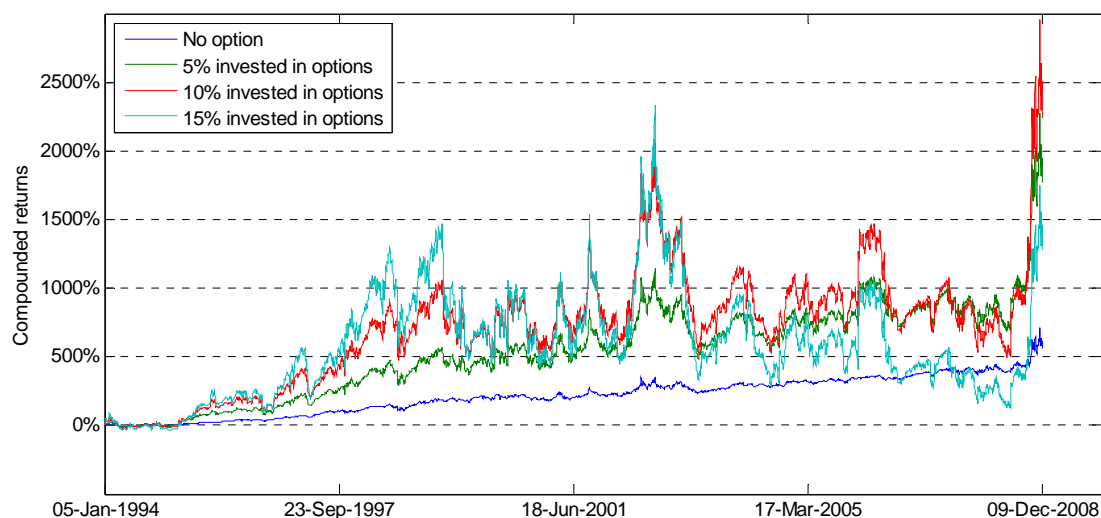
To summarize, the use of options may be valuable only for trading rules with high forecasting abilities, and the investor should also tolerates a large increase in volatility.

4.4   Leverage with debt

4.4.1   Methodology and data

The second method to add leverage in the complex trading rules relies on debt. The new investment strategy consists in borrowing 100% of the capital to double the amount invested after a buy signal, and similarly, a sell signal implies to short 200% of the capital. We assume that the capital is sufficient to cover the shorting requirements as collateral, and therefore, no other cost is taken into account. The returns of a strategy with debt leverage, $R_{DL,t}$, including transaction costs, are computed as

$$R_{DL,t+1} = \begin{cases} S_t \cdot 2 \cdot \left[\left(\dfrac{P_{t+1}-P_t}{P_t} \right) - \dfrac{R_{B,t+1}}{2} - |S_{t+1}-S_t| \cdot TC \right] & \text{if } 0 < S_t \leq 1 \\[2ex] S_t \cdot 2 \cdot \left[\left(\dfrac{P_{t+1}-P_t}{P_t} \right) - |S_{t+1}-S_t| \cdot TC \right] & \text{if } -1 \leq S_t < 0 \\[2ex] R_{L,t+1} & \text{if } S_t = 0, \end{cases} \qquad (1.19)$$

where P_t the index price at time t, $R_{B,t}$ is the borrowing rate and $R_{L,t}$ the lending rate. S_t represents the trading signal and it takes 1 (-1) for a buy (sell) trading signal, or the fractional trading signal for the Partial strategy, and zero for a neutral signal. TC is the one-way transaction cost, which is arbitrarily set to 0.2%. In the first line associated with long positions, the borrowing rate is divided by two in order to consider this cost only on the debt and not on the capital.

As different investors have different borrowing costs, we compute the results with three different levels: a higher borrowing rate for retail investors, the lending rate which may be realistic for large investors and zero borrowing cost, which may proxy the profitability of investing with futures. The US Bank prime loan rate is the borrowing rate, which yields on average 2.3% p.a. more than the lending rate. The two interest rates are also extracted from *Thomson Reuters Datastream.*

4.4.2   Performance

Table 7 displays the performance analysis for our complex trading strategies with the three levels of borrowing costs. It also includes the buy-and-hold performance obtained with leverage in order determine if a strategy higher return with leverage is due solely to it or to its forecasting power. First, we want to highlight the influence of using either the simple or the compounded mean returns to quantify the usefulness of using debt leverage. For instance, the buy-and-hold simple mean return without leverage is 6.16%, which is lower than the ones obtained with debt, even with the highest level of borrowing cost, as it is 7.3%. Nonetheless, the compounded returns with leverage are lower than those without leverage with the two highest borrowing cost levels, and they are only slightly higher without cost, i.e. 5% compared with 4.4% in the standard investment setting.

The complex trading systems offer a different picture. Indeed, even with high borrowing costs, the compounded average return is higher with debt leverage than in the standard investment setting for three out of the four rules. For the other one, the Opt_all strategy, they are lower with debt than in the standard investment context, except when borrowing costs are set to zero. In this case, the annual mean compounded return is slightly higher at 13.8% compared with 13.3% without leverage. For the Opt_4 strategy, which has the highest predictive power, the difference is large. The annual mean compounded returns are 19.1%, 21% and 24% according to the three level of borrowing costs, while it is only 13.62% in the standard investment setting without leverage. This is the first evidence that adding leverage is valuable only if the trading rule has forecasting abilities.

The second one is provided by comparing strategies compounded returns with their buy-and-hold counterparts. The complex strategies generate annual mean compounded returns ranging from 9.2% to 19.1% with the highest borrowing rate, while the buy-and-hold statistics is zero. When the risk free rate is used as the borrowing cost, they generate returns ranging from 10.9% to 21% while the buy-and-hold is only 2%. The difference is lower when no borrowing costs are considered, but they remain large, between 13.8% and 24% for the strategies and 5% for the buy-and-hold. This difference is due to the return calculation method with the borrowing costs. Indeed, the buy-and-hold is always long, and thus, has to pay the borrowing costs every day. On the other hand, trading strategies have a significant amount of short positions during which no interest is paid. This indicates that if the strategies performance is higher with leverage, this is not due to the leverage itself but to the timing abilities.

Table 7: Complex rules returns with debt leverage

Note: *Buy*, *Sell* and *Strategy* are the annual simple mean returns of the long positions, sell positions and the whole strategy. *Compound* is the annual compounded mean return of the strategy. *Beta* and *Alpha* are obtained with an OLS regression with excess returns in the static CAPM setting. The alpha is annualized. The *t*-statistics associated with the strategy mean return results from a Student *t*-test about equal means between the strategy and the leveraged buy-and-hold returns with the corresponding borrowing cost. Those associated with the alphas determine whether they are different from zero. Sharpe is the annual Sharpe ratio.

Panel A: borrowing rate = US Bank prime loan

	Investment strategies					
	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy	0.073	0.213	0.166	0.219	0.18	0.18
Sell		0.354	0.195	0.358	0.262	0.226
Strategy	0.073	0.244	0.162	0.249	0.197	0.187
t-statistic		*1.21*	*0.63*	*1.25*	*0.88*	*0.83*
Compound	0.000	0.185	0.092	0.191	0.13	0.127
Volatility	0.384	0.383	0.384	0.383	0.384	0.367
Beta	2.000	-0.152	-0.09	-0.186	-0.069	-0.06
Alpha	-0.02	0.22	0.136	0.226	0.169	0.16
t-statistic	*-74.63*	*2.22*	*1.37*	*2.28*	*1.71*	*1.68*
Sharpe	0.11	0.56	0.346	0.573	0.436	0.43

Panel B: borrowing rate =risk free rate

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy	0.093	0.234	0.186	0.239	0.2	0.199
Sell		0.354	0.195	0.358	0.262	0.226
Strategy	0.093	0.26	0.178	0.264	0.212	0.202
t-statistic		*1.18*	*0.6*	*1.22*	*0.85*	*0.79*
Compound	0.020	0.204	0.109	0.21	0.148	0.144
Volatility	0.384	0.383	0.384	0.383	0.384	0.367
Beta	2.000	-0.152	-0.09	-0.186	-0.069	-0.06
Alpha	0.00	0.235	0.151	0.241	0.185	0.175
t-statistic	*-1.00*	*2.37*	*1.52*	*2.44*	*1.87*	*1.84*
Sharpe	0.17	0.6	0.387	0.613	0.477	0.471

Panel C: no borrowing cost

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy	0.123	0.266	0.22	0.272	0.233	0.231
Sell		0.354	0.195	0.358	0.262	0.226
Strategy	0.123	0.285	0.203	0.289	0.238	0.227
t-statistic		*1.15*	*0.57*	*1.19*	*0.82*	*0.76*
Compound	0.050	0.235	0.138	0.24	0.178	0.173
Volatility	0.384	0.383	0.384	0.383	0.384	0.367
Beta	2.000	-0.152	-0.089	-0.186	-0.069	-0.06
Alpha	0.03	0.26	0.177	0.266	0.211	0.2
t-statistic	*87.17*	*2.63*	*1.78*	*2.69*	*2.12*	*2.1*
Sharpe	0.24	0.666	0.454	0.678	0.544	0.539

This is confirmed with the strong positive returns generated by the short side of the strategies. The Opt_4 rule produces a massive 35.8% annual simple mean return on the short side, which is higher than the long side with 27.2% without borrowing costs.

Figure 8 displays these returns over the entire sample for the Opt_4 strategy and for various specifications of the buy-and-hold. This graph indicates that the higher performance of leveraged strategies is not due to the sole fact of using leverage, but genuinely to the forecasting power of this trading system.

Figure 8: Opt_4 strategy compound returns with debt leverage

Note: *BH, BH leverage loan* and *BH leverage no cost* are respectively the buy-and-hold without leverage, with leverage when the US bank loan rate is used as borrowing rate and when no borrowing cost is considered. The three last series are those of the Opt_4 strategy with the same 3 levels of borrowing costs.
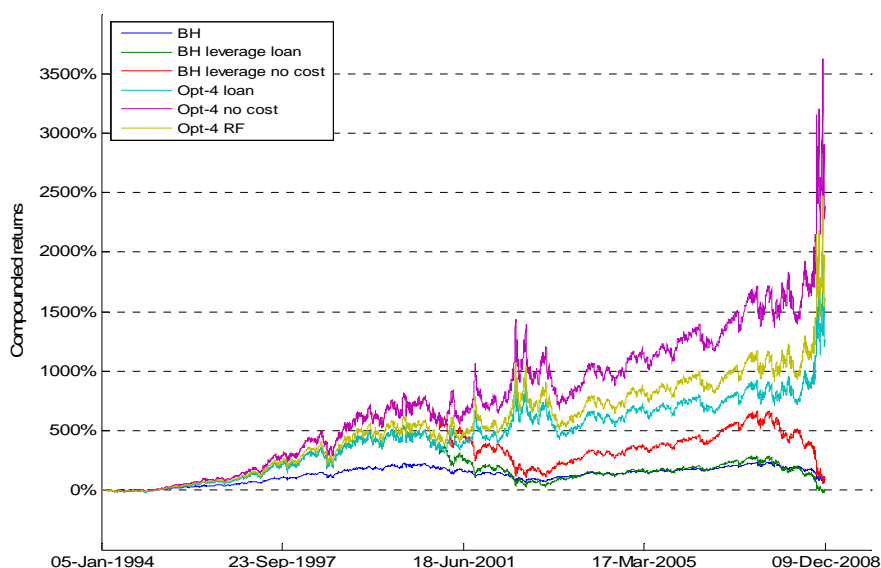
This is confirmed by analysing alphas. Indeed, using leverage without forecasting power would only increase the beta, and the normal return for bearing more market risk, but not the alpha. However, we observe that the alpha of the leverage buy-and-hold without borrowing cost is positive and strongly significant. The alphas are derived from the static CAPM, which is expressed in term of returns in excess of the risk-free rate. Thus, considering a zero borrowing cost implies an abnormal return equal to the risk-free rate. This coincides with the buy-and-hold alpha that is equal to the average risk-free rate over the test sample. The average annualized alpha for the four complex strategies without leverage is 9.4%, while it is between 1.83 and 2.2 times higher for the leveraged strategies. They are all higher than the 2% associated with the leverage

buy-and-hold without borrowing costs. This indicates that risk is not the primary reason for these high abnormal returns.

On the other hand, the leveraged strategies Sharpe ratios do not increase. They are very close to those obtained in the standard investment setting. Nevertheless, they consider the standard deviation as the risk measure, and thus, require a normal distribution. This is not the case when leverage is used[34].

Finally, it is interesting to compare the merit of including leverage with the two methods, i.e. with debt or exchange-traded options. We find that debt is much more suitable to improve the performance of the complex trading rules than options. Especially, the increase in volatility is more limited with debt. While the two methods provide similar return for the Opt_4 strategy, the volatility with debt is much lower, as it is 38% compared with 89% with 15% of the capital invested in options. Another danger of using options is that if the trading rule does not have a strong forecasting power, the profitability drops sharply. On the other hand, debt improves the performance for the four rules, and not only for the best-performing rule, the Opt_4.

5.   Other risk measures

The risk measures that we use so far to examine the strategies performance rely on the returns normality for the Sharpe ratio, and on the CAPM framework for the alphas. Thus, they are not suitable for our leverage strategies, which have non-normal returns. We propose to examine other measures linked to the downside risk and the coskewness.

First, the Sharpe ratio uses the return volatility as the risk measure, which comprises returns both on the upside and on the downside. Thus, the risk refers to the uncertainty in returns, while investors may be more likely to consider risk as the probability of facing large losses. Another statistical issue is that the volatility is usually larger during down markets than when they rise. Thus, the Sortino ratio uses the downside standard deviation as the risk measure. Compared with the volatility, the downside volatility is an un-centred moment, which means that the mean is not subtracted in the computation. We also present the upside and downside betas, β^- and β^+, calculated in the same CAPM setting but when the returns series are divided according to the

---

[34] The mean skewness of strategies with debt and options is respectively -0-21 and 2.99, whereas the average kurtosis is 13.40 and 33.85.

sign of the market return. They shed lights on whether strategies take more risk when the market rises or declines.

To cope with non normal returns, especially for leveraged strategies, we compute the coskewness and the downside coskweness. This first measure is the third co moment between the strategy and the market returns, while the second one considers only returns when those of the market are negative. The objective is to analyse how the strategy performs when the market undergoes large movement. The downside coskewness is particularly interesting as it only considers returns when the market return is negative. The intuition is the following: During a bear market, it is rather likely that the majority of assets hold by an investor declines in value, thus, he may want to hold some assets that decline less, or even rise during such times. In consequence, these assets, with a high downside coskewness, are considered as less risky and holding them should yield a lower return, and conversely, those with low downside coskewness should have a higher expected return. One possibility to explain our strategies high returns is that they are performing badly during downside markets periods, and thus, the return premium is justified by more risk and not by good forecasting abilities. For instance, Ang, Chen and Xing (2006) find that these two risks, the downside risk and the coskewness exposure, are priced independently in the cross-sections of stocks returns, and they both bear a statistically and economically significant risk premium.

The various risk measures, the downside and upside betas (β^- and β^+), the Sortino ratio, the unconditional coskewness (*cos*) and the downside coskewness (*cos -*) are estimated as

$$\beta^- = \frac{\text{cov}\left(R_{Strat}, R \mid R < 0\right)}{\text{var}\left(R \mid R < 0\right)} \tag{1.20}$$

$$\beta^+ = \frac{\text{cov}\left(R_{Strat}, R \mid R > 0\right)}{\text{var}\left(R \mid R > 0\right)} \tag{1.21}$$

$$Sortino = \frac{\overline{R}_{Strat}}{\sqrt{\sum_{t=1}^{N}\left(R_{Strat} \mid R_{Strat} < 0\right)^2 \Big/ \sum_{t=1}^{N} \mathbf{1}\{R_{Strat} < 0\}}} \tag{1.22}$$

$$\cos = \sum_{t=1}^{N}\left(R_{Strat,t} - \overline{R}_{Strat}\right)\left(R_t - \overline{R}\right)^2 \Big/ N \tag{1.23}$$

$$\cos^- = \sum_{t=1}^{N}\left[\left(R_{Strat,t} - \overline{R}_{Strat}\right)\left(R_t - \overline{R}\right)^2 \mid R_t < 0\right] \Big/ \sum_{t=1}^{N} 1\{R_t < 0\}, \tag{1.24}$$

where R_{Strat} and R are the strategy and buy-and-hold returns series. Table 8 shows that the conclusions reached either with the Sharpe or the Sortino ratios are similar, as the strategies yield consistently higher ratios than the benchmark. Thus, the performance can not be attributed to bearing more downside risk. This is confirmed by the downside betas that are consistently lower than their upside betas. This indicates that the strategies do not load market risk when the market follows a downside trend. Regarding the coskewness, it appears that strategies with debt have a negative coskewness, while it is positive for strategies using options. This means that strategies with options tend to increase the skewness of the returns while strategies with debt tend to reduce the skewness of returns distributions. Nonetheless, the latter have a mean or median return higher than the strategies with options, which may make this conclusion confuse. Indeed, the computation of the coskewness implies to remove the mean returns from the daily observations. Moreover, leveraged strategies have a positive downside coskewness, which is higher than in the standard investment setting. This indicates that they have positive returns when the buy-and-hold returns are on the far left side of the distributions. Thus, leverage, especially with debt, increases returns and reduces at the same time the risk associated with skewness when the market declines. In short, combining our strategies with the buy-and-hold would generate a portfolio with a higher skewness when returns are negative, which is sought by risk averse investors. Our strategies are especially interesting for investors as this "skewness insurance" does not imply lower returns, as it is even the opposite.

To summarize, these other risk measures do not support the hypothesis that the large returns of our trading strategies compensate for more risk bearing when the hypothesis of normal return is relaxed.

Table 8: Other risk measures

Note: The five first lines grouped under the Standard label correspond to strategies without financial leverage. The six subsequent groups refer to the debt leverage strategies according to the three level of borrowing costs and to the strategies with options. Sortino ratios are annualized. Values in the two last columns are multiplied by 10,000.

		Risk measures				
		β^-	β^+	Sortino	Cos	Cos$^-$
Standard	BH	1.000	1.000	0.310	-0.007	-0.250
	Opt_all	-0.053	-0.036	0.543	-0.024	0.179
	Voting	-0.047	-0.022	0.608	-0.026	0.176
	Opt_4	-0.124	-0.061	0.745	-0.020	0.188
	Partial	-0.044	-0.015	0.612	-0.023	0.177
Debt Borrowing cost	BH	2.004	1.996	0.186	-0.013	-0.499
	Opt_all	-0.103	-0.075	0.413	-0.048	0.359
	Voting	-0.091	-0.045	0.500	-0.051	0.352
	Opt_4	-0.245	-0.123	0.638	-0.040	0.378
	Partial	-0.086	-0.032	0.500	-0.045	0.355
Debt RF	BH	2.000	2.000	0.236	-0.014	-0.499
	Opt_all	-0.106	-0.072	0.453	-0.048	0.359
	Voting	-0.094	-0.042	0.540	-0.051	0.352
	Opt_4	-0.248	-0.120	0.677	-0.040	0.377
	Partial	-0.088	-0.029	0.540	-0.045	0.355
Debt no cost	BH	1.994	2.007	0.310	-0.014	-0.499
	Opt_all	-0.110	-0.067	0.517	-0.048	0.359
	Voting	-0.098	-0.037	0.603	-0.052	0.352
	Opt_4	-0.252	-0.115	0.738	-0.041	0.377
	Partial	-0.093	-0.024	0.605	-0.046	0.355
Options 5%	BH	1.726	1.974	0.209	0.052	-0.321
	Opt_all	0.044	0.373	0.498	0.055	0.281
	Voting	0.034	0.401	0.604	0.057	0.280
	Opt_4	-0.115	0.309	0.819	0.057	0.304
	Partial	0.063	0.396	0.635	0.051	0.277
Options 10%	BH	2.452	2.949	0.162	0.111	-0.392
	Opt_all	0.142	0.781	0.459	0.134	0.382
	Voting	0.115	0.824	0.576	0.140	0.384
	Opt_4	-0.107	0.679	0.812	0.135	0.419
	Partial	0.169	0.806	0.615	0.125	0.376
Options 15%	BH	3.179	3.923	0.138	0.169	-0.464
	Opt_all	0.239	1.189	0.437	0.213	0.484
	Voting	0.196	1.246	0.557	0.223	0.488
	Opt_4	-0.099	1.049	0.802	0.212	0.534
	Partial	0.275	1.217	0.600	0.199	0.476

6.   A market timing test based on trading positions

6.1   Intuitions and methodology

A multitude of market timing tests have been proposed in the literature. We provide an overview in the literature review of the third Part, which is dedicated to test procedures. They usually examine whether the trading position is in line with the next period return, i.e. in case of market timing abilities, a long position should be followed by a positive return. Nonetheless, we argue that they are not suitable for our strategies, which follow long-term trends in the market. Suppose for instance that a bull market is occurring; however there are 45% of negative daily returns. If the trading rule is designed to invest according to the market phases and it is successful, it will generate a buy signal over the whole bull market phase. With the commonly used tests, this means that 45% of the trading signals are not right. Thus, we propose a test that relies on these market phases, and not on the next period return. We could have performed standard tests on yearly returns instead of daily returns for instance. However, this implies to reduce sharply the number of data available and this has an impact on the statistical properties of the tests.

The first step consists in defining the market phases, or in other words, bull and bear markets. Visual inspection could be used; nonetheless, this approach would be subjective. Instead, we consider a variation of the algorithm proposed by Pagan and Sossounov (2003) designed to identify turning points inside various cycles. A cycle is defined as two subsequent phases, a bull market following a bear market or the opposite. A central issue is to separate local peaks or troughs from turning points in the cycle. Indeed, a short decrease (increase) in prices during a bull (bear) market should not indicate a change in the primary trend. This should be considered as a correction (rally) in a bull (bear) market. The algorithm proceeds as follows: First, it identifies the highest and lowest points over a 30 months window. The next step is to ensure that each market phase persists for at least nine months, or if it is not the case, the difference between the highest and lowest price should be larger than 25% in absolute value. The last stage of the algorithm warrants that a market cycle lasts for two years at least. Figure 9 displays the phases identified by the algorithm for the S&P 500 over the test sample.

Once the bull and bear phases are identified, the trading signals of the complex and the simple MA rules are analysed in this framework. Four different statistics describing whether long (short) positions coincide with bull (bear) markets are computed.

Figure 9: Bull and Bear markets

Note: This graph presents the separation between bull markets in blue and bear markets in red according to the algorithm presented above.
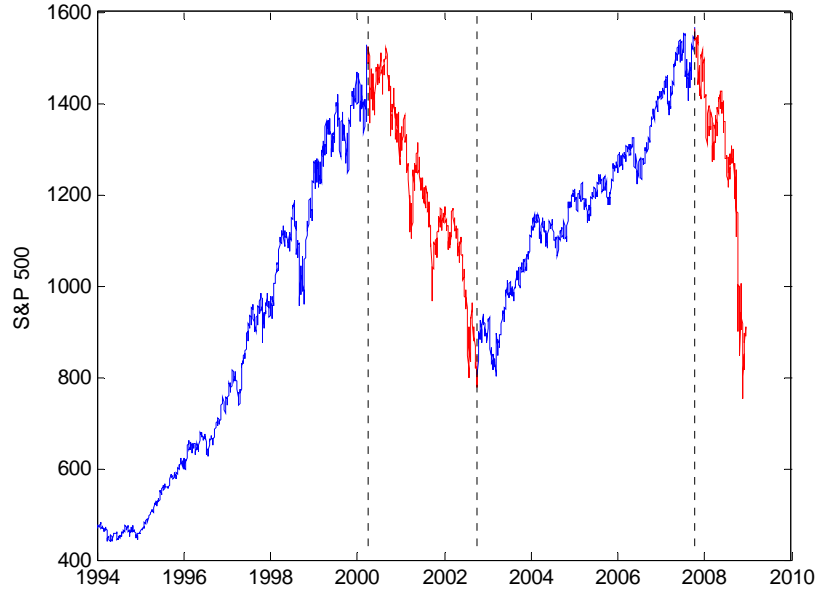
For each series, we compute the following statistics: the percentage of long (short) positions during bull (bear) market and the total percentage of right or wrong[35] signals. They are computed as:

$$\% \text{ buy-Bull} = \frac{\sum_{t=1}^{N}\left(S_t = 1 \middle| S_{BH,t} = 1\right)}{\sum_{t=1}^{N}(S_{BH,t} = 1)} \tag{1.25}$$

$$\% \text{ sell-Bear} = \frac{\sum_{t=1}^{N}\left(S_t = -1 \middle| S_{BH,t} = -1\right)}{\sum_{t=1}^{N}(S_{BH,t} = -1)} \tag{1.26}$$

$$\% \text{ right} = \frac{\sum_{t=1}^{N}\left(S_t = S_{BH,t}\right)}{N} \tag{1.27}$$

---

[35] Here, right (wrong) signals are defined as long (short) position during a bull (bear) market, whatever the market return.

$$\% \text{ wrong} = \frac{\sum_{t=1}^{N}\left(S_t \neq S_{BH,t}\right)}{N}, \tag{1.28}$$

where S_t is the trading rule signal at time t, $S_{BH,t}$ takes the value of 1 or -1 whether the market is in a bull or bear phase and N is the total number of trading days. Let us denote one of these statistics as V. In order to determine the significance of these statistics, we use a bootstrap methodology.[36] The idea is to compare them with those obtained by randomly generated trading signals. For each strategy, 500 (N) random trading signals series (V^*) are constructed according to a block bootstrap procedure in order to keep, to some extent, the same structure as the original series. Then, the four statistics presented above are computed with each of this 500 artificial series in conjunction with the original market phases detailed in Figure 9. Finally, they are ranked such as $V_1^* < V_2^* < \ldots < V_N^*$. We calculate the empirical p-value, P, as:

$$V_m^* \leq V < V_{(m+1)}^* \tag{1.29}$$

$$P = 1 - M / N \tag{1.30}$$

Intuitively, this p-value corresponds to the percentage of simulated trading signals series that have a higher value than the original statistic. This test differs from standard market timing test, such as the Henriksson and Merton (1981) test, by using market phases instead of market returns as turning points. Therefore, the proposed test should not be used to detect short-term market timing abilities.

6.2   Results

We perform this test for the four complex trading rules. As the test requires only the trading signals and not the strategies returns, the investment approach, with or without leverage, is irrelevant. We also provide these statistics for standard simple MA rules widely used in the literature, which are presented in the last four columns of the table.

---

[36] We choose to use non overlapping blocks, which means that each observation of the original series is only in one block. Then the procedure is with replacement, which implies that a specific block may appear more than once in the simulated series. These choices are subjective, as the issue of determining the best bootstrap procedure is still debated among econometricians and most of the papers are linked with the estimation of parameters and not for its use in computing empirical p-values.

Before turning to the results, Panel A of Table 9 documents the bootstrap process. It displays the median, minimum and maximum value of the four statistics computed for 500 artificial signals series. The length of the block is set to 40 days, thus, 94 blocks are available for the bootstrap. First, the standard MA rules indicate that there is a clear relationship between the parameters length used in the MA rules and the percentage of right signal. The longer (shorter) the length of the MA window is, the higher (lower) the percentage of long position in bull markets is. Nevertheless, the opposite is true for the percentage of short positions during bear markets.

For example, the artificial series based on the (2,10) MA rule have a median percentage of 57.2% of long position during the two bull markets, while it increases up to 69.4% for the (50,200) MA rule. This panel also supports the bootstrap methodology (i.e. performing simulations based on each original signals series) used for the test, as the simulated statistics distribution varies widely according to the original structure of trading signal. Indeed, comparing a strategy which is approximately in and out of the market 50% of the time with simulated series being in the market much more often would clearly biased the results[37].

---

[37] Indeed, comparing the (2, 10) strategies with simulations based on the Opt_4 strategy would be problematic. As simulations have more long positions than the (2, 10) strategy (2871 days for the former and only 2329 for the latter), it would be almost impossible for the (2, 10) strategy to have a higher percentage of long positions during bull markets, even if it has good timing abilities.

Table 9: Rules positions during Bull and Bear markets

Note: The first five columns refer to the best in-sample rule and our four complex trading systems, while the last four correspond to standard simple MA rules. The first (second) number in parenthesis is the length of the short (long) MA. No bandwidth is used. The statistics in panel A describe the distribution of the 500 simulated series constructed with the corresponding trading signals. The *p*-values in panel B are the percentage of simulated series that have a statistic higher than the original.

Panel A: Block Bootstrap estimation (block length of 40 days)

	Investment strategies								
	Best	Opt_all	Opt_4	Voting	Partial	(2,10)	(10,50)	(5,100)	(50,200)
% buy-Bull									
median	77.4	77.6	76.5	78.0	77.8	57.2	62.2	64.6	69.4
min	58.4	62.9	61.9	61.4	62.9	51.6	44.1	52.5	52.5
max	91.8	91.8	89.4	90.0	92.1	62.3	72.5	77.1	83.7
% sell-Bear									
median	23.1	22.4	23.3	21.5	22.2	42.6	37.9	34.9	29.0
min	2.3	1.6	0.0	3.2	0.0	31.7	20.8	9.9	7.5
max	52.0	52.8	51.5	48.0	49.9	51.9	62.0	56.7	52.1
% right									
median	63.8	63.8	63.4	64.0	64.0	53.6	56.4	57.4	59.6
min	47.4	50.3	50.4	48.9	49.1	48.0	39.2	44.8	46.4
max	77.9	80.7	74.1	75.1	75.9	58.2	65.8	67.4	73.4
% wrong									
median	36.2	36.2	36.6	36.0	36.0	46.4	43.6	42.6	40.4
min	22.1	19.3	25.9	24.9	24.1	41.8	34.2	32.6	26.6
max	52.6	49.7	49.6	51.1	50.9	52.0	60.8	55.2	53.6

Panel B: Statistics of strategies

	Best	Opt_all	Opt_4	Voting	Partial	(2,10)	(10,50)	(5,100)	(50,200)
Nb buy	2892	2906	2871	2923	2923	2329	2323	2426	2632
Nb sell	868	853	889	837	837	1431	1437	1334	1128
% buy-Bull	92.4	90.2	93.0	91.8	91.8	62.2	71.0	76.7	85.9
p-value	*0*	*0*	*0*	*0*	*0*	*0.01*	*0.01*	*0.002*	*0*
% sell-Bear	70.2	61.9	74.0	65.0	65.0	57.9	66.1	72.4	78.1
p-value	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
% right	86.9	83.2	88.3	85.2	85.2	61.1	69.8	75.6	83.9
p-value	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*	*0*
% wrong	13.1	16.8	11.7	14.8	14.8	38.9	30.2	24.4	16.1
p-value	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*	*1*

The results in Panel B provide strong evidence that our complex rules exploit long-term trends in the market cycle to provide their excess returns. Each complex rule is long during more than 90% of bull markets days. The results are somewhat weaker during the bear markets, but the percentages, between 61% and 74%, are much higher than what it could be achieved by luck, as proved by p-values of 0. Furthermore, the total percentage of trading signals corresponding to the various market phases ranges between 83% and 88%. This is higher than the 75% achieved by the buy-and-hold. For instance, the Opt_all rule has long (short) positions during 90.2% (61.9%) of the days the when market rises (declines). The two associated p-values are zero, which means that such high percentages are never obtained with the simulated signals series based on the original one. Thus, these results support the timing abilities of complex rules with respect to long-term market trends, as these statistics cannot be re produced by luck. The results for simple MA rules are also significant; however, percentages levels indicate that complex strategies follow more closely and accurately long-term markets trends. In addition, rules based on very short trends have lower percentages than the buy-and-hold.

The distribution of the simulated statistics presented in Panel A may also allow comparing the standard simple MA rules with the complex systems. For instance, the median of the simulated % right statistic based on the Opt_4 rule is 63.4%, while those of the original MA (2,10) is only 61.1%. This means that this latter rule trading positions coincide with the market phases to a lesser extent than the simulated series obtained according to the Opt_4 rule. We can indirectly reject the null hypothesis that the MA(2, 10) rule has an equal market timing ability to the Opt_4 rule. Indeed the MA (2, 10) original statistic is 61.1%, and thus, is lower than the median of the simulated statistics based on the Opt_4 rule, while the original Opt_4 rule statistic is higher than all the simulated series based on it. This proves statistically that the Opt_4 rule follows the market phases more closely than the standard MA rules specifications.

One may argue that the aforementioned results are biased as the block length used in the bootstrap process does not correspond to the trends. Indeed, the complex rules keep their trading positions for much longer than our 40 days blocks. To mitigate this issue, Table 10 presents similar tests with a block length of 470 days. The main difference in simulated statistics distributions lies in the range of values. Indeed, the minimum (maximum) values are systematically lower (higher) than those obtained with a 40 days block. Nonetheless, the medians are only marginally influenced by the change of the length. As the new simulated series

correspond more closely to the originals, the *p*-values associated with the first statistic rise[38]. Nonetheless, the evidence in favour of market timing abilities is not invalidated, as the *p*-values of the last two statistics, which describe the whole strategy (both long and short positions), remain very significant.

Finally, to illustrate these results, Figure 10 shows the S&P500 evolution and the position taken by the Opt_4 rule. The segments in blue (red) correspond to long (short) positions taken by the Opt_4 rule. It is unquestionable that the rule identifies the trends reversals with a lag, but as these trends are long-term trends, the rule is still able to profit from them.

Figure 10: Opt_4 positions and the market phases

Note: This graph represents the evolution of the S&P500 prices. The segments in blue (red) correspond to long (short) positions taken by the Opt_4 rule.

---
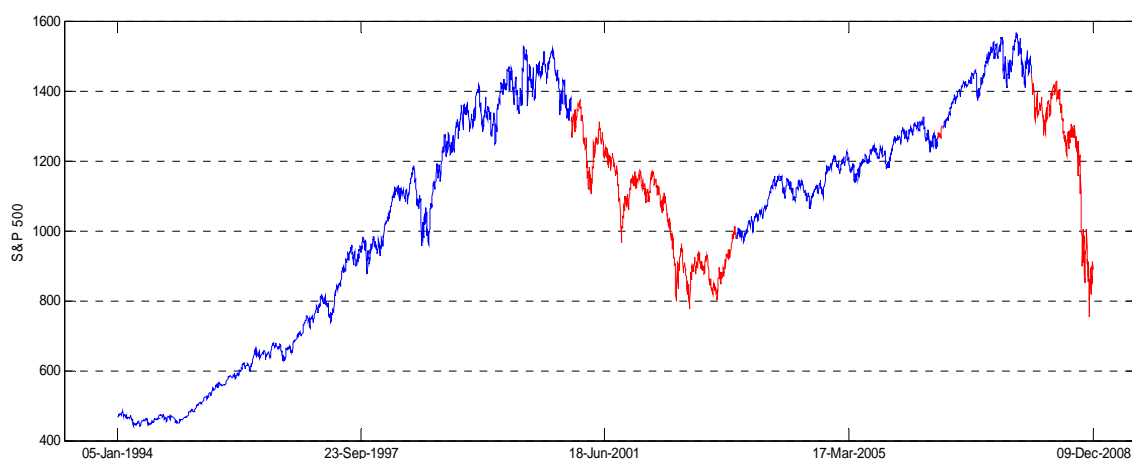
[38] As the block bootstrap is performed with replacement, a few of the simulated trading signal series consist only in long positions. However, for these series, the overall performance is not positively biased as they have no short position in bear markets. By investigating every simulated statistics, we find a negative correlation coefficient between the first 2 statistics. Reading *p*-value for the whole strategy (the last two) provide a wider picture and thus mitigate this issue. For very long length of the block (such as the 470 days), a permutation method (i.e. without replacement) might be more appropriate. However, we doubt that such a method would have strong impacts on our results as the *p*-values for the overall strategy are systematically very close to zero.

Table 10: Rules positions during Bull and Bear markets

Note: The notation is similar to Table 9. The length of the block is 470 days and the number of simulations is 1,000.

Panel A: Block Bootstrap estimation (block length of 470 days)

	Investment strategies								
	Best	Opt_all	Opt_4	Voting	Partial	(2,10)	(10,50)	(5,100)	(50,200)
% buy-Bull									
median	78.4	78.7	77.6	78.3	78.1	57.8	62.5	65.3	71.0
min	22.8	21.1	28.5	31.4	22.8	46.4	40.7	39.0	23.8
max	100.0	100.0	99.3	100.0	100.0	65.6	78.9	86.1	95.2
% sell-Bear									
median	21.5	17.5	23.6	22.7	21.8	43.2	38.5	37.7	31.3
min	0.0	0.0	0.0	0.0	0.0	31.8	14.7	10.2	0.0
max	89.8	85.8	100.0	100.0	100.0	59.5	70.9	76.1	100.0
% right									
median	64.7	64.6	64.4	65.1	64.5	54.1	56.6	58.4	60.9
min	30.6	22.9	30.1	30.3	20.9	45.9	38.5	36.4	25.3
max	86.4	86.4	88.8	86.5	88.9	60.4	71.6	79.1	82.7
% wrong									
median	35.3	35.4	35.6	34.9	35.5	45.9	43.4	41.6	39.1
min	13.6	13.6	11.3	13.5	11.1	39.6	28.4	20.9	17.3
max	69.4	77.1	69.9	69.7	79.1	54.1	61.5	63.6	74.7

Panel B: Statistics of strategies

	Best	Opt_all	Opt_4	Voting	Partial	(2,10)	(10,50)	(5,100)	(50,200)
Nb buy	2892	2906	2871	2923	2923	2151	2323	2426	2632
Nb sell	868	853	889	837	837	1609	1437	1334	1128
% buy-Bull	92.4	90.2	93.0	91.8	91.8	62.2	71.0	76.7	85.9
p value	*0.111*	*0.145*	*0.091*	*0.123*	*0.137*	*0.055*	*0.095*	*0.062*	*0.066*
% sell-Bear	70.2	61.9	74.0	65.0	65.0	57.9	66.1	72.4	78.1
p value	*0.015*	*0.042*	*0.007*	*0.045*	*0.045*	*0.012*	*0.011*	*0.004*	*0.005*
% right	86.9	83.2	88.3	85.2	85.2	61.1	69.8	75.6	83.9
p value	*0*	*0.009*	*0.001*	*0.002*	*0.001*	*0*	*0.005*	*0.001*	*0*
% wrong	13.1	16.8	11.7	14.8	14.8	38.9	30.2	24.4	16.1
p value	*1*	*0.991*	*0.999*	*0.998*	*0.999*	*1*	*0.995*	*0.999*	*1*

7.   Sign prediction and profitability

The last Section of the first Part aims to shed some light on the relationship between the percentage of right signals and the profitability. Indeed, the results presented in Section 4.2.3 show that even a small increase in this percentage, compared with the buy-and-hold strategy for instance, can have a strong impact on the strategy profitability. For example, the Opt_all strategy has a percentage of right signals of 53.4%, which is only 0.3% higher than for the buy-and-hold, but it generates a mean return of 10.7%, while those of the buy-and-hold is 6.16% only. In addition, the Opt_4 strategy percentage of right signals is 54%, and it is associated with a 14.61% mean return. Thus, we propose a simulation to determine whether the strategies high performances, while having only a slightly higher predictability, are due to luck or not.

7.1   Intuitions and methodology

The simulation consists in generating artificial trading positions, or strategies, with a specific percentage of right signals[39]. This enables us to examine the relation between these percentages and the strategies performance. Or in other words, we create a large number of artificial strategies with a similar average level of predictability to our complex trading rules, and then, we compare the mean performance of these artificial strategies with the original ones. The percentage of right signals can be used to asses the forecasting ability of a trading rule as it does not depend on returns, but solely whether the strategy takes the right position at the right time.

The procedure, which is illustrated in Figure 11, is the following;

1) First, we divide the 3757 S&P 500 returns that are not equal to zero from 1994 to 2008 according to their daily returns sign[40]. Thus, we obtain a series of 1998 ($Nb+$) positive returns, R^+, and 1759 ($Nb-$) negative returns, R^-.

2) Then we construct the artificial strategies, or trading positions series, with a predetermined level of right signal. The first step is to resample randomly, without replacement, the original

---

[39] Note that the way we presented the percentage of right signal for the overall strategy in chapter 4 is equal to computing the percentage of positive returns. We use this latter method in this chapter to ease the computation and as we calculate only this statistic for the overall strategy and not for buy and sell signals separately.

[40] This ensures that the proportion of right signal is evenly distributed between positive and negative returns.

positive and negative return series, R^+ and R^- to get new returns series. Thus, the distribution is similar to the original, only the ranking changes;

$$R^+_{resampled} = \begin{bmatrix} R_1^{+,boot} \\ \vdots \\ \vdots \\ R_{Nb+}^{+,boot} \end{bmatrix} \qquad R^-_{resampled} = \begin{bmatrix} R_1^{-,boot} \\ \vdots \\ \vdots \\ R_{Nb-}^{-,boot} \end{bmatrix}, \qquad (1.31)$$

where the elements of the $R^+_{resampled}$ vector contains the same values as R^+ with a different ranking, and thus, we add the *boot* extension in the notation.

.3) Then, we constructed the two sides of the strategies directly by multiplying the series obtained in the second point by a vector containing ones and minus ones with a proportion corresponding to the predetermined level of predictability. Thus, for the side of the strategy with positive (negative) returns, a right signal is a (minus) one. This vector is the simulated trading positions, which allow us to compute directly the strategy returns series as

$$R^+_{artificial} = \begin{bmatrix} R_1^{+,boot} \\ \vdots \\ R_{P*Nb+}^{+,boot} \\ \vdots \\ R_{Nb+}^{+,boot} \end{bmatrix} \left\{ \begin{bmatrix} \mathbf{1}_{1,P*Nb+} \end{bmatrix}, \begin{bmatrix} -\mathbf{1}_{1,(Nb+)-(P*Nb+)} \end{bmatrix} \right\} \qquad (1.32)$$

$$R^-_{artificial} = \begin{bmatrix} R_1^{-,boot} \\ \vdots \\ R_{P*Nb-}^{-,boot} \\ \vdots \\ R_{Nb-}^{-,boot} \end{bmatrix} \left\{ \begin{bmatrix} \mathbf{1}_{1,P*Nb-} \end{bmatrix}, \begin{bmatrix} -\mathbf{1}_{1,(Nb-)-(P*Nb-)} \end{bmatrix} \right\} \qquad (1.33)$$

$$R_{artificial} = \begin{bmatrix} R^+_{artificial} \\ R^-_{artificial} \end{bmatrix}, \qquad (1.34)$$

where $R^+_{artificial}$ and $R^-_{artificial}$ are the two sides of the artificial strategy, which do not contain only positive and negative return anymore, as they have been multiplied by the trading positions with the predetermined level of predictability. $R_{artificial}$ is the entire artificial strategy obtained by adding its two sides. P is the level predictability stated in percentage. Note that $P \cdot Nb^+$ is rounded to the

nearest integer. For example, with *P* equal to 60% this value is *round (0.6*1,998) = 1,199*. Thus, the artificial trading positions vector is composed of 1,199 ones and 799 minus ones. Finally, $\mathbf{1}_{1, P \cdot Nb+}$ and $-\mathbf{1}_{1,, P \cdot Nb+}$ are vectors containing either ones or minus ones, or in our example, 1,199 ones and 799 minus ones.

4) Compute the annual mean return of the artificial strategy by multiplying its simple mean by 252 and the *t*-statistic under the null hypothesis that the artificial mean return is similar to the original buy-and-hold mean return, as detailed in appendix I-A. They should provide some insights about the link between the economic and statistic performance.

5) Repeat the points 3) and 4) for each of the 142 levels of predictability, or right signals, *P*. They range from 50% to 100% with an incremental step that is smaller for levels up to 60%[41]. Thus, we generate 142 artificial annual mean returns with the same original return series resampled in point 2).

6) Repeat the points 2) to 5) 5,000 times to obtain 5,000 artificial annual mean returns for each level of predictability, resulting in a total of 710,000 artificial strategies.

7) These artificial returns can be used to compute empirical *p*-values for the following test; The high returns and only slightly superior levels of predictability generated by our complex strategies is only due to luck and the predictability level should be higher in order to justify such high returns.

---

[41] Precisely, this step is 0.001 up to 60% and then 0.01.

Figure 11: Predictability and profitability: an illustration of the procedure

Note: This figure illustrates the points 1 to 4 of the simulation presented above in this Section with an illustrative market series, R, comprising 10 observations and a predictability level of 60.% Note that in this simple example, percentage of positive artificial returns is far from 60% as the number of observation is very limited. This is not an issue with our sample of 3757 observations.

$$
R = \begin{bmatrix} 0.002 \\ -0.0015 \\ -0.002 \\ 0.001 \\ 0.0035 \\ -0.0025 \\ 0.003 \\ -0.001 \\ 0.0015 \\ 0.0025 \end{bmatrix} \xrightarrow{\text{Step 1}} R^{+} = \begin{bmatrix} 0.002 \\ 0.001 \\ 0.0035 \\ 0.003 \\ 0.0015 \\ 0.0025 \end{bmatrix} \text{ and } R^{-} = \begin{bmatrix} -0.0015 \\ -0.002 \\ -0.0025 \\ -0.001 \end{bmatrix}
$$

$$
\xrightarrow{\text{Step 2}} R^{+}_{resempled} = \begin{bmatrix} 0.0025 \\ 0.0015 \\ 0.003 \\ 0.0035 \\ 0.001 \\ 0.002 \end{bmatrix} \text{ and } R^{-}_{resempled} = \begin{bmatrix} -0.001 \\ -0.002 \\ -0.0025 \\ -0.0015 \end{bmatrix} \xrightarrow{\text{Step 3}} \begin{bmatrix} Nb+{*}P = 4 \\ Nb-{*}P = 2 \end{bmatrix}
$$

$$
R^{+}_{artificial} = \begin{bmatrix} 0.002 \\ 0.001 \\ 0.0035 \\ 0.003 \\ 0.0015 \\ 0.0025 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}' = \begin{bmatrix} 0.002 \\ 0.001 \\ 0.0035 \\ 0.003 \\ -0.0015 \\ -0.0025 \end{bmatrix} \text{ and } R^{-}_{artificial} = \begin{bmatrix} -0.0015 \\ -0.002 \\ -0.0025 \\ -0.001 \end{bmatrix} \cdot \begin{bmatrix} -1 \\ -1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.0015 \\ 0.002 \\ -0.0025 \\ -0.001 \end{bmatrix}
$$

$$
R_{artificial} = \begin{bmatrix} R^{+}_{artificial} \\ R^{-}_{artificial} \end{bmatrix} \quad \text{with an annual mean return of 13.86\%}
$$

7.2   Results

The Panels A and B of Figure 12 show that even a small increase in the percentage of right signal is sufficient to generate very high excess returns. Indeed, a level of 55% generates, on average, annual mean excess returns of 15%. As a reference, the percentage of right signal for the buy-and-hold during the 1994-2008 sample is 53.12%[42]. On average, we observe an increase of 4.1% in mean annual excess return for an increase of 1% in the proportion of right signals. These results suggest that the performance of our strategies is not due to luck as they are in line with those obtained in this simulation.

The last two panels examine whether these excess returns are statistically significant from zero with Student *t*-tests. Panel C shows that percentages that are only slightly higher than the buy-and-hold are sufficient to generate statistically significant excess returns. However, they are also linked with very high excess return, 15% on average.

Panel D provides a different picture of the same question; 95% of the 5000 tests would result in a significant difference when the percentage of right signals is 56.4%. Nevertheless, this level corresponds, on average, to annual mean excess of 20.1%. In comparison the annual mean buy-and-hold return is 6.16%. This simulation experiment suggests that the Student *t*-test is not a powerful testing procedure, at least with returns series. These findings are in line with Xu (2004), who also shows that a small increase in predictability can lead to large economical gains. This link between excess returns and the statistical tests results is extensively examined in the third Part of this thesis.

---

[42] This level corresponds to the percentage of positive daily return over the sample.

Figure 12: Percentage of right signals: a simulation

Note: This figure presents the results of the simulation study between the percentage of right signals and mean returns in excess of the buy-and-hold. The first two panels display the mean and the 5% and 95% deciles of the distribution of the simulated annual mean excess returns. The third panel summarizes the relation with mean *t*-statistics obtained with a standard bilateral Student *t*-test of equal means. Panel D provides the percentage of these tests that result in a significant difference between mean returns at a 5% confidence level. For each level of right signals, 5,000 simulations are conducted.

Panel C

Panel D
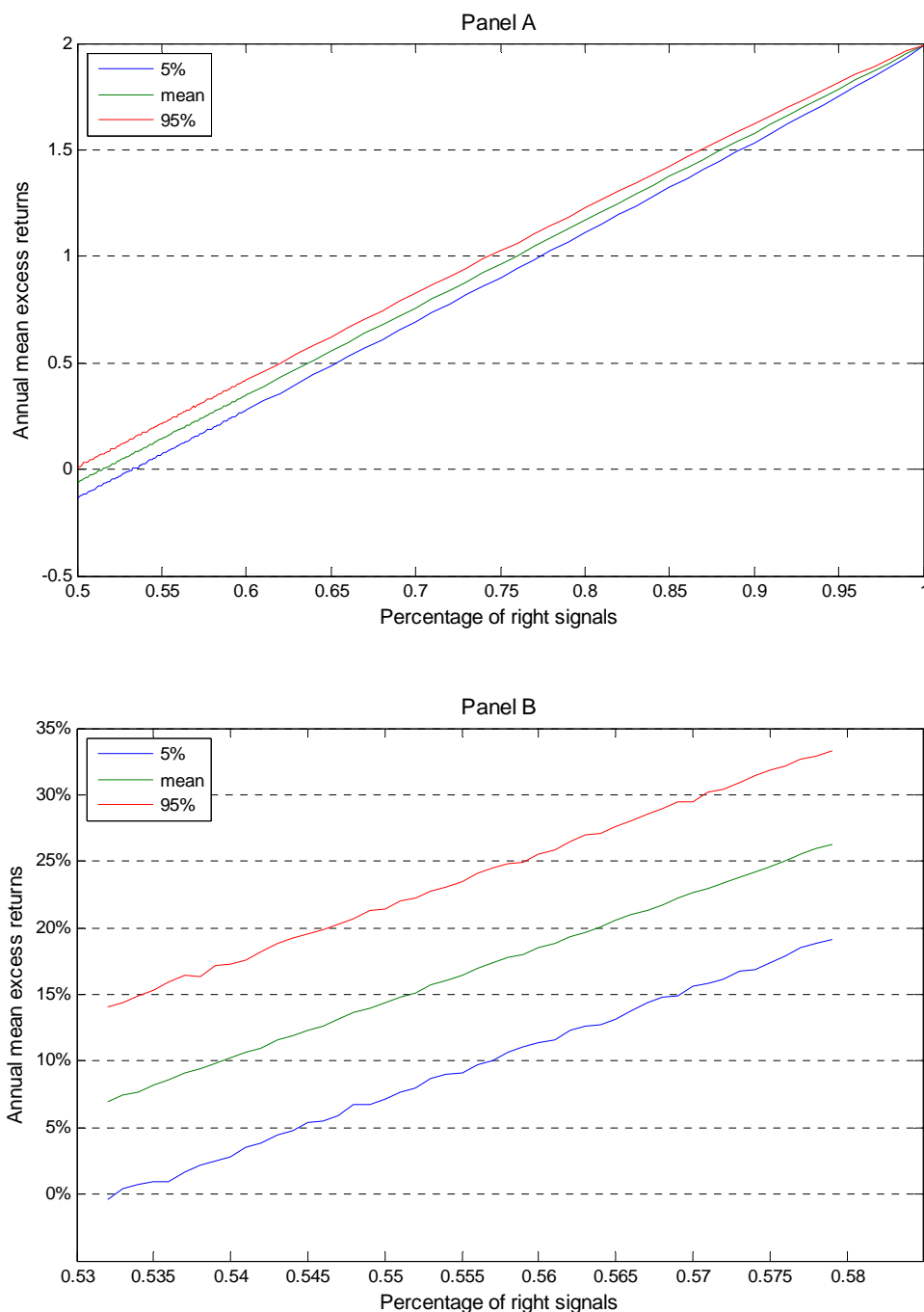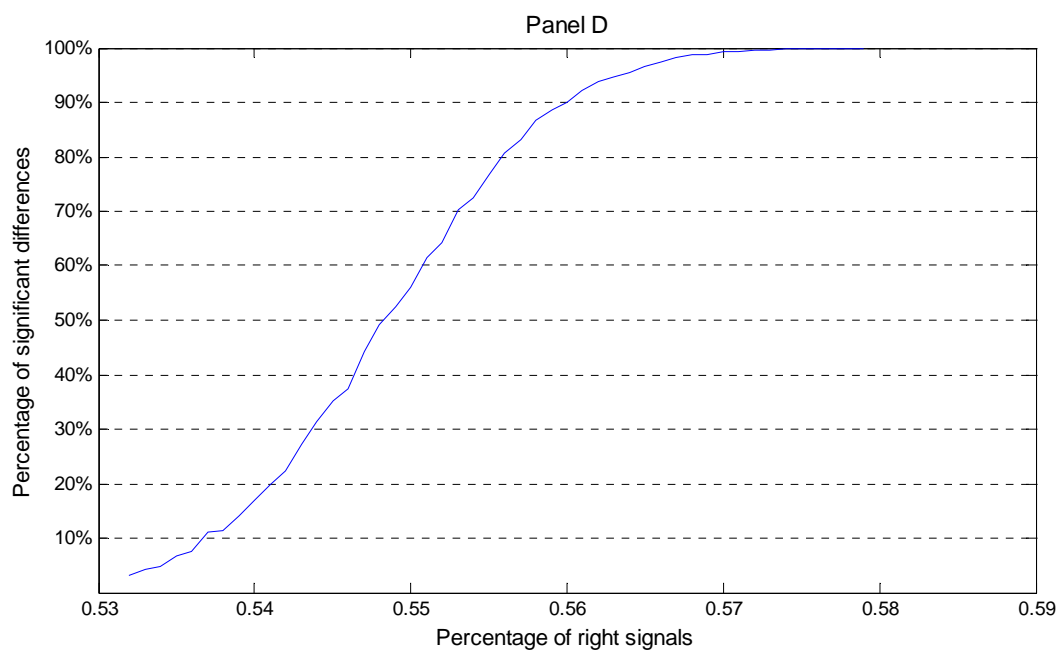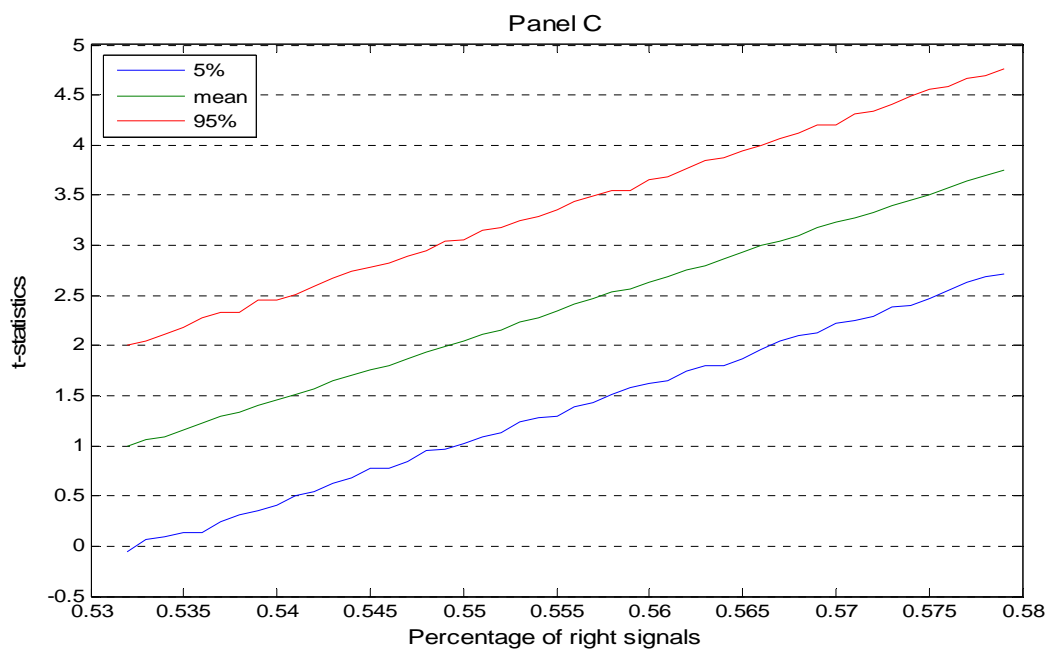
8.   Conclusion

No conclusive agreement has been reached in the academic literature about the usefulness of technical analysis as an efficient predictive tool. However, it is obvious that these methods are widely used in practice. This first Part adds new insights about the ongoing debate by examining three related issues. First, we examine trading systems based on MA rules that are not restricted to forecast short-term trends, as it is usually the case. Indeed, we consider a much wider range of parameters that may also be used to exploit long-term trends. Based on recent studies, we use various specifications of complex trading strategies in order to use more information, and to mitigate the data-snooping issue. Then, we examine the use of financial leverage in the trading strategies, as out performing a market that follows a strong upward trend would require a very high level of predictability. To this purpose, we evaluate the performance of standard strategies combined with debt leverage and with exchange-traded options. Finally, we propose a new market timing test in order to determine whether these new trading strategies tend to follow long-term trends in the market. We would like to highlight the following findings:

1. Designing MA rules to exploit long-term trends improves dramatically their performance, especially over the most recent subsample. This contracts the majority of studies that find that technical analysis has lost its predictive power recently.

2. The complex trading rules produce a total compounded return ranging from 274% to 572%, while the market yields only 90%. These trading systems rely on rules with a long MA length higher than the 200 or 250 days used in other studies.

3. These returns are not due to additional risk bearing, and transaction cost is not an issue, as the complex rules switch position rarely. Furthermore, this lack of frequent trading also reduces the probability that the strategies face microstructure problems, such as non synchronous trading. Thus, they may challenge the EMH.

4. However, the strategies mean returns are not significantly different from the buy-and-hold, but we argue that this is due to the low power of the Student t-test procedure.

5. The systems are free from any look-ahead bias.

6. The formal tests show that long and short positions coincide strongly with bull and bear market phases, to an extent that can not be replicated by luck. Thus, they possess significant market timing abilities.

7. These abilities justify the use of leverage. However, using debt is more valuable than exchange-traded options. The latter induce a lost of time-value and increase the volatility dramatically.

8. The strategies can reduce skewness risk without reducing the returns.

9. A simulation experiment shows that a small increase in the level of right signal induces in large increase in mean returns. In addition, very large excess returns are required to reject the null hypothesis of equal means with Student t-tests.

In the third Part of this thesis, we further investigate the link between the economic and statistical significance of the strategies excess returns. We propose a new testing procedure, based on returns to assess the economic performance of the strategies. In order to keep their structure, we generate artificial signal series with Markov-chains. We show that this test is more powerful than commonly used procedures, and the strategies returns in excess of the buy-and-hold are significant.

As the strategies investigated in the first Part are successful, the next one consists in applying some of their characteristics to optimized mean-variance portfolios.

9.   Appendices Part I

Appendix I-A: A statistical reminder: the Student *t*-test and *p*-values

<u>Student *t*-test</u>

A Student *t*-test is statistical hypothesis testing procedure that investigates whether a null hypothesis, *H0*, should be accepted against an alternative one, *H1*. The procedure consists of a test statistic, *t*, and a confidence interval, also named the critical region that is computed with a predetermined confidence level, 5% being the most commonly used. If the test statistic is not included in this interval, the null hypothesis is rejected in favour of the alternative. In other words, it is very unlikely that the random variable under investigation is issued from the distribution stated in the null hypothesis. Note that in the opposite case, the null is not accepted, but the test is unable the reject the null.

In this statistical reminder, we present the Student *t*-test in the following context:

1) As this thesis is mainly empirical, we describe this test when all distributions moments, such as means and variances, have to be empirically estimated from a data sample.

2) All tests are two-tailed tests. This implies that the alternative hypothesis has no direction and is constructed with a "$\neq$" sign, and neither with ">" nor "<". In other words, the test is about if a random variable is equal to zero (H0) or not (H1) and not if it is higher than zero (H1 of a one-sided test). Note that the null is always stated as an equality in order to perform the test.

3) In order to give intuitive examples instead of a theoretical explanation, I restrict this presentation to the three main uses in this thesis; to test whether a random variable (a regression coefficient or the mean of a returns series) differs statistically from a value specified in the null hypothesis or if two random variables are issued from the same distribution, and thus, are not statistically different.

The general form of the test statistic, *t*, is

$$t = \frac{A}{s_A},\tag{1.35}$$

where A includes the random variables of interest minus the value defined in the null hypothesis, S_A is it sample standard deviation. *A* follows a normal distribution, but as its standard deviation has to be estimated, the resulting test statistic is distributed as a Student *t*-distribution with a

number of degrees of freedom depending on the number of observations used to estimate the random variable. Note that if this number is large, the distribution tends towards a standardized normal distribution, with a zero mean and a variance of one.

The following figure provides an illustration, when the number of observation is large, and thus, the test statistic distribution is approximated by a normal distribution with a zero mean and variance of one.

Figure 13: Hypothesis test
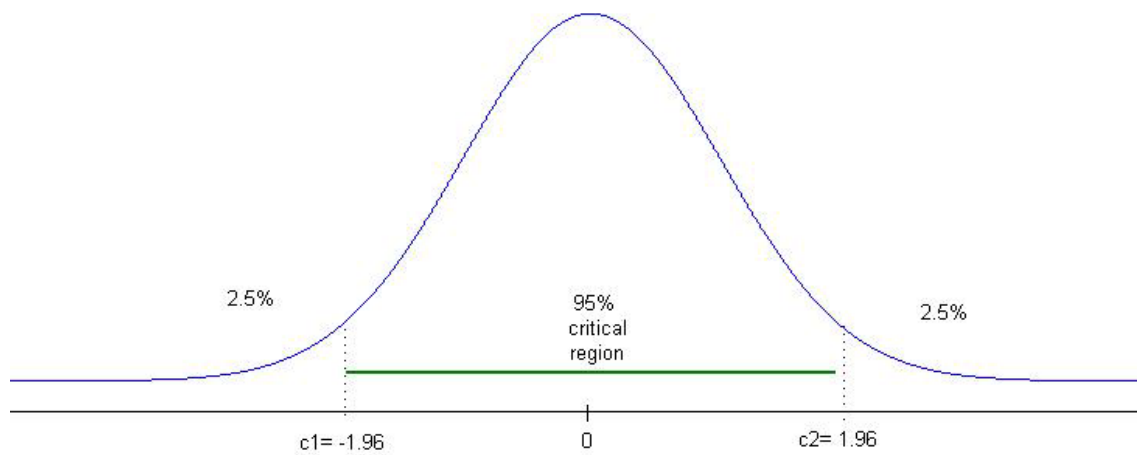
This figure displays a standardized normal distribution with the two critical values, $c1$ and $c2$, that delimit the critical region between -1.96 and 1.96 for a 95% confidence level. This means that if a test statistic lies within these bounds, the null hypothesis is not rejected. The test statistic is standardized in order to have always the same critical values with the same confidence level.

Thereafter, I detail the test for its three main uses in this thesis;

1) Regression coefficient: In this case, the objective is to determine if a regression coefficient is statistically different from zero, and thus, whether the associated explanatory variable Y influences the dependant variable X;

$$X = \alpha + \beta Y + e$$
$$\text{H0}: \beta = 0 \quad \text{H1}: \beta \neq 0 \tag{1.36}$$
$$t = \left(\frac{\beta - 0}{\sigma_\beta} \right) \sim T^{n-2},$$

where the above equation first line is the regression equation with β as the coefficient of interest, the second line displays the two hypothesis and the last one is the test statistic where σ_β is the coefficient estimated standard deviation, which depends on the method used to estimate the regression. The test statistic is distributed as a Student t-distribution with a number of degrees of freedom equal to the number of observations used in the regression minus two, as there are two coefficients in this example. The distribution is denoted T^{n-2} in the equation.

2) Mean return: The second case investigates if a return series mean is statistically different from zero, or any other value;

$$\bar{\mu} = \frac{1}{n}\sum_{t=1}^{n} R_t$$

$$H0: \bar{\mu} = 0 \quad H1: \bar{\mu} \neq 0 \tag{1.37}$$

$$t = \left(\frac{\bar{\mu}}{s/\sqrt{n}}\right) \sim T^{n-2},$$

where R_t is the return at time t, s is the unbiased estimation of the return series standard deviation and n the number of observations.

3) Difference in mean returns. Finally, this test determines if two means returns are statistically different from each other.

$$\bar{\mu}_1 = \frac{1}{n}\sum_{t=1}^{n} R_{1,t} \qquad \bar{\mu}_2 = \frac{1}{n}\sum_{t=1}^{n} R_{2,t}$$

$$H0: \bar{\mu}_1 = \bar{\mu}_2 \quad H1: \bar{\mu}_1 \neq \bar{\mu}_2 \tag{1.38}$$

$$t = \left(\frac{\bar{\mu}_1 - \bar{\mu}_2}{\sqrt{\dfrac{s_1^2 + s_2^2}{n}}}\right) \sim T^{n-2},$$

where R_t is a the return at time t, s is the unbiased estimation of the return series standard deviation and n the number of observations. Here, we assume that the covariance between the two series is equal to zero.

P-values

A drawback of the previously explained test procedure is that the critical values have to be re-computed for different confidence levels. A way to cope with this issue is to use *p*-values to determine if the null is rejected or not. They are defined as the probability to have a value at least as large as the test statistic in absolute value (for two-tailed tests) under the null hypothesis, and they are computed as

$$p \text{ value} = 2\big(1 - student\text{'s } t \ cdf\big(|tstat|, df\big)\big) \tag{1.39}$$

where *student's t cdf* is the cumulative distribution function of the *T* distribution with *df* degrees of freedom and *tstat* is the test statistic. For example, if a test statistic is associated with a *p*-value of 0.075, the null is rejected for a 10% confidence level, but not if this level is reduced to 5%. Actually, it is rejected for confidence levels higher than 7.5% for a large enough sample.

Note that in this thesis, we often use simulation methods to conduct statistical tests that results in simulated *p*-values. While their computation is obviously different than those Student *t*-test *p*-values, their interpretation is similar.

Appendix I-B: An overview of popular trading rules

Here, we present a short description of the most widely used trading rules. Many others can be found in practitioners books such as Achelis (2000) for technical indicators or Edward and Magee (1996) for an in-depth description of charting.

Moving averages rules

This rule uses two main parameters, a short and a long-term MA to smooth noisy price series in order to detect market trends and to generate trading signals. They are defined as

$$M_{t,S} = \frac{1}{S} \sum_{j=1}^{S} P_{t-S+j} \quad \text{and} \quad M_{t,L} = \frac{1}{L} \sum_{j=1}^{L} P_{t-L+j}, \tag{1.40}$$

where S and L are respectively the length of the short and the long MA. The series will be smoother as the length of the MA window increases. $S \leq 50$ and $L \leq 250$ are usually chosen as the maximum length of the short and long window. Then, the relative difference between these two MA is defined as

$$D_t = \frac{M_{t,S} - M_{t,L}}{M_{t,L}}. \tag{1.41}$$

Buy (sell) signals are generated when the short-term MA is above (below) the long-term MA, that is to say when D_t is positive (negative). The underlying intuition is that when the short MA crosses the long MA from below (above) an upward (downward) trend is initiated. An improvement of this trading rule is to introduce a security band to avoid non informative signals, i.e. when the two MA are very close from each other and consequently no clear trend can be detected. Defining the bandwidth as B (1% is generally chosen in practical implementations), the various signals are generated in the following way: A buy signal is generated whether $Dt > B$, a sell signal if $D_t < -B$ and a neutral signal otherwise, that is when $-B \leq D_t \geq B$.

The trading range break-out rule

The trading range break-out (TRB) is based on support and resistance levels. The intuition is that investors want to sell their stocks at a peak price, and thus, there is a resistance at the level of the previous peak price. However if the price crosses this level, an upward trend is initiated. According to this rule, a buy signal is generated when the price crosses from below a resistance

level, defined as a local maximum over a certain time period, and a sell signal is emitted when the price drops below the support level, which is the local minimum. Mathematically, the resistance and support levels are defined as

$$res_t(m) = \max\left(P_{t-1}, ..., P_{t-m}\right)$$
$$sup_t(m) = \min\left(P_{t-1}, ..., P_{t-m}\right),$$

(1.42)

A buy signal is generated whether $P_t > res_t(m)$ and a sell signal if $P_t < sup_t(m)$. A security band, to avoid noisy signals, could also be considered. Usually, the position is kept for a fixed period, such as 10 days, once the signal is emitted, as this rule emphasises the crossing of the price and the support/resistance levels.

The filter rule

This rule is supposed to detect trend reversals. A buy (sell) signal is generated when the closing price increases by x % above (below) its most recent low (high). Then, for a trend beginning at time s (when a long or short position has been taken), the high and low are defined as

$$m_{t-1} = (1 - x)\max\left(P_s, ..., P_{t-2}, P_{t-1}\right)$$
$$M_{t-1} = (1 + x)\min\left(P_s, ..., P_{t-2}, P_{t-1}\right),$$

(1.43)

where x is the filter, usually around 5% or 10%. s and x are the two parameters to define. Expecting a change in trend, trading signals are generated according to the following rules: If day t is a buy, the long position is kept as long as the price has not declined by at least x% from a subsequent high. Mathematically, the trend reversal point is reached whether $P_t < m_{t-1}$. At this point, the long position is liquidated and a short one is initiated and it is kept until the price increases by at least x% above a previous low, that is to say, when $P_t > M_{t-1}$. At this point, a new long position is opened. To avoid noisy signals, a security band may be introduced.

The channel rule

In the same way as the filter rule, the channel rule also aims at detecting trend reversal points. These two rules share some similarities, however, they use different parameters and trend reversal points. First, the channel rule is not defined according to a filter parameter, and second, the period upon which the low and high prices are computed is fixed, and they are defined as

$$m_{t\text{-}1} = \min\left(P_{t-L}, ..., P_{t\text{-}2}, P_{t\text{-}1}\right)$$
$$M_{t\text{-}1} = \max\left(P_{t-L}, ..., P_{t\text{-}2}, P_{t\text{-}1}\right).$$

(1.44)

L is the only parameter to define and it represents the period of time over which the minimum and maximum prices are computed. This rule predicts that an upward (downward) trend will reverse when the latest price is lower (higher) than the minimum (maximum) price over the L previous prices. Mathematically, the long position is reversed whether $P_t < m_{t\text{-}1}$ and in the same manner, the short position is switch to a long one when $P_t > M_{t\text{-}1}$. It is also possible to introduce a bandwidth to avoid noisy signals.

The relative strength index

The relative strength index (RSI) is one of the most popular oscillators. These tools are supposed to detect trend reversals, in the same manner as the filter rules, by defining whether an asset is overbought or oversold. The index is obtained by comparing the ratio of up-closes (U_t) and down-closes prices (D_t) during a predetermined period of d days and they are defined as

$$U_t = \begin{cases} P_t - P_{t-1} & \text{if} \quad P_t > P_{t-1} \\ 0 & \text{otherwise} \end{cases}$$

(1.45)

$$D_t = \begin{cases} P_{t-1} - P_t & \text{if} \quad P_{t-1} > P_t \\ 0 & \text{otherwise} \end{cases}.$$

(1.46)

Then, the RSI is computed as

$$RSI_t = 100 - \frac{100}{1 + RS_t} \quad \text{where} \quad RS_t = \frac{\sum_{n=t}^{t-d} U_n}{\sum_{n=t}^{t-d} D_n}.$$

(1.47)

The RSI can get values from zero, which indicates a pure downward trend, to 100, which indicates a pure upward trend in prices. Buy and sell signals could be generated according to various rules. Perhaps the most widely used method is to define an upper and a lower bound. When the RSI touches the upper bound (usually set at 70), the market (or the individual asset) is considered to be overbought and thus, a sell signal is generated. On the contrary, when the RSI touches the lowest bound (usually set a 30), the market is considered to be oversold and a buy signal is emitted.

The Advance/Decline Ratio and the Short-Term Trading index

These two indicators consider the behaviour of individual securities in a market to detect trends in the whole group formed by them, such as an index. The first one, the Advance/Decline Ratio, is the proportion of advancing security prices during a period of time, such as a trading day. The ratio is defined as

$$ADR_t = \frac{ADC_t}{\left(ADC_t + DCL_t\right)},$$ (1.48)

where ADC_t and DCL_t represent the number of advancing and declining security prices during the interval t (the interval may be a trading day, from close to close). A ratio higher than 0.5 means that rising securities outnumber the falling securities and this indicates a rising trend.

The second ratio, the Short-Term Trading Index, considers the volume of the rising stocks proportional to the volume of all stocks in the market and is defined as

$$STTI_t = \frac{UPVOL_t}{\left(UPVOL_t + DNVOL_t\right)},$$ (1.49)

where $UPVOL_t$ and $DNVOL_t$ are respectively the total volume of all rising and declining stocks during the trading interval t. It is worth noting that these two indicators can generate buy and sell signals according different values of the ratio, which have to be chosen subjectively.

Appendix I-C: Summary of studies finding profitable technical trading strategies

Authors	Asset	Sample	Adjustments	Trading rules	Test methodology	Results
Pruitt and White (1988)	Individual US stocks	1976-1985	Transaction costs (0.5 or 1%) and risk	CRISMA trading system	Standard tests	Excess returns obtained by the trading system are statistically and economically significant when transaction costs of 1% (or 2% to a lesser extent) are considered.
Pruitt and White (1989)	Options on individual US stocks	1976-1985	Transaction costs (maximum retail costs)	CRISMA trading system	Binomial proportionality test	Even considering maximum retail transaction costs, they report an average return of 12% for options hold on average 25 days.
Pruitt, Tse and White (1992)	Individual US stocks (CRISP)	1986-1990	Transaction costs (0.5 or 1%) and risk	CRISMA trading system	Student tests	Results consistent with those found by Pruitt and White (1988), including during the last 5 years sub-sample.
Brock, Lakonishok and LeBaron (1992)	DJIA	1897-1986		MA and TRB	Student tests and bootstrap	On average, the strategies are profitable. Long (short) positions mean returns are higher (lower) than unconditional returns. Commonly used return generating models cannot explain the rules profitability.
Gençay (1998)	DJIA	1963-1988	Transaction costs and risk (Sharpe ratio)	Feed forward network model	Student tests and sign predictions tests	Returns of the trading strategy (between 7% and 35%) are higher than those of the buy-and-hold for all sub-samples considered (between -20% and 17%). The strategy has market timing ability according to sign predictions tests.
Cooper (1999)	Large US stocks	1962-1993	Transaction costs (0.5%) and risk (beta, Sharpe and alpha)	Filters rules using prices and volume	Student tests	The filter rules produce significant returns that are higher when volume is also taken into account. This strategy outperforms the "standard" contrarian strategy. Risk and transaction cost cannot explain the profits.
Isakov and Hollistein (1999)	Swiss index (SBC) and 5 individual stocks	1969-1997	Transaction costs (0.3% and 1.6%)	MA, RSI and a stochastic indicator	Student tests and bootstrap	They find that a MA rule produces a yearly excess return of 18%, which might not be matched by common returns generating models. However, only institutional investors with low transaction costs ($\leq$ 0.3%) can benefit from this trading strategy.
Fernandez-Rodriguez and al. (2000)	Madrid General Index	1966-1997	Risk (Sharpe ratio)	Artificial Neural Network	Student tests and sign predictions tests	The strategy outperforms the buy-and-hold for bear and stable markets, but not during the bull sub-sample.

Appendix I-C (continued)

Authors	Asset	Sample	Adjustments	Trading rules	Test methodology	Results
Skouras (2001)	DJIA	1962-1986	Transaction costs	Recursive MA	Student tests and maximization of utility equations	Profitability found by BLL is significantly improved by the use of a recursive MA specification. Using past prices information increases utility for risk-neutral and mean-variance investors, as well as for some risk-averse investors.
Fong and Ho (2001)	DJ Internet Composite Index	1997-2000	Transaction costs (0.5% single trip)	MA	Student tests and bootstrap	Statistically and economically significant returns of the buy-sell strategy, even considering transaction cost. Dynamic CAPM is able to explain only a part of this profitability (39%).
Fang and Xu (2003)	DJIA DJ Transport DJ Utilities	1896-1996 1929-1996	Transaction costs (break-even cost)	MA rules, Time series models and combinations	Student tests	Both technical and time series models have predictive power. The profitability of strategies combining the two forecasting methods is significantly higher (between 92% and 142%) than those of the individual strategies. Break-even transaction costs are between 1% and 2%
Wong, Manzur and Chew (2003)	Singapore Index	1974-1994		MA rules and the RSI	Student tests	On average, the strategies are profitable or when they are not, they have some predictive power. Nevertheless, they show that the confidence levels of excess returns decrease from between 1% and 5% during the first half of the sample to a level of 10% during the most recent period.
Szakmary, Shen and Sharma (2010)	28 commodity futures	1959-2009	Transaction costs risk	MA, Momentum	Student tests and bootstrap	They find that MA rules generate similar returns to momentum strategies. They are profitable, but the evidences are somewhat weaker in the last subsample.
Savin, Weller and Zvingelis (2007)	S&P 500 Russell 2000	1990-1999	Transaction costs (0.5%) and risk (Fama-French 3 factors model)	head-and-shoulders	Student tests	Raw excess returns are negative whereas risk adjusted returns are positive and statistically significant.

Appendix I-D: Summary of studies finding no profitability of technical trading strategies

Authors	Asset	Sample	Adjustments	Trading rules	Test methodology	Results
Hudon and al. (1996)	FT 30 (UK)	1935-1994	Transaction costs	MA and TRB rules	Student tests	On average, the profitability decreases over time. Over the most recent sub-sample, the strategies are not profitable anymore when transaction costs are considered. However, buy returns are higher than sell returns, hence support predictability.
Brown, Goetzmann and Kumar (1998)	VW index of US stocks	1902-1929 1930-1997	Risk (beta, Sharpe ratio and Jensen's alpha)	Dow Theory	Student tests and bootstrap	First, they show that the Dow Theory possesses predictability and that risk adjusted returns are significant during the period 1902-1929. Then, using a neural network model, they infer predictions of this theory and show that it has some forecasting power during the out-of-sample period. However, during the last 20 years, it is rather unlikely that a profitable strategy might have been generated.
Allen and Karjalainen (1999)	S&P 500	1928-1995	Transaction costs (0.25%)	Genetic algorithm based on MA and TRB rules	Out-of-sample student tests	The vast majority of the strategies are not profitable when transaction costs are considered and when a one day lag is introduce to avoid the nonsynchronous trading bias. However, buy returns are higher than sell returns, hence support predictability.
Goodacre, Bosher and Dove (1999)	UK stocks Options	1987-1996	Transaction costs (0.25, 0.5 or 1%) and risk	CRISMA trading system	Student tests	Raw returns of the trading system are statistically significant, even with transaction costs. However, excess returns from various return generating models are negative. Thus, risk explains profits from the trading system. However, the signals have some predictive power. Using options might be profitable.
Taylor (2000)	FT all shares 12 UK stocks S&P500 DJIA	1972-1991 1982-1992 1897-1988	Transaction costs	MA rules	Student tests and random walk tests	Evidences against the random walk hypothesis are found for almost all assets and sub-samples considered. However, the strategies are not profitable when transaction costs are considered, especially during the recent period with the exception of the DJIA which generate break-even transaction costs of 1.1%.

Appendix I-D (continued)

Authors	Asset	Sample	Adjustments	Trading rules	Test methodology	Results
Goodacre and Kohn-Speyer (2001)	333 US stocks	1988-1996	Transaction costs (0.25, 0.5 or 1%) and risk	CRISMA trading system	Student tests	The profitability of the trading system is due to risk, as when excess returns from popular return generating models are considered, the profits are not significant anymore, whereas raw returns were significant.
Cheng, Cheung and Yung (2003)	37 indices	1990-2001	Transaction costs (0.125%) and risk	CRISMA trading system	Student tests	On average, the trading system is not profitable on country indices. It might be profitable only on HK stocks with high turnover. Furthermore, the system produces very few trading signals and its performance might be improved by using other parameters.
	Hong Kong stocks	1988-2000				
Cesari and Cremonini (2003)	MSCI World, Europe, US and Pacific	1997-1999	Transaction costs and risk (various measures)	MA rules	Historical and Monte Carlo simulations	They compare the risk adjusted performance of various dynamic asset allocation strategies, including one based on MA rules. This strategy is optimal only on the Pacific market.
Neely (2003)	S&P 500	1929-1995	Transaction costs (0.25%) and risk (various measures)	Genetic algorithm based on MA and TRB rules	Out-of-sample Student tests and market timing tests	The trading strategy does not produce positive significant risk adjusted returns, even when risk is considered in the fitness criterion definition. However, predictability is found according to a market timing test.
Potvin, Soriano and Vallée (2004)	14 Canadian stocks	1992-2000		Genetic algorithm based on MA, ROC and the RSI	Out-of-sample Student tests	On average, the strategy does not generate positive excess returns. However, its performance depends on the market state, as the rules are profitable in a bear or a stable market but not during a bull period.
Fong and Yong (2005)	US internet stocks	1998-2002	Transaction costs (3% or 0.5%)	Recursive MA rules	Student tests and bootstrap and random walk tests	The strategy is neither profitable over the whole sample nor during the bull period. Moreover, the random walk hypothesis might not be rejected. However, the buy-and-hold is outperformed during the bear period as the strategy stays outside the market most of the time.

Appendix I-E: Summary of studies considering technical trading rules profitability on different markets

Authors	Asset	Sample	Adjustments	Trading rules	Test methodology	Results
Bessembinder and Chan (1995)	Malaysia Japan Taiwan Thailand Hong Kong Korea	1975-1989	Transaction costs	MA, TRB	Student tests and bootstrap	The strategies produce, on average, a significant return of 26.8% on a yearly basis. However, the profitability of rules is concentrated on emerging markets (Malaysia, Thailand and Taiwan). It is reduced, but not eliminated, when transaction costs and a one day lag are included. On the other hand, the profits are not significant for the others mature markets.
Tian, Wan and Guo (2002)	DJIA 4 China indices	1926-2000 1992-2000	Transaction costs	MA, TRB	Student tests and bootstrap	They first illustrate the decrease in trading rules profitability on the DJIA. Then, comparing the 2 markets during the nineties, they show that technical strategies produce significant profits only on the Chinese market and conclude that it might not be fully efficient compared with the DJIA.
Fifield, Power and Sinclair (2005)	11 European stock indices	1990-2000	Transaction costs	MA, filter	Student tests	First, the various strategies perform very differently across the markets considered. On average, technical trading rules generate significant excess returns only on the less developed European markets, such as Greece, Hungary, Portugal and Turkey, as they might be less efficient than the large mature markets (UK, France and Germany).
Hsu and Kuan (2005)	DJIA, S&P500 NASDAQ Russell2000	1990-2002 (2 last years for out-of-samle tests)	Transaction costs (0.05%)	39'832 rules (simple and complex)	White and Hansen bootstrap	They show that technical trading strategies are, on average, only profitable for the two more recent stock indices (NASDAQ COMP and Russell 2000). These profits decrease during the out-of-sample test interval but do not disappear. However, they point out that liquidity might explain these results.

Appendix I-F: Summary of studies considering the temporal evolution of the technical trading rules profitability

Authors	Asset	Sample	Adjustments	Trading rules	Test methodology	Results
Mills (1997)	FT 30	1935-1994		MA, TRB	Student tests and bootstrap	The strategies produce economically and statistically significant excess returns over the entire sample and the two first sub-samples whereas they are, on average, negative during the last sub-sample, ranging from 1975 to 1994. Furthermore, transaction costs are not considered.
Bessembinder and Chan (1998)	DJIA	1926-1991	Transaction costs and dividend	MA, TRB (same as BLL)	Student tests and bootstrap	They show that both technical trading rules profitability and break-even transaction costs decline over time. While the strategies produce economically and statistically significant returns over the entire sample, the obtained break-even costs are in line with those of institutional traders during the most recent sub-sample (1976-1991).
Sullivan, Timmermann and White (1999)	DJIA S&P500 futures	1897-1996 1984-1996	Risk (Sharpe ratio)	MA, filter, TRB, channel, On balance volume (7846 specifications)	Student tests and White (2000) bootstrap	Investigating the sample used by BLL, they found that profitability is not due to data-snooping, as both standard and White (2000) bootstrap p-values are equal to zero. However, during the 10 years following BLL, they find that theirs best rule profitably is due to data-snooping, as the standard p-value equals zero, whereas the White (2000) p-value is above conventional significance levels (0.34). The profit found on the futures is half of those on the DJIA and the difference in p-values are even larger (0.04 against 0.9).
LeBaron (2000)	DJIA	1897-1999	Risk (Sharpe ratio)	MA (only 1 rule)	Student tests and bootstrap	The MA rule produces statistically significant returns over the whole sample. However, he shows that the rule is not profitable over the interval not investigated by BLL (1988-1999). Moreover, its return is slightly negative, whereas the buy-and-hold return is positive. He concludes that either BLL results are due to data-snooping or the price process changes during the nineties.

Appendix I-F (continued)

Authors	Asset	Sample	Adjustments	Trading rules	Test methodology	Results
Day and Wang (2002)	DJIA (various computations)	1962-1996	Transaction costs and dividend	MA, TRB	Student tests and bootstrap	Until 1986, the profits of MA rules are statistically significant only for the price-weighted DJIA and not for the value-weighted index. Trading rules over the last 10 years of the sample are never profitable. The increase in volume has eliminated the nonsynchronous bias, which was the cause of anterior technical trading profits.
Olson (2004)	18 major currencies	1971-2000	Transaction costs	MA	Student tests	Using a recursive strategy, they show that excess return of the currencies portfolio decreases from a significant 3.34% during the first 5 years sub-sample to -0.12% over the last one. Moreover, excess returns are significant only until 1985 and they conclude that foreign exchange markets inefficiencies have been corrected during the nineties.
Kidd and Brorsen (2004)	17 futures series (commodities and financial)	1975-2001		Structural change tests	Bootstrap tests for correlations and other measure	The sample is divided in two in order to investigate whether a structural change happened which might explain why technical trading profits decreased since 1990.
Summer, Griffiths and Hudson (2004)	FT 30	1935-1994		Neural network	Student tests	Trading rules are not profitable since 1976. However, the rules developed during the first period (up to 1950) have predicative power over the last sub-sample and produce significant returns. The increase in volatility might makes detection of forecasting factors more difficult, however they are constant.
Dueker and Neely (2007)	DEM, EUR, JPY, GBP, CHF	1974-2005 (test from 1982)	Transaction costs and risk (various measures)	MA, filter, Markov rules and combination	Student tests	They first show that complex rules combining the two classes of rules provide higher excess returns that those taken individually. They also find that Markov rules profitability is stable over time (including during the last sub-sample between 2002 and 2005) whereas MA and filter rules are not profitable anymore from the nineties.

Appendix I-G: Using a weekly frequency: A selection of results

Data and methodology

We considered also the investment strategy exposed in Section 3 in a weekly frequency. The universe of simple MA rules is composed of 678 various specifications with short MA ranging from one to six weeks and long MA from two to 208 weeks. Each of the resulting combinations is evaluated with and without a 1% bandwidth.

The complex strategies are constructed in a similar manner as with the daily frequency. Thus, the four years test sample used to select parameters correspond to 208 weeks instead of the 1008 daily observation.

Investment strategies results

The following figure displays the simple MA rules performance over the entire sample. It is similar to Figure 2.

Figure 14: Simple MA rules returns: Weekly frequency

Note: These figures report the annual mean simple returns of all individual MA rules over the entire sample ranging from 1990 to 2008. The horizontal plane is the average buy-and-hold return over this period, i.e. 6.1%. The X and Y axes are stated in terms of weeks.
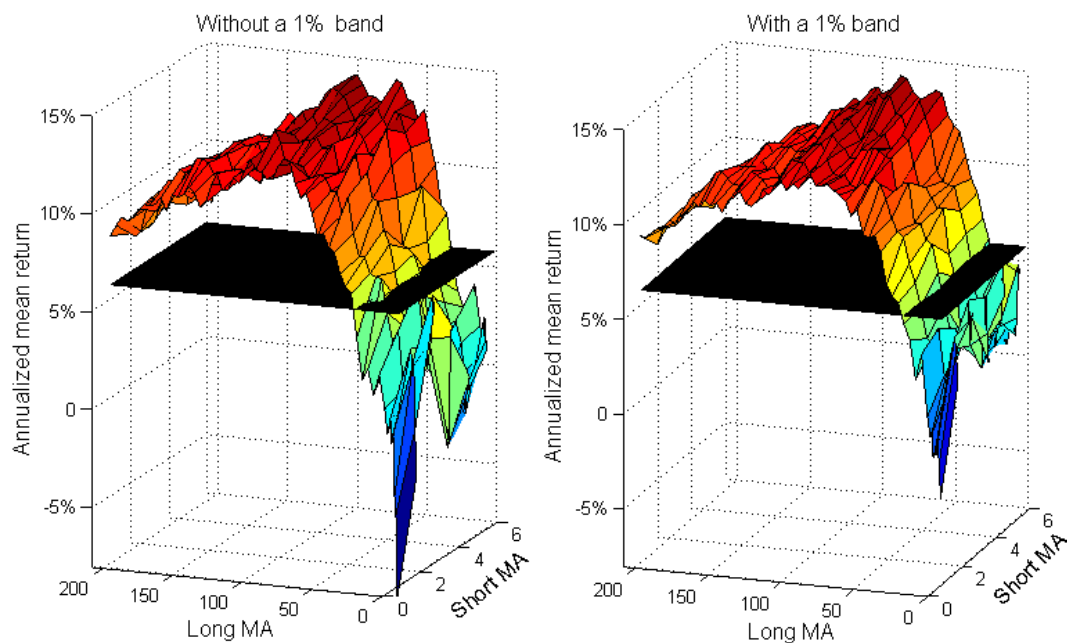
Table 11: Complex rules returns: Weekly frequency

Note: This table reports the results of three benchmark strategies: buy-and-hold (BH), random walk (RW) and best in-sample rule (BEST), as well as the four complex strategies over the entire evaluation period, from January 1994 to December 2008. *Nb Buy* and *Nb Sell* are the number of long and short weekly positions generated by the strategies, *% Right* corresponds to the percentage of right positions as described above. *Buy*, *Sell* and *Strategy* are the annualized mean return of long, short positions and the overall strategy. The *t*-statistics test the null hypothesis that the mean of the specific series is significantly different from the mean return of the buy-and-hold strategy. Strategy compounded is the mean annualized return of the strategy in term of compounded returns. *Nb Trades* is the number of trades generated by the strategy. *Break Even TC* is the level of transaction costs that makes the excess return of the buy-and-hold strategy equal to zero. *Volatility Buy, Sell* and *Strategy* are respectively the annualized volatility of the long, short and overall positions. *Beta* and *Alpha* are estimated in the static CAPM framework. The *t*-statistics of the alpha and beta test the null hypothesis that these parameters are equal to 0. *Alphas* and *Sharpe Ratios* are expressed in annual terms.

| | Investment strategies | | | | | | |
	BH	RW	Best	Opt_all	Opt_4	Voting	Partial
Nb buy	779	438	568	604	596	591	592
% right buy	56.2%	55.7%	59.7%	58.6%	59.4%	59.4%	59.5%
Nb sell		341	178	174	183	182	187
% right sell		43.1%	55.1%	51.7%	54.1%	53.8%	54.0%
% right strategy		50.2%	58.6%	57.1%	58.2%	57.6%	58.2%
Buy		-0.0235	0.1310	0.1103	0.1228	0.1220	0.1195
t-statistic		*-1.16*	*1.17*	*0.86*	*1.06*	*1.04*	*1.00*
Sell		-0.1618	0.1909	0.1241	0.1547	0.1388	0.1440
t-statistic		*-2.62*	*1.19*	*0.59*	*0.88*	*0.73*	*0.80*
Strategy	0.0576	-0.0840	0.1412	0.1133	0.1303	0.1250	0.1254
t-statistic		*-2.28*	*1.35*	*0.90*	*1.17*	*1.09*	*1.10*
Strategy compounded	0.0439	-0.0939	0.1354	0.1038	0.1227	0.1169	0.1183
Volatility buy		0.145	0.139	0.138	0.138	0.139	0.135
Volatility sell		0.197	0.248	0.251	0.247	0.245	0.236
Volatility strategy	0.170	0.170	0.168	0.170	0.170	0.169	0.165
Nb trade		775	11	23	15	9	13.83
Break even TC		-0.27%	11.38%	3.63%	7.26%	11.22%	7.35%
Beta	1	-0.18	0.00	0.03	0.01	0.02	0.01
t-statistic		*-5.14*	*0.10*	*0.71*	*0.19*	*0.53*	*0.34*
Alpha		-0.1196	0.1021	0.0738	0.0912	0.0857	0.0862
t-statistic		*-2.77*	*2.35*	*1.68*	*2.08*	*1.96*	*2.03*
Sharpe	0.108	-0.724	0.608	0.437	0.537	0.507	0.524

The performance in term of returns is very similar according to the two data frequencies. In addition, all tests generate the same conclusions except for the Jensen's alpha whose *t*-statistics are higher than 1.96 for three strategies compared with only two in the daily frequency. The major difference can be found in the percentage of right signals. Indeed, these statistics are consistently higher in the weekly frequency. For instance, our strategies have a percentage of right signals ranging from 54.25% to 54.7% in a daily frequency compared with 57.1 % and 58.2%. This may seem to be in contradiction with the simulation performed in Section 7, as these higher percentages are not associated with higher returns. However, these two percentages can not be compared directly. Consider for instance that the complex strategies take a long position during a week and the weekly return is positive. The percentage in a weekly frequency is thus 100%, while it may be lower in the daily frequency if some daily returns are negative. The level may be as low as 20% if there are four negative returns and one positive large enough to make the weekly return positive.

To further investigate this issue, we conduct the same simulation as in Section 7 in the weekly frequency. The first difference is the proportion of positive buy-and-hold returns[43] in the two frequencies, 53.12% for daily data and 56.23% in the weekly frequency. We show in Section 7 that a level of 55% would generate, on average, mean excess returns of 15%. To reach the same amount in a weekly frequency, a level of around 70% is required. Consequently, the average change in mean excess return following a 1% increase in the level of right signals is much smaller with weekly data with 0.85% compared with 4.1% previously. The results are similar with Student *t*-tests aimed at detecting whether abnormal returns are statistically different from zero. Indeed, at least 95% of the tests would reject the null hypothesis when strategies have a percentage of right signals higher than 63% compared with 56.4% in a daily frequency.

This selection of results shows that there is no significant difference in the complex strategies performance according to the investment frequency. This is not surprising as they rely on long-term trends, and they do not switch trading position often.

---

[43] This also represents the buy-and-hold percentage of right signal.

Figure 15: Percentage of right signals: Weekly frequency

Note: This figure presents the results of the simulation study between the percentage of right signals and annual mean returns in excess of the buy-and-hold. The first two panels display the mean and the 5% and 95% deciles of the distributions of the simulated mean excess returns. The third panel summarizes the relation with mean *t*-statistics obtained with a standard bilateral Student *t*-test of equal means. Panel D provides the percentage of these tests that result in a significant difference between mean returns. For each level of right signals, 5,000 simulations are conducted.

Panel C

Panel D
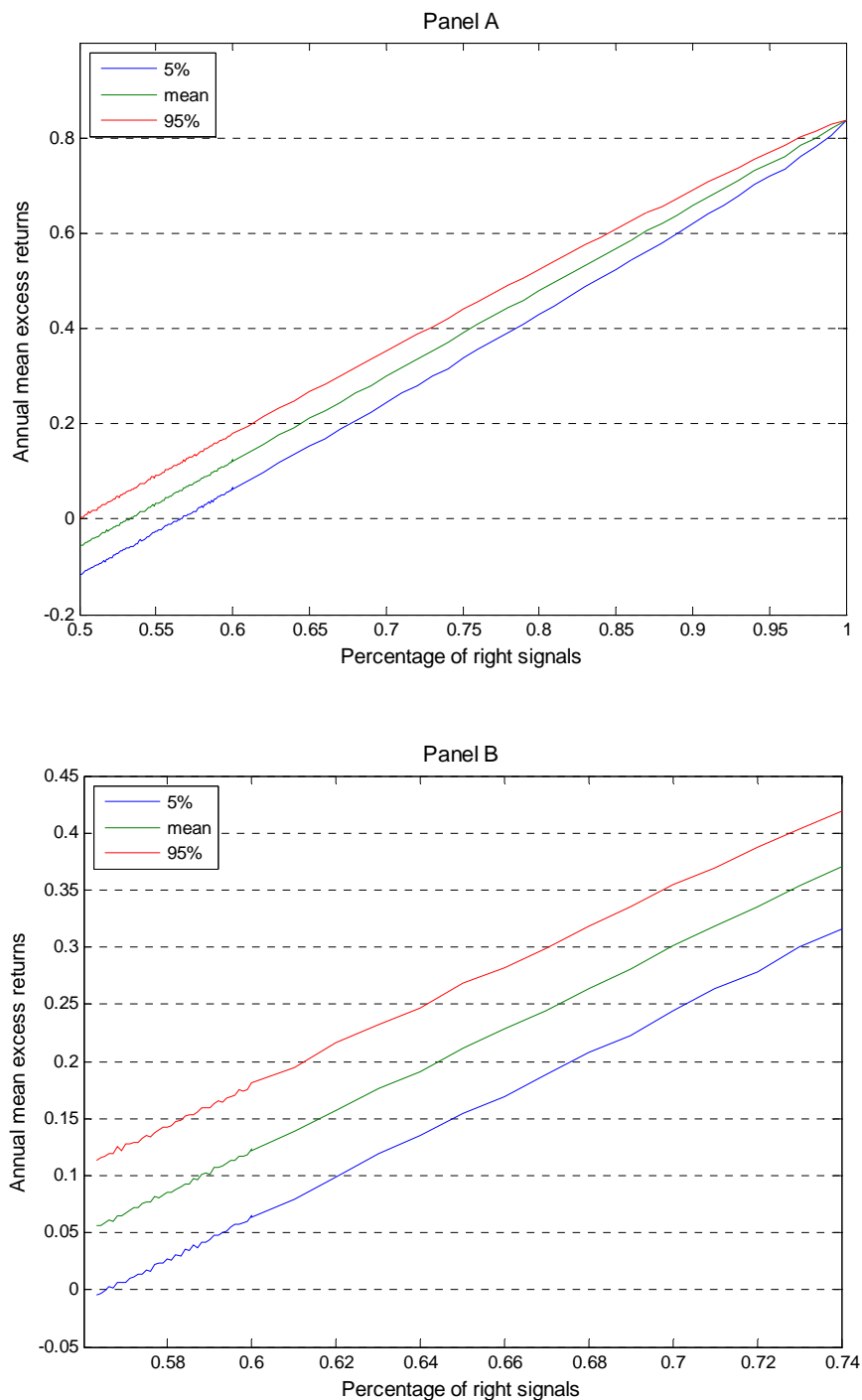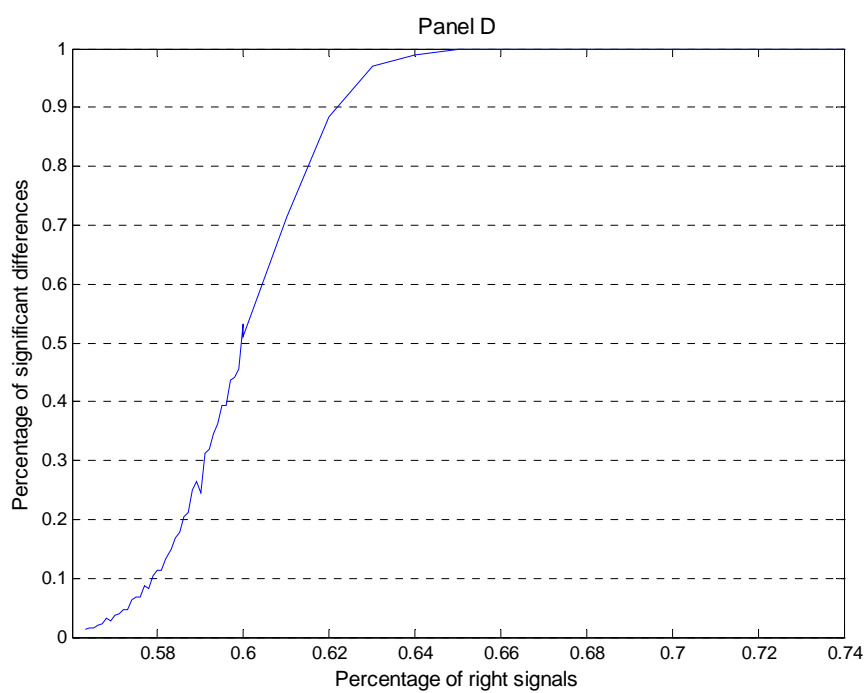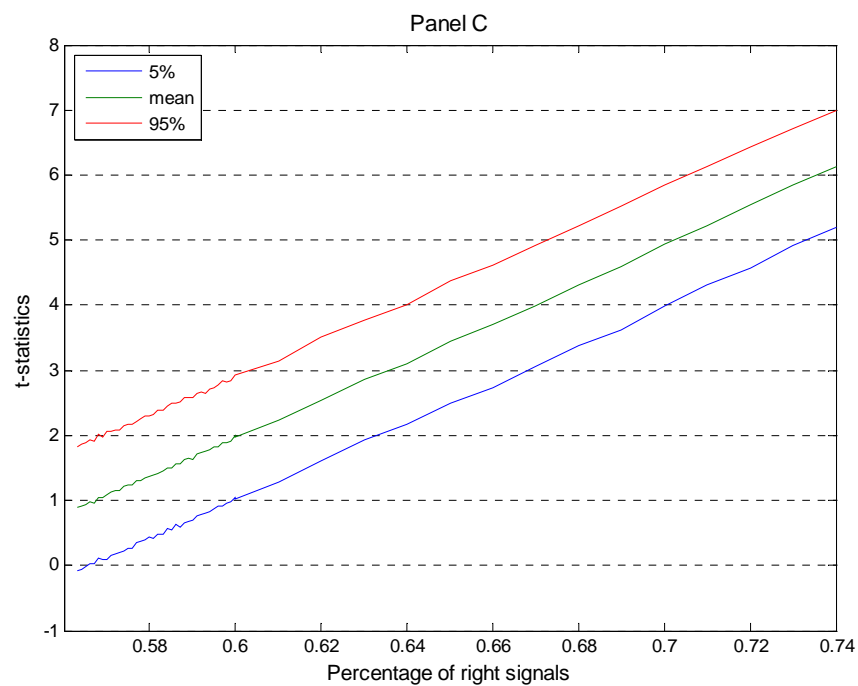
Appendix I-H: Strategies with options that have a longer time to maturity

In this Appendix, we present the performance of the strategies constructed with options as in Section 4.3. However, instead of using options that have a relative short time to maturity, we select options with a farther maturity date. Table 12 displays the statistics of the options time-series selected according to the same process, but they should have a time to maturity between 50 and 180 days.

Table 12: Statistics of selected call and put options

Note*:* This table describes the options selected for our investment strategies. *Mean return* is the simple daily mean return, Open interest is the average number of contracts that are neither closed nor delivered. *Daily volume* is the average number of contract traded during a day, *Relative BA spread* is the relative bid-ask spread, *Moneyness* is the average moneyness as calculated in equation (1.15), *Maturity* is the average time to maturity in days and *BS beta* is the mean beta obtained under the Black and Scholes assumptions.

	Call	**Put**
Mean return	0.0028	-0.0037
Median return	0.0000	-0.0160
Open interest	38692.3	43719.1
Daily volume	1378.2	1957.1
Relative BA spread	0.0564	0.0582
Moneyness	-0.0092	-0.0274
Maturity	100.6	98.2
BS beta	14.5	-17.9

Compared with their short time-to-maturity counterparts, the call options have a lower mean return, but a higher median return. The lower mean return can be explained by a lower leverage, which is also reflected in the lower mean beta, i.e. 14.5 compared with 20.5 for the short maturity options.

In Table 13, we present the performance analysis of the strategies with the three different percentages of capital invested with options. The use of this new options set does not modify the conclusions. The only difference is the decrease in leverage, compared with the options selected according to short time-to-maturity. However, the strategies are not more profitable.

Table 13: Strategies with option selected with a longer time to maturity

Note: This table presents the performance of the trading strategies with traded options according to three levels of leverage. *Buy*, *Sell* and *Strategy* are respectively the simple mean annual return of the long, short and overall positions. *Strategy compound* is the annual mean compounded return. *Volatility buy, sell* and *strategy* are the annualized volatilities of the two components of the strategies and the global one

Panel A: 5% of options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy		0.1940	0.1479	0.2075	0.1614	0.1699
Sell		0.3028	0.0429	0.2869	0.1878	0.1577
Strategy	0.0609	0.2192	0.1240	0.2263	0.1673	0.1672
Strategy compound	-0.0003	0.1767	0.0717	0.1857	0.1168	0.1264
Volatility buy		0.303	0.309	0.301	0.307	0.292
Volatility sell		0.438	0.405	0.435	0.437	0.374
Volatility strategy	0.353	0.339	0.333	0.338	0.340	0.312

Panel B: 10% of options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy		0.2555	0.1893	0.2799	0.2065	0.2249
Sell		0.4309	-0.0004	0.3982	0.2468	0.2038
Strategy	0.0604	0.2959	0.1462	0.3079	0.2155	0.2202
Strategy compound	-0.0797	0.1847	0.0252	0.2009	0.0916	0.1220
Volatility buy		0.481	0.490	0.477	0.487	0.462
Volatility sell		0.616	0.533	0.612	0.616	0.483
Volatility strategy	0.548	0.515	0.500	0.512	0.518	0.466

Panel C: 15% of options

	BH	Best	Opt_all	Opt_4	Voting	Partial
Buy		0.317	0.231	0.352	0.252	0.280
Sell		0.559	-0.044	0.510	0.306	0.250
Strategy	0.060	0.373	0.168	0.390	0.264	0.273
Strategy compound	-0.183	0.158	-0.047	0.181	0.035	0.091
Volatility buy		0.664	0.676	0.658	0.672	0.636
Volatility sell		0.806	0.668	0.800	0.807	0.598
Volatility strategy	0.751	0.699	0.674	0.694	0.704	0.628

Part II: Portfolio optimization and parameter selection

1. Introduction

The Markowitz (1952) mean-variance portfolio optimization is a milestone of the modern financial theory. It enables investors to select objectively a portfolio that maximizes their returns for a predetermined level of risk, which is stated in term of variance. However, despite its theoretical appeal, this asset management method is not widely used in the financial industry. For instance, Michaud (1989) presents a wide range of issues that may explain the skepticism of investment professionals; the resulting portfolio is not well diversified or with a composition far from standard benchmarks, the estimation of the required parameters and the organisational structure or the conservatism of investment firms.

A major issue with the practical implementation of mean-variance optimization is that the returns distribution moments are not stable over time. For instance, it is rather unlikely that an asset with low volatility and high mean returns during the estimation period will have the same properties during the out-of-sample evaluation period. This implies that the in-sample performance, i.e. those resulting from the optimization with past data, is not a reliable proxy for the out-of-sample performance. Indeed, the theory is stated in terms of expected returns and covariance matrix. In practice, we have to use estimation methods based on past data to proxy these expectations as they are not known. In addition, the optimized portfolio may have an unintuitive composition and may be poorly diversified, when for instance, a small proportion of assets under consideration represents the majority of the optimized portfolio capitalisation.

Most of the recent academic studies address this issue by examining alternative[44] parameters estimation models. In the current Part of this thesis, we follow a different approach, as we focus on the length of the estimation window used to estimate the parameters. Indeed, virtually all studies related to the mean-variance optimization use a five years window. This procedure may make sense from a statistical point of view; indeed, it allows having enough data to efficiently estimate the parameters, especially as the optimization is often performed with monthly data. However, we argue that shorter lengths may be more in line with market trends. Thus, we compare the performance of various optimization specifications that consist in an estimation length and a choice of models to estimate the parameters. Note that the portfolio rebalancing is made on a monthly basis, but the parameters are estimated with daily data. We find that the length has a stronger impact on the optimization performance than the set of estimation models. Furthermore, shorter lengths improve the out-of-sample performance dramatically. However, considering a large set of specifications is subject to data-snooping issues, as in the technical analysis setting presented in the first Part. Hence, we propose to construct new portfolios with similar selection processes to those applied to simple MA rules. The findings are in close agreement with those presented in the first Part; the complex portfolios generate large excess returns, when short-sales are allowed, but the Student t-tests do not consistently conclude in favour of a statistically significant difference in returns between these portfolios and a benchmark.

This part is structured as follows; the Section 2 presents a literature review with a focus on the optimization performance and the proposed methods to improve it. In Section 3, we detail both the optimization with various specifications and the parameters selection processes. Section 4 presents the data, Section 5 focuses on the empirical results and Section 6 concludes.

2.   Literature review

We divide this literature review according to three main directions. The first presents the early studies about the mean-variance optimization empirical implementation and the related issues. The second focuses on the proposed methods to overcome these problems, such as the use of

---

[44] By alternative, we mean other method than the mean return for the expected return and the sample covariance matrix.

alternative parameter estimation techniques and constraints. Finally, the last part is dedicated to the comparison of these different methods to improve the optimization.

2.1   The empirical implementation of mean-variance optimization

Mao and Särndal (1966) and Fried (1970) are among the first to point out the impact of using parameters estimates instead of their "true values" in the mean-variance framework. Frankfurter, Phillips and Seagle (1971) provide an empirical illustration. They consider an investment universe made up of three assets, which follow a multivariate normal distribution. They use empirical data to estimate the means vector and the covariance matrix, and they suppose that these estimates are the true values. Then, they simulate returns series that represent the data available to the manager to estimate the parameters. Thus, their experiment consists in comparing the composition, or in other words the weights of each asset, of the efficient portfolio obtained with the true parameters with those constructed with the estimates resulting from the simulated returns series. We call them the true optimal portfolios and the estimated optimal portfolios. They find that these two set of portfolios have a different composition. Indeed, the efficient set varies substantially when the estimated parameters are used compared with the true values. In addition, some of the inefficient true portfolios are optimal when the frontier is constructed with the estimated portfolios. Conversely, former true efficient portfolios are dominated by their estimated counterparts. They conclude that the error in estimates issue is serious enough to challenge the usefulness of the mean-variance optimization in a realistic setting. Jobson and Korkie (1981) provide further evidence in a simulation setting as well. They examine a wider investment universe of 20 assets, and they also compare the financial performance of various portfolios. They find that the Sharpe ratio drops sharply when estimates are used instead of the true parameters. In addition, they suggest that naive portfolios, such as the equally weighted method, can outperform mean-variance optimal portfolios with estimated parameters.

Best and Grauer (1991a) perform a sensitivity analysis to show how a variation in the parameters influences the composition, and the performance, of a mean-variance optimal portfolio. They show that a small increase in the mean of one out of 100 assets implies a large difference in the optimal portfolio weights, but the portfolio mean and volatility are almost not affected. They further illustrate this issue in Best and Grauer (1991b). For example, increasing the mean of one asset by 11.6%, i.e. from an 18% annual mean to 20.1%, is enough to drive 50% of the assets out of the mean-variance optimal portfolio composition. They also argue that the mean-variance optimization may produce portfolios that are not well diversified. Chopra and

Ziemba (1993) examine this issue but also for variances and covariances estimates and not only for the mean. They conduct a simulation experiment with the cash equivalent loss[45] (CEL) as the performance measure, which takes the risk aversion into account. They find that the CEL is approximately 11 (20) times higher for errors in means than in variances (covariances) estimates. In addition, these proportions rise for investors with a lower risk aversion, as they target portfolios with higher expected returns. To conclude, they suggest that the minimum-variance portfolio may be a solution for investors without superior mean estimates, since it does not require them.

The potential lack of diversification of the mean-variance optimal portfolio is in line with Kallberg and Ziemba (1983), who compare the portfolios obtained with various utility functions. They also document that the optimal portfolio weights are similar for different investor utility functions. Green and Burton (1992) suggest that this issue is not only due to estimation errors, but also to a dominant factor in the assets returns covariance matrix. This implies that the betas' cross-sections do not show large variations. Thus, to minimize the variance, large short positions in some assets are required to finance large long positions in other assets. They also show that if the betas are constant across assets, any portfolio is well diversified, but they also bear a significant factor risk. On the other hand, a reasonable variation in betas can eliminate this risk, but at the expense of extreme weights. They provide an illustration with minimum-variance portfolios. First, they show that as the number of assets increases, the correlation between assets that have a long or a short position increases. This provides evidence for the single factor hypothesis. Second, the weight concentration in the minimum-variance portfolio is not significantly lower than those of other portfolios along the efficient frontier. Since the former does not use the expected returns, they conclude that estimation errors are not the primary driver for extreme weights. Chopra (1993) sheds light on the "near-optimal" portfolio issue. He illustrates that portfolios with a large difference in their composition can have a very similar risk-return profile to the optimum portfolio.

---

[45] The cash equivalent, CE, of a portfolio with unknown, or risky, returns is the value for which a risk averse investor would exchange it for cash and having the same utility. The CEL is defined as the difference between the CE resulting from holding a random portfolio and the one from holding the optimal one. This difference is divided by the first value (CE from non optimal portfolio) to have a percentage. This measure allows taking into account risk aversion in holding a non optimal portfolio.

Finally, Michaud (1989) reviews the reasons why mean-variance optimization is not widely used by asset managers. First, he suggests that the main reason is political, as the composition of mean-variance optimized portfolios are not in line with the structure and the management of investment firms. In addition, the optimal portfolio composition may be rather unintuitive and with extreme weights, specially compared with a standard stock index. This is the fundamental question: Do portfolio optimization possess any intrinsic value, but they are only difficult to understand and explain? An argument against a positive answer is the estimation errors issue. Indeed, he defines the optimizers as "estimation-errors optimizers", because they give more weight to assets with the highest expected returns, which are more likely plagued by the largest estimation errors. He also argues that extreme weights can lead to liquidity problems in case of a small-cap fund, for example. Finally, he points out some technical issues, such as the method used to perform the optimization and the initial values. He concludes that one should be careful when using optimizers, but the practical implementation can be enhanced with constraints and alternative parameters estimation techniques. The next part of this literature review focuses on these aspects.

2.2   Enhancing the optimization

2.2.1   The use of constraints

Two kinds of constraints can be implemented in the optimization process. The first restricts short-sales, i.e. negative weights. The second is the diversification constraint, which defines a maximum weight for each asset. It is obvious that adding them in the portfolio optimization reduces its in-sample performance, as fewer combinations are available. However, when the parameters are not known and have to be estimated, they may improve the out-of-sample performance of the optimal portfolio.

In a simulation study, Frost and Savarino (1988) examine the impact of these two constraints on the performance of optimal portfolios. First, they show that a diversification constraint reduces the estimation bias between the true parameter and its estimate. The impact is larger for the mean returns vector than the covariances matrix. In addition, the constraints are economically valuable, as the certainty equivalent return increases, especially for the most risk-averse investors. Secondly, they show that the short-sale constraint reduces the in-sample expected return massively, from 212.9% without constraint to 1.9% with the constraint. Nonetheless, the unconstrained out-of-sample return drops sharply from 212.9% to 5.68%, while the constrained one is reduced from 1.9% to 1%. In other words, the expected return is 37.5 (1.9)

times higher than the actual return when short-sales are (not) allowed. They also show that the out-of-sample variance explodes from 49% in the estimation process to 1882% when short-sales are allowed. This contrasts with the constrained optimization, as the variances are almost similar between the in-sample estimation and the out-of-sample evaluation measures. Grauer and Shen (2000) propose a different diversification constraint. Instead of using an equal limit for each stock, they restrict each asset weight to be within a plus or minus 5% or 10% margin from their percentage in a value-weighted portfolio. They find that these constraints reduce both the portfolio return and the variance, compared with the standard diversification constraint. However, this may only benefit to the most risk-averse investors.

Jagannathan and Ma (2003) provide further evidence supporting the constrained optimization, even if it is not judicious when true parameters values are known. Finally, DeMiguel, Garlappi, Nogales and Uppal (2009) propose to constrain the total of weights together instead of their individual values. For example, the sum of the weights in absolute values has to be smaller than one. They argue that this novel method to constrain the optimization makes practical sense, as it is similar to "130-30" portfolios, which allow the manager to short some assets up to 30% of the capital and to invest 130% in other assets. Finally, these two studies demonstrate that constraining the optimization is comparable with shrinkage estimators, which we detail in the next Section of this review.

2.2.2   The use of alternative parameters estimation methods: The means vector

Stein (1955) points out that the usual mean estimator is not accurate in a multivariate setting, as it does not minimize the expected sum of squared errors. James and Stein (1961) propose the so-called James-Stein estimator that shrinks individual means to the global mean, i.e. the mean of all random variables means. The shrinkage factor is the intensity according to which an asset individual mean is shrunk to the global mean, and it depends on the sample covariance matrix and the number of variables. If this factor is equal to one, all variables have the same mean, which results in the minimum-variance portfolio in the mean-variance setting. Jobson and Korkie (1981) and Grauer and Hakansson (1995) estimate the means vector according to this method. Jorion (1986) proposes a slightly different estimator, the Bayes-Stein estimator, which considers the return of the minimum-variance portfolio as the shrinkage target. This method is used, among others, by Jorion (1991) or DeMiguel, Garlappi and Uppal (2009). Grauer and Hakansson (1995) compute mean returns according to the CAPM. Garlappi, Uppal and Wang (2007) extend the mean-variance framework to take into account the estimation uncertainty, and the aversion to

this uncertainty. This enables to diminish the portfolio weights variation over time, and thus, this provides better out-of-sample results.

Solnik (1993) and Harvey (1995) suggest that using conditional information to obtain better returns forecasts improves the performance of mean-variance optimization. This is based on the return predictability literature (see for example Keim and Stambaugh (1986), Campbell and Hamao (1992), Ferson and Harvey (1993), or more recently Ang and Bekaert (2007)). They consider the standard set of known information, such as the interest-rate term structure, the credit-risk spread or the dividend yield.

2.2.3   The use of alternative parameters estimation methods: The covariance matrix

The estimation of the covariance matrix can be problematic when the number of assets grows. Indeed, for an investment universe composed of *N* assets, there are *N(N-1)/2* elements to estimate with finite time-series of length *T*. The estimates are noisy when *T* is not very large compared to *N*. A solution is to impose some structure on the covariance matrix. The market model, introduced by Sharpe (1963), reduces the number of estimates. According to this model, the common variation of asset returns depends on one factor only, the market return. The resulting covariance matrix requires each asset beta, obtained by regressing the asset returns on the market return, and the variance matrix of these regressions residuals. Chan, Karceski and Lakonishok (1999) further develop this approach with various multi-factors models.

The next group of covariance matrix estimates consists of shrinkage estimators. They are similar to those presented above for the means vector. First of all, Elton and Gruber (1973) suppose that all correlation coefficients are equal to the same value, their means across all assets. Thus, the covariance matrix is constructed with this average correlation and the assets individual standard deviation. This is the constant-correlation model. Aneja, Chandra and Gunay (1989) propose a portfolio approach to estimate the average correlation, without having to estimate all pair-wise correlation coefficients. Ledoit and Wolf (2003) develop a covariance matrix shrinkage estimator similar to the James-Stein mean estimator. They shrink the sample covariance matrix to the one resulting from the market model. The shrinkage factor determines the structure of the covariance matrix, ranging from no structure, i.e. the sample covariance matrix, to a one-factor

structure in case of complete shrinkage. In Ledoit and Wolf (2004), they use the constant-correlation model as the shrinkage target. The RiskMetrics[46] methodology proposes to estimate the covariance matrix by weighting the squared returns exponentially to give more weight to recent returns.

Laloux, Cizeau, Potters and Bouchaud (2000) and Rosenow, Plerou, Gopikrishnan and Stanley (2002) apply the random matrix theory to improve the correlation matrix estimation. The concept is to compare a sample covariance matrix with random matrices that are obtained with independent simulated time series. Thus, the correlations in these matrices are spurious, and they result from finite time series. If the original covariance matrix does not correspond to the random ones, this means that there is reliable information that can be used in the optimization process. Then, a filter is used to clean the sample covariance matrix by removing the noisy part of the correlations estimates. Tola, Lillo, Gallegati and Mantegna (2008) review and compare the impact of various filtering procedures on the portfolio optimization. Bai, Liu and Wong (2009) suggest that bootstrapped parameters estimates are less subject to estimation errors than their sample counterparts.

The majority of these alternative estimators are solely based on past data. Nonetheless, a vast literature focuses on one-step ahead volatility forecasts. Poon and Granger (2003) provide a comprehensive review. Pojarliev and Polasek (2001) consider the BEKK model, defined by Engle and Kroner (1995), to forecast the covariance matrix. Pojarliev and Polasek (2003) and Specht and Gohout (2003) focus on univariate volatility processes. However, these techniques are usually designed to model the volatility with daily or intraday data, while portfolio optimization is performed at best with a monthly rebalancing. Hlouskova, Schmidheiny and Wagner (2009) propose a closed-form solution for multiple-steps forecast when the rebalancing frequency is lower than the one used to estimates the models.

2.2.4   A comparison of alternative parameters estimation methods

As mentioned earlier, Jagannathan and Ma (2003) suggest that constraining the weights in the optimization program is similar to use shrinkage estimators. They focus on the covariance matrix, and thus, examine the minimum-variance portfolio. First, they show theoretically that constraining the weights and the shrinkage estimators leads to a similar reduction in the out-of-

---

[46] This method comes from J.P Morgan and it is released in a technical document published in 1996 in New York.

sample variance, compared with the sample approach. Then, they examine the variance of the out-of-sample portfolio constructed with various estimation methodologies[47]; the sample covariance matrix, factors models, the Ledoit and Wolf (2003) shrinkage estimator and the sample covariance matrix estimated with daily returns. The comparison is based on the out-of-sample variance obtained with the various methods. As they consider the minimum-variance portfolio, the best performing method is the one that generates the lowest out-of-sample variance. They perform the optimization with and without the non-negative weight constraint. They find that the constrained sample covariance matrix performs almost as well as those obtained with the alternative estimation methods. However, adding the constraint in conjunction with these models does not improve the optimization performance. In addition, they show in a simulation experiment that the sample covariance matrix with the constraint provides a similar performance as factors models. This holds even if the returns are generated, in the simulation, with a factor structure.

DeMiguel, Garlappi, Nogales and Uppal (2009) further investigate the link between weights constraints and parameters estimation techniques. They show that constraining the square of the assets weights sum to be smaller than a certain threshold leads to portfolios constructed with shrinkage covariance matrix estimators. In addition, if this threshold is smaller than $1/N$, where N is the number of assets, the optimal portfolio corresponds to the equally-weighted portfolio. Finally, they provide empirical evidence supporting their proposed norm-constraint portfolio. Board and Sutcliffe (1994) also point out that the choice of the parameters estimation models has a larger influence on the optimal portfolio performance when the optimization is unconstrained. Indeed, the average correlation coefficient among 11 optimization specifications is 0.14, compared with 0.77 when short-sales are restricted. Golosnoy and Okhrin (2007) propose to shrink the optimal weights vector to the portfolio current weights. They argue that when uncertainty about optimized weights is large, an investor should keep his current portfolio holdings. To the contrary, when the uncertainty is reduced, he should allocate a higher proportion of his wealth according to these optimum weights. They show that shrinking weights, instead of estimated parameters, is optimal in term of utility, while Ledoit and Wolf (2003) and Jorion (1986) models are optimal according to a quadratic loss criterion only.

---

[47] They consider a large investment universe of 500 stocks. In this context, the sample covariance matrix should be affected by estimation errors to a large extent.

Among others, Chopra, Hensel and Turner (1993), Jorion (1985), Jorion (1986) and Ledoit and Wolf (2003) compare the performance of alternative parameters estimation methods, including shrinkage estimators. Unanimously, they find that they out-perform the portfolios generated with the sample mean and covariance estimates. To compare the portfolios, various methods are used, such as the terminal wealth of a USD invested at the beginning, the loss of utility of investing in a non optimal portfolio (similar to the CEL detailed above) or the direct comparison of out-of-sample mean returns and variance[48]. On the other hand, DeMiguel, Garlappi and Uppal (2009) find that none of the optimization specification considered is consistently better than the equally-weighted portfolio, both in terms of Sharpe ratio and certainty equivalent returns. However, Kritzman, Page and Turkington (2010) argue that the results of the previous study are due to short rolling-samples of five years to estimates the parameters. On the contrary, they use up to 20 years to estimate the covariance matrix, and they employ long-term risk premium, which are supposed to be constant, for expected returns. They simulate more than 50'000 optimal portfolios, and show that both the minimum-variance portfolio and the one maximizing the Sharpe ratio consistently generate a higher performance than the equally-weighted portfolio.

Pafka and Kondor (2004) simulate artificial returns series in order to isolate the impact of noise in the correlations matrix estimation. Thus, this noise arises from the use of finite time series lengths for estimation purposes and not from the uncertainty about the "real" correlation structure. They simulate returns series with various covariance structures, and then, they compare three estimation methods; the sample covariance matrix, a one-factor model and the filtered matrix obtained with the random matrix theory. The latter generates the smallest difference between the true values and the estimates. This is in line with Rosenow, Plerou, Gopikrishnan and Stanley (2002).

In contradiction with most studies, Liu and Lin (2010) find that simple methods to estimate the covariance matrix, such as a single-factor model, perform better than more sophisticated methods, the Bayesian shrinkage or random matrix estimators. This ranking is obtained according to the root mean squared error criterion, but it differs significantly when comparing the out-of-sample standard deviations. They argue that these results differ from the related literature because they have a relative high "sample length/number of assets" ratio.

---

[48] The drawback of this method is that a portfolio has to yield a higher mean return and a lower variance in order to out perform another one.

3.   The investment strategy

In this Section, we extend the literature on mean-variance optimization in two directions. First, we examine the influence of using various estimation window lengths to estimate the parameters, i.e. the expected return and the covariances. Then, we propose to implement processes to select objectively the optimization specifications that consist of four elements; a model to estimate the expected returns, one to estimate the covariance matrix, the window length on which they are estimated[49] and whether short-sales are allowed or not. These selection processes are similar to those used by complex strategies developed in the technical analysis framework. To avoid confusion, we refer to the portfolios generated by these selection processes as complex portfolios.

We divide the description of these portfolios in two parts, which are in line with their empirical implementation. In the first, we present the optimization process with various specifications. Thus, we obtain "simple" portfolios[50] that are used to construct the complex portfolios. We propose two different approaches to objectively select the optimization specification used by the complex portfolios, which are described in the second part.

3.1   The optimization process with various optimization specifications

3.1.1   The optimization

The Markowitz (1952) develops a investment theory that aims to maximize a portfolio return for a given level of risk, which is represented by the volatility of its returns. The efficient frontier is a set of mean-variance efficient portfolios that have the following characteristic; for each level of volatility (expected return), the efficient portfolio is the one that has the highest (lowest) expected return (variance). Thus, the first decision is to select a specific portfolio on this frontier. Two popular choices are the minimum-variance portfolio and the one that maximizes the Sharpe ratio. We choose a different approach. We select the portfolio with the highest expected return among those that have a volatility lower than or equal to 20% in annual terms. The idea is to choose a portfolio with a variance similar to the long-term market average, but with the

---

[49] For example, the basic specification consists in using the assets simple mean and the sample covariance matrix estimated over the last five years of data.

[50] They are similar to the simple MA rules in the technical analysis setting presented above.

opportunity to have a lower variance during bear markets[51]. This approach also makes sense for investors, as they can select volatility levels in line with their risk aversion. Nonetheless, this method does not ensure that the volatility constraint is respected in the out-of-sample evaluation period. Pojarliev and Polasek (2001) use a similar approach, but they select the portfolio with the smallest variance for an expected return target. In addition, we want to compare estimation models for the mean return and the variance, thus the minimum-variance portfolio is not suitable. Our sample under consideration includes both the early 2000 bear market and the late 2008 financial crisis, and thus, using the Sharpe ratio with stock indices mean returns only slightly higher than those of bonds is not very attractive as the optimum portfolios would consist mostly in bonds[52].

The first step is to compute the portfolio weights, for each specification, with the following standard optimization program;

$$
\begin{aligned}
&Max_x \quad x_{t,sp}{}'\mu_{t,sp} \\
&\text{s.t.} \quad x_{t,sp}{}'\Sigma_{t,sp}x_{t,sp} \leq \text{var\_max} \\
&\qquad x_{t,sp}{}'\boldsymbol{1} = 1 \\
&\qquad x_{t,sp} \geq -0,5 \ \ (\text{or } 0 \text{ when short sales are not allowed}),
\end{aligned}
\tag{2.1}
$$

where $x_{t,sp}$ is the optimal weights vector obtained according the sp^{th} specification at time t, $\mu_{t,sp}$ and $\Sigma_{t,sp}$ are the assets expected returns vector and the covariance matrix. var_max is the maximum variance target and $\boldsymbol{1}$ is a vector of ones with a number of components equal to the number of assets. The second constraint states that the sum of weights has to add up to one, and the last restricts individual weights to be higher than or equal to -50%[53] when taking short positions is

---

[51] The optimization constraint states that the portfolio volatility should be equal or lower than 20% in an annual frequency. In case of an equity bear market, when most of indices expected returns are negative, the optimization might allocate the whole capital in the bonds indices, and thus, the portfolio volatility might be lower than 20%. If the volatility constraint were "equal" instead of "lower or equal", the optimization would have to invest in equity indices in order to reach the volatility limit, even if this implies in a lower expected return.

[52] We have constructed these portfolios and find that bonds are very often the highest weighted assets. They are also much less interesting to use in the last part of this thesis, which focuses on new simulation based market-timing tests.

[53] This limit is set arbitrarily in order to avoid extremes weights which might generate problems for a practical implementation.

allowed or 0 otherwise. We perform the optimization on a monthly basis, and then, we compute the returns over the next month. Thus, there is no look-ahead bias in the process.

We consider a set of 180 optimization specifications. They consist of a model to estimate the expected returns vector, the covariance matrix, an estimation window length used to estimate these parameters, which ranges from 126 to 1260 days, and whether short-sales are allowed or not.

3.1.2   The optimization inputs estimation

As in Jagannathan and Ma (2003), we estimate the optimization inputs with daily data. This enables us to use short estimation window lengths. We consider 10 estimation window lengths that range from 126 to 1260 days with an incremental step of 126 days. The number of assets, n, is 18.

3.1.2.1   The assets means vector

We use three different models to estimate the expected returns vector: The sample mean (S), the James-Stein (JS) and the Bayes-Stein (BS) estimators.

The ($n \times 1$) sample means vector, $\mu_{S,t,j}$, is estimated as

$$\mu_{S,t,j} = \frac{1}{j}\sum_{t=t-j}^{t} R_{assets,t}, \tag{2.2}$$

where R_{assets} is the ($n \times 1$) vector of assets returns at time t, j is one of the 10 estimation window lengths and $\mathbf{1}$ is a ($n \times 1$) vector of ones.

The James-Stein estimator is a shrinkage estimator, which tends the individual sample means toward a global mean to limit estimation errors. The aim is to reduce problems raised by the instability of means estimates during the in-sample optimization process and the out-of-sample evaluation:

$$\mu_{JS,t,j} = \mu_{S,t,j}(1 - w_{t,j}) + \mu_{G,t,j}w_{t,j}$$

$$w_{t,j} = \min\left[1, \frac{n-2}{j\left(\mu_{S,t,j} - \mu_{G,t,j}\mathbf{1}\right)'\Sigma_{S,t,j}^{-1}\left(\mu_{Si,t,j} - \mu_{G,t,j}\mathbf{1}\right)}\right], \tag{2.3}$$

where $\mu_{H,t,j}$ is the ($n \times 1$) vector of assets sample means at time t calculated over the last j observations, $\mu_{G,t,j}$ is the average of the n assets sample means and $\Sigma_{S,t,j}$ is the sample covariance matrix. $\mathbf{1}$ is a ($n \times 1$) vector of ones. Finally, $w_{t,j}$ is the shrinkage factor and it ranges between zero

and one. In these two extreme cases, a single asset expected return is respectively its sample mean or the average of all assets means.

The last estimator is introduced by Jorion (1986). It is similar to the James-Stein estimator, but uses the minimum-variance portfolio return, $\mu_{MV,t,j}$, as the shrinkage target[54].

$$\mu_{BS,t,j} = \mu_{S,t,j}(1-w_{t,j}) + \mu_{MV,t,j}w_{t,j} \quad with \ \mu_{MV,t,j} = \frac{\mathbf{1}'\Sigma_{S,t,j}^{-1}}{\mathbf{1}'\Sigma_{S,t,j}^{-1}\mathbf{1}}\mu_{S,t,j} \tag{2.4}$$

$$w_{t,j} = \frac{\lambda_{t,j}}{\lambda_{t,j} + j} \tag{2.5}$$

$$\lambda_{t,j} = \frac{(n+2)(j-1)}{\left(\mu_{S,t,j} - \mu_{MV,t,j}\mathbf{1}\right)'\Sigma_{S,t,j}^{-1}\left(\mu_{S,t,j} - \mu_{MV,t,j}\mathbf{1}\right)(j-n-2)}, \tag{2.6}$$

where $\mu_{G,t,j}$ is the in-sample minimum variance portfolio return.

Note that for the James-Stein and Bayes-Stein estimators, the previous models are used only for equity indices and not for bond indices in order to avoid the latter to influence the global mean estimation. More information about the data used in the empirical part is given in the next Section.

3.1.2.2   The assets covariance matrix

The covariance matrix is also estimated with three different models, the sample covariance matrix, the market model and the Ledoit and Wolf (2003) shrinkage estimator.

The sample covariance matrix is estimated as

$$\Sigma_{S,t,j} = \left(r_{t,j}'r_{t,j}\right)/j-1, \tag{2.7}$$

where $r_{j,t}$ is the (n x j) assets return matrix at time t containing the j last observations minus each asset respective mean calculated over these last j observations.

The market model, introduced by Sharpe (1963), is a single-factor model that limits the number of parameters to estimate, and the resulting covariance matrix is computed as

---

[54] This target is not an objective in itself, but results from the model derivation when a different loss function is used in the basic Stein estimator. The loss function used in Jorion (1986) is the square of the quadratic loss function used in the Stein estimator. It is more flexible to uncertainty about the exact from of the true loss function.

$$\Sigma_{MM,t,j} = \sigma_{m,t,j}^2 \beta_{t,j} \beta_{t,j}' + \sigma_{t,j}^2, \tag{2.8}$$

where $\sigma_{m,t,j}^2$ is the market variance at time t estimated over the last j days. $\beta_{t,j}$ is the vector of regression coefficients between each asset returns series and the market and $\sigma_{t,j}^2$ is the variance matrix of the regression residuals. For a specific estimation window length, the market model regressions are the following;

$$R_n = \alpha_n + \beta_n R_m + e_n \quad \text{for } n = 1,...,n, \tag{2.9}$$

where R_n and R_m are the return vector of asset n and the market and e_n is the vector of residuals.

The Ledoit and Wolf (2003) estimator has a similar structure to the James-Stein and Bayes-stein models for expected returns. It shrinks the sample covariance matrix toward the one resulting from the market model:

$$\Sigma_{LW,t,j} = w_{t,j} \Sigma_{MM,t,j} + (1 - w_{t,j})\Sigma_{S,t,j}, \tag{2.10}$$

where $\Sigma_{MM,t,j}$ is the market model covariance matrix calculated at time t over the last j days and $\Sigma_{S,t,j}$ is its sample one defined in equation (2.7). The shrinkage coefficient[55], $w_{t,j}$, define the weight given to each of these covariance matrices, diverging on the amount of structure given.

Finally, we use 10 different estimation window lengths, which range from six months to five years. They represent the interval used to estimate the parameters with the various models presented above. Using intervals shorter than five years could have some advantages. First, it is well known that returns distributions are time-varying. Thus, using shorter lengths may be more in line with economic cycles and market phases. This is also why we do not consider longer estimation lengths than 5 years[56]. Secondly, this may results in more dynamic strategies, which may exploit shorter trends and momentum in stock indices returns. Table 14 presents the 18 combinations regardless of the estimation window lengths in Panel A, while Panel B displays the set of specifications without short-sales at all.

---

[55] As the description of how to compute the shrinkage factor take about three pages, we refer the interested reader directly to the Appendix B in Ledoit and Wolf (2004).

[56] In addition, our database time span is limited, as bond indices data are not available before 1988. Using estimation lengths up to 10 years would limit our test interval to only 10 years.

Table 14: Optimization specifications

This table reports optimization specifications, without the estimation window lengths, *S* is respectively the sample mean or the sample covariance matrix, *MM* is the market model, *LW* the Ledoit and Wolf (2003) model, *JS* correspond to the James-Stein estimator and *BS* the Bayes-Stein estimator proposed by Jorion (1985). *Short* indicates whether short-sales are allowed in the optimization.

Panel A: With and without short sales

N°	E(R)	Cov	Short	N°	E(R)	Cov	Short	N°	E(R)	Cov	Short
1	H	H	no	7	JS	H	no	13	BS	H	no
2	H	MM	no	8	JS	MM	no	14	BS	MM	no
3	H	LW	no	9	JS	LW	no	15	BS	LW	no
4	H	H	yes	10	JS	H	yes	16	BS	H	yes
5	H	MM	yes	11	JS	MM	yes	17	BS	MM	yes
6	H	LW	yes	12	JS	LW	yes	18	BS	LW	yes

Panel B: Without short sales only

N°	E(R)	Cov	Short	N°	E(R)	Cov	Short	N°	E(R)	Cov	Short
1	H	H	no	4	JS	H	no	7	BS	H	no
2	H	MM	no	5	JS	MM	no	8	BS	MM	no
3	H	LW	no	6	JS	LW	no	9	BS	LW	no

However, we do not use these portfolios directly in the performance analysis. Indeed, selecting a few successful specifications would be strongly affected by data-mining issues. Thus, complex portfolios are constructed in a similar manner to the complex technical trading rules presented in the first Part of this thesis.

3.2   The construction of complex portfolios

In the portfolio optimization framework, the complex portfolios are mean-variance portfolios that employ different optimization specifications over time. We propose two out-of-sample selection processes to determine which of the 180 specifications is used every month.

This first step is similar for the two processes and thus, we do not repeat the computation. It consists in computing a time series of returns for the 180 "simple" portfolios optimized according to the specifications presented above. At the beginning of each month, we estimate the

parameters with the individual assets data up to the last day of the previous month[57]. As the shortest estimation length is 126 days, we only use daily returns to estimate the mean returns and the covariance matrices. Then we perform the optimization with these daily parameters estimations and obtain the weights. Finally, we compute the "simple" portfolios returns by holding these weights during the month. Note that we estimate the parameters with daily data, but we have a monthly rebalancing frequency. The process is repeated each month, which provides a "simple" portfolio monthly returns matrix, named *RM*, composed of 180 columns, the number of specifications, and 204 lines representing the number of months from July 1993 to December 2009[58]. This matrix actually includes all the required data to construct the complex portfolios returns, as the selection process will choose, each month, one return among the 180.

PF_ALL: At the beginning of each month of the test sample, this process selects the specification that yields the highest cumulated return up to the previous month. The cumulated returns are computed with the *RM* matrix detailed above. Then we compute the weights with the selected specification and hold them over the current month[59]. The selection sample consists in returns computed over all past returns, which means that the selection sample extends over time. The first year of data, from July 1993 is reserved for the selection process, thus, the test sample spans the period from July 1994 to December 2009. Note that the selection criterion does not include risk, as it is already controlled in the optimization algorithm with the maximum level of variance. Thus, more risk averse investors should set a lower level of maximum variance and then try to maximize the return.

For example, we are currently at the end of month *T*. First, we compute the cumulated return of each portfolio obtained with the 180 specifications in the matrix *RM* from the month 0, July 2003, to the end of month *T*. Then, we select the specification with the highest return. Finally, we compute the portfolio weights with this specification and use them as the actual investment strategies over the month *T*+1. Then at the end of the month *T*+1, we repeat the procedure. Thus, the new test sample lasts from month 0 to the end of the month *T*+1.

---

[57] This process insures that the optimization is not affected by any look-ahead bias.

[58] As the "simple" portfolios generated with a short estimation length have a longer return time series than those with a longer one, we only start recording the returns when all specifications generate returns from July 1993.

[59] Note that we could also take the returns in the *RM* matrix directly. However, we perform the optimization again to control that we have a similar result.

PF_4: The second process follows a similar procedure. However, instead of selecting the "simple" portfolio specification that yields the highest cumulated return over the entire returns history, the selection sample consists in the four previous years[60]. Thus, the selection sample do not expend over time. Note that a new specification can be selected each month; we do not keep it for an extended test sample, as it is the case for the technical analysis strategy.

Finally, these two complex portfolios are also examined when short-sales are entirely forbidden. Thus, the number of optimization specifications is reduced to 90, and we refer to them as PF_ALL_NS and PF_4_NS.

4.   Data

Our data set consists of 11 national equity indices, the world equity index and six bond indices. In order to have similar series, we use the equity indices constructed by *Thomson Reuters Datastream*. The equity indices are: USA, UK, France, Germany, Switzerland, Italy, Spain, Netherland, Sweden, Hong Kong, Japan and the world index. The national indices comprise stocks that represent a minimum of 75% or 80% of the total market capitalisation. 6330 stocks are included in the world index in 2008. The inclusion of a stock depends on its market value and its data availability; however, liquidity and off-exchange shares holdings are not taken into account. Financial or special stocks such as warrants, units trust, investment funds or foreign listings are not taken into account neither. The indices composition is re evaluated on a quarterly basis. Finally, they are value weighted, which means that each stock weight in the index is proportional to its market value.

The first bond index is the Bank of America Merrill Lynch US Corp AA-AAA 5-10Y. It consists of all US corporate bonds rated AA or AAA, which represent the highest rating, with a maturity between five and 10 years. The second one also regroups US corporate bond, but with a BBB rating, which implies that the companies are able to meet their financial obligation but they are more risky in case of difficult economic situation. This is the Citigroup US BIG Corp BBB 1-10Y. BIG stands for Broad Investment Grade and means that bonds are at least rated BBB- and have a maturity of a least one year. The third one, also provided by Citigroup, is the US BIG

---

[60] This four year selection interval is chosen in order to match the technical analysis strategies, so it is subjective. In addition, the two processes are similar during the first four years of the test sample.

Gvt/Corp 1-3Y. It combines both bonds emitted by the US government and by US corporations. The last three bond indices are exclusively composed of bonds issued by national governments. They are the JPMorgan France bond index, Bank of America Merrill Lynch UK Gilts 5-7Y and the Swiss Gvt 1-10Y. Note that these bond indices are more selected according to their data availability than any other criteria. Nonetheless, we think that they are fairly representative of the main investment opportunities in fixed income.

For these 18 assets, we use the total returns indices that include dividend payment in the price series used to compute the simple daily returns. As mentioned before, the test sample lasts from the end of June 1994 to the end of December 2009. Data from September 1988 is used for estimation purposes. As a benchmark, we construct an equally weighted portfolio with the 12 equity indices. It corresponds to the 1/N strategy, which is found to be better than most optimum portfolios examined in DeMiguel, Garlappi and Uppal (2009). We choose not to include bonds in the benchmark strategy, as we consider them in order to give the optimization algorithm an investment opportunity when equity markets face a bear market. Thus we argue that an all equity benchmark is more representative for comparing our complex portfolios. All price series are extracted from *Thomson Reuters Datastream* and are stated in USD.

Table 15 displays the descriptive statistics for each individual asset over the test sample. The statistics are computed with daily simple returns, while they are reported on an annual basis. Note that all *Jarque-Bera* tests reject the null hypothesis that daily returns follow a normal distribution at very high confidence levels.

Table 15: Descriptive statistics with daily returns

This table reports the descriptive statistics for all indices over the test period ranging from June 1994 to December 2009. Computation are performed with daily data, however the presentation of results varies as stated thereafter. *Mean* is the annualized mean return, *Min* and *Max* are the minimum and maximum daily return over the sample, *Vol* is the annualized volatility, *Sk* is the skewness and *Ku* the kurtosis, which respectively measure the asymmetry of the distribution and the thickness of its tails. They are estimated with daily returns. A normal distribution has a skewness equal to 0 and a kurtosis equal to 3. We test whether our assets returns series follow a normal distribution with the Jarque-Bera statistic, *JB*, reported in the last column.

	Statistics						
	Mean	Min	Max	Vol	Sk	Ku	JB
Equity indices							
USA	0.100	-0.090	0.115	0.197	-0.033	11.548	12275
UK	0.098	-0.099	0.126	0.201	0.076	14.328	21561
France	0.119	-0.101	0.113	0.215	0.123	10.429	9282
Germany	0.095	-0.082	0.177	0.213	0.473	14.964	24199
Switzerland	0.091	-0.103	0.119	0.229	0.105	9.447	6990
Italy	0.108	-0.108	0.108	0.218	-0.075	11.552	12289
Spain	0.137	-0.091	0.109	0.213	0.018	9.290	6647
Netherlands	0.150	-0.098	0.143	0.280	0.238	8.729	5552
Sweden	0.108	-0.068	0.095	0.181	0.085	8.464	5021
Hong Kong	0.116	-0.127	0.169	0.261	0.273	12.720	15922
Japan	0.011	-0.084	0.122	0.230	0.196	7.162	2936
World	0.080	-0.064	0.085	0.149	-0.267	11.463	12081
Bond indices							
US AA-AAA	0.066	-0.037	0.030	0.058	-0.481	9.822	7973
US Big	0.066	-0.018	0.014	0.042	-0.467	7.260	3196
US Gvt	0.051	-0.006	0.009	0.016	-0.024	8.053	4289
JPMorgan France	0.076	-0.039	0.056	0.107	0.309	6.598	2239
UK Gilts	0.076	-0.039	0.048	0.097	0.086	7.778	3840
Swiss Gvt	0.063	-0.044	0.058	0.116	0.307	6.292	1884

Table 16: Descriptive statistics with monthly returns

This table reports the same statistics as in Table 15 but they are computed with monthly returns. Thus, Min and Max are the minimal and maximal monthly return over the test sample from 1994 to 2009.

	Statistics						
	Mean	Min	Max	Vol	Sk	Ku	JB
Equity indices							
USA	0.095	-0.281	0.208	0.187	-0.688	8.903	312.3
UK	0.100	-0.276	0.209	0.192	-0.804	7.575	199.9
France	0.122	-0.263	0.209	0.214	-0.649	5.422	64.2
Germany	0.105	-0.263	0.215	0.217	-0.542	5.208	51.4
Switzerland	0.112	-0.297	0.341	0.261	-0.033	6.139	83.8
Italy	0.121	-0.367	0.232	0.233	-1.096	8.821	328.9
Spain	0.146	-0.220	0.239	0.224	-0.083	4.921	31.6
Netherlands	0.161	-0.287	0.301	0.275	-0.117	5.187	41.1
Sweden	0.122	-0.228	0.179	0.189	-0.547	5.329	56.3
Hong Kong	0.132	-0.226	0.282	0.280	0.043	4.136	11.0
Japan	0.017	-0.229	0.181	0.214	-0.041	3.932	7.4
World	0.089	-0.278	0.215	0.182	-0.704	8.568	280.4
Bond indices							
US AA-AAA	0.062	-0.136	0.051	0.067	-2.083	15.909	1563.9
US Big	0.063	-0.114	0.061	0.055	-1.887	18.871	2262.0
US Gvt	0.049	-0.013	0.020	0.017	-0.266	3.615	5.6
JPMorgan France	0.080	-0.072	0.101	0.112	0.295	2.918	3.0
UK Gilts	0.075	-0.061	0.076	0.096	-0.052	2.718	0.8
Swiss Gvt	0.071	-0.064	0.111	0.126	0.520	2.823	9

5.   Empirical results

5.1   The optimization specifications

The sole objective of these preliminary results is to show the influence of the length of the estimation window used in the optimization problem. Hence, we do not provide all detail results for each optimization specification, but we only show them graphically. Furthermore, the "simple" portfolios returns are only indicative and do not represent an investment strategy, as they are in-sample.

Panel A of Figure 16 displays the annual mean returns for each of 180 portfolios over the July 1993 to December 2009 sample. They lie between 6.53% for the (JS-LW-S-756)[61] specification and 28.06% for the (JS-MM-S-252). This figure suggests that the length of the estimation window influences to a larger extent the performance of optimized portfolio than the parameters estimation models. Panel B shows the same results with two subsamples to highlight this finding. The first (second) one consists in specifications that (do not) consider the possibility of shorting each asset up to 50% of the entire capital. The respective importance between the estimation window length and the various estimation models is even more striking. It is also important to note that restricting short-sales has a strong negative impact of the optimized portfolios performance.

---

[61] The specifications are presented in the following manner: The expected return and covariances estimation models as in Table 14, whether short-sales are allowed (S) or not (NS) and the estimation window length in days.

Figure 16: Estimation window lengths and parameters estimation models

Estimation window refers to the number of days used to estimate the parameters, and Model is the combination of estimation methods used. Annual mean returns are computed over the July 1993 – December 2009 sample. Transaction costs are ignored. Note that in Panel B, the scales are not similar in order to keep the visual representation meaningful.

Panel A: All specifications
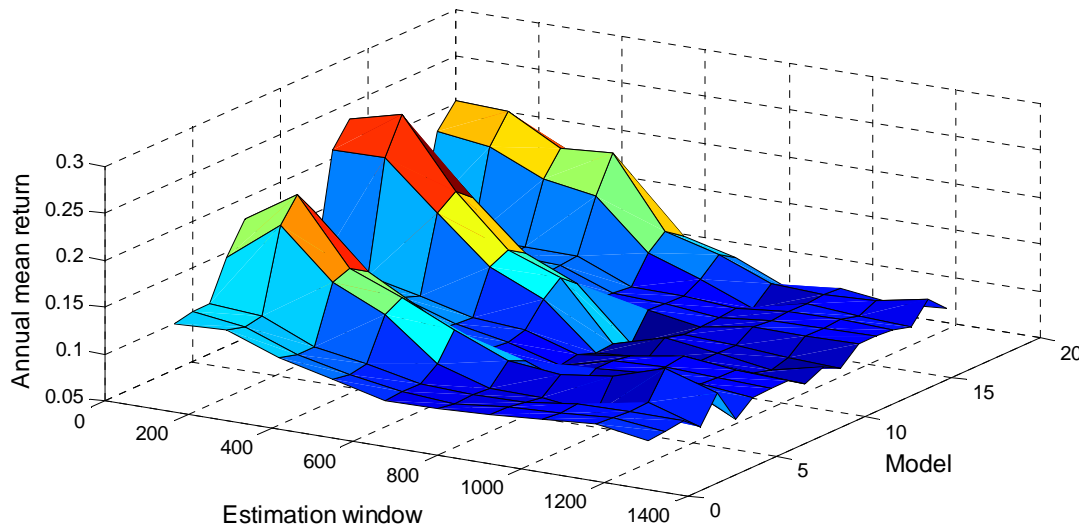
Panel B: With or without short-sales
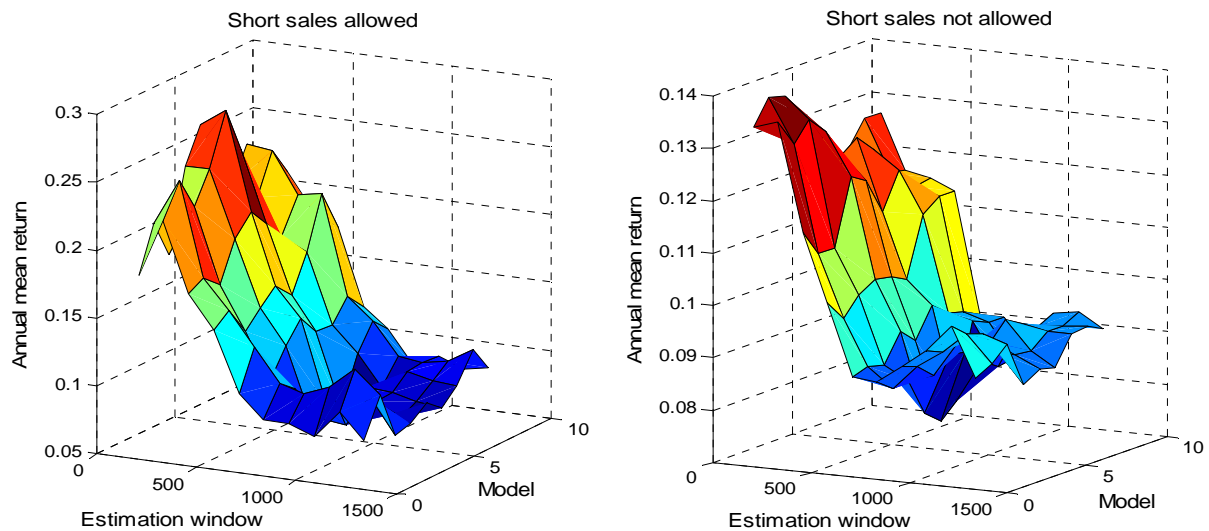
In order to verify the relative importance between the estimation window length and the various estimation models explicitly, we perform portfolios based tests. We combine the portfolios together according to their estimation window length in four groups. The first consists of all portfolios with the five shortest estimation windows, while the second groups the ones with

the five longest together. Thus, each of them is made up of 90 specifications. The third considers only the 18 specifications with the shortest length (6 months), and the last one, the 18 specifications with the longest length (5 years). They are all equally weighted. Table 17 contains their annual mean returns and the associated Student t-test p-values resulting from tests of equal means. We find that the two groups of portfolios with shorter window lengths yield higher returns than their longer lengths counterparts. The returns differences are economically large, as they amount to 5.97% and 8.56%, compared with 7.98% for the World index during this sample. In addition, they are all significant at the 1% level. The results are also consistent for the subsample analysis displayed in Panel B and C.

The last Panel reports the results for similar groups of portfolios constructed with the set of 90 specifications that do not allow short-sales at all. The conclusion is similar; nonetheless, the differences in annual means drop sharply. Indeed, they are 2.33% and 3.70% respectively. Nevertheless, p-values of 0.000 indicate that they are still significant at very high confidence levels. These findings contrast with Garlappi, Uppal and Wang (2007) and Kritzman, Page and Turkington (2010) who argue that a long estimation window provides the best performance, and with other studies that follow the standard methodology with a fixed five years estimation window.

Table 17: Estimation lengths versus parameters estimation models

Computations are performed with daily return. Mean return is the annualized simple mean return of a portfolio. The 5 short (long) portfolio contains the specifications with the 5 shortest (longest) window lengths and all estimation models. The 1 short (long) portfolio includes only the estimation models with the shortest (longest) estimation window length. *P*-values are obtained by standard bilateral Student *t*-tests.

Panel A: Entire sample July 1993 - December 2009

	5 short	5 long	Diff	1 short	1 long	Diff
Mean return	0.1484	0.0887	0.0597	0.1734	0.0879	0.0856
P-value			*0.000*			*0.000*

Panel B: First subsample July 1993 - October 2001

	5 short	5 long	Diff	1 short	1 long	Diff
Mean return	0.1579	0.0878	0.0701	0.1959	0.0848	0.1111
P-value			*0.000*			*0.000*

Panel C: Second subsample October 2001 - December 2009

	5 short	5 long	Diff	1 short	1 long	Diff
Mean return	0.1390	0.0897	0.0493	0.1510	0.0909	0.0601
P-value			*0.000*			*0.000*

Panel D: Entire sample: No short sales allowed

	5 short	5 long	Diff	1 short	1 long	Diff
Mean return	0.1130	0.0897	0.0233	0.1271	0.0900	0.0370
P-value			*0.000*			*0.000*

We also form similar portfolios according to the parameters estimation models specifications reported in Table 14. To isolate the effect of the estimation models, we compute the portfolios either with the set of nine specifications that allow or restrict short-sales. Table 18 reports the differences in annual mean returns. First of all, the means of these differences, in absolute value, are only 1.19% and 0.18% respectively when short sales are allowed or not. This first result indicates that when using a similar set of estimation window lengths, the models used to estimate the mean return and the covariance matrix have no influence, statistically or economically on the returns generated. Indeed, these differences are much lower than differences between portfolios formed according to the estimation lengths. Furthermore, none of these differences is statistically significant according to bilateral Student *t*-tests at the usual 5% confidence level. This is in line with Liu and Lin (2010) who also find basic model to perform well when the number of asset is small relative to the sample length.

Table 18: Difference in portfolios returns formed according to the estimation models

This table reports the annualized differences in mean simple returns between the portfolios formed according to the specification denoted PF A and those denoted PF B. The first set of letters indicates the method to calculate the mean returns, *H* being the historic mean, *JS* the James-Stein estimator and *BS* the Jorion (1985) model. The second set relates to the covariance matrix estimation, with *H* the sample covariance matrix, *MM* is those resulting from the market model and *LW* from the Ledoit and Wolf (2003) model. For instance, the portfolio denoted H-H regroups the 10 simple portfolios generated with the historic mean and the sample covariance matrix with the 10 windows estimation lengths. In panel A, short sales are allowed, but not in panel B. None of these differences is statistically different from zero at a 5% confidence level.

Panel A: Only the specifications with short sales

		PF B							
		H-H	H-MM	H-LW	JS-H	JS-MM	JS-LW	BS-H	BS-MM
	H-MM	-0.001							
	H-LW	-0.001	0.000						
P	JS-H	-0.003	-0.001	-0.002					
F	JS-MM	-0.004	-0.003	-0.004	-0.002				
	JS-LW	-0.004	-0.003	-0.003	-0.002	0.000			
A	BS-H	-0.001	0.001	0.000	0.002	0.004	0.004		
	BS-MM	-0.001	0.000	0.000	0.002	0.003	0.003	-0.001	
	BS-LW	-0.001	0.000	0.000	0.002	0.003	0.003	-0.001	0.000

Panel B: Only the specifications without short sales

		PF B							
		H-H	H-MM	H-LW	JS-H	JS-MM	JS-LW	BS-H	BS-MM
	H-MM	-0.001							
	H-LW	-0.001	0.000						
P	JS-H	-0.003	-0.001	-0.002					
F	JS-MM	-0.004	-0.003	-0.004	-0.002				
	JS-LW	-0.004	-0.003	-0.003	-0.002	0.000			
A	BS-H	-0.001	0.001	0.000	0.002	0.004	0.004		
	BS-MM	-0.001	0.000	0.000	0.002	0.003	0.003	-0.001	
	BS-LW	-0.001	0.000	0.000	0.002	0.003	0.003	-0.001	0.000

Finally, we report the estimation lengths selected by the four selection processes used to construct the complex portfolios. Figure 17 presents these lengths as a histogram and shows clearly that the two processes result in selecting short estimation window lengths very often. The two longest lengths are even never chosen. This suggests that the selection process is accurate in the light of these preliminary results, as portfolios with a short estimation window statistically outperform those with a long estimation window. However, more formal evidence is given at the end of the next Section.

Figure 17: Estimation window lengths selected

This figure illustrates, with histograms, the frequency of the estimation window lengths selected by the four procedures used to create complex portfolios explained in chapter 3.2 of this current part. The total amount of selection periods is 192.
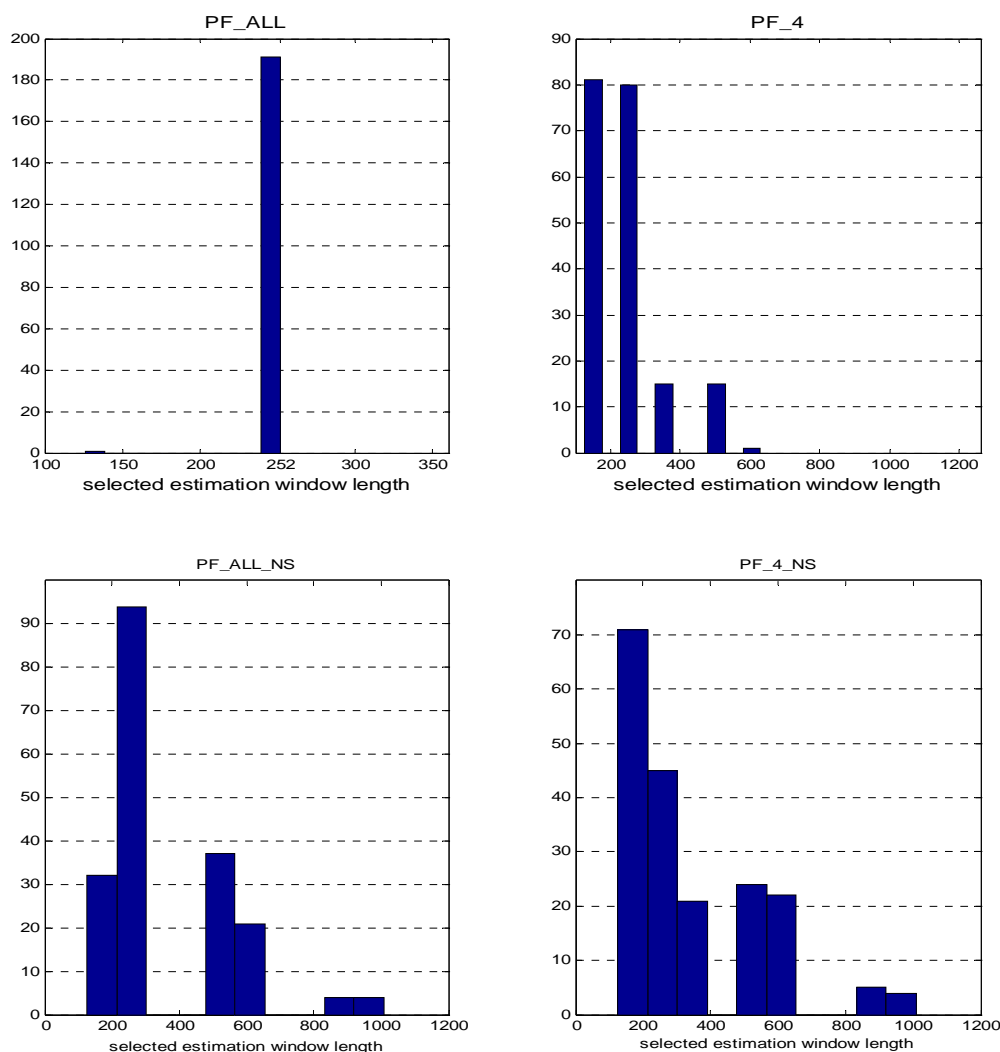
5.2 Complex portfolios performance

Table 19 displays the performance analysis for the four complex portfolios. Panel A presents various performance measures for the entire sample, while Panel B and C display those over two subsamples. The two first portfolios, PF_ALL and PF_4, produce returns that are more than twice as high as the equally weighted benchmark strategy, 1/N, over a 16 years sample. Indeed, the annual mean buy-and-hold return is 10.31% compared with 25.68% and 24.56% respectively for these portfolios.

Table 19: Portfolios performance analysis

1/N, the benchmark strategy, is an equally weighted portfolio including only equity indices. The four other columns are the complex portfolios described in Section 3.2. All results are given in an annual frequency (except *Nb* trade and *BE TC*, which represent respectively the number of transactions and break even transaction costs). *P*-values associated with simple mean returns are obtained with bilateral Student *t*-tests between the respective portfolio and the buy-and-hold. In contrast, the Jensen's alphas tests determine whether coefficients are statistically different from zero.

Panel A: Entire sample June 1994 - December 2009

	1/N	PF_ALL	PF_4	PF_ALL_NS	PF_4_NS
Simple mean return	0.1031	0.2568	0.2456	0.0965	0.0965
P-value		*0.0350*	*0.0562*	*0.9054*	*0.9031*
Compounded mean return	0.0928	0.2566	0.2401	0.0903	0.0913
Volatility	0.1690	0.2376	0.2462	0.1415	0.1350
Jensen's alpha		0.1963	0.1772	0.0207	0.0202
P-value		*0.0008*	*0.0030*	*0.4727*	*0.4471*
Sharpe ratio	0.3098	0.8672	0.7916	0.3236	0.3390
Nb trade		311.49	365.44	115.73	117.33
BE TC		0.0079	0.0062	-0.0009	-0.0009
Beta		0.1856	0.3381	0.4792	0.4882

Panel B: First subsample June 1994 - March 2002

	1/N	PF_ALL	PF_4	PF_ALL_NS	PF_4_NS
Simple mean return	0.0995	0.2621	0.2755	0.0810	0.1093
P-value		*0.1108*	*0.0815*	*1.2054*	*0.8887*
Compounded mean return	0.0940	0.2587	0.2766	0.0731	0.1044
Volatility	0.1390	0.2527	0.2497	0.1444	0.1413
Jensen's alpha		0.2008	0.2072	-0.0017	0.0272
P-value		*0.0222*	*0.0152*	*1.0374*	*0.4500*
Sharpe ratio	0.4310	0.8805	0.9448	0.2871	0.4933
Nb trade		167.03	176.00	62.85	66.53
BE TC		0.0078	0.0080	-0.0023	0.0012
Beta		0.1075	0.2767	0.3613	0.3804

Panel C: Second subsample March 2002 - December 2009

	1/N	PF_ALL	PF_4	PF_ALL_NS	PF_4_NS
Simple mean return	0.1067	0.2515	0.2157	0.1120	0.0837
P-value		*0.1646*	*0.3212*	*0.9494*	*1.2196*
Compounded mean return	0.0917	0.2545	0.2045	0.1078	0.0784
Volatility	0.1944	0.2215	0.2427	0.1386	0.1284
Jensen's alpha		0.2047	0.1576	0.0482	0.0186
P-value		*0.0084*	*0.0600*	*0.2515*	*0.6155*
Sharpe ratio	0.3452	0.9566	0.7259	0.5229	0.3439
Nb trade		141.30	189.77	53.20	51.11
BE TC		0.0082	0.0046	0.0008	-0.0036
Beta		0.1075	0.2767	0.3613	0.3804

Their associated *p*-values indicate that they are statistically different from the benchmark return at high confidence levels, i.e. 3.5% and 5.62%. Nonetheless, the Pf_4 *p*-value is higher than the standard 5% level, despite its large excess return. This is in close agreement with the technical analysis strategies that generate an economically significant performance, but not statistically. This contrast is strengthened in the subsample analysis. Over the first subsample, the difference between the PF_ALL and the benchmark portfolio returns is even slightly higher than during the whole sample; however, the *p*-value is 11.08%, a sharp increase from those of the entire sample, which is 3.5%. For the PF_4 portfolio, the difference[62] in mean annual returns between the strategy and the 1/N benchmark increase from 14.25% over the entire sample to 17.6% over the first subsample. However, for this subsample, the *p*-value is 8.15%, while it is 5.62% over the entire sample and with a smaller difference in mean returns. As shown in the beginning of the next part, this is due to the Student *t*-test lack of power, especially as the number of observation decreases. The PF_4 portfolio over the second subsample illustrates this issue well; its return is twice as high as the benchmark, but the *p*-value is 32.12%, which is very far from standard confidence levels. In other words, the standard Student *t*-test requires very large excess return in order to conclude in favour of statistical significance between the mean of two return series. In the third part of this thesis, we also develop a simulation test based on artificial portfolio weights to address this issue.

In the previous subsection, we find that restraining short-sales decreases the "simple" portfolios returns sharply. This also holds for the complex portfolios, as the PF_ALL_NS and the PF_4_NS mean returns are slightly lower than the 1/N benchmark strategy. Nonetheless, these two portfolios volatility is lower than the benchmark as well, which is reflected in slightly higher Sharpe ratios. Thus, these strategies may appeal to more risk averse investors.

Jensen's alphas and Sharpe ratios provide evidence supporting that the performance of the two complex portfolios with short-sales is not due to additional risk bearing. First, the alphas amount to 19.62% and 17.72% in annual terms when negative weights are allowed in the optimization process. In addition, they are statistically different from zero at very high confidence levels. The subsamples show that the strategies positive alphas are consistent over time. Furthermore, they provides an interesting comparison between *p*-values obtained with the Student *t*-test of equal means and those applied to determine whether alphas are significantly

---

[62] The difference in mean returns is simply the difference between the annual mean return of the 1/N and PF_4 strategy, i.e., 24.56%-10.31% over the entire sample and 27.55%-9.95% over the first subsample.

different from zero. While we find here above that the difference in mean returns between the PF_ALL and PF_4 and the 1/N benchmark is not statistically different over the subsamples, the alphas p-value are still below the usual 5% limits over the subsamples, except 6% for PF_4 over the second subsample. Jensen's alphas for strategies that do not allow short sales confirm the lack of excess return of the two strategies over the benchmark, as their p-values are well above the usual 5% limit, even if most of the alphas are slightly positive.

These results are confirmed by Sharpe ratios. For the two complex portfolios that allow short-sales, they are 0.86 and 0.79, while the benchmark ratio is 0.31 only. Nonetheless, the variance constraint is not respected with annual values of 23.7% and 24.6%, while it is limited to 20% in the optimization process. This is due to non stationary return distributions between the in-sample optimization and the actual out-of-sample performance. Restraining short-sales seems to resolve this problem, but at a high cost in term of returns.

In contrast with the technical analysis strategies, transaction costs suggest that the complex portfolios can not challenge the EMH. Indeed, they would reduce returns to a significant extent, as indicated by relatively low break even transaction costs[63] of 0.79% and 0.62%. The number of trades, or transactions, is calculated as the sum of the differences between the assets weights vectors at time t and $t-1$ over all periods. Note that a vector of zero is used as the first value in order to consider the first investment in the number of trades, and thus, in the break-even transaction cost as well. This also implies that the number of trades over the entire sample is not equal to the sum of the two subsamples. For instance, 150 trades mean that the capital is traded 150 times.

In order to provide an investor perspective, i.e. using compounded returns, Figure 18 displays the value of one USD invested in the benchmark and in the first two complex portfolios. The final value is 4.13 USD for the benchmark, which is much lower than the 38.6 and 31.2 USD for the two portfolios. Nonetheless, transaction costs are not taken into account in this figure.

---

[63] For a practical implementation of the proposed optimization process, transaction cost might be limited by using fewer indices as underlying assets or limiting trading until differences in weights reach a minimum level.

Figure 18: Value of 1 USD calculated with compounded returns

1/N, PF_ALL and PF_4 represent the respective value of 1 USD invested in the 1/N benchmark portfolio or according to the two proposed optimization processes with short-sales. Compounded returns are used but neither transaction costs nor foreign exchange influences.
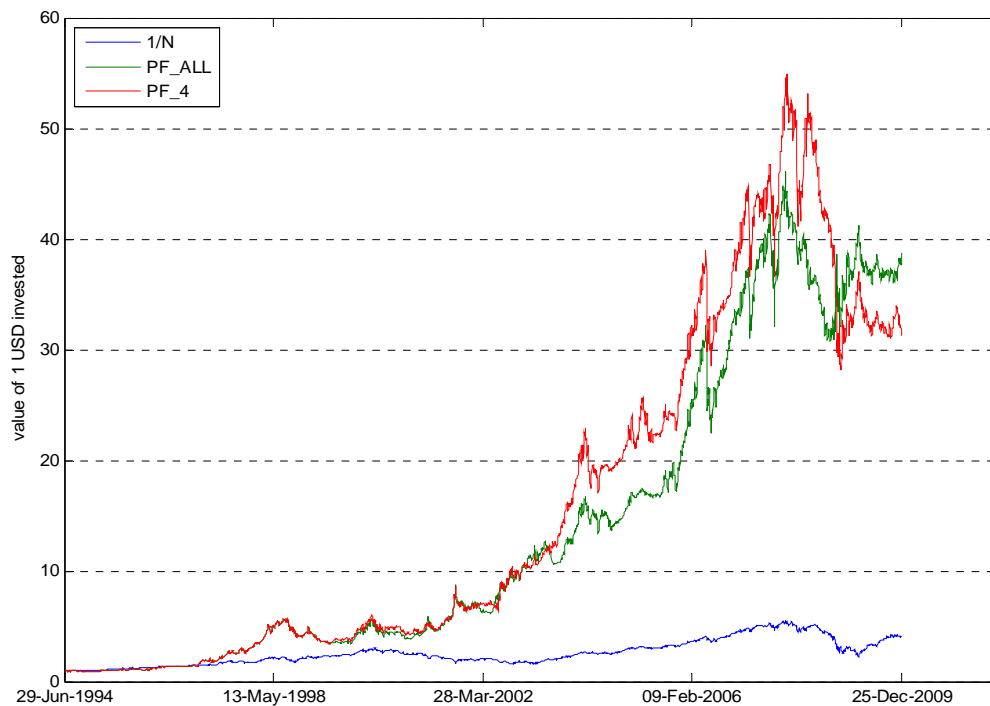
To further illustrate the complex portfolios, Figure 19 reports the percentage invested in all equity indices, for the PF_4 portfolio, and the evolution of the world equity index. It suggests that the complex portfolio invests according to the market phases. Indeed, it switches to a global short position at the beginning of the two bear markets, in early 2000 and in late 2008. However, it misses the market rebound that started in late March 2009. This means that using short estimation intervals correspond to follow market phases, at least to some extent. This is in close agreement with the results presented in the first part about MA rules, which also profit from market phases. Note that more than 300% of the capital is invested in equities at some times, and this is possible as the diversification constraint states that shorting an individual asset is possible up to 50% of the capital. Thus, as the data set contains six bond indices, a maximum of 400% of the capital may be invested in equities. However, the issue of extreme weights is moderate to some extent as the volatility is constrained in the optimization process, thus shorting bonds to invest in equity is possible as long as the total portfolio volatility does not exceed the maximum level in the optimization process.

Figure 19: Allocation in equity and the World index

The first graph displays the sum of the PF_4 portfolio weights invested in equity indices, while the second present the evolution of the World index.
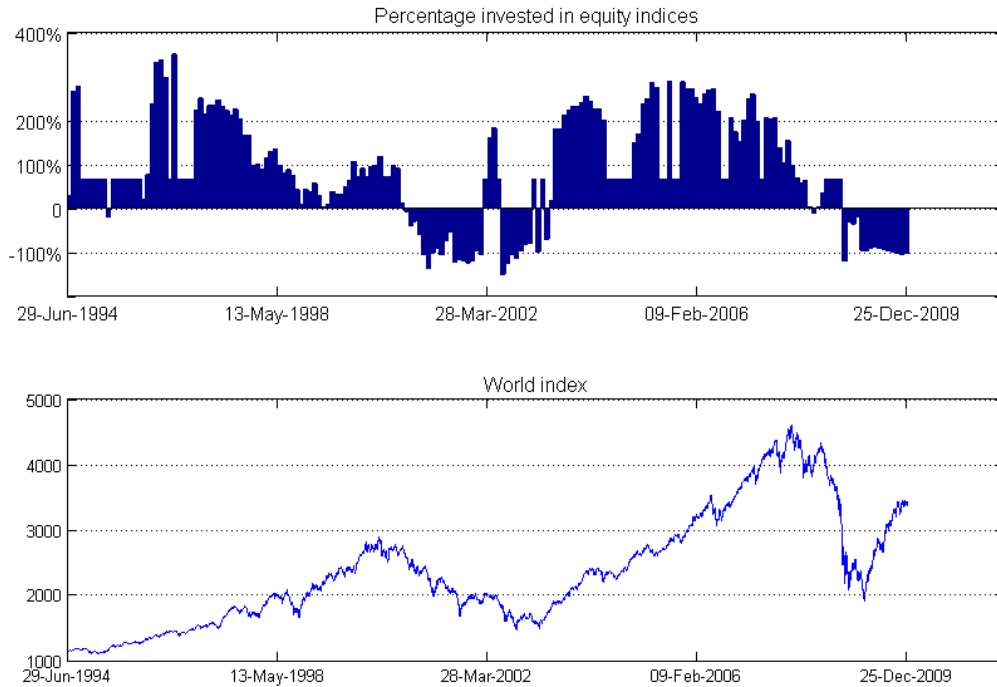
Finally, we propose the following method to evaluate the parameters selection procedure used by the PF_ALL and the PF_4 complex portfolios. As explained in chapter 3.2 of the current part, the complex portfolios consist in selecting each month one of the 180 available specifications. Thus, we propose to compare the actual selection made by these two complex portfolios with random choices. We base this test on a performance statistic named *Perf*. For each month, we compute the percentage of non selected specifications that yield a higher return than the selected one over the next month. Then, the *Perf* measure is the median of these percentages[64]. Obviously, the lower is this statistic; the more successful is the selection procedure. These statistics are 21.79% for PF_ALL and 21.51% for PF_4. This means that on average about 27% of the specifications not selected by the complex portfolios yield a return higher than the selected one over the next month. To determine whether these levels are lower that what could be archived by luck, we propose a bootstrap technique to calculate *p*-values. First, we generate 1'000 artificial

---

[64] For instance, a value of 0.4 means that, on average, 40% of the non selected optimization specifications produce a return higher than the strategy.

portfolios by choosing, each month, one of the 180 specifications randomly. Then, the *Perf* statistic is computed for each of these portfolios[65]. They represent the statistic distribution under the null hypothesis that the original strategy is not better than a random selection. Thus, the *p*-value is the percentage of simulated *Perf* statistics with a lower value than the original one. We find that the two selection processes performs better than what could be expected by luck. Indeed, the *p*-values are 0 for the two complex portfolios, meaning that none of the artificial strategies could yield a *Perf* statistic lower than the actual two complex portfolios. The performance is similar for the two processes, which is consistent with results obtained for technical analysis. To further illustrate our simulation test, Figure 20 presents the histogram of the *Perf* statistics of the simulated portfolios.

Figure 20: Histogram of simulated *Perf* statistics

This histogram presents the frequency of the 1'000 Perf statistics computed with simulated portfolios. None of them are lower than the current statistic of 21.79% for PF_ALL and 21.51% for PF_4.

---
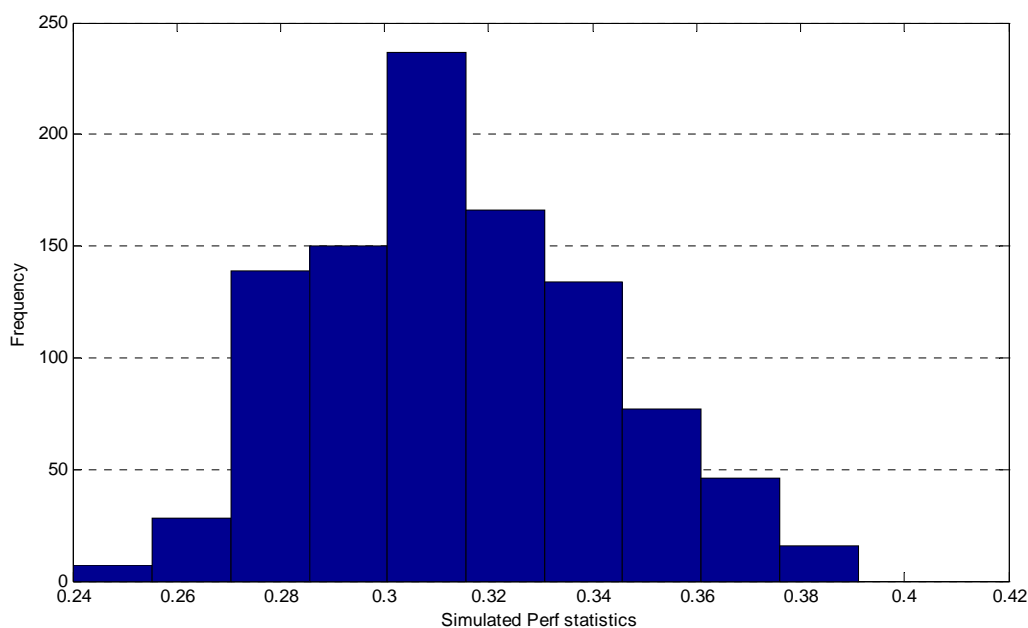
[65] The median *Perf* statistic obtained by the simulations is lower than 50%, i.e. 31.28%, as some specifications have similar returns. This downside bias is due by computing the percentage of the non selected specifications that have a *higher* return and not a higher or equal return. However, doing this would introduce an upside bias in the statistics. The test is repeated with this second calculation. The *Perf* statistics increase for the original methods (53.91% and 56.15%) and for the simulated statistics (median of 67.68%). Nonetheless *p*-values are still 0.

6.   Conclusion

To conclude this Section about portfolio optimization, we want to emphasise that this part is not a comprehensive guide about how implementing portfolio optimization in practice. We rather investigate issues that are not treated often in the related literature. First, the estimation window length in portfolio optimization should not be overlooked, and second, using complex procedures to select the specifications to be used in the optimization algorithm can be useful to incorporate more information. Indeed, the proposed complex portfolios, which rely on short estimation window lengths, are able to generate large annual mean returns, i.e. 25.6% and 24.5%, over a long sample compared with the equally weighted benchmark portfolio, which yields only 10.3%. However, these results hold within our subjective investment universe and time period only when short-sales are allowed and transaction costs are ignored. Furthermore, our portfolios weights may seem unrealistic for professional fund managers, as they short bond indices to leverage up to three times the available capital in equity indices at some points. There are various leads to address these issues. For instance, we may use less indices to reduce the transaction costs influence and add further diversification constraints to avoid extreme weights over the two assets classes.

Our results are in contrast with some recent studies such as DeMiguel, Garlappi and Uppal (2009), who conclude that portfolio optimization is not able to generate consistently higher returns than an equally weighted portfolio. Contrary to this study, they use longer estimation window lengths, either 60 or 120 months. Secondly, it is noteworthy that the results about portfolio optimization are in line with those of the first part of this thesis about technical analysis. Both consider trends ranging from one to two years, and they find significant higher performance than their respective benchmarks. This contrasts with the vast majority of studies about technical analysis and portfolio optimization. The former consider only short-term trends (only up to one year), and the latter estimate the parameters almost always over a five years estimation window.

Finally, we also show that large economical differences are not always confirmed statistically. Indeed, even if our complex portfolios produces annual mean returns that more than twice higher than the benchmark, Student *t*-test of equal means are usually not able to conclude in favour of a statistically difference in means. This issue is much more pronounced over shorter subsamples. Thus, in the last part of this thesis, we develop a more powerful test procedure, and we re-examine the findings presented so far, both for technical analysis and portfolio optimization.

7.   Appendices part II

Appendix II-A: Estimation lengths and models: Subsamples

Estimation window refers to the number of days used to estimate the parameters and *Model* is the combination of estimation methods used. Subsample 1 lasts from 13 July 1994 to 01 October 2001 while the second subsample finishes on the 25 December 2009.

Subsample 1: 13 July 1994 - 01 October 2001

Subsample 2: 02 October 2001 - 25 December 2009
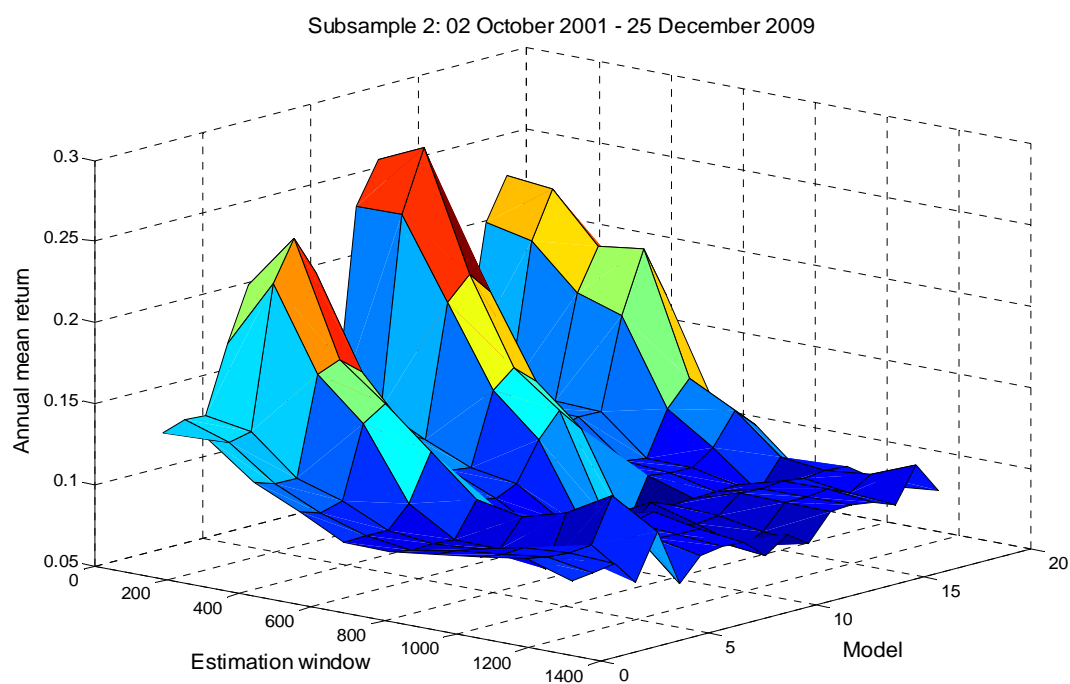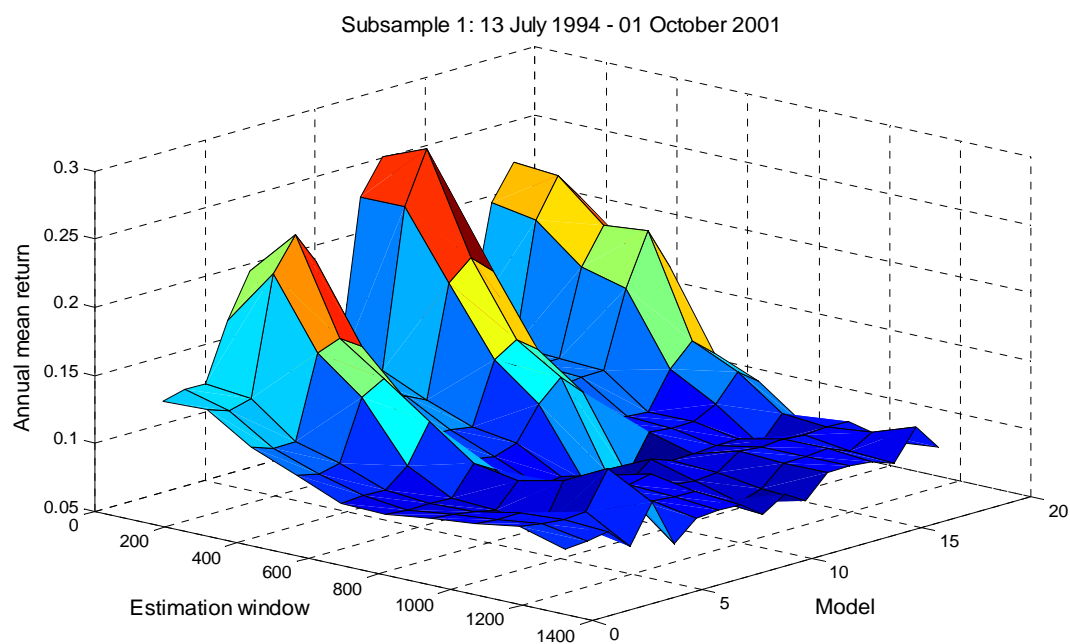
Appendix II-B: With emerging markets indices: A selection of results

In this Appendix, we present some results related to a larger investment universe, which include some emerging market equity indices. The following Table displays the new indices descriptive statistics. Compared with the other indices presented in Table 15, the addition of these countries should not be particularly beneficial. Indeed, their annual returns are even lower than the developed markets, as they dropped sharply at the end of our sample during the 2008-2009 financial crisis. This clearly contrasts with earlier studies supporting the profitability of investing in emerging markets. Nonetheless, this new sample may also be considered as a robustness analysis.

Table 20: Descriptive statistics

Mean is the annual mean return, *Min* and *Max* are the minimum and maximum daily return, *Vol* is the annualized volatility, *Sk* the skewness, *Ku* the kurtosis and *JB* the Jarque-Bera statistic.

	Statistics						
	Mean	Min	Max	Vol	Sk	Ku	JB
Equity indices							
Korea	0.1347	-0.1947	0.3083	0.3998	0.8254	19.8866	48364.3
Malaysia	0.0672	-0.3076	0.2585	0.2762	0.1858	59.4654	535665
Singapore	0.0759	-0.0909	0.1121	0.2159	0.1067	9.6386	7411.6
Thailand	0.0520	-0.1629	0.1779	0.3332	0.5248	11.2962	11747.9
South Africa	0.1324	-0.1348	0.1288	0.2716	-0.3952	9.6615	7560.1
Australia	0.1364	-0.1475	0.0876	0.2226	-0.6885	14.2653	21639.0

First, we show that the relative importance between the estimation window length and the model used to estimate the optimization inputs is similar.

Table 21: Estimation lengths versus parameters estimation models

The notation is similar to Table 17.

Panel A: Entire sample July 1993 - December 2009

	5 short	5 long	Diff	1 short	1 long	Diff
Mean return	0.1665	0.1057	0.0608	0.1903	0.0912	0.0991
P-value			*0.000*			*0.000*

Panel B: Entire sample: No short sales allowed

	5 short	5 long	Diff	1 short	1 long	Diff
Mean return	0.1179	0.1016	0.0163	0.1441	0.1010	0.0431
P-value			*0.000*			*0.000*

Figure 21: Estimation lengths versus parameters estimation models

Note: Estimation window refers to the number of days used to estimate the parameters, and Model is the combination of estimation methods used. Annual mean returns are computed over the July 1993 – December 2009 sample. Transaction costs are ignored. Note that in Panel B, the scales are not similar in order to keep the visual representation meaningful:
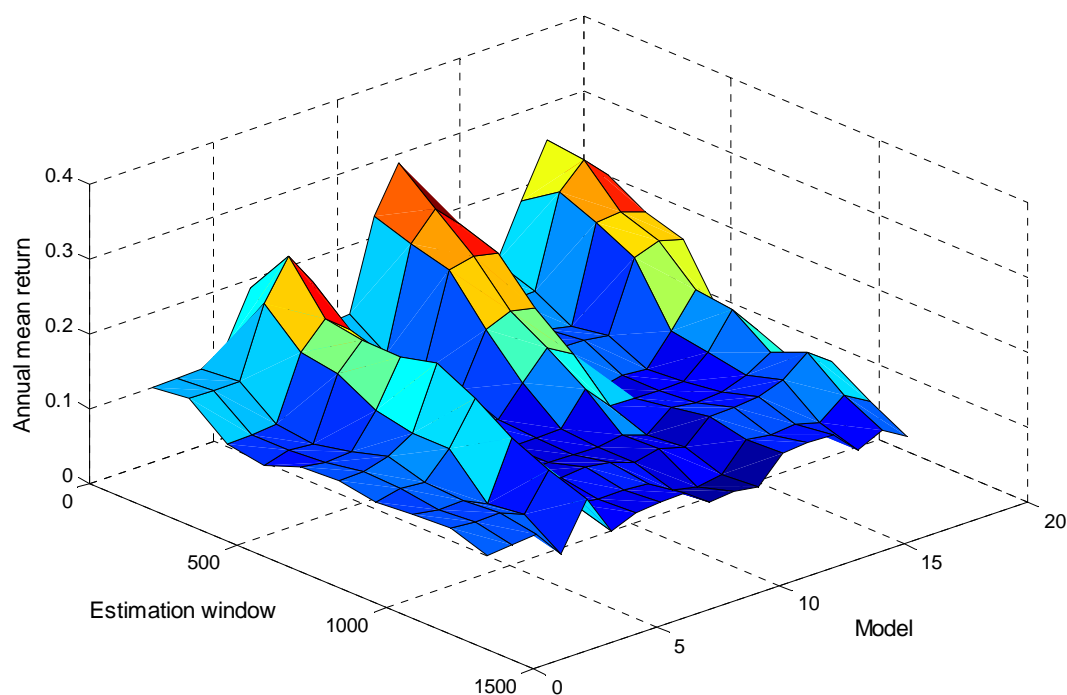
Panel A: All specifications

Panel B: With or without short-sales
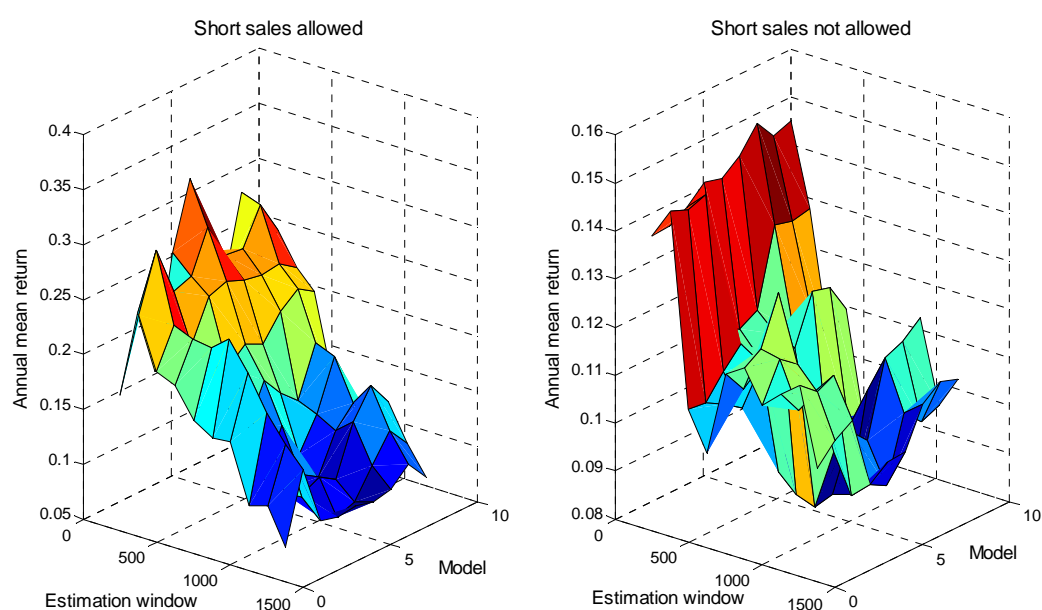
Figure 21 and Table 21 show that adding emerging markets indices does not modify the finding presented above: The estimation window length has a stronger impact on the optimized portfolio return than using different parameters estimation models. In addition, restraining short-sales influence negatively its performance.

Table 22: Portfolios performance analysis

1/N is the benchmark strategy represented by an equally weighted portfolio including only equity indices. The four other columns are the complex portfolios described in Section 3.2. All results are given in an annual frequency (except *Nb* trade and *BE TC*, which represent respectively the number of transactions and break even transaction costs). *P*-values associated with simple mean returns are obtained with bilateral Student *t*-tests between the respective portfolio and the benchmark. The Jensen's alphas tests focus on whether coefficients are statistically different from zero.

Entire sample June 1994 - December 2009

	Investment strategies				
	1/N	PF_ALL	PF_4	PF_ALL_NS	PF_4_NS
Simple mean return	0.1019	0.2717	0.2665	0.1095	0.1237
P-value		*0.0376*	*0.0574*	*0.8912*	*0.6912*
Compouned mean return	0.0930	0.2598	0.2449	0.1027	0.1191
Volatility	0.1608	0.2844	0.3069	0.1531	0.1492
Jensen's alpha		0.2147	0.2034	0.0294	0.0442
P-value		*0.0024*	*0.0074*	*0.3341*	*0.1357*
Sharpe ratio	0.3183	0.7770	0.7032	0.3839	0.4890
Nb trade		452.8	459.0	117.9	118.4
BE TC		0.0060	0.0057	0.0013	0.0033
Beta		0.1230	0.2431	0.5733	0.5626

Part III: A new simulation based market timing test

1. Introduction

Parallel to the literature presented in the two first parts of this thesis about investment strategies, a myriad of testing procedures has also been proposed, ranging from basic performance ratios to complex simulation methods. However, these two aspects, the investment strategies and the testing methods, are usually investigated separately. A major characteristic of a test is its power, or in other words, the amount of abnormal return required to reject the null hypothesis of equal performance when they are not. Ignoring the power of a statistical test used to assess the profitability of investment strategies may results in a misleading conclusion. Consider, for example, an investment strategy that posses a "real" forecasting ability and generates large abnormal returns. However, a low-power procedure is used to perform the statistical test, and thus, fails to reject the null hypothesis of equal performance. If one does not examine the test properties, the logical conclusion is to deny that the investment strategy is successful.

In this last part, we present a new testing method based on the simulation of trading positions. This approach is not new for strategies that invest in one asset only, usually an equity index, such as technical analysis trading rules. Cowles (1933) proposes to shuffle and draw cards manually to generate artificial strategies, which are then compared with financial advisors records. Brown, Goetzmann and Kumar (1998) improve the methodology with computer simulations. The main difference with our proposed test method resides in the assumption about the artificial trading signals. Brown, Goetzmann and Kumar (1998) draw randomly, with replacement, trading positions with the only restriction that long, short and neutral positions have the same frequency

as the original series. Thus, this method does not keep the structure[66] of the original position series. Hence, we propose two different techniques to construct artificial position series with a similar structure to the original one. The question addressed by the proposed test is the following: Is the performance of a strategy due to its specific positions in the market, or is it possible to replicate it with random trading signals that have a similar structure to the original one? The test can be summarized as a two stages procedure. The first one consists in generating several artificial investment strategies, or trading positions. Then, during the second stage, we compare the original strategy return with the return of these artificial strategies. Hence, the test *p*-value indicates the percentage of simulated series that have a return, or any other performance measure, higher or equal to the original value. Using a method that generates artificial strategies that share a similar structure with the original one is of particular interest if the investment strategy is designed to capture long-term trends in the market. Indeed, such a strategy generates relatively few variations in trading signals. Suppose, for example, that a strategy consists in detecting and investing according to bull and bear markets, without trying to time the market on a daily basis. If the strategy is successful, the trading positions are composed of long positions over a few years, as long as the bull market lasts, and then reverse to short positions when the market enters a bear phase. If artificial trading signals are constructed with a standard bootstrap, i.e. drawing randomly with or without replacement, the original structure is ignored. Thus, the artificial trading signals are not comparable, as they switch from a long to a short position much more often[67]. In addition, if the performance measure includes transaction costs, the artificial series are clearly disadvantaged.

In this final part, we re-examine the investment strategies presented so far with the proposed simulation test. We propose to construct artificial trading signals series as first-order Markov chains in the single-asset setting. The advantage is to limit the number of assumption to only one; the signals series follow a first-order process, i.e., the next-period state depends only on the present state. In the multiple-assets case, we simulate portfolio weight with a block bootstrap.

In the first part, we show that technical analysis strategies generate economically significant returns in excess of the buy-and-hold over a 15 years sample. The annual mean returns lie

---

[66] By structure, we denotes whether a position is likely to last over many periods, or if signals reversion is frequent.

[67] On the other hand, if the investment strategy focuses on short-term trends, both methods should provide similar results.

between 10.7% and 14.6% compared with 6.1% for the benchmark. However, standard Student t-tests fail to reject the null hypothesis of equal means; the highest t-statistic is 1.2, which corresponds to a p-value of 0.23. This is in sharp contrast with the proposed test methodology that generates p-values consistently lower than 5%. This means that random trading signals, with a similar structure[68], can not yield returns as high as the original series. This indicates that the trading rules possess significant market timing abilities. We also apply the proposed test to risk-adjusted measures, the Jensen's alpha and the Sharpe ratio, and we reach similar conclusions.

Regarding strategies that invest in more than one asset, such as the complex portfolios presented in the second part, we propose to construct artificial trading positions, or more precisely portfolios weights, with a block bootstrap. We reach similar conclusions as all simulation-based p-values are lower than their Student counterparts, but the difference with the standard 5% level is less pronounced than for technical analysis strategies.

Finally, we conduct a comprehensive Monte-Carlo experiment to assess the power of standard performance measures compared with our proposed simulation-based test. For technical analysis strategies, they are the simulation test introduced by Brock, Lakonishok and LeBaron (1992) the Student t-test of equal means, the Jensen's alpha and the statistical difference in Sharpe ratios. We compare the power of these tests and this does not mean that they are interchangeable as the test questions are not the same. First, we show that the proposed simulation test has a similar power to the BLL test, even though the latter does not test the performance directly. Second, these two simulation tests are much more powerful than the three other measures. For example, the proposed test generates a median p-value of 1.4% and a null hypothesis rejection frequency of 87% for yearly mean excess returns of 6% over a 16 years sample, while none of the Student t-test rejects the null. The difference in Sharpe ratios presents a power as low as Student t-tests, but Jensen's alphas are slightly more powerful with a rejection frequency of 37%. Finally, we show that the proposed test is well specified, as it does not reject the null when it is true. We find similar results in the multiple-assets setting, in which we consider as well the two popular Henriksson and Merton (1981) and Treynor and Mazuy (1966) market timing tests, HM and TM thereafter. The only exception is that Jensen's alphas have a similar power to the proposed procedure. We also illustrate the issues raised by the interpretation of the HM and TM tests. First of all, the correlation coefficient between the selection and market timing parameters is almost

---

[68] The strategies rely on long-term trends and change trading position only a few times over the 15 years sample.

always equal to minus one. In addition, the distinction between selection and market timing abilities leads to interpretation problems, especially for the HM model. For around 5% of the simulations, we obtain both a significant negative selection parameter and a positive market timing coefficient. Finally, their power is relatively low for high levels of abnormal returns. We also find that the proposed simulation test is well specified in case of no abnormal returns.

The next Section of this third part is dedicated to a literature review, which presents the most widely used performance measures and their drawbacks. The third one consists in a brief illustration about the lack of power of Student *t*-tests, while in Section four we describe the proposed test methodology theoretically. In the next two Sections we re-examine the results obtained in the first two parts of this thesis with the proposed simulation test. Then, we present the Monte-Carlo experiment conducted to compare the power of various testing procedures. Finally, Section eight concludes this last part.

2.   Literature review

We arbitrarily choose to summarize performance measures[69] in five categories: Tests based on the equality of mean returns between a strategy and a benchmark, risk-adjusted performance ratios, excess returns from asset pricing models, market timing tests focusing on returns and trading positions and simulation techniques.

2.1   Student *t*-tests

It determines whether the strategy returns distribution differs from those of a benchmark[70]. We present them in detail in the appendix I-A. They are widely used to test the profitability of technical analysis since the BLL study. Among others, Bessembinder and Chan (1995), Allen and Karjalainen (1999), Skouras (2001), Day and Wang (2002), Fang and Xu (2003) or Fong and Yong (2005) consider this test. It is straightforward to implement, as it requires only the returns means, variances and covariances estimates. Nevertheless, BLL argue that this procedure depends

---

[69] A comprehensive review is behind the scope of this paper,. We present performance measure widely used by practitioners or academics and the methods related to our test. Thus, this review ignores measures based on investor utility functions and those developed in the Bayesian framework.

[70] A reference return of zero might also be considered, especially for zero investment strategies.

on unrealistic assumptions such as normal, stationary and time independent returns distributions. Furthermore, another aspect often overlooked is that they are not particularly powerful. An illustration is provided at the beginning of the next Section.

2.2   Performance ratios

They usually contain the strategy mean return in excess of the risk-free rate in the numerator and a risk measure in the denominator. Thus, they can be interpreted as the excess return per unit of risk. Among them, the Sharpe ratio[71] is probably the most widely used both by the financial industry and academics. Jobson and Korkie (1981), Memmel (2003) and Ledoit and Wolf (2008) propose various procedures to assess whether two Sharpe ratios are statistically different. A similar measure is the Treynor ratio, which uses the CAPM beta as the risk measure. Among others, Jorion (1985), Chan, Karceski and Lakonishok (1999) or Jagannathan and Ma (2003) conduct performance analyses with them. However, they suffer from some well documented drawbacks. First of all, they suppose that the respective risk measure describes it adequately, and consequently, that returns are normally distributed. Obviously, the Treynor measure depends on the CAPM empirical validity. Dowd (2000) argues that the Sharpe ratio can be used only if the evaluated fund or investment strategy is not correlated with other assets owned by an investor. To address some of these issues, several enhancements have been proposed. Hodges (1998) introduces a Generalized Sharpe Ratio to take into consideration higher moments of the returns distribution. Zakamouline and Koekebakker (2009) derive an Adjusted for Skewness Sharpe Ratio in an investor utility framework. They note that their formula is only an approximation, which limits its practical use. Darolles and Gourieroux (2010) argue that conditionally fitted Sharpe ratios can be used, at least to some extent, with returns that depart from normality. Researchers have proposed alternative measures that consider different risk dimensions. Kaplan and Knowles (2004) show that their Kappa measure regroups two others: the Sortino ratio by Sortino and Van Der Meer (1991) and the Omega, introduced by Keating and Shadwick (2002). They focus on downside risk and represent the risk of not achieving a determined threshold. The Calmar ratio considers the maximum drawdown as the risk measure. Leland (1999) proposes to modify the CAPM beta to take into account the distribution skewness. Dowd (1999) uses the

---

[71] It is defined as the relation between the mean return in excess of the risk free rate divided by the volatility.

Value-at-Risk, and Agarwal and Naik (2004) extend this framework with a conditional version. This list is obviously not exhaustive.

An interesting question is to determine whether the performance assessment differs with these various measures. Eling and Schuhmacher (2007) provide some insights based on a hedge funds performance analysis. They compare 12 alternative measures with the standard Sharpe ratio. They find that, even if returns are not normal, the rankings obtained with the various measures are very close. This is in close agreement with Eling (2008) who compares performance measures applied to a large funds universe, which includes various asset classes (stocks, bonds, real estate, funds of hedges funds or commodity trading advisers). Finally, a note of caution is given by Zakamouline and Koekebakker (2009) who argue that most of these measures lack a sound theoretical basis.

2.3   Tests based on asset pricing models

This approach requires calculating normal returns according to various risk factors considered by an asset pricing model. Thus, the abnormal return consists in the difference between the strategy return and its normal return. The performance assessment of a fund, or an investment strategy, proceeds as follows: First, a regression is run between the strategy returns[72] and the dependent variables defined by the selected model, including a constant. This latter term is the performance measure, which is commonly denominated *alpha*. Finally, Student *t*-tests are performed to determine whether the alpha is statistically different from zero. The following discussion concerns two aspects: The asset pricing model choice and the testing procedures.

Asset pricing model based performance analysis was first introduced by Jensen (1968), who uses the Sharpe (1964) and Lintner (1965) CAPM to estimate normal returns. This measure is used for all investment strategies, but especially for mutual funds. Among many others, we may cite Jensen (1968), Lehmann and Modest (1987), Malkiel (1995) or Elton, Gruber and Blake (1996). Nevertheless, it depends on the CAPM empirical validity to explain returns adequately. Researchers have documented some patterns in asset returns that are not in line with the CAPM. Therefore, other asset pricing models have been proposed, such as the Fama and French (1993) and Carhart (1997) models. They include risks factors related to individual stocks, the size (SMB) and the book-to-market (HML) effects. The latter model includes, as well, a factor capturing the

---

[72] Or with returns in excess of the risk-free rate.

momentum in stocks returns. These models are commonly used in the most recent studies: Wermers (2000) re-examines US mutual funds performance with holding data. Otten and Bams (2002) investigate the performance of mutual funds in the five most important European markets, while Cesari and Panetta (2002) focus on the Italian equity funds market. Bauer, Koedijk and Otten (2005) test the performance of international ethical funds and Comer, Larrymore and Rodriguez (2009) add fixed-income related factors to the Carhart (1997) measure. Another approach to capture determinants of assets returns is to focus on higher co-moments between the strategy and the market returns. Harvey and Siddique (2000), Barone Adesi, Gagliardini and Urga (2004) or Smith (2007) add co-skewness and co-kurtosis related factors in asset pricing models. Moreno and Rodríguez (2009) find that adding a co-skewness factor in the CAPM modifies the mutual fund performance analysis. Indeed, many funds with a negative (positive) Jensen's alpha have a positive (negative) alpha when the asset pricing model includes the co-skewness. Ranaldo and Favre (2005) and Ding and Shawky (2007) use them in hedge funds performance evaluations, as returns usually depart from normality.

The second issue concerns the testing procedures, which depend on the research question. First, one can be interested in testing a single fund/strategy performance. In this case, Kosowski, Timmermann, Wermers and White (2006) point out that bootstrap techniques generate more accurate test for an individual alpha. Second, a study could also focus on the average performance of fund managers to determine whether active management adds value. Barras, Scaillet and Wermers (2010) propose a simple method to isolate true positive alphas from those resulting from luck, or estimation errors. They also document the impact of using several asset pricing models to determine normal returns. They report evidence of skill among mutual fund managers with the CAPM, but not with more sophisticated models.

2.4   Market timing measures

Market timing consists in increasing (decreasing) the portfolio exposure to the market when it follows an upward (downward) trend. The tests can be conducted either on returns or trading positions series.

Treynor and Mazuy (1966) and Henriksson and Merton (1981), referred to as TM and HM, are the most commonly used models to assess the market timing ability with returns series. They are similar to the CAPM with an additional factor to capture market timing. Thus, they establish a dichotomy between stocks selection and market timing abilities. They are widely used in the mutual funds industry, for example, in Henriksson (1984), Chang and Lewellen (1984), Cumby

and Glen (1990) or Coggin, Fabozzi and Rahman (1993). Most of these studies find a negative correlation between the selection (positive) and market timing (negative) coefficients. This means that the fund exhibits negative market timing, as they have higher market exposure when returns are low. Henriksson (1984) argues that this may be explained by the parametric testing method. Indeed, the procedure could be biased by errors in variables, misspecification of the market portfolio or using a single-factor model. Ferson and Schadt (1996) document that models conditioned on publicly known information may control these biases. In addition, Jiang, Yao and Yu (2007) argue that estimating the fund beta as the weighted average of betas of individuals stocks in the portfolio instead of those of the portfolio itself helps to mitigate some issues. First, tests are more powerful, and thus, they find evidence of mutual funds timing abilities. Second, they are less subject to "artificial timing bias". For instance, Jagannathan and Korajczyk (1986) show that buying option-like-securities, such as highly leveraged firms stocks, generates a negative correlation between the two coefficients as well. Bollen and Busse (2001) propose to add the three other factors from Carhart (1997) to the HM and TM models. Jiang (2003) introduces a new nonparametric market timing test. It has the advantage to separate timing abilities from the response given in term of market exposure. He applies this test to a large sample of US mutual funds, but he does not find more evidence of market timing abilities than with the TM and HM models. Cuthbertson, Nitzsche and O'Sullivan (2010) find similar results on the UK market. Finally, Admati, Bhattacharya, Pfleiderer and Ross (1986) question the possibility, even theoretically, to separate objectively the selection and timing abilities.

Grinblatt and Titman (1989a) are among the first to propose a market timing test based on portfolio holdings. In Grinblatt and Titman (1989b), they develop the widely used *positive period weighting measure*. Wermers (2006) provides a comprehensive survey of studies that use portfolio holdings to test the performance.

Testing the performance of technical analysis strategies with trading positions has a substantial advantage about the quantity of information. Indeed, mutual funds holdings are available usually in a quarterly frequency, while all daily trading positions are known when testing a technical analysis strategy. In addition, most of them are, by construction, pure market timing strategies, whereas the performance of mutual funds or mean-variance optimal portfolios depends also on selection abilities. Cumby and Modest (1987) propose regressing market returns in excess of the risk-free rate on lagged trading signals. Thus, a positive and significant coefficient indicates positive market timing, i.e. a long (short) position is followed by a positive (negative) excess market return during the next trading interval. Pesaran and Timmermann (1992) develop a nonparametric procedure to test the proportion of correctly predicted trading signals. The

Diebold and Mariano (1995) test focuses on the difference in predictive accuracy between two competing forecasting procedures. In Section 6 of the first part, we propose to examine whether trading signals are consistent with long-term market phases, or in other words, bull and bear markets. This test may be more appropriate if the investment strategy is stable and does not try to time the market in the short-term[73]. The *p*-values are obtained with a bootstrap procedure. A drawback of these measures is that they do not consider the economic performance of the trading strategies. Indeed, some studies find significant evidence of timing abilities, while the strategies do not yield significant abnormal risk-adjusted returns. Among others, Allen and Karjalainen (1999), Fernandez-Rodriguez, Gonzalez-Martel and Sosvilla-Rivero (2000) or Neely (2003) reach such a conclusion.

2.5   Simulation methods

They are typically used for technical analysis, as they require an investment rule that can be completely summarized in a mathematical formula. BLL introduce one of the most commonly used simulation-based tests. It consists in applying trading rules to a set of artificial prices series obtained by bootstrapping the original returns according to a return generating process[74]. Empirical *p*-values are then obtained by comparing the original strategy mean return, or another performance measure, with those obtained with the artificial prices series. Many studies employ this test, or some variations, such as Bessembinder and Chan (1998), Isakov and Hollistein (1999), Fong and Ho (2001), Day and Wang (2002) or Gençay, Dacorogna, Olsen and Pictet (2003). The latter extend the procedure in an intraday data setting. Nevertheless, it is important to note that this bootstrap method does not test directly for abnormal returns[75]. First of all, Student *t*-tests are conducted to detect abnormal returns, and then, the bootstrap method helps to relax some assumptions of these tests. The bootstrap test does not shed light on the trading rule profitability, but whether the rule relies on returns stylised facts in order to yield higher returns.

---

[73] Indeed, negative daily market returns happens rather often in a bull market. Having a long position during these days would be considered as negative market timing by the before mentioned tests, not in the one proposed in the first part of this thesis.

[74] Autoregressive processes, the random walk model and several GARCH specifications are usually considered.

[75] Suppose for instance that a strategy yields a return higher than the buy-and-hold. However, no other test of abnormal returns is conducted. Bootstrap *p*-values higher than 5% would mean that the trading rule applied to artificial series is also able to yield high return. Does is means that the trading rule is not profitable?

Furthermore, this method could be time consuming if the trading rule is complex, such as genetic algorithms. Indeed, the rule has to be computed on a multitude of artificial prices series.

Levich and Thomas (1993) implement a similar bootstrap test, but without relying on a return-generating model. Brown, Goetzmann and Kumar (1998) propose to bootstrap trading positions instead of returns series. They use a standard bootstrap approach, which ignores any structure in positions series, such as long-term trends.

White (2000) proposes a testing procedure that takes into account the data-snooping bias, the reality check (RC) test. This bias arises when several forecasting models are considered simultaneously. Indeed, it is rather likely that a few specifications of a forecasting method are successful by luck. The superior predictive ability (SPA) test, introduced by Hansen (2005), slightly modifies the RC test to make it more powerful. Sullivan, Timmermann and White (1999) and Hsu and Kuan (2005) show that previous results in favour of technical analysis profitability are, at least to some extent, due to this bias. Qi and Wu (2006) reach a similar conclusion in the foreign exchange market. Romano and Wolf (2005) and Hsu, Hsu and Kuan (2010) further extend the RC and SPA tests with a stepwise multiple testing approach. These new tests are able to identify all models that outperform the buy-and-hold, while the RC and SPA tests determine only if at least one specification is successful. Neuhierl and Schlusche (2011) apply these four tests to market timing rules[76], and they find that these strategies are not successful anymore when the bias is controlled in the tests.

Finally, a note of caution is given by Goetzmann, Ingersoll, Spiegel and Welch (2007) who show that some performance measures presented in this review[77] can be easily manipulated to produce a significant performance without any forecasting abilities.

---

[76] They consider a wide set of financial or sentimental instruments that are supposed to predict future stock market returns, such as ratios (the price-earning, book-to-market or the dividend yield), credit spread, implied volatility or the interest rate differential.

[77] Specifically the Sharpe Ratio, the Generalized Sharpe Ratio, Jensen's alpha, the Sortino Ratio, the Henriksson and Merton (1981) and the Treynor and Mazuy (1966) market timing models.

3.   Student *t*-tests: a low-power testing procedure

Before turning to our proposed test methodology, we illustrate the Student *t*-tests low power. Figure 22 reports the level of abnormal return required for a bilateral Student *t*-test to reject the null hypothesis of equal means at the standard 5% level. We compute them, for various volatility levels and time series lengths, by inverting the test. Thus, instead of calculating the test statistic, we give it the standard value of 1.96 and we use the standard Student *t*-test equation to determine the required difference in mean returns according to the selected number of observations and the variance, such as

$$\frac{X}{\sqrt{2 \cdot \dfrac{Var}{Nb\_obs}}} = 1.96 \tag{3.1}$$

where *X* is the minimal difference in mean returns required to reject the null hypothesis, *Var* is the series variance[78] and *Nb_obs* is the number of observations used to compute the means. The results are annualized in order to show the magnitude, but calculations are carried out with daily data (i.e. a year comprises 252 trading days). The results, i.e. the required abnormal returns, in the following figures are calculated for various chosen levels of volatility and sample lengths. Thus, equation 3.1 is solved many times with *X* being the unknown variable and with several chosen values for *Var* and *Nb_obs*. Panel A presents all results for annual volatilities ranging from 10% to 30% and time series lengths from two to 30 years. The minimum value of the required difference in mean returns is 5.06%, which is not far from the long-term market average return. This means that an investment strategy has to yield, roughly, a mean return that is twice as high as the buy-and-hold to reject the null. Clearly, the test power drops as the volatility increases, or when the sample length decreases.

Panel B displays the results for three common levels of volatility. For the 20% volatility level, a mean abnormal return of 10% is the minimum to obtain a significant difference, and furthermore, this level increases sharply for shorter samples. This illustration indicates that Student *t*-test based performance analysis must be interpreted cautiously, as its power is low. Obviously, these findings are in line with those presented in the Section 7 of the first part, which

---

[78] We assume an equal variance and a zero correlation. A positive (negative) correlation would decrease (increase) the required difference in mean returns.

consists in the simulation experiment between the percentage of right signal and the excess returns.

Figure 22: Student *t*-test power

This figure presents the levels of abnormal returns required to reject the null hypothesis of equality in means between two returns series. The test is the standard bilateral Student *t*-test at a 5% confidence level. Volatility is the return series volatility, which is supposed to be the same for the two series. The resulting required abnormal returns and the volatility are presented in an annual frequency, but calculations are performed in a daily frequency, thus 10 years correspond to 2520 days for instance.
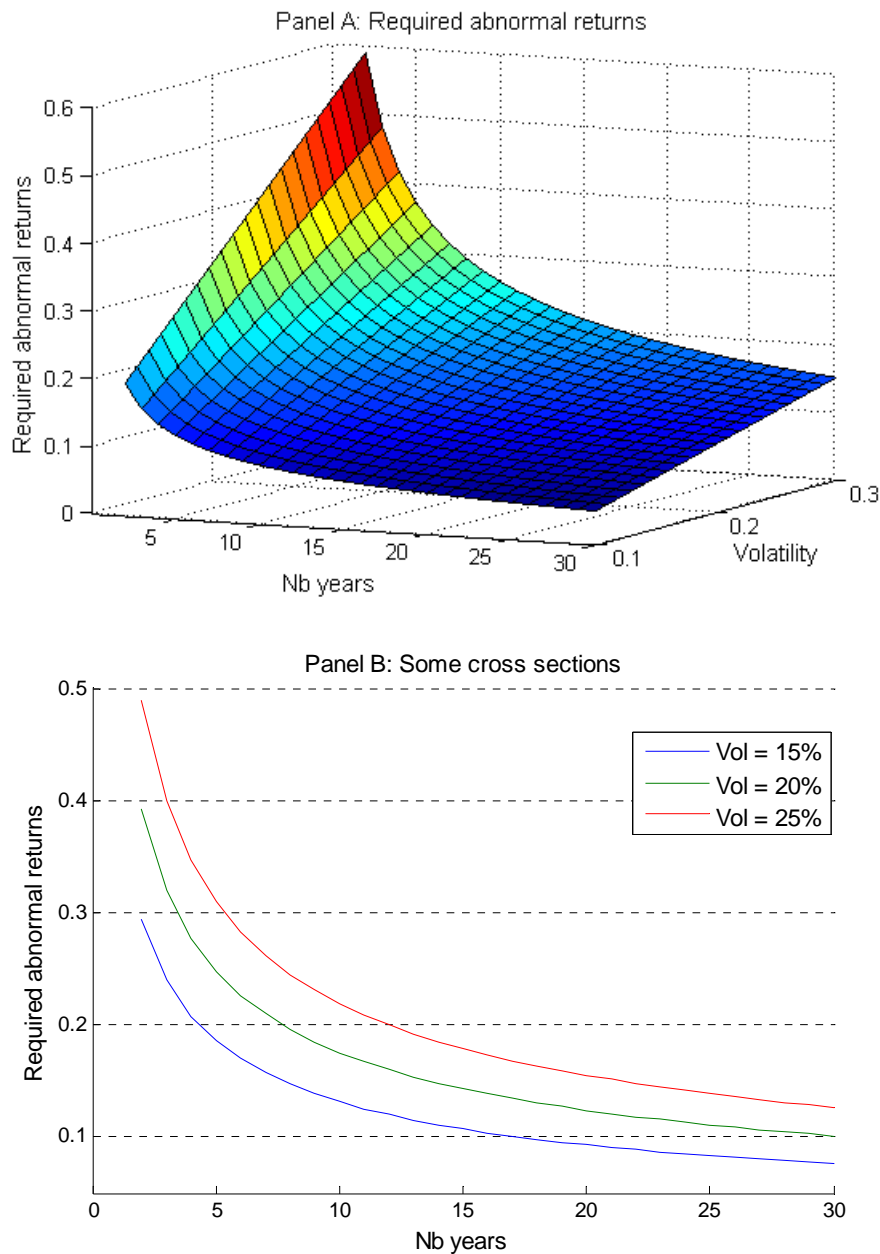
4.   The simulation test

4.1   Introduction and intuition

The proposed test compares the performance of an original trading strategy with simulated strategies constructed with random trading signals (or portfolio weights). Other performance measures than the average return can be tested, such as ratios or measures not directly related to the returns distribution, such as Shape ratios, Jensen's alphas. The question addressed by this test is the following: Does the strategy performance depend on its specific trading positions, or is a random signals series able to reproduce it? Thus, this is a market timing measure, as it determines whether the trading positions have been taken at the right time. This test can be used, both, to test the economic performance of a strategy or its forecasting ability with, for example, the percentage of correct signals. It is an extension of the Brown, Goetzmann and Kumar (1998) methodology, which does not consider the structure of the trading positions.

The first step is to construct artificial signals series that have a comparable structure to the original. Then, returns are obtained by applying these signals to the original market (i.e. stock index) returns series. The next step consists in computing the performance measures for these artificial returns series. It is worth emphasizing that this procedure does not generate new prices series, but it resamples only the original trading positions, which are then applied to the original price series. Finally, p-values are calculated by comparing the original performance measure(s) with those obtained in the simulation. They represent the percentage of simulated series with an equal or higher value than the original[79]. As mentioned earlier, we examine strategies that invest either in only one or in several assets. The test procedure is similar for both cases; nonetheless, the way artificial trading signals are generated differs.

4.2   Single-asset case

This case is mainly suitable for technical analysis strategies, as they usually consist of long, neutral[80] or short positions in an asset, such a market index or a currency. Hence, the trading

---

[79] For example, a p-value of 0.03 means that only 3% of the simulated signals series are able to produce a performance as high as the original. This would provide strong evidence of market timing abilities

[80] The neutral signal usually implies to invest the capital in the risk free rate.

signals can be expressed as a vector containing ones, zeroes and minus ones[81]. In order to keep the original trading positions structure, we propose to model them as a first-order Markov chain. This is a random process characterised by the Markov property, i.e. the next state depends only on the current state. In our setting, a state represents a trading position. The first step is to compute the transition probability matrix. It describes the probability of moving from a state to another over the next period, and it is defined as

$$M = \begin{bmatrix} p_{1,1} & p_{1,2} & p_{1,3} \\ p_{2,1} & p_{2,2} & p_{2,3} \\ p_{3,1} & p_{3,2} & p_{3,3} \end{bmatrix} \tag{3.2}$$

where $p_{i,j}$ is the probability of moving from the state i to the state j. Thus, $p_{1,3}$ is the probability of moving from a long to a short position during the following day. As we know the original states series (i.e. the original trading signals series), they are calculated as

$$p_{ij} = P(S_{t+1} = j | S_t = i) \tag{3.3}$$

where S_t is the trading signal at time t. The transition matrix has to be squared, and the elements of each line have to add up to one. These probabilities define whether the trading signals series is stable. For instance, a strategy designed to follow long-term market phases should have $p_{1,1}$ and $p_{3,3}$ close to one. This means that if the current state is a long (short) position, the next one will be, as well, a long (short) position with a high probability. To construct the artificial series, we use the following method: First of all, the initial state is randomly drawn from the original series. Then, the next states are obtained according to

$$S_{t+1} = rando(P(S_t)) \tag{3.4}$$

where $P(S_t)$ is the transition probability matrix line corresponding to the trading position at time t. *Rando* is a function that generates random variables according to the probability expressed in $P(S_t)$ [82]. We simulate artificial states until the series has the same number of observations as the original strategy. Finally, we apply a filter to select only a few of these simulated positions in order to have similar series; we keep a simulated signals series only if it has the same proportion of long and

---

[81] The proposed method is also adequate for long only strategies and for those without the neutral position. We decide to present the process with three trading positions.

[82] If the strategy consists in the three trading positions, i.e. including neutral positions, the *Rando* function generates either a 1, a 2 or a 3. Thus, the 2 will be replaced with 0 and the 3 with -1.

short positions as the original one, with an error margin of 5%[83]. This ensures that the artificial series have trading positions with a similar structure than the original in real and not only in the process used to generate them. We repeat this process until N suitable artificial signals series are obtained. Then, they are used in conjunction with the original index returns to generate artificial strategies returns series as

$$R^{*}_{i,t+1} = S^{*}_{i,t} \cdot R_{t+1} \quad \textit{for } i = 1,...,N \tag{3.5}$$

where R is the original index return, R^{*}_{i} is the artificial return series obtains with the simulated signal series, S^{*}_{i}. N is the number of artificial series, and we set it to 5'000. The next step is to compute, for each of these return series, the performance measures. Finally, we calculate the simulation p-values by comparing the original performance measures with those resulting from the 5'000 simulated return series, St^{*}_{i}, for $i=1,..., 5'000$. First, we order these statistics such as

$$St^{*}_{1} \leq St^{*}_{2} \leq ... \leq St^{*}_{N} \tag{3.6}$$

then, we find M such as

$$St^{*}_{M} \leq St_{org} < St^{*}_{M+1} \tag{3.7}$$

where St_{org} is the original strategy performance measure. Thus, the simulation p-value is $1-M/N$. It represents the percentage of simulated statistics that have a higher or equal value to the original. In this setting, the simulated performance measures represent the distribution under the null hypothesis that the strategy has no timing ability. Figure 23 provides an illustration of artificial signals series compared with those of the Opt_4 strategy described in the first part.

---

[83] Suppose that the original strategy is composed of 2170 long and 1850 short positions. An accepted artificial series should have between 2061 and 2278 buy signals and short signals ranging from 1758 to 1942.

Figure 23: Original and simulated trading signals

This figure displays the original trading position (the first line from above) and 10 artificial series simulated as Markov chains. Long (short) positions are in blue (red) and neutral positions in black.
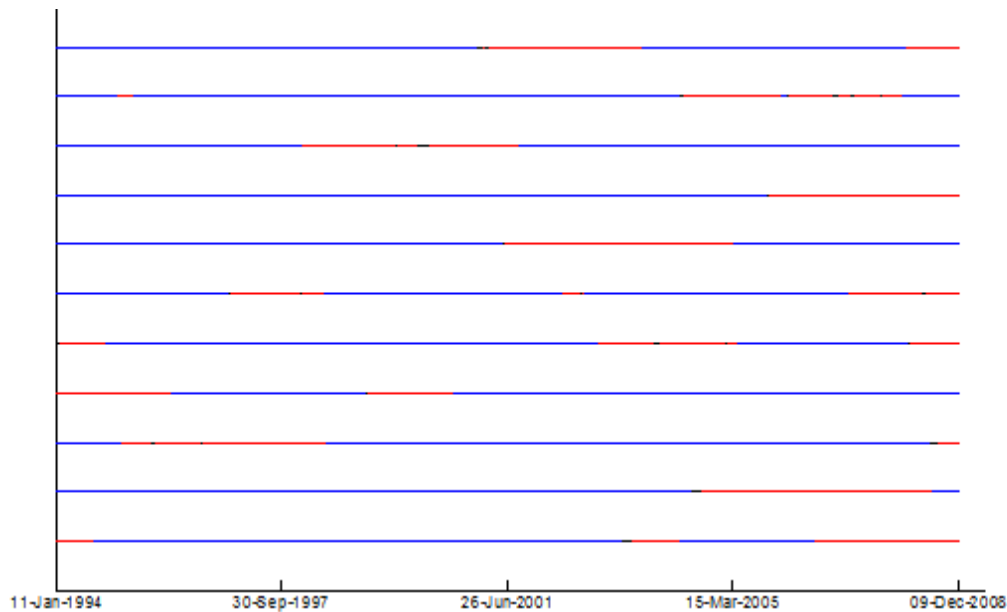
4.3   Multiple-assets case

When the strategy allocates the capital in more than one asset at the same time, it is not possible to use Markov chains to construct random positions series. First, if artificial portfolio weights were generated as multivariate Markov chains, there is no guarantee that the weights of each asset add up to one. Thus, they would not be comparable with the original weights. Secondly, portfolio weights do not fit the inherent nature of Markov chains, which are characterized by a set of a few, well-defined states[84].

The only difference compared with the single-asset case is the manner artificial signals, or portfolio weights in this case, are generated. Instead of Markov chains, we implement a block bootstrap approach. This method ensures that weights sum up to one and dependencies among assets returns are conserved, at least to some extent. However, this method requires choosing the block length arbitrarily. The procedure is the following: First, a block length, *Bl*, is chosen and the

---

[84] Indeed, it is rather likely that portfolio weights change in every period and none of them are equal over time.

portfolio weights matrix, W_{TxN} where T is the number of periods and N the number of assets, is divided in Nb blocks[85] such as:

$$Nb = T / Bl \qquad (3.8)$$

The choice of the block length implies a trade-off. As the length increases, the time dependency of the original data is conserved to a larger extent than blocks of smaller length. On the other hand, the number of blocks available for the simulation process decreases. This choice is also limited insofar as Nb has to be an integer.

The next step consists in generating N_{sim} artificial weights matrices, W^*_i where i=1, …, N_{sim}, by random permutations, without replacement, of the original weights blocks. Finally, we calculate artificial returns series as

$$R^*_{i,t+1} = W^*_{i,t} \cdot R_{t+1} \quad for \ i = 1,...,N_{sim} \qquad (3.9)$$

where R_{t+1} is the original assets return matrix at time $t+1$. Finally, the simulation p-values for the performance measures are computed in a similar manner to the single-asset case described in equations (3.6) and (3.7).

Compared with the various performance measures presented in the literature review, the proposed test has the following merits: First of all, the simulation process is based on a simple concept, which makes economic sense as a market timing test. In addition, the results interpretation is straightforward, as there is no dichotomy between market timing and selection abilities. Another advantage is the limited number of assumptions required, compared with measured based on asset pricing models or the returns distributions. The former depends on the empirical validity of the underlying model, which should be tested before conducting a performance analysis. Finally, our procedure requires less information than other simulation-based tests[86]. Indeed, it needs only trading positions series and not a mathematically well defined investment process. Thus, the proposed test is also appropriate for "subjective" investment strategies, as long as one know the trading positions.

---

[85] Thus, one block consists of all assets weights over a number of periods defined by the block length, Bl.

[86] Obviously, this is an advantage only compared with other simulation based tests, as the vast majority of other performance measures require only the return series. Thus, the proposed test might clearly not be used to examine mutual fund performance in an outsider setting, but it can be employed by the fund manager itself.

5.   An application to technical analysis

5.1   Percentage of right signals

First, we examine the percentage of right signals with the p-values computed with the proposed test. These measures are not directly related to the performance, but they are useful to shed light on the strategies forecasting ability. They are defined as

$$Right\_buy_j = \frac{\sum_{t=1}^{Nb_j} 1\{SB_{j,t} | R_{t+1} > 0\}}{Nb_j} \tag{3.10}$$

$$Right\_sell_j = \frac{\sum_{t=1}^{Ns_j} 1\{SS_{j,t} | R_{t+1} < 0\}}{Ns_j} \tag{3.11}$$

$$Right\_strat_j = \frac{\sum_{t=1}^{Nb_j} 1\{SB_{j,t} | R_{t+1} > 0\} + \sum_{t=1}^{Ns_j} 1\{SS_{j,t} | R_{t+1} < 0\}}{N} \tag{3.12}$$

where Nb_j, Ns_j are the jth strategy number of buy and sell signals, $SB_{j,t}$ and $SS_{j,t}$ represent the buy or sell signals at time t, depending on the position taken by the strategy at this time. R_t is the buy-and-hold return at time t and N the total number of observations. 1 is a function that generates a one if the condition in brackets is filled.

The results in Table 23

Table 23 indicate that the three strategies have slightly higher percentages than the buy-and-hold for long positions[87]. Even if the values are lower for the short positions, we find stronger evidence of timing on the short side. Indeed, shorting continuously the market would only result in 46.88% of winning trades, while the percentage associated with the strategies lies between

---

[87] In an unreported simulation study, we find that only a small increase in the percentage of right signals, compared with the buy-and-hold, is sufficient to generate large excess returns. On average, a 1% increase corresponds to a 4% annual excess return. Further details are available upon request.

50.53% and 51.74%. We find evidence of a significant forecasting ability for the first two strategies, as the simulation p-values are lower than the standard 5% level. This means that these percentages cannot be achieved by luck. Results are weaker for the Voting strategy with p-values ranging from 7.1% to 7.2%.

Table 23: Percentage and right signals

This table presents the percentages of right signals, respectively for long and short positions and for the whole strategy. *BH* is the buy-and-hold, *Opt_all, Opt_4* and *Voting* are the three complex strategies. Simulation p-values are obtained with the proposed simulation test with 5'000 trials.

| | Investment strategies | | | |
	BH	Opt_all	Opt_4	Voting
Nbj	3761	2907	2872	2924
Right_buy	53.12%	54.25%	54.70%	54.38%
Simulation p-value		*0.012*	*0.028*	*0.072*
Nsj		853	889	837
Right_sell		50.53%	51.74%	51.14%
Simulation p-value		*0.012*	*0.029*	*0.071*
Right_strat		53.40%	54.00%	53.66%
Simulation p-value		*0.012*	*0.028*	*0.072*

5.2   Complex trading rules in the standard investment setting

In Table 24, we display the same technical trading strategies performance analysis presented in Section 4.2.3 of the first part, and in addition, we perform the statistical tests with the proposed simulation method. As mentioned earlier, the strategies generate large abnormal returns; however, the difference is not statistically significant according to Student t-tests. The t-statistic associated with the *Opt_4* strategy, which yields an average return more than twice as high as the buy-and-hold, is only 1.2. This corresponds to a p-value of 0.23, which is far from standard confidence levels of 5% or 10%. On the other hand, all three simulation p-values are smaller than 5%. This means that the strategies returns can not be replicated by taking random positions in the market, and thus, the performances are due to market timing abilities, rather than luck. The contrast between the conclusions derived from the two testing procedures is striking. Indeed, we would have concluded that the forecasting ability is not sufficient to generate statistically significant abnormal returns with standard Student t-tests., while our test provide statistical evidence in favour of significant market timing abilities. These results are also in line with those

presented in the first part with our long-term market timing test, which focuses of trading position instead of returns.

Table 24: Complex rules performance

This table reports the performance analysis for the three investment strategies and for the buy-and-hold, named *BH*. *t*-statistics associated with mean returns are calculated with standard Student *t*-tests between the stated return (buy, sell or strategy) and the mean buy-and-hold. Simulation *p*-values are computed with 5'000 trials. *SR* is the Sharpe ratio. All values are in an annual frequency.

| | Investment strategies | | |
	BH	Opt_all	Opt_4	Voting
Simple mean return	0.0616	0.1070	0.1461	0.1200
Simulation *p*-value		*0.028*	*0.007*	*0.040*
t-statistic		*0.65*	*1.20*	*0.83*
Simple mean buy return		0.1098	0.1360	0.1167
Simulation *p*-value		*0.027*	*0.007*	*0.040*
t-statistic		*0.70*	*1.09*	*0.80*
Simple mean sell return		0.0975	0.1789	0.1312
Simulation *p*-value		*0.030*	*0.008*	*0.041*
t-statistic		*0.28*	*0.92*	*0.54*
Compounded mean return	0.0441	0.0925	0.1362	0.1068
Simulation *p*-value		*0.028*	*0.007*	*0.040*
Volatility	0.192	0.192	0.192	0.192
Simulation *p*-value		*0.954*	*0.993*	*0.960*
Beta	1	-0.04	-0.09	-0.03
Simulation *p*-value		*1.000*	*0.998*	*0.993*
t-statistic		*-2.75*	*-5.74*	*-2.12*
Alpha		0.0790	0.1198	0.0917
Simulation *p*-value		*0.020*	*0.009*	*0.029*
t-statistic		*1.59*	*2.42*	*1.85*
SR	0.168	0.404	0.609	0.472
Simulation p-value		*0.032*	*0.010*	*0.039*

The other measures suggest that the performance is not due to risk bearing, at least with the considered indicators. Indeed, strategies volatilities and betas are lower than those of random strategies. In addition, simulation p-values associated with risk-adjusted performance measures, i.e. the Jensen's alphas and the Sharpe ratios, are in line with those of mean returns. It is also noteworthy that Jensen's alpha Student t-statistics are sharply higher than those associated with mean returns. However, only one out of the three coefficients is statistically different from zero with the Student t-tests, while the two others are also economically large, with respective values of 7.9% and 9.1% in annual term. According to our simulation test, they are all statistically different from zero.

We further examine this issue with the Monte-Carlo experiment displayed in the last Section. These results suggest that the proposed simulation test generates statistical conclusions in line with the strategies large economical out-performance.

Finally, we test whether the simulation p-values are consistent over various trials. We compute them 500 times for the Opt_all strategy, with several numbers of internal simulations, defined as N in the test description, or in other words, the number of artificial trading signals constructed to estimate the p-value. While we can not objectively define a maximum acceptable level of the variation in the p-value resulting from several test trials, Table 25 shows that using 5'000 internal simulations is enough to have a consistent p-value in the standard 5% confidence setting. Indeed, every single test among the 500 trials generates a p-value lower than 5%. Nonetheless, the results obtained with less than 2'000 simulations may lead to different conclusions as some p-values are higher than 5%.

Table 25: The simulation p-values distribution for the Opt_all strategy

This table reports the distribution of simulation p-values obtained by running the test 500 times. Number of internal simulation is the number of trials used to compute each p-value.

	N=	Number of internal simulations						
		100	300	500	700	900	2000	5000
min		0.0000	0.0100	0.0140	0.0143	0.0178	0.0235	0.0240
5%		0.0100	0.0167	0.0220	0.0229	0.0244	0.0270	0.0288
mean		0.0318	0.0329	0.0336	0.0335	0.0335	0.0331	0.0334
median		0.0300	0.0333	0.0330	0.0329	0.0333	0.0330	0.0332
95%		0.0600	0.0500	0.0460	0.0457	0.0444	0.0400	0.0380
max		0.0700	0.0600	0.0620	0.0543	0.0567	0.0470	0.0430

In Appendix I-F, we present the complex rules performance according to the standard investment setting with a weekly frequency. Here, we present again these results, but with *p*-values computed according to the simulation test as well. Table 26 indicates that the small increase in the percentage of right signal between the buy-and-hold and the trading rules is sufficient to generate *p*-values lower than 5%.

Table 26: Percentage and right signals – weekly frequency

This table presents the percentages of right signals, respectively for long and short positions and for the whole strategy. *BH* is the buy-and-hold, *Opt_all, Opt_4* and *Voting* are the three complex strategies. Simulation *p*-values are obtained with the proposed simulation test with 5'000 trials.

	Investment strategies			
	BH	Opt_all	Opt_4	Voting
Nbj	779	604	596	591
Right_buy	56.23%	58.61%	59.40%	59.39%
Simulation *p*-value		*0.030*	*0.012*	*0.021*
Nsj		174	183	182
Right_sell		51.72%	54.10%	53.85%
Simulation *p*-value		*0.033*	*0.013*	*0.021*
Right_strat		57.07%	58.15%	57.64%
Simulation *p*-value		*0.030*	*0.011*	*0.035*

In Table 27, we also find very similar results to those associated with the daily frequency. Indeed, the Student *t*-tests do not reject any null hypothesis of equal means between any parts of the strategies returns and the buy-and-hold, while simulation *p*-values are consistently lower than the 5% limit. These conclusions hold for all performance measures.

Table 27: Complex rules performance – weekly frequency

This table reports the performance analysis for the three investment strategies and for the buy-and-hold, named *BH*. The *t*-statistics associated with mean returns are calculated with standard Student *t*-tests between the stated return (buy, sell or strategy) and the mean buy-and-hold. Simulation *p*-values are computed with 5'000 trials with the proposed simulation test. *SR* is the Sharpe ratio. All values are in an annual frequency.

	Investment strategies				
	BH	Best	Opt_all	Opt_4	Voting
Simple mean return	0.0576	0.1412	0.1133	0.1303	0.1250
Simulation *p*-value		*0.011*	*0.032*	*0.023*	*0.033*
t-statistic		*1.35*	*0.90*	*1.17*	*1.09*
Simple mean buy return		0.1310	0.1103	0.1228	0.1220
Simulation *p*-value		*0.016*	*0.032*	*0.023*	*0.030*
t-statistic		*1.17*	*0.86*	*1.06*	*1.04*
Simple mean sell return		0.1909	0.1241	0.1547	0.1388
Simulation *p*-value		*0.009*	*0.032*	*0.022*	*0.035*
t-statistic		*1.19*	*0.59*	*0.88*	*0.73*
Compounded mean return	0.0439	0.1354	0.1038	0.1227	0.1169
Simulation *p*-value		*0.011*	*0.033*	*0.023*	*0.033*
Volatility	0.170	0.168	0.170	0.170	0.169
Simulation *p*-value		*0.454*	*0.925*	*0.977*	*0.849*
Beta	1	0.00	0.03	0.01	0.02
Simulation *p*-value		*0.995*	*0.999*	*0.997*	*0.996*
t-statistic		*0.10*	*0.71*	*0.19*	*0.53*
Alpha		0.1021	0.0738	0.0912	0.0857
Simulation *p*-value		*0.011*	*0.025*	*0.010*	*0.032*
t-statistic		*2.35*	*1.68*	*2.08*	*1.96*
SR	0.108	0.608	0.437	0.537	0.507
Simulation p-value		*0.017*	*0.039*	*0.015*	*0.033*

5.3   Complex trading rules with options

In the first part of this thesis, we show that using options to improve the strategies performance is problematic, as the loss of time value lessens the profitability, unless the trading rule possesses strong forecasting abilities. In Table 28, we report the simulation *p*-values associated with the daily mean returns of the option part of the strategies, i.e. when the whole capital is invested in options. To construct the artificial returns series, we proceed as follows; for each trading rule, we generate 5'000 artificial signals series as in the standard investment setting. Then, we compute options returns with these signals. This means that the options are not the same as the original series, but they correspond to the artificial signals series. The *p*-values associated with the call options taken after buy signals and those of the global strategy are lower than 5%. This indicates that the high mean returns generated by the trading rules are not due to the options leverage, but to their forecasting abilities. Nonetheless, the returns from the sell side with put options are not higher than those obtained by the random strategies. We also observe that options increase the returns volatility to a large extend, but not only for the original strategies. Indeed, the associated *p*-values lie between 31.3% and 87.1%.

We also provide the simulation test for strategies that invest only a part of the capital in options, while the remaining amount is invested according to the standard setting. The *p*-values computation process is similar to the above-mentioned one for an investment in options of 100%. In this case, the artificial strategy consists in investing a percentage in options according to the simulated signals series. As expected, the results displayed in Table 29 are in line with the previous ones. When options increase the performance, especially for the Opt_4 strategy, this can neither be attributed by the leverage offered by options, nor by higher risk.

Table 28: Leverage with options

BH is the buy-and-hold strategy with the selected options. Mean return is the daily simple mean return for the options taken according to the various trading rules signals. Volatility is the daily volatility of the options used in the trading strategies. Here, the whole capital is invested in options. Simulation *p*-values are computed with 5'000 trials.

	Investment strategies					
	BH	Best	Opt_all	Voting	Opt_4	Partial
Long position - Calls						
Mean return	-0.0042	0.0030	0.0018	0.0019	0.0045	0.0026
Simulation *p*-values		*0.0160*	*0.0190*	*0.0330*	*0.0040*	*0.0140*
Volatility	0.3363	0.3300	0.3413	0.3421	0.3312	0.3295
Simulation *p*-values		*0.672*	*0.385*	*0.313*	*0.649*	*0.368*
Short position - Puts						
Mean return	-0.0166	0.0132	-0.0019	0.0066	0.0136	0.0058
Simulation *p*-values		*0.0370*	*0.1050*	*0.1650*	*0.0200*	*0.1290*
Volatility	0.3413	0.3303	0.2793	0.3294	0.3272	0.2634
Simulation *p*-values		*0.483*	*0.871*	*0.489*	*0.476*	*0.619*
Strategy - Calls and puts						
Mean return	-0.0042	0.0054	0.0010	0.0030	0.0066	0.0033
Simulation *p*-values		*0.0240*	*0.0350*	*0.0560*	*0.0080*	*0.0330*
Volatility	0.3363	0.339	0.330	0.328	0.339	0.330
Simulation *p*-values		*0.654*	*0.649*	*0.437*	*0.615*	*0.628*

Table 29: Strategies with options

This table presents the performance of the trading strategies with traded options according to the three level of leverage. Mean buy, sell and strategy are respectively the simple mean of the long, short and global positions. Compound is the mean compound return of the strategies and Annual compound is it annualized value. Vol strategy is the daily volatility. Simulation p-values are computed with 5'000 trials.

Panel A: 5% of options

	BH	Best	Opt_all	Voting	Opt_4	Partial
Mean buy	0.0156	0.0009	0.0007	0.0008	0.0010	0.0008
Simulation p-value		*0.014*	*0.02*	*0.035*	*0.005*	*0.016*
Mean sell	-0.0171	0.0015	0.0005	0.0010	0.0015	0.0008
Simulation p-value		*0.025*	*0.072*	*0.094*	*0.011*	*0.091*
Mean strategy	0.0003	0.0010	0.0007	0.0008	0.0011	0.0008
Simulation p-value		*0.015*	*0.038*	*0.052*	*0.006*	*0.029*
Compound	0.0000	0.0007	0.0004	0.0005	0.0008	0.0005
Simulation p-value		*0.015*	*0.036*	*0.051*	*0.006*	*0.03*
Vol strategy	0.0257	0.0253	0.0253	0.0257	0.0253	0.0238
Simulation p-value		*0.733*	*0.732*	*0.542*	*0.717*	*0.748*

Panel B: 10% of options

	BH	Best	Opt_all	Voting	Opt_4	Partial
Mean buy	0.0233	0.0013	0.0011	0.0011	0.0014	0.0011
Simulation p-value		*0.015*	*0.021*	*0.035*	*0.005*	*0.015*
Mean sell	-0.0257	0.0022	0.0005	0.0014	0.0023	0.0012
Simulation p-value		*0.032*	*0.091*	*0.124*	*0.013*	*0.115*
Mean strategy	0.0003	0.0015	0.0009	0.0012	0.0016	0.0012
Simulation p-value		*0.02*	*0.038*	*0.058*	*0.006*	*0.032*
Compound	-0.0005	0.0007	0.0002	0.0004	0.0008	0.0005
Simulation p-value		*0.017*	*0.036*	*0.057*	*0.007*	*0.03*
Vol strategy	0.0413	0.0406	0.0405	0.0415	0.0406	0.0379
Simulation p-value		*0.698*	*0.701*	*0.486*	*0.669*	*0.729*

Panel C: 15% of options

	BH	Best	Opt_all	Voting	Opt_4	Partial
Mean buy	0.0310	0.0016	0.0014	0.0014	0.0018	0.0015
Simulation p-value		*0.015*	*0.02*	*0.036*	*0.005*	*0.016*
Mean sell	-0.0344	0.0030	0.0006	0.0019	0.0031	0.0016
Simulation p-value		*0.036*	*0.101*	*0.15*	*0.018*	*0.119*
Mean strategy	0.0004	0.0019	0.0012	0.0015	0.0021	0.0015
Simulation p-value		*0.024*	*0.04*	*0.058*	*0.007*	*0.034*
Compound	-0.0011	0.0005	-0.0002	0.0000	0.0007	0.0003
Simulation p-value		*0.017*	*0.035*	*0.063*	*0.007*	*0.028*
Vol strategy	0.0572	0.0562	0.0560	0.0576	0.0563	0.0524
Simulation p-value		*0.687*	*0.692*	*0.477*	*0.657*	*0.72*

5.4   Complex trading rules with debt

The results regarding investment strategies leveraged with debt, presented in Table 30, confirm two above-mentioned findings. First the difference in conclusions about the excess return significance, performed either with Student t-test or the proposed simulation test, is similar to the standard investment setting. Indeed, we find that Student t-tests are unable to reject the null of equal means, while simulation p-values are consistently lower than 5%, both for simple and compounded returns. In addition, the high performance can not be explained by a higher volatility, as the associated p-values are close to one for the first three strategies. This indicates that they are less risky than their simulated counterparts. Second, the strategies large excess returns obtained with debt leverage can not be attributed to the leverage itself. If it was the case, the simulated strategies would be able to generate returns as high as the original, which clearly contrasts with very low p-values.

To summarize, this application of our proposed test to technical analysis strategies shed lights of the following aspects; first, it is more powerful than standard Student t-test and the difference between these two testing procedures is striking. Student t-tests fail to reject the null, despite economically large excess returns. On the other hand, the simulation test provides statistical conclusions in line with these returns. Second, the use of leverage does not bias the results. Adding leverage without reliable forecasting abilities does not lead to a reduction in p-values. Third, the test is also valid with weekly data, options and compounded returns, as it does not rely on any predetermined returns distribution.

Table 30: Strategies with debt leverage

Simple mean and *Compounded mean* are the strategies mean returns in annual terms. The simulation *p*-values are computed with 5'000 trials. The *t*-statistic results from a Student *t*-test about equal means between the strategy and the buy-and-hold returns. *Volatility* is the annualized volatility.

Panel A: borrowing rate = US Bank prime loan

	Investment strategies					
	BH	Best	Opt_all	Voting	Opt_4	Partial
Simple mean	0.073	0.244	0.162	0.197	0.249	0.187
Simulation *p*-values		*0.011*	*0.034*	*0.041*	*0.010*	*0.000*
t-statistic		*1.21*	*0.63*	*0.88*	*1.25*	*0.83*
Compounded mean	0.000	0.185	0.092	0.13	0.191	0.127
Simulation *p*-values		*0.011*	*0.034*	*0.041*	*0.010*	*0.000*
Volatility	0.384	0.383	0.384	0.384	0.383	0.367
Simulation *p*-values		*0.991*	*0.894*	*0.979*	*0.991*	*0.368*

Panel B: borrowing rate =risk free rate

	BH	Best	Opt_all	Voting	Opt_4	Partial
Simple mean	0.093	0.26	0.178	0.212	0.264	0.202
Simulation *p*-values		*0.010*	*0.037*	*0.033*	*0.009*	*0.001*
t-statistic		*1.18*	*0.6*	*0.85*	*1.22*	*0.79*
Compounded mean	0.020	0.204	0.109	0.148	0.21	0.144
Simulation *p*-values		*0.010*	*0.037*	*0.033*	*0.009*	*0.002*
Volatility	0.384	0.383	0.384	0.384	0.383	0.367
Simulation *p*-values		*0.992*	*0.909*	*0.981*	*0.992*	*0.389*

Panel C: no borrowing cost

	BH	Best	Opt_all	Voting	Opt_4	Partial
Simple mean	0.123	0.285	0.203	0.238	0.289	0.227
Simulation *p*-values		*0.011*	*0.032*	*0.039*	*0.009*	*0.000*
t-statistic		*1.15*	*0.57*	*0.82*	*1.19*	*0.76*
Compounded mean	0.050	0.235	0.138	0.178	0.24	0.173
Simulation *p*-values		*0.011*	*0.032*	*0.039*	*0.009*	*0.000*
Volatility	0.384	0.383	0.384	0.384	0.383	0.367
Simulation *p*-values		*0.991*	*0.905*	*0.980*	*0.991*	*0.372*

6.   An application to mean-variance portfolios

Table 31 displays the performance analysis for the four complex portfolios, with an emphasis on the proposed simulation and Student *t*-tests. Panel A presents various performance measures for the entire sample, while Panel B and C display only mean returns, and the associated *p*-values, over two subsamples. As mentioned in the second part, the first two portfolios, PF_ALL and PF_4, produce returns that are more than twice as high as the benchmark over a 16 years sample. Indeed, the annual mean 1/N strategy return is 10.31% compared with 25.68% and 24.56% respectively for our complex portfolios.

Turning to the test procedures, we find results in line with technical analysis, even though the difference between our proposed simulation method and the standard Student *t*-tests is slightly less marked. Indeed, simulation *p*-values are lower than their Student counterparts, and the test conclusion regarding the PF_4 strategy would differ at the standard 5% level: The returns are not significantly different according to the Student *t*-test, even if the strategy yields an average return more than twice higher than the benchmark. On the other hand, the simulation *p*-value of 2.82% means that it is unlikely that the strategy performance is due to random positions in our universe of assets. With the proposed test, we conclude in a significant higher performance. The subsample analysis presented in Panel B and C shows that the simulation *p*-values are affected to a lesser extent by the shorter time interval. The differences in means are similar, but the simulation *p*-values increase less than the Student ones. However, they are not statistically different anymore at the 5% level, even with the proposed test. This relationship between the tests power and the sample length is further investigated in the Monte-Carlo experiment presented in the next Section.

The two simulation *p*-values of 0.527 and 0.5 associated with the long-only portfolios are also noteworthy. PF_ALL_NS and PF_4_NS mean returns are very close to the benchmark strategy. This means that returns from artificial series used to compute these *p*-values are evenly distributed between returns that are higher and lower than the original ones. In other words, they correspond to our expectations for strategies that are neither better nor worse than the benchmark. The difference between the two kinds of *p*-values[88] associated with portfolios that neither out-perform nor under-perform makes statistical sense.

---

[88] As a reminder, appendix I-A proposes a summary of the Student *t*-test *p*-value.

Table 31: Portfolios performance analysis

1/N is the benchmark strategy represented by an equally weighted portfolio including only equity indices. The four other columns are the complex portfolios described in Section 3.2 of the part II. All results are given in an annual frequency (except *Nb trade* and *BE TC*, which respectively represent the number of transactions and break even transaction costs). Simulation *p*-values are estimated with 5'000 replications and a block length of 1 month. Student *p*-values associated with simple mean returns are obtained with bilateral Student *t*-tests between the respective portfolio and the buy-and-hold. The Student *t*-test for Sharpe ratios follows Memmel (2003), and the null corresponds to two ratios with the same value. In contrast, the Jensen's alphas tests focus on whether coefficients are statistically different from zero.

Panel A: Entire sample June 1994 - December 2009

	1/N	PF_ALL	PF_4	PF_ALL_NS	PF_4_NS
Simple mean return	0.1031	0.2568	0.2456	0.0965	0.0965
Simulation *p*-value		*0.0116*	*0.0282*	*0.5272*	*0.5002*
Student *p*-value		*0.0350*	*0.0562*	*0.9054*	*0.9031*
Compouned mean return	0.0928	0.2566	0.2401	0.0903	0.0913
Simulation *p*-value		*0.0082*	*0.0160*	*0.4654*	*0.4476*
Volatility	0.1690	0.2376	0.2462	0.1415	0.1350
Simulation *p*-value		*1.0000*	*0.9998*	*1.0000*	*0.9990*
Jensen's alpha		0.1963	0.1772	0.0207	0.0202
Simulation *p*-value		*0.0048*	*0.0092*	*0.2926*	*0.2788*
Student *p*-value		*0.0008*	*0.0030*	*0.4727*	*0.4471*
Sharpe ratio	0.3098	0.8672	0.7916	0.3236	0.3390
Simulation *p*-value		*0.0038*	*0.0060*	*0.4062*	*0.3840*
Student *p*-value		*0.0909*	*0.1205*	*0.9525*	*0.8948*
Nb trade		311.4866	365.4386	115.7292	117.3309
BE TC		0.0079	0.0062	-0.0009	-0.0009
Beta		0.1856	0.3381	0.4792	0.4882
Simulation *p*-value		*1*	*1*	*1*	*1*

Panel B: First subsample June 1994 - March 2002

	1/N	PF_ALL	PF_4
Simple mean return	0.0995	0.2621	0.2755
Simulation *p*-value		0.0752	0.0556
Student *p*-value		0.1108	0.0815

Panel C: Second subsample March 2002 - December 2009

	1/N	PF_ALL	PF_4
Simple mean return	0.1067	0.2515	0.2157
Simulation *p*-value		0.0632	0.1778
Student *p*-value		0.1646	0.3212

Student p-values should be close to one, as it is very likely that the mean returns are from a similar distribution. On the other hand, a simulation p-value of 1 would mean that all simulated series have a higher return than the original, which means that the portfolio under-perform statistically.

Turning to the other performance measures, we find mixed evidence about our test superior power compared with the standard procedure. First, we reach opposite conclusions regarding the Sharpe ratios with the two tests. We use the Jobson and Korkie (1981) methodology with the Memmel (2003) correction to compute the Student p-value about the equality of two ratios[89]. They are estimated as:

$$z = \frac{\sigma_1 \mu_1 - \sigma_2 \mu_2}{\sqrt{\vartheta}} \tag{3.13}$$

$$\vartheta = \frac{1}{N} \left(2\sigma_1^2 \sigma_2^2 - 2\sigma_1 \sigma_2 \sigma_{1,2} + \frac{1}{2}\mu_1^2 \sigma_2^2 + \frac{1}{2}\mu_2^2 \sigma_1^2 - \frac{\mu_1 \mu_2}{\sigma_1 \sigma_2} \sigma_{1,2}^2 \right) \tag{3.14}$$

$$\text{Student } p\text{-value} = 2\left(1 - student's\ t\ cdf\left(|z|, df\right)\right) \tag{3.15}$$

where σ_i and μ_i are the estimated standard deviation and mean return of the ith portfolio computed over N observations and $\sigma_{1,2}$ is the estimated covariance between the two portfolios returns. The two complex portfolios that allow short-sales, PF_ALL and PF_4, have a Sharpe ratio of 0.86 and 0.79, while the benchmark ratio is only 0.31. The associated Student p-values (0.091 and 0.12) are much higher than those resulting from the proposed test methodology (0.004 and 0.006). Furthermore, the differences in ratios are significant with the proposed test method, but not with standard Student t-tests at the usual 5% level. This contrasts with Jensen's alphas that are statistically different from zero, whatever the test considered. It is also worth noting that the small positive difference between the two long-only portfolios and the benchmark Sharpe ratios (0.32 and 0.34 for the strategies and 0.31 for the benchmark) is not enough to reject the null hypothesis. This means that the proposed test methodology also requires an economically significant difference in order to generate a statistical one.

Table 32 and Table 33 shed some light on the robustness of the simulation test regarding the length of the block bootstrap and the number of internal simulations. In Table 32, we report the

---

[89] We refer to this test as a Student t-test as it follows the same procedure than the standard Student t-test about the equality in mean returns.

simulation *p*-value obtained with various block lengths, which are used to construct the artificial portfolio weights series

Table 32: Simulation *p*-values and the length of the block bootstrap

This table reports the *p*-values obtained with the proposed simulation method when various block lengths are considered in the simulation process. All tests are performed with 5'000 replications.

Panel A: PF_ALL

	block length in months					
	1	2	3	4	8	12
Simple mean *p*-value	0.0136	0.016	0.0136	0.0094	0.0058	0.0214
Compounded mean *p*-value	0.0078	0.011	0.0078	0.0054	0.0034	0.0162
Jensen's alpha *p*-value	0.0046	0.0064	0.0054	0.003	0.003	0.0114

Panel B: PF_4

	block length in months					
	1	2	3	4	8	12
Simple mean *p*-value	0.0266	0.0318	0.0226	0.0246	0.012	0.0376
Compounded mean *p*-value	0.0146	0.0202	0.017	0.0162	0.009	0.029
Jensen's alpha *p*-value	0.0084	0.0132	0.011	0.0114	0.007	0.021

. We find that this length does not influence the simulation *p*-values to a large extent. Indeed, they are close for the three test statistics considered, the simple or compounded mean return and the Jensen's alpha. In addition, all tests reach in the same conclusion at the 5% confidence level.

Second, Table 33 displays the characteristics of the distribution obtained by running the simulation tests 500 times with various numbers of internal simulations. We show that using 5'000 internal simulation is sufficient to obtain consistent *p*-values. In other words, running the simulation tests many times provides the same conclusion regarding the significance of the abnormal returns.

Table 33: Simulation p-values and the number of internal simulations

This table describes the distribution of p-values obtained with the proposed simulation method applied to the PF_4 strategy. The number of internal simulations indicates with how many trials the p-value is computed. We perform the simulation test 500 times for each number of internal simulations.

|	Number of internal simulations						
	100	300	500	700	900	2000	5000
min	0.0000	0.0033	0.0080	0.0057	0.0111	0.0130	0.0192
5%	0.0000	0.0100	0.0140	0.0143	0.0167	0.0190	0.0211
mean	0.0258	0.0244	0.0245	0.0245	0.0244	0.0245	0.0245
median	0.0200	0.0233	0.0240	0.0243	0.0244	0.0245	0.0244
95%	0.0550	0.0433	0.0360	0.0343	0.0328	0.0300	0.0279
max	0.0900	0.0567	0.0440	0.0457	0.0378	0.0345	0.0302

7.   A Monte-Carlo experiment

The results presented so far show that statistical tests performed with the proposed simulation method are in line with the large excess returns produced by our investment strategies. This contrasts with the results obtained in the first two parts that mainly use Student t-tests that usually fail to reject the null hypothesis of equal mean returns. However, using only a few series is not enough to reach robust conclusions about the various tests power. Thus, in this Section, we compare the performance measures in a simulation study. It focuses on strategies returns in excess of the buy-and-hold and their associated p-values obtained with several testing procedures. We also consider various time series lengths, as this may affect the tests power. The simulation proceeds as follows: The first step is to generate an artificial buy-and-hold series, based on the original one. Then, we construct several investment strategies, but we only select those with abnormal returns that correspond to a desired target. Finally, various performance measures are computed and compared. The main objective is to compare the chosen level of excess return of an investment strategy and the results of several performance measures. We want to stress that in this Section we really focus on the power of various performance measures, not weather investment strategies are successful. Indeed, we simulate trading rules (i.e. technical analysis ones and mean variance portfolios) as long as we find some that have a specified abnormal returns. This is a pure data-mining procedure that should not be use to draw conclusions about these strategies and market efficiency.

The first element to define in our Monte-Carlo experiment is the measure to calculate excess returns between a simulated benchmark and an investment strategy. We choose two different excess returns measures for the single and multiple-assets cases; the difference in returns between

the strategies and the buy-and-hold for the first, and the Jensen's alpha for the second. The Jensen's alpha has the advantage to include risk; however, its use induces a bias against the Student *t*-test. Indeed, a strategy may have a positive alpha, while its mean return is similar to the buy-and-hold. Hence, Student *t*-test would not reject the null hypothesis correctly even if we would consider this strategy to out perform the benchmark. Nonetheless, Figure 24 shows that the difference between alphas and the median difference in returns is relatively small in the multiple-assets setting, while it is sharply larger for the single-asset case. For instance, when a single-asset strategy has an annual alpha of 2%, the median difference of mean between the strategy and the benchmark is around -2%. Thus, we use the alpha as the measure of excess return only in the multiple-assets case. The technical analysis strategies have a beta close to zero on average.

Figure 24: Median differences in returns and alpha

Note: This graph shows the median difference between the two excess return measures, the Jensen's alpha and the difference in mean returns. For each level of alpha, the median difference in mean is calculated over 500 simulations of both an investment and benchmark strategy. Both the Jensen's alpha and the median difference in means are annualized. The simulated series are those used in the Monte-Carlo experiment.
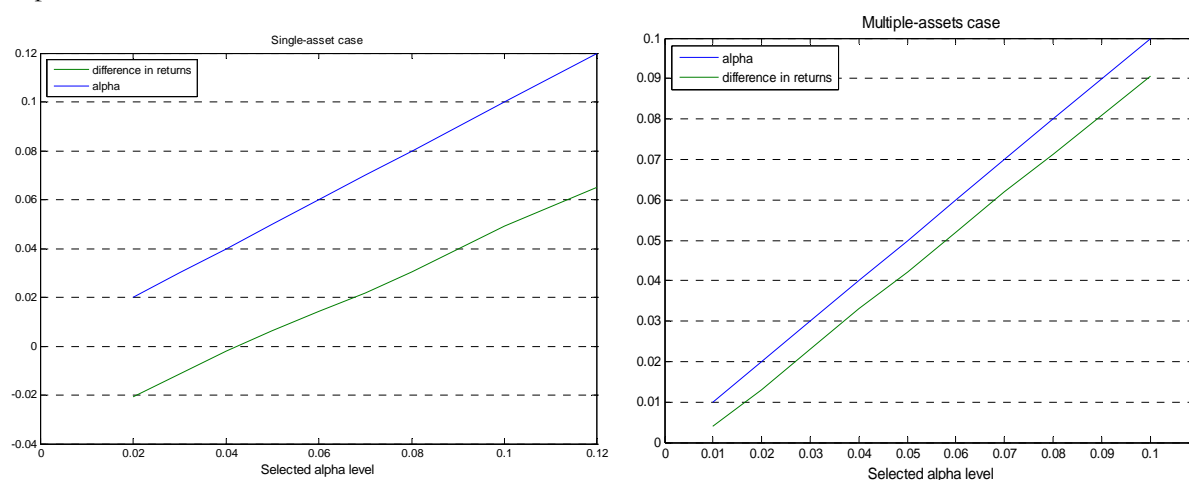
7.1   Single-asset case

This Section aims to compare the power of the proposed simulation test with other performance measures, which are widely used in the technical analysis setting; the bootstrap method introduced by BLL, the Student *t*-test of equal means, the Jensen's alpha and the Sharpe ratio. The overall procedure is the following:

1. First, we generate a four years artificial prices series by resampling with replacement the original Standard & Poor 500 returns with a block bootstrap[90]. The price series is obtained by setting the first price at 100, and then according to the resampled returns. We also include the early estimation period, which starts in January 1985. Thus, the unconditional resampled buy-and-hold returns distribution is not similar to the original, which is used in the performance analysis presented in Section 4.2.3 of the first part. This enables the simulation experiment to take into account various market situations, as the buy-and-hold return is not necessary positive in all trials.

2. We generate investment strategies with a subset of the 1'876 simple MA rules described in the first part. We use a subset of 690 simple MA rules with a short (long) window ranging from one (10) to 100 (990) days. The scope of window lengths are similar than the full set used in the first part, but we consider less values between the minimum and maximum. A 1% bandwidth is also used. Not using the whole set of MA rules speed up the simulation.

3. Then we calculate the difference in return for each of these 690 strategies with the benchmark and select those that have a difference between mean returns within a 4% margin[91] from an abnormal return target. These targets range from 4% to 18% in an annual frequency with an incremental step of 2%[92]. This margin aims to speed up the simulation process. Indeed, it can be extremely time consuming, or even impossible, to find strategies with the exact desired level of abnormal return.

4. Selected strategies should also fit the following requirements[93]: They should move from a long to a short position, or conversely, at least four times, and it should not

---

[90] The block length is set to 21 observations corresponding to one trading month. Results are similar with 42 observations, however results are not reported. This value is subjective, but should ensure to keep at least some of the dependencies in the simulation process.

[91] For instance, for the 6% abnormal return level, only strategies with an annual difference in mean return between the strategy and the benchmark that ranges between 5,76% and 6,24% are selected.

[92] We only consider abnormal returns up to 10% when for series longer than 10 years as it is impossible, or extremely time consuming, to get strategies with such high abnormal return over a long sample. Indeed, it took us around four months, 24 hours a day, to run this Monte-Carlo experiment with Matlab on a recent laptop.

[93] As this simulation aims to compare performance measures and not the profitability of investment strategies, the data mining bias introduced by selecting strategies with a specific level of abnormal return is not relevant.

have the same position more than 70% of the time. This ensures that there is a minimum of trading activities. Finally, we include only one strategy for each abnormal return target and artificial price series. If there is more than one strategy at this stage, we select only one randomly.

5. We compute the proposed simulation and the BLL[94] tests with 500 "internal" simulations and other standard performance measures, the Student t-test for the difference in mean returns, the Jensen's alpha and the Sharpe ratio. These last two measures are also tested for statistical significance with the proposed simulation test.

6. Repeat the procedure from the first step until 250 strategies are selected and performance measures are computed for each level of abnormal return target.

7. Repeat the entire procedure with six, eight, 10, 12, 14, 16 years of returns in the first step.

Figure 25 provides a schematic overview of this procedure for the four years interval.

Figure 26 displays the median p-value obtained in the experiment for the Student t-test, the proposed simulation and the BLL tests. First, it shows that the Student t-tests power is much lower than the two simulation-based methodologies. Even at the highest abnormal return level, they are not able to reject on average the null hypothesis of equal means at the 5% level. On the other hand, the two simulation-based tests generate consistently lower p-values. Second, these two tests produce similar p-values, but the BLL test generates slightly lower p-values. However, we can not claim that this test has a higher power than the proposed one, as it does not test directly for abnormal returns.

---

[94] This simulation test proceeds as follow: First, we estimate an AR(1) model on the buy-and-hold series. Then, new artificial returns and prices series are obtained by resampling the residuals issues from the AR(1) model. The next step is to evaluate the MA rule on this new prices series and to compute the mean return. This process is repeated 500 times, and thus, 500 mean returns are generated. Finally, p-values are calculated by comparing the original mean return with the simulated ones. We use only the AR(1) model as return generating process.

Figure 25: Summary of the Monte-Carlo simulation

Note: This graph presents a summary of the simulation procedure for the first time interval of four years.

1:    | Generate random buy-and-hold series with 2'520 returns from the original |

2:    | Generate 690 investments strategies based on simple MA rules. |

3:    | Calculate the excess return for each of these strategies and select those that correspond to an abnormal return target. |

4:    | If no strategy is selected, go back to step 1, otherwise check that they have a minimum amount of trading. If more than one strategy is selected for an abnormal return target, select only one of them randomly. |

5:    | If no strategy is selected, go back to step 1, otherwise compute the various performance measures for the selected strategy, including the proposed simulation test with 500 internal simulations. |

6:    | Repeat the procedure from step 1 until 250 strategies are selected and performance measures are computed for each level of abnormal return target. |

Figure 26: Median *p*-values

Note: This figures shows the median *p*-values obtained during this Monte-Carlo experiment for various levels of abnormal returns and series lengths. The median is used as, by definition, *p*-values are not normally distributed. We use different scale on the Z axes as the range of values differs largely among the three methods.
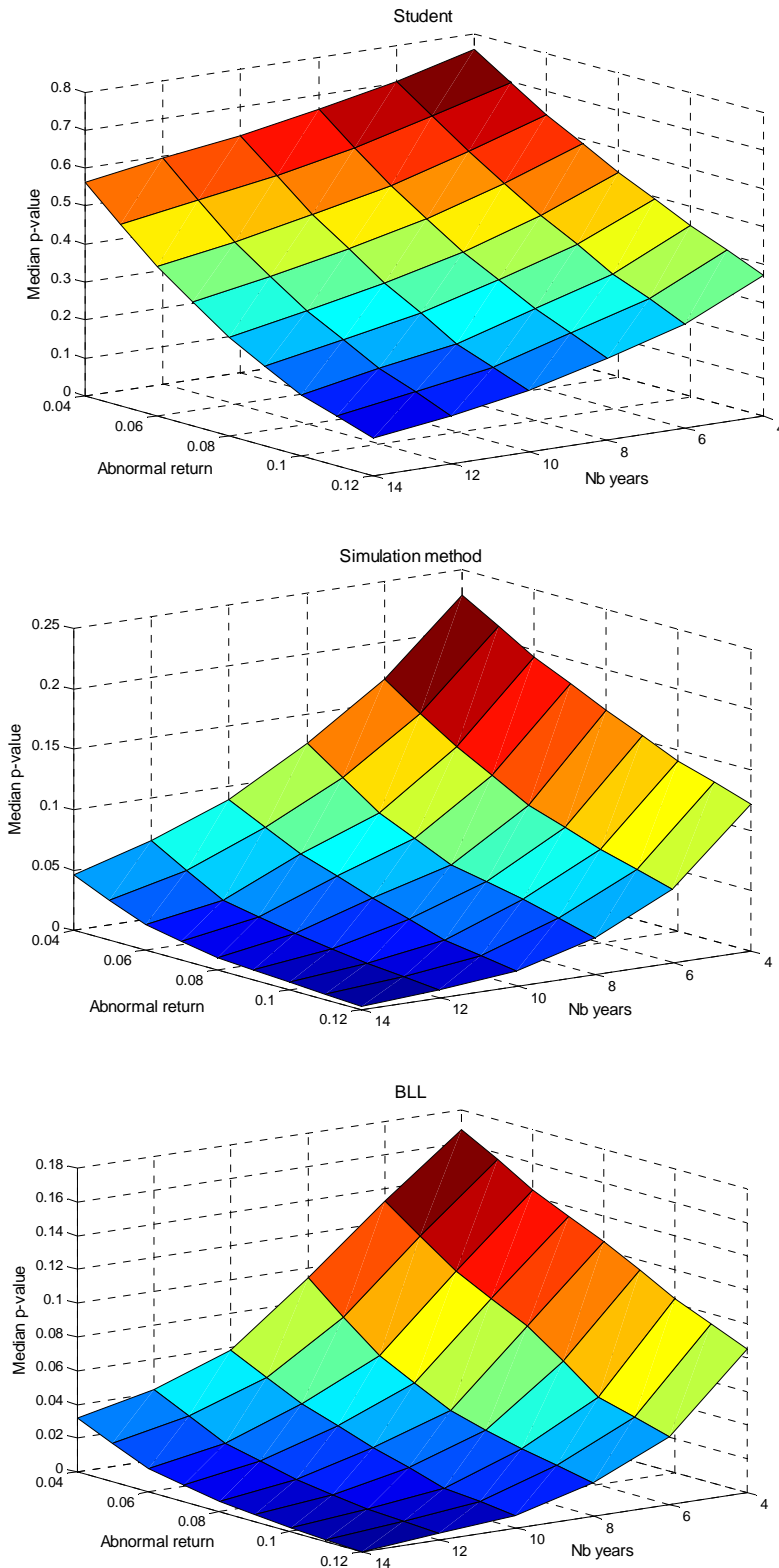
Figure 27 presents some cross-sections. For example, the proposed simulation (BLL) test requires an abnormal return of 10% (8%) to generate a median *p*-value smaller than 5% for eight years samples. On the contrary, the Student *t*-test median *p*-value amounts to 0.30 with a 10% abnormal return and only 0.068 at the 18% level. For 12 years samples, the proposed test generates a median *p*-value lower than 5% with 5% abnormal returns, while the Student one is still higher than 10% with 12% abnormal returns.

Figure 27: Some cross-sections
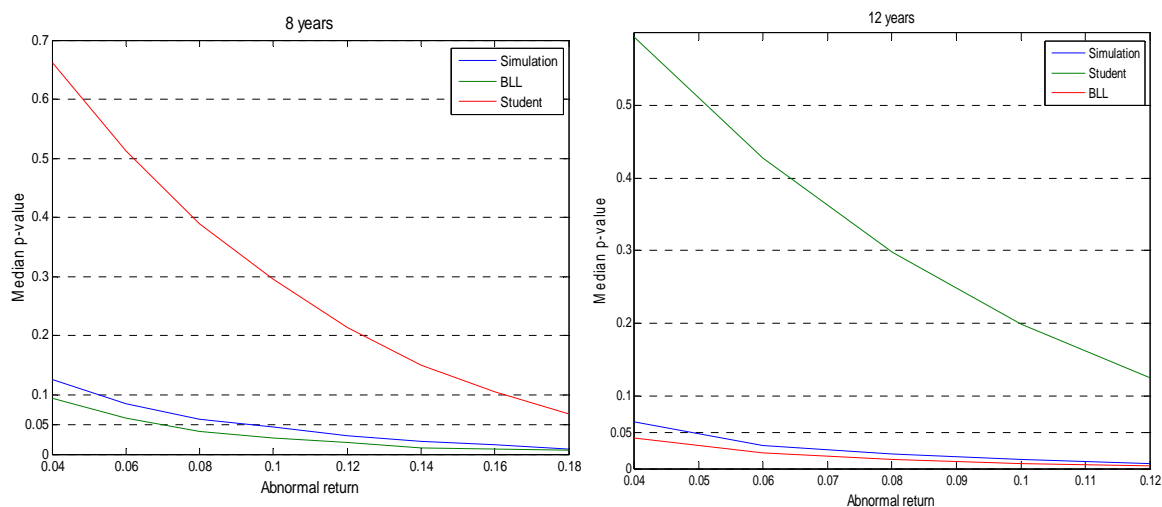
These findings are confirmed by the results summarized in Figure 28, which reports the percentage of tests that generate a *p*-value lower than 5%. The Student *t*-test power is strikingly low, even for 14 years samples, as only a few tests generate a significant difference with abnormal returns as high as 12%. For eight years samples, only 20% of the tests reject the null for 18% abnormal returns, while none of them reject it with four years sample. On the other hand, the rejection frequency is sharply higher for the two simulation-based tests. As noted above, the BLL test produces consistently lower *p*-values. While more powerful than Student *t*-test, our test requires nonetheless economically large abnormal returns to find a significant performance.

In addition, the sample length has a significant impact on the three testing procedures. The explanation for the Student *t*-test is straightforward; as the number of observation used to

compute the means grows, the means variance decreases[95], and thus, the *t*-statistic increases. For the proposed simulation test, we argue that it is less likely that a random strategy generates large returns over an extended period than for short samples. Hence, the artificial returns dispersion is larger for short samples, resulting in higher *p*-values associated with the same level of abnormal return.

Figure 28: Percentage of significant tests

Note: This figure presents the percentage of tests that result in a rejection of the null hypothesis that the simulated strategy has a similar mean return than the benchmark. Simulation is our proposed test, BLL is the Brock, Lakonishok and LeBaron (1992) and Student is the test for the difference in mean returns performed with bilateral Student *t*-tests. Note that for the 14 years interval, the abnormal return is at most 12% while it is 18% for others lengths.
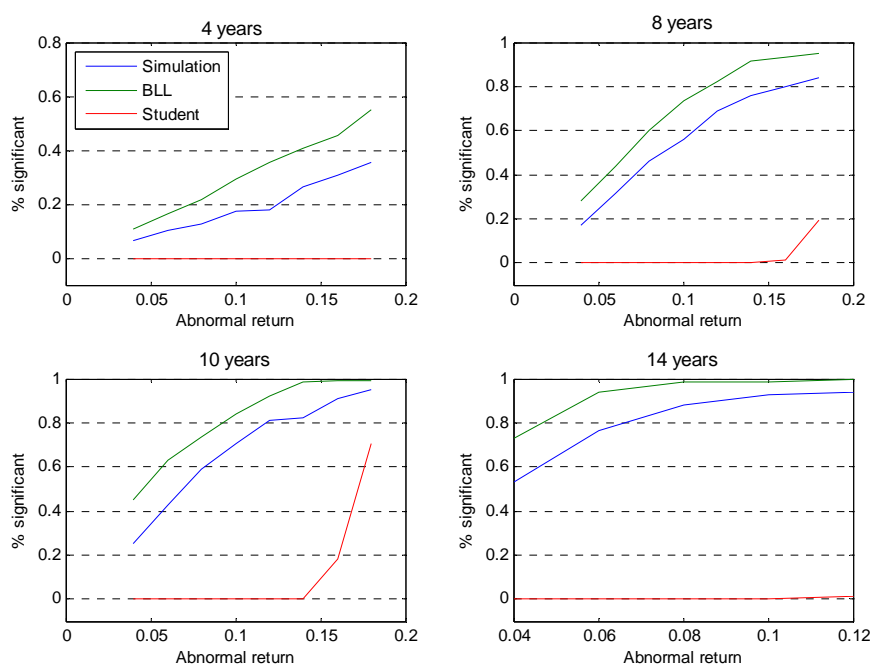
Table 34 and Table 35 present some of these results numerically, and those related to the two other performance measures, the Jensen's alpha and the difference in Sharpe ratio. First, the vast majority of the tests resulting in a significant difference in returns, with the proposed simulation tests, reach a similar conclusion with the BLL test, as well. For example, at the 16% (8%) abnormal return level and a sample length of 8 (16) years, there are 197 (231) tests out of the 200 (232) significant differences with the proposed simulation method that also result in a significant performance with the BLL test. Thus, even if the research question differs between the two testing procedures, the results are very similar. However, using the BLL test alone for strategies

---

[95] Indeed, the mean variance is the series variance divided by the number of observation.

with a relatively small superior performance may be misleading, as this procedure does not test for abnormal returns. Indeed, there are 30 cases when the BLL test generates p-values lower than 5% for strategies with a 4% abnormal return but without any indication that the performance is higher than the buy-and-hold according to the proposed simulation test (and neither according to the other standard measures of performance).

It is interesting to note that there are more significant Jensen's alphas than Student t-tests for differences in mean returns. However, testing whether these alphas are significant with our simulation method is more powerful than performing Student t-tests, especially for the lowest abnormal return levels. Furthermore, the simulation test generates almost identical results when we test the performance either with abnormal returns or Jensen's alphas. However, the median value of alphas differs slightly from the returns in excess of the buy-and-hold, and thus, the two testing procedures are not directly comparable.

If we compare the strategies with a median alpha of 6.4% with those that have a difference in mean returns of 6%, we find that 23 (176) alphas are significant with Student t-tests (the proposed simulation method) while none of the differences in mean are statistically different with the Student t-tests. This observation also holds for the 8 years samples.

Turning to the difference in Sharpe ratios, we find that Student t-tests following the Memmel (2003) procedure have an equally low power as tests about the difference in mean returns. For 16 years samples and abnormal returns of 10%, the median strategies ratio is 0.597 compared with 0.063 for the simulated buy-and-hold, nonetheless, only 20 differences are found to be statistically significant with Student t-tests. It is also worth noting that the proposed simulation test generates almost the same number of significant differences in means and Sharpe ratios. This indicates that this test reach a similar conclusion for the three performance measure used in this experiment.

Table 34: A comparison of tests power – 16 years

This table reports various performance measures computed on 250 artificial investment strategies over 16 years samples. The abnormal returns are the differences in mean returns between the strategy and the buy-and-hold, in annual terms. ⊿ *in returns* means that the three testing procedure (the proposed simulation test, the BLL test and Student *t*-tests) assess whether means are statistically different. *Med p-value* is the median *p*-value obtained over the 250 series for each respective test, and *Nb sig* is the number of significant measures. Tests on Jensen's alphas determine whether the coefficient is statistically different from zero. *Med alpha* is the median annual alpha associated with strategies according to each level of abnormal returns. Similarly, *Med SR strategy* and *BH* are the median annual Sharpe ratios for the strategies and the corresponding buy-and-hold series. ⊿ *SR* refers to the difference between the strategy and the buy-and-hold Sharpe ratios. Student *t*-tests are performed according to Memmel (2003). Jensen's alphas and Sharpe ratios are annualized.

	Abnormal returns			
	0.04	0.06	0.08	0.1
Δ in returns				
Med p-value Simulation	0.034	0.014	0.006	0.002
Nb sig simulation	179	218	232	225
Med p-value BLL	0.018	0.008	0.002	0
Nb sig BLL	211	244	249	250
Nb sig BLL and Simulation	165	213	231	225
Med p-value Student	0.541	0.364	0.230	0.132
Nb sig Student	0	0	0	0
Jensen's alpha				
Med alpha	0.064	0.081	0.097	0.110
Med p-value Simulation	0.034	0.014	0.006	0.002
Nb sig simulation	176	216	229	230
Med p-value Student	0.154	0.067	0.033	0.016
Nb sig Student	23	93	164	240
Sharpe ratio				
Med SR strategy	0.373	0.454	0.525	0.597
Med SR BH	0.150	0.135	0.095	0.063
Δ SR - Med p-value Simulation	0.034	0.014	0.006	0.002
Δ SR - Nb sig Simulation	176	218	232	225
Δ SR - Med p-value Student	0.508	0.340	0.228	0.142
Δ SR - Nb sig Student	0	0	0	20

Table 35: A comparison of tests power – 8 years

This table reports various performance measures computed on 250 artificial investment strategies over 8 years samples. The notation is similar to Table 34.

	Abnormal returns						
	0.04	0.08	0.1	0.12	0.14	0.16	0.18
Δ in returns							
Med p-value Simulation	0.126	0.058	0.046	0.03	0.021	0.016	0.008
Nb sig simulation	43	116	140	172	190	200	210
Med p-value BLL	0.094	0.039	0.028	0.02	0.01	0.008	0.006
Nb sig BLL	71	151	184	206	229	233	237
Nb sig BLL and Simulation	41	101	133	165	188	197	210
Med p-value Student	0.661	0.389	0.295	0.214	0.151	0.106	0.068
Nb sig Student	0	0	0	0	0	3	48
Jensen's alpha							
Med alpha	0.049	0.077	0.090	0.096	0.105	0.109	0.121
Med p-value Simulation	0.124	0.06	0.045	0.034	0.022	0.019	0.012
Nb sig simulation	46	112	137	163	182	190	199
Med p-value Student	0.405	0.213	0.164	0.126	0.101	0.090	0.068
Nb sig Student	7	30	39	63	83	87	107
Sharpe ratio							
Median SR strategy	0.287	0.436	0.482	0.528	0.589	0.617	0.668
Median SR BH	0.053	-0.015	-0.053	-0.109	-0.153	-0.208	-0.236
Δ SR - Med p-value Simulation	0.126	0.06	0.047	0.03	0.022	0.018	0.009
Δ SR - Nb sig Simulation	42	113	137	170	190	199	210
Δ SR - Med p-value Simulation	0.643	0.394	0.314	0.235	0.179	0.135	0.103
Δ SR - Nb sig Student	0	0	1	1	1	8	34

Finally, we also perform the simulation by selecting strategies that have a similar return than the buy-and-hold, with an error margin of 0.5% in annual term. The purpose is to examine whether the proposed tests do not reject the null hypothesis when it is true (type I error)

.

Table 36: Type I error for simulations tests

Note: This table reports various performance measures computed on 250 artificial investment strategies that yields a return similar to the benchmark, with a 0.5% margin in annual term. The notation is similar to Table 34.

	Interval lengths	
	8 years	14 years
Δ in returns		
Med p-value Simulation	0.194	0.0995
Nb sig simulation	10	13
Med p-value BLL	0.175	0.074
Nb sig BLL	15	31

We find that it is well specified as the median p-value is 19.4% (0.095%) for eight (14) years samples, and it generates only 10 (13) out of 250 p-values lower than 5% in this case. This is still lower than the 12.5 false rejections that we should obtain with this confidence level for the short sample, and the rejection rate is only slightly higher for the long samples. On the other hand, the lower p-values reported for the BLL test are more problematic. Indeed, this test rejects the null 15 (31) times. According to these results, this means that using the BLL test may provide statistical evidence that strategy outperform its benchmark while it is not the case. Thus, this also confirms that the BLL test should not be used as a primary performance measure, but only to assess whether excess returns found by Student t-test are due to the test hypotheses about returns series such as the normal distribution.

This Monte-Carlo experiment suggests that tests based on the returns distribution moments, such as the Student t-test of equal means and the Sharpe ratio, have a very low power. They need particularly high levels of abnormal returns to conclude in favour of a significant out-performance. As expected, the power diminishes for shorter samples. Standard tests performed on the alpha are only slightly more powerful. On the other hand, the two simulation testing procedures provide statistical results that are more in line with the economically large abnormal returns. In addition, our proposed simulation test provides consistent conclusions with the three performance measures, and we find that it is well specified.

7.2   Multiple-assets case

This Monte-Carlo experiment follows a similar procedure to those presented above for the single-asset case. One of the differences, as justified in the beginning of this Section, is that we use the Jensen's alpha, and not the difference in mean returns, as the performance measure to define the excess return of the investment strategy compared with its benchmark:

1. The same bootstrap procedure is used to create artificial buy-and-hold series. In this setting, one block consists in one of month data across all assets. Note that each return series is generated independently from the others assets.

2. In contrast with the single-asset case, we do not generate artificial strategies, i.e. portfolios, with the same optimization program presented in the second part of this thesis. Indeed, this would be extremely time consuming as a very large amount of portfolios are simulated in order to get the 250 strategies with excess returns corresponding to the various target. In order to generate these artificial portfolios, we use another block bootstrap procedure to resample the original portfolio weights from the PF_4 portfolio. These new weights series are then used in conjunction with the resampled buy-and-hold returns series to generate the artificial portfolios returns.

3. To ensure that the simulation comprises various market situations, only 100 artificial portfolios are created with the same buy-and-hold return series. Then, a new series is generated from the original data.

4. Then, the Monte-Carlo experiment does not differ anymore with the single-asset case. We compute the artificial portfolio excess return, and if it corresponds to one of the target, ranging from 2% to 20%, we compute the various performance measures. In addition to our proposed simulation test, the difference in mean returns with the associated Student t-test and the Share ratio, we also consider HM and TM test. As these tests divide the performance between selection and market-timing abilities, they are not suitable for the single-asset case where our strategies are pure market timing strategies.

This process is repeated until 1'000 observations for each time interval and abnormal return level are obtained. This represents 98'000 artificial scenarios[96].

Panel A of Figure 29 presents the median p-value obtained across the 1'000 artificial investment strategies for each level of abnormal return and time interval. The proposed simulation test is consistently more powerful than standard Student t-tests. The cross sections presented in Panel B show that the difference is more pronounced for relative low abnormal returns and shorter time intervals.

To have a more complete picture regarding the power of the two testing procedures, Table 37 contains the percentage of tests that result in a significant difference[97]. The results confirm the above-stated finding. Indeed, even with annual alpha as high as 20%, almost no Student t-test reject the null hypothesis of equal means when the samples are shorter than 10 years. In line with the single-asset case, the proposed simulation test, even if it is more powerful than Student t-tests, requires relatively large abnormal return to reject the null. For instance, it requires at least 14% abnormal returns to reject the null for 95% of scenarios with the 16 years samples. For these levels, none of the Student t-test rejects the null, and even with 20% abnormal returns.

---

[96] The number of trial is higher than for the single-asset case as the simulation of investment strategies is much faster in the multiple-assets setting.

[97] Results are only presented for abnormal levels with at least one statistically significant test.

Figure 29: Simulations and Student *t*-tests: median *p*-values

This figure presents the median p-value obtained respectively with the proposed simulation and standard bilateral Student *t*-tests. Abnormal returns are stated in term of annualized Jensen's alpha and Nb years represents the length, in years, of the series. For each level of abnormal return and each length, 1000 artificial series are tested. The simulation test method employs 1000 simulations to generate *p*-values.

Panel A: All abnormal return levels and time intervals
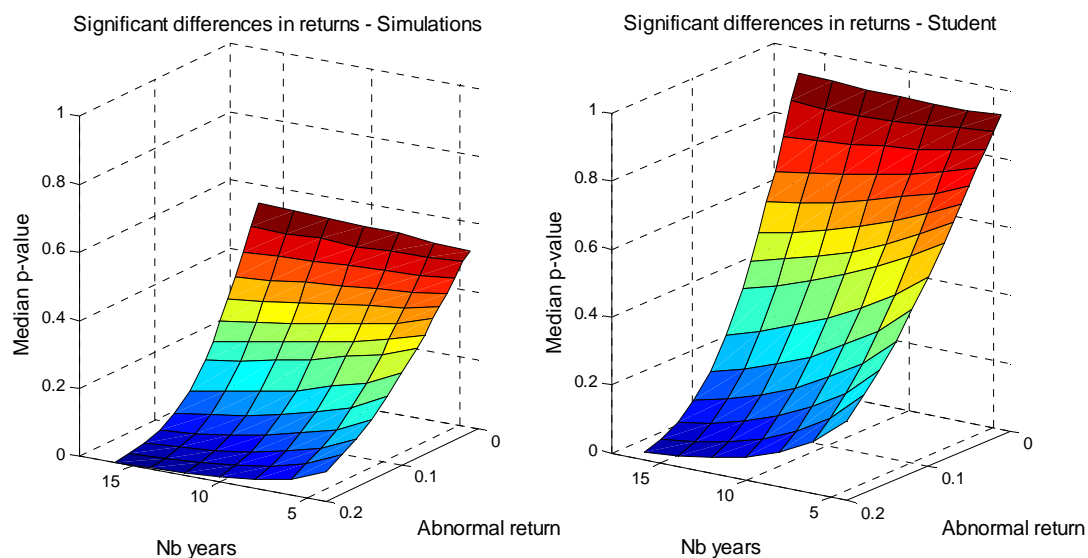
Panel B: Some cross sections
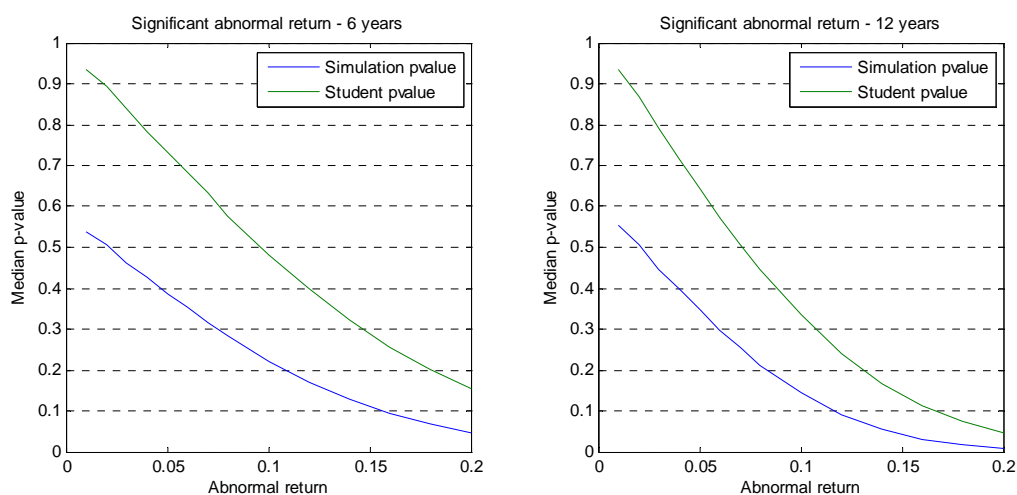
Table 37: Simulations and Student *t*-tests: Significant differences in returns

This table shows the percentage of simulations with a significant difference in returns, at a 5% confidence level, for the tests conducted either with the proposed simulation method or bilateral Student *t*-tests. *Ab ret* is the level of abnormal return stated in term of annualized Jensen's alpha.

Panel A: Simulation method

Ab ret	Series length in years						
	4	6	8	10	12	14	16
0.08	0.002	0.001	0	0	0	0	0
0.1	0.001	0	0.001	0	0.004	0.001	0.004
0.12	0.002	0.003	0.007	0.012	0.035	0.116	0.279
0.14	0.008	0.02	0.061	0.161	0.428	0.76	0.95
0.16	0.032	0.083	0.262	0.616	0.916	0.993	1
0.18	0.076	0.281	0.649	0.938	0.994	1	1
0.2	0.186	0.6	0.891	0.988	0.999	1	1

Panel B: Student test

Ab ret	Series length in years						
	4	6	8	10	12	14	16
0.14	0	0	0	0	0	0	0
0.16	0	0	0	0	0	0	0.023
0.18	0	0	0.001	0.006	0.07	0.378	0.915
0.2	0.003	0.003	0.015	0.157	0.609	0.975	1

The last part of this experiment focuses on the comparison between the proposed simulation method and Jensen's alphas, Sharpe ratios and the two HM and TM market timing tests. Table 38 shows that, on average, methods based on regressions are more powerful than Student *t*-test of equal means. First, testing the significance of Jensen's alpha with either the proposed simulation method or with Student *t*-tests produces similar results. This contrasts with the single-asset case. We also find that testing Jensen's alphas with the simulation method display a similar power to the tests about the difference in mean returns.

For the Sharpe ratios, the findings are similar to the ones presented above; when the statistical tests are performed with the simulation method, the power is comparable with the Jensen's alphas and the difference in means. On the contrarily, testing with Student *t*-tests produces only a few significant differences, especially for the shorter samples.

Turning to the HM and TM tests, this experiment illustrates the issue regarding the negative correlation between the selection and market timing parameters α and γ. Indeed, the correlation

coefficient is almost equal to minus one. A closer inspection of this table suggests that this issue is more pronounced for the HM test. Indeed, the majority of positive and significant γ are associated with negative α. Furthermore, approximately one out of 10 of these negative α coefficients are also significant. This means that the strategies have positive market timing and negative selection abilities. On the other hand, this situation almost never occurs with the TM test, as only six out of 7000 α coefficients are statistically negative. It is also worth noting that this negative correlation differs from the one found in empirical mutual fund studies, such as in Henriksson (1984) or Cumby and Glen (1990). Indeed, they find negative market timing coefficient, while we find that the negative correlation is mainly due to positive market timing and negative selection abilities.

The conclusions about the power of these tests are less homogenous than for the proposed simulation method or Student *t*-tests. Indeed, market timing abilities are detected with lower levels of abnormal returns, but for the highest levels of abnormal returns, results are less consistent with only 84.6% (68.9%) of significant market timing coefficients for the HM (TM) model. In addition, with abnormal returns of 2%, there are almost as many significant negative timing coefficients as positive ones. Table 39 reports these results for eight years samples. It is worth noting that the shorter time interval reduces the HM and TM tests power to a lesser extent than for Student *t*-tests or the proposed simulation method.

Finally, we also perform the simulation for a zero alpha target, within a 1% margin, to control for error type I. Our proposed simulation test generates median *p*-values of 0.58 and 0.61 for the eight and 16 years sample, and there is no value under the critical 5% level. Hence, we conclude that the proposed test methodology is also well specified for the multiple-assets case. This contrasts with the HM and TM tests that have respectively 23.2% and 44.2% of significant market timing coefficients, which are mostly negative.

Table 38: Simulation method and other performance measures – 16 years

This table reports various performance measures computed on 1000 artificial investment strategies with seven level of abnormal returns, stated in term of annual alpha. Only the longest time interval of 16 years is reported. *⊿ in returns* is the testing procedure between strategies and buy-and-hold mean returns. *Nb sig* is the number of significant measures out of the 1000 for each abnormal returns level. Simulation method and Student indicate how tests are performed. *⊿ SR* is the number of significant differences between strategies Sharpe ratios with test done either with the proposed simulation method or Student *t*-test adjusted according to Memmel (2003). Henriksson and Merton (1981) and Treynor and Mazuy (1966) coefficients are estimated, respectively, with the following models:

$$R_{S,t} - R_{rf,t} = \alpha + \beta(R_{BH,t} - R_{rf,t}) + \gamma \max(0, R_{BH,t} - R_{rf,t}) + \varepsilon$$

$$R_{S,t} - R_{rf,t} = \alpha + \beta(R_{BH,t} - R_{rf,t}) + \gamma(R_{BH,t} - R_{rf,t})^2 + \varepsilon$$

Corr(a,γ) is the median correlation between the two coefficients, *Nb sig a/γ + (-)* is the number of statistically significant positive (negative) coefficients and *Nb sig γ+ and sig a –* is the number of series with simultaneity statistically significant positive γ and significant negative α coefficients, with bilateral Student *t*-test at the 5% level.

	Abnormal returns						
	0.02	0.06	0.1	0.14	0.16	0.18	0.2
Δ in returns							
Simulation method	0	0	4	950	1000	1000	1000
Student	0	0	0	0	23	915	1000
Jensen's alpha							
Nb sig simulation method	0	0	0	1000	1000	1000	1000
Nb sig student	0	0	0	999	1000	1000	1000
Sharpe ratio							
Median SR strategy	0.23	0.37	0.51	0.65	0.72	0.80	0.87
Median SR BH	0.31	0.31	0.31	0.31	0.31	0.31	0.31
Δ SR - simulation method	0	0	0	969	1000	1000	1000
Δ SR - student	0	0	0	0	11	365	872
Henriksson and Merton (1981)							
corr (α,γ)	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
Nb sig γ+	139	265	444	575	724	774	846
Nb sig γ-	137	66	12	6	2	1	0
Nb sig α+	84	77	51	60	33	50	33
Nb sig α-	32	40	40	38	52	60	49
Nb sig γ+ and α-	139	265	444	483	528	509	578
Nb sig γ+ and sig α-	32	40	40	38	52	60	49
Treynor and Mazuy (1966)							
corr (α,γ)	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
Nb sig γ+	267	348	456	519	608	619	689
Nb sig γ-	205	156	78	67	26	25	17
Nb sig α+	19	63	133	340	374	528	604
Nb sig α-	4	2	0	0	0	0	0
Nb sig γ+ and α-	267	212	88	21	22	7	4
Nb sig γ+ and sig α-	4	2	0	0	0	0	0

### Table 39: Simulation method and other performance measures – 8 years

This table reports various performance measures computed on 1000 artificial investment strategies with seven level of abnormal returns, stated in term of annual alpha. Only the time interval of 8 years is reported. The notation is similar to Table 38.

	Abnormal returns						
	0.02	0.06	0.1	0.14	0.16	0.18	0.2
Δ in returns							
Simulation method	0	0	1	61	262	649	891
Student	0	0	0	0	0	1	15
Jensen's alpha							
Nb sig simulation method	0	0	0	29	249	708	947
Nb sig student	0	0	0	5	285	794	974
Sharpe ratio							
Median SR strategy	0.21	0.37	0.51	0.66	0.73	0.80	0.87
Median SR BH	0.29	0.31	0.32	0.33	0.33	0.30	0.31
Δ SR - simulation method	0	0	0	46	274	650	888
Δ SR - student	0	0	0	0	13	46	135
Henriksson and Merton (1981)							
corr (α,γ)	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00	-1.00
Nb sig γ+	125	196	271	426	504	558	647
Nb sig γ-	156	83	30	27	14	12	6
Nb sig α+	94	70	56	62	53	53	51
Nb sig α-	30	33	43	51	39	50	74
Nb sig γ+ and α-	125	196	271	424	488	488	533
Nb sig γ+ and sig α-	30	33	43	51	39	50	74
Treynor and Mazuy (1966)							
corr (α,γ)	-0.99	-0.99	-0.99	-0.99	-0.99	-0.99	-0.98
Nb sig γ+	249	302	329	431	497	481	550
Nb sig γ-	209	158	101	93	67	66	53
Nb sig α+	28	39	58	145	172	243	289
Nb sig α-	9	1	1	1	0	0	0
Nb sig γ+ and α-	249	261	155	110	74	54	69
Nb sig γ+ and sig α-	9	1	1	1	0	0	0

These results are in line with the single-asset case. Nonetheless, the proposed test is less powerful in the multiple-assets setting. Abnormal returns of 14% are required to reject the null most of the time, compared with 8% in the first case. This can be explained by the range of artificial strategies returns obtained in the simulation. Figure 30 displays 5'000 of these artificial returns for an original strategy that generate a 10% abnormal return over 16 years, both for the single and multiple-assets cases.

Figure 30: Histogram of artificial returns

The histograms display 5'000 artificial annual mean returns obtained by the simulation tests for two strategies with an abnormal return of 10%. In the single (multiple) asset case, the original annual mean return is 14.22% (19.84%).
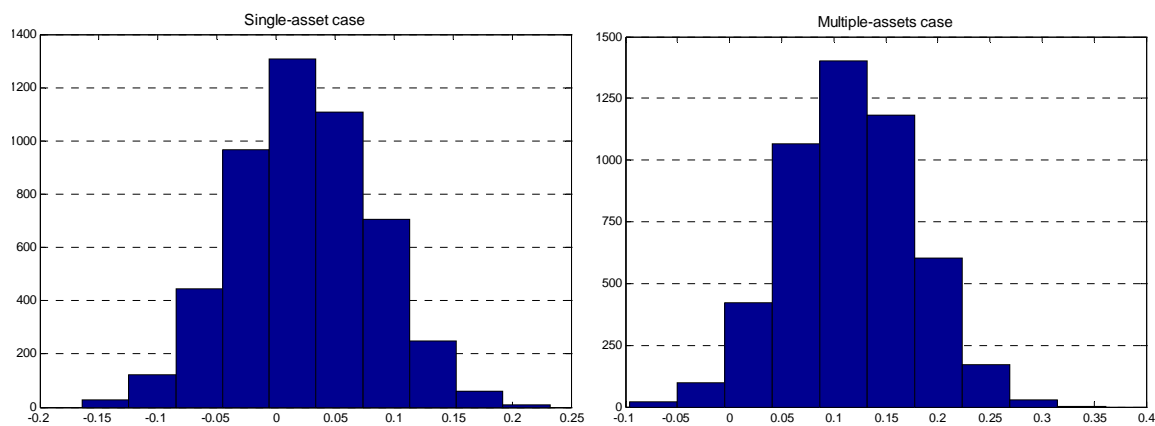
We find that the artificial returns distribution is wider for the multiple-asset case than the single-asset case. The differences between the minimum and the maximum artificial mean returns are respectively 39.6% and 45.7%. The standard deviation is also higher for the second case, i.e. 6.1% compared to 5.8% for the single-asset case. This indicates that if the artificial returns are more spread out, it is more likely to find a performance as high as the original. Thus, p-values are higher. In this illustration, the first (second) strategy mean return is associated with a p-value of 0.021 (0.095).

To conclude, this experiment shows that, even if our proposed simulation test is more powerful than standard procedures, large abnormal returns are required to conclude in favour of statistically significant abnormal returns. We think this in an important point to consider when conducting performance analyses.

8.   Conclusion

In this last part of this thesis, we develop a new performance test and apply it to both technical trading strategies and mean-variance optimized portfolios. The proposed test is based on the simulation of trading positions and is designed to keep the original positions structure. Specifically, we propose to use first-order Markov chains to simulate artificial trading signals when the strategy invests only in one asset and a block bootstrap otherwise. The test result, i.e. the p-values, is the percentage of simulated series that have a return, or any other performance measure, higher or equal to the original one. In other words, the test determines whether a strategy performance measure can be achieved with random trading positions, which have a similar structure to the original. Compared with other popular simulation-based tests, this procedure has the following advantages: First, its economical interpretation is straightforward. Second, it does not rely on any return-generating models, and the only assumption is how artificial trading signals are constructed. In addition, it only requires information on trading positions, and not the investment process according to which they are generated. Thus, we can use this test with strategies that can not be summarized as a specific mathematic formula. Finally, unlike the widely used BLL methodology, the proposed one tests directly for an abnormal performance.

Then, we apply this test to successful investment strategies and compare p-values associated with various performance measures with those resulting from standard Student t-tests. We find that Student t-tests fail to reject the null of equal performance, even if the strategies yield returns that are twice as high as the benchmark. On the other hand, the proposed test generates p-values lower than 5%, which indicates that these returns can not be replicated by taking random positions. The results are consistent for all performance measures considered, the difference in means, the Jensen's alpha and the Sharpe ratio.

Finally, a Monte-Carlo experiment is run to compare the power of various performance measures. It focuses on the level of abnormal return and the corresponding p-values. We find that the proposed simulation test is more powerful than other measures, except the Jensen's alpha when the strategy invests in more than one asset. However, annual abnormal returns ranging between 8% and 14% are still required to reject the null hypothesis of equal performances. This finding suggests that it is essential to consider the test power before to conclude that a strategy is not able to out-perform its benchmark.

Thesis conclusion

The main finding of this thesis that we would like to highlight is the importance of considering the testing procedure power before performing an investment strategy performance analysis. Indeed, we show that the usual Student t-test requires very large returns in excess of the benchmark in order to provide a statistically significant difference. Thus, we propose a new test based on the simulation of trading positions. It differentiates from the other procedures proposed in the literature as it is designed to keep the structure of the original position in the simulation process. We present two versions of this test, one based on the simulation of Markov chains for strategies that invest only in one asset and another one, which use a block bootstrap procedure, when the investment universe is not restraint to one asset. In a Monte-Carlo experiment, we show that the proposed test is much more powerful than commonly used procedures, but nonetheless, it still require large excess returns to reject the null hypothesis of equal performance.

We also extend the literature on investment strategies based on technical analysis and portfolio optimization. First, we show that using moving-average trading strategies in a long-term setting improve its performance dramatically. Indeed, these rules are usually constructed with a long MA that is no longer than 250 days. In this thesis, we analyse them with window lengths as long as four years of data. First, we show that these new long term based MA rules generate higher returns than the commonly used specifications. Then, we construct several complex rules to use information from more than one simple rule and to address the data-mining issue in an out of sample setting. They generate mean returns that ranges from an annual 10.7% to 14.61%, while the buy-and-hold returns is 6.16%. However, even if their excess returns are economically significant, they are not statistically different from the buy-and-hold according to Student t-tests. On the other hand, the proposed simulation test concludes that their performances can not be replicated by taking random positions, with a similar structure, in the market. Thus, the excess

returns are significant according to our proposed test. The difference in conclusions is striking. Indeed, with Student *t*-test, we would conclude that the strategies possess some forecasting abilities but not enough to generate significant excess returns. In addition, leveraging these strategies with debt generate high returns that are not due to the leverage itself, but to the predictive power of these rules. We also develop a new market timing test that provides evidence that they invest according to bull and bear market phases to an extent that can not be replicated by random positions. We also investigate whether transaction costs and risk can prevent them to challenge the efficient market hypothesis, but we find no evidence.

Finally, we provide two extensions concerning mean-variance optimized portfolios. First, we show that the length of the estimation window has an impact the optimization performance to a greater extent than using various parameters estimation models. In addition, we propose to use similar selection processes than those used in the technical analysis setting to select the optimization specification objectively. While the performance of these new strategies is higher than the corresponding benchmark, it is less likely that they may challenge the efficient market hypothesis. Indeed, transaction costs would diminish their returns noticeably and they require that the investor can to short some assets. Finally, we adapt the proposed simulation test to these portfolios. To the best of our knowledge, we are the first to examine their performance with a simulation method. The results are in line with those reported for technical analysis strategies; the *p*-values are consistently lower than those resulting from Student *t*-tests. For one specific portfolio, we also find that the latter would not reject the null hypothesis of equal performance, while the simulation test does.

Bibliography

Admati, Anat R., Sudipto Bhattacharya, Paul C. Pfleiderer, and Stephen A. Ross, 1986, On timing and selectivity, *Journal of Finance* 41, 715-730.

Agarwal, Vikas, and Narayan Y. Naik, 2004, Risks and portfolio decisions involving hedge funds, *The Review of Financial Studies* 17, 63-98.

Allen, Franklin, and Risto Karjalainen, 1999, Using genetic algorithms to find technical trading rules, *Journal of Financial Economics* 51, 245-271.

Allen, Helen, and Mark P. Taylor, 1990, Charts, noise and fundamentals in the london foreign exchange market, *The Economic Journal* 100, 49-59.

Amenc, Noël, Felix Goltz, and Abraham Lioui, 2011, Practitioner portfolio construction and performance measurement: Evidence from europe, *Financial Analysts Journal* 67, 39-50.

Aneja, Yash P., Ramesh Chandra, and Erdal Gunay, 1989, A portfolio approach to estimating the average correlation coefficient for the constant correlation model, *The Journal of Finance* 44, 1435-1438.

Ang, Andrew, and Geert Bekaert, 2007, Stock return predictability: Is it there?, *The Review of Financial Studies* 20, 651-707.

Ang, Andrew, Joseph Chen, and Yuhang Xing, 2006, Downside risk, *The Review of Financial Studies* 19, 1191-1239.

Bai, Zhidong, Huixia Liu, and Wing-Keung Wong, 2009, Enhancement of the applicability of markowitz's portfolio optimization by utilizing random matrix theory, *Mathematical Finance* 19, 639-667.

Barone Adesi, Giovanni, Patrick Gagliardini, and Giovanni Urga, 2004, Testing asset pricing models with coskewness, *The Journal of Business and Economic Statistics* 22, 474-485.

Barras, Laurent, Oliver Scaillet, and Russ Wermers, 2010, False discoveries in mutual fund performance: Measuring luck in estimated alphas, *The Journal of Finance* 65, 179-216.

Bauer, Rob, Kees Koedijk, and Rogér Otten, 2005, International evidence on ethical mutual fund performance and investment style, *Journal of Banking & Finance* 29, 1751-1767.

Bessembinder, Hendrik, and Kalok Chan, 1995, The profitability of technical trading rules in the asian stock markets, *Pacific-Basin Finance Journal* 3, 257-284.

Bessembinder, Hendrik, and Kalok Chan, 1998, Market efficiency and the returns to technical analysis, *Financial Management* 27, 5-17.

Best, Michael J., and Robert R. Grauer, 1991a, Sensitivity analysis for mean-variance portfolio problems, *Management Science* 37, 980-989.

Best, Michael J., and Robert R. Grauer, 1991b, On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results, *The Review of Financial Studies* 4, 315-342.

Blanchet-Scalliet, Christophette, Awa Diop, Rajna Gibson, Denis Talay, and Etienne Tanré, 2007, Technical analysis compared to mathematical models based methods under parameters mis-specification, *Journal of Banking & Finance* 31, 1351-1373.

Board, John L. G., and Charles M. S. Sutcliffe, 1994, Estimation methods in portfolio selection and the effectiveness of short sales restrictions: Uk evidence, *Management Science* 40, 516-534.

Bollen, Nicolas P. B., and Jeffrey A. Busse, 2001, On the timing ability of mutual fund managers, *The Journal of Finance* 56, 1075-1094.

Brabazon, Anthony, and Michael O'Neill, 2004, Evolving technical trading rules for spot foreign-exchange markets using grammatical evolution, *Computational Management Science* 1, 311-327.

Brock, William, Josef Lakonishok, and Blake LeBaron, 1992, Simple technical trading rules and the stochastic properties of stock returns, *The Journal of Finance* 47, 1731-1764.

Brown, Stephen J., William N. Goetzmann, and Alok Kumar, 1998, The dow theory: William peter hamilton's track record reconsidered, *The Journal of Finance* 53, 1311-1333.

Campbell, John Y., and Yasushi Hamao, 1992, Predictable stock returns in the united states and japan: A study of long-term capital market integration, *The Journal of Finance* 47, 43-69.

Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57-82.

Cesari, Riccardo, and David Cremonini, 2003, Benchmarking, portfolio insurance and technical analysis: A monte carlo comparison of dynamic strategies of asset allocation, *Journal of Economic Dynamics and Control* 27, 987-1011.

Cesari, Riccardo, and Fabio Panetta, 2002, The performance of italian equity funds, *Journal of Banking & Finance* 26, 99-126.

Chan, Louis K. C., Jason Karceski, and Josef Lakonishok, 1999, On portfolio optimization: Forecasting covariances and choosing the risk model, *The Review of Financial Studies* 12, 937-974.

Chang, Eric C., and Wilbur G. Lewellen, 1984, Market timing and mutual fund performance, *Journal of Business* 57, 73-96.

Chang, P. H. K., and C. L. Osler, 1999, Methodical madness: Technical analysis and the irrationality of exchange-rate forecasts, *The Economic Journal* 109, 636-661.

Cheng, Wai-Yan, Yan Leung Cheung, and Haynes H. M. Yung, 2003, Profitability of the crisma system: From world indices to the hong kong stock market, *Asia-Pacific Financial Markets* 10, 45-57.

Chopra, Vijay K, 1993, Improving optimization, *The Journal of Investing* 2, 51-59.

Chopra, Vijay K., Chris R. Hensel, and Andrew L. Turner, 1993, Massaging mean-variance inputs: Returns from alternative global investment strategies in the 1980s, *Management Science* 39, 845-855.

Chopra, Vijay K., and William T. Ziemba, 1993, The effect of errors in means, variances, and covariances on optimal portfolio choice, *The Journal of Portfolio Management* 19, 6-11.

Clyde, William C., and Carol L. Osler, 1997, Charting: Chaos theory in disguise?, *Journal of Futures Markets* 17, 489-514.

Coggin, T. Daniel, Frank J. Fabozzi, and Shafiqur Rahman, 1993, The investment performance of u.S. Equity pension fund managers: An empirical investigation, *The Journal of Finance* 48, 1039-1055.

Comer, George, Norris Larrymore, and Javier Rodriguez, 2009, Controlling for fixed-income exposure in portfolio evaluation: Evidence from hybrid mutual funds, *The Review of Financial Studies* 22, 481-507.

Constantinides, George M., Jens Carsten Jackwerth, and Stylianos Perrakis, 2009, Mispricing of s&p 500 index options, *The Review of Financial Studies* 22, 1247-1277.

Cooper, M., 1999, Filter rules based on price and volume in individual security overreaction, *The Review of Financial Studies* 12, 901-935.

Cowles, Alfred, 1933, Can stock market forecasters forecast?, *Econometrica* 1, 309-324.

Cumby, R. E., and D. M. Modest, 1987, Testing for market timing ability: A framework for forecast evaluation, *Journal of Financial Economics* 19, 169-189.

Cumby, Robert E., and Jack D. Glen, 1990, Evaluating the performance of international mutual funds, *The Journal of Finance* 45, 497-521.

Cumby, Robert E., and David M. Modest, 1987, Testing for market timing ability: A framework for forecast evaluation, *Journal of Financial Economics* 19, 169-189.

Cuthbertson, Keith, Dirk Nitzsche, and Niall O'Sullivan, 2010, The market timing ability of uk mutual funds, *Journal of Business Finance & Accounting* 37, 270-289.

Darolles, Serge, and Christian Gourieroux, 2010, Conditionally fitted sharpe performance with an application to hedge fund rating, *Journal of Banking & Finance* 34, 578-593.

Day, Theodore E., and Pingying Wang, 2002, Dividends, nonsynchronous prices, and the returns from trading the dow jones industrial average, *Journal of Empirical Finance* 9, 431-454.

DeMiguel, Victor, Lorenzo Garlappi, Francisco J. Nogales, and Raman Uppal, 2009, A generalized approach to portfolio optimization: Improving performance by constraining portfolio norms, *Management Science* 55, 798-812.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *Review of Financial Studies* 22, 1915-1953.

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal, 2009, Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy?, *The Review of Financial Studies* 22, 1915-1953.

Dewachter, H., 2001, Can markov switching models replicate chartist profits in the foreign exchange market?, *Journal of International Money and Finance* 20, 25-41.

Dewachter, Hans, and Marco Lyrio, 2005, The economic value of technical trading rules: A nonparametric utility-based approach, *International Journal of Finance & Economics* 10, 41-62.

Diebold, F. X., and R.S. Mariano, 1995, Comparing forecasting accuracy, *Journal of Business and Economic Statistics* 13, 253-263.

Ding, Bill, and Hany A. Shawky, 2007, The performance of hedge fund strategies and the asymmetry of return distributions, *European Financial Management* 13, 309-331.

Dowd, Kevin, 1999, A value at risk approach to risk-return analysis, *The Journal of Portfolio Management* 25, 60-67.

Dowd, Kevin, 2000, Adjusting for risk: An improved sharpe ratio, *International Review of Economics & Finance* 9, 209-222.

Dueker, Michael, and Christopher J. Neely, 2007, Can markov switching models predict excess foreign exchange returns?, *Journal of Banking & Finance* 31, 279-296.

Eling, Martin, 2008, Does the measure matter in the mutual fund industry?, *Financial Analysts Journal* 64, 54-66.

Eling, Martin, and Frank Schuhmacher, 2007, Does the choice of performance measure influence the evaluation of hedge funds?, *Journal of Banking & Finance* 31, 2632-2647.

Elton, Edwin J., and Martin J. Gruber, 1973, Estimating the dependence structure of share prices--implications for portfolio selection, *The Journal of Finance* 28, 1203-1232.

Elton, Edwin J., Martin J. Gruber, and Christopher R. Blake, 1996, Survivorship bias and mutual fund performance, *The Review of Financial Studies* 9, 1097-1120.

Engle, Robert F., and Kenneth F. Kroner, 1995, Multivariate simultaneous generalized arch, *Econometric Theory* 11, 122-150.

Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.

Fang, Yue, and Daming Xu, 2003, The predictability of asset returns: An approach combining technical analysis and time series forecasts, *International Journal of Forecasting* 19, 369-385.

Fernandez-Rodriguez, Fernando, Christian Gonzalez-Martel, and Simon Sosvilla-Rivero, 2000, On the profitability of technical trading rules based on artificial neural networks: Evidence from the madrid stock market, *Economics Letters* 69, 89-94.

Fernández-Rodríguez, Fernando, Christian González-Martel, and Simón Sosvilla-Rivero, 2000, On the profitability of technical trading rules based on artificial neural networks:: Evidence from the madrid stock market, *Economics Letters* 69, 89-94.

Ferson, Wayne E., and Campbell R. Harvey, 1993, The risk and predictability of international equity returns, *The Review of Financial Studies* 6, 527-566.

Ferson, Wayne E., and Rudi W. Schadt, 1996, Measuring fund strategy and performance in changing economic conditions, *The Journal of Finance* 51, 425-461.

Fifield, Suzanne, David Power, and Donald Sinclair, 2005, An analysis of trading strategies in eleven european stock markets, *The European Journal of Finance* 11, 531-548.

Fong, Wai Mun , and Yat Wai Ho, 2001, Simple trading rules and the market for internet stocks, *International Review of Finance* 2, 247-268.

Fong, Wai Mun, and Lawrence H. M. Yong, 2005, Chasing trends: Recursive moving average trading rules and internet stocks, *Journal of Empirical Finance* 12, 43-76.

Frankfurter, George M., Herbert E. Phillips, and John P. Seagle, 1971, Portfolio selection: The effects of uncertain means, variances, and covariances, *The Journal of Financial and Quantitative Analysis* 6, 1251-1262.

Fried, Joel, 1970, Forecasting and probability distributions for models of portfolio selection, *The Journal of Finance* 25, 539-554.

Frost, Peter A., and James E. Savarino, 1988, For better performance: Constrain portfolio weights, *The Journal of Portfolio Management* 15, 29-34.

Garlappi, Lorenzo, Raman Uppal, and Tan Wang, 2007, Portfolio selection with parameter and model uncertainty: A multi-prior approach, *The Review of Financial Studies* 20, 41-81.

Gehrig, Thomas, and Lukas Menkhoff, 2006, Extended evidence on the use of technical analysis in foreign exchange, *International Journal of Finance & Economics* 11, 327-338.

Gençay, Ramazan, 1996, Non-linear prediction of security returns with moving average rules, *Journal of Forecasting* 15, 165-174.

Gençay, Ramazan, 1998, Optimization of technical trading strategies and the profitability in security markets, *Economics Letters* 59, 249-254.

Gençay, Ramazan, 1999, Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules, *Journal of International Economics* 47, 91-107.

Gençay, Ramazan, Michel Dacorogna, Richard Olsen, and Olivier Pictet, 2003, Foreign exchange trading models and market behavior, *Journal of Economic Dynamics and Control* 27, 909-935.

Goetzmann, William, Jonathan Ingersoll, Matthew Spiegel, and Ivo Welch, 2007, Portfolio performance manipulation and manipulation-proof performance measures, *The Review of Financial Studies* 20, 1503-1546.

Golosnoy, Vasyl, and Yarema Okhrin, 2007, Multivariate shrinkage for optimal portfolio weights, *The European Journal of Finance* 13, 441-458.

Goodacre, A., J. Bosher, and A. Dove, 1999, Testing the crisma trading system: Evidence from the uk market, *Applied Financial Economics* 9, 455-468.

Goodacre, Alan, and Tessa Kohn-Speyer, 2001, Crisma revisted, *Applied Financial Economics* 11, 221-230

Grauer, Robert R., and Nils H. Hakansson, 1995, Stein and capm estimators of the means in asset allocation, *International Review of Financial Analysis* 4, 35-66.

Grauer, Robert R., and Frederick C. Shen, 2000, Do constraints improve portfolio performance?, *Journal of Banking & Finance* 24, 1253-1274.

Green, Richard C., and Hollifield Burton, 1992, When will mean-variance efficient portfolios be well diversified?, *The Journal of Finance* 47, 1785-1809.

Grinblatt, Mark, and Sheeridan Titman, 1989b, Portfolio performance evaluation: Old issues and new insights, *The Review of Financial Studies* 2, 393-421.

Grinblatt, Mark, and Sheridan Titman, 1989a, Mutual fund performance: An analysis of quarterly portfolio holdings, *The Journal of Business* 62, 393-416.

Hansen, Peter Reinhard, 2005, A test for superior predictive ability, *Journal of Business and Economic Statistics* 23, 365-380.

Harvey, Campbell R., 1995, Predictable risk and returns in emerging markets, *The Review of Financial Studies* 8, 773-816.

Harvey, Campbell R., and Akhtar Siddique, 2000, Conditional skewness in asset pricing tests, *The Journal of Finance* 55, 1263-1295.

Henriksson, Roy D., 1984, Market timing and mutual fund performance: An empirical investigation., *Journal of Business* 57, 73-96.

Henriksson, Roy D., and Robert C. Merton, 1981, On market timing and investment performance. Ii. Statistical procedures for evaluating forecasting skills, *The Journal of Business* 54, 513-533.

Hlouskova, Jaroslava, Kurt Schmidheiny, and Martin Wagner, 2009, Multistep predictions for multivariate garch models: Closed form solution and the value for portfolio management, *Journal of Empirical Finance* 16, 330-336.

Hodges, S., 1998, A generalization of the sharpe ratio and its applications to valuation bounds and risk measures, *Financial Options Research Centre, University of Warwick.*

Hsu, Po-Hsuan, Yu-Chin Hsu, and Chung-Ming Kuan, 2010, Testing the predictive ability of technical analysis using a new stepwise test without data snooping bias, *Journal of Empirical Finance* 17, 471-484.

Hsu, Po-Hsuan, and Chung-Ming Kuan, 2005, Reexamining the profitability of technical analysis with data snooping checks, *Journal of Financial Econometrics* 3, 606-628.

Hudson, Robert, Michael Dempsey, and Kevin Keasey, 1996, A note on the weak form efficiency of capital markets: The application of simple technical trading rules to uk stock prices - 1935 to 1994, *Journal of Banking & Finance* 20, 1121-1132.

Isakov, D., and M. Hollistein, 1999, Application of simple technical trading rules to swiss stock prices: Is it profitable?, *Finanzmarkt und Portfolio Management* 1, 9-26.

Isakov, Dusan, and Marc Hollistein, 1999, Application of simple technical trading rules to swiss stock prices: Is it profitable?, *Finanzmarkt und Portfolio Management* 13, 9-26.

Jagannathan, Ravi, and Robert A. Korajczyk, 1986, Assessing the market training performance of managed portfolios, *The Journal of Business* 59, 217-235.

Jagannathan, Ravi, and Tongshu Ma, 2003, Risk reduction in large portfolios: Why imposing the wrong constraints helps, *The Journal of Finance* 58, 1651-1684.

James, W., and Charles Stein, 1961, Estimation with quadratic loss, Proceedings of the fourth Berkeley symposium on probability and statistics I (University of California Press, Berkley).

Jensen, Michael C., 1968, The performance of mutual funds in the period 1945-1964, *The Journal of Finance* 23, 389-416.

Jiang, George J., Tong Yao, and Tong Yu, 2007, Do mutual funds time the market? Evidence from portfolio holdings, *Journal of Financial Economics* 86, 724-758.

Jiang, Wei, 2003, A nonparametric test of market timing, *Journal of Empirical Finance* 10, 399-425.

Jobson, J. D., and Bob Korkie, 1981, Putting markowitz theory to work, *The Journal of Portfolio Management* 7, 70-74.

Jobson, J. D., and Bob M. Korkie, 1981, Performance hypothesis testing with the sharpe and treynor measures, *The Journal of Finance* 36, 889-908.

Jobson, J.D., and Robert. M. Korkie, 1981, Putting markowitz theory to work, *The Journal of Portfolio Management* 7, 70-74.

Jorion, Philippe, 1985, International portfolio diversification with estimation risk, *The Journal of Business* 58, 259-278.

Jorion, Philippe, 1986, Bayes-stein estimation for portfolio analysis, *The Journal of Financial and Quantitative Analysis* 21, 279-292.

Jorion, Philippe, 1991, Bayesian and capm estimators of the means: Implications for portfolio selection, *Journal of Banking & Finance* 15, 717-727.

Kallberg, Jarl G., and William T. Ziemba, 1983, Comparison of alternative utility functions in portfolio selection problems, *Management Science* 29, 1257-1276.

Kaplan, Paul, D., and James Knowles, A. , 2004, A generalized downside risk-adjusted performance measure, *Morningstar Associates and York Hedge Fund Strategies* January 2004.

Kavajecz, Kenneth A., and Elizabeth R. Odders-White, 2004, Technical analysis and liquidity provision, *The Review of Financial Studies* 17, 1043-1071.

Keating, Con , and William Shadwick, 2002, A universal performance measure, *Journal of Performance measure* 6, 59-84.

Keim, Donald B., and Robert F. Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357-390.

Kidd, Willis V., and B. Wade Brorsen, 2004, Why have the returns to technical analysis decreased?, *Journal of Economics and Business* 56, 159-176.

Kosowski, Robert, Allan Timmermann, Russ Wermers, and Hal White, 2006, Can mutual fund "stars" really pick stocks? New evidence from a bootstrap analysis, *The Journal of Finance* 61, 2551-2595.

Kraus, Alan, and Robert H. Litzenberger, 1976, Skewness preference and the valuation of risk assets, *The Journal of Finance* 31, 1085-1100.

Kritzman, Mark, Sébastien Page, and David Turkington, 2010, In defense of optimization: The fallacy of 1/n, *Financial Analysts Journal* 66, 31-39.

Kwan, Josephine W. C., K. Lam, Mike K. P. So, and Philip L. H. Yu, 2000, Forecasting and trading strategies based on a price trend model, *Journal of Forecasting* 19, 485-498.

Laloux, Laurent, Pierre Cizeau, Marc Potters, and Jean-Philippe Bouchaud, 2000, Random matrix theory and financial correlations, *International Journal of Theoretical and Applied Finance* 3, 391-397.

LeBaron, Blake 2000, The stability of moving average technical trading rules on the dow jones index *Derivatives Use, Trading and Regulation* 5, 324-338.

Ledoit, Oliver, and Michael Wolf, 2008, Robust performance hypothesis testing with the sharpe ratio, *Journal of Empirical Finance* 15, 850-859.

Ledoit, Olivier, and Michael Wolf, 2003, Improved estimation of the covariance matrix of stock returns with an application to portfolio selection, *Journal of Empirical Finance* 10, 603-621.

Ledoit, Olivier, and Michael N. Wolf, 2004, Honey, i shrunk the sample covariance matrix, *The Journal of Portfolio Management* 30, 110-119.

Lee, Chun I., and Ike Mathur, 1996, Trading rule profits in european currency spot cross-rates, *Journal of Banking & Finance* 20, 949-962.

Lehmann, Bruce N., and David M. Modest, 1987, Mutual fund performance evaluation: A comparison of benchmarks and benchmark comparisons, *The Journal of Finance* 42, 233-265.

Leland, Hayne E., 1999, Beyond mean-variance: Performance measurement in a nonsymmetrical world, *Financial Analysts Journal* 55, 27-36.

Levich, Richard M., and Lee R. Thomas, 1993, The significance of technical trading-rule profits in the foreign exchange market: A bootstrap approach, *Journal of International Money and Finance* 12, 451-474.

Lintner, John, 1965, The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *The Review of Economics and Statistics* 47, 13-37.

Liu, Lan, and Hao Lin, 2010, Covariance estimation: Do new methods outperform old ones?, *Journal of Economics and Finance* 34, 187-195.

Lo, A. W., and A. C. MacKinlay, 1990, Data-snooping biases in tests of financial asset pricing models, *The Review of Financial Studies* 3, 431-467.

Lui, Yu-Hon, and David Mole, 1998, The use of fundamental and technical analyses by foreign exchange dealers: Hong kong evidence, *Journal of International Money and Finance* 17, 535-545.

Maillet, Bertrand, and Thierry Michel, 2000, Further insights on the puzzle of technical analysis profitability, *The European Journal of Finance* 6, 196 - 224.

Malkiel, Burton G. , 1995, Returns from investing in equity mutual funds 1971 to 1991, *The Journal of Finance* 50, 549-572.

Mao, James C. T., and Erik Särndal, 1966, A decision theory approach to portfolio selection, *Management Science* 12, 323-333.

Markowitz, Harry, 1952, Portfolio selection, *The Journal of Finance* 7, 77-91.

Markowitz, Henry M., 1952, Portfolio selection, *The Journal of Finance* 7, 77-91.

Marshall, Ben R., Rochester H. Cahan, and Jared M. Cahan, 2008, Does intraday technical analysis in the u.S. Equity market have value?, *Journal of Empirical Finance* 15, 199-210.

Memmel, Christoph, 2003, Performance hypothesis testing with the sharpe ratio, *Finance Letters* 1, 21-23.

Michaud, Richard O., 1989, The markowitz optimization enigma: Is 'optimized' optimal?, *Financial Analysts Journal* 45, 31-42.

Mills, Terence C., 1997, Technical analysis and the london stock exchange: Testing trading rules using the ft30, *International Journal of Finance & Economics* 2, 319-331.

Moreno, David, and Rosa Rodríguez, 2009, The value of coskewness in mutual fund performance evaluation, *Journal of Banking & Finance* 33, 1664-1676.

Nam, Kiseok, Kenneth M. Washer, and Quentin C. Chu, 2005, Asymmetric return dynamics and technical trading strategies, *Journal of Banking & Finance* 29, 391-418.

Neely, C. J., and P. A. Weller, 2003, Intraday technical trading in the foreign exchange market, *Journal of International Money and Finance* 22, 223-237.

Neely, Christopher J., 2003, Risk-adjusted, ex ante, optimal technical trading rules in equity markets, *International Review of Economics & Finance* 12, 69-87.

Neely, Christopher, Paul Weller, and Rob Dittmar, 1997, Is technical analysis in the foreign exchange market profitable? A genetic programming approach, *The Journal of Financial and Quantitative Analysis* 32, 405-426.

Neftci, Salih N., 1991, Naive trading rules in financial markets and wiener-kolmogorov prediction theory: A study of "technical analysis", *The Journal of Business* 64, 549-571.

Neuhierl, Andreas, and Bernd Schlusche, 2011, Data snooping and market-timing rule performance, *The Journal of Financial Econometrics* forthcoming.

Oberlechner, Thomas, 2001, Importance of technical and fundamental analysis in the european foreign exchange market, *International Journal of Finance & Economics* 6, 81-93.

Olson, Dennis, 2004, Have trading rule profits in the currency markets declined over time?, *Journal of Banking & Finance* 28, 85-105.

Osler, Carol L., 2003, Currency orders and exchange rate dynamics: An explanation for the predictive success of technical analysis, *The Journal of Finance* 58, 1791-1820.

Otten, Roger, and Dennis Bams, 2002, European mutual fund performance, *European Financial Management* 8, 75-101.

Pafka, Szilárd, and Imre Kondor, 2004, Estimated correlation matrices and portfolio optimization, *Physica A: Statistical Mechanics and its Applications* 343, 623-634.

Pagan, Adrian R., and Kirill A. Sossounov, 2003, A simple framework for analysing bull and bear markets, *Journal of Applied Econometrics* 18, 23-46.

Pesaran, M. H., and A. Timmermann, 1992, A simple nonparametric test of predictive performance, *Journal of Business and Economic Statistics* 10, 461-465.

Pesaran, M. Hashem, and Allan Timmermann, 1992, A simple nonparametric test of predictive performance, *Journal of Business and Economic Statistics* 10, 461-465.

Pojarliev, Momtchil, and Wolfgang Polasek, 2001, Applying multivariate time series forecasts for active portfolio management, *Financial Markets and Portfolio Management* 15, 201-211.

Pojarliev, Momtchil, and Wolfgang Polasek, 2003, Portfolio construction by volatility forecasts: Does the covariance structure matter?, *Financial Markets and Portfolio Management* 17, 103-116.

Poon, Ser-Huang, and Clive W. J. Granger, 2003, Forecasting volatility in financial markets: A review, *Journal of Economic Literature* 41, 478-539.

Potvin, Jean-Yves, Patrick Soriano, and Maxime Vallée, 2004, Generating trading rules on the stock markets with genetic programming, *Computers & Operations Research* 31, 1033-1047.

Pruitt, Stephen , and Richard White, 1988, The crisma trading system: Who says technical analysis can't beat the market?, *The Journal of Portfolio Management* 14, 55-58.

Pruitt, Stephen, Maurice Tse, and Richard White, 1992, The crisma trading system: The next five years, *The Journal of Portfolio Management* 18, 22-25.

Pruitt, Stephen, and Richard White, 1989, Exchange-traded options and crisma trading, *The Journal of Portfolio Management* 15, 55-56.

Qi, Min., and Yangru. Wu, 2006, Technical trading-rule profitability, data snooping, and reality check: Evidence from the foreign exchange market, *Journal of Money, Credit, and Banking* 38, 2135-2158.

Ranaldo, Angelo, and Laurent Favre, 2005, Hedge fund performance and higher-moment market models, *The Journal of Alternative Investments* Winter 2005, 37-51.

Ratner, Mitchell, and Ricardo P. C. Leal, 1999, Tests of technical trading strategies in the emerging equity markets of latin america and asia, *Journal of Banking & Finance* 23, 1887-1905.

Reitz, Stefan, 2006, On the predictive content of technical analysis, *The North American Journal of Economics and Finance* 17, 121-137.

Romano, Joseph P., and Michael Wolf, 2005, Stepwise multiple testing as formalized data snooping, *Econometrica* 73, 1237-1282.

Rosenow, B., V. Plerou, P. Gopikrishnan, and H.E. Stanley, 2002, Portfolio optimization and the random magnet problem, *Europhysics Letters* 59, 500-509.

Sharpe, William F., 1963, A simplified model for portfolio analysis, *Management Science* 9, 277-293.

Sharpe, William F., 1964, Capital asset prices: A theory of market equilibrium under conditions of risk, *The Journal of Finance* 19, 425-442.

Skouras, Spyros, 2001, Financial returns and efficiency as seen by an artificial technical analyst, *Journal of Economic Dynamics and Control* 25, 213-244.

Smith, Daniel R., 2007, Conditional coskewness and asset pricing, *Journal of Empirical Finance* 14, 91-119.

Solnik, Bruno, 1993, The performance of international asset allocation strategies using conditioning information, *Journal of Empirical Finance* 1, 33-55.

Sortino, Frank A., and Robert Van Der Meer, 1991, Downside risk, *The Journal of Portfolio Management* 17, 27-31.

Specht, Katja, and Wolfgang Gohout, 2003, Portfolio selection using the principal components garch model, *Financial Markets and Portfolio Management* 17, 450-458.

Stein, Charles, 1955, Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, Proceedings of the 3rd Berkeley Symposium on Probability and Statistics I (University of California Press, Berkeley,).

Sullivan, Ryan, Allan Timmermann, and Halbert White, 1999, Data-snooping, technical trading rule performance, and the bootstrap, *The Journal of Finance* 54, 1647-1691.

Summers, Barbara, Evan Griffiths, and Robert Hudson, 2004, Back to the future: An empirical investigation into the validity of stock index models over time, *Applied Financial Economics* 14, 209-214.

Szakmary, Andrew C., Qian Shen, and Subhash C. Sharma, 2010, Trend-following trading strategies in commodity futures: A re-examination, *Journal of Banking & Finance* 34, 409-426.

Taylor, S. J., 2000, Stock index and price dynamics in the u.K. And the u.S.: New evidence from a trading rule and statistical analysis, *The European Journal of Finance* 6, 39-69.

Taylor, Stephen J., 1980, Conjectured models for trends in financial prices, tests and forecasts, *Journal of the Royal Statistical Society. Series A (General)* 143, 338-362.

Tian, Gary Gang, Guang Hua Wan, and Mingyuan Guo, 2002, Market efficiency and the returns to simple technical trading rules: New evidence from U.S. Equity market and chinese equity markets, *Asia-Pacific Financial Markets* 9, 241-258.

Tola, Vincenzo, Fabrizio Lillo, Mauro Gallegati, and Rosario N. Mantegna, 2008, Cluster analysis for portfolio optimization, *Journal of Economic Dynamics and Control* 32, 235-258.

Treynor, Jack L., and Kay K. Mazuy, 1966, Can mutual funds outguess the market?, *Harvard Business Review* 44, 131-136.

Wermers, Russ, 2000, Mutual fund performance: An empirical decomposition into stock-picking talent, style, transactions costs, and expenses, *The Journal of Finance* 55, 1655-1703.

Wermers, Russ, 2006, Performance evaluation with portfolio holdings information, *The North American Journal of Economics and Finance* 17, 207-230.

White, Halbert, 2000, A reality check for data snooping, *Econometrica* 68, 1097-1126.

Wilkens, Sascha, 2007, Option returns versus asset-pricing theory: Evidence from the european option market, *J Deriv Hedge Funds* 13, 170-176.

Wong, Wing-Keung , Meher Manzur, and Boon-Kiat Chew, 2003, How rewarding is technical analysis? Evidence from singapore stock market, *Applied Financial Economics* 13, 543-551.

Xu, Yexiao, 2004, Small levels of predictability and large economic gains, *Journal of Empirical Finance* 11, 247-275.

Zakamouline, Valeri, and Steen Koekebakker, 2009, Portfolio performance evaluation with generalized sharpe ratios: Beyond the mean and variance, *Journal of Banking & Finance* 33, 1242-1254.

Zhu, Yingzi, and Guofu Zhou, 2009, Technical analysis: An asset allocation perspective on the use of moving averages, *Journal of Financial Economics* 92, 519-544.