

On the Relevance of Optimizing Technical Indicators on U.S. Stock Markets

John BERGERAT

Master of Sciences in Finance - 2020
Quantitative Asset and Risk Management
University of Lausanne

Abstract:

This study attempts to assess the propensity of an optimization process to overfit when it is applied to technical indicators on the U.S. stock market. The PBO of Bailey et al. (2015) shows that six out of ten and six out of seven indicators applied to the S&P 500 and the Nasdaq 100, respectively, do not lead to backtest overfitting at the 5% significance level. Moreover, first order stochastic dominance evaluated through a CSCV procedure shows that optimization on technical indicators brings an added value in regard to a simple random parametrization. Additionally, two out of six strategies controlled for overfitting for both the S&P 500 and the Nasdaq 100 show statistically and economically significant out-of-sample predictive power for the daily equity risk premium, according to the MSPE-adjusted test of Clark and West (2006). Finally, three out of six strategies applied to the S&P 500 and controlled for overfitting show statistically and economically significant out-of-sample daily abnormal returns, after controlling for multiple-testing in a model adjusted to correlations, according to the FDR at 5% of Harvey et al. (2015) with a baseline of $r = 1/2$ and a correlation of $\rho = 0.2$, and this, without taking into account transaction costs. The control of backtest overfitting and the control of a multiple-testing approach adjusted to correlations allow to assert that these discoveries are statistically and significantly not of type I error. However, the level of transaction costs contemporaneous of the data would not have permit to benefit from the alphas, and the current emergence of brokers that propose zero transaction costs reinforces the risk for structural signals to be absorbed in the future.

Keywords:

Technical Indicators, Technical Analysis, Market Efficiency, Stock Market Forecasting, Backtest, Optimization, Overfitting

JEL Classification: C61, G11, G17

Résumé:

Cette étude tente d'évaluer la propension d'un processus d'optimisation à suroptimiser lorsqu'il est appliqué à des indicateurs techniques sur le marché boursier américain. Le PBO de Bailey et al. (2015) établit que six indicateurs sur dix et six indicateurs sur sept appliqués respectivement au S&P 500 et au Nasdaq 100 ne conduisent pas à une suroptimisation du backtest au seuil de significativité de 5%. De plus, la dominance stochastique de premier ordre évaluée par une procédure CSCV montre que l'optimisation d'indicateurs techniques apporte une valeur ajoutée par rapport à une simple paramétrisation aléatoire. En outre, deux des six stratégies contrôlées de la suroptimisation pour le S&P 500 et le Nasdaq 100 montrent un pouvoir prédictif hors échantillon statistiquement et économiquement significatif pour la prime de risque journalière sur actions, selon le test ajusté MSPE de Clark et West (2006). Enfin, trois des six stratégies appliquées au S&P 500 et contrôlées de la suroptimisation présentent des rendements anormaux journaliers hors échantillon statistiquement et économiquement significatifs, après contrôle de tests multiples dans un modèle ajusté aux corrélations, ceci selon le FDR à 5 % de Harvey et al. (2015) avec une base de référence de $r = 1/2$ et une corrélation de $\rho = 0.2$, et cela, sans prendre en compte les coûts de transaction. Le contrôle de la suroptimisation du backtest et le contrôle d'une approche de tests multiples ajustée aux corrélations permettent d'affirmer que ces découvertes ne sont statistiquement et significativement pas des erreurs de type I. Cependant, le niveau des coûts de transaction de l'époque n'auraient pas permis de bénéficier des alphas et l'émergence actuelle de courtiers proposant des coûts de transaction nuls renforce le risque que les signaux structurels soient absorbés à l'avenir.

Mots-Clés:

Indicateurs Techniques, Analyse Technique, Efficience de Marché, Prédicibilité du Marché actions, Backtest, Optimisation, Suroptimisation

Classification JEL: C61, G11, G17

TABLE OF CONTENTS

Introduction	5
1 Outperforming the Market with Technical Analysis	6
1.1 Outperforming the Market in Classical Finance	6
1.1.1 Active vs Passive Investing	6
1.1.2 The ‘Random Walk Theory’	6
1.1.3 The ‘Efficient Market Hypothesis’	7
1.1.4 Behavioral Finance	7
1.2 The Need for Identifying Robust Predictors	8
1.2.1 Fundamental Analysis	8
1.2.2 Technical Analysis	9
1.3 The Forecasting Power of Technical Analysis	10
1.4 A Practical Challenge : Find Technical Strategies that Outperform Standard Market Indices	11
1.4.1 The Process of Generating Strategies with Technical Indicators.....	11
1.4.2 The Risk of False Positive	11
1.4.3 The Risk of Backtest Overfitting	13
2 Review of the Literature	14
2.1 Profitability of Technical Analysis : A Review - Cheol-Ho Park and Scott H. Irwin, 2004	14
2.1.1 Introduction.....	14
2.1.2 Early Empirical Studies : from 1960 to 1987.....	15
2.1.3 Modern Empirical Studies : from 1988 to 2004	16
2.1.4 Summary on Empirical Studies.....	19
2.2 The Probability of Backtest Overfitting – Bailey, Borwein, De Prado, Zhu - 2013	20
3 Generate and Evaluate Trading Strategies : Methodology	25
3.1 Review of the Main Technical Indicators	25
3.1.1 Technical Indicators : A Presentation	25
3.1.2 Bears Power	27
3.1.3 Bulls Power.....	27
3.1.4 Commodity Channel Index	27
3.1.5 DeMarker	28
3.1.6 Moving Average Convergence/Divergence adapted to Price	28
3.1.7 Momentum	28
3.1.8 Percent Bandwidth.....	29

3.1.9	Relative Strength Index	29
3.1.10	Relative Vigor Index	30
3.1.11	Stochastic	30
3.2	Generating Strategies : Methodology	31
3.2.1	Full Market Exposure	31
3.3	Evaluating Strategies : Methodology	32
3.3.1	Data	32
3.3.2	Hold-Out Method and Out-Of-Sample Validation	33
3.3.3	Probability of Backtest Overfitting : a CSCV Approach.....	33
3.3.4	Equity Risk Premium Predictability of Fitted Technical Indicators	34
3.3.5	Market Outperformance and Significance on OOS Data	36
4	Full Market Exposure : Analysis on S&P 500 and on Nasdaq 100	38
4.1	Optimization Results: Fitted Parameters and Generated Strategies	38
4.1.1	Generated Strategies	38
4.1.2	A view on fitted Technical Indicators	39
4.2	Hold-Out Analysis: IS and OOS Performance Results	42
4.2.1	Performance Measures: Analysis	42
4.2.2	A view on Strategies' Performance	45
4.3	Probability of Backtest Overfitting.....	45
4.3.1	Probability of Backtest Overfitting on Optimization Procedures: Analysis	45
4.3.2	Performance Degradation: Analysis	52
4.3.2	Stochastic Dominance: Analysis	52
4.4	Equity Risk Premium Predictability of Fitted Strategies : Analysis	57
4.5	Statistical Significance of Market Outperformance	59
5	Practical Implementation and Risks	60
	Conclusion	61
	References	64

INTRODUCTION

This study is primarily motivated by the will to bring a quantitative approach to what would be widely considered by Technical Analysis as a field of visual and graphical interpretation. Usually intensively used by practitioners such as traders and desk traders, the purpose would be to assess the predictive power of technical indicators through long-term quantitative strategies, in a given market and at a given timeframe.

To detect and assess potential predictability, an optimization procedure seems to be the most inclined to find positive results. However, applying optimization also raises the concerns of backtest overfitting and the discovery of false positives. This problem must be tackled in order to provide legitimacy for the use of such procedure. This will be treated with the Probability of Backtest Overfitting of Bailey et al. (2015). Also, the first order stochastic dominance in their context will state if an optimization brings a real added value, with regards to a random or a standard parametrization of technical indicators.

Additionally, it is the U.S. stock market that will be retained thanks to its propensity to represent a growing market. This will allow a technical strategy to follow the growing path of the market, where a long-term investment can be envisaged. A strategy applied to a security that has no priority direction would make less sense in the long run (commodities, foreign exchange markets).

The predictability will be assessed either through levels of optimized technical indicators that will be considered as predictors, either through the ability of fitted technical strategies to generate abnormal returns or alphas. The weak stage of market efficiency will be tested without taking into account transaction costs that are contemporaneous to the period of the data, since only prices are involved in the EMH. The question to *ex ante* be able to profit or not from potential alphas when including contemporaneous transaction costs belongs to another perspective. However, it is unclear if potential alphas owe their existence to the amount of transaction costs; and if current zero commission rates proposed since 2019 may nullify observed abnormal returns in the future, by increasing the efficiency of the market.

The problematic can be then decomposed into two parts such as : (1) to know if it exists a predictability from technical indicators that rejects the weak form of the EMH and (2) if this predictability could have been exploited in practice.

Concerning the conduct of the survey, this latter will examine in a first part a majority of theoretical models developed over time that characterize the stock markets. These models may reject or accept the fact that Technical Analysis can develop a forecasting power. A comparison between Technical and Fundamental Analysis will then be carried out, in order to assess their differences and the degrees of market efficiency they contradict. It will then be discussed the practical aspects of generating strategies with technical indicators and the issues that such procedures can lead to. Finally, the first part will gather all the ingredients one needs to take into account when searching for technical strategies that want to outperform a market's benchmark. The second part will be a literature review that first assesses Park and Irwin (2004) that is a general overview of how academic articles deal with Technical Analysis and their profitability. A second part will include Bailey et al. (2015), which addresses the probability of backtest overfitting in order to tackle the overfitting bias enounced previously. The third part will bring together all the methodology applied to develop strategies and how to evaluate them and the fourth part will sum up all the results obtained. Finally, the fifth part will attempt to have a better view on what a practical implementation would look like and what risks are to take into account when trying to follow investment strategies that rely on technical analysis.

1 OUTPERFORMING THE MARKET WITH TECHNICAL ANALYSIS

1.1 Outperforming the Market in Classical Finance

1.1.1 Active vs Passive Investing

Since the dawn of financial markets, scholars and practitioners have relentlessly searched for alternatives to the simple yet efficient ‘buy-and-hold the market’ strategy. For all the research, a fundamental question remains with regard to outperforming the market: Is it conceivable to consistently outperform the market in the long run? Much recent research on the performance of institutional investors raises the concern regarding the utility of the efforts placed on elaborating complex and structured strategies. Many of them have been found eventually to underperform when tested on the market, even when taking implementation costs and transaction costs into account. Thus, the eternal debate between the proponents of passive and active investing continues.

According to the S&P Dow Jones Indices LLC, the U.S. SPIVA (S&P Indices Vs. Active), scorecard through mid-year 2019¹ records an outperformance of 87% of all U.S. domestic funds by their own benchmark for the last 15 years and of approximately 71% for 2018. Fama and French (2009) have explored the concern about the relevance of skill as opposed to luck from funds and asset managers. Their findings have shown no evidence that the U.S. active mutual funds have sufficient skills to cover their costs in the period from 1984 to 2006. Moreover, according to Bloomberg² assets under management (AuM) from the U.S. passive equity funds have surpassed for the first time AuM from active ones in August 2019. According to Morningstar Inc.³, the assets in index-tracking U.S. equity funds reached \$4.27 trillion compared to \$4.25 trillion managed by stock-pickers.

Overall, passive and active investing differ fundamentally in their views on whether it is possible to outperform the market, in other words, on whether it is possible to anticipate market returns in one way or another. Subsequently, the related question is whether the market is predictable. If it is not, selecting between different stocks without firm views or strong assumptions on their future performance relies on sheer luck alone. It is in this context that the present study is embedded. It explores market predictability based on an econometrical and statistical approach, and if some predictive factors may be able to capture and differentiate some signals from random noise, in different market time-series.

1.1.2 The ‘Random Walk Theory’

The theory of no predictability of the market is referred to as the random walk theory (RWT) according to which all and any successive changes in security prices are de facto independent random variables. ‘Independent’ here implies that prices have no chronological relationship with one another and ‘random’ means that changes in prices are not related by any deterministic model with forecasting power. Introduced initially by the French broker Jules Regnault in 1863, and later by the French mathematician Louis Bachelier, the concept was finally popularised by Burton Malkiel in 1973 in its book *A Random Walk Down Wall Street*. Academic research has empirically tested this theory;

¹ <https://us.spindices.com/search/?query=US+spiva&Search=Go&sortType=Relevance&resultsPerPage=25&Region=region-americas>

^{2,3} <https://www.bloomberg.com/news/articles/2019-09-11/passive-u-s-equity-funds-eclipse-active-in-epic-industry-shift>

Alexander (1961; 1964) and Fama and Blume (1966) have analysed returns from trading rules based on past prices and they found that these rules were not able consistently to produce returns beyond a simple buy-and-hold strategy. However, Theil and Leenders (1965) found a significantly positive dependence in successive prices of securities on the Amsterdam stock exchange from November 1959 to October 1963. Also, the studies conducted by Robert Levy (1966; 1967), who has tested a number of additional trading rules based on past prices, yielded results that seem to be inconsistent with the RWT. Taking into consideration the results of empirical studies, there is thus no consensus on the applicability of the RWT among academics.

1.1.3 The 'Efficient Market Hypothesis'

The efficient market hypothesis (EMH) is another concept the aim of which is to explain or at least justify the RWT (cf. Samuelson 1965). Thus, in his research entitled *Proof That Properly Anticipated Prices Fluctuate Randomly* published in 1965, Samuelson argues that in a fully informational efficient market, price changes have to be unforecastable. If prices fully incorporate all the expectations and all the information of all the market participants, they should no more be forecastable at a certain point and have to follow a random walk. He also relies on the randomness of news and events and on the fact that randomness is an important factor that introduces the notion of risk, and where the return is simply understood as the reward for taking this risk, that is, uncertainty. Fama (1970), on the other hand, has depicted three stages of efficiency for the EMH, namely, (1) weak efficiency, i.e. when predictability is not achievable according to past prices, (2) semi-strong efficiency, i.e. when public information does not allow predictability, and (3) strong efficiency, i.e. when all available information including inside information is not sufficient to forecast future market prices.

Generally, the EMH has been widely tested and the results have largely and consistently supported the hypothesis through a variety of datasets from a number of different markets including the New York and American stock exchange, the Australian, English, and German stock markets, different commodity futures markets, OTC markets, and corporate and government bond markets. As a result, from its early days when it was seen as a curious concept taken seriously by only a few economists, the EMH has then become the dominant paradigm in finance. From a macroeconomic point of view, the EMH can be seen, in fact, as an extension of the zero-profit and perfect competitive equilibrium condition. The generally accepted version of the EMH is the second or the semi-strong state hypothesis, which involves all publicly available information since insider traders are able to take advantage on the market by themselves. This includes then the first stage of market efficiency that asserts that past prices are not valuable to predict future prices.

The results of some studies, however, are not in line with, or at least call into question, the semi-strong version of the EMH by pointing out the anomalies that should describe market inefficiencies. Ball (1978) and Watts (1978), for instance, have examined stock price reaction to earnings announcements and systematically found non-zero abnormal returns after the announcement. They concluded that some of these returns were either due to inadequacies in the asset pricing model for the former, or to market inefficiencies for the latter.

1.1.4 Behavioral Finance

The most severe critique comes however from behavioral economists who argue that the EMH is based on counterfeit assumptions regarding human behavior which is rationality. This led to the advent of a new paradigm in finance known as the modern Behavioral Finance theories, which try to describe market anomalies such as severe rises or falls in the stock prices as a result of psychological influences and biases that affect the behavior of investors and financial practitioners. There is then Ricciardi and Simon (2000) which try to make a global assessment on what is Behavioral Finance in

general, trying to define some key concepts and pointing out the importance of sociology and psychology in the whole concept. The conclusion to retain after the previous review is that even after several decades of research and thousands of published studies, economists and theoreticians have not yet reached a consensus about whether markets are in fact truly efficient. That means that there is no absolute truth that drive the consideration of a single market's model, but more some assumptions based on theories which train their own proponents.

1.2 The Need for Finding Robust Predictors

1.2.1 Fundamental Analysis

Fundamental analysis refers to the use of financial statements metrics, in order to predict changes in price of a security in the medium to the long-term. It posits that the market may incorrectly price a security and, therefore, is in conflict with the EMH. The intrinsic value of an asset such as a stock can then be evaluated based on fundamental metrics, which are tangible economic variables extracted from the real economy perspective. It is mean by real economy all what concerns the production, purchase and flow of goods and services and the financial economy, all what concerns transactions of securities and financial assets that represent ownership and claims. It can be said that the latter reflects the former and translates physical transactions into financial ones, this, based on financial and accounting measures.

In matters of asset valuation, the discounted cash-flow model (DCM) and the dividend discount model (DDM) are relevant as they respectively use previsions of future discounted cash-flows and dividends in order to valuate a company based on actual values. This activity is conducted by practitioners and financial analysts who assess the value of companies in order to assist clients in making investment decisions. Warren Buffet is a prime and well-known example of a fundamental investor whose strategy has essentially relied on the value-based investing model. In other words, such strategy involves buying only stocks with strong fundamentals such as robust earnings and long-term growth potential. Another approach that is one of the most widespread investment strategies is to rely on accounting metrics as factors to select stocks to invest in. Examples include, among others, earnings-to-price ratio, dividend-to-price ratio, cash-flow-to-debt ratio, sales-to-cash ratio. These are all potential screeners that will allow the investor to construct portfolios by weighting stocks regarding to these screeners' values.

Another strategy explored in the academic literature and commonly used by hedge funds is sorting stocks into deciles according to a factor value and going long for the best decile stocks and short for the worst ones, on the condition that average returns are negative for the losers. This is quantitative factor investing. On a related note, Feng, Giglio and Xiu (2019) summarise numerous factors (altogether 150 of them in the 2019 version of their study) under what they called the 'zoo factor'. Also, Harvey and Liu (2019) listed as many as 382 factors in their meta-analysis of studies published in top academic journals until 2019. The factors are considered to be significant when they include significant alphas according to market models such as the capital asset pricing model (CAPM), the Fama and French 3 or 5-factors models. In this context, alpha is understood as a constant 'abnormal return' not explained by the market model used whether it is the CAPM or a Fama-French model. If the alpha is positive and statistically significant, the factor used for investment earns consistently an extra return above the standard market index and, consequently, this factor may be used to beat the market.

In summary, however, one of the main drawbacks of fundamental analysis as an investment approach is the frequency of gathering available data, which are mostly quarterly according to the general

accounting guidelines followed by companies, in publishing their results and financial statements. This therefore narrows the opportunity for rebalancing in the short-run and benefitting more substantially from an eventual alpha generated by the factor, under the condition that this alpha is also available for shorter time spans, namely, months, weeks or days. This low frequency in signals opens a potential advantage for technical analysis, despite the fact that this latter has to counter a more restrictive form of the EMH.

1.2.2 Technical Analysis

Compared to fundamental analysis, technical analysis relies exclusively on price, more specifically, on past price movements and, sometimes, on volume. It counters the first stage of market efficiency, namely the weak form of the EMH based on past information from historical prices. Accordingly, for Technical Analysis, past price movements may provide an insight into the expected future prices when the first stage of the EMH says the absolute contrary. In this view, if short-term predictability would be confirmed for some cases, could it be hypothetically explained through behavioural finance? In the sense that if all main market actors rely on technical analysis such as trends, key price levels (such as supports and resistances) and other signals, the market would likely evolve in the direction of a consensus shared among market practitioners. The question that also arises is whether technical analysis is taken seriously by institutional and big market actors such as pension funds, hedge funds, and the like, and what are the reason if not?

Returning to the factor zoo introduced by Feng, Giglio and Xiu (2019), despite the majority of fundamental factors, it can still be mentioned the momentum factors (1-year, 6-month, 36 months), long and short-term reversals factors, and few other factors based exclusively on past price values, as being technical factors. According to the listed table of factors presented in Feng et al. (2019), which includes monthly average returns and Sharpe ratios of all listed factors, it can be noted that the technical factors mentioned previously perform quite well. There is for example the one year momentum factor (Carhart, 1997), which is the third most profitable factor among the 150 of those listed in the zoo factor (Feng et al., 2019), with a monthly average return of 0.63% for an annual Sharpe ratio of 50.2%, against a 0.91% monthly average return and a 92.8% Sharpe ratio for the 'betting against beta' best performing factor of Frazzini and Pedersen (2014). This precision is made here in order to replace technical analysis at the hearth of investing through the momentum strategy, among the slew of fundamental factors, and not to be confused by not making any difference between technical and fundamental factors. Thus, it has to be recognised that fundamental factors are of preponderant in investment strategies and research, when looking at their ratio comparing to technical ones. From an academic perspective also, they keep an important weight, and influence more institutions through their strategic approach.

In the context of discussing Technical Analysis, it is important to introduce Technical Indicators to better understand the issues related to them. These are in fact standardised formulas which are derived from the price of the concerned underlying and sometime its volume. Several of these factors have their own proper designations and are intensively used by technical analysts or chartists. In spite of their popularity among practitioners, traders, and desk traders, however, technical indicators (except for momentum and several other factors as seen previously) are not often mentioned in the academic research and are predominantly not the subject of any in-depth studies. Also, it is generally admitted by practitioners that technical analysis can be used only for short-term predictability, which explains why only active traders and investors rely on. This is at odds with the above-mentioned results regarding the momentum and reversals factors, even though mentioned indicators are not the popular and mostly used ones in academic research, such as Moving Averages, Filters, etc. A review on the profitability of Technical Analysis through the Academic Literature is thus carried out in section 2 in order to have a better overview on what should be expected. It is also in this purpose that main

Technical Indicators will follow an in-depth assessment in the present study, when looking at their potential forecasting power and the significance of an eventual generated alpha.

As mentioned above, Technical Indicators are derived from the price and, for some, from the volume. They were generally developed by different well-known practitioners, among them the renowned technical analyst and trader J. Welles Wilder Jr. (1930s–), who is one of the pioneers of technical analysis and the architect of today's core indicators such as the relative strength index (RSI), the average true range (ATR) or the average directional index (ADX). Another that can be mentioned is the American trader Bill M. Williams (1932–), who developed several other indicators such as the accelerator/decelerator oscillator (AD), the alligator and the awesome oscillator (AO). Generally, all of these indicators are constructed in a similar way since they are derived from the price and all of them aim to translate the notions of momentum and market reversal. The fact that they are derived from price and, therefore, have a correlation to the momentum and to other reversal indicators can explain why specific and targeted indicators, except for moving averages, are not well incorporated into academic literature. Sections 3 and 4 aim to demonstrate a difference, if any, between these specific indicators. If any are shown to yield benefits, this would call for a more detailed study and understanding of these indicators. Going back to their function, the aim of these indicators is to recognise market patterns, whether the market price is into a trend, a steady range, or ready to prime a reversal. Consequently, there are different families of indicators which can be separated into trend indicators, oscillator indicators, volatility indicators, and volume indicators. Section 3 introduces a more exhaustive and extensive study of a precise list of different technical indicators, for which category they stand and how they could be used for.

Another important difference between technical indicators and fundamental factors is that the former may take a wide range of values according to their initial parameters, while the latter are solely unique values which are given. Technical indicators thus can be calibrated so as to include the number of past days used to calculate their value, which then offers a wide range of potential values.

1.3. The Forecasting Power of Technical Analysis

There are mainly two ways of evaluating the market predictability of technical indicators. The first is to explore whether a potential linear relation exists between values of an indicator and logarithmic market returns at time t . This is obtained by regressing the values of the indicator one period ahead to the logarithmic equity risk premium, which is the logarithm of the market return minus the risk-free rate. If a statistically significant relationship is found through this regression, at a given significance level, indicator values at time $t - 1$ are said to have a forecasting power on prices at time t . Logarithmic equity premium is used for econometric properties convenience of log-returns, in order to get a log-linear model that is linear in parameters. The more important the R-squared, the more robust the relationship is considered to be. The methodology used in section 4.2 to evaluate the equity risk premium predictability of technical trading strategies explains more in detail the procedure in section 3.3.5.

The second method is to regress excess market returns to excess returns obtained from a strategy built with technical indicators. One needs then to test the significance of a positive constant alpha in the regression. Subsequently, the strategy may generate, at a certain significance level, an abnormal constant return apart from the market one.

If both tests yield significant findings, the strategy built with technical indicators is able to significantly outperform the market. However, the choice to build strategies by optimising parameters which render a high number of possible strategies, brings some issues regarding the principle of significance in a multiple testing framework. These issues are addressed in section 1.4.2 and 1.4.3. It

can also be specified that these tests need to be conducted in Out-Of-Sample data in priority, in order to become the more consistent.

1.4. A Practical Challenge: Find Technical Strategies That Challenge Standard Market Indices

1.4.1. The Process of Generating Strategies from Technical Indicators

As mentioned above, before measuring predictability one has to calibrate the technical indicator such as the past lagging period(s) used to calculate the indicator's value. Another variable to calibrate for a technical strategy is the indicator's value, that can represent an entry point for a buy/sell position. It is noteworthy to specify that a full exposition to the market is sought after, in order to maximise chances to outperform it by maximizing the time of exposure. To achieve a complete exposure, a unique indicator value needs to be selected which will switch exposure from long to short.

An intuitive approach would be to use an optimisation of these two variables in order to seek the best combination that generates the best performance measure in a given period. The selection process is described in detail in section 3.2.

Another possible approach is to select extreme values of an indicator to enter the market specifically when these values are reached. One would end up with two entry/exit levels instead of one as previously: one indicator value for a buy signal and another for a sell signal. This would leave the possibility of no positions opened when the market would navigate between these two levels. That is how a 'targeted market exposure' would be characterized. The strategy therefore can decide to sell when the market is considered to be in the state of 'overbuying' and buy when considered as 'oversold'. Inversely, it can decide to open positions in the sense of a trend when this trend is clearly defined.

This approach, however, can prevent the strategies to outperform the market in term of cumulative return, since no full market exposure is maintained. This will lead to some positive profits to be left behind. Indeed, knowing that the stock market is a growing market, not being exposed is a risk to give up certain profits.

Nevertheless, the main advantage of this approach may be the reduction in risk since only good quality signals may be considered. This would have more accuracy to be successful than a single signal that wants to separate long and short exposure. Chances are higher to find better Sharpe ratios and lower maximum drawdowns in the OOS period and these variables could be used to optimise the strategies. Also, a solution to overcome the gap in cumulative return would be to leverage positions.

However, making the choice of this targeted approach would increase the number of parameters to optimise since, two indicator levels have to be fit, one for a buy signal and one for a sell signal. Also, chances to find false positives and the risk to be exposed to backtest overfitting may also increase in that sense.

1.4.2. The Risk of False Positives

Problem arises when assessing the reliability of significant alphas discoveries in an optimisation framework. If 'x' number of lagging days and 'y' number of indicators entry points are tested, then 'x' times 'y' of potential strategies need to be tested. That to say, all possible combinations of the two variables. Also, if 20 lagging days and 100 indicator levels from 1 to 100 are combined, this would result to 2'000 strategies to be tested. This is an issue that is most likely not accounted for in the content of numerous publications. This is p-hacking in a multiple testing framework.

Many publications include significant factors but omit to specify that these discoveries are made by testing numerous other factors. This calls into question the statistic accuracy of their findings. More specifically, what is not considered in their tests is that they are conditional to the fact that 'x' factors have already been tested. There is thus a publication bias where only significant results are published, and non-significant results are purposely omitted.

When testing the significance of a factor in a multiple testing approach, the t-statistic or p-value of the test must suffer a penalty according to the number of factors ultimately tested in order to be statistically relevant. If no penalty is accounted for, one would face a potential type I error (a false positive or false discovery) where the null hypothesis of $\alpha \leq 0$ is wrongfully rejected.

Harvey and Liu (2015) assessed several ways of addressing the type I error when facing a multiple testing approach, such as applying a haircut either to the strategy's Sharpe ratio, either to the t-statistic or p-value. They introduced two gauges, the Family-Wise Error Rate (FWER) and False Discovery Proportion (FDP) in order to quantify the potential false discoveries. More precisely, and explored in detail below, the FWER is the probability to make at least one false discovery among the N significant strategies tested to generate a positive alpha. The FDP, on the other hand, focuses on the entire proportion of false rejections. Harvey and Liu (2015) introduce then some p-value adjustment under the two previously described measures, introducing several adjustment processes such as the Bonferroni, Holmes and BHY (Benjamini, Hochberg and Yekutieli) methods. Also, it has to be noticed that these adjustments are quite severe and do apply uniquely for independent trials or low correlated ones according to Harvey and Liu (2015).

A question that may arise, is to know whether strategies built from Technical Indicators may apply to a multiple testing framework and would require some p-values adjustments. As specified, p-values penalties are quite severe and are motivated by the independence of tested strategies. Also, when a researcher is manually testing among few dozens of factors to see if one would be positively significant, such a penalty would be justified. However, it is more unclear when searching among thousands of combinations of a unique configurable strategy. Correlation of returns between combinations in this case can be high. Looking for example the FWER Bonferroni method, it would require a confidence level of 0.05 divided by the amount of total trials in order to make the test significant. Applying the test to 2'000 combinations for a single technical indicator would lead to consider a confidence level of 0.0025% which would be much severe, especially considering a potential high correlation. Considering the FDP BHY formula, it can be noted that it is less stringent than the Bonferroni, since it reduces the hurdle the more strategies are progressively tested. For the last strategy to be tested for example, it can be expected a confidence level of 0.66% but this uniquely for the last one. For the second strategy to be tested for example, one would have end up with a confidence level required of 0.0007% which is even more stringent than the Bonferroni test. For this purpose, these statistics will not be used, but instead, it is the model developed by Harvey et al. (2015) that accounts for correlation that will take place. Results will then rely on t-statistics that control for the FWER and the FDR at 5%, with a baseline of $r = 1/2$ and $\rho = 0.2$. The baseline is defined at section 4.4.

It can also be recalled that technical indicators are substantially different from fundamental factors since factors do not vary depending on certain parameters. Indeed, they always have a unique given value. Technical indicators, to the contrary, allow for the calibration of their levels and therefore become an adjustable variable. There is a need therefore to find a way to deal with the discovery of false positives in the framework of optimization. Also, it can intuitively be inferred that strategies generated by a unique indicator but from different combinations of parameters may be closely related, specifically if increments in the parameters are close one to another. It can be presumed that strategies returns may be highly correlated in this sense. A way to contest this claim is to assess whether cumulative returns of combinations would look substantially different between them. The question

then is to see if there are some significant jumps and divergences across performance curves; which would lead to interpret results as potentially being the product of pure chance.

1.4.2 The Risk of Backtest Overfitting

Backtest overfitting can be defined as the risk of finding a false positive or type I error by over-training a finite set of data to an important amount of independent parameters. The risk of backtest overfitting increases with the number of parameters and the number of trials carried out on the same data sample. The case is when an overfitted strategy captures only random noise or residual variations instead of a structural signal in the data. This brings it to overperform an asset uniquely by chance. Also, a high independence of the fitted parameters increases the risk to imply randomness. When comparing the notions of backtest overfitting and the risk of finding false positives, they still intrinsically the same with the difference that the first one can be a practice that can imply the second. Also, Backtest overfitting can involve the intervention of computers and statistical tools for generating false successful strategies.

Backtest overfitting can also describe the fact to accumulate more parameters than the data can justify, leading to an increase of possible combinations and mechanically to an increase of type I errors.

Overfitting mostly occurs in machine learning and data mining when trading strategies are set by applying a heavy training on data. They can thus test millions, if not billions, of combinations in purpose to increase the chance to find positive output.

Therefore, and in the case in question, the optimisation of technical indicators' parameters could lead to a major issue when considering the risk of backtest overfitting. Question arises whether optimizing all parameters may not result in a false positive that may essentially be due to chance.

A way to assess if optimisation is the result of random noise is to carefully assess all outputs from this optimisation process. The cumulative returns from all possible combinations of parameters need to be surveyed to explore whether the distribution of these performances follows a random path in particular.

Another way to control the risk of backtest overfitting is to use the *holdout method* that consists of splitting data into two parts such as an in-sample (IS) part which is the training set and an out-of-sample (OOS) part which will officiate as a confirmation set. If OOS results are congruent to IS results, it can be concluded that the pattern found in the IS part tends to persist in the OOS one, and that backtest overfitting may not apply. The more important the proportion of untrained data, the more robust obtained results should be. Also, practitioners may usually use an 80/20 percent proportion of respectively IS/OOS data, but a 50/50 or even less may produce more robust results.

However, the hold-out method has its weaknesses as the confirmation of a trend can be the result of chance again. A process needs to be run several times in order to assess whether the outputs of the process may be the fruit of the hazard. This is the logic behind the cross-validation process, which, instead of separating data into two unique parts, it proposes a scheme where the data is divided into 'k' parts. For example, a k-fold cross-validation scheme will result in a prior random shuffle of the entire dataset. Then the shuffled data will be splitted into k-parts, and a single part will officiate as the confirmation set when all others are training sets. The k-fold CV procedure will designate subsequently all k-partitions as training sets, resulting in k overall tests. A times-series cross-validation could be more appropriate in financial strategies backtesting since, unlike the k-fold procedure, the data is not shuffled randomly at first but directly splitted.

Bailey, Borwein, de Prado, and Zhu (2013) proposed a solution to quantify if a strategy generating process may lead to overfitting. The procedure in this case is defined as the optimisation of the parameters of a technical indicator. They proposed a measure that estimates the tendency for a procedure to overfit and developed the so-called probability of backtest overfitting (PBO). More

specifically, this measure applies to a strategy selection procedure within a combinatorially symmetric cross-validation (CSCV) procedure. The procedure uses a cross-validation scheme that has been mentioned earlier but being a more specific and particular case (combinatorially symmetric). It is also a times-series CV since it does not propose to shuffle the data first.

‘Symmetric’ is defined as in the sense that all training sets from the CSCV procedure are re-used as testing sets and vice versa, and this with equal weights. Secondly, the term ‘combinatorially’ is understood as choosing to separate the data into k partitions to form all possible combinations of them, without repetitions. The role of running the PBO is to see whether the OOS performance metric (the cumulative return, the Sharpe ratio, etc) tends to be lower than the median of all OOS trials. Applied to an optimization framework, the procedure seeks to detect if OOS results are consistent to IS ones, and if the optimal OOS is consistently above all OOS results of other non-optimal trials in the optimization process. This procedure needs to be repeated for all combinations of partitioned data from the CSCV procedure. The PBO is in finality the ratio of the number of cases where the OOS performance measure underperforms the OOS median, divided by the total number of combinations of partitioned data. An in-depth study on the PBO of Bailey, Borwein, de Prado, and Zhu (2013) is practiced in section 2.2 of the literature review and in the 3.3.3 methodology section. Finally, the PBO will represent the main tool for quantifying and controlling for Backtest overfitting in this survey. Park and Irwin (2004) will allow to verify if results obtained are in line with most of published articles.

2 LITERATURE REVIEW

2.1 Profitability of Technical Analysis : A Review - Cheol-Ho Park and Scott H. Irwin, 2004

2.1.1 Introduction

One of the aims of this study is to review the academic literature on the profitability of technical analysis. Its starting point is the survey carried out by Cheol-Ho Park and Scott H. Irwin (2004), the main purpose of which was to ‘review the evidence on the profitability of technical analysis’ across theoretical and empirical studies on technical trading strategies. The authors attempted to provide a general overview of the potential results obtained through technical analysis based on different approaches.

The two authors argue that technical analysis has generally been widely used by market participants in futures markets and foreign exchange markets. In 2004, as many as 30% to 40% of practitioners believed technical analysis to be a reliable tool for determining price movement within short time horizons of up to six months. The paper also reports on the main technical trading systems usually employed in academic tests such as the dual moving average crossover, the outside price channel, the relative strength, and the Alexander’s filter rule.

The dual moving average crossover is the rule that implies a sell signal when a short-term moving average cross below a longer-term one, and a long signal in the opposite case. In other words, it accounts for the start of a momentum signal. The outside price channel is a tool that signals a break in channel range. The price evolves between two lines traced according to a support and a resistance level, and a break in one of them may denote an opportunity to buy or short the market.

The relative strength or a momentum filter aims to detect oversold and overbought signals depending on the value of the relative strength. Finally, the Alexander’s filter rule is a system for detecting

change in momentum as it introduces a buy signal when a one-day close is above its most recent low by x%, and the inverse also holds for a short signal.

Park and Irwin (2004) moreover categorised studies into two distinct types according to the time of their publication. They distinguished between the so-called early empirical studies (1960–1987), which represent all technical trading strategies that ‘were profitable in foreign exchange markets and futures markets, but not in stock markets before the 1980s’, and modern empirical studies (1988–2004), which showed that ‘technical trading strategies consistently generated profits in a variety of speculative markets at least until the early 1990s’.

2.1.2 Early Empirical Studies (1960–1987)

Some early empirical studies include Fama and Blume (1966), Mandelbrot (1963), Alexander (1961; 1964), Stevenson and Bear (1970), and Sweeney (1986).

According to Alexander (1961), small filter rules from 24 Alexander’s own filter rules, ranging from 0.5% to 50%, outperform the buy-and-hold strategy, and this accounting for transactional costs. Mandelbrot (1963), however, found Alexander’s computations of returns to include some biases as he failed to take into account potential gaps between the signal and the entry price. In response to this criticism, Alexander (1964) reiterated his computations for S&P Industrials, this time using closing prices of the day as entry prices once the signal was generated. He discovered that, after accounting for commissions, only the largest filter (45.6%) was able to beat the buy-and-hold strategy. Fama and Blume (1966) also found the results presented in Alexander (1961; 1964) to be biased but, this time, because dividend payments were not accounted for, which considerably lowered profits. By reproducing the same tests but correcting for Alexander’s shortcomings, Fama and Blume concluded that, overall, despite few marginal positive cases, the filter technique could not be used to increase the expected profits after accounting for all costs, such as brokerage fees and implementation costs.

Stevenson and Bear (1970) carried out a comparable study on corn and soybean futures from 1957 to 1968. They assessed three trading systems regarding the filter technique: stop-loss orders from Houthakker (1961), filter rules by Alexander, Fama and Blume, and some combinations of both. They found that stop-loss orders and combinations of filters and stop-loss consistently outperformed the buy-and-hold strategy, severely questioning the applicability of the random walk hypothesis. Also, Sweeney (1986) tested several foreign exchange rates considering risk, transaction costs, out-of-sample performance, and statistical tests. For that purpose, he assumed that the capital asset pricing model (CAPM) can explain excess returns for both the filter rules as well as the buy-and-hold strategies. His model implies that expected excess returns to the CAPM should be equal to zero. However, the discovery of high significant returns of the filter rules implies that the CAPM is not an appropriate model in foreign exchange markets. Sweeney then concluded that major currency markets indicated important signs of inefficiency over the first eight years beginning in March 1973, assuming a constant risk premium. Nevertheless, he stated that the results could be consistent with the efficient market hypothesis (EMH) if risk premia vary over time.

Moreover, Park and Irwin (2004) surveyed in their paper 42 early empirical studies on technical analysis published between 1961 and 1987 and summarised them in a detailed, tabular format. They mentioned, next to those reviewed above, studies by Donchian (1960), Levy (1967a; 1967b), James (1968), Stevenson and Bear (1970), Martell (1976), Arnott (1979), Dooley and Shafer (1983), Taylor (1983; 1986), Thompson and Waller (1987), and others. All of them present different technical trading systems such as channels, filter rules, stop-loss orders, moving averages, momentum oscillators, relative strength, non-weighted and exponentially weighted moving averages, forms of five points chart patterns, adaptive filter models and pure information models, beta-modified relative strength, slope (trendline) method, and statistical price trend model. Most used as their benchmark the buy-and-hold strategy on the S&P 500, NYSE stocks, various commodities and other underlying futures.

In summary, early empirical studies tested different markets, yielding highly varying results from one market to the next. Park and Irwin (2004) conclude that ‘overall, in the early studies, very limited evidence of the profitability of technical trading rules was found in stock markets (e.g., Fama and Blume 1966; Van Horne and Parker 1967; Jensen and Benington 1970), while technical trading rules often realized sizable net profits in futures markets and foreign exchange markets (e.g., for futures markets, Stevenson and Bear 1970; Irwin and Uhrig 1984; Taylor 1986; for foreign exchange markets, Poole 1967; Cornell and Dietrich 1978; Sweeney 1986). Thus, stock markets appeared to be efficient relative to futures markets or foreign exchange markets during the time periods examined.’

These studies, however, may have still been subject to limitations in testing procedures.

First, most of them tested for only one or two well-known trading systems, such as filters or moving averages, potentially resulting in snooping and survivorship bias. It is worth noting that, for popular trading systems, numerous variants could have been tested among many different markets and some profitable strategies could have been the result of chance only. In fact, Dooley and Shafer (1983) and Tomek and Querin (1984) revealed that to be the case when technical trading rules were applied to randomly generated price series. Survivorship bias, on the other hand, refers to the practice of publishing only significant results among numerous tests that could have yielded non-significant results. This practice, consequently, highlights statistical exceptions rather than representative cases. Second, the riskiness of technical trading rules was arguably often ignored in the context of excess-return on benchmark. Thus, trading rule returns may not necessarily refute market efficiency if they are at a higher risk, which has to imply higher returns. Therefore, a careful assessment of the risk-reward profile on the strategy also needs to be carried out.

Third, most early studies do not systematically use statistical tests to assess their results seeing that technical returns may not follow a normal distribution. Thus, only four studies (James 1968; Peterson and Leuthold 1982; Bird 1985; Sweeney 1986) measured the statistical significance of returns on technical trading rules by using Z- or t-tests, under the assumption of normality. Also, Lukac and Brorsen (1990) demonstrated that returns on technical rules were positively skewed and leptokurtic (kurtosis above 3), which led them to conclude that the use of t-tests for technical rules could be biased.

Fourth, according to Taylor (1986), there is a chance that most published studies apply doubtful optimisation with the risk of the identifying false positives. He thus recommended splitting datasets into training and testing segments in order to confirm optimisation results obtained in the training segment. This process does not guarantee, however, that the results found are more significant but merely that more confidence should be placed in the results. According to Park and Irwin (2004), only three studies among the early empirical ones apply an in-sample and an out-of-sample procedure, namely, Irwin and Uhrig (1984) and Taylor (1983;1986).

2.1.3 Modern Empirical Studies (1988–2004)

The first modern empirical study was carried out by Lukac, Brorsen, and Irwin (1988), and it aimed to explain most of the limitations enumerated in the early period. Most modern studies improved their testing procedure, the treatment of transaction costs, the risk, the parameter optimisation, out-of-sample tests, statistical tests, and data snooping issues, but they did so in different ways.

Park and Irwin (2004) thus divided modern studies into seven distinct categories according to their testing procedures, namely, standard, model-based bootstrap, genetic programming, reality check, chart patterns, nonlinear, and other studies.

In what follows, these different categories are explored along with the review presented by Park and Irwin (2004) of altogether 92 modern studies from Lukac, Brorsen, and Irwin (1988) to Sapp (2004).

Standard studies are those that include adjustment for transaction cost and risk, statistical tests, parameter optimisation, and out-of-sample tests. According to Jensen (1967) and Taylor (1986),

parameter optimisation and out-of-sample tests are significant improvements compared to early studies since they may reflect traders' behaviour and partly address data snooping. Lukac, Brorsen, and Irwin (1988) applied them to test three hypotheses: (1) the random walk model, (2) the market efficient model with a traditional test, and (3) the market efficient model with the Jensen's test. They tested 12 technical trading systems on 12 different commodities future markets from 1975 to 1984. They found that seven out of 12 systems generated statistically significant monthly gross returns (1), with four of seven yielding statistically significant monthly adjusted returns to transaction costs (2) and risk (3). They concluded that the disequilibrium pricing model suggested by Beja and Goldman (1980) could be a more appropriate pricing model than the EMH for describing futures markets for the 1978–1984 period.

Besides commodities futures markets, many standard studies also explore foreign exchange markets. Taylor and Tari (1989), Taylor (1992,1994), Silber (1994), and Szakmary and Mathur (1997) found technical rules to yield annual net returns between two and 10 percent, when looking at major currency futures from the late 1970s to the early 1990s. Others—such as Menkoff and Schlumberger (1995), Lee and Mathur (1996a;1996b), Maillet and Michel (2000), Lee, Gleason, and Mathur (2001), Lee, Pan, and Liu (2001), and Martin (2001)—investigated different sample periods. They found that technical trading rules generated profits for some spot currencies for their respective testing periods. However, it seems that the profits generated by technical rules in currency markets tend to decrease over time, which has been confirmed by, for instance, Olson (2004) and Kidd and Brorsen (2004). Overall, standard studies demonstrated that technical trading rules generated statistically significant profits principally in foreign exchange and futures markets.

Model-based bootstrap studies apply the bootstrap process proposed by Brock, Lakonishok, and LeBaron (1992) to test statistical significance of trading profits. The authors tested technical trading rules without optimisation and out-of-sample confirmation. Their study served as a reference point in modern studies since they tried to detect and adjust for data snooping biases by using datasets covering longer periods and model-based bootstrap methods. According to them, there are several advantages to the bootstrap method such as using a joint test of significance (multiple testing). In applying their bootstrap method consisting of data resampling from different stochastic models, they introduced null models to their testing procedure such as random walk with a drift, AR(1), and GARCH-M model. The basic method in the bootstrap procedure is to compare conditional returns from then the original dataset to conditional returns from a simulated dataset. They generated 500 bootstrap samples by applying the technical trading rule on each of them.

They concluded that '[i]n sum, this paper shows that the returns generating process of stocks is probably more complicated than suggested by the various studies using linear models. It is quite possible that technical rules pick up some of the hidden patterns.'

Also, Levich and Thomas (1993), found some statistically significant results in the average profits of four currencies, namely, the Deutsche mark, the Japanese yen, the British pound, and the Swiss franc from 1976 to 1990. Overall, results differed in model-based bootstrap studies, which depended on markets and sample periods tested. Generally, technical trading strategies were profitable in several emerging stock markets. For example, Ratner and Leal (1999) discovered that the moving average rules generated statistically significant returns in the stock markets of Mexico, Taiwan, Thailand, and the Philippines from 1982 to 1995. Others—namely, Bessembinder and Chan (1995), Raj and Thurston (1996), Ito (1999), Coutts and Cheung (2000), and Gunasekarage and Power (2001)—verified that observation. Despite being profitable in emerging markets, developed markets seemed either to show negligible profits after transaction costs or seriously declined over time. This finding is supported by Hudson, Dempsey, and Keasey (1996), Mills (1997), Bessembinder and Chan (1998), Ito (1999), and Day and Wang (2002). On another note, LeBaron (1999), Neely (2002), and Saacke (2002) reported the profitability of moving average rules in currency markets where the origin of profitability remained unknown regardless of whether they ensued from Central Banks' interventions or not.

Genetic programming, introduced by Koza (1992), is an intensive and heuristic computer process that searches for the fittest (best performance criterion) among a number of possible generated outputs. The solution candidates are functions represented by different nodes that form a tree structure. The structure is evaluated in a recursive manner starting from the roots of the tree where the combinations of different parameters lead to a unique solution (the top of the tree). The combination is introduced then of several functions such as real functions including arithmetic operators (plus, minus, divide) and Boolean functions like logical functions (*and*, *or*, *not*) or comparisons (*greater than*, *less than*). Allen and Karjalainen (1999) used a genetic programming approach to test the profitability of technical rules for the S&P 500 index from 1928 to 1995 in a daily frequency. They found that technical rules failed to generate consistent excess return after transaction costs for an out-of-sample data set which is in line with the EMH. Ready (2002) and Neely (2003) also documented genetic trading rules underperforming buy-and-hold strategies in the long-run for the S&P 500 or the DJIA indices. However, Neely, Weller, and Dittmar (1997) carried out tests on six foreign exchange rates from 1974 to 1995 and found that technical rules on each currency earned statistically significant positive net returns above buy-and-hold returns. However, generated trading rules appear to be excessively complex in their structures and realised profits did not seem to compensate for systematic risk. Genetic programming is subject to serious limitations, such as the utilization of a technology that was not available in the time the data is tested, which would not make sense according to Cooper and Gulen (2003). Moreover, the complexity of generated trading rules and the propensity of genetic programming may lead to overfitted solutions because of high combinations of rules and parameters.

Reality check studies principally adopt White's bootstrap reality check method to deal with data snooping and directly quantify its effects. White's bootstrap reality check method delivers a data snooping adjusted p-value by testing the performance of an optimised solution while taking into account the dependencies across all tested trading rules. Sullivan, Timmermann, and White (1999; 2003) demonstrated, therefore, that technical trading rules may have been profitable in the stock market until the mid-1980s but not afterwards. Also, Qi and Wu (2002) obtained statistically significant technical trading profits from 1973 to 1998 in the foreign exchange markets for major currencies. However, this method too suffers from drawbacks such as the difficulty to assess properly the entire universe of trading rules when calculating the p-value.

Chart patterns are technical figures that can be drawn for the diachronic evolution of an asset price. The pattern and its clear characteristics are generally identifiable, and they represent a potential signal of a trend, a reversal or a range structure in the market. Examples include figures such as gaps, spikes, flags, pennants, wedges, saucers, triangles, head-and-shoulders, and different tops and bottoms. Chang and Osler (1999) demonstrated that the head-and-shoulders pattern earned statistically significant returns of about 13% and 19% per year for the mark and the yen respectively from 1973 to 1994, which is higher than a buy-and-hold strategy on most stock indices. Also, Caginalp and Laurent (1998) showed that down-to-up candlestick reversal patterns for S&P 500 stocks earned statistically significant returns after transaction costs from 1992 to 1996. Lo, Mamaysky, and Wang (2000) also tested 10 different chart patterns and concluded that technical chart patterns provided incremental information that could add value to investment processes without the ability to generate systematically excess returns. Dawson and Steeley (2003) confirmed the results presented by Lo, Mamaysky, and Wang (2000) but found negative average market adjusted returns when applied on UK stocks data. Generally, results obtained differ significantly across patterns, markets, and sample periods tested. For all that, some showed statistically significant positive results in certain stock and foreign exchange markets. Most studies, however, did not use optimisations, out-of-sample testing, and did not take snooping biases into account.

Nonlinear studies tried to assess either the profitability or the informational usefulness of technical

rules derived from nonlinear models, such as the feedforward network regressions and the nearest neighbour regressions. Overall, nonlinear studies assert that technical trading rules based on nonlinear models show some evidence of profitability or predictability in both stock and foreign exchange markets. Gençay (1998a) showed that some simple technical trading rules based on a feedforward network for the DJIA index generated annual net returns of 7% to 35% across different subsamples in the period from 1963 to 1988. Therefore, technical trading rules based on a feedforward network easily outperformed a buy-and-hold strategy after accounting for transaction costs and risk. Fernández-Rodríguez, González-Martel, and Sosvilla-Rivero (2000) also found technical trading rules to outperform the buy-and-hold strategy before transaction costs when applied to the Madrid stock index. Again, Fernández-Rodríguez, Sosvilla-Rivero, and Andrada-Félix (2003) showed that trading rules based on the nearest neighbours model generated superior returns than moving average rules in the European exchange markets from 1978 to 1994. Timmermann and Granger (2004), however, found it inappropriate to apply nonlinear techniques unavailable in the period when the tested data was generated. Also, most studies ignored statistical tests for trading profits and they be subject to data snooping due to the use of simple and popular trading rules.

The studies in the ‘other’ category comprise all those not reviewed in the preceding ones. They are, in fact, quite similar to early empirical studies as they, with few exceptions, did not include optimisation or out-of-sample verification. Neely (1997) tested the profitability of filter rules and moving averages on four major exchange rates over the period from 1974 to 1997. He found that technical trading rules earned positive and significant net returns for 95% of the rules after deducting transaction costs. He admitted, however, that the results are not proof of market inefficiency because of possibly biased testing procedures relative to data snooping.

Neftci (1991) tried to propose a more theoretical approach. By applying the Markov times approach, he found some predictive power of the moving average rule on the Dow-Jones Industrials. Also, Pruitt and White (1988) and Pruitt, Tse, and White (1992) showed that a combination system consisting of cumulative volume, relative strength, and moving average (CRISMA) is a trading strategy that earns consistently higher returns than market indices in the period from 1986 to 1990. Overall, studies in this category performed quite well in stock, foreign exchange, and futures markets. But again, they generally did not account for trading rule optimisation, out-of-sample verification, and data snooping issues.

2.1.4 Summary on Empirical Studies

Park and Irwin (2004) attempted to provide a general overview of the profitability of technical analysis in academic publications from 1960 to 2004. In general, technical analysis has been widely used particularly by participants in futures and foreign exchange markets. As many as 30 to 40% of practitioners considered technical analysis to be a determinant factor in forecasting future price movement in short-time periods up to six months. Conventional theoretical models, such as the martingale and the random walk models, which deny technical trading profits, are seriously challenged by those models that assume that prices adjust gradually to information and that irrational behaviour could lead to mispricing on the market. Among these are noisy rational expectations models, feedback models, disequilibrium models, herding models, agent-based models, and chaos theory.

More than 130 empirical studies have been examined in terms of profitability of technical trading rules in the period specified previously. Studies were thus separated into two categories, the early empirical studies and modern empirical studies, depending on their testing procedures and the complexity of signals. Among modern studies, the number of trading systems increases, transaction costs and risk are incorporated, parameter optimisation and out-of-sample verification are carried out, and statistical tests, either conventional or more sophisticated, are applied. Park and Irwin (2004) classified modern studies into seven subcategories based on the different testing procedures, namely, standards, model-based bootstraps, genetic programming, reality checks, chart patterns, nonlinear, and others.

Results of early studies varied depending on the applied market tests. Overall, they found some evidence of profitability from technical rules on foreign exchange markets and futures markets but limited evidence on stock markets. Modern studies showed some economic profits for technical trading rules in the U.S. stock markets until the late 1980s, according to Bessembinder and Chan (1998), Sullivan, Timmermann, White (1999), and Ready (2002). Several studies also found economic profits in emerging stock markets (Bessembinder and Chan 1995; Ito 1999; Ratner and Leal 1999). In foreign exchange markets, technical trading strategies have made economic profits over the last few decades, but some studies suggest that trading profits have declined or disappeared in recent years (Marsh 2000; Neely and Weller 2001; Olson 2004). With regard to futures markets, technical trading strategies seemed to be profitable between the mid-1970s and the mid-1980s. According to Park and Irwin (2004), *'among a total of 92 modern studies, 58 studies found profitability or predictability in technical trading strategies, while 24 studies reported negative results.'* As many as 63% of the respective articles found technical trading rules to be profitable whereas 37% did not.

2.3 The Probability of Backtest Overfitting

David H. Bailey, Jonathan M. Borwein, Marcos Lopez de Prado, and Qiji Jim Zhu (2015) proposed a general framework to assess the probability of backtest overfitting. They have argued that standard statistical techniques that seek to prevent regression overfitting, such as the holdout method, are unreliable and inaccurate when building investment strategies. They set out to address the alarming rise of false positives in related academic publications, which is a consequence of the algorithmic and computational research processes that nowadays allow testing millions, if not billions, of investment strategies in a finite dataset of financial securities. They thus propose a computational procedure for detecting false positives when developing investment strategies. The risk of discovering false positives is high, the authors claim, since combining and fitting several parameters results in billions of possible outcomes. Even a simple process, such as one based on moving averages, may require fitting five parameters such as the two sample lengths for the moving averages, an entry threshold, an exit threshold, and a stop-loss. Also, when calculating the statistical significance of the Sharpe ratio or any other performance statistic from the fittest strategy, the Neyman-Pearson framework of hypothesis testing still remains consistent with the approach. The probability of finding false positives has to be mechanically greater than 5% since, in fact, the test would then be applied several times, or billions of times in some cases, for each strategy on the same data sample.

The core question posed in the respective study is: What constitutes a legitimate empirical finding in the context of investment research?

According to Miller (1981), the probability of finding false positives is positively related to the number of tests conducted on the same data. In addition, Sala-i-Martin (1997) claims, millions of regressions may be carried out by researchers. The notion of legitimacy is thus questioned here, since it is reported that most published and significant strategies do not account for multiple testing, or, in the words of Ioannidis (2005), *'most published research findings are false'*. Potential outcomes are then overfit backtests and strategies. An overfitting process occurs when the signal-to-noise ratio in a dataset is too weak, resulting in strategies that have been subject to past noise, rather than those based on a real signal.

Most backtests presented in academic and practitioners' publications typically do not specify the number of conducted trials before a positive discovery. Thus, according to Bailey, Borwein, Lopez de Prado, and Zhu (2015): *'Because those researchers have most likely not controlled for the number of*

trials, it is highly probable that their findings constitute false positives.’ They add that, in spite of their awareness of the issue of backtest overfitting, academic and investment institutions ‘have little incentive to expose them’, since their motivations to receive a mandate or to raise funds for research can lead them to intentionally hiding or ignoring this issue. Some may even pretend to control overfitting but do so with inappropriate techniques and tools, benefiting from an asymmetry of access to information.

The goal of their respective paper was to ‘*develop computational techniques to control for the increased probability of false positives as the number of trials increases, applied to the particular field of investment strategy research.*’ They introduce the probability of backtest overfitting (PBO) as a reference measure to control for overfitting and for computing when publishing positive results. In doing so, they first precisely account for the event of backtest overfitting. Overfitting occurs, the authors claim, when the strategy systematically underperforms the out-of-sample (OOS) set, given the fact that it carries the parameters configuration that delivers the best performance in the in-sample (IS) set. This underperformance is then explained by the fact that the IS optimal strategy is too closely tied to the noise of the training set (overfitted), leading then to the impossibility of extracting the signal in the validation test.

Second, they establish a general framework for assessing the probability of the actual occurrence of backtest overfitting. They set up an abstract probability space where the sample space contains IS and OOS test results.

Third, they introduce the null hypothesis that backtest overfitting has occurred and developed an algorithm for testing the hypothesis. More precisely, PBO is evaluated as the conditional probability that a strategy performs under the median performance measure in the OOS set while being optimal in the IS one. This framework can also be used to measure performance decay, probability of loss, and stochastic dominance of a strategy.

According to them, backtest overfitting is a deterministic fact (the model is either overfit or not) and associating a probability to a non-random event may seem incongruous. In a Bayesian sense, given some empirical evidence and priors, the posterior probability of overfitting can be inferred, and MacKay (2003) shows some examples of such reasoning.

For them, a generic, model-free, and nonparametric testing algorithm is required, since backtests can be applied to strategies generated by a large variety of methods and models. They thus introduce the so-called combinatorially symmetric cross-validation (CSCV) procedure, which ‘*produces reasonable estimates of PBO for several useful examples.*’ This CSCV implementation ‘*draws from elements in experimental mathematics, information theory, Bayesian inference, machine learning and decision theory*’ to address the difficulty of assessing the real representativeness of a backtest. Indeed, there is a gap in the academic literature regarding the representativeness of backtests, which is a significant omission seeing that practitioners rely heavily on backtests. The authors chose, therefore, to swap all IS and OOS datasets, in order to lower the credibility issue on *fixed* hold-out methods, where the IS sets do not take the place of the OOS ones. The procedure is described as ‘generic’ as no information on strategy development is necessary, making it applicable to all types of strategies. Finally, the general output is a bootstrapped distribution of OOS performance measure. Also, the CSCV implementation is not the only one, but one among many applicable techniques that fit the general framework when computing the PBO.

The overall procedure is derived from the simplest yet alternative way to control for overfitting. The principle of the so-called *hold-out* or *test set* method is quite simple and results in separating a data set into a training (IS) and a validation sets (OOS), and the strategy is related to the IS part. If the performance of the strategy in the OOS part is congruent with the one in the IS part, then the hypothesis of backtest overfitting can be rejected. However, despite the simplicity of this method, which can be seen as its advantage, it also suffers from limitations.

First, if the analysed data is publicly available, there is a chance that the supposed validation set (OOS) has been used in the training one (IS) to conduct data snooping.

Second, even if the OOS set is not used, information on existing data may be used unconsciously or not in accordance with Schorfheide and Wolpin (2012).

Third, the hold-out method could not adapt for too small datasets; the IS dataset would thus be too small to fit correctly and the OOS one too small to validate any assumption with enough confidence. According to Weiss and Kulikowski (1990), the hold-out method should be used only for those datasets that contain more than 1,000 observations which would equal to approximately four years of daily data, 20 years of weekly data, and 83 years for monthly ones. Van Belle and Kerr (2012) have also argued that a hold-out set can lead to validating an invalid strategy and vice versa because of the high variance of hold-out estimation errors.

Fourth, even if sufficient amount of data is collected, the OOS part still influences the latter, and, depending on whether the data is taken at the beginning or at the end of a time series, it may respectively be more or less representative of future data. This can lead to a handicap in two cases, first, by not being sufficiently representative of actual data and, second, taking the most representative part that could have been employed for the strategy's fit.

Fifth, according to Bailey et al. (2015), as long as several strategies are tested, overfitting will always be present. The hold-out method does not take into account this parameter, so it cannot really assess the representativeness of a backtest.

The review of these drawbacks shows that the question whether a backtest is overfitted is not the right one to ask. The question needs to be refined and the degree at which overfitting occurs needs to be assessed.

An alternative approach is to model an underlying financial asset by generating pseudorandom scenarios and then measure the performance of given strategies against these different scenarios. Carr and López de Prado (2014) present an application of this technique. The advantage of this over the hold-out method is that it relies on a distribution of several outcomes rather than on a single dataset. However, there is a possibility that these models are themselves overfitted or may be not be completely statistically representative of the underlying data. An example of these techniques is the pseudorandom generation of data by fractal interpolation.

Several other procedures have been proposed to determine whether an econometric model is overfit. For instance, White (2000), Romano et al. (2005), and Harvey et al. (2013) assessed econometric models that rely essentially on applying a penalty to the *p-value* to account for a multiplicity of trials. However, these approaches are mainly applicable to econometrical models that involve algebraic equations, which thus do not represent all possible investment strategies. The PBO is thus an alternative to these approaches.

Bailey et al. (2015) also set up a 'measure theoretic framework' in which the PBO and other statistics could be clearly defined.

They considered a probability space $(T, F, Prob)$ where Ω represents a sample space composed of pairs of IS and OOS samples. The purpose is to evaluate the probability of overfitting for the following *backtest strategy selection process*: select from N strategies labelled as $(1, 2, \dots, N)$ the strategy that brings the best performance measure (Sharpe Ratio, Cumulative Return, etc) when backtesting to the examined, underlying security. The random vectors $\mathbf{R} = (R_1, R_2, \dots, R_N)$ and $\bar{\mathbf{R}} = (\bar{R}_1, \bar{R}_2, \dots, \bar{R}_N)$ represent respectively IS and OOS performance measure of the N strategies on (T, F, P) . For a given sample $c \in T$ that represent a corresponding pair of IS and OOS datasets, by \mathbf{R}^c and $\bar{\mathbf{R}}^c$ to signify respectively IS and OOS sets that belong to the same pair c . For most applications, T will be finite, and the power set T can be used as F . It is often reasonable to assume that *Prob* is uniform on elements in T . However, no specific assumptions are made at this stage of the general discussion in order to allow for some flexibility in some particular applications.

The purpose is to compare the ranking of the best selected strategies IS to their corresponding OOS results. Considering the ranking space Ω of the $N!$ permutations of the strategies $(1, 2, \dots, N)$, the random vectors r and \bar{r} are obtained that represent respectively the rankings of \mathbf{R} and $\bar{\mathbf{R}}$, mapping $(T, F, Prob)$ to Ω . It gives then r^c and \bar{r}^c representing respectively the rankings of \mathbf{R}^c and $\bar{\mathbf{R}}^c$, with c signifying the same pair for IS and OOS sets. For example, if $N = 3$ and the performance measure is the Total Return, for a particular sample $c \in T$, $\mathbf{R}^c = (5, 11, 7)$ and $\bar{\mathbf{R}}^c = (6, 7, 13)$, this leads to $r^c = (1, 3, 2)$ and $\bar{r}^c = (1, 2, 3)$.

The following subset of Ω needs to be introduced $\Omega_n^* = \{f \in \Omega \mid f_n = N\}$ in order to define backtest overfitting in the context of an investment strategy selection process.

Bailey et al. (2015) state the following concerning the definition of backtest overfitting for a strategy selection process : *We say that the backtest strategy selection process overfits if a strategy with optimal performance IS has an expected ranking below the median OOS. By the Bayesian formula and using the notation above that is*

$$\sum_{n=1}^N E[\bar{r}_n \mid r \in \Omega_n^*] Prob[r \in \Omega_n^*] \leq N/2$$

They then continue to define PBO as *A strategy with optimal performance IS is not necessarily optimal OOS. Moreover, there is a non-null probability that this strategy with optimal performance IS ranks below the median OOS. This is what we define as the probability of backtest overfit (PBO). More precisely,*

$$PBO = \sum_{n=1}^N Prob[\bar{r}_n < N/2 \mid r \in \Omega_n^*] Prob[r \in \Omega_n^*]$$

In other words, it can be said that a strategy selection process does overfit if the expected performance OOS of the strategies chosen IS falls under the median performance rank OOS of all strategies.

Having provided a general definition of PBO, practical guidelines for its application are necessary. This requires a scheme for selecting different pairs of IS/OOS datasets. Bailey et al. (2015) developed the CSCV procedure for that purpose. This procedure is a way to divide a single set of data into several by cross-validating partitions of data and combining symmetrically IS/OOS obtained pairs. The CSCV procedure is explained in detail in section 1.4.2 and its precise methodology is described in section 3.3.4. Therefore, the CSCV procedure allows to assess the PBO of a strategy generating process. According to Bailey et al. (2015), their developed framework allows for evaluating the reliability of a strategy's backtest through four complementary measures:

“ 1. **Probability of Backtest Overfitting (PBO):** *The probability that the model configuration selected as optimal IS, will underperform the median of the N model configurations OOS.*

2. **Performance degradation:** *This determines to what extent greater performance IS leads to lower performance OOS, an occurrence associated with the memory effects discussed in Bailey et al. [1].*

3. **Probability of loss:** *The probability that the model selected as optimal IS will deliver a loss OOS.*

4. **Stochastic dominance:** *This analysis determines whether the procedure used to select a strategy IS is preferable to randomly choosing one model configuration among the N alternatives. ”*

In the CSCV method, the PBO is characterised by $\varphi = \int_{-\infty}^0 f(\lambda)d\lambda$. It represents the rate at which optimal IS strategies underperform the median of the OOS trials. In accordance with standard applications of the Neyman-Pearson framework, a standard approach is to reject models that have a PBO greater than 5%. Bailey et al. (2015) also believe that the CSCV method is more appropriate than a simple *k-fold cross-validation* (**K-FCV**) or a *leave-one-out cross-validation* (**LOOCV**) for evaluating backtest overfitting. First, the **CSCV** procedure ensures equal sizing of IS and OOS sets, which then ensures the comparability between IS and OOS performances. Second, CSCV is symmetric, meaning that all training sets (IS) are re-used as validation sets (OOS) and vice versa, which would again improve the comparability of generated strategies through all data. Third, CSCV respects the time-dependence and other season-dependent potential features present in the data since it does not require random data shuffling but only recombines the S subsamples initially created. Fourth, CSCV derives a non-random distribution of logits, in the sense that each logit is deterministically derived from one item in the set of C combinations. The CSCV approach provides, then, a single and unique result that can be independently replicated and verified by another user. Fifth, the distribution of logits gives some relevant information concerning the robustness of the strategy selection procedure. A robust strategy generation procedure must imply consistent OOS performance rankings that give similar logits. Sixth, the overall procedure to estimate PBO is model-free since it does not require one to “specify a forecasting model or the definitions of forecasting errors”. It is also non-parametric, as no distributional assumptions on PBO are made prior to the test. This is possible by using the concept of logits, which is the logarithm of odds where the odds are represented by relative ranks. Odds for IS optimized strategies to underperform OOS median are assessed. The distribution of logits may approach the standard Normal distribution if the selection procedure is information-less. On the contrary, if the selection procedure generates some information from the data, the distribution of logits would be centred in a positive value and the left tail would cover a small part of the negative value region. The PBO (φ) will be, then, approximatively close to zero.

Another important part in the CSCV procedure is the choice of the S subsamples that will deliver C combinations of market data samples to work on. S has to be large enough to create sufficient combinations that will allow to draw inferences. On the other hand, S cannot be too large since potential seasonal effects in the original data have to be maintained. Additionally, too many subsamples would drastically increase the number of combinations created; this must be taken into account for practical implementation. Bailey et al. (2015) recommends $S = 16$ is a “reasonable value to use in most cases”, specifically for data more than 6 years old. This will preserve daily, weekly, monthly, and quarterly effects. The number of combinations can be obtained by the following formula:

$$C_k^S = \binom{S}{S/2} = \binom{S}{k} = \frac{S!}{k!(S-k)!}$$

One would end up, therefore, with 12,870 logits with a standard deviation of the relative frequency $\sigma[f(\lambda)] < 0.0045$, and less than a 0.01 estimation error (at 95% confidence level), according to Bailey et al. (2015). Another important value is the number of trials (N) of the strategies' selection procedure that will be carried on independently in the 12,870 set of data created with the CSCV procedure. N must be large enough to make the selection procedure representative in its potential added informational value. The number of resultant logits must, therefore, be significantly higher than 10 if a PBO threshold of 10% is retained by the investment researcher.

However, as with every implementation type, the CSCV approach has some limitations and potential misuse. First, the CSCV method may not be suitable for every investment strategy; sometimes the K-FCV may suit better. Also, the division of the data into several subsamples may lead auto-correlated time-series to lose their characterization. In another part, the researcher must not hide any information in trials in order to avoid the file drawer problem. “The test is only as good as the completeness of the underlying information”, accordingly to Bailey et al. (2015). Indeed, hiding trials may lead one to underestimate the probability of overfit, since each logit will be computed under a biased relative rank. Here is the importance to correctly parameterize upstream the strategy selection procedure in order to launch the CSCV procedure only after this stage. Second, this procedure does not evaluate the

correctness of a backtest in regards to assumptions that have been made on it (like the use of incorrect transaction costs or the use of data that were not available at the time of investment decision). Third, the CSCV procedure can take into account structural breaks in the data as long as they are present in the used set. If change in paradigm or change in the underlying's model are not present in the set of data used in the procedure, the latter will not take into account these types of structural breaks. It is, therefore, advisable to use data that is expected to behave in the same model's configuration in future states. Fourth, ending up with a high PBO for a strategy selection procedure among N trials do not mean that no skilful strategies do not exist in the range of trials. Fifth, Bailey et al. (2015) warn the reader that the use of the PBO as a measure to fit some selection process is a *gross misuse of our method*. They add that *any counter-overfitting technique used to select an optimal strategy will result in overfitting*. They also quoted Strathern (1997) who advised, *when a measure becomes a target, it ceases to be a good measure*. To resume, a measure to counter overfitting must not be applied in a fitting process in order to achieve the optimal outcome. To conclude, Bailey, D. J. Borwein, M. López de Prado and J. Zhu added, *We believe that CSCV provides both a new and powerful tool in the arsenal of an investment and financial researcher, and that it also constitutes a nice illustration of our general framework for quantitatively studying issues related to backtest overfitting. We certainly hope that this study will raise greater awareness concerning the futility of computing and reporting backtest results, without first controlling for PBO and MinBTL (Minimum Backtest Length)*.

2 GENERATING AND EVALUATING STRATEGIES: METHODOLOGY

3.1 Review of the Main Technical Indicators

3.1.1 Technical Indicators : A Presentation

Technical indicators are tools used in technical analysis, which is the field of interpreting the market's evolution through price and volume. Technical analysis attempts to predict short-term price evolution and it is antinomic to the weak form of the EMH, according to which past prices do not allow to forecast future prices. Principal technical indicators explored in the early studies mainly involve moving averages, filters, and chart patterns. However, several other indicators that are less popular in literature but more so among practitioners made their appearance through time and some at an early stage, even though they did not undergo several tests. These technical indicators are mainly extracted from open, high, low, and close (OHLC) prices, and are a combination of mathematical formulas that are based in a given logic which gives sense to the indicator. This is why financial charts of practitioners include OHLC components and indicators in order to detect useful information on the asset.

Classical technical indicators can be categorised into different types, namely, trend indicators, oscillators indicators, volatility indicators, and finally volume indicators.

Trend indicators are those that translate the market's dynamic and its direction as either bullish (increasing price) or bearish (decreasing price). They tend to indicate to the investor when to eventually enter into a long or a short position according to the level or the behaviour of the indicator. Trend indicators include, among others, moving averages, the moving average convergence divergence (MACD), Bollinger bands, the directional movement index (DMI), which contains the average directional index (ADX) as its component, and the parabolic stop and reversal (SAR). Oscillators indicators are advanced indicators that seek to detect potential trend reversals or corrections by examining whether the price is overbought or oversold. Generally, tension zones are

detected by the indicator's level, where a high level denotes an overbought zone and a low level an oversold one. Oscillators are effective in both ranging and directional markets. They can be bounded within a given range of values or not limited at all. Oscillators indicators include, among others, the commodity channel index (CCI), the relative strength index (RSI), the stochastic, and the Williams percent range (WPR).

Volatility indicators attempt to measure the market volatility by observing price changes in a given time period. They translate the nervousness and the uncertainty of the actors. The average true range (ATR), Bollinger bands, Donchian channels, and Keltner channels are examples of volatility indicators in the market.

Finally, volume indicators include in their computation the volume of the underlying asset in addition to the price. They express the strength of market moves and the investor's psychology. The higher the volume and movement's strength, so is the certainty of the direction of the market. Low volumes indicate breathlessness and a potential correction. Volume indicators must include the price to have any predictive power. Volume can be used as a confirmation parameter that comes to confirm potential buy or sell signals generated by indicators. Examples of volume indicators are, among others, the accumulation/distribution (AC), the money flow index (MFI), the on-balance volume (OBV), and the force index (FI).

Among these technical indicators, 10 are selected each of which belongs to at least one of class of indicators (trend, oscillators, volatility, volume) except for the volume because of lacking data. Thereafter, a survey will be conducted on each of these indicators. The number of ten is chosen in order to assess a sufficient number of indicators that can represent technical analysis in general. Then, the selection process of indicators is built around two major factors: 1) the familiarity and representativeness of technical indicators in technical analysis and (2) the presence of interesting features in their computation.

A first approach consist to select only principal and main technical indicators that represent an undisputable reference in their own categories (trend, oscillators, volatility) and will therefore act as representative of Technical Analysis.

Fixing the initial population of indicators and restraining them into a smaller sample would still close to the population number in finality, since existing Technical Indicators are not that numerous. This limitation in available indicators reduce the risk of publication bias. Thus, the initial fixed population will consist of 20 indicators : the Average Directional Index (ADX), the Alligator, the Awesome Oscillator (AO), the Average True Range (ATR), Bollinger Bands (BB), Bears Power, Bulls Power, the Commodity Channel Index (CCI), the DeMarker, the Directional Indexes (DI +/-), the Donchian Channel, the Envelopes, the Momentum ,Moving Averages (MA), the Moving Average Convergence Divergence (MACD), the Momentum, the Oscillator of a Moving Average (OsMA), the Relative Strength Index (RSI), the Relative Vigor Index (RVI) and finally the Williams Percent Range (WPR).

However, there exists a constraint for several indicators that do have several components. This is the case of indicators that are not oscillators but intervals which consists of two lines framing the underlying's price in a chart. It is the case for the Bolling Bands, the Donchian Channels, and the Envelopes. Also, these indicators need to be adapted in a quantitative manner in order to make the strategies viable.

From the list above, the purpose is to firstly withhold the known and representative indicators that are : Moving Averages, the CCI, Bollinger Bands, the RSI and the Stochastic. These are main and general indicators known by practitioners in the market.

All others can be chosen by looking at the composition of their computation if they include some attractive features such as simple moving averages or more complex structures that would make sense.

After assessing those features, 10 selected indicators will be finally chosen: the Bears Power, the Bulls

Power, the Commodity Channel Index (CCI), the DeMarker, the Momentum, the personalized Moving Average Convergence Divergence adapted to close price (MACDp), the Percent Bandwidth (%B) implying the Bolling Bands, the Relative Strength Index (RSI), the Relative Vigor Index (RVI) and the Stochastic.

There will be since now no change in the composition of this panel and a survey will be conducted on their ability or not to outperform the market. In what follows, a more precise presentation is made for each retained indicator by detailing their history, the way they may be used and their computation.

3.1.2 Bears Power

Bears power is an oscillator indicator that attempts to determine the strength of the sellers in the market and, particularly, the capacity of bears forces to push prices below the moving average. It was created by Alexander Elder, a USSR-born American trader, in 1989 as a part of his complex Elder-Ray indicator, which metaphorically attempts to view the market 'like an X-ray machine'. Next to bears power, both bulls power and an N-period exponential moving average on the close price are part of the Elder-Ray indicator. The more negative the bears power, the higher the strength of sellers in the market. It is, in fact, a very simply computed indicator, the Low price minus a N-period exponential moving average (EMA) on the close price. The bears power formula is then:

$$\alpha = 2/(n + 1)$$

$$EMA_t(Close, n) = \begin{cases} Close_1 & , \quad t = 1 \\ ((Close_t - EMA_{t-1}) \times \alpha + EMA_{t-1}) & , \quad t > 1 \end{cases}$$

$$\mathbf{Bears\ Power}_t = \mathbf{Low}_t - \mathbf{EMA}_t$$

3.1.3 Bulls Power

Bulls power, like bears power, is an oscillator indicator that determines the implication of buyers in the market. Also introduced by Alexander Elder as part of his Elder-Ray indicator in 1989, the more positive it is, the greater the advantage of the buyers over sellers. Like bears power, this indicator too is computed simply: high price minus a N-period EMA on the close price. The formula is then:

$$\alpha = 2/(n + 1)$$

$$EMA_t(Close, n) = \begin{cases} Close_1 & , \quad t = 1 \\ ((Close_t - EMA_{t-1}) \times \alpha + EMA_{t-1}) & , \quad t > 1 \end{cases}$$

$$\mathbf{Bulls\ Power}_t = \mathbf{High}_t - \mathbf{EMA}_t$$

3.1.4 Commodity Channel Index

The commodity channel index (CCI) is also an oscillator, but which accounts simultaneously for both buying and selling forces. Developed by Donald Lambert in 1980 and highlighted in the *Commodities* magazine in the same year, the CCI is an unbound indicator that measures the difference between the current price and the historical average price. The indicator oscillates around a zero line that represents the historical average price and levels as high as 100 or as low as -100 generally signify respectively overbought and oversold levels, where a potential reversal signal could be considered possible. Initially developed for commodities, as its name suggests, it has been gradually extended to all asset classes. The formula for the CCI is:

$$\mathbf{Typical\ Price}_t = \frac{\mathbf{High}_t + \mathbf{Low}_t + \mathbf{Close}_t}{3}$$

$$SMA_t(\text{Typical Price}, n) = \frac{1}{n} \sum_{i=1}^n \text{Typical Price}_{t-i}, \quad \forall t \geq n$$

$$\text{Mean Deviation}_t = \frac{1}{n} \sum_{i=1}^n |\text{Typical Price}_{t-i} - SMA_t|, \quad \forall t \geq n$$

$$CCI_t = \frac{\text{Typical Price}_t - SMA_t}{0.015 \times \text{Mean Deviation}_t}, \quad \forall t \geq n$$

3.1.5 DeMarker

The DeMarker Indicator (DeM) is an oscillator that compares the most recent to the previous high and low prices, in order to evaluate market demand. Created by Thomas De Mark, a special consultant and the founder and CEO of DeMark Analytics LLC, the indicator, as all oscillators, tends to determine overbought and oversold levels according to a historical average on the price. Ranging from 0 to 1, the values of 0.3 and 0.7 usually represent key levels while 0.5 represents a status quo. It is considered a 'leading indicator' in that it may anticipate imminent changes in the market. It includes a risk component in its interpretation whose value above 0.60 indicates lower volatility and risk, while a level below 0.40 signals increasing risk, which is consistent with bearish, more risky markets. The DeMarker indicator is calculated as follows:

$$DeMax_t = \begin{cases} High_t - High_{t-1}, & DeMax_t > 0 \\ 0, & DeMax_t \leq 0 \end{cases}$$

$$DeMin_t = \begin{cases} Low_{t-1} - Low_t, & DeMin_t > 0 \\ 0, & DeMin_t \leq 0 \end{cases}$$

$$SMA_t(DeMax, n) = \frac{1}{n} \sum_{i=1}^n DeMax_{t-i}, \quad \forall t \geq n$$

$$SMA_t(DeMin, n) = \frac{1}{n} \sum_{i=1}^n DeMin_{t-i}, \quad \forall t \geq n$$

$$DeMarker_t = \frac{SMA_t(DeMax, n)}{SMA_t(DeMax, n) + SMA_t(DeMin, n)}$$

3.1.6 Moving Average Convergence/Divergence on Price

The Moving Average Convergence Divergence (MACD) is a trend-following indicator created by Gerald Appel in the late 1970s which translates the distance between two moving averages of a security's price. The indicator is calculated as the difference between the 26-period exponential Moving Average (EMA) and the 12-period EMA, and it measures their convergence when moving averages tend to get closer to each other, signifying the shortness of a trend, or the divergence when the emerging of a trend is in an acceleration phase. In this case, a personalized version of the MACD is used; the minuend Moving Average is replaced by the closing price of a security. This allows to assess the distance between the closing price and its n-period Moving Average, with the advantage of quantifying this value while knowing that Moving Average rules mostly rely on graphical interpretation and do not allow the generation of quantitative strategies. Also, a main and simple trading rule based on Moving Averages is to enter into long and short positions respectively when an upward/downward crossover from the slow MA is observed. This adapted MACD on the Price will allow to generate trading signals for different values of the obtained indicator and differentiate from a lagging MA crossover strategy that can have some delay in the detection of a trend. This also allows to place one of the main and the simplest technical indicator in the panel for generating strategies. The

adapted MACD on Price can be computed as follow:

$$MACDp_t(n) = Close_t - \frac{1}{n} \sum_{i=0}^{n-1} Close_{t-i}$$

3.1.7 Momentum

The Momentum indicator is the measurement of the speed or the velocity of price changes. It shows the difference between today's closing price and the close n-periods ago. It is calculated as the difference or a ratio of these two variables. The Momentum indicator is similar to the Rate-of-Change (ROC) indicator which is simply the Momentum divided by the closing price n-periods ago. When the Momentum shows an absolute change in price, the ROC shows this change in relative terms (such as in percentage). The Momentum signals a continuing uptrend by remaining positive and a sustained downtrend if it remains negative. Crossing up through zero can be seen as a signal to buy; inversely, a crossing down through zero generates a sell signal. The formula is the following:

$$Momentum_t(n) = Close_t / Close_{t-n}$$

3.1.8 Percent Bandwidth

The Bollinger Bands are technical analysis tools defined by a set of trendlines that are plotted two standard deviations away from a simple moving average of a security's price. Developed by the financial analyst John Bollinger in the 1980s, the Bollinger Bands consist of three lines: the middle band is a n-period moving average, the upper band is shifted positively by two standard deviations above the SMA and the lower band is offset negatively by two standard deviations under the SMA. The price hitting the lower band may be interpreted as oversold and can be considered as a buy signal; inversely, the price reaching the upper band can be considered a sell signal. The percent bandwidth (or %b) is an indicator introduced by John Bollinger in 2010 in order to normalize the width of the bands over time and to quantify what began as a graphical indicator. Using this quantified indicator allows to implement one of the most known indicators in the future strategies (as the MACD is applied to the price). The formula of the %b is the following:

$$Middle Band_t(n) = \frac{1}{n} \sum_{i=0}^{n-1} Close_{t-i}$$

$$Upper Band_t(n) = Middle Band_t(n) + 2 \times \sigma_t(n, Close)$$

$$Lower Band_t(n) = Middle Band_t(n) - 2 \times \sigma_t(n, Close)$$

$$\%b_t(n) = \frac{Close_{t-i} - Lower Band_t(n)}{Upper Band_t(n) + Lower Band_t(n)}$$

3.1.9 Relative Strength Index

The relative strength index (RSI) is among the most known and popular of all technical indicators. Developed by J. Welles Wilder and published in his book *New Concepts in Technical Trading Systems* and in *Commodities* magazine in 1978, the indicator's purpose is, as its name suggests, to measure the strength of a price movement and assess whether it is running out of steam. As an oscillator, it allows, as its peers do, the detection of overbought and oversold price levels, calling for some potential trend corrections. A 14-period is used as an indicator by practitioners and recommended by Wilder. The

indicator takes values between 0 and 100 and levels above 70 and under 30 are usually taken to signal overbought and oversold price zones. The formula of the RSI is:

$$\begin{aligned}
 Gain_t &= Close_t - Close_{t-1}, \quad \forall (Close_t - Close_{t-1}) > 0 \\
 Loss_t &= Close_t - Close_{t-1}, \quad \forall (Close_t - Close_{t-1}) < 0 \\
 Average\ Gain_n(n) &= \frac{1}{n} \sum_{i=0}^{n-1} Gain_{n-i} \\
 Average\ Loss_n(n) &= \frac{1}{n} \sum_{i=0}^{n-1} Loss_{n-i} \\
 Average\ Gain_t(n) &= \frac{1}{n} [Average\ Gain_{t-1} \times (n-1) + Gain_t], \quad \forall t > n \\
 Average\ Loss_t(n) &= \frac{1}{n} [Average\ Loss_{t-1} \times (n-1) + Loss_t], \quad \forall t > n \\
 Relative\ Strength_t &= \frac{Average\ Gain_t(n)}{Average\ Loss_t(n)}, \quad \forall t > n \\
 Relative\ Strength\ Index_t &= 100 - \frac{100}{1 + Relative\ Strength_t}
 \end{aligned}$$

3.1.10 Relative Vigor Index

The relative vigour index (RVI) is a technical indicator that measures the strength of a trend by comparing a security's closing price to its trading range. It assesses the conviction of a recent price action and the likelihood that this price action will continue. The Relative Vigor Index (RVI) was first introduced in the January 2002 edition of *Technical Analysis of Stocks and Commodities* magazine by John Ehlers. The concept behind this indicator is that prices tend to close at higher levels than they open during bull trends and close at lower levels than they open during bear trends. The RVI does also contain a RVI signal which is computed by using the simple RVI but this latter will not be employed in this research. The RVI signal allows to generate trading signals when it is crossed by the RVI line. The formula of the simple RVI is:

$$\begin{aligned}
 RVI\ Numerator_t(n) &= [(Close_t - Open_t) + 2 \times (Close_t - Open_{t-1}) + 2 \times (Close_t - Open_{t-2}) + \\
 &\quad (Close_t - Open_{t-3})]/6 \\
 RVI\ Denominator_t(n) &= [(High_t - Low_t) + 2 \times (High_t - Low_{t-1}) + 2 \times (High_t - Low_{t-2}) \\
 &\quad + (High_t - Low_{t-3})]/6 \\
 RVI_t(n) &= \frac{\frac{1}{n} \sum_{i=0}^{n-1} RVI\ Numerator_{t-i}}{\frac{1}{n} \sum_{i=0}^{n-1} RVI\ Denominator_{t-i}}
 \end{aligned}$$

3.1.11 Stochastic

The Stochastic Oscillator is a momentum indicator developed by George C. Lane in the late 1950s that shows the location of the close relative to the high-low range for a given number of periods.

According to Lane, the indicator follows the speed or the momentum of price. It can also be used to identify overbought and oversold levels in order to anticipate a future reversal; this is helped by the fact that the indicator is bounded. The Stochastic is composed by the %K and the %D which is only the 3-period SMA of the %K. The %K of the Stochastic can also be considered as the percent bandwidth of the Donchian Channels since it retakes the highest high and lowest low, which are respectively the Upper and Lower Donchian channels. The formula for the stochastic are the

following:

$$\%K_t(n) = \frac{Close_t - Lowest\ Low_t(n)}{Highest\ High_t(n) - Lowest\ Low_t(n)} \times 100$$

$$\%D_t(m) = \frac{1}{n} \sum_{i=0}^{n-1} \%K_{t-i}(n)$$

3.2 Generating Strategies : Methodology

3.2.1 Full Market Exposure

The purpose is to target a full exposure on the market when developing trading strategies from technical indicators. The reason is to facilitate the generation of strategies that outperform the market by maximizing their exposure time in this latter. Indeed, a partial exposure on the market would considerably reduce the chances of finding strategies that outperform the market in terms of cumulative returns. Also, the introduction of technical indicators in the implementation of trading strategies requires the ability to short the market, since they generate sell signals. Exchange Traded Funds (ETFs) and futures markets on indices are financial instruments that can track stock indices and allow one to enter into short positions on them.

Concretely, Technical Indicators do generate sell and buy signals, especially for oscillators indicators that detect overbought and oversold price zones. Signals generated in overbought and oversold zones suggest that a potential reversal is imminent. However, the problem with seeking a full exposure on the market is that it does not allow the consideration of two different levels for a Buy and a Sell signal. Instead, a unique level that roughly signals a paradigm shift in the trend is required. There is, therefore, a risk to not use technical indicators at their full potential, such as taking extreme values in overbought or oversold conditions.

Besides the Buy/Sell signal, another parameter to consider in technical indicators is the lagging period used in their calculation. Though usually a 14-day indicator, this period can be reduced or raised, thereby impacting the sensitivity and the consistency of the signals they generate. This period is used, for example, to calculate the RSI.

After defining the nature of the two parameters required in the elaboration of trading strategies comes the problem to know which values must be applied to them. A natural and intuitive answer is that one would seek parameters that give the best results in terms of cumulative returns. However, these results can only be assessed in past historical data, which lead to seek values that best fit past historical data. This brings the application of an optimization process regarding the cumulative returns. However, as mentioned earlier, applying an optimization process can lead the strategy to be overfitted, in the sense that results applied on a finite set of data may be the result of chance. For this reason, it can be used a *test set validation* or a *holdout method* in order to reduce the risk of overfitting. Applied by numerous 'Early Empirical Studies' described by Park and Irwin (2004) from 1988 to 2004, it consists to divide the data into two sets; one is used as a training set (in-sample data) where the optimization is performed, and the other as a validation set (out-of-sample data) where the optimized parameters are tested for confirmation. For this survey, a 50% training set and a 50% validation test will be retained in order to achieve an equilibrium in the division of the data. This symmetrical separation will also be necessary when a *Combinatorially Symmetric Cross-Validation* (CSCV) will be practiced for controlling backtest overfitting.

A 50% training set in the data is used in order to extract the two optimal values for each indicator, that renders the highest cumulative return in the in-sample period.

A period range of 14 days is arbitrarily chosen since it represents the period that is the most used and recommended among practitioners. Also, testing period will start from 2 to 14 days since testing for a 1-day period seems to make nonsense. Regarding the second parameter, the purpose is to choose the range of values to be tested that seem the more pertinent. The task becomes even more complicated when looking at certain indicators. For the one being bounded, the ranges of values between these boundaries are taken, and for the ones that are not bounded, the selected range is defined according to values between the maximum and minimum values that a 14-period gives.

For example, if the RSI is optimized, one would end up testing the cumulative return on combinations of strategies that got a period from 2 to 14 (13 values) and levels from 1 to 100 (100 values). This gives a total number of 13 times 100, equalling 1,300 strategies to be tested.

Parameters that lead to the strategy with the best cumulative return in the training sample (IS) are retained and applied in the OOS validation sample. This will allow assess out-of-sample cumulative return regarding to its benchmark. A battery of tests will be applied in order to assess the statistical significance of the results and evaluate if the optimization process has a pronounced tendency to overfit.

3.3 Evaluating Strategies : Methodology

3.3.1 Data

Data from Yahoo Finance is used to conduct a survey on the S&P 500 (code ^GSPC) and the Nasdaq 100 (code ^NDX) indexes from the beginning of May 1988 until the end of April 2020. This set of 32 years of data is double that of the selected number of partitions in the CSCV procedure ($S = 16$, cf. section 3.3.3). This amount of data is also consistent with the *minimum backtest length* (MinBTL) presented by Bailey et al. (2014), which requires a minimum of 12 years of data for thousands of trials (asymptotic curve). The accuracy of Yahoo Finance data is not in question and its provider for U.S. equities and global index historical data is Commodity Systems, Inc.⁴. In September 1999, the latter was identified by Futures Magazine⁵ as having the lowest defect quantity among other data providers.

Data from Yahoo Finance has also been chosen since they provide all Open, High, Low, Close, Volume (OHLCV) prices from the required date from March 1988. Such data were not available in Datastream from Thomson Reuters, or at least not for the desired period. As the risk-free rate, 13-week U.S. Treasury Bills from Yahoo Finance (code ^IRX) are acquired. To run regressions, the risk-free rate from data provided by Kenneth R. French on his website⁶ is used.

Concerning transaction costs, taking costs that are contemporaneous to data period can harden if not suppress any chance to find structural signals from technical indicators in a daily time frame. Also, since priority will be given to a purely statistically and theoretical assessment of abnormal returns in this survey, no transaction costs will be integrated at first for a raw analysis. This framework is consistent with the actual one where the broker Charles Schwab introduced for the first time a zero

⁴ <https://help.yahoo.com/kb/SLN2310.html?redirect=true>

⁵ http://www.csidata.com/?page_id=856

⁶ <https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/>

commission rate on stocks and ETFs in November 2019⁷, followed soon by E-Trade, TD Ameritrade, Interactive Brokers and Fidelity⁸, in which was considered as a ‘race to zero’. The framework is however not consistent with trading costs contemporaneous to the data period and a sensitivity analysis will be carried out according to different levels of transaction costs to assess what impact transaction costs will have on discovered alphas.

3.3.2 Hold-Out Method and Out-Of-Sample Validation

The hold-out-method is the use of a partial set of data (out-of-sample) to validate the performance of a strategy constructed in-sample (training set). Data is divided into two parts, usually 80% and 20% for the training and the validation sets, respectively. The use of 50% data partitioning may strengthen the significance of out-of-sample results since the period of validation is equal to the period of training. The purpose is to evaluate the strategy in the out-of-sample set, according to different performance measures including the risk component, such as the cumulative return, yearly average return, volatility, Sharpe ratio, and the maximum drawdown. An in-sample review can also be conducted in order to compare results from both parts.

3.3.3 Probability of Backtest Overfitting : A CSCV Approach

The probability of backtest overfitting (PBO) of Bailey, Borwein, De Prado, and Zhu (2015) is applied to assess the statistical significance of the optimisation and backtesting approach. The purpose is to evaluate the risk of overfitting the data, when implementing an optimization process that leads to testing hundreds or thousands of strategies to overcome the strategy yielding the best performance measure. A significance threshold is thus chosen to prove or reject the null hypothesis of backtest overfitting and therefore, forsake a strategy generating a PBO value above that threshold. In what follows, the method for applying the CSCV procedure is described, mainly by following the description provided by Bailey, Borwein, De Prado, and Zhu (2015).

First, a matrix \mathbf{M} is formed by collecting the performance series from the N strategies tested, representing all the possible combinations of values from days and entry levels of the technical indicator. In particular, each column $n = 1, \dots, N$ represents a vector of returns over $t = 1, \dots, T$ observations of a strategy applied to a predetermined asset. \mathbf{M} is, therefore, a real-valued matrix of returns of order $(T \times N)$. The only imposed conditions are that:

- i) \mathbf{M} is a true Matrix, i.e. with the same number of rows for each column, where observations are synchronous for every row across the N trials, and
- ii) the performance evaluation metric used to choose the “optimal” strategy can be estimated on subsamples of each column.

For example, if that metric was the Sharpe ratio, one would expect that the i.i.d. Normal distribution assumption could be maintained on various slices of the reported performance.

Second, \mathbf{M} is partitioned across rows into an even number S of disjoint submatrices of equal dimensions. Each of these submatrices \mathbf{M}_s , with $s = 1, \dots, S$, is of order $\left(\frac{T}{S} \times N\right)$.

Third, all combinations C_s of \mathbf{M}_s are formed and taken in groups of size $S/2$. This gives a total number of combinations

⁷ <https://www.cnbc.com/2019/11/21/schwabs-zero-fee-trading-kicked-off-a-price-war-and-could-pay-off-with-the-deal-of-a-lifetime.html>

⁸ <https://www.cnbc.com/2019/10/10/fidelity-joins-the-stampede-to-eliminating-fees-for-online-trading.html>

$$\binom{S}{S/2} = \binom{S-1}{\frac{S}{2}-1} \frac{S}{\frac{S}{2}} = \cdots = \prod_{i=0}^{\frac{S}{2}-1} \frac{S-i}{\frac{S}{2}-i}$$

For instance, if $S=16$, 12,870 combinations are formed. Each combination $c \in C_S$ is composed of $\frac{S}{2}$ submatrices \mathbf{M}_S .

Fourth, for each combination $c \in C_S$:

a) The *training set* J is formed by joining the $S/2$ submatrices \mathbf{M}_S that constitute c in their original order. J is a matrix of order $\binom{T}{S} \binom{S}{2} \times N = \frac{T}{2} \times N$.

b) The *testing set* \bar{J} is formed as the complement of J in \mathbf{M} . In other words, \bar{J} is the $\frac{T}{2} \times N$ matrix formed by all rows of \mathbf{M} that are not part of J also in their original order. (The order in forming J and \bar{J} does not matter for some performance measures such as the Sharpe ratio but does matter for others such as the return maximum drawdown ratio).

c) Form a vector \mathbf{R}^c of performance statistics of order N , where the n th component R_n^c of \mathbf{R}^c reports the performance associated with the n th column of J (the testing set). As before rank of the components of \mathbf{R}^c is denoted by r^c , the IS ranking of the N strategies.

d) Repeat c) with J replaced by \bar{J} (the test set) to derive $\overline{\mathbf{R}}^c$ and \overline{r}^c , the OOS performance statistics and rank of the N strategies, respectively.

e) Determine the element n^* such that $r_{n^*}^c \in \Omega_{n^*}^c$. In other words, n^* is the best performing strategy IS.

f) Define the relative rank of $\overline{r}_{n^*}^c$ by $\overline{\omega}_c = \frac{\overline{r}_{n^*}^c}{N+1} \in (0, 1)$. This is the relative rank of the OOS performance associated with the strategy chosen IS. If the strategy optimization procedure is not overfitting, one should observe that $\overline{r}_{n^*}^c$ systematically outperforms OOS, just as $r_{n^*}^c$ outperformed IS.

g) The logit $\lambda_c = \ln \frac{\overline{\omega}_c}{(1-\overline{\omega}_c)}$ is defined. High logit values imply a consistency between IS and OOS performances, which indicates a low level of backtest overfitting.

Fifth, the distribution of ranks OOS is computed by collecting all the λ_c , for $c \in C_S$. Define the relative frequency at which λ occurred across all C_S by

$$f(\lambda) = \sum_{c \in C_S} \frac{\chi\{\lambda\}(\lambda_c)}{\#(C_S)}$$

where χ is the characterization function and $\#(C_S)$ signifies the number of elements in C_S . Then $\int_{-\infty}^{\infty} f(\lambda) d\lambda = 1$. This concludes the procedure.

The relative frequency $f(\lambda)$ designates the PBO and the principle is then to reject the strategies that have a PBO above a given threshold of 1%, 5%, or 10% depending on the tolerance level, even though a level of 5% is preconised by Bailey, Borwein, De Prado, and Zhu (2013).

3.3.5 Equity Risk Premium Predictability of Fitted Technical Indicators

The PBO allows to put a value on the overfitting risk, and to statistically discard it if the PBO is less than 5%. The statistical significance of the strategy's returns can be appraised, knowing that the Backtest is not overfitted (e.g. the optimal discovery is not a false positive). This is a conditional statistical assessment. This gives credit to the following statistical assessment and it rejects the possibility that positive outcome may be due to chance.

One thing that can be assessed is the equity risk premium predictability of technical indicators used to build the strategies. Christopher J. Neely, David E. Rapach, Jun Tu and Guofu Zhou (2014) who measure the role of technical indicators in forecasting the equity risk premium and Goyal and Welch (2003, 2008) will be the principal references.

Also, framework will be adapted for analysing equity risk premium predictability based on macroeconomic variables for technical indicators as predictors.

The conventional framework for analysing equity risk premium predictability is the following bivariate predictive regression model:

$$R_{t+1} = \alpha_i + \beta_i X_{i,t} + \varepsilon_{i,t+1}$$

In this model, the equity risk premium R_{t+1} is the log-return on a stock market index in excess of the risk-free rate from period t to $t + 1$; $X_{i,t}$ is the predictor (technical indicator value) available at t ; $\varepsilon_{i,t+1}$ is a zero-mean error term. Under the null hypothesis of no predictability, there is $\beta_i = 0$, and the equation will reduce to the constant expected equity risk premium model. Because the theory would imply the sign of β_i , Inoue and Kilian (2004) recommend a one-sided alternative hypothesis to increase the power of tests of predictability. $x_{i,t}$ is defined such that β_i is expected to be positive under the alternative. The test is: $H_0 : \beta_i = 0$ against $H_A : \beta_i > 0$. T-statistics are used to test for the significance of $\hat{\beta}_i$, this latter being the ordinary least squares (OLS) estimate of β_i in the regression. A small difference regarding Neely et al. (2014) is that predictive regression estimation results are shown for Out-of-Sample sets instead of simply a full sample set.

As a robustness check, In-Sample OLS estimates from the prior regression are used where $\{R_s\}_{s=2}^T$ is regressed on a constant and on $\{x_{i,s}\}_{s=1}^{T-1}$ (with T being the last day of the In-Sample data set) in order to forecast the equity risk premium for day $t + 1$ in the out-of-sample data set. The out-of-sample forecast based on an individual technical indicator is given by:

$$\hat{R}_{t+1} = \hat{\alpha}_i + \hat{\beta}_i X_{i,t}$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are the OLS estimates from regressing $\{R_s\}_{s=2}^T$ on a constant and on $\{x_{i,s}\}_{s=1}^{T-1}$.

These forecasts are then compared to the historical average of the forecasts, which is:

$$\hat{R}_{t+1}^{HA} = \frac{1}{t} \sum_{s=1}^t \hat{R}_s$$

This popular benchmark forecast (e.g., Goyal and Welch 2003, 2008; Campbell and Thompson 2008; Ferreira and Santa-Clara 2011) assumes a constant expected equity risk premium ($R_{t+1} = \alpha + \varepsilon_{t+1}$). Goyal and Welch (2003, 2008) show that this historical average of forecasts is a very stringent out-of-sample benchmark; they found that predictive regression forecasts on individual macroeconomic variables typically fail to outperform the historical average. Forecasts are analysed in terms of the Campbell and Thompson (2008) out-of-sample R_{OOS}^2 and the Clark and West (2007) MSPE-adjusted statistic.

OOS statistics are:

$$\begin{aligned} e_{A,t} &= R_t - \hat{R}_t, \\ e_{N,t} &= R_t - \hat{R}_t^{HA}, \end{aligned}$$

$$\begin{aligned} MSPE_A &= \frac{1}{T} \cdot \sum_{t=1}^T e_{A,t}^2, \\ MSPE_N &= \frac{1}{T} \cdot \sum_{t=1}^T e_{N,t}^2, \end{aligned}$$

$$R_{OOS}^2 = 1 - \frac{MSPE_A}{MSPE_N},$$

$$\Delta RMSE = \sqrt{MSPE_N} - \sqrt{MSPE_A},$$

The R_{OOS}^2 statistic measures the proportional reduction in mean squared forecast error (MSPE) of the predictive regression forecast relative to the historical average. It denotes economic significance in the regressions. A R_{OOS}^2 positive indicates that the predictive regression forecast outperforms the historical average in terms of MSFE, and a negative value signals the opposite. Like In-Sample R^2 , R_{OOS}^2 statistics may appear small because there exists a large unpredictable component in returns.

The MSPE-adjusted statistic tests the null hypothesis that the historical average MSPE is equal to the predictive regression MSPE against the one-sided alternative hypothesis that the historical average MSPE is greater than the predictive regression MSPE. It gives: $H_0: MSPE_A = MSPE_N$ against $H_A: MSPE_A > MSPE_N$. The asymptotic critical values for MSPE tests provided by Clark and West (2006) are used to reject or not the null hypothesis. The statistic is calculated by first defining:

$$f_{t+1} = (R_{t+1} - \hat{R}_t^{HA})^2 - [(R_{t+1} - \hat{R}_{t+1})^2 - (\hat{R}_t^{HA} - \hat{R}_{t+1})^2].$$

The vector f_{t+1} is then regressed in a constant and the resulting t-statistic is used to test for a zero coefficient. Reject if this statistic is greater than +1.282 (for a one sided 0.10 test) or +1.645 (for a one sided 0.05 test).

3.3.6 Statistical Significance of Market Outperformance

The outperformance of a strategy on a market index can be measured from the abnormal return it generates in excess of that index. This abnormal return can be assessed by regressing excess market returns to excess returns of the strategy. A simple model developed by Sharpe (1963) for portfolio analysis over the Markowitz's model is the single-index model (SIM) which can measure both the risk and the return of a stock (or a portfolio of stocks). It mainly takes its basis from the Capital Asset Pricing Model (CAPM). Mathematically the SIM is expressed as:

$$\begin{aligned} r_{i,t} - r_f &= \alpha_i + \beta_i (r_{m,t} - r_f) + \epsilon_{i,t}, \\ \epsilon_{i,t} &\sim N(0, \sigma_i^2) \end{aligned}$$

Where $r_{i,t}$ is the return of the stock i at time t ; r_f is the risk-free rate; $r_{m,t}$ is the return to the market index; α_i is the stock's abnormal return; β_i is the responsiveness of the stock to the market return; $r_{i,t} - r_f$ is the excess return of the strategy; $r_{m,t} - r_f$ is the excess market return; and $\epsilon_{i,t}$ are the residual returns which are assumed to be independent and normally distributed with mean zero and standard deviation σ_i .

The key assumptions on which the econometrical model is based on are:

- (1) The error term ($\epsilon_{i,t}$) is zero mean and has finite variance.
- (2) Securities are related through common response to return of market index only. Meaning there by error term for one security is not correlated with error term of any other security: $Cov(\epsilon_i, \epsilon_j) = 0$
- (3) There is no correlation between error term and return on market index: $Cov(\epsilon_i, r_m) = 0$

The SIM assumes that there is only one macroeconomic factor that causes the systematic risk for all stock returns. It assumes too that the return of any stock can be decomposed into the expected excess return of the individual stock due to firm-specific factors, the return of the market due to macroeconomic events, and the unexpected microeconomic events that affect only the firm. The alpha denotes the expected excess return of the individual stock and the beta the sensitivity of the stock to the return of the market.

By applying this model initially proposed for stocks or stock portfolios to strategies generated by technical indicators, the alpha can be interpreted as the expected excess return due to potential indicators market predictability, and the beta to the sensitivity of the strategy to the return of the market. Because returns of technical trading strategies developed in this case are the same in absolute terms as the market returns, since only their signs are changed by introducing short selling (cf. section 3.2), the sensitivity of the strategy to the market returns β the has to be high and close to one. It is also interesting to note that volatility of returns would be exactly the same between strategies returns and market returns since again, returns are the same in absolute terms. This implies that if a strategy performs better than the market index, it does by having the same amount of risk in terms of volatility. However, another notion of risk that can increase for such strategies is the uncertainty of their future performance, which that would place such strategies not of the same risk as the market.

The purpose would be to reject the null hypothesis of no market outperformance for a strategy at a given statistical level if the *p-value* is lower than this level (e.g. 1%, 5% or 10%). The test would be $H_0 : \alpha = 0$ against $H_A : \alpha \neq 0$.

The multiple-testing approach that would increase the chance to find a false positive can also be dealt with t-statistics corrected for multiple-testing but in a model with correlations, in Harvey et al. (2015). Thus, significance of alphas will also be assessed regarding inflated critical values under family-wise error rate (FWER) and false discovery rate (FDR), depending on a given level of correlation among strategies' returns.

Thus, Harvey et al. (2015) argue:

Across various correlation specifications, our estimates show that in general a t-statistic of 3.9 and 3.0 is needed to control FWER at 5% and FDR at 1%, respectively. Notice that these numbers are not far away from our previous estimates of 3.78 (Holm adjustment that controls FWER at 5%) and 3.38 (BHY adjustment that controls FDR at 1%). However, these similar numbers are generated through different mechanisms. Our current estimate assumes a certain level of correlation among returns and relies on an estimate of more than 1,300 for the total number of factor tests.

Tests can be then relied in these statements and more particularly in given critical values adjusted to correlations, that are present in *Table 5.* of Harvey et al. (2015).

This test of market outperformance has again to be brought conditionally on the PBO (cf. section 3.3.3) in order to have more statistical power and to deal with publication bias and multiple testing of factors issues that increase the probability to find false positives. The FWER and FDR of Harvey et al. (2015) accounting for correlation between strategies (cf. section 1.4.2) will complete the PBO assessment of overfitting and will correct for it.

4 FULL MARKET EXPOSURE : ANALYSIS ON S&P 500

4.1 IS Optimization Results: Fitted Parameters and Generated Strategies

4.1.1 Generated Strategies

As described in section 3.2.1, strategies based on technical indicators depend on two optimized parameters. Table 1 shows optimization results on the S&P 500 and the Nasdaq 100. Optimization has been brought in-sample, and results will be used out-of-sample.

Tested days are generally from 2 to 14 for convenience, in accordance with the 14-period generally used by practitioners. The range is reduced to 2 to 7 days or to 7 to 14 days if no positions are taken in the optimization process from the original approach (or only few/insufficient positions are taken). This means that the initial range of 2 to 14 days is not adapted to the indicator for this particular time-series. The range can also be increased if optimization results are 14 days precisely, which means that the initial range may be bounding potential outcomes. The range is then increased by a step of 7 more days. The choice of the range of levels of indicators to be tested is made by looking at the distribution of the indicator with a 14-day period. Levels are chosen so as not to account for extremes and rare values but to include only extrema that are recurrent. Levels of bounded indicators are simply the bounds of these indicators. This process generates a high number of strategies that are the results of different combinations of the tested days and levels. The number of tested days is multiplied by the number of tested levels in order to get the total number of trials in the optimization process.

The RSI for the S&P 500 requires a cut in the tested levels because it presents only few and insufficient signals that do not mark a significant difference from the market. The Percentage Bandwidth (*normalized Bollinger Bands*) requires reducing the range of tested days to 7 to 14, which seems a more appropriate range to optimize the difference in the market (no signals produced from 2 to 14 and from 2 to 7). The DeMarker applied to the Nasdaq 100 requires increasing the range of tested days by 7 since one of the boundaries of the range is the optimization output (14). Concerning the number of tested strategies, the number of 2'000 trials is not surpassed (2'000 trials for the DeMarker on the Nasdaq 100). Only 273 combinations are tested for the Momentum indicator for both markets. This represents the minimum number of trials for an indicator.

It can be observed that the average number of signals through the ten strategies attains a level of few hundreds for the S&P 500 (362 in average) and few dozens for the Nasdaq 100 (an average of 149). It is recalled that a signal accounts for two trade orders, such as one order to close a position and one to open a new one. The change in positions can reach a maximum of 648 for the %B and a minimum of 46 for the Momentum, when applied to the S&P 500 and a maximum of 240 for the Bears power and a minimum of 16 for the DeMarker, when applied to the Nasdaq 100. The average number of trades corresponds to a trade every 22 days for the S&P 500 and a trade every 54 days for the Nasdaq 100, considering 252 trading days in a year and 32 years of data. This is quite low, and this may induce that the strength of the strategy should reside in its short selling capability, more than in its frequency.

Table 1. Tested Parameters and IS Optimization Results for *Technical Trading Strategies*

S&P 500 - From 02/05/1988 to 23/04/2004						
<i>Indicators</i>	Tested Days	Tested Levels	# Tested Strategies	# Signals	Optimal # Days	Optimal Level
BeP	2:14	-45:35	1'053	120	11	18
BuP	2:14	-40:45	1'118	152	12	30
CCI	2:14	-25:20	598	464	12	13
Dmk	2:14	1:100	1'300	482	10	56
MACDp	2:14	-50:50	1'313	66	12	41
Mom	2:14	90:110	273	46	12	108
%B	7:14	1:100	700	648	13	84
RSI	2:7	1:100	600	558	4	77
RVI	2:14	-50:50	1'313	372	8	12
Stoch	2:14	1:100	1'300	718	13	94
Nasdaq 100 - From 02/05/1988 to 23/04/2004						
BeP	2:14	-20:10	403	240	4	3
BuP	2:14	-20:20	520	166	4	6
CCI	2:14	-20:20	520	-	-	-
Dmk	2:21	1:100	2'000	16	17	96
MACDp	2:14	-20:20	533	210	4	5
Mom	2:14	90:110	273	204	3	104
%B	2:14	1:100	1'300	78	4	87
RSI	2:14	1:100	1'300	-	-	-
RVI	2:14	-50:50	1'313	132	11	34
Stoch	2:14	1:100	1'300	-	-	-

Looking at the optimization results, outputs are generated from a range from 3 days (Momentum on Nasdaq 100) to 17 days (DeMarker on Nasdaq 100). It seems that more lower periods are found for the Nasdaq 100. Also, optimization on the Nasdaq 100 generates the market itself for three indicators: the CCI, the RSI and the Stochastic. In these cases, no signals are generated even when reducing or increasing the range of tested days. There are no strategies that outperform the benchmark (Nasdaq 100) in-sample for these three indicators. Such results are given by a dash in Table 1.

Optimum levels also vary widely depending on the indicator. They can range from 3 (Bears Power on Nasdaq 100) to 108 (Momentum on S&P 500). This can be the consequence of the difference in their calculation, or because of the cuts applied to them in order to lower their values and therefore lower the range of tested levels (cf. section 3.1.2 to 3.1.11).

4.1.2 A View on Fitted Technical Indicators

Figures 1.1 and 1.2 show a visual interpretation of Technical Indicators applied to the S&P 500 and the Nasdaq 100, respectively. These graphs assess the congruence of market paradigms in-sample and out-of-sample. It may be noted that indicators that are calculated with ratios tend to stay homogeneous over time and keep the same distribution of values from in-sample to out-of-sample periods for both

Figure 1.1 In-Sample/Out-of-Sample visualization of Fitted Technical Indicators and their Turnover Signal applied to the S&P 500

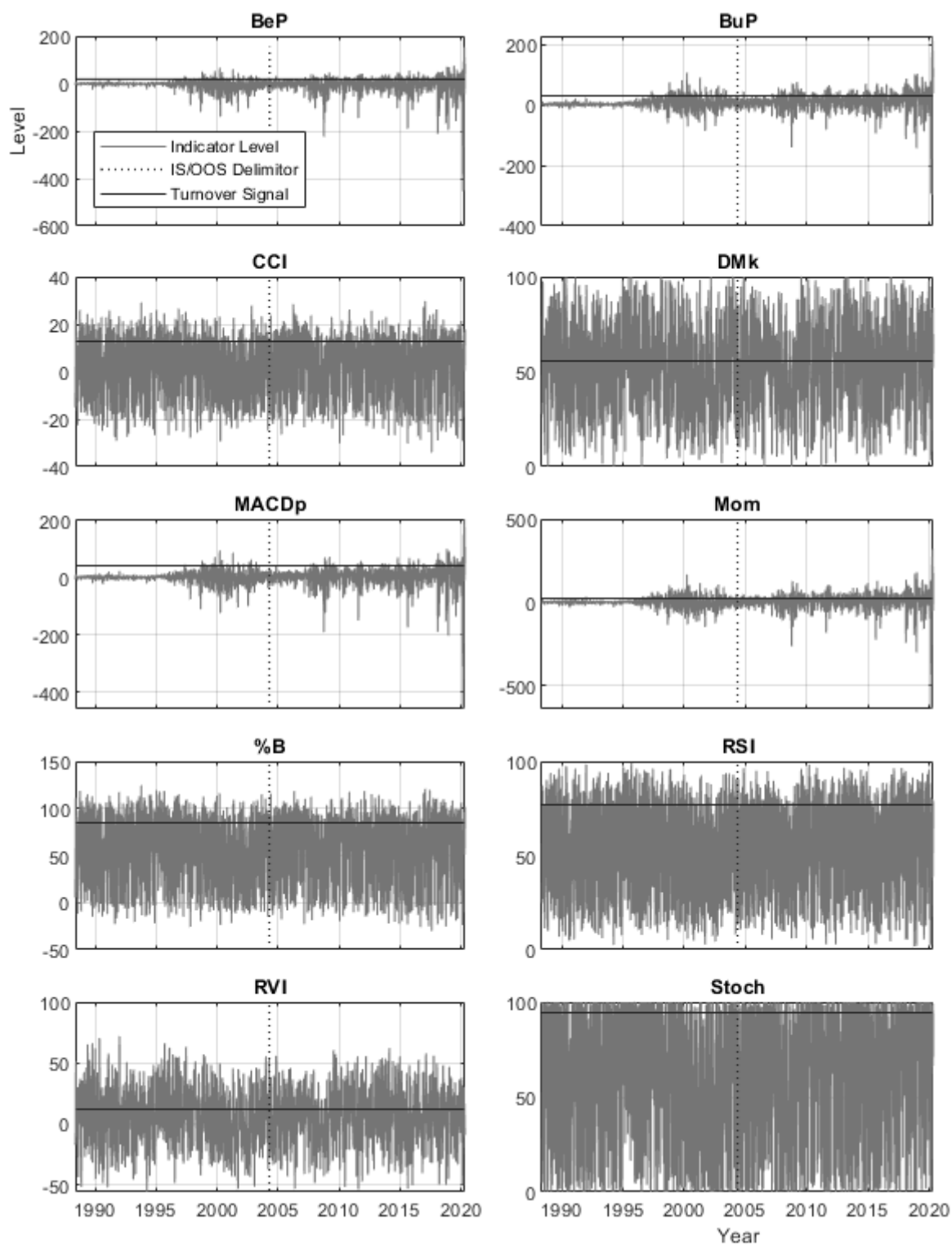
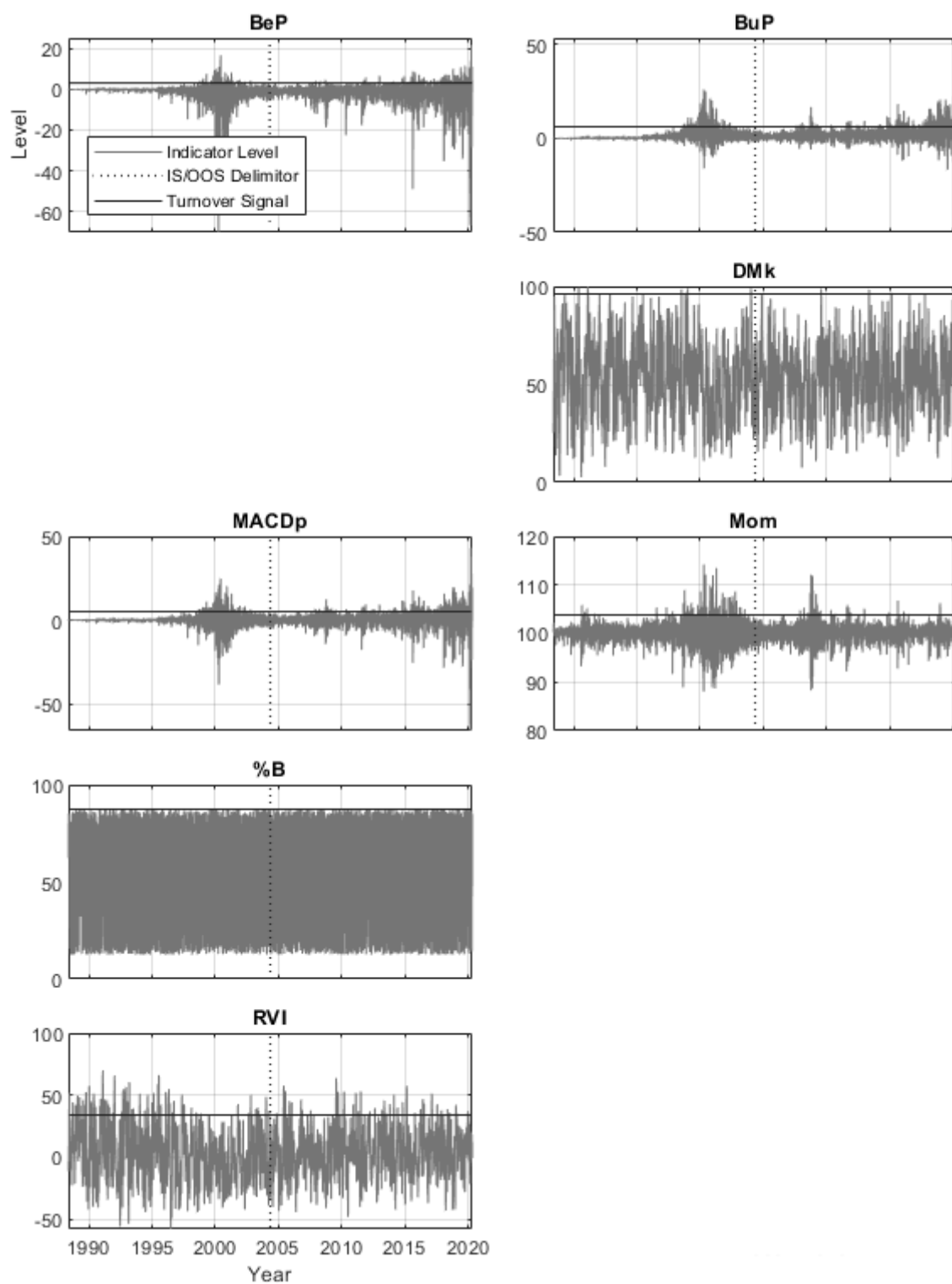


Figure 1.2 In-Sample/Out-of-Sample visualization of Fitted Technical Indicators and their Turnover Signal applied to the Nasdaq 100



the S&P 500 and the Nasdaq 100 markets. Indicators in question here are the CCI, the DeMarker, the Momentum, the %B, the RSI, the RVI and the Stochastic. It may be inferred from that

However, indicators that are not calculated as ratios but from subtractions tend to show changing paradigms in the S&P 500 and the Nasdaq 100. Indicators in question are the Bears Power, the Bulls Power and the MACD applied to the close price. Specifically, the period from 1988 to 1996 tends to show flagrant lower absolute values than the rest of the period from 1997 to 2020. However, similarities in the out-of-sample period from 2004 to 2020 for the S&P 500 can be observed, even if values look to increase through time in absolute terms. The discrepancy looks more obvious and marked in the Nasdaq 100 market. This questions the fact whether an in-sample optimization of an indicator, which is not homogeneous out-of-sample, may lead to congruent results.

However, such indicators may prove useful by detecting spikes in volatility, such as the dot-com bubble in the late 1990s and the financial crisis of 2007-2008. These indicators may provide valuable information when trying to detect signals in the dataset.

Looking at the trading level that signal a shift in the position (either long or short), it is interesting to note that these levels tend to place themselves in an extremum value of the indicators, except for the RVI applied to the S&P 500 which seems to take a central place in the distribution of the RVI. It can also be noted that these levels are positive and recurrent extrema which render a frequent production of signals. This describes the characteristic of a growing stock market where bullish phases tend to be more frequent than bearish ones. The extremity in these signals often allows strategies to remain in long positions, switching to short positions only when these extrema are reached.

4.2 Hold-Out Analysis: IS and OOS Performance Results

4.2.1 Performance Measures

Table 2.1 and 2.2 show the in-sample and out-of-sample performance results of in-sample optimized strategies for the S&P 500 and the Nasdaq 100, respectively. These tables evaluate the continuity and the homogeneity of the in-sample results where the optimization has been taking place on the out-of-sample part where the in-sample parameters are used. For a strategy to be congruent in both data parts, it is expected to conclude with results that are consistent with each other in terms of performance measures. Known as the Hold-out method (described in 3.3.2), it is a simple but controversial method of validation because cross-validation may imply multiple runs on multiple datasets, while the holdout method involves a single run. This issue will be addressed lately with the computation of the PBO.

Also, it has to be recalled that, since contemporaneous transaction costs have not been taken into account, the following out-of-sample performances do not represent what an investor could have earned if he had optimized from 1988 to 2004 and had invested from May 2004 to 2020. It is in fact a pure theoretical view of the potential of technical indicators. A sensitivity analysis will be carried out later to assess the impact of trading costs on the profitability of fitted technical strategies.

First, for the S&P 500 case, it can be observed in Table 2.1 that all indicators except the RVI outperformed the market benchmark in terms of returns and returns-adjusted-to-risk in the out-of-sample part from 2004 to 2020. The maximum drawdown has been reduced for all indicators, and the annual average return can be as high as two times the average return in the market for the %B. Some interesting values are written in bold type to convey their differences according to the benchmark.

Table 2.1 Performance Measures on *Technical Strategies* optimized on the **S&P 500** with respect to Cumulative Return

<i>Strategies</i>	Number Of Signals	Cumulative Return (%)	Max DD (%)	Sharpe Ratio	Annual Av. Return (%)	Volatility (%)
In-Sample - From 02/05/1988 to 23/04/2004						
BeP	120	589	40.94	0.59	19.09	-
BuP	152	706	51.25	0.65	20.80	-
CCI	464	374	41.34	0.45	15.14	-
Dmk	482	461	35.81	0.52	16.90	-
MACDp	66	605	42.61	0.60	19.34	-
Mom	46	513	48.13	0.54	17.84	-
%B	648	332	39.06	0.42	14.17	-
RSI	558	398	36.41	0.47	15.64	-
RVI	372	878	35.65	0.73	22.93	-
Stoch	718	491	44.65	0.53	17.45	-
S&P 500	-	336	49.15	0.59	14.26	16.32
Out-of-Sample - From 26/04/2004 to 30/04/2020						
BeP	345	324	37.47	0.51	13.98	-
BuP	349	109	45.41	0.28	6.90	-
CCI	508	400	44.03	0.56	15.69	-
Dmk	497	58	43.36	0.19	4.26	-
MACDp	193	478	42.31	0.60	17.22	-
Mom	36	197	58.27	0.39	10.35	-
%B	710	589	34.82	0.66	19.09	-
RSI	560	515	41.22	0.63	17.88	-
RVI	348	27	50.78	0.12	2.20	-
Stoch	756	249	47.91	0.44	11.97	-
S&P 500	-	157	56.78	0.35	8.90	19.50

The volatility is exactly the same of the one of the markets since no extra return is created from the benchmark, but instead is simply inverted when entering short positions. It is represented by a dash in the tables. Note the consistency of the strategies to outperform the market out-of-sample and in-sample. It can be observed that, despite the difference in performance of the market in-sample and out-of-sample (14.26% annual average return against 8.90%, respectively), returns of fitted strategies have globally remained the same, which means that the outperformance is even better out-of-sample than in-sample. Sharpe ratios, too, have remained the same (around 0.50 and more). The strategy that has underperformed the market the most is the RVI, which presents a poor cumulative return of 27% from 2004 to 2020, a low Sharpe ratio of 0.12 and a weak annual average return of 2.20%. It should be noted that the RVI looked to be the more promising strategy in-sample, with an average return of 22.93% and a Sharpe ratio of 0.73. The Bulls Power and the DeMarker have also failed to outperform the market.

Looking now to the Nasdaq 100 case in table 2.2, it can be noted that available optimized strategies do

Table 2.2 Performance Measures on *Technical Strategies* optimized on the **Nasdaq 100** with respect to Cumulative Return

<i>Strategies</i>	Number Of Signals	Cumulative Return (%)	Max DD (%)	Sharpe Ratio	Annual Av. Return (%)	Volatility (%)
In-Sample - From 07/05/1985 to 30/11/2000						
BeP	240	781	63.96	0.53	21.78	-
BuP	166	1'915	51.94	0.75	31.25	-
CCI	-	-	-	-	-	-
Dmk	16	489	77.93	0.43	17.41	-
MACDp	210	1'242	43.39	0.64	26.51	-
Mom	204	1'099	52.84	0.61	25.22	-
%B	78	454	75.31	0.41	16.77	-
RSI	-	-	-	-	-	-
RVI	132	479	75.77	0.42	17.24	-
Stoch	-	-	-	-	-	-
<i>Nasdaq 100</i>	-	440	77.93	0.40	16.49	19.72
Out-of-Sample - From 01/12/2000 to 27/07/2016						
BeP	231	363	55.01	0.51	14.90	-
BuP	397	258	44.11	0.43	12.23	-
CCI	-	-	-	-	-	-
Dmk	6	389	55.63	0.53	15.46	-
MACDp	375	410	47.61	0.54	15.89	-
Mom	88	564	36.90	0.62	18.70	-
%B	74	456	53.26	0.56	16.80	-
RSI	-	-	-	-	-	-
RVI	96	322	55.63	0.48	13.93	-
Stoch	-	-	-	-	-	-
<i>Nasdaq 100</i>	-	336	55.63	0.49	14.27	21.14

perform better than their benchmark for the indicators except for the Bulls Power and the RVI. However, this outperformance is of lower value for the S&P 500 case. Thus, the only strategy that seems to generate respectable annual average returns when compared to the S&P 500 is the Momentum indicator, but this excess return still does not reach twice as the one of the market (18.70% vs. 14.27%, respectively) as for the S&P 500.

It must be noted that IS results on the Nasdaq 100 from 1988 to 2004 are more similar than the ones OOS from 2004 to 2020, with a Sharpe ratio of 0.40 against 0.49, respectively. However, even if OOS results are better than one of the benchmarks for 8 strategies on 10, they do not look homogeneous to IS results where cumulative returns of the strategies reached extremely high levels. This is not the case for the validation set. It can be concluded that OOS results obtained for the Nasdaq 100 look poor when compared to IS results and when comparing the case for the S&P 500, even if a majority of strategies beat the market.

Concerning the number of trades, it can be observed that indicators that are computed as ratios are quite congruent from in to out-of-sample. Technical indicators that are calculated as differences between moving averages and prices do not look congruent from Figure 1.1 and 1.2 and they show a

higher number of signals in the out-of-sample dataset.

4.2.2 A view on Fitted OOS Strategies' Performance

Figure 2.1 and Figure 2.2 show the out-of-sample cumulative return of fitted technical strategies applied to the S&P 500 and Nasdaq 100, respectively, in comparison to their benchmark. These figures also visually interpret the performance measures reported in Table 2.1 and 2.2 and provide an overview of the OOS outcomes.

Looking at the case of the S&P 500 in Figure 2.1, it can be observed that the spread in the cumulative return between the strategy and its market benchmark is quite important for six indicators: the Bears Power, the Bulls Power, the CCI, the MACD adapted to price, the %B and the RSI. Two strategies show a lower spread (the DeMarker and the Stochastic), one strategy is close to the market (the Momentum) and the last strategy (the RVI) has an important negative spread.

It can also be noticed that, among the strategies that have a consequent spread, some have a smoothed curve while the others have some irregularities in their progression. The %B, for example, has a smooth progression; a regression curve may almost be drawn. The CCI comes next, followed by the Bulls Power, which presents a few drawdowns from 2018 to 2020. The MACD adapted to the close price also appears quite smooth but presents more bumps and drawdowns than the enumerated previous ones. Even though some present more or less pronounced drawdowns, these strategies tend to smooth their benchmark's drawdowns that are related to the crisis this latter had experienced through time, namely the dot-com bubble in the late 1990s and the financial crisis of 2007-2008. In contrast, the Bears Power is the strategy among the six ones previously listed that presents a larger drawdown between 2013 and 2019. The DeMarker and the Momentum closely follow the market, while the Stochastic is more irregular. The RVI follows the market until 2013, then deviates from it completely, proposing a progression that does not grow.

In regard to the Nasdaq 100 case in Figure 2.2, one may notice that all strategies generated follow the path of the market except for the Bulls Power, which experiences a stall as of 2017. The Momentum strategy is the one that has the largest spread in cumulative return to its benchmark, while the RVI is the strategy that most closely follows the market. Overall, except for the Momentum, fitted strategies show poor performance in regard to their benchmark and may be of little additional value to the simple Buy-and-Hold strategy.

4.3 Probability of Backtest Overfitting

4.3.1 Probability of Backtest Overfitting on Optimization Procedures: Analysis

A computational procedure known as the probability of backtest overfitting (PBO) developed by Bailey et al. (2015) (cf. sections 2.2 and 3.3.3) calculates the odds of a strategy's selection procedure to overfit (i.e., to determine the chances of finding false positives). This addresses the shortcomings and the insufficiency of the precedent holdout method, where a cross-validation is integrated in the PBO computation (CSCV procedure, section 3.3.3). To review, choosing a divisor of 16 gives sixteen different partitions of data that, when being combined symmetrically, end up with C_2^{16} or 12,870 market simulated data samples. These simulated data will then be used to run the procedure that has been used in the holdout method in order to evaluate the propensity of this procedure to lead to overfitting. Having used an Optimization procedure in the holdout method, 12,870 optimizations are

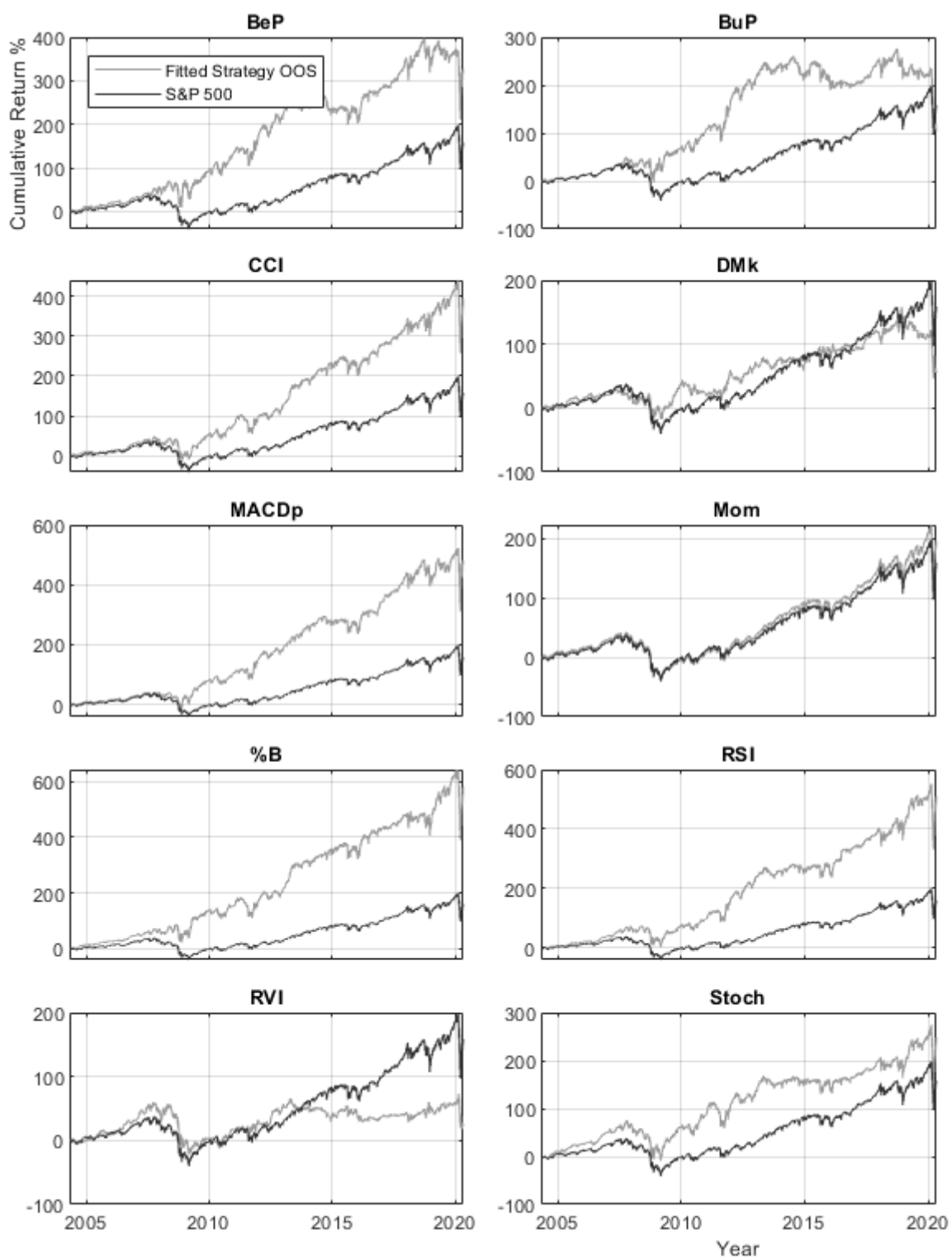
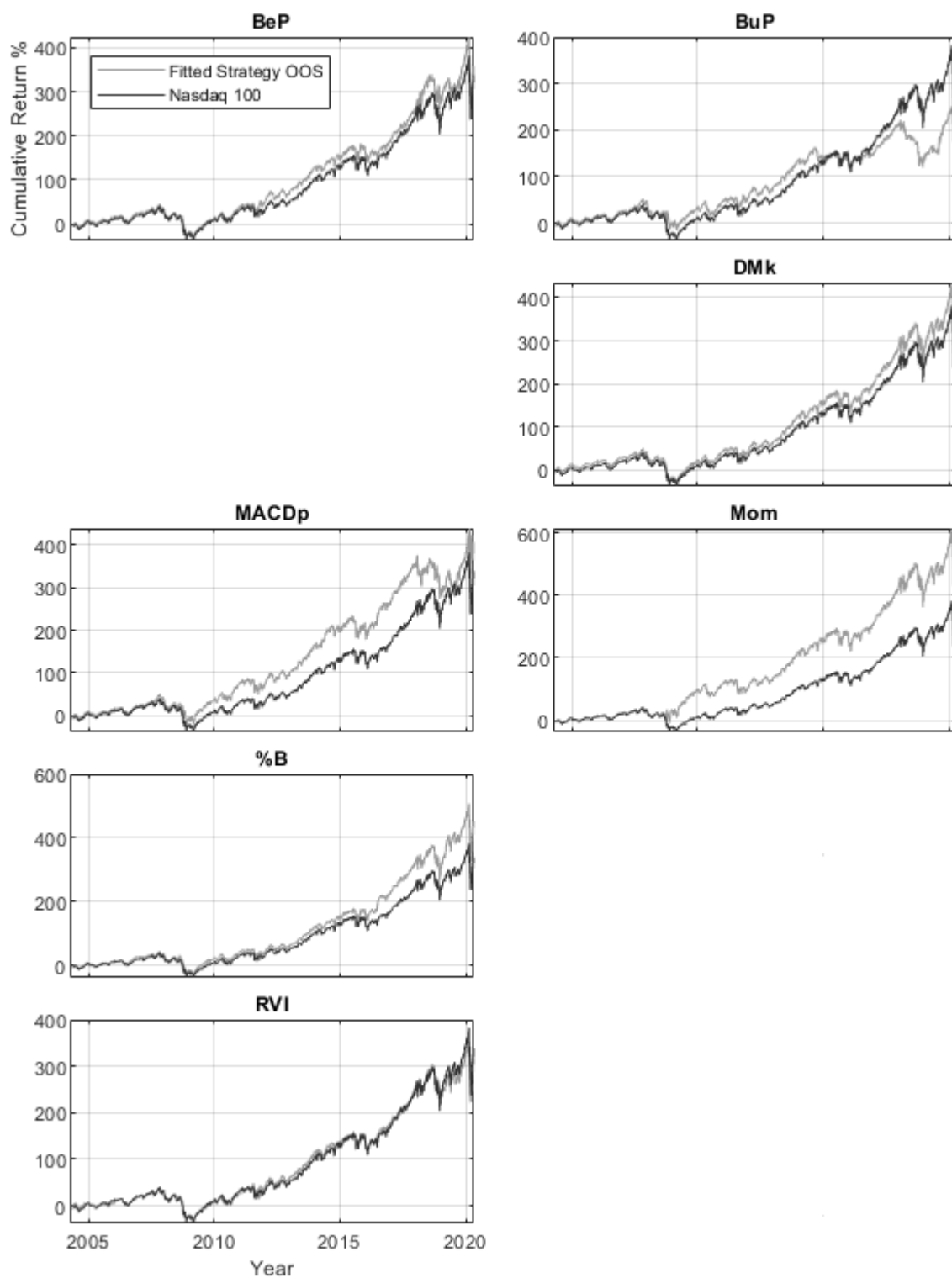
Figure 2.1 Out-of-Sample Cumulative Return of Fitted Strategies applied to the S&P 500

Figure 2.2 Out-of-Sample Cumulative Return of Fitted Strategies applied to the Nasdaq 100

run in all simulated data sets with 50% in-sample and 50% out-of-sample sets. The PBO is the number of times that the OOS cumulative return is lower than the OOS median cumulative return of all trials that have been brought depending on the indicator's parameters (cf. 3.2.1). Results are reported in Table 3:

Table 3. Probability of Backtest Overfitting in a CSCV Procedure for *Technical Strategies Selection Procedure*

S&P 500 - From 02/05/1988 to 30/04/2020					
<i>Optimization Process</i>	# Trials	Median	# of Failed Tests (Rank \leq Median)	PBO (%)	Prob. of Loss (%)
BeP	1'053	526	57	0.44	1.24
BuP	1'118	559	108	0.84	1.84
CCI	598	299	86	0.67	1.31
Dmk	1'300	600	2'046	15.90	13.96
MACDp	1'313	656	301	2.34	4.41
Mom	273	136	446	3.47	4.39
%B	700	350	354	2.75	5.56
RSI	426	213	719	5.59	6.82
RVI	1'313	656	1'056	8.21	10.15
Stoch	1'300	650	1'053	8.18	11.22

N = 12'870 tests per process

Nasdaq 100 - From 01/12/2000 to 27/07/2016					
<i>Optimization Process</i>	# Trials	Median	# of Failed Tests (Rank \leq Median)	PBO (%)	Prob. of Loss (%)
BeP	403	201	49	0.38	1.09
BuP	520	260	18	0.14	0.19
CCI	-	-	-	-	-
Dmk	2'000	1'000	199	1.55	4.39
MACDp	533	266	94	0.73	1.28
Mom	273	136	67	0.52	1.24
%B	1'300	650	897	6.97	8.92
RSI	-	-	-	-	-
RVI	1'313	656	42	0.33	1.79
Stoch	-	-	-	-	-

N = 12'870 tests per process

Table 3 shows that, among all 12,870 tests practiced on each indicator for both the S&P 500 and the Nasdaq 100 markets, the number of failed tests (OOS cumulative return \leq OOS Median cumulative return) remains low. Optimization procedures for the S&P 500 based on 6 out of 10 indicators (Bears Power, Bulls Power, CCI, MACD adjusted to close price, Momentum and Percentage Bandwidth) and optimization procedures for the Nasdaq 100 based on 6 out of 7 indicators (Bears Power, Bulls Power, DeMarker, MACD adjusted to close price, Momentum and RVI) present a PBO less than 5%, rejecting the null hypothesis of backtest overfitting at a 5% significance level.

Moreover, 3 of 6 indicators for the S&P 500 (Bears Power, Bulls Power and CCI) and 5 of 6 indicators for the Nasdaq 100 (Bears Power, Bulls Power, MACD adjusted to close price, Momentum and RVI) give a PBO less than 1% to their optimization procedures, rejecting the null hypothesis of backtest overfitting at the 1% significance level.

With this in mind, it can be interpreted that the optimization procedures described in 3.2.1 (and in the cases enumerated previously) do not lead to backtest overfitting, and that potential discoveries (statistically significant abnormal returns) are not likely to be false positives, but true discoveries.

In regard to the PBO intrinsic values, these studies show that the PBO can reach values as low as 0.11% for the Bulls Power optimization applied to the S&P 500 and 0.14% for the Bulls Power procedure applied to the Nasdaq 100. The highest PBO observed is for the procedure based on the DeMarker applied to the S&P 500, which reaches a level of 15.90% with 2'046 failures on 12'870 tests. Then comes the PBO on the RVI and the Stochastic which reaches levels of 8.21% and 8.18%, respectively. The highest value for the Nasdaq 100 is the percentage bandwidth, which reaches a level of 6.97%, with 897 failures on 12,870 tests.

Another measure that can be assessed is the Probability of Loss presented in Table 3, which is the probability that the strategy selected as being optimal in-sample will deliver an out-of-sample loss. This loss, if it occurs, must be considered as occurring for a reason other than overfitting. It can be seen that the majority of PL follow the PBO value, but by being slightly higher than this latter. It can also be observed that the majority of PBO under 5% present a PL under 5%; this excludes the %B applied to the S&P 500, where the PL reaches a value of 5.56% for a PBO of 2.75%. Overall, the PL seems to remain low across optimizations, with a maximum value reaching 12.75%. It can be interpreted that the null hypothesis of finishing with a loss (by applying such an optimization procedure for indicators that show a PBO less than 5%) can be rejected at the 5% significance level. This does not apply at the 1% significance level, except for the Bulls Power applied to the Nasdaq 100.

The number of failed tests can also be represented by logits, as presented by Bailey et al. (2015) and where their computation is available at section 2.3. These logits are the distribution of relative ranks of out-of-sample cumulative return, and they translate failed tests by values being less than or equal to zero. All positive logits represent relative ranks that have successfully passed the test of being above the median OOS rank (which also represents the rank of the OOS median cumulative return). The visual representation of logits, according to histograms, allows one to have a better overview of the dispersion of the distribution of these logits that *convey relevant information regarding the robustness of the strategy selection procedure*, according to Bailey et al. (2015). A robust strategy selection procedure must lead to a consistent OOS cumulative return ranking where this is represented by similar logits. Also, if the relative ranks are distributed close to uniformly, the distribution of the logits will approximate the standard Normal distribution with a 0 mean and a standard deviation of 1. This is the case when the backtest lacks information and does not detect any signal in the data. According to Bailey et al. (2015), *if good backtesting results are conducive to good OOS performance, the distribution of logits will be centered in a significantly positive value, and its left tail will marginally cover the region of negative logit values, making $\varphi \approx 0$* , where φ is the PBO.

This is what can be observed in Figure 3.1 and Figure 3.2, which represent the PBO logits of optimization procedures applied to the S&P 500 and the Nasdaq 100, respectively. It can be seen, as expected from results presented in the Table 3, that the distribution of logits is more or less centred in a positive value for the majority of implied indicators, and that the left tail of the distributions is marginally spreading negative values. In general, it is the covering of this negative values region and the skewness of the distribution that give an insight into what will be the PBO (more than the mean of the distribution).

Figure 3.1 PBO Logits of Optimization Procedures applied on the S&P 500

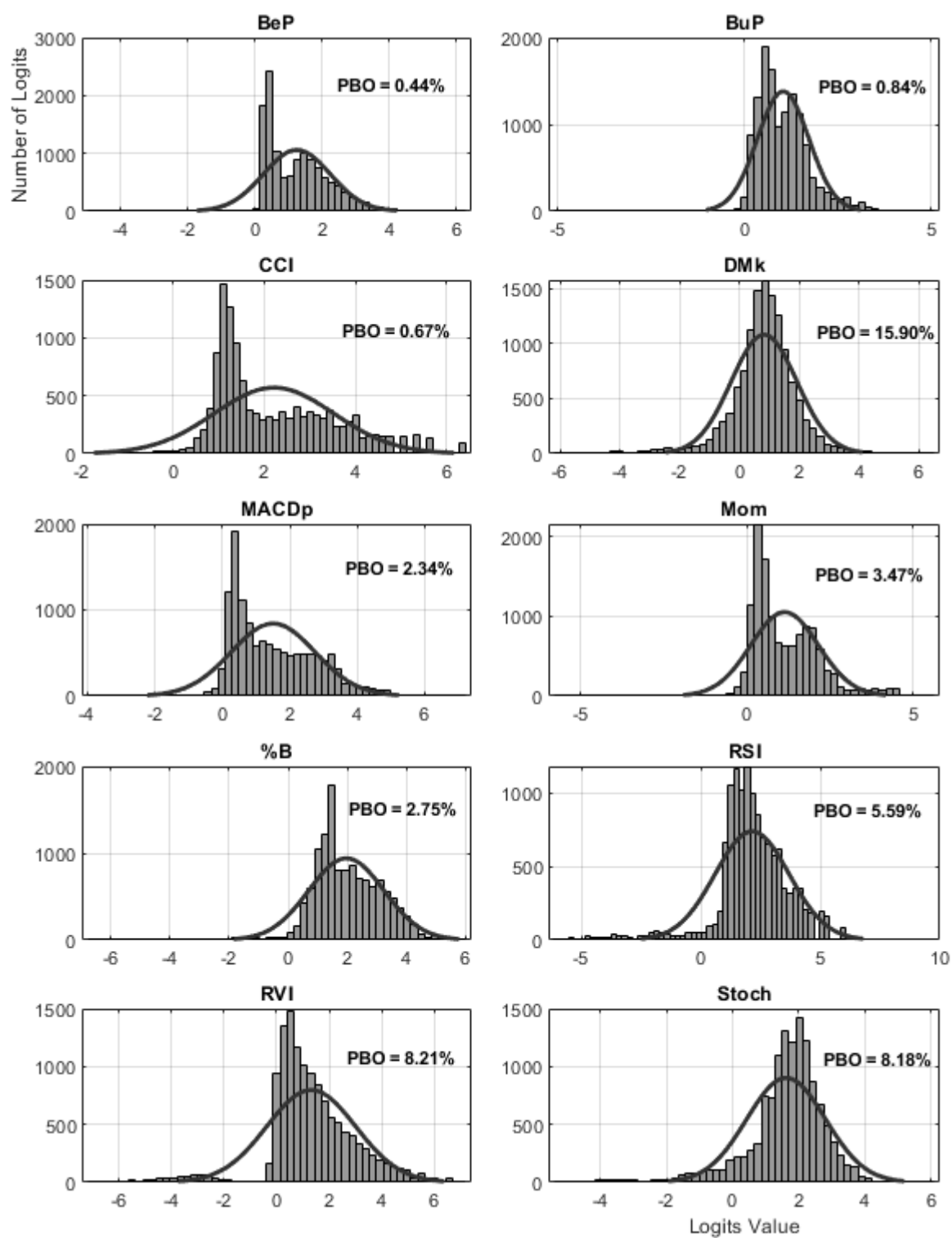
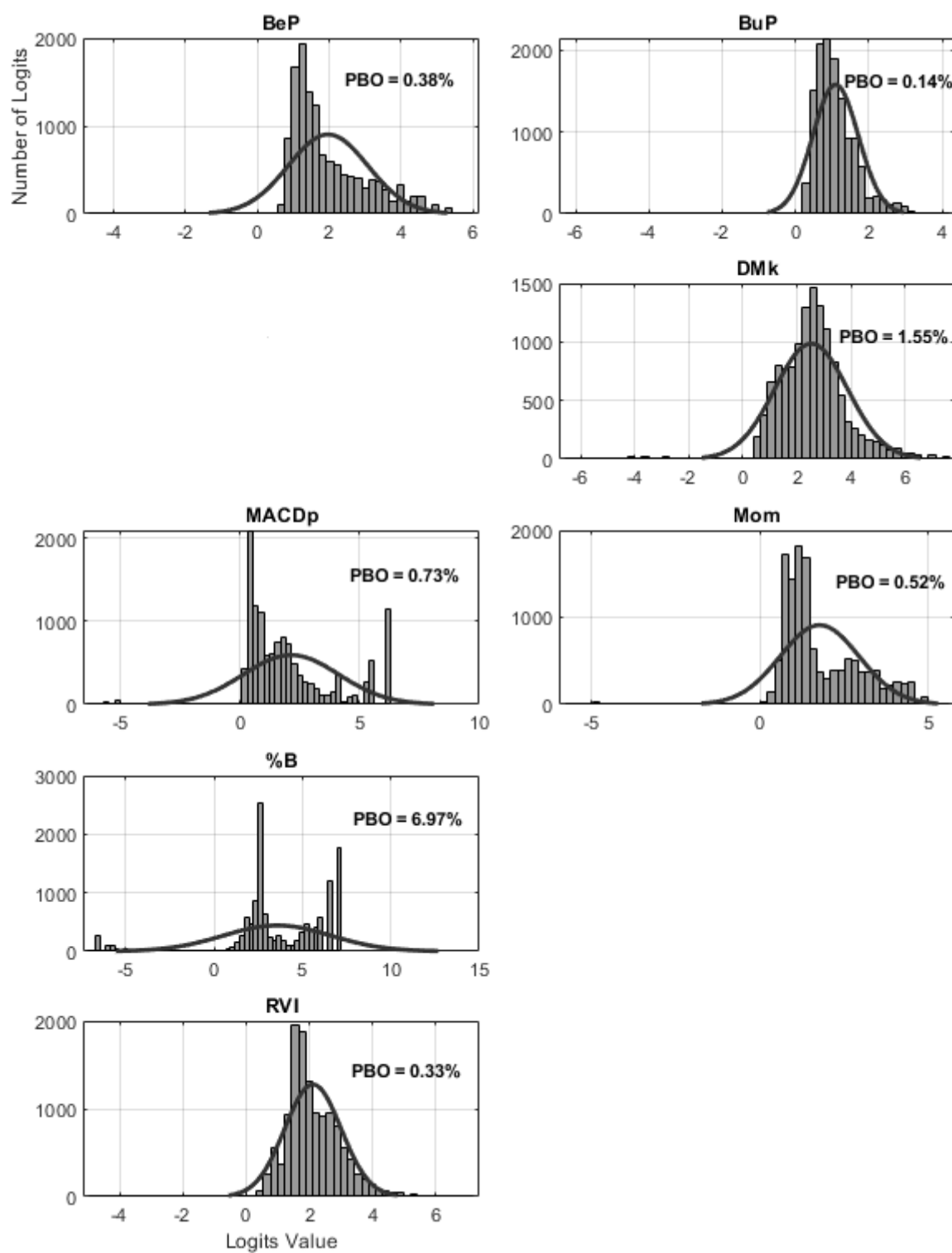


Figure 3.2 PBO Logits of Optimization Procedures applied on the Nasdaq 100



logits take an important place in the left tail of the distribution. Again, the 6.97% PBO for the %B on the Nasdaq 100 also shows extreme negative values. It may also be noted that all optimization procedures that present a PBO of less than 5% are positively skewed; on the contrary, PBO above 5% have a tendency to be negatively skewed. This is less obvious for the RVI and the RSI cases for the S&P 500 and for the %B case for the Nasdaq 100 where their distributions differ and where the PBO of the RSI applied to the S&P 500 is close to 5% (5.59%).

4.3.2 Performance Degradation : Analysis

The performance degradation *determines to what extent greater performance IS leads to lower performance OOS* according to Bailey et al. (2015). Figure 4.1 and Figure 4.2 show the performance degradation of technical strategies fitted on the S&P 500 and on the Nasdaq 100. These figures plot out-of-sample cumulative returns as a function of in-sample cumulative returns obtained in the CSCV procedure (12,870 values). The PBO computation looks similar to the way OOS performance appears when compared to overall OOS performance, when performance degradation measures how OOS performance behaves in regard to IS performance. Like the holdout method, it measures the congruence between IS and OOS performance.

The first thing that may be noted for the two markets is the propensity of strategies to generate high cumulative returns by following the market, but also their propensity to issue negative cumulative returns, which are of lower amplitude than the gains. It may also be observed that OOS cumulative returns reach higher level than IS returns for both markets, but at a smaller frequency in the extremes. One can observe that a polynomial regression would fit better than a linear regression, leading one to conclude that extremes findings IS are likely to fail OOS and, inversely, that congruence between IS and OOS is predominantly reached when IS performance attains the centre of the distribution of all IS performances. This can be a good way to prevent OOS failures: by checking if the optimized IS cumulative return is not that of an extreme of the IS overall distribution, in a CSCV procedure. This can also be explained by the fact that the relationship between IS and OOS performance seems to be negative for all cases since a linear regression would lead to a negative slope. This means that, at some point, looking for the optimal performance becomes detrimental.

It can also be observed that the Momentum, the %B, the RSI and the Stochastic applied to the S&P 500 have a distribution more pronounced in the negative cumulative return area, and that, in general, losses are less represented in the Nasdaq 100 than in the S&P 500.

4.3.1 Stochastic Dominance : Analysis

Stochastic Dominance determines whether the procedure used to select a strategy IS (optimization) is preferable to a randomly selected strategy among the N alternatives. This analysis seeks to answer whether it would make more sense to optimize or randomly parameterize a technical indicator. The test consists of seeing if the distribution of OOS fitted cumulative returns stochastically dominates the entire set of OOS performances in the CSCV procedures (the number of all OOS cumulative returns in a CSCV procedure can be obtained by multiplying 12,870 by the number of N trials of a strategy). Looking back to the notations from section 2.3, and recalling that \mathbf{R} represents the vector of OOS ranks of the N trials specific to a technical indicator and $\bar{\mathbf{R}}_{n^*}$ is the OOS rank of the optimal strategy selected, the first-order stochastic dominance is verified if $Prob[\bar{\mathbf{R}}_{n^*} \geq x] \geq Prob[Mean(\bar{\mathbf{R}}) \geq x]$ for all x , and for some x , $Prob[\bar{\mathbf{R}}_{n^*} \geq x] > Prob[Mean(\bar{\mathbf{R}}) \geq x]$, according to Bailey et al. (2015).

Figure 4.1 Performance Degradation of Strategies Fitted on the S&P 500 - IS vs OOS Cumulative Return %

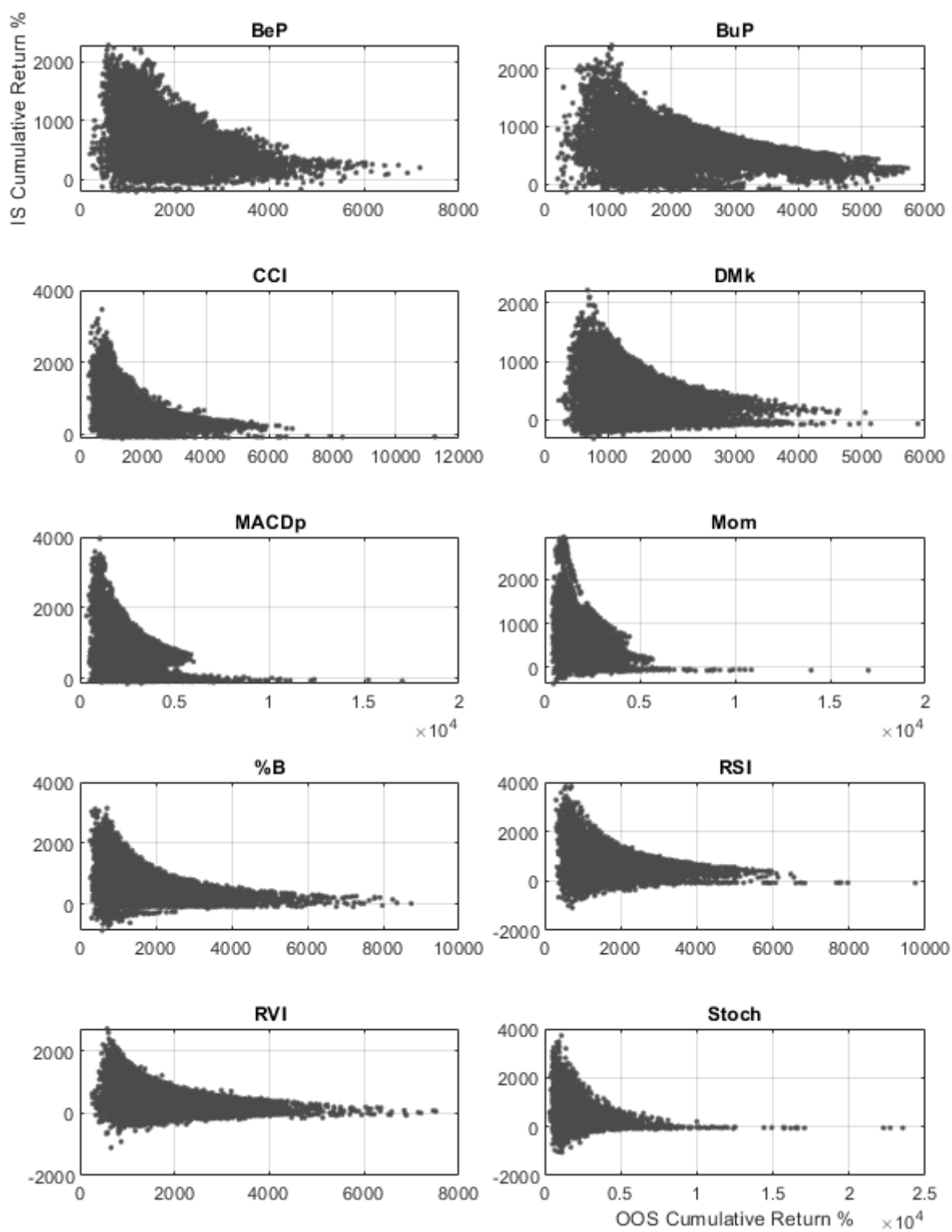


Figure 4.2 Performance Degradation of Strategies Fitted on the Nasdaq 100 - IS vs OOS Cumulative Return %

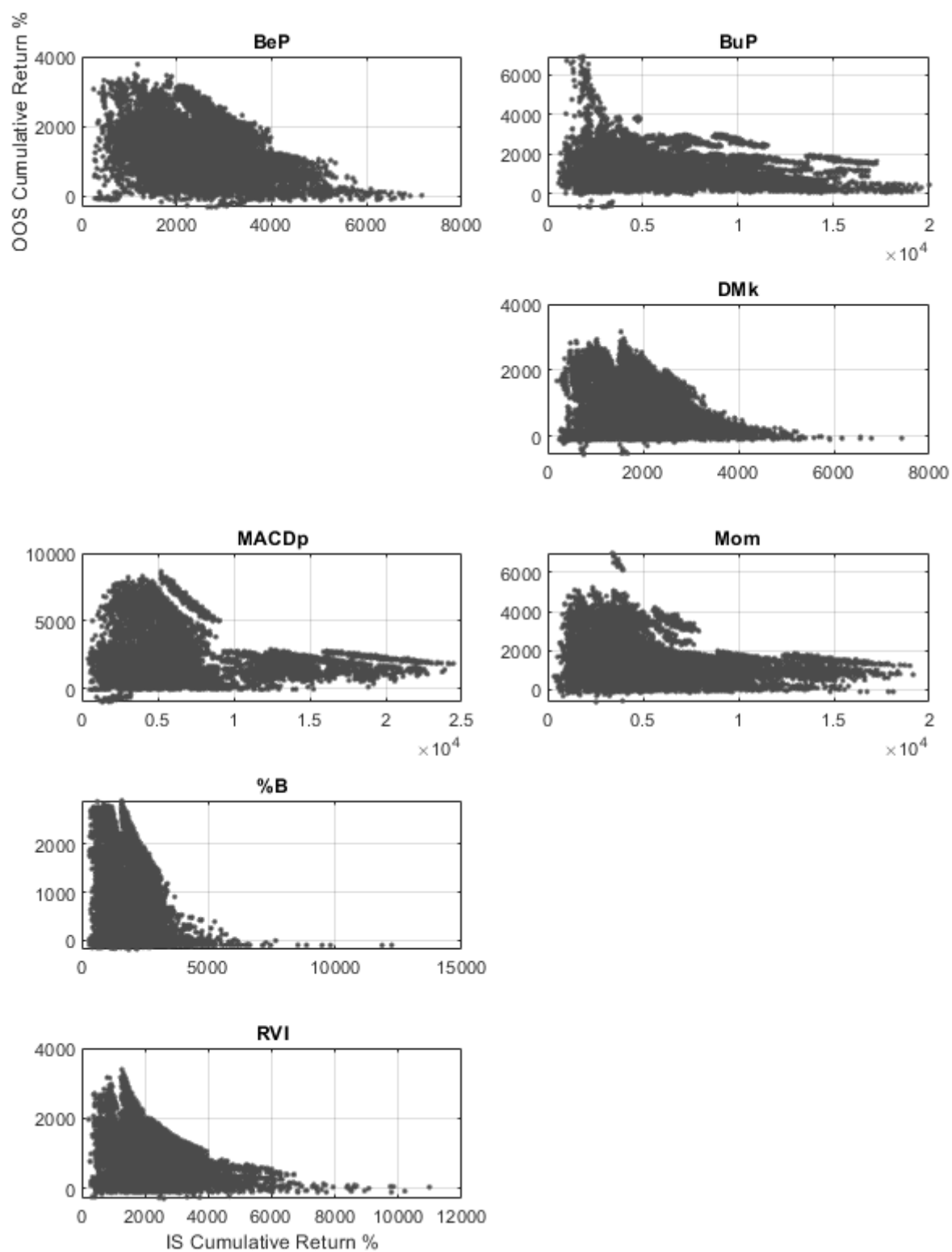


Figure 5.1 Stochastic Dominance of Strategies Fitted on the S&P 500 - CDF of all OOS Cumulative Returns in the CSCV procedure vs CDF of OOS fitted Cumulative Returns

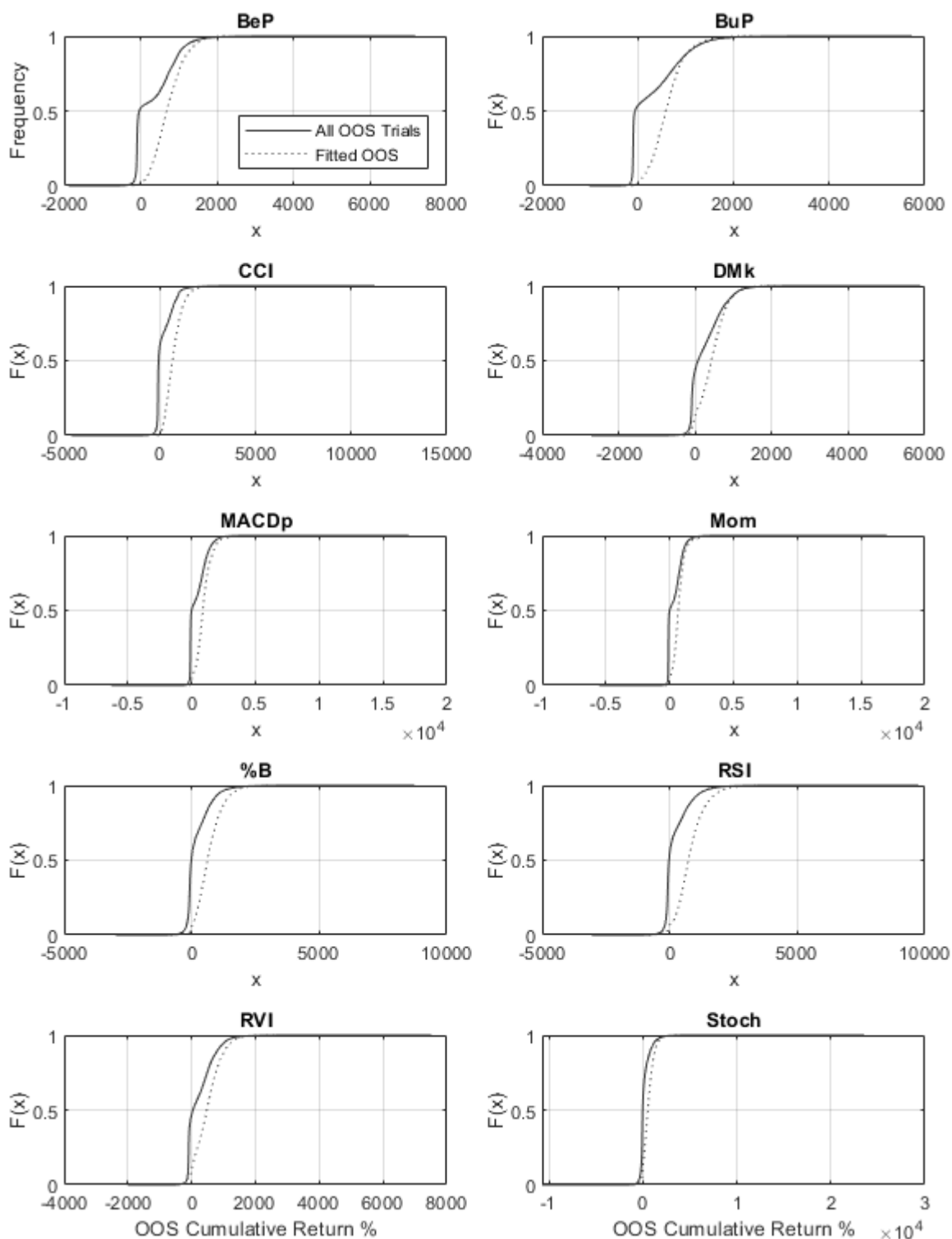
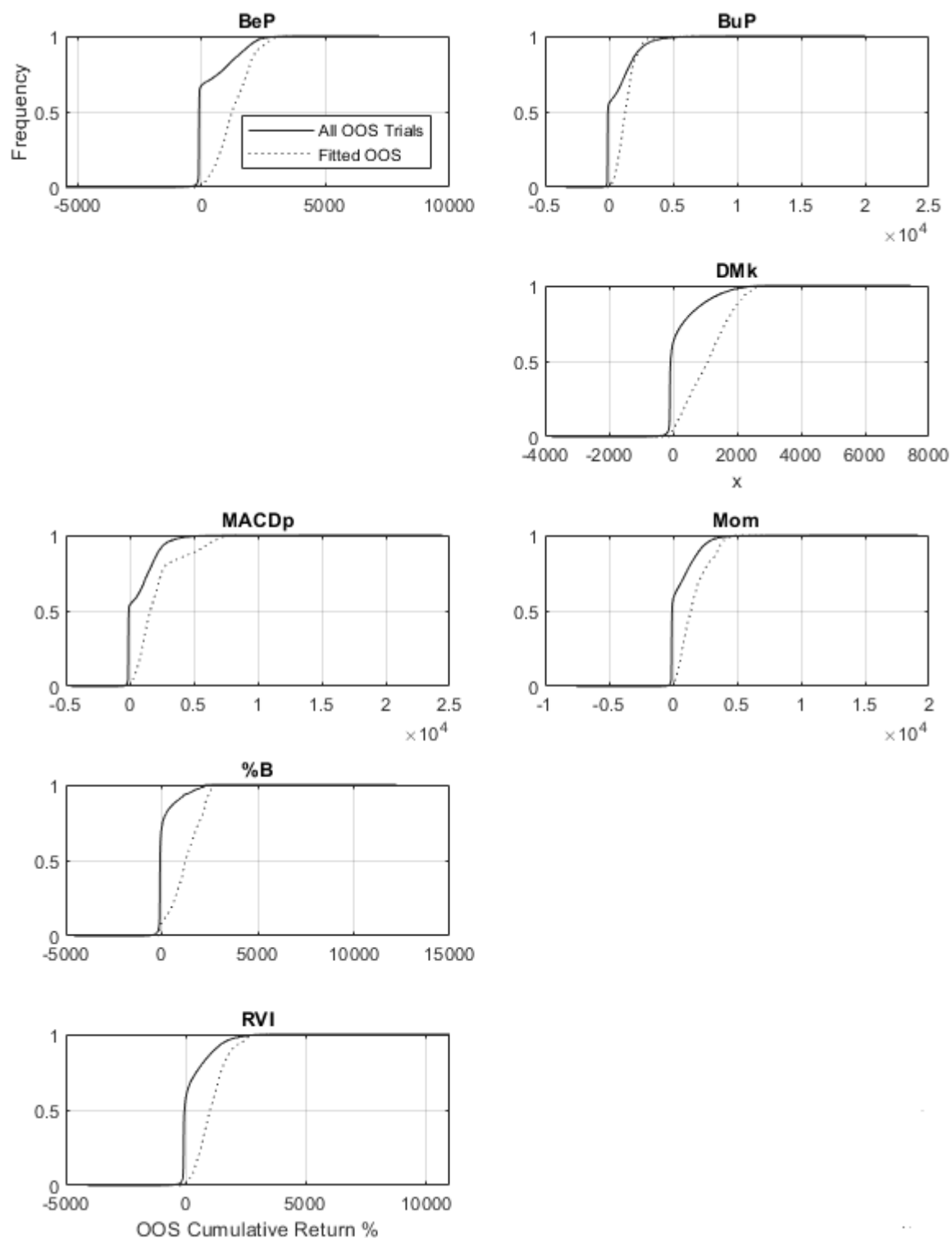


Figure 5.1 Stochastic Dominance of Strategies Fitted on the Nasdaq 100 - CDF of all OOS Cumulative Returns in the CSCV procedure vs CDF of OOS fitted Cumulative Returns



They add: *It can be verified visually by checking that the cumulative distribution function of $\bar{\mathbf{R}}_{n^*}$ is not above the cumulative distribution function of $\bar{\mathbf{R}}$ for all possible outcomes, and at least for one outcome the former is strictly below the latter. Under such circumstances, the decision maker would prefer the criterion used to produce $\bar{\mathbf{R}}_{n^*}$ over a random sampling of $\bar{\mathbf{R}}$, assuming only that her utility function is weakly increasing.*

This is likely as verifying if the CDF of selected OOS cumulative returns is on the right of the CDF of all OOS cumulative returns. Figure 5.1 and Figure 5.2 represent the stochastic dominance of fitted strategies both on the S&P 500 and the Nasdaq 100, respectively. For all cases, it can be observed that all CDF of cumulative returns of OOS optimized strategies are placed on the right of the CDF of cumulative returns for all OOS strategies. It can be concluded that all strategies for both markets verify the first-order stochastic dominance, and that it is preferable to optimize technical indicators in this sense rather than to randomly choose some of their parameters.

Another thing that can be noted is the propensity for OOS strategies to reach a cumulative return around zero for indicators such as the Bears Power, the Bulls Power, the MACD adjusted to price and the Momentum for both markets, and the CCI for the S&P 500 in addition. This gives the illusion that two types of distributions are registered: one around zero, and a new one starting from zero. A possible interpretation may be the relevance of indicators to follow the market in its path at certain precise levels and a uniform distribution for a given range of levels that do not follow their benchmarks.

4.4 Equity Risk Premium Predictability of Fitted Strategies : Analysis

Equity Risk Premium Predictability is assessed here according to Neely et al. (2014) and Goyal and Welch (2003, 2008) for certain parameters. Table 4 shows the results of the assessment of the equity risk premium predictability of fitted technical indicators when applied to the S&P 500 and the Nasdaq 100. It presents out-of-sample predictive regression estimation, such as the OOS beta obtained for the predictor, its student t-statistic and its p-value. As a robustness check, out-of-sample forecasting statistics are reported: the R_{OOS}^2 in percent, the $\Delta RMSE$, and the MSPE-adjusted statistic by Clark and West (2006). Normal critical values used by Clark and West (2006) are reported in the table and are 1.282 for the 10% significance level and 1.645 for the 5% significance level. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively. For the S&P 500, Table 4 shows that four on ten fitted indicators present statistically significant beta at the 1% level (MACD adjusted to price, the Momentum, the RSI and the Stochastic); two indicators (CCI, %B) are significant at the level of 5% and one indicator (Bulls Power) is significant at the 10% level. Regarding the R_{OOS}^2 , all strategies present positive values except for the Bears Power. R-squared positive indicates that the predictive regression forecast outperforms the historical average in terms of MSPE. The $\Delta RMSE$ measures the difference between the square roots of the $MSPE_N$ minus the $MSPE_A$.

Four indicators (CCI, %B, RSI and Stochastic) have a MSPE test statistic that is statistically significant at the 5% level regarding normal critical values used by Clark and West (2007). The null hypothesis of equal MSPE ($H_0: MSPE_A = MSPE_N$ against $H_A: MSPE_A > MSPE_N$) is rejected at the 5% significance level, and these fitted technical indicators outperform the historical average benchmark in terms of MSPE.

Looking at the Nasdaq 100, Table 4 shows that three on seven fitted technical indicators present statistically significant beta at the 1% level (MACD adjusted to price, the Momentum and the %B), and all three have a MSPE test statistic statistically significant at the 5% level, which means that these three fitted technical indicators outperform the historical average benchmark in terms of MSPE again.

Table 4. Equity Risk Premium Predictability of fitted Technical Indicators applied on the S&P500 and the Nasdaq 100

S&P 500 - Out-of-Sample - From 26/04/2004 to 30/04/2020						
<i>Strategies</i>	OOS Predictive Regression Estimation			Forecasting Results		
	β	$ t\text{-stat} $	$p\text{-value}$	R_{OOS}^2 (%)	$\Delta RMSE$	$MSPE\text{-Statistic}$
BeP	-0.000005	0.96	0.336	-0.57	-0.000035	0.28
BuP	-0.000010	1.48	0.138	-0.32	-0.000020	0.50
CCI	-0.000037	1.98**	0.047	0.09	0.000006	1.73**
Dmk	-0.000011	1.27	0.203	0.03	-0.000002	1.04
MACDp	-0.000021	3.83***	0.000	0.25	0.000015	1.22
Mom	-0.000176	3.10***	0.002	0.24	0.000015	1.14
%B	-0.000016	2.53**	0.011	0.13	0.000008	2.04**
RSI	-0.000026	2.95***	0.003	0.12	0.000007	2.62**
RVI	-0.000011	1.20	0.232	0.02	0.000001	1.02
Stoch	-0.000016	2.65***	0.008	0.16	0.000010	2.30**
MSPE-adjusted <i>Normal Crit. Values</i> : 1.282 (10%), 1.645 (5%)						
Nasdaq 100 - Out-of-Sample - From 26/04/2004 to 30/04/2020						
BeP	0.000007	0.17	0.861	-0.04	-0.000002	0.01
BuP	-0.000060	1.10	0.272	-0.16	-0.000011	0.51
CCI	-	-	-	-	-	-
Dmk	-0.000005	0.44	0.659	-0.19	-0.000013	0.37
MACDp	-0.000216	5.08***	0.000	0.19	0.000013	2.16**
Mom	-0.000604	5.10***	0.000	0.03	0.000002	2.54**
%B	-0.000024	2.74***	0.006	-1.19	-0.000079	2.72**
RSI	-	-	-	-	-	-
RVI	-0.000004	0.40	0.692	-0.02	-0.000001	0.38
Stoch	-	-	-	-	-	-
MSPE-adjusted <i>Normal Crit. Values</i> : 1.282 (10%), 1.645 (5%)						

Overall, it can be concluded that the fitted MACD adjusted to price, Momentum, RSI and Stochastic applied to the S&P 500 and the fitted MACD adjusted to price, Momentum and %B applied to the Nasdaq 100 all exhibit statistically and economically significant out-of-sample predictive power for the daily equity risk premium.

4.5 Statistical Significance of Market Outperformance

Market outperformance is measured by the alpha of the regression between excess strategies' returns and excess market returns, following the single-index model (SIM) described in section 3.3.6.

Table 5 presents results of regressions applied on the S&P 500 and on the Nasdaq 100. Significance levels at 10%, 5%, and 1% are denoted by one, two, and three stars, respectively. Critical values are the ones of the t-Student distribution. Critical values in a multiple-testing approach adjusted for correlation $\rho = 0.2$ are presented in Table 5 of Harvey et al. (2015). The baseline is $r = 1/2$, and it corresponds to the fraction of the sample of factors they use to elaborate their model hidden by publishers; it considers statistics from 1.96 and 2.57. This adjustment is made in order to overcome unpublished t-statistics and missing factors.

Table 5. Market Outperformance Analysis of *Technical Strategies* optimized on the S&P 500 and the Nasdaq 100 with respect to Cumulative Return

S&P 500 - Out-of-Sample - From 26/04/2004 to 30/04/2020						
<i>Strategies</i>	β	α	$ t\text{-stat} $	<i>p-value</i>	FWER(5%) $\rho = 0.2$	FDR(5%) $\rho = 0.2$
BeP	0.80	0.00017	1.45	0.147	-	-
BuP	0.67	0.00003	0.20	0.845	-	-
CCI	0.92	0.00017	2.32**	0.021	-	-
Dmk	0.59	-0.00002	0.13	0.894	-	-
MACDp	0.88	0.00022	2.46**	0.014	-	-
Mom	0.93	0.00004	0.65	0.517	-	-
%B	0.87	0.00027	2.83***	0.005	-	-
RSI	0.89	0.00024	2.79***	0.005	-	-
RVI	0.62	-0.00008	0.55	0.583	-	-
Stoch	0.91	0.00010	1.24	0.215	-	-
					3.91	2.27
Nasdaq 100 - Out-of-Sample - From 02/05/1988 to 23/04/2004						
BeP	0.95	0.00002	0.33	0.742	-	-
BuP	0.78	0.00003	0.24	0.807	-	-
CCI	-	-	-	-	-	-
Dmk	1.00	0.00001	1.22	0.223	-	-
MACDp	0.88	0.00008	0.79	0.429	-	-
Mom	0.89	0.00014	1.49	0.136	-	-
%B	0.99	0.00006	1.71*	0.088*	-	-
RSI	-	-	-	-	-	-
RVI	0.97	-0.00001	0.18	0.854	-	-
Stoch	-	-	-	-	-	-
					3.91	2.27

The model with correlation has been chosen in order to account for correlation in the returns of strategies optimized with the same technical indicator. The value of $\rho = 0.2$ has been chosen because it is low enough to be more stringent, even if a higher correlation may be expected from different trials.

For the S&P 500, two fitted strategies (%B and RSI) exhibit a statistically significant daily alpha at the 1% significance level, and two (CCI, MACD) show a statistically significant alpha at the 5% level, according to t-Student critical values. All four strategies are significant at 5% to the statistic that controls the FDR at 5% from Harvey et al. (2015), but do not satisfy the FWER requirement. The Nasdaq 100 shows only one strategy that present a significant alpha at the level of 10%, which is the Percent Bandwidth, but it does not satisfy the FDR and the FWER requirement at 5%.

It can be observed that the value of the beta for all strategies applied to both markets are, as expected, close to one, since excess strategies' returns are only excess market returns adjusted by short positions (depending on bearish signals reported by fitted technical indicators).

Overall, it can be concluded that strategies optimized with the %B, the RSI, the CCI and the MACD adjusted to price, show statistically and economically significant out-of-sample daily abnormal returns on the S&P 500, after accounting for multiple testing.

6 PRACTICAL IMPLEMENTATION AND RISKS

When considering an investment simulation, one would want to consider strategies that present statistically significant results on all tests passed previously, whether it be the PBO, the Equity risk premium predictability or the market outperformance accounting for multiple testing.

The only strategies that meet all the conditions among both studied markets are the strategies that are fitted on the S&P 500 according to the **Commodity Channel Index**, the **Moving Average Convergence Divergence adapted to close price** and the **Percent Bandwidth**.

For this reason, a sensitivity analysis on transaction costs can be carried out in order to assess their impact on trading strategies that have passed all the tests of the survey. Two-sided trading orders are accounted such as closing the position and opening a new one for every trade signal.

Table 6 - Maximum Commission Rate to still Outperform the Market Benchmark for *fitted Technical Strategies* that have passed all the Tests

<i>S&P 500</i>			
<i>Strategies that have passed the tests</i>	CCI	MACDp	%B
α (annualized) %	6.77	8.56	10.46
Max. Commission Rate to Still Outperform the Market %	0.0083	0.0101	0.0123

Table 6. shows the maximum commission rate for strategies that have successfully passed the tests to keep outperforming the market. Therefore, a commission rate of 0.005% is required to approximately divide the alpha generated of the MACDp and the %B by two for example. The issue is that this amount of transaction costs was not available in the market in 2004 (*start of the out-of-sample period*) for most of institutions. Indeed, according to Box, Davis and Fuller (2019), the minimum amount proposed by the six leading fund sponsors during each year from 2004 to 2018 was between 0.01% and 0.02% (10-20 basis points), with an average of 0.04%.

It seems clear that, even the weak form of the EMH has been rejected for the S&P 500, the high number of transactions and the high amount of fees would have not allow an investor to take advantage of the observed abnormal returns.

Also, it is unclear if observed alphas owe their persistence through time to the high amount of transaction costs.

Also, it can be supposed that the emergence and the spreading of current brokers that propose a zero-commission rate for ETFs since 2019, may nullify observed abnormal returns in the future by increasing the efficiency of the market.

Concerning other risks of strategies based on technical analysis or technical indicators in this actual framework, it has already been seen that the volatility would remain the same and that an investor would be exposed to the volatility of the market, but with returns of opposite signs when short selling positions. The betas exposed in Table 5 and gravitating around one give a good insight into the market exposition of the strategies.

Another risk that such strategies based on technical indicators carry is the risk of a changing paradigm in the market. When an in-sample optimization tries to detect structural patterns that tend to persist in an out-of-sample data set of the same length (characterized by a significant alpha), it does not control the risk that such a structural signal may disappear, evolve in another way or progressively decline in the future.

However, an argument that tends to contradict this belief is that a large set of data has been used in this study; exactly sixteen years of daily data for each in-sample and out-of-sample set, which provides a total set of 32 years. This may increase the robustness of the results and sustain the persistence of such signals, knowing that it had continued for 16 years after having been detected from 1988 to 2004. Also, a change in market paradigm would mean a general change of market actors' behaviour, since market prices are driven by their actions. It can be seen that even important macroeconomic events, such as the dot-com bubble in the late 1990s and the financial crisis of 2007-2008, did not changed the behaviour of the retained strategies in the out-of-sample test.

This problem could also be approached by applying a daily rolling optimization that updates the outcome of the optimization to the investment strategy; but knowing that the PBO for a particular indicator on a particular market is less than 5%. Indeed, it can be recalled that the PBO measures the probability of backtest overfitting for a particular strategies selection process that does not depend on a particular set of data, since it is supposed to include thousands of simulated datasets (in this case, 12,870) thanks to the CSCV procedure. This, however, does not take into account the limitations of the CSCV procedure to properly forecast market returns.

CONCLUSION

This research attempted to assess if technical indicators can generate abnormal returns in regard to the

U.S. stock market, and particularly in the S&P 500 and the Nasdaq 100, in a pure statistical and theoretical point of view, without taking account transaction costs. Also, an optimization procedure was set up first in order to find abnormal returns, and this procedure was assessed through the probability of backtest overfitting (referencing studies in 2013 by Bailey et al.) in order to control for backtest overfitting. Second, an assessment of equity risk premium predictability was conducted using Neely et al. (2014) with respect to the MSPE-adjusted test of Clark and West (2006). Third, an analysis was run to detect potential abnormal returns on fitted technical strategies; alphas were controlled for a multiple-testing approach that represents an optimization but were adjusted by correlations among trials for a single indicator. The FDR at 5% of Harvey et al. (2015) with a baseline of $r = 1/2$ and a correlation of $\rho = 0.2$ was used for this purpose.

Referring the PBO of Bailey et al. (2015), results found that 6 on 10 indicators (Bears Power, Bulls Power, CCI, MACD adjusted to price, Momentum and Percentage Bandwidth) applied to the S&P 500, and 6 on 7 indicators (Bears Power, Bulls Power, CCI, MACD adjusted to price, Momentum and RVI) applied to the Nasdaq 100, do not lead to backtest overfitting at the 5% significance level. Moreover, first order stochastic dominance evaluated through a CSCV procedure found that optimization on all ten technical indicators for both markets bring an added value in regard to a simple random parametrization.

Also, 2 out of 6 strategies (CCI and %B) controlled for overfitting for both the S&P 500 and the Nasdaq 100 showed statistically and economically significant out-of-sample predictive power for the daily equity risk premium, according to the MSPE-adjusted test of Clark and West (2007).

Two other indicators (RSI and Stochastic) showed significant out-of-sample predictive power but were not controlled for backtest overfitting.

Finally, 3 out of 6 strategies (CCI, MACD and %B) applied to the S&P 500 and controlled for overfitting showed statistically and economically significant out-of-sample daily abnormal returns after controlling for multiple-testing in a model adjusted to correlations, according to the FDR at 5% of Harvey et al. (2015) with a baseline of $r = 1/2$ and a correlation of $\rho = 0.2$.

The RSI applied to the S&P 500 did not reject the null hypothesis of backtest overfitting but showed some statistically significant daily abnormal returns. No significant abnormal returns were found for strategies applied to the Nasdaq 100.

It can then be asserted that strategies controlled for backtest overfitting and for a multiple-testing approach adjusted to correlations are not of type I error.

Also, results obtained reject the weak form of the EMH that past prices do not allow to predict future prices for the S&P 500.

Overall, results demonstrated that optimization of technical indicators on the S&P 500 and Nasdaq 100 can be considered as a good way to generate trading signals when applied to indicators that do not lead to backtest overfitting. The stochastic dominance, moreover, encourages the optimization of any of the 10 indicators listed above when applied to the specific markets tested (S&P 500, Nasdaq 100), on a daily basis. Further studies can be conducted on other markets and timeframes in order to assess the utility of optimization in particular cases.

Finally, the discovery of positive statistically-significant abnormal returns is in line with discoveries made by *Modern Empirical Studies* on the stock markets mentioned by Park and Irwin (2004), from 1988 to 2004, which is the in-sample period used in this study. To recall, 41 studies have been reviewed across worldwide stock markets; 24 reported positive results, 5 reported mixed results and 12 reported negative results (cf. Table 10; Park and Irwin, 2004).

Concerning the predictability, it is unclear if observed alphas and observed predictability owe their persistency through time to the high amount of contemporaneous transaction costs that would have made an investor incapable to take advantage of them. Also, it can be supposed that the emergence and the spreading of current brokers that propose a zero-commission rate for ETFs since 2019 could nullify such abnormal returns in the future, by increasing the efficiency of the market. This, at the condition that the zero-transaction cost paradigm generalizes through the market and may be accessible for big actors such as pension funds, mutual funds and other financial institutions. It depends also if some of these funds would invest into such technical strategies, a sign that they grant a higher consideration and legitimacy for Technical Analysis in general.

REFERENCES

- Allen, F., and R. Karjalainen. "Using Genetic Algorithms to Find Technical Trading Rules." *Journal of Financial Economics*, 51(1999):245-271.
- Alexander, S. S. "Price Movements in Speculative Markets: Trends or Random Walks." *Industrial Management Review*, 2 (1961) : 7-26.
- Alexander, S. S. "Price Movements in Speculative Markets: Trends or Random Walks No. 2." *Industrial Management Review*, 5(1964):25-46.
- Arnott, R. D. "Relative Strength Revisited." *Journal of Portfolio Management*, (Spring1979):19- 23.
- Bailey, David H. and Borwein, Jonathan and López de Prado, Marcos, and Zhu, Qiji Jim, "The Probability of Backtest Overfitting" (February 27, 2015). *Journal of Computational Finance* (Risk Journals), 2015, Forthcoming.
- Ball, Ray. 1978. "Anomalies in Relationships Between Securities' Yields and Yieldsurrogates." *Journal of Financial Economics* 6, no. 2/3: pp 103-126.
- Beja, A., and M. B. Goldman. "On the Dynamic Behavior of Prices in Disequilibrium." *Journal of Finance*, 35(1980):235-248.
- Bessembinder H., and K. Chan. "The Profitability of Technical Trading Rules in the Asian Stock Markets." *Pacific-Basin Finance Journal*, 3(1995):257-284.
- Bessembinder H., and K. Chan. "Market Efficiency and the Returns to Technical Analysis." *Financial Management*, 27(1998):5-17.
- Bird, P. J. W. N. "The Weak Form Efficiency of the London Metal Exchange." *Applied Economics*, 17(1985):571-587.
- Box, Travis and Davis, Ryan and Petrie Fuller, Kathleen, The Dynamics of ETF Fees (November 1, 2019). Travis Box, Ryan Davis & Kathleen Fuller (2020) The Dynamics of ETF Fees, *Financial Analysts Journal*, 76: 1, 11-18.
- Brock, W., J. Lakonishock, and B. LeBaron. "Simple Technical Trading Rules and the Stochastic Properties of Stock Returns." *Journal of Finance*, 47(1992):1731-1764.
- Burton Malkiel (2011) "A random walk down Wall Street: the time-tested strategy for successful investing"
- Caginalp, G., and H. Laurent. "The Predictive Power of Price Patterns." *Applied Mathematical Finance*, 5(1998):181-205.
- Campbell, John Y. and Thompson, Samuel Brodsky, "Predicting Excess Stock Returns Out of Sample: Can Anything Beat the Historical Average?" (July 2008). *The Review of Financial Studies*, Vol. 21, Issue 4, pp. 1509-1531, 2008.

- Carhart, M. M. (1997). "On persistence in mutual fund performance." *The Journal of Finance*, 52(1):57–82.
- Carr, P. and M. López de Prado, "Determining Optimal Trading Rules without Backtesting", (2014)
- Chang, P. H. K., and C. L. Osler. "Methodical Madness: Technical Analysis and the Irrationality of Exchange-Rate Forecasts." *Economic Journal*, 109(1999):636-661.
- Clark, T., and K.D. West, 2007. "Approximately Normal Tests for Equal Predictive Accuracy in Nested Models." *Journal of Econometrics* 138, pp. 291-311.
- Cooper, M., and H. Gulen. "Is Time-Series Based Predictability Evident in Real-Time?" Working Paper, Krannert Graduate School of Management, Purdue University, 2003.
- Cornell, W. B., and J. K. Dietrich. "The Efficiency of the Market for Foreign Exchange under Floating Exchange Rates." *Review of Economics and Statistics*, 60(1978):111-120.
- Coutts, J. A., and K. Cheung. "Trading Rules and Stock Returns: Some Preliminary Short Run Evidence from the Hang Seng 1985-1997." *Applied Financial Economics*, 10(2000):579- 586.
- Dawson, E. R., and J. Steeley. "On the Existence of Visual Technical Patterns in the UK Stock Market." *Journal of Business Finance & Accounting*, 30(January/March 2003), 263-293.
- Day, T. E., and P. Wang. "Dividends, Nonsynchronous Prices, and the Returns from Trading the Dow Jones Industrial Average." *Journal of Empirical Finance*, 9(2002):431-454.
- Donchian, R. D. "High Finance in Copper." *Financial Analysts Journal*, (Nov./Dec. 1960):133- 142.
- Dooley, M. P., and J. R. Shafer. "Analysis of Short-Run Exchange Rate Behavior: March 1973 to November 1981." In D. Bigman and T. Taya, (ed.), *Exchange Rate and Trade Instability: Causes, Consequences, and Remedies*, Cambridge, MA: Ballinger, 1983.
- Fama, E. F. "Efficient Capital Markets: A Review of Theory and Empirical Work." *Journal of Finance*, 25(1970):383-417.
- Fama, E. F., and M. E. Blume. "Filter Rules and Stock Market Trading." *Journal of Business*, 39 (1966):226-241.
- Feng, Guan hao and Giglio, Stefano and Xiu, Dacheng, "Taming the Factor Zoo: A Test of New Factors" (June 29, 2019). *Fama-Miller Working Paper*; Chicago Booth Research Paper No. 17-04.
- Fernández-Rodríguez, F., C. González-Martel, and S. Sosvilla-Rivero. "On the Profitability of Technical Trading Rules Based on Artificial Neural Networks: Evidence from the Madrid Stock Market." *Economic Letters*, 69(2000):89-94.
- Fernández-Rodríguez, F., S. Sosvilla-Rivero, and J. Andrada-Félix. "Technical Analysis in Foreign Exchange Markets: Evidence from the EMS." *Applied Financial Economics*, 13(2003):113-122.
- Ferreira, M.A., and Santa-Clara, P., 2011, "Forecasting Stock Market Returns: The Sum of the Parts is More than the Whole", *Journal of Financial Economics* 100, 514–537.

Frazzini, A. and Pedersen, L. H. (2014). "Betting against beta." *Journal of Financial Economics*, 111(1):1–25.

Gençay, R. "Optimization of Technical Trading Strategies and the Profitability in Security Markets." *Economic Letters*, 59(1998a):249-254.

Goyal, A., Welch, I., 2003. "Predicting the equity premium with dividend ratios." *Management Science* 49(5), 639–654.

Gunasekarage, A., and D. M. Power. "The Profitability of Moving Average Trading Rules in South Asian Stock Markets." *Emerging Markets Review*, 2(2001):17-33.

Harvey, Campbell R. and Liu, Yan, "Backtesting" (July 28, 2015).

Harvey, Campbell R. and Liu, Yan, "A Census of the Factor Zoo" (February 25, 2019).

Harvey, Campbell R. and Liu, Yan and Zhu, Caroline, "...and the Cross-Section of Expected Returns" (February 3, 2015).

Hudson, R., M. Dempsey, and K. Keasey. "A Note on the Weak Form Efficiency of Capital Markets: The Application of Simple Technical Trading Rules to UK Stock Prices - 1935 to 1964." *Journal of Banking & Finance*, 20(1996):1121-1132.

Houthakker, H. "Systematic and Random Elements in Short-Term Price Movements." *American Economic Review*, 51(1961):164-172.

Ioannidis, J.P.A., "Why most published research findings are false." *PloS Medicine*, Vol. 2, No. 8,(2005) 696-701.

Irwin, S. H., and J. W. Uhrig. "Do Technical Analysts Have Holes in Their Shoes?" *Review of Research in Futures Markets*, 3(1984):264-277.

Ito, Takatoshi, *Capital Flows in Asia* (May 1999).

James, F. E. Jr. "Monthly Moving Averages - An Effective Investment Tool?" *Journal of Financial and Quantitative Analysis*, (September 1968):315-326.

Jensen, M. C. "Random Walks: Reality or Myth - Comment." *Financial Analysts Journal*, 23 (1967):77-85.

Jensen, M. C., and G. A. Benington. "Random Walks and Technical Theories: Some Additional Evidence." *Journal of Finance*, 25(1970):469-482.

Kidd, W. V., and B. W. Brorsen. "Why Have the Returns to Technical Analysis Decreased?" *Journal of Economics and Business*, 56(2004):159-176.

Koza, J. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. Cambridge, MA: MIT Press, 1992.

LeBaron, B. "Technical Trading Rule Profitability and Foreign Exchange Intervention." *Journal of International Economics*, 49(1999):125-143.

- Lee, C. I., and I. Mathur. "Trading Rule Profits in European Currency Spot Cross-Rates." *Journal of Banking & Finance*, 20(1996a):949-962.
- Lee, C. I., and I. Mathur. "A Comprehensive Look at the Efficiency of Technical Trading Rules Applied to Cross-Rates." *European Journal of Finance*, 2(1996b):389-411.
- Lee, C. I., K. C. Gleason, and I. Mathur. "Trading Rule Profits in Latin American Currency Spot Rates." *International Review of Financial Analysis*, 10(2001):135-156.
- Lee, C. I., M. Pan, and Y. A. Liu. "On Market Efficiency of Asian Foreign Exchange Rates: Evidence from a Joint Variance Ratio Test and Technical Trading Rules." *Journal of International Financial Markets, Institutions and Money*, 11(2001):199-214.
- Levich, R. M., and L. R. Thomas. "The Significance of Technical Trading Rule Profits in the Foreign Exchange Market: A Bootstrap Approach." *Journal of International Money and Finance*, 12(1993):451-474.
- Levy, Robert A. 1966. "An Evaluation of Selected Applications of Stock Market Timing Techniques, Trading Tactics and Trend Analysis." Unpublished Ph. D. dissertation, the American University.
- Levy, Robert A. 1967a. "The Principle of Portfolio Upgrading." *Industrial Management Review*: pp 82-96.
- Levy, Robert A. 1967b. "Random Walks: Reality or Myth." *Financial Analysts Journal* November/December.
- Levy, Robert A. 1967c. "Relative Strength as a Criterion for Investment Selection." *Journal of Finance* 22: pp 595-610.
- Lo, A., H. Mamaysky, and J. Wang. "Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation." *Journal of Finance*, 55 (2000):1705-1765.
- Lukac, L. P., and B. W. Brorsen. "A Comprehensive Test of Futures Market Disequilibrium." *Financial Review*, 25(1990):593-622.
- Lukac, L. P., B. W. Brorsen, and S. H. Irwin. "A Test of Futures Market Disequilibrium Using Twelve Different Technical Trading Systems." *Applied Economics*, 20(1988):623-639.
- MacKay, D.J.C. "Information Theory, Inference and Learning Algorithms", *Cambridge University Press*, First Edition, 2003.
- Maillet, B., and T. Michel. "Further Insights on the Puzzle of Technical Analysis Profitability." *European Journal of Finance*, 6(2000):196-224.
- Mandelbrot, B. "The Variation of Certain Speculative Prices." *Journal of Business*, 36(1963): 394-419.
- Marsh, I. W. "High-Frequency Markov Switching Models in the Foreign Exchange Market." *Journal of Forecasting*, 19(2000):123-134.

- Martell, T. F. "Adaptive Trading Rules for Commodity Futures." *Omega*, 4(1976):407-416.
- Martin, A. D. "Technical Trading Rules in the Spot Foreign Exchange Markets of Developing Countries." *Journal of Multinational Financial Management*, 11(2001):59-68
- Menkhoff, L., and M. Schlumberger. "Persistent Profitability of Technical Analysis on Foreign Exchange Markets?" *Banca Nazionale del Lavoro Quarterly Review*, No. 193(1997):189- 216.
- Miller, R.G., "Simultaneous Statistical Inference", 2nd Ed. *Springer Verlag*, New York, 1981. ISBN 0-387-90548-0.
- Mills, T. C. "Technical Analysis and the London Stock Exchange: Testing Trading Rules Using the FT30." *International Journal of Finance and Economics*, 2(1997):319-331.
- Neely, C. J. "Technical Analysis in the Foreign Exchange Market: A Layman's Guide." *Review*, Federal Reserve Bank of St. Louis, September/October, (1997):23-38.
- Neely, C. J. "The Temporal Pattern of Trading Rule Returns and Exchange Rate Intervention: Intervention Does Not Generate Technical Trading Profits." *Journal of International Economics*, 58(2002):211-232.
- Neely, C. J. "Risk-Adjusted, Ex Ante, Optimal Technical Trading Rules in Equity Markets." *International Review of Economics and Finance*, 12(2003):69-87.
- Neftci, S. N. "Naïve Trading Rules in Financial Markets and Wiener-Kolmogorov Prediction Theory: A Study of 'Technical Analysis.'" *Journal of Business*, 64(1991):549-571.
- Neely, Christopher J. and Rapach, David and Tu, Jun and Zhou, Guofu, "Forecasting the Equity Risk Premium: The Role of Technical Indicators" (August 3, 2013). Federal Reserve Bank of St. Louis
- Neely, C. J., and P. A. Weller. "Technical Analysis and Central Bank Intervention." *Journal of International Money and Finance*, 20(2001):949-970
- Neely, C. J., P. A. Weller, and R. Dittmar. "Is Technical Analysis Profitable in the Foreign Exchange Market? A Genetic Programming Approach." *Journal of Financial and Quantitative Analysis*, 32(1997):405-426.
- Olson, D. "Have Trading Rule Profits in the Currency Markets Declined over Time?" *Journal of Banking and Finance*, 28(2004):85-105.
- Park, Cheol-Ho and Irwin, Scott, "The Profitability of Technical Analysis: A Review" (October 2004). *AgMAS Project Research Report No. 2004-04*.
- Peterson, P. E., and R. M. Leuthold. "Using Mechanical Trading Systems to Evaluate the Weak form Efficiency of Futures Markets." *Southern Journal of Agricultural Economics*, 14(1982):147-152.
- Poole, W. "Speculative Prices as Random Walks - An Analysis of Ten Time Series of Flexible Exchange Rates." *Southern Economic Journal*, 33(1967):468-478.
- Pruitt, S. W., and R. E. White. "The CRISMA Trading System: Who Says Technical Analysis Can't Beat the Market?" *Journal of Portfolio Management*, (Spring 1988):55-58.

- Pruitt, S. W., K. S. Maurice Tse, and R. E. White. "The CRISMA Trading System: The Next Five Years." *Journal of Portfolio Management*, (Spring 1992):22-25.
- Qi, M., and Y. Wu. "Technical Trading-Rule Profitability, Data Snooping, and Reality Check: Evidence from the Foreign Exchange Market." *Working Paper*, 2002.
- Raj, M., and D. Thurston. "Effectiveness of Simple Technical Trading Rules in the Hong Kong Futures Markets." *Applied Economics Letters*, 3(1996):33-36.
- Ratner, M., and R. P. C. Leal. "Tests of Technical Trading Strategies in the Emerging Equity Markets of Latin America and Asia." *Journal of Banking & Finance*, 23(1999):1887- 1905.
- Ready, M. J. "Profits from Technical Trading Rules." *Financial Management*, 31(2002):43-61.
- Ricciardi, Victor and Simon, Helen K., "What is Behavioral Finance?". *Business, Education & Technology Journal*, Vol. 2, No. 2, pp. 1-9, Fall 2000.
- Romano, J. and M. Wolf, "Stepwise multiple testing as formalized data snooping", *Econometrica*, 73 (2005), 1273–1282.
- Saacke, P. "Technical Analysis and the Effectiveness of Central Bank Intervention." *Journal of International Money and Finance*, 21(2002):459-479.
- Sala-i-Martin, X., "I just ran two million regressions." *American Economic Review*. 87(2), May (1997).
- Samuelson, P. A. (1965). "Proof that Properly Anticipated Prices Fluctuate Randomly." *Industrial Management Review*, 6, 41-49.
- Sapp, S. "Are All Central Bank Interventions Created Equal? An Empirical Investigation." *Journal of Banking & Finance* 28(2004):443-474.
- Schorfheide, F. and K. Wolpin, "On the Use of Holdout Samples for Model Selection," *American Economic Review*, 102 (2012), 477–481
- Sharpe, W.F., "A simplified model for portfolio analysis." *Management Science*, 9:277–293, 1963.
- Silber, W. L. "Technical Trading: When It Works and When It Doesn't." *Journal of Derivatives*, 1(1994):39-44.
- Stevenson, R. A., and R. M. Bear. "Commodity Futures: Trends or Random Walks?" *Journal of Finance*, 25(1970):65-81.
- Strathern, M., "Improving Ratings: Audit in the British University System," *European Review*, 5, (1997) pp. 305-308.
- Sullivan, R., A. Timmermann, and H. White. "Data Snooping, Technical Trading Rule Performance, and the Bootstrap." *Journal of Finance*, 54(1999):1647-1691.
- Sullivan, R., A. Timmermann, and H. White. "Forecast Evaluation with Shared Data Sets."

International Journal of Forecasting, 19(2003):217-227.

Sweeney, R. J. "Beating the Foreign Exchange Market." *Journal of Finance*, 41(1986):163-182.

Taylor, S. J., and A. Tari. "Further Evidence against the Efficiency of Futures Markets." In R. M. C. Guimaraes et al., (ed.), *A Reappraisal of the Efficiency of Financial Markets*, Heidelberg: Springer-Verlag, 1989.

Taylor, S. J. "Trading Rules for Investors in Apparently Inefficient Futures Markets." *In Futures Markets - Modeling, Managing and Monitoring Futures Trading*, Oxford, UK: Basil Blackwell, 1983.

Taylor, S. J. *Modelling Financial Time Series*. Chichester, England: John Wiley & Sons, 1986.

Taylor, S. J. "Rewards Available to Currency Futures Speculators: Compensation for Risk or Evidence of Inefficient Pricing?" *Economic Record*, 68(1992):105-116.

Taylor, S. J. "Trading Futures Using a Channel Rule: A Study of the Predictive Power of Technical Analysis with Currency Examples." *Journal of Futures Markets*, 14(1994): 215-235.

Theil, H. and Leenders, C.T. (1965) Tomorrow on the Amsterdam Stock Exchange, *Journal of Business*, 38, pp. 227-284.

Thompson, S. R., and M. L. Waller. "The Execution Cost of Trading in Commodity Futures Markets." *Food Research Institute Studies*, 20(1987):141-163.

Timmermann, A., and C. W. J. Granger. "Efficient Market Hypothesis and Forecasting." *International Journal of Forecasting*, 20(2004):15-27.

Tomek, W. G., and S. F. Querin. "Random Processes in Prices and Technical Analysis." *Journal of Futures Markets*, 4(1984):15-23.

Van Belle, G. and K. Kerr, *Design and Analysis of Experiments in the Health Sciences*, John Wiley and Sons, 2012.

Van Horne, J. C., and G. G. C. Parker. "The Random-Walk Theory: An Empirical Test." *Financial Analysts Journal*, 23(1967):87-92.

Watts, R.L., 1978. "Systematic 'Abnormal' Returns after Quarterly Earnings Announcements." *Journal of Financial Economics*, volume 6, p. 127 – 150

Weiss, S. and C. Kulikowski, *Computer Systems That Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning and Expert Systems*, Morgan Kaufman, 1st Edition, 1990.

Welch, Ivo and Goyal, Amit, "A Comprehensive Look at the Empirical Performance of Equity Premium Prediction" (July 2008). *The Review of Financial Studies*, Vol. 21, Issue 4, pp. 1455-1508, 2008.

Welles Wilder, J., *New Concepts in Technical Trading Systems, Commodities*, 1978

White, H., "A Reality Check for Data Snooping," *Econometrica*, 68 (2000), 1097– 1126.