THESES OF DOCTORAL (PHD) DISSERTATION

Cseszník Zoltán

Gödöllő

2024

Hungarian University of Agriculture and Life Sciences

Analysis of the functionality of algorithmic trading systems

DOI: 10.54598/002810

Csesznik Zoltán

Gödöllő

2024

Doctoral School

Name:	Doctoral School of Economic and Regional Sciences
Discipline:	Management and Business Administration
Head of School:	Prof. Dr. Bujdosó Zoltán PhD.
Supervisors:	Prof. Dr. Zéman Zoltán, PhD
	Dr. habil. Bárczi Judit PhD

Approval by the Head of School

Approval of Supervisor

Approval of Supervisor

Table of contents

1. BACKGROUND OF THE WORK, OBJECTIVES

I came into contact with financial markets in 1999, after graduating with my first degree. Looking at the evolution of almost a quarter of a century since then, the world has undergone technological advances almost incomprehensible on a human scale. At the same time, the essence of stock trading, speculation and the desire for profit have not changed over the millennia and still dominate financial markets today. During my years in the field, every single day has been a challenge, and for many years I have consistently tried to estimate the unknown or calculate the unpredictable.

Stock trading has experienced and continues to experience a renaissance in the past five years, thanks to technological innovation and legal harmonization (Aslam et al., 2023). One of the main motivations of millions of new entrants is to make money, preferably as much and as quickly as possible. The world has not changed in this regard. People have always had exuberant expectations of the stock market, hoping for financial independence and enrichment from it (Nassar, 2006). Most of us are naturally easy to manipulate and surrender to our desire to trade our usual office life for independence. According to Nassar (2006) the reality is much more nuanced and the internet is a great breeding ground for urban legends.

The motivation for trading is very diverse, it can be for investment purposes, it can be speculative or it can be aimless (Fry-McKibbin & McKinnon, 2023).

My dissertation deals with speculative trading, specifically the creation, analysis, execution and reliability of strategies driven by computer algorithms.

In my opinion, successful trading requires finding price patterns in the past that repeat periodically and have similar outcomes (Chen & Tsai, 2020).

1.1 Hypotheses and objectives of the research

During my professional work, I experienced the operation of financial markets, the process of creating financial instruments, almost all the nuances of trading stocks and futures. I got to know the public and less public operating mechanisms of financial funds, their structure and the background of their legal environment.

C1: Due to the multifaceted nature of the topic, I will systematize the literature. I collect and present conflicting scientific trends and theories. During the literary processing, I will place great emphasis on ensuring that it is of such a scale and direction as to help us understand how financial markets work.

C2: In my research, I highlight the methods that can be used, summarize their most important information and their effectiveness. I determine the effectiveness of the indicators to be applied on the given tool, the optimal trading time slots.

C3: My goal is to create a diversified trading model that can be traded successfully in the long run. In it, I will use program codes generated by genetic algorithms developed based on my research.

C4: My goal is to test and validate the algorithms forming the model with statistical methods. Furthermore, I examine the long-term profitable operation of the portfolio made up of them.

When formulating the objectives and hypotheses of my research, I relied on what was written in the literature, but at the same time I used my knowledge and experience gained in the financial world over the past two decades.

H1A: Using the traditional investment model – buy and hold – makes it possible to trade futures markets more successfully than using a structured trading algorithm.

H1B: The algorithms created by the machine model can be traded more profitably than a simple, commonly known strategy (moving average crossing).

H2A: A set of indicators and parameters can be defined that allow a computer-controlled algorithmic model to trade successfully in the long run.

H2B: It can be demonstrated that the aggregate performance of systems based on indicators that perform better in a given market will be higher than the combined performance of systems generated using all indicators.

H3: It can be demonstrated that the aggregate out-of-sample trading outcome of diversified systems based on different indicators is more balanced than the result of individual systems.

2. MATERIALS AND METHODOLOGY

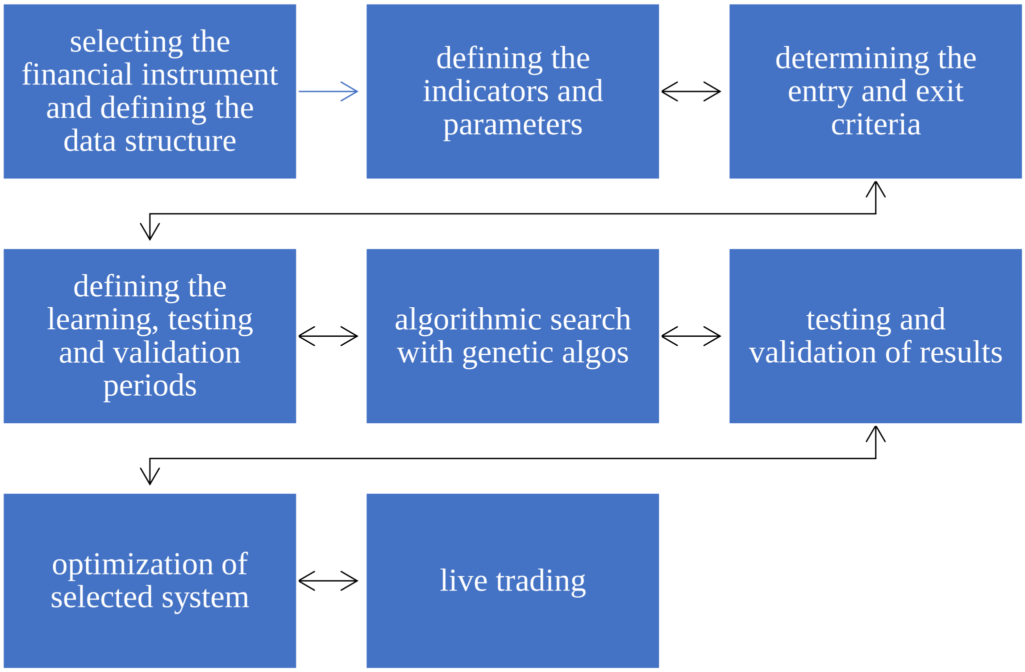
Below I detail my methodology used in the research, which I divide into three chapters. The first chapter details the process of machine model building, while the second chapter covers the development of the traditional investment model – buy and hold. In the third part, I present a simple and generally known strategy.

2.1. Definition of the machine system building model used in the research

Today, artificial intelligence (AI) excites both the scientific world and ordinary people. We hear a lot about AI infiltrating all areas of life. The truth is that AI has been with us since the 1950s, and its uses and developments are constantly changing (Huang et al., 2020). Machine learning and deep learning are subfields of artificial intelligence. AI systems make predictions or classifications based on input data (Kühl et al., 2022).

The role of artificial intelligence in stock exchange trading is in the execution of predefined systematic algorithms, and it can be used efficiently to perform complex data processing tasks, which coincides with the objectives of complex machine system building, during which models have to be selected from a large number of combinations that are able to generate profits similar to those of the past on future, unknown exchange rate data (Kühl et al., 2022). To achieve this goal, narrow artificial intelligence systems (ANI), developed to perform a specific task, are suitable. These systems are robust and can work very efficiently with large amounts of data, so we can choose trading algorithms that will generate profit in the future (Kühl et al., 2022). Of course, AI systems can also be used effectively in testing, validating and selecting trading algorithms. They can also be used in portfolio compilation, as they process data much faster and more efficiently.

With the computing power we currently have, it would be impossible to simulate every single combination and sequence the results. With such a large number of combinations, we use powerful search algorithms, with which we do not look at every single case, so we cannot be sure that we have found the global optimum, but the results provided by these genetic algorithms are considered good, and by choosing a sufficiently large generation and population, we can say with certainty that we will get results approaching the global optimum (Nicholls & Engelbrecht, 2019).



1. Figure: Key aspects of system building

Source: Genetic System Builder, self-edited

Figure 1 shows the most important steps of system building. Of course, if the results obtained for the given step are not satisfactory, the step must be repeated after making the appropriate adjustments.

Although machine model building has been dealt with in many different ways and its processes can be found in many literature, the uniqueness of system building is that the subprocesses are interchangeable, and changing their parameters and methods of systematization provides researchers with new and novel results. It is considered a novel result if we manage to develop profitable systems based on a given methodology. The process and steps were developed on the basis of the literature, but the order and variables of them I determined myself.

A. Defining markets: Trading systems can be built on any liquid financial market (asset) with some efficiency, in my dissertation I focus on commodity futures markets. One reason for this is personal preference. I think that because these markets are smaller, they are less interesting for multi-billion dollar funds and investment banks, so speculative demand is less. At the same time, hedge traders present themselves in these markets with well-defined and visible seasonal demands, so that more frequent price patterns can be better defined. When selecting the instruments, I take into account that they should be independent of each other, that economic and political events of a given product have a minimal influence on the other product, and that geopolitical events should also be given less weight in trading. Stable markets that have existed for many decades, allowing proper analysis with the help of historical data. Another consideration was to keep the daily trading time short, which allows for faster analysis. Based on the above, I select the following futures contracts:

- Lean hog (LH)
- Wheat (W)
- Coffee (KC)
- Frozen orange juice (OJ)

B. Portfolio definition: The portfolio will consist of a maximum of the top eight systems per market, considering point A), we will have a maximum of 32 systems in total. This will ensure diversification and we expect systems to have low correlation.

C. Database definition: Means all the data that we use in research. The Global Database from 01/01/2010 to 01/09/2023 contains data on the markets defined in point A). We divide our database into the following periods:

- Training period: 01.01.2010 – 31.12.2021 with eighty days breaks
- Test period: 01.01.2010 – 31.12.2021 with eighty days gaps.
- Validation period: 01.01.2022 – 01.09.2023

Training (within sample) and test (outside sample) periods alternate every eighty days. The significance of this is that we use data from all parts of the database, including the latest exchange rate, to train the system. Thus, both during the training and test periods, the exchange rate characteristics are diverse. The purpose of the validation period is to simulate live trading on this dataset once the portfolio has been assembled.

Data structure definition: The basic data is organized over time according to the 30-minute open-high-low-close principle, while the verification data structure: 29- and 31-minute opening-high-low-closing, and 30-minute opening-high-low-closing data by adding 1 tick random noise and 30-minute opening-high-low-closing data by adding 2 tick random noise.

D. Indicators and their parameters: For system building we use GSB software, which contains 85 indicators, which are listed in Table 1 organized into categories.

To generate systems from the set of indicators, 4 indicators are used at once. The parameters of indicators can move from 0 to 100 by 0.5 units.

1. Table: System building indicators

Source: GSB, own editing

Trendfollowing	Mean reversion	Momentum	Volume based	Volatility based	Level based
AccumDistClose	Bollinger Lower Band	AccumDist Mom	OBV	ADXR	Close
AccumDistCloseUpDn	Bollinger Upper Band	AccumDist MomClose	VWAP	ATR	CloseLessLowestLow

AccumDistDR	CCI	ADX		Chaikin Oscillator	CloseLessOpenDBpv
AccumDistDR0	DeCycler	FastK		Chaikin Oscillator Difference	CloseLessPrevCloseD
AccumSwingIndex	DeCycler Oscillator	MACD difference		ChaikinVolatility	CloseLessPrevCloseDBpv
Adaptive Moving Avgerage	Hurst	Momentum		Title:	Doctoral School of Economic and Regional Sciences
Discipline:	Business and Management Sciences	Head:		Supervisors:	Prof. Dr. Zoltán Zemman, PhD, professor,
Average Fc	Dr. habil. Judit Bárczi PhD, associate professor,	Rate of change		Range	CloseLessPrevR3
Counter Trend	Approval of the head of school	RSI		Standard deviation	Approval of the supervisor
Counter Trend Median	RoofingFilter1Pole2	SlowK		True Range	CloseOverPrevCloseD
CounterTrend2	ZeroCrossings	Stochastic			CloseOverPrevHighD
Dmi		SuperSmoother			CloseOverPrevLowD
Dmi Minus					CloseToHighLow5V4Pos
Dmi Plus					CloseToHighLow6V3Neg
Forward Reverse EMA					CloseToHighLow9
HighestFc					CloseToHighLow9V2Neg
LowestFc					CloseToPrevHighestHighLowestLow
TrendBiasExp					HighLow9LessClose
XAverage					HighLowC
					HighLowLvl
					Traffic-based
					OBV
					VWAP
					R2
					R3
					S1
					S1R1
					S2
					S2R2
					S3
					S3R3

E. Decision criterion for buy position: In the GSB software I defined the following decision condition for opening buy and sell positions:

[Sign(indicator1(parameter1)) * Power(Absolute value(indicator1(parameter1)),0)] * [Sign(indicator2(parameter2)) * Power(Absolute value(indicator2(parameter2)),0)] * [Sign(indicator3(parameter3)) * Power(Absolute value(indicator3(parameter3)),0)] Crosses upwards 0 & (indicator4(parameter4)) > X

The first part of the decision criterion consists of determining the sign of the value of the indicators, which can be negative or positive. Since exponentiation always raises to the zero power, exponentiation returns 1. By changing the power, indicators can be weighted. However, I dispense with this, so all indicators are given equal weight in the case of systems. The "Crosses upwards" formula means that a condition is true if the result was negative for the previous interval and positive for the current interval.

And the second part of the logical formula is true if the value of the indicator is above a certain value.

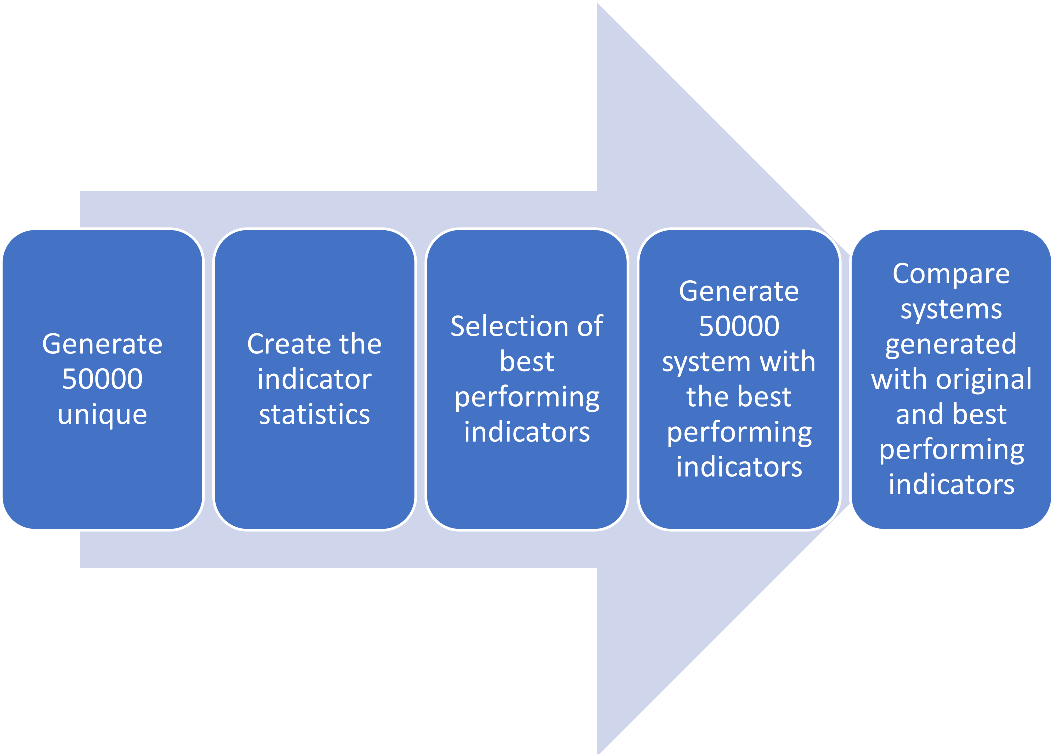
The decision condition for a sell position is the buy inverse.

The decision criterion includes determining the trading volume, which means 1 contract for each trade. In addition, we define a force majeure exit condition: A trade will be closed if your open loss exceeds $2,000 at any time.

F. Fitness criteria: The profit factor and the Pearson correlation are used as minimum criteria. During the test period, systems are expected to have a minimum profit factor of 1.5 and a correlation of 0.9. Only systems that meet these performance criteria will be saved for further testing.

G. Determining an optimal set of indicators: My assumption is that certain indicators in a given market generate systems with better results than other indicators. Based on Figure 2, as a first step, we generate 50,000 unique systems in which we include all our indicators (Table 1).

Then we aggregate the results. It is assumed that if an indicator occurs more often in systems with good performance than in systems with lower performance, then better systems can be built using that indicator. In practice, we do this by sorting the systems according to fitness criteria, and then splitting them into two groups (Top 50%, Bottom 50%) halfway through. Next, we look at how many times an indicator has occurred in the Top 50% and Bottom 50% groups. From the two frequencies, the incidence rate is calculated. It is assumed that an indicator that occurs more often in systems with a higher performing system, i.e. an indicator with a prevalence rate above 0.5, performs better.



2. Figure: Determining the optimal set of indicators

Source: Own editing

H. Selection of systems: From the 50000 systems generated based on the optimal set of indicators, we will select 8 systems:

- 2 systems according to the highest Pearson correlation

- 2 systems based on highest net profit
- 2 systems based on highest profit factor
- 2 systems based on maximum fitness criteria

The method is simple, but by choosing according to different criteria, it will be more likely that the best systems will be different from each other.

I. The selected 8 systems are verified on the 4 data streams specified in point C: The verification rate must be at least 50%. A system on a data stream is considered verified if it has at least a profit factor of 1.2 and a Pearson coefficient of 0.9. The 50% verification rate implies that a system is considered verified if the profit factor and the Pearson correlation on at least two data streams exceeded the minimum.

J. Walk-forward analysis of verified systems: We perform a fixed walk-forward analysis for systems with a verification rate of at least 50%. The recorded walk-forward analysis efficiency rate shall be at least 50%. Efficiency rate refers to the ratio of yield outside the sample to within the sample.

Q. Process repetition for next market

L. Portfolio validation: The portfolio consists of successfully verified systems with a WF efficiency of more than 50%. The portfolio is validated using the validation data set set in point C to determine the Pearson correlation coefficient for the validation period and the net profit.

2.2. Methodology of the traditional investment model

In the traditional investment model, the asset is purchased and held in the portfolio for a certain period of time (Thomsett, 1998). The purchase may be preceded by an analysis of the asset or its price according to different methodologies.
I use the markets described in Chapter 2.1.A (lean pork, wheat, coffee, frozen orange juice) for the traditional investment model as well. The comparison of the

two models only makes sense during the validation period (01.01.2022 – 01.09.2023), since the machine model is generated during the training and test periods.

Based on the methodology of the traditional model, on 01.01.2022 in all four asset classes, by analogy with the machine model, as many contacts will be purchased as there are systems in the machine model. A futures contract always means the contract of the nearest month. When the market moves from the current expiry month to the next expiry period, we roll over our contract.

At the end of the period – 01.09.2023 – the positions will be liquidated and profit statistics will be prepared for the period. Our main indicator is the net profit for the entire period, which will be decisive for both machine model building and traditional models. The correlation of the profit curve and the retraction, which are also decisive in machine model building, will not be useful for comparison due to one entry and one exit point of the traditional model.

2.3. Simple strategy model

With machine model building, I will also compare a simple but widely known trading strategy by technical analysts. During the moving average crossover, the 50 and 200 moving averages are calculated for each 30-minute time interval. If the 50 moving averages cross the 200 moving average from below, we open a buy position and/or close a sell position, if the 50 moving averages cross the 200 moving averages from above, we open a sell position and/or close a buy position. I wrote the code of the strategy in EasyLanguage programming language, I use the TradeStation trading platform for the simulation.

The programme code of the strategy:

```
FastAvg = AverageFC( Close, 50 );

SlowAvg = AverageFC( Close, 200 );
```

```
if CurrentBar > 1 and FastAvg crosses over SlowAvg then
    Buy ( !( "MA2CrossLE" ) next bar at market;
if CurrentBar > 1 and FastAvg crosses under SlowAvg then
    Sell Short ( !( "MA2CrossSE" ) next bar at market;
```

I run the model for the validation period (01.01.2022 – 01.09.2023) and in the four asset classes, analogously with the machine model, as many contacts are purchased as there are systems in the machine model.

3. RESEARCH RESULTS

Based on the methodology described, I build and verify systems on futures contracts using Genetic System Builder (GSB), TradeStation and Portfolio Analyst Pro software. The portfolio compiled from the results will be tested on the data hidden before the system building. Furthermore, I determine the results of the same portfolio according to the traditional model – buy and hold – and a simple strategy – moving average crossing.

3.1. Results of the machine model

In my research, I examined four futures markets. As a first step, I built 50,000 systems, analyzed and systematized the indicators used, and calculated the average metrics. Then I built another 50,000 systems using the better performing indicators, after which I also calculated the average metrics. For all four futures markets, I found that systems generated using better-performing indicators produced higher average metrics.

The final portfolio consists of a total of 20 schemes (Table 2), of which:

- 8 lean hogs (LH)
- 1 wheat (W)
- 5 coffee (KC)
- 6 frozen orange juice (OJ)

2. Table: Successfully verified and tested systems

Source: Own editing

ID	System
LH.20230909-151352-585540-7cTi4	((Sign(GSB_FastK(14) of Data1) * Power(Absvalue(GSB_FastK(14) of Data1), 0)) * ((Sign(GSB_HighLow9LessClose(22) of Data1) * Power(Absvalue(GSB_HighLow9LessClose(22) of Data1), 0)) * (Sign(GSB_HighLow9LessClose(30) of Data1) * Power(Absvalue(GSB_HighLow9LessClose(30) of Data1),

	1.75)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 0
LH.20230909-153133-651956-hFY41	((Sign(GSB_FastK(24) of Data1) * Power(Absvalue(GSB_FastK(24) of Data1), 0)) * ((Sign(GSB_CCI(46) of Data1) * Power(Absvalue(GSB_CCI(46) of Data1), 0)) * (Sign(GSB_CloseToHighLow5v4Pos(29) of Data1) * Power(Absvalue(GSB_CloseToHighLow5v4Pos(29) of Data1), 1.75)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 32.5
LH.20230909-154219-614805-Z8t7l	((Sign(GSB_CloseToHighLow5v4Pos(31) of Data1) * Power(Absvalue(GSB_CloseToHighLow5v4Pos(31) of Data1), 0)) * ((Sign(GSB_CCI(53) of Data1) * Power(Absvalue(GSB_CCI(53) of Data1), 0)) * (Sign(GSB_TrendBiasExp(24) of Data1) * Power(Absvalue(GSB_TrendBiasExp(24) of Data1), 1.75)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 15
LH.20230909-154304-998445-BVp3v	((Sign(GSB_CloseToHighLow5v4Pos(31) of Data1) * Power(Absvalue(GSB_CloseToHighLow5v4Pos(31) of Data1), 0)) * ((Sign(GSB_CCI(53) of Data1) * Power(Absvalue(GSB_CCI(53) of Data1), 0)) * (Sign(GSB_TrendBiasExp(24) of Data1) * Power(Absvalue(GSB_TrendBiasExp(24) of Data1), 1.75)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 45
LH.20230909-154651-313421-Uctor	((Sign(GSB_FastK(25) of Data1) * Power(Absvalue(GSB_FastK(25) of Data1), 0)) * ((Sign(GSB_HighLow9LessClose(21) of Data1) * Power(Absvalue(GSB_HighLow9LessClose(21) of Data1), 0)) * (Sign(GSB_CloseOverPrevLowD of Data1) * Power(Absvalue(GSB_CloseOverPrevLowD of Data1), 1.25)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 50
LH.20230909-154717-342029-5YbCi	((Sign(GSB_FastK(25) of Data1) * Power(Absvalue(GSB_FastK(25) of Data1), 0)) * ((Sign(GSB_HighLow9LessClose(21) of Data1) * Power(Absvalue(GSB_HighLow9LessClose(21) of Data1), 0)) * (Sign(GSB_CloseOverPrevLowD of Data1) *

	Power(Absvalue(GSB_CloseOverPrevLowD of Data1), 1.25)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 42.5
LH.20230909-163551-960930-aGEIh	((Sign(DMI(25) of Data1) * Power(Absvalue(DMI(25) of Data1), 0)) * ((Sign(GSB_HighLowLvlNeg(28) of Data1) * Power(Absvalue(GSB_HighLowLvlNeg(28) of Data1), 0)) * (Sign(GSB_SlowK(11) of Data1) * Power(Absvalue(GSB_SlowK(11) of Data1), 1.25)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 10
LH.20230909-163602-983301-NzJxo	((Sign(DMIPlus(22) of Data1) * Power(Absvalue(DMIPlus(22) of Data1), 0)) * ((Sign(GSB_LessCloseS2R2v2 of Data1) * Power(Absvalue(GSB_LessCloseS2R2v2 of Data1), 0)) * (Sign(AdaptiveMovAvg(Close, 10, 4, 20) of Data1) * Power(Absvalue(AdaptiveMovAvg(Close, 10, 4, 20) of Data1), 1.75)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 20
W.20230917-045839-115959-MAQJ4	((Sign(GSB_CloseToHighLow5v4Pos(35) of Data1) * Power(Absvalue(GSB_CloseToHighLow5v4Pos(35) of Data1), 0)) * ((Sign(GSB_CloseToHighLow9v3(124) of Data1) * Power(Absvalue(GSB_CloseToHighLow9v3(124) of Data1), 0)) * (Sign(GSB_KeltnerChannelv2(Close, 105, 2) of Data1) * Power(Absvalue(GSB_KeltnerChannelv2(Close, 105, 2) of Data1), 2)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 22.5
KC.20230919-224241-493439-KFE1M	((Sign(GSB_SS_RSI(26) of Data1) * Power(Absvalue(GSB_SS_RSI(26) of Data1), 0)) * ((Sign(GSB_Highest(High, 106) of Data1) * Power(Absvalue(GSB_Highest(High, 106) of Data1), 0)) * (Sign(GSB_AccumDistMomv2(Ticks) of Data1) * Power(Absvalue(GSB_AccumDistMomv2(Ticks) of Data1), 0.75)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 55
KC.20230919-200905-422000-BuuXP	((Sign(StandardDev(Close, 4, 1) of Data1) * Power(Absvalue(StandardDev(Close, 4, 1) of Data1), 0)) * ((Sign(GSB_KeltnerChannelv2(Close, 18, 3.5) of Data1) * Power(Absvalue(GSB_KeltnerChannelv2(Close, 18, 3.5) of Data1), 0)) * (Sign(GSB_CounterTrendMedian(Close, 12, 2.5) of Data1) * Power(Absvalue(GSB_CounterTrendMedian(Close, 12, 2.5) of Data1), 0.5)))) Cross 0 & 0 &

	GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 0
KC.20230919-153229-046788-ZTBD5	((Sign(GSB_ChaikinVolatilityv2(3, 3) of Data1) * Power(Absvalue(GSB_ChaikinVolatilityv2(3, 3) of Data1), 0)) * ((Sign(StandardDev(Close, 8, 3) of Data1) * Power(Absvalue(StandardDev(Close, 8, 3) of Data1), 0)) * (Sign(GSB_AvgTrueRange(28) of Data1) * Power(Absvalue(GSB_AvgTrueRange(28) of Data1), 1.5)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 25
KC.20230919-233504-836152-eyBkv	((Sign(GSB_CounterTrend(Close, 150, 1) of Data1) * Power(Absvalue(GSB_CounterTrend(Close, 150, 1) of Data1), 0)) * ((Sign(GSB_AveLessMedianv2(Close, 114, 5) of Data1) * Power(Absvalue(GSB_AveLessMedianv2(Close, 114, 5) of Data1), 0)) * (Sign(GSB_Highest(High, 30) of Data1) * Power(Absvalue(GSB_Highest(High, 30) of Data1), 0.75)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 30
KC.20230919-210702-646795-NtQek	((Sign(GSB_SS_RSI(32) of Data1) * Power(Absvalue(GSB_SS_RSI(32) of Data1), 0)) * ((Sign(GSB_SS_RSI(7) of Data1) * Power(Absvalue(GSB_SS_RSI(7) of Data1), 0)) * (Sign(GSB_CounterTrend(Close, 199, 4) of Data1) * Power(Absvalue(GSB_CounterTrend(Close, 199, 4) of Data1), 1.5)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 67.5
OJ.20230922-045352-796815-KIvBH	((Sign(GSB_ChaikinVolatilityv2(13, 140) of Data1) * Power(Absvalue(GSB_ChaikinVolatilityv2(13, 140) of Data1), 0)) * ((Sign(GSB_CCI(4) of Data1) * Power(Absvalue(GSB_CCI(4) of Data1), 0)) * (Sign(GSB_CloseLessPrevLowDv2 of Data1) * Power(Absvalue(GSB_CloseLessPrevLowDv2 of Data1), 0)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 7.5
OJ.20230922-065359-186625-D0RI0	((Sign(GSB_CloseLessPrevHighDv2 of Data1) * Power(Absvalue(GSB_CloseLessPrevHighDv2 of Data1), 0)) * ((Sign(GSB_CCI(17) of Data1) * Power(Absvalue(GSB_CCI(17) of Data1), 0)) * (Sign(GSB_HighLow9LessClose(11) of Data1) * Power(Absvalue(GSB_HighLow9LessClose(11) of Data1), 0))))

	Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 60
OJ.20230922-064325-146027-RBWAc	((Sign(GSB_DeCyclerOscillator(35, 45) of Data1) * Power(Absvalue(GSB_DeCyclerOscillator(35, 45) of Data1), 0)) * ((Sign(GSB_HighLow9LessClose(36) of Data1) * Power(Absvalue(GSB_HighLow9LessClose(36) of Data1), 0)) * (Sign(GSB_DMI(48) of Data1) * Power(Absvalue(GSB_DMI(48) of Data1), 0)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 40
OJ.20230922-064550-334839-fAonB	((Sign(GSB_DeCyclerOscillator(35, 45) of Data1) * Power(Absvalue(GSB_DeCyclerOscillator(35, 45) of Data1), 0)) * ((Sign(GSB_HighLow9LessClose(36) of Data1) * Power(Absvalue(GSB_HighLow9LessClose(36) of Data1), 0)) * (Sign(GSB_DMI(79) of Data1) * Power(Absvalue(GSB_DMI(79) of Data1), 0)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 40
OJ.20230922-045427-522624-OtxU2	((Sign(GSB_ChaikinVolatilityv2(13, 140) of Data1) * Power(Absvalue(GSB_ChaikinVolatilityv2(13, 140) of Data1), 0)) * ((Sign(GSB_CCI(4) of Data1) * Power(Absvalue(GSB_CCI(4) of Data1), 0)) * (Sign(GSB_CloseLessPrevLowDv2 of Data1) * Power(Absvalue(GSB_CloseLessPrevLowDv2 of Data1), 0)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 20
OJ.20230922-234520-464434-4x49z	((Sign(GSB_MedianBand(Close, 43, 1) of Data1) * Power(Absvalue(GSB_MedianBand(Close, 43, 1) of Data1), 0)) * ((Sign(GSB_CloseOverPrevHighDv2 of Data1) * Power(Absvalue(GSB_CloseOverPrevHighDv2 of Data1), 0)) * (Sign(GSB_CloseLessPrevLowDv2 of Data1) * Power(Absvalue(GSB_CloseLessPrevLowDv2 of Data1), 0)))) Cross 0 & 0 & GSB_Norm5(GSB_CloseToHighLow3v5(iSFLength) of Data(iSFData), 13, 100) of Data1 > 25

During the validation procedure, I proved that these systems produced reliable, stable results during the period under review (01.01.2022 – 01.09.2023), which was unknown to the systems before.

The task is to stitch together the validation results, calculate the portfolio-level results, and draw the appropriate conclusions from them.

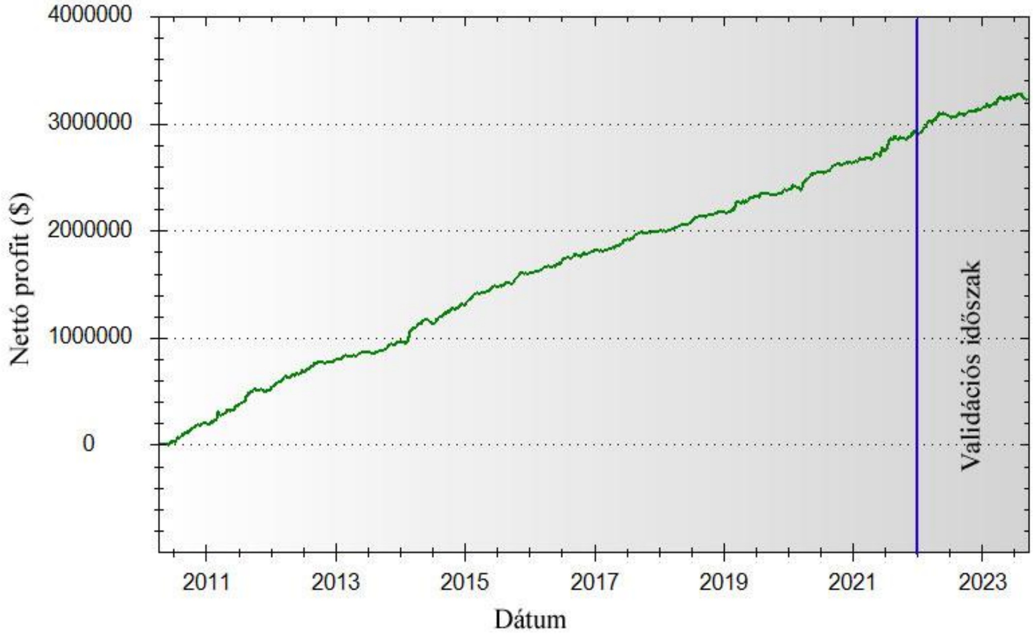
3. Table: Aggregate results of the portfolio

Source: Own editing

Training and test period		**Validation period**	
Net profit ($)	Pearson correlation	Net profit ($)	Pearson correlation
2891390	0,997	290667	0,967

Table 3 shows that during the validation period, which was completely unknown to the systems but contained real market data, net profit was proportional to the values calculated during the training and test period. This proves that we are able to build algorithmic trading systems using machine model building methods that can hold their own during live trading, with results showing profits based on future market price data. In this respect, it is irrelevant that we did not trade on an actual trading account during the validation period. Of course, trading enemies that only occur during live trading, such as technological glitches, human interactions, slippage would naturally have affected the results. At the same time, the validation took into account the average slippage and trading commission typical of the given market. These ensure that the results of the validation process would have been close to the net profit achieved in actual live trading.

Figure 3, showing the overall profit curve, is almost a 45-degree straight line. During the validation period, the characteristics and slope of the curve are very similar to those of the training and test periods, so it has been proven that it is possible to build a portfolio that results in balanced profits and continuation of the profit curve on exchange rate data unknown to the system.



3. Chart: Aggregate profit curve of the portfolio

Source: Portfolio Analyst Pro, Own editing

3.2. The traditional model and its results

The machine model portfolio consists of a total of 20 systems (8 LH, 1 W, 5 KC, 6 OJ). Maintaining consistency and comparability requirements in the traditional - buy and hold - portfolio, at the beginning of the validation period, as many contacts from each market are purchased as there are systems in the machine model – given that each system trades with one contract. We open on 01.01.2022 and close all trades on 01.09.2023. The transaction cost is negligible as there is only one transaction per market. The price movement per instrument during the validation period is shown in Chart 4.

4. Chart: Price development of lean pig (LH), wheat (W), coffee (KC), frozen orange juice (OJ) during the validation period

Source: TradeStation Inc.

The results of the traditional portfolio corresponding to the parameters are summarised in Table 4. From this it can be seen that in more than one and a half years we can only realize profits on the frozen orange juice market, in which the exchange rate has increased almost fourfold. As a result, the net profit of the frozen orange juice market was able to offset the losses of other markets and we ended the validation period with a profit of $60160.

4. Table: Results of the traditional model

Source: own editing

	Entry date	**Entry price ($)**	**Contract size**	**Quantity**	**Market value ($)**	**Exit date**	**Exit price ($)**	**Market value ($)**	**Profit/loss**
Lean hog (LH)	2022-01-01	91,425	40000 font	8	292560	2023-09-01	85,725	274320	-18240
Wheat (W)	2022-01-01	946,25	5000 véka	1	47312,5	2023-09-01	628,75	31437,5	-15875
Coffee (KC)	2022-01-01	202,90	37500 font	5	380437,5	2023-09-01	145,90	273562,5	-106875
Frozen orange juice (OJ)	2022-01-01	82,60	15000 font	6	74340	2023-09-01	306,10	275490	201150
Total									60160

3.3. Simple trading model results

I run the moving average crossing strategy described in Chapter 2.3 using EasyLanguage and TradeStation trading platform. Trading volumes per market are adjusted to machine model: We trade a total of 20 contracts (8 LH, 1 W, 5 KC, 6 OJ). The transaction costs are the same as described in machine model building.

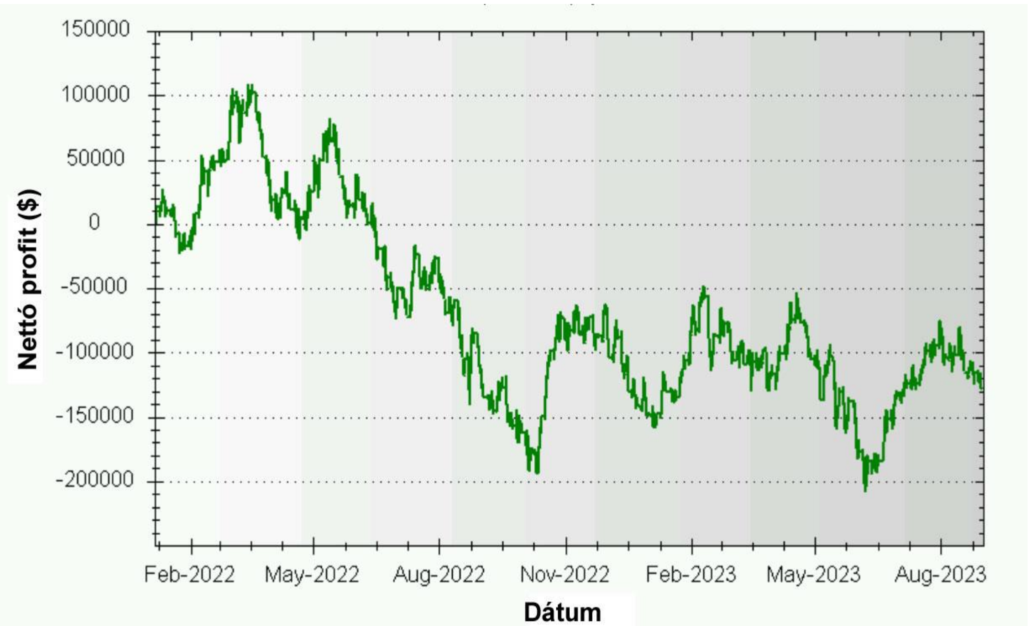
Table 5 shows the results achieved by the simple model. It can be seen that the strategy could not generate profit during the validation period.

5. Table: Results of a simple trading strategy during the validation period

Source: Own editing

	Number of contracts	Gross profit ($)	Transaction cost ($)	Net profit ($)	Number of trades	Profit factor
Lean hog (LH)	8	-112968	6032	-119000	29	0.57
Wheat (W)	1	9801	2759	7042	89	1.11
Coffee (KC)	5	58156	10875	47281	50	1.16
Frozen orange juice (OJ)	6	-60069	3996	-64065	37	0.71
Total	20	-105080	23662	-128742	205	0.85

After interpreting Figure 5, it is clear that the profit curve is hectic, after an initial upward phase, the strategy continues to generate losses.



5. Figure: Profit curve of the simple model during the period under review

Source: Portfolio Analyst Pro

4. CONCLUSIONS AND PROPOSALS

It has been proven that machine model building is a successful approach to stock trading, in which the use of artificial intelligence is unavoidable. Advanced search algorithms and machine learning allow you to select algorithms that were profitable during the study periods from a large number of combinations. With the help of verification procedures, we can increase the likelihood that selected systems will successfully trade on future data. From the results, it became clear that these machine models are more stable and can achieve higher returns than the traditional model – buy and hold – and trading according to a simple strategy.

Of course, the key to success lies in following the methodology precisely, executing trades systematically and instilling a certain degree of trust in trading models, since it is impossible to predict how long a system will operate in the future and generate the expected profit. Verification and validation procedures are of great importance here, as they allow us to model changes in the characteristics and internal life of the exchange rate and calculate how our systems react to them. Robust systems give traders more confidence as they deliver stable results in multiple situations. At the same time, it is also possible that a strategy generates losses from the first day of live trading, so it is definitely necessary to develop a methodology that makes decisions based on objective metrics about how long a given system can remain part of the portfolio, where is the point after which we have to delete it and replace it with another system. This was not the goal of my research, the focus was on the diversity of machine model building, and on the possibility of building a portfolio that can generate profit even in the long run, without human intervention, and is able to overcome the traditional investment model. I used diversification so that the performance of a system only burdens the overall performance with its own weight, so it was actually not important when a system did not bring the expected results. I think the whole process is reproducible, so if I repeated machine model building, the results would be similar.

It was important to calculate how the generated systems relate to each other. We have repeatedly encountered that some systems are strongly correlated with each other. Taking a closer look at this phenomenon, we found that the structure of these systems is very similar, their set of indicators is the same in a large percentage, they differ mainly in their parameters. It is advisable to filter out these systems, since, as I mentioned earlier, we cannot know how a system will behave on unknown exchange rate data, so it is advisable to include systems in the portfolio that have a maximum slight correlation between them – that is, their set of indicators and parameters is different.

Using the methodology detailed in the machine model building literature, I successfully built a portfolio, which I did not run in live trading, but on data unknown to him. Although this test was successful, it should be remembered that during live trading, there are several unforeseen factors that make execution more difficult, such as slippage, i.e. the difference between the desired and actual price. In live trading, you need to know and, above all, accept the potential risks. It is even more important to accept the risks that arise from portfolio failure and result in real financial losses.

5. NEW AND NOVEL SCIENTIFIC DEVELOPMENTS

During my research, I built stock trading systems using genetic algorithms. I compared these with the traditional – buy and hold – investment model and a simple and generally known strategy – moving average crossing. I found that machine model building can produce better results and higher net profit.

The main steps and processes of system building are predefined and can be found in many literatures, however, the process with which I built the systems also contains unique steps and parameters defined by me. I consider the process of system building defined by me (Chapter 3.1) as an achievement, since with this methodology I successfully built trading systems that brought the expected results in the long run, and the Pearson correlation of the profit curve during the validation period differed only slightly from the correlation of the profit curve during training and test periods.

It has been proven that among the indicators of technical analysis and their parameters, there are those that can be used to create valid trading systems and that are also suitable for live trading. The technology provides an opportunity to find these indicators and their parameters much more efficiently and last but not least faster by using advanced search algorithms and artificial intelligence.

I found that there are computer-controlled algorithmic models in financial markets that can generate profits in the long run. The portfolio built on futures markets produced reliable and stable profits during the period under review. The process can be repeated and the results reproducible.

I determined that the aggregate out-of-sample trading results of diversified systems based on different indicators are more balanced than the individual results of individual systems. The correlation of the profit curve of each system was lower than that of the entire portfolio, regardless of the time period in which it was examined. Thus, the need for diversification was demonstrated and that diversification is one of the simplest and most important tools of risk management.

I also found that different technical indicators can be used with different degrees of effectiveness in different markets. In the futures markets I examined, I found that indicators showing better results in a given market can help generate more profitable trading algorithms.

6. PUBLICATIONS OF THE AUTHOR RELATED TO THE TOPIC OF THE THESIS

1. Csesznik, Zoltán (2021). Az algoritmikus kereskedés anatómiája
   CONTROLLER INFO 9 : 4 pp. 40-43. Paper:
   10.24387/CI.2021.különszám.7 , 4 p.
   Közlemény:32494880 Nyilvános Forrás Idéző Folyóiratcikk (Szakcikk)

2. Csesznik, Zoltán (2023). Creating trading algorithms using machine model building methods
   CONTROLLER INFO 11 : Special Issue pp. 57-61. Paper:
   10.24387/CI.2023.SPECIAL ISSUE.9 , 5 p.
   Közlemény:34525433 Nyilvános Forrás Folyóiratcikk (Szakcikk)

3. Csesznik, Zoltán (2023). Határidős piacok és szerződések működési mechanizmusai
   CONTROLLER INFO 11 : 3 pp. 39-42. Paper: 10.24387/CI.2023.3.8 , 4 p.
   Közlemény:34525377 Nyilvános Forrás Folyóiratcikk (Szakcikk)

4. Csesznik, Zoltán (2021). Irányítatlan árfolyamkereskedés – stabil profit a pénzügyi piacokon
   CONTROLLER INFO 9 : 4 pp. 44-48. Paper: 10.24387/CI.2021.4.8 , 5 p.
   Közlemény:32538622 Nyilvános Forrás Folyóiratcikk (Szakcikk)

5. Csesznik, Zoltán (2022). A kereskedési algoritmusok ellenségei
   CONTROLLER INFO 10 : 4 pp. 52-56. , 5 p.
   Közlemény:34525295 Nyilvános Forrás Idéző Folyóiratcikk (Szakcikk)

6. Zoltán, Csesznik ; Sándor, Gáspár ; Gergő, Thalmeiner ; Zoltán, Zéman (2021). Examining the effectiveness of fundamental analysis in a long-term stock portfolio
ECONOMIC ANNALS-XXI 190 : 5-6(2) pp. 119-127. , 9 p.
Közlemény:32556708 Admin láttamozott Forrás Idéző Folyóiratcikk (Szakcikk)

7. Gáspár, Sándor ; Thalmeiner, Gergő ; Csesznik, Zoltán ; Zéman, Zoltán (2022). Részvényár előrejelzés neuro-fuzzy modell alkalmazásával, pénzügyi-számviteli mutatószámok felhasználásával
Közlemény:33568016 Nyilvános Forrás Könyvrészlet (Absztrakt / Kivonat)

8. Csesznik, Zoltán (2022). Kripto a jövőbe csomagolva
Pécs, Magyarország : Pécsi Tudományegyetem Állam- és Jogtudományi Kar (PTE ÁJK) pp. 40-48. , 9 p.
Közlemény:33070335 Nyilvános Forrás Könyvrészlet (Konferenciaközlemény)

9. Zoltán, Csesznik (2022). Enemies of trading algorithms
In: Zsolt, Pál (szerk.) New aspects of financial literacy : Collection of Studies
Miskolc, Magyarország : FINTELLIGENCE Financial Literacy Center 186 p. pp. 26-38. , 13 p.
Közlemény:33366588 Munkapéldány Nem nyilvános Forrás Könyvrészlet (Szaktanulmány)

10. Suhajda, Ádám ; Csesznik, Zoltán (2021). Tőzsdék és ECN kereskedési rendszerek versenyhelyzete az algoritmikus kereskedés piacán, valamint a BÉT stratégiájának revíziója a témában
CONTROLLER INFO 9 : 4 pp. 53-58. , 6 p.

Közlemény:32536606 Nyilvános Forrás Idéző Folyóiratcikk (Szakcikk)

11. Csesznik, Zoltán (2022). Beating The Indexes Using Fundamental Analysis In Long-Term Trading: Searching For Alpha By Using Company Fundamentals
Technical Analysis of Stocks&Commodities 40 : 5 pp. 26-31. , 6 p.
Közlemény:32788214 Nyilvános Forrás Folyóiratcikk (Szakcikk)

12. Zoltan, Csesznik (2021). Fundamental risk management in trading stocks
Aspects of Financial Literacy : Proceedings of the International Scientific and Practical Conference March 22-23, 2021
Sumy, Ukrajna : Sumy State University (SSU) 417 p. pp. 389-402. , 14 p.
Közlemény:32534330 Nyilvános Forrás Idéző Könyvrészlet (Konferenciaközlemény)

13. Csesznik, Zoltán (2012). Matching strategy and stocks
Active Trader 13 : 7 pp. 28-31. , 4 p.
Közlemény:32290898 Nyilvános Forrás Folyóiratcikk (Szakcikk)

14. Csesznik, Zoltán (2012). Defeating Slippage
Technical Analysis of Stocks&Commodities 30 : 5 pp. 38-40. , 3 p.
Közlemény:32290890 Nyilvános Forrás Folyóiratcikk (Szakcikk)

15. Csesznik, Zoltán (2021). A vállalatértékelés fontossága a kereskedési döntések meghozatalának előkészítésében
CONTROLLER INFO 9 : 2 pp. 59-62. , 4 p.
Közlemény:32122168 Nyilvános Forrás Idéző Folyóiratcikk (Szakcikk)

16. Csesznik, Zoltán (2021). Tőzsdei indikátorok hatékonysága
CONTROLLER INFO 9 : 3 pp. 28-31. , 4 p.
Közlemény:32290793 Nyilvános Forrás Folyóiratcikk (Szakcikk)

7. BIBLIOGRAPHY

1. Aslam, F., Mohti, W., Ali, H., & Ferreira, P. (2023). Interplay of multifractal dynamics between shadow policy rates and stock markets. *Heliyon*, *9*(7), e18114. https://doi.org/10.1016/j.heliyon.2023.e18114

2. Chen, J.-H., & Tsai, Y.-C. (2020). Encoding candlesticks as images for pattern classification using convolutional neural networks. *Financial Innovation*, *6*(1), 26. https://doi.org/10.1186/s40854-020-00187-0

3. Fry-McKibbin, R., & McKinnon, K. (2023). The evolution of commodity market financialization: Implications for portfolio diversification. *Journal of Commodity Markets*, *32*, 100360. https://doi.org/10.1016/j.jcomm.2023.100360

4. Huang, S.-C., Chiou, C.-C., Chiang, J.-T., & Wu, C.-F. (2020). Online sequential pattern mining and association discovery by advanced artificial intelligence and machine learning techniques. *Soft Computing*, *24*(11), 8021–8039. https://doi.org/10.1007/s00500-019-04100-5

5. Kühl, N., Schemmer, M., Goutier, M., & Satzger, G. (2022). Artificial intelligence and machine learning. *Electronic Markets*, *32*(4), 2235–2244. https://doi.org/10.1007/s12525-022-00598-0

6. Nassar, D. S. (2006). *Ordinary people, extraordinary profits: How to make a living as an independent stock, options, and futures trader*. Wiley.

7. Nicholls, J. F., & Engelbrecht, A. P. (2019). Co-evolved genetic programs for stock market trading. *Intelligent Systems in Accounting, Finance and Management*, *26*(3), 117–136. https://doi.org/10.1002/isaf.1458

8. Thomsett, M. C. (1998). *Mastering fundamental analysis*. Dearborn Finanical Pub.