CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Gambler's Fallacy in Investor's Decision-making

Master's thesis

Author: Bc. Tereza Javůrková

Study program: Economics and Finance

Supervisor: PhDr. Jiří Kukačka, Ph.D.

Year of defense: 2023

Declaration of Authorship

The author hereby declares that he or she compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain any other academic title.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis in whole or in part and agrees with the thesis being used for study and scientific purposes.

Prague, April 30, 2023 Bc. Tereza Javůrková

Abstract

This thesis focuses on the Gambler's Fallacy and its effect on the behavior of investors operating in the stock market. The aim is to incorporate the psychological findings about this behavioral phenomenon to the field of finance. This allows us to analyze the dynamics of the stock market that results from human misconceptions about the probabilities of independent events. More specifically, we analyze the profitability of two types of virtual investors whose decision-making is affected by distorted probabilities based on the Gambler's Fallacy. We further define two other trivial benchmark investors' strategies with different levels of randomness. We examine investors' gains in a simulated efficient market as well as in the real S&P 500 index constituents. Our analysis builds on three different approaches: simulation analysis, empirical frequency analysis, and asset pricing models.

By applying the simulation approach together with frequency analysis on the historical stock prices, we find that investors affected by the Gambler's Fallacy gain statistically higher returns than a random investor. Then, we apply both the three-factor and five-factor Fama & French asset pricing model to stocks sorted into portfolios based on their previous earnings per share evolution. Our findings reveal a negative excess return for stocks that are, based on their recent evolution, more likely to be purchased by investors exhibiting a bias towards the Gambler's Fallacy. These results are also consistent with our novel asset pricing approach based directly on psychological findings.

JEL Classification	F12, F21, F23, H25, H71, H87
Keywords	Gambler's Fallacy, Law of Small Numbers, Fama & French model, Efficient Market Hypothesis
Title	Gambler's Fallacy in Investor's Decision-making

Abstrakt

Tato práce se zabývá konceptem Gambler's Fallacy a jeho vlivem na chování investorů při působení na akciovém trhu. Jejím hlavním cílem je propojit výsledky z oblasti psychologie a financí, a v důsledku analyzovat dynamiku akciového trhu, která pramení z mylného chápání pravděpodobností nezávislých jevů. Konkrétně zkoumáme zisk myšlených investorů, kteří používají různé strategie. Rozhodovací proces dvou z nich vychází z psychologických výzkumů zaměřujících se na Gambler's Fallacy. Další dva používají triviální strategie, které slouží k následnému porovnání. V analýze simulujeme ziskovosti jednotlivých strategií nejprve na teoretickém eficientním trhu, a následně na reálných datech indexu S&P 500. Pro naše měření aplikujeme postupně tři různé metody: simulační metodu, empirickou frekvenční analýzu a metodu oceňování aktiv.

Výsledky simulační metody i frekvenční analýzy na historických datech ukazují, že investoři ovlivnění Gambler's Fallacy mají konzistentně vyšší výnos než investor používající čistě náhodnou strategii. Pokročilejší metoda oceňování aktiv ovšem ukazuje, že portfolia—z pohledu investora ovlivněného Gambler's Fallacy vyhodnocená jako výhodná—mohou mít negativní ziskový potenciál. Obdobný výsledek získáváme také oceňovacím přístupem, který nově zavedeme, a který vychází přímo z popisu chování takového investora z psychologického hlediska.

Klasifikace JEL	F12, F21, F23, H25, H71, H87
Klíčová slova	Gambler's Fallacy, Zákon Malých Čísel, Fama & French model, Teorie eficientního trhu

Název práce	Gambler's Fallacy v Rozhodování Investorů

Acknowledgments

I would like to express my gratitude especially to my supervisor, PhDr. Jiří Kukačka, Ph.D. for his expertise, useful comments, and willingness to help.

Also, I would like to thank my family and the closest friends for their permanent support during the whole studies and mainly during the process of writing this thesis.

Typeset in LaTeX using the IES Thesis Template.

Bibliographic Record

Javůrková, Tereza: *Gambler's Fallacy in Investor's Decision-making.* Master's thesis. Charles University, Faculty of Social Sciences, Institute of Economic Studies, Prague. 2023, pages 65. Advisor: PhDr. Jiří Kukačka, Ph.D.

Contents

List of Tables viii

List of Figures ix

Acronyms x

Thesis Proposal xi

1 Introduction 1

2 Literature Review 4

2.1 Psychological biases 5

2.2 Efficient Market Hypothesis 7

3 Simulation of Investors' Decision-making 9

3.1 Description of Investors 9

3.1.1 Random Investors 10

3.1.2 Investors with the Gambler's Fallacy Bias 10

3.1.3 Investors According to Boynton (2003) 12

3.1.4 Buyers 13

3.2 Description of the Random Walk 13

3.2.1 Rules of Investments 14

3.2.2 Theoretical Approach 15

3.2.3 Simulation Based Approach 19

4 Evaluation of Historical Data Distribution 22

4.1 Theory 22

4.2 Description of the Data 24

4.3 Methodology 24

4.4 Results 25

4.5 Frequency Analysis 27
4.5.1 Methodology 27
4.5.2 Results of Frequency Analysis 28
4.5.3 Particular Stocks 30
4.6 Summary 30

5 Fama & French Models 34
5.1 Asset Pricing Models 35
5.2 Methodology 36
5.2.1 Quintiles 37
5.2.2 Signs 38
5.2.3 Probabilities 39
5.3 Description of the Data 40
5.4 Results 41
5.4.1 Comparison of Models 43
5.4.2 Summary 44

6 Conclusion 45

Bibliography 51

List of Tables

3.1 Conditional probability based on the length of previous series of the same color. 11
3.2 Conditional probability of obtaining 1, based on the previous outcome. 12
3.3 Payoffs of an investor operating on a random walk with drift equals to q. 16
3.4 Probability of a success and a failure of an investor operating on a random walk with drift equals to q. 17
3.5 Average final payoff of different investor types in the market simulated by random walks with different drifts. 19
3.6 P-values of two-sided t-test evaluating the equality of final payoff of random investors and others. 20

4.1 Sum of final wealth of each investor type operating in S&P 500 index constituents. 26

5.1 Alphas [%] of Fama & French models, Quintile approach. 41
5.2 Alphas [%] of Fama & French models, Sign approach. 41
5.3 Alphas [%] of Fama & French models, Probability approach. . . 42

List of Figures

3.1 Distribution of final wealth of each non-deterministic investor type operating on a random walk with different drifts. 21

4.1 Distribution of average final wealth of each investor type operating in S&P 500 index constituents. 25

4.2 Average gains of investments into stocks from different industries of S&P 500 index. 26

4.3 Proportion of patterns of length 5 in the S&P 500 constituent stock prices evolution between 2012 and 2022, grouped by number of '1'. 29

4.4 Proportion of patterns of length five and price evolution of AMZN stock between 2012 and 2022. Patterns are sorted and groupped by the number of '1'. 32

4.5 Proportion of patterns and price evolution of NVR stock between 2012 and 2022. Patterns are sorted and groupped by the number of '1'. 33

Acronyms

EMH Efficient Market Hypothesis

CLT Central Limit Theorem

i.i.d. Independent, identically distributed events

CAPM Capital Asset Pricing Model

EPS Earnings Per Share

SURP Earnings Per Share surprise

SMB Small Minus Big factor

HML High Minus Low factor

RMW Robust Minus Weak factor

CMA Conservative Minus Aggresive factor

Master's Thesis Proposal

Author	Bc. Tereza Javůrková
Supervisor	PhDr. Jiří Kukačka, Ph.D.
Topic	Gambler's Fallacy in Investor's Decision-making

Motivation In general, many traditional financial models assume all agents to be rational. This assumption is challenged by the field of behavioral finance, which emerged in the late seventies. A human element became more and more used in theories trying to understand the financial market biased by the erroneous intuition of its agents. Since its beginnings, researchers of this field have revealed a number of influences and biases that can be the source for the explanation of different types of market anomalies.

This work will focus especially on the Gambler's fallacy. It is a bias caused by omitting the theory of independent, identically distributed events that one intuitively expects to be correlated with each other. People tend to anticipate a reversal of chances while long series of gains or losses occur. Their confidence about the decision-making thereby increases with every new observation. This behavior is caused by people's expectancy that even a small sample of events should be representative according to the overall probabilities. This bias is also called the Law of small numbers as a paraphrase of the Law of large numbers used in statistics. Those psychological effects are mostly observed in gambling, sports bets, and many other fields. There are also pieces of evidence of Gambler's fallacy occurrence in investment (Rabin 2002b).

To examine the existence of this effect, a number of experiments on the decision-making of investors focusing on their motivation and confidence while deciding were run (Stöckl et al. 2015). As a result, it was found that diverse biases affect the investor's intuition, however, in comparison with herding or status quo bias, the Gamblers' fallacy is a highly significant one. In this work, I will study its impact on investor's behavior by two approaches. In the first one, I will use the results of psychological experiments to parametrize a simulation-based model of biased decision making in order to better understand the effect of Gamblers fallacy on the investor's gain. The second one will be data-based and will examine the occurrence of Gambler's fallacy in the real world. The significance of this effect will be studied using the quarterly EPS dataset. Dataset used will be representative for a market of less developed countries and I will make a comparison with the S&P 500 index representative for the US market.

Hypotheses

> Hypothesis #1: The difference between the long-term gain of a rational and biased investor in the simulation-based model is not statistically significant.
>
> Hypothesis #2: Will the investors more influence by the Gambler's fallacy have lower return?
>
> Hypothesis #3: There is no significant difference in the level of Gambler's fallacy affecting developed and less developed countries' markets.

Methodology In the first step, I will clarify and sum up the notion of the Gambler's fallacy from the mathematical and psychological point of view. The mathematical approach will be based on the theory of posterior probability of dependent and independent events. The erroneous thinking of investors will be shown using the theorem of Bayes which proves that the posterior probabilities of independent identically distributed events are equal to the initial probabilities of those events determined by their distribution (Rabin 2002b).

The empirical part will be divided into two subparts, a simulation-based and a data-based one.

Many psychological papers are focused on the Gambler's fallacy in fields of gambling, sports pools, or coin toss, however, only a few of them map the occurrence of this effect in finance. For that reason, in the first step, I will make a model based on empirical psychologic studies (Barron, Leider, 2010, Ayton, Fischer 2005). I will simulate the evolution of stock prices by a random walk with a zero mean and standard deviation of 1. The model will simulate the decision-making of potential investors on this artificial market. In every period, an investor will choose how to deal with his wealth. Three choices will be possible. In every time period, for every unit of wealth, he will decide whether to invest it into the stock, not invest or (short-)sell already owned stock. This simulation will be done for two investors separately. The first one will choose between the three choices randomly, without any bias. The second will rely on his intuition and self-confidence, which will make the bias occur. His intuition will be modeled according to the psychological results of empirical experiments concerning the presumed posterior probability of price growth or decrease after the sequence of previous gains/losses of the stock. The results of the investment of those two investors will be compared and described.

The second half of the empirical part will be based on (Igboekwu 2015, Rabin 2002b, Rabin, Vayanos 2010). In this thesis, the time series of quarterly earnings per share data of S&P500 is used to measure the "earnings surprise", which is defined as the difference between the observed EPS and the expected EPS. The prediction is based on the precedent year observed EPS for the corresponding quarter. The paper examines whether the conditional EPS follows the Bayesian distribution as it is supposed to, or whether a bias is observed. A potential shift of probabilities distribution could signify the Gambler's fallacy effect (Rabin 2002b). I will use a similar approach to study the market of South America and China. For that, I will use the data of Brazil Bovespa Stock Index and Shangai Composite available through Eikon. I will measure all the posterior probabilities of the sample time series and compare them with its theoretical probability distribution.

Expected Contribution I will set up a new model simulating the investor's decision-making as a response to the generated stock price evolution. This model will be based on the psychological findings and apply them to the stock market functioning in order to simulate the Gambler's fallacy bias in this field. I will compare the long-term returns of biased and unbiased investor's decision-making for different parameters. The data-based empirical part will compare the results of Gambler's fallacy significance measured for developed US market already explained in the literature and the developing market measured by myself by a similar approach.

Outline

1. Introduction: There is evidence, that investors are not rational as supposed in the traditional financial models. A part of the irrational behavior of investors could be explained by the Gambler's fallacy effect.

2. Mathematical background: I will describe the mathematical and psychological background of the Gambler's fallacy effect.

3. Simulation-based model: I will simulate investor's decision making by a model based on psychological experiment results.

4. Data-based model: I will describe the dataset and methodology used in the second model evaluating real stock market data.

5. Results: I will discuss trends for different parameters of simulation-based model and conclude on results of the data-based approach.

6. Conclusion: I will summarize my findings and propose possibilities for future research.

Core bibliography

Rabin, M. (2002b) Inference by believers in the law of small numbers. The Quarterly Journal of Economics. 117 (3). p.775–816

Rabin, M. & Vayanos, D. (2010) The gambler's and hot-hand fallacies: Theory and applications. The Review of Economic Studies. 77 (2). p.730–778.

Igboekwu, A. Representative agent earnings momentum models: the impact of sequences of earnings surprises on stock market returns under the influence of the Law of Small Numbers and the Gambler's Fallacy. Loughborough, 2014. Doctoral thesis. Loughborough University

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. (1998) A Model of Investor Sentiment. Journal of Financial Economics 49 (3): 307–343.

Barron, Greg & Leider, Stephen. (2010). The role of experience in the Gambler's Fallacy. Journal of Behavioral Decision Making. 23. 117-129. 10.1002/bdm.676.

Ayton, Peter & Fischer, Ilan. (2005). The Hot Hand Fallacy and the Gambler's Fallacy: Two faces of Subjective Randomness?. Memory & cognition. 32. 1369–78. 10.3758/BF03206327.

Stöckl, Thomas & Huber, Jürgen & Kirchler, Michael & Lindner, Florian. (2015). Hot hand and gambler's fallacy in teams: Evidence from investment experiments. Journal of Economic Behavior & Organization. 117. 327–339. 10.1016/j.jebo.2015.07.004.

Author

Supervisor

Chapter 1

Introduction

Traditional financial models generally assume that the agents are fully rational and they use all freely available information to maximize their utility. At the same time, agents are supposed not to take into consideration any irrelevant information that does not affect their situation. However, this assumption has been increasingly challenged by researchers since the 70s, leading to the emergence of behavioral finance as a new area of research.

In behavioral finance, the assumption of full rationality is considered unrealistic and not corresponding to human behavior. The main idea of this field is that the agents' rationality is bounded due to the limited human mental capacity. The agents are not seeking the best, optimal solution but the one with an adequate level of acceptability. To make the decision simpler and comprehensible for them, the agents become less precise. They tend to ignore some information or might use irrelevant information to simplify the problem. This may result in systematic biases that behavioral finance focuses on and tries to explain.

Our main focus is on investigating two commonly observed biases in human decision-making: the Gambler's Fallacy and the Law of Small Numbers. These biases stem from the inability of individuals to fully consider the probabilities of independent events in everyday life. As a result, despite attempting to optimize their utility, individuals base their decision-making on erroneous assumptions. While these biases are most commonly observed in gambling, coin tossing, and sports betting, we can detect them in any situation where individuals try to predict the future outcome of independent events. Evidences, such as Loh & Warachka (2012) or Huber *et al.* (2010), suggest that even investors in the stock market might be subject to these biases. The objective of our analyses is to evaluate the investment returns resulting from strategies influenced by the Gambler's Fallacy. Specifically, we aim at testing the hypothesis that an investor who exhibits this bias does consistently loose while operating in the market.

Another key topic addressed in this thesis is the Efficient Market Hypothesis. The Gambler's Fallacy as well as the Law of Small Numbers assume that the underlying events are independent. In the case of stock prices evolution, this assumption is only met if the market is efficient and it satisfies the Random Walk Theory.

To test the Efficient Market Hypotheses, we combine the findings of financial theory about market evolution together with psychological research fields describing human thinking and behavior. We draw on well-established research on human reasoning when making predictions about the outcomes of random events, and we apply these findings to evaluate the specific case of stock prices. For this purpose, we utilize quantified results of psychological experiments from Boynton (2003) and Barron & Leider (2010) that analyze the behavior from the probabilistic point of view.

To simulate these concepts, we define two non-trivial investment strategies that are influenced by the Gambler's Fallacy. In addition, we instantiate one investor with fully random decision-making and one other deterministic strategy that involve only buying. These two trivial investors' strategies serve as benchmark bases for performance evaluation of more complex, probabilistic strategies.

Besides the four theoretical investors' strategies, we analyze four distinct approaches to estimate the returns of each investor. In general, the simulation-based estimates of theoretical returns enables us to compare the profitability of each strategy and to determine whether there is a significant difference in profitability among various investment strategies. In the first simulation-based approach, we evaluate the hypothesis from a theoretical perspective, assuming that the market is efficient. In the second simulation-based analysis, we simulate the simplified operations of biased investors in the market, using real historical data to test the feasibility of these strategies. Thirdly, we analyze historical stock prices using frequency analysis. Finally, we apply the Fama & French three-factor and five-factor asset pricing models to assess the extent to which the stocks, invested in by biased agents, are overpriced or underpriced. These approaches for the estimation of returns provide insight into the broader implications of investors' strategies for market efficiency.

This thesis presents two significant contributions that are both empirical and theoretical in nature. The primary empirical contribution involves the use of the more powerful five-factor Fama & French model to replicate the results of Loh & Warachka (2012). Additionally, we propose a novel approach to evaluate the profitability of investors affected by the Gambler's Fallacy based on the Fama & French asset pricing methods. The main theoretical contribution of this thesis lies in the incorporation of psychological research data into this novel approach. Our model builds upon the Fama & French asset pricing framework while integrating insights from psychological research on the Gambler's Fallacy and the Law of Small Numbers as outlined in Barron & Leider (2010). Through this approach, we are able to assess the extent to which stocks favored by biased investors may be overpriced or underpriced in the market. Our findings confirm those of Loh & Warachka (2012), demonstrating that the strategy employed by investors influenced by the Gambler's Fallacy generates a significant negative excess return. However, the simulation-based approach as well as the frequency analysis indicate that using this strategy might be in some cases profitable for investors.

This thesis is structured as follows: In Chapter 2, we discuss the notion of Gambler's Fallacy and the Law of Small Numbers from the probability point of view, and we explain the characteristics of independent, identically distributed events. In Chapter 3, we introduce strategies of four investors with different biases motivated by psychological literature as well as their theoretical gains in an efficient market. We apply the same approach using historical S&P 500 data in Chapter 4. Finally, in Chapter 5, we evaluate the performance of different investors' strategies using the methods of asset pricing and we compare the results of different approaches. Chapter 6 concludes the thesis.

Chapter 2

Literature Review

Tversky & Kahneman (1971) present results of an experiment where an audience was shown three possible sequences of sex of six babies born that day in a local hospital, where 'B' means a boy and 'G' means a girl: 'BBBGGG', 'GGGGGG', 'BGBBGB'. They asked the audience to guess whether these observations are equally likely to be observed, and if not, which one is the most likely one. The results show that the majority of the audience intuitively answered that these sequences are not equally likely to be observed. It is well known that the event of giving a birth to a boy or to a girl is random with the same probability for both and the sex of one of them does not influece the others in any way. However, the third presented sequence is judged to be more likely, than others.

In another experiment, Barron & Leider (2010) invite participants to predict the next outcome of a roulette wheel spin. Given the colors of previous outcomes, people were asked to predict whether this time red or black color would be dropped. Authors observed that long streaks of one color are missing and people rather predict the trends to reverse.

In Ayton & Fischer (2004), authors show that the probability of considering, whether a sequence of results of coin tossing or roulette spinning is random, depends on the alternation rate of the sequence. It means that it depends on the number of times when we observe two consequent events with different results.

Kahneman (2011) presents examples of sequences 'HTHHHT', 'HHHTTT' and 'HHTHTH', where 'H' stands for head and 'T' for tail, as results of a coin tossing with the equal probabilities to be observed. However, for the majority of people, the last one seems to be more 'random' and more likely to be generated from a binomial distribution. This is also a consequence of the belief that long sequences would rather reverse than continue, which is however incorrect.

2.1 Psychological biases

All the above presented experiment results are clear examples of the Gambler's Fallacy and the Law of Small Numbers, the two biases this thesis focuses on.

Gambler's Fallacy

The Gambler's Fallacy is a psychological bias which was first observed in Monte Carlo casino at the beginning of the 20^{th} century and subsequently described by Ayton & Fischer (2004), Tversky & Kahneman (1971), Rabin (2002) and others. It is based on human inability to fully consider the probabilities of random events while making decisions. This bias is an erroneous belief that a result of a random event is based on the outcome of the previous independent and identically distributed events. Based on the definition of independent events, we know that this believe is incorrect because the past events do not change the probability of observing one specific result in the next trial. Therefore, the probabilities subjectively interpreted by the agent are different from the real ones. Consequently, while people predict the future outcome, the alternation rate is high and long streaks of only one output are missing, compared to a random generator.

Law of Small Numbers

For the first time, the Law of Small Numbers was mentioned in Tversky & Kahneman (1971). The name is a paraphrase of the Law of Large Numbers, a key theorem that is used in probability theory, explained later in Section 3.2.2. In a nutshel, the Law of Large Numbers states that repeating the same experiment of independent events with the same distribution for a large number of times, the average of the obtained results should be close to the mean of the original distribution. The larger the sample is, the closer the result is expected to converge to the mean value.

The Law of Small Numbers is a psychological bias based on misunderstanding of the concept of probability in a real life. It states that some people tend to think that the Law of Large Numbers applies also to small samples. People therefore assume small samples to be much more representative than they are in reality. Consequently, sequences with significantly higher proportion of one color seem to be less likely to be produced from a Binomial distribution with zero mean.

In general, we may consider the Gambler's Fallacy as a consequence of individual's believe in the Law of Small Numbers. Indeed, people tend to believe that even small subsamples should have the same proportion of outcomes as the overall distribution, they are generated from. As a consequence, they tend to predict an output with a higher alternation rate to keep the outputs' proportion all the time close to the mean of the distribution.

Independent, Identically Distributed Events

All of the presented examples have a common property of independent, identically distributed (i.i.d.) events. It means that the events are produced by the same distribution with same parameters and one observation does not affect any other. E.g., for a fair coin tossing, we assume that observations come from the same distribution because we are supposed to use the same perfectly balanced coin, and we toss it in the same way. Such observations are also independent of each other, since the coin has no 'memory' that would affect the current trial by the preceeding ones.

On the grounds of these properties, there is a 50% chance to obtain a head as well as a tail in each period. All of the a posterior probabilities are equal to the respective a priori ones. As a result, all of the sequences of observations of the same length appear with the same probability.

However, Rabin (2002) shows that people, who are influenced by the Gambler's Fallacy and by the Law of Small Numbers, fail to take these properties fully into consideration. In reality, they believe that coin toss corresponds more to drawing balls of two colors from an urn without replacement. By picking a ball of one color, the proportion of the balls inside the urn changes in the way that it is more likely to pick the other color in the next trial. In this case, the a priori probability differs from the a posteriori one, and therefore, the events are not independent of each other.

In Chapter 3, to examine the evolution of stock prices, we are also interested in the distribution of cummulative results for a sequence of a certain length. Using the coin tossing example, we are interested in the proportion of heads and tails observed during a certain number of identical repetitions of this event. This is a different case. Even if the events are independent, the probability of obtaining one head and three tails in four tosses is higher than the prbability of obtaining for example four heads. This is because there is only a single unique outcome of the series that gives four heads, which is the trivial one 'HHHH'. On the other hand, there exist more possible sequences of obtaining exactly one head in four tosses, which are 'HTTT', 'THTT', 'TTHT', 'TTTH'. Therefore, in the cummulative point of view, the one tail and three heads are four times more likely to be observed than four heads.

Kahneman (2011) and Ayton & Fischer (2004) show that the Gambler's Fallacy and the Law of Small Numbers can be observed in gambling or sports bets as well as in daily life situations. In general, both can be observed wherever people aim at predicting the results of future events or where they evaluate the probability of independent, identically distributed events. This thesis focuses on evaluating how the Gambler's Fallacy affects investors' decision-making, what effect it would have on individual's gain and generally on the evolution of a market.

2.2 Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) was first used by Fama (1970) where the author explains the importance of ownership allocation of the capital stock. The aim of this hypothesis is to find how market prices reflect the market valuation, and whether they incorporate all of the available information. Malkiel (1989) states that under the EMH, all available information is already incorporated in the market prices. As a result, there should not exist any overvalued or undervalued securities because the prices are supposed to be a result of all available information. In other words, there is no possibility for market anomalies to occur because they would be immediately arbitraged away as the market has no more information than the investor.

Fama (1998) explain that in efficient market, the overreaction to a new information is as frequent as the underreaction. Therefore, we cannot observe long-term anomalies in efficient markets. The EMH states that the only way of consistently obtaining a higher gain than the market is by buying riskier investments. Malkiel (1989) explains that this is not possible by a precise stock selection or choosing the time of the investment. Therefore, investors who believe in this hypothesis tend to invest in low-cost passive portfolio, which is expected to generate at least the market return. A more recent examination of the EMH is presented in Malkiel (2003)

Three Forms of Efficient Market Hypothesis

Fama (1970) suggest three forms of the EMH: the Week, the Semi-strong and the Strong. Malkiel (1973) focuses on the Weak form and he states that all relevant and publicly available historical financial information is incorporated in the market prices. According to this hypothesis, the past evolution of the stock prices and the volumes of respective trades do not affect the market value of the stocks. The only way how an investor can make a profit is to possess an insider information that is not incorporated in the price.

In addition to the assumptions of the Week form, the Semi-strong form of the hypothesis assumes that all publicly available information is at any time incorporated in the stock prices. Finally, the Strong form of the efficient market theory is also the narrowest one. It claims that the prices incorporate all of the information in the market, public as well as private. It includes all the information of the Weak and Semi-strong forms and the private information regarding a financial asset.

Malkiel (1989) as well as Ţiţan (2015) show that both the Strong and Semi-strong forms of the EMH have been largely invalidated by historical evidence and are not supported by financial data. The Weak form of EMH, however, has been the subject of various studies with mixed results, none of which strongly predominates. Therefore, it is worth testing it once more, this time from simulation based approach.

Random Walk Theory

One assumption of the Weak form of EMH is the Random Walk Theory, first studied by Bachelier (1900) and further developped for example by Malkiel (1973), or more recently by (Fama 1995). This theory states that no investor can gain a consistent in an efficient market, because its prices are mutually independent and come from the same probability distribution . Therefore, they behave like a random walk and they possess all of the properties of independent identically distributed events. (Horne & Parker 1967) explain that due to this properties, no investor can gain a consistent abnormal return in an efficient market. The Gambler's Fallacy, as well as the Law of Small Numbers are biases based on misunderstanding of independent events, therefore, the assumption of the EMH, at least in its Weak form, is essential for this thesis. This hypothesis allow us, in the theoretical part, to simulate the evolution of stock prices by a random walk.

Chapter 3

Simulation of Investors' Decision-making

In this chapter, we simulate the behavior of several types of investors with different biases, which affect their decision-making while investing into stocks, and we evaluate their performance. This analysis combines the theory of stock markets and psychology findings about human understanding of probabilities in everyday life to examine investors' behavior. Barberis *et al.* (1998) and Rabin (2002) show that the psychological biases explained in Chapter 2, which are typical for gamblers are also observed among investors operating on financial markets. Therefore, to analyze investor's behavior, we apply the findings of psychological research about human understanding of probabilities of i.i.d. events. We run simulations, where we compare the final wealth of investors with different psychological biases, to examine whether they affect investors' long-term success. The main aim is to analyze how an investor affected by the Gambler's Fallacy performs in the market in a long run. Possible types of investors, mostly inspired by psychological research by Barron & Leider (2010) and Boynton (2003), are presented in detail in the next section.

3.1 Description of Investors

Out of the four investors' strategies presented in this thesis, two are inspired by psychological findings about the Gambler's Fallacy. One generates his decision randomly and two of them are trivial deterministic investors who make no decision. Next, we describe their description in detail.

3.1.1 Random Investors

Random investors, as their name indicates, decide in each period on a perfectly random basis. Their decision-making has a binomial distribution with the probability of one half. We can imagine that in the beginning of each period, they toss a perfectly calibrated coin. If the result is a head, they buy, if it is a tail, they sell the stock. These investors are fully random and are not affected by any previous evolution of the stock prices. In our analysis, we use random investors as a benchmark to compare and contrast with investors who are affected by the Gambler's Fallacy.

3.1.2 Investors with the Gambler's Fallacy Bias

For simplicity, investors affected by the Gambler's Fallacy are called 'Gamblers' in this thesis. Behavior of such investors, is inspired by Barron & Leider (2010). The authors describe investors who believe that the paths of the price evolution are not independent of each other, therefore, they do not follow a random walk. Gamblers perceive one-period investment as drawing balls of two different colors from an urn without replacement. They also believe in the so called Law of Small Numbers, as explained in Chapter 2.

As a consequence, they assume that each subsample of the time series should have the same distribution as the whole price evolution time series. In case of binomial distribution, it can be translated as having the same parameter of the probability of success. Due to this bias, Gamblers feel that continuing trends should reverse. I.e., after observing several periods of growing prices, Gambler's perceived probability of observing a decrease in the stock price in the next period is higher than for an increase.

To quantify this bias, we use the results of Barron & Leider (2010). Authors experimentally examine how people are affected by the Gambler's Fallacy in predicting random outcomes. Participants of their research were asked to predict the next result of a virtual roulette wheel spins, whose outcomes could be one of two colors. In every period, participants were shown the most recent results of the roulette wheel spins and they were asked to make their predictions based on these results. Authors studied the probabilities of predicting certain result depending on the preceeding outputs. The findings of this paper could be found in Table 3.1.

It is well known that the results of a roulette wheel spins are independent of each other and both colors have the probability of 50% to be observed in each

Table 3.1: Conditional probability based on the length of previous series of the same color.

Length of previous series of one color	1	2	3	4
Probability of the same color in the next turn	0.438	0.456	0.369	0.381

trial. Number zero, which has a different color is neglected in this experiment. However, even after just one observation of a red number, the expected probability of observing it once more decreases to 0.438, according to the experiment. The results also show that the longer the previous streak of one color is, the lower is the perceived probability of observing the same color in the next spin. The longest pattern evaluated by Barron & Leider (2010) is of length four. If the whole pattern consists of results of one color only, the perceived probability of observing the same color once again decreases to 0.381.

In this simulation, we suppose that the evolution of the stock prices is independent of its history, and its increases and decreases have the Binomial distribution with the probability of one half. This assumption is discussed in Section 2.2. What more, investors as well as roulette players are trying to maximize their wealth by predicting the result of the observed random variable. Thanks to these characteristics of the experiments, we can compare the evolution of the stock with the roulette wheel spin and we can use the results of the paper to simulate the probabilities' perception of the investor.

We use these results to simulate such a behavior in the market. Before every decision-making, we observe how long pattern of only '1' (growths) or only '0' (decreases) of the stock prices has been produced prior to current period. This observation is used to choose the corresponding probability parameter of the binomial distribution to generate the next decision, i.e., whether to buy, or sell this stock. For example, if the previous evolution of the stock prices is '1011', the number of previous consecutive growths is 2. Therefore according to Table 3.1, the perceived probability of observing '1' in the following period is 0.456. Consequently, the decision to invest or not in the market simulated by a random wallk is generated from the Bernoulli distribution with parameter $p = 0.456$. Similarly, the perceived probability of observing '1' after the sequence '1000' is calculated as $p = 1 - 0.369 = 0.631$. Therefore, in this situation, the decision of a Gambler is generated from the Bernoulli distribution with the corresponding parameter. In the first four periods the previous evolution paterns of length of four cannot be fully observed. In this situation, Gamblers do not have enough information to make decision based on their strategy. Instead, they decide on a random basis like the random investors in Section 3.1.1.

3.1.3 Investors According to Boynton (2003)

We call 'Boyntons' the investors, whose decision-making is inspired by Boynton (2003). The participants of this study were USA college students of a psychology class. They were shown 100 cards with either a pink, or a blue dot. And then they were asked to predict the color of the dot on the following card, based on the colors of the previous cards. This experiment is similar to that of Gambler with slight differences.

The contribution of this paper is that authors distinguish two cases: whether the strategy was a success in the previous time period or when it was a failure. The probabilities of observing one color after a certain previous output are presented in Table 3.2. For simplicity, we encode pink as '1' and the blue as '0'.

Table 3.2: Conditional probability of obtaining 1, based on the previous outcome.

Previous outcome	0001	1001	0101	1101	0011	1011	0111	1111
$\Pr[1 \mid success]$	0.69	0.70	0.59	0.65	0.66	0.61	0.59	0.59
$\Pr[1 \mid failure]$	0.70	0.71	0.59	0.51	0.46	0.34	0.33	0.17

This time, we focus once more on the number of the four previous cards. Table 3.2 shows the probability that a student will predict the same number as in the previous trial based on the fact whether the last guess was a success or not. We can see a similar trend as for the Gambler: the more '1' were observed in the previous pattern, the lower is the perceived probability of Boynton to see it also in the next period. At the same time, these investors tend to learn from their experience: in case they predicted one and it was a success, they expect that it will be profitable again. E.g., in case there are four previous '1', such an experience can make a difference of $0.59 - 0.17 = 0.42$ in the probability of observing another one. For example, if the previous cards were '0010' and the previous estimation of the color was a success, the probability that a student will estimate '0' is 0.65. However, if the previous estimate was a failure, the probability is 0.51.

To simulate this behavior, in every step, we observe whether the stock prices increases or decrease in the four previous periods and whether the last investment was a success or not. These two variables identify 32 cases, an investor can be in: the 16 presented in Table 3.2 and the 16 situations with exact invers patterns. Using these results, in each step, we generate the investor's decision for the next period from the Bernoulli distribution with the parameter equal to the corresponding probability from Table 3.2. During the first four periods, where the history cannot be observed, Boyntons behave as a random investors presented in Section 3.1.1.

3.1.4 Buyers

For a global comparison of our simulations, we also use the trivial investing strategy of 'Buyers'. As their names imply, these investors do not make any decision about whether to buy or sell: no matter what happens, Buyers always buy the stock.

These investors would never exist in the reality, but for our experimental purposes, they help us understand the effect of only buying strategy that we can use as a benchmark. What more, these extreme strategies with no randomness can reveal some anomalies in the stock prices. If the prices follow a random walk, Buyers should have on average zero gain in a long run. On the other hand, if their gain is non zero, it helps us estimate the long term mean of the stock price changes.

We can also define an oposite strategy based uniquely on shortselling the stocks. However, due to the symetry of payoffs, the gain of this strategy is exacly the same as the Buyers' one, with the oposite sign. Therefore, for simplicity, we present only the results for Buyers, in the further analysis.

3.2 Description of the Random Walk

To evaluate the final wealth of investors on an efficient market, we also have to simulate the underlying market with its evolution of stock prices. In Section 2.2, we have shown that the market evolution can be simplified to a series of random independent events. In this case, by the event, we mean a change in the stock price during one single time period that can be either positive, or negative. Therefore, the stock prices can be simulated by a random walk with a finite mean and variance. Our aim is to analyze the performance of all investor types presented in Section 3.1, and to compare their final payoffs.

As the stock prices have an increasing trend in the long term, we can use a positive mean corresponding approximately to the level of inflation. In our simulation, we choose six different values of a drift, both, positive and negative (0, 0.02, 0.5, -0.02, -0.5) to evaluate how its magnitude changes the investors' performance. The drift of 0.02 is the one corresponding to the theoretical long term inflation on a global market. The value of 0.5 is used to emphasize the drift and see better its effect. We simulate the random walk as follows:

$$x_t = x_0 + t \cdot \mu + \sum_{i=1}^{t} z_i, \tag{3.1}$$

where the variable x_t denotes the stock price in time t. μ is the value of the potential drift and z is a random variable following a normal distribution with zero mean and unit variance.

We are interested in the average final wealth of each investor if the time series is sufficiently long and the number of observations sufficiently large. In our setup, the final wealth of the investors is a sum of every single investment outcome during the time period observed. This sum can be positive as well as negative.

3.2.1 Rules of Investments

To simplify the simulations, in the whole analysis, we consider the investors to have a sufficiently large wealth, so that they cannot go bankrupt, they can only get themselves into debt. At the same time, we allow them to sell a stock that they do not own at the moment, in other words, shortselling is allowed. In each period, investors choose whether they are willing to buy, or sell the observed stock. They make this decision according to the stock's historical evolution, their type and psychological bias. In the basic setup, investors do not have the possibility not to invest, i.e., in each period, they need to either buy or sell the stock. Also, the transaction cost and other administrative costs are neglected.

In the simplified approach we use, investors cannot reinvest the money, they gain. They can either gain or lose exactly the difference between the current price and the price in the most recent period. More precisely, if investors decide to buy the stock while the price increases, or they sell the stock while the price decreases, they gain the difference. If they buy the stock while it dectreases or sells it while the price increases, they lose exactly the same amount.

However this approach might seem unrealistic, as the fees are neglected, the real life 'buy and hold' strategy is feasible. It is only split into several consecutive actions. For example the strategy of buying a stock in 9 consecutive periods and sell it the 10^{th} one is equivalent to the the realistic strategy of buying it in the first period and holding it till the 9^{th} period, when finally selling it.

The approach we use, can be in reality compared to a situation, where investors are given one unit of wealth at the beginning of each period. They directly decide whether they want to buy or shortsell the stock for the following period. However, at the end of the period, they must give us back the one unit of wealth they received at the beginning. Their gain or loss is just the interest resulting from the investment.

Although we employ a simplified investment simulation, it remains a reliable representation for stock market transactions for our analysis. Using the same rules for each investor allow us to compare the final outcomes between them, which is the aim of our analysis. Furthermore, using a unitary wealth enables us to evaluate the direct impact of buying or selling a specific stock, independent of previous investment successes or failures. On top of that, this simplification does not influence the sign of the gain from the particular investment so we can still asses if the decision to buy or to sell was a good choice or not. Consequently, we are able to determine which investor did the right or wrong decision more often.

3.2.2 Theoretical Approach

While operating in a random walk, intuitively, the random investors should attain zero gain on average. The cumulative gain/loss of Buyers is the sum of every single difference between the prices in two consequent periods of time. Less intuitive results would be observed for the Gamblers and Boyntons, as their decision-making is more complex and takes also the previous evolution of the stock into consideration.

By the definition of the random walk, single steps are independent events. This means that their conditional probabilities of a growth based on the evolution in the previous period are equal to the posterior probability of a growth.

Therefore, we can define q as a probability of grow and $1 - q$ the probability of increase.

No matter which decision-making process is behind the investors' behavior, they finish by choosing between buying and selling the stock. Once they choose, they can observe how the price have changed. Therefore, every decision of the investor to buy or sell the stock for a given period can be described as a random event with a probability p to buy and $1 - p$ to sell. These properties allow us to analyze all combinations of outcomes of these two variables and their corresponding probabilities. We provide the overview with respective payoffs in Table 3.3.

Table 3.3: Payoffs of an investor operating on a random walk with drift equals to q.

Price change	Buy ($P = p$)	Sell ($P = 1 - p$)
Increase by X	X	$-X$
Decrease by X	$-X$	X
Payoffs of an investor	$X(2q - 1)$	$X(1 - 2q)$

We can see that the success of an investor in every period is also an i.i.d. random variable which has a Binomial distribution with parameter q for buying strategy and parameter $1 - q$ for selling strategy. In case of a basic random walk with no drift, the probability q is equal to 0.5. Then we can easily see that both strategies, buying and selling, has the same expected payoff equal to 0.

For further analysis of different strategies, we define a new dummy variable which shows whether the investor had a gain or loss in the last period. As well as the payoff, this variable also depends on the market evolution and decision made by the investor. On the other hand, it focuses only on the fact if the investment generated a gain or a loss and not it's value. Investors make a gain if they are buying a stock growing in price or selling a stock which's price decreased and they make a loss in the opposite situations. In this example, we once more denote by q the probability that the stock price grows and by p the probability, that an investor decides to buy the stock as we did in the previous table. As we explained in Section 3.1, p depends on the type of investor and the evolution of the stock prices in recent periods. All the possible cases and corresponding probabilities of the success variable are shown in Table 3.4.

Based on the last row of Table 3.4, we can conclude that the variable of

Table 3.4: Probability of a success and a failure of an investor operating on a random walk with drift equals to q.

Strategy	Success	Failure
Buy	pq	$p(1-q)$
Sell	$(1-p)(1-q)$	$q(1-p)$
Sum of probabilities	$1-(p+q-2pq)$	$p+q-2pq$

success has a Bernoulli distribution with the parameter equals to $1-(p+q-2pq)$. Therefore, the dummy variable of success is an independent, identically distributed variable with a finite mean and variance. This property allow us to use the Law of Large Numbers and the Central Limit Theorem for further analysis.

Application of Law of Large Numbers

The Law of Large Numbers is the theorem that gave name to the Law of Small Numbers bias explained in Chapter 2. This theorem describes what happens when we repeat a large number of independent and identically distributed random events. Essentially, the theorem tells us that as the number of repetitions increases, the mean of the outcomes will converge in probability to the expected value of the underlying variable. Proof of this theorem can be found for example in Anděl (2002) and it was further studied in Mason (1982).

Thanks to the assumption that the result of the investment in every single period is an i.i.d. variable, we can appy the Law of Large Numbers to this problem. We can conclude that the average final wealth of one type of investor after a sufficient large number of realizations converges to the expected wealth corresponding to his strategy on such market. Therefore, the arithmetic average over all the random walks and all the realizations is a good estimate for the expected value of a chosen strategy in a certain type of market.

Application of Central Limit Theorem

As explained in Anděl (2002), Central Limit Theorem states that if we observe a large number of i.i.d. random variables, each with a finite mean and variance, their sum will tend to follow a normal distribution, regardless of the distribution of the original variables. In other words, as the number of observations increases, the distribution of the sum of the variables will converge to a normal distribution.

Aplying this theorem to the case of investors' gain, we can conclude that the sum of gains resulting from their investments has a normal distribution. This statement is particularly useful while comparing the results of distinct types of investors. Thanks to the normal distribution, we can also use the asymptotic t-test to evaluate the hypothesis of equal means of the distributions of investors' gains. For proof of this theorem, we refer to, e.g., Anděl (2002). More detailed explanation of the Central Limit Theorem for Binomial distribution can also be found in Kwak Sang Gyu (2017).

Thanks to the assumption that the gain of an investor in every period is an i.i.d. random variable, we can apply the Law of Large Numbers and the Central Limit Theorem to our problem. This allow us to conclude that the final gain of an investor operating on one market, represented by one random walk, has a normal distribution with a mean equal to the expected value of the relative strategy. Applying these theorems to the results presented in Table 3.3, we can observe that in case of a random walk with no drift no one of our investors could make a long term profit. Their average payoff has a normal distribution with mean equal to 0.

Random Walk with Drift

We can suppose that in reality, stock prices do not follow a basic random walk with zero mean. Mainly due to inflation, the prices seem to have a certain long-term growing trend. This effect can be incorporated in our analysis through a drift, represented by the parameter q in Table 3.3. In case of a random walk with a drift, this parameter can be derived from the actual value of the drift. Using (3.1), we can see that every path of the stock price is normally distributed around the mean equal to the drift. The stock grows when $z_t + \mu > 0$. Consequently, the value of q can be calculated as the quantile corresponding to the value of 0 of the normal distribution with mean μ. This is equivalent to the quantile of $-\mu$ of the normal distribution with mean 0. For example, in case of the drift equal to 0.5 the expected value of the probability, that the stock will grow in a period is equal to 0.69.

If the drift is positive, the q is greater and consequently, according to Table 3.3, the buying strategy gets more profitable. From the symmetry of payoffs of the buying and the selling strategy, we can see that random investors end up on average with no gain or loss as in the previous case. In other words, in 50% of cases, their gain is higher by the drift and in the other 50%, their looss is greater by the drift.

Other investors' payoff depends mostly on how often they make the decision of buying the stock. In Table 3.1 and Table 3.2, we can see that continuing, only increasing trends, decrease the probability that the investor will choose to buy that stock. In the situation of a negative drift, they contrarily tend more to choose to buy the stock that is the losing strategy. Therefore, those two investors should in theory perform worse than the random investor, while operating on a random walk with a drift.

3.2.3 Simulation Based Approach

To examine and quantify the final payoffs of different investors empirically, we simulate their behavior on a market following a random walk, by the approach explained in Subsection 3.2.1. To neglect extreme realizations, we generate 500 random walks with the same parameters of length 500 time periods. Then, based on the behavior of each investor type presented in Section 3.1, we simulate 500 times their operation on each market represented by the random walk. Finaly, we calculate the mean of all 250 000 trial's payoffs to obtain the average gain of a particular type of investor.

We simulate each type of investor on random walks with zero, positive and negative drift. To see better the effect of the drift on the final wealth, we choose the drift of 0.02 and 0.5 in the simulation. The first corresponds to the average inflation and the second is chosen significantly higher to show better the effect of the drift. The final average payoffs are shown in Table 3.5 and their distribution is presented in Figure 3.1.

Table 3.5: Average final payoff of different investor types in the market simulated by random walks with different drifts.

drift	0	0.02	−0.02	0.5	−0.5
Random	0.006	0.064	0.074	0.035	0.004
Gambler	−0.158	0.111	−0.344	−19.711	−20.191
Boynton	−0.092	0.224	−0.028	−26.796	−26.669
Buyer	1.329	10.482	−10.513	250.002	−250.033

We may notice that regardless the drift, a random investors end up with zero gain on average. Buyers gain or lose on average an amount equal to the drift times number of periods. For Gamblers and Boyntons, the result supports our hypothesis that due to their erroneous understanding of probabilities, they end up with a loss. Due to the Gambler's Fallacy affecting their decision-making, both Gamblers and Boyntons underperform random investors. This corresponds to the theoretical findings from Subsection 3.2.2.

Thanks to the property of normal distribution of payoffs, we can test their equality by two-sampled t-test. As the random investors are our benchmark, we test the equality of their payoffs with each other to determine if their final payoffs significantly differ.

Table 3.6: P-values of two-sided t-test evaluating the equality of final payoff of random investors and others.

drift	0	0.02	-0.02	0.5	-0.5
Gambler	0.527	0.780	0.015	0.000	0.000
Boynton	0.894	0.555	0.695	0.000	0.000
Buyer	0.199	0.000	0.000	0.000	0.000

We can see that while operating in the market with no drift, we cannot reject the hypothesis of equality of payoffs of the random investors and the others. Their differences are not statistically significant on the 0.05 confidence level. On the other hand, if we consider a random walk with a non-zero mean, the difference became more significant. We can see that for both, positive and negative drift, the p-values are close to zero. Consequently, in this case, we can reject the hypothesis of equality of final payoffs of the random investors and the others. Therefore, we can conclude that the drift plays a significant role while comparing different investor types. This analysis shows us an evidence that the Gambler's Fallacy causes a loss while the stock price follows a random wallk with a non-zero drift.

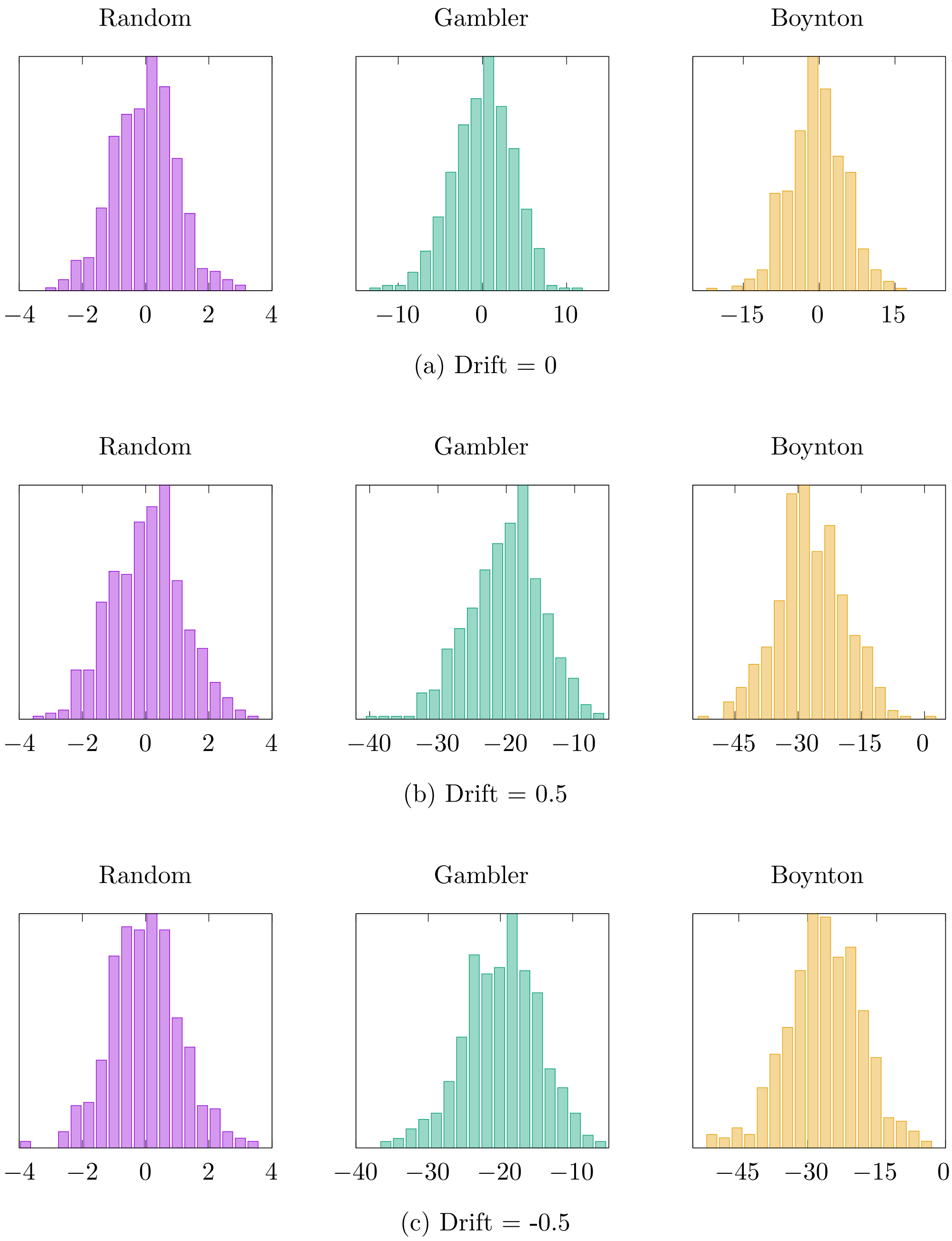


Figure 3.1: Distribution of final wealth of each non-deterministic investor type operating on a random walk with different drifts.

Chapter 4

Evaluation of Historical Data Distribution

In Chapter 3, we study the Gambler's Fallacy by the theoretical and simulation based approach using random walk to represent stock price evolution. In this one, we no longer use a generated random walk. Instead, we use the real S&P 500 historical data and calculate potential gain of each investor on this market to evaluate their performance. We also study the historical market data by the frequency analysis to explain profitability of different strategies on this specific market. The aim of this chapter is to find the evidence whether the Gambler's Fallacy thinking is or is not supported by the real historical data. Based on these results we are also able to evaluate if the S&P 500 stocks seem to follow a random walk presented in Chapter 3 and therefore, if the market can be considered efficient.

4.1 Theory

In Rabin (2002), author introduces a situation of generating number of observations with two possible outcomes '0' and '1' with the respective probabilities of θ and $1-\theta$ to occur. Those events are independent and identically distributed which means that for all $t = 1, 2, 3...$, $\Pr[s_t = 0] = \theta$ and $\Pr[s_t = 1] = 1-\theta$. The author presents an imaginary individual who is trying to predict the evolution of other outcomes according to his beliefs about the probabilities. The individual is fully Bayesian. It means that he associate a probability, represented by a number between '0' and '1', to each event he is facing. This representation corresponds to the plausibility of the event to hapen. However, he is a believer in the Law of Small Numbers. He believes that the events are not independent in reality. In his beliefs, we are not observing a subsequence of results of an infinite number of mutually independent trials. From his point of view, the number of realizations of the repetition is a certain integer N and the realization is a ball picking without replacement.

Random Walk Theory and Weak form of Efficient Market Hypothesis described in Chapter 2 present one possible explanation about market prices evolution. Random Walk Theory states that changes in the stock markets are random and that past evolution prices cannot be used to explain the current price. Random Walk Theory implies that the market is efficient. It means that all available information is incorporated in the price. If we supose those two theories to be valid, the stock prices have no memory. Therefore, future stock prices can't be explined by their past evolution and consequently the changes are independent on each other. One of the properties of mutually independent data prooved in Chapter 2 is that their conditional probabilities are equal to their posterior probabilities. In our case of generating an infinite series of '0' and '1' from the Bernoulli distribution with the probability parameter θ, it can be mathematically written as follows. The numbers in the condition are the results of previous realization of the random effect.

$$
\begin{aligned}
\Pr[0 \mid 0] &= \Pr[0 \mid 1] = \Pr[0] \\
\Pr[0 \mid 00] &= \Pr[0 \mid 01] = \Pr[0 \mid 10] = \Pr[0 \mid 11] = \Pr[0] \\
\Pr[0 \mid 000] &= \Pr[0 \mid 001] = \Pr[0 \mid 010] = \ldots = \Pr[0 \mid 111] = \Pr[0] \\
&\ldots
\end{aligned}
\tag{4.1}
$$

As a consequence, in a sufficiently long period, all possible sequences of the same length should be observed with the same frequency. For example there should be approximately the same number of sequences '000', '001', '010' and '011' in the whole sample. If there are some of them considerably missing or observed with a lower frequency, it can mean that the market is not evolving as a perfect random walk. This would signify that the Weak form of the Efficient Market Hypothesis is not satisfied and investors have some more information that is not incorporated in the past prices and returns. In this case, there would be an opportunity for an investor to consistently earn more than the market by using a strategy based on his additional information.

4.2 Description of the Data

In our sample, we use weekly stock prices data form April 2012 to April 2022 of all the S&P 500 constituents downloaded from Eikon (2022). To ensure the consistency of the data, we use only the stocks with existing data for the whole observed period. This means that we remove 51 companies with missing values and our final dataset consists of 455 companies. For further results we subdivide the companies into 11 industry sectors to test for their characteristics.

4.3 Methodology

Similarly as in Chapter 3, we simulate 1 000 investment trials of all the investor types on the same dataset of S&P 500 historical prices. For that, we use the approach presented in Section 3.2.2. In each period, investors decide separately for each stock, whether they want to buy or sell it according to its previous evolution. In our setting, investors absorb the particular gains and losses and they do not reinvest their wealth. The wealth, they gain or lose is exactly the difference of the price in the previous and current period. We also assume investors to have enough money, so they cannot go bankrupt and they always have the possibility to buy and also to sell. At the same time, in each period, they can invest only one unit of wealth.

Investors decide separately for each stock in every period whether to buy or sellit, so that we can calculate their final wealth for each stock. As their decision-making is stochastic, we repeat this procedure 1 000 times to obtain 1 000 results for each investor and stock combination. By calculating means of these results, we obtain an average payoff of one type of investor investing in one particular stock. Finally, we calculate means through all the stocks to obtain an average gain of each investor operating on all of the S&P 500 constituents. These results can be compared to those in Table 3.5.

In the second step, we evaluate how the results vary among different industries. We divide the S&P 500 stocks into 11 categories according to the industry, they are operating in. Then, we calculate an average gain of each investor resulting from investment into stocks of different groups. The aim is to see whether the success of Gambler's strategy can differ among industries and if there are industries with price evolution suitable for Gamblers.

4.4 Results

The distribution of average gains resulting from investments into S&P 500 index constituents are shown on Figure 4.1 for each investor type. The results of the random investor, Gambler and Boynton represent the mean calculated through all the 1 000 trials on each stock. On the other hand, the decision-making of Buyers are deterministic and there is no element of randomness. Therefore, Figure 4.1 shows the distribution of the exact values, they earn while operating on different stocks of S&P 500. The final equally weighted average gains or losses of each type of investor operating on all constituent stocks are shown in Table 4.1.

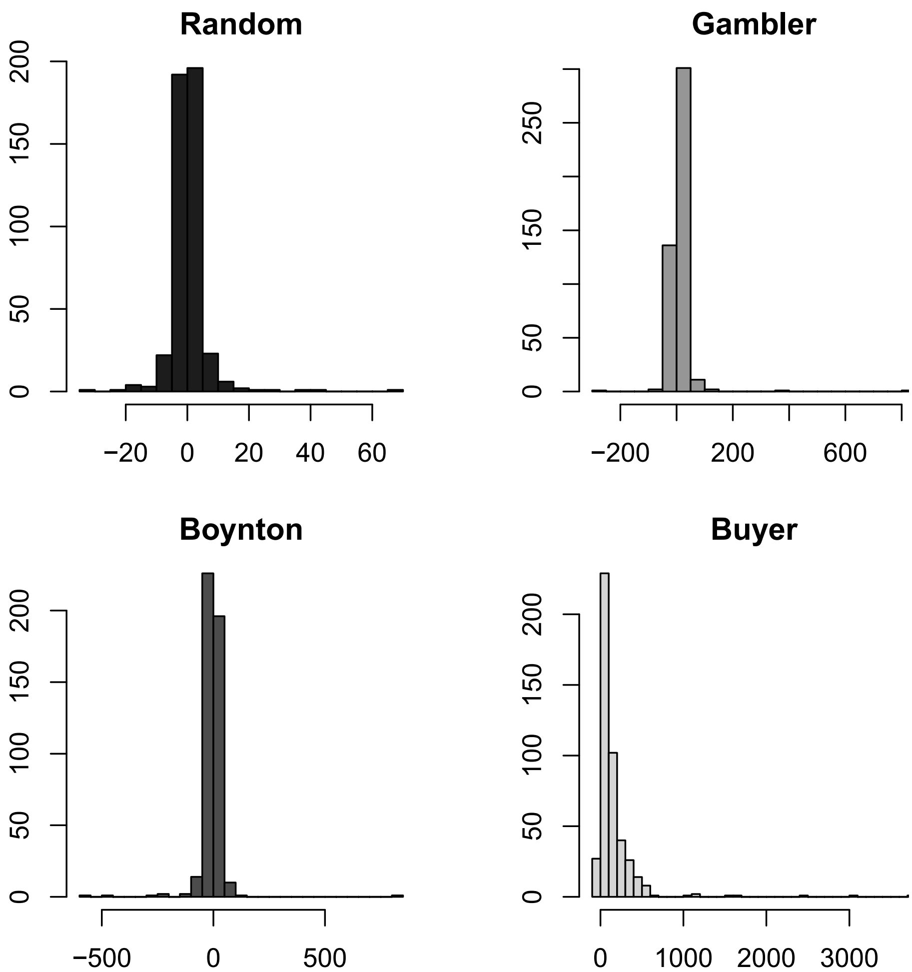


Figure 4.1: Distribution of average final wealth of each investor type operating in S&P 500 index constituents.

We can conclude that based on Table 4.1 and Figure 4.1, if we disregard the trivial strategies of Buyers, Gamblers have the most profitable one. By

Table 4.1: Sum of final wealth of each investor type operating in S&P 500 index constituents.

Random	Gambler	Boynton	Buyer
−74.17	3 767.43	−1 960.41	70 481.32

using two sided t-test, we test the hypothesis of equal final wealth of Gambler and random investor and we can reject it on the significancy level of 0.05. The same results are obtained for the Buyers. Only the equality of distribution of final wealth of random investor and Boynton cannot be rejected on this level.

The results also significantly differs from the theoretical ones from Section 3.2.2 and the simulation based ones from Section 3.2.3. We can therefore doubt if the S&P 500 market follows a random walk as we assumed in Chapter 3.

Industries results

From the values of the Buyer shown on the top of Figure 4.2, we can observe that all the industries were on average growing during last ten years. The average stock prices have grown the most in the Consumer discretionary and Healthcare sectors. Those are also the ones where the Gambler has the highest final wealth.

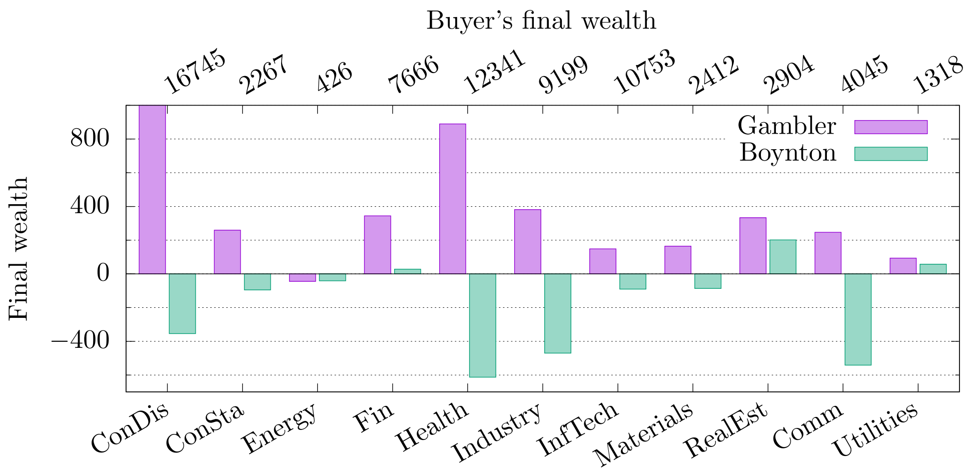


Figure 4.2: Average gains of investments into stocks from different industries of S&P 500 index.

We may notice that the only industry where the Gambler has final loss is the Energy sector. This signifies that evolutions of stocks operating in this sector is different from the others and that this industry was not a good investment choice for a believer in the Law of Small Numbers in the past years. At the same time, we can see that this is the industry with generally the lowest growth during the observed period.

Interesting result is also observed for the Real Estate sector, in this industry, all stochastic investors end up with a positive gain. This is the one that fits the best the decision-making of the Boynton, however it is not the most growing one. These inconsistencies can be explained by different distributions of the stock prices increases and decreases in corresponding industries. It can signal that stock prices in these sectors do not follow a random walk, some patterns appear more often and therefore there exists a strategy based on historical prices that generates a positive gain.

Generally, we can see that Gambler ends up with a positive gain, however the random investor has long term gain of 0. In Chapter 3, we showed that if the prices follow a random walk with or without a drift, Gamblers cannot have a positive gain. These results therefore contradict the Random Walk Theory so we can also doubt about efficiency of the S&P 500 market. It seems that not all available information is incorporated in the price and that its historical evolution has an effect of the curent prices. This information can be partialy used to predict the future prices. The results show us that these anomalies in prices evolution are generally profitable for investors with a strategy driven by Gambler's Fallacy.

4.5 Frequency Analysis

To study the theory presented in Section 4.1, we also use the frequency analysis of price decreases and increases. For the whole analysis, we use the same dataset as described in Section 4.2.

4.5.1 Methodology

To simplify the situation, we transform the series of weekly prices to a sequence of '0' and '1'. The observation is '1' if the price increased compared to its value in a previous period and we associate number '0' to periods when the variable decreased in the weekly comparison.

First, we count the number of '0' and '1' in the sequence, which tells us the proportion of these two numbers in the outcome. Then, we observe how many times is each of the possible subsequences of length 2 present in the series. By this analysis, we are able to determine the alternation rate, whether it contains the same number of periods when the direction of the evolution did change (‘01’, ‘10’) and when it did not (‘00’, ‘11’). Significant dominance of alternating subsequences could signify an opportunity for investors affected by the Gambler’s Fallacy believing in the Law of Small Numbers. It would mean that the stock prices significantly oscillate and reversals are more likely to be observed than streaks.

We continue with the frequency analysis of longer subsequences till the length that makes sense according to the overall length of the output. By the characteristics of independent events described in Section 4.1, all outputs of the same length should have the same probability to occur. Thus they should be observed approximately with the same proportion in the sample. If it is not the case, it means that the property of conditional probability of independent events does not hold. Consequently, the evolution of the prices could not be an independent event following a random walk with zero mean.

Random Walk with Drift

As we can see from the result of Buyers in Table 4.1, there is a strong probability that the historical prices follow a random walk with a positive drift described in (3.1). In this case, there shoul be a constant difference between the same patterns differing only in the last number. Probability of observing one particular patern would depend on the number of ones or zeroes contained in this patern. As we showed in Chapter 3, if the prices followed a random wallk with a drift, the strategies of Gamblers and Boyntons should be less profitable than the one of Random investors, whose gain would be zero.

The longer the analyzed subsequence is, the less it is likely to observe the same proportion of possible outcomes. However, any large anomaly or shift of these probabilities could examine the validity of the Weak form of the efficient market theory and it could represent an opportunity for an investor to beat the market. We run more detailed analysis of distribution of different paterns using frequency analysis.

4.5.2 Results of Frequency Analysis

From the result of all stocks together presented in Figure 4.3, it is visible that most of the stocks were on average growing, because the subsequences containing more ‘1’ than ‘0’ are more present. In theory, this could correspond

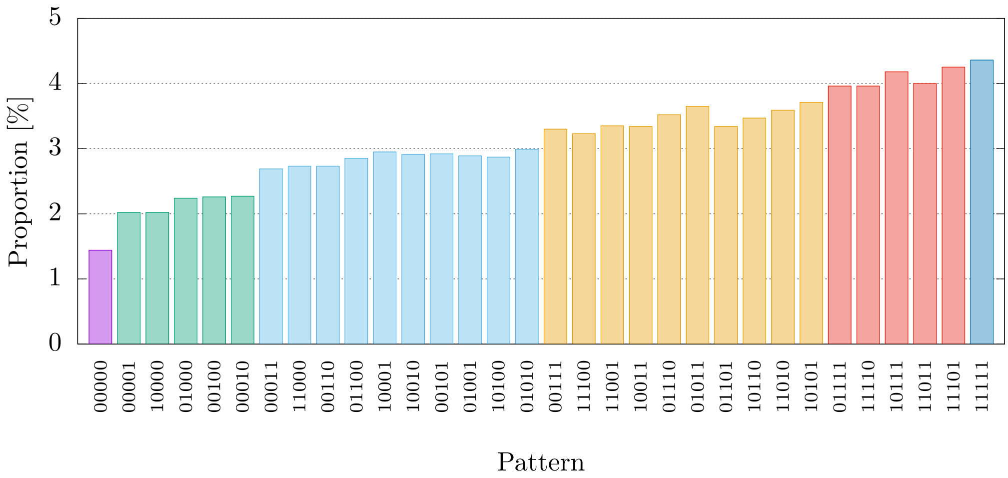


Figure 4.3: Proportion of patterns of length 5 in the S&P 500 constituent stock prices evolution between 2012 and 2022, grouped by number of '1'.

to the random walk with a positive drift. The proportion of frequences of length two and three support this hypothesis as all patterns with the same nuber of ones and zeros have almost the same frequency.

Both, Gamblers and Boyntons, observe four periods back into the history to make their investment decision. Therefore, the analysis of subsequences of length four, together with Table 3.1 and 3.2, show us how many times each of the investors decided with which probability of choosing to buy the stock. The most important graph for this analysis is Figure 4.3 and it presents the frequency of five-period subsequences. In this graph, we can see how many times an investment of Gamblers and Boyntons was a success or a failure.

It is at the level of patterns of length five, where we can doubt if the properties of i.i.d. events are satisfied. If the evolution of stock was a sequence of i.i.d. events followin a random walk with a drift, the probabilities of each pattern should be exactly the product of individual events' probabilities.

By observing the histogram of subsequences of length five, we may notice that the paterns with lower number of reversals have lower probability to be observed. For example, the paterns '00001' and '10000' were observed 4 802 times in our sample, whereas all other paterns containing one '1' have more than 5 300 observations. Same results could be observed for example for paterns containing three '1'. Here, we may observe that paterns with two '1' on adjacent positions have lower probability to be observed compared to patterns, where '1' are separated by '0'.

This observation explains the significant gain of Gamblers in Table 4.1. Gamblers believe in more reversals and adapt their strategy to their beliefs. That is why they profit from this overall distribution of all S&P 500 stocks. Based on these observations, we may conclude that the market does not seem to follow a random walk – neither with, nor without a drift. Therefore, this result is in accordance with the one obtained in Section 4.5. It also contradicts the Random Walk Theory and makes doubts about the efficiency of S&P 500 stocks.

4.5.3 Particular Stocks

Using this method, we can deeply study particular gains of different investors and compare their succes on different stocks. Looking at the results of the most growing industry , which is Consumer discretionary, we can focus on the Amazon and the NVR stocks. Their prices grew by a similar amount during the observed period. However, the average result of a Gambler investing in these stocks differ from 816.22 for NVR to -261.92 for Amazon. This is a result of differently distributed dectreases and increases of these two stock prices. Graphs of their evolution and frequency analysis are presented in Figure 4.4 and 4.5.

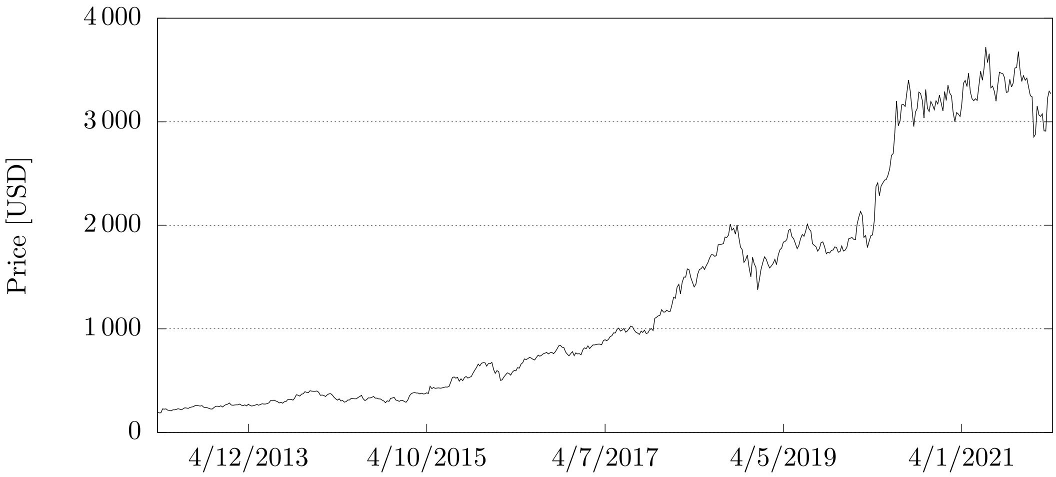
We can see that the evolution of Amazon stock price contains more patterns with only growing trend, compared to NVR. The price of NVR stock oscilated significantly more so the reversals were more present. This evolution fits well the Gamblers' strategy because they believe in reversals more than in continuing trends. This explains the large difference in Gamblers' gains while investing in these two stocks.

4.6 Summary

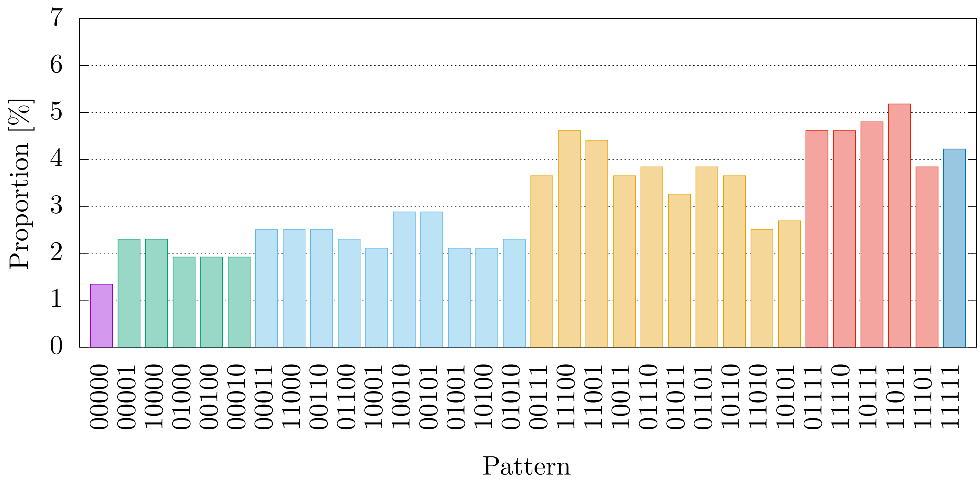
Based on both analyses, we can conclude that the S&P 500 index constituents have generally properties that are profitable for Gamblers. It means that the Gamblers' strategy is also more profitable than the one of random investors, whose average gain is in the long run zero.

The average gains of investors differ among different industries. For example, we determined that the energy sector was not profitable for an investor with the Gambler's strategy in the past years. On the other hand, this strategy can be highly profitable for Consumer discretionary sector. At the same time, there is a difference between the profitability of different stocks within the industries.

These results contradict the theory from Chapter 3, which shows that the more the stock grows, the lower the gain of a Gambler should be if the prices follow a random walk. We can therefore doubt whether the stock prices evolution really follows a random walk and if the market satisfies the Weak form of the Efficient Market Hypothesis.

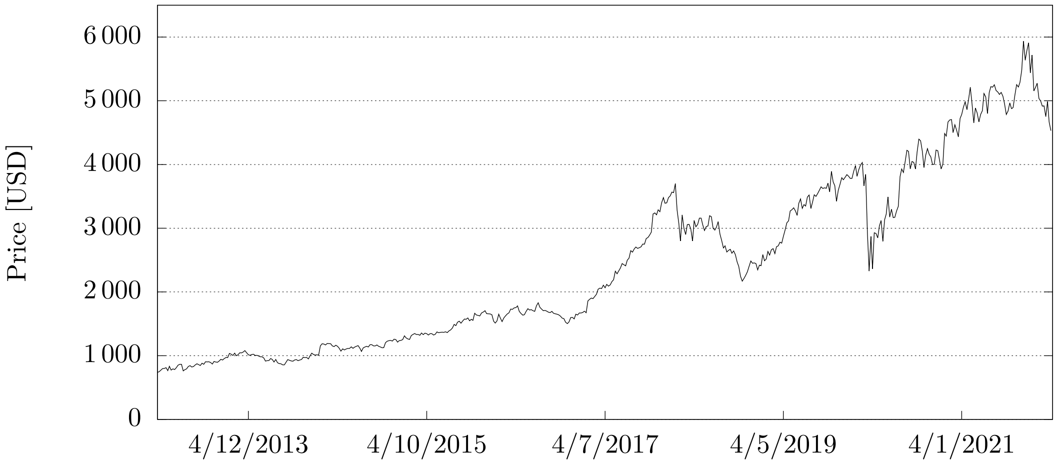


(a) Stock price evolution

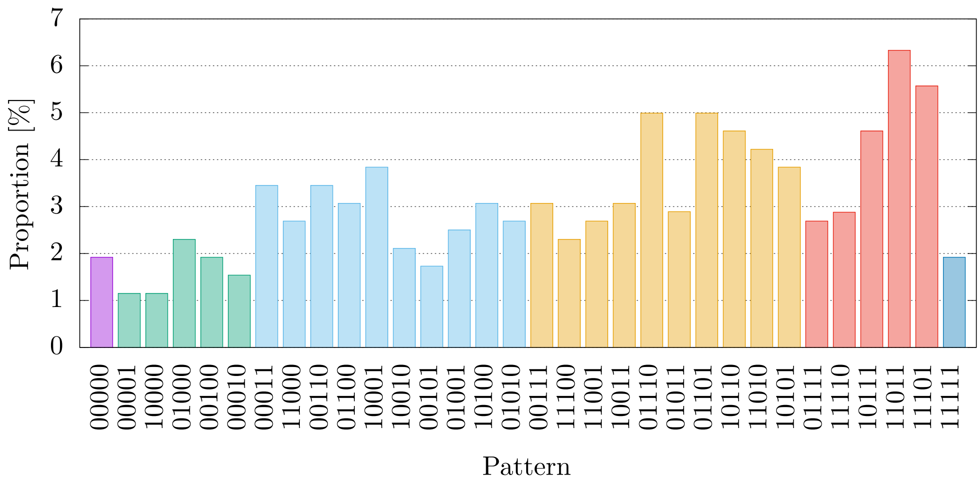


(b) Proportion of patterns of length five

Figure 4.4: Proportion of patterns of length five and price evolution of AMZN stock between 2012 and 2022. Patterns are sorted and groupped by the number of ‘1’.



(a) Stock price evolution



(b) Proportion of patterns of length five

Figure 4.5: Proportion of patterns and price evolution of NVR stock between 2012 and 2022. Patterns are sorted and groupped by the number of '1'.

Chapter 5

Fama & French Models

In this chapter, we study the presence of Gambler's Fallacy in the market using asset pricing models. We simulate different investment strategies based on available historical data and focus on their corresponding excess return. More specificaly, we are interested in the strategies based on historical quarterly EPS surprises (SURP) that detect how the forecast differed from the real price in the history. This difference serves as an indicator for investors while making investment decisions.

Similarly to previous analysis, our aim is to examine the strategies of all investor types presented in Section 3.1 and determine how profitable are the ones affected by the Gambler's Fallacy. We analyze whether the portfolios, created based on the Gambler's probability perception, can generate an abnormal return or not. This analysis can help us to examine whether the market evolves as a random walk or if taking the most recent evolution into consideration has some effect on the excess returns.

Our method is mostly inspired by Loh & Warachka (2012). Authors use the Fama & French three-factor model and Cahar four-factor model to measure excess returns of different strategies. We expand this analysis by more powerful five-factor model presented in Fama & French (2015). At the same time, we also test for different portfolio characteristics that better explain the effect of historical EPS evolution patterns on investment profitability.

In this chapter, we introduce a novel approach that links these models with Chapter 3 and Barron & Leider (2010). For this approach, all the patterns are classified according to their probabilities to be observed from the Gambler's perspective. Using this classification, we can test also the excess return of the Gamblers' strategy and the strategy resulting from the opposite intuition.

5.1 Asset Pricing Models

In the 1960s, Capital Asset Pricing Model (CAPM), the first model relating the required return on an investment to the systemic risk of that investment, was developed. It is a linear model that calculates expected returns of a portfolio based on the risk-free rate, market risk and market returns. This method is widely used in research for securities pricing and estimating excess returns of assets, given their risk. The main idea of CAPM is that investors should be compensated for the systemic risk and time value of money. Therefore, more risky assets are priced higher than the ones with a low risk.

Fama & French Three-factor Model

Fama & French (1993) expanded CAPM by including factors of size and value risk. This was a reaction to the finding that publicly traded companies with small market capitalization generaly outperform the large-cap stocks, and those with high book to market ratios generate in general higher returns compared to those with low book to market value. Consequently, a portfolio consisting of stocks with low market capitalization or high value would have higher returns than estimated by CAPM. The Fama & French model uses three factors explaining the returns of assets: Small minus big (SMB), High minus low (HML) and the portfolio's return minus the risk-free rate of return. The entire equation of the Fama & French three-factor model is presented in (5.1).

To calculate these factors, Fama & French (1993) rank the stocks based on their size and book-to-market value. They include all NYSE, Amex and NASDAQ stocks and classify them as small or big by comparing their size with the NYSE median size. As NYSE contains generaly bigger stocks, the 'Small' group contains a larger number of stocks. They range the stocks also by the book-to-market value and create 'Low', 'Middle' and 'High' groups containing respectively the bottom 30%, midle 40% and top 30%. By intersection of these criteria, they create six portfolios (S/L, S/M, S/H, B/L, S/L, S/H).

SMB factor estimates the risk factor of returns related to the size. It is defined as the difference of the simple average returns of the three groups of stocks classified as 'Small' (S/L, S/M, S/H) and the three groups of stocks classified as 'Big' (B/L, B/M, B/H). By this aproach, we obtain a difference in returns of 'Big' and 'Small' stocks with the same book-to-market value.

HML factor estimates the risk factor of returns on investment related to the book-to-market value of the stock. We calculate it by a similar approach as the

SMB factor. It is the difference between the simple average returns of the group of stocks classified into the same size group with high and low book-to-market value.

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \epsilon_{it} \tag{5.1}$$

In this equation, R_{it} denotes total return of a stock i at time t, R_{ft} represents the risk-free rate of return at time t and R_{Mt} the market portfolio return at time t. SMB_t and HML_t denotes the size premium and value premium respectively. $\beta_{1,2,3}$ are the factor coefficients and coefficient α_{it} measures the abnormal profit resulting from the coresponding investment compared to the market return. α_{it} is the value, we use to compare the performance of different strategies. Later, Carhart (1997) introduced a new fourth factor that helps to explain the returns by using momentum data.

Fama & French Five-factor Model

In Fama & French (2015) authors of the original three-factor model introduce its another improvement. They extend the original model by using also profitability factor and investment factor. They rely on the empirical findings that the return of an asset depends also on the firm's operating profitability and their investing strategy. The formal form of the five-factor model is presented in (5.2). The Robust minus weak (RMW) factor is calculated as a difference between the returns of portfolios of stocks with robust and weak profitability. Conservative minus aggressive (CMA) is a difference of returns of portfolios of the stocks of low and high investment firms. Thanks to these additional factors, this model has higher power in explaining the crossection variance of expected returns.

$$R_{it} - R_{ft} = \alpha_{it} + \beta_1(R_{Mt} - R_{ft}) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \epsilon_{it} \tag{5.2}$$

5.2 Methodology

Our analysis is inspired by Loh & Warachka (2012). The authors compare the expected excess return of investments into stocks based on their historical earnings surprises. Their aim is to determine whether there exists a strategy that can lead to an abnormal excess return. To evaluate this strategy, authors use the three-factor and the four-factor Fama & French models for portfolios sorted according to the sign and size of SURPs.

In our analysis, we focus mainly on the sign of the SURPs of a particular stock in particular time. Our aim is to determine if the difference between the concensus forecast and real EPS in the most recent periods somehow affects the excess return of a stock. In other words, we want to evaluate if the excess return differs between stock portfolios composed of stocks sorted according to their previous SURP evolution. We focus on SURPs, because it can be considered as an indicator of underpriced or overpriced stock. A positive SURP signifies that the stock was underpriced and the consensus forecast was lower, than the real value at the time of quarterly EPS announcment. By the same logic, a negative SURP is a sign of overpriced stock. In this case, the predicted price of the stock is higher than the real price.

5.2.1 Quintiles

For the first analysis, we proceed month by month through the whole time period. In every month, we divide the stocks into five quintiles according to the value of their latest SURP. The first portfolio then contains the most overpriced stocks with the lowest, potentially the most negative, SURPs and the fith contains the most underpriced stocks with the highest SURPs in the most recent observation.

As well as Loh & Warachka (2012), we work with quarterly SURPs data and monthly stock prices, therefore each time three months of prices are sorted according to the same SURP observation. Once a new, more recent, quarterly observation is available, we update the sorting procedure to this one.

Consequently, we determine whether the stocks are in a streak or reversal. We use the definition of Loh & Warachka (2012) and define a streak as two or more consecutive SURPs with the same signs. A reversal is defined as a SURP following a streak of opposite sign. Using this definition, it is evident that some of the patterns are not defined neither as a streak nor a reversal. Same as for the quintiles, the streak/reversal dummy variable changes with each new quarterly SURP observation, so each time three months of stock prices have the same classification. Based on these two criteria, we divide stocks into ten groups. We create five SURP quintiles of stocks which are in the streak and five SURP quintiles which are in reversal.

In the next step, we calculate the average return for each of the ten groups at each month. It is an equally weighted average of returns of all group constituents during that month. By proceeding like this, we obtain a monthly time series of average returns for each of the ten groups. As the SURPs change every three months, the quintile constituents also change. Therefore, the average is calculated from different companies in every quarter. Finally, we run the Fama & French three-factor and five-factor regressions explained in (5.1) and (5.2) for each of these time series. We use monthly Fama & French three and five factors to find the excess return correspondning to investments into the stocks with corresponding characteristics defined by the ten groups.

We are mainly interested in the results for the difference between the first and fifth quantile and how it changes if the stock is in a streak or reversal. For that reason, we simulate a behavior of investors taking a short position in a portfolio consisting of the stocks from the first quintile, and a long position in those from the fifth quintile. If they consider in their decision-making that high SURP stock would generate a higher profit, this strategy would be the best suitable for them.

To measure this, we create another monthly time series that contains in each period an average return of the spread. This return is calculated as a difference of average abnormal returns of portfolios from the fifth and the first quintile each month. By running the three-factor and the five-factor Fama & French regressions using this time series, we obtain the excess return that is likely to be gained by investing into the spread.

5.2.2 Signs

We also use a second, slightly different, approach to test these excess returns. The methodology for this one is exactly the same except one difference. In the first step we do not divide the stocks into quintiles. Instead of that, we are just interested in the sign of the SURP and we use a binary sorting into two groups. In the second step, we sort the stocks within the groups into ones in streaks and reversals exactly as in the previous analysis. Consequently, we obtain four groups of stocks that are in positive streak, negative streak, positive reversal and negative reversal. By the same procedure as in the previous approach, we create a time series of average returns for each group. Finally, we run the three-factor and the five-factor Fama & French method using these time series to compare the excess returns of stocks in different groups.

To quantify the abnormal return of an investor choosing a strategy based on the SURP sign, we calculate the return of the spread. Similarly as in the previous approach, we create a time series of average returns of the spread as a difference between the average of stocks with a positive SURP and a negative SURP. By using this time series in the Fama & French models, we obtain the excess return resulting from investing into the spread.

5.2.3 Probabilities

Inspired by Loh & Warachka (2012), we introduce a third, novel, approach based on the findings of Barron & Leider (2010), that have been already used in previous chapters to define a Gamblers' behavior. In order to better simulate the Gambler's behavior, we do not use the definition of streaks and reversals. Instead of that, we divide the stocks based on the Gambler's perception of probabilities explained by Barron & Leider (2010). In each period, we observe four previous values of SURP variable of a company and classify it as one of the possible patterns of the SURP evolution. Using Table 3.1, we associate a probability of a next growth perceived by the Gambler in this situation. For each month, we sort the stocks based on these probabilities into two groups. If the corresponding probability is lower than 45%, it is classified as 'low probability' group, and if the probability is higher than 55%, it is in the 'high probability' group. The cases, for which the associated probability of the growth falls between these two levels, are not taken into consideration in this analysis.

Each group is divided into two others according to the sign of the most recent SURP. Consequently, we create four groups of 'high positive probability', 'high negative probability', 'low positive probability' and 'low negative probability'. Similarly as in the previous approaches, the SURP evolution is updated quarterly. Each time, particular stock is classified into the same group during three months. As in the previous approaches, by calculating the monthly group return averages, we obtain four time series of average returns corresponding to each group. Finally, we run the Fama & French three-factor and five-factor regressions.

This approach helps us better evaluate the profitability of the Gambler's strategy. We are mainly interested in the difference of excess returns between the 'low probability' and 'high probability' classified stocks. A portfolio of stocks classified by a Gambler as less likely to grow has a high excess return. It means that Gamblers' behavior in the market is not profitable. On the other hand, if none of these groups of stocks have a significant excess return, it would signify that the return does not depend on the EPS surprises. To quantify possible gain resulting from the Gambler's strategy, we analyze the return of a spread between the 'low probability' and 'high probability' stocks. We create a monthly time series consisting of difference between the average returns of stocks from these groups. We use this time series in the Fama & French models to obtain the returns of the spread.

5.3 Description of the Data

Our dataset includes S&P 500 stocks' specific data during the period from June 2002 to June 2022. It contains monthly stock prices as well as quarterly EPS and their consensus forecast. Monthly stock prices of all S&P 500 constituents are downloaded from Eikon (2022). This resource is also used to download the quarterly EPS data (EPS) and the 12-month forward consensus forecast of quarterly EPS (DIEP). This variable is created using I/B/E/S 12-month forward earnings per share estimates for each of the index constituents. Using these data, we derive a new variable of Earning surprises (SURP) defined as the difference between the actual EPS and the most recent EPS consensus forecast.

To run the three-factor and the five-factor Fama & French models, we need to download the monthly dataset of the corresponding factors also in the time range from June 2002 to June 2022. Those monthly data are downloaded from the official freely available online data library French & Fama (2022). To work with this dataset, we transform it into a quarterly time series by selecting only data from March, June, September and December.

Three stocks (News Corporation Class B, Amcor, Fox Corporation Class B) are excluded from the dataset due to the missing stock prices data. At the same time, some of the S&P 500 stocks emerged between years 2002 and 2022. We include data about these companies since the year they entered the index and data are available. Same as Loh & Warachka (2012), we also remove all the stocks whose book value decreased under the value of 5$ during the observed period. It helps us to keep only stable stocks in our dataset. This requirement eliminates 46 stocks and results in a final dataset containing 455 S&P 500 index stocks.

Table 5.1: Alphas [%] of Fama & French models, Quintile approach.

	Three-factor model			Five-factor model		
quintile	streaks	reversals	difference	streaks	reversals	difference
Q1	0.128	−0.144	0.272	0.189	−0.114	0.303
Q3	0.412*	0.799*	−0.387	0.434**	0.701	−0.267
Q5	0.685**	0.657*	0.028	0.751**	0.717**	0.034
Q5 – Q1	0.556**	0.801		0.561**	0.831	

Asterisks describe the significancy of each result: '***' signifies the p-value lower than 0.001, '**' p-value lower than 0.01 and '*' lower than 0.05.

Table 5.2: Alphas [%] of Fama & French models, Sign approach.

	Three-factor model			Five-factor model		
SURP	streaks	reversals	difference	streaks	reversals	difference
positive	0.576*	0.519*	0.057	0.657**	0.537*	0.057
negative	0.383*	0.393	−0.010	0.414*	0.445	−0.010
difference	0.194	0.126		0.342*	0.092	

Asterisks describe the significancy of each result: '***' signifies the p-value lower than 0.001, '**' p-value lower than 0.01 and '*' lower than 0.05.

5.4 Results

Results of the three approaches are presented in Table 5.1, 5.2 and 5.3. All tables present and compare the three-factor as well as the five-factor Fama & French model results with their significancy.

In Table 5.1, we may spot that the strategy of shorting the stocks in streaks from the first quintile and buying those from the last quintile, generates a significant excess return of 0.561%, using the five-factor model. The same strategy for reversals generates a return of 0.831% with lower significancy. Similar results are also obtained by the three-factor model. However, none of them presents a significant gain from the spread between the streaks and reversals among the same quintile.

Similar results can be seen for the strategy sorting the stocks by the sign of the most recent SURP presented in Table 5.2. For streaks, the spread long positive, short negative generates a significant return of 0.342% for the five-factor model. The result for reversals is not significant and equals to 0.092%. The three-factor model generates only insignificant results for the spread gain.

These results partially confirm the finding of Loh & Warachka (2012). It

Table 5.3: Alphas [%] of Fama & French models, Probability approach.

	Three-factor model			Five-factor model		
SURP	high	low	spread	high	low	spread
positive	0.464	0.623**	−0.159	0.465	0.623**	−0.158
negative	0.337*	0.694***	−0.357*	0.336*	0.694***	−0.357*

Asterisks describe the significancy of each result: '***' signifies the p-value lower than 0.001, '**' p-value lower than 0.01 and '*' lower than 0.05.

supports the belief that the strikes of SURPs have mostly a continuing trend in the market. Consequently, the strategy of buying the stocks currently in positive streak and selling those in negative streak generates a consistent gain. This strategy is an opposite of the Gamblers' behavior, because they believe that the streaks will not continue, as explained in Chapter 3.

On the other hand, we get less significant results for reversals. This should be mostly due to the fact that we use a different dataset than Loh & Warachka (2012). First, our dataset consists of 455 companies, compared to 9 706 in their paper. Due to a more strict definition of a reversal, it is also rarer to observe a reversal than a streak. Therefore, in our smaller dataset, we do not observe enough reversals to obtain some significant results. Second, our dataset focuses on S&P 500 index constituents. Characteristics of those stocks are not representative for all US companies, on the other hand, by their size, they represent a high percentage of the the whole US market. However, the dataset used by Loh & Warachka (2012) contains much larger sample of US firms across different sizes and industries. On top of that, if a company encounters some financial problems, it is replaced in the S&P 500 index. That is why we are also facing a survivorship bias in our case which can also result in the lower number of observed reversals.

Results opposed to the Gambler's intuition were also obtained using the new, probability-based method. In columns 'high' and 'low', Table 5.3 shows excess returns resulting from investing into stocks that Gamblers perceive as highly probable to grow and highly probable to decrease in price, respectively. We can see a highly significant returns for the portfolios of 'low probability' stocks, using both, three-factor and five-factor models. Investing into stocks classified by a Gambler as low probable to grow leads to a significant excess return of 0.694% per month in case of negative SURP, and 0.623% in case of positive SURP. This result is the same for the three-factor as well as the five-factor method. On the other hand, stocks favorized by a Gambler generate on average an excess return of only 0.465% for positive SURP and 0.336% for negative SURPs. These returns almost correspond with the average abnormal return of the whole S&P 500 index for the observed period. This one was calculated to 0.392% using the five-factor Fama & French model with monthly time series containing the average return of the whole index for each month. These results show that the Gambler's strategy does not yield any extra profit. On top of that, it can be even loss-making compared to the investment into the whole index.

We also calculate the excess return of potential Gambler spread strategy. It means taking the long position in the 'high probability' stocks and short in the 'low probability' stocks. We can see that both, for the most recent positive and negative SURPs, the spread has a negative excess return. In particular, in case the stock has been previously over-priced, the excess return of the Gambler's strategy is significant and negative, equal to -0.357%. This result is also generated by both, three factor as well as the five-factor models. Therefore, the Gamblers' strategy is not optimal and the exacly oposite one generates a consistent profit thanks to the symetry of gains.

5.4.1 Comparison of Models

Although both, the approach mentioned earlier and the one described in Chapter 4 focus on the same index, they yield different results. Specifically, the Fama three-factor and five-factor models produce findings that contradict those obtained by the simulation-based approach. As detailed in Chapter 4, our analysis shows that investing in S&P 500 stocks using the Gambler's strategy can lead to higher returns than the one of random investors. However, our findings in this chapter suggest that the Gambler's strategy results in consistent losses.

The discrepancy in results can largely be attributed to the type of data utilized in each model. In the simulation approach, investors make decisions based solely on recent observations of relevant stock prices. On the other hand, the asset pricing models use more complex data, incorporating the dynamics of the entire market through the use of five factors. On top of that, investors in these models do not make decisions based on the evolution of price changes. Instead, they use SURP evolution, which is defined as the difference between the current EPS and the most recent EPS consensus forecast. Therefore, this approach is based not only on factual information about the market's evolution but also on subjective human predictions.

In addition, we must consider the different granularities of the datasets used in both approaches. In Chapter 4, we use weekly stock prices of last ten years. As the EPS forecasts are published quarterly, the asset pricing aproach uses quarterly data for last twenty years. It means that this approach, compared to the other one includes also data from the period of financial crisis that can cause differences in results. Another difference is that inspired by Loh & Warachka (2012), we remove all stocks whose book value decreased under the value of 5$ during the observed period as explained in Section 5.3. Therefore, there is also a difference in number of stocks included in the analysis.

5.4.2 Summary

One of the explications of the result given by Loh & Warachka (2012) is that the prices are influenced by the investors operating in the market. As they belief in the Law of Small Numbers and in the dependency of the price's evolution, Gamblers tend to underreact to trends. This underreaction is then reflected into the price evolution through a subsequent drift in prices. This means that our result could be caused by a significant proportion of Gamblers investing in the S&P 500 index constituents, which results in a price shift. Investors with the opposit strategy then benefit from this reaction.

By these results, we show that making decision based on observing how underpriced or overpriced a stock was in the history can lead to a consistent gain. It means that there exist a strategy, based only on historical SURP data that is profitable in a long term. This finding signifies that the SURPs of S&P 500 constituents do not follow a random walk and some patterns of the evolution are more likely to be observed than others. Due to the existence of a consistently profitable startegy, we can also question the efficiency of this market and doubt about validity of the Weak form of EMH.

Chapter 6

Conclusion

The primary objective of this thesis is to bridge the gap between finance and psychology by examining the impact of psychological biases on investors' profitability. Specifically, the Gambler's Fallacy and the Law of Small Numbers are evaluated. In addition, this thesis aims to quantify the gains of investors in a theoretical efficient market and compare them with the ones of the real market to examine its efficiency.

To achieve these goals, we analyze the profitability of four different types of theoretical investors, which are defined based on the results of psychological research related to the Gambler's Fallacy. One investor is fully random, two are biased by the Gambler's Fallacy and the last one use trivial startegies of only buying. The random one is considered as a benchmark and the two trivial ones are used to represent the extreme strategies and to show the dynamics of the market.

First, we evaluate the theoretical payoffs for investors operating in an efficient market that is represented by random walks with different drifts. This analysis serves as a foundation for examining the real market's efficiency in subsequent analyses. By employing both mathematical and simulation approaches, we demonstrate that, regardless of the drift's value, fully random investors achieve a long-term average gain of zero. For investors biased by the Gambler's Fallacy, the results depend on the value of the drift. Specifically, if the drift is zero, biased investors perform identically to random investors and, on average, they obtain no gain. However, due to their belief in the Law of Small Numbers, biased investors achieve less gain than random investors in both positively and negatively drifting markets.

To determine the applicability of these findings to real markets, we simulate the investment strategies of all investor types by using the S&P 500 constituents as the underlying market. We then compare each investor's potential average gain to the theoretical values obtained from an efficient market. Our simulations reveal that, over the long term, the strategy based on the Gambler's Fallacy is more profitable than that of a random investor. This outcome contradicts the theoretical prediction for an efficient market and raises doubts about the efficiency of the S&P 500 market.

This observation is supported also by the frequency analysis results, using the same data. It shows that the S&P 500 constituents' prices generally increased, indicating that the market does not conform to a random walk with zero drift. At the same time, we observe a low alternation rate between price increases and decreases, which favors the Gambler's strategy. This suggests that the market may not adhere to any random walk model, casting further doubt on its efficiency.

To further examine the profitability of investment strategies influenced by the Gambler's Fallacy and the Law of Small Numbers, we apply various asset pricing models. Specifically, we use the Fama & French three-factor and five-factor models to evaluate the excess return of S&P 500 constituents, which we sort into portfolios by their most recent SURP. This approach aligns with the psychological definition of investors biased by the Gambler's Fallacy and enables us to estimate the potential returns for these investors in the underlying market.

Our results show that the strategy of buying the highest SURP stocks in streaks and selling the lowest SURP stocks in streaks generates an excess return of 0.561%. A slightly lower return is obtained by focusing on buying stocks in a positive SURP streak and selling those in a negative SURP streak. On the other hand, we observe no significant return from aplying these strategies to stocks in SURP reversal.

By their psychological definition, Gamblers believe that continuing trends of i.i.d. events should revers. Therefore, their startegy is rather based on reversals than streaks. Using this logic, we can conclude, that the Gamblers' startegy is not profitable on the S&P 500 market. In contrast, we observe that the exact opposite strategy, i.e., focusing on streaks rather than reversals, can generate a significant excess return.

To refine the implementation of the Gamblers' strategy in the asset pricing model, we introduce a novel asset pricing approach. Instead of sorting stocks according to SURP strikes and reversals, we examine how Gamblers percieve the most recent SURP pattern. This approach involves sorting stocks by the likelihood of having a positive or negative SURP in the next time period, as perceived by Gamblers based on their psychological biases. This approach provides a more accurate representation of how Gamblers trade in the market and helps to evaluate the performance of their strategy in a more realistic setting.

As a result, we confirm that the Gamblers' strategy is significantly loss making. We find that a strategy of buying stocks perceived as likely to have a positive SURP and selling those perceived as likely to have a negative SURP generates a significant negative excess return of -0.357%. From the symetry of results, we can conclude that the oposite strategy is consistently profitable.

This result can be explained by the behavioral biases of investors operating in the market, which can influence stock prices. Due to their belief in the Gambler's Fallacy and the Law of Small Numbers, Gamblers tend to underreact to continuing trends and overreact to reversals. This behavior is reflected in stock prices and causes price changes. According to Loh & Warachka (2012), this result suggests that a significant proportion of Gamblers operates in the S&P 500 market, leading to opportunities for profitable trading strategies that exploit their biases.

This thesis explores the impact of the Gambler's Fallacy and the Law of Small Numbers on investors' decision-making and their profitability when investing in S&P 500 constituents. To accomplish this, we employ three distinct analytical approaches that utilize the same simplified rules of investment and focus on the same stock index. Additionally, we maintain consistent definitions of investors' decision-making and their perception of probabilities throughout all analyses. Despite these similarities, the findings of each approach differ significantly

The disparate findings among the three approaches can be largely attributed to the use of different type of data sets in each analysis. In the simulation approach as well as in the frequency analysis, investors take into consideration only the historical stock prices. They decide whether to buy or sell the stock by predicting the future evolution based on the historical observations.

In contrast, the asset pricing models employ a different methodology altogether. These models employ three and five key factors to assess general market dynamics as an input. Additionaly, we use EPS values together with concensus EPS forecasts to sort the stocks into portfolios. Therefore, this approach incorporates not only historical market data, but also publicly available expert forecasts that brings a subjective human opinion into the analysis and that can influence the result.

Future Work

As this thesis focuses on the S&P 500 index, our results are representative of larger, stable US firms. However, to further analyze this phenomenon, we could use a larger dataset containing more stocks such as NYSE. This would enable us to determine if these findings are applicable to smaller, less stable companies as well.

It would be also interesting to compare indexes representative of markets on different continents. This would allow us to evaluate if there exists a market more suitable for Gamblers than the US market. Additionally, it would be possible to estimate the proportion of Gamblers on different markets and make assumptions about theoretical investor behavior in different locations. Furthermore, examining similar experiments testing the Gambler's Fallacy from different continents could provide evidence to define more types of investors influenced by this bias according to their continent.

Bibliography

ANDĚL, J. (2002): *Základy matematické statistiky*. Univerzita Karlova v Praze, Matematicko-fyzikální fakulta.

AYTON, P. & I. FISCHER (2004): "The Hot Hand Fallacy and the Gambler's Fallacy: Two faces of subjective randomness?" *Memory & Cognition* **32**: pp. 1369–1378.

BACHELIER, L. (1900): "Théorie de la spéculation." In "Annales scientifiques de l'École normale supérieure," volume 17, pp. 21–86.

BARBERIS, N., A. SHLEIFER, & R. VISHNY (1998): "A model of investor sentiment." *Journal of Financial Economics* **49(3)**: pp. 307–343.

BARRON, G. & S. LEIDER (2010): "The role of experience in the Gambler's Fallacy." *Journal of Behavioral Decision Making* **23**: pp. 117 – 129.

BOYNTON, D. M. (2003): "Superstitious responding and frequency matching in the positive bias and Gambler's Fallacy effects." *Organizational Behavior and Human Decision Processes* **91(2)**: pp. 119–127.

CARHART, M. M. (1997): "On persistence in mutual fund performance." *The Journal of Finance* **52(1)**: pp. 57–82.

EIKON, T. R. (2022): "Refinitiv Eikon, database of financial data." `https://eikon.thomsonreuters.com/index.html`. Accessed on Jul. 25, 2022.

FAMA, E. F. (1970): "Efficient Capital Markets: A Review of Theory and Empirical Work." *The Journal of Finance* **25(2)**: pp. 383–417.

FAMA, E. F. (1995): "Random walks in stock market prices." *Financial Analysts Journal* **51(1)**: pp. 75–80.

FAMA, E. F. (1998): "Market efficiency, long-term returns, and behavioral finance." *Journal of Financial Economics* **49(3)**: pp. 283–306.

FAMA, E. F. & K. R. FRENCH (1993): "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics* **33(1)**: pp. 3–56.

FAMA, E. F. & K. R. FRENCH (2015): "A five-factor asset pricing model." *Journal of Financial Economics* **116(1)**: pp. 1–22.

FRENCH, K. R. & E. FAMA (2022): "U.S. Research Returns Data." `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html`. Accessed: 2022-07-25.

HORNE, J. C. V. & G. G. PARKER (1967): "The random-walk theory: An empirical test." *Financial Analysts Journal* **23(6)**: pp. 87–92.

HUBER, J., M. KIRCHLER, & T. STÖCKL (2010): "The Hot Hand Belief and the Gambler's Fallacy in Investment Decisions under Risk." *Theory and Decision* **68**: pp. 445–462.

KAHNEMAN, D. (2011): *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

KWAK SANG GYU, K. J. H. (2017): "Central limit theorem: the cornerstone of modern statistics." *kja* **70(2)**: pp. 144–156.

LOH, R. K. & M. WARACHKA (2012): "Streaks in Earnings Surprises and the Cross-Section of Stock Returns." *Management Science* **58(7)**: pp. 1305–1321.

MALKIEL, B. G. (1973): *A Random Walk Down Wall Street.* Norton, New York.

MALKIEL, B. G. (1989): *Efficient Market Hypothesis*, pp. 127–134. London: Palgrave Macmillan UK.

MALKIEL, B. G. (2003): "The Efficient Market Hypothesis and its critics." *Journal of economic perspectives* **17(1)**: pp. 59–82.

MASON, D. M. (1982): " of Large Numbers for Sums of Extreme Values." *The Annals of Probability* **10(3)**: pp. 754–764.

RABIN, M. (2002): "Inference by Believers in the Law of Small Numbers." *The Quarterly Journal of Economics* **117(3)**: pp. 775–816.

TVERSKY, A. & D. KAHNEMAN (1971): "Belief in the Law of Small Numbers." *Psychological Bulletin* pp. 105–110.

Ţiţan, A. G. (2015): "The Efficient Market Hypothesis: Review of Specialized Literature and Empirical Research." *Procedia Economics and Finance* **32**: pp. 442–449.