THESES OF DOCTORAL (PHD) DISSERTATION

Barta Ákos

GÖDÖLLŐ

2024

MagyarHungarian University of Agriculture and Life Sciences

OIL PRICE FORECASTING USING ARTIFICIAL NEURAL NETWORKS BASED ON WALL STREET JOURNAL ARTICLES

DOCTORAL (PHD) DISSERTATION

BARTA ÁKOS

GÖDÖLLŐ

2024

Doctoral School

Name: Doctoral School of Economic and Regional Sciences

Discipline: management and organizational science

Head: **Dr. Bujdosó Zoltán PhD**
university teacher
Hungarian University of Agricultural and Life Sciences
Institute of Sustainable Development and Management

Supervisor(s): **Dr. habil Molnár Márk PhD**
university associate professor
Lóránd Eötvös University
Faculty of Economics
Department of Comparative Economics
Dr. Naárné Dr. Tóth Zsuzsanna Éva PhD
university associate professor
Budapest Metropolitan University
Institute of Economics

....

Approval of Head of Department Approval of Supervisor(s)

Content

1. Background and objectives of the work.

More than 40 years have passed since the first oil price shock in 1973. During this period, global demand for oil increased dramatically, while new energy technologies and new sources of energy made global consumers more resilient to oil shocks. Since the oil shocks of the 1970s, the role of emerging economies has increased significantly in global energy consumption. The share of the People's Republic of China, for example, is five times larger than it was in the 1970s. On the other hand, the shares of the largest (USA) and currently the third largest (Japan) oil consumers have decreased since the 1970s, from 32% to 21% for the United States and from 10% to 5% for Japan.

Following the oil crises of the 1970s and subsequent economic recessions, several studies found that oil price shocks played a significant role in economic downturns. In recent years, the sharp rise in oil prices that began in 2001 and the sharp drop in 2008 following the subprime mortgage crisis have rekindled interest in the effects of oil prices on the macro economy.

Fossil fuels continue to cover a significant part of current global energy consumption, with a corresponding share of 80% in 2020. (IEA, 2020). Oil remains the world's leading fuel, accounting for 31.2% of global energy consumption in 2020, indicating that crude oil remains important in international factor markets. Therefore, understanding the historical development of crude oil price determinants is an extremely important research question for economic policy planning. In this regard, the price movements on the crude oil market in the first decade of the new millennium attracted attention mainly for two reasons: on the one hand, the price rose for several years from the low level of the 1980s and 1990s, setting records. This trend in prices has sparked widespread debate about another oil crisis, referring to the two oil crises of the 1970s and 1980s. Second, and more importantly, unlike the first two oil crises, the reason for this current price peak is not clear: several potentially relevant developments took place at the same time, which makes it difficult to identify their impact on the price. More and more articles in the economic literature bear witness to this. Academic discourse generally oscillates between three explanations, reflecting the market forces affecting the price of crude oil: first, it is argued that price increases reflect the finiteness of crude oil reserves and the inability to further expand production capacities (supply-driven price increases). (Kaufmann, 2011) Second, it is hypothesized that the unexpectedly strong economic growth of emerging countries such as China and India resulted in an unexpected increase in demand for crude oil, leading to a squeeze on the spot supply of crude oil and an increase in price (demand-driven price inflation). (Hamilton, 2008) (Kilian, 2009).

Based on the fact that crude oil is one of the most important raw materials, price fluctuations have a significant impact on the global economy. As a result, economic actors try to predict exchange rate changes, trends, and future values. Since the majority of economic actors are not in a position to influence the market, they try to extract data and information from the information they receive.

In today's fast-paced world, we are faced with an enormous mass of information. There are thousands of clickbait articles and sources believed to be authentic, from which we try to gather useful information for us. The key is therefore not necessarily to acquire information or news, but to filter the relevant information content from the flow of information and then process it efficiently. The problem is complex and economic actors are applying their increasingly complex information processing and filtering methods, trying to filter out noise, that is, false information and additional facts, from increasingly complex information networks and large amounts of data.

The topic of this thesis is the analysis of the oil exchange rate using a more complex approach. Since it affects everyone and indirectly affects all industries, it plays a big role. It operates in an oligopoly market, which means that the actors pay a lot of attention to the decisions of the other actors. Since each player has a relatively large market share, they can have a big impact. As a user, as an end consumer, there is no impact on the market, but we need to apply a special filter from the large mass of information in order to know which news is worth taking seriously.

In recent years, neural networks have gained more and more space, which are able to recognize correlations and conclusions, connections between the points of data sets, such as news, on non-statistical bases. Neural networks began to develop around 1940, parallel to the development of computer technology. At first there were capacity limitations, but with the current technological development, such developments are enjoying their heyday. The method of information processing in the case of neural networks is very similar to the network present in the human brain, it does not necessarily work on a statistical basis. In addition, it is much faster and more efficient than what a person will be able to do.

My assumption is that based on the content of newspaper articles related to economic and political news, oil price changes can be predicted, at least to some extent. With the help of economic and political information in newspaper articles, you can answer the question of what kind of change is expected in oil prices. Based on the information in the news, it is predictable, in my opinion, when an increase or decrease in oil prices is expected, as well as the extent of the change.

It is likely that if positive or negative opinions regarding oil prices are published in the written or electronic press, this may affect the decisions of buyers and managers, which may ultimately affect the change in oil prices. Based on the news, the market participants make decisions based on speculative grounds, which will result in an actual price change due to a change in the demand or supply side, so that the originally commented and probable result may not even occur.

It is important to state that economic actors mean households and companies that have information, but are not "close to the fire", that is, they can get information secondhand and rely on the credibility of the trade press.

In the current information dumping, economic actors try to process the available information. During information processing, it is necessary to filter out invalid information and to weigh the influence of individual facts or opinions. In this way, we can assume that the news and information have a manipulative nature. Regarding the oil price, which is a typical indicator of an oligopoly market where production decision-making is not public, economic actors can rely heavily on the news.

Methodologically, the thesis is basically based on the use of artificial neural networks, which are currently enjoying their heyday, becoming more and more organically incorporated into our lives. In my opinion, the usefulness, structure and development of the methodology will be one of the main topics of the coming period. As a result, I am investigating the applicability of neural networks in science and daily life, more precisely how well they can help our daily activities and decision-making methods.

My goal in my research is to prove that by analyzing the flow of information (i.e. news, press information, not fundamental data) there is a connection between the oil price and press news, thus it can be used as a forecast on the one hand, and the exchange rate movement based on speculation can be detected on the other hand.

In connection with the topic and the research, I formulate the following hypotheses, which I will investigate during my research:A mesterséges neurális hálózatok képesek hatékony információfeldolgozásra, vagyis nagy adattömegek (Big Data) gyors és hatékony elemzésére is használhatóak.

[1] A correlation can be shown between the examined journals and the exchange rate, i.e. the speculative exchange rate movement with respect to the oil price can be clearly proven.
[2] By analyzing the content of the Wall Street Journal (WSJ) newspaper articles with an artificial neural network (ANN), the next day's oil price change can be determined with sufficient accuracy.
[3] By summarizing, or compressing, newspaper articles, the efficiency of forecasting with an artificial neural network can be greatly increased.
[4] By analyzing the sentiment of newspaper articles, the prediction efficiency with an artificial neural network can be greatly increased.
[5] The next day's oil price change can be determined with sufficient accuracy by examining the oil exchange rate with a feedback neural network (RNN), i.e. only by analyzing the historical movement of the exchange rate.
[6] By greatly increasing the hidden layers and neurons of the artificial neural network, i.e. the network part, the efficiency can be significantly increased.
[7] The analysis of WSJ newspaper articles with an artificial neural network is more effective than the analysis of the exchange rate with a feedback neural network, that is, speculation has a greater role in the movement of the exchange rate than fundamentals.
[8] The forecast obtained with the help of an artificial neural network is more effective than forecasting exchange rate changes with specific mathematical stock market models.

2. Material and methods

An oil price forecasting neural network (NN) uses Wall Street Journal articles that are about oil prices for analysis and processes this information using the neural network.

Using the Neural Network, which provides oil price forecasting, helps to understand the effect of market conditions on the price of oil. Researchers may be able to understand correlations between speculative market conditions and Wall Street Journal articles, thereby better understanding how markets work and predicting future oil prices.

An Artificial Neural Network (ANN) learns the information summarized by the article regarding market movements and applies it to draw conclusions. ANN considers market trends, market tensions, highs and lows, and other useful information that can be gleaned from Wall Street Journal articles. ANN predicts the price of oil and uses different assumptions to understand different market situations. ANN also takes into account technical analysis and, in addition, other techniques that can be used in forecasting.

In the thesis, I explained in detail the operation of ANN and its partners. The neural network allows to improve a poor quality solution; we can refine it on the mesh, then run the model again and trust that the result will improve.

There is no uniform oil price worldwide, since the quality of the extracted oil and oligopoly markets determine prices separately. At the same time, the direction and magnitude of the price change is very similar, so the different prices move together.

In reality, there are many different types and grades of crude oil—the thick, unprocessed liquid that drilling rigs extract from deep in the earth—and some are more desirable than others. For example, refiners find it easier to produce gasoline and diesel from low-sulfur or "sweet" feedstock than from high-sulfur oil. Low-density or "light" crude oil generally favors the high-density variety for the same reason. (Jiang, An, Jia, & Sun, 2017)

It also matters where the oil comes from if you are a buyer. The cheaper the delivery of the product, the cheaper it is for the consumer. From a transportation perspective, offshore oil has certain advantages over onshore supplies that depend on pipeline capacity.

Because of these factors, crude oil buyers, along with speculators, need a simple way to value the commodity based on its quality and location. Benchmarks such as Brent, WTI and Dubai/Oman serve this important purpose. When refiners buy the Brent contract, they have a firm idea of how good the oil will be and where it will come from. Today, most of the global trading takes place in the futures market, and each contract is tied to a certain oil category.

Due to the dynamic nature of supply and demand, the value of each benchmark is constantly changing. In the long term, a marker sold at a premium to another index may become available at a discount. (Arshad, Rizvi, Haroon, Mehmood, & Gong, 2021)

Roughly two-thirds of all crude oil contracts in the world are for Brent Crude, making it the most widely used indicator. Today, "Brent" actually refers to oil from four different fields in the North Sea: Brent, Forties, Oseberg and Ekofisk. Crude oil from this region is light and sweet, making it ideal for refining diesel fuel, gasoline and other high-demand products. Since the supply is by water, it can easily be transported to remote locations.

WTI refers to oil produced from wells in the United States and sent by pipeline to Cushing, Oklahoma. The product itself is very light and very sweet, making it particularly ideal for gasoline refining. WTI remains the main benchmark for oil consumed in the United States.

Middle East crude is a useful benchmark for oil slightly weaker than WTI or Brent. A "basket" of raw materials from Dubai, Oman or Abu Dhabi, it is slightly heavier and has a higher sulfur content, so it can be classified as a "sour". Dubai/Oman is the main point of reference for Persian Gulf oil delivered to the Asian market. The oil extraction market forms an oligopoly market, that is, the decisions, statements, forecasts or even local crises of certain actors can have a great impact on market prices.

Although Brent better reflects the change in oil prices on the world market, taking into account the other parts of the research, i.e. the fact that I examined an American-published newspaper that deals heavily with domestic oil production, references to this may often appear in the articles, henceforth as an oil price benchmark the I used WTI price.

During the analysis of the websites, the Wall Street Journal (hereinafter: WSJ) proved to be the most suitable for data collection during the research to be carried out, so the scraper was run here.

During the research, I downloaded, or rather scraped, the articles published by the WSJ between 2000-2020. In the case of the articles, I summarized them for possible future use, that is, I also saved the abbreviated version.

Web scraping i.e. web data collection or web data extraction, is a data acquisition method used to extract data from websites. Web scraper software can directly access the World Wide Web using the Hypertext Transfer Protocol or a web browser. While data collection can be done manually by the software user, this term usually refers to automated processes implemented with a bot or web crawler. It is a form of copying in which specific data is collected and copied from the web, typically into a central local database or table, for later retrieval or analysis.

The complete execution of the code line took 1604 hours, that is, it ran continuously for almost 67 days. On average, he reviewed and saved a month's worth of articles in 401 minutes. The period between 2000-2020, i.e. 21 years, represents a total of 330,435 articles. The .xlsx file created in this way was 558 MB in size.

Data processing systems based on neurological networks in the brain and implemented in a programming environment according to their structure and pattern. The systems are primarily used for pattern identification and processing, and are able to gradually improve performance based on the analysis results of previous tasks. (Jain, Mohiuddin, & Mao, 1996)

Neural organization and networking can basically be explained with the multilayer perceptron model. In this model, neural networks take the form of layers that make connections in one direction, also known as feedforward neural networks. Nodes have several layers: input, hidden, and output. The connections between different nodes change the behavior of networks. The input layers receive information, then connections are established between the input and hidden layers. The hidden layers then process the information, which in turn goes to the output layers. Finally, the output layers become the input to the next layer and the sequence continues. (Xin, 1999)

Artificial intelligence is a computer program that can organize information in a way similar to the human brain. Artificial intelligence, a combination of neural networks, emerged as a result of research on cognitive talent and machine design. (Kutsurelis, 1998) The history of artificial intelligence goes back to Aristo. It is known that Aristo worked on the algorithm of thinking and discussed its difficulties. In the modern sense, artificial intelligence entered the scientific world when the first electronic computer was put into operation in the 1940s (note: there were computers before, e.g. the German Konrad Zusé in the 30s, but this we'll leave it at that for now) and Alan Turing developed the first software. Artificial neural networks, the most significant sub-segment of artificial intelligence, are a statistical approach created to develop predictive models. Artificial neural networks consist of processing devices and data processing similar to the design of the human brain. (Blackard & Dean, 1999)

During the examination with the neural network, I apply several settings, expand them or take them away, thereby slimming the result, according to the more accurate and more efficient estimation of future changes. First of all, I focus on percentage changes in the exchange rate, that is, based on the content of the articles, I can tell how likely it is that the direction and magnitude of the next day's exchange rate change will be.

I used two different methods. In the first case, I took into account every single date and exchange rate change. In the second case, only those that have reached a certain indicator variable. I counted the occurrence of given indicator words in the articles and only considered them valid and examined the article further if they occurred in a given number. Otherwise, I did not consider it, skipped the article and did not analyze it. Examining it, and avoiding the distorting effect of words with few articles and examining it not from the exchange rate, but only from the article side.

For articles, I use indicators and keywords. In this case, the indicator words are words whose occurrence makes it likely that the topic of the article is related to the price of oil, so it is recommended for further investigation. Keywords are the words whose quantitative occurrence I examine in the articles, I look for a connection between their frequency of occurrence and previous exchange rate changes, which can provide a pattern for future changes. The composition of the words is based on your own choice, and applies to both price increases and political influence, as well as production words. On the fly, ANN weighs exactly which one is needed and how much it should take into account. Also, due to the running of several versions, the utility and the final formula are formed according to plans.

Also, in the case of the neural network, I constantly use new methods and variants during the research, such as neural network, deep learning method with feedback neural network, introduction of stock market moving average in the input side of the neural network, sentiment analysis, artificial summary of articles.

I examine how each expansion and modification increases efficiency, and which version can work most effectively, and how much more efficient it is than other predictions.

3. Results

In the first round, I used a model without deletion, that is, when I also analyzed the days where there was a change in the oil exchange rate, but there were not enough indicators among the articles to assume an article related to oil. In other words, all keywords were calculated with 0 units.

In the further part of the study, I ran an analysis in which I deleted the days that did not have the sufficient indicator number. That is, I tested the hypothesis that days without adequate data could be misleading, so I ignored them. In this case, a significant part of the days are lost, so the white noise affecting the ANN, or more precisely the large changes but runs without keywords, are reduced to a minimum, thereby promoting the cleanest possible analysis. In the future, it is necessary to examine how much can be examined quantitatively, whether we have, so to speak, a sufficient amount of examined days.

From the results obtained in this way, it is clear that the model is not capable of accurate results with its current development and settings. In other words, you can only determine the next day's exchange rate with a rough estimate, with a large error. It is important to clarify what we mean by an accurate estimate. If we accept the deviation or accuracy within 1.2% for the next exchange rate change, the model calculates correctly in ~45% of the cases. Which is not necessarily a bad thing.

In addition, the SIGN (sign prediction) efficiency, i.e. the determination of the sign (direction) of the next day's exchange rate change, functions with an accuracy of around 60%, which clearly indicates that there is a connection between the articles and the exchange rate forecast. During the examined period, the exchange rate rose by 51.77%. The rise surplus is justified taking inflation into account. However, if we say that the exchange rate will rise the next day, versus the SIGN results of the neural network, the model predicts the next day's exchange rate more effectively.

In the next test, I ran an RNN, or feedback neural network. In the first round, I only estimated the one-day price, and later the change as well. Also, I put the results obtained in this way back into the modified ANN without deletion. I also examined the possibility of how much the 12-day and 26-day moving average, or MACD, thus increases or decreases the effectiveness of Signal.

That is, I used a double neural network to improve the results. This is a more complex Deep Learning method. Based on the numbers of the original ANN layer, we can already speak of a deep learning method, as a result of this expansion, a more complicated deep learning method is created, neural network results are embedded in the neural network.

Efficiency improved in some cases, but not in others. Currently, we cannot say whether time series analysis increases or decreases efficiency. For the time being, I can state that it shows no significant change in either a positive or negative direction.

In addition, it is important to note that at the moment I tried to forecast from pure prices, we did not examine any kind of stock market analysis, such as the previously explained MACD analysis. It is a question of whether this increases efficiency in any given case or cannot make any significant changes.

In the first version, I searched the entire article and examined the articles. As a second setting, I summarized the articles and analyzed the keywords within them. Investigating whether block-building and highlighting increases efficiency. In this way, what has to be said is concentrated, the vocabulary presumably decreases, which increases the occurrence of relevant words, thus the complete analysis.

Based on the examination of the results, it is clear that the sensitivity of the model is greatly increased by analyzing only an extract, rather than the entire article. In other words, the "concentration" of articles increases the effectiveness of keyword research.

We can declare that it is necessary to further examine the abstract of the articles during the subsequent development and efficiency increase tests.

As a supplement to the keyword analysis, we also performed the mood or sentiment analysis of the articles using the Nltk VADER lexicon. Sentiment analysis is a procedure that can be used to determine the state of mind, emotions and opinions of written texts, usually articles, comments or tweets appearing in the media. Nltk (Natural Language Toolkit) is a library written in Python that contains many tools and English-language corpora for computational linguistics research. Nltk has tools such as tokenization, lemmatization, and stop word removal that can be used to clean data to make it more manageable for analysis. The VADER (Valence Aware Dictionary and sEntiment Reasoner) lexicon is a tool specially developed for sentiment analysis of social media content.

The addition of sentiment analysis clearly resulted in an increase in efficiency. Although the overall results resulted in only a small increase, they can clearly be said to be more effective. It is interesting to observe that the results are significantly better for indicator number 1, as well as for indicator number 2. For three indicators, the previous analysis is slightly more efficient, and both methods prefer exactly the same results by increasing the number of indicators. However, since we prefer fewer indicator numbers, i.e. more analyzed article event series, we always consider the increase in efficiency in the case of fewer indicator numbers to be more useful.

I am expanding the existing neural network with the MACD method used in stock exchange rate analysis. I have detailed this before. In short, the MACD is monitoring the changes and crossings of a 12- and 26-day moving average, which probably indicates the current trend, i.e. the rising or falling trend.

I continuously expanded the neural network, each additional set of relevant data entered increased the efficiency to a greater or lesser extent.

The Sign value works with an efficiency of almost 75%, which means that in 3 out of 4 days I can tell whether the next day's exchange rate will rise or fall. Which is an acceptable result.

The most effective results can be found with a total of 4 indicator numbers. This seems to be effective, as it will see relatively few articles as invalid. It will not take into account very few articles that should have been taken into account, so it will process the majority of relevant articles.

It is absolutely necessary to investigate how efficiently the neural network would have worked if we only worked with exchange rate-based data, so it can be found out whether and how much extra the neural network based on keyword research means. More specifically, is it worth using an ANN, or is the same or even better efficiency achieved if only a mathematical model is used.

Thus, examining the validity and strength of the keyword research, I ran the neural network without the keyword research, that is, I only examined the numerical data that could be extracted from the changes in the oil exchange rate. This is shown by the results of the neural network forecast based on exchange rate analysis in Table 1.

ACCEPTED MAXIMUM ABSOLUTE DEVIATION	SUCCESSFUL ESTIMATION
0,02	0,91%
0,04	1,52%
0,07	2,86%
0,1	3,65%
0,15	6,27%
0,25	10,28%
0,4	16,67%
0,8	34,06%
1,2	48,30%
1,6	61,25%
2	70,13%
2,5	77,37%
3	84,49%
3,5	88,69%
4	91,91%
SIGN	73,97%

1. Table Results of neural network prediction based on exchange rate analysis

ACCEPTED MAXIMUM ABSOLUTE DEVIATION	OIL MATH	SUCCESSFUL ESTIMATE DAY: t, IND: 4	ANN ind: 4 - OIL MATH difference		ANN Best Value	ANN Best Value - OIL MATH difference	
0,02	0,91%	0,48%	-0,43%	-47,25%	1,13%	0,22%	24,18%
0,04	1,52%	1,40%	-0,12%	-7,89%	1,92%	0,40%	26,32%
0,07	2,86%	3,32%	0,46%	16,08%	3,32%	0,46%	16,08%
0,1	3,65%	4,80%	1,15%	31,51%	4,80%	1,15%	31,51%
0,15	6,27%	7,03%	0,76%	12,12%	7,03%	0,76%	12,12%
0,25	10,28%	11,52%	1,24%	12,06%	11,52%	1,24%	12,06%
0,4	16,67%	18,33%	1,66%	9,96%	18,46%	1,79%	10,74%
0,8	34,06%	36,67%	2,61%	7,66%	36,67%	2,61%	7,66%
1,2	48,30%	50,50%	2,20%	4,55%	50,50%	2,20%	4,55%
1,6	61,25%	61,02%	-0,23%	-0,38%	61,59%	0,34%	0,56%
2	70,13%	70,01%	-0,12%	-0,17%	71,06%	0,93%	1,33%
2,5	77,37%	78,52%	1,15%	1,49%	79,18%	1,81%	2,34%
3	84,49%	84,59%	0,10%	0,12%	85,16%	0,67%	0,79%
3,5	88,69%	89,13%	0,44%	0,50%	90,22%	1,53%	1,73%
4	91,91%	91,84%	-0,07%	-0,08%	93,06%	1,15%	1,25%
SIGN	73,97%	72,94%	-1,03%	-1,39%	73,29%	-0,68%	-0,92%

2. Table ANN's most efficient results and based only on oil exchange rate

Table 2 shows the differences between the most effective results of ANN and those based only on the oil exchange rate, and the real differences. As usual, the tested efficiency is the accepted maximum deviation, i.e. what percentage of the predicted results is within the defined margin of error in the case of the actual results. The OIL MATH column is the accuracy of the predicted results from the time series analysis of the daily closes of the oil price only with a neural network. According to what was discussed in the previous chapters, I considered the analysis with 4 indicator numbers to be the most valid. In addition, the efficiency increases only slightly, and when using so many indicators, not too many articles are dropped. When using too high an indicator number, there is a risk that too many relevant articles will not be taken into account. The following two columns are the numerical deviation in favor of the ANN based on keyword research, and the deviation compared to the ANN base. After that, I selected the best result per line for each keyword research ANN, and then compared it with the results of the Neural Network based on the oil price using the previous method.

The analysis of time series synchronicity is nowadays increasingly prominent in scientific research, as it can be applied in many interdisciplinary fields, such as economics, meteorology, biology or even social sciences

(Brockwell & Davis, 2016). In this brief summary, we present four main methods used to analyze time series synchronicity: Pearson correlation, time-lagged cross-correlation, dynamic time warping, instantaneous phase synchrony, Wilmott's concordance index, and the R2 indicator.

According to the Pearson correlation results, a certain correlation can be detected. An r value of 0.115 means that there is a weak positive linear relationship between the two variables. In other words, the relationship between the two variables is not strong, but when the value of one variable increases, the value of the other variable will also increase somewhat. However, this correlation is not strong enough to draw firm conclusions about the relationship between the two variables, and other factors may influence the behavior of the variables.

Az időkésleltetett keresztkorreláció során (TLCC) eredménye 0,12. Azonban a 0,12-es érték továbbra is gyenge pozitív összefüggésre utal. Ez azt jelenti, hogy a két idősor közötti kapcsolat nem erős, de még mindig van némi összefüggés a két idősor között.

A Dinamikus idővetemítés (DTW) eredménye 3219,41. Ha a DTW eredménye 3219,41, az azt jelenti, hogy a két idősor között van némi eltérés, és a hasonlóságuk nem teljesen magas. A kapott érték arra utal, hogy az idősorok alakja, mintázata vagy viselkedése eltér egymástól, de nem olyan mértékben, mint amikor a DTW érték nagyon magas.

Az azonnali fázisszinkron (IPS) vizsgálatánál az eredmény azt mutatja, hogy a neurális háló hatékonyan előrejelzi az olajár változásokat, és magas fokú szinkronizációt ér el a valós adatokkal.

A Wilmott-féle egyezési index értéke 0,57, az arra utal, hogy az előrejelzési modell mérsékelt, de a gyakorlatban már viszonylag elfogadható pontosságú. Ez az érték azt jelzi, hogy a modell képes a valós adatok mintázatainak megragadására és viszonylag jól összehangolja az előrejelzéseket a tényleges megfigyelésekkel.

A 0,53-as R^2 értékű modell azt jelzi, hogy a modell már hatékonyan képes megragadni az adatok mögötti mintázatokat, és az előrejelzései többnyire megbízhatók. Az eredmények alapján a modell már használható sok gyakorlati alkalmazásban, különösen olyanokban, ahol a magas fokú pontosság nem elengedhetetlen.

4. Conclusions and suggestions

During the research, I discovered that keyword research with a neural network increases the efficiency in the field of oil price forecasting, more precisely, it is more effective than the analysis of oil price changes only with time series and MACD data. It is important to note that I also used a duplicate neural network during the exclusively exchange rate-based forecast, i.e. I predicted the expected value with an RNN, and on top of that, I expanded the expected value with other MACD data and thus it was included in a second ANN, which indicated based on these the result.

Examining the change in oil prices, it is clear that the price rose by approximately the same amount as it fell, according to the daily analysis. Typically, its daily price change was between 0-1%. In other words, the exchange rate is volatile, but rather rising. An increase was observed in 54% of the cases.

In this way, we can more accurately evaluate the results predicted by the neural network.

The neural network also put the most frequent occurrence between 0 and 1 percent, but it worked with drastically less standard deviation. The neural network expected an increase in 58.9% of the cases, which is very similar, but it only managed to hit the correct sign of the change in about three quarters of the cases. In fact, the most important point of the research is that it was able to determine the trend in a large number of cases, so we were able to determine with a high degree of certainty, from the user's point of view, whether we are in an increasing or decreasing oil price trend, as well as to determine trend reversals, which can form the basis of important economic decisions.

Despite all of this, we can make it clear that co-movement exists in connection with the results of Pearson correlation, TLCC, DTW and especially IPS. Wilmott's match index and R2 only confirm these. In other words, neural network predictions are suitable for the decision preparation stage!

5. New scientific results

In the course of the research, I investigated how much the speculative nature of the speculative stock market can be examined, that is, how much certain news and information influence it. In the course of the research, a clearly demonstrable connection is the analysis of the contents of the news and the subsequent exchange rate changes. In the present research, I examined it in relation to the price of oil, but we can assume that it is valid in other cases as well, and may be even more effective in the case of exchange rates more sensitive to speculation.

Market players who have non-inside information or who do not participate in decision-making are also able to calculate and infer trends, increases or decreases in the price changes of certain raw materials based on open information, thus including it in their economic decisions, which can reduce the risk independent of them.

There is an extremely high amount of noise in current news feeds and in the flow of information concerning economic actors, which distorts understanding. In other words, it is necessary to analyze the news and information in a special and very sensitive way, and to filter out the essential and truly newsworthy data.

We are able to make analytical decisions using machine learning methods, that is, to analyze the content of news written by real people and its reality. This can be made even more effective with much more sophisticated methods than the present method. It can be stated that the printed press can also be interpreted using automated methods. As a result, hundreds of articles can be processed in a very short time, which we would not be able to do as humans.

As a market participant, there is a decision-making function that is based on the analysis of Wall Street Journal articles and we can determine with sufficient certainty the oil exchange rate trend and the size of the price increase or decrease in the next period.

I can state, and we can confirm, that artificial neural networks are capable of efficient information processing, that is, they can also be used for fast and efficient analysis of Big Data.

I proved that a correlation can be shown between the examined journals and the exchange rate, that is, the speculative exchange rate movement with regard to the price of oil is clear. The extent of the relationship is still a question, as the accuracy of the forecast could still be made more precise. However, I determined the exchange rate based on the keywords found in the articles, so a clear connection can be demonstrated.

I can also state that by analyzing the content of the Wall Street Journal (WSJ) newspaper articles with an artificial neural network (ANN), the next day's oil price change can be determined with sufficient accuracy.

During the application of the optimization and speed-increasing algorithms, I proved that by summarizing, or compressing, newspaper articles, the prediction efficiency with artificial neural networks can be greatly increased.

It has also been proven that the prediction effectiveness of artificial neural networks can be greatly increased by analyzing the sentiment of newspaper articles.

In the course of the research, I investigated how the next day's oil price change can be determined with sufficient accuracy by examining the oil exchange rate with a feedback neural network (RNN), i.e. only by analyzing the historical movement of the exchange rate. He also won a certificate.

During the operation of the ANN, I checked that the efficiency can be significantly increased by greatly increasing the hidden layers and neurons of an artificial neural network, i.e. the network part. I can only accept this statement with reservations, since increasing it increased the efficiency exponentially for a while, and after that it did not mean any improvement, so there is an upper efficiency limit for a given task.

I analyzed, compared and verified that the analysis of WSJ newspaper articles with an artificial neural network is more effective than the analysis of the exchange rate with a feedback neural network, that is, speculation has a greater role in the movement of the exchange rate than fundamentals.

During the research, it was hypothesized that the prediction obtained with the help of an artificial neural network is more effective than the exchange rate change prediction with specific mathematical stock market models. I cannot fully accept this statement, since the forecast worked with sufficient accuracy, but there is not enough material available to prove that it is more effective than any existing mathematical model. Thus, it is conceivable that the statement is true.

6. Scientific publications related to the topic of the dissertation

Journal article

1. THALMEINER, G. – **BARTA, Á**. – GÁSPÁR, S. (2024): Extending the Hamiltonian operator with market information, ECONOMICS & WORKING CAPITAL
2. **BARTA, Á. –** MOLNÁR, M. (2023): Keyword research-based stock market oil price forecast validity test with neural network, GRADUS 10: 1 Paper: 2023.1.ECO.008, 9 p. (2023)
3. GÁSPÁR, S. - MUSINSZKI, Z. - HÁGEN, I.Z. - **BARTA, Á**. - BÁRCZI, J. - THALMEINER, G (2023): Developing a Controlling Model for Analyzing the Subjectivity of Enterprise Sustainability and Expert Group Judgments Using Fuzzy Triangular Membership Functions, SUSTAINABILITY 15: 10 Paper: 7981, 26 p. (2023)
4. GÁSPÁR, S. - PATAKI, L. - **BARTA, Á**. - THALMEINER, G - ZÉMAN, Z. (2023): Consumer Segmentation of Green financial Products Based on Sociodemographic Characteristics, JOURNAL OF RISK AND FINANCIAL MANAGEMENT 16: 2 Paper: 98 , 19 p.
5. GÁSPÁR, S. - THALMEINER, G. - **BARTA, Á**. - ZÉMAN, Z. (2022): Development of a Fuzzy Controlling Model to Measure the Leanness of Manufacturing Systems, ACTA POLYTECHNICA HUNGARICA 19: 4 pp. 189-207., 18 p.
6. THALMEINER, G. - GÁSPÁR, S. - **BARTA, Á**. - ZÉMAN, Z. (2022): Prediktív KPI-ok osztályozási módszerének ismerete, CONTROLLER INFO X: 3 pp. 2-6., 5 p. (2022)
7. **BARTA, Á**. - MOLNÁR, M. (2021): Crude oil stock market trend reversal forecast based on Wall Street Journal articles with keyword indication, GRADUS 8: 3 pp. 118-125., 8 p.
8. **BARTA, Á**. - MOLNÁR, M. (2021): Indication of organizational collusion by examining dynamic market indicators, GRADUS 8: 1 pp. 160-165., 6 p.
9. **BARTA, Á**. - MOLNÁR, M. (2021): Forecasting oil price based on online occurence volume, MODERN SCIENCE / MODERNI VEDA № 1 - 2021 pp. 5-11., 7 p.
10. THALMEINER, G. - GÁSPÁR, S. - **BARTA, Á**. - ZÉMAN, Z. (2021): Application of Fuzzy Logic to Evaluate the Economic Impact of COVID-19: Case Study of a Project-Oriented Travel Agency, SUSTAINABILITY 13: 17 Paper: 9602, 19 p.

11. MOLNÁR, M. - **BARTA, Á**. - BÁN, E. - VILLÁNYI, J. (2019): A novel approach to market anomaly sensing using neural networks, MECHANICAL ENGINEERING LETTERS: R AND D: RESEARCH AND DEVELOPMENT 19 pp. 103-112., 10 p

Conference publication or conference announcement

1. **BARTA, Á**. - MOLNÁR, M. - NAÁRNÉ TÓTH, ZS. (2023): Idősor szinkron validitás elemzés kulcsszókutatáson alapuló neurális háló által előrejelzett olajárfolyam változáson, MULTIDISZCIPLINÁRIS KIHÍVÁSOK SOKSZÍNŰ VÁLASZOK: 3 pp. 3-23., 21 p.
2. **BARTA, Á**. - NAÁRNÉ TÓTH, ZS. - MOLNÁR, M. (2023): Neurális háló alkalmazásával történő olajárfolyam előrejelzés eredményeinek idősor szinkron validitás elemzése / Time Series Synchronous Validity Analysis of the Results of Oil Exchange Forecasting using a Neural Network, In: Vágány, Judit; Fenyvesi, Éva (szerk.) Multidiszciplináris kihívások, sokszínű válaszok: 11. Tudományos Szimpózium: absztrakt füzet, Budapest, Magyarország: Budapesti Gazdasági Egyetem (BGE) (2023) pp. 52-53., 2 p.
3. MOLNÁR, M. - **BARTA, Á**. - VILLÁNYI, J. (2023): A piaci bizonytalanság becslése keresésintenzitás-alapú megközelítésben / Estimating Market Uncertainty using search volume indicators, In: Vágány, Judit; Fenyvesi, Éva (szerk.) Multidiszciplináris kihívások, sokszínű válaszok: 11. Tudományos Szimpózium: absztrakt füzet, Budapest, Magyarország: Budapesti Gazdasági Egyetem (BGE) (2023) pp. 54-55., 2 p.
4. **BARTA, Á**. - MOLNÁR, M. - NAÁRNÉ TÓTH, ZS. (2022): Investigation of the Online Press and Commodity Exchange Using Neural Networks, In: [S., n.] (szerk.) Strategic Management Proceedings: 27th International Scientific Conference and Decision Support Systems in Strategic Management, Subotica, Szerbia: University of Novi Sad, Faculty of Economics in Subotica (2022) 518 p. pp. 337-342., 6 p.
5. MOLNÁR, M. - **BARTA, Á**. (2021): Global market collusion analysis using artificial neural networks, In: Kinga, Pázmándi; Kinga, Pétervári (szerk.) Space – Time – Market– Economy, Budapest, Magyarország: HVG-ORAC (2021) 395 p. pp. 295-307., 13 p.
6. **BARTA, Á**. (2021): Tőzsdei trendfordulók előrejelzése pénzügyi folyóiratokban történő kulcsszókutatással, AGTECO 2021
7. **BARTA, Á**. - MOLNÁR, M. (2020): Piactorzító hatások elemzése a gépi tanulás eszközeivel, In: Horváth, Bálint; Kápolnai, Zsombor; Földi, Péter (szerk.) Közgazdász Doktoranduszok és Kutatók VI. Nemzetközi Téli

Konferenciája: Konferenciakötet, Budapest, Magyarország: Doktoranduszok Országos Szövetsége (DOSZ) (2020) 373 p. pp. 32-40., 9 p.

8. **BARTA, Á**. (2020): Kartellműködés kimutatás piaci dinamizmusok alapján neurális hálók alkalmazásával, AGTEDU 2020
9. **BARTA, Á**. (2020): Indication of cartel activity using neural networks, In: Horváth, Bálint; Földi, Péter; Kápolnai, Zsombor (szerk.) VI. Winter Conference of Economics PhD Students and Researchers: Book of Abstracts, Gödöllő, Magyarország: Szent István University, Doktoranduszok Országos Szövetsége, Közgazdaságtudományi Osztály (2020) 128 p. pp. 21-21., 1 p.

7. Bibliography

Arshad, S., Rizvi, S. A., Haroon, O., Mehmood, F., & Gong, Q. (2021). Are oil prices efficient? *Economic Modelling, Vol. 96.*, 362-370. doi:10.1016/j.econmod.2020.03.018

Blackard, J. A., & Dean, D. (1999). Comparative Accuracies of Artificial Neural Networks and Discriminant Analysis in Predicting Forest Cover Types From Cartographic Variables. *Computers and Electronics in Agriculture Vol 24 Issue 3*, 131-151.

Brockwell, P. J., & Davis, R. A. (2016). *Introduction to Time Series and Forecasting.* Springer. doi:10.1007/978-3-319-29854-2

Fattouh, B., Kilian, L., & Mahadeva, L. (2013). The role of speculation in oil markets: what have we learned so far. *Energy Journal Vol 34 Issue 3*, 7-33.

Hamilton, J. D. (2008). Understanding crude oil prices. *NBER Working Paper No. 14492*.

IEA. (2020). *Global Energy Review 2020*. Forrás: https://www.iea.org/reports/global-energy-review-2020

Jain, A. K., Mohiuddin, K. M., & Mao, J. (1996). Artificial neural networks: a. *Computer Vol. 29. Issue 3.*, 31-44.

Jiang, M., An, H., Jia, X., & Sun, X. (2017). The influence of global benchmark oil prices on the regional oil spot. *Energy, Vol. 118, 1.*, 742-752. doi:10.1016/j.energy.2016.10.104

Kaufmann, R. K. (2011). The role of market fundamentals and speculation in recent price changes for crude oil. *Energy Policy Vol 39*, 105-115.

Kilian, L. (2009). Not all oil price shocks are alike: disentangling demand and supply shocks in the crude oil market. *American Economic Review Vol 99 Issue 3*, 1053-1069.

Kutsurelis, J. (1998). *Forecasting Financial Markets Using Neural Networks: An Analysis Of Methods And Accuracy, Thesis.* Naval Postgraduate School Monterey: California.

Xin, Y. (1999). Evolving artificial neural networks. *Proceedings of the IEEE, vol. 87, no. 9.*, 1423-1447. doi:10.1109/5.784219