

The Efficient Market Hypothesis  
Through the Eyes of an Artificial Technical Analyst

als Inaugural-Dissertation  
zur Erlangung des akademischen Grades eines Doktors  
der Wirtschafts- und Sozialwissenschaften  
der Wirtschafts- und Sozialwissenschaftlichen Fakultät  
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Dipl. Ing. Timur Yusupov, M. Sc.  
aus Potsdam

Kiel, 2006

# ABSTRACT

The Efficient Market Hypothesis Through the Eyes of an Artificial Technical Analyst

Timur Yusupov

The academic approach to model and analyze financial markets has been reluctant to accept technical analysis. Recent publications show that practitioners do not share the skepticism of academicians. In practice traders and analysts heavily use technical analysis to make investment decisions when dealing with the high frequency finance. To resolve this inconsistency the aim of this dissertations is to translate technical analysis into a rigorous formal framework and to investigate its potential failure or success.

To avoid subjectivism the Artificial Technical Analyst is designed. This system is based on methods and techniques used in technical analysis. Given that the main criticism comes from reigning market efficiency and the corresponding impossibility of technical analysis the empirical part presents the evidence of past market inefficiencies observed on the Tokyo Stock Exchange.

With a diversity of transaction costs some investors with small transaction costs observe and exploit market inefficiencies, which are inaccessible for other investors with more pronounced transaction costs. To capture the diversity the analysis involves the schedule of transaction costs, which ranges from zero to the highest transaction cost. In this way one can identify under which transaction costs a market is perceived as inefficient, and what is the breakeven transaction costs in terms of the aggregate return. The aggregate return adjusted for transaction costs is taken also as the main criterion of market inefficiency.

The study uses several active and passive benchmark mechanisms along with the Artificial Technical Analyst. The passive strategy is buy-and-hold. Active strategies can be ordered according to the underlying intelligence of the mechanism. The highest intelligence is attributed to the perfect forecast strategy. The Artificial Technical Analyst is the second intelligent strategy. A less intelligent strategy is based on random walk, and the random strategy is the least intelligent among active strategies. The magnitude of their aggregate returns supports the ordering.

The overall conclusion is that technical analysis can provide a substantial aggregate return even after adjustment for transaction costs. The higher is the level of intelligence in the application of technical analysis, the higher is the aggregate return. The market can be perceived as inefficient if technical analyst's transaction costs are below the breakeven level, derived from technical analysis.

## Acknowledgements

Before starting with the dissertation I would like to use this excellent opportunity to thank those people who made the time I spent working on the dissertation such an instructive and enjoyable experience and contributed to completing this dissertation.

First of all, I want to thank my advisor Prof. Dr. Thomas Lux. His guidance, advice, detailed comments and numerous consultations have greatly improved the quality of this dissertation. His enthusiasm for research and willingness to discuss all problems of my dissertation has not only been contagious to me, but also to his other Ph.D. students.

I also greatly appreciate the help of my second advisor, Prof. Dr. Roman Liesenfeld. His thoughtful and constructive comments on my research were very important for the quality of this dissertation and his consultations always inspired me and moved my work one step further.

I want to thank the members and participants of European Union project on Complex markets. Their comments substantially helped me to remove weak points from this work.

I am grateful to my colleagues, Mishael, Simone and Markus. They provided great moral and informational support. Also they were a source of motivation which helped me to bear better the hours spent in the office.

I also want to thank my parents. Their faith and support encouraged me and always gave me the confidence so important for succeeding in any aspect of my life.

And finally I want to thank my beloved spouse Elena for consultations, help with editing as well as for her shared passion for research, fruitful discussions, great sense of humor and for creating a stable environment that enabled me to successfully complete this work.

## Contents

ABSTRACT	ii
Acknowledgements	iii
List of Tables	vi
List of Figures	vii
INTRODUCTION	1
Chapter 1. Efficient Market Hypothesis vs. Technical Analysis	5
1.1. Efficient Market Hypothesis	6
1.2. Technical Analysis	19
1.3. Discussion	37
Chapter 2. Building an Artificial Technical Analyst	41
2.1. Data Preprocessing	42
2.2. Piecewise Linear Approximation	53
2.3. Learning Classifier Systems	65
2.4. Discussion	97
Chapter 3. Application of the ATA on The Tokyo Stock Exchange Data	101
3.1. Decision-Making Function and Its Informational Base	102
3.2. Performance of ATA Trading Strategy	104
3.3. Performance of Benchmark Trading Strategies	110
3.4. Discussion	113
CONCLUSIONS	115
Future Work	118
Appendix A. Statistical Background	119
A.1. Monte Carlo Method	119
A.2. Kolmogorov-Smirnov Test	123
A.3. BDS Test	124
Appendix B. Genetic Algorithms	126
B.1. GA Technique	126
B.2. $\mu$ GA: Micro Genetic Algorithm	130
B.3. Algorithm Selection	131

Appendix C. Revealed Transaction Costs	132
C.1. Case (i): Expected Increase in Asset Price at Time $t + 1$	133
C.2. Case (ii): Expected Decrease in Asset Price at Time $t + 1$	133
Appendix D. Tokyo Stock Exchange: Market Microstructure and Data Description	135
D.1. Market Microstructure	135
D.2. Data Description	139
Appendix E. Comparative Analysis of PLA Algorithms	142
Appendix F. Application of TiCS and Benchmark Trading Strategies on the TSE data: Empirical Results	157
Appendix G. Abbreviations	167
Appendix. References	168
Appendix. Vita	184

## List of Tables

2.1	Averages and slope estimates of BTC for aggregate return with perfect forecast	49
2.2	Optimal subsample size	52
2.3	Preliminary ranking of PLA algorithms.	62
2.4	PLA specification of the Monte Carlo method	63
2.5	TRA ranking of PLA algorithms	64
2.6	Final ranking of PLA algorithms	65
2.7	Example of environmental records	71
2.8	Exemplary population of classifiers in LCS	71
2.9	Sample realization of environment conditions	72
2.10	LCS by hand - matching and payments	74
2.11	XCS parameters and their values	78
2.12	TiCS by hand - learning 2-bits mapping	82
2.13	Illustration of pattern-to-bits encoding mechanism	83
2.14	Illustration of bits-to-pattern decoding mechanism	84
3.1	Application of the ATA trading strategy	105
3.2	Application of the benchmark trading strategies	112
B.1	Example of initial population of trial solutions in GA	127
B.2	Illustration of a roulette-wheel mechanism	130
D.1	Tick size at the TSE	136
D.2	Special quote parameters at the TSE	137
D.3	Daily price limits at the TSE	138
D.4	List and description of stocks used in the empirical study	140

## List of Figures

2.1	Previous-tick vs. linear interpolation	45
2.2	Results of Kolmogorov-Smirnov test over the set of returns time series	46
2.3	Profitability analysis with respect to the frequency of homogenization and TC	48
2.4	Looking for the optimal subsample size	51
2.5	Illustration of Sliding Window algorithm	56
2.6	Illustration of iterative procedure with the Sliding Window algorithm	58
2.7	Illustration of Top-Down algorithm	59
2.8	Illustration of Bottom-Up algorithm	60
2.9	Illustration of GA-based algorithm	61
2.10	Illustration of signal flow in ANN	67
2.11	Illustration of crossover operator in GP	68
2.12	Illustration of mutation operator in GP	69
2.13	Holland's classifier system and its environment	73
2.14	Schematic illustration of XCS	76
2.15	TiCS and its environment	80
2.16	XCS vs. TiCS in a 6-multiplexer problem	87
2.17	XCS vs. TiCS in a 11-multiplexer problem	88
2.18	XCS vs. TiCS in a 20-multiplexer problem	88
2.19	XCS vs. TiCS in a 37-multiplexer problem	89
2.20	XCS vs. TiCS in a 70-multiplexer problem	89
2.21	TiCS in learning k-bits mapping	91
2.22	Periodic functions	93
2.23	Sample time series generated by the random walk process	94
2.24	Performance of the TiCS in learning simple periodic function	95
2.25	Active population size of the TiCS in learning simple periodic function	96
2.26	Performance of the TiCS in learning complex periodic function	97
2.27	Active population size of TiCS in learning complex periodic function	98
2.28	Performance of the TiCS in learning a random walk process	98

2.29	Performance of ARMA(1, 1) in learning complex periodic function	99
3.1	Application of ATA trading strategies on the pool of high liquidity stocks	106
3.2	Application of ATA trading strategies on the pool of medium liquidity stocks	107
3.3	Application of ATA trading strategies on the pool of low liquidity stocks	108
3.4	Time span distribution analysis of returns and rolling BDS test	109
A.1	Examples of histograms for common distributions	120
A.2	Examples of empirical CDF for different distributions	122
A.3	Example of the overlay chart of several empirical CDF	123
B.1	Illustration of mutation operator	128
B.2	Illustration of reproduction operator	128
B.3	Illustration of single-point crossover operator	128
B.4	Illustration of two-point crossover operator	129
D.1	Equally weighted indexes for liquidity pools of TSE stocks	141
E.1	GA-based approximation of 10-points time series	142
E.2	GA-based approximation of 50-points time series	143
E.3	GA-based approximation of 250-points time series	143
E.4	Sliding Window approximation of 10-points time series	144
E.5	Sliding Window approximation of 50-points time series	144
E.6	Sliding Window approximation of 250-points time series	145
E.7	Top-Down approximation of 10-points time series	145
E.8	Top-Down approximation of 50-points time series	146
E.9	Top-Down approximation of 250-points time series	146
E.10	Bottom-Up approximation of 10-points time series	147
E.11	Bottom-Up approximation of 50-points time series	147
E.12	Bottom-Up approximation of 250-points time series	148
E.13	PLA with 3 linear segments	148
E.14	PLA with 4 linear segments	149
E.15	PLA with 5 linear segments	150
E.16	PLA with 6 linear segments	150
E.17	PLA of 10-points time series with 3 linear segments	151
E.18	PLA approximation of 50-points time series with 3 linear segments	151



E.19	PLA approximation of 250-points time series with 3 linear segments	152
E.20	PLA approximation of 10-points time series with 4 linear segments	152
E.21	PLA approximation of 50-points time series with 4 linear segments	153
E.22	PLA approximation of 250-points time series with 4 linear segments	153
E.23	PLA approximation of 10-points time series with 5 linear segments	154
E.24	PLA approximation of 50-points time series with 5 linear segments	154
E.25	PLA approximation of 250-points time series with 5 linear segments	155
E.26	PLA approximation of 10-points time series with 6 linear segments	155
E.27	PLA approximation of 50-points time series with 6 linear segments	156
E.28	PLA approximation of 250-points time series with 6 linear segments	156
F.1	Application of perfect forecast trading strategies on the pool of high liquidity stocks	158
F.2	Application of perfect forecast trading strategies on the pool of medium liquidity stocks	159
F.3	Application of perfect forecast trading strategies on the pool of low liquidity stocks	160
F.4	Application of random walk trading strategies on the pool of high liquidity stocks	161
F.5	Application of random walk trading strategies on the pool of medium liquidity stocks	162
F.6	Application of random walk trading strategies on the pool of low liquidity stocks	163
F.7	Application of random trading strategies on the pool of high liquidity stocks	164
F.8	Application of random trading strategies on the pool of medium liquidity stocks	165
F.9	Application of random trading strategies on the pool of low liquidity stocks	166

## INTRODUCTION

We are living in a world where the financial market health determines the health of economies and, as a consequence, well-being of people. Therefore it is crucial to understand and to control the processes in financial markets. The efficient market hypothesis (EMH) has been long a dominant paradigm in describing behavior of prices in financial markets. Its weak form postulates that in a competitive market it should not be profitable to use investment decisions based on information obtained from past prices or returns of publicly traded securities. Numerous empirical studies and practice among financial professionals, on a contrary, show that technical analysis, which directly contradicts the weak form of the EMH, exploits predictability of past prices for profit. To avoid the joint hypothesis problem with verifying the EMH, an artificial technical analyst is created to conduct the test. This approach has two advantages. First, it is free of equilibrium model limitations, and second, technical analysis can be tested in a robust way, which should validate its existence.

The EMH is the cornerstone of the financial theory. The paradigm was coined in the 1960-70s by Harry Roberts [208] and formalized by Eugene Fama [71]. They identified three forms of market efficiency distinguished by which information prices of securities should correctly incorporate. The weak form of market efficiency postulated that past prices or returns should have no information, which can be used to predict next period values. This form was linked to the random walk hypothesis, which constituted the majority of tests performed that time. The definition of the weak was supported by empirical studies conducted before and shortly after the 1970s. Its association to the random walk allowed development of analytical tools in financial theory. The most famous example is the application of the random walk hypothesis by Myron Black and Fischer Scholes [30] to derive the option pricing formula, which caused boom of derivative markets and further development of financial theory.

With the development of statistical techniques more and more deviations from the random walk hypothesis were observed in time series of prices. Finally, Andrew Lo and Craig MacKinlay [156] used a simple specification test to reject the random walk hypothesis for stock prices and returns. The test was based on the variance properties of random walk time series. It is robust to different heteroskedasticities and non-normality of data. In reaction to this the assumptions of the random walk hypothesis were relaxed, first, to allow only independence of increments, and, later, to require only zero correlation of increments.

In the 1980s technical analysis appeared as a new trend in the empirical literature on testing the EMH. This technique, reported to be wide-spread among professional financial practitioners, exploits predictability of prices for profit, and thus is in direct contradiction

with the weak form of market efficiency. For example, in 1989 Helen Allen and Mark Taylor [5] surveyed that at the shortest horizons, intraday to one week, approximately 90% of respondents [professional traders and analysts at foreign exchange markets] used some chartist input in forming their exchange rate expectations, with 60% judging charts to be at least as important as fundamentals. Moreover, there appeared to be a persistent 2% of presumably "pure" chartists, who only used technical analysis. Numerous empirical studies reported either profitability of technical analysis or its positive added value to investment decision making.

Along with empirical studies financial theory turned its attention to technical analysis. This paradigm was incorporated to capture in models observed empirical anomalies, like volatility clustering, high transaction volumes and erratic behavior of prices [25]. It was persuasively shown that technical analysis could bring additional information, and as such should be used by rational investors [31, 38]. Moreover, the application of technical analysis for speculation can produce stable equilibria in an economy [178]. In the extreme case, the dominance of "irrational" technique, which technical analysis was referred to be, in the market can create its own space, where application of other "rational" techniques is suboptimal [63]. At present technical analysis and market efficiency are studied from behavioral prospective, which promises to resolve their inconsistency.

Since the original formulation of the weak form of the efficient market hypothesis became outdated, it experienced several transformations. First, unpredictability of prices was replaced by inability to outperform passive benchmarks, such as the buy-and-hold strategy, and later, by adding the aspect of unprofitability. In this formulation a financial market is weakly efficient if an outcome of market interactions does not contain any information, which can be persistently and profitably exploited by predicting the next period prices or returns. The notion of profitability is taken in a strict sense, that is all transaction costs should be accounted for.

With the first formulation of the EMH, it became obvious that its tests might be sensitive to the joint hypothesis problem. It manifests in ambiguity of test results, which could be either due to a wrong equilibrium model or due to market inefficiency. One way to avoid this problem is to construct the test which does not assume any underlying model.

Technical analysis is a counterpart of the weak form of the EMH, and as such can be directly used for its testing. The design of the test is as follows. One should process past stock prices with technical analysis to obtain next period price predictions. Predicted values should be used in hypothetical investment decisions. Associated returns, adjusted for transaction costs, should be aggregated and can be used as a measure of financial market inefficiency if they are in excess to a passive benchmark.

Normally the application of technical analysis suffers from subjectivism. Technical analysis is taken rather as an art than a precise science. The application of principles of technical analysis in an autonomous or artificial decision-making system should eliminate the subjective factor. Additionally, the system should be relatively simple to insure its

robustness, and transparent to provide its understandability. Spyros Skouras [227] proposed artificial technical analysts as a quantifiable measure of market efficiency. Reflecting his idea, an artificial technical analyst (ATA) will be used for the test.

My idea of creating and employing the ATA originates from the seminal paper by Arthur et al. [10]. The authors have created the Artificial Stock Market (ASM), known as the Santa Fe ASM. The market is populated with artificial agents involved in stock trading. The agents use technical analysis to screen the market and an implementation of artificial intelligence, Learning Classifier Systems (LCS), for the optimal application of technical trading rules.

The success of the Santa Fe ASM inspired me to create the ATA. Unfortunately, the Santa Fe ASM had a few shortcomings: computational and algorithmic limitations of that time, short and fixed list of technical trading rules. Additionally, the mechanism of forming the traders' expectations was based on undeveloped concept of LCS. From the methodological point of view, the results lacked reality due to the distillation effect of laboratory experiments. My implementation takes into account the shortcomings of the Santa Fe ASM. It incorporates technical analysis in an adaptive way, where the core of the ATA is driven by a new implementation of LCS.

A new implementation of the ATA should incorporate three main components: data-preprocessing, pattern-recognition and decision-making under transaction costs. The first component would insure that raw input data are homogenized in a way that maximizes the informational content. Homogenization itself would allow to reduce the complexity of pattern-recognition, by focusing attention only to levels in time series. The pattern-recognition should be driven by an implementation of artificial intelligence, which would allow transparent structure of results. LCS can provide a possible candidate, but their performance needs to be improved. The decision-making under transaction costs should insure optimality of investment decisions.

Once the ATA is designed and implemented its abilities in technical analysis should be tested on some test-bed problems. These problems can include periodic functions augmented with noise and random walk series. Periodic functions can help in tuning parameters of the ATA. Random walk time series can be used to verify mistakes in programming. Correct ATA programming should result in high performance to predict periodic functions, and in low performance for the random walk problems.

In the final stage, the ATA should be applied to historical time series of security prices. Since a part of transaction costs is an impact on a price, in a study of hypothetical investments with historical data the aspect of profitability should be inverted. Instead of measuring the aggregate return adjusted for specified transaction costs, the value of revealed transaction costs, which insure positive aggregate return, should be used. This would allow to derive breakeven transaction costs, which could be later used for benchmarking different investment strategies.

This work differs from others in the extent to which the design of the market efficiency test is free of equilibrium model limitations. By applying technical analysis one can see whether price time series have predictable patterns. In case persistent patterns are detected, their profitability will be studied under the schedule of transaction costs. In this way an absence of patterns will be a clear indication of the weak form of market efficiency, which was advocated in the theoretical literature of the 1970s. Otherwise, the profitability of detected patterns can be assessed through revealed transaction costs. When the level of breakeven transaction costs is below market transaction costs, a market can be perceived as efficient, according to the pertinent literature. At the same time a comparison to market transaction costs leaves a speculative space for those market participants, whose correctly accounted transaction costs are below a breakeven value.

The novelty of this work comes, first, from using an adaptive method of technical analysis, which allows to extract uncharted patterns; second, from developing pattern-recognition engine, which makes detected patterns to be accessible afterwards; and, third, from testing market efficiency under a wide schedule of transaction costs, which allows to identify boundaries of market efficiency perception.

The thesis is organized as follows. Chapter 1 presents an overview of the efficient market hypothesis and technical analysis literature. A genesis of each paradigm is presented along its empirical tests. In Chapter 2 a design of the artificial technical analyst is worked out, which is used in Chapter 3 for testing the EMH and validating TA. Conclusion finalizes the thesis.

## CHAPTER 1

### **Efficient Market Hypothesis vs. Technical Analysis**

This chapter presents the overview of the efficient market hypothesis and technical analysis literature. The importance of their study comes from the fact that even though they both deal with the same subject, prices of securities, their proponents - academicians for the EMH and practitioners for TA - have opposed to each other's positions on price predictability. The aim of this literature overview is, first, to reveal the genesis of the EMH and TA and, second, to show what is the up-to-date position of practitioners and academicians on the subject. The latter should help to identify how to settle the existing dispute.

The EMH is advocated by academicians. It originates from the work of Louis Bachelier [13], where seeming price unpredictability of financial securities was assumed for the first time. In the later empirical studies this observation was confirmed and associated with the random walk hypothesis, which defines behavior of price increments. Harry Roberts [208] and Eugene Fama [71], on bases of the RWH, gave a formal definition of the EMH and its forms. The EMH states that in an active market that includes many well-informed and intelligent investors, securities will be appropriately priced and will reflect all available information. The forms of market efficiency are distinguished by progressively inclusive types of information, which a market should correctly price: past prices, all public and private information.

The weak form is of particular interest since it requires that in weakly efficient markets past prices should provide no information to gain profit. Initial empirical studies backed academicians in their view that financial markets are efficient. However, later studies in support of technical analysis, which directly contradicts the weak form of market efficiency, and its wide-spread use among practitioners forced the academic community to rethink the EMH. Recent empirical and theoretical studies show that the original formulation of market efficiency needs to be augmented with impact of markets microstructure, transaction costs and behavioral aspects of financial markets.

Technical analysis is one of the oldest methods of forecasting development of security prices. It dates back to the 1600s when it was used to forecast rice prices in Japan. In the western world it is known since the 1900s, when Charles Dow laid its fundamentals. For more than one hundred years the techniques of technical analysis were refined. At the end of the 1980s they were merged with the Japanese technique of technical analysis - candlesticks. Since the moment of its introduction practitioners started to use it, first, to forecast economic development, and later, to forecast development of securities prices. The importance of technical analysis comes from its ability to account for other than economic

factors in price formation. An overwhelming flow of recent empirical studies shows its abilities and limitations.

The overview of the TA and EMH literature should help to design the mechanism, which will allow to settle the existing dispute on price predictability between academicians and practitioners.

The chapter is organized as follows. First, Section 1.1 presents the concept of the EMH and a literature overview of its empirical tests. It is followed by Section 1.2, which introduces technical analysis along with an overview of the results of its empirical applications. The chapter is concluded by a discussion of presented findings.

### 1.1. Efficient Market Hypothesis

The efficient market hypothesis is an issue that is a subject of intense debate among academicians and financial practitioners. In short, the EMH states that at any given time, security prices fully reflect all available information.

The notion of "efficient market hypothesis" was coined by Harry Roberts and popularized by Eugene Fama. In his Ph.D. dissertation Fama convincingly made the argument that in an active market that includes many well-informed and intelligent investors, securities will be appropriately priced and will reflect all available information. If a market is efficient, no information or analysis can be expected to result in outperformance of an appropriate benchmark.

It has been customary since Harry Roberts [208] to distinguish three levels of market efficiency by considering three different types of information sets:

- *The weak form* of the EMH asserts that prices fully reflect the information contained in the historical sequence of prices. Thus, investors cannot devise an investment strategy to yield abnormal profits on the basis of an analysis of past price patterns (a technique known as technical analysis).<sup>1</sup>
- *The semi-strong form* of the EMH asserts that current stock prices reflect not only historical price information but also all publicly available information relevant for company securities. If markets are efficient in this sense, then an analysis of balance sheets, income statements, announcements of dividend changes or stock splits or any other public information about a company will not yield abnormal economic profits.
- *The strong form* of the EMH asserts that all information that is known to any market participant about a company is fully reflected in market prices. Hence, not even those with privileged information can make use of it to secure superior investment results. There is perfect revelation of all private information in market prices [164].

---

<sup>1</sup>The definition of Mark Rubinstein in [215] and William Beaver in [23] requires that publishing the information does not change equilibrium prices. Rubinstein's notion allows one to ask only if the market is efficient with respect to all information. William Beaver provides a definition with respect to information set that might the information contained in historical prices.

The clear definitions of the EMH forms allow to construct tests focused on their specific aspects. Acceptance or rejection of some form of market efficiency allows to identify the level of financial market development. Obtained results can be used to generate recommendations and policies to improve efficiency of financial markets.

Theoretical arguments for the efficient market hypothesis are based on three assumptions. These are the rationality of investors, the irrationality of investors and the randomness of trades and, finally, the presence of rational arbitrageurs. Each assumption is progressively weaker but their combination allows to justify market efficiency in the most of market situations.

*Rationality of investors.* Investors are assumed to be rational and hence to value securities rationally. Investors value each security for its fundamental value. When investors learn something new about fundamental values of securities, they quickly respond to the new information by bidding up prices (buying securities) when the news is good and bidding them down (selling) when the news is bad. As a result, security prices incorporate all the available information almost immediately and adjust to new levels corresponding to the new net present values of expected cash flows.

*Irrationality of investors and randomness of trades.* Some investors are not rational, and they are trading randomly. When there is a large number of such investors and when their trading strategies are uncorrelated, their trades are likely to cancel each other.<sup>2</sup>

*Presence of rational arbitrageurs.* Although some investors might be irrational in a similar way there are some rational arbitrageurs in the market. Arbitrage<sup>3</sup> is one of the most intuitively appealing and plausible arguments in all economics. The main condition for arbitrage is the existence of over- or undervalued security with a close substitute.<sup>4</sup> This could be the case when the trade involves irrational investors. Noting the overpricing the arbitrageur would sell (or even short sell) the overpriced security and simultaneously purchase another, “essentially similar” but truly valued, security to hedge any risk. The effect of this arbitrage is to bring the price of the overpriced security down to its fundamental value.

The arbitrage argument allows to cover the most complex case - the existence of irrational investors. To the extent that the securities that irrational investors are buying are overpriced and the securities they are getting rid of are undervalued, such investors earn lower returns than either passive investors or arbitrageurs. Relative to their peers, irrational investors lose money and in a long run leave the market. Thus, not only investor rationality, but also the market forces bring about the efficiency of financial markets [226].

Since this work is presenting the efficient market hypothesis from the point of view of technical analysts this section focuses on the weak form of efficiency and related tests.

---

<sup>2</sup>In such market, there will be substantial trading volume as the irrational investors exchange securities with each other, but the prices are nonetheless close to their fundamental values.

<sup>3</sup>That is the simultaneous purchase and sale of the same, or essentially similar, security in two different markets at advantageous different prices.

<sup>4</sup>In some cases the access to alternative markets can be taken as a form of substitute.



The section is organized as follows. Subsection 1.1.1 presents historical development of the EMH. In Subsection 1.1.2 the problem of joint hypothesis testing is discussed. Subsection 1.1.3 outlines the link between the weak form of the EMH and the random walk hypothesis. It is followed by Subsection 1.1.4, where the results of various tests on the EMH are presented. Subsection 1.1.5 discusses the role of transaction costs and market microstructure in market efficiency. The section is concluded by Subsection 1.1.6, which briefly reviews recent updates in the EMH research brought by behavioral finance literature.

### 1.1.1. History of the EMH

The concept of financial market efficiency, as it is known today, is a product of numerous empirical and theoretical studies. It started with the observation of random development in asset prices, followed by verification and analytical work. This allowed to understand the nature of the observation and to formulate a theory which is known today as the efficient market hypothesis. It postulates that in informationally efficient markets all prices have already incorporated all information and immediately incorporate any coming information. So, if markets are efficient then the market price provides the best estimate of value. Thus, starting with a simple observation the idea was shaped into the efficient market hypothesis that became a central element of financial theory.

The origin of the EMH dates back to 1900, when Louis Bachelier [13] first introduced the idea that the stock market fluctuations follow a stochastic process for which the future does not depend on the past except through the present and the best prediction of the subsequent price is the value of the current price.<sup>5</sup> That is if all the information necessary for a successful trade is already contained in the quoted prices, the only cause of new variations could be elements of information that are not predictable and thus the next period prices. Bachelier stresses importance of the information concept, on which the efficiency concept is developed. However, as Paul Cootner [49] comments - Bachelier's work received little attention from academicians and was forgotten for almost fifty five years. His work was rediscovered in 1955 by Leonard Jimmie Savage.

In the beginning of the 1920s many economists started to question predictability of stock prices, which one would expect with respect to the laws of supply and demand [236]. In the 1930s the first tests started in two directions. In the first one the independence of successive price changes was tested. In the second one researchers sought to characterize the form of return dispersion and their empirical distribution. The random walk hypothesis was validated by Alfred Cowles [50] for American stock prices. He argued that there is no evidence of ability to outguess the market. The same results were obtained for other price series including British stock and commodity prices by Holbrook Working [255], Alfred Cowles and Herbert Jones [51] and later by Maurice G. Kendall [122] and Harry Roberts [207] using the serial correlation tests, momentum or "runs" tests, filter tests and persistence tests.

---

<sup>5</sup>This notion was later branded by Karl Pearson as the random walk hypothesis [198, 199].

At the end of the 1950s M. F. M. Osborne [193] complemented Bachelier’s model, using the developments in statistical mechanics. He introduced the first hypothesis of Brownian motion in the logarithms of prices. This was an important contribution to the application of models to stock market movements.

In the beginning of the 1960s the interest in the stock market research started to increase. Holbrook Working [256], Sidney Alexander [3], Hendrik Houthakker [109], Paul Cootner [48], Arnold Moore [177], Clive Granger and Oskar Morgenstern [96] and others again found support for the random walk hypothesis in the stock market data series and confirmed that successive price changes resemble in their first approximation Brownian motion.

The Bachelier-Osborne model became a paradigm for the study of market behavior in academic research [247]. The assumption of normal distribution of the stock market series and the hypothesis of random walk enabled practical application of portfolio theory created by Harry Markowitz [168] and simplified by William Sharpe [222], who made it more practical given the limited computational power available in the 1960s.

However, these findings turned out to be in contrast with the practices of financial experts and asset managers who reported better performance than the market itself [3]. These opponents came from two distinct camps. The first camp presented financial analysts (called “fundamentalists”) looking for additional information in financial statements, macroeconomic and econometric forecasts. The second included technical analysts seeking the structure of movements and fluctuations by examining the past time series of underlying assets.

At the beginning of the 1960s the confrontation between supporters and opponents of the random walk hypothesis intensified. In response the academic research started to pay attention to the claims of professional investors. At the end of the 1960s M. F. M. Osborne [194] found some non-random properties of tested price series primarily in short interval data and individual stock prices. John Evans [69] also showed an existence of trading rules which, when applied to a portfolio, leads to a greater expected return than buy-and-hold strategy, even after including transaction costs. However, he still considered the found non-randomness too small to be meaningful for investors.

In 1965 Paul Samuelson [218] provided the first formal economic argument for “efficient markets”. His contribution is neatly summarized by the title of his article: “Proof that Properly Anticipated Prices Fluctuate Randomly”. He focused on the concept of a martingale, rather than a random walk. Samuelson showed that in an informationally efficient market price changes must be unforecastable [200].

At the same time Eugene Fama [70] came up with a compromise stating that to accept an independence assumption consistent with the random walk theory the dependence of successive price changes must not be above some “minimum acceptable” level, which he defines as the dependence in the series of price changes is not sufficient to allow the past history of the series to be used to predict the future in a way which makes expected profits greater than they would be under a buy and hold model [70].

After validating the random walk hypothesis, it was necessary to find economic reasons for these results. This was achieved with the introduction of the informational market efficiency concept, which established a connection between the allocation role of capital markets, and the informational role of prices [247]. In 1967 Harry Roberts [208] coined the term “efficient markets hypothesis” and made the distinction between weak and strong form tests, which got the classic taxonomy by Eugene Fama [71].

Eugene Fama popularized the efficient market hypothesis and refined its definition in his publication in 1976 [72]. The underlying rationale was that the theory of market efficiency cannot be separated from its probability base. Only with the probability underpinning, the theory acquires capacity for verification and testing. In other words, to know whether the quoted price is a true representation of the theoretical equilibrium price, and thus whether the market is informationally efficient, it is necessary to attach to this postulate a complementary piece of information describing the way how this representativeness is defined. In particular, the nature of the information, measurement and the appropriate must be defined [247]. In response to this issue, Eugene Fama [71] formalized three forms of informational efficiency, which are based upon what information is reflected in prices.

The introduction of the three forms of the EMH constituted the most important moment of the concept development. From that moment tests went in three directions trying to empirically verify the validity of each form.

### 1.1.2. Joint Hypothesis Problem

In the process of refining the EMH a fundamental problem of testing the efficient market hypothesis became obvious. It is known as the joint hypothesis test of the EMH.

The problem comes from underlying probabilistic assumptions. One cannot speak of efficiency by itself, except through a model that defines the generation of prices with a representative probability system. Eugene Fama [71, 73] stressed that the market efficiency per se is not testable. One can test whether information is properly reflected in prices in the context of a pricing model. It means that when one finds anomalous evidence on the behavior of prices or returns it is ambiguous if this is caused by market inefficiency or/and a bad model of market equilibrium. This leads to the conclusion that the efficiency tests are always joint tests on the market efficiency and the model and its probabilistic assumptions.

In [247] Christian Walter considers this overlapping as a common cause of misinterpretations and errors, leading to rejection of efficiency when there is only a misspecification of the stochastic process.

### 1.1.3. Weak Form Efficiency and Random Walk Hypothesis

A close analysis of the weak form of the EMH shows its similarity to the random walk hypothesis. Indeed, both concepts postulate unpredictability of the next observation and that the best prediction is its current value. Moreover, the origin of the EMH comes from the random walk hypothesis. With a progress of statistical tools it became obvious

that financial data deviate from the strong assumptions of the random walk hypothesis. To keep the tandem of the two the random walk hypothesis obtained weaker, but more realistic, forms. These forms allow to preserve a formal link between the EMH and random walk hypothesis, which, in turn, provides a base for constructing tests of the weak form of efficiency in financial markets.

For the random walk hypothesis empirical investigations invariably find that actual returns are too fat-tailed to be lognormal, which is mainly due to the hierarchical quality of shocks which are not homogeneous [97]. Benoit Mandelbrot [165] pointed out this problem and introduced the concept of “Pareto’s randomness”, arguing that the Gaussian randomness did not stand for a good candidate in respect to probability distribution.<sup>6</sup>

Under the assumption of risk neutrality of investors, the weak form of the EMH is equivalent to the random walk hypothesis - the statement that stock returns are to some extent unpredictable based on past returns:

$$p_t = \mu + p_{t-1} + \varepsilon_t, \quad (1.1)$$

where  $p_t$  is stock price at time  $t$ ,  $\mu$  is a drift term and  $\varepsilon_t$  is a stochastic component.

The behavior of the stochastic component is central to defining the form of the random walk hypothesis. Specifically, each form requires a number of restrictive assumptions, which the stochastic component should satisfy:

- *Random walk hypothesis 1* requires identical and independent distribution of the stochastic component under homoskedastic variance:

$$\varepsilon_t \sim IID(0, \sigma^2). \quad (1.2)$$

- *Random walk hypothesis 2* requires independent distribution of the stochastic component with the provision of heteroskedastic variance:

$$\varepsilon_t \sim indep(0, \sigma_t^2). \quad (1.3)$$

- *Random walk hypothesis 3* requires only no correlation in the stochastic component:

$$cov(\varepsilon_t, \varepsilon_{t-k}) = 0 \text{ for } \forall k \neq 0. \quad (1.4)$$

Each form of the random walk hypothesis is progressively less restrictive and closer to reality. The test for weak form of the EMH is the test of whether the stochastic components satisfies one of forms of the random walk hypothesis.

#### 1.1.4. Testing the Efficient Market Hypothesis

One of the main problems with the EMH is its rejection by practitioners in spite of its wide acceptance in financial theory. To settle down this controversy the EMH should be verified empirically. This subsection presents results of EMH empirical tests.

---

<sup>6</sup>These ideas attracted attention mainly in the second half of the 1980s, especially after the market crash in 1987, when numerous bankruptcies resulting from risk management and hedging techniques based on a Gaussian concept of stock market fluctuations caused a new surge in scientific research into finance [247].

**1.1.4.1. Short-Horizon Tests.** The earliest tests were concerned with short horizon returns. These tests typically assumed that in the efficient market the expected rate of return was constant through time and the realized returns should not be serially correlated.

Before 1970, the work focused on forecasting returns from past returns. More recent studies started to use variables such as dividend yields, earnings/price ratios and interest rate term-structure. While the early work concentrated on predictability of daily, weekly and monthly returns, the recent tests also examine predictability of returns at longer horizons.

Eugene Fama [70] finds that the first-order autocorrelation of daily returns are positive. In [166] and [70] it is also recognized that returns are characterized by volatility clustering and leptokurtic unconditional distributions. Lawrence Fisher [86] suggests that autocorrelations of monthly returns of a diversified portfolio are bigger than those of individual stocks. However, the evidence often lacked statistical power and the EMH was not rejected.<sup>7</sup>

Later research used daily and weekly NYSE or AMEX data. Andrew Lo and Craig MacKinlay [156] find that weekly returns on portfolios of NYSE stocks show reliable positive autocorrelation, which is stronger mainly for portfolios of small stocks. This can be due to their smaller liquidity and consequent non-synchronous trading effect discussed already in [86]. In [47] Jennifer Conrad and Gautam Kaul mitigated this problem, examining the autocorrelations of Wednesday-to-Wednesday returns for size-grouped portfolios. They also found positive autocorrelation especially in portfolios of small stocks.

However, in [119] Shmuel Kandel and Robert Stambaugh show that stock return predictability, which seems weak when evaluated by classical statistical criteria, may nevertheless be economically important in the sense that a rational Bayesian investors would substantially alter portfolio holdings in response to the current values of predictive variables.

As noted by Francis Diebold [66] and Robert Cumby and John Huizinga [53], the presence of conditional heteroskedasticity or excess kurtosis biases the test towards rejection of the null hypothesis of uncorrelated returns. In contrast to the weak evidence for autocorrelation in returns, Tim Bollerslev and Robert Hodrick [33] stress the importance of conditional heteroskedasticity. One should say that finding a strong dependence in the even ordered moments does not necessarily imply market inefficiency, which is consistent with a martingale hypothesis for stock prices. The latter stipulates that the innovations can be serially uncorrelated or be white noise.

Generally, the researches show that daily and weekly returns are predictable from past returns. Thus, the empirical findings rejects the weak form of the EMH. At the same time the estimated autocorrelations are typically found to be very small and the variation of these returns is a small part of the overall return variance.

---

<sup>7</sup>Reported  $R^2$  was often less than 0.01 for individual stocks.

In the present time the return predictability research gradually moved in the direction of higher frequency time series. In [120] Ludwig Kanzler developed a new version of the BDS test<sup>8</sup> to test the EMH on ultra-high frequency foreign exchange market data. He found that the EMH holds only in some periods, in particular, when a release of important news takes place.

The more striking evidence on the predictability of returns from past returns comes from tests on predictability of long-horizon returns.

**1.1.4.2. Long-Horizon Tests.** The evidence on autocorrelation of short-horizon returns against the hypothesis of a constant conditional mean return is not very strong. Robert Shiller [225] and Lawrence Summers [233] challenge the validity of EMH, based on low autocorrelations of short-horizon returns. They provide evidence of stock market inefficiency by using models in which stock prices take large slowly decaying swings, even though the short-term returns have little autocorrelation.

In [60, 61] Werner DeBondt and Richard Thaler attack market efficiency in similar manner, trying to unmask irrational bubbles. They find that the NYSE stocks identified as the most extreme losers over a 3- to 5-year period tend to have strong returns relative to the market during the following years. The stocks identified as extreme winners tend to have, on the contrary, weak returns relative to the market. They attribute these results to market overreaction to extreme news.

In [115] Narasimhan Jegadeesh and Sheridan Titman observed that past winners realized consistently higher returns around their earnings announcements in the first 7 months following the portfolio formation date than do past losers. They argue that to attribute the results to underreaction is overly simplistic. Buying past winners and selling past losers, consistent with positive feedback trading, move prices further from their long-run values and thereby causes price to overreact. The interpretation is consistent with DeLong et al. [62] who explore the implications of positive feedback trading on market prices. Louis Chan [43] and Ray Ball and S. P. Kothari [15] argue that these results are due to a failure to risk-adjust returns.

James Poterba and Lawrence Summers [201] and Eugene Fama and Kenneth French [76, 77] also realized that the negative serial correlation in returns would manifest more transparently at longer horizons. Evidence in [201] and [76], using multi-period regressions and variance ratio statistics, suggests that for longer return horizons a large proportion of returns is explainable from the history of past returns alone.<sup>9</sup> James Poterba and Lawrence Summers [201] argue that asset prices are characterized by speculative fads, in which market prices experience long systematic swings away from their rational fundamental values.

However, whether the longer-horizon mean reversion really exist is controversial, since these results are not observed over time. For example, Narasimhan Jegadeesh [114], Kim

<sup>8</sup>For details on the BDS test see Appendix A.3.

<sup>9</sup>The estimated in [76] for monthly U.S. stock returns imply that for 3- to 5-year returns up to 40% of variability is predictable. However, this does not necessarily imply market inefficiency, since the variation could be due to a time-varying risk premium.

et al. [128], Mankiw et al. [167], Richardson et al. [205], all argue that the case for predictability of long-horizon stock returns is weak when one corrects for the small sample biases in test statistics. In addition, [76] make a counterargument that irrational bubbles and swings in stock prices are indistinguishable from rational time-varying expected returns.

The subsequent work showed, that the apparent predictability of long-horizon returns should be interpreted very carefully. As [33] point out, overlapping nature of the data in the multi-year return regressions gives rise to a non-standard small sample distribution of test statistics, which appear to be better approximated by the alternative asymptotic distribution derived by Richardson et al. [205].<sup>10</sup>

In [33] Tim Bollerslev and Robert Hodrick developed tests based on the iterated version of the null hypothesis using Hansen's GMM [101] and found some improvement in the small sample performance of the test statistics. However, they conclude that there is still little evidence for predictability of returns.

To sum up, the degree of predictability is generally small compared to the high variability of returns. In [74] Eugene Fama supports this argument saying that market anomalies are chance events, i.e. they split randomly between overreaction and underreaction to news (see for example [2, 11, 175, 212]).

**1.1.4.3. Volatility Tests.** In the 1970s researchers interested in the efficiency of asset markets shifted their focus from the predictability of returns to the volatility of prices. The main reason was that price fluctuations seemed to be too large to be justified by the subsequent variation in dividend payments. The EMH could not be tested directly but only as a part of a joint hypothesis. Researchers were still required to specify a particular model of expected returns. In addition, the predictions of price volatility depended on the assumed time series properties of dividend process and information sets of economic agents [33].

Stephen LeRoy and Richard Porter [152] and Robert Shiller [223, 224] introduced another important class of tests for market efficiency: the volatility or variance bounds tests. They assumed a constant expected rate of return model and reported overwhelming rejections of market efficiency since excess price volatility was supposed to imply market inefficiency.

In the first generation of volatility tests the null hypothesis was taken to be the standard present value model with a constant discount rate. The vast majority of these tests resulted in clear rejections of market efficiency, with actual asset prices being excessively volatile compared to the implied price series calculated from the discounted value of the expected or actual future fundamentals. One possible explanation was the idea that asset prices may be characterized by self-fulfilling speculative bubbles that earn the fair rate of return but cause prices to differ from their rational fundamentals.

---

<sup>10</sup>The unsatisfactory spurious results led to the studies by John Campbell [41], Robert Hodrick [104] and Geert Bekaert and Hodrick [26]. They used other forecasting variables, e.g. dividends and interest rates. The studies support the economically significant long-horizon return predictability.

However, as Charles Nelson [187] and Eugene Fama and William Schwert [78] showed in their works, the assumption of constant expected return was unjustified. In response to this problem, subsequent research, in particular by Marjorie Flavin [87], Allan Kleidon [131, 132] and Eugene Fama [73] questioned the small sample statistical properties of these analyses.<sup>11</sup>

The volatility tests thus clearly show that expected returns vary through time, but give no help on the central issue of whether the variation in returns is rational.<sup>12</sup>

The next challenge for economists in the 1990's was to develop models that are consistent with the observed variability in the distributions of returns and to understand the link between the conditional volatility of fundamentals and variation in required expected returns that arises from risk aversion of economic agents.

**1.1.4.4. Tests of Market Efficiency in Derivative Markets.** Introduction of the option pricing theory by Myron Black and Fischer Scholes [30] turned in the 1970s and 1980s the attention of financial community to derivative markets.

A. L. Tucker [245] studied the currency option markets. After accounting for transaction costs and bid-ask spreads he reported no possibility to earn riskless arbitrage profits on the currency options market. James Bodurtha and Georges Courtadon [32] achieved similar results. The currency option market was inefficient only before adjusting for transaction costs. However, the violations of the EMH disappear when transaction costs are taken into account.

Y. P. Chung [44] investigated the efficiency of the market for stock index futures and the profitability of index arbitrage. The results indicate that the size and frequency of boundary violations decreased significantly over the sample years for all levels of transaction costs, which indicates maturing of the futures market in which arbitrage trading has tended to correct mispricing.

### 1.1.5. Transaction Costs, Market Microstructure and Market Efficiency

At the end of the 1970's and beginning of the 1980's an increasing number of studies reported the violation of the weak form of market efficiency. In-depth analyses showed that the EMH does not rule out small abnormal returns before accounting for transaction costs. Given that collecting and processing information is a costly process, prices are expected to reflect information to the point where the marginal benefits of acting on information do not exceed the marginal costs [99, 117].

---

<sup>11</sup>For surveys of this literature see [92].

<sup>12</sup>The efforts of Sanford Grossman and Robert Shiller in [98] and John Campbell and Shiller in [42] to resolve this issue ran into the joint hypothesis problem of testing market efficiency jointly with the hypothesis that their consumption-based asset pricing model capture all rational variation in expected returns.



Steven Thorley [243] defines four components of transaction costs: *brokerage commissions, bid-ask spreads, taxes, suboptimal diversification and research*.<sup>13</sup> Based on the performed simulations he states that 67% of portfolios are underperforming when transaction costs are correctly taken into account. Thus, bravado reports of practitioners beating the market could be just a result of myopic accounting, but not of market inefficiency.

John Hussman [112] argues that transaction costs create a region in which the market may be inefficient while still excluding the possibility of abnormal risk-adjusted returns. If sufficiently high trading costs reduce long-term returns below those of a passive approach, an active approach may still be optimal from the standpoint of utility maximization for “myopic” investors whose utility is defined over the sequence of returns during individual holding periods, instead of terminal wealth.

Analysts could therefore still have an incentive to obtain and act on valuable information. As Elroy Dimson and Massoud Mussavian [67] suggest, time-varying expected returns could also explain these patterns. In addition, there is a growing literature that seeks to explain observed patterns in terms of sentiments of non-rational noise traders.

In response to these problems, a new strand of research focused on market microstructure developed. The seminal paper on microstructure is the article of Webster Bagehot<sup>14</sup> [14] who provided an early insight into the way information is incorporated into security prices through the activities of investors and how a market structure can have an impact on the efficiency of the stock market. In particular, the author explains why informed investors can win over uninformed ones.

Albert Kyle [139] formalized the story of Bagehot. He developed a model showing that only informed traders achieve a profit, and they do this at the expense of noise traders. Werner DeBondt and Richard Thaler [60] suggested that irrational investors putting too much weight on recent events fail to make proper Bayesian forecasts and thus lose money. Steven Thorley [243] provides the empirical evidence to support this argument. He points out some investing mistakes, which make the prices predictable, e.g. putting too much stress on recent information and events.

Sanford Grossman and Joseph Stiglitz [99] extended the Bagehot’s idea. They constructed a model which shows that prices cannot perfectly reflect the information which is available since if it did, those, who spent resources to obtain it would receive no compensation. In other words, in a world with costly information, it is impossible for markets to be informationally efficient.

Ludwig Kanzler [120] pointed out that the markets are more efficient when a release of important news is present because participants pay close attention to the news. However, at times of absence of new information, the market exhibits conditional mean dependence. Alex Tung and James Marsden [246], using laboratory experiments, also found support for positive relationship between information quality and trading profits in a market.

<sup>13</sup>One would additionally include *impact on a price* since in a low liquidity market an actual transaction can dramatically shift the price level.

<sup>14</sup>A pseudonym for Jack Treynor.

Lawrence Glosten and Paul Milgrom [93] showed that the possibility of trading on information could be sufficient to induce a positive bid-ask spread. Together with Demsetz's [65] order processing costs, this provided a framework that is now used for analyzing the bid-ask spread faced by investors.

To sum up, transaction costs, in particular bid-ask spreads, are one of the main reasons for rejecting the EMH. That is the stock market is efficient when transaction costs are considered [12]. Correct accounting for transaction costs removes perception of market inefficiency. As Richard Roll [211] stipulates it is hard to profit from even the most extreme violations of market efficiency. Moreover, the empirical evidence suggests that the degree of predictability had diminished over time [24]. On the other hand, the lack of intertemporal arbitrage, which Spyros Skouras [227] uses as a measure of market efficiency, can be rejected for investors with low transaction costs. So, the magnitude of transaction costs and the generality of preferences is crucial for measuring market efficiency.

### 1.1.6. Behavioral Finance and Market Efficiency

The empirical research has documented dynamic return phenomena that appear to be inconsistent with the weak form of the EMH. Numerous studies reported price momentum and reversals, which both suggest intertemporal predictability in stock returns based on past price information. At the same time psychology literature cites a number of persistent behavioral patterns attributed to financial investors. Under the hypothesis that violation of the weak form of market efficiency can be a result of human behavioral interference a new strand of literature emerged - behavioral finance.<sup>15</sup> Findings of behavioral finance are expected to finalize the quest for financial markets efficiency.

Drawing on behavioral decision theory by Daniel Kahneman and Amos Tversky [118], a number of behavioral models were proposed, which try to explain the deviations of price developments from the EMH. The models explain these phenomena by assuming investors make specific types of information-processing errors that have been well documented in psychology literature. Kent Daniel et al. [56] show that overconfidence and biased self-attribution can generate short horizon momentum followed by reversals at longer horizons. In Nicholas Barberis et al. [18] investors behave in a manner consistent with representativeness, which is in violation of the probabilistic laws governing the event process. Albert Kyle and F. Albert Wang [140], Terrance Odean [190] and Simon Gervais and Terrance Odean [90] consider models, in which agents are overconfident about the quality of the information they trade on.

Generally, in behavioral models, rational traders are limited in their ability to exploit the mispricings generated by irrational investors, either due to high transaction costs or

---

<sup>15</sup>It is early today to determine by which route financial theory will rethink the efficiency concept because several competitive currents of thought coexist. Eugene Fama [74] predicts a surge in behavioral models. Doyne Farmer and Andrew Lo [81] find the most promising direction to view financial market from a biological perspective and, specifically, within an evolutionary framework in which markets, instruments, institutions and investors interact and evolve dynamically according to the "law" of economic selection.

myopic behavior or nonexistence. Steve Slezak [228] proposes a model, in which rational traders are long-lived. He shows that the weak-form market efficiency will be violated under rather weak conditions: rational investors are risk-averse and the fundamental value of the asset is risky. In this situation, price predictability is not removed and irrational traders persist.

Since this group of models is vulnerable to the joint hypothesis problem, Steve Hogan et al. [105] introduce the concept of statistical arbitrage – a long horizon trading opportunity that generates a riskless profit and is designed to exploit persistent anomalies – to circumvent the joint hypothesis dilemma of the EMH tests. In spite of adjusting for transaction costs, the influence of small stocks, margin requirements and liquidity buffers, they find evidence that value and momentum strategies generate opportunities for a statistical arbitrage.

Although the EMH is a powerful idea, the concept of informational efficiency has, as Doyne Farmer and Andrew Lo [81] put it, a counterintuitive and seemingly contradictory flavor in it. On the one hand, in the efficient market, participation of many active traders ensures totally random and unpredictable prices, but at the same time, one of the central tenets of modern financial economics is the necessity of a trade-off between risk and expected return. Therefore, the prices need not be absolutely random, the fact supported by many empirical studies, even if markets operate efficiently. Moreover, they argue that to make the EMH operational, one must specify additional structure, e.g. investors' preferences, information structure etc. As a result, a test of the EMH becomes a test of several auxiliary hypotheses and its rejection will tell us little about which aspect of the joint hypothesis is inconsistent with the data [81].

The desire to build financial theories based on more realistic assumptions has led to several new strands of literature, which can be divided into

- (1) *psychological approaches* to risk-taking behavior departing from Daniel Kahneman and Amos Tversky [118], which focus on the manner in which human psychology influences the economic decision-making process as an explanation of apparent departures from rationality;
- (2) *evolutionary game theory*, studying the evolution and steady-state equilibria of populations of competing strategies and
- (3) *agent-based modeling* of financial markets capturing complex learning behavior and dynamics in financial markets by using more realistic markets, strategies and information structures. The strategy is analogous to biological specie and the total capital deployed by agents following a given strategy is analogous to the population of that species. Although, the agent-based models are still in their infancy, the simulations and related theory have already demonstrated an ability to understand many aspects of financial markets [81].

Michel Dacorogna et al. [55] consider the concept of heterogeneous markets, which is supported by empirical facts, the most elegant way to reconcile the EMH and price

predictability. Thomas Lux and Michele Marchesi [162] developed simulation models of financial markets that include agents with different strategies. They show that this model can reproduce most of the empirical regularities: fat tails, long memory and scaling law.

Differentiating expectations according to their time dimension distinguishes the market participants with different time scales. The interaction of components with different time scales gives rise to relativistic effects, i.e. the dynamic interaction between different market components relative to each other [55]. As a result, specific price movements cannot lead to a uniform reaction, instead they lead to individual reactions of different components. Blake LeBaron [146] shows that introducing agents with different time horizons gives rise to long-term clustering effects in observed price volatility. With the arrival of new faster technology, users are able to identify additional trading opportunities at very short intervals. This accelerates the pace of trading and contributes to higher market volume and liquidity, which in turn decreases the bid-ask spreads. Lower transaction costs open new horizons for profitable trading. In this way, according to Michel Dacorogna et al. [55] new technology introduces a shift in perspective, with components focusing on more numerous time intervals. The idea of their argument is that more trading does not necessarily trade away all anomalies. The described development indicates that the weak form of efficiency, as conventionally defined, cannot be attained.

Andrew Lo [159] introduces a new framework, the adaptive markets hypothesis (AMH), in which the traditional models of financial economics can coexist alongside behavioral model in a consistent manner. The AMH implies that the degree of market efficiency is related to environmental factors characterizing market ecology as well as the number of competitors in the market, the magnitude of profit opportunities and the adaptability of market participants. Many behavioral patterns taken as violations of rationality are actually consistent with an evolutionary model of individuals adapting to changing environment. This new framework is based on some principles of evolutionary biology – competition, mutation, reproduction and natural selection. Although, the AMH is still primarily qualitative and descriptive framework, it yields surprisingly concrete insights.

At present, there seems to be little consensus as to what empirical properties an efficient market should display. Generally the researchers do not abandon the concept of market efficiency. It just should be adapted and a good measure of how well a market operates should be found. High-frequency data research, extensive computing power and focus on dynamic and nonlinear processes in price series should help us to achieve this aim.

## 1.2. Technical Analysis

There are two competing ways to forecast the price development of financial instruments: fundamental and technical analysis. The fundamental analysis relies on the fundamental attributes of the instrument, such as price/earning ratio, return on investment and associated economic statistics. The aggregation of these measures provides an intrinsic value of the instrument, which in an efficient financial market should be equal to the trading price

of the instrument. Unfortunately, this is not the case observed in reality. The existence of a human factor brings distortions causing deviations of the trading price from its intrinsic value. Technical analysis is aimed at detecting a psychological component of financial trading and consequently converting findings into profit.

Technical analysis is a technique to draw investment decisions for individual financial instruments based on the development of their recent price, traded volume and other quantitative measures. The first application of technical analysis dates back to the 1600s. It was used on the Dojima Rice Exchange to forecast the price development of rice and its futures. Due to the cultural isolation technical analysis was rediscovered by western investors only in the early 20th century. This principle was initially branded as chartism, after the technique to analyze charts and draw investment decision rules. Since then, this development has been enriched by the introduction of computers which took tedium out of complex mathematical manipulations [188, 254].

There is a growing evidence that many contemporary professional investors use technical analysis. The profitability of technical analysis comes in contradiction with the Efficient Market Hypothesis introduced by Roberts [208] and propagated by Fama [73]. The EMH postulates that in efficient markets it is impossible to profit by predicting price development based on its past performance.

Although there is a vast amount of research investigating different aspects of the EMH, only recently a controversial issue has been pointed out in the literature, that whereas in the informationally efficient market, asset prices should evolve as a random walk (in line with the EMH), practitioners kept on using the technical trading rules in making their investment decisions, which is justified only if dependences in conditional returns exist [81, 55]. The return and price predictability is evident not only in the stock markets, where irrational traders can drive the prices from their equilibrium values and increase the price predictability, but also in the currency markets dominated by professional investors [5, 240]. Even though, Bollerslev and Hodrick [33] summarize in their work, that high frequency returns are approximately linearly unpredictable, they are characterized by predictable volatility clustering. Moreover, Kandel and Stambaugh [119] note that evidence of stock return predictability that seems weak when evaluated by classical statistical criteria may be important if a rational Bayesian investor would substantially alter portfolio holdings in response to the evidence.

Based on this, one might argue that individual investors are aware of short-term stock market returns and price predictability and try to exploit the trading opportunities by using technical analysis.

This section is organized as follows. First, Subsection 1.2.1 presents the history of technical analysis. Subsection 1.2.2 discusses the importance of technical analysis perceived by practitioners, which justifies its existence. It is followed by Subsection 1.2.3, where a classification and outline of most prominent technical trading rules are presented. Subsection

1.2.4 describes implementations of mechanisms to generate technical trading rules. The section is concluded by Subsection 1.2.5 on results of technical analysis validation.

### 1.2.1. History of Technical Analysis

Technical analysis accompanies investors for as long as people publicly trade investment instruments. With the first application dated back to the 1600s the TA was rediscovered in the 20th century to monitor the economic activity with the Dow theory on market indexes. Refocusing the attention from aggregate indexes to individual financial instruments produced technical analysis which is known today. With many supporters and as many opponents the history of technical analysis is the refinement of original ideas and their rejection and acceptance by financial community.

The Japanese were the first to use technical analysis to trade rice on the Dojima Rice Exchange in Osaka as early as the 1600s. A Japanese man called Munehisa Homma who traded in the futures markets in the 1700s discovered that although there was a link between supply and demand of rice, the markets were also strongly influenced by emotions of traders. As a result there could be a vast difference between the value and price of rice. Homma realized that he could benefit from understanding the emotions to help predict the future prices. He formulated his trading principles in two books, *Sakata Senho* and *Soba Sani No Den*, which were said to have been written in the 1700s. His work, as applied to the rice markets, evolved into the candlestick methodology currently used in Japan. Unfortunately, the results of four hundred years old studies were isolated by cultural and language barriers from the western world up to a moment when they have been rediscovered in the second half of 20th century [188, 254].

In the western world technical analysis starts in the early 20th century with the Dow theory. The theory was developed by Charles Dow based on his analysis of market price action in the late 19th century. Charles Dow never wrote a book or scholarly article on his theory. Instead, he put down his ideas of stock market behavior in a series of editorials that *The Wall Street Journal* published around the turn of the century. In 1903, the year after Dow's death, S. A. Nelson compiled these essays into a book entitled *The ABC of Stock Speculation*. In this work, Nelson first coined the term "Dow's Theory". In 1922, William P. Hamilton categorized and published Dow's tenets in a book titled *The Stock Market Barometer*. Robert Rhea developed the theory even further in the *Dow Theory* (New York: Barron's), published in 1932 [181].

The Dow theory addresses the fundamentals of technical analysis as well as general principles of financial markets, which are primarily applied to stock indexes. The theory assumes impossibility of manipulating the primary trend<sup>16</sup>, while at short time intervals or with individual stocks it could be prone to manipulation by large institutional investors, speculators,

---

<sup>16</sup>*Primary trend* is a long-running (up to five years) general movement in price data.

breaking news or rumors [206]. It assumes that the market should reflect all available information, i.e. prices represent the total sum of all hopes, fears and expectations of all participants.

The theory identifies three types of index price movements:

- *Primary movements* last from a few months to several years and represent a broad underlying trend of the market. These movements are typically referred to as *bull* and *bear* markets. Through a set of guidelines, the Dow theory enables investors to identify the primary trend and invest accordingly.<sup>17</sup>
- *Secondary (or reaction) movements* last from a few weeks to a few months and move counter to the primary trend. In a bull market a secondary move is considered a correction. In a bear market, secondary moves are sometimes called *reaction rallies*. William Hamilton noted some characteristics that were common to many secondary moves in both bull and bear markets. These characteristics should not be considered as rules, but rather as loose guidelines to be used in conjunction with other analysis techniques.
- *Daily fluctuations* can move with or against the primary trend and last from a few hours to a few days, but usually not more than a week. Daily price movements are important, but only when grouped with other days to form a pattern for analysis. The study of daily price action can add a valuable insight, but only when taken in the context of a larger picture.

William Hamilton identified three stages to both primary bull and bear markets. A primary bull market is defined as a long sustained advance marked by improving business conditions that elicit increased speculation and demand for stocks. A primary bear market is defined as a long sustained decline marked by deteriorating business conditions and a subsequent decrease in demand for stocks.

In the Dow theory Robert Rhea [206] identified four separate theorems that addressed trend identification, buy and sell signals, volume, and trading ranges. The first two were deemed the most important and served to identify the primary trend as bullish or bearish. The second two theorems, dealing with volume and trading ranges, were not considered instrumental in primary trend identification by Hamilton. Volume was looked upon as a confirming statistic and trading ranges were thought to identify periods of accumulation and distribution.

William Hamilton and Charles Dow openly admit that the Dow theory is not a sure-fire means of beating the market. It is looked upon as a set of guidelines and principles to assist investors with their own study of the market. The Dow theory provides a mechanism to help make decisions less ambiguous [206].

During the 1920s and 1930s, Richard W. Schabacker reopened the subject of the Dow theory in a somewhat new direction. He realized that whatever significant action appeared

---

<sup>17</sup>Trying to predict the length and the duration of the trend is an exercise in futility. Hamilton and Dow were mainly interested in catching the big movements of the primary trend. Success, according to Hamilton and Dow, is measured by the ability to identify the primary trend and stay with it [206].

in a stock index it must derive from similar action in constituent stocks. In his books, *Stock Market Theory and Practice*, *Technical Market Analysis* and *Stock Market Profits*, Schabacker showed how the principles of the Dow theory can be applied to the charts of individual stocks [68].

Further development of technical analysis was pretty straightforward. First, Richard Schabacker, the interpreter of the Dow theory, was joined by Robert D. Edwards. Then, in 1942 John Magee joined the study of technical analysis. With his participation the entire process of technical evaluation became more scientific. As a result of their research from 1942 to 1948, Edwards and Magee developed new technical methods of technical analysis. They put these methods to practical use in actual market operation. And eventually, in 1948, these findings were published in their definitive book, *Technical Analysis of Stock Trends* [68]. The 8th edition of this book was published in 2001. It demonstrates strong interest of investors in methods of technical analysis.

### 1.2.2. Importance of Technical Analysis: The Prospective of Technical Analyst

*"There is nothing so disastrous as a rational investment policy in an irrational world."*

John Maynard Keynes [230]

This citation is a key to understanding the perspective of technical analysts. As pure practitioners they do not have illusions about the true nature of financial markets. Markets are irrational and they learn how to deal with it. The history of technical analysis reveals that its rationale is to capture the irrational or emotional component of financial markets.

Following Steve Nison [188], a recognized expert of technical analysis, technical analysts distinguish five points, which define the importance of technical analysis:

- (1) While fundamental analysis may provide a gauge of the supply/demand situations, price/earnings ratios, economic statistics, and so forth, there is no psychological component involved in such analysis. Technical analysis provides the only mechanism to measure the "irrational" (emotional) components present in all markets.
- (2) The application of technical analysis allows investors to separate investment decisions from investors sentiments and to see the market without the prism of subjectivity.
- (3) Following technical analysis is important even if one does not fully believe in it. This is because, at times, technical analysts are the major reason for a market move. Since they are a market moving factor, they should be watched.
- (4) People remember prices from one day to the next and act accordingly. Peoples' reaction affect prices, but prices also affect peoples' reactions. Thus, price, itself, is an important component in market analysis.
- (5) The price action is the most direct and easily accessible method of seeing combined action of different factors.



All but the second point seem to be acceptable. The second point is unrealistic since it requires enormous self-control of a technical analyst. To make it valid one would need a machine with intelligence and expertise of a technical analyst and zero whatsoever emotions.

### 1.2.3. Taxonomy of Technical Analysis

Technical analysis is the practice of identifying recurring patterns in historical prices in order to forecast future price trends.<sup>18</sup> The technique relies on the idea that, as Martin Pring [204] puts it - prices move in trends which are determined by the changing attitudes of investors toward a variety of economic, monetary, political and psychological forces. Detection of trends is performed through indicators or technical rules which are aimed to capture underlying dependencies.

Following the classification by Christopher Neely [183] the methods of technical analysis attempt to identify trends and reversals of trends. To distinguish trends from shorter-run fluctuations, technical analysts employ two types of analysis: *charting* and *technical (mechanical) rules*. Charting, the older of the two, involves graphing the history of prices over some period - determined by a practitioner - to predict future patterns in the data from the existence of past patterns.<sup>19</sup> The second type of methods, technical rules, imposes consistency and discipline on technical analysts by requiring them to use rules based on mathematical functions of present and past prices.

**1.2.3.1. Charting.** To identify trends through the use of charts, technical analysts must first find *peaks* and *troughs* in the price series. A peak is the highest value of the price within a time interval under consideration (a local maximum), while a trough is the lowest value the price has taken on within the same time period (a local minimum). A series of peaks and troughs establishes downtrends and uptrends, respectively.

Detecting a trendline allows technical analysts to issue a short-term investment recommendation. Usually if an uptrend is detected the recommendation is a long position, alternatively, for the downtrend it is a short position.

Spotting the reversal of a trend is just as important as detecting trends. Peaks and troughs are important in identifying reversals too. Local peaks are called *resistance levels*, and local troughs are called *support levels*. If the price fails to break a resistance level (a local peak) during uptrend period, it may be an early indication that the trend may soon reverse.

Technical analysts identify several patterns that are said to foretell a shift from a trend in one direction to a trend in the opposite direction. The best known type of reversal formation called "head and shoulders". The head and shoulders reversal following an uptrend is characterized by three local peaks with the middle peak being the largest of the three. The line between the troughs of the shoulders is known as the "neckline". When the price

<sup>18</sup>John Murphy [180] defined the technical analysis, as a study of market action, primarily through the use of charts, for the purpose of forecasting future price trends.

<sup>19</sup>Its advocates admit that this subjective system requires analysts to use judgement and skill in finding and interpreting patterns [183].

penetrates the neckline of a heads and shoulders, technical analysts confirm a reversal of the previous uptrend and issue a recommendation to take short position.<sup>20</sup>

Another method of charting is the *candlestick technique* developed in Japan more than four centuries ago for rice and its futures market. The technique is based on the recognition of visual patterns that take the shape of candlesticks. Every candle includes information on the high, low, opening and closing prices of a particular time interval. The figurative "body" shows the difference between opening and closing prices, and its length depends on this difference. If the closing price is higher than the opening price, the body is white, which signals rising prices. If the opening price is higher than the closing price, the body is black, which signals falling prices. Above and below the candle's body are the "shadows", called upper shadow and lower shadow. They depict the high and the low of the trading interval [88]. In general, the candlestick technique consists of a set of patterns, defined by candlesticks, and respective expectations of market reaction.

The advantage of candlestick technique is that, first, it allows to express several relative to each other values within one graphical symbol.<sup>21</sup> Second, this technique can be easily combined with other charting methods or with technical rules.

In general the problem with charting is that it is very dependent on the interpretation of a technical analyst who is drawing the charts and interpreting the patterns. Subjectivity can permit emotions like fear or greed to affect the trading strategy. Technical rules make the analysis more consistent and disciplined and thus allow to avoid the problem of subjective analysts' judgment [183].

**1.2.3.2. Technical Rules.** There are many types of technical rules. In general, they aim at identifying the initiation of new trends. The best known technical rules are the following:

- (1) *Filter rules* - buy when the price rises by a given proportion above a recent through.
- (2) *Trading Range Break* or *Channel rules* - buy when the price rises by a given proportion above a recently established trading range.
- (3) *Moving Average rules* - buy when the current price level is above the moving average.
- (4) *Moving Average Intersection rules* - buy when a shorter moving average penetrates a longer moving average from below. They can have a form of Variable Length Moving Average or Fixed Length moving average, which differs in the number of days during which the buy or sell signal is assumed to be issued.
- (5) *Oscillator rules* - buy (sell) when the oscillator index takes an extremely low (high) value. A simple type of oscillator index is a difference between two moving averages: with short and long horizons [147, 183].
- (6) *Statistical rules* are based upon *ARMA*-family models for rescaled returns. The rules rely on a standardized forecast, given by the one-period-ahead forecast divided

<sup>20</sup>For more details and examples of charting technique see [68, 181].

<sup>21</sup>In this way it is similar to hieroglyphics, where each symbol is a word or a combination of words.

by an estimate of its standard error. For example, for  $ARMA(1, 1)$  an upward trend is predicted when the value of the standardized forecast is positive [242].

- (7) *Other rules.* Many technical analysts assign a special role to round numbers in support or resistance levels, and to historical record prices.<sup>22</sup> Other prominent types of technical analysis use exotic mathematical concepts such as Elliot wave theory and/or Fibonacci numbers.<sup>23</sup> Finally, technical analysts sometimes use technical analysis of one market's price history to take positions in another market, a practice called intermarket technical analysis [183].

Each rule has a mirror equivalent, which suggests short position. In each case a technical analyst has to choose the time horizon over which troughs and peaks are identified and moving averages calculated as well as the threshold before a decision is made.

#### 1.2.4. Implementation of Technical Analysis

Today there are two ways of implementing the mechanism of technical analysis: tabulation and adaptive evolution. Classical technical analysis is basically a tabulation of visual patterns and mathematical formulas. As such one way to implement the mechanism of technical analysis is to check data against all known chart patterns and technical rules. Unfortunately, their public knowledge plays against them. Competing technical analysts immediately detect and "cash" on these patterns. As a result the known chart patterns or technical rules lose original profitability.

With the development and spread of high-power computers technical analysts get access to new methods of technical analysis which are either adaptive to changing environment or in addition able to evolve new rules in on-line fashion. These methods constitute the second way of implementing the mechanism of technical analysis. Since this is the most promising way the rest of this subsection is devoted to it.

**1.2.4.1. Non-Linear Adaptive Methods.** In this strand of literature financial time series are modeled and treated as chaotic time series, which are assumed to have some hidden structure. To show the applicability of this method Doyne Farmer and John Sidorowich [80] presented a forecasting technique for chaotic data. After embedding a time series in a state space using delay coordinates, they obtained the induced nonlinear mapping using a local approximation. Then they used this technique to make short-term predictions of the future behavior of a time series, using information based only on past values. The results demonstrated its effectiveness on several examples, including data from Mackey-Glass delay differential equation, Rayleigh-Benard convection and Taylor-Couette flow.

Other examples of applying non-linear adaptive methods are presented by William Clyde and Carol Osler [45], who developed a method for nonlinear forecasting on high dimension

---

<sup>22</sup>One can argue that this rule captures the effect of the "psychological" barrier, which market has to overcome.

<sup>23</sup>To get more details on Elliot wave theory and Fibonacci numbers see [180].

systems, or by Lo et al. [158], who proposed a systematic and automatic approach to technical pattern recognition using nonparametric kernel regression.

Fernandez-Rodriguez et al. [83] applied a nearest-neighbour prediction method. This method, based on the literature on forecasting in non-linear systems, allows to make short-term forecasting of future behavior of a time series using information based on past values only.<sup>24</sup>

Spyros Skouras [227] developed artificially intelligent agents who learn from experience - artificial technical analysts. Skouras showed that artificial technical analysts are able to select trading rules, which can lead to the recognition of subtle regularities in return processes.

Walid Ben Omrane and Herve Van Oppens [192] used a smoothing technique to identify regularities in the time series of currency prices. The smoothing technique allows to identify significant price movements which are only characterized by sequences of extrema.

**1.2.4.2. Methods Based on Artificial Neural Networks.** Artificial neural networks (ANN) allow to detect patterns in time series without their prior specification. The method has an advantage that it can be easily extended to multivariate case but it requires enormous amount of data for training and learned patterns are hidden in a "black box".<sup>25</sup>

Recent examples of ANN application are Fernandez-Rodriguez et al. [84], where artificial neural networks generate simple trading rules, or Monica Lam [142], who applied neural networks with the backpropagation to integrate fundamental and technical analysis for financial performance prediction.<sup>26</sup>

**1.2.4.3. Methods Based on Genetic Algorithms.** Genetic Algorithms (GA) allow to solve complex non-linear problems within acceptable time. This feature is used to find new technical rules which are initially set as a random combination of elements describing time series patterns. By using the evolution principles GA recombine the best elements of trial technical rules to find superior specimen. After sufficiently large number of iterations GAs produce a population of technical rules which best fits the sample data.<sup>27</sup>

Franklin Allen and Risto Karjalainen [6] demonstrated a classical example of using GA to find a set of technical rules. Mariano Matilla-Garcia [171] used the concept of GA to approximate the equation, which described a given time series, in the specific case of daily security prices.<sup>28</sup>

**1.2.4.4. Methods Based on Learning Classifier Systems.** Learning classifier systems (LCS) are an extension of GA, where optimization is conditional on matched environment

<sup>24</sup>For details on the univariate nearest-neighbour prediction method see [80] and for the multivariate case see [82]. For more examples of nearest-neighbour in technical analysis see [85].

<sup>25</sup>To overcome the latter limitation a modification of artificial neural networks was proposed - self-organizing maps (SOM). For details on SOM see [102, 133, 134].

For more details on ANN see Section 2.3.1.

<sup>26</sup>For more examples of ANN in technical analysis see [9, 91, 149, 258].

<sup>27</sup>For more details on GAs see Appendix B.

<sup>28</sup>For other examples of using GA in technical analysis see [135].

conditions. As such LCS are able to identify several sub-problems and to find optimal solutions for each of detected scenarios.<sup>29</sup>

The work of Arthur et al. [10] is one of the pioneering attempts to apply LCS in technical analysis. They designed the implementation of LCS to find a combination of technical rules out of fixed predefined set. They demonstrated the advantages of LCS, first, that it was able to skip some rules if their presence was not necessary and, second, after learning the importance of each technical rule could be easily accessed.<sup>30</sup>

**1.2.4.5. Methods Based on Genetic Programming.** Genetic Programming (GP) is another extension of GA. The idea of GP is to write a program which best fits the problem by using the principles of GA.<sup>31</sup>

Recent example of implementing GP for technical analysis is Neely et al. [182]. They used this method to find novel technical trading rules in foreign exchange markets.<sup>32</sup>

### 1.2.5. Validation of Technical Analysis

This subsection presents the validation of technical analysis from empirical and theoretical points of view. In the first part the empirical validation, listed in chronological order, shows the struggle of technical analysis, first, against the random walk theory and, later, to prove its profitable application. The second part investigates whether technical analysis can be consistent with prevailing economic and financial theory.<sup>33</sup>

**1.2.5.1. Empirical Validation of Technical Analysis.** In the western world the history of technical analysis starts with the Dow theory developed by Charles Dow in the beginning of twentieth century. In 1925 the Dow theory and its application became a hot topic for debates in the American Statistical Association (ASA). The ideas of Charles Dow were either unconditionally accepted to, for example, monitor economic activity with the Dow Jones Industrial Average (DJIA) or inspired to invent similar barometer indexes. The attitude that time was to rely on economic factors to perform prediction rather than on the market itself [220].

Even though the Dow theory was not meant to be used for forecasting trends with individual stocks, in the following years the original idea was scaled down and adopted for gaining profit. In 1932 at the meeting of the ASA several examples of deriving empirical rules to predict price movements for profit were reported. Here technical analysis was taken as a trigger mechanism: economics may load the gun, select the mark, but technics [technical analysis] pulls the trigger [129].

Alfred Cowles [50] was one of the first scholars who analyzed the profitability of technical analysis. In 1933 he reported results of a hypothetical investment strategy based on market

<sup>29</sup>For more details on LCS see Section 2.3.3.

<sup>30</sup>For other examples of using LCS in technical analysis see [9, 219].

<sup>31</sup>For more details on GP see Section 2.3.2.

<sup>32</sup>For other examples of using GP in technical analysis see [202, 209].

<sup>33</sup>For additional information see overviews conducted by Cheol-Ho Park and Scott Irwin [197] and Stephen Taylor [242].

forecasts of William Hamilton in his editorials to *The Wall Street Journal*. The hypothetical application of published forecasts of the stock market based on the Dow theory over a period of 26 years, from 1904 to 1929, achieved a result better than what would be ordinarily regarded as a normal investment return, but poorer than the result of a continuous outright investment in representative common stock for this period.

A year later, the topic of the ASA meeting was "Technical Methods of Forecasting Stock Prices". Here the link between the Dow theory and technical analysis was recognized. The latter was accepted as a mechanism to capture other than economic influence on stock prices. It was stressed that the risk of financial investments can be reduced by technical analysis, which should indicate timing of investment decisions [130].

In 1955 Charles Roos [213] conducted a survey of economic forecasting techniques known at that time. Technical analysis was included in the category of naive methods of forecasting. In spite of his scepticism he stressed that forecast techniques, including technical analysis, should concentrate on predicting turning points rather than the trend prevailing in the market.

The study of Harry Roberts [207] conducted in 1959 on American data, for both indexes and individual stocks, questioned the applicability of technical analysis since time series of prices seemed to follow an extremely simple chance model. He referred to Maurice G. Kendall [122], who obtained the same results for British stock indexes and American commodity prices in 1953. Moreover, Roberts found that even in 1934 Holbrook Working [255] achieved the same conclusion: that [financial] time series commonly possess in many respects the characteristics of series of cumulated random numbers.

In contrast to these results Hendrik S. Houthakker [109] in 1961 found elements of non-randomness in speculative price movements. He presented evidence that stop orders gave rise to a non-random profit. Sidney S. Alexander [3] by using 5-percent filtering of noise showed that after filtering large changes are more likely to continue than to reverse: in speculative markets price changes appear to follow a random walk over time, but a move, once initiated, tends to persist.<sup>34</sup>

Robert Weintraub [249] analyzed the pertinent literature of that time on testing technical analysis. He found that the studies up to 1963 were using too restrictive assumptions about the behavior and abilities of technical analysts. For example, Weintraub argued that Kendall's assumption of fixed interval between trades did not reflect the reality and reduced potential profit opportunities. By using more realistic varying waiting time Weintraub obtained results which speak more in favor of technical analysis than the random walk hypothesis. He concluded that the lack of serial correlation between first differences of closing prices simple meant that speculators [technical analysts] who were supposed to smooth out price movements over time were doing their job well.

---

<sup>34</sup>One can question his methodology and the choice of filter value of 5 percent.

Later, the general methodology of a filter rule has often been used in academic studies to separate days into two sets based upon a trader's market position [242].

In 1964 Sidney S. Alexander [4] tested a number of filter rules. Although they appeared to yield returns above the buy-and-hold strategy for the DJIA and S&P stock indexes, he concluded that adjusted for transaction costs, the filter rules were not profitable. Eugene Fama [70] came with even more restrictive conclusion: the data seem to present consistent and strong support for the [random walk] model. This implies, of course, that chart reading [technical analysis], though perhaps an interesting pastime, is of no real value to the stock market investor. In [75] Eugene Fama and Marshall Blume achieved similar conclusions,<sup>35</sup> which in 1970 led Eugene Fama [71] to dismiss technical analysis as a futile activity.

In 1967, in spite of the tendency to reject technical analysis, M. F. M. Osborne [194] found that applicability of the random walk theory and technical analysis can be dependent on used time frequency of prices. He concluded: in general shorter time interval data (daily, weekly) tend to show more "non-random walk" properties than for longer intervals (monthly).

James Jr. [113] extended the moving-average tests for monthly data to include averaging by exponential smoothing. To test the significance of findings he calculated for each sample stock the difference between the moving average and the buy-and-hold returns. The results showed that none of the rules or various smoothing coefficients was more profitable than the buy-and-hold strategy unless the data were unadjusted for dividends or recorded at higher than monthly frequency.

Robert Levy [154] tested the Relative Strength technique.<sup>36</sup> He found that the relative strength portfolios had higher returns than the buy-and-hold strategy, but also the risk was higher.

After Eugene Fama silenced empirical studies of technical analysis for almost twenty years, in the second half of the 1980s the interest of academic community returned to the topic. New empirical studies either found a proof of applicability of technical analysis or defined segments and markets, where the weak form of market efficiency prevails and technical analysis brings a small added value.

In 1984 Salih Neftci and Andrew Policano [185] used the moving-average and slope method to analyze the futures market. Their results suggested that by characterizing the actual behavior of market participants, improved price predictions can be obtained in the futures markets. A year later Richard Sweeney [234, 235] used the filter rule to challenge market efficiency, respectively, in the foreign exchange and US stock market. He showed using statistical tests that filter rules found in foreign exchange markets kept persistence. These tests were consistent with, but independent of, a wide variety of asset pricing models. Achieved profits could not be explained by risk if risk premia are constant over time. Similar results were obtained for the US stock market.

---

<sup>35</sup>They conducted a study of the thirty Dow Jones Industrial Stocks using 24 different filter rules, ranging in size from 0.5% to 50% for the time period from 1957 to 1962. They concluded that the filter rules were not profitable when the effect of interim dividends and brokerage commissions were considered [75].

<sup>36</sup>This technique selects those stocks that were performing best, relative to their average price of the previous 26 weeks [154].

Trading rules based upon  $ARMA(1, 1)$  models are investigated in several publications, commencing with Stephen Taylor [237, 238]. Stephen Taylor [239] compared filter, channel and moving-average trading rules with rules that use  $ARIMA$  price forecasts by evaluating their ex ante performance for currency futures transactions from December 1981 to November 1987. He found the support of profitability of the studied trading rules. Applied Monte Carlo studies strongly suggested that the trading profits were too large to be explained by the elusive, time-varying risk premium sought in forward market literature.

Lukac et al. [160] found that a channel rule performed best in their study of several technical trading systems.<sup>37</sup> Later, Louis Lukac and Wade Brorsen [161] found that many rules applied to commodity and financial futures prices made profits, but they did not conclude that these futures markets were inefficient. David Brown and Robert Jennings [38] showed in their model that the use of technical analysis has a value when the prices are not fully revealing and traders have rational conjectures about the relation between prices and signals. Richard Levich and Lee Thomas [153] and Carol Osler and Kevin Chang [195] selected and tested the technical trading rules on the US dollar exchange rates and found statistically significant excess returns.

Steven Dawson [57] studied 292 technical-analysis-based investment recommendations made over a five year period, November 1979 to April 1984, by a Singapore investment advisory firm. The objective was to test whether this firm was able to use technical analysis to select common share investments which would allow investors to earn an excess return. After adjusting for trading commissions, market trends and risk, the recommended shares did not outperform the market. He concluded that technical analysis could not reliably forecast future prices and lead investors to returns which exceed the market.<sup>38</sup>

In 1992 the study by Brock et al. [36], demonstrated that a relatively simple set of technical trading rules possess significant forecast power for changes in the DJIA over a long sample period. They used trading rules to learn about the conditional means and variances of future returns. In particular, various technical rules were used to identify Buy and Sell days for the DJIA from 1897 to 1986. The results showed significant, positive values of the mean return. They concluded that buy signals consistently generate higher returns than sell signals,<sup>39</sup> returns following sell signals are negative, which is not easily explained by any of the currently existing equilibrium models, returns following buy signals are less volatile than returns following sell signals.<sup>40</sup> In general, the first and second moments of

<sup>37</sup>Their version of the channel trading rule assumes that a futures trader is always in the market.

<sup>38</sup>These results correspond with what would be expected in a weakly efficient market. At the same time one can question proficiency of the company in technical analysis.

<sup>39</sup>The average returns on Buy and Sell days, across all parameter combinations, are respectively equivalent to 12% and -7% per annum. Tests on the differences between Buy and Sell average returns, using the  $z$ -statistics, provide highly significant values of  $z$  for each of the ten parameter combinations evaluated, ranging from 3.79 to 6.04. The null hypothesis is also rejected at the 5% significance level for each subperiod considered: 1897-1914, 1915-1938, 1939-1962 and 1962-1986. The respective subperiod difference between annualized Buy and Sell returns are 18%, 27%, 11% and 12% [36].

<sup>40</sup>Standard deviations of 0.89% and 1.34% are reported, respectively for Buy and Sell days [36].



these returns were concluded to have been predictable to some degree and hence technical rules have been informative.<sup>41</sup>

Hendrik Bessembinder and Kalok Chan [27] reported that the rules presented by Brock et al. [36] were useful also for forecasting index returns for a group of Asian Stock markets. They documented the success of similar rules in forecasting changes in currency exchange rates.<sup>42</sup> In the consequent study [28] they stated that the forecast power was only partially attributable to return measurement errors arising from nonsynchronous trading.<sup>43</sup> However, Bessembinder and Chan argued that this evidence can coexist with the notion of market efficiency since breakeven trading costs were computed and found to be smaller than actual transaction costs.

In 1996 Hudson et al. [110] applied the methodology of Brock et al. [36] to UK stock index, FT-30, and the time frame from 1935 to 1994. All their average buy returns were significantly higher than average sell returns, at low significance levels. The overall averages were equivalent to annual rates of 16% on Buy days and -6% on Sell days. Any profits from trading rules were reported to be less than transaction costs.

Jonathan Batten and Craig Ellis [20] studied Australian All Ordinaries Share Price Index time series data over the period 1987-1991. They considered the ability of technical trading systems to generate returns greater than the buy-and-hold strategy. The results showed that none of the trading systems employed were able to earn a return greater to the buy-and-hold strategy once transactions costs were taken into consideration.<sup>44</sup>

Fernandez-Rodriguez et al. [83] applied the nearest-neighbour prediction method to the General Index of the Madrid Stock Market for the period 1968 to 1994. By using Theil's  $U$ -statistic to measure the forecasting performance, they reported that the nearest-neighbour prediction performed marginally better than the random walk strategy, outperforming the random walk directional forecast. When they were accessing the economic value of the nearest-neighbour prediction, the results of this method as a filter technique were always superior to the buy-and-hold strategy assuming zero transaction costs.<sup>45</sup>

Terence Mills [174] analyzed daily data of the London Stock Exchange FT30 index for the period 1935 to 1994. By assessing the statistical significance of technical rules via  $AR-ARCH$  models and bootstrapping techniques he found that technical analysis produced a

<sup>41</sup>Acar [1] has presented several related results for the single moving average rule without a band. He shows that the single moving-average rule has optimal properties when returns follow a Gaussian  $ARMA(1,1)$  process.

<sup>42</sup>Over the period from 1975 to 1989 the annualized difference between Buy and Sell returns for the moving-average rule averages 8% for Hong Kong, Japan and Korea and a massive 52% for the emerging markets of Malaysia, Thailand and Taiwan [27].

<sup>43</sup>In 2002 Theodore Day and Pingying Wang [59] showed that the evidence of predictability may arise because the prices of component stocks in the index are not always synchronous. They found that the difference between buy and sell average returns from 1962 to 1986 was much smaller for a value-weighted index constructed from the DJIA stocks, which gives less weight to smaller firms whose stocks may trade less often [242].

<sup>44</sup>As the second goal the authors tested the weak-form efficiency of the Australian Stock Market, which is considered in both the statistical context and in terms of the trading system net returns. Statistical test results showed that the Australian Stock Market was weak-form efficient [20].

<sup>45</sup>For more recent results see [85].

return greater than the buy-and-hold strategy, for most of the sample period, at least up to the early 1980s. Since then Terence Mills reported a clear dominance of the buy-and-hold strategy.

Diminishing usefulness of technical analysis was observed as well for the US stock market. Although Sullivan et al. [232] reported significant excess returns to technical trading rules, their out-of-sample performance was not satisfactory. Sullivan et al. [232] found statistically significant evidence that trading rules provided information about the conditional mean of the DJIA until 1986. However, they found no evidence that the rules were informative during the subsequent decade until 1996, for both the DJIA and futures on the S&P500 index [242].

Franklin Allen and Risto Karjalainen [6] used a genetic algorithm to learn technical trading rules for S&P composite stock index. They found that in the out-of-sample test period 1970-1989 the rules were able to identify periods to be in the index when returns were positive and volatility was low, and out when the reverse was true. Using data for other periods since 1929, the rules could identify high returns and low volatility but did not lead to excess returns after transaction costs.

Neely et al. [182], using genetic programming to generate technical trading rules, found strong evidence of economically significant out-of-sample excess returns. The found rules were applied to six exchange rates, over the period 1981-1995. When the rules generated for USD/DEM exchange rate were allowed to determine trades in the other markets, there was a significant improvement in performances in all cases, except for DEM/YEN. Betas calculated for the returns according to various benchmark portfolios provided no evidence that the returns to these rules were a compensation for bearing systematic risk.

Blake LeBaron [145] showed that simple rules used by technical traders have some predictive value over the future movement of foreign exchange prices. He explained that the reason can be in the nature of foreign exchange markets, where there are several major players whose objectives may differ greatly from those of maximizing economic agents. The results of the study showed that this predictability was greatly reduced, if not eliminated, when the days in which the Federal Reserve was actively intervening were removed.<sup>46</sup>

In 2000 Stephen Taylor [241] applied moving-average rules to several UK series recorded from 1972 to 1991. The moving-average rule produced values of test statistic that are significant at the 5% level for the FTA index<sup>47</sup>, the 12-share indexes and four of the twelve firms. The largest predictability was found in the FTA index.

Several rules based upon technical analysis were defined rigorously and investigated by Lo et al. [158]. They stated that some technical patterns provided useful information about the prices of US stocks.<sup>48</sup> However, expected returns following these patterns were

<sup>46</sup>Similar topic is discussed by Christopher Neely [184].

<sup>47</sup>FTA is the Financial Time All-Share index, calculated from the prices of more than 600 stocks.

<sup>48</sup>By comparing the unconditional empirical distribution of daily stock returns to the conditional distribution - conditioned on specific technical indicators such as head-and-shoulders or double-bottoms - they found that over the 31-year sample period, several technical indicators did provide incremental information and might have some practical value.

similar to unconditional expectations [116] and the same conclusion applied to UK firms [58].

Fernandez-Rodriguez et al. [84] by using artificial neural networks to generate simple trading rules for the General Index of the Madrid Stock Market suggested that in the absence of transaction costs the technical trading rules were always superior to the buy-and-hold strategy for both bear market and "stable" market episodes. For the bull market they found that the buy-and-hold strategy generated higher returns than the trading rules based on artificial neural networks.

Giles et al. [91] used neural networks and grammatical inference to predict daily foreign exchange rates for German Mark, Japanese Yen, Swiss Franc, British Pound and Canadian Dollar over time period of September 3, 1973 - May 18, 1987. They reported the correct prediction of change for the next day with an error rate of 47.1%.<sup>49</sup> They concluded that a particular recurrent neural network combined with symbolic methods could generate useful predictions and rules for such predictions on a non-stationary time series.

Bertrand Maillet and Thierry Michel [163] studied daily main European exchange rates from January 1989 to November 1996. Unlike the findings for main dollar exchange rates, there was no significant technical analysis profitability for the European currencies. Moreover, the significance did not depend on the length of the period over which the moving averages were computed, although the return differential between the naive and chartist strategies was positive for almost every combination of short and long moving average on some European currencies. They concluded that with low-volatile European currencies technical analysis had no added value.

Walid Ben Omrane and Herve Van Oppens [192] investigated the existence of chart patterns in the euro/dollar intraday foreign exchange market. They used two identification methods of the different chart patterns: one built on 5-min close price only, and one based on both 5-min low and high prices. To obtain statistical significance of results they used Monte Carlo simulation. With the first method they found an apparent existence of chart patterns in the currency market. Six out of twelve detected chart patterns showed significant predictability. But only two patterns implied significant profitability, which was still insufficient to cover transaction costs. The second method provided higher but more risky profits than with the first one.

Armano et al. [9] developed an artificial technical analysis expert, which combined learning classifier systems and artificial neural networks. They performed experiments on the major Italian stock market COMIT index (from April 1992 to May 2000) in the presence of trading commissions. The results pointed to the good forecasting capability of the proposed approach, which allowed outperforming the buy-and-hold strategy as well as predictions obtained by using just recurrent neural networks.

Jerzy Korczak and Patrick Roger [135] used genetic algorithms to search a set of technical rules which gives buying and selling signals. They tested the approach on a sample

---

<sup>49</sup>The error rate reduced to around 40% when rejecting examples where the system had low confidence in its prediction.

of 24 French stocks traded on the Paris Stock Exchange in the period of 1997-1999. They showed that in most cases their method outperformed the simple buy-and-hold strategy.

Sanjay Sehgal and Anurag Garhyan [221] evaluated whether stock investment recommendations based upon technical analysis can provide abnormal returns in the Indian capital market. They employed several return measures including those adjusted for market trend, risk and transaction costs. The study covered 21645 recommendations for 21 companies using 13 technical indicators. The mean return was found to be statistically significant even after the adjustment for transaction costs, but the gains disappeared once the market-adjusted measures were used. Mitra [176] examined the applicability of moving-average-based techniques and filter rule technique in the same market. He found that profit was high in moving-average crossover with periods of 2 and 10 days. All low value filters gave profits, however, the profitability became negative with high value filters. The filter of 1 percent of stock price was found suitable for all examined series and the use of filter method was found to be less risky.

In 2003 Fang and Xu [79] compared DJIA average returns to the moving-average rule with a time series rule based upon forecasting returns using an  $AR(1)$  process, from 1896 to 1996. The average difference between buy and sell returns for the time series rule was more than double the figure for the moving-average rule.

Potvin et al. [202] studied the applicability of genetic programming on 14 Canadian stocks listed on the Toronto Stock Exchange in the period of June 30, 1992 - June 30, 2000. Data included the stock price and the transaction volume for each working day. The results showed that the trading rules generated by genetic programming were generally beneficial when the market was bearish or when it was stable. At the same time these results did not match the buy-and-hold strategy when the market was bullish.

Monica Lam [142] applied neural networks with backpropagation algorithm to integrate fundamental and technical analysis for financial performance prediction. The study covered 365 common stocks constituent of S&P500 for the period of 1985-1995. The performance of neural networks was compared with the average return from the top one-third returns in the market (maximum benchmark) that approximated the return from perfect information as well as with the overall market return (minimum benchmark) that approximated the return from highly diversified portfolios. The results showed that neural networks using one year or multiple year financial data consistently and significantly outperformed the minimum benchmark, but not the maximum one.

Matthew Roberts [209] employed genetic programming on futures prices from 24 markets. He found out that only two of the markets were capable of generating profits at the 5% level of significance using out-of-sample data.

Mariano Matilla-Garcia [171] used a genetic algorithm to generate technical trading rules. The method was applied on the Spanish IBEX-35 index over the period 2 January 1990 - 2 November 1999. The results suggested that in general the profitability of generated

trading rules was superior to the buy-and-hold strategy. This conclusion is clearer in bull, bear and volatile market episodes.

**1.2.5.2. Theoretical Validation of Technical Analysis.** Up to the 1990s there were close to none studies devoted to theoretical validation of technical analysis. The situation changed after numerous empirical examples showed that technical analysis was real and important phenomenon. The following part surveys studies which provide theoretical validation of technical analysis.

One of the first theoretical treatments of technical analysis is the discussion of the Dow theory by Robert Rhea [206] in 1932.

In the early 1980s a broad strand of research started to explore asset price models with speculators pursuing "non-rational" chartist and "rational" fundamentalist strategies. A typical example of this body of literature is a study of Avraham Beja and Barry Goldman [25]. They showed that incorporating of technical analysis could create models, which were more realistic and thus could help to explain empirical properties of financial markets.

In 1989 David Brown and Robert Jennings [38] constructed a two-period dynamic equilibrium model to demonstrate that rational investors should use historical prices in forming their demands.

Two years later Salih Neftci [186] attempted a formal study of technical analysis. He devised formal algorithms to represent various forms of technical analysis in order to see if these rules are well defined. He showed that most patterns used by technical analysts need to be characterized by appropriate sequences of local minima and/or maxima and will lead to nonlinear prediction problems. Since the theory of the minima and maxima of stochastic processes can be tedious [144], technical analysis may serve as a practical way of using the information contained in such statistics.<sup>50</sup>

In 1994 Blume et al. [31] investigated the informational role of volume and its applicability for technical analysis. They developed a new equilibrium model in which aggregate supply is fixed and traders receive signals with differing quality. Blume et al. concluded that volume provided information on information quality that could not be deduced from the price statistic. Traders who used information contained in market statistics did better than traders who did not. Technical analysis, thus, arose as a natural component of the agents' learning process.

In 1997 Arthur et al. [10] created a model, populated with artificial technical analysts, known as Santa Fe artificial stock market (ASM). The model allowed the authors to explain the contradiction between the theoretical literature and practitioners' view on the market efficiency. The simulation results showed that both views were correct, but within different regimes of the markets. The market settled into the rational-expectations equilibrium of the efficient market literature when the ASM agents had enough time to accumulate and process information. Otherwise, a "rush" in accumulating and processing market information, which is a property of real stock market investors, resulted in the consistence of the ASM

---

<sup>50</sup>Tests done using the DJIA for 1911-76 suggested that this might indeed be the case for the moving-average rule [186].

with the practitioners' believe of market predictability, and gave rise to technical trading. As a result, the market dominated by technical trading experienced temporary bubbles and crashes, which made the ASM under that regime to be very similar to the real stock market.

At the same time William Clyde and Carol Osler [45] provided theoretical foundations for technical analysis as a method for doing nonlinear forecasting on high dimension systems. They argued that traditional graphical technical modeling methods might be viewed as an equivalent to nonlinear methods that use the Taken method of phase space reconstruction combined with local polynomial mapping techniques for nonlinear forecasting. The study presented evidence in support of this hypothesis in the form of an application of the head-and-shoulders formation identification algorithm to high-dimension nonlinear data, resulting in successful pattern identification and prediction.

Alexandra Ilinskaia and Kiril Ilinski [141] used the framework of Gauge Theory of Arbitrage to show that technical analysis and market efficiency corresponded to different time regimens. From this point of view, technical analysis predictions exist due to internal deterministic dynamics, which brings the market to equilibrium. They showed that technical analysis indicators and the predictions existed for short time horizons while for long time horizons the model produced EMH state with realistic statistical data. The investors aware of this fact might increase the trading frequency to exploit these opportunities.

Spyros Skouras [227] analyzed the relationship between the efficiency of financial markets and efficiency of technical analysis. He used artificial technical analysts to provide a quantifiable measure of market efficiency. After application to the DJIA index from 1962 to 1986 Skouras derived the implications for the behavior of traditional agents.

Alan Morrison and Nir Vulkan [178] studied a version of the standard Kyle [139] model with endogenous information acquisition. They found that there was a robust equilibrium, which allowed free entry and in which speculators made positive profit.

### 1.3. Discussion

This chapter presented the overview of two competing concepts: efficient market hypothesis and technical analysis. The overview showed that from the very beginning technical analysis was "right on the money". For more than one hundred years in the western world and more than four centuries in Japan financial practitioners were successfully using technical analysis to process historical prices for profit. At the same time it took almost one hundred years for academicians to understand and to accept technical analysis as a valid tool in financial markets. The main argument against technical analysis is the EMH. Its proponents claimed that in competing markets any profitable information should be and already was incorporated into prices. As a result historical time series of prices should contain no information, and as such should be useless for technical analysis. In spite of this hypothesis substantial empirical support was presented for technical analysis as well as for market efficiency. Only recently analytical studies partially explained this phenomenon.

The first part of this chapter presented the concept of market efficiency. It was shown that in competing markets any information, which could bring profit, should be and already was incorporated into prices. As a result historical time series of prices should have no information, which could be turned into profit. Theoretical literature distinguishes three forms of market efficiency depending on which type of information - past prices, public or private information - should be incorporated into prices. The case, when past prices should be incorporated into prices, defines the weak form of the EMH.

Since this work investigates the weak form of efficiency the overview covers empirical studies where it is tested. Numerous studies found limited support for the weak form of the EMH, which varied across different markets, financial market segments and for different financial instruments. Apparent problems in testing were attributed to the joint hypothesis problem, where ambiguity of result interpretation could be either due to underlying model or market inefficiency. Growing evidence against market efficiency forced academicians to relax restricting assumptions and to turn attention to behavioral aspects of financial markets.

The second part of the overview covers technical analysis. This technique is designed to analyze historical prices in order to draw investment decisions. The rationale of this technique is based on capturing behavioral components in financial markets. The history of technical analysis is more than four centuries old. This allowed the technique to tabulate observed regularities in time series of asset prices. These regularities are recorded either in the form of chart patterns or as mathematical formulas. In addition to tabulation advances in science brought new methods of implementing technical analysis, including application of non-linear methods and artificial intelligence.

For a long time technical analysis was perceived as a "voodoo" of financial markets. Its explicit contradiction to the EMH was the main argument against its acceptance as a valid financial tool to analyze development of prices. Only an avalanche of empirical studies in support of technical analysis and a wide acceptance among practitioners<sup>51</sup> drew to it attention of financial theory. Many theoretical models showed that technical analysis could bring additional information to support investment decision and thus rational investor should incorporate it into arsenal of financial tools.

However, does the successful application of the technical trading rules mean that the financial markets are not efficient? The available evidence suggests that financial market returns are partly predictable, which is in conflict with the efficient market hypothesis. There are several responses to this evidence.

First, technical analysts support this conclusion by saying that even if all relevant information is actually reflected in the quoted price, market fluctuations are not exclusively

---

<sup>51</sup>In 1989 Helen Allen and Mark Taylor [5] surveyed that at the shortest horizons, intraday to one week, approximately 90% of respondents [professional traders and analysts at foreign exchange markets] used some chartist input in forming their exchange rate expectations, with 60% judging charts to be at least as important as fundamentals. Moreover, there appeared to be a persistent 2% of presumably "pure" chartists, who only used technical analysis.

exogenous (as the fundamentalists argue). There may also be endogenous irrational phenomena or speculative crises, and these might be sources of important fluctuations, which occurred in the past and can be repeated in the future.

Second, Andrew Lo and Craig MacKinlay [157] stressed the problem of data-snooping since most of the academic research uses the same dataset. Fama [74] makes the related point that many anomalies are sensitive to the research methodology used and disappear when changes in technique are applied.

Third, observed predictability may reflect time variation in the size of the risk premium [33]. The problem with this argument is that as Karen Lewis [155] points out, there are no satisfactory models of risk premia. Without such model, an explanation is empirically empty [24].

Fourth, in many cases market efficiency or applicability of technical analysis is conditional on transaction costs. While the first empirical studies were concentrated on rejecting predictability of prices, more recent publications showed that financial community finally realized presence of some degree price predictability, which is sustained as long as it is economically not profitable to exploit.

The reconciliation of the EMH and technical analysis became imminent. The main aim was to accept applicability of market analyses and at the same time not to discard the general concept of market efficiency. Alexandra Ilinskaia and Kiril Ilinski [141] concluded that technical analysis corresponded to different time regimens. From this point of view, technical analysis predictions exist due to an internal deterministic dynamics, which brings the market to equilibrium. They showed that technical analysis indicators and predictions existed for short time horizon while for long time horizon the model produces EMH state with realistic statistical data. The investors aware of this fact might increase trading frequency to exploit these opportunities. Similarly, Fernanda Strozzi and Jose-Manuel Zaldivar [231] used the state space reconstruction techniques to test the EMH for 18 high-frequency currency exchange rate series. They concluded that positive profits might be attained, so certain determinism was embedded. Although, the forecastability quickly disappears [127], one can still model high-frequency series better than with the random walk.

The cited studies generally use ad hoc specification of trading rules. As concluded in Section 1.1 on market efficiency, the proposed new directions of market efficiency research are more in favor of adjusting to changing environment. Spyros Skouras [227] argues that technical analysts should use different rules in different times and in different markets to be successful and one needs a model of how analysts adapt to the market environment. He provides such model by introducing artificial technical analysts who are artificial intelligent agents learning from their environment. The agent chooses among technical trading rules and his actions are the outcome of an explicit decision problem. Since the optimality of a trading rule can be judged only in the context of a specific decision problem, a generally optimal trading rule, as suggested in previous studies, does not exist. He refers to this version of the EMH as the *lack of intertemporal arbitrage hypothesis*, which has the advantage



of being testable independently of equilibrium model. In this way, he formalizes a way in which a market can be characterized as inefficient and provides a quantifiable measure of efficiency.

To conclude, the overview showed that to resolve the dispute between the proponents of market efficiency and technical analysts a well designed mechanism is required. To avoid the problem of joint hypothesis testing the mechanism should be build upon principles of technical analysis. To avoid the problem of subjectivity it should be independent of human influence but as intelligent, adaptive and proficient in technical analysis as professional technical analysts. Finally, test data should be free of any selection bias and the test should incorporate a variety of transaction costs.

## CHAPTER 2

### **Building an Artificial Technical Analyst**

The first chapter concluded that to correctly resolve dispute between practitioners and academicians on efficiency of financial markets one needs to design a model-free autonomous mechanism. Spyros Skouras [227] proposed a solution, which allows to test market efficiency independently of equilibrium model, - artificial technical analysts. His concept of the ATA will be used to design my artificial technical analyst, which will incorporate latest advances in artificial intelligence, pattern recognition and data preprocessing. This chapter is devoted to design of a new implementation of the ATA.

A blueprint of the ATA has two major components: data preprocessing and pattern recognition mechanism. The components will be introduced according to their appearance. First component is data preprocessing. It is followed by pattern recognition component, which, in its turn, includes piecewise linear approximation mechanism and the implementation of artificial intelligence - learning classifier systems.

The first component in the ATA is data preprocessing mechanism. Since the idea of technical analysis is the interpretation of past financial data, its quality and consistency crucially depends on the quality of data. To minimize the interference of errors due to the outsourcing of data preprocessing, raw data will be used as an input. For consistency raw data will be preprocessed. Two aspects of data preprocessing will be taken into account. The first one is a reduction of problem complexity. And the second one is a maximization of the informational content of data. Both aspects can be resolved through the mechanism of data homogenization.

The second component starts from piecewise linear approximation (PLA) mechanism. This mechanism serves to identify important peaks and troughs in time series of prices. Since the ATA should have as good proficiency in detecting peaks and troughs as expert technical analysts the quality of piece wise approximation has a crucial role. To pick its best implementation several competing alternatives will be considered. The Monte Carlo studies should insure that chosen implementation is the best for building the ATA.

The last component in the ATA is responsible for pattern recognition. It will use peaks and troughs, identified by PLA, to learn patterns in the data. The pattern recognition of the ATA is similar in a sense to the candlestick technique. Like with the candlestick patterns actual values of prices has no importance, but their relative values. In addition, the ATA will incorporate relative waiting time of appearance of each price observation. This approach will allow to see which patterns can predict price behavior.

The cognitive mechanism of the ATA is based on artificial intelligence. Since learning classifier systems (LCS) are one of few implementations of artificial intelligence which have

transparent structure of solutions, they will be used for building cognitive mechanism. In LCS there are many implementations, but all of them are dominated by a XCS, developed by Stewart Wilson [251, 252]. Unfortunately, the XCS does not have required power to solve aimed problem within reasonable time. To overcome this limitation a new implementation of LCS will be introduced - a true classifier system (TiCS), - developed by author.

This chapter is organized as follows. First, Section 2.1 presents the data preprocessing component. It is followed by Section 2.2, where different options of piecewise linear approximation are analyzed and the best is selected. Section 2.3 shows the selection process of the best implementation of artificial intelligence, and results of its trial tests. A discussion of the results concludes the chapter.

## 2.1. Data Preprocessing

Data preprocessing is an important component of the artificial technical analyst. The nature of technical analysis is the interpretation of past financial data. Thus, the quality and the consistency of data analysis crucially depends on the quality of data. To minimize the interference of errors due to the outsourcing of data preprocessing, raw data will be used as an input. For consistency raw data will be preprocessed. There are two aspects of data preprocessing, which should be taken into account. The first one is a reduction of problem complexity. And the second one is a maximization of the informational content of data. Both aspects can be resolved through the mechanism of data homogenization.

The issue of problem complexity arises from the structure of raw financial data. Normally raw data comes from financial markets at irregular intervals. This is a result of irregular news arrival to the market. Once the news reaches the market it causes asset re-evaluation, which, in turn, results in a new price tick.<sup>1</sup> Thus, every observation has two descriptors: the time stamp and the value of the observed variable. Irregularly spaced data requires attention as to the observed value, as well as, to the waiting time. In a high-frequency environment the irregular waiting time can be approximated by a fixed value. This approximation gives the reduction of problem complexity. The reduction is achieved through time series homogenization. With homogeneous time series one can concentrate on the analysis of values, leaving the timing issue to a more complex research.

For homogenization one can use either linear, or previous-tick interpolation. Linear interpolation fills the gaps between the observations by linearly interpolating two values. Previous-tick interpolation fills the gaps by repeating the last known value until the arrival of a new one. There is no consensus on which method to choose. The analysis of distributional properties of resulting price series does not show any significant difference. So the choice would be governed by convenience.

In either case one has to pick the frequency of homogenization at which data shows the highest past market inefficiency. The grid analysis can provide the answer. For this one has to search through the possible choices to find the frequency, at which revealed transaction

---

<sup>1</sup>This sort of data is known as a tick or tick-by-tick data.

costs (RTC, for details see Chapter C) is at maximum. The positive difference between RTC and actual transaction costs indicates observed market inefficiency, which could be exploited by technical analysis.

Technical analysis requires a subsample of recent data points. While the ATA can filter out observations irrelevant for technical analysis, the size of the subsample should be defined. By fixing the size of the subsample the consistency of the analysis is preserved. With the goal of maximizing the informational content to reveal market inefficiency, one can employ the same approach as for identifying the optimal frequency of homogenization. Moreover, the search for the optimal frequency and the size of the subsample can be combined. The grid analysis will go through the set of frequencies, and the set of subsample sizes.

This section has the following structure. Subsection two presents the linear and previous tick interpolation. Subsection three investigates the efficiency issue of homogenization. Subsections four and five describe the analysis performed to find the optimal frequency of the homogenization, and the optimal subsample size. A conclusion closes the section.

### 2.1.1. Time Series Homogenization

Most theoretical models and time series analysis takes a homogeneous time series for granted. In reality financial data comes from a stock market in inhomogeneous form. That is with times  $t_j$  and values:

$$z_j = z(t_j), \quad (2.1)$$

where index  $j$  refers to the irregularly spaced sequence of the raw series.<sup>2</sup> With the raw data in hand one has to construct the homogeneous time series. The values in the homogeneous time series are regularly spaced by  $\Delta t$  and rooted at a time  $t_0$ :

$$z_i = z(t_0 + i\Delta t), \quad (2.2)$$

where index  $i$  refers to the homogeneous series.

There are two scenarios of homogenization. According to the first scenario the inhomogeneous time series contains all observations of the homogeneous one. To obtain homogeneous time series one has to filter observations, which satisfy the following condition:

$$t_j = t_0 + i\Delta t \text{ for } \forall i. \quad (2.3)$$

Second scenario applies when condition (2.3) is not satisfied for at least one  $i$ . In this case there is one or more missing observation. Assuming the missing observation at time  $t_0 + i\Delta t$  is bracketed by two observations of inhomogeneous series:

$$t_j \leq t_0 + i\Delta t < t_{j+1}, \quad (2.4)$$

one can use either linear or previous tick interpolation to fill the gaps.

---

<sup>2</sup>For more details see [55].

*Linear-tick interpolation* fills the gaps between the observations by linearly interpolating two values,  $z_j$  and  $z_{j+1}$ :

$$z(t_0 + i\Delta t) = z_j + \frac{t_0 + i\Delta t - t_j}{t_{j+1} - t_j} (z_{j+1} - z_j). \quad (2.5)$$

Linear-tick interpolation produces series with a "smooth" look, but the need to know the future values for interpolation makes the homogenization process discrete. Additionally, one has to be cautious of synthetic values, as they may not reflect unobserved transition. Panels B and D on Figure 2.1 demonstrate the example of the linear-tick interpolation.

*Previous-tick interpolation* fills the gaps by repeating the last known value until the arrival of a new one<sup>3</sup>:

$$z(t_0 + i\Delta t) = z_j. \quad (2.6)$$

The drawback of previous-tick interpolation is the possibility of spurious jumps in the homogenized data. This, in turn, can affect the analysis of extreme returns. On the other hand, the method is not computationally intensive, and it does not require the values from the future to fill the gaps. This makes it good candidate for the online application. Panels A and C on Figure 2.1 demonstrate the example of the previous-tick interpolation.

In the high-frequency environment with a liquid asset the interpolation of missing values produces visually the same result. Figure 2.1, panel A and B pictures linear and previous-tick interpolation of the price series of the TSE:7723 stock over the time period from 1996/03/11 to 1998/06/22. The target frequency of homogenization is 60 minutes. Visual inspection suggests that time series of prices on both panels are almost identical. To support the conclusion the statistical test is required to check the distribution of prices obtained for different frequencies of homogenization.

Kolmogorov-Smirnov test can provide quantitative base for the analysis.<sup>4</sup> The test was applied over a grid of the homogenization frequencies. The frequencies range from 1 to 60 minutes with an increment step of 5 minutes. The Kolmogorov-Smirnov test compared the distribution of the homogeneous time series of prices obtained by the previous- and linear-tick interpolation from the tick-by-tick data of the TSE:7723 stock. For all frequencies of homogenization the null hypothesis of identical distribution had to be accepted. At the same time, relatively low  $p$ -values suggest slight difference in series.<sup>5</sup> Though the investigation of one stock is hardly sufficient for generalization, one can see that previous-tick and linear interpolation can produce series with similar distributional properties.

An overview of the pertinent literature does not suggest the dominance of either homogenization technique too. The linear interpolation is advocated in [179]. It is considered to be appropriate for a random process with identically and independently distributed increments. [55] suggests that with a high-frequency data statistical studies and model estimation can be alternatively done with both interpolation methods, though one can

<sup>3</sup>The method was proposed by [248].

<sup>4</sup>For details on the Kolmogorov-Smirnov test see Appendix A.2.

<sup>5</sup>Test reported the  $p$ -value of 0.844, and test statistics of 0.5 for all frequencies of homogenization.

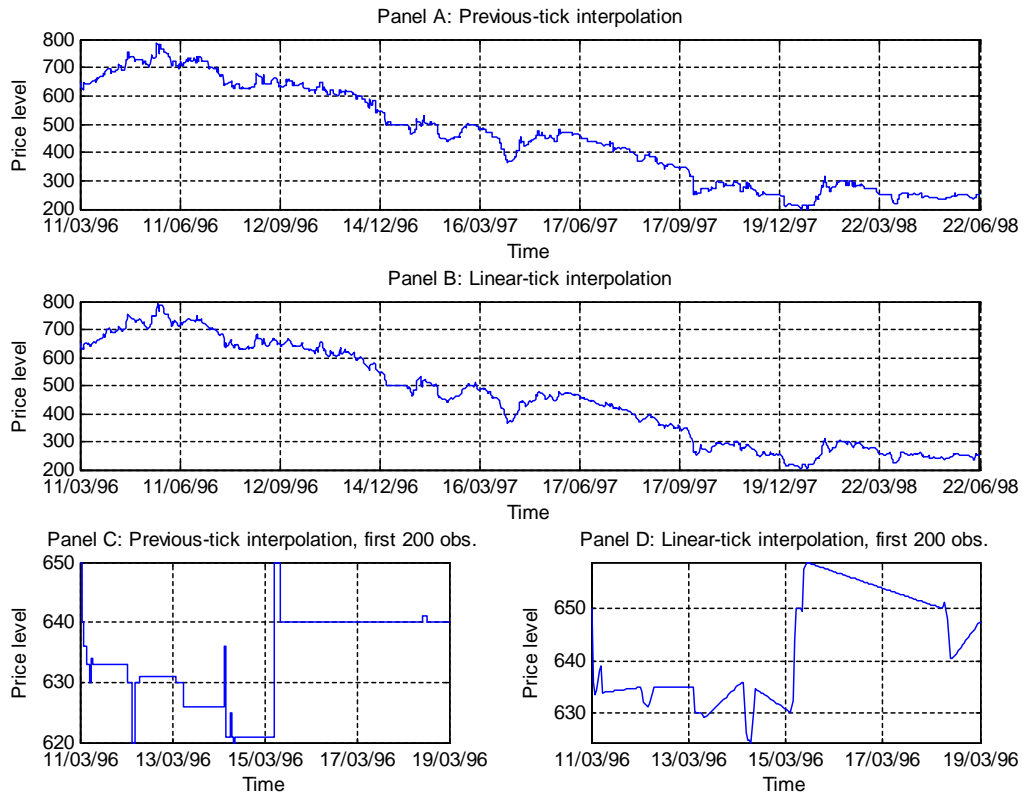


Figure 2.1: Previous-tick vs. linear interpolation

Example on TSE:7723 (see Appendix D on data description). Raw data covers from 1996/03/11 to 1998/06/22. The frequency of homogenization is 60 minutes. Panel C and D is the subsample of first 200 observations from, correspondingly, previous- and linear-tick interpolation.

observe negligible differences. [19] used the Monte Carlo simulation to show that linear interpolation brings bias into realized volatility.

To conclude the analysis there is no pronounced difference between previous-tick or linear interpolation technique. Due to the low computational requirement the choice is in favor of the previous-tick interpolation.

### 2.1.2. Efficiency Issue of Homogenization

The aim of this subsection is to discuss the efficiency issues related to the homogenization of time series. Efficiency in this context means that data preprocessing preserves the maximum of information, i.e. the quantity of data points obtained through homogenization should not be reduced.

One way to do so is to derive several homogenous time series from one inhomogeneous one. Each homogeneous series will have constant time phase shift, relative to its neighbor. Thus, the total number of data points in derived time series can be kept close to the number in the inhomogeneous one.

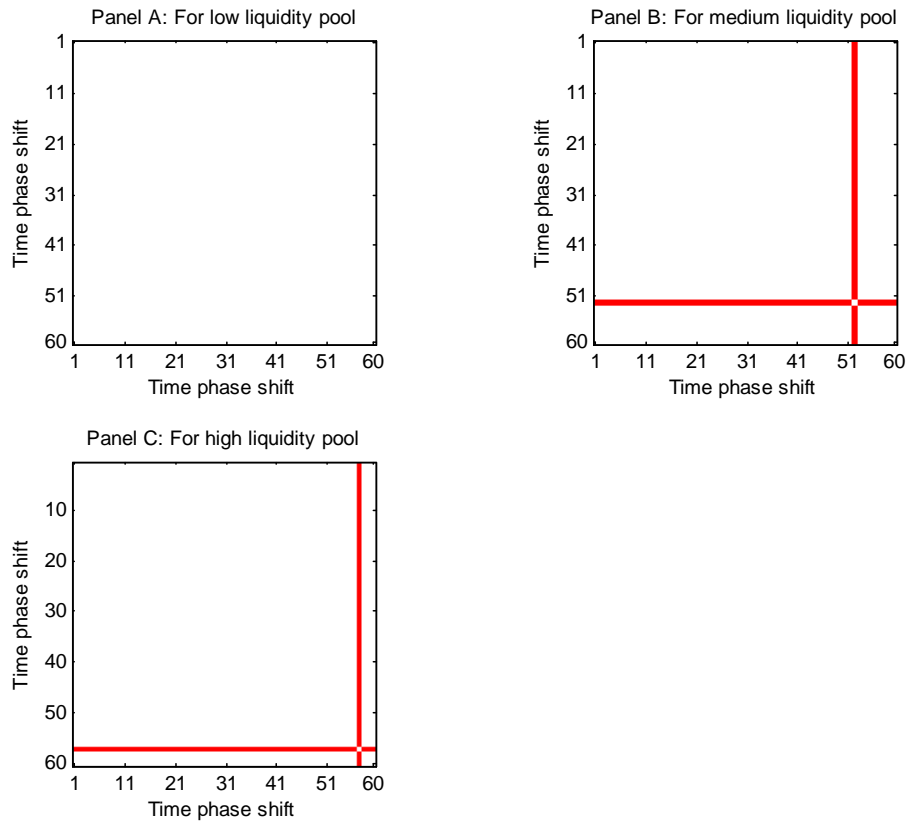


Figure 2.2: Results of Kolmogorov-Smirnov test over the set of returns time series

Panel A is for TSE:6396, panel B is for TSE:7723, and panel C is for TSE:4401. White cells stands for combinations of series (ordinate and abscissa) which passed the test, and red - did not pass.

**Example 1.** *For illustration, assume one has inhomogeneous time series, where the precision of the time stamps is 1 minute. The homogeneous time series should have frequency of observations 60 minutes. Assume the first time stamp in the raw data is 09:00. With either technique homogenization uses only those observations, where time stamps is close to  $9 + 60' \cdot i$ . This leaves unutilized around  $59/60 \approx 98\%$  of raw observations. If one would produce second homogeneous time series, by assuming the first time stamp in the raw data is 09:01, the utilization would be  $2/60 \approx 3\%$ . Obviously, one can produce another 58 homogeneous time series, resulting in 100% utilization or efficiency.*

To prove the legitimacy of this approach one has to show that the returns for each series have the same distribution. To do so, the Kolmogorov-Smirnov test will be applied over the combination of time series pairs. The results will be accumulated in a symmetric matrix, where the cells will assume zeros if the ordinate and abscissa series have identical distributions, and ones otherwise.

Figure 2.2 presents the results. The Kolmogorov-Smirnov test was applied on a set of returns time series. Returns were derived from price series of TSE:6396 (for panel A),

TSE:7723 (for panel B), and TSE:4401 (for panel C).<sup>6</sup> In every case the homogenization produced 60 time series of prices with the frequency of 60 minutes. Each time series has the time phase shift of 1 minute relative to its neighbor. Panel A shows that TSE:6396 has identical distribution of return for all phase shifts. Panel B indicates that TSE:7723 has identical distribution of returns only up to the phase shift of 52 minutes. Similar result is for panel C, where identity for TSE:4401 returns holds only up to the phase shift of 57 minutes.

These results suggest one can increase the efficiency of homogenization by deriving several time series from one inhomogeneous series. At the same time, one has to be careful on how many series can be produced, as the number of identical series in a set is not the same for different levels of liquidity.

### 2.1.3. Looking for Optimal Frequency of Homogenization

The aim of this subsection is to help in picking the optimal frequency of homogenization. With the assumption the choice depends on profitability, one has to search through the possible choices to find the frequency, which maximizes the revealed transaction cost.<sup>7</sup>

Financial time series, used for the analysis, comes in raw, tick form.<sup>8</sup> In case one would like to make use of this data without preprocessing, one has to incorporate in the model both the values and the timing of observations. To avoid that cumbersome task, one employs the preprocessing, which converts inhomogeneous time series into homogeneous one. This allows considerable simplification of the model, but one has to exogenously specify the frequency of homogenization.

One way to find the optimal frequency is to use the grid search over the possible choices. To produce maximum cover three groups of frequencies are chosen:

- Ultra-high frequency: 1, 2, 3, ... 10 minutes,
- High frequency: 10, 20, 30, ... 1440 minutes,<sup>9</sup> and
- Medium frequency: 1, 2, 3, ... 30 days.

In the absence of information on actual transaction cost, the analysis was extended to study the impact of transaction cost on profitability. The study covers transaction cost from 0% to 10% with the increment step of 0.01%.

The profitability was measured as an aggregated return adjusted for transaction costs:

$$R = \prod_{t=1}^T \max \left\{ 1, \frac{p_t - c(p_t + p_{t-1})}{p_{t-1}}, \frac{p_{t-1} - c(p_t + p_{t-1})}{p_t} \right\}, \quad (2.7)$$

<sup>6</sup>Those stocks are medians in the one of three groups of stocks, which were sorted based on liquidity. TSE:6396 is median for low liquidity stocks, TSE:7723 is median for medium liquidity, and TSE:4401 - for high liquidity. The choice of this mechanism of selecting stocks allows to avoid problem of biased data-mining.

For data description see Appendix D.

<sup>7</sup>For relation of revealed transaction cost and profitability see Chapter C.

<sup>8</sup>For data description see Appendix D.

<sup>9</sup>1440 minutes is equivalent to 1 day.



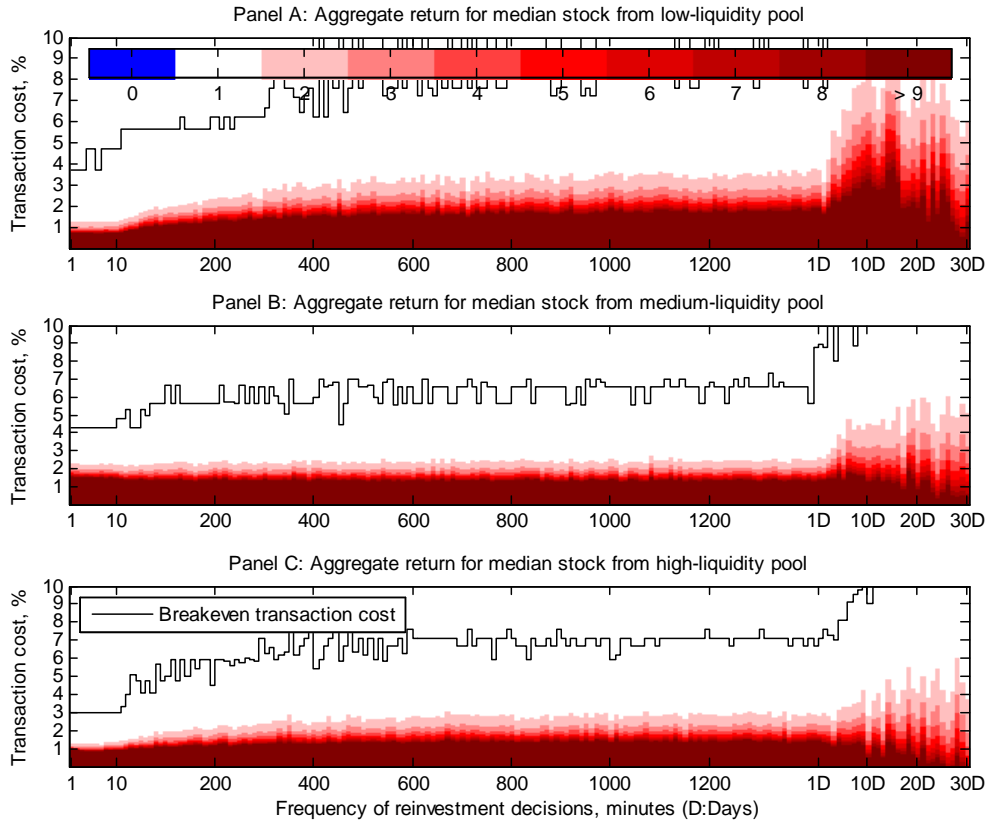


Figure 2.3: Profitability analysis with respect to the frequency of homogenization and TC

where  $p_t$  is the level of price at time  $t$ ,  $c$  is transaction cost. Equation (2.7) assumes possibility of short selling. At the end of each period short or long positions should be cleared. In case transaction in current period cannot provide sufficient return to cover transaction costs no position is taken. Dividends, inflation, taxes and impact on prices are assumed negligible.

Figure 2.3 exhibits the results. Analysis was performed for three time series: TSE:6396 (panel A), TSE:7723 (panel B), and TSE:4401 (panel C). Aggregated return is presented as specter: from the lowest return (in cold color) to the highest return (in worm color). For all levels of liquidity the curve of breakeven transaction costs (BTC) at different frequencies has well pronounced linear parts, which correspond to the defined three sub-intervals of frequencies. Estimated average values and standard deviations of TC and slope coefficients of BTC curves are summarized in Table 2.1. BTC are increasing with a decrease in frequency. This is supported by an increase in average values, as well as, the positive slope coefficients. The implication is that if investor's TC are above the breakeven curve the rational investor should adjust the frequency of trading up to a point where TC are at or below the breakeven level. A similar principle can be applied if an investor targets for specific rates of return (on graphs they are denoted by borders between color shades). Since frontiers of these returns vary with frequency, one can target for selected return by adjusting frequency of trades.

Frequency	Ultra-high		High		Low	
	Aver. TC ( $\sigma_M$ )	TC Slope (p-value)	Aver. TC ( $\sigma_M$ )	TC Slope (p-value)	Aver. TC ( $\sigma_M$ )	TC Slope (p-value)
Median stock from						
low-liquidity pool	4.30 (0.16)	0.12 (0.02)	8.29 (0.54)	0.03 (0.00)	9.92 (0.14)	0.02 (0.09)
medium-liquidity pool	4.37 (0.05)	0.03 (0.12)	6.26 (0.22)	0.01 (0.00)	9.86 (0.14)	0.02 (0.01)
high-liquidity pool	3.02 (0.00)	0.00 (0.00)	6.57 (0.28)	0.02 (0.00)	9.52 (0.29)	0.08 (0.00)

Table 2.1: Averages and slope estimates of BTC for aggregate return with perfect forecast

That is for the same level of TC, say 3% and medium liquidity, at the frequency above 1 day the aggregate return is less than 2, while for low frequency it takes values of 2, 3 and sometimes 4 (see Figure 2.3).

In addition to profitability description, Figure 2.3 reveals three important facts. First, given the method of returns accumulation, one cannot increase the profit beyond that reported on a graph for a given transaction cost. Equation (2.7) works with historical data. In online environment it is equivalent to have 100% accurate forecast of the next period price. Any forecast mechanism have or will have at maximum 100% accuracy. Thus, the aggregate return adjusted for transaction cost cannot be higher for reported stocks over the covered time span.<sup>10</sup> Second, the market maker can clearly see at which transaction cost the investors will be attracted by low or high aggregated return.<sup>11</sup> Third, the data clearly indicates the increase of high aggregated returns with increase of the frequency. The implication is to pick the highest trading frequency at which actual transaction costs are below BTC. Figure 2.3 indicates that for the actual transaction costs below 4% it is optimal to pick the highest possible trading frequency. Given that in real world transactions take some time, one would pick the frequency which is a bit higher than average transaction time.

#### 2.1.4. Looking for the Optimal Subsample Size

This subsection describes the procedure to select the optimal subsample size for application with the ATA. The necessity to define the optimal subsample size comes from the nature of technical analysis and the ATA autonomy. Technical analysis processes recent subsample of observations to form next period price expectations. Normally there is no issue with the optimal subsample size, as the rules of technical analysis are provided *a priori* and used under human supervision. In case of the ATA the rules of technical analysis are generated endogenously in autonomous fashion. With the undefined size of subsample there could be a

<sup>10</sup>This study does not claim that all stocks from low, medium or high liquidity will have exactly the same pattern. At the same time, the choice of using median stocks in each group suggests that these three stocks are quite representative.

<sup>11</sup>One has to take into account that these price series were produced under definite transactions costs. Once transaction cost changes this picture will change as well.

case when the ATA cannot apply the derived rule due to insufficient number of observations. Thus, the subsample size should be defined and fixed.

The selection of the optimal subsample size should take into account two points. First, too high subsample size can slow down the computation time of the ATA, which is crucial in the ultra-high frequency environment. With the availability of powerful computational resources the importance of this argument can be not so pronounced. At the same time, the possibility of high subsample size can be limited by the microstructure of the market where the ATA is intended to operate. Second, too small subsample size can reduce the informational content of subsample, which will result in a low quality of forecasts. Thus, the selection of the optimal subsample should go along the maximization of the informational content of the subsample within the maximum possible subsample size.

One way to find the optimal subsample size is to use a grid search within feasible limits. An upper limit will be defined by the microstructure of the market, while the lower limit is the minimum number of observations required for a fault-free operation of the ATA. The microstructure of the TSE gives roughly 350 minutes of two-session continuous trading in one day [244]. The available tick data has the accuracy of 1 minute. This gives at maximum 350 trading opportunities per day.<sup>12</sup> Say  $N$  first observations are in the subsample for the forecast, then the rest,  $350 - N$ , observations are available for the trading. To keep it at realistic level there should be at least 100 trading opportunities per day, which makes the upper limit for grid search to be 250 observations. The ATA mechanism requires at least 50 observations for the fault-free operation, making the lower limit to be 50 observations. Given the computational intensity the pitch in the grid search is selected to be 10 observations.

Figure 2.4 shows the results of the grid search. As in the previous subsection three median stocks representing three liquidity pools were used in the analysis. To obtain each graph the following procedure was employed. For the homogenization the domain of the grid search was defined to be from 1 to 10 minutes with an increment of one minute. For the subsample size the domain was defined to be from 50 to 250 observations with an increment of 10 observations. For each combination of the frequency and the subsample size the ATA performed sequential forecast of the next period prices. The forecast results were accumulated for every time series of the considered median stocks. In the next step a schedule of transaction costs was used to calculate the aggregate return. The schedule ranges from 0 to 2% with an increment of 0.001%. Returns were calculated according to (2.7). Next, starting from the smallest TC the first breakeven TC has been determined for every combination of the frequency and the subsample size. Resulting breakeven TC were drawn as a separation surface on Figure 2.4, in the first column of the panels. Every combination of TC, the frequency and the subsample size, which is below the separation surface, indicates past profitable opportunities, while every combination above the separation surface gives breakeven or losing scenarios. The second and third columns of the

---

<sup>12</sup>The assumption is the information from the previous days has little or no influence on a next day prices.

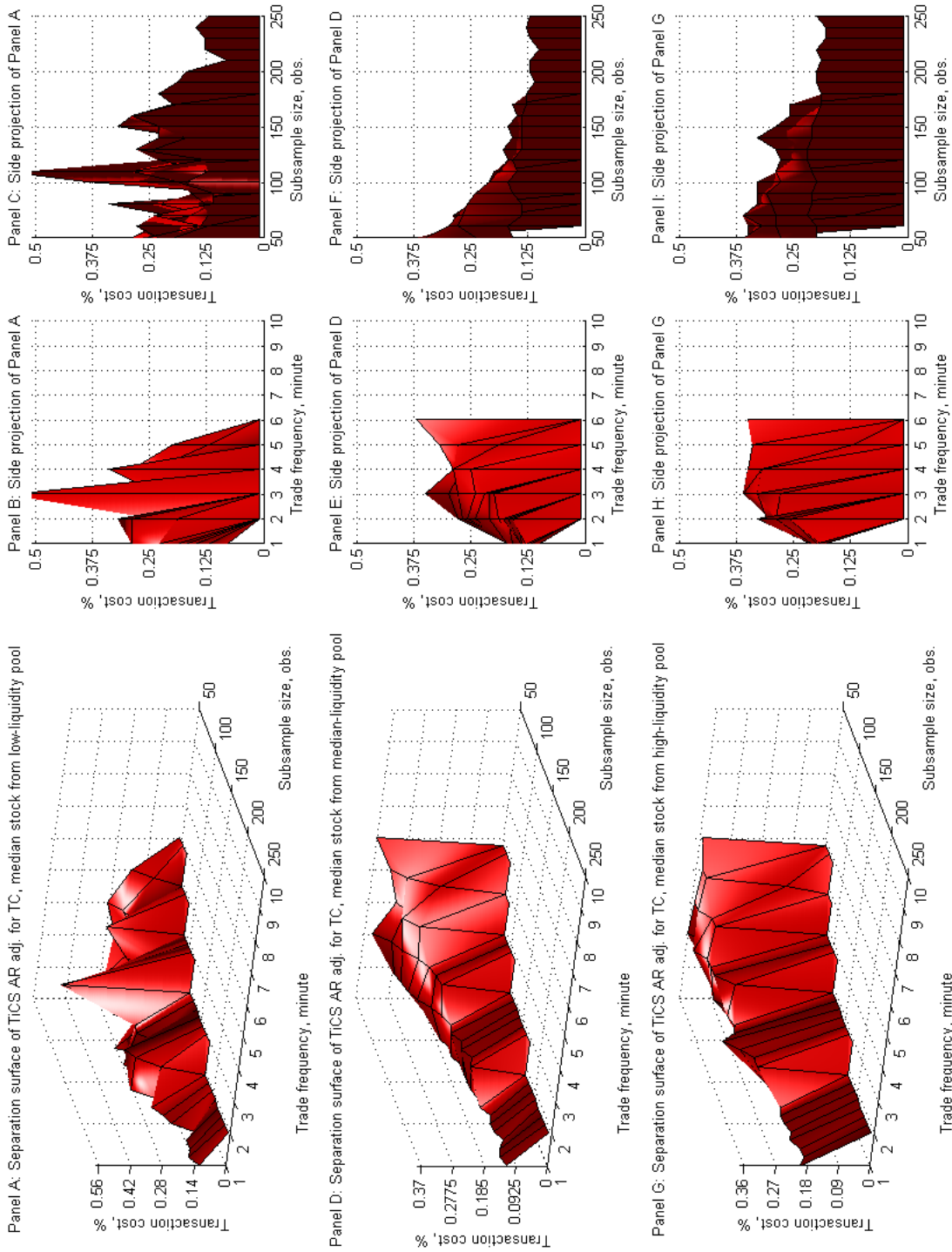


Figure 2.4: Looking for the optimal subsample size

Median stock for	Value of breakeven TC, %	Optimal subsample size, obs.
low-liquidity pool	0.56	110 ( at 3 minute)
medium-liquidity pool	0.37	50 (at 6 minute)
high-liquidity pool	0.36	50 (at 3 minute)

Table 2.2: Optimal subsample size

panels show the projections from the side of the subsample size. Graphs clearly indicates the optimal subsample size for each case of liquidity.

Table 2.2 summarizes values of breakeven TC and the optimal subsample size. Analysis indicates the optimal subsample size is 110 observations for low-liquidity pool, and 50 observations for other liquidity. In addition, the optimal recorded frequency of trades is 3 minute. One has to take into account the local nature of results. Observed values can differ from one stock to another, though the closeness of results and robust selection of sample data indicates observed persistence.

### 2.1.5. Discussion

This section deals with the preprocessing of the data. With the data coming in raw, tick-by-tick form one has to apply some homogenization procedure to reduce the data and the problem complexity. There are two methods of homogenization. That is previous- and linear-tick interpolation. Empirical analysis on one of the median stocks showed no pronounced difference in the distribution of price time series obtained by either method. The Kolmogorov-Smirnov test did not indicate significant difference in price distribution for all frequencies of homogenization. Due to simplicity and fast computability the previous-tick interpolation was selected.

The drawback of homogenization is inefficient use of data. High-frequency data has much more observations than it is required for some fixed frequency of homogenization. At the same time the ATA needs as much data as possible for fast learning. One way to "feed" the ATA and to improve the efficiency of data processing is to use sequential root time shift in the homogenization procedure. To see within how many shift the distribution of derived time series are identical the Kolmogorov-Smirnov test was applied again. Results showed the root time shift is eligible up to 50 minutes for an hour frequency of homogenization. This means the ATA can have at least 50 similar time series for training.

One important aspect of homogenization is a necessity to exogenously specify the optimal frequency. The criterion of optimality is profitability. To get the sign and the magnitude of the profit one has to provide transaction costs. With no knowledge of TC the notion of revealed transaction costs can shed light on hypothetical profits. Aggregate returns adjusted for a schedule of TC, a direct indicator of profitability, were computed for a grid of frequencies of homogenization. The median stocks, selected for the analysis, showed identical pattern. With decreasing TC it is rational to increase the frequency of trades, and thus, the frequency of homogenization. At the ultra-high frequency the breakeven TC

were around 4%. This indicated the optimal frequency of homogenization to be around 1 minute or so.

Another related aspect of data preprocessing is selection of the optimal subsample size. For consistent application of technical trading rules the ATA should have defined and fixed subsample of recent observations. Like in the previous case the grid analysis and revealed TC were used for the search. The results indicate the optimal subsample size of 110 observation for low-liquidity stocks, and 50 observation for other liquidity. Those were observed at 3 and 6 minutes frequency of trades.

To conclude, there are two things worth of mentioning. First, the empirical examples, used for the analysis or demonstration, show results for selected stock only. Though one can argue that the procedure to select the stocks, i.e. pick the median for each group of liquidity pool, can insure some persistence across similar in liquidity stocks. And second, the optimal value of the subsample size, or the frequency of homogenization can vary from one stock to another, but wrong selection of these parameters can destine the ATA performance to be very low from the very beginning. Thus, one has to apply the procedures, described in this section, to insure high data potential.

## 2.2. Piecewise Linear Approximation

Looking for regularities in the data, one has to use some data mining technic. Data mining, also known as knowledge-discovery in databases, is the practice of automatic searching for patterns in large stores of data. To do this, data mining uses computational techniques from statistics and pattern recognition. With the pattern recognition the important ingredient of the data mining is a Piecewise Linear Approximation. Specifically, for data mining, the piecewise linear representation has been used to:

- support fast exact similarity search [124],
- support novel distance measure for time series, including dynamic time warping [196] and relevance feedback [126],
- support novel clustering and classification algorithms [125],
- support change point detection [216, 89].

Here the piecewise linear approximation is used to get the maximum representation of original time series by several linear segments. Assuming the ends of linear segments are connected, one can use those joints as identification points of the underlying time series. Based on these identification points, the classifier systems (for details see Chapter 2.3) perform similarity search and subsequent forecast of the next observation in the time series.

The formal definition of the *piecewise linear approximation* is an approximation of a time series of the length  $n$  by  $k$  linear segments. Normally, one chooses  $k$  to be much smaller than  $n$ , which makes the storage, transmission and computation of the data more efficient [123]. The abbreviation of the piecewise linear approximation is PLA. One can refer to it as *segmentation algorithm* or *approach*.

This section reviews four major segmentation approaches known in the literature, including the Genetic-Algorithm-based approach.

### 2.2.1. Classification

Though the segmentation algorithms play an important role in datamining, they are almost not classified or documented. Often one can find the same algorithm under different names in different fields and publications. Their low publicity forced researchers to rediscover the same algorithm several times. One of the first attempts to classify and compare different algorithms was undertaken by [123]. Following [123] one can group the time series segmentation approaches into the following categories:

- *Sliding Windows*: A segment is growing until it exceeds some error bound. The process repeats itself with the next data point not included in the newly approximated segment.
- *Top-Down*: Time series is recursively partitioned until some stopping criteria.
- *Bottom-Up*: Starting from the finest possible approximation, the segments are merged until some stopping criteria.

The advance in Genetic Algorithms and accessibility of computational power allowed to incorporate the fourth category:

- *GA-Based*: One segments the time series based on the breaking points found by a Genetic Algorithm.

One can also classify the algorithms in *online* and *batch* algorithms. Intuitively, the online algorithms insure dynamics by one-flow processing of data. The batch algorithms often work iteratively. Thus, the computational time can vary dramatically. Normally, the batch algorithms are superior in quality of linearization. With increasing availability of powerful computational resources the processing time of batch algorithms shortens, making them present day favorites.

One can use either *linear interpolation* or *linear regression* to approximate the time series by several linear segments. Linear interpolation approximates the part of original time series by a straight line connecting the first and the last data point in the part. Linear regression employs the least square approach, which for Euclidian distance guaranties superior fit. The drawback of the regression approach is the close ends of linearized segments can be disjoin. The threat of discontinuity makes the linear interpolation approach a reasonable choice. In addition, one can get the interpolation in a constant time, while regression time is linearly increasing with the segment's length [123].

The important part of classification is the quality of approximation. Goodness of fit allows to classify the algorithm based on their performance of approximation. One can use the sum of squares of the residual errors, or use  $R^2$ . Another commonly used measure of goodness of fit is the distance between the best fit line and the data point furthest

away in the vertical direction. The prospect of online application justifies to measure the performance as the elapsed time.

The classification presented in this section does not pretend to be the most complete. One can use more detailed classification and include more exotic algorithms.

### 2.2.2. Sliding Window Algorithm

The first algorithm to be analyzed is the Sliding Window algorithm. It belongs to the family of online algorithms. The Sliding Window algorithm works by linearly linking the first data point in the current segment with the consequent ones. Every time the goodness of fit of the current linear segment is evaluated in comparison with the predetermined benchmark. The benchmark is target goodness of fit, which, for example, can be measured as  $R^2$ . Once the goodness of fit reaches the value of the benchmark, the segment stops growing. Next data point, not included in the previous segment, is counted as the first one, and the process repeats.

Figure 2.5 displays the steps of segmentation. Each panel presents one iteration. Here, after five iterations the original time series is approximated by 5 linear segments. The measure of the goodness of fit is the sum of squared errors.<sup>13</sup>

The strong points of this algorithm are simplicity and online performance. One can gain a higher speed of segmentation by using the step of increment greater than one. Each segment's length depends on the threshold goodness of fit. Thus the produced approximation can be highly close to the original time series, though the number of linear parts can be numerous.

The weak point of the Sliding Window algorithm is its myopia to find optimal partitioning. The application of predetermined benchmark constrains the algorithm so the goodness of fit of each segment is slightly smaller than the value of the benchmark. Imagine the time series originally has two adjoint linear parts. The algorithm would not stop enlarging the first segment at the original kink point, but would continue until the goodness of fit of the first part reaches the benchmark value.

There is another weak feature of the algorithm. One cannot constrain it to produce fixed number of linear segments. That means that visually similar time series can have approximations with different number of linear segments. Thus, one cannot directly compare several time series approximated by the Sliding Window algorithm. To overcome this limitation, an iterative procedure can be used. When one would like to have a fixed number of linear segments, the dichotomy algorithm can find the best benchmark goodness of fit.

To start, one needs to introduce the *dichotomy algorithm*. Here the principle of dichotomy is used to obtain recursive solution of continuously increasing monotonic function. Assuming the function:

$$y = f(x), \tag{2.8}$$

<sup>13</sup>Here and in the following illustrations of piecewise linear approximation algorithms the goal is to obtain an approximation with the given number of linear segments.



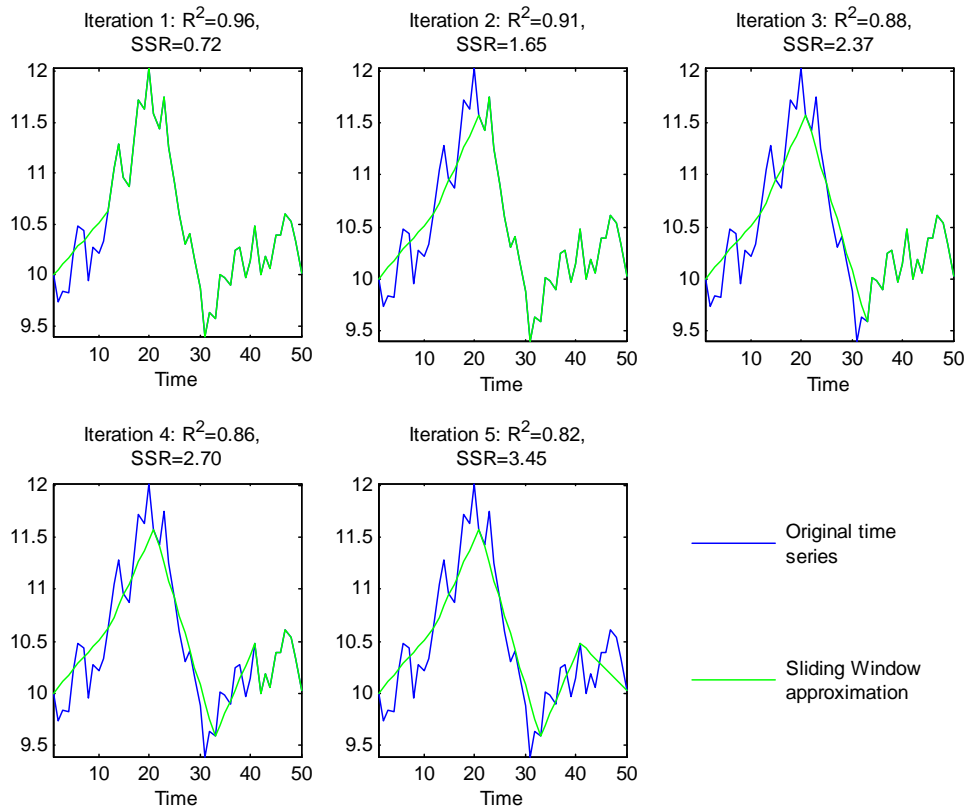


Figure 2.5: Illustration of Sliding Window algorithm

The length of time series is 50. Approximation is with 5 linear segments.

where  $x$  is an argument, the recursive solution has the form:

$$x = f^{-1}(y). \quad (2.9)$$

Given, one does not know the functional form, the analytical solution is not available. Thus, one would look for numerical solution of (2.9) for  $y = y^*$ . Because of the continuity and monotonicity of (2.8) the dichotomy algorithm gets the solution in the most efficient way. The search of solution starts with defining the most wide search bounds  $[\underline{x}, \bar{x}]$ . Next, one randomly picks from interval  $[\underline{x}, \bar{x}]$  a trial solution  $x_1$ . It corresponds to  $y_1 = f(x_1)$ , or in general case, to  $y_i = f(x_i)$ . In the next and all consecutive steps the relation of  $y_i$  with  $y^*$  defines the direction of the search. That is if  $y_i > y^*$  then:

$$\bar{x} = x_i, \quad (2.10)$$

$$x_{i+1} = \frac{x_i + \underline{x}}{2}. \quad (2.11)$$

And if  $y_i \leq y^*$  then:

$$\underline{x} = x_i, \quad (2.12)$$

$$x_{i+1} = \frac{x_i + \bar{x}}{2}. \quad (2.13)$$

After the redefinition of the search bounds and next trial solution the procedure is repeated.

Equations (2.11) and (2.13) explains the name of the algorithm. Each time when the algorithm over- or undershoots, the next trial value is on a halfway to upper bound,  $\underline{x}$ , or lower one,  $\bar{x}$ . The algorithm converges when the maximum distance between two consecutive trial solutions is less then some predefined threshold of error,  $\varepsilon$ :

$$\varepsilon < |x_{i+1} - x_i|. \quad (2.14)$$

**Example 2.** *To better understand the mechanism one can use the following example. A subsample of 50 time series observations has to be approximated with 5 linear segments. The goodness of fit,  $R^2$ , is selected as a benchmark. Since  $R^2$  is defined between 0 and 1, these values are used for initial search bounds, i.e.  $[0, 1]$ . A random drawing from this interval gives, say, 0.56, which is used as the benchmark value at the first iteration with the Sliding Window algorithm. The resulting number of linear segments is 3. Since 3 is less than the target value 5, the next period trial solution will be defined according to equation (2.13), while the lower search bound - according to equation (2.12). Thus, for the next iteration the benchmark value is  $(0.56 + 1) / 2 = 0.78$ , and the search interval is  $(0.56, 1]$ . Evaluation of 0.78 results in 6 linear segments. Now one should use equations (2.11) and (2.10), which give  $(0.78 + 0.56) / 2 = 0.67$  and  $(0.56, 0.78)$ . New evaluation of 0.67 results in 5 linear segments. One can stop at this stage or proceed if more precise benchmark value is required.*

Figure 2.6 sketches the work of the dichotomy algorithm. The threshold error is  $\varepsilon = 10^{-5}$ . That is 5 digits accuracy after comma. Here  $x_i$  is a sum of squared errors, and  $y^* = 5$  is the target number of linear segments. The dichotomy algorithm gets 5-segment approximation in 10 iterations with the sum of squared errors equal to 139.23, and  $R^2 = 0.81$ .

One can characterize the time performance of the algorithm as follows. The computational time is increasing with the number of observations in the time series,  $n$ . The increase of the threshold error,  $\varepsilon$ , decreases the computational time.

### 2.2.3. Top-Down Algorithm

Unlike the Sliding Window algorithm the Top-Down approach requires a complete time series for approximation. On the one hand it makes it difficult to apply the Top-Down approach in online context, but, on the other hand, it allows to predefine either the target goodness of fit of approximation, or the number of linear segments.

In the Top-Down algorithm the first approximation is one line, which connects the first and the last point in the original time series. To find better approximation one evaluates every possible partitioning of the previous approximation. The partitioning which provides the best goodness of fit locates next split. Newly segmented approximation is reexamined for a new partitioning and the process repeats. The algorithm runs until either the benchmark goodness of fit is reached, or an approximation gets enough linear segments. Figure 2.7

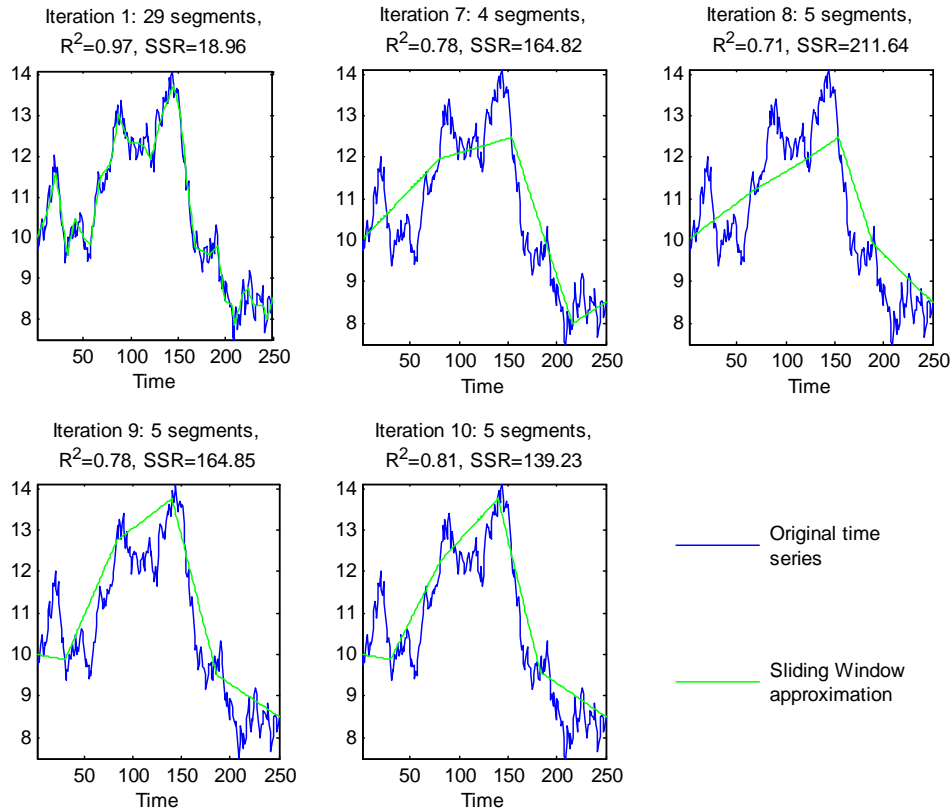


Figure 2.6: Illustration of iterative procedure with the Sliding Window algorithm

The dichotomy algorithm is applied to obtain segmentation with 5 linear segments and the threshold error  $\varepsilon = 10^{-5}$ . Time series has 250 observations.

shows the process of the Top-Down approximation. Here the stopping criterion is the creation of 5-segment approximation. It is performed in 5 iterations. The goodness of fit is the sum of squared errors equal to 231.76, and  $R^2 = 0.69$ .

Because of the design of the algorithm the computational time is increasing with the number of observations in the time series,  $n$ . But, the increase of the target number of linear segments,  $k$ , makes the computation faster. Thus,  $k$ -segment approximation of  $n$ -length time series takes only  $k$  iterations, or  $(k - 1)n - \frac{1}{2}k(2 + k)$  evaluations of goodness of fit.

#### 2.2.4. Bottom-Up Algorithm

Similar in the initial setting is the Bottom-Up algorithm. It is a batch type, that is it requires the complete time series. It can produce either the approximation with the predefined goodness of fit, or with the desired number of linear segments.

In the first step one creates the finest approximation of  $n - 1$  segments from the original  $n$ -points time series. Next, the goodness of fit of merging each neighboring segments is evaluated. The pair with the best goodness of fit merges. With a new segment one repeats

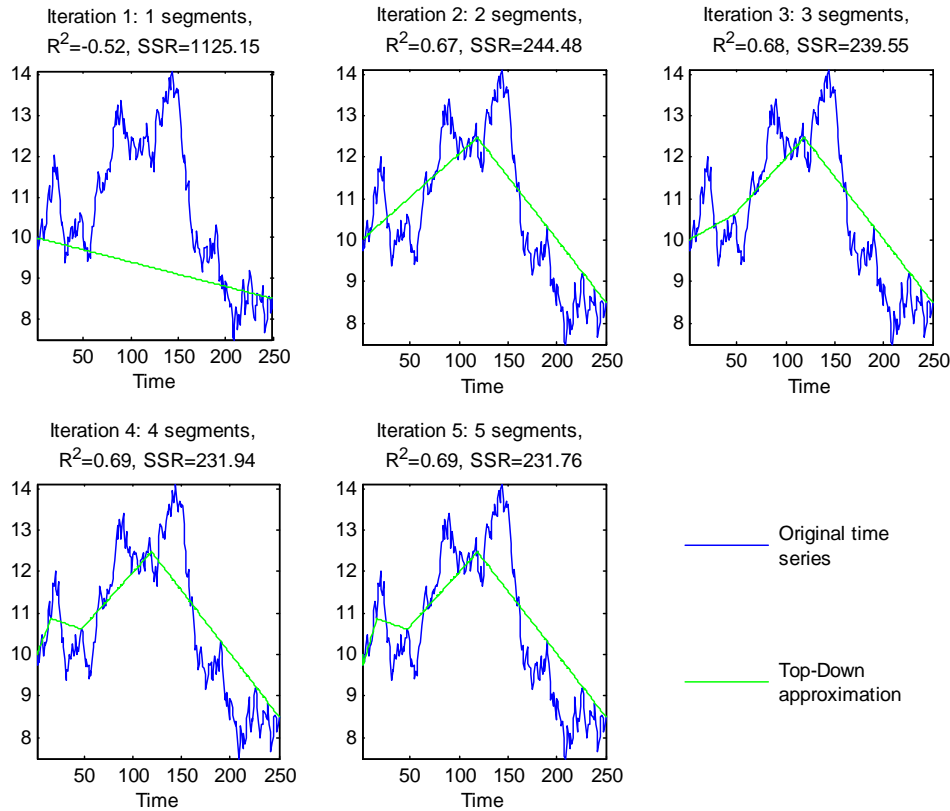


Figure 2.7: Illustration of Top-Down algorithm

Time series is of length 250. Approximation is with 5 linear segments.

with evaluating the goodness of fit and merging the best pair. To stop one should reach the benchmark goodness of fit, or get enough linear segments.

Figure 2.8 presents iterations of the Bottom-Up algorithm. The goodness of fit is the sum of squared errors. The algorithm takes 245 iteration to produce 5-segment approximation, where the sum of squared errors is 132.34, and  $R^2$  is 0.82. In general it takes  $n - k$  iterations to segment  $n$ -length time series with  $k$  linear parts, or  $\frac{1}{2}(n - k)(n + k - 1)$  evaluations of goodness of fit.

### 2.2.5. GA-Based Algorithm

The fourth algorithm is the Genetic-Algorithm-based approach. It does not differ in the initial settings from the Top-Down or the Bottom-Up algorithm. It is a batch type. The advantage of Genetic-Algorithm-based approach is the possibility to find the best possible approximation given the required goodness of fit, or the desired number of linear parts.<sup>14</sup>

To apply GA for the piecewise linear approximation one needs to design the form of a solution. There are two hints for doing it. First, the original time series is in homogeneous form. That is the time elapsed between neighbor data points is a constant. Second, one

<sup>14</sup>For details on GAs see Appendix B.

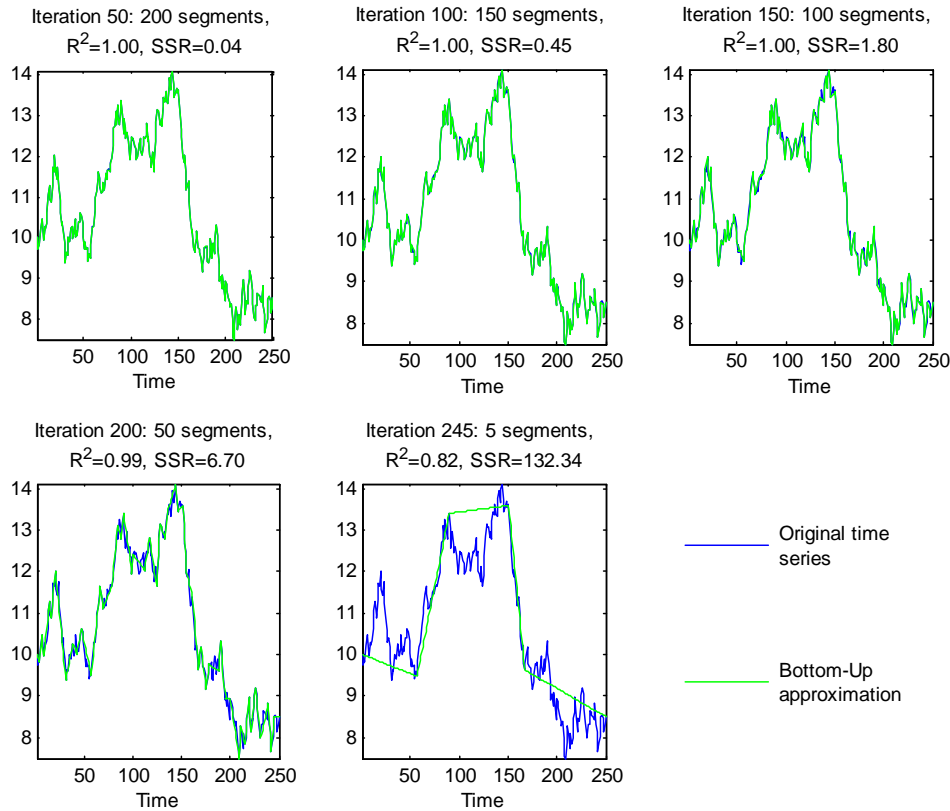


Figure 2.8: Illustration of Bottom-Up algorithm

Time series is of length 250. Approximation is with 5 linear segments.

would use linear interpolation, to avoid disjoints of linear parts. This implies the solution should outline only the break points. Thus, one form of the solution can be a string of zeros and ones. The string's length should be the same as the number of observations in the original time series. The ones would mark the break points, and the zeros - the linear part. The number of ones in each string should be kept constant and equal to  $k + 1$ , where  $k$  is the target number of linear segments<sup>15</sup>.

As GA are a maximization technique, one needs to define the objective function. It can be any suitable measure of the goodness of fit. One uses the value of the function to find the best solution. Figure 2.9 displays the iterations of GA in producing 5-segment approximation. The objective function is the sum of squared errors<sup>16</sup>. The stopping criteria are the threshold error,  $\varepsilon = 10^{-5}$ , and the maximum number of iterations, 200. The GA-based algorithm produced 5-segment approximation with the sum of squared errors equal to 77.15, and  $R^2 = 0.90$ .

<sup>15</sup>In case one would like to produce maximum goodness of fit irrespective of the number of linear segments, one should relax this constraint. At the same time, the objective function should incorporate negative influence of the linear segments' number. Otherwise the maximization would produce the approximation which is identical with the original time series.

<sup>16</sup>Naturally, one would like to get the approximation with minimized sum of squared errors. That is why the objective function takes the sum with negative sign.

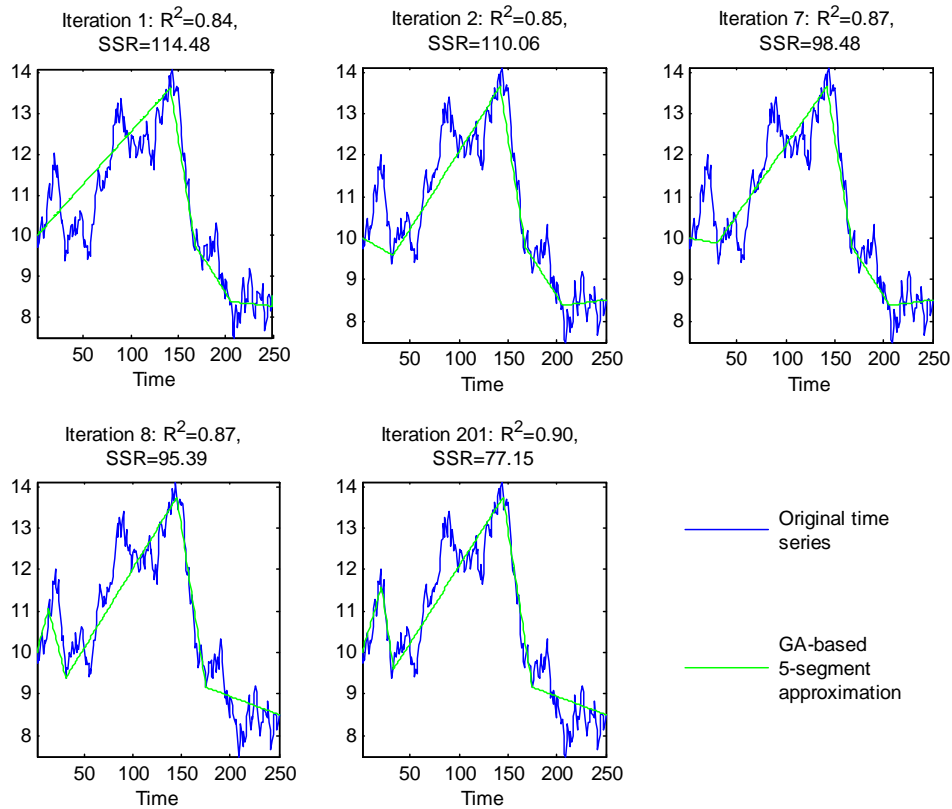


Figure 2.9: Illustration of GA-based algorithm

Time series is of length 250. Approximation is with 5 linear segments.

The drawback of the GA-based approach is its stochastic nature. Each time the algorithm starts with a random population. This makes the probability of finding the optimal segmentation conditional on the initial realization. With the increasing number of iterations that probability converges to 1.

### 2.2.6. Monte-Carlo Simulation: Analysis of PLA Algorithms' Performance

Attentive reader would notice that in the previous section two measures of fit accompany each illustration, SSE and  $R^2$ . Both of them indicate the goodness of fit for all approaches. The preliminary analysis suggests that with a time series of length 250 the best approximation is obtained by GA-based algorithm, then by Bottom-Up, by Sliding Window, and, finally, by Top-Down algorithm (see Table 2.3). The ranking is the same for both indicators.

To draw robust conclusion, one would study those algorithms under several settings. In particular, the size of time series and the target number of linear segments could influence the performance of the algorithms. The sum of squared errors (SSE), and  $R^2$  will measure the performance. Conditioned on possible online application of the algorithms, one would add the time measure. It should show how fast the algorithm performs the segmentation relative to its rivals. Of course, one should compare the timing under the same settings.

In the literature the comparison of the algorithms' performance was performed using empirical time series. For example, [123] used 7 different sources of time series, including the data from space shuttle telemetry, stock markets, and noisy periodic functions. Here, the analysis makes one more step towards specificity. The study is focused on the performance of the algorithms in the stock market environment. Namely one is interested to test the algorithms on high frequency stock prices.

Following the random walk hypothesis, the stock market prices can be best described as:

$$p_{t+1} = p_t + \varepsilon_t, \quad (2.15)$$

where  $p_t$  is a current price,  $p_{t+1}$  is the next period price, and  $\varepsilon_t$  is a stochastic component. One can use equation (2.15) to produce artificial data with the "idealistic" properties of high frequency stock prices. With high frequency data one can expect high returns with the same probability as low ones. To capture this effect the distributional assumption of stochastic component is the uniform distribution. With this data generating process one can produce as many and as long time series, as one would need for the analysis.

The backbone of the analysis is the Monte Carlo Method (for details see Appendix A.1). One used the Monte Carlo Method to reveal important statistical properties of the algorithms in the stochastic environment. Here, the stochastic environment consist of an artificially created high frequency stock prices. After evaluating the statistical properties one can use them for the qualitative analysis of the segmentation algorithms. The goal of the analysis is to select the best segmentation algorithm for its further application in the model of the artificial technical analyst.

Table 2.4 specifies the Monte Carlo Method. Each algorithm should be tested against each series' length. Each length of the time series should be analyzed for each number of linear segments. For each specification there should be  $10^5$  repetitions. In each repetition the first observation of the artificially generated price series is 10, and the consequent increments have uniform distribution within interval  $[-0.5, +0.5]$ .

Appendix E presents the graphical results of the Monte Carlo studies. Generally each figure displays the statistical properties of the algorithms either in different settings, or in a comparison with one another. Panel A illustrates the sample time series (in blue) and the examples of linear approximation, distinguished by colors. Panels B, C, and D presents the empirical cumulative distribution functions (CDF) for SSE,  $R^2$ , and the time required for approximation (TRA), correspondingly. The color of CDF and the example linear

##	Algorithm	SSE	Rank based on SSE	$R^2$	Rank based on $R^2$
1.	Sliding Window	139.23	3	0.81	3
2.	Top-Down	231.76	4	0.69	4
3.	Bottom-Up	132.34	2	0.82	2
4.	GA-based	77.15	1 (best)	0.90	1

Table 2.3: Preliminary ranking of PLA algorithms.

##	Parameters	Values
1.	Algorithms to be tested	A. GA-based algorithm B. Sliding Window algorithm C. Top-Down algorithm D. Bottom-Up algorithm
2.	Number of repetitions	$10^5$
3.	Length of time series	10, 50, 250
4.	Number of linear segments	3, 4, 5, 6
5.	Price formula	$p_{t+1} = p_t + \varepsilon_t$
6.	$p_1$	10
7.	$\varepsilon_t$	$\sim U[-0.5, +0.5]$

Table 2.4: PLA specification of the Monte Carlo method

approximations corresponds for each panel. So, one can easily trace the approximation, say distinguished in green, through all panels in a figure.

While panel A only graphically illustrates the segmentation abilities of the algorithms, panels with CDF provides their qualitative properties. Naturally, one would prefer the algorithm, which provides on average smaller SSE and TRA, and higher  $R^2$ . Thus, the CDF curve for SSE and TRA should be maximally on the left, and the CDF curve for  $R^2$  should be maximally on the right. Normally the panels with CDF contain several overlay graphs. The graphs are either for the same algorithm, but with different settings, or for several algorithm under one setting. One can use the relative position of CDF curves for qualitative ordering of the algorithms. Depending on the intersection of the CDF curves one can apply either by first-order, or by second-order stochastic dominance (see Appendix A.1 for definitions) to order algorithms. In most of the cases the CDF curves are not intersecting, which suggests the applicability of the first-order stochastic dominance. That is for SSE and TRA the CDF curves of the better algorithm should be on the left, and for  $R^2$  - on the right.

First 12 figures (from E.1 to E.12) presents the overview of distributions properties of the algorithms in different settings. For each algorithm there are 3 figures, which differ by the size of the underlying time series: 10, 50, and 250 data-points. On panel A the approximations for 3, 4, 5, and 6 linear segments overlay the sample time series. Their corresponding CDF are presented on panels B, C, and D. For all algorithms there is a strong continuity. This means that with increasing number of segments the quality of approximation continuously increases. Graphically, the CDF curves of SSE and TRA are ordered from the left (the best) to the right. The CDF curves for  $R^2$  are ordered from the right (the best) to the left. With the Top-Down algorithm for 50 and 250-points time series the incremental improvement of SSE and  $R^2$  is negligible, while the increment of the time of approximation is still significant.

Figures E.13-E.16 shows the time performance of the algorithms. The figures differ by the number of linear parts: from 3 to 6 segments. In each figure there is one panel



for each algorithm, where the CDF curves of TRA overlay. The reported measurement is in second. Depending on the computational power one can get different timing<sup>17</sup>. Thus, one would consider the relative speed rather than the absolute value. The computational time is continuously increasing with the increasing number of linear segments. Though for the Sliding Window algorithm the time increment is relatively high. Table 2.5 summarizes the ranking. Absolute leader is the Top-Down algorithm. The slowest is the GA-based algorithm. Bottom-Up takes the second position for series length of 250, but not of 50. Here the Sliding Window algorithm holds the second position.

Segments number	Series length	Algorithm			
		Sliding W.	Top-Down	Bottom-Up	GA-based
3	10	1 (best)	1	1	4
	50	2	1	3	4
	250	3	1	2	4
4	10	1	1	1	4
	50	2	1	3	4
	250	3	1	2	4
5	10	1	1	1	4
	50	2	1	3	4
	250	3	1	2	4
6	10	1	1	1	4
	50	2	1	3	4
	250	3	1	2	4

Table 2.5: TRA ranking of PLA algorithms

Figures E.17-E.28 display the performance of the algorithms on the overlay graphs. For each combination of series' length and the number of segments there is one figure. The relative position of each CDF curves suggests the ranking of the algorithms under different settings. Table 2.6 provides the summary. One can see that GA-based algorithm provides superior fit according to  $R^2$ . Second is the Sliding Window algorithm. Then, depending on the number of segments or the series' length there is Top-Down or Bottom-Up algorithm.

### 2.2.7. Discussion

Piecewise linear approximation is an important ingredient of the Artificial Technical Analyst. Without this mechanism the system is insensitive to breaking points in the time series. The segmentation algorithm is an effective way to perform this task. The analysis performed in this chapter gives the number of choices. Depending on speed or accuracy needed one can pick the most appropriate algorithm. To find the properties of each algorithm the Monte Carlo simulations were used. Since the aim of piecewise linear approximation is to identify the relative position of key points in time series, the random walk was selected for

<sup>17</sup>This computation was performed on Intel 2.8 GHz processor with activated hyperthreading (for details see <http://www.intel.com/technology/hyperthread>). For maximum performance the Matlab code was compiled.

Segments number	Series length	Rank based on	Algorithm			
			Sliding Window	Top-Down	Bottom-Up	GA-based
3	10	$SSE$ $R^2$	2 2	3 3	4 4	1 (best) 1
	50	$SSE$ $R^2$	2 2	3 3	4 4	1 1
	250	$SSE$ $R^2$	3 3	2 2	4 4	1 1
4	10	$SSE$ $R^2$	2 2	4 3	3 4	1 1
	50	$SSE$ $R^2$	2 2	3 3	4 4	1 1
	250	$SSE$ $R^2$	2 2	3 3	4 4	1 1
5	10	$SSE$ $R^2$	2 2	4 3	3 4	1 1
	50	$SSE$ $R^2$	2 2	4 4	3 3	1 1
	250	$SSE$ $R^2$	2 2	4 4	3 3	1 1
6	10	$SSE$ $R^2$	2 2	4 4	3 3	1 1
	50	$SSE$ $R^2$	2 2	4 4	3 3	1 1
	250	$SSE$ $R^2$	2 2	4 4	3 3	1 1

Table 2.6: Final ranking of PLA algorithms

data generation. The data produced by the random walk have advantages of been close to financial time series and easy generation of as many observations as the Monte Carlo studies require. Tables 2.5 and 2.6 can help with the right choice. For example, if one is indifferent about the computational time, the wise choice is GA-based algorithm. If one would need to perform the task in a constant time, the choice is either the Top-Down, or the Bottom-Up algorithm. For long time series the Bottom-Up algorithm provides superior fit compared to the Top-Down approach.

### 2.3. Learning Classifier Systems

The artificial technical analyst is designed to replicate the professional expertise of a technical analyst. The quality of an analysis crucially depends on the cognitive ability of the ATA. Imitation of human-like cognition is a very cumbersome task, since one needs to match the abilities of human brain. The sophistication of the human brain allows it to instantaneously process incoming information, to associate it with past experience, and to make sensible prognoses. As a result we are able to learn from our experience and to

generalize it to new, unseen situations. With the help of Artificial Intelligence (AI) the ATA should be able to replicate the process of human-like cognition.

In the area of AI there are several mechanisms of replicating the cognitive process. Artificial Neural Networks (ANN), Genetic Programming (GP) and Learning Classifier Systems (LCS) are the most prominent. Each mechanism has its own merits and drawbacks. Given that the author has subjective preference for LCS, it will be selected as *a priori* choice. To make the picture complete, this section starts with an overview of ANN and GP. It is followed by the analysis of LCS, and its favorite implementation - XCS. An introduction and discussion of a new implementation, TiCS, completes the section.

### 2.3.1. Artificial Neural Networks

The origin of ANN comes from the formulation of the Neural Doctrine in 1836. It states that the nervous system of living organisms is a structure consisting of many simple elements (neurons) working in parallel and interconnection. While each neuron can hold only a limited number of states, the interconnection of neurons can hold much wider diversity of states. As a result such system can learn complex paradigms, and to exhibit sophisticated behavior.

The design of ANN reflects the Neural Doctrine.<sup>18</sup> Figure 2.10 illustrates the layout of ANN. They have several layers of neurons depicted as circles. All neurons are interconnected with each other, creating multiple chains - neural networks. The connection of neurons allows to carry through the system input signals  $x_1, x_2, \dots, x_N$ , which are eventually transformed into output signals  $y_1, y_2, \dots, y_M$ .

While neural networks do not have to be adaptive *per se*, their practical use comes with algorithms designed to alter the strength (weights) of the connections in the network to produce a desired signal flow. The general mathematical model of ANN defines a function  $f : X \rightarrow Y$ . The function  $f(x)$  is defined as a composition of other functions  $g_i(x)$ , which can be further defined as compositions of other functions. In Figure 2.10 each neuron is a representation of one basic function. One way to compose  $f(x)$  is to use the nonlinear weighted sum:<sup>19</sup>

$$f(x) = K \sum_i w_i g_i(x), \quad (2.16)$$

where  $K$  is some predefined function,  $w_i$  is a weight of the  $i$ -th connection and  $x = (x_1, x_2, \dots, x_N)$  is an input vector.

To measure how far away is the output of ANN from the desired signal flow there is a cost function. ANN search through the solution space in order to minimize the cost. In the process of training, low-cost connections gets stronger (their weight increases), while others gets weaker. Chains of strong connections allows ANN to learn, recognize and react to specific patterns in incoming information.

<sup>18</sup>For the introduction, design and application of ANN see [8, 100, 34].

<sup>19</sup>This is one of the compositions used in ANN. For more details and examples see [8, 100, 34].

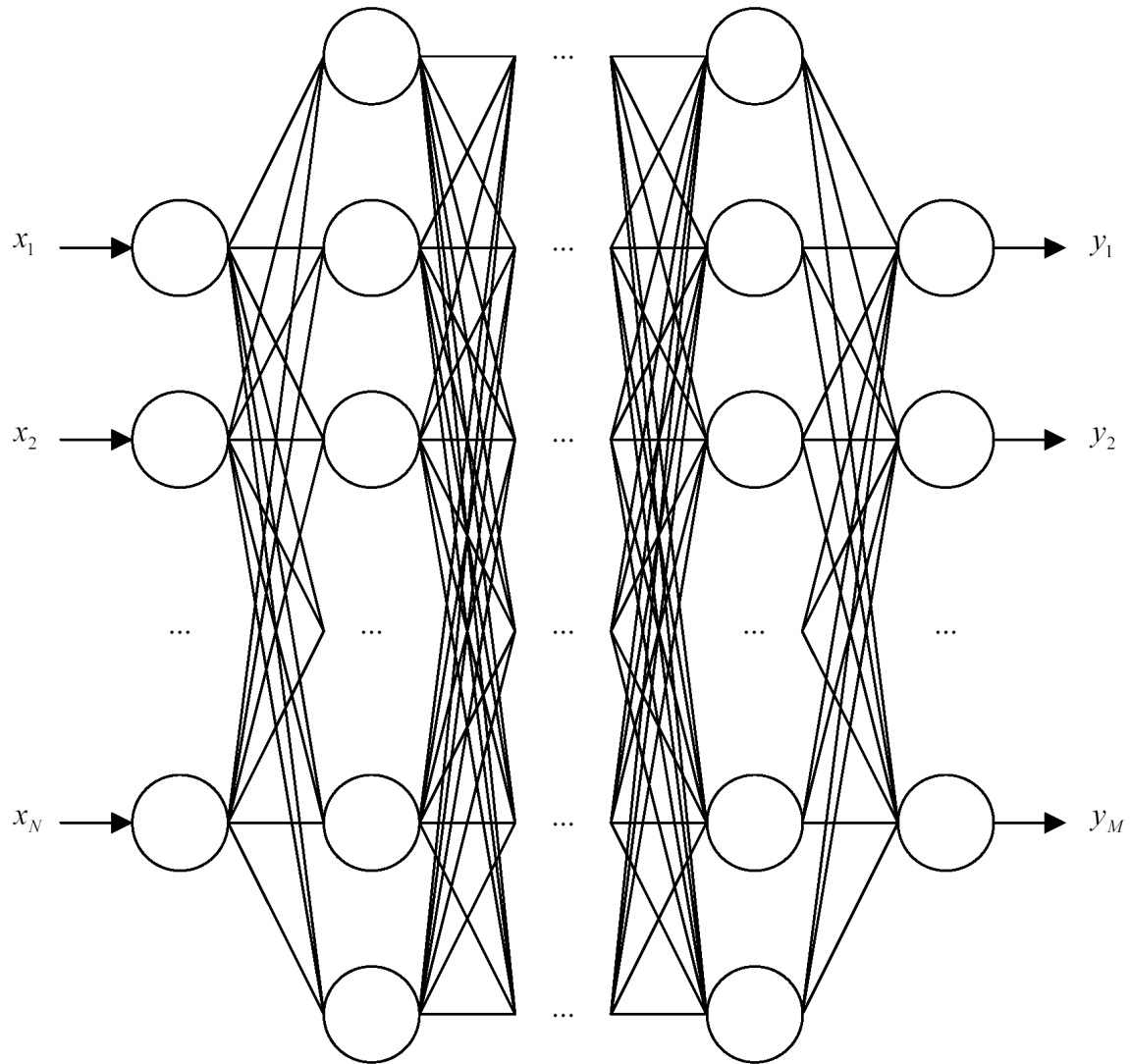


Figure 2.10: Illustration of signal flow in ANN

The diagram shows the signal flow in ANN. Neurons are depicted as circles. Input signals  $x_1, x_2, \dots, x_N$  are carried through interconnected neurons and transformed to output signals  $y_1, y_2, \dots, y_M$ .

The close resemblance of ANN to nervous system comes with a price. The use of ANN and the control of its cognitive process are not straightforward. This is attributed to a non-transparent structure of ANN and their sensitivity to algorithm and parameter selection. First, the complex interconnections between neurons make tracing the cognitive process virtually impossible. Second, one has to appropriately specify a model, cost function and learning algorithm to have robust ANN. Third, selecting and tuning parameters for training on unseen data requires a significant amount of experimentation and is not protected from a subjective influence.

### 2.3.2. Genetic Programming

GP is an automated methodology to write computer programs that best perform a user-defined task. The origin of GP is the work of J. Holland on Genetic Algorithms (GA).<sup>20</sup> While in original GA the alphabet for a bit-string is either binary or continuous variable, in GP the alphabet includes symbolic and numerical values, logical and mathematical operators, as well as, complex functions. Additionally, the length of the bit-string is not fixed. As in GA there is a population of trial solutions. Fitness of each solution is evaluated through the user-specified criteria. A repeated application of genetic operations over the population of trial solutions constitutes the search for optimal solution.<sup>21</sup>

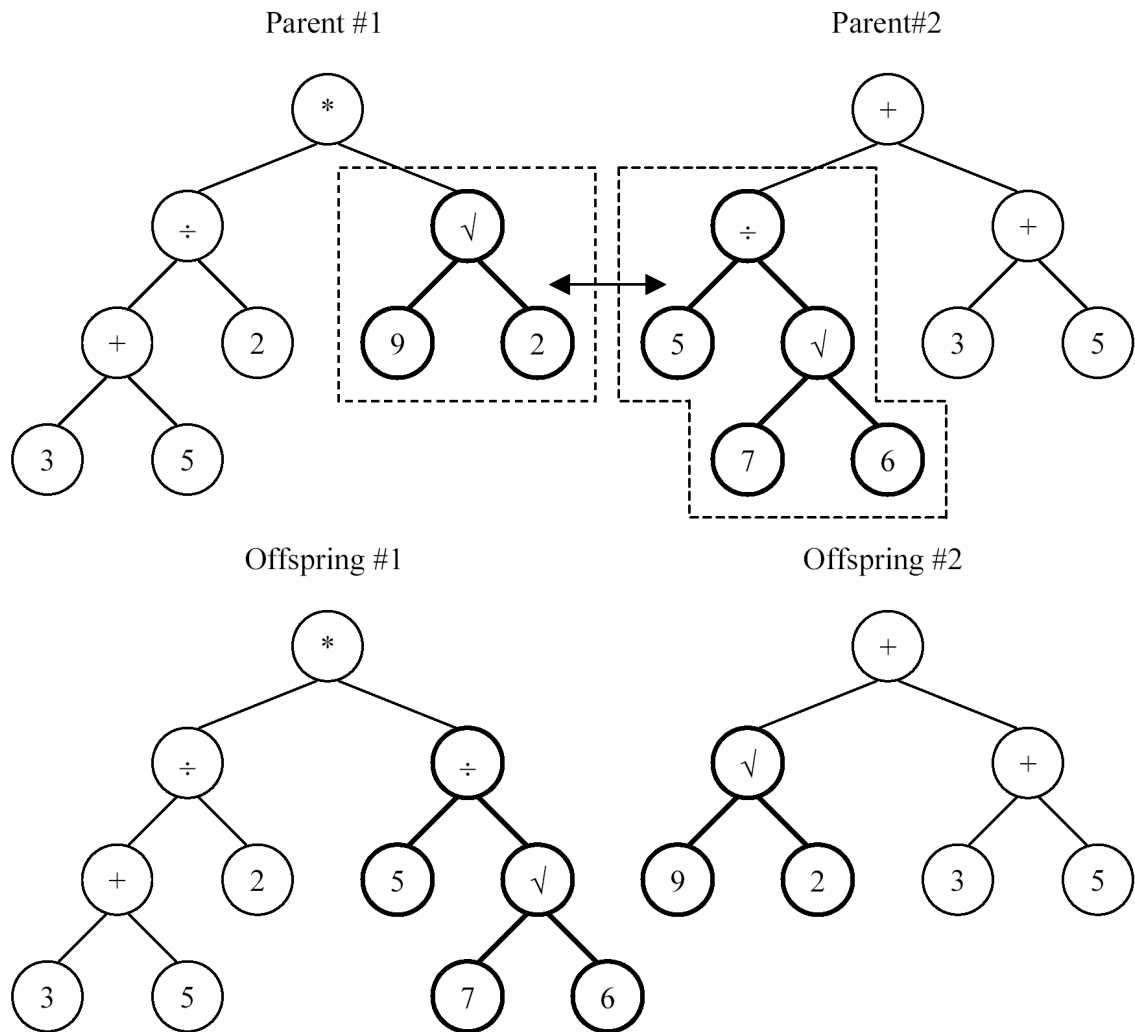


Figure 2.11: Illustration of crossover operator in GP

Figure 2.11 shows two exemplary trial solutions selected for a crossover operator. They are graphically depicted as Parent #1 and Parent #2. These tree structures can be interpreted as sequences of variables (numerical values) and mathematical operators, including

<sup>20</sup>For details on GA and genetic operators see Appendix B.

<sup>21</sup>For more details see [137].

predefined complex functions. The tree representation of trial solutions allows to swap nodes and complete branches. The result of this crossover operator are two new trial solutions, Offspring #1 and Offspring #2. The application of the crossover operator allows to exchange valuable information within the population of trial solutions. The crossover operator can be applied to any node including the terminal one.

**Example 3.** In Figure 2.11 the graphical representation of the trial solution Parent #1 expresses  $(3 + 5) \div 2 * \sqrt[3]{9}$ , for Parent #2 it is  $5 \div \sqrt[6]{7} + (3 + 5)$ . The crossover operator swaps  $\sqrt[3]{9}$  from Parent #1 with  $5 \div \sqrt[6]{7}$  from Parent #2. A new trial solution Offspring #1 gives expression  $(3 + 5) \div 2 * 5 \div \sqrt[6]{7}$ . Offspring #2 gives  $\sqrt[3]{9} + (3 + 5)$ .

Figure 2.12 shows an example of a mutation operator. Mutation can be applied to any node. For the terminal nodes mutation changes the value by a random increment. In the case mutation is applied to a node containing a mathematical operator or a function, the value of the node is randomly replaced by one of the available alternatives from the alphabet of operators and functions.

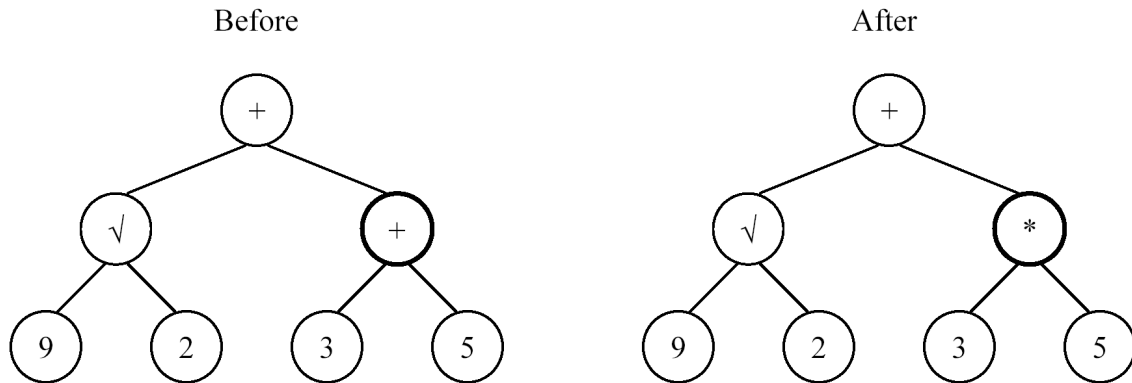


Figure 2.12: Illustration of mutation operator in GP

**Example 4.** In Figure 2.12 a node containing an additive operator (it sums 3 and 5) was randomly selected. Mutation of this node produced a new trial solution  $\sqrt[3]{9} + (3 * 5)$ .

The first experiments with GP were reported by [229] and [52] in the second half of the 1980s. In the following decade GP showed high potential to solve complex problems with human-competitive results<sup>22</sup>. Computational intensity constrains GP to solving relatively simple problems. With growing computational power the scope of GP steadily increases.

<sup>22</sup>The result is “human-competitive” if it satisfies one or more of the eight criteria below [257]:

- (A) The result was patented as an invention in the past, is an improvement over a patented invention, or would qualify today as a patentable new invention.
- (B) The result is equal to or better than a result that was accepted as a new scientific result at the time when it was published in a peer-reviewed scientific journal.
- (C) The result is equal to or better than a result that was placed into a database or archive of results maintained by an internationally recognized panel of scientific experts.
- (D) The result is publishable in its own right as a new scientific result - independent of the fact that the result was mechanically created.

### 2.3.3. Learning Classifier Systems

In the area of Machine Learning, LCS is a mechanism meant to imitate the human ability for classification, learning and generalization. The Encyclopedia Britannica defines these abilities as follows. *Classification* is the ability to systematically arrange in groups or categories according to established criteria. *Learning* is the ability to adapt to the environment and to alternate behavior as a result of individual experience. And *generalization* is the ability to respond in the same way to different but similar environmental conditions. To meet these criteria, LCS should possess the following characteristics.

- (i) the ability for on-line classification and establishing the patterns of the different environmental situations.
- (ii) the ability to distinguish and preserve the most persistent patterns.
- (iii) the ability to ignore any irrelevant or noise information.

Mimicking the human process of learning, the evolutionary approach constitutes LCS.

The first LCS were introduced by J. Holland [106, 107].<sup>23</sup> LCS were designed to read current environment state in terms of a fixed number of predetermined conditions and to provide the most adequate mapping into the space of coming events. For this purpose during the training period LCS employ Genetic Algorithms (GA)<sup>24</sup> to identify a correct mapping from the combination of predetermined conditions to the most probable event. In the process, LCS should identify irrelevant or noise conditions and distinguish persistent combinations of conditions.

**Example 5.** *To better understand the problem one can think of the following scenario. A fisherman keeps the weather records of his fishing trips along with the records of his catch. He records whether that day it was windy, sunny or rainy. Each is binary, encoded with 1 if the event was present, and 0 if not. The records of catch are in quantity units. Table 2.7 gives an example of environmental records.*

*After collecting sufficiently long records the fisherman wants to know which factors and their combination had the influence on the quantity of the catch given that his fishing methods, skills, and other factors were the same for all trips. Second question is what will be the expected catch if weather conditions are known right before the trip and are expected to stay the same all the day?*

---

(E) The result is equal to or better than the most recent human-created solution to a long-standing problem for which there has been a succession of increasingly better human-created solutions.

(F) The result is equal to or better than a result that was considered an achievement in its field at the time it was first discovered.

(G) The result solves a problem of indisputable difficulty in its field.

(H) The result holds its own or wins a regulated competition involving human contestants (in the form of either live human players or human-written computer programs).

<sup>23</sup>For the introduction to LCS see [95, 173], for the recent survey on application and development see [143].

<sup>24</sup>For introduction to GA see Appendix B.

Trip Number	Wind	Sun	Rain	Catch
1	1	1	0	3
2	0	0	1	10
3	0	0	0	2
...	...	...	...	...

Table 2.7: Example of environmental records

##	Condition Bits	Action Bits	Classifier Fitness
1	01#0#10##10	0010	1.000
2	0#0#1101###	1010	0.998
3	1#0#11#1###	1100	0.884
...	...	...	...
$N$	####001##01	0110	0.001

Table 2.8: Exemplary population of classifiers in LCS

In general any LCS have the population of  $N$  randomly initiated classifiers. Each classifier has a condition part composed of condition bits, an action part and some fitness value. Table 2.8 demonstrates an exemplary population of classifiers.

The number of condition bits is fixed and the same for all classifiers in the population. The encoding of each bit follows the rule:<sup>25</sup>

$$c_i = \begin{cases} 1 & \text{if condition}_i, \\ 0 & \text{if not condition}_i, \\ \# & \text{if "don't care" about condition}_i, \end{cases}$$

where  $c_i$  is the bit at position  $i$  in the condition part of the classifier. This part is responsible for matching the particular input signal encoded as:

$$s_i = \begin{cases} 1 & \text{if condition}_i, \\ 0 & \text{if not condition}_i, \end{cases}$$

where  $s_i$  is the bit at position  $i$  in the encoded input signal. The number of bits in the condition part should be the same as the number of bits in the encoded input signal. This allows bit-by-bit comparison of the conditional part of the classifier and the input signal. For example, the encoded input signal 01001101010 will match classifier 1 and 2 (see Table 2.8). Each bit in classifier 1 or 2 matches the corresponding input bit, or has a

<sup>25</sup>In the literature one can find as well symbolic encoding. That is each event can be described in more than two states (in general any real number). The simplest example is color notation: W, B, G and so on:

$$c_i = \begin{cases} W_s & \text{if condition}_i \text{ is White,} & \dots 00_2 \\ B_s & \text{if condition}_i \text{ is Black,} & \dots 01_2 \\ G_s & \text{if condition}_i \text{ is Green,} & \dots 10_2 \\ \dots & & \\ \# & \text{if "don't care" about condition}_i, & \dots \#\# \end{cases},$$

where the subscript  $s$  stands for the symbolic system, and 2 for binary system. Given that any real number can be written in binary, one can use only binary alphabet.



"don't care" symbol, which instructs to ignore the condition<sup>26</sup>. The first bit in classifier 3 does not match the first bit in the encoded input signal. It indicates the classifier 3 does not match the current input signal.

With  $N$  classifiers in hand LCS should learn the mapping from the environmental state to the space of coming events or actions. For this each classifier has an action part. Each bit in the action part can be encoded in the way similar to the condition bits, but with a different set of conditions. The past success of each classifier is reflected in fitness.

To learn the mapping LCS should process records, where encoded input signal is followed by an environmental state<sup>27</sup>. Each time LCS get from the population the subset of classifiers, whose condition part matches the current input signal. The subset of matched classifiers is ordered based on its fitness. The action part of the classifier with the highest fitness suggests the possible coming event. After observing the actual outcome the fitness of all classifiers in the current subset is adjusted. After that, classifiers strong in fitness participate in the GA. The GA produces the new generation of classifiers which replace the current subset or worst classifiers in the overall population. With the new input signal the process is repeated.

##	Condition Bits	Action Bits	Classifier Fitness	Decoded Catch
1	0#0	00101 <sub>2</sub>	1.000	5 <sub>10</sub>
2	0#1	10100 <sub>2</sub>	0.998	20 <sub>10</sub>
3	1#1	11001 <sub>2</sub>	0.884	25 <sub>10</sub>
...	...	...	...	...
$N$	###	01011 <sub>2</sub>	0.001	11 <sub>10</sub>

Table 2.9: Sample realization of environment conditions

**Example 6.** *Proceeding with the fisherman example one can use accumulated weather and catch records for the CS. The population has  $N$  classifiers. Each classifier has 3 bits in the condition part and 5 bits in the action part.<sup>28</sup> The number of bits in the action part defines limits of the catch:  $00000_2 = 0_{10}$ , and  $11111_2 = 31_{10}$ , where the subscripts stand for binary and decimal system. After running the CS though all records the population of classifiers can be as in the example, Table 2.9. Now the fisherman can see, first, that the sun does not influence his catch; and second, he can have the prognoses of the catch by*

<sup>26</sup>With the total of  $K$  conditions the population should have  $2^K$  classifiers to match any possible input signal. With the "don't care" symbol, #, some classifiers match more than one input signal. This allows, first, to dramatically reduce the population, and, second, to identify rare or unimportant conditions by marking the corresponding position with the "don't care" symbol.

<sup>27</sup>In the fisherman example that is the weather record and the catch:

Trip Number	Wind	Sun	Rain	Encoded Input Signal	Catch	Encoded Catch
1	1	1	0	110	3	00011
2	0	0	1	001	10	01010
3	0	0	0	000	2	00010
...	...	...	...	...	...	...

<sup>28</sup>Even though the notion "action part" is rather misleading here, since it is only descriptive, it is preserved to be according to LCS literature formulation.

observed weather: with the wind and the rain he should expect the catch of 25, with only the rain it is 20, and 5 with no wind and no rain.

Figure 2.13 summarizes the learning process of LCS. Two additional comments are required on the Bucket Brigade Algorithm (in the *Credit Allocation*) and the *Genetic Algorithm*.

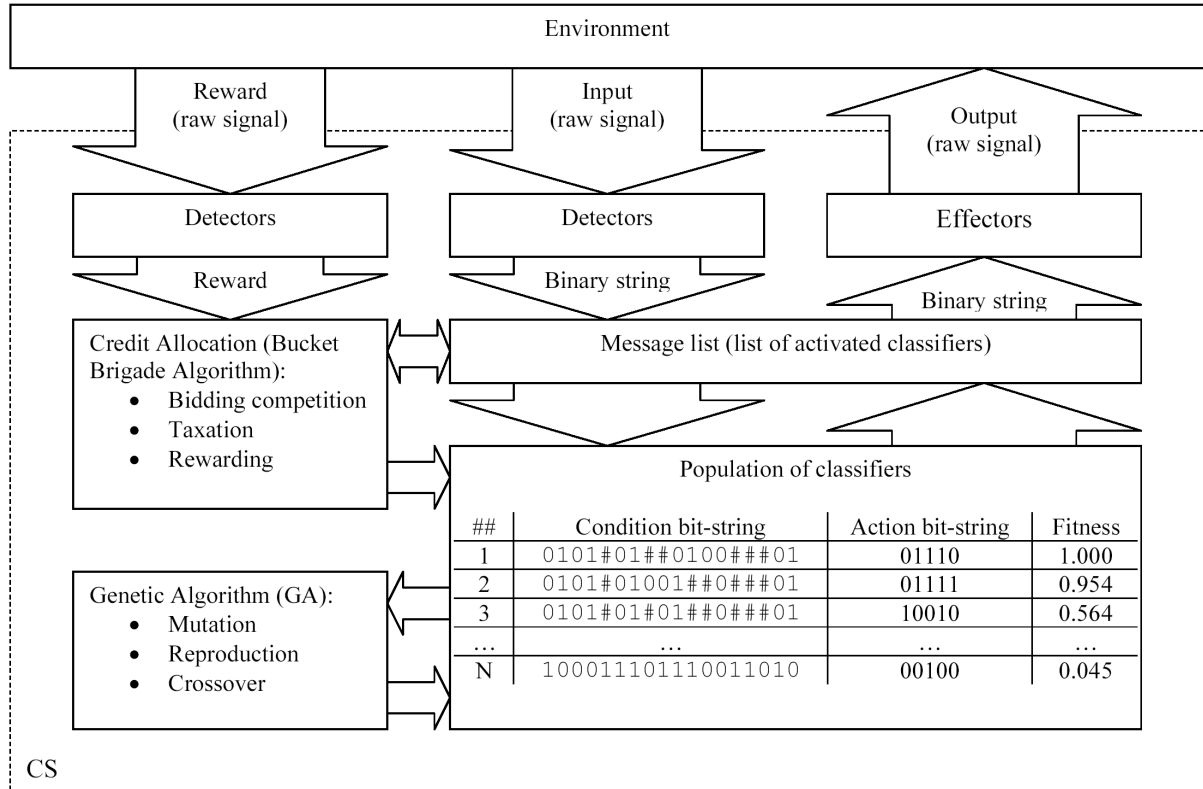


Figure 2.13: Holland's classifier system and its environment

The Credit Allocation serves to rank individual classifiers according to a classifier's role. In learning multi-step environment the notion of Credit Allocation is substituted by the Bucket Brigade Algorithm, which is aimed at establishing the link between interdependent classifiers.<sup>29</sup> The idea behind is an information economy where the right to trade information is bought and sold by classifiers in the population. In each round all matched classifiers first participate in an activation auction. Each matched classifier makes a bid  $B$ , which is a function of its current fitness.<sup>30</sup> In this way classifiers that are highly fit are given preference over other classifiers. Once a classifier is selected for activation, its fitness is reduced by the amount of a bid. The bid, in turn, is in some way distributed among recently activated classifiers. The division of the bid among contributing classifiers insures a survival of fit classifiers and their chains in interspecies competition. [95].

<sup>29</sup>One example of the multi-step environment is the search for food in woods. For details on this problem see [39]. In a single-step environment the Bucket Brigading can be omitted from the Credit Allocation without disturbing the cognitive properties of LCS.

<sup>30</sup>The exact functional form depends on the realization of LCS. Normally it is proportional.

##	C. Bits	A. Bits	$t = 0$				$t = 1$				
			Fitness	Input	Match to	Bid	Fitness	Input	Match to	Bid	
(1)	01##	0000	200	0111	(Env.)	20	180	0000			
(2)	00#0	1100	200				200		(1)	20	
(3)	11##	1000	200				200				
(4)	##00	0001	200				200		(1)	20	
Initial Input				0111	(Env.)	20				(Env.)	20

##	C. Bits	A. Bits	$t = 2$				$t = 3$				
			Fitness	Input	Match to	Bid	Fitness	Input	Match to	Bid	
(1)	01##	0000	220				220				
(2)	00#0	1100	180	1100			218				
(3)	11##	1000	200		(2)	20	180	1000			
(4)	##00	0001	180	0001	(2)	18	162	0001	(3)	16	
					(Env.)	20				(Env.)	20

##	C. Bits	A. Bits	$t = 4$				$t = 5$
			Fitness	Input	Match to	Bid	Final Fitness
(1)	01##	0000	220				$200 - 20 + 20 + 20 = 220$
(2)	00#0	1100	218				$200 - 20 + 20 + 18 = 208$
(3)	11##	1000	196				$200 - 20 + 16 = 196$
(4)	##00	0001	156	0001			$200 - 20 - 18 - 16 = 156$
					(Env.)	20	

Table 2.10: LCS by hand - matching and payments

Source: Adopted from [95].

**Example 7.** To illustrate the Credit Allocation one can use a following problem. In the first step LCS face an initial input signal. Next, matched classifiers participate in the auction. The winner provides its action bits as the next step input signal. At each step classifiers are paying bids either to the environment (at the first step) or to the classifier which provided the input signal. Here the bid is proportional to the classifier fitness:

$$B_i(t) = C_{bid}F_i(t), \quad (2.17)$$

where index  $i$  denotes the  $i$ -th classifier in the population;  $F_i(t)$  is the fitness of the  $i$ -th classifier at time  $t$ ;  $C_{bid}$  is a bid coefficient. The bid coefficient is  $C_{bid} = 0.1$ .

Table 2.10 shows the process of credit allocation. In the initial time step ( $t = 0$ ), classifier 1 is matched. It bids 20 units of fitness and sends its action bits as the next step input signal. In this case the environment is assigned of 20 units to be distributed at the end. In subsequent time steps. activated classifiers make their payment to previously active classifiers.<sup>31</sup>

Within the Credit Allocation there is a taxation mechanism. It is used to prevent biasing the population only towards recently active classifiers. Taxation reduces fitness of

<sup>31</sup>Adopted from [95].

active classifiers, allowing them to have only a slight advantage in comparison to currently inactive classifiers. With no pronounced dispersion in fitness all classifiers in population have similar chances to participate in the reproduction process.

The key element in the Credit Allocation and in LCS is a reward mechanism. It provides a quantitative feedback to LCS. Through the reward mechanism the fitness of each active classifier is adjusted according to its performance. Fitness is increased for a good performance and reduced for a bad one. Formally the fitness evolution of each classifier can be described as:

$$F_i(t+1) = F_i(t) - B_i(F_i(t)) - T_i(F_i(t)) + R_i(t), \quad (2.18)$$

where  $T_i(\cdot)$  is a classifier's tax, and  $R_i(t)$  is a classifiers reward. Like in the case of the bid, the tax is a function of fitness. The functional form of the tax and reward is specific to the realization of LCS.

While the Credit Allocation is a clean procedure for evaluating and deciding among competing classifiers, one has to devise a way of injecting new and possibly better classifiers into the system. This is done through the application of GA. There are a few modifications which are required for its successful application. First, only a fraction of the population is replaced through GA. This permits diversity in covering the solution space. Second, the invocation of GA takes place at predefined intervals. In this way classifiers have time to accumulate and average their fitness. That is classifiers should get sufficient statistical significance of condition, which they match. The intervals of the invocation can be either deterministic or stochastic or conditional on particular events such as lack of match or poor performance. Third, mutation must be modified because LCS use a ternary alphabet. When mutation is called for a bit its character changes to one of the other two with equal probability ( $0 \rightarrow \{1, \#\}$ ,  $1 \rightarrow \{0, \#\}$ ,  $\# \rightarrow \{0, 1\}$ ) [95].

#### 2.3.4. XCS

In the seminal work of J. Holland and in the related literature of the following decades the adequacy of each classifier was measured by the predetermined criteria, known as strength or fitness. This parameter was serving both as a predictor of future payoff and as classifier's fitness for the genetic reproduction. Unfortunately, this primitive aggregation resulted in a low performance of LCS. As a result, the considerable enthusiasm of the 1980's turned to decline in the early 1990's. LCS seemed too complicated to be studied, with only few successful applications reported. In the mid 1990's the field appeared almost at dead end [108].

In response to this situation Wilson [251] introduced a XCS. The primary distinguishing feature of the XCS is that classifier fitness is based on the accuracy of classifier payoff prediction rather than on payoff prediction (strength) itself. Figure 2.14 gives an overall picture of the XCS, which is shown in interaction with an environment via detectors for sensory inputs and effectors for actions.

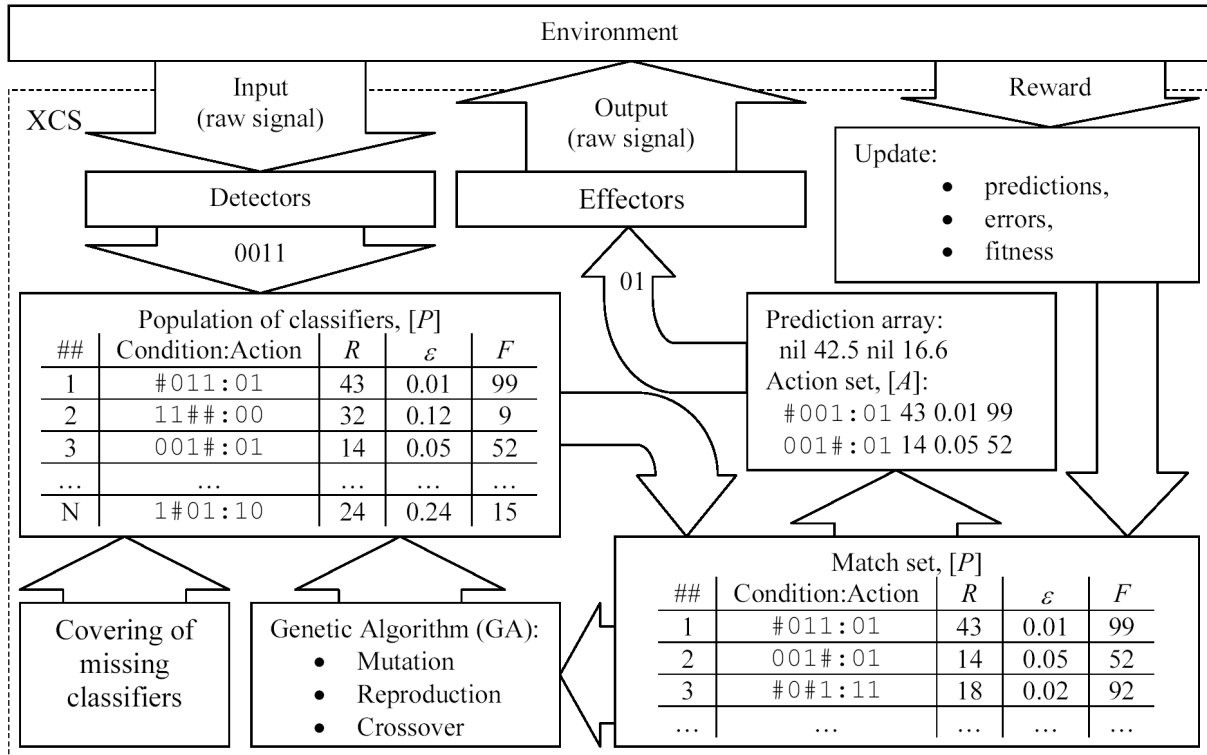


Figure 2.14: Schematic illustration of XCS

The main component of the XCS is a population of classifiers. Each classifier carries a set of specific conditions and a corresponding action. In the process of learning, classifiers are becoming specialized in a problem solution. Successful learning makes each classifier to be taken as an expert in its problem. To identify the degree of an "expert" knowledge each XCS classifier has five main components and several additional estimates [39]:

- (1) *Condition part*,  $C$ , specifies when the classifier is applicable. In the binary case it is coded by  $C \in \{0, 1, \#\}^L$ , where  $L$  is the length of condition bit-string.  $C$  identifies a hypercube in which the classifier is applicable.<sup>32</sup>
- (2) *Action part*,  $A$ , specifies a proposed action or classification or solution.
- (3) *Reward prediction*,  $R$ , estimates the average reward received. Reward prediction  $R \in \mathfrak{R}$  is iteratively updated resulting in a moving average measure of encountered reward received in the recent problem instances in which condition  $C$  was applicable and action  $A$  was executed.
- (4) *Reward prediction error*,  $\varepsilon$ , estimates the absolute difference between the reward prediction and the actual reward from the environment.
- (5) *Fitness*,  $F$ , estimates the moving average of the accuracy of the classifier's reward prediction relative to other classifiers that are applicable at the same time.
- (6) *Action set size*,  $as$ , estimates the moving average of the action set size.
- (7) *Time stamp*,  $ts$ , specifies the time when the classifier was a part of GA competition.

<sup>32</sup>The notion of applicability is taken in the sense of matching a condition part with the input signal.

- (8) *Experience counter*,  $Exp$ , counts the number of parameter updates the classifier underwent so far.
- (9) *Numerosity*,  $num$ , specifies the number of classifiers in the population which are identical.

The beginning of learning process starts either with an empty or with randomly generated population of classifiers. At each iteration step detectors bring into the system a coded instance of learning problem  $s \in S$ , where  $S$  is a problem space. After matching a match set  $[M]$  is created. In the case there are no classifiers in  $[M]$ , or the actions are not well represented the process of covering completes the set  $[M]$ .<sup>33</sup> Given the match set, the XCS estimates the payoff for each possible action by forming a prediction array  $P(A)$ . The prediction array reflects the fitness-weighted average of all reward prediction estimates for the classifiers in  $[M]$  that suggest action  $A$ . A set of currently available actions is represented by set  $[A]$ . The maximum value in the prediction array defines the action to be executed out of  $[A]$ . A set of effectors carry out the selected action to the environment.

After the execution a feedback is received in a form of reward  $R$ . Its value is used to update the parameters of involved classifiers. Updating completes the current iteration. At the next step the process is repeated with a new coded instance of learning problem  $s$ .

Beside the covering mechanism the XCS applies GA to the evolution of rules. GA are invoked in the current action set  $[A]$  if the average time since the last GA application upon the classifiers in  $[A]$  exceeds some threshold. GA selects two parental classifiers using a roulette-wheel mechanism with probabilities defined by classifiers' relative fitness in  $[A]$ . Offsprings are generated using reproduction, crossover and mutation operators. Normally, genetic operators change the condition part of the classifiers. In the niche case new classifiers are required to match a current problem instance. Prediction, error and fitness parameters of offspring classifiers are set to be the same as the respective parameters of parents. Newly created classifiers then join the population where they have to compete with their parents. In the XCS the population of classifiers is of maximum size  $N$ . Excess classifiers are deleted from the population if their experience does not result in high performance.

Although the XCS is currently a favorite, it has some negative features. Its overcomplicated structure slows down the algorithm in finding solutions. The XCS originates from a "zeros level" classifier system, which was intended to simplify Holland's canonical framework while retaining the essence of the classifier system idea [251]. At some step the intention to simplify turned out to overcomplicate. To illustrate this point Table 2.11 lists the 28 parameters and switches, which should be endogenously specified. Their numerosity and vague explanation in the source literature<sup>34</sup> makes tuning of XCS into an art than a precise science. This parameters are claimed to change the XCS's behavior, adjust it to the current problem and specify output characteristics [40]. Unfortunately, low transparency of

<sup>33</sup>Covering creates classifiers that match  $s$  by copying  $s$  to the condition part of new classifiers and replacing some condition bit with "don't care" symbol #. The corresponding actions are randomly generated.

<sup>34</sup>There are only few articles where some of these parameters are studied in detail. The best description is provided by [40, 39], which is summarized in Table 2.11.

##	XCS parameter	Value	Explanation [40, 39]
1	$N$	$10^4$	Maximum population size
2	$\alpha$	0.100	Accuracy determination parameter
3	$\beta$	0.200	Learning rate
4	$\gamma$	0.950	Reward discount factor
5	$\epsilon$	0.010	$\epsilon$ -greedy strategy parameter
6	$\nu$	5.000	Accuracy determination parameter
7	$\theta_{GA}$	25.00	Threshold that controls GA invocation
8	fitnessReduction	1.000	Specifies the fitness fraction that an Offspring receives from its parents (1 = no fitness reduction)
9	$\tau$	0.400	Tournament size (proportion of the current action set size)
10	selectTolerance	0.001	Difference in fitness or error that is neglected in the tournament selection
11	forceDifferentInTournament	NO	Enforces two different classifiers to be selected in the tournament selection
12	crossoverType	0	Specifies the crossover type: 0 = uniform, 1 = 1-point, or 2 = 2-point crossover
13	$\chi$	1.000	Probability of applying the chosen crossover operator
14	$\mu$	0.010	Probability of mutating a condition attribute (or the action)
15	doGeneralizationMutation	NO	Specifies if purely generalizing mutation should be applied
16	doNicheMutation	NO	Specifies if niche mutation should be used
17	$\delta$	0.100	Fraction of mean fitness below which deletion probability is further decreased
18	$\theta_{del}$	20	Threshold that requires minimum experience for fitness influence during deletion
19	deletionType	1	Specifies the deletion type: 0 = random deletion, 1 = fitness and action set size ( $ [A] $ ) estimate bias, 2 = fitness bias only, 3 = $ [A] $ estimate bias only, 4 = error bias only, 5 = error and $ [A] $ estimate bias
20	doMAM	YES	Specifies if the moyenne adaptive modifiée technique should be applied
21	$P_{\#}$	0.660	"Don't care" probability

Table 2.11: XCS parameters and their values

Reported values are used in the study presented in Subsection 2.3.6.

##	XCS parameter	Value	Explanation [40, 39]
22	doErrorBasedSelect	NO	Specifies if selection should be based on the error estimate of the classifiers instead of the fitness estimate
23	doGASubsumption	YES	Specifies if GA subsumption should be applied
24	doActionSetSubsumption	NO	Specifies if action set subsumption should be applied
25	$\theta_{sub}$	20	Threshold that requires minimal experience for subsumption
26	exploreProb	1.000	Specifies the probability of choosing an action at random during exploration. Otherwise, the best action in the prediction array is executed
27	teletransportation	50	Specifies the maximum number of steps in one exploration trial
28	initializePopulation	NO	Determines if the population should be initialized with randomly generated classifiers

Table 2.11 (continued) XCS parameters and their values

the algorithm and a lack of theoretical studies complicate the re-specification of the endogenous parameters of the XCS.<sup>35</sup> As a result the XCS shows high inertia in learning relatively simple problems. To overcome those limitations a new classifier system was introduced here - a True Classifier System (TiCS).

### 2.3.5. True Classifier System: TiCS

After almost thirty years since the invention of LCS the field of Artificial Intelligence is overwhelmed with the spreading diversity of their new types. Through time the development of LCS went in the direction of enhancing the original Holland's LCS, which led to a substantial complication of the systems. The leading implementation, the XCS, is a clear example. The overcomplicated structure, low transparency and numerous tuning parameters make the application of contemporary LCS hard task. An additional "nail in the coffin" is a fixation to target LCS at solving complex behavioral problems, while LCS are not able to perform simple classification tasks in a feasible time frame.

To overcome the limitations of LCS, the original idea of J. Holland as well as the application domain of LCS have to be reconsidered. The first point is the application domain. Classifier Systems, as it follows from their name, are meant for classification tasks, i.e. for a recognition of individual patterns in information flows. As such, the classification should be the main application of LCS. The second point is a new implementation of LCS. It should be extremely simple to provide high transparency and a minimum of tuning parameters. To meet these conditions a new classifier system is introduced - the True Classifier System (TiCS).

<sup>35</sup>For this reason the comparative study uses the default parameters of the XCS.



The key element of the TiCS is the Micro-Genetic Algorithm ( $\mu$ GA)<sup>36</sup>. Unlike conventional algorithms the  $\mu$ GA requires specification of only 2 parameters. The same 2 parameters are used for the TiCS activation and run. The first parameter  $N_g$  defines the global population size of classifiers. The second parameter  $N_s$  instructs the algorithm on how many classifiers should match the input signal to form the sub-population.

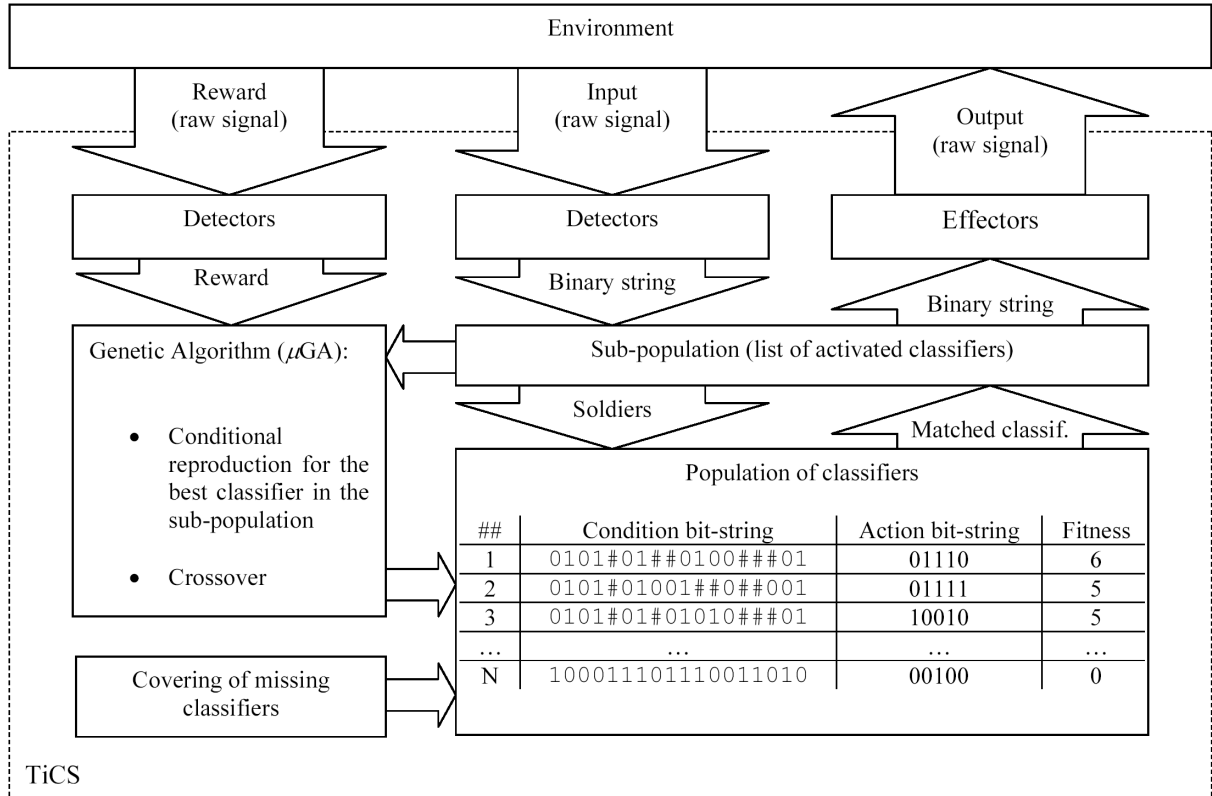


Figure 2.15: TiCS and its environment

Figure 2.15 outlines the operation of the TiCS. It follows the original idea of Holland's LCS. From a dynamic environment the TiCS extracts static sub-problems to provide an adequate response. For this purpose an input signal is matched against the condition part of each classifier. The matched classifiers form a current sub-population. Each classifier in the population should be as general as possible, i.e. it should correctly respond to the maximum number of states. A matched classifier is assumed to have the correct response. Given the goal of generalization, the measure of fitness is the quantity of "don't care" symbols in the condition part. This allows to rank the classifiers. If the number of matched classifiers is more than  $N_s$ , the classifiers with the lowest fitness are excluded from this sub-population.

The classifiers in the sub-population are referred to as **Soldiers**, since they are too **S**pecialized. The best in fitness classifier is referred to as a **G**eneral, since it achieved the most **G**eneral level without failing. The General forms the TiCS response to the input

<sup>36</sup>For details on the  $\mu$ GA see Appendix B.2.

signal. Meanwhile, the classifiers in the sub-population share the "experience" through the  $\mu$ GA. The probability of being selected is proportional to fitness. Selected classifiers crossover their condition and action parts to form a new sub-population. The General enters the new sub-population only if his response was adequate. The new sub-population then replaces the current sub-population in the global population.

In the case there is not sufficient number of classifiers in the sub-population a covering mechanism fills missing positions. Covering creates classifiers that match a current input signal by copying it to the condition part of new classifiers and replacing some condition bits with "don't care" symbol #. The corresponding actions are randomly generated.

**Example 8.** *Table 2.12 illustrates the first three iterations of learning 2-bits mapping with two followed noise bits. Successful learning should manifest in ability to map first two bits to the action and mark the last two bits with "don't care" symbol #. In the beginning the TiCS has 10 randomly initiated classifiers. In the first iteration the most general classifier from the matched subpopulation is classifier #5. Its action is adequate, so it is copied to the new subpopulation as it is and in addition it participates in the genetic reproduction process. In the second iteration the most general classifier from the matched subpopulation is classifier #5. Now it does not provide adequate action, so it is eliminated. In the third iteration there are two classifiers with the same highest generality. Out of them only the second one has an adequate response, so it is preserved, while the first one is eliminated. At this stage after only three iterations the TiCS learned already one correct mapping, presented by classifier #9.*

**2.3.5.1. Pattern Encoding Mechanism.** Since the TiCS is intended to be used for pattern recognition in time series it requires encoding mechanism. This mechanism allows to encode time series into strings of conditional bits.

At the first step a subsample of time series is demeaned and normalized to have unit variance. This allows to recognize similar patterns irrespective of current level or variance. Next, the PLA algorithm identifies positions of key points describing the subsample of time series. Each key point connects linear approximations found by PLA. With  $n$  predefined linear segments there are  $n + 1$  key points.

Application of demeaning and normalization allows to describe the position of each key point within the same two-dimensional coordinate space. An ordinate dimension corresponds to the level in time series and is constrained to be between  $[-2, 2]$ . A coordinate dimension corresponds to time stamps of time series and is constrained to be between  $[1, L]$ , where  $L$  is the length of the subsample.

At the second step the encoding takes place. First, two equal segments of a coordinate and ordinate space are defined. In the beginning they are  $[-2, 0)$  and  $[0, 2]$  for the coordinate axe, and  $[1, L/2)$  and  $[L/2, L]$  for the ordinate axe. Next a coordinate (ordinate) of a key point is checked to belong to one of the two segments. If it belongs to the below (left) segment of the coordinate (ordinate) then the first condition bit is assumed to be 1, otherwise it is assumed to be 0. Next the segment, where the key point is located is divided

$t = 0$ : Input signal 0100. General provides <b>correct</b> response.						New subpopulation	
##	C. Bits	A. Bits	Fitness	Match & Rank	Action	C. Bits	A. Bits
(1)	#1#0	10	2	S1		#1#0	01
(2)	0010	11	0				
(3)	0###	10	2	S1		0###	01
(4)	1###	00	3				
(5)	####	01	4	G	01	####	01
(6)	#01#	11	2				
(7)	1100	01	0				
(8)	01#0	11	1	S2		###0	01
(9)	11#0	00	1				
(10)	#0##	01	3				

$t = 1$ : Input signal 1110. General provides <b>wrong</b> response.						New subpopulation	
##	C. Bits	A. Bits	Fitness	Match & Rank	Action	C. Bits	A. Bits
(1)	#1#0	01	2	S2		#1#1	00
(2)	0010	11	0				
(3)	0###	01	3				
(4)	1###	00	3	S1		11#0	10
(5)	####	01	4	(G)	01	#1#0	01
(6)	#01#	11	2				
(7)	1100	01	0	(S4)			
(8)	###0	01	3	S1		###0	10
(9)	11#0	00	1	S3		11##	11
(10)	#0##	01	3				

$t = 2$ : Input signal 1101. 2nd general provides <b>correct</b> response.						New population	
##	C. Bits	A. Bits	Fitness	Match & Rank	Action	C. Bits	A. Bits
(1)	#1#1	00	2	(G)	C	#101	10
(2)	0010	11	0		C	##01	11
(3)	0###	10	3			0###	10
(4)	11#0	10	1		C	1##0	11
(5)	#1#0	01	2			#1#0	01
(6)	#01#	11	2			#01#	11
(7)	1100	01	0		C	###1	11
(8)	###0	10	3			###0	10
(9)	11##	11	2	G	11	11##	11
(10)	#0##	01	3			#0##	01

Table 2.12: TiCS by hand - learning 2-bits mapping

The table uses following notations:  $S\#$  – Soldier, where  $\#$  denotes common rank in the subpopulation with 1 as the highest. G – General. Letter in brackets denotes that this classifier was excluded from subpopulation at some stage. C – Classifiers are created by covering, where classifiers to be replaced are selected according to lowest generality.

into two new equal size segments. And again, the key point is checked to belong to one of the two segments and corresponding bit is added to condition bit. After predefined number

of iterations the coordinate (ordinate) position of the key point is encoded by a sequence of zeros and ones.

**Example 9.** *One needs to encode a position of a key point from demeaned and normalized subsample of 50 observations. The level of a key point is 1.17, and the time stamp is 11. Table 2.13 provides the results for encoding the first 5 bits. After performing iterative process the level is encoded as 11000, and the time stamp is encoded as 00111.*

Step ##	Ordinate segment		Ordinate bits	Coordinate segment		Coordinate bits
	Left	Right		Low	Up	
1.	[01, 25)*	[25, 50]	1	[-2.00, 0.000)	[0.000, 2.000]*	0
2.	[01, 12)*	[12, 25)	1	[0.000, 1.000)	[1.000, 2.000]*	0
3.	[01, 06)	[06, 12)*	0	[1.000, 1.500)*	[1.500, 2.000]	1
4.	[06, 09)	[09, 12)*	0	[1.000, 1.250)*	[1.250, 1.500)	1
5.	[09, 10)	[10, 12)*	0	[1.000, 1.125)*	[1.125, 1.250)	1
Final ordinate bit-string:			1 1 0 0 0	Final coordinate bit-string:		0 0 1 1 1

Table 2.13: Illustration of pattern-to-bits encoding mechanism

A subsample of time series has 50 observation (ordinate axe). Values are demeaned and normalized. Encoding of a key point with the level of 1.17 and the position of 11. Asterisk denotes that the position is within marked segment.

The encoding mechanism allows decoding of condition bit-string to get coordinate and ordinate limits. Since the TiCS has in the alphabet of condition bits the "don't care" symbol #, during decoding the process stops either when all bits are processed or the first symbol # is found. The latter case allows the TiCS have varying area within which key points should be.

**Example 10.** *Using the same settings as in the previous example one needs to decode bit-strings of coordinates and ordinates: 010## and 11010, respectively. Table 2.14 provides the results for decoding these bit-strings. After performing iterative process the level is decoded to be within [0.500, 1.000), and the time stamp is decoded to be within [07, 09).*

### 2.3.6. XCS vs. TiCS in Learning Single-Step Problems

This subsection investigates the performance of two implementations of LCS in learning simple classification problems. The results should indicate whether the XCS or the TiCS is the most appropriate for further application in the ATA. The first part presents test problems, which are designed to reveal the maximum potential of each implementation. The second part presents comparative results.

**2.3.6.1. Single-Step Problems: Introduction.** An ultimate goal of LCS is a cognition of environment. The quality and speed of the cognitive process differ from one implementation of LCS to another. Moreover, the same implementation but with a different set of tuning parameters can show dramatic differences in the learning process. To pick the best

Step ##	Ordinate bit-string	Ordinate segment		Coordinate bit-string	Coordinate segment	
		Left	Right		Low	Up
1.	1	[01, 25)*	[25, 50]	0	[-2.00, 0.000)	[0.000, 2.000]*
2.	1	[01, 12)*	[12, 25)	1	[0.000, 1.000)*	[1.000, 2.000]
3.	0	[01, 06)	[06, 12)*	0	[0.000, 0.500)	[0.500, 1.000)*
4.	1	[06, 09)*	[09, 12)	#		
5.	0	[06, 07)	[07, 09)*	#		
Final ordinate interval:		[07, 09)		Final coordinate interval:		[0.500, 1.000)

Table 2.14: Illustration of bits-to-pattern decoding mechanism

A subsample of time series has 50 observation (ordinate axe). Values are demeaned and normalized. Decoding of coordinate bit-string 010###, and ordinate bit-string 11010. Asterisk denotes selected segment.

implementation and an optimal set of tuning parameters one has to define selection criteria. They should clearly measure the quality of the learning process and the rate of learning. In addition, one can measure the LCS effectiveness in terms of computational resources allocation. To correctly use the selection criteria there is a number of test bed problems. They provide a controlled environment of experiments, where the selection and fine-tuning of LCS are done in a robust and efficient way.<sup>37</sup>

Test problems are designed to address specific aspects in LCS operation. The most important aspects are performance and generalization. The performance of LCS can be measured as a percentage of correct predictions over a fixed time interval. To monitor the learning process one normally gets averages in a fixed width sliding window. The generalization attributes to LCS the ability to correctly identify irrelevant information. The measure of generalization is a ratio of irrelevant bits to the overall number of bits in a condition bit-string. Normally, to measure the generalization the best classifiers are used. To address these aspects there is a class of single-step problems through which LCS should learn either simple boolean functions or some binary mapping. Along with relevant information there is a noise in the input to the system. The addition of the noise allows to control for the generalization ability of LCS in the experiment.

Test problems include multiplexer and  $k$ -bits mapping. The multiplexer problems have an important role in the area of machine learning, and especially, in the design and tuning of LCS. The success was attributed to its ability to provide a transparent and, at the same time, challenging environment. The complexity of the problem is  $2^{k+1}$ . In the early steps of LCS development the complexity of the multiplexer function was a sufficient challenge. Now there is a demand to solve more realistic, and thus, more complex problems. The multiplexer does not meet the demand for increasing complexity since the function output is only boolean, and the level of complexity is fixed for a defined bit-string length.

<sup>37</sup>For the extensive overview of problems related to testing machine learning techniques and, in particular, LCS see [39].

To be close to reality one would expect a quantitative or qualitative variable instead of the boolean answer. The  $k$ -bits mapping well satisfies this requirement. The output of the  $k$ -bits mapping is a variable, which can take  $2^k$  diverse answers. For a fixed length of the condition part, the problem complexity can vary with parameter  $k$ , reaching at maximum  $2^N$ .<sup>38</sup> Finally, a condensed location of control bits makes the  $k$ -bits mapping more transparent than the multiplexer function.

All test problems are reserved to measure the relative rate of learning. To access the relative rate one has to compare the steepness of learning curves. An overlay chart of performance curves can provide a simple visual solution to identify a superior LCS implementation. Every test problem also allows to measure the effectiveness of computational resources allocation. It can be proxied by the maximum number of classifiers defined in an implementation, since in sequential calculations there is almost linear dependence of the calculation time on the number of classifiers in the system. With scarce computational resources one would prefer the implementation which requires for the same performance and learning rate less classifiers.

**2.3.6.2. Single-Step Problems:  $N$ -Multiplexer.** Pertinent literature finds multiplexer problems to be challenging for machine learning [39]. In addition they provide a simple and intuitive way to test LCS in a single-step environment. In the LCS community the multiplexer problem has a celebrity position, as LCS proved to be superior in multiplexer learning relative to other machine learning techniques [64].

A  $N$ -multiplexer is a boolean function  $f(N_2)$  of length:

$$N = k + 2^k, \quad (2.19)$$

where  $k$  is an integer. Each realization of the  $N$ -multiplexer is a string of bits taking value 0 or 1. Depending on the value of  $k$ , the length of the  $N$ -multiplexer can be of  $N \in \{3, 6, 11, 20, 37, 70, \dots\}$ . The first  $k$  bits are used to encode an address into the remaining  $2^k$  bits. To find the address referred to by address bits their binary value should be translated into the decimal base. Then the actual location of an addressed bit is counted starting from  $k + 1$  bit, which has the decimal address of 0. The value of the function is the value of the addressed bit [136].

**Example 11.** *In the 6-multiplexer ( $N = 6$ ), the input to the system consists of a string of six binary digits, of which the first  $k = 2$  bits represent an index into the remaining  $2^k = 4$  bits. Let the realization of the 6-multiplex be  $100010_2$ . The decimal value of the first two bits is  $10_2 = 2_{10}$ . Starting from the third bit the value of the second part is  $1_2$ , which is the value of this realization,  $f(100010_2) = 1_2$ .*

To use a  $N$ -multiplexer as a test, the length of the binary string,  $N$ , is selected and fixed. On each iteration a random binary string of  $N$  bits is generated and used as the input to the classifier system. The classifier system generates a response which is compared

<sup>38</sup>This is not a case for the multiplexer problem, where for a given  $N$  the complexity is fixed at  $2^k$ . As a result, the peak complexity of the  $k$ -bits mapping with the same  $N$  is much higher.

to a correct value of the boolean function. The success of learning is reported back to the classifier system before the next iteration.

The complexity of the  $N$ -multiplexer is defined by the number of maximally general classifiers. For the multiplexer problem it is  $2^{k+1}$  maximally general classifiers.<sup>39</sup> To be maximally general a classifier should have all but address and addressed bits filled with "don't care" symbol #.

**Example 12.** *For the 6-multiplexer problem the input space can be completely covered by  $2^{k+1}|_{k=2} = 8$  maximally general classifiers: 000### : 0, 001### : 1, 01#0## : 0, 01#1## : 1, 10##0# : 0, 10##1# : 1, 11###0 : 0, 11###1 : 1.*

**2.3.6.3. Single-Step Problems:  $k$ -Bits Mapping.** The main drawback of  $N$ -multiplexer problems is the simplistic output. The multiplexer function defines just two possible outcomes: true (1) or false (0). In reality, single-step problems rarely have the same boolean output. To be realistic the output should provide a quantitative or qualitative value with a preference for continuous unbounded realization. This requires more complex test functions in term of the output. To get a more realistic test environment one can use a  $k$ -bits mapping.

The  $k$ -bits mapping is a binary function  $f(N_2)$  of length:

$$N = k + d, \quad (2.20)$$

where  $k$  is an integer defining the length of mapping and  $d$  is an integer defining the number of control bits. Every bit in a bit-string takes the value 0 or 1. The first  $k$  bits define the output of mapping or the value of the binary function. In a regular case the mapping is "one" to "one", i.e. the exact value observed within  $k$  bits is mapped into the output. In a more complex settings the mapping is a transformation from base 2 to base 10.

**Example 13.** *In the 6-bits mapping problem ( $N = 6$ ), the input to the system consists of a string of six binary digits. With  $k = 6$ , the first 6 bits are used for mapping. Let the realization of the 6-bits mapping be  $100010_2$ . Then the value of the binary function is  $f(100010_2) = 100010_2$  for the regular case, and  $f(100010_2) = 34_{10}$  for the complex case.*

To control the generalization ability of LCS there are  $d$  residual bits. They do not have any influence on the value of the function. A classifier system with good generalization ability should recognize those bits as noise and mark them with "don't care" symbol #.

**Example 14.** *In the 4-bits mapping problem ( $N = 6$ ) with  $k = 4$ , the first 4 bits are used for mapping and the residual 2 bits are used for generalization control. In the realization of the 4-bits mapping of  $100010_2$ , the value of the function is  $f(100010_2) = 1000_2$  for the regular case and  $f(100010_2) = 8_{10}$  for the complex case. A classifier system should recognize the position of noise bits. In this example bit 5 and 6 should be marked with "don't care" symbol #.*

<sup>39</sup>An increase of a bit string with noise bits does not affect complexity, since the maximum number of maximally general classifiers stays unchanged.

To use a  $k$ -bits mapping as a test, the length of the binary string,  $N$ , is selected and fixed. On each iteration a random binary string of  $N$  bits is generated and used as the input to the classifier system. The classifier system generates a response which is compared to a correct value of the binary function. The success of learning is reported back to the classifier system before the next iteration.

The complexity of  $k$ -bits mapping is defined by the number of maximally general classifiers,  $2^k$  maximally general classifiers.<sup>40</sup> To be maximally general a classifier should have at the end of condition part  $d$  bits filled with "don't care" symbol #.

**Example 15.** *In the 3-bits mapping problem ( $N = 6$ ,  $k = 3$ ) the input space can be completely covered by  $2^k|_{k=3} = 8$  maximally general classifiers:  $000### : 000_2/0_{10}$ ,  $001### : 001_2/1_{10}$ ,  $010### : 010_2/2_{10}$ ,  $011### : 011_2/3_{10}$ ,  $100### : 100_2/4_{10}$ ,  $101### : 101_2/5_{10}$ ,  $110### : 110_2/6_{10}$ ,  $111### : 111_2/7_{10}$ .*

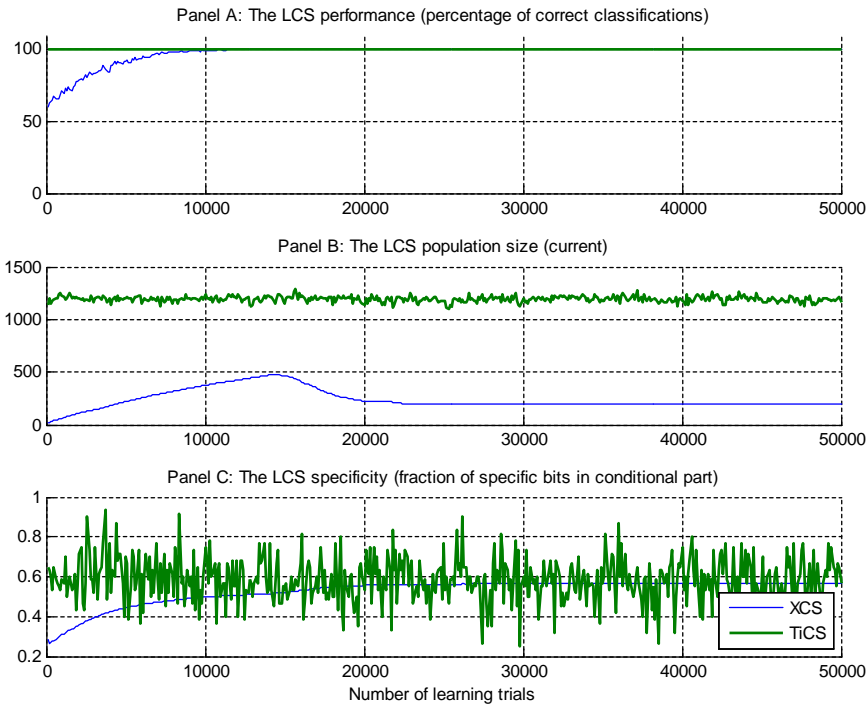


Figure 2.16: XCS vs. TiCS in a 6-multiplexer problem

For XCS parameter selection and their explanation see Subsection 2.3.4. The TiCS has  $N_g = 10^4$  and  $N_s = 25$ . For the explanation of its parameters see Subsection 2.3.5.

On Panel A one can see a chart of performance, measured as a percentage of correct classifications over a sliding window of 100 trials. On Panel B an active population size in each system is shown. The maximum population size is restricted to  $10^4$  classifiers. On Panel C it is depicted a specificity of the best classifiers, measured as a fraction of specific bits to the length of the condition part. All results are averaged over 10 runs.

<sup>40</sup>In contrast to the multiplexer problem the complexity of the mapping problem can be changed without affecting the length of the bit-string,  $N$ .



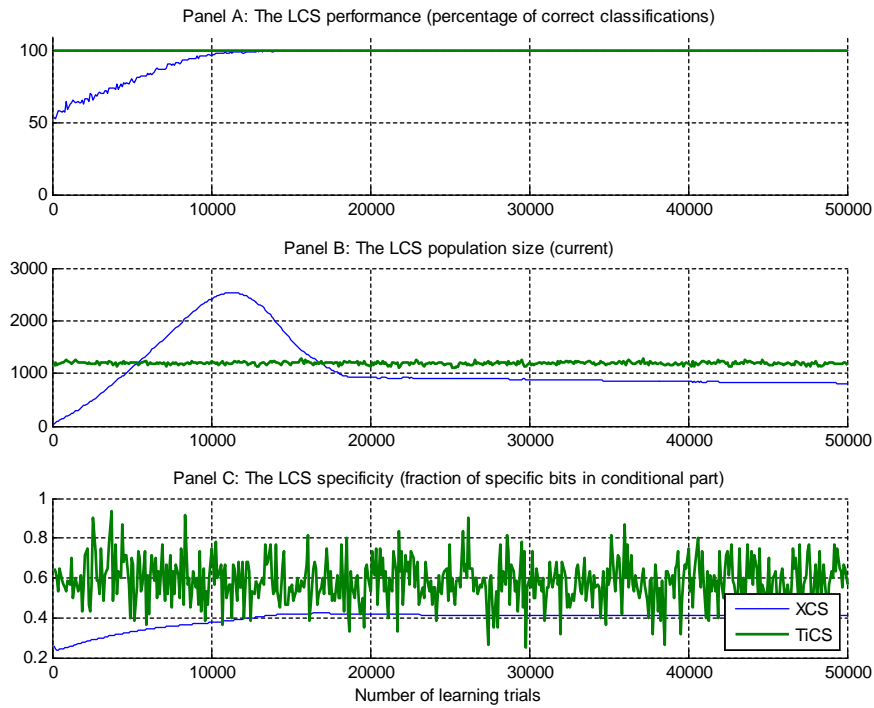


Figure 2.17: XCS vs. TiCS in a 11-multiplexer problem

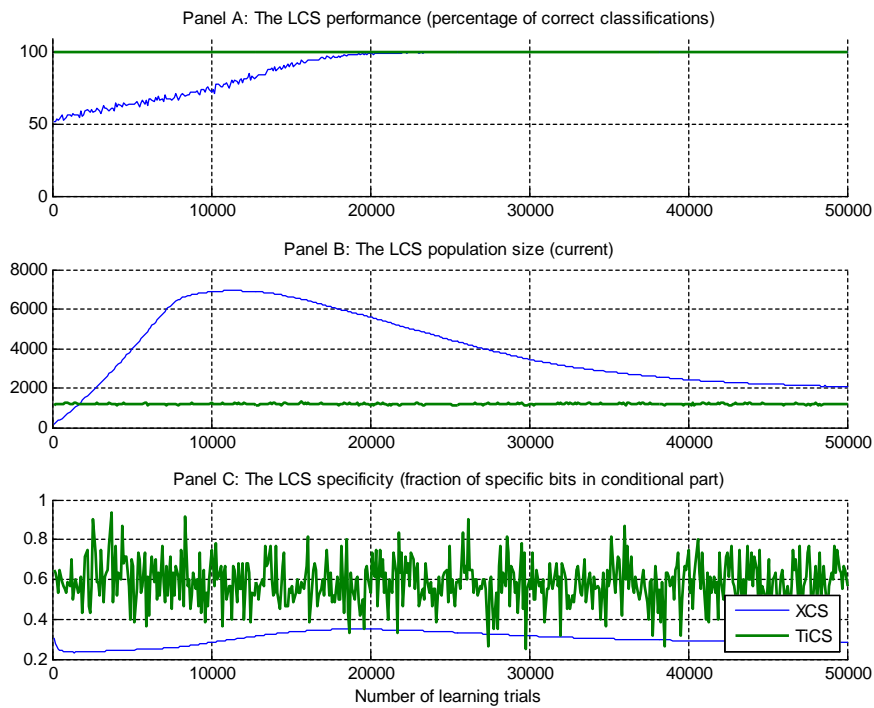


Figure 2.18: XCS vs. TiCS in a 20-multiplexer problem

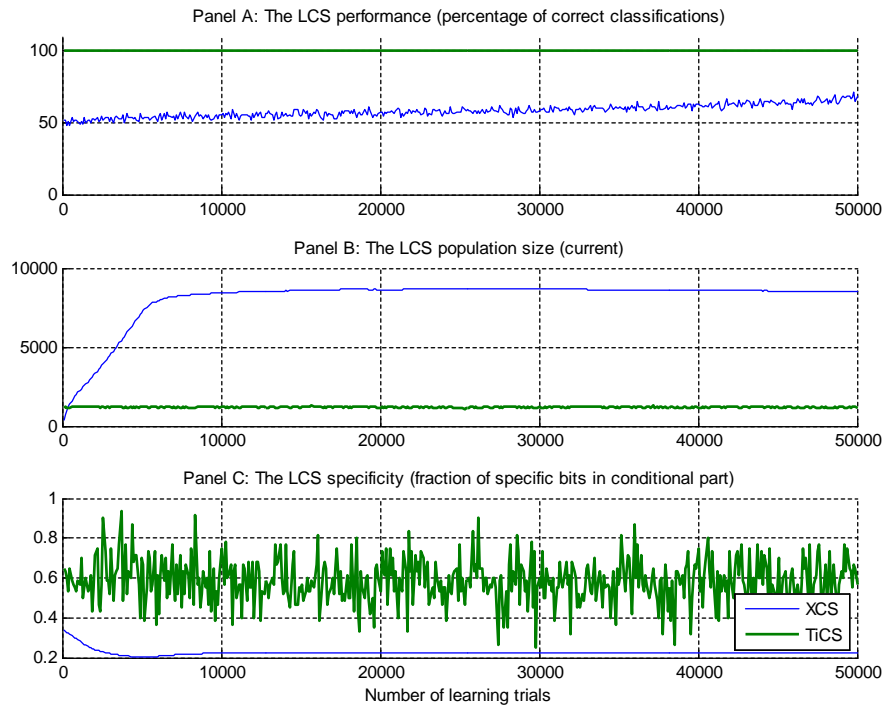


Figure 2.19: XCS vs. TiCS in a 37-multiplexer problem

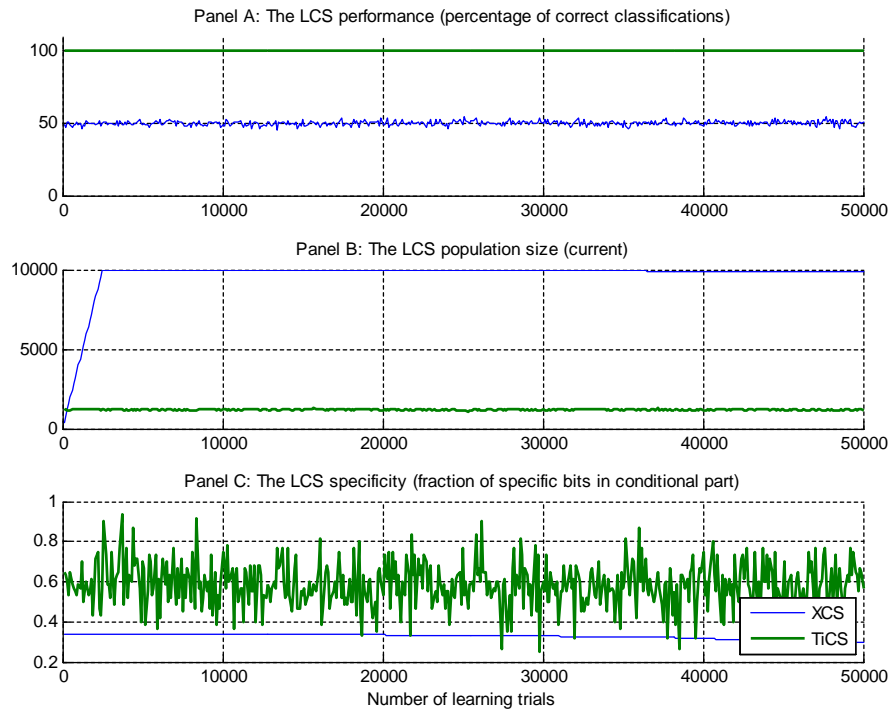


Figure 2.20: XCS vs. TiCS in a 70-multiplexer problem

**2.3.6.4. XCS vs. TiCS: Results.** The test is based on a  $N$ -multiplexer problem. Multiplexer problems are considered to be challenging for machine learning. In the LCS community the multiplexer problem has a celebrity position, since LCS proved to be superior in multiplexer learning relative to other machine learning techniques [64].

There are five specifications selected for the  $N$ -multiplexer problem. The classifier systems should learn the 6-, 11-, 20-, 37- and 70-multiplexer functions. To insure the robustness of results all simulations were repeated 10 times. Figures 2.16-2.16 report the results averages over 10 simulations. Parameter selection for the XCS is presented in Table 2.11, Subsection 2.3.4. For the TiCS the parameters are  $N_g = 10^4$  and  $N_s = 25$ . The maximum population size of both classifier systems is restricted to  $10^4$  classifiers.

The chart analysis of performance<sup>41</sup>, LCS population size<sup>42</sup> and LCS specificity<sup>43</sup> shows a clear dominance of the TiCS in solving the multiplexer problem of any considered difficulty. It achieves the maximum 100% performance steadily within the first 500 iterations. As for the XCS, it is able to achieve 100% performance within  $1 \cdot 10^4$ ,  $1.5 \cdot 10^4$  and  $2 \cdot 10^4$  iterations for the 6-, 11- and 20-multiplexer problems, respectively. For the 37- and 70-multiplexer problems the XCS is not able to achieve even 75% performance within  $5 \cdot 10^4$  iterations.

From the computational resources allocation the TiCS dominates as well. For the selected specifications it requires at maximum  $1.5 \cdot 10^3$  classifiers in the active population. Unlike the TiCS, the XCS has a varying size of the active population. In the 6- and 11-multiplexer problems it peaks at  $0.5 \cdot 10^3$  and  $2.5 \cdot 10^3$ , respectively. Then the population diminishes to  $0.25 \cdot 10^3$  and  $0.8 \cdot 10^3$  classifiers. With the 20- and 37-multiplexer problems the XCS requires at least  $6 \cdot 10^3$  and  $9 \cdot 10^3$  classifiers, respectively. For the 70-multiplexer problem the XCS needs more than  $1 \cdot 10^4$  classifiers.

From the specificity perspective the TiCS performs slightly worse than the XCS. Besides, in comparison to the TiCS, the XCS specificity has low volatility, which is attributed to averaging in the reported XCS specificity values.

To sum up, the TiCS dominates in two the most important measures of performance. It is able to learn the multiplexer problem in the shortest time. The results are robust to increasing complexity of the problems. The computational resources allocation is steady and highly efficient. The TiCS should be the LCS implementation of choice to build the ATA.

To finalize the choice of the TiCS it should be tested in learning more complex problems than the  $N$ -multiplexer. The multiplexer problem has only a boolean function output, which does not resemble the complexity of real-world problems. To overcome this limitation a new test problem was introduced - a  $k$ -bits mapping. Figure 2.21 illustrates simulation results of learning the 6-bits mapping. The simulations are performed for the population

<sup>41</sup>*Performance* measures a percentage of correct classifications over the last 100 iterations.

<sup>42</sup>*Population size* measures a number of active classifiers in the overall population.

<sup>43</sup>*Specificity* measures a fraction of specific bits in condition part of the best classifier relative to the overall length of condition part. It shows the ability of LCS to filter out irrelevant information in the information flow.

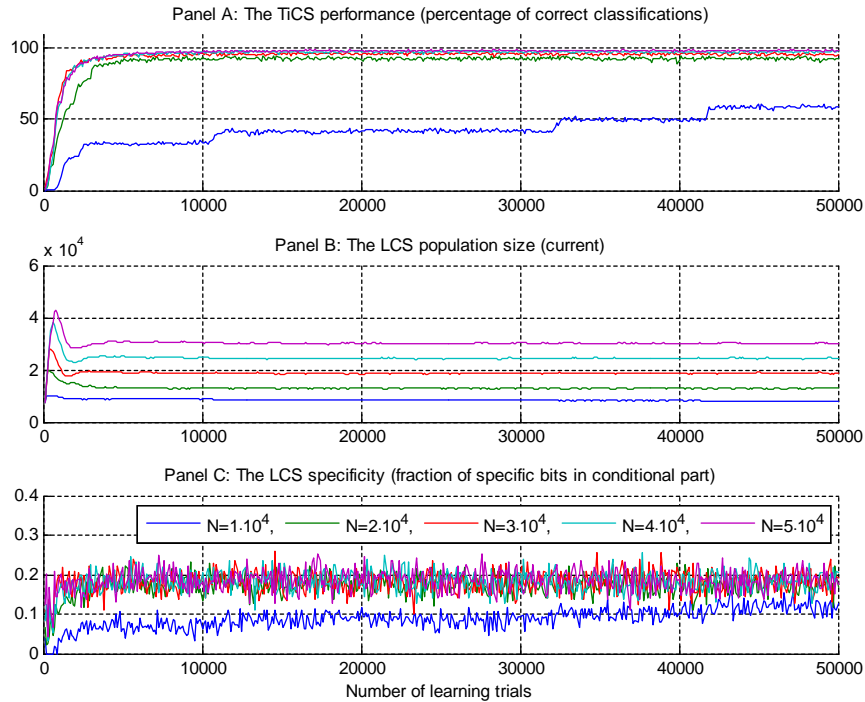


Figure 2.21: TiCS in learning k-bits mapping

Each panel on the overlay chart shows the results for the TiCS with the maximum population size of  $1 \cdot 10^4$ ,  $2 \cdot 10^4$ ,  $3 \cdot 10^4$ ,  $4 \cdot 10^4$  and  $5 \cdot 10^4$  classifiers.

On Panel A one can see a chart of performance, measured as a percentage of correct classification over a sliding window of 100 trials. On Panel B an active population size in each system is shown. On Panel C it is depicted a specificity of the best classifiers, measured as a fraction of specific bits to the length of the condition part. All results are averaged over 10 runs.

size of  $1 \cdot 10^4$ ,  $2 \cdot 10^4$ ,  $3 \cdot 10^4$ ,  $4 \cdot 10^4$  and  $5 \cdot 10^4$  classifiers. Reported results are averaged over 10 runs.

The performance of the TiCS is almost identical except for the case when the population size is  $1 \cdot 10^4$  classifiers. Here, the performance is increasing in a stairs-like manner. It reaches 50% within the first  $3.2 \cdot 10^5$  iterations. At the same time efficiency of filtering out irrelevant information is the highest for the small population size. In the rest of the cases the performance is quickly reaching 100% within the first  $0.5 \cdot 10^5$  iterations.

The overall conclusion stays intact. The TiCS should be used for building the ATA.

### 2.3.7. TiCS Parameter Selection

This subsection describes the TiCS parameter selection for the application to the ATA. As the TiCS will be responsible for detecting patterns in time series the parameter selection should be performed in the corresponding environment. The first part presents

pattern-recognition problems. The second part report results of testing the TiCS in this environment.

**2.3.7.1. Pattern-Recognition Problems: Introduction.** In this project LCS are intended to be used for time series pattern recognition. The pattern recognition in a time series is a process of identification of statistically significant patterns. The pattern is understood to be a sequence of observations which identify some graphical form like, for example, a straight line, convex or concave curve and so on. Statistical significance manifests in the possibility to forecast the value of the last observation in the identified pattern through the knowledge of constituent values prior to the forecast.<sup>44</sup>

To test LCS in this environment a time series should contain well pronounced patterns. A class of periodic functions can produce a required time series. In the first step a simple sine function is used to test LCS. Once it has been successfully learned, in the second step a more complex function is implemented in the test. It is designed to produce more diverse patterns. The problem complexity is augmented by a linear trend, periodic amplitude and variable phase. In both cases the functions are complemented by slight random noise to insure the robustness of LCS results.

To verify the legitimacy of the pattern recognition results a test on a time series generated by the random walk process is required, since high performance in learning predetermined patterns can be attributed to a mistake in programming or a conceptual problem in LCS implementation. To rule out this scenario the classifier system should be applied to a time series with no deterministic patterns. One way to generate such time series is to use the random walk data generating process. It creates time series by adding to the previous observation some stochastic increment. In this environment no or low performance in detecting data patterns is an indication of a problem-free implementation of LCS.

**2.3.7.2. Pattern-Recognition Problems: Periodic Functions.** To test LCS in terms of pattern recognition there are several test problems, where periodic patterns are simulated by simple periodic functions. To get diverse cases the problems are differenced by the presence of a trend, variation in amplitude and variation in phase. The presence of a noise component allows to check the robustness of LCS.

The first periodic function has the following form:

$$y(t) = \sin(0.1 \cdot t) + \varepsilon, \quad (2.21)$$

where  $t$  is the time and  $\varepsilon \sim N(0, 0.1)$  is a stochastic component. Panel A in Figure 2.22 shows a sample realization of this periodic function. The solid line outlines the function with added random noise. The dotted line depicts the underlying periodic function.

---

<sup>44</sup>One way to see it is to imagine a linear pattern. With a clear established knowledge of the identified pattern, one can easily fill any gaps within the pattern. The last observation in the pattern can be filled in the same way as long as the position of the pattern end is clearly identified. The pattern end is determined by the significance of the pattern. For a statistically significant pattern its end position is well identified, and thus, one can easily extrapolate the value.

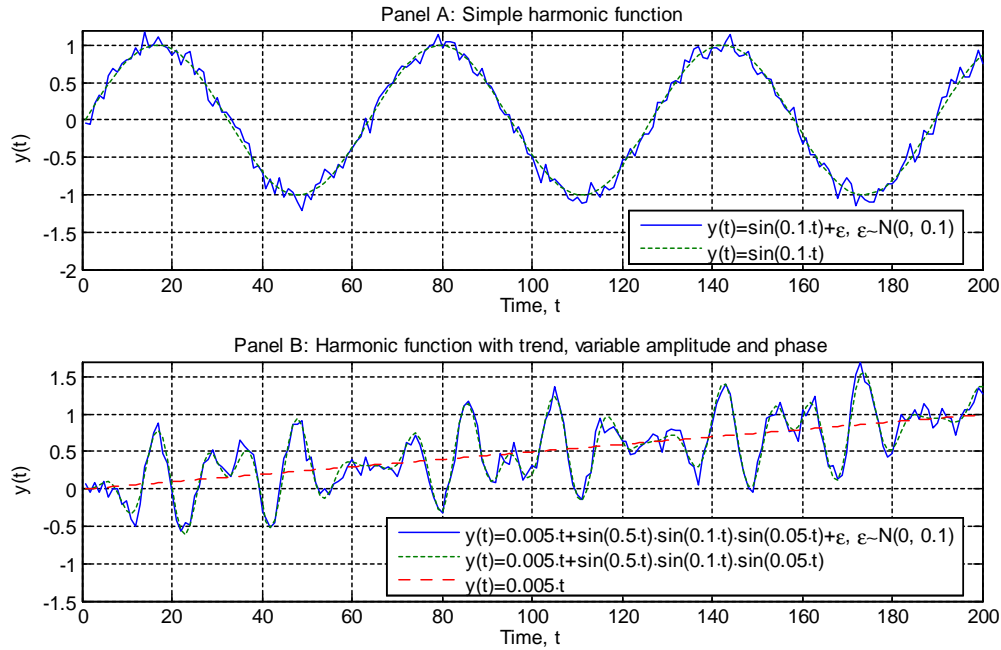


Figure 2.22: Periodic functions

The second periodic function has a more complicated structure. It has the following functional form:

$$y(t) = 0.005 \cdot t + \sin(0.5 \cdot t) \cdot \sin(0.1 \cdot t) \cdot \sin(0.05 \cdot t) + \varepsilon. \quad (2.22)$$

First, it includes the trend component,  $0.005 \cdot t$ . Second, the amplitude and the phase of periodic oscillations are varying with the time. Panel B in Figure 2.22 shows a sample realization of this periodic function. The solid line outlines the function with added random noise. The dotted line depicts the underlying periodic function. The dashed line reflects the trend.

To test the pattern recognition ability of LCS a time series is generated with a selected periodic function. The ability to recognize patterns in a time series is equivalent to the ability to predict a next value after observing some previous values. A classifier system requires some fixed length window of recent observations to make a prediction. So the length of generated time series should be at least of the length of the window. With the generated time series on each time iteration,  $t$ , a subsample of recent observations, excluding the one at time  $t$ , is used as an input to the classifier system. The classifier system generates a response as an expected value which is compared to a correct value realized at time  $t$ .<sup>45</sup> The success of learning is reported back to the classifier system before the next time iteration.

<sup>45</sup>Depending on a mechanism for time series encoding, the output can be a scalar or boundaries of the interval in which the value at time  $t$  is expected to be.

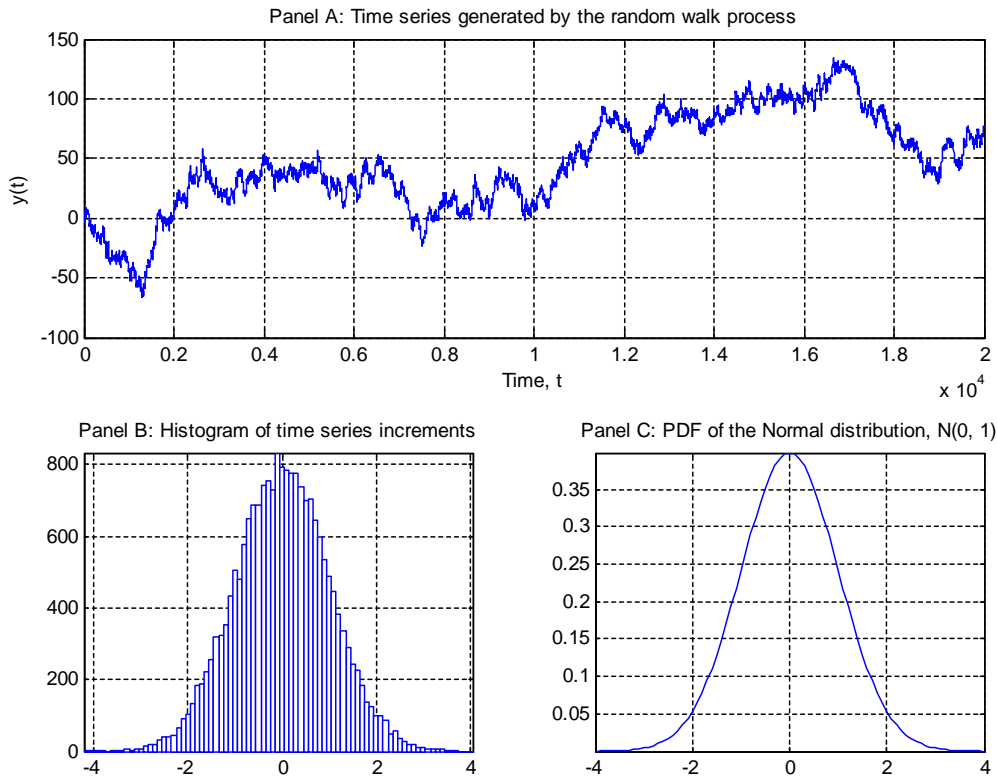


Figure 2.23: Sample time series generated by the random walk process

**2.3.7.3. Pattern-Recognition Problems: Random Walk Time Series.** To rule out the case when high performance of LCS in determining patterns in a time series is attributed to programming or implementation error the following test is proposed. A time series is generated by adding random increments to the last observed value. Increments are assumed to be stochastic and drawn from the Normal distribution:

$$y(t) = y(t-1) + \varepsilon, \quad (2.23)$$

where  $\varepsilon \sim N(0, 1)$  is a stochastic component.<sup>46</sup> Panel A in Figure 2.23 gives a sample realization of a time series generated by the random walk process. Panels B and C show that the distribution of the time series increments is identical to the theoretical Normal distribution  $N(0, 1)$ . Since the data generating process is stochastic the application of problem-free LCS on this time series should result in no or low performance to detect patterns in the data.

**2.3.7.4. TiCS in Pattern-Recognition Problems.** To test the TiCS for pattern recognition, two types of time series are selected. The first type represents the family of periodic functions, where some predetermined periodic pattern is masqueraded with a noise, trend

<sup>46</sup>This data-generating process is known as the random walk.

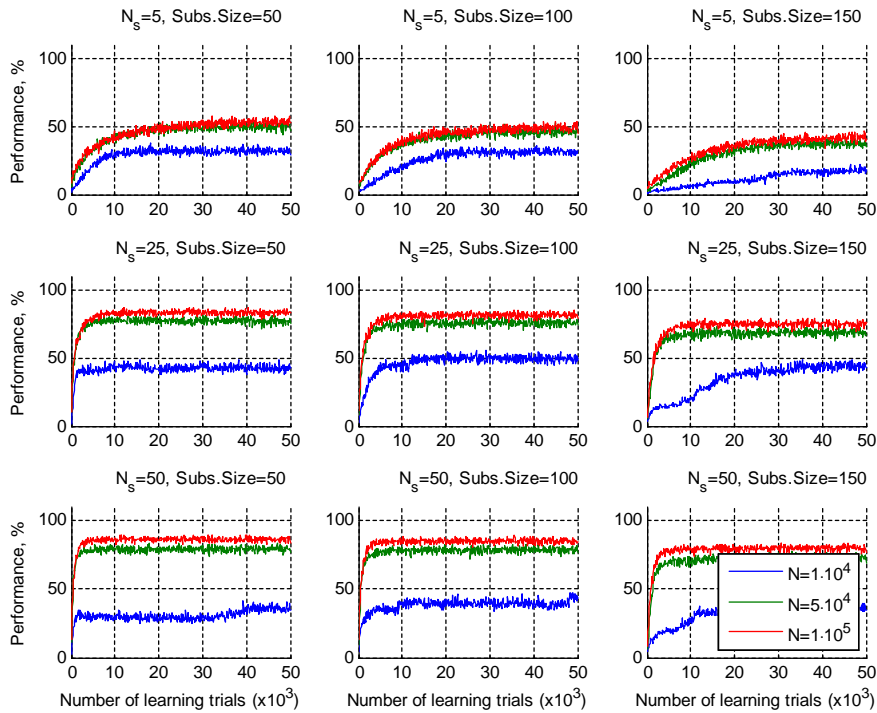


Figure 2.24: Performance of the TiCS in learning simple periodic function

The overlay charts show the results for the TiCS with the maximum population size of  $1 \cdot 10^4$ ,  $5 \cdot 10^4$  and  $1 \cdot 10^5$  classifiers. In each chart panels are arranged to see the impact of an increase in the subpopulation size (from the top to the bottom row: 5, 25 and 50 classifiers), and of an incremental increase in the subsample size (from the left to the right column: 50, 100 and 150 observations). All results are averaged over 10 runs. The

measure of performance is a count of correct interval predictions within last 100 iterations. The TiCS uses the Bottom-Up algorithm for piecewise linear approximation.

The level and time stamp are described by 5 bits each, which corresponds to the minimum width of a prediction interval to be of 0.125 for normalized subsample. For levels normalization brings the mean to zero and variance - to one. As for the time stamps a subsample is assumed to be within  $[-2, 2]$ , where -2 is assigned to the first observation in the subsample and 2 - to the last one.

and changing amplitude. A grid search of the TiCS parameters should maximize the performance of the TiCS in detecting and identifying those patterns. The second type of the test time series is generated by the random walk process. It is used to verify the legitimacy of the pattern-recognition results, since the high performance in learning predetermined patterns can be attributed to a mistake in programming or a conceptual problem in the TiCS implementation. To rule out this scenario the TiCS should be tested on a time series with no deterministic patterns. In this environment no or low performance in detecting data patterns will an indication of a problem-free implementation of the TiCS.

In the pattern recognition the TiCS has only three specification parameters, i.e. the global population size  $N_g$ , the subpopulation size  $N_s$  and the subsample size.<sup>47</sup> The grid

<sup>47</sup>The *subsample size* defines how many recent observations should the TiCS use for pattern recognition.



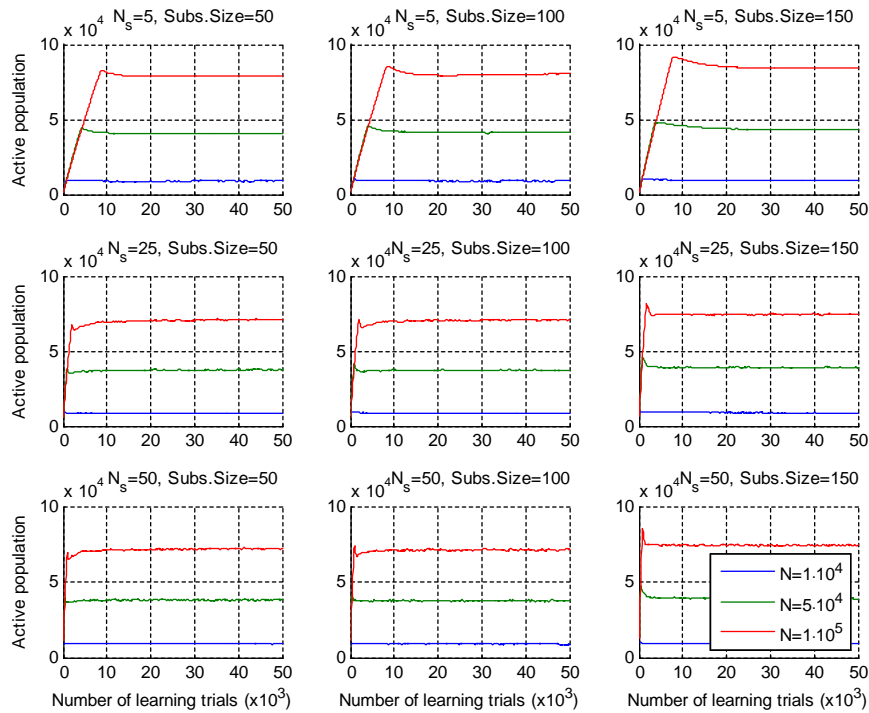


Figure 2.25: Active population size of the TiCS in learning simple periodic function

search is performed along these three parameters. For the global population size the grid includes  $1 \cdot 10^4$ ,  $5 \cdot 10^4$  and  $1 \cdot 10^5$  classifiers. For the subpopulation size it includes 5, 25 and 50 classifiers. For the subsample size it includes 50, 100 and 150 observations. Ten simulations were performed for each combination of the parameters. Figures 2.24-2.28 present the averaged results.

In learning simple and complex periodic functions the TiCS gained the maximum performance of 85% within  $4 \cdot 10^4$  iterations and 75% within  $6 \cdot 10^4$  iterations, respectively.<sup>48</sup> It is achieved with  $N_g = 1 \cdot 10^5$ ,  $N_s = 25$  and the subsample size of 50 observations. The same parameter selection provides the most efficient computational resources allocation.<sup>49</sup>

The pattern recognition in the time series generated by the random walk process achieved 40% performance of the TiCS.<sup>50</sup> This is a clear indication of the legitimacy of the previous results with the TiCS. The positive value of performance can be explained by the TiCS detection of short local trends present by chance in the time series. The argument is supported by the fact that the maximum performance was achieved with the smallest subsample size of 50 observations.

<sup>48</sup>The measure of performance is a count of correct interval predictions within the last 100 iterations.

<sup>49</sup>The comparison is performed for the fixed value of  $N_g = 1 \cdot 10^5$  and varying  $N_s$  and the subsample size.

<sup>50</sup>The same measure of performance is used here.

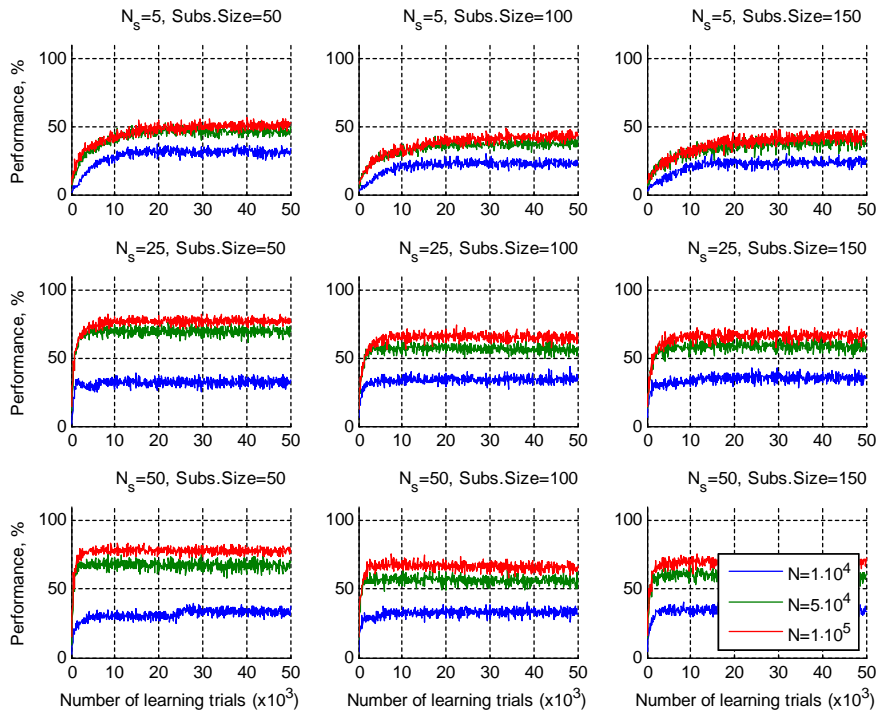


Figure 2.26: Performance of the TiCS in learning complex periodic function

In comparison to  $ARMA(1, 1)$  model the TiCS has clear dominance since for the same complex periodic function  $ARMA(1, 1)$  achieved only the performance of 5% (see Figure 2.29) while the TiCS gets the performance of up to 80% (see Figure 2.26).<sup>51</sup>

The results of the parameter selection are as follows. The TiCS can detect patterns in time series. The performance of pattern recognition increases with an increase in the global population size. With  $N_g = 1 \cdot 10^5$  the performance reaches 85% for the simple periodic function. The optimal parameters are  $N_s = 25$  and the subsample size of 50 observations.

## 2.4. Discussion

This chapter presented the design of the key components of the artificial technical analyst. Following the recommendations, derived in the first chapter, the ATA include the data-preprocessing mechanism and pattern-recognition engine comparable with experienced technical analysts. Data preprocessing incorporates time series homogenization and piecewise linear interpolation algorithms, while pattern-recognition is built upon learning classifier systems.

The first chapter concluded that to correctly resolve dispute between practitioners and academicians on efficiency of financial markets one needs to design a model-free autonomous

<sup>51</sup>One should note that performance was measured as a count of correct interval predictions within the last 100 iterations. Since  $ARMA$  model produces point estimates its results were modified to be comparable with the TiCS. For this an interval equal to the variance was divided into bins, where their number is identical to the number of minimum bins of the TiCS. A bin within which  $ARMA$  model estimate is located defines an interval prediction of  $ARMA$  model.

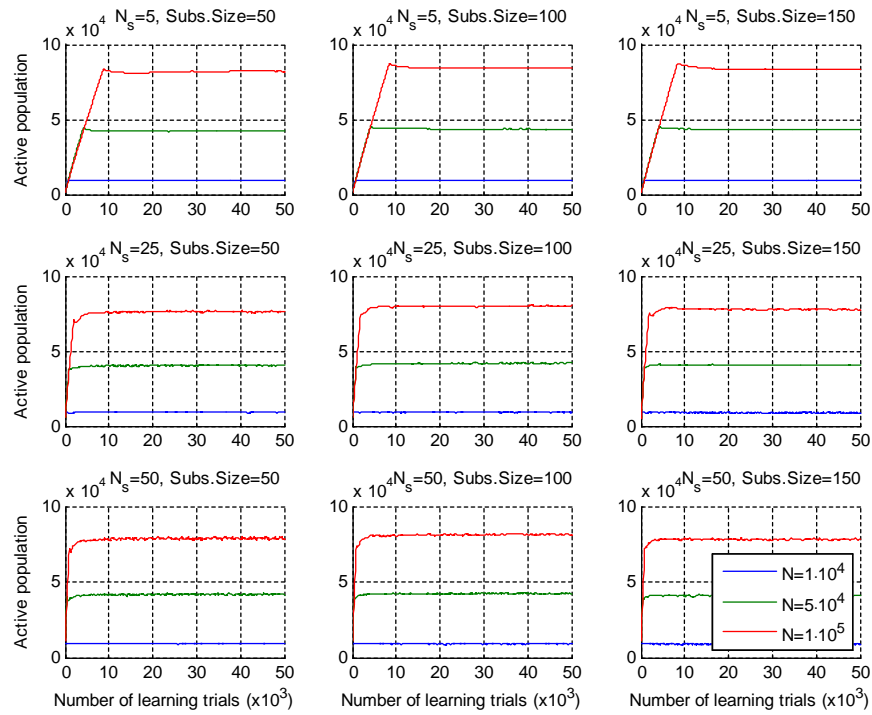


Figure 2.27: Active population size of TiCS in learning complex periodic function

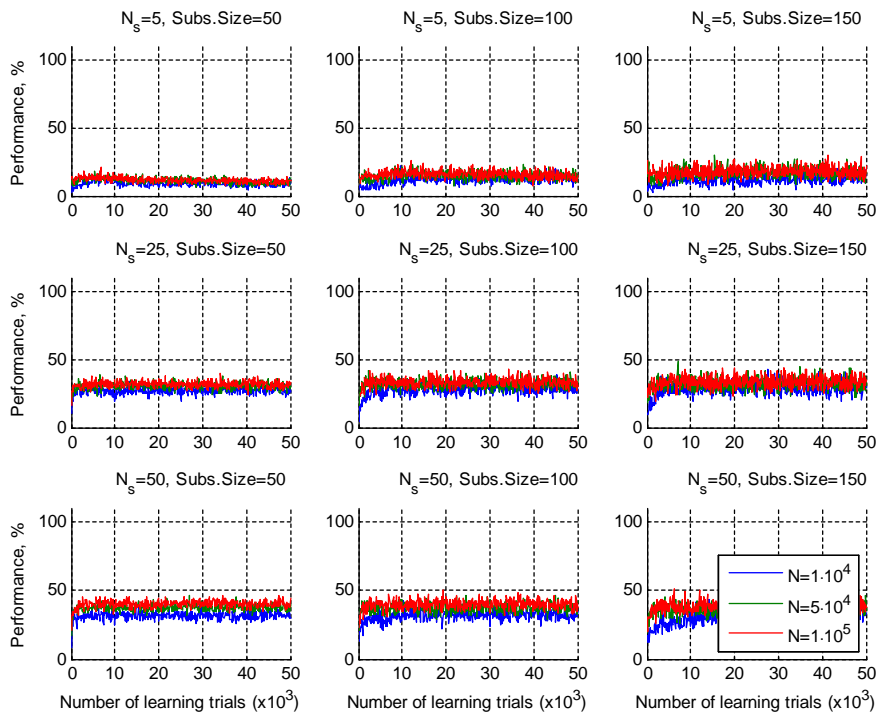


Figure 2.28: Performance of the TiCS in learning a random walk process

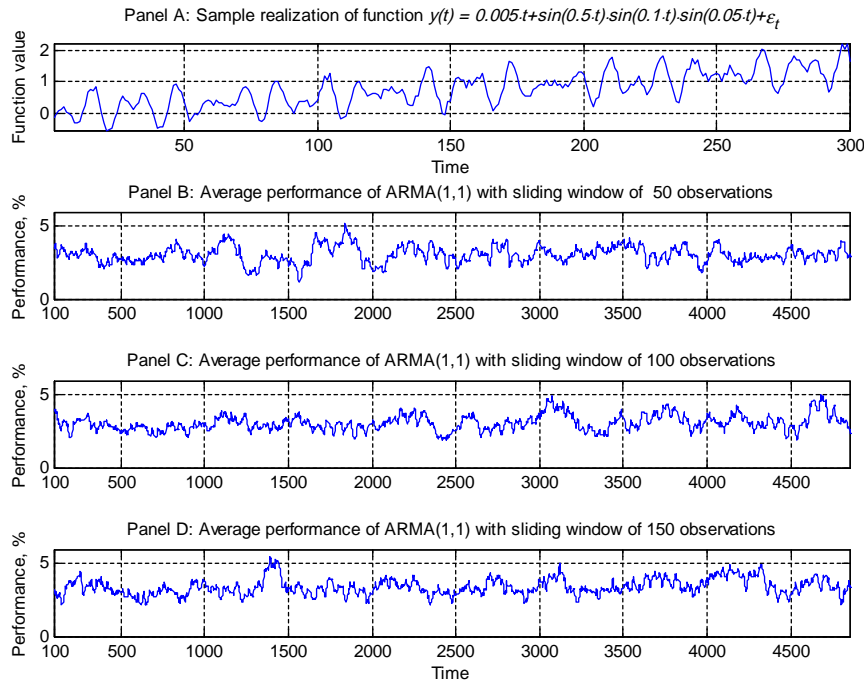


Figure 2.29: Performance of ARMA(1, 1) in learning complex periodic function

mechanism. Following the idea of Spyros Skouras [227] to test market efficiency independently of equilibrium model one can use the artificial technical analyst. The blueprint of the ATA has two major components: data preprocessing and pattern recognition mechanism.

The first component of the ATA is data preprocessing mechanism. To minimize the interference of errors due to outsourcing of data preprocessing, raw data were chosen as an input. At the first stage data should be homogenized. This step allows in the future to reduce complexity of pattern-recognition problem, where one needs to focus only on price levels observed at a predefined frequency. Given two methods of homogenization - linear- and previous-tick interpolation, - the latter method was selected due to its computational simplicity and similarity of the results of both methods.

In the original form the homogenization has low efficiency of data utilization, which is crucial in consequent pattern-recognition. Since the pattern-recognition requires as much data as possible, one can augment the homogenization procedure with time shifts. As a result the homogenization produces several diverse time series, which have an identical underlying distribution. This allows to maximize efficiency of data utilization and allows the ATA to learn diversity of patterns.

The homogenization requires exogenous specification of its frequency. The grid search was used to pick the frequency, which maximizes the informational content of homogenized data. To proxy the quantity of informational content the criterion of profitability was used. That is the ATA was fed with the data homogenized at different frequencies. Obtained

results were used to derive breakeven transaction costs. The frequency, which provided the highest value of breakeven transaction costs, was chosen for homogenization. The same approach was used for selecting the optimal subsample size, which the ATA needs for effective operation. Preliminary results show that for the medium- and high-liquidity pools of stocks the optimal subsample size is 50 observations, which was observed at 6 and 3 minute frequencies of homogenization, respectively. For the low-liquidity pool the optimal subsample size is 110 observations, which was observed at 3 minute frequency.

The second component of the ATA starts with the piecewise linear approximation mechanism. This mechanism serves to identify important peaks and troughs in time series of prices. To pick the best implementation several competing alternatives were tested on artificial data with the Monte Carlo studies. The test included sliding windows, top-down, bottom-up and GA-based algorithms. Due to their high speed and acceptable level of approximation the bottom-up algorithm have been selected for application in the ATA.

The last component of the ATA is responsible for pattern recognition. To match the abilities of professional technical analysts the cognitive mechanism of the ATA is based on artificial intelligence. The choice of learning classifier systems is explained by their ability to present transparent structure of solutions. Since the best implementation of learning classifier systems does not have required power to solve such a complex problem within a reasonable time, the new implementation of LCS was introduced - the TiCS. In a series of test problems it was shown that the TiCS has indeed superior performance. In additional experiments its optimal input parameters were selected. The TiCS has the best performance in detecting non-linear patterns with the subsample size of 50 observations and the subpopulation size of 25.

The fusion of these components constitutes the ATA, which will be used in the empirical study of the Tokyo Stock Exchange data to show the technical analysts' perception of the EMH.

## CHAPTER 3

### **Application of the ATA on The Tokyo Stock Exchange Data**

The aim of this work is to show how does a technical analyst perceive the EMH. Its necessity comes from an existing conflicting views on validity of the efficient market hypothesis and applicability of technical analysis.

Technical analysis<sup>1</sup>, i.e. detection of patterns in time series of prices, is a direct contradiction to the weak form of the EMH. This hypothesis is a corner stone of most theories of financial markets. The EMH postulates that in weakly efficient financial markets it is impossible to profit by predicting next period price developments based on time series of past prices or returns.<sup>2</sup> The contradiction turns technical analysis into an "outlaw" of financial theory separating the financial community into two camps: the camp of academicians, who advocate the EMH; and the camp of practitioners, who keep using technical analysis.

The overview of practices in financial industry shows that many investors, including professional traders and analysts, are widely using technical analysis [5]. Previous investigations demonstrate that under some conditions technical analysis can provide substantial returns.<sup>3</sup> Unfortunately, the findings are undermined by two limitations. Technical analysis is a complex method, which includes rigorous mathematical formulas as well as abstract visual patterns of time series. As a result, researches use only a fraction of the methodology, which dramatically constrains the potential of technical analysis. On the other hand, when the ability of an expert, who is knowledgeable of and fluent in all aspects of technical analysis, is studied one cannot distinguish whether the results are solely due to technical analysis application or a subjective interference of an analyst. Thus, the studies, up to now, lacked robustness and "unleashed" power of technical analysis.

The ATA overcomes these drawbacks. The ATA is a robust implementation of the technical analyst expertise.<sup>4</sup> It follows the aim of technical analysis to detect patterns in price time series. The cognitive mechanism of the ATA, the TiCS, uses the methodology of technical analysis to generate online forecasts of next period price developments in a dynamic environment. To insure the robustness and effectiveness of generated forecasts the ATA performs a sequence of data preprocessing steps. Investment decisions are generated to obtain aggregated return adjusted for transaction costs.

To verify that the extracted positive returns are not just a result of defective programming several benchmark trading strategies are applied to the same data set. The benchmark

---

<sup>1</sup>For the introduction to the technical analysis see Section 1.2.

<sup>2</sup>For the introduction to the EMH and its tests see Section 1.1.

<sup>3</sup>For the literature overview of technical analysis application see Subsection 1.2.5.

<sup>4</sup>For the introduction to the ATA and its components see Chapter 2.

trading strategies include a perfect forecast, random walk, random and buy-and-hold trading strategies.

Empirical results show that the ATA is able to extract positive aggregate returns. The benchmark strategies support the credibility of results, which show that under low transaction costs a technical analyst can perceive the market as weakly inefficient, while at higher transaction costs the perception of inefficiency vanishes, since there are no profitable opportunities.

In the process of calculation several assumptions are made. First, the impact of inflation, dividends and splits is disregarded, since at high frequency their impact is negligible. Second, the aggregate return is computed separately for every time series in the study. Third, trading strategies are applied at every period, i.e. at the beginning of each period the previous position should be closed. Correspondingly the aggregate return is a product of returns in all time periods. And finally, the decision-making process and calculation of returns take into account round-trip TC.<sup>5</sup>

In this study we analyze the time series of stock prices collected on the Tokyo Stock Exchange (TSE) over a time period from 11/03/1996 to 23/06/1998. The pool of time series covers 2273 stocks listed on the TSE. Considering computational intensity only 30 time series were selected for the study. For this all stocks were sorted into 3 pools according to their liquidity, which can be described as low-, medium- and high-liquidity pools of stocks. In each pool 10 stocks were randomly selected. Their corresponding time series of prices were used for computations.<sup>6</sup> Before reporting the computed results are averaged within each segment of liquidity.

This chapter is organized as follows. Section 3.1 presents a decision-making function and required informational base. Section 3.2 specifies the ATA decision-making function and presents results of the ATA empirical application. Section 3.3 presents the benchmark trading strategies and results of their application. Discussion of results in Section 3.4 concludes the chapter.

### 3.1. Decision-Making Function and Its Informational Base

This section presents a general form of decision-making functions and its informational base. With minor modifications the function is used in the ATA as well as in the benchmark trading strategies.

The informational base of the decision-making process includes two components. The first component are expected transaction costs. Transaction costs have a direct influence on trading decisions, since the value of extracted return should be sufficient to cover round-trip transaction costs and with a stochastic return distribution the probability of trading is higher under small transaction costs than under high transaction costs. Due to its hypothetical nature this empirical study uses a set of discrete values of transaction costs

<sup>5</sup>By design the random and buy-and-hold strategies do not take transaction costs into account during the decision-making process.

<sup>6</sup>For details on the TSE market microstructure and description of the TSE data see Appendix D.

taken in a grid from an interval  $[0, 2]$  of percent of transaction value with an increment of 0.001 percent.

The second component of the informational base is a subsample of recent prices observations. Technical analysis requires this subsample for forecasting the next period prices. Specifically to the ATA, it requires specification of the subsample size.<sup>7</sup> The ATA is able to identify relevant price observations and filter out the rest. But, in the case subsamples are too small or too big, the ATA does not have enough price observations to perform forecast or, respectively, it would take more time to detect the pattern than the duration of this pattern phenomena in the market. To cover this aspect the study uses a set of discrete subsample sizes taken in a grid from an interval  $[50, 250]$  of price observations with an increment of 10 observations.

Along with the two explicit components of the informational base for a decision making there is a third component - the frequency of decision making.<sup>8</sup> It has an implicit influence on outcomes of decision making. A low frequency allows traders to avoid frequent expenses on transaction costs. At the same time, in a volatile market it constrains traders to take a speculative advantage from local trends. The study uses a set of discrete frequencies taken in a grid from an interval  $[1, 10]$  of minutes with an increment of a minute.

In general, the process of decision-making is formalized through a function  $D()$ , which maps the arguments into the space of trading actions:

$$D(c, p_t, p_{t-1}, \dots, p_{t-n}) = \{B, S, N\}, \quad (3.1)$$

where  $c$  stands for transaction costs expressed in fractions of transaction value;  $\{p_t, p_{t-1}, \dots, p_{t-n}\}$  are current and lagged asset prices;  $B$  denotes the buy transaction;  $S$  is the short-selling; and  $N$  means to pass up trading.

Normally, the decision-making process includes two steps: forecasting of a next period prices and a maximization of a next period return adjusted for transaction costs. Forecasting is specific for each strategy since in every case it has specific forecast mechanism. In general forecasting provides an estimate of the next period asset price conditional on informational base,  $E[p_{t+1} | \{p_t, p_{t-1}, \dots, p_{t-n}\}]$ . The second step, return maximization, performs a search through the outcomes of possible trading actions to find a highest return. This step is based on the notion of the revealed transaction cost presented in Appendix C. Thus, the general form of the decision-making functions is as follows:

$$D_t(c, p_t, p_{t-1}, \dots, p_{t-n}) = \begin{cases} B & \text{if } c < \frac{E[p_{t+1}] - p_t}{E[p_{t+1}] + p_t}, \\ S & \text{if } c > \frac{p_t - E[p_{t+1}]}{p_t + E[p_{t+1}]}, \\ N & \text{if otherwise.} \end{cases} \quad (3.2)$$

For specific trading strategies the estimate of the next period asset price,  $E[p_{t+1}]$ , should be replaced in (3.2) with an appropriate equivalent.

<sup>7</sup>For more details on the analysis of the optimal subsample size see Section 2.1.4.

<sup>8</sup>For more details on the analysis of the optimal frequency of trading see Section 2.1.3.



### 3.2. Performance of ATA Trading Strategy

The ATA is a robust implementation of the technical analyst expertise. It pursues the aim of technical analysis to detect patterns in time series of prices. The cognitive mechanism of the ATA, the TiCS, uses the methodology of technical analysis to generate online forecasts of next period price development in a dynamic environment. Investment decisions are generated to obtain aggregated return adjusted for transaction costs.

The importance of the ATA comes from its ability to replicate the expertise in technical analysis in a robust way. The success of the ATA in extracting positive returns will allow to justify the methodology of technical analysis and eventually to incorporate it in a theory of financial markets.

Unlike the conventional forecast methods the ATA performs interval price forecasts.<sup>9</sup> This allows to convey the forecast of price developments and the ATA confidence in those forecasts.<sup>10</sup> The confidence is indicated by the difference between the upper and lower limits of the forecast. A high confidence is reflected in a small difference, while a low confidence is conveyed through a high difference. The smallest difference is defined to be one basic point in the market. The highest difference is infinity.

The ATA uses technical analysis to obtain the interval forecast, which is constrained by the upper and lower limits:  $\bar{p}_{t+1}$  and, respectively,  $\underline{p}_{t+1}$ .<sup>11</sup> Once the interval is defined one of the three possible trading actions is selected to maximize a current return. This is formalized in the ATA decision-making function:

$$D_t^I(c, p_t, p_{t-1}, \dots, p_{t-n}) = \begin{cases} B & \text{if } c < \frac{\underline{p}_{t+1} - p_t}{\underline{p}_{t+1} + p_t}, \\ S & \text{if } c > \frac{p_t - \bar{p}_{t+1}}{p_t + \bar{p}_{t+1}}, \\ N & \text{if otherwise.} \end{cases} \quad (3.3)$$

The aggregate return adjusted for transaction costs is calculated as:

$$R = \prod_{t=1}^T r_t, \quad (3.4)$$

<sup>9</sup>For details on the interval forecasts produced by the ATA see Subsection 2.3.5.1, and for examples of previous ATA application see Subsection 2.3.7.

<sup>10</sup>The ATA does not distinguish whether a low confidence is only due to a lack of experience in a new situation or/and due to a presence of risk associated with an underlying asset, since either case leads to uncertainty.

<sup>11</sup>These limits are the decoded action part suggested by the TiCS. Since the ATA operates with homogenized data only the limits of levels (in this case prices) are returned. In the future the ATA can be extended to work with irregularly spaced time series. In this case the forecast of the ATA will include additionally the time interval within which the next level is predicted.

where the time is defined on the interval from 0 to  $T$ . The return in each time period,  $r_t$ , is calculated as:

$$r_t(c, p_t, p_{t-1}) = \begin{cases} \frac{p_t - (p_{t-1} + p_t)c}{p_t - (p_{t-1} + p_t)c} & \text{if } D_{t-1}() = B, \\ \frac{p_{t-1} - (p_{t-1} + p_t)c}{p_t} & \text{if } D_{t-1}() = S, \\ 1 & \text{if } D_{t-1}() = N. \end{cases} \quad (3.5)$$

Figures 3.1 - 3.3 present the results of hypothetical application of the ATA trading strategy on the TSE data. The first column of each figure shows breakeven separation surfaces. Columns 2 - 3 display the slice views corresponding to the first column. Each row represents the time span over which the returns are aggregated. The separation surface shows at which transaction costs and at which combination of trade frequency and subsample size the application of the ATA trading strategy is breakeven. Any point above the surface indicates a loss, while any point below is a gain. On the slice view fat dots shows breakeven transaction costs at either fixed subsample size (column 2) or trade frequency (column 3). Color notation distinguishes the magnitude of gains (in shades of red) or losses (in shades of blue).

##	Stock pool liquid.	All time spans			Time span #1			Time span # 2			Time span #3		
		TC %	@ obs.	@ min.	TC %	@ obs.	@ min.	TC %	@ obs.	@ min.	TC %	@ obs.	@ min.
1	High	0.28	50	3	0.15	50	4	0.20	50	4	0.33	50	3
2	Medium	0.32	50	4	0.20	50	6	0.37	60	3	0.51	60	3
3	Low	0.25	50	4	0.25	50	4	0.31	50	4	0.32	50	2

Table 3.1: Application of the ATA trading strategy

All time spans cover data from 11/03/1996 till 22/06/1998. Time spans #1, #2 and #3 are 11/03/1996-13/12/1996, 14/12/1996-17/09/1997, and 18/09/1997-22/06/1998, respectively. Each time span on average covers 277 days.

Table 3.1 presents the summary of the figures. It lists the maximum breakeven transaction costs obtained for three groups of stock liquidity. Each value of breakeven transaction costs is accompanied by the frequency of trading and the subsample size at which it was observed.

The first conclusion is that the ATA can successfully apply technical analysis to extract positive returns adjusted for transaction costs. The best performance is achieved at trading frequency of 4 minutes, the subsample size of 50 observations, with the medium liquidity stocks. Here the revealed transaction costs are 0.32 percent of the transaction value. If

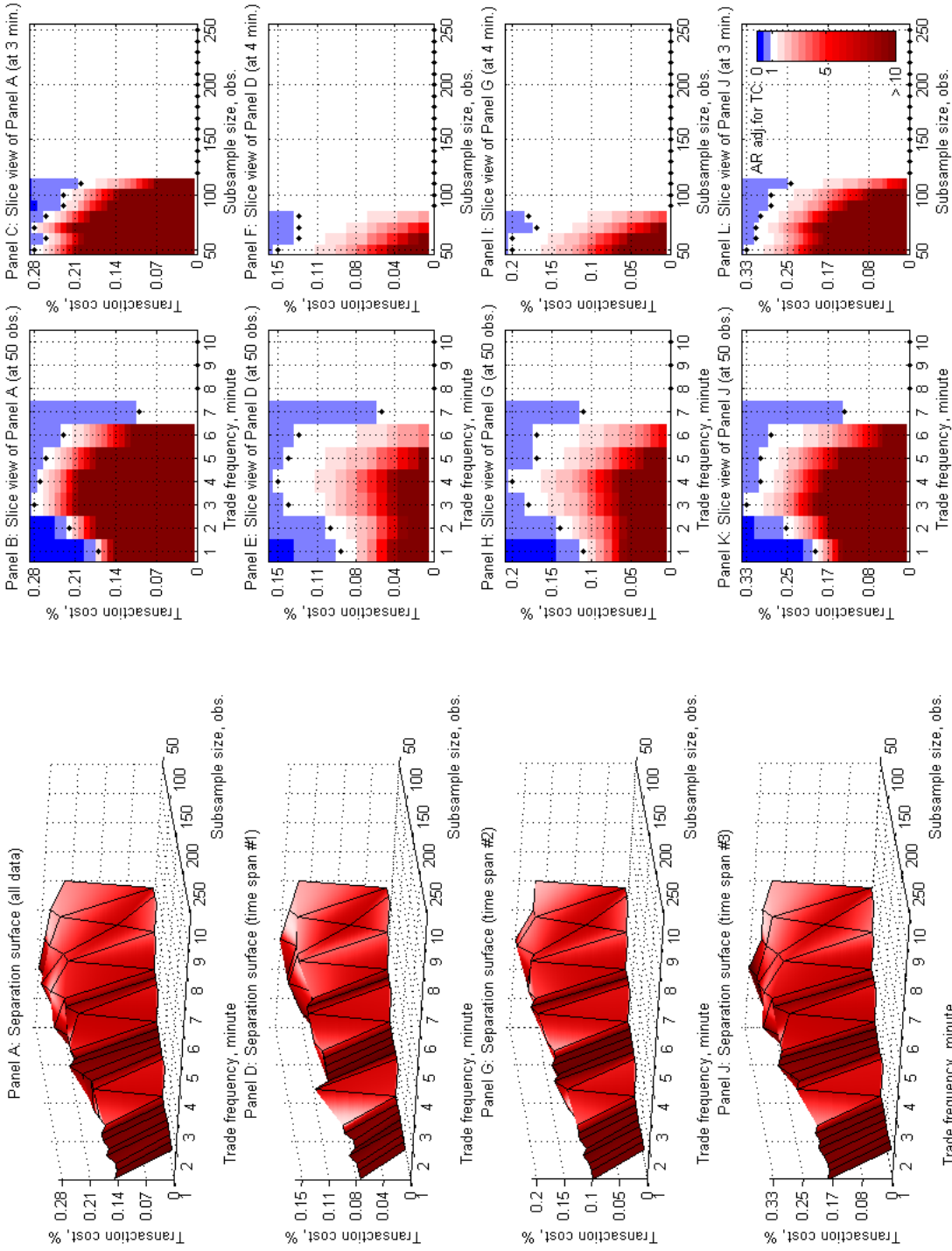


Figure 3.1: Application of ATA trading strategies on the pool of high liquidity stocks

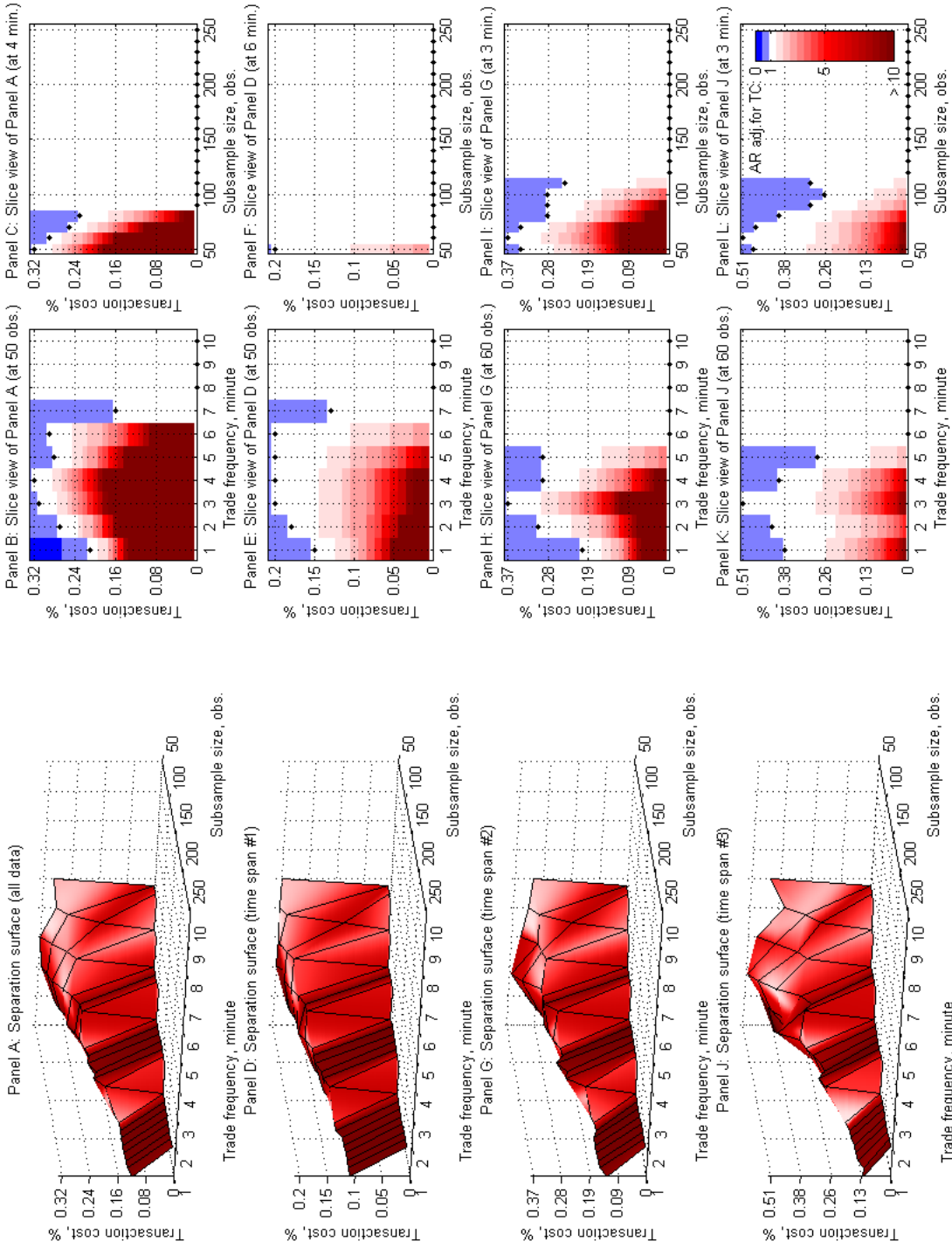


Figure 3.2: Application of ATA trading strategies on the pool of medium liquidity stocks

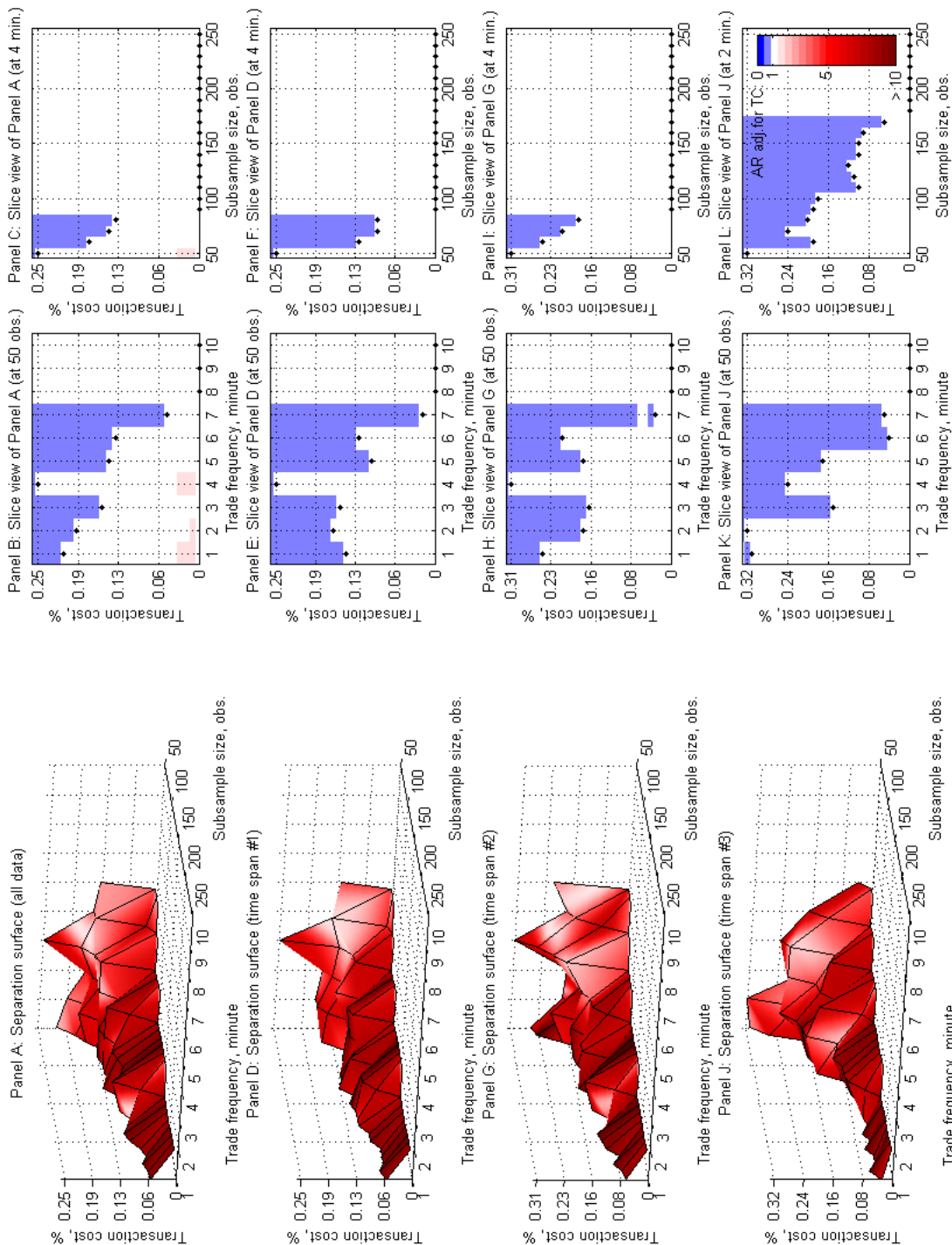


Figure 3.3: Application of ATA trading strategies on the pool of low liquidity stocks

under these settings the actual transaction costs would be strictly less than 0.32 percent of the transaction value, then a technical analyst would perceive this market segment at that time as weak-form inefficient.

The analysis of specific time spans reveals that the values of maximum breakeven transaction costs are increasing through the time spans.<sup>12</sup> This observation is valid for all liquidity pools. There are two possible explanations. The first one is that the data have different structure in each time span, which, in relation to the ATA forecasting, translates to different degree of predictability and different distribution of returns in each time span. The second explanation is a learning ability of the ATA. That is if the distribution of prices and the degree of predictability is the same for each time span, then the increase in the maximum breakeven transaction costs is due to the ATA learning.

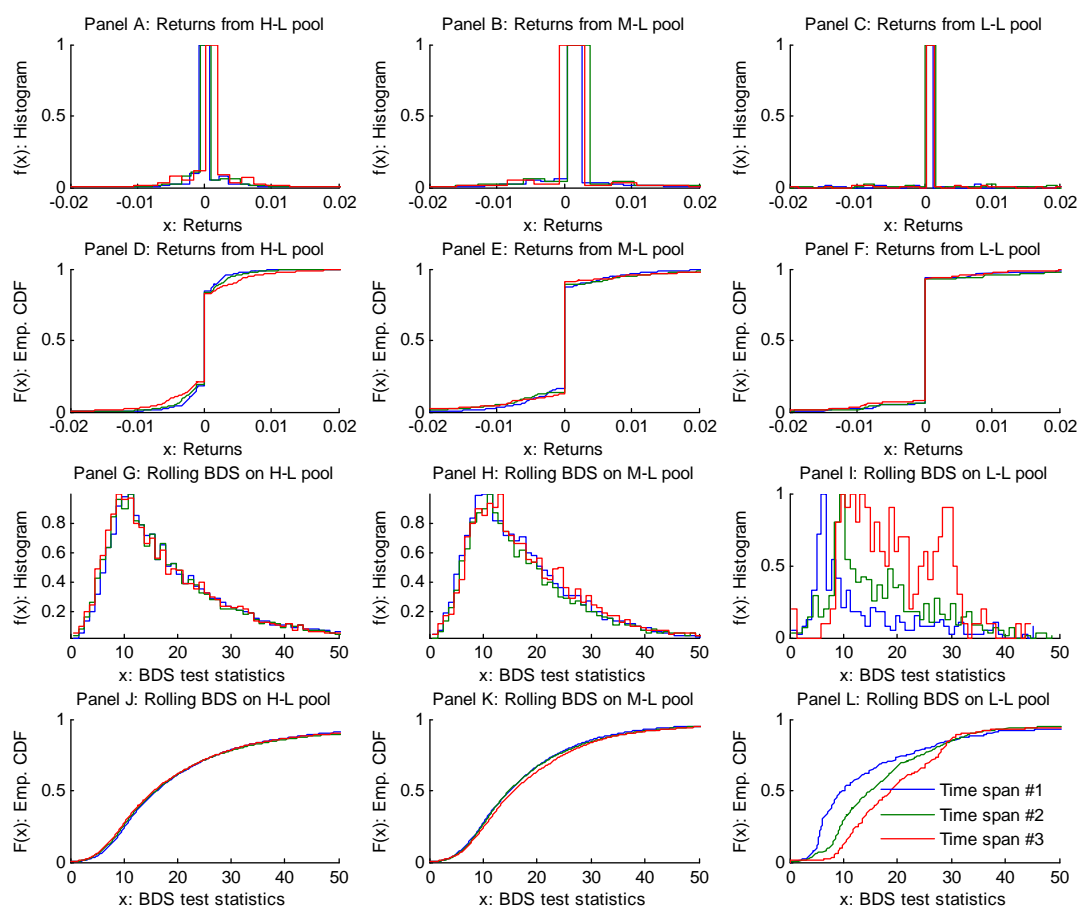


Figure 3.4: Time span distribution analysis of returns and rolling BDS test

To test the validity of the first explanation the histogram and empirical CDF of returns were produced for each time span and liquidity pool. The analysis of returns instead of raw

<sup>12</sup>That is in the first time span the value of maximum break-even transaction costs is less than in the second one, and in the second time span the value is less than in the third time span.

prices allows to avoid the problem of comparing time series with different levels of prices.<sup>13</sup> Figure 3.4 presents the histograms and empirical CDF of returns. Columns from left to right correspond to high-, medium- and low-liquidity pools, respectively. Each panel shows the overlay of histograms and empirical CDF for all three time spans. Charts show that the distribution of returns are almost identical for each time span.<sup>14</sup>

The BDS test is applied to compare degrees of predictability throughout all time spans.<sup>15</sup> Its advantage is that it tests the null hypothesis of no dependence in the data against any possible alternative. In this way the approach of the BDS test is similar to the ATA, which also looks for any possible dependence in the data. Since the ATA uses rolling subsamples of observations to perform forecasts the same subsamples are used in the BDS test. The histograms and empirical CDF of the BDS test statistics are presented in the last two rows of Figure 3.4. In the high- as well as in the medium-liquidity pools the histograms and empirical CDF are identical for all time spans.

The identity of the return distributions and degree of predictability clearly supports the second explanation. The increase in the value of maximum breakeven transaction costs is indeed due to ATA learning (see Table 3.1).

The conclusion is that the ATA is able to extract positive returns adjusted for and conditional on transaction costs by processing past prices. The returns, proxied by the maximum breakeven transaction costs, are increasing through time, which demonstrates the ability of the ATA to master technical analysis.

### 3.3. Performance of Benchmark Trading Strategies

This section presents the benchmark trading strategies and the results of their application on the TSE data. The benchmark strategies allow to evaluate the relative performance of the ATA and to detect problems, which can occur due to defective strategy implementation. The applied benchmark strategies include the perfect forecast, random walk, random trading and buy-and-hold strategy.

The *perfect forecast trading strategy* is based on the assumption that the next period price is known with one hundred percent accuracy. The decision-making of the perfect forecast strategy can be formalized in the following way:

$$D_t^{II}(c, p_t) = \begin{cases} B & \text{if } c < \frac{p_{t+1} - p_t}{p_{t+1} + p_t}, \\ S & \text{if } c > \frac{p_t - p_{t+1}}{p_t + p_{t+1}}, \\ N & \text{if otherwise.} \end{cases} \quad (3.6)$$

<sup>13</sup>Returns are unit-free, which makes them easy to compare and aggregate.

<sup>14</sup>Here and in the comparison of the BDS test statistic distributions in different time spans the analysis omits the results of the statistical tests. Even though the Kolmogorov-Smirnov test rejected the hypothesis of identical distributions the high number of observations present in each sample (on average 27000 observations) questions the validity of this nonparametric test. Similar problem was observed with other nonparametric tests, which were used to compare the distribution of samples.

<sup>15</sup>For details on the BDS test see Appendix A.3.

The perfect forecast trading strategy allows to calculate the maximum hypothetical aggregate return adjusted for transaction costs, which is possible in a given market segment at a given time. Since it is calculated from the current and next period prices, no other trading strategies can generate a higher aggregate return on the same time span at the given costs and frequency of transaction. A strategy, which generates a higher aggregate return adjusted for the same transaction costs, would indicate a problem in its implementation.

The *random walk strategy* exploits the random walk hypothesis. In short the hypothesis postulates that the best prediction of the next period price is the current price plus some stochastic component:

$$p_{t+1} = p_t + \varepsilon_t, \quad (3.7)$$

where the stochastic component is assumed to be normally distributed with zero mean and a variance measured over a fixed number of previous observations:

$$\varepsilon_t \sim N(0, \sigma_t^2) \quad (3.8)$$

and

$$\sigma_t^2 = \text{Var}[\varepsilon_t | p_t, p_{t-1}, \dots, p_{t-n}]. \quad (3.9)$$

The decision-making process of the random walk strategy takes two steps. First, the estimate of next period price,  $\hat{p}_{t+1}$ , is calculated using equation (3.7). To obtain the next period price estimate a realization of random variable, distributed according to  $N(0, \sigma_t^2)$ , is added to the current price,  $p_t$ . Second, based on this forecast the decision is made:

$$D_t^{III}(c, p_t, p_{t-1}, \dots, p_{t-n}) = \begin{cases} B & \text{if } c < \frac{\hat{p}_{t+1} - p_t}{\hat{p}_{t+1} + p_t}, \\ S & \text{if } c > \frac{p_t - \hat{p}_{t+1}}{p_t + \hat{p}_{t+1}}, \\ N & \text{if otherwise.} \end{cases} \quad (3.10)$$

The random walk strategy is characterized by a simple structure and minimum intelligence. With the explicit exploitation of the last price and implicit use of several lagged prices, this strategy can represent the simplest technical analysis, where the variance of lagged prices gives a pattern and the last price gives a level. Comparison to the random walk strategy allows to estimate relative efficiency of a strategy in processing data. A strategy with bad efficiency should have performance lower than of the random walk strategy.

The *random trading strategy* represents the situation when the decision-making is based on no information. This can be the case when investor has wrong, noisy or no information, but due to false confidence in his knowledge he trades. In comparison to the previous strategies this one has no intelligence in decision-making. The decision-making of the random trading strategy can be formalized as follows:

$$D_t^{IV} = \begin{cases} B & \text{with } \Pr(B) = \frac{1}{3}, \\ S & \text{with } \Pr(S) = \frac{1}{3}, \\ N & \text{with } \Pr(N) = \frac{1}{3}. \end{cases} \quad (3.11)$$



##	Stock liquidity pool	TC, %			
		All time spans	Time span #1	Time span # 2	Time span #3
1. The perfect forecast trading strategy					
1.1	High	> 2.00	> 2.00	> 2.00	> 2.00
1.2	Medium	> 2.00	> 2.00	> 2.00	> 2.00
1.3	Low	> 2.00	> 2.00	> 2.00	> 2.00
2. The random walk trading strategy					
2.1	High	0.01	0.01	0.00	0.00
2.2	Medium	0.01	0.01	0.01	0.01
2.3	Low	0.01	0.01	0.01	0.02
3. Random trading					
3.1	High	0.00	0.01	0.01	0.00
3.2	Medium	0.01	0.01	0.01	0.01
3.3	Low	0.01	0.02	0.01	0.00
4. Buy-and-hold trading strategy					
4.1	High	0.00	0.00	0.00	0.00
4.2	Medium	0.00	0.00	0.00	0.00
4.3	Low	0.00	0.00	0.00	0.00

Table 3.2: Application of the benchmark trading strategies

That is the probability to pick one of the three strategies is unconditional and equal to one third.

The random trading strategy can be justified only if the market is efficient and as Thorley [243] puts it, outperforming the market index is a matter of luck, so the best strategy is tossing a coin. Therefore, the dominance of the random trading strategy would clearly indicate market efficiency.

For the perfect forecast, random walk and random trading strategy the aggregate return and the return in each period are calculated as in equation (3.4) and equation (3.5), respectively.

The *buy-and-hold strategy* is included since it is a widely used benchmark in the literature claiming that even if inefficiencies are present, aggregate returns achieved with active trading and application of the technical rules do not exceed the returns when the assets are simply bought and held [6]. Naturally, the decision is only to buy. This strategy has the following aggregate return adjusted for transaction costs:

$$R^V = \frac{p_T - (p_T + p_0) c}{p_0}, \quad (3.12)$$

where  $p_0$  is the price level at the beginning of an investment period and  $p_T$  is the price level at a terminal time.

Figures F.1 - F.9 in Appendix F present the results of the hypothetical application of the benchmark trading strategies on the TSE data. The layout of the figures is the same as for the ATA strategy results. The only exception is that here the slice views are taken at the same trading frequency and subsample size as for the ATA strategy. This allows

to see the alternative results, which are provided by the benchmark strategies at the given conditions.

Table 3.2 presents the summary of the figures. It lists the maximum breakeven transaction costs obtained for the three groups of stock liquidity. The values of the breakeven transaction costs are recorded at the same trade frequency and subsample size as for the ATA strategy reported in Table 3.1.

The best benchmark trading strategy is the perfect forecast. It has the maximum revealed breakeven transaction costs, well above 2 percent of the transaction value. In relative comparison to each other the perfect forecast strategy has a clear dominance since its intensity of returns is much higher.<sup>16</sup> The worst benchmark trading strategy is the buy-and-hold strategy since it is used in the market at the time when it has clear negative slope.

The random walk and random trading strategies have a similar low performance.

To sum up, the analysis of the application of the perfect forecast strategy shows that the data have a high return potential. The random walk and random trading strategies strategy with low or no intelligence, respectively, do not promise substantial returns adjusted for transaction costs.

### 3.4. Discussion

This section presents the discussion of the comparative analysis of the ATA and benchmark trading strategies. The analysis shows that the results of the ATA are credible and second only to the perfect forecast strategy.

Based on the maximum revealed breakeven transaction costs the best strategy is the perfect forecast. It is closely followed by the ATA strategy. After a long blank there are the random walk and random strategy. The buy-and-hold concludes ranking (see Table 3.1 and 3.2).

The position of the ATA strategy in ranking shows that there is no obvious mistake in programming. Its large distance from the random walk and random strategies indicate that the data exhibit the properties of a weakly inefficient market. The application of intelligent technique, like the ATA, allows to extract significant returns conditional on the value of transaction costs. The studied time series show that under the transaction costs of less than 0.32 percent the application of technical analysis produces a gain. On specific time spans gains are possible even under the transaction costs of less than 0.51 percent of the transaction value (see Table 3.1).

From the intelligence aspect the ATA strategy is the closest to the perfect forecast. Even though the ATA has a much lower level of the maximum revealed breakeven transaction costs, the intensity of returns and their persistence clearly resemble the results of the perfect forecast strategy (compare Figure 3.2 and F.2). Since the perfect forecast is a proxy of the

---

<sup>16</sup>That is for any combination of transaction costs, trade frequency and subsample size the perfect forecast strategy provides higher aggregate returns adjusted for transaction costs than any other strategy.

most intelligent strategy, the close resemblance to its results is an indication of the high ATA intelligence level.

An additional argument in favour of the ATA strategy comes from the way it processes data. The ATA is an online technique, which does not have a classical in-sample training. Each iteration is an out-of-sample analysis of the time series. Moreover, it operates only with a fixed number of recent observations. Information about the past is stored as patterns in the memory of the ATA. The ability of the ATA to derive positive returns even after adjusting for transaction costs is a clear indication of past predictability of time series of prices.

One can compare obtained values of breakeven transaction costs with values derived in other markets. Hendrik Bessembinder and Kalok Chan [27, 28] used the same rules and parameter combinations as Brock et al. [36] to estimate breakeven transaction costs for the DJIA. For the period from 1926 to 1991 the estimated value is 0.39 percent, while for a more recent period, from 1976 to 1991, it was reduced to 0.22 percent. With optimized parameters of the Brock et al. [36] methodology Stephen Taylor [241] estimated breakeven transaction costs of 1.07 percent for the DJIA, from 1968 to 1988. For twelve UK firms the value of breakeven transaction costs is 0.08 percent for the period from 1972 to 1991. More recent work of Stephen Taylor [242] reports that the average breakeven transaction costs equals to 0.31 percent for sixteen financial series (including US stock indexes, individual US stocks, and commodities), and in a range from 0.59 to 2.25 percent for four currency exchange rate series (DM/USD, GBP/USD, SF/YEN, and YEN/USD) over the same period. Thus, from the point of view of breakeven transaction costs the degree of TSE market inefficiency is comparable to US stock market.

To conclude, the application of the ATA strategy shows weak inefficiency of the TSE observed on the historical data. The results are genuine since the analysis of individual time spans did not show significant changes in the data. The degree of inefficiency is conditional on the level of transaction costs. Under the low transaction costs technical analysts can perceive the market as inefficient, while at the higher transaction costs the perception of efficiency prevails.

## CONCLUSIONS

The thesis presented a study of how market efficiency is perceived by technical analysts. Its necessity comes from an existing contradiction between the EMH and technical analysis. To clarify the situation, the artificial technical analyst is designed, to replicate the expertise of technical analysts in a robust and objective way. Its application to the Tokyo Stock Exchange data under a variety of transaction costs revealed that when transaction costs were sufficiently low the market could be perceived as inefficient and technical analysis could be rewarded with substantial profits, while under higher transaction costs the perception of market efficiency prevailed. Comparative analysis of the ATA strategy shows that only the perfect forecast strategy provides better reward.

The efficient market hypothesis is one of the central elements of financial theory. Coined by Harry Roberts and formalized by Eugene Fama it postulates impossibility of sustainable gaining profit on information contained in past prices or returns of financial securities. In spite of the advantages which it gives to financial theory empirical studies persistently show the violation of market efficiency. Moreover, the technique used by financial practitioners, technical analysis, directly contradicts the EMH.

This contradiction turns technical analysis into an "outlaw" of financial theory separating the financial community into two camps: the camp of academicians, who advocate the EMH; and the camp of practitioners, who keep using technical analysis. Suppression of technical analysis has a negative impact on the industry since the academicians are missing the ground of practice and the practitioners are deterred from the benefits of financial theory. The possibility of validating technical analysis and justifying a conditional violation of the weak form of the EMH allows to bring together these two camps and to enjoy the benefits of their symbioses.

To resolve the dispute between proponents of market efficiency and technical analysts a well designed mechanism is required. Following ideas of Spyros Skouras [227] to test market efficiency one can use the artificial technical analyst. This approach allows to construct the test independently of equilibrium model. The implementation of technical analysis can be performed in a robust and objective way. The mechanism can be designed to dynamically adjust to changing environment, and to produce transparent results.

The blueprint of the ATA has two major components: data preprocessing and pattern recognition mechanism. The first component of the ATA is the data preprocessing mechanism. To minimize the interference of errors due to outsourcing of data preprocessing, raw data were chosen as an input. At the first stage data should be homogenized. This step allows in the future to reduce complexity of pattern-recognition problem, where one

needs to focus only on price levels observed at predefined frequency. Given two methods of homogenization - linear- and previous-tick interpolation, - the latter method was selected due to its computational simplicity and similarity of results after both methods.

In the original form the homogenization has low efficiency of data utilization, which is crucial in consequent pattern-recognition. Since the pattern-recognition requires as much data as possible, one can augment the homogenization procedure with time shifts. As a result the homogenization produces several diverse time series, which have an identical underlying distribution. This allows to maximize efficiency of data utilization.

The homogenization requires an exogenous specification of its frequency. The grid search was used to pick the frequency, which maximizes the informational content of homogenized data. To proxy the quantity of informational content the criterion of profitability was used. That is the ATA was fed with the data homogenized at different frequencies. The obtained results were used to derive breakeven transaction costs. The frequency, which provided the highest value of breakeven transaction costs, was chosen for homogenization. The same approach was used for selecting the optimal subsample size, which the ATA needs for effective operation. The preliminary results show that for the medium- and high-liquidity pools of stocks the optimal subsample size is 50 observations, which was observed at 6 and 3 minute frequencies of homogenization, respectively. For the low-liquidity pool the optimal subsample size is 110 observations, which was observed at 3 minute frequency.

The second component of the ATA starts with the piecewise linear approximation mechanism. This mechanism serves to identify important peaks and troughs in time series of prices. To pick the best implementation several competing alternatives were tested on artificial data with the Monte Carlo studies. The tests included the sliding windows, top-down, bottom-up and GA-based algorithms. Due to their high speed and acceptable level of approximation the top-down and bottom-up algorithms were selected for application in the ATA.

The last component of the ATA is responsible for pattern recognition. To match the abilities of professional technical analysts the cognitive mechanism of the ATA is based on artificial intelligence. The choice of learning classifier systems is explained by their ability to present a transparent structure of solutions. Since the best implementation of learning classifier systems does not have required power to solve such a complex problem within a reasonable time, the new implementation of LCS was introduced - the TiCS. In a series of test problems it was shown that the TiCS had indeed superior performance. In additional experiments its optimal input parameters were selected. The TiCS has the best performance in detecting non-linear patterns with the subsample size of 50 observations and the subpopulation size of 25.

In this study the time series of stock prices, collected on the Tokyo Stock Exchange (TSE) over the time period from 11/03/1996 to 23/06/1998, are analyzed. The pool of time series covers 2273 stocks listed on the TSE. Considering computational intensity only 30 time series were selected for the study. For this all stocks were sorted into 3 pools

according to their liquidity, which can be described as low-, medium- and high-liquidity pools of stocks. In each pool 10 stocks were randomly selected. Their corresponding time series of prices were used for computations. Before reporting the computed results are averaged within a corresponding liquidity pool.

In the process of calculation several assumptions were made. First, the impact of inflation, dividends and splits is disregarded, since at high frequency their impact is negligible. Second, the aggregate return is computed separately for every time series in the study. Third, trading strategies are applied at every period, i.e. at the beginning of each period the previous position should be closed. Correspondingly the aggregate return is a product of returns in all time periods. And finally, the decision-making process and calculation of returns take into account round-trip TC.

To verify that the extracted positive returns are not just a result of defective programming several benchmark trading strategies are applied to the same data set. The benchmark trading strategies include the perfect forecast, random walk, random and buy-and-hold trading strategies. Based on the maximum revealed breakeven transaction costs the best strategy is the perfect forecast. It is closely followed by the ATA strategy. After a long blank the random walk, random strategy and the buy-and-hold strategy concludes ranking.

The position of the ATA strategy in ranking shows that there is no obvious mistake in programming. Its large distance from the random walk and random strategies indicates that the data exhibit the properties of a weakly inefficient market. The application of an intelligent technique, like the ATA, allows to extract significant returns conditional on the value of transaction costs. The studied time series show that under the transaction costs of less than 0.32 percent the application of technical analysis produces a gain. In specific time spans gains are possible even under the transaction costs of less than 0.51 percent of the transaction value.

From the intelligence aspect the ATA strategy is the closest to the perfect forecast. Even though the ATA has a much lower level of the maximum revealed breakeven transaction costs, the intensity of returns and their persistence clearly resemble the results of the perfect forecast strategy. Since the perfect forecast is a proxy of the most intelligent strategy, the close resemblance to its results is an indication of the high ATA intelligence level.

To conclude, the analysis shows that the results of the ATA are credible. The application of the ATA strategy shows weak inefficiency of the TSE observed on the historical data. The results are genuine since the analysis of individual time spans did not show significant changes in the data. The degree of inefficiency is conditional on the level of transaction costs. Under the low transaction costs technical analysts can perceive the market as inefficient, while at the higher transaction costs the perception of efficiency prevails.

## **Future Work**

In the future one can extend the analysis of breakeven transaction costs. In the current form reported values lack confidence intervals, which can be obtained through bootstrapping. The methodology can be extended to European and American markets, including foreign exchanges. Since the ATA is flexible to take several input time series one can extend the methodology to accommodate order book data. In this case comparative results of processing mid prices and order book data can reveal whether inclusion of additional information can improve the ATA performance. Finally, a population of the ATA can be used to study crisis prevention and control in financial markets. Here one can use a swarm of artificial technical analysts to exploit the effect of herding among market participants to defend against market manipulations.

## APPENDIX A

### Statistical Background

#### A.1. Monte Carlo Method

The Monte Carlo method is a statistical technique to study the behavior of a model in the stochastic environment. Stanislaw Ulam and Nicolas Metropolis introduced this method in 1946. It is coined as the Monte Carlo method after Stanislaw Ulam's uncle who had propensity to gamble on the games of chance in Monte Carlo [172]. The method has the same spirit as the roulette wheel game in the casino of Monte Carlo. The random outcome of the roulette wheel defines the winning and losing in the game. In spite of its stochastic nature, the rules of the game insure a positive profit for the casino in the long run. In the short run the clients of the casino have a high chance of winning, which insures their involvement. In a similar way the Monte Carlo method studies the statistical properties of model outcomes, obtained by repeatedly feeding the model with stochastic inputs.

Assume that one needs to study the following model:

$$[y_1, y_2, \dots, y_n] = f(x_1, x_2, \dots, x_m), \quad (\text{A.1})$$

where  $x_1, x_2, \dots, x_m$  are some input parameters;  $f(\cdot)$  is a conversion function, which captures the transformation of the input parameters in the model into the output ones:  $y_1, y_2, \dots, y_n$ . One is interested in the behavior of the model, when some or all input parameters are uncertain. That is one has knowledge only about their statistical and distributional properties. Often it is difficult or impossible to find analytical solution of  $f(\cdot)$ , due to its nonlinear nature or unknown functional form. In those cases one can use a computational method to accumulate the outcomes under random inputs. With the statistical analysis of the outcomes one can obtain the range of possible values of  $y_1, y_2, \dots, y_n$  and their probabilities.

As a rule, the first step in the Monte Carlo method is the analysis of the input parameters. To simulate a real-life situation one has to feed into the model the realizations of parameters which have the same statistical and distributional properties as the input parameters observed in reality. After quantifying the statistical and distributional properties one uses random number generator to produce samples with those characteristics.

Next, the artificially created parameters are fed into the model. The model is evaluated and the outcomes are stored. One has to repeat this step at least  $10^5$  times to collect convincing statistics.

Once the simulation is complete, the outcomes are the subject of statistical and distributional analysis. To describe the resulting distribution it is customary to report [253]:



- *Histogram*, which allows visual identification of the distribution type and critical statistical parameters of the sample. Figure A.1 exhibits several examples of histograms for common distributions.

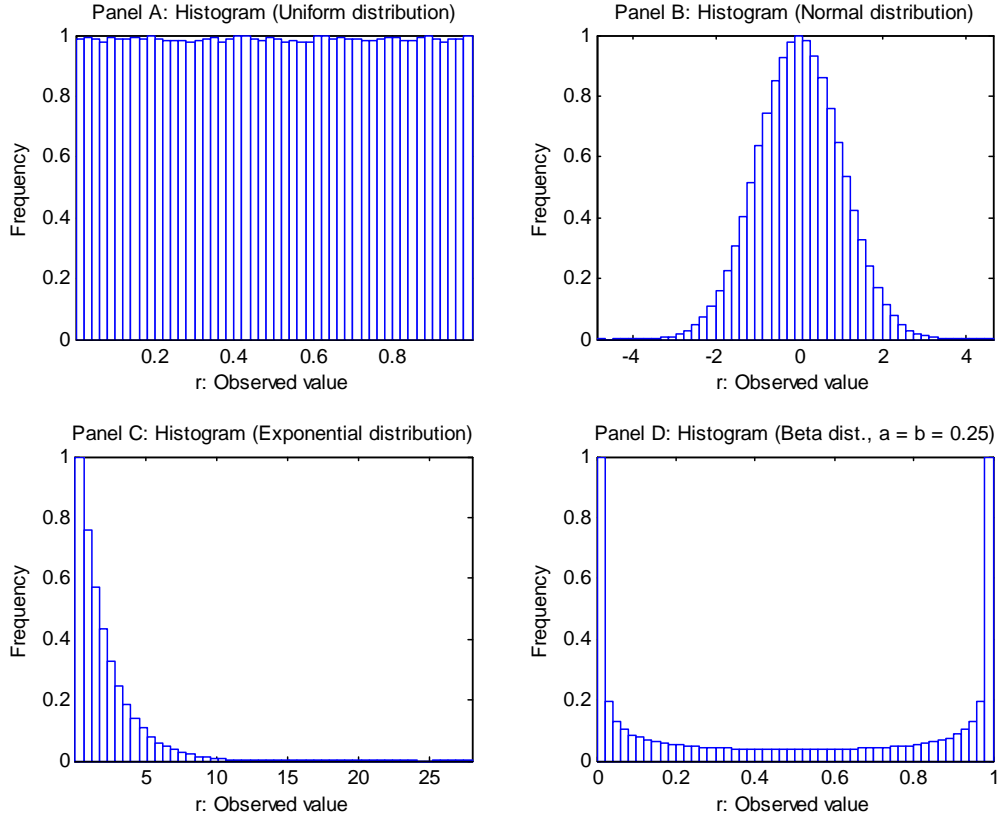


Figure A.1: Examples of histograms for common distributions

- The *sample size*,  $n$ , which is the number of observations or data points in a Monte Carlo simulation. One observes smaller difference between the repeated simulations with a larger sample size.
- *Mean*,  $\bar{y}_i$  which describes central tendency of the distribution, is the average value of observations:

$$\bar{y}_i = \frac{1}{n} \sum_{j=1}^n y_{i,j}. \quad (\text{A.2})$$

The standard deviation of mean in the finite sample is:

$$\sigma_{M,i} = \frac{s_i}{\sqrt{n}}. \quad (\text{A.3})$$

- *Median*, which is the middle value (or average of the two middle values) of the ordered observations, i.e. 50% of the observations from the simulation are smaller than the median. The median is a robust measure of the center of distribution that is less sensitive to outliers than the mean.
- *Minimum* and *maximum*, which define the extreme observations.

- *Standard deviation*, which is a measure of dispersion or spread in observations:

$$s_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (y_{i,j} - \bar{y}_i)^2}. \quad (\text{A.4})$$

- *Skewness*, which is a measure of distribution asymmetry around its mean:

$$S_i = \frac{1}{n} \sum_{j=1}^n \left( \frac{y_{i,j} - \bar{y}_i}{\hat{\sigma}_i} \right)^3, \quad (\text{A.5})$$

where  $\hat{\sigma}_i$  is an estimator of the standard deviation that is based on the unbiased estimator of the variance:

$$\hat{\sigma}_i = s_i \sqrt{\frac{n-1}{n}}. \quad (\text{A.6})$$

Positive skewness means the distribution has a long right tail and negative skewness implies the distribution has a long left tail.

- *Kurtosis*, which is a measure of distribution flatness:

$$K_i = \frac{1}{n} \sum_{j=1}^n \left( \frac{y_{i,j} - \bar{y}_i}{\hat{\sigma}_i} \right)^4. \quad (\text{A.7})$$

If kurtosis exceeds 3, the distribution is peaked (leptokurtic) relative to the normal; if the kurtosis is less than 3, the distribution is flat (platykurtic) relative to the normal.

- *Tail index*, which describes the decay of the tails of a distribution. The value of the tail index,  $\alpha$ , can be obtained by the Hill estimator [103]:

$$\hat{\gamma}_{n,m}^H = \frac{1}{m-1} \sum_{j=1}^{m-1} (\ln(y_{(j)}) - \ln(y_{(m)})), \quad (\text{A.8})$$

where  $m > 1$  is the number of statistics and  $y_{(1)} \geq y_{(2)} \geq \dots \geq y_{(n)}$  are descending order statistics from an underlying sample. This estimator was proven to be a consistent estimator of  $\gamma = \frac{1}{\alpha}$  for fat-tailed distributions [169].<sup>1</sup>

One can augment the analysis of the distributional properties by a graph of estimated or empirical *cumulative distribution function* (CDF). The CDF is defined as the probability of observing a value from a series not exceeding a specified value  $r$ :

$$F_y(r) = \Pr(y \leq r). \quad (\text{A.9})$$

Figure A.2 displays empirical CDF for corresponding histograms in Figure A.1.

To compare different model specification one can use an *overlay chart*. It displays multiple CDF on the same axis. With the overlay chart, one can compare and select the best alternatives.

Figure A.3 offers two examples of overlay charts. In the simplest case, shown on Panel A, the graphs of the CDF are not intersecting. This is the case of the so called *first-order*

<sup>1</sup>The significance of the estimator can be accessed through  $(\hat{\gamma}_{n,m}^H - \gamma) \sqrt{m} \rightarrow N(0, \gamma^2)$ .

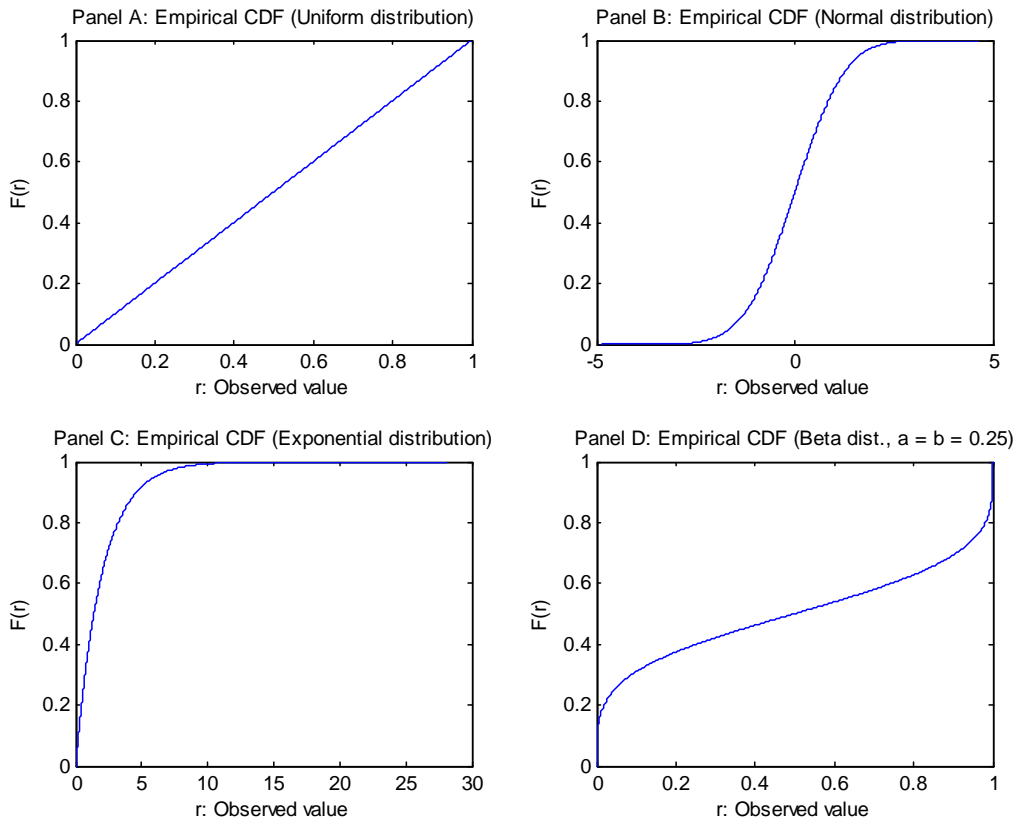


Figure A.2: Examples of empirical CDF for different distributions

*stochastic dominance*. The definition of the first-order stochastic dominance is following: *when there are two specifications of the model, and the possible outcomes are described, correspondingly, by statistical distributions  $F(\cdot)$  and  $G(\cdot)$ , then the distribution  $F(\cdot)$  first-order stochastically dominates distribution  $G(\cdot)$  if for every possible value, the probability of getting a value that high is never greater in  $G(\cdot)$  than in  $F(\cdot)$* . Clearly series  $y_1$  first-order stochastically dominates series  $y_2$ .

Panel B of Figure A.3 presents a more complicated case. The graphs of CDF are intersecting. To perform the comparison one needs to introduce the *second-order stochastic dominance*. With the previous setup and notations, one defines the second-order stochastic dominance as following:  *$F(\cdot)$  second-order stochastically dominates  $G(\cdot)$  if:*

$$\int_a^r [G(t) - F(t)] dt \geq 0 \text{ for all } r \in [a, b], \quad (\text{A.10})$$

where  $a$  and  $b$  are the minimum and maximum values of the model outcomes. One would need a numerical evaluation of equation (A.10) to gain a sound conclusion. With the method proposed in [259] it is a simple task. Another way is to overlay the graphs of integrated CDF:

$$f(r) = \int_a^r F(t) dt. \quad (\text{A.11})$$

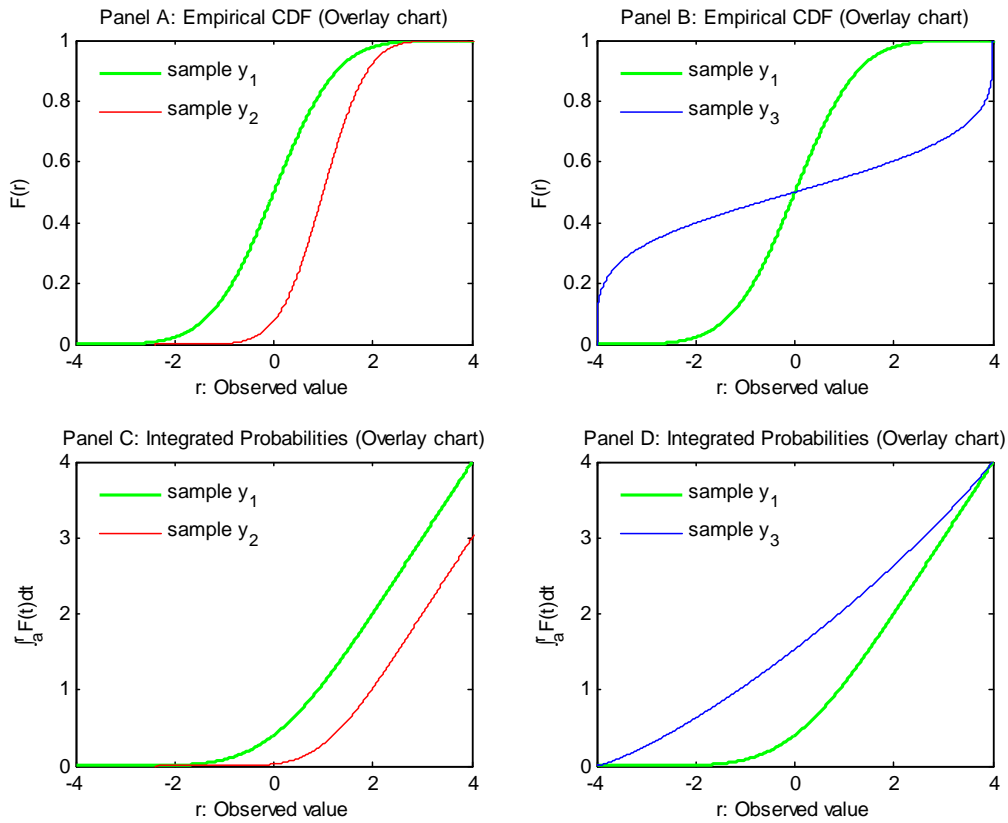


Figure A.3: Example of the overlay chart of several empirical CDF

Panels C and D show graphs of integrated CDF which correspond to graphs on Panels A and B, respectively. Now one can immediately see that series  $y_3$  second order stochastically dominates series  $y_1$ .

## A.2. Kolmogorov-Smirnov Test

Kolmogorov-Smirnov test performs a two-sided comparison of two sample series. The null hypothesis for the test is that series  $X_1$  and series  $X_2$  are drawn from the same continuous distribution. For this, the Kolmogorov-Smirnov test statistics is defined as the maximum value of the absolute difference between two cumulative distribution functions [203]:

$$D = \max_{-\infty < x < \infty} |F_{X_1}(x) - F_{X_2}(x)|, \quad (\text{A.12})$$

where  $F_{X_1}(x)$  and  $F_{X_2}(x)$  are cumulative distribution functions for  $X_1$  and  $X_2$ , respectively.

Critical values for  $D$ -statistics are tabulated in [170].

### A.3. BDS Test

BDS test<sup>2</sup> performs the analysis for independently and identically distributed time series based on correlation integrals at various dimensions. The null hypothesis for the test is that there is no dependence in the data. The alternative hypothesis is that the data can have any linear or non-linear dependence [37].<sup>3</sup>

The test requires an endogenous specification of dimensional distance  $\varepsilon$ , which should not exceed the spread of the time series (if it exists):

$$0 < \varepsilon < \max(X) - \min(X), \quad (\text{A.13})$$

where  $X$  is a time series, which follows some distribution  $F$ .

Since the test is based on correlation integrals at various dimensions, the maximum number of dimension,  $m$ , should be endogenously specified as well.

The BDS test checks for independently and identically distributed time series in the following way [121]. The probability of any pair of observations  $\{x_i, x_j\}$  lying within  $\varepsilon$  of each other is:

$$P_1 \equiv \Pr(|x_i - x_j| \leq \varepsilon) \text{ for any integers } i \neq j. \quad (\text{A.14})$$

A similar relationship in dimensions 2 is defined by any two observations and their neighbors directly preceding them. The probability of a history of two observations being within  $\varepsilon$  of each other is:

$$P_2 \equiv \Pr(|x_i - x_j| \leq \varepsilon \ \& \ |x_{i-1} - x_{j-1}| \leq \varepsilon) \text{ for any integers } i \neq j. \quad (\text{A.15})$$

If the time series  $X$  is independently and identically distributed (*IID*), then the probability of a two-observation history being within  $\varepsilon$  of each other is equal to the square of the probability of any two observation being within  $\varepsilon$ :

$$P_2 = P_1^2 \text{ if } X \sim F(\text{IID}). \quad (\text{A.16})$$

The power relationship generalizes for any dimension  $m$ , and the BDS test for embedding dimension  $m$  is a test of the null hypothesis that the probabilities for dimension 1 in power  $m$  and for dimension  $m$  are equal:

$$\begin{aligned} H_0: P_m &= P_1^m, \\ H_1: P_m &\neq P_1^m. \end{aligned} \quad (\text{A.17})$$

The BDS statistic for embedding dimension  $m$  and dimensional distance  $\varepsilon$  is estimated consistently on a sample of  $n$  observations by:

$$w_{m,n}(\varepsilon) = \sqrt{n - m + 1} \frac{c_{m,n}(\varepsilon) - [c_{1,n-m+1}(\varepsilon)]^m}{\sigma_{m,n}(\varepsilon)}, \quad (\text{A.18})$$

<sup>2</sup>The test is called after its original authors William Brock, Davis Dechert and Jose Scheinkman, who developed it in 1986.

<sup>3</sup>Thus, the BDS test may help to identify the existence of non-linear dependence, but not its type [121].

where the correlation integral  $c_{m,n}(\varepsilon)$  is defined as:

$$c_{m,n}(\varepsilon) \equiv \frac{2}{(n-m+1)(n-m)} \sum_{s=mt=s+1}^n \sum_{t=s+1}^n \prod_{j=0}^{m-1} I_\varepsilon(x_{s-j}, x_{t-j}), \quad (\text{A.19})$$

and the Heavside function  $I_\varepsilon(x_i, x_j)$  is defined as:

$$I_\varepsilon(x_i, x_j) = \begin{cases} 1 & \text{if } |x_i - x_j| \leq \varepsilon, \\ 0 & \text{otherwise.} \end{cases} \quad (\text{A.20})$$

The estimated variance of  $c_{m,n}(\varepsilon) - [c_{1,n-m+1}(\varepsilon)]^m$  is given by:

$$\sigma_{m,n}^2(\varepsilon) = 4 \left\{ k^m + 2 \sum_{j=1}^{m-1} k^{m-j} c^{2j} + (m-1)^2 c^{2m} - m^2 k c^{2m-2} \right\}, \quad (\text{A.21})$$

where the parameter  $c \equiv c_{1,n}(\varepsilon)$  is the first-dimensional correlation integral. Parameter  $k \equiv k_n(\varepsilon)$  is the probability of any triplet of observations lying within distance  $\varepsilon$  of each other:

$$k_n(\varepsilon) \equiv \frac{2}{n(n-1)(n-2)} \sum_{t=1}^n \sum_{s=t+1}^n \sum_{r=s+1}^n \{ I_\varepsilon(x_t, x_s) I_\varepsilon(x_s, x_r) + I_\varepsilon(x_t, x_r) I_\varepsilon(x_r, x_s) + I_\varepsilon(x_s, x_t) I_\varepsilon(x_t, x_r) \}. \quad (\text{A.22})$$

In large samples the BDS statistic follows the normal distribution. For small samples of at least 50 observations the critical values are tabulated in [121].

## APPENDIX B

### Genetic Algorithms

Genetic Algorithms (GA) are optimization techniques based on the idea of natural selection and survival of the fittest. The inventor of GA is John Holland [106, 107]. He came with an idea to use the principles of genetic reproduction and natural selection in automated problem solving. An algorithm iteratively selects from a small population a subset of fittest candidate solutions. These solutions are used in repopulation either directly, or through genetic reproduction. In genetic reproduction parents contribute best features to create offsprings with a superior fit. With a new population the cycle is repeated until a stopping criterion is fulfilled. In the final population, members with the highest fitness carry the problem solution.

The genetic nature of the procedure requires two necessary criteria to be satisfied. The first one is the possibility of a genetic representation of solutions. Here the term genetic stands for the possibility to split any solution into a number of diverse components. Without a component structure of the solution the algorithm is not able to get offsprings, which carry a combination of parents' features. The second criterion is the existence of a fitness function to evaluate each candidate solution. The existence of a fitness function allows to select superior candidates and to remove the weak.

Function optimization naturally satisfy both criteria, which makes it a main application domain of GA. GA can take as input arguments numerical and non-numerical variables, which greatly widens the GA scope of application. Since no restrictions on fitness functions exist, GA can be applied to optimize non-linear problems including those, which do not have analytical solutions.

The following sections describe in detail main principles and aspects of GA. Section B.1 describes a general mechanism of GA, proposed by J. Holland. Section B.2 presents its contemporary modification -  $\mu$ GA. A discussion of algorithm selection in Section B.3 closes Appendix B.

#### B.1. GA Technique

Genetic Algorithms is an optimization procedure based on evolution principles adopted by J. Holland. The working principle of GA is the search for the best solution through a competition of trial solution, and subsequent breeding of new solutions from current winners.

Two criteria have to be satisfied to apply GA. That is the possibility of genetic representation of trial solutions, and the existence of a value function, which can quantitatively distinguish different trial solutions.

Genetic representation enables to transform the value of the trial solution in the most detailed way. With a genetic representation the solution can be analyzed in a way similar to the factor analysis, i.e. by altering one element at a time and observing its influence on the outcome. Naturally, one would require as many elements in the genetic representation as possible. For this reason in the numeric optimization problems real variables are presented in a binary form, which provides the most detailed representation.

The value function is determined by the objective of GA. In order to achieve the goal of optimization the value function should have at least one maximum or minimum defined within a search domain. The search domain as well as the direction of the search (i.e. the goal is maximization or minimization) should be explicitly defined.

With the existing value function the search for the optimal solution starts from creating an initial population of trial solutions. Each solution is randomly generated and is assigned the strength or fitness using the value function. The value function should assign higher strength to a better solution, and lower to a worse one. By ordering with respect to strength the population can be ranked. The best solutions with the highest strength will get the highest ranks.

**Example 16.** *One has to maximize the function  $f(x) = x^2$ , where  $x$  is allowed to vary between 0 and 255. The random population of trial solutions is presented in Table B.1. Given the goal of function maximization, the value function is  $V(x) = x^2$ .<sup>1</sup> After evaluating the strength of each trial solution the population can be ranked. The best solution is #1 with  $x = 163$ .*

##	$x$ value,		evaluated value, strength	current rank
	in binary form	in decimal form		
1	10100011	163	26569	1
2	00111010	58	3364	3
3	01110011	115	13225	2

Table B.1: Example of initial population of trial solutions in GA

A clear distinction of good and bad solutions allows to use the principle of evolution and survival of the fittest. The evolution aspect is implemented through several genetic operators, which allow to cross-mate and modify solutions in the population. The survival of the fittest is performed through replacing bad solutions with better ones.

Genetic operators allow to explore the solution space, to carry on current successful solutions and to exchange valuable elements between solutions. Genetic operators include mutation, reproduction and crossover.

*Mutation* is a genetic operator, which randomly alters one bit in a trial solution. The role of mutation is to explore the solution space without significant deviation from the

<sup>1</sup>In case of function minimization the value function would be  $V(x) = -x^2$ .



current trial solution. In case of the binary representation the value of a bit changes either from 0 to 1, or from 1 to 0.

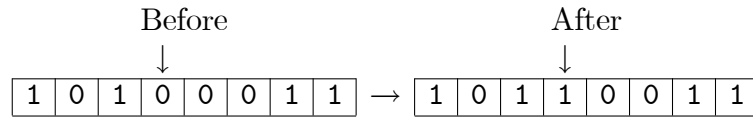


Figure B.1: Illustration of mutation operator

**Example 17.** Continuing with the previous example Figure B.1 shows an exemplary application of the mutation operator to the trial solution  $10100011_2 = 163_{10}$ . First, the position to mutate the bit is randomly selected (here it is #4 from the left). Second, the value of the 4th bit is changed to one of the possible alternatives (in this case  $0 \rightarrow 1$ ). The mutation operator produced a new trial solution with the value  $10110011_2 = 179_{10}$  and greater strength.

*Reproduction* is a genetic operator which copies the best solutions from the old population to the new one without any changes. The role of reproduction is to preserve the best solutions in the population. Figure B.2 exhibits the process of reproduction.

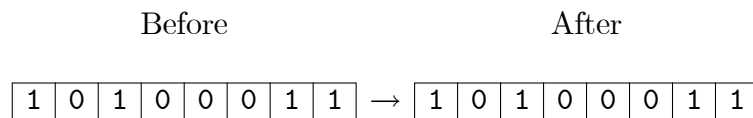


Figure B.2: Illustration of reproduction operator

*Crossover* is a genetic operator which swaps parts of two trial solutions. This is the most important operator in GA since it allows to exchange valuable information within the population of trial solutions. It is based on the idea that two highly strong parents (trial solutions) through mating (information exchange) can produce even stronger offsprings (new trial solutions). There are two types of crossover, that is a single- and two-point crossover. With the single-point crossover a new generation of offsprings is created by swapping one continuous part of elements between two parents. The two-point crossover swaps two disjoint continuous parts of elements.

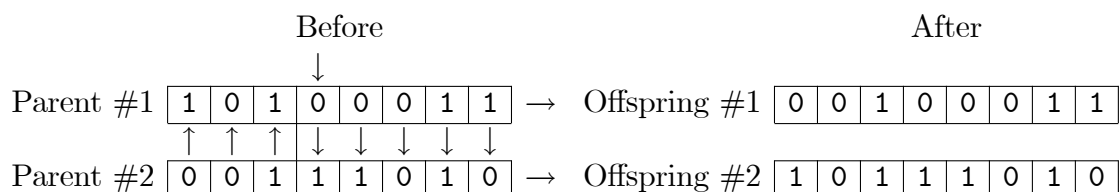


Figure B.3: Illustration of single-point crossover operator

**Example 18.** Figure B.3 shows the single-point crossover applied to trial solutions #1 and #2 (see Table B.1). At the first step the point of a swap is randomly chosen (here it is bit #4 from the left). At the second step two offsprings are created. Offspring #1 has the first 3 bits from parent #2, and the rest from parent #1. Offspring #2, on the opposite, has the first 3 bits from parent #1, and the rest from parent #2. New offsprings are  $00100011_2 = 35$  and  $10111010_2 = 186$ . The application of the one-point crossover operator produced two new trial solutions, where one has the strength superior to its parents (see Table B.1).

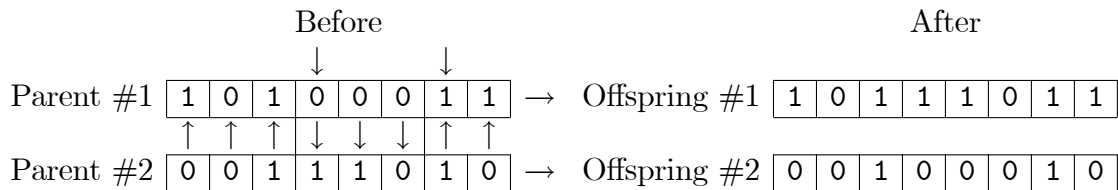


Figure B.4: Illustration of two-point crossover operator

**Example 19.** Figure B.4 shows the two-point crossover applied to trial solutions #1 and #2 (see Table B.1). At the first step two points of a swap are randomly chosen (here they are bits #4 and #7 from the left). At the second step two offsprings are created. Offspring #1 has the first 3 and the last 2 bits from parent #1, and the rest from parent #2. Offspring #2, on the opposite, has the first 3 and the last 2 bits from parent #2, and the rest from parent #1. New offsprings are  $10111011_2 = 187$  and  $00100010_2 = 34$ . The application of the two-point crossover operator produced two new trial solutions, where one has the strength superior to its parents (see Table B.1).

Each genetic operator is invoked one at a time and in a stochastic way. The probabilities to apply each operator are selected *a priori* and sum up to one. A *roulette-wheel* mechanism stochastically picks one of the operators. The name of the roulette wheel describes its working principle. Three slots of the hypothetical roulette wheel are sized according to the values of the probabilities. By drawing from a uniformly distributed realization of random variable the interval  $[0, 1]$ , a position of a hypothetical ball is defined. This position determines which genetic operator should be used in the current round.

**Example 20.** To illustrate the roulette-wheel mechanism one can assume the following probabilities: for mutation operator  $P_M = 0.1$ , for reproduction operator  $P_R = 0.4$  and for crossover operator  $P_C = 0.5$ . Three probabilities sum up to one  $P_M + P_R + P_C = 1$ . Table B.2 shows the realizations of a random variable (uniformly distributed on the interval  $[0, 1]$ ) and selected genetic operators.

The fusion of genetic representation and genetic operators constitutes the essence of GA. The first implementation, proposed by J. Holland, was later improved and now is known as a Simple Genetic Algorithm (SGA).<sup>2</sup> It is a serially implemented algorithm, which works

<sup>2</sup>For more details see [95].

##	Realization of random variable	Selected genetic operator
1	0.453	reproduction
2	0.687	crossover
3	0.048	mutation

Table B.2: Illustration of a roulette-wheel mechanism

with the binary coded population of trial solutions. It starts with  $N$  randomly generated trial solutions. In a series of consequent iterations the following sequence is repeated:

- (1) A mating pool of trial solutions is selected by a roulette wheel.<sup>3</sup>
- (2) Using a roulette wheel a genetic operator is selected.
- (3) In case of invoking the mutation or reproduction operator one trial solution is randomly selected from the mating pool. Otherwise, two parents are randomly selected (with no replacement).
- (4) The selected genetic operator is applied to create new trial solutions.
- (5) Steps 2 - 4 are repeated until the required number of new trial solutions is achieved.
- (6) The strength of every new trial solution is evaluated.
- (7)  $N$  best-in-fitness trial solutions out of the new and old population are left for the next iteration.

There are two condition terminating the SGA search. These two conditions are achieving convergence and/or the maximum predefined number of iterations. The condition of convergence is fulfilled when the algorithm cannot improve the solution for a sufficiently long time. In case the problem has several local optima this condition can cause a premature convergence. The premature convergence is one of the main drawbacks of the SGA design. The existing work-around is to increase the population size, but it results in a longer time required to find the optimal solution.

The threat of premature convergence leads to an increase in the population size. The consequently increased calculation time constrains the application domain of SGA to static problems where the elapsed time is not crucial. To overcome this obstacle a new implementation of GA was proposed by [138] coined as a Micro-Genetic Algorithm -  $\mu$ GA.

## B.2. $\mu$ GA: Micro Genetic Algorithm

There are many problems where the functions to be optimized are evolving faster than the SGA can find an optimum. In this environment the application of the SGA is meaningless. Problems of this type can be found in most of the real-world situations, including problems related to the replication of behavioral and professional qualities of financial market participants. Under these conditions the  $\mu$ GA<sup>4</sup> reaches the near-optimal region much earlier than the SGA [138].

<sup>3</sup>The roulette wheel works in the same way as in the case of selecting a genetic operator but the number of slots is equal to  $N$  and each slot is sized according to the strength of the corresponding trial solution.

<sup>4</sup> $\mu$  is the author's notation for "micro".

The  $\mu$ GA works with binary coded populations. The main difference between the SGA and the  $\mu$ GA is in the population choice. In the  $\mu$ GA the population size is fixed at  $N = 5$ . Minor differences come from using only crossover operator. The aspect of mutation is covered by frequent generation of new random solutions. The step by step procedure for the  $\mu$ GA implementation is presented below:<sup>5</sup>

- (1) A population is selected either of size 5 randomly or of size 4 randomly and the best trial solution from any previous search (reshuffling strategy).
- (2) The strength of each new solution is evaluated.
- (3) The best trial solution in the population is determined (elitist strategy).<sup>6</sup>
- (4) Four trial solutions are chosen for reproduction (the best trial solution also competes for a copy in reproduction) based on the deterministic tournament selection strategy.<sup>7,8</sup>
- (5) A new generation of trial solutions is obtained through the application of the crossover operator.<sup>9</sup>
- (6) The nominal convergence is checked.<sup>10</sup> If convergence condition is fulfilled steps 1 - 6 are repeated.
- (7) Steps 2 - 7 are repeated.<sup>11</sup>

The reshuffling procedure of the  $\mu$ GA helps avoiding the premature convergence and allows the  $\mu$ GA to look for a better solution. Since the interest in using the  $\mu$ GA lies purely in finding the optimum as quickly as possible, the performance measure for the algorithm should be based on the best-so-far string, rather than on any average performance [138].

### B.3. Algorithm Selection

The SGA and  $\mu$ GA are only a small part of the family of GA. The presentation and analysis of the SGA is explained by necessity to demonstrate the main building blocks of GA. As for the  $\mu$ GA, its presentation is solely due to the author's subjective preference. It is considered to be the best and the most elegant implementation of GA. A minimum of tuning parameters makes the use of this optimization procedure easy. In fact, by manipulating the population size (the only parameter) the  $\mu$ GA can be adjusted to solve extremely dynamic as well as complex stationary problems.

---

<sup>5</sup>For more details see the source [138].

<sup>6</sup>In this way there is a guarantee that the information about the best trial solution is not lost.

<sup>7</sup>In the tournament selection strategy, the trial solutions are grouped randomly and adjacent pairs are made to compete for the final four. Care should be taken to avoid two copies of the same solution mating for the next generation.

<sup>8</sup>Since the population is small, the law of averages does not hold and the selection strategy is purely deterministic.

<sup>9</sup>The mutation rate is zero since it is clear that enough diversity is introduced after every convergence through a new population of trial solutions.

<sup>10</sup>A reasonable measure is based on either genotype convergence or phenotype convergence.

<sup>11</sup>In my experience the algorithm is able to find acceptable solution in a dynamic environment within 30 iterations.

## Revealed Transaction Costs

This section of the Appendix presents the concept of revealed transaction costs and the technique to calculate them.

Long and short selling are the most basic transactions in financial markets. Any complex arbitrage strategy is just their combination. The applicability of either transaction normally depends on the expectation about the future price movement and transaction costs. Wrong assessment of either of them can cause a financial loss. With endogenous price expectations one can evaluate maximum transactions costs at which arbitrage strategies generate positive profits.

Transaction costs include commissions paid per each transaction, a bid-ask spread and the impact on a price by performing a transaction. Every component complicates measuring actual transaction costs. If, for example, there is only a time series of past prices, then the bid-ask spread, past commissions and the impact on a price have to be estimated from the data. Roll [211] introduced an implicit measure of effective bid-ask spread that can be estimated from the univariate time series. Lesmond et al. [150] used time series observations of zero returns to estimate transaction costs of a marginal investor.<sup>1</sup> Unfortunately, for the precise estimation of transaction costs one needs a complete data set to capture all the components, which are usually not accessible on time or have confident nature.

Known transaction costs enable to reveal the return sufficient to insure a positive profit. If, for some reason, the exact value of transaction costs is not known, a reverse solution can be found, i.e. one can find transaction costs that insure non-negative profit. Those costs can be derived from the arbitrage strategies mentioned above and are referred to as *revealed transaction costs*. Bessembinder et al. [28] used a similar concept, known as breakeven costs. In relation to our definition the breakeven transaction costs are maximum costs that insure zero profit.

Following the strategies differentiation, one can derive revealed transaction costs for the case of expected increase and decrease in the asset's price. Given the high-frequency nature of data and transactions, one can neglect interest rates as well as the influence of dividends and splits.

---

<sup>1</sup>For the latest review of literature on transaction costs see [151].

### C.1. Case (i): Expected Increase in Asset Price at Time $t + 1$

At time  $t$  one can buy  $N$  units of an asset at price  $p_t$ . With the limited budget,  $m_t$ , and no investment alternatives one faces the following budget constraint:

$$Np_t + Np_t c = m_t. \quad (\text{C.1})$$

where  $0 < c < 1$  is a proportionate transaction cost. At time  $t + 1$  all  $N$  units of the asset should be sold at price  $E[p_{t+1}]$  with the expected revenue of  $NE[p_{t+1}]$ . The expected profit from closing the long position is:

$$\pi_{t+1} = NE[p_{t+1}] - NE[p_{t+1}]c - m_t. \quad (\text{C.2})$$

Substituting (C.1) into (C.2) gives:

$$\pi_{t+1} = N[E[p_{t+1}] - p_t - (E[p_{t+1}] + p_t)c]. \quad (\text{C.3})$$

The sign of the profit depends only on the value in the brackets. Thus the condition for the positive profit is:

$$E[p_{t+1}] - p_t - (E[p_{t+1}] + p_t)c > 0. \quad (\text{C.4})$$

By solving for  $c$  one gets:

$$c < \frac{E[p_{t+1}] - p_t}{E[p_{t+1}] + p_t} \equiv \widehat{c}_t^L, \quad (\text{C.5})$$

where  $\widehat{c}_t^L$  stands for revealed transaction costs from the long position, which has the following interpretation: *if revealed transaction costs are above the market transaction costs,  $c$ , then the long position will be profitable.*

### C.2. Case (ii): Expected Decrease in Asset Price at Time $t + 1$

At period  $t$  one can short  $N$  units of the asset at price  $p_t$ , with revenue  $Np_t$ , and to carry the transaction cost of  $Np_t c$ . At time  $t + 1$  one has to buy back  $N$  units of the asset at price  $E[p_{t+1}]$  to replace the borrowed quantity of the asset, with the total cost of  $NE[p_{t+1}] + NE[p_{t+1}]c$ . The expected profit after this transactions is:

$$\pi_{t+1} = N[p_t - E[p_{t+1}] - (E[p_{t+1}] + p_t)c]. \quad (\text{C.6})$$

Again looking for the positive profit condition and solving for  $c$  one gets:

$$c < \frac{p_t - E[p_{t+1}]}{E[p_{t+1}] + p_t} \equiv \widehat{c}_t^S. \quad (\text{C.7})$$

where  $\widehat{c}_t^S$  stands for revealed transaction costs from the short position, such that *if revealed transaction costs are above the actual market transaction costs,  $c$ , then the short position will be profitable.*

To conclude, one can use the concept of revealed transaction costs to derive the investment decision making function:

$$D_t(c, p_t, E[p_{t+1}]) = \begin{cases} B & \text{if } c < \frac{E[p_{t+1}] - p_t}{E[p_{t+1}] + p_t}, \\ S & \text{if } c < \frac{p_t - E[p_{t+1}]}{E[p_{t+1}] + p_t}, \\ N & \text{otherwise,} \end{cases} \quad (\text{C.8})$$

where  $B$  denotes a decision to buy,  $S$  - a decision to short-sell and  $N$  - to do nothing.

## APPENDIX D

### Tokyo Stock Exchange: Market Microstructure and Data Description

This section of the Appendix describes the data used in Chapter 2 and 3. The data cover historical stock prices collected on the Tokyo Stock Exchange (TSE).

The TSE is a classical example of order-driven market. It operates as a continuous auction, where buy and sell orders interact directly with one another. The study of the order-driven activity of the TSE did not show significant differences from the New York Stock Exchange or from stocks markets with designated market-makers [7, 22, 148]. At the same time the TSE market has some specific features, which can have impact on empirical results. The TSE microstructure is outlined in Section D.1.

The results of empirical analysis can be sensitive to the method of data selection. Section D.2 describes the method of data selection, which minimizes the impact of selection bias on empirical results. The section is concluded with a statistical description of the data.

#### D.1. Market Microstructure

The TSE is an entirely order-driven market. It operates as a continuous auction, where buy and sell orders interact directly with one another.<sup>1</sup>

The TSE has no market-makers. All orders, whether limit or market orders, are placed by broker/dealer trading participants (see Table D.1) and matched in accordance with *price priority* and *time priority* rules. Under the price priority rule, a sell (buy) order with the lowest (highest) price takes precedence. Under the time priority rule, an earlier order takes precedence over others at the same price. Thus, when the lowest sell and highest buy orders match in price, the transaction is executed at the price.

At the TSE there are two transaction methods: the *itayose* and *zaraba*. The *itayose* method is used mainly to decide opening and closing prices. Under the *itayose* method, the time priority rule is not applied and numerous orders placed before price setting are matched in aggregate.<sup>2</sup> In contrast, under the *zaraba* method, both the price priority and time priority rules are applied, and pairs of buy and sell orders are matched continuously.

The TSE adopts the following measures to prevent wild short-term fluctuations in prices. These measures do not only help ensure price continuity, but also in effect work as "circuit breakers" in an emergency:

---

<sup>1</sup>Unless it is explicitly stated this subsection is a digest of [244].

<sup>2</sup>In other words, if a sell (buy) order with a lower (higher) price than the opening price is placed before the market opens, it will be executed at the opening price regardless of size.



Stock price per share				Tick size
		Up to	¥2,000	¥1
More than	¥2,000	Up to	¥3,000	¥5
More than	¥3,000	Up to	¥30,000	¥10
More than	¥30,000	Up to	¥50,000	¥50
More than	¥50,000	Up to	¥100,000	¥100
More than	¥100,000	Up to	¥1,000,000	¥1,000
More than	¥1,000,000	Up to	¥20,000,000	¥10,000
More than	¥20,000,000	Up to	¥30,000,000	¥50,000
More than	¥30,000,000			¥100,000

Table D.1: Tick size at the TSE

Source: [244].

- (1) *Special Bid & Ask Quotes.* When there is a major imbalance in orders, special bid or ask quotes are indicated by the TSE. Special quotes are disseminated publicly through the TSE market information system. If counter orders come into the market and the orders are matched at that price, the quote is withdrawn. Conversely, if the imbalance continues, the special quotes are revised up or down within certain parameters (see Table D.2), at intervals of at least five minutes until the imbalance is resolved.
- (2) *Daily Price Limits.* The TSE sets daily price limits for individual stocks to prevent day-to-day wild swings in stock prices and provide for "time-out" in the event of a sharp rise or decline in price and the resulting reaction from the investing public. Daily price limits are set in terms of absolute yen values according to the price range of each stock (see Table D.3). Since the price limits prohibit bids and offers at prices beyond the set limits, the market for a stock is open for trading within these limits, even though the stock may have hit a limit. Daily price limits also apply to special quotes. Consequently, special quotes cannot be indicated outside the daily price limit.
- (3) *Trading Units.* Trading units are the minimum amounts of each individual stock that may be traded. With an amendment made to the Commercial Code in October 2001 the number of shares which constitute one unit may now be determined under the constitution of each listed company.
- (4) *Margin Transactions.* Margin transactions are the purchase, sale or other transactions of securities effected on credit extended to the customer by a securities company. Customers buying or selling stocks on margin must deposit a warranty deposit equivalent to at least 30% of the transaction value or 300,000 yen (whichever is greater) with the securities company by noon of the trading day. When the deposited margin rate to the trading value is depleted below 20% due to price fluctuation, the customer must deposit additional margin to increase the rate to at least 20%.

Current stock price per share				Parameters ( $\pm$ )
		Less than	¥500	¥5
Equal to or more than	¥500	Less than	¥1,000	¥10
Equal to or more than	¥1,000	Less than	¥1,500	¥20
Equal to or more than	¥1,500	Less than	¥2,000	¥30
Equal to or more than	¥2,000	Less than	¥3,000	¥40
Equal to or more than	¥3,000	Less than	¥5,000	¥50
Equal to or more than	¥5,000	Less than	¥10,000	¥100
Equal to or more than	¥10,000	Less than	¥20,000	¥200
Equal to or more than	¥20,000	Less than	¥30,000	¥300
Equal to or more than	¥30,000	Less than	¥50,000	¥400
Equal to or more than	¥50,000	Less than	¥70,000	¥500
Equal to or more than	¥70,000	Less than	¥100,000	¥1,000
Equal to or more than	¥100,000	Less than	¥150,000	¥2,000
Equal to or more than	¥150,000	Less than	¥200,000	¥3,000
Equal to or more than	¥200,000	Less than	¥300,000	¥4,000
Equal to or more than	¥300,000	Less than	¥500,000	¥5,000
Equal to or more than	¥500,000	Less than	¥1,000,000	¥10,000
Equal to or more than	¥1,000,000	Less than	¥1,500,000	¥20,000
Equal to or more than	¥1,500,000	Less than	¥2,000,000	¥30,000
Equal to or more than	¥2,000,000	Less than	¥3,000,000	¥40,000
Equal to or more than	¥3,000,000	Less than	¥5,000,000	¥50,000
Equal to or more than	¥5,000,000	Less than	¥10,000,000	¥100,000
Equal to or more than	¥10,000,000	Less than	¥15,000,000	¥200,000
Equal to or more than	¥15,000,000	Less than	¥20,000,000	¥300,000
Equal to or more than	¥20,000,000	Less than	¥30,000,000	¥400,000
Equal to or more than	¥30,000,000	Less than	¥50,000,000	¥500,000
Equal to or more than	¥50,000,000			¥1,000,000

Table D.2: Special quote parameters at the TSE

When there is a major imbalance in orders, special bid or ask quotes are indicated by the TSE. Special quotes are disseminated publicly through the TSE market information system. If counter orders come into the market and the orders are matched at that price, the quote is withdrawn. Conversely, if the imbalance continues, the special quotes are revised up or down within reported parameters, at intervals of at least five minutes until the imbalance is resolved. Source: [244].

In addition to the measures listed above the TSE uses *off-auction trading* to handle large block orders. The off-auction trading system allows to accommodate large block order and basket order transactions, the trading of which is difficult under the competitive trading principals of auction trading.

*Clearing & Settlement.* The TSE utilizes the central counter-party (CCP) system and the clearing participant system for the settlement of trading on its exchange. In the CCP system a clearing organization (Japan Securities Clearing Corporation, JSCC) acts as the settlement counter-party to clearing participants for all trades on TSE's market, thereby guaranteeing settlement. This system eliminates the risk of trade counterparty default.

Previous day's closing price or special quote per share				Daily price limits ( $\pm$ )
		Less than	¥100	¥30
Equal to or more than	¥100	Less than	¥200	¥50
Equal to or more than	¥200	Less than	¥500	¥80
Equal to or more than	¥500	Less than	¥1,000	¥100
Equal to or more than	¥1,000	Less than	¥1,500	¥200
Equal to or more than	¥1,500	Less than	¥2,000	¥300
Equal to or more than	¥2,000	Less than	¥3,000	¥400
Equal to or more than	¥3,000	Less than	¥5,000	¥500
Equal to or more than	¥5,000	Less than	¥10,000	¥1,000
Equal to or more than	¥10,000	Less than	¥20,000	¥2,000
Equal to or more than	¥20,000	Less than	¥30,000	¥3,000
Equal to or more than	¥30,000	Less than	¥50,000	¥4,000
Equal to or more than	¥50,000	Less than	¥70,000	¥5,000
Equal to or more than	¥70,000	Less than	¥100,000	¥10,000
Equal to or more than	¥100,000	Less than	¥150,000	¥20,000
Equal to or more than	¥150,000	Less than	¥200,000	¥30,000
Equal to or more than	¥200,000	Less than	¥300,000	¥40,000
Equal to or more than	¥300,000	Less than	¥500,000	¥50,000
Equal to or more than	¥500,000	Less than	¥1,000,000	¥100,000
Equal to or more than	¥1,000,000	Less than	¥1,500,000	¥200,000
Equal to or more than	¥1,500,000	Less than	¥2,000,000	¥300,000
Equal to or more than	¥2,000,000	Less than	¥3,000,000	¥400,000
Equal to or more than	¥3,000,000	Less than	¥5,000,000	¥500,000
Equal to or more than	¥5,000,000	Less than	¥10,000,000	¥1,000,000
Equal to or more than	¥10,000,000	Less than	¥15,000,000	¥2,000,000
Equal to or more than	¥15,000,000	Less than	¥20,000,000	¥3,000,000
Equal to or more than	¥20,000,000	Less than	¥30,000,000	¥4,000,000
Equal to or more than	¥30,000,000	Less than	¥50,000,000	¥5,000,000
Equal to or more than	¥50,000,000			¥10,000,000

Table D.3: Daily price limits at the TSE

Source: [244].

Settlement for normal domestic equity transactions in the exchange market is made on the fourth business day starting from the transaction date ( $T + 3$ ). Delivery and receipt of domestic equities during settlement is made by an account transfer between JSCC's account and the clearing participant's account at Japan Securities Depository Center, Inc. (JASDEC). Transfer of funds in settlement is carried out between the clearing participant's bank account and JSCC's bank account at either the Bank of Japan or a bank it designates as a fund settlement bank. Delivery versus payment settlement is utilized via a netting basis in JSCC for all regular traded domestic stocks. In netting, only the net amount of securities bought and sold and net funds arising from these transactions are transferred between accounts.

The TSE has three trading sessions per day: from 8:20 till 9:00, from 11:00 till 12:30 and from 15:00 till 16:30. For half-trading days only the first two sessions take place. The total

trading time sums up to 240 minutes over 490 minutes between the opening and closing time.

## D.2. Data Description

The empirical part of the dissertation studies the data collected on the TSE. Data comes in a raw, tick-by-tick form, and includes mid-prices and volumes of each transaction. The precision of time stamps is one minute.<sup>3</sup> The dataset covers 2273 stocks traded at the TSE during a period from 11/03/1996 till 22/06/1998, i.e. 833 trading days.

Due to computational limitations<sup>4</sup> three subsets of time series were selected for the empirical study presented in Chapter 3. Under the assumption that the return (or price) predictability and, correspondingly, investors' attitude might be different for different levels of liquidity the dataset was divided into three groups based on the liquidity level of the underlying stocks.<sup>5</sup> Each subset was randomly (with withdrawing) populated by 9 time series, which belong to stocks with the same level of liquidity. For each group of liquidity median stocks provided the tenth time series, thus creating a high-, medium and low-liquidity pools of time series.

Table D.4 lists selected stocks and provides statistical properties of their time series of prices and returns. Statistical properties are obtained for daily data over the time period covered in the study.<sup>6</sup> A short description of the occupational activity of underlying companies accompanies each record.<sup>7</sup> Median in liquidity stock is marked by the asterisk next to a company name.

Figure D.1 presents equally weighted indexes for each liquidity pool of TSE stocks (thick black line). Prior to aggregation each constituent time series was rebased<sup>8</sup> by its first observation in the series (thin cyan lines).

<sup>3</sup>At the same there are many observations stamped with the same time, which allows to use time-shift technique to maximize data utilization.

<sup>4</sup>Using optimized and compiled C++ code it takes Pentium 2.8 GHz to perform one data subsample in 0.5 seconds. With the methodology to maximize the efficiency of data utilization (see Section 2.1.2) there are approximately  $833 \text{ days} \times 240 \text{ minutes} \times 60 \text{ seconds} = 12 \cdot 10^6$  data-points for each stock in the study. Thus, processing one stock takes around 70 days of non-stop computing.

<sup>5</sup>The criterion of liquidity is the number of trades over the covered time period.

<sup>6</sup>Statistical properties covers the following measures:  $\mu$  - mean;  $\sigma$  - sample variance;  $\beta$  - sample slope;  $K$  - kurtosis;  $S$  - skewness;  $\alpha$  - tail index. Tail index is computed for 5% of extreme values using the Hill estimator [103].

<sup>7</sup>The table uses the following sector codes: (1) - Nonlife Insurance; (2) - Industrial Metals; (3) - Leisure Goods; (4) - Automobiles & Parts; (5) - Technological Hardware & Equipment; (6) - Electrical, Electronic Equipment; (7) - Beverages; (8) - Chemicals; (9) - Food Producers; (10) - Industrial Engineering; (11) - Personal Goods; (12) - Pharmacology, Biotechnology; (13) - Health Equipment & Services; (14) - General Retailers; (15) - Food & Drug Retailers; (16) - General Financial.

<sup>8</sup>Rebasing means that the level of time series was shifted up or down by equal amount. This allows to see the trend development of several time series.

##	TSE code	Company name	Sector (code)	Statistics of price series			and returns		
				$\mu$	$\sigma$	$\beta$	$K$	$S$	$\alpha$
1. High-liquidity pool									
1.1	8753	Sumitomo Marine & Fire Insurance	(1)	814.98	90.26	-0.27	2.40	-0.20	3.24
1.2	5407	Nisshin Steel	(2)	288.24	97.34	-0.54	1.81	0.17	3.39
1.3	6765	Kenwood	(3)	524.83	122.22	-0.59	2.07	-0.16	2.72
1.4	7262	Daihatsu Motor	(4)	613.29	71.99	-0.32	2.05	-0.38	4.55
1.5	6913	Melco	(5)	3053.31	917.88	-4.39	1.72	-0.47	3.32
1.6	7752	Ricoh	(5)	1386.98	211.53	0.82	2.06	0.24	3.68
1.7	6501	Hitachi	(6)	1056.57	107.76	-0.13	3.34	0.90	4.49
1.8	2503	Kirin Brewery	(7)	1140.67	119.82	-0.38	2.16	0.19	5.00
1.9	6752	Matsushita Elec. Industrial	(3)	1992.27	178.00	0.57	2.46	0.60	3.82
1.10	4401	Adeka*	(8)	772.51	164.61	-0.87	1.83	-0.31	2.79
2. Medium-liquidity pool									
2.1	1351	Hoko Fishing	(9)	245.24	112.93	-0.61	1.70	0.22	2.54
2.2	7738	Chinon Industries	(10)	857.62	480.91	-2.62	1.55	0.32	2.55
2.3	8112	Tokyo Style	(11)	1517.73	239.48	-1.23	1.79	0.29	3.01
2.4	4544	Miraca Holdings	(12)	710.59	212.03	-1.10	1.86	-0.39	3.68
2.5	4003	CO-OP Chem.	(8)	256.82	123.70	-0.70	1.74	0.19	3.06
2.6	7739	Canon Electr.	(5)	818.12	81.70	0.00	2.60	0.26	3.87
2.7	6772	Tokyo Cosmos Electric	(6)	326.55	110.50	-0.59	1.84	0.32	3.34
2.8	2571	Chunkyo Coca-Cola	(7)	1054.55	114.99	-0.46	2.31	0.24	3.29
2.9	6910	Hitachi Medical	(13)	1569.24	196.23	-0.86	2.30	-0.33	4.86
2.10	7723	Aichi Tokei Denki*	(6)	458.38	171.07	-0.96	1.67	0.16	2.45
3. Low-liquidity pool									
3.1	8291	Tonichi Carlife GP	(14)	503.26	164.28	-0.86	1.96	0.58	3.33
3.2	2898	Sonton Food Ind.	(9)	1296.90	146.96	-0.72	2.32	-0.32	2.91
3.3	6776	Tensho Elec. Ind.	(10)	352.64	120.43	-0.60	1.84	-0.05	2.81
3.4	8229	CFS	(15)	1121.76	253.27	-1.42	1.84	0.05	3.18
3.5	6360	Tokyo Auto Machine Works	(10)	472.81	156.88	-0.86	1.84	-0.03	1.68
3.6	8623	Smbc Friend Sec.	(16)	295.27	146.52	-0.80	1.99	0.51	3.66
3.7	2893	Lohmeyer	(9)	269.37	109.71	-0.60	1.72	0.17	2.43
3.8	5610	Daiwa Heavy Ind.	(10)	278.31	110.56	-0.61	1.86	0.38	2.91
3.9	7942	JSP	(8)	1021.06	207.80	-1.07	2.55	-0.61	3.29
3.10	6396	Unozawa-Gumi Iron Works*	(10)	426.58	118.27	-0.56	3.91	1.02	2.07

Table D.4: List and description of stocks used in the empirical study

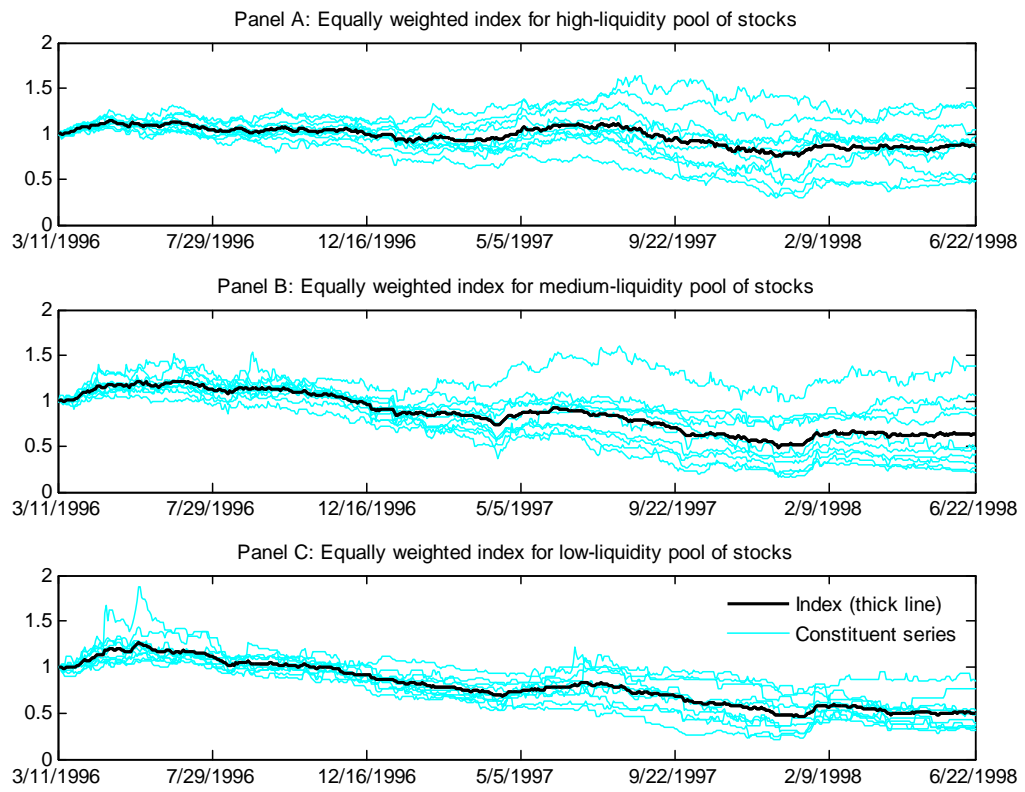


Figure D.1: Equally weighted indexes for liquidity pools of TSE stocks

## APPENDIX E

### Comparative Analysis of PLA Algorithms

This appendix presents the graphs to illustrate the analysis of different piecewise linear approximation algorithms (for details see Chapter 2.2.6).

Each figure displays either the statistical properties of the algorithms, or their comparison. Normally, panel A illustrates the sample time series (in blue) and the examples of linear approximation, distinguished in colors. Panels B, C, and D presents the empirical cumulative distribution functions (CDF) for the sum of squared errors (SSE),  $R^2$ , and the time required for approximation, correspondingly. The color of CDF and the example linear approximations corresponds for each panel. The empirical CDF are computed after  $10^5$  repetitions.

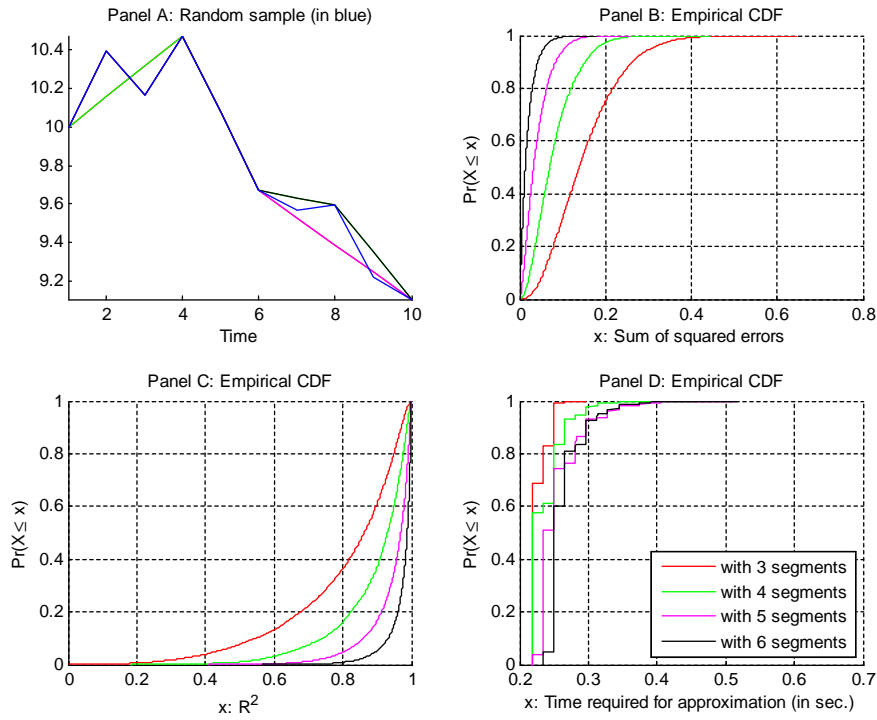


Figure E.1: GA-based approximation of 10-points time series

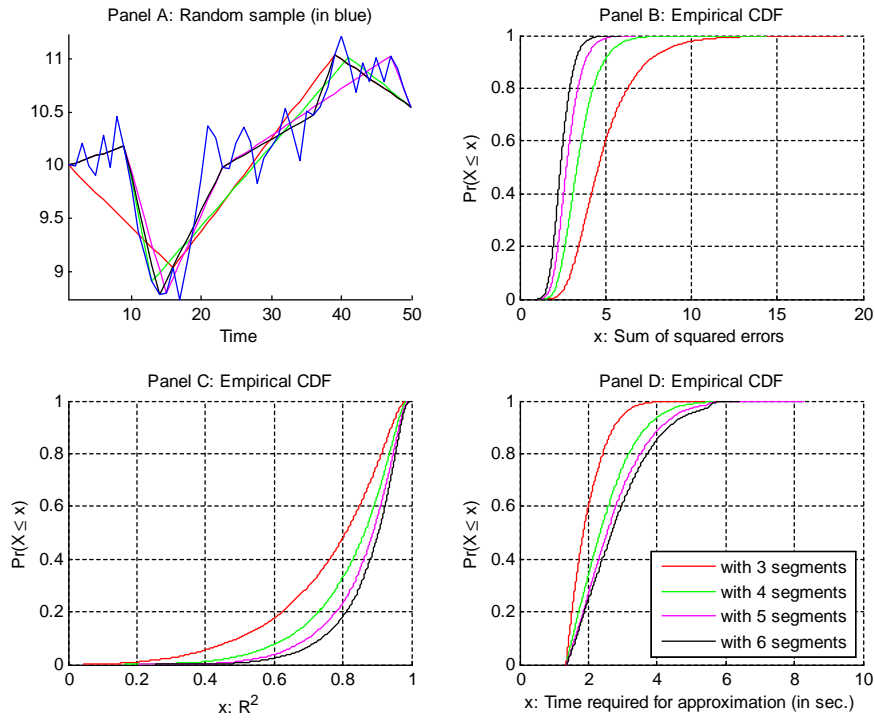


Figure E.2: GA-based approximation of 50-points time series

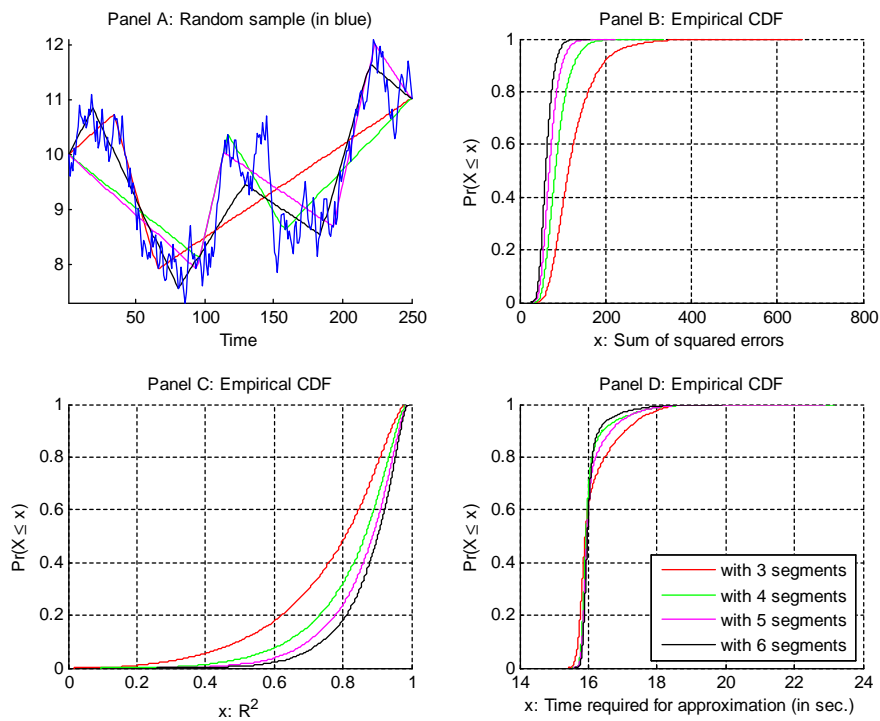


Figure E.3: GA-based approximation of 250-points time series



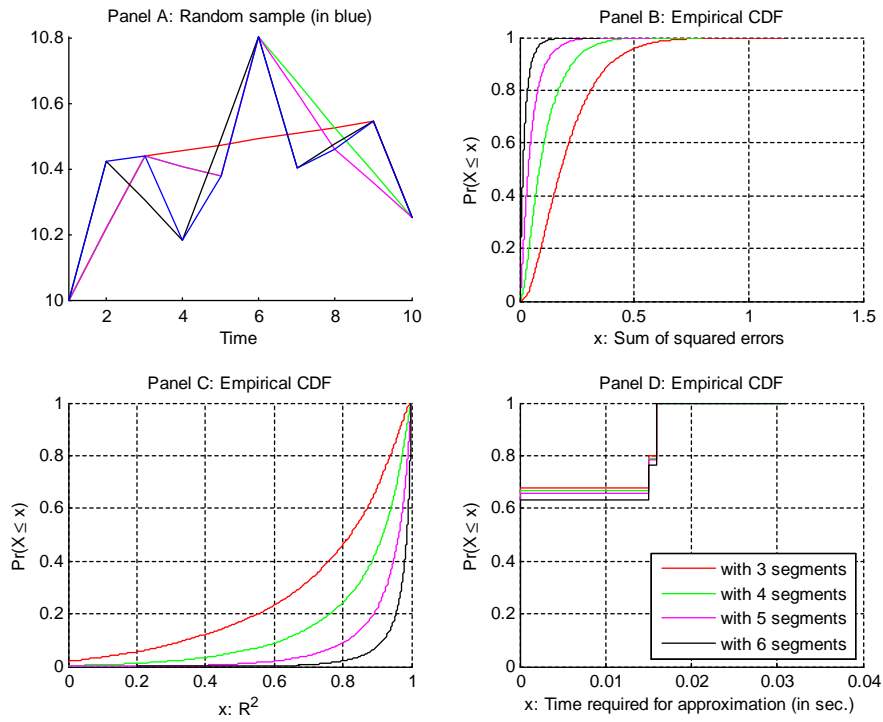


Figure E.4: Sliding Window approximation of 10-points time series

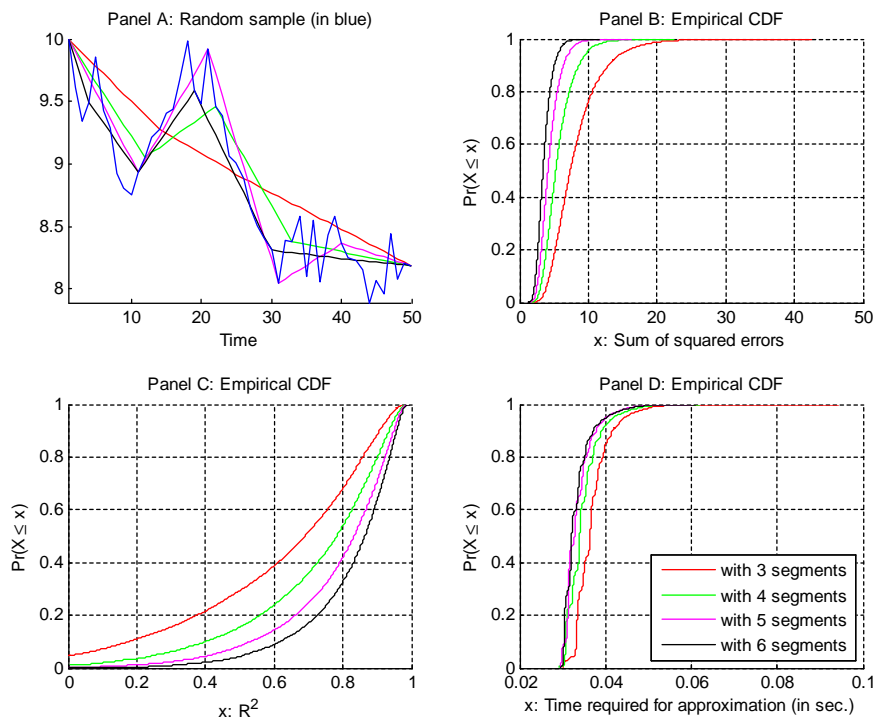


Figure E.5: Sliding Window approximation of 50-points time series

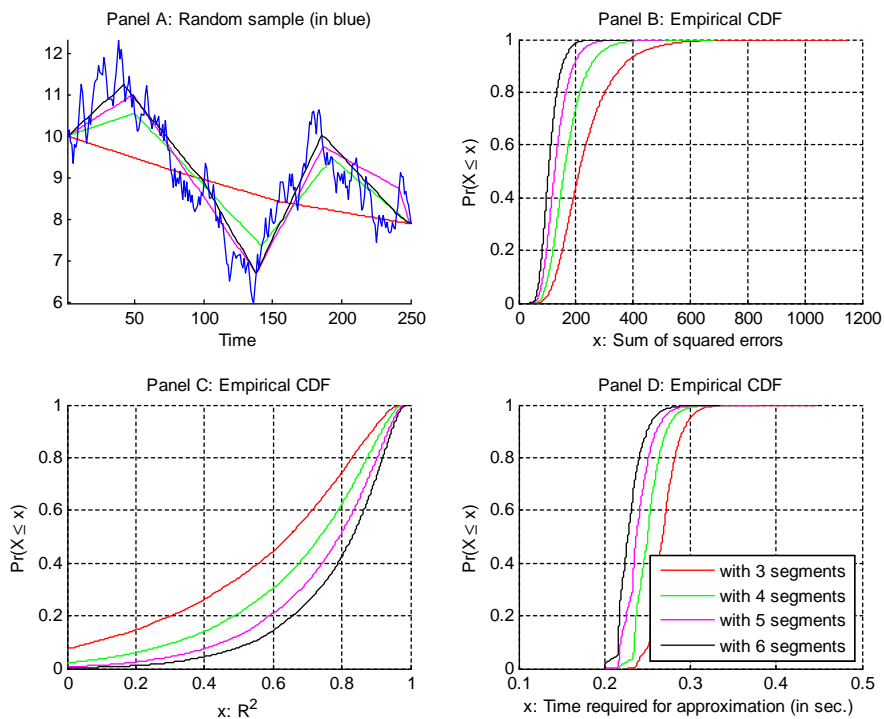


Figure E.6: Sliding Window approximation of 250-points time series

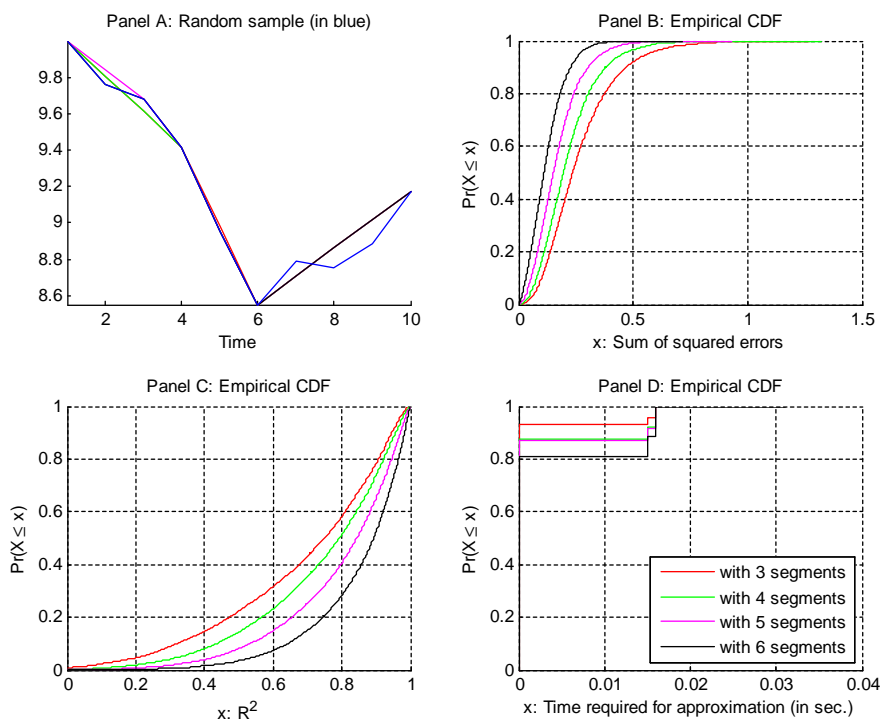


Figure E.7: Top-Down approximation of 10-points time series

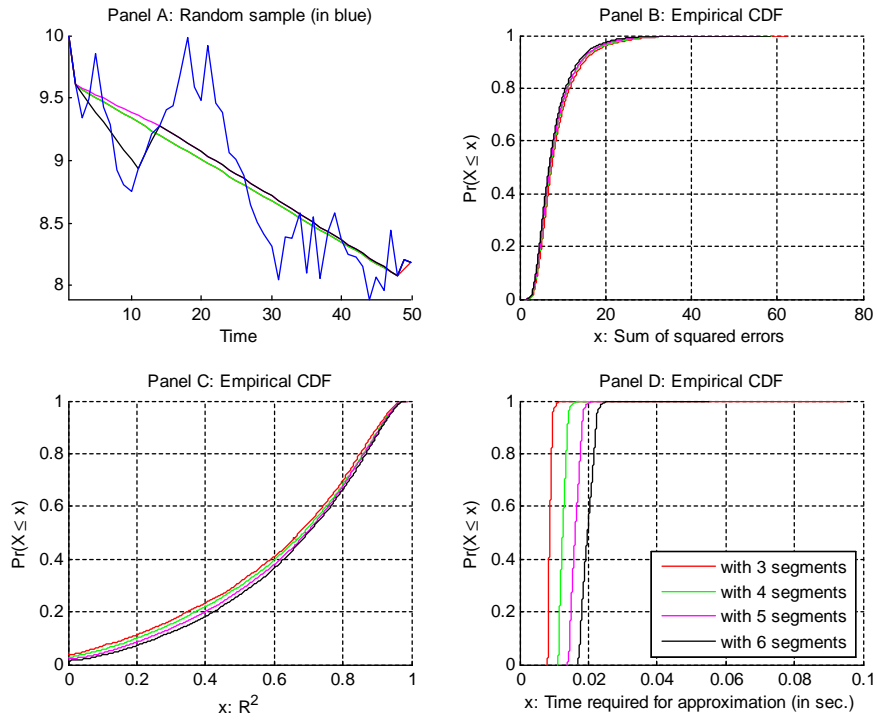


Figure E.8: Top-Down approximation of 50-points time series

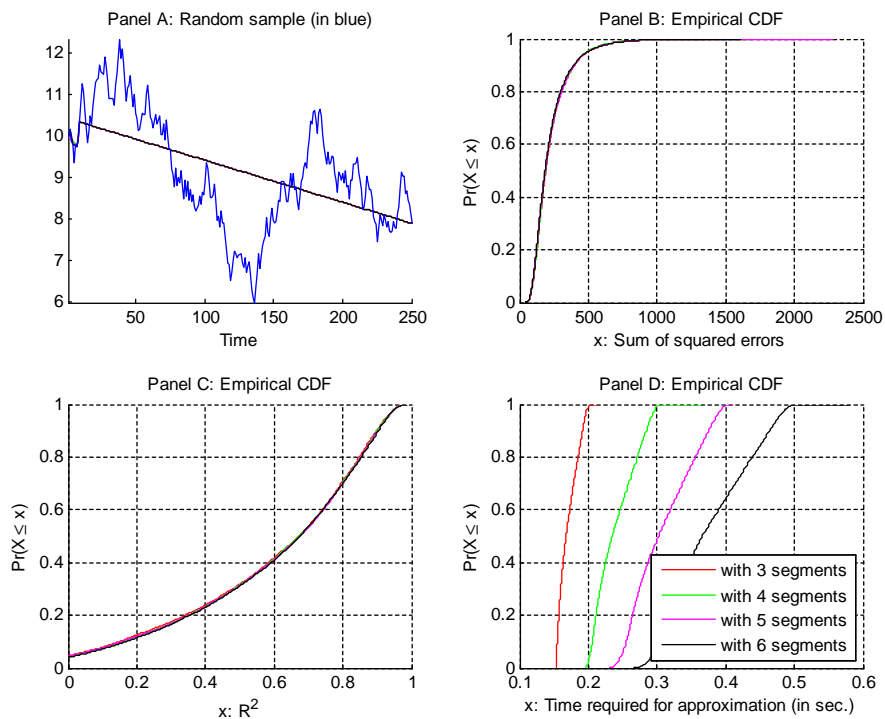


Figure E.9: Top-Down approximation of 250-points time series

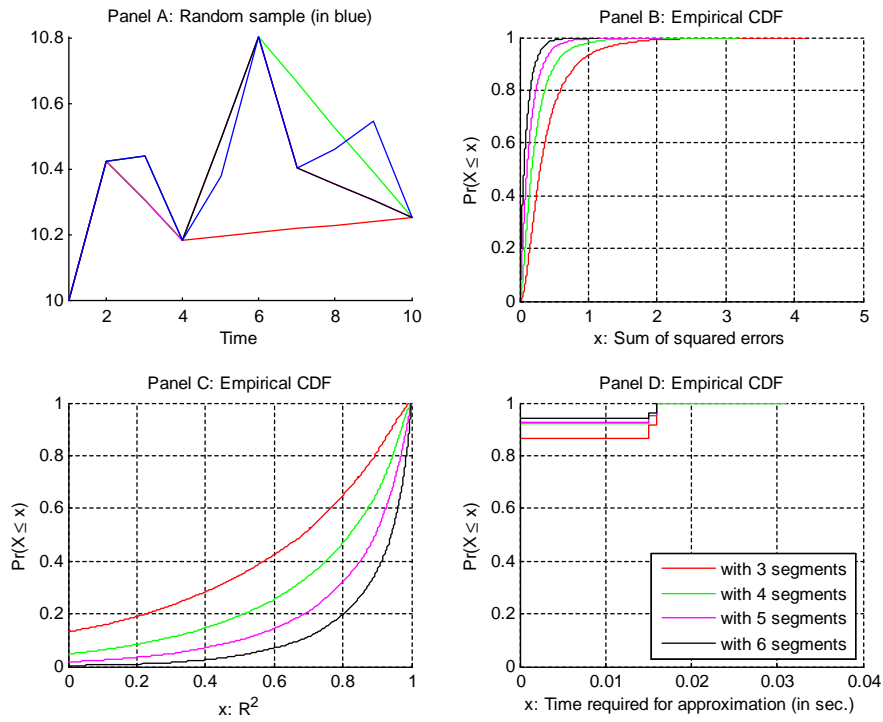


Figure E.10: Bottom-Up approximation of 10-points time series

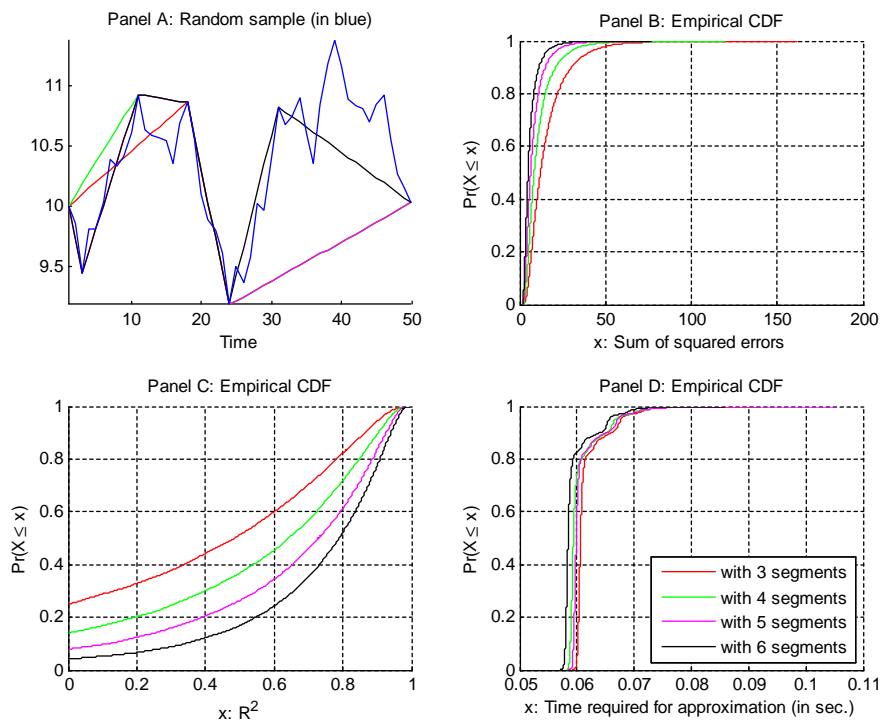


Figure E.11: Bottom-Up approximation of 50-points time series

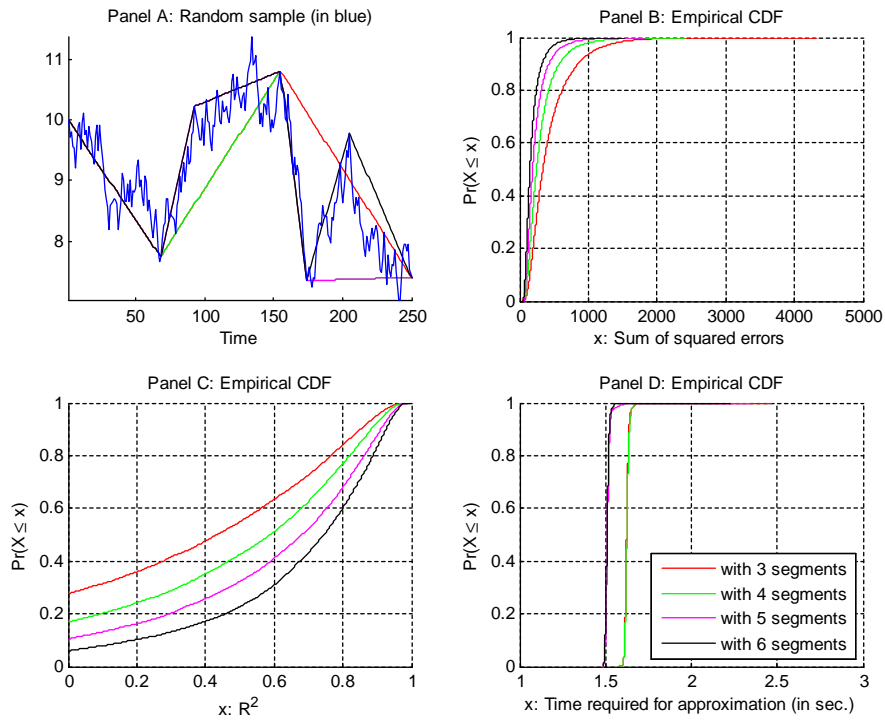


Figure E.12: Bottom-Up approximation of 250-points time series

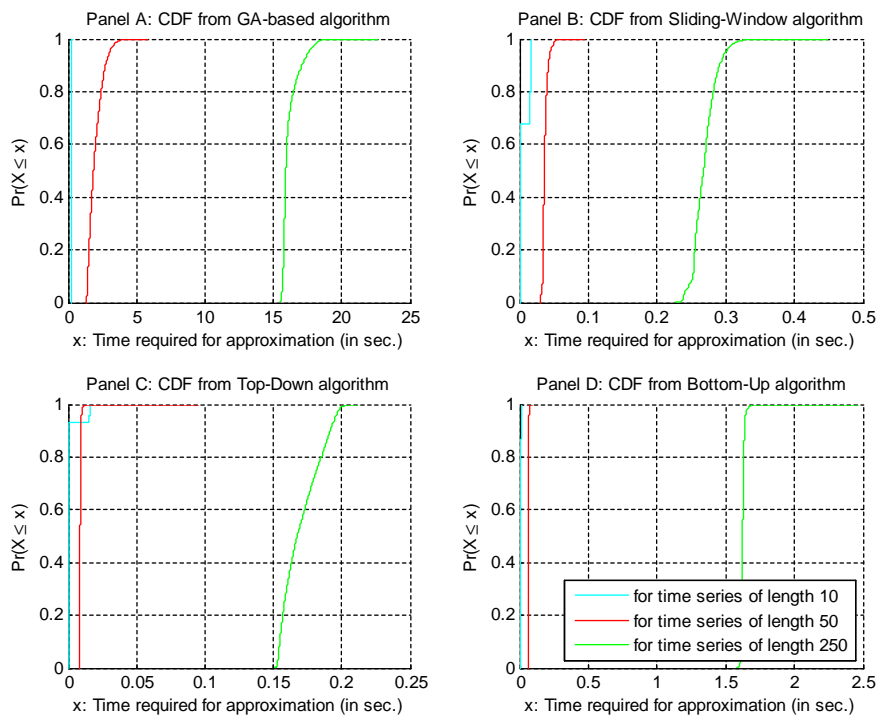


Figure E.13: PLA with 3 linear segments

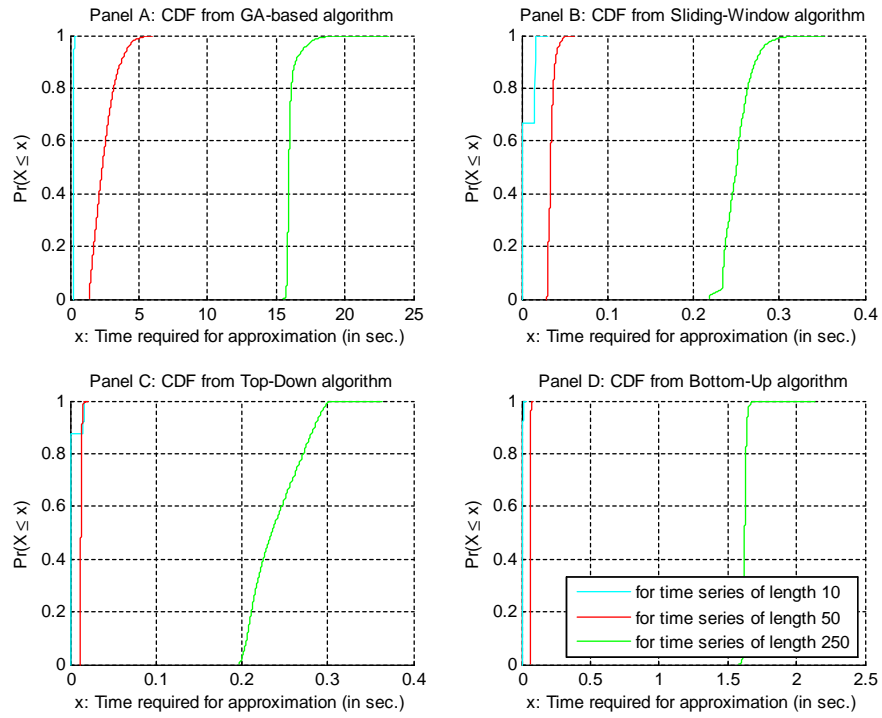


Figure E.14: PLA with 4 linear segments

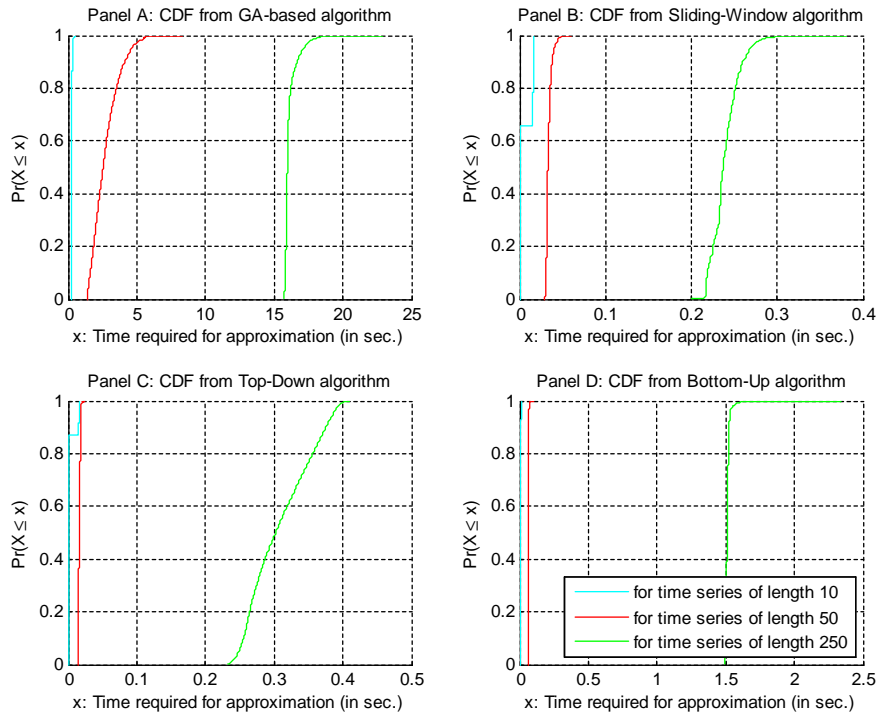


Figure E.15: PLA with 5 linear segments

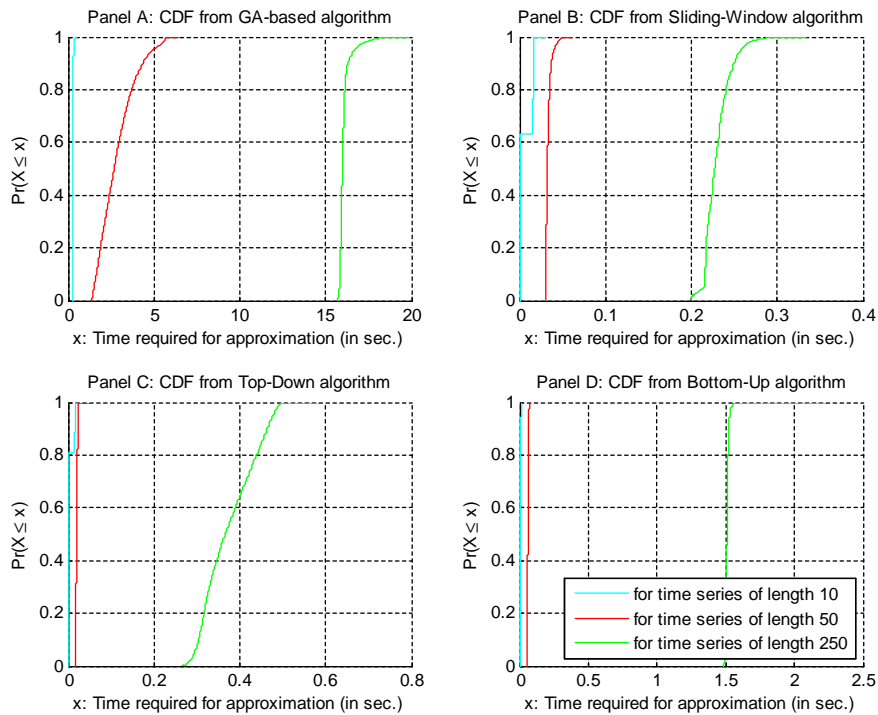


Figure E.16: PLA with 6 linear segments

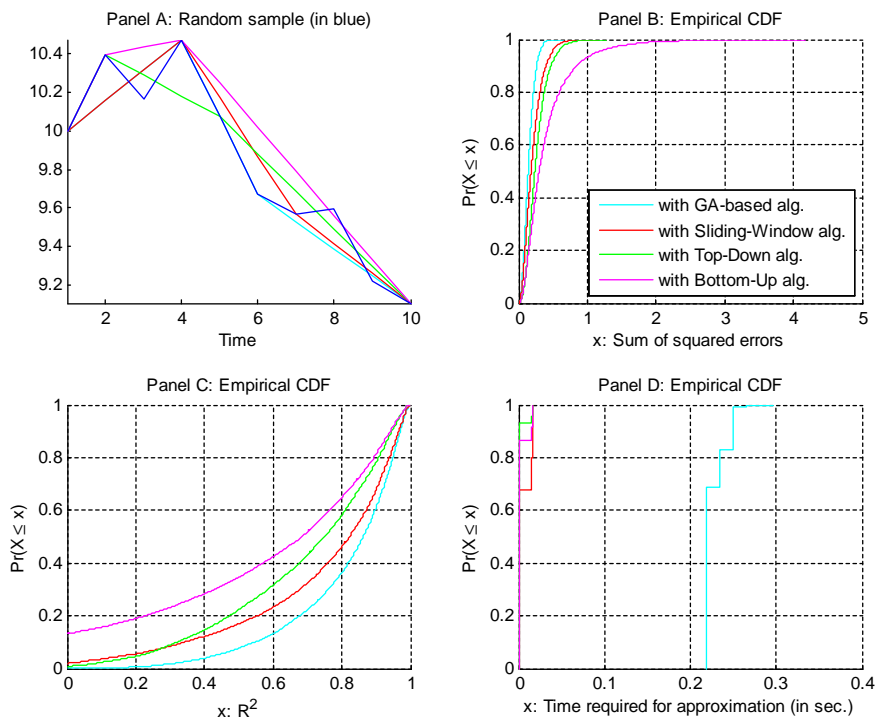


Figure E.17: PLA of 10-points time series with 3 linear segments

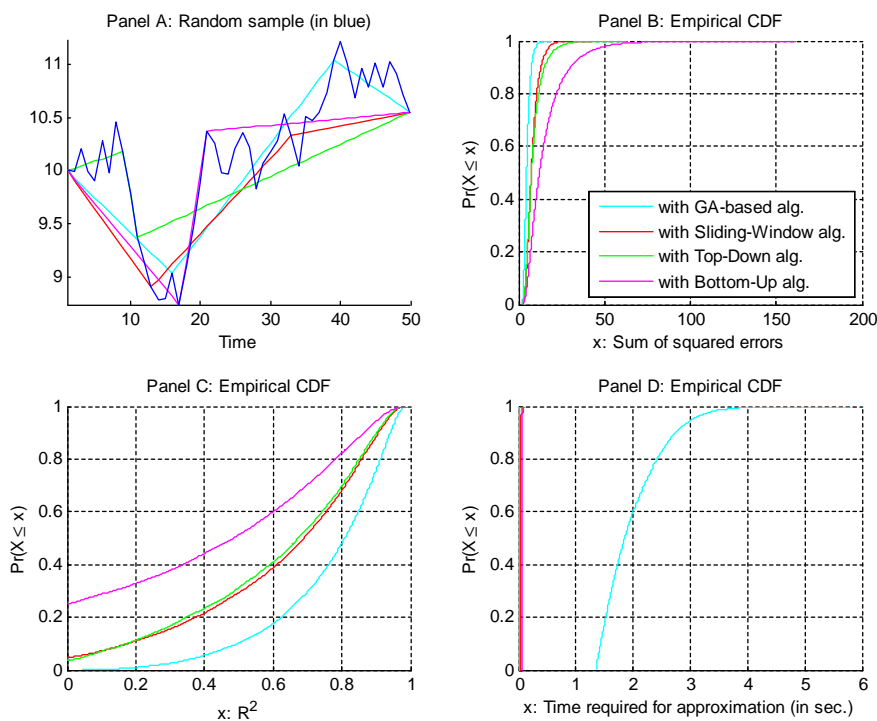


Figure E.18: PLA approximation of 50-points time series with 3 linear segments



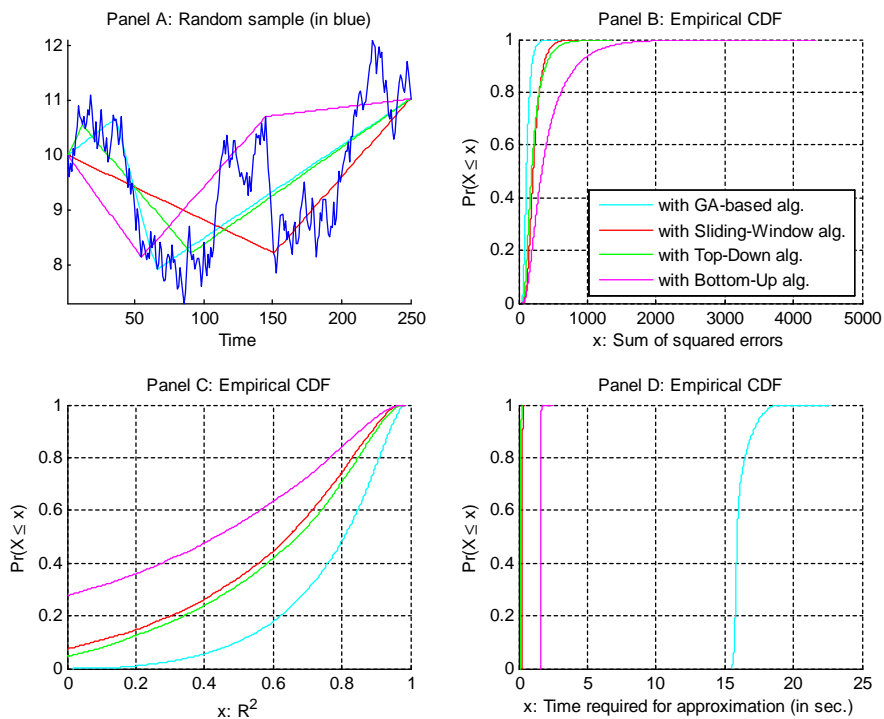


Figure E.19: PLA approximation of 250-points time series with 3 linear segments

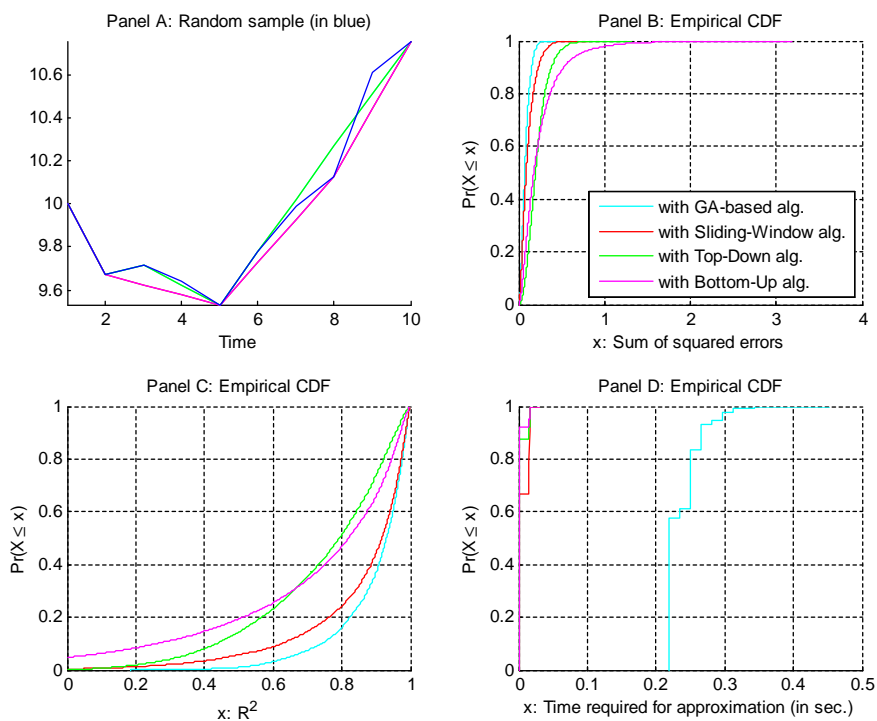


Figure E.20: PLA approximation of 10-points time series with 4 linear segments

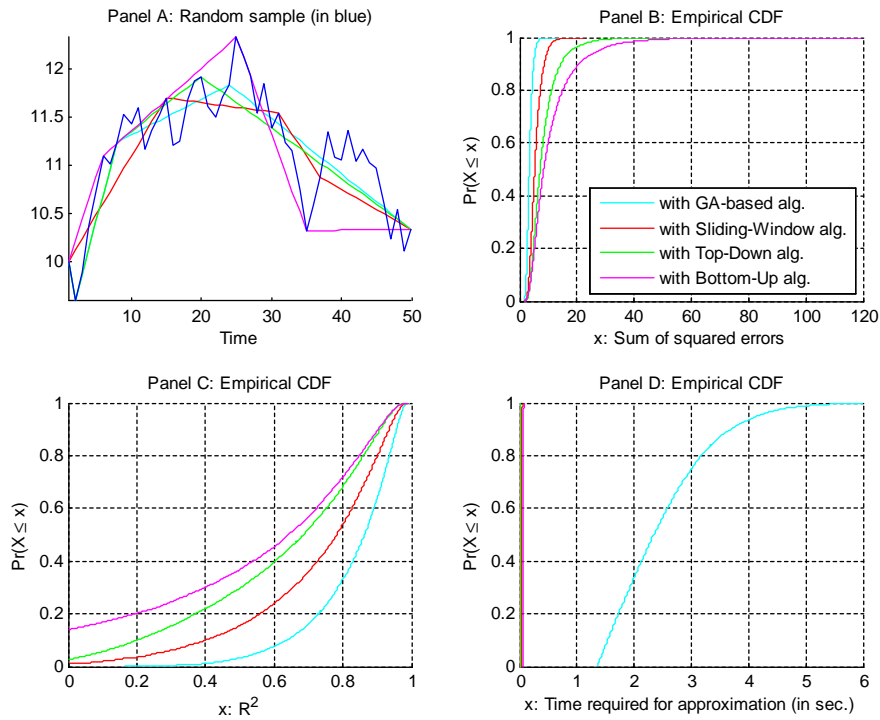


Figure E.21: PLA approximation of 50-points time series with 4 linear segments

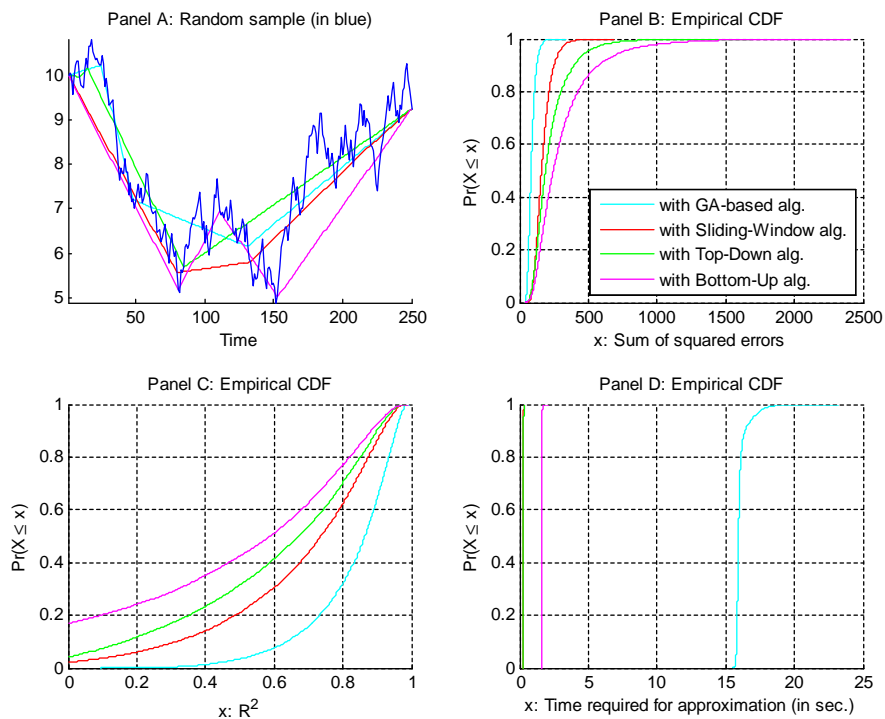


Figure E.22: PLA approximation of 250-points time series with 4 linear segments

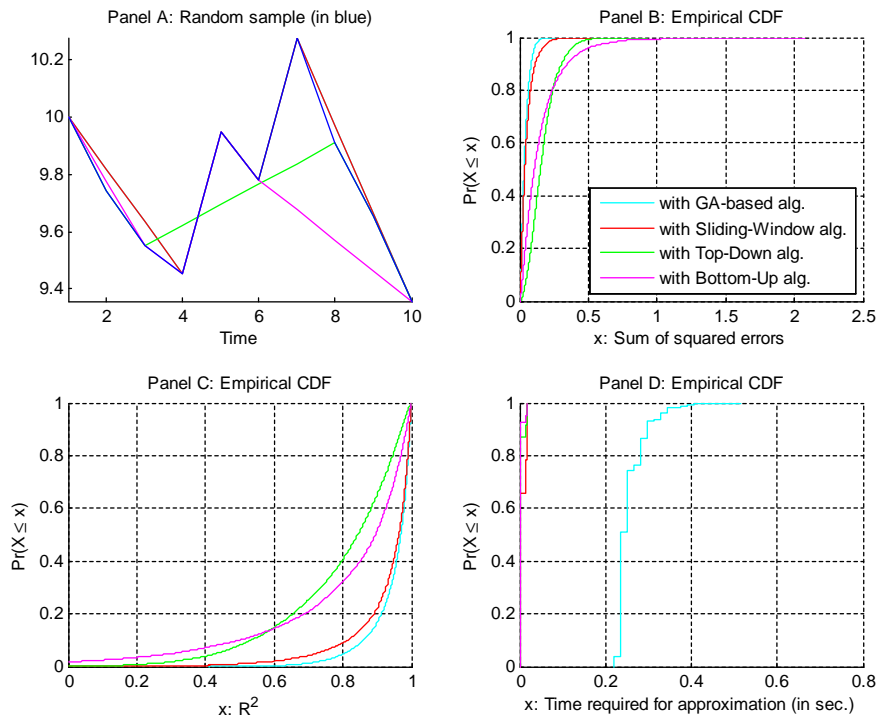


Figure E.23: PLA approximation of 10-points time series with 5 linear segments

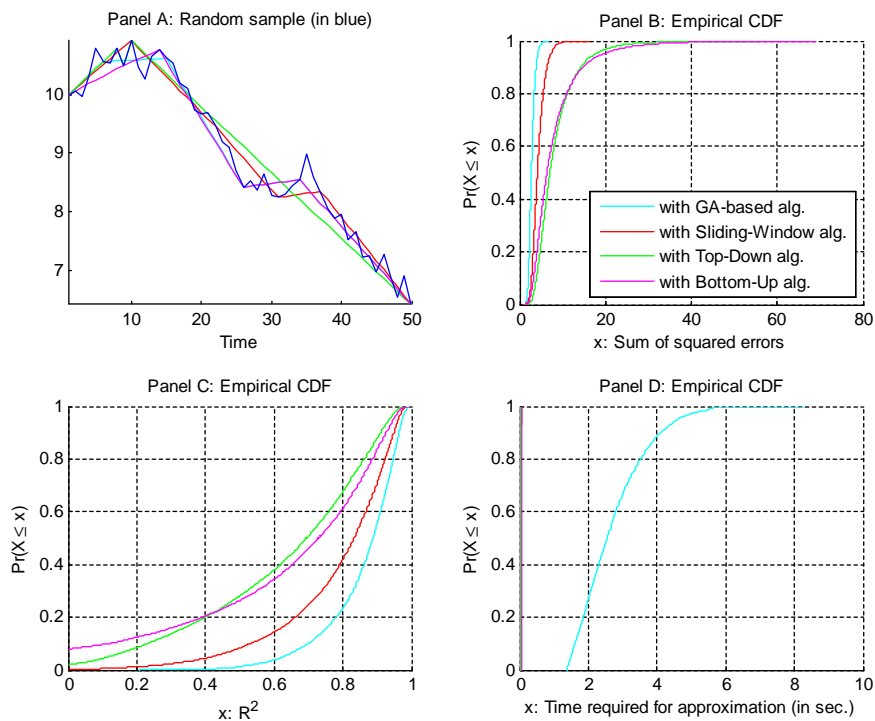


Figure E.24: PLA approximation of 50-points time series with 5 linear segments

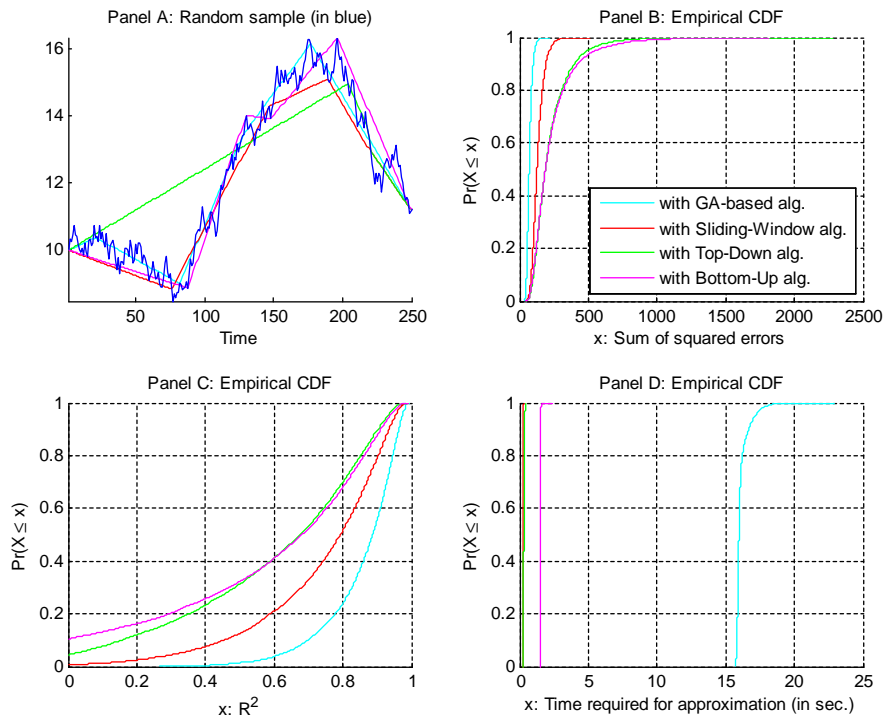


Figure E.25: PLA approximation of 250-points time series with 5 linear segments

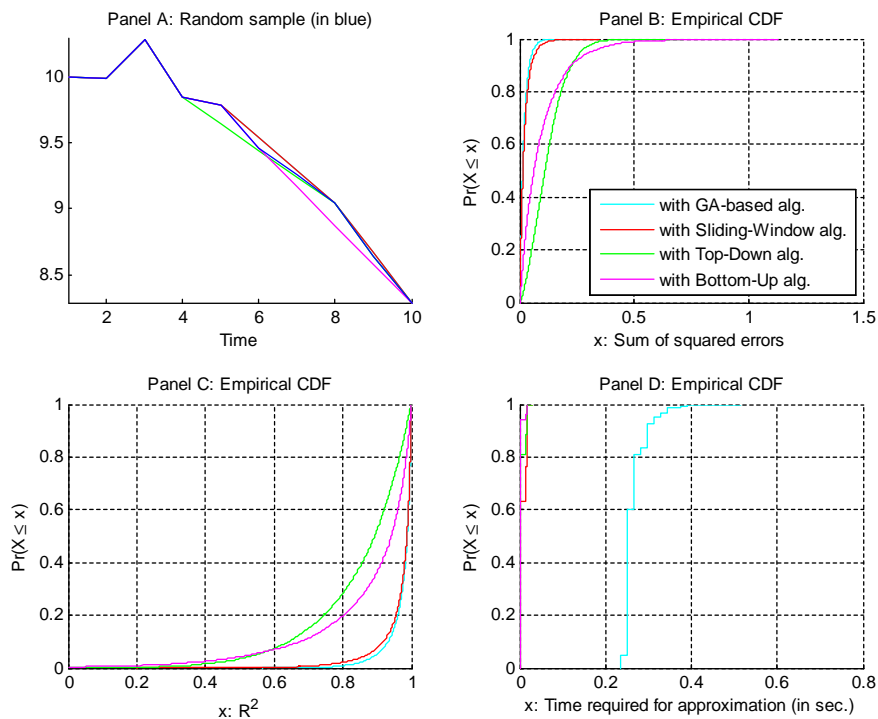


Figure E.26: PLA approximation of 10-points time series with 6 linear segments

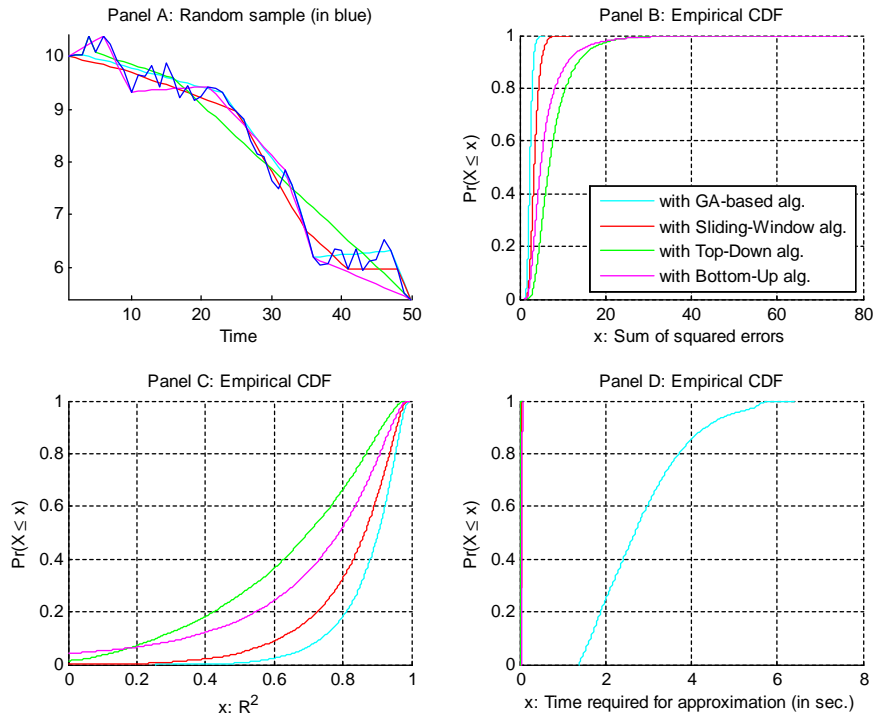


Figure E.27: PLA approximation of 50-points time series with 6 linear segments

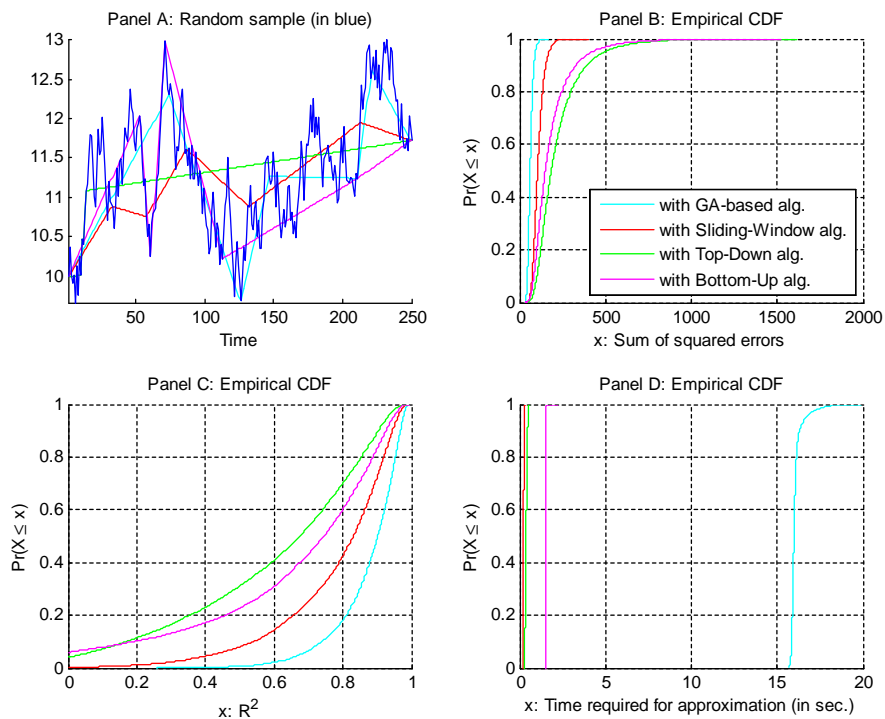


Figure E.28: PLA approximation of 250-points time series with 6 linear segments

## APPENDIX F

### **Application of TiCS and Benchmark Trading Strategies on the TSE data: Empirical Results**

This appendix presents figures to illustrate the application of the benchmark trading strategies on the TSE data. The description of the data and their market of origin is presented in Appendix D. The set up of the trading strategies and the interpretation of results is presented in Chapter 3.

Each figure displays breakeven separation surfaces and slice views of aggregate return (AR) adjusted for transaction costs (TC). The values of AR adjusted for TC are averaged over a pool of stocks of high, medium or low liquidity. Each row of a figure illustrates results for a different investment time span. The first rows are for the whole time span. Rows 2 - 4 are for the first, second and third time spans, respectively.

In each figure the first column of panels illustrate breakeven separation surfaces in three-dimensional space. The surface shows at which level of transaction costs, trade frequency and subsample size the AR adjusted for TC equals to one. Any point above (below) the surface indicates the combination at which the AR adjusted for TC is less (more) than one, i.e. loss (gain) on investments. The top point of the separation surface indicates the maximum breakeven transaction costs and the corresponding optimal trade frequency and subsample size.

The second and third columns of panels illustrate slice views of corresponding separation surfaces. In the second (third) column the views are sliced through the respective optimal subsample sizes (frequencies).<sup>1</sup> In color notation the AR adjusted for TC are showed. Panels L contain the colorbar to interpret color-denoted aggregate returns. The location of thick dots indicates the breakeven TC at given trade frequencies or subsample sizes.

---

<sup>1</sup>The titles of these panels contain in brackets the optimal values of subsample sizes and frequencies.

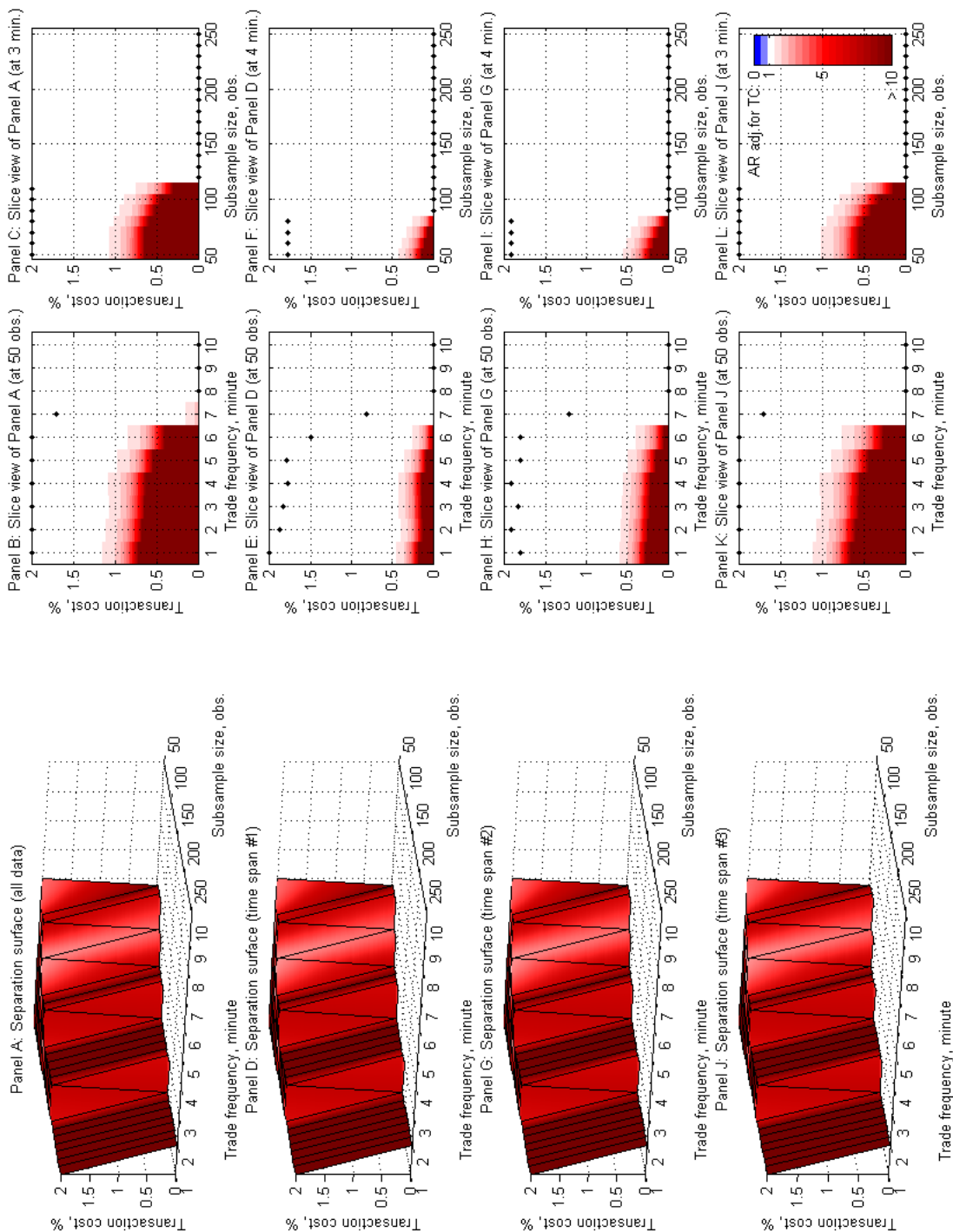


Figure F.1: Application of perfect forecast trading strategies on the pool of high liquidity stocks

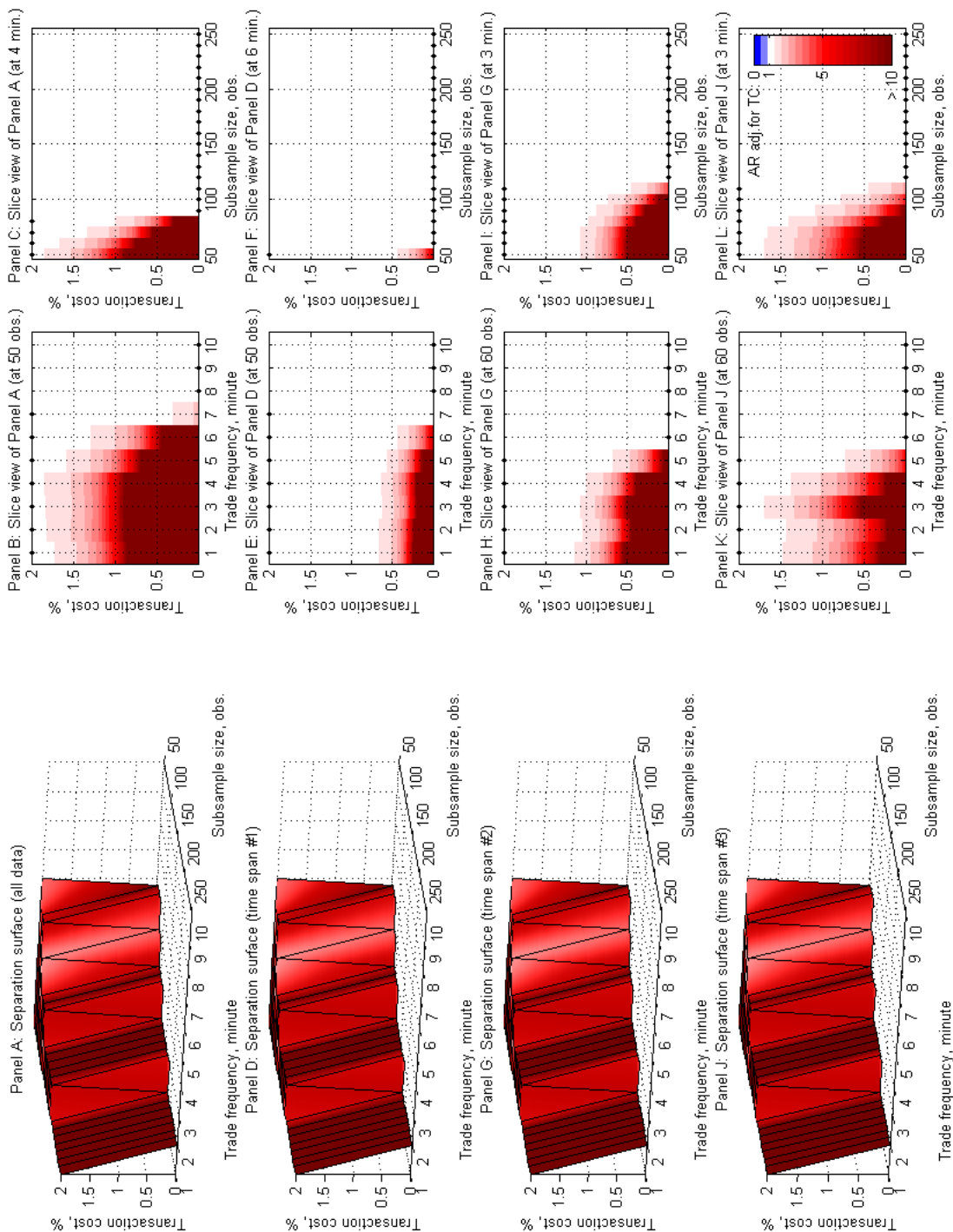


Figure F.2: Application of perfect forecast trading strategies on the pool of medium liquidity stocks



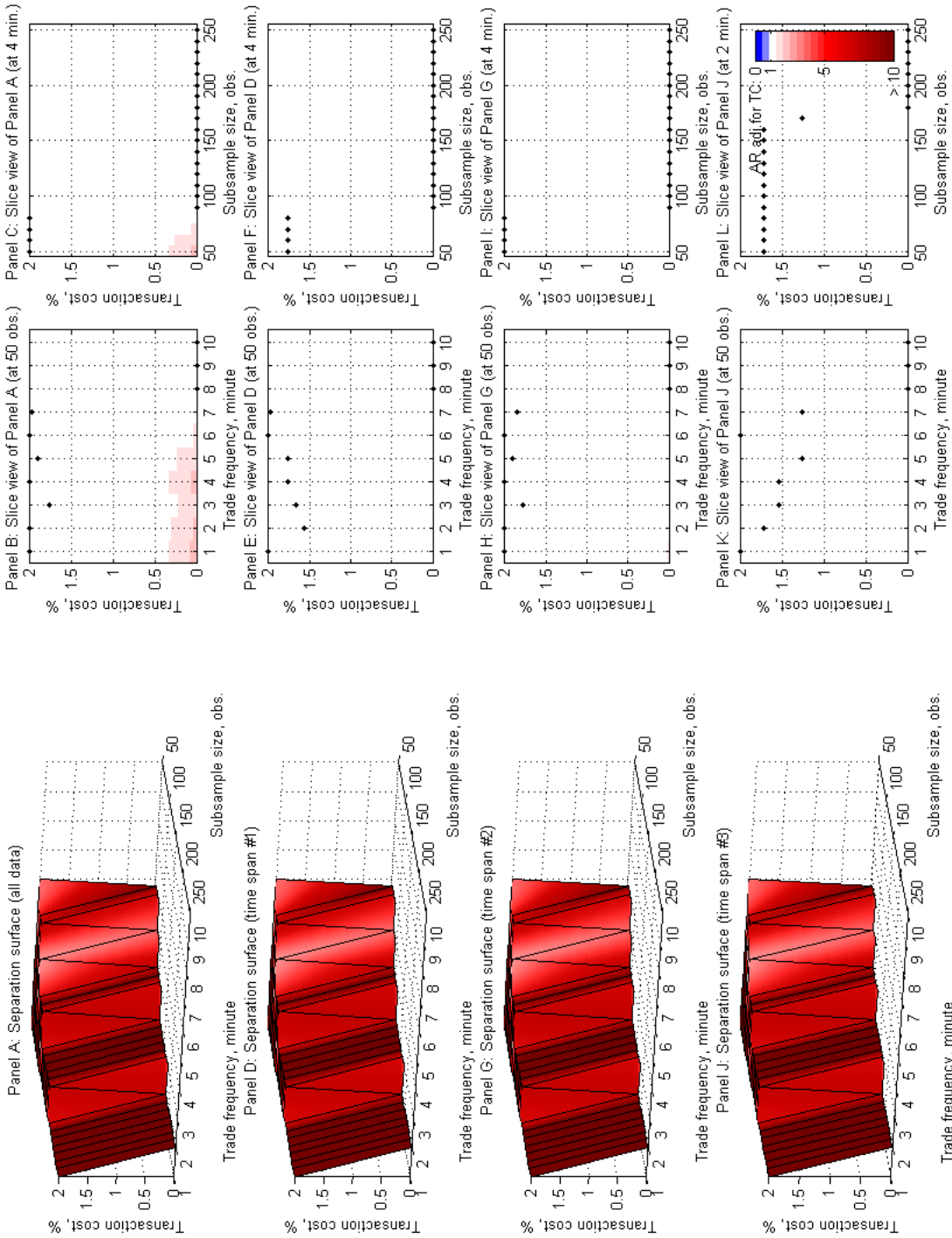


Figure F.3: Application of perfect forecast trading strategies on the pool of low liquidity stocks

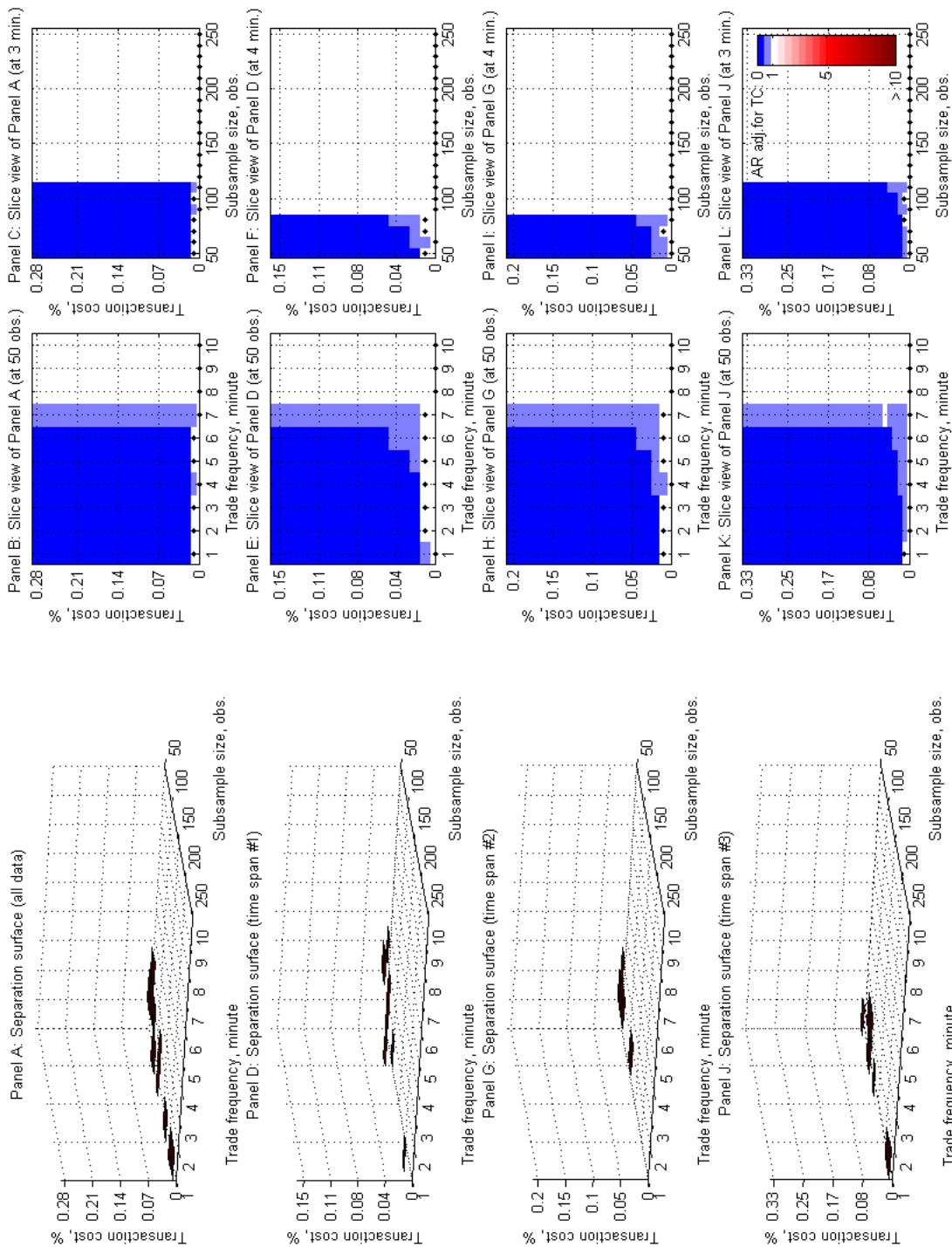


Figure F.4: Application of random walk trading strategies on the pool of high liquidity stocks

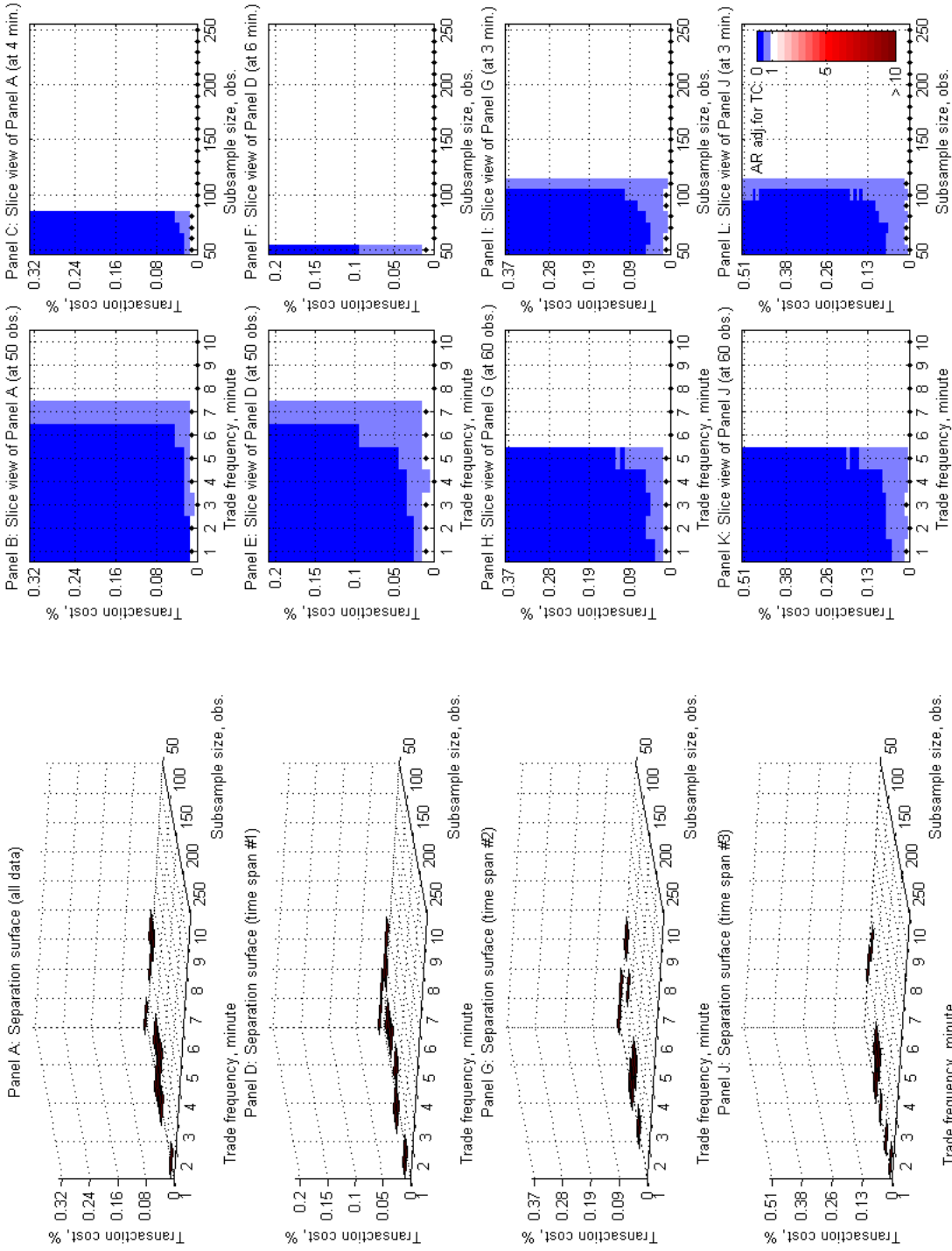


Figure F.5: Application of random walk trading strategies on the pool of medium liquidity stocks

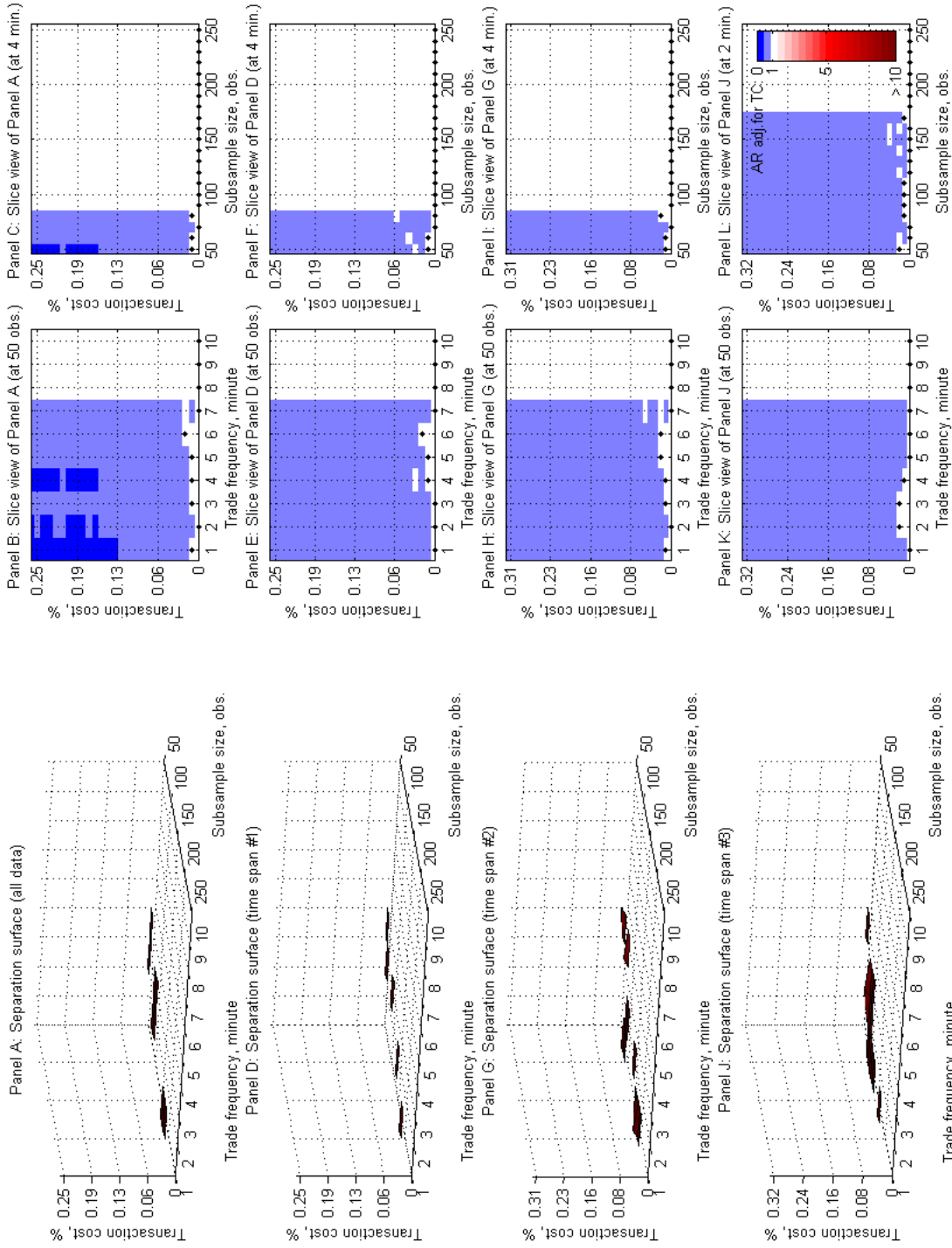


Figure F.6: Application of random walk trading strategies on the pool of low liquidity stocks

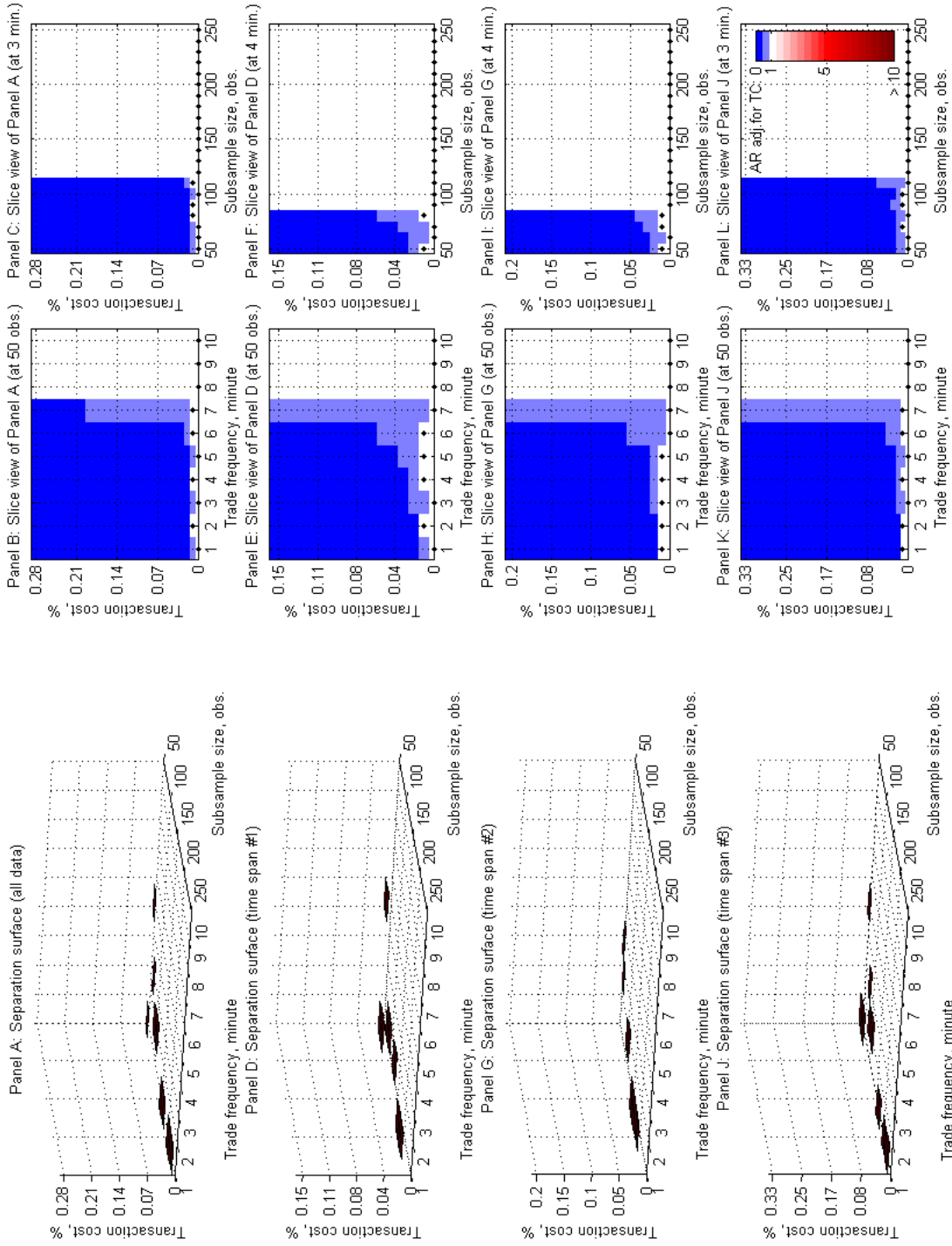


Figure F.7: Application of random trading strategies on the pool of high liquidity stocks

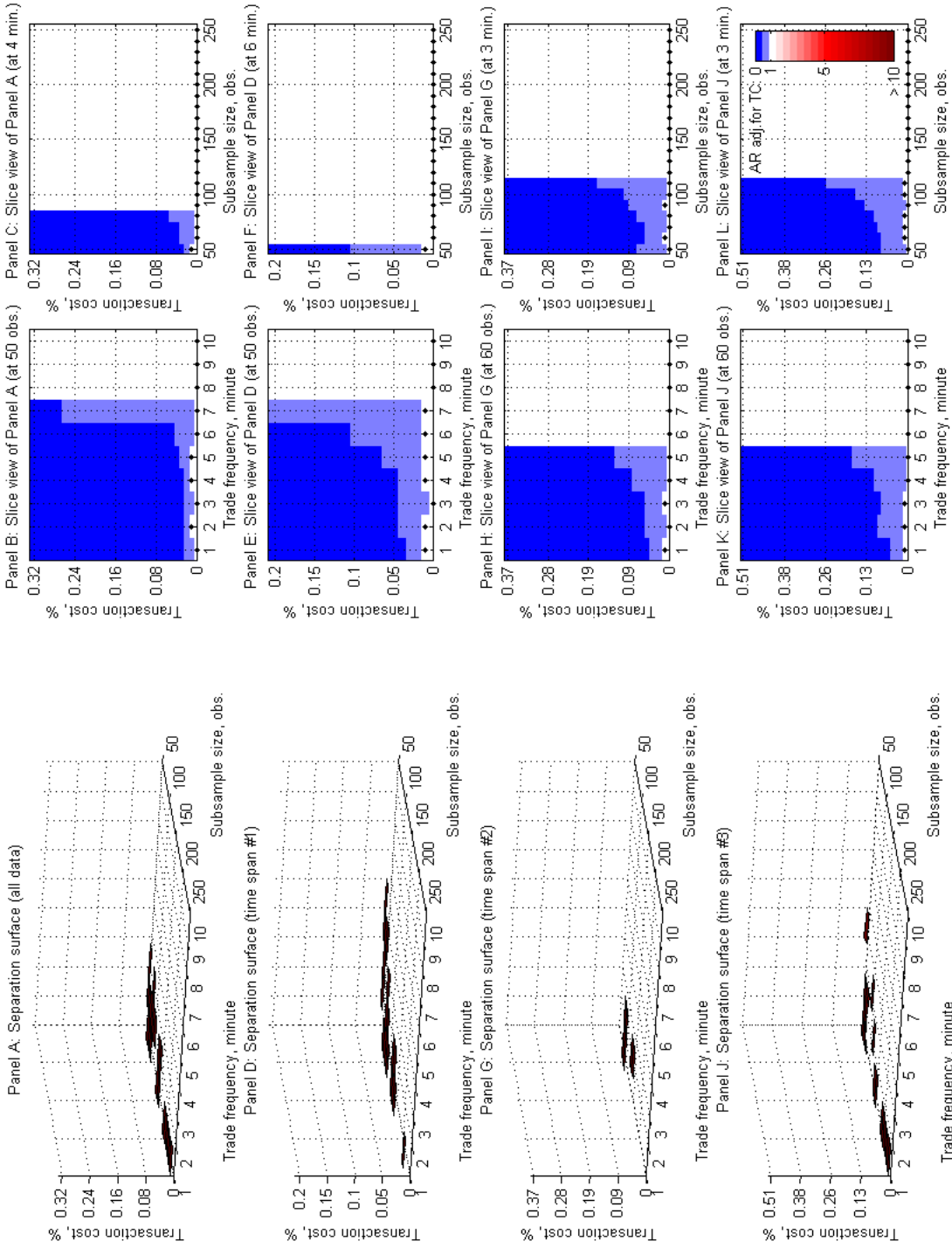


Figure F.8: Application of random trading strategies on the pool of medium liquidity stocks

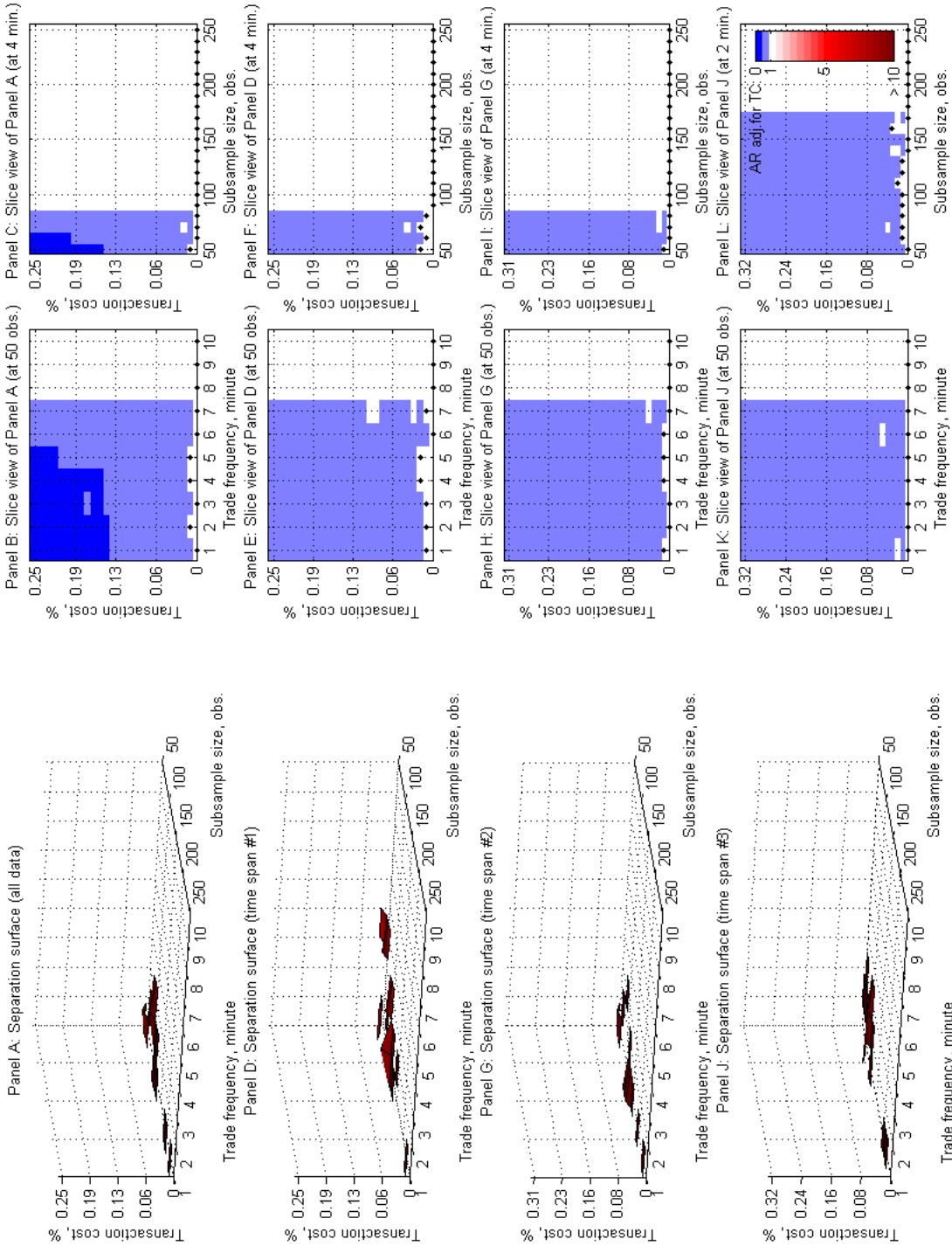


Figure F.9: Application of random trading strategies on the pool of low liquidity stocks

## APPENDIX G

### Abbreviations

AI	Artificial Intelligence
AMEX	American Stock and Options Exchange
AMH	Adaptive Market Hypothesis
ANN	Artificial Neural Networks
ASA	American Statistical Association
ASM	Artificial Stock Market
AR	Aggregate Return
ATA	Artificial Technical Analyst
BTC	Breakeven Transaction Costs
CDF	Cumulative Distribution Function
CS	Classifier Systems
DJIA	Dow Jones Industrial Average
EMH	Efficient Market Hypothesis
G	General
GA	Genetic Algorithms
GP	Genetic Programming
LCS	Learning Classifier Systems
$\mu$ GA	Micro-Genetic Algorithm
NYSE	New York Stock Exchange
PLA	Piecewise Linear Approximation
RTS	Revealed Transaction Costs
RW	Random Walk
RWH	Random Walk Hypothesis
S	Soldier
SGA	Simple Genetic Algorithm
SSE	Sum of Squared Errors
TA	Technical Analysis
TC	Transaction Costs
TiCS	True Classifier System
TRA	Time Required for Approximation
TSE	Tokyo Stock Exchange
XCS	Willson's X-Classifier System



## References

- [1] Acar, E., *Expected Returns of Directional Forecasters*, in *Advanced Trading Rules*, ed. E. Acar and S. E. Satchell, 2nd ed., Oxford: Butterworth Heinemann, pp. 51-80, 2002.
- [2] Agrawal, A., J. F. Jaffe, G. N. Mandelker, *The Post-Merger Performance of Acquiring Firms: A Reexamination of an Anomaly*, *Journal of Finance* 47, pp. 1605-1621, 1992.
- [3] Alexander, S. S., *Price Movements in Speculative Markets: Trends or Random Walks?*, *Industrial Management Review* 2, pp. 7-26, 1961.
- [4] Alexander, S. S., *Price Movements in Speculative Markets: Trends or Random Walks*, in Cootner, P. (eds.) *The Random Character of Stock Market Prices*, Vol. 2, MIT Press, Cambridge, pp. 338-372, 1964.
- [5] Allen, H. and M. P. Taylor, *Charts, Noise and Fundamentals in the London Foreign Exchange Market*, *The Economic Journal* 100, pp. 49-59, 1990.
- [6] Allen, F. and R. Karjalainen, *Using Genetic Algorithms to Find Technical Trading Rules*, *Journal of Financial Economics* 51 (2), pp. 245-271, 1999.
- [7] Amihud, Y. and H. Mendelson, *Volatility, Efficiency, and Trading: Evidence from the Japanese Stock Market*, *Journal of Finance* 46 (5), pp. 1765-1789, 1991.
- [8] Anderson, A., *An Introduction to Neural Networks*, MIT Press, Boston, 1995.
- [9] Armano, G., A. Murru and F. Roli, *Stock Market Prediction by a Mixture of Genetic-Neural Experts*, *International Journal of Pattern Recognition and Artificial Intelligence* 16 (5), pp. 501-526, 2002.
- [10] Arthur, W. B., J. H. Holland, B. LeBaron, R. Palmer and P. Tayler, *Asset Pricing Under Endogenous Expectations in an Artificial Stock Market*, *The Economy as an Evolving Complex System II*, Eds. Arthur, Durlauf, and Lane SFI Studies in the Science of Complexity, vol. XXVII, Addison-Wesley, pp. 12-44, 1997.
- [11] Asquith, P., *Merger Bids, Uncertainty and Stockholder Returns*, *Journal of Financial Economics* 11, pp. 51-83, 1983.
- [12] Atkins, A. B. and E. A. Dyl, *Price Reversals, Bid-Ask Spreads and Market Efficiency*, *Journal of Financial and Quantitative Analysis* 25 (4), pp. 535-547, 1990.
- [13] Bachelier, L. , *Theory of Speculation*, A thesis presented to the Faculty of Sciences of the Academy of Paris, *Annales de l'Ecole Normal Superieure* 27, pp. 21-86, 1900.

- [14] Bagehot, W., *The Only Game in Town*, Financial Analysts Journal 27, pp. 12-17, 1971.
- [15] Ball, R. and S. P. Kothari, Nonstationary Expected Returns: Implications for Tests of Market Efficiency and Market Correlation in Returns, Journal of Financial Economics 25, pp. 51-74, 1989.
- [16] Barber, B. M. and T. Odean, *Trading Is Hazardous to Your Wealth: The Common Stock Investment Performance of Individual Investors*, Journal of Finance 55 (2), pp. 773-806, 2000.
- [17] Barber, B. M. and T. Odean, *Online Investors: Do the Slow Die First?*, The Review of Financial Studies, Special Issue: Conference on Market Frictions and Behavioral Finance 15 (2), pp. 455-487, 2002.
- [18] Barberis, N., A. Shleifer and R. Vishny, *A Model of Investment Sentiment*, Journal of Financial Economics 49, pp. 307-343, 1998.
- [19] Barucci, E. and R. Reno, *On Measuring Volatility of Diffusion Processes with High Frequency Data*, Economic Letters 74, pp. 371-378, 2002.
- [20] Batten, J. and C. Ellis, *Technical Trading System Performance in the Australian Share Market: Some Empirical Evidence*, Asia Pacific Journal of Management 13 (1), pp. 87-99, 1996.
- [21] Bauwens, L. and P. Giot, *Econometric Modelling of Stock Market Intraday Activity*, Boston: Kluwer Academic Publishers, 2001.
- [22] Bauwens, L., *Econometric Analysis of Intra-daily Trading Activity on Tokyo Stock Exchange*, IMES Discussion Paper, No. 2005-E-3, Institute for Monetary and Economic Studies, Bank of Japan, 2005.
- [23] Beaver, W. H., *Information Efficiency*, Accounting Review 56, pp. 23-27, 1980.
- [24] Beechey, M., D. Gruen and J. Vickery, *The Efficient Market Hypothesis: A Survey*, Research Discussion Paper Series, No. 2000-01, Economic Research Department, Reserve Bank of Australia, 2000.
- [25] Beja, A. and M. B. Goldman, *On the Dynamic Behavior of Prices in Disequilibrium*, Journal of Finance 35 (2), pp. 235-248, 1980.
- [26] Bekaert, G. and R. J. Hodrick, *Characterizing Predictable Components in Equity and Foreign Exchange Rates of Return*, Journal of Finance 47, pp. 467-509, 1992.
- [27] Bessembinder, H. and K. Chan, *The Profitability of Technical Trading Rules in the Asian Stock Markets*, Pacific-Basin Finance Journal 3, pp. 257-284, 1995.
- [28] Bessembinder, H. and K. Chan, *Market Efficiency and the Returns to Technical Analysis*, Financial Management 27 (2), pp. 5-17, 1998.
- [29] Beyer, H. G. and H. P. Schwefel, *Evolution Strategies: A Comprehensive Introduction*, Journal Natural Computing 1 (1), pp. 3-52, 2002.

- [30] Black, F. and M. Scholes, *The Pricing of Options and Corporate Liabilities*, Journal of Political Economy 81, pp. 637-659, 1973.
- [31] Blume, L., D. Easley and M. O'Hara, *Market Statistics and Technical Analysis: The Role of Volume*, Journal of Finance 49, pp. 153-181, 1994.
- [32] Bodurtha, J. N. and G. R. Courtadon, *Efficiency Tests of the Foreign Currency Options Market*, The Journal of Finance 41 (1), pp. 151-162, 1986.
- [33] Bollerslev, T. and R. J. Hodrick, *Financial Market Efficiency Tests*, NBER Working Paper Series, No. 4108, Cambridge: National Bureau of Economic Research, 1992.
- [34] Bose, N. K., P. Liang, *Neural Network Fundamentals with Graphics, Algorithms and Applications*, New York McGraw-Hill Inc., 1996.
- [35] Bray, M., *Learning, Estimation, and Stability of Rational Expectations*, Journal of Economic Theory 26, pp. 318-339, 1995.
- [36] Brock, W., J. Lakonishok and B. LeBaron, *Simple Technical Trading Rules and the Stochastic Properties of Stock Returns*, Journal of Finance 47 (5), pp. 1731-1764, 1992.
- [37] Brock, W., D. Dechert, J. Scheinkman and B. LeBaron, *A Test for Independence Based on the Correlation Dimension*, Econometric Review 15 (3), pp. 197-235, 1996.
- [38] Brown, D. P. and R. H. Jennings, *On Technical Analysis*, Review of Financial Studies 2 (4), pp. 527-551, 1989.
- [39] Butz, M. V., *Rule-Based Evolutionary Online Learning Systems: Learning Bounds, Classification, and Prediction*, IlliGAL Report No. 2004034, 2004.
- [40] Butz, M. V., *Documentation of XCS+TS C-Code 1.2*, IlliGAL Report No. 2003023, 2003.
- [41] Campbell, J. Y., *A Variance Decomposition of Stock Returns*, Economic Journal 101, pp. 152-179, 1991.
- [42] Campbell, J. Y. and R. Shiller, *The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors*, Review of Financial Studies 1, pp. 195-228, 1988.
- [43] Chan, K. C., *On the Contrarian Investment Strategy*, Journal of Business 61, pp. 147-163, 1988.
- [44] Chung, Y. P., *A Transaction Date Test of Stock Index Futures Market Efficiency and Index Arbitrage Profitability*, Journal of Finance 46 (5), pp. 1791-1809, 1991.
- [45] Clyde, W. C. and C. L. Osler, *Charting: Chaos Theory in Disguise?*, Journal of Futures Markets 17 (5), pp. 489-514, 1997.
- [46] Cohen, P. R., *Empirical Methods for Artificial Intelligence*, Cambridge: MIT Press, 1995.
- [47] Conrad, J. and G. Kaul, *Time-Variation in Expected Returns*, Journal of Business 61, pp. 409-425, 1988.

- [48] Cootner, P. A., *Stock Prices: Random vs. Systematic Changes*, Industrial Management Review 3, pp. 24-45, 1962.
- [49] Cootner, P. A. *Refinement and Empirical Testing: Introduction*, in Cootner (1967), *The Random Character of Stock Market Prices*, Cambridge: MIT Press, 1967.
- [50] Cowles, A., *Can Stock Market Forecasters Forecasts?*, *Econometrica* 1, pp. 309-324, 1933.
- [51] Cowles, A. and H. Jones, *Some Posteriori Probabilities in Stock Market Action*, *Econometrica* 5 (3), pp. 280-294, 1937.
- [52] Cramer, N. L., *A Representation for the Adaptive Generation of Simple Sequential Programs*, in *Proceedings of an International Conference on Genetic Algorithms and the Applications*, Grefenstette, John J. (ed.), Carnegie Mellon University, 1985.
- [53] Cumby, R. E. and J. Huizinga, *Testing the Autocorrelation Structure of Disturbance in Ordinary Least Squares and Instrumental Variables Regressions*, *Econometrica* 60, pp. 185-195, 1992.
- [54] Dacorogna, M., V. Muller, R. B. Olsen and O. V. Pictet, *Defining Efficiency in Heterogeneous Market*, *Quantitative Finance* 1 (2), pp. 198-201, 2001.
- [55] Dacorogna, M., R. Gencay, U. A. Mueller, R. B. Olsen and O. V. Pictet, *An Introduction to High-Frequency Finance*, Academic Press, 2001.
- [56] Daniel, K., D. Hirshleifer and A. Subrahmanyam, *A Theory of Overconfidence, Self-Attribution and Security Market Under- and Over-reaction*, *Journal of Finance* 53, pp. 1839-1885, 1998.
- [57] Dawson, S. M., *Singapore Share Recommendations Using Technical Analysis*, *Asia Pacific Journal of Management* 2 (3), pp. 180-188, 1985.
- [58] Dawson, E. R. and J. M. Steeley, *On the Existence of Visual Patterns in the UK Stock Market*, *Journal of Business Finance and Accounting* 30, pp. 263-293, 2003.
- [59] Day, T. E. and P. Wang, *Dividends, Nonsynchronous Prices and Returns from Trading the Dow Jones Industrial Average*, *Journal of Empirical Finance* 9, pp. 431-454, 2002.
- [60] DeBondt, W. F. M. and R. H. Thaler, *Does the Stock Market Overreact?*, *Journal of Finance* 40, pp. 793-805, 1985.
- [61] DeBondt, W. F. M. and R. H., Thaler, *Further Evidence on Investors Overreaction and Stock Market Seasonality*, *Journal of Finance* 42, pp. 557-581, 1987.
- [62] DeLong, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann, *Positive Feedback Investment Strategies and Destabilizing Rational Speculation*, *Journal of Finance* 45, pp. 379-395, 1990.
- [63] DeLong, J. B., A. Shleifer, L. H. Summers and R. J. Waldmann, *Noise Trader Risk in Financial Markets*, *Journal of Political Economy* 98 (4), pp. 703-738, 1990.

- [64] De Jong, K. A. and W. M. Spears, *Learning Concept Classification Rules Using Genetic Algorithms*, IJCAI-91 Proceedings of the Twelfth International Conference on Artificial Intelligence, pp. 651-656, 1991.
- [65] Demsetz, H., *The Cost of Trading*, Quarterly Journal of Economics 82, pp. 33-53, 1968.
- [66] Diebold, F. X., *Testing for Serial Correlation in the Presence of ARCH*, Proceedings of the American Statistical Association, Business and Economic Statistics Association, pp. 323-328, 1986.
- [67] Dimson, E. and M. Mussavian, A Brief History of Market Efficiency, European Financial Management 4 (1), pp. 91-193, 1998.
- [68] Edwards, R. D., Magee, J. and W. H. C. Bassetti, *Technical Analysis of Stock Trends*, 8th ed., St. Lucie Press, 2001.
- [69] Evans, J. L., *The Random Walk Hypothesis, Portfolio Analysis and the Buy and Hold Criterion*, Journal of Financial and Quantitative Analysis 3 (3), pp.327-342, 1968.
- [70] Fama, E. F., *The Behavior of Stock Market Prices*, Journal of Business 38, pp. 34-105, 1965.
- [71] Fama, E. F., *Efficient Capital Market: A Review of Theory and Empirical Work*, Journal of Finance 25, pp. 383-417, 1970
- [72] Fama, E. F., *Foundations of Finance*, New York: Basic Books, 1976.
- [73] Fama, E. F., *Efficient Capital Markets: II*, The Journal of Finance 46 (5), pp. 1575-1617, 1991.
- [74] Fama, E. F., *Market Efficiency, Long-Term Returns and Behavioral Finance*, Journal of Financial Economics 49, pp. 283-306, 1998.
- [75] Fama, E. F. and M. E. Blume, *Filter Rules and Stock Market Trading. Security Prices: A Supplement*, Journal of Business 39, pp. 191-225, 1966.
- [76] Fama, E. F. and K. R. French, *Permanent and Temporary Components of Stock Prices*, Journal of Political Economy 96, pp. 246-273, 1988.
- [77] Fama, E. F. and K. R. French, *Dividend Yields and Expected Stock Returns*, Journal of Financial Economics 22, pp. 427-465, 1988.
- [78] Fama, E. F. and W. Schwert, *Asset Returns and Inflation*, Journal of Financial Economics 5, pp. 115-146, 1977.
- [79] Fang, Y. and D. Xu, *The Predictability of Asset Returns: An Approach Combining Technical Analysis and Time Series Forecasts*, International Journal of Forecasting 19, pp. 369-385, 2003.
- [80] Farmer, J. D. and J. J. Sidorowich, *Predicting Chaotic Time Series*, Physical Review Letters 59 (8), pp. 845-848, 1987.

- [81] Farmer, J. D. and A. W. Lo, *Frontiers of Finance: Evolution and Efficient Markets*, Proceeding of the National Academy of Science 96, pp. 9991-9992, 1999.
- [82] Fernandez-Rodriguez, F., S. Sosvilla-Rivero and J. Andrada-Felix, *Combining Information in Exchange Rate Forecasting: Evidence from the EMS*, Applied Economic Letters 4 (7), pp. 441-444, 1997.
- [83] Fernandez-Rodriguez, F., S. Sosvilla-Rivero and M. D. Garcia-Artiles, *Using Nearest-Neighbour Predictors to Forecast the Spanish Stock Market*, Investigaciones Economicas 21 (1), pp. 75-91, 1997.
- [84] Fernandez-Rodriguez, F., C. Gonzalez-Martel and S. Sosvilla-Rivero, *On the Profitability of Technical Trading Rules Based on Artificial Neural Networks: Evidence from the Madrid Stock Market*, Economic Letters 69, pp. 89-94, 2000.
- [85] Fernandez-Rodriguez, F., S. Sosvilla-Rivero and J. Andrada-Felix, *Nearest-Neighbour Predictions in Foreign Exchange Markets*, FEDEA Working Paper No. 2002-05, 2002.
- [86] Fisher, L., *Some New Stock-Market Indexes*, Journal of Business 39, pp. 191-225, 1966.
- [87] Flavin, M., *Excess Volatility in Financial Markets: A Reassessment of the Empirical Evidence*, Journal of Political Economy 91, pp. 929-956, 1983.
- [88] Fock, J. H., C. Klein and B. Zwergel, *Performance of Candlestick Analysis on Intraday Futures Data*, Journal of Derivatives, pp. 28-40, Fall 2005.
- [89] Ge, X., P. Smyth, *Segmental Semi-Markov Models for Endpoint Detection in Plasma Etching*, in IEEE Transactions on Semiconductor Engineering, 2001.
- [90] Gervais, S. and T. Odean, *Learning to Be Overconfident*, Review of Financial Studies 14, pp. 1-27, 2001.
- [91] Giles, C. L., S. Lawrence and A. C. Tsoi, *Noisy Time Series Prediction Using a Recurrent Neural Network and Grammatical Inference*, Machine Learning 44 (1/2), pp. 161-183, 2001.
- [92] Gilles, C. and S. F. LeRoy, *Econometric Aspects of the Variance-Bounds Tests: A Survey*, Review of Financial Studies 4, pp. 753-791, 1991.
- [93] Glosten, L. and P. Milgrom, *Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders*, Journal of Financial Economics 14, pp. 71-100, 1985.
- [94] Goldberg, D. E., *Sizing Populations for Serial and Parallel Genetic Algorithms*, TCGA Report No. 88004, University of Alabama, The Clearinghouse for Genetic Algorithms, 1988.
- [95] Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Reading MA: Addison-Wesley, 1989.

- [96] Granger, W. J. and O. Morgenstern, *Spectral Analysis of New York Stock Market*, *Kyklos* 16, pp. 162-188, 1963.
- [97] Green, T. and S. Figlewski, *Market Risk and Model Risk for a Financial Institution Writing Options*, *Journal of Finance* 54, pp. 1465-1499, 1999.
- [98] Grossman, S. J. and R. J. Shiller, *The Determinants of the Variability of Stock Market Prices*, *American Economic Review* 71, pp. 222-227, 1981.
- [99] Grossman, S. J. and J. Stiglitz, *On the Impossibility of Informationally Efficient Markets*, *American Economic Review* 70, pp. 393-408, 1980.
- [100] Hagen, M. T., Demuth H. B. and M. Beale, *Neural Network Design*, Boston: PWS Publishing Co., 1996.
- [101] Hansen, L. P., *Large Sample Properties of Generalized Method of Moments Estimators*, *Econometrica* 50, pp. 111-120, 1982.
- [102] Haykin, S., *Self-Organizing Maps*, in *Neural Networks - A Comprehensive Foundation*, 2nd edition, Prentice-Hall, 1999.
- [103] Hill, B. M., *A Simple General Approach to Inference about the Tail of a Distribution*, *Annals of Statistics* 3, pp. 1163-1174, 1975.
- [104] Hodrick, R. J., *Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement*, *Review of Financial Studies* 5 (3), pp. 357-386, 1992.
- [105] Hogan, S., R. Jarrow, M. Teo and M. Varechka, *Testing Market Efficiency Using Statistical Arbitrage with Applications to Momentum and Value Strategies*, *Journal of Financial Economics* 73, pp. 525-565, 2004.
- [106] Holland, J. H., *Adaptation in Natural and Artificial Systems*, Ann Arbor, MI: University of Michigan Press, 1975.
- [107] Holland, J. H. and J. S. Reitman, *Cognitive Systems Based on Adaptive Algorithms*, in D. A. Waterman and F. Hayes-Roth (Eds.), *Pattern-directed Inference Systems*, New York: Academic Press, pp. 313-329, 1978.
- [108] Holmes, J. H., Lanzi, P. L., Stolzmann, W., and Wilson, S. W., *Learning Classifier Systems: New Models, Successful Applications*, *Information Processing Letters* 82 (1), pp. 23-30, 2002.
- [109] Houthakker, H., *Systematic and Random Elements in Short Term Price Movements*, *American Economic Review* 51, pp. 164-172, 1961.
- [110] Hundson, R., M. Dempsey and K. Keasey, *A Note on the Weak Form Efficiency of Capital Markets: The Application of Simple Technical Trading Rules to UK Stock Prices - 1935 to 1994*, *Journal of Banking and Finance* 20, pp. 1121-1132, 1996.
- [111] Hull, J. C., *Options, Futures, and other Derivatives*, 5th Edition, Prentice-Hall, 2003.
- [112] Hussman, J. P., *Time Variation in Market Efficiency: A Mixture-of-Distribution Approach*, Hussman Funds, mimeo, 1998.

- [113] James, Jr. F. E., *Monthly Moving Averages - An Effective Investment Tool?*, Journal of Financial and Quantitative Analysis 3 (3), pp. 315-326, 1968.
- [114] Jegadeesh, N., *Evidence of Predictable Behavior of Security Returns*, Journal of Finance 45, pp. 881-898, 1990.
- [115] Jegadeesh, N. and S. Titman, *Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency*, Journal of Finance 48, pp. 65-91, 1993.
- [116] Jegadeesh, N., *Discussion*, Journal of Finance 55, pp. 1765-1770, 2000.
- [117] Jensen, M., *Some Anomalous Evidence Regarding Market Efficiency*, Journal of Financial Economics 6, pp. 95-101, 1978.
- [118] Kahneman, D. and A. Tversky, *Intuitive Predictions: Biases and Corrective Procedures*, Reprinted in Kahneman, Slovic and Tversky, *Judgement under Uncertainty: Heuristics and Biases*, Cambridge (England): Cambridge University Press, 1982.
- [119] Kandel, S. and R. Stambaugh, *On the Predictability of Stock Returns: An Asset Allocation Perspective*, Journal of Finance 51, pp. 385-424, 1996.
- [120] Kanzler, L., *A Study of the Efficiency of the Foreign Exchange, Through Analysis of Ultra-high Frequency Data*, Ph.D. Dissertation, Oxford University, 1998.
- [121] Kanzler, L., *Very Fast and Correctly Sized Estimation of the BDS Statistic*, Department of Economics of Oxford University, Available at <http://users.ox.ac.uk/~econlrk>, 1999.
- [122] Kendall, M.G., *The Analysis of Economic Time Series - Part 1: Prices*, Journal of Royal Statistical Society (Series A) 96, pp. 11-25, 1953.
- [123] Keogh, E., S. Chu, D. Hart and M. Pazzani, *An online algorithm for segmenting time series*, in Proceedings of the IEEE International Conference on Data Mining, pp. 289-296, 2001.
- [124] Keogh, E., K. Chakrabarti, M. Pazzani and Mehrotra, *Dimensionality Reduction for Fast Similarity Search in Large Time Series Databases*, Journal of Knowledge and Information Systems, 2000.
- [125] Keogh, E. and M. Pazzani, *An Enhanced Representation of Time Series Which Allows Fast and Accurate Classification, Clustering and Relevance Feedback*, Proceedings of the 4th International Conference of Knowledge Discovery and Data Mining, AAAI Press, pp. 239-241, 1998.
- [126] Keogh, E. and M. Pazzani, *Relevance Feedback Retrieval of Time Series Data*, Proceedings of the 22th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, 1999.
- [127] Kidd, W. V. and B. W. Brorsen, *Why Have the Returns to Technical Analysis Decreased?*, Journal of Economics and Business 56, pp. 159-176, 2004.



- [128] Kim, M. J., C. R. Nelson and R. Startz, *Mean Reversion in Stock Prices? A Reappraisal of the Empirical Evidence*, Review of Economic Studies 58, pp. 515-528, 1991.
- [129] King, W. I., *Forecasting Methods Successfully Used Since 1928*, Journal of the American Statistical Association 27 (179), pp. 315-319, 1932.
- [130] King, W. I., *Technical Methods of Forecasting Stock Prices*, Journal of the American Statistical Association 29 (187), pp. 323-325, 1934.
- [131] Kleidon, A. W., *Variance Bounds Tests and Stock Price Valuation Models*, Journal of Political Economy 94, pp. 953-1001, 1986.
- [132] Kleidon, A. W., *Bias in Small Sample Tests of Stock Price Rationality*, Journal of Business 59 (2), pp. 237-262, 1986.
- [133] Kohonen, T., *Self-Organization and Associative Memory*, New York: Springer-Verlag, 1988.
- [134] Kohonen, T., *Self-Organizing Maps*, New York: Springer-Verlag, 1997.
- [135] Korczak, J. and P. Roger, *Stock Timing Using Genetic Algorithms*, Applied Stochastic Models in Business and Industry 18, pp. 121-134, 2002.
- [136] Kovacs, T. and M. Kerber, *High Classification Accuracy does not Imply Effective Genetic Search*, in K. Deb et al., GECCO-2004, Proceedings of the 2004 Genetic and Evolutionary Computation Conference, Lecture Notes in Computer Science (3102), Springer, pp. 785-796, 2004.
- [137] Koza, J. R. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*, MIT Press, 1992.
- [138] Krishnakumar, K., *Micro-Genetic Algorithms for Stationary and Non-Stationary Function Optimization*, SPIE Proceedings: Intelligent Control and Adaptive Systems, 1196, pp. 289-296, 1989.
- [139] Kyle, A., *Continuous Auctions and Insider Trading*, Econometrica 53, pp. 1315-1335, 1985.
- [140] Kyle, A. and F. A. Wang, *Speculation Duopoly with Agreement to Disagree: Can Over-Confidence Survive the Market Test?*, Journal of Finance 52, pp. 2073-2090, 1997.
- [141] Ilinskaia, A. and K. Ilinski, *How to Reconcile Market Efficiency and Technical Analysis*, <http://xxx.lanl.gov/abs/cond-mat/9902044>, 1999.
- [142] Lam, M., *Neural Network Techniques for Financial Performance Prediction: Integrating Fundamental and Technical Analysis*, Decision Support Systems 37, pp. 567-581, 2004.
- [143] Lanzi, P. L., W. Stolzmann, S. W. Wilson (Eds.), *Learning Classifier Systems: From Foundations to Applications*, Vol. 1813 of LNAI, Berlin: Springer-Verlag, 2000.

- [144] Leadbetter, M. R., G. Lindgren and H. Rootzen, *Extremes and Related Properties of Random Sequences and Processes*, New York: Springer-Verlag, 1983.
- [145] LeBaron, B., *Technical Trading Rule Profitability and Foreign Exchange Intervention*, *Journal of International Economics* 49 (1), pp. 125-143, 1999.
- [146] LeBaron, B., *Evolution and Time Horizons in an Agent Based Stock Market*, *Macroeconomic Dynamics* 5 (2), pp. 225-254, 2001.
- [147] LeBaron, B., *Technical Trading Rules and Regime Shifts in Foreign Exchange*, in *Advanced Trading Rules*, edited by E. Acar and S. Satchell, 2nd Edition, Butterworth-Heinemann, pp. 5-40, 2002.
- [148] Lehmann, B. N. and D. M. Modest, *Trading and Liquidity on the Tokyo Stock Exchange: A Bird's Eye View*, *Journal of Finance* 49 (3), pp. 951-984, 1994.
- [149] Leigh, W., R. Purvis and J. M. Ragusa, *Forecasting the NYSE Composite Index with Technical Analysis, Pattern Recognizer, Neural Networks, and Genetic Algorithm: A Case Study in Romantic Decision Support*, *Decision Support Systems* 32, pp. 361-377, 2002.
- [150] Lesmond, D. A., J. P. Ogden and C. A. Trzcinka, *A New Estimate of Transaction Costs*, *Review of Financial Studies* 12, pp. 1113-1141, 1999.
- [151] Lesmond, D. A., M. J. Schill and C. Zhou, *The Illusory Nature of Momentum Profits*, *Journal of Financial Economics* 71, pp. 349-380, 2004.
- [152] LeRoy, S. F. and R. D. Porter, *The Present-Value Relation: Test Based on Implied Variance Bounds*, *Econometrica* 49, pp. 555-574, 1981.
- [153] Levich, R. M. and L. R. Thomas, *The Significance of Technical Trading-Rule Profits in the Foreign Exchange Market: A Bootstrap Approach*, *Journal of International Money and Finance* 12 (5), pp. 451-474, 1993.
- [154] Levy, R. A., *Relative Strength as a Criterion for Investment Selection*, *Journal of Finance* 22 (4), pp. 595-610, 1967.
- [155] Lewis, K. K., *Puzzles in International Financial Markets*, in G. Grossman and K. Rogoff (eds.), *Handbook of International Economics*, Vol. III, Elsevier Science, pp. 1913-1971, 1995.
- [156] Lo, A. W. and A. C. MacKinlay, *Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test*, *Review of Financial Studies* 1 (1), pp. 41-66, 1988.
- [157] Lo, A. W. and A. C. MacKinlay, *Data-Snooping Biases in Tests of Financial Asset Pricing Models*, *Review of Financial Studies* 3, pp. 431-467, 1990.
- [158] Lo, A. W., H. Mamaysky and J. Wang, *Foundations of Technical Analysis: Computational Algorithms, Statistical Inference and Empirical Implementation*, *Journal of Finance* 55, pp. 1705-1765, 2000.

- [159] Lo, A. W., *Reconciling Efficient Market with Behavioral Finance: The Adaptive Markets Hypothesis*, MIT Sloan School of Management, mimeo, 2005.
- [160] Lukac, L. P., B. W. Brorsen and S. H. Irwin, *A Test of Futures Market Disequilibrium Using Twelve Different Trading Systems*, *Applied Economics* 20, pp. 623-639, 1988.
- [161] Lukac, L. P., B. W. Brorsen, *A Comprehensive Test of Futures market Disequilibrium*, *Financial Review* 25, pp. 593-622, 1990.
- [162] Lux, T. and M. Marchesi, *Volatility Clustering in Financial Markets: A Microsimulation of Interacting Agents*, *International Journal of Theoretical and Applied Finance* 4, pp. 1-20, 2000.
- [163] Maillet, B. and T. Michel, *Technical Analysis Profitability when Exchange Rates are Pegged: A Note*, *European Journal of Finance* 11 (6), pp. 463-470, 2005.
- [164] Malkiel, B., *Efficient Market Hypothesis*, in Newman, P., M. Milgate, and J. Eatwell (eds.), *New Palgrave Dictionary of Money and Finance*, London: Macmillan, 1992.
- [165] Mandelbrot, B., *Sur Certains Prix Speculatifs: Faits Empiriques et Modele Base sur les Processus Stables Additives non Gaussiens de Paul Levy*, *Compte-rendus a l'Academie des Sciences*, pp. 3968-3970, 1962.
- [166] Mandelbrot, B., *The Variation of Certain Speculative Prices*, *Journal of Business* 36, pp. 394-419, 1963.
- [167] Mankiw, N. G., D. Romer and M. D. Shapiro, *Stock Market Forecastability and Volatility: A Statistical Appraisal*, *Review of Economic Studies* 58, pp. 455-477, 1991.
- [168] Markowitz, H., *Portfolio Selection*, *Journal of Finance* 12, pp. 77-91, 1952.
- [169] Mason, D. M., *Laws of Large Numbers for Sums of Extreme Values*, *Annals of Probability* 10, pp. 754-764, 1982.
- [170] Massey, F. J. Jr., *The Kolmogorov-Smirnov Test of Goodness of Fit*, *Journal of the American Statistical Association* 46, p. 70, 1951.
- [171] Matilla-Garcia, M., *Are Trading Rules Based on Genetic Algorithms Profitable?*, *Applied Economic Letters* 13, pp. 123-126, 2006.
- [172] Metropolis, N., *The Beginning of the Monte Carlo Method*, *Los Alamos Science* 15, pp. 125-130, 1987.
- [173] Michalewicz, Z., *Genetic Algorithms + Data Structures = Evolution Programs*, 3rd Ed., Berlin: Springer-Verlag, 1999.
- [174] Mills, T. C., *Technical Analysis and the London Stock Exchange: Testing Trading rules Using the FT30*, *International Journal of Finance & Economics* 2, pp. 319-331, 1997.
- [175] Mitchell, M. and E. Stafford, *Managerial Decisions and Long-term Stock Price Performance*, *Journal of Business* 73, pp. 287-329, 2000.

- [176] Mitra, S. K., *Profiting from Technical Analysis in Indian Stock Market*, Finance India 16 (1), pp. 109-120, 2002.
- [177] Moore, A., *A Statistical Analysis of Common Stock Prices*, Ph.D. thesis, University of Chicago, Graduate School of Business, 1962.
- [178] Morrison, A. D. and N. Vulkan, *Making Money out of Publicly Available Information*, Economics Letters 89 (1), pp. 31-38, 2005.
- [179] Mueller, U. A., M. Dacorogna, R. B. Olsen, O. V. Pictet, M. Schwarz and C. Morgenegg, *Statistical Study of Foreign Exchange Rates, Empirical Evidence of a Price Changing Law, and Intraday Analysis*, Journal of Banking and Finance 14, pp. 1189-1208, 1990.
- [180] Murphy, J. J., *Technical Analysis of the Future Markets*, New York: Prentice Hall, 1986.
- [181] Murphy, J. J., *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*, New York: Prentice Hall, 1999.
- [182] Neely, C. J., P. Weller and R. Dittmar, *Is Technical Analysis in the Foreign Exchange Market Profitable? A Genetic Programming Approach*, Journal of Financial and Quantitative Analysis 32 (4), pp. 405-426, 1997.
- [183] Neely, C. J., *Technical Analysis in the Foreign Exchange Market: A Layman's Guide*, Federal Reserve Bank of St. Louis: Review, pp. 23-38, September 1997.
- [184] Neely, C. J., *Technical Analysis and Profitability of US Foreign Exchange Intervention*, Federal Reserve Bank of St. Louis: Review, pp. 3-17, July 1998.
- [185] Neftci, S. N. and A. J. Policano, *Can Chartists Outperform the Market? Market Efficiency Tests for Technical Analysis*, Journal of Futures Markets 4 (4), pp. 465-478, 1984.
- [186] Neftci, S. N., *Naive Trading Rules in Financial Markets and Weiner-Kolmogorov Prediction Theory: A Study of "Technical Analysis"*, Journal of Business 64 (4), pp. 549-571.
- [187] Nelson, C. R., *Inflation and Rates of Return on Common Stocks*, Journal of Finance 31, pp. 471-483, 1976.
- [188] Nison, S., *Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Techniques of the Far East*, NYIF, 2nd Ed., 2001.
- [189] Nocedal, J. and S. J. Wright, *Numerical Optimization*, Springer-Verlag, 1999.
- [190] Odean, T., *Volume, Volatility, Price and Profit When All Traders Are Above Average*, Journal of Finance 53, pp. 1887-1934, 1998.
- [191] Odean, T., *Do Investors Trade Too much?*, American Economic Review 89 (5), pp. 1279-1298, 1999.

- [192] Omrane, W. B. and H. Van Oppens, *The Performance Analysis of Chart Patterns: Monte Carlo Simulation and Evidence from the Euro/Dollar Foreign Exchange Market*, Empirical Economics 30, pp. 946-971, 2006.
- [193] Osborne, M. F. M., *Brownian Motion in the Stock Market*, Operations Research 7, pp. 145-173, 1959.
- [194] Osborne, M. F. M., *Some Quantitative Tests for Stock Price Generality Models and Trading Folklore*, Journal of the American Statistical Association 62 (318), pp. 321-340, 1967.
- [195] Osler, C. L and P. H. K. Chang, *Head and Shoulders: Not Just a Flaky Patter*, Federal Reserve Bank of New York Staff Report, No. 4, Federal Reserve Bank of New York, 1995.
- [196] Park, S., D. Lee and W. Chu, *Fast Retrieval of Similar Subsequences in Long Sequence Databases*, Proceedings of the 3rd IEEE Knowledge and Data Engineering Exchange Workshop, 1999.
- [197] Park, C.-H. and S. H. Irwin, *The Profitability of Technical Analysis: A Review*, AgMAS Project Research Report No. 2004-04, 2004.
- [198] Pearson, K., *The Problem of the Random Walk*, Nature 72 (1865), p. 294, 1905.
- [199] Pearson, K., *The Problem of the Random Walk*, Nature 72 (1867), p. 342, 1905.
- [200] Pesaran, M. H., *Market Efficiency Today*, IEPR Working Paper 05.41, 2005.
- [201] Poterba, J. and L. H. Summers, *Mean Reversion in Stock Prices: Evidence and Implications*, Journal of Financial Economics 22, pp. 27-59, 1988.
- [202] Potvin, J.-Y., P. Soriano and M. Vallee, *Generating Trading Rules on the Stock Markets with Genetic Programming*, Computers & Operations Research 31, pp. 1033-1047, 2004.
- [203] Press, W. H., B. P. Flannery, S. A. Teukolsky and W. T. Vetterling, *Numerical Recipes in C: The Art of Scientific Computing*, 2nd Ed., Cambridge University Press, 1992.
- [204] Pring, M. J., *Technical Analysis Explained: the Successful Investor's Guide to Spotting Investment Trends and Turning Points*, 2nd Ed., New York: McGraw Hill, 1985.
- [205] Richardson, M. and J. H. Stock, *Drawing Inferences from Statistics Based on Multi-Year Asset Returns*, Journal of Financial Economics 25, pp. 323-348, 1989.
- [206] Rhea, R., *The Dow Theory*, Fraser Publishing, 1993.
- [207] Roberts, H., *Stock-Market Patterns and Financial Analysis: Methodological Suggestions*, Journal of Finance 14, pp. 1-10, 1959.
- [208] Roberts, H., *Statistical versus Clinical Prediction of the Stock Market*, unpublished manuscript, 1967.

- [209] Roberts, M. C., *Technical Analysis and Genetic Programming: Constructing and Testing a Commodity Portfolio*, Journal of Futures Markets 25 (7), pp. 643-660, 2005.
- [210] Roll, R., *What Every CEO Should Know About Scientific Progress in Economics: What is Known and What Remains to be Resolved*, Financial Management 23, pp. 69-75, 1984.
- [211] Roll, R., *A Simple Implicit Measure of the Effective Bid-Ask Spread in an Efficient Market*, Journal of Finance 39 (4), pp. 1127-1139, 1984.
- [212] Roll, R., *The Hubris Hypothesis of Corporate Takeovers*, Journal of Business 59, pp. 197-216, 1986.
- [213] Roos, C. F., *Survey of Economic Forecasting Techniques: A Survey Article*, Econometrica 23 (4), pp. 363-395, 1955.
- [214] Rosenblatt, F., *The Perception: A Probabilistic Model for Information Storage and Organization in the Brain*, Psychological Review 65 (6), pp. 386-408, 1958.
- [215] Rubinstein, M., *Security Market Efficiency in an Arrow-Debreu Economy*, American Economic Review 65, pp. 812-824, 1975.
- [216] Sugiura, N. and R. T. Ogden, *Testing Changepoints with Linear Trend*, Communications in Statistics B: Simulation and Computation 23, pp. 287-322, 1994.
- [217] Samuel, A. L., *Some Studies in Machine Learning Using the Game of Checkers*, IBM Journal 3, pp. 211-229, 1959.
- [218] Samuelson, P. A., *Proof that Properly Anticipated Prices Fluctuate Randomly*, Industrial Management Review 6, pp. 41-49, 1965.
- [219] Schulenburg, S. and P. Ross, *Explorations in LCS Models of Stock Trading*, in Advances in Learning Classifier Systems, ed. Lanzi, P. L., W. Stolzmann and S. W. Wilson, volume 2321 of Lecture Notes in Artificial Intelligence, Berlin: Springer-Verlag, pp. 150-179, 2002.
- [220] Schultz, H., *Forecasting Security Prices*, Journal of the American Statistical Association 20 (150), pp. 244-249, 1925.
- [221] Sehgal, S. and A. Garhyan, *Abnormal Returns Using Technical Analysis: The Indian Experience*, Finance India 16 (1), pp. 181-203, 2002.
- [222] Sharpe, W. F., *A Simplified Model for Portfolio Analysis*, Management Science 9, pp. 277-293, 1963.
- [223] Shiller, R. J., *The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure*, Journal of Political Economy 87, pp. 1190-1219, 1979.
- [224] Shiller, R. J., *Do Stock Prices Move too much to be Justified by Subsequent Changes in Dividends?*, American Economic Review 71, pp. 421-436, 1981.
- [225] Shiller, R. J., *Stock Prices and Social Dynamics*, Brookings Papers on Economic Activity 2, pp. 457-510, 1984.

- [226] Schleifer, A., *Inefficient Markets : An Introduction to Behavioral Finance*, Oxford University Press, 2000.
- [227] Skouras, S., *Financial Returns and Efficiency as Seen by an Artificial Technical Analyst*, Journal of Economic Dynamics and Control 25 (1-2), pp. 213-44, 2001.
- [228] Slezak, S. L., *On the Impossibility of Weak-Form Efficient Markets*, Journal of Financial and Quantitative Analysis 38 (3), pp. 523-554, 2003.
- [229] Smith, S. F., *A Learning System Based on Genetic Adaptive Algorithms*, Ph.D. Thesis, Univ. of Pittsburgh, December, 1980.
- [230] Smith, A., *The Money Game*, New York, New York: Random House, p. 154, 1986.
- [231] Strozzi, F. and J. M. Zaldivar, *Non-linear Forecasting in High-Frequency Financial Time Series*, Physica A 353, pp. 463-479, 2005.
- [232] Sullivan, R., Timmerman A. and H. White, *Data-Snooping, Technical Trading Rule Performance and the Bootstrap*, CEPR Discussion Paper, No. 1976, Center for Economic Policy Research, Washington D.C., 1998.
- [233] Summers, L. H., *Does the Stock Market Rationally Reflect Fundamental Values?*, Journal of Finance 41, pp. 591-601, 1986.
- [234] Sweeney, R. J., *Beating the Foreign Exchange Market*, Journal of Finance 41, pp. 163-182, 1986.
- [235] Sweeney, R. J., *Some New Filter Rule Tests: Methods and Results*, Journal of Financial and Quantitative Analysis 23, pp. 285-300, 1988.
- [236] Taussig, F. W., *Is Market Price Determinate?*, Quarterly Journal of Economics 35 (3), pp. 394-411, 1921.
- [237] Taylor, S. J., *Strategies for Investors in Apparently Inefficient Futures Markets*, in Futures Markets - Modelling, Managing and Monitoring Futures Trading, ed. M. E. Streit, Oxford: Basil Blackwell, pp. 165-198, 1983.
- [238] Taylor, S. J., *Modelling Financial Time Series*, Chichester: Wiley, 1986.
- [239] Taylor, S. J., *Rewards Available to Currency futures Speculators: Compensation for Risk or Evidence of Inefficient Pricing?*, Economic Record 68 (supplement), pp. 105-116, 1992.
- [240] Taylor, M. P., *The Use of Technical Analysis in the Foreign Exchange Market*, Journal of International Money and Finance 11, pp. 304-314, 1992.
- [241] Taylor, S. J., *Stock Index and Price Dynamics in the UK and the US: New Evidence from a Trading Rule and Statistical Analysis*, European Journal of Finance 3, pp. 39-69, 2000.
- [242] Taylor, S. J., *Asset Price Dynamics, Volatility, and Prediction*, Princeton University Press, 2005.

- [243] Thorley, S., *The Inefficient Markets Argument for Passive Investing*, The Marriott School at BYU, mimeo, 1998.
- [244] Tokyo Stock Exchange, *Tokyo Stock Exchange Fact Book 2006*, 2006.
- [245] Tucker, A. L., *Empirical Tests of the Efficiency of the Currency Option Market*, The Journal of Financial Research 8 (4), pp. 275-285, 1985.
- [246] Tung, Y. A. and J. R. Marsden, *Test of Market Efficiency Using Experimental Electronic Markets*, Journal of Business Research 41, pp. 135-151, 1998.
- [247] Walter, C., *The Efficient Market Hypothesis, the Gaussian Assumption, and the Investment Management Industry*, Institut d'Etudes Politiques, mimeo, 2003.
- [248] Wasserfallen, W., and H. Zimmermann, *The Behavior of Intradaily Exchange Rates*, Journal of Banking and Finance 9, pp. 55-72, 1985.
- [249] Weintraub, R. E., *On Speculative Prices and Random Walks: A Denial*, Journal of Finance 18 (1), pp. 59-66, 1963.
- [250] Wilson, S. W. and D. E. Goldberg, *A Critical Review of Classifier Systems*, in Proceedings of the Third International Conference on Genetic Algorithms. Fairfax, VA: Morgan Kaufmann, pp. 244-255, 1989.
- [251] Wilson, S. W., *Classifier Fitness Based on Accuracy*, Evolutionary Computation 3, 1995.
- [252] Wilson, S. W., *Generalization in the XCS Classifier System*, 1998.
- [253] Wittwer, J. W., *Monte Carlo Simulation in Excel: A Practical Guide*, Vertex42.com, June 1, <http://vertex42.com/ExcelArticles/mc>, 2004.
- [254] Wong, W.-K., M. Manzur and B.-K. Chew, *How Rewarding is Technical Analysis? Evidence from Singapore Stock Market*, Applied Financial Economics 13 (7), Taylor and Francis Journals, pp. 543-551, 2003.
- [255] Working, H., *A Random Difference Series for Use in the Analysis of Time Series*, Journal of American Statistical Association 29 (185), pp. 11-24, 1934.
- [256] Working, H., *Note on the Correlation of First Differences of Averages in a Random Chain*, Econometrica 28, pp. 916-918, 1960.
- [257] <http://www.genetic-programming.com/humancompetitivedefinition.html>, 2006.
- [258] Van Eyden, R. J., *The Application of Neural Networks in the Forecasting of Share Prices*, Haymarket, VA: Finance and Technology Publishing, 1995.
- [259] Xu, K., G. Fisher and D. Willson, *Testing First- and Second-Order Stochastic Dominance*, Canadian Journal of Economics 29 (1), Canadian Economics Association, pp. 562-64, 1996.
- [260] Zadeh, L. A. *Fuzzy sets*, Information and Control 8, pp. 338-353, 1965.



## Vita

Timur Yusupov was born in Potsdam, Germany on July 23, 1976. He graduated from High School Gymnasium in Izhevsk (Russia) in June, 1993. During high-school studies he graduated from a distance learning project of Moscow State University: Distance Learning High School of Mathematics. Timur then commenced his undergraduate studies in Engineering at the Izhevsk State Technical University (IzhSTU) in fall 1993. After the graduation in 1998 he continued his education in the Magister program at the IzhSTU with the specialization in Information Technologies, which he graduated with honors in 2000. After studying one year in the Ph.D. program of the IzhSTU he changed specialization and got admitted to the Ph.D. program in Economics at the Center for Economic Research and Graduate Education in Prague, the Czech Republic. Since July 2003, he has been a doctoral student and research assistant at the Department of Economics at the University of Kiel. Following the completion of his doctoral degree, Timur will engage in postdoctoral research focusing on artificial intelligence and crisis prevention and control in financial markets.

I hereby affirm that I have completed my doctoral thesis entitled "The Efficient Market Hypothesis Through the Eyes of an Artificial Technical Analyst" entirely on my own and unassisted, and that I have specially marked all of the quotes I have used from other authors as well as those passages in my work that are extremely close to the thoughts presented by other authors, and listed the sources in accordance with the regulations I have been given.

Date

Signature