A Professional Journal Published by the International Federation of Technical Analysts

IFTA Journal

24

Inside This Issue

"Even if I knew that tomorrow the world would go to pieces, I would still plant my apple tree"

—Martin Luther

ISSN 2409-0271

Letter From the Editor

By Dr. Rolf Wetzer, CFTe, MFTA

Dear IFTA Colleagues and Friends,

This year, for the first time, the Journal is being delivered after the conference. The 2023 conference was held in Jakarta. It took place for the 36th time and was hosted by AATI, the association of our Indonesian colleagues. As always—since 1988, the presentations were informative and of high quality, and the time was characterised by hospitality and meeting new and old friends.

And yet, the year was much different than usual. The world seems to be recovering slowly from COVID. All national associations reported that direct meetings around the world are only slowly starting up again. We will regain our former strength, but the disease has left its mark on people's behaviour. In addition, the world is slipping into a phase of armed conflict the likes of which we have not seen for a long time. We will also overcome this together, but a veil of uncertainty is slowly settling over societies. The easy and carefree days seem further away at the moment.

It is therefore important that we do our part to promote friendship, cooperation, and togetherness in our sector. IFTA provides the platform for this, and many nations and countries contribute to it. This Journal is another example. The articles, and the work on the journal itself, come from people all over the world. Technical analysis unites!

As every year, I would like to thank everyone who has contributed to its creation. First and foremost, of course, to NAAIM and to Susan Truesdale. For years we have been allowed to publish articles from their programme. This year, it is the contribution of Andrew Thrasher, the winner of the NAAIM Founders Award.

We would also like to thank all the authors, the Linda Bernetich team, who year after year bring the Journal into a form that never ceases to amaze me. Many thanks for that. And finally, I would like to thank Regina Meani, who has been providing valuable contributions and book reviews for over a decade now.

I wish the entire IFTA family a blessed and happy 2024. May it be peaceful and bring many new friendships.

Best regards,
Dr. Rolf Wetzer, CFTe, MFTA

It is therefore important that we do our part to promote friendship, cooperation, and togetherness in our sector.

IFTA Journal

The IFTA Journal is published annually by the International Federation of Technical Analysts, 1300 Piccard Dr., Suite LL 14 Rockville, MD 20850 USA.

The Impact of Freely Accessible Large Language Models on Market Sentiment Monitoring

Andrea Bolognesi

University College of London

andrea.bolognesi11@gmail.com

Abstract

This article explores how freely accessible large language models capable of performing natural language processing on publicly available financial documents can be applied to technical analysis. I argue that, because these artificial intelligence models only have access to publicly available financial information such as news reports, news articles, investment forums, public discussions, and earnings transcripts, they can constitute a proxy to gauge retail investors sentiment's strength and direction (i.e., either positive or negative), thus providing valuable information to a technical analyst. This paper proposes possible applications in trend analysis (to anticipate trend continuations, reversals, or breakouts) as well as other areas of technical analysis such as testing support and resistance areas and trading bands. Lastly, this paper suggests the use of retail investors' sentiment analysis can be extended to other aspects of interest to a technical analysis such as event studies or uncovering retail information asymmetry.

Introduction

This paper takes as a starting point the release on the eighth of March 2023 of Google Bard, the first large language model (LLM) chatbot developed by Google AI. This comes at a point in time where the exponential growth in the availability and size of unstructured textual data sets (articles, reports, forum discussions) coincided with the absence of tools freely available to all technical analysts (regardless of their coding background) to analyze these large amounts of unstructured textual data. Additionally, in recent years, the unbiased analysis of large amounts of this type of data has gained recognition to understand market sentiment around certain securities or events, thus constituting a potential competitive advantage.

To uncover the possible applications of Google Bard in technical analysis, this article starts by introducing natural language processing (NLP) and its applications to the financial markets as well as how the release of LLMs capable of performing NLP on publicly available news reports, news articles, investment forums, public discussions, earnings transcripts can be used as a proxy to gage retail investors attention and sentiment around certain security and/or event. I argue for the inclusion of LLMs such as Google Bard in a technical analyst's toolkit by highlighting how it can be used as an additional data point when investigating price trends, chart patterns, and signals validity.

Understanding NLP Within Large Language Models

What Is Natural Language Processing

Natural language processing is a subfield of artificial intelligence and computational linguistics that focuses on the interaction between computers and human language. It involves the development and application of algorithms and models to enable computers to understand and manipulate natural language. In this particular case, I will focus on two of the tasks it can perform: Information extraction and sentiment analysis. Information extraction (IE) is the process of automatically extracting structured information from unstructured or semi-structured text data (such as news articles). By applying information extraction techniques to a news article, website, or report, we can extract key entities, events, facts & data, thereby gaining structured and meaningful information from the unstructured text. Similarly, sentiment analysis aims to analyze sentence structure, word choice and other elements to understand the tone of the article and identify positive or negative sentiments expressed towards specific entities or topics.

The interesting aspect of the results of information extraction and sentiment analysis is that it can inform us on possible developments in stock returns and trading volumes before they happen. For instance, sentiment studies on articles published on a specific company often have contemporaneous or short-term effects on stock prices and trading volumes (Kearney and Liu, 2014). Similarly, particular types of information published in public news is associated to abnormal returns following its release (Chan, 2003).

NLP Integration in Large Language Models

In the past, NLP algorithms were often inaccessible to most investors as they necessitated coding experience and technical expertise to implement and maintain effectively. Even commercially available options were inaccessible as subject to licensing restrictions imposed by large companies or research institutions. As a result, investors were largely excluded from accessing NLP models and integrating sentiment analysis into their toolkit.

However, a significant shift occurred with the release of public LLMs. The introduction of GPT-3 in November 2022 marked a significant turning point, providing broader access to NLP capabilities. More importantly, the launch of public LLMs with real-time data access, exemplified by Google Bard in March 2023, democratized NLP on press releases and publicly available documents to all investors. This accessibility to NLP, coupled with real-time data, opened new avenues for investors to gain insights and make informed decisions based on sophisticated language analysis.

Literature Review

Given the recent emergence of freely accessible LLMs, very limited research has been published on the specific impact of rendering NLP available to all investors and the overall impact of being able to freely run NLP algorithms on the entirety of publicly available financial data & news. Nonetheless, considering the significant attention cast on NLP in the past decade, researchers have taken interest in the new data provided by complex NLP analysis to relatively small batches of data over a limited time span.

Researchers have shown a strong interest in exploring the correlation between sentiment expressed in annual reports and future earnings/stock performance. Li (2006) conducted a study utilizing a linear regression sentiment-based trading strategy on annual company reports spanning nine years. The study concluded that an increase in risk sentiment in annual reports is associated with significantly lower future earnings.

While most of the previous research has focused on sentiment analysis of news articles, Tetlock (2007) found that high levels of media pessimism exert temporary downward pressure on market prices and induce high market trading volume. Tetlock et al. (2008) further refined the analysis and concluded that stock market prices respond to negative words with a small, one-day delay. This finding was supported by Carretta et al. (2011) and Engelberg et al. (2012), who designed a news sentiment-based trading strategy centered around short selling and negative news. Their study showed an impressive 180% return over a two-and-a-half-year sample period (2005–2007). Ferguson et al. (2012) reaffirmed and extended this research by demonstrating that the predictive relationship between media content and firms' returns is particularly significant for low-visibility firms with low market capitalization and high book-to-market ratio.

In contrast, other researchers have focused on sentiment analysis of information available only to professional investors, such as analysts' reports. These studies have indicated that sentiment analysis on analysts' reports seems to be less effective in establishing a relationship with stock market returns. Kothari et al. (2009) found that disclosures from analysts appear to be heavily discounted by the market.

Furthermore, researchers have applied NLP techniques to other financial documents, such as earnings press releases (Henry, 2008) or earnings conference calls and Q&A sessions (Doran et al., 2010). These studies have shown that both negative and positive sentiments in these documents have significant explanatory power for accompanying abnormal returns.

Relevance to Financial Markets

The current research seems to point out a strong predictive relationship between negative sentiment in news articles, earnings press releases & earnings calls, and abnormal firms' returns. Taking this into account, this paper will now investigate how the NLP data produced by freely accessible LLMs can help us as technical analysts.

This paper suggests these freely available LLMs measure the attention, intention, and consensus amongst retail investors, the roots of this interpretation are suggested by Ren (2023) which highlights Internet search as the "most important channel through which retail investors gather and process information". Freely available LLMs, such as Google Bard, differ from non-publicly available NLP models due to the fact they can only access and perform sentiment analysis and information extraction on publicly available information (i.e., News reports, news articles, investment forums, public discussions, earnings transcripts... but not analysts' reports). This means their results suffer from the same information disadvantages as retail investors are subject to compared to institutional investors who have access to more researched information via analyst reports (Cheng et al., 2016; Jung et al., 2018; Soltes, 2014). These new accessible LLMs, therefore, reflect the sentiment of the content and discussions available to retail investors and thus, following the literature mentioned, allow us to anticipate an increase in interest from these retail investors on the concerned stocks. Given this, Barber and Odean (2011) provides a helpful insight into the behavioral characteristics of these retail investors. They note the tendency of retail investors to be influenced by various behavioral biases (such as being influenced by past experience and exposure to media) that often lead them to contribute to a short-term overreaction in stock prices paving the way to an eventual reversal.

This paper will thus contend accessible LLMs can be potentially relevant to technical analysis by using Google Bard sentiment analysis data on publicly available news, articles, and retail investors' discussions as a proxy for retail investors' interest and thus a tool to identify the short-term overreaction characterizing retail investors. In this case, we refer to the situation where the Google Bard retail sentiment shows an opposite trend to the price movement (i.e., retail sentiment is falling while the price is moving in an uptrend and/or the retail sentiment is rising while the price is moving in a downtrend) as a price trend driven by non-retail investors (i.e., professional or institutional investors) who are likely to be more informed, less subject to the afore-mentioned behavioral biases and hence lead to a more enduring price trend less subject to reversals.

Table 1

Table 1: Retail investors sentiment evolution around Tesla from Jan-22 to Jan-23. Prompt used: " By performing sentiment analysis on News reports, news articles, investment forums, public discussions, earnings transcripts available to retail investors, create a table for retail investors sentiment around Tesla stock since January 2022 by month"

Month	Google Bard sentiment score*
Jan-22	75.00
Feb-22	60.00
Mar-22	55.00
Apr-22	40.00
May-22	35.00
Jun-22	20.00
Jul-22	15.00
Aug-22	10.00
Sep-22	-5.00
Oct-22	-10.00
Nov-22	-15.00
Dec-22	-20.00
Jan-23	-25.00

**Note: in this specific table, each month is given a value between -100 and 100 based on the sentiment analysis algorithm results with 100 being extremely positive, 0 neutral and -100 extremely negative*

Applications to Technical Analysis

For the sake of simplicity, this paper will base its analysis on the results of a single LLM, Google Bard. The usefulness of Google Bard will vary depending on the specific prompt used. In this study, prompts of the type "what is the sentiment from retail investors around y stock since [insert timeframe] month by month" or "by performing sentiment analysis on news reports, news articles, investment forums, public discussions, earnings transcripts available to retail investors, create a table for retail investors sentiment around y stock since [insert timeframe] by month" will be used to retrieve tangible numerical data from 0 to 100 to display the intensity and direction of retail investors sentiment; for the purpose of this study any value below 40 is considered as "negative sentiment" from retail investors.

Analysis of Trends

In technical analysis, the notion of a trend holds significant importance in understanding and interpreting price movements. A trend, in essence, refers to the prevailing direction of price movement over a certain period. Technical analysts employ a range of tools and methodologies to discern two fundamental aspects related to price trends: the direction in which the price is currently moving and the probability of its continuation.

The main way Google Bard can be applied to technical analysis is by helping us determine the probability of trend continuation; for this use, Google Bard can help us track the strength and direction of retail investors' attention by analyzing news, forum conversations and all other material published by and available to retail investors. This can then be compared to the direction and strength of the current trend to better understand which types of investors are driving the current trend and thus evaluate its sustainability. This paper will use the average directional indicator (ADX), outlined in Wilder (1978) to help determine the strength of a trend regardless of its direction (upward or downward). The ADX provides a value ranging from 0 to 100, with higher values (>40) indicating a stronger trend.

The analysis of ADX value in conjunction with the retail investors' attentions' strength values yields two possible scenarios with separate interpretations: On the one hand, if ADX rises or remains constant while Google Bard indicates retail investor's attention is on the rise then it can be concluded that it is very likely the current price trend is driven by retail investors and thus subject to a reversal as the current price trend is likely to be driven by short-term sentiment rather than long-term fundamentals. Inversely, if the ADX remains constant or rises while retail investor's attention declines it can be concluded that it is very likely the current trend is driven by institutional investors and is thus less subject to reversal and instead much more likely to be sustainable.

Given the ADX is a directionless indicator, one of the possible limitations of this analysis is the incorrect identification of the drivers behind the price trend. In other words, in the current state of the analysis, an increase in retail investors' interest doesn't necessarily mean they are pushing the price in the same direction as the trend. To reduce the likelihood of incorrect driver identification, the trend analysis can be refined with an analysis of the direction of the security's price trend in conjunction with Google Bard's retail investors' sentiment analysis. If the direction of the retail investor's attention coincides with the current price trend (i.e., upwards price trend accompanied by an increase in retail investor's attention characterized by positive sentiment) then it can be concluded with even more precision that the trend is indeed driven by retail investors and thus subject to a reversal as the current price trend is likely to be driven by short-term sentiment rather than long-term fundamentals.

This can be seen in the period highlighted in red in Figure 1, the strong upwards price trend from Nov-21 to Jan-22 is likely to be driven by retail investors as Google Bard sentiment analysis highlights a rise in positive sentiment around PFE. Given that this trend is likely to be driven by retail investors, a subsequent reversal is likely to happen as seen from Jan-22 to Mar-22.

Table 2

Table 2: Retail investors sentiment evolution around Pfizer from Feb-21 to May-22. Prompt used: "What is the sentiment from retail investors around Pfizer (PFE) stock since February 2021 by month"

Month	Google Bard sentiment score*
Feb-21	64.20
Mar-21	75.40
Apr-21	61.60
May-21	59.80
Jun-21	62.40
Jul-21	60.80
Aug-21	47.80
Sep-21	57.60
Oct-21	58.80
Nov-21	60.80
Dec-21	61.60
Jan-22	63.20
Feb-22	64.20
Mar-22	65.20
Apr-22	66.20
May-22	67.20

**Note: each month is given a value between 0 and 100 based on the sentiment analysis algorithm results with 100 being extremely positive, below 40 being considered negative sentiment*

Figure 1

Figure 1: demonstration of the reversal of a trend led by retail investors

Determining Entry Points

The identification of an optimal entry point is very important for a technical analyst, as it facilitates the initiation of a trade at a favorable price level, thereby maximizing potential gains and mitigating associated risks.

This paper suggests that an analysis of the relative strength index (RSI) in conjunction with Google Bard retail investors' sentiment can offer insight into whether the price of overbought or oversold security is fueled by retail investors or institutional investors to help ascertain whether the current price presents the characteristics of a good entry-point for either short or long positions. The RSI is a momentum oscillator that measures the speed and change of price movements. It provides a numerical value between 0 and 100, suggesting whether the stock is overbought (When RSI>70) and thus indicating that buying pressure has driven the price to an unsustainable level or conversely, that the stock is oversold (when RSI<30) potentially presenting a buying opportunity.

Once again, two possible scenarios can be observed, if the RSI indicator rises above 70 signaling a security is overbought and Google Bard indicates a rise in positive sentiment amongst retail investors, then it can be concluded that it is very likely that the buying pressure that has driven the price to an unsustainable level is coming mainly from retail investors thus signaling that the current stock price may be a good entry point for a short position given the likelihood of a subsequent normalization in price levels. Inversely, when the RSI indicator rises above 70 signaling a certain security is overbought and Google Bard indicates a decrease in positive sentiment amongst retail investors, it is likely that in this case, the buying pressure originates amongst institutional investors possessing better market knowledge suggesting the reason behind the price movement should be further investigated as the reversal may not be immediate.

An instance of this can be seen in Figure 2 at the orange marker line. As the RSI reaches above 70, while retail investors' sentiment declines, this leads to uncertain price fluctuations through the following month (Dec-21).

The inverse reasoning can be applied when the RSI indicator drops below 30.

Support and Resistance Areas

Support and resistance areas refer to specific price levels or range on a chart where buying or selling pressure is expected to be significant. Support levels are considered the price levels where demand is expected to be strong enough to prevent

Figure 2

Figure 2: demonstration of the identification of a trade entry-point requiring further investigation

further price declines, while resistance levels are price levels where selling pressure is anticipated to be strong enough to impede further price increases. The utilization of moving average charts offers a method for identifying support and resistance levels. Specifically, when the price approaches a moving average from below and subsequently rebounds, it implies the presence of potential support at that level. Conversely, when the price approaches a moving average from above and encounters resistance, it signifies a potential resistance area.

That said, different resistance support areas have different strengths, some support or resistance levels may be more robust and respected, while others may be more easily breached.

The strength of a support or resistance level is determined by multiple factors such as the number of times it has been tested, the volume of trading at that level, the psychology of the market... This paper aims to use Google Bard retail investors' sentiment analysis to identify the types of buyers and sellers behind a certain support or resistance level to attempt to determine its strength. Following the same assumptions as in the previous parts, if it could be concluded resistance (support) levels backed by retail investors might be considered less strong as they might have less researched underlying reason why the price of the security is not likely to rise above (drop below) that level or simply be fueled by news overreaction.

When analyzing moving averages in conjunction with Google Bard retail sentiment, If the price approaches a moving average from above (below) in a way suggesting the existence of a support/resistance area while Google Bard indicates positive retail (negative) investors sentiment, this may indicate the support (resistance) area is backed by institutional investors. As such it is likely to be tested and easily pierced as it is unlikely to be backed by strong fundamental analysis. if Google Bard indicates opposite retail investors' sentiment to the aforementioned, the observed support/resistance area is likely to be stronger and thus lead to a price reversal.

Analysis of Trading Bands

Google Bard can be utilized in conjunction with trading bands to enhance the reliability of their signals. Trading bands encompass bands that are plotted above and below the price line, aiming to offer a relative delineation of high and low prices. Common examples of trading bands include the Bollinger and Donchian bands. The application of trading bands facilitates the identification of transient price overreactions, typically discerned when the price surpasses the upper band or falls below the lower band. In tandem with trading bands, Google Bard can provide valuable insights regarding the underlying participants driving the observed price dynamics, thus contributing to the evaluation of the signal's validity.

If the price moves above the upper band (below the lower band) in conjunction with an increase in positive (negative) sentiment from retail investors, this confirms the signal is likely highlighting a short-term overreaction. Consequently, an impending reversal of the prevailing trend would be anticipated. This is shown in Figure 3 as the price moves below the lower Bollinger band accompanied by a sudden drop in retail investors sentiment; this price movement is very likely to have been a short-term overreaction fueled by retail investors because, as expected, price bounces back up immediately after.

However, if the signal arises with retail investors either exhibiting indeterminate sentiment or expressing sentiment counter to the price movement's direction, this suggests that the bands may have identified the initiation of a new trend rather than a transient overreaction. Consequently, further analysis of the fundamental drivers underpinning the recent price fluctuations is advised.

Figure 3

Figure 3: Demonstration of price moving below the lower band of Bollinger trading bands in conjunction to an increase in negative sentiment from retail investors anticipating the correction of the short-term overreaction

Extensions

Retail Information Asymmetry

Understanding the extent of information asymmetry is vital as it aids investors in assessing whether they possess unique insights conferring them with a competitive edge or if it aligns with the general knowledge already priced into the market. For technical analysts, understanding market developments can help in price analysis and trend recognition. This paper suggests Google Bard as a tool to check for information asymmetry and verify whether retail investors and/or the street has picked up on specific news or market developments. Given an event, new data point, or piece of information x, a prompt of the type "has x been cited in articles, reports, retail investors forums?" or "provide the number of news articles, reports, retail investors forums… citing x in the past" allows technical analysts to check for these information asymmetries before entering a trade in order to avoid sudden shifts in price, volume, or trading activity as x gets priced in by retail investors.

Additionally, through the same types of prompts, Google Bard allows technical analysts to check for the reactivity of the market by comparing the time from retail investors pick up certain pieces of information and the following increase in trading volume driven by these retail investors.

Event Studies and Market Sentiment

Google Bard can also be used as a source of information as part of an event study. By prompting it to perform sentiment analysis on all documents related to a certain event y, it is possible to gather important information in certain types of event studies as well as plot the evolution of sentiment following the event y. This paper suggests prompts of the type "by performing sentiment analysis on all articles, reports, retail investors forums released in the past [select timeframe], what has been the sentiment around an event y and how has it evolved?" where y can be a certain company's product launch, marketing campaign, management communication, legal issues, defective products… This can then be cross-checked with specialized industry reports to confront retail investor's reactions to certain event is in line with more knowledgeable and researched positions held by institutional investors.

For instance, if we were to enquire about retail investors sentiment around a company's product launch and it returned an extremely positive sentiment from retail investors, this could then be compared to specialized industry reports which may display different sentiments by offering a more complete picture. It is then in the best interest of the technical analyst to keep this information in mind when entering specific positions that may be affected by the possible shift in sentiment from retail investors towards the institutional consensus as they bridge the informational gap.

Rising (and Falling) Wedges

A rising wedge is a technical pattern that frequently appears in bear markets and serves as an indicator of a potential reversal. This pattern is characterized by the upward movement of prices, where pivot highs and lows gradually converge towards a singular point referred to as the apex on charts. Correctly identifying the rising wedge pattern is extremely important for technical analysts as it serves as an indicator of a potential downside movement as the price is expected to break out of the narrowing range and undergo a bearish reversal. Inversely, a falling wedge on the other hand suggests a potential upside. The rising wedge pattern (falling wedge pattern) is usually accompanied by signals that improve the accuracy of its recognition such as a decrease (increase) in trading volume as the pattern progresses. In this case, volume decrease shows sellers are consolidating their energy before they start pushing the price action lower towards the breakout.

This paper suggests Google Bard can be used to verify whether retail investors or institutional investors are behind the decrease in trading volume signaling the potential breakout of the pattern. In other words, knowing the investors behind the decrease in trading volume may play an important role in the correct identification of the pattern, if it is retail investors driving the decrease in trading volume this might lead to a failed pattern due to a lack of genuine market conviction or a temporary disruption in the anticipated price movement. This paper thus suggests a prompt of the type "by analyzing retail investor sentiment, retail-specific news, retail investors forum posts, market participation analysis and trading activity data, are retail investors behind the notable decrease in trading volume in the past [insert timeframe]?" in order to ascertain the type of groups of investors behind the accompanied trading volume changes.

Conclusion

In this paper, we have explored various ways to integrate Google Bard in a technical analyst tool kit to complement and improve trend analysis, price analysis and other chart pattern recognition. Nevertheless, this paper recommends technical analysts to exercise a high degree of caution when using Google Bard as a proxy for retail investors sentiment as attention due to the lack of transparency behind Bard's NLP algorithms. Technical analysts are thus encouraged to use Google Bard's applications listed throughout this paper as an additional data point in the context of a more in-depth analysis.

References

Barber, B. M., & Odean, T. (2011). The behavior of individual investors. Handbook of the Economics of Finance, 2, 1533-1570.

Carretta, A., Farina, V., Graziano, E. A., & Reale, M. (2011). Does investor attention influence stock market activity? The case of spin-off deals. SSRN. Retrieved from http://ssrn.com/abstract=1930861

Cheng, Q., Du, F., Wang, X., & Wang, Y. (2016). Seeing is believing: Analysts' corporate site visits. Review of Accounting Studies, 21(4), 1245-1286. doi:10.1007/s11142-016-9368-9

Colm Kearney, & Sha Liu. (2014). Textual sentiment in finance: A survey of methods and models. International Review of Financial Analysis, 33, 171-185. doi:10.1016/j.irfa.2014.02.006

Doran, J. S., Peterson, D. R., & Price, S. M. (2010). Earnings conference call content and stock price: The case of REITs. Journal of Real Estate Finance and Economics, 45(2), 402-434.

Durnev, A., & Mangen, C. (2011). The real effects of disclosure tone: Evidence from restatements. SSRN. Retrieved from http://ssrn.com/abstract=1650003

Engelberg, J., Reed, A. V., & Ringgenberg, M. C. (2012). How are shorts informed? Short sellers, news, and information processing. Journal of Financial Economics, 105(2), 260-278.

Henry, E. (2008). Are investors influenced by how earnings press releases are written? The Journal of Business Communication, 45(4), 363-407.

Jung, M. J., Wong, M. H. F., & Zhang, X. F. (2018). Buy-side analysts and earnings conference calls. Journal of Accounting Research, 56(3), 913-952. doi:10.1111/1475-679X.12180

Kothari, S. P., Li, X., & Short, J. E. (2009). The effect of disclosures by management, analysts, and business press on cost of capital, return volatility, and analyst forecasts: A study using content analysis. The Accounting Review, 84(5), 1639-1670.

Li, F. (2006). Do stock market investors understand the risk sentiment of corporate annual reports? SSRN. Retrieved from http://ssrn.com/abstract=898181

N.J. Ferguson, J. Guo, H.Y.T. Lam, & D. Philip. Media content and stock returns: The predictive power of press.

N.J. Ferguson J. Guo H.Y.T. Lam D. Philip Media content and stock returns: The predictive power of press

Ren, W., & Zhou, J. (2019). Stock price reaction to CEO messages on social media: Evidence from Twitter. Journal of Accounting and Economics, 68(2-3), 101253. doi:10.1016/j.jacceco.2019.06.001

Soltes, E. (2014). Private interaction between firm management and sell-side analysts. Journal of Accounting Research, 52(1), 245-272. doi:10.1111/1475679X.12037

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. Journal of Finance, 62, 1139-1168.

Tetlock, P. C., Saar-Tsechansky, M., & Macskassy, S. (2008). More than words: Quantifying language to measure firms' fundamentals. Journal of Finance, 63, 1437-1467.

Wentao Ren. (2023). Retail investors' accessibility to the internet and firm-specific information flows: Evidence from Google's withdrawal. International Review of Economics & Finance, 86, 402-424. doi:10.1016/j.iref.2023.03.019

Wesley S. Chan. (2003). Stock price reaction to news and no-news: drift and reversal after headlines. Journal of Financial Economics, 70(2), 223-260. doi:10.1016/S0304-405X(03)00146-6.

Wilder, J. W. (1978). New concepts in technical trading systems. Greensboro, N.C.: Trend Research.

Endnotes

Retail investors Sentiment data courtesy of Google Bard (https://bard.google.com/?hl=en) and Stock charts courtesy of StockCharts (www.stockcharts.com).

Multiple Regression Analysis A Forecasting Technique for the Financial Markets

Davide Pandini, PhD, MFTA, CFTe, CMT, CSTA

Milano, Italy

+39 348 397 3247
davide.pandini@hotmail.com

"… There are three types of lies: lies, damn lies and statistics …"
Benjamin Disraeli (1804-1881), Prime Minster of Great Britain from 1874 to 1880

Abstract

In financial markets forecasting is an essential aspect of making investment and trading decisions, managing risk, or optimizing portfolios. Accurate predictions of stock prices and returns, bond yields, and other financial indicators can provide valuable insights into market trends and help investors and traders identify potentially profitable opportunities. However, it is well known that forecasting in financial markets is a very challenging task, as these markets are complex, dynamic, and subject to a wide range of economic, political, and social factors. An investor might simply want to find the strongest stock in a given sector or financial index, while a trader may want to forecast future price movements of a security based on a variety of factors, such as economic indicators, market sentiment, or company-specific information. Or, she/he might need to identify the factors influencing the risk of a particular investment to assess the impact of interest rates, currency fluctuations, or commodity prices on the value of an asset.

Multiple regression is a statistical technique commonly used in finance to predict the future values of financial assets. It involves analyzing the relationship between one dependent variable and several independent variables, and then using this information to make predictions. Multiple regression models are useful because they consider the effects of several factors simultaneously and can provide more accurate forecasts than simple linear models. By including a range of economic and financial indicators in the regression model, it is possible to identify the factors that are most likely to impact the dependent variable and use this insight to make informed investment and trading decisions.

Many of these concepts are also important in technical analysis and should be understood by all technicians to increase the probability of successful trades, since multiple regression can be a useful tool to estimate the underlying price direction for a given security.

In this work, a rigorous statistical assessment of multiple regression is presented and a regression model to forecast the behavior of the stock market based on financial and economic indicators is developed and thoroughly tested by verifying the statistical assumptions. Furthermore, statistical accuracy metrics have been used to compare the regression model against other forecasting techniques.

1. Introduction

The financial markets and the price of the asset classes are complex systems that are influenced by a several factors, ranging from economic and financial indicators to geopolitical global events. Hence, predicting the behavior of financial markets and asset classes is a challenging task [1], but one that is of utmost importance to investors and traders.

Regression techniques are a widely used statistical analysis tool and have been used extensively in financial analysis and investment management for several years, with many researchers finding evidence of their effectiveness in predicting the behavior of asset prices and returns in the financial markets [2]. In [3] a trading approach based on linear regression was presented. In this work, linear regression was considered as a reliable method to represent the underlying trend of a given security.

Multiple regression involves modeling the relationship between a dependent variable (such as stock price or bond yield) and several independent variables (such as interest rates, economic indicators, or company-specific data). By estimating the coefficients of these independent variables, multiple regression allows the analysts to identify the most influential factors driving the movement of asset prices, and to make predictions about future performance. This approach has proved particularly useful in analyzing composite relationships between financial variables and in identifying key drivers of asset returns. For example, multiple regression analysis has been used to study the impact of macroeconomic factors such as inflation and interest rates on asset prices, as well as to analyze the performance of individual securities and portfolios. One key advantage of multiple regression is its ability to incorporate many elements into the model, allowing for a more comprehensive understanding of the variables influencing the asset prices and the financial markets.

Despite its widespread use, however, multiple regression analysis is not without some limitations. One challenge is the potential for overfitting, where the model becomes too complex and begins to fit noise rather than signal. Another challenge is the dynamic nature of financial markets, where the factors that influence the asset values may change over time, making it difficult to build a model that can accurately predict future market and price movements. Furthermore, the technique assumes that the relationship between the dependent variable and the independent variables is linear (or quite close to linear), and that data are normally distributed. In practice, these assumptions may not hold in the financial markets, and analysts and traders must take care to avoid the potential pitfalls of using multiple regression inappropriately. In order to avoid these drawbacks, a rigorous assessment of all the statistical assumptions is mandatory.

This work presents an overview and assessment of multiple regression forecasting for the financial markets, in particular for the stock market represented by the S&P 500 index, based on some interest rates, prices, and economic indicators. Moreover, it includes a thorough discussion of its applications, benefits, and limitations.

The paper begins by describing the historical data and lookback period analyzed to develop the stock market multiple regression model. Next, it continues by introducing the basic principles of multiple regression analysis and explaining how it can be applied in the financial markets. Then, it discusses some of the key issues and challenges associated with multiple regression, such as data quality, model specification, and interpretation of results. Overall, this work aims to provide a comprehensive introduction to multiple regression, highlighting its strengths and limitations, and providing guidelines on its appropriate use in financial analysis and forecasting.

This paper is organized as follows: In Section 2, the stock market dependent (or explained) variable and the independent (or explanatory) variables are defined through their rates, prices, and macroeconomic areas. Next, descriptive statistics on the historical data are calculated. Section 3 discusses multiple regression statistics, analysis of variance or ANOVA, and coefficients analysis.

In Section 4, multiple regression correct specification is evaluated through the individual coefficients statistical significance and corrected through backward elimination stepwise regression. After that, multiple regression independent variables linear independence is evaluated by means of multicollinearity test and corrected with a specification re-evaluation. Next, the multiple regression correct functional form is evaluated through a linearity test and corrected through non-linear quadratic variable transformations. Then, multiple regression residuals autocorrelation is evaluated and adjusted by including lagged dependent variable data in the original regression. Later, multiple regression residuals homoscedasticity is evaluated and improved through heteroscedasticity consistent standard errors estimation. Finally, multiple regression residuals normality is assessed. It is worth noting that to maximize the effectiveness and robustness of the multiple regression analysis forecasting, a rigorous assessment of all the assumptions must be performed, to understand and evaluate the capabilities and limitations of this statistical analysis method.

In Section 5, multiple regression forecasting accuracy is evaluated by dividing the historical data into training and testing ranges. The training range is used for the best model fitting by going through the steps described in the previous Section. Next, the best fitting model coefficient values are used to forecast the stock market behavior within the testing range. Finally, the forecasted values accuracy in the testing range is evaluated by comparing it with random walk and arithmetic mean benchmarks using statistical accuracy metrics. Section 6 introduces an alternative time series forecasting technique like Exponential Triple Smoothing. This method consists of fitting a curve to the historical data and extrapolating it into the future, assuming that the future values will follow the same pattern as the past values, with some degree of randomness. Finally, Section 7 presents some conclusive remarks.

2. Data Analysis and Descriptive Statistics

To create a robust stock market (S&P 500) model that explains the changes in S&P 500 price using economic data, it is necessary to select the most meaningful data. Lincoln [4] suggested to use both corporate and Treasury bonds, price indexes, and other economic indicators.

The historical data used in this work is based on more than 26 years of arithmetic monthly returns of the S&P 500 State Street Exchange Traded Fund (ETF ticker symbol: SPY) adjusted close prices, effective monthly yield of 1 Year U.S. Treasury Bill Yield, 10 Years U.S. Treasury Note Yield, U.S. High Yield Corporate Bond Index Yield, arithmetic monthly inflation or deflation of U.S. Consumer Price Index, U.S. Producer Price Index, arithmetic monthly returns of West Texas Intermediate Oil prices, arithmetic monthly changes of U.S. Industrial Production Index value, and U.S. Personal Consumption Expenditures amount, from January 1st, 1997, to February 1st, 2023 corresponding to 314 observations for each time series.

The historical data were downloaded from *Yahoo! Finance* [8] and from the *Federal Reserve Economic Data* (FRED) database of economic time series, maintained by the Federal Reserve Bank of St. Louis [9], and are summarized in Table 1.

Table 1. Asset classes and data sources

	Asset Class	Symbol	Data source
	S&P 500 index (SPY ETF)	stocks	Yahoo! Finance
Rates	1 Year U.S. Treasury Bill Yield	t1y	FRED
	10 Years U.S. Treasury Note Yield	t10y	FRED
	U.S. High Yield Corporate Bond Index Yield	hyield	FRED
Prices	Consumer Price Index	cpi	FRED
	Producer Price Index	ppi	FRED
	West Texas Intermediate Oil Prices	oil	FRED
Macro	Industrial Production Index	indpro	FRED
	Personal Consumption Expenditures	pce	FRED

For multiple regression forecasting, monthly data is divided into a training range for best model fitting, and a testing range for evaluating best fitting model forecasting accuracy. The training range goes from January 1st,1997, to December 31st, 2012, including the first 16 years of data corresponding to 192 observations for each time series. The testing range goes from January 1st, 2013, to February 1st, 2023, with more than 10 years of data corresponding to 122 observations for each time series.

The S&P 500 index (i.e., *stocks*) is the dependent (or explained variable) and consists of the SPY ETF adjusted close prices arithmetic monthly returns. The eight independent (or explanatory) variables consist of rates, prices, and macroeconomic indicators:

- Rates
  - 1 Year Treasury (*t1y*) consists of effective monthly yield of 1 Year U.S. Treasury Bill bond equivalent yield or quoted annual yield;
  - 10 Year Treasury (*t10y*) consists of effective monthly yield of 10 Year U.S. Treasury Note bond equivalent yield or quoted annual yield;
  - High Yield (*hyield*) consists of effective monthly yield of U.S. High Yield Corporate Bond Index equivalent yield or quoted annual yield;
- Prices
  - Consumer Prices (*cpi*) consists of U.S. Bureau of Labor Statistics Consumer Price Index arithmetic monthly inflation or deflation;
  - Producer Prices (*ppi*) consists of U.S. Bureau of Labor Statistics Producer Price Index arithmetic monthly inflation or deflation;
  - Oil Prices (*oil*) consists of U.S. Energy Information Administration West Texas Intermediate oil prices arithmetic monthly returns;
- Macro
  - Industrial Production (*indpro*) consists of U.S. Federal Reserve Economic Data Industrial Production Index value arithmetic monthly change;
  - Consumption Expenditure (*pce*) consists of U.S. Bureau of Economic Analysis Personal Consumption Expenditure amount arithmetic monthly change.

All the variables are expressed in terms of the arithmetic monthly returns:

$$\Delta x_t = \frac{x_t}{x_{t-1}} - 1.$$

2.1 Descriptive Statistics

Descriptive statistics were considered to perform an early analysis of the dataset. The variables used for this preliminary assessment are:

- Arithmetic mean, consisting of the data average:

$$\mu = \frac{1}{n}\sum_{t=1}^{n} x_t$$

- Standard deviation, consisting of the data amount of dispersion or variation from its arithmetic mean:

$$\sigma = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(x_t - \mu)^2}$$

- Skewness, consisting of the data probability distribution asymmetry from its arithmetic mean:
  - Zero skew indicates symmetric data probability distribution;
  - Positive skew indicates data probability distribution with asymmetric tail extending towards positive values;
  - Negative skew indicates data probability distribution with asymmetric tail extending towards negative values;

$$s = \frac{\frac{1}{n}\sum_{t=1}^{n}(x_t - \mu)^3}{\sigma^3}$$

- Excess kurtosis, consisting of the data probability distribution peakedness or flatness compared with a normal probability distribution:
  - Zero excess kurtosis indicates data probability distribution peakedness or flatness similar to a normal probability distribution;
  - Positive excess kurtosis indicates data probability distribution more peaked than a normal probability distribution;
  - Negative excess kurtosis indicates data probability distribution flatter than a normal probability distribution;

$$ek = \frac{\frac{1}{n}\sum_{t=1}^{n}(x_t - \mu)^4}{\sigma^4} - 3.$$

The descriptive statistics results on the complete historical data from 1997 to 2023 of the asset classes conveyed in Table 1 are summarized in Table 2.

Table 2. Descriptive statistics results

		rates			prices			macro	
	stocks	t1y	t10y	hyield	cpi	ppi	oil	indpro	pce
mean	0.64%	1.24%	0.14%	0.21%	0.20%	0.22%	0.87%	0.09%	0.39%
standard deviation	4.53%	16.08%	7.77%	7.20%	0.30%	0.96%	10.12%	1.13%	1.14%
skewness	-0.56	1.37	-0.17	1.82	-0.87	-0.68	0.55	-5.18	-3.47
excess kurtosis	0.79	9.02	4.19	10.17	7.78	3.06	9.53	67.92	67.37

All the explanatory variables have a probability distribution more peaked than a normal distribution, while the distribution asymmetry is either positive or negative. Based on these results, the probability distributions of the independent variables deviate from the typical shape of a normal probability distribution.

3.Multiple Regression Analysis

Multiple regression analysis is a statistical method used to analyze the relationship between one dependent variable (i.e., explained variable or predicted variable) and a set of multiple independent variables (i.e., explanatory variables or predictors). It consists of fitting the best model within the training data range, then using the values of the best fitting model coefficients to forecast the dependent variable value in the testing data range and evaluating its prediction accuracy. In other words, multiple regression analysis allows us to predict how the changes in one or more independent variables affect the dependent variable. The output of multiple regression analysis consists of regression statistics, analysis of variance (i.e., ANOVA) and regression coefficients analysis.

ANOVA is a statistical technique used to test the overall significance of a multiple regression model. ANOVA evaluates whether there is a significant relationship between the dependent variable and the set of independent variables in the model. The ANOVA table shows the variation in the dependent variable that is explained by the independent variables, as well as the dependent variable variation that is not explained by the independent variables (i.e., the residual variation). ANOVA generates the *F-statistic* and the associated p-value (i.e., *significance F*), which are used to test the null hypothesis that all the regression coefficients in the model are equal to zero, thus indicating that there is no significant relationship between the dependent variable and the independent variables. If the p-value of the F-statistic is less than the significance level (e.g., 0.05 at 95% of confidence level), then we reject the null hypothesis and conclude that at least one of the independent variables in the model is significantly related to the dependent variable. In other words, the regression model as a whole is significant. It is important to note that ANOVA only tests the overall significance of the model and does not provide information about the statistical significance of the individual independent variables. To determine which independent variables are significant, we need to examine the individual t-statistic for each regression coefficient. In summary, ANOVA is a useful tool for assessing the overall significance of a multiple regression model and can provide important information about the relationship between the dependent variable and the set of independent variables.

In this work we analyze three areas: regression statistics, ANOVA, and regression coefficients. Although the formulas of the statistical parameters reported in this Section are well known and can be found in any statistical analysis textbook [5][6][7], nevertheless, they are summarized below to make the reading of this paper self-contained.

3.1 Regression Statistics

The regression statistics consist of coefficient of determination, adjusted coefficient of determination, and regression standard error:

- Coefficient of determination or R^2 (R Square) consists of the percentage of the variance from the dependent variable explained by its relationship with the independent variables:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

- Adjusted coefficient of determination or *adj* R^2 (Adjusted R Square) consists of the percentage of the variance from the dependent variable explained by its relationship with the independent variables adjusted by the degrees of freedom:

$$adj\ R^2 = 1 - \frac{SS_{res}}{SS_{tot}} * \frac{df_{tot}}{df_{res}}$$

- Regression standard error consists of the standard deviation estimation of the regression residuals or forecasting errors and is a measure of the accuracy of the regression model's predictions. It is also known as the standard error of the estimate, and it represents the average distance that the observed values fall from the regression line:

$$SE_{reg} = \sqrt{\frac{SS_{res}}{df_{res}}}.$$

3.2 Regression Statistics Results

The regression statistics results obtained from the historical data set reported in Table 1 for the entire backtesting period (monthly data from January 1997 to February 2023) are summarized in Table 3.

Table 3. Regression statistics

Regression Statistics	
Multiple R	0.68
R Square	0.47
Adjusted R Square	0.45
Standard Error	0.03
Observations	314

The multiple R parameter is a measure of the strength and direction of the linear relationship between the dependent variable and the set of independent variables included in the regression model and is expressed by the relationship: $multiple\ R = \sqrt{R^2}$.

The R Square value of 0.47 reported in Table 3 means that the independent variables in the regression model explain 47% of the variation in the dependent variable. In other words, the R^2 value of 0.47 indicates that the independent variables are able to explain a moderate amount of the variability in the dependent variable, but that there is still a significant amount of variation in the dependent variable that is not explained by the independent variables in the model.

3.3 Analysis of Variance (ANOVA)

The analysis of variance, or ANOVA, consists of the following parameters: regression degrees of freedom, residuals degrees of freedom, regression sum of squares, residuals sum of squares, total sum of squares, regression mean square error, residuals mean square error, regression F-statistic, and regression p-value (i.e., significance F):

- Regression degrees of freedom consists of the number of independent or explanatory variables, and is used in ANOVA to compute the F-statistic to test the overall significance of the regression model;

$$df_{reg} = v$$

- Residuals degrees of freedom consists of the number of observations minus the number of regression coefficients including the constant or intercept term, and it represents the number of degrees of freedom associated with the unexplained error in the regression model. It is used in the calculation of the regression standard error;

$$df_{res} = n - df_{reg} - 1$$

- Total degrees of freedom consist of the number of observations minus the constant or intercept term, and represent the total variation in the dependent variable explained by the regression model;

$$df_{tot} = df_{reg} + df_{res} = n - 1$$

- Regression sum of squares consist of the explained sum of squares, and is used in regression analysis to quantify the amount of variation in the dependent variable explained by the independent variables;

$$SS_{reg} = \sum_{t=1}^{n} (\hat{y}_t - \mu)^2$$

- Residuals sum of squares consists of the unexplained sum of squares, and quantifies the amount of unexplained variation in the dependent variable;

$$SS_{res} = \sum_{t=1}^{n} (y_t - \hat{y}_t)^2 = \sum_{t=1}^{n} e_t^2$$

- Total sum of squares consists of the explained and unexplained sum of squares, and quantifies the total variation in the dependent variable;

$$SS_{tot} = SS_{reg} + SS_{res} = \sum_{t=1}^{n} (y_t - \mu)^2$$

- Regression mean square error consists of determining the explained regression variance, and measures the amount of variation in the dependent variable explained by the independent variables;

$$MS_{reg} = \frac{SS_{reg}}{df_{reg}}$$

- Regression mean square error consists of determining the explained regression variance, and measures the amount of variation in the dependent variable explained by the independent variables;

$$MS_{res} = \frac{SS_{res}}{df_{res}}$$

- Regression F-statistic consists of determining the relationship between explained and unexplained regression variance. The F-statistic is a single value calculated from the ratio of two variances, and it is used in statistical tests such as ANOVA to test the null hypothesis that all the regression coefficients (except the intercept) are equal to zero, against the alternative hypothesis that at least one of the coefficients is different from zero. The regression F-statistic is computed by dividing the regression mean square error by the residual mean square error. If the regression F-statistic is large enough, it indicates that the regression model is a better fit for the data than the null model, which assumes that all the coefficients are zero. The significance of the F-statistic can be assessed using an F distribution with df_{reg} degrees of freedom for the numerator (where df_{reg} is the number of independent variables) and $n\text{-}df_{reg}\text{-}1$ degrees of freedom for the denominator (where n is the number of observations);

$$F_{reg} = \frac{MS_{reg}}{MS_{res}}$$

- Regression p-value (F-statistic p-value or significance F) consists of determining whether regression coefficients are jointly statistically significant. This formula is used in regression analysis to assess the significance of the regression model. It calculates the probability of observing an F-statistic as large as the one computed from the data, under the assumption that the null hypothesis (all regression coefficients are zero) is true:

$$p\text{–}value = Pr\left(F\left(df_{reg}, n - df_{reg} - 1\right) > F_{reg}\right),$$

3.4 ANOVA Results

The ANOVA results on the historical dataset are reported in Table 4.

Table 4. ANOVA results

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	0.2992	0.0374	33.1848	2.14E-37
Residual	305	0.3437	0.0011		
Total	313	0.6430			

The multiple regression p-value (i.e., significance F) is used to determine whether the coefficients of the independent variables in the model are jointly statistically significant, that is, whether they have a statistically significant effect on the dependent variable. A p-value less than 0.05 (at the 95% level of statistical confidence) indicates that the coefficients are statistically significant. This means that there is strong evidence suggesting that the independent variables are associated with the dependent variable and that the relationship is unlikely to have occurred by chance.

3.5 Regression Coefficients Analysis

The regression coefficients analysis consists of the following parameters: regression coefficients, standard error, t-statistic, and p-value for each coefficient:

- Regression coefficients are the values β $\beta_1, \cdots, \beta_v$ which minimize the residuals sum of squares SS_{res}: $\sum_{t=1}^{n} e_t^{\,2} = \sum_{t=1}^{n} (\hat{y}_t - y_t)^2$. The SS_{res} is a measure of the difference between the observed values of the dependent variable and the predicted values (i.e., fitted values) from a regression model. It is used to evaluate the goodness of fit of a regression model, where a smaller SS_{res} indicates a better fit;

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_v x_{v,t} + e_t$$

- Regression coefficients standard error consists of the standard deviation estimation of regression residuals or forecasting errors for each coefficient;

$$SE_{coeff} = \sqrt{MS_{res,coeff}}$$

- Regression coefficients t-statistic consists of determining the relationship between coefficient value and standard error;

$$t_{coeff} = \frac{\beta_{coeff}}{SE_{coeff}}$$

- Regression coefficients p-value consists of determining whether regression coefficients are individually statistically significant.

$$t_{pval} = t_{dist,2tails}\left(abs\left(t_{coeff}\right), df_{res}\right)$$

3.6 Regression Coefficient Results

Table 5. Multiple regression coefficients

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0093	0.0026	3.5548	0.0004
t1y	0.0087	0.0144	0.6038	0.5464
t10y	0.0649	0.0308	2.1088	0.0358
hyield	-0.4374	0.0280	-15.6041	0.0000
cpi	-0.5201	1.0400	-0.5001	0.6174
ppi	0.2327	0.3621	0.6426	0.5210
oil	-0.0146	0.0286	-0.5106	0.6100
indpro	-0.0472	0.2514	-0.1878	0.8512
pce	-0.3815	0.2839	-1.3437	0.1800

The p-values reported in Table 5 are related to the statistical significance of each regression coefficient individually (while the p-value reported in Table 4 relates to the joint statistical significance of all the regression coefficients). From the individual p-values it is possible to assess whether there are some explanatory variables that are not statistically significant (with the p-value > 0.05) at the 95% confidence level. Therefore, it is important to differentiate between the individual statistical significance of each coefficient and the joint statistical significance.

4. Multiple Regression Assumptions

The assumptions for multiple regression consist of independent variables correct specification, independent variables with no linear dependence, regression correct functional form, residuals with no autocorrelation, residuals homoscedasticity, and residuals normality. All these assumptions must be rigorously tested in order to have a good assessment of the regression model for an accurate interpretation of the results.

The correct specification refers to the appropriate selection of the explanatory variables (or independent variables) that will be included in the regression model. The key aspect of correct specification is to include in the model only those variables that have a statistically significant impact on the dependent variable, while excluding those that do not significantly contribute to the model statistical accuracy. Including statistically irrelevant explanatory variables in a multiple regression model can lead to several drawbacks, such as:

- *Reduced precision of coefficient estimates:* statistically insignificant variables may reduce the precision of coefficient estimates for the significant variables. This happens because including irrelevant variables introduces additional variability into the model, which can obscure the relationships between the explanatory variables and the dependent variable;
- *Increased risk of overfitting*: Irrelevant variables can increase the risk of overfitting, which occurs when a model fits the noise in the data rather than the underlying relationships. Overfitting leads to poor generalization performance when the model is applied to new data;
- *Increased complexity of the model:* Irrelevant variables can increase the complexity of the model without adding any explanatory power. This can make the model more difficult to interpret;
- *Potential violation of assumptions:* Irrelevant variables can potentially violate the assumptions of the model, such as normality, linearity, and homoscedasticity. These violations can lead to biased and unreliable coefficient estimates and statistical inference.

It is important to consider the theoretical and empirical evidence when selecting the independent variables for inclusion in the model. Additionally, it is essential to evaluate the robustness of the results by performing sensitivity analyses, such as testing the stability of the coefficients when adding or removing variables from the model. Overall, the correct specification of a multiple regression model is critical for obtaining valid and reliable results that can be used for informed trading and investment decisions.

Correct specification is done through a variable selection procedure, which consists of determining the individually and jointly statistically significant regression coefficients through backward elimination or forward selection stepwise regressions:

- *Backward elimination stepwise regression* consists of removing individually non-statistically significant variables one step at a time until all variables are individually and jointly statistically significant;
- *Forward selection stepwise regression* consists of adding individually statistically significant variables one step at a time while all variables are individually and jointly statistically significant.

4.1 Correct Specification: Backward Elimination Stepwise Regression

In this work we consider the *backward elimination stepwise regression*. This method starts with a model that includes all the independent variables, and iteratively removes the least significant variable until only the most significant variables remain in the model. At each step, the statistical significance of each independent variable is evaluated, and the variable with the highest p-value is removed if its p-value exceeds a predefined threshold (i.e., 0.05). This process is repeated until all remaining independent variables have p-values below the threshold, or until there are no more variables to remove. The final model includes only the independent variables that are statistically significant predictors of the dependent variable.

The multiple regression coefficients of the complete model are reported in Table 5, where the least significant explanatory variable (with the highest p-value) is *indpro*, which is removed from the model. After step 1 of the backward elimination stepwise procedure, the remaining independent variables are summarized in Table 6.

Table 6. Variable selection (step 1)

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0094	0.0026	3.6556	0.0003
t1y	0.0084	0.0143	0.5902	0.5555
t10y	0.0652	0.0307	2.1231	0.0345
hyield	-0.4377	0.0279	-15.6782	0.0000
cpi	-0.5112	1.0373	-0.4928	0.6225
ppi	0.2285	0.3608	0.6333	0.5270
oil	-0.0141	0.0285	-0.4960	0.6203
pce	-0.4181	0.2060	-2.0297	0.0433

After step 1, the least statistically significant independent variable (with the highest p-value) is *cpi,* which is eliminated in step 2.

Table 7. Variable selection (step 2)

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0087	0.0021	4.1917	3.6246E-05
t1y	0.0069	0.0139	0.4926	0.6227
t10y	0.0661	0.0306	2.1610	0.0315
hyield	-0.4373	0.0279	-15.6898	3.7922E-41
ppi	0.1071	0.2634	0.4068	0.6845
oil	-0.0137	0.0284	-0.4817	0.6303
pce	-0.4208	0.2057	-2.0456	0.0416

After step 2, the least statistically significant independent variable is *ppi,* which is eliminated in step 3.

Table 8. Variable selection (step 3)

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0088	0.0020	4.3698	1.7017E-05
t1y	0.0071	0.0139	0.5126	0.6086
t10y	0.0680	0.0302	2.2495	0.0252
hyield	-0.4355	0.0275	-15.8418	9.3221E-42
oil	-0.0075	0.0239	-0.3124	0.7549
pce	-0.4208	0.2054	-2.0488	0.0413

After step 3, the least statistically significant independent variable is *oil*, which is eliminated in step 4.

Table 9. Variable selection (step 4)

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.0089	0.0020	4.4171	1.3851E-05
t1y	0.0075	0.0138	0.5418	0.5883
t10y	0.0655	0.0291	2.2497	0.0252
hyield	-0.4336	0.0268	-16.2087	3.4466E-43
pce	-0.4518	0.1797	-2.5146	0.0124

After step 4, the least statistically significant independent variable is *t1y*, which is eliminated in step 5.

Table 10. Variable selection (step 5)

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.0089	0.0020	4.4398	0.0000
t10y	0.0729	0.0257	2.8310	0.0049
hyield	-0.4322	0.0266	-16.2534	0.0000
pce	-0.4397	0.1781	-2.4692	0.0141

After step 5, the remaining independent variables are shown in Table 10, where all the predictors are statistically significant at the 95% confidence level. To confirm this conclusion, the ANOVA results are reported in Table 11, where the significance F value is < 0.05, verifying that the explanatory variables in Table 10 are also jointly statistically significant with 95% confidence level.

Table 11. ANOVA results (step 5)

	df	SS	MS	F	Significance F
Regression	3	0.2983	0.0994	89.4197	1.03648E-41
Residual	310	0.3447	0.0011		
Total	313	0.6430			

At this stage, the backward elimination stepwise procedure can be considered terminated, and the multiple regression model can be expressed as:

$$stocks = \alpha + \beta_1 * t10y + \beta_2 * hyield + \beta_3 * pce,$$

where the regression statistics are summarized in Table 12.

Table 12. Regression statistics after step 5

Multiple R	0.6811
R Square	0.4639
Adjusted R Square	0.4587
Standard Error	0.0333
Observations	314

According to the coefficient of determination (R Square), 46.39% of the variance in the dependent variable is explained by the independent variables. The adjusted R Square takes into account the number of independent variables and penalizes the addition of irrelevant or redundant independent variables in the model. Therefore, the adjusted R Square is a more appropriate measure of the model goodness of fit when comparing models with different numbers of independent variables, and after step 5 it explains 45.87% of the dependent variable variance. It is worth noting that by construction, adjusted R Square is always lower than R square.

To further simply the model and perform linearity, autocorrelation, and homoscedasticity tests, the backward elimination stepwise procedure is continued to reduce the number of independent variables. After step 5, the least statistically significant independent variable is *pce*, which is eliminated in step 6.

Table 13. Variable selection (step 6)

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.0072	0.0019	3.7880	0.0002
t10y	0.0515	0.0244	2.1079	0.0358
hyield	-0.4206	0.0264	-15.9399	3.1496E-42

The regression statistics after step 6 are reported in Table 13, and by eliminating the explanatory variable *pce*, there is only a negligible reduction in the model predictive power of less than 2% with respect to step 5, as reported in Table 14.

Table 14. Regression statistics after step 6

Multiple R	0.6733
R Square	0.4534
Adjusted R Square	0.4498
Standard Error	0.0336
Observations	314

Finally, in step 7 after eliminating the independent variable *t10y*, the only remaining explanatory variable is *hyield*:

$$stocks = \alpha + \beta * hyield,$$

as shown in Table 15.

Table 15. Variable selection (step 7)

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.0073	0.0019	3.8060	0.0002
hyield	-0.4200	0.0265	-15.8343	7.4188E-42

The regression statistics are summarized in Table 16, and the loss of the model predictive power stemming from removing the *t10y* independent variable is still less than 2%.

Table 16. Regression statistics after step 7

Multiple R	0.6675
R Square	0.4456
Adjusted R Square	0.4438
Standard Error	0.0338
Observations	314

The overall loss of predictive power from a model with three predictors (*t10y, hyield, pce*) shown in Table 10, to a model with only one independent variable (*hyield*) is ~3%, but a simple model will allow to more easily assess other critical features such as linearity, autocorrelation, homoscedasticity, normality.

4.2 Linear Independence

Linear independence consists of the regression explanatory variables not being highly correlated. When two or more explanatory variables are perfectly correlated, then they are linearly dependent, and it is impossible to estimate the unique coefficients for each of them. This situation is known as perfect multicollinearity, and as a result, the regression model becomes unstable and unreliable. To avoid this problem, we need to ensure that the explanatory variables are linearly independent, which means that they are not perfectly correlated with each other. This allows us to estimate unique regression coefficients for each variable, so that the model remains stable and reliable.

Linear independence is evaluated through the multicollinearity test, which consists of calculating the inverted correlation matrix of the independent variables and assessing whether the main diagonal values are < 10. If the main diagonal values are > 10, then the variance of the estimated regression coefficients is inflated, thus leading to inflated standard errors, making it difficult to distinguish between the effects of the different independent variables on the dependent variable. To address multicollinearity, it may be necessary to consider dropping one or more of the correlated independent variables from the model or using techniques such as principal component analysis (PCA) or ridge regression to reduce the impact of multicollinearity on the model estimates. The multicollinearity test must be performed when there are two or more explanatory variables. When all the elements on the principal diagonal of the correlation matrix are < 10 like in Table 17, then there is no multicollinearity (or linear dependence) among the explanatory variables.

Table 17. Inverted correlation matrix

	t1y	t10y	hyield	cpi	ppi	oil	indpro	pce
t1y	1.49	-0.71	-0.09	-0.45	0.28	0.14	-0.17	-0.08
t10y	-0.71	1.59	-0.06	0.13	-0.36	-0.25	0.09	-0.18
hyield	-0.09	-0.06	1.13	0.05	-0.26	0.43	-0.11	0.15
cpi	-0.45	0.13	0.05	2.67	-2.05	0.09	0.11	-0.14
ppi	0.28	-0.36	-0.26	-2.05	3.38	-1.17	-0.17	0.17
oil	0.14	-0.25	0.43	0.09	-1.17	2.33	0.21	-0.94
indpro	-0.17	0.09	-0.11	0.11	-0.17	0.21	2.26	-1.76
pce	-0.08	-0.18	0.15	-0.14	0.17	-0.94	-1.76	2.91

The explanatory variables are not linearly dependent and to further confirm this conclusion, the correlation matrix reported in Table 18 shows a low correlation among all the explanatory variables, except for *pce* vs *indpro*, which will not be included in the model.

Table 18. Correlation matrix

	t1y	t10y	hyield	cpi	ppi	oil	indpro	pce
t1y	1.00	0.51	0.07	0.31	0.22	0.16	0.24	0.26
t10y	0.51	1.00	0.01	0.33	0.38	0.37	0.23	0.33
hyield	0.07	0.01	1.00	-0.03	-0.03	-0.26	-0.05	-0.16
cpi	0.31	0.33	-0.03	1.00	0.78	0.48	0.22	0.32
ppi	0.22	0.38	-0.03	0.78	1.00	0.63	0.27	0.38
oil	0.16	0.37	-0.26	0.48	0.63	1.00	0.35	0.56
indpro	0.24	0.23	-0.05	0.22	0.27	0.35	1.00	0.74
pce	0.26	0.33	-0.16	0.32	0.38	0.56	0.74	1.00

4.3 Correct Functional Form

The correct functional form consists of evaluating whether the optimal relationship among the regression independent variables with the dependent variable is linear or non-linear. One common approach to selecting the correct functional form is to begin with a simple linear regression model and then test for deviations from linearity using various diagnostic tools. For example, we can use residual plots, scatterplots, or tests of linearity. If there is evidence of non-linearity, we can consider using polynomial regression, logarithmic transformation, or other functional forms that better capture the shape of the relationship between the dependent variables and the independent variables.

The Ramsey RESET (*Regression Equation Specification Error Test*) linearity test [10] consists of adding squared fitted regression data as independent variable, and assessing if it is individually statistically significant. The test was developed by the economist James Ramsey in 1969 and is based on the idea that if a regression model is correctly specified, then any function of the independent variables (such as their squared or cubed values) should not have a significant impact on the dependent variable. In contrast, when such a function is found to be significant, this suggests that the model may be misspecified. The RESET test involves adding a set of squared (or cubed) terms of the original independent variables to the regression model and then examining whether these new terms are statistically significant. If they are significant, then the model may be misspecified, and it should be considered adding additional variables to the model. Overall, the RESET test is a useful tool for identifying potential problems with a regression model and for improving its accuracy. If the squared fitted regression coefficient is individually statistically significant, then non-linear functional form transformations of dependent and/or independent variables need to be evaluated.

A quadratic functional form consists of performing a quadratic transformation to the independent variables:

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{1,t}^2 + \beta_3 x_{2,t} + \beta_4 x_{2,t}^2 + \cdots + \beta_{2v-1} x_{v,t} + \beta_{2v} x_{v,t}^2 + e_t.$$

In the Ramsey RESET linearity test we evaluate if the addition of the squared fitted values independent variable, computed with the model obtained after variable selection step 7 represented in Table 15.

$$\widehat{fitted}_t = \alpha + \beta * hyield_t$$

has a p-value > 0.05. The model evaluated in the RESET test is given by:

$$\widehat{stocks}_t = \alpha + \beta_1 * hyield_t + \beta_2 * \widehat{fitted}_t^2$$

and the regression results are reported in Table 19.

Table 19. Ramsey RESET linearity test regression results

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0060	0.0020	2.9807	0.0031
hyield	-0.4447	0.0292	-15.2308	1.5991E-39
fitted(var sel step 7)^2	1.3976	0.7059	1.9798	0.0486

Since the p-value of the squared fitted independent variable is < 0.5, this variable is statistically significant at the 95% confidence level, thus indicating that the model obtained after variable selection step 7 may be misspecified, and that a liner form might not be the best predictive model. Moreover, the ANOVA results shown in Table 20 report that the significance F is < 0.05, confirming that the coefficients are jointly statistically significant. In the Ramsey RESET linearity test, the null hypothesis is that the multiple regression model is correctly specified with respect to linearity, meaning that there is no omitted non-linear relationship between the independent variables and the dependent variable. In other words, the null hypothesis states that the functional form of the regression model is correctly specified and that there is no need to include additional non-linear terms. The alternative hypothesis is that the model is misspecified, and that there is a need to include additional non-linear terms to better capture the relationship between the independent variables and the dependent variable. A significance F < 0.05 rejects the null hypothesis.

Table 20. Ramsey RESET linearity test ANOVA results

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.2909	0.1455	128.4951	2.10975E-41
Residual	311	0.3520	0.0011		
Total	313	0.6430			

The results of the RESET test suggest that there is evidence of non-linearity in the relationship between the dependent variable and the independent variables in the regression model. Hence, a non-linear functional form transformation of the dependent and independent variables, like a quadratic transformation, needs to be evaluated.

4.4 Residuals No Autocorrelation

Residuals autocorrelation occurs when the forecasting errors (or residuals) of a regression model are not statistically independent and are correlated with each other. The residuals no autocorrelation test consists of evaluating whether the residuals (or forecasting errors) have a constant mean value through the Breusch-Godfrey autocorrelation test [11][12], also known as the Breusch-Godfrey LM (Lagrange Multiplier) test, which is a statistical test used to detect the presence of autocorrelation in a regression model.

The test checks whether the residuals are significantly correlated with their lagged values, thus indicating the presence of autocorrelation. It consists of considering the regression residuals data as the dependent variable, then adding the lagged regression residuals data as independent variables and assessing if they are individually or jointly statistically significant. The Breusch-Godfrey autocorrelation test verifies the null hypothesis that there is no autocorrelation in the residuals of a regression model. Specifically, the null hypothesis states that the residuals of the regression model are not correlated with each other over time or across different observations. The test is performed by:

- Estimating the residuals of the regression model;
- Using these residuals to construct an auxiliary multiple regression equation;
- The auxiliary regression equation includes the lagged values of the residuals as explanatory variables;
- The null hypothesis states there is no significant autocorrelation in the residuals of the auxiliary regression equation over time or across different observations;
- The test statistic for the Breusch-Godfrey autocorrelation test is the LM (i.e., Lagrange Multiplier) test statistic, which measures the difference in the sum of squared residuals between the null and alternative hypotheses;
- The LM-statistic follows a chi-squared distribution with the number of lagged residuals used as degrees of freedom;
- If the calculated LM-statistic is greater than the critical value at the desired level of statistical significance, the null hypothesis of no autocorrelation is rejected, and the alternative hypothesis of autocorrelation is accepted.

The Breusch-Godfrey autocorrelation test is commonly used in econometric analysis to test for autocorrelation in time series. The test can be applied to both linear and nonlinear regression models, and it can be used to test for autocorrelation up to a specified number of lags.

The current period, previously fitted model forecasting errors or residuals are given by the following expression:

$$\hat{\varepsilon}_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_v x_{v,t} + \gamma_1 \hat{\varepsilon}_{t-1} + \cdots + \gamma_p \hat{\varepsilon}_{t-p} + e_t,$$

where $x_{(i,t)}$ is the *i-th* explanatory variable at time *t*, $\hat{\varepsilon}_{t-p}$ is the previously fitted model forecasting error or residual at period *t-p, p* is the number of lags included in the test, and e_t is the regression forecasting error or residual at time *t*. If γ_1 indicates that we include only 1 lag, and γ_p indicates we include *p* lags, then we evaluate the $\gamma_1,\cdots,\beta\gamma_p$ joint statistical significance by testing if either their t-statistic or F-statistic p-value (i.e., significance F) is < 0.05:

$$\gamma_1, \cdots, \gamma_p \rightarrow t_{pvalue} \; or \; F_{pvalue} < 0.05.$$

If lagged regression residuals coefficients $\gamma_1,\cdots,\beta\gamma_p$ are individually or jointly statistically significant, then lagged dependent variables data needs to be added as independent variables in the original regression model. Therefore, we have that the current period regression fitted or forecasted values (in case of 2 explanatory variables and p-lagged residuals) are given by:

$$y_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + AR_1 y_{t-1} + \cdots + AR_p y_{t-p} + e_t,$$

where $AR_1,\cdots,AR_p$ are the autoregressive coefficients multiplied by the previous period dependent or explained variable data all the way to the *p* autoregressive coefficients plus the regression forecasting error or residual at time *t*.

We performed the Breusch-Godfrey test including only 1-lagged (or the previous period) residuals. As dependent variables we consider the corresponding regression fitted values forecasting errors or residuals, and as independent variables we have the same variables included in the corresponding model (i.e., *hyield*) and the previous period residuals. The regression results are reported in Table 21.

Table 21. Breusch-Godfrey test regression results

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	2.7289E-06	0.0019	0.0014	0.9989
hyield	-0.0009	0.0266	-0.0336	0.9732
resid(var sel step 7)(-1)	-0.0364	0.0568	-0.6408	0.5221

The p-value of the lagged residuals is > 0.05 showing that the 1-lagged residuals independent variable is not statistically significant; therefore, we can conclude there is no 1-lagged residuals autocorrelation with 95% of statistical confidence. Moreover, the F-statistics p-value (i.e., significance F) reported in Table 22 confirms that the 1-lagged residuals are not jointly statistically significant.

Table 22. Breusch-Godfrey test ANOVA results

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	0.0005	0.0002	0.2053	0.8145
Residual	311	0.3560	0.0011		
Total	313	0.3565			

In conclusion, there is no need to add lagged dependent variables data in the original multiple regression model.

4.5 Residuals Homoscedasticity

Residuals homoscedasticity consists of evaluating whether the regression residuals or forecasting errors have a constant variance. Two main homoscedasticity tests can be performed:

- *Breusch-Pagan* test [14]: This is a statistical test that checks whether the variance of the residuals is constant across all levels of the independent variables. The null hypothesis of the test is that the variance of the residuals is constant (homoscedastic). If the p-value is less than the significance level (i.e., 0.05 at 95% of statistical confidence), then the null hypothesis is rejected, indicating that the assumption of homoscedasticity does not hold;
- *White* test [13]: This test is similar to the Breusch-Pagan test but is more robust to violations of normality and homoscedasticity. The test checks whether the residuals are uncorrelated with the squared values of the independent variables. If the significance F value is < 0.05, then the null hypothesis of homoscedasticity is rejected at 95% of statistical confidence.

In this work we performed the White test, which is commonly used to test for heteroscedasticity in regression analysis. It was proposed by Halbert White in 1980 and is sometimes referred to as the "*generalized heteroscedasticity test*". The White test is a robust version of the Breusch-Pagan test, which assumes that the errors in the regression model are normally distributed. In contrast, the White test is more general as it can be used when the errors are not normally distributed. The test is based on the following steps:

- Obtain the multiple regression model residuals;
- Square the residuals to obtain the new dependent variable;
- Estimate a new regression model using the squared residuals as dependent variable and the independent variables and their squared values as additional independent variables;
- Test the null hypothesis that the coefficients of the squared independent variables are equal to zero, thus indicating homoscedasticity.

There are two versions of the White homoscedasticity test: the *White(no-cross terms)* and the *White(cross terms)* tests, which consist of using squared regression residuals data as dependent variable and doing quadratic (no-cross terms) or quadratic and cross products (cross terms) transformations to the independent variables in order to assess if the squared independent variables are jointly statistically significant. If the independent variables are jointly statistically significant, heteroscedasticity consistent standard errors estimation needs to be done (and the null hypothesis of homoscedasticity must be rejected).

In this work we performed the White(no-cross terms) test. The following regression model is an example of the White(no-cross terms) homoscedasticity test with two independent or explanatory variables, in which the current period, previously fitted model squared residuals (to the power of two) are expressed by the following formula:

$$White(no-cross\ term)\ \rightarrow \hat{\varepsilon}_t^2 = \alpha + \beta_1 x_{1,t} + \beta_2 x_{1,t}^2 + \beta_3 x_{2,t} + \beta_4 x_{2,t}^2 + e_t.$$

For this test the regression explained variable is the squared residuals and the independent variables are the quadratic transformation of the explanatory variables. The residuals homoscedasticity (or constant variance) is evaluated by means of the regression F-statistic p-value (significance F). If the significance F is < 0.05, meaning that the independent or explanatory variables are jointly statistically significant, then we need to reject the null hypothesis (homoscedasticity of the residuals), the heteroscedasticity (i.e., non-constant variance) of the residuals cannot be rejected, and we conclude that the residuals are heteroscedastic.

Table 23. White(no-cross terms) homoscedasticity test ANOVA results

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2.4593E-05	1.2296E-05	3.0945	0.0467
Residual	311	0.0012	3.9736E-06		
Total	313	0.0013			

In Table 23 the significance F value is < 0.05, therefore the corresponding regression residuals are not homoscedastic. The residuals are heteroscedastic with 95% of statistical confidence according to the White(no-cross terms) test. Hence, it is necessary to perform an estimation of heteroscedasticity consistent standard errors and the associated F-statistic. Moreover, the p-values of the White(no-cross term) test regression results shown in Table 24 confirm that regression coefficients of the linear and squared independent variables are not statistically significant with a 95% of statistical confidence.

Table 24. White(no-cross terms) homoscedasticity test regression results

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0011	0.0001	9.1250	9.2639E-18
hyield	0.0019	0.0019	1.0021	0.3171
hyield^2	0.0102	0.0074	1.3847	0.1671

4.6 Heteroscedasticity Consistent Standard Errors

Heteroscedasticity consistent (HC) standard errors (also known as robust standard errors) is a method to adjust the standard errors in regression analysis to account for possible heteroscedasticity in the data. Heteroscedasticity means that the variance of the residuals is not constant across all levels of the independent variables.

HC standard errors use a different method to estimate the standard errors that does not rely on the assumption of homoscedasticity. Instead, it is based on the variance-covariance matrix of the coefficients using a *sandwich* estimator, which takes into account possible heteroscedasticity. This approach leads to standard errors that are robust to heteroscedasticity, thus producing more accurate inference in the presence of heteroscedasticity. The sandwich estimator involves *sandwiching* the original estimator of the standard errors between two estimates of the variance-covariance matrix. This generates a robust estimator of the variance-covariance matrix, which can then be used to calculate HC standard errors. The sandwich estimator is a popular method for estimating HC standard errors because it is relatively simple to calculate and is robust to violations of the homoscedasticity assumption. This method is commonly used in regression analysis and is an important tool for conducting reliable statistical inference.

In the previous section we evaluated the residuals homoscedasticity by means of the White(no-cross terms) test. In Table 23 we observed that the F-statistic associated p-value (i.e., significance F) was < 0.05, and we concluded that those residuals were not homoscedastic, but were heteroscedastic with 95% of statistical confidence, thus highlighting the need to estimate the heteroscedasticity consistent (HC) standard errors. The HC standard errors obtained with a sandwich estimator are reported in Table 25.

Table 25. Heteroscedasticity consistent standard errors

	Coefficients	*HC Standard Error*
Intercept	0.0073	0.0019
hyield	-0.4200	0.0352

After estimating the HC standard errors, we can estimate the HC standard errors F-statistic, which is reported in Table 26.

Table 26. Heteroscedasticity consistent standard errors F-statistic

Regression	142.0528

We can compare the HC standard errors and F-statistic values with the values from the original step 7 regression model summarized in Table 27 and Table 28 respectively

Table 27. Step 7 regression results

	Coefficients	*Standard Error*	*t Stat*	*P-value*
Intercept	0.0073	0.0019	3.8060	0.0002
hyield	-0.4200	0.0265	-15.8343	7.4188E-42

Table 28. Step 7 ANOVA results

	df	F	Significance F
Regression	1	250.7244	7.4188E-42
Residual	312		

The HC standard errors t-statistic is computed in Table 30 as the regression coefficients divided by the HC standard errors reported in Table 25, then the p-values are calculated as the two-tails t distribution of the absolute value of the t-statistic with the residual degrees of freedom reported in Table 29.

Table 29. HC standard errors ANOVA results

	df	F	Significance F
Regression	1	142.0528	3.0436E-27
Residual	312		

Table 30. HC standard errors regression results

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.0073	0.0019	3.8394	1.4933E-04
hyield	-0.4200	0.0352	-11.9186	3.0436E-27

The HC standard errors p-value (i.e., significance F) in Table 29 is computed through the F distribution of the F-statistic, the regression degrees of freedom, and the residuals degrees of freedom given in Table 29. The original regression F-statistic in Table 28 is compared with the F-statistic in Table 29 obtained with the HC standard errors.

The regression p-value (i.e., significance F) in Table 28 is < 0.05 and since we only have one independent variable (Table 27), it is used to evaluate whether the regression that includes that explanatory variable is a better fit than the one that only includes the constant or intercept term. Since the p-values is < 0.05, we conclude that the regression with the explanatory variable is a better fit with 95% of statistical confidence.

We confirm this result with the HC standard errors regression in Table 29, where the p-value (i.e., significance F) is < 0.05, and the regression with the independent variable is a better fit than the one that only includes the constant or intercept term with a 95% of statistical confidence. The HC standard errors in Table 30 are modified with respect to the original standard errors reported in Table 27. The HC standard errors regression p-value of the *hyield* explanatory variable is < 0.05, thus confirming that the *hyield* independent variable is statistically significant with 95% of statistical confidence.

4.7 Residuals Normality

Residuals normality consists of evaluating whether the regression residuals or forecasting errors skewness and kurtosis match a normal distribution. This is evaluated through the Jarque-Bera normality test [15], which is a statistical test used to determine if a given dataset follows a normal distribution. The Jarque-Bera test consists of calculating the test statistic and assessing if its value is less than six. If the test statistic value is greater than six, the dataset does not comply with the normality assumption. The Jarque-Bera normality test was introduced by Jarque and Bera in 1980 and is based on the idea that if a dataset is normally distributed, then its skewness and kurtosis should match the theoretical values of 0 and 3 respectively. Skewness is a measure of the asymmetry of the dataset, while kurtosis measures the degree of peakedness of the distribution. The Jarque-Bera test uses the JB test statistic, which follows a chi-squared distribution with two degrees of freedom under the null hypothesis of normality:

$$JB = \frac{n}{6}\left(s^2 + \frac{1}{4}ek^2\right),$$

where n is the sample size of the forecasting errors or residuals, s is the skewness of the dataset, and ek is the excessive kurtosis. To perform the test, the JB test statistic is computed and compared against the critical value of the chi-squared distribution. If the JB test statistic is greater than the critical value (i.e., six), then the null hypothesis of normality is rejected, indicating that the dataset does not follow a normal distribution. If JB value is less than or equal to the critical value, the null hypothesis of normality cannot be rejected. As shown in Table 31, the p-value from the chi-squared distribution with two degrees of freedom of the JB test statistic is < 0.05, then the null hypothesis of normality is rejected with 95% of statistical confidence. The Jarque-Bera test results on the forecasting errors or residuals of the step 7 regression analysis are summarized in Table 31, and confirm that the residuals do not follow a normal distribution.

Table 31. Jarque-Bera normality test results

	resid(var sel step 7)
skewness	-0.2856
excess kurtosis	1.1516
jarque-bera	21.6175
P-value (2 df)	2.0222E-05

The Jarque-Bera test is widely used in finance and economics to test the normality of financial and economic data. However, it should be noted that the test can sometimes give false positives or false negatives, especially for small sample sizes. The Jarque-Bera test is sensitive to sample size, and the trustworthiness of the test increases with larger sample sizes. Additionally, the test may not be appropriate for small sample sizes or non-normal distributions with heavy tails. Overall, the Jarque-Bera test is a useful tool for assessing the normality of a dataset, but it should be used in conjunction with other diagnostic tests to ensure accurate results.

4.8 Multiple Regression Assumptions Summary

All the multiple regression assumptions have been thoroughly evaluated on the complete historical data set of the independent variables, after a correct regression model has been obtained by means of the backward elimination stepwise regression. In summary:

- The explanatory variables are linearly independent and there is no multicollinearity;
- The correct functional form of the regression model was estimated through the Ramsey test, and the results suggested that there is evidence of non-linearity in the relationship among the independent variables and the dependent variable. Hence, a quadratic transformation of the explanatory variables was considered;
- The residuals autocorrelation was analyzed with the Breusch-Godfrey test, whose results demonstrated there is no statistically significant residuals autocorrelation and no need to add lagged dependent variables data in the original multiple regression model;
- The White(no-cross terms) test demonstrated that the residuals are heteroscedastic (with no constant variance across all levels of the explanatory variables). In this case the estimated standard errors of the regression coefficients can become biased and inconsistent;
- The heteroscedasticity consistent (HC) standard errors procedure used a different method to estimate the standard errors and confirmed that the independent variables coefficients in the original regression model are statistically significant;
- The normality of the residuals was evaluated through the Jarque-Bera test. The results confirmed that the residuals do not follow a normal distribution.

5. Multiple Regression Forecasting and Results

Multiple regression forecasting consists of fitting the best model within the *training data* subset and using the best fitting model to forecast within the *testing data* subset. Forecasting can be classified into:

- *Ex-ante forecasting* consists of dependent or explained variable current period forecast using previous periods independent or explanatory variables data only:

$$\hat{y}_t = \alpha + \beta_1 x_{1,t-1} + \beta_2 x_{2,t-1} + \cdots + \beta_v x_{v,t-1} + e_t$$

- *Ex-post forecasting* consists of dependent or explained variable current period forecast using current period independent or explanatory variables data:

$$\hat{y}_t = \alpha + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_v x_{v,t} + e_t.$$

Ex-ante forecasting in multiple regression refers to predicting the values of the dependent variable before the data is collected, based on the values of the previous period independent variables. In other words, it involves using a multiple regression model to make predictions about future values of the dependent variable. In contrast, ex-post forecasting refers to predicting the values of the dependent variable based on the actual values of the independent variables. In other words, it involves using multiple regression to predict the values of the dependent variable after the data has been made available.

These predictions can be compared to the actual values of the dependent variable to evaluate the accuracy of the model. Ex-ante forecasting allows to anticipate future outcomes based on current data. This can be particularly valuable in finance, where investors and traders need to make predictions about future trends and performance. However, it is important to note that ex-ante forecasts are inherently less accurate than ex-post forecasts, which are obtained by using actual data. This is because ex-ante forecasts are based on assumptions about future trends and relationships between the independent and dependent variables, which may not hold true in practice.

In this work we perform ex-ante forecasting based on 1-lagged explanatory or independent variables.

5.1 Training and Testing Range

The historical monthly data range is divided into:

- *Training range* spanning the period from Jan 1997 till Dec 2012;
- *Testing range* spanning the period from Jan 2013 till Feb 2023.

Multiple regression best fitting model evaluation is first performed on the training range, by obtaining the best fitted model and evaluating all the assumptions analyzed in Section 4. The forecasting accuracy of the model is then assessed on the testing range.

5.2 Forecasting Model Correct Specification (in the Training Range)

In this Section the best fitting of the regression model within the training range is performed to evaluate all the multiple regression assumptions studied in Section 4. The initial regression model includes all the independent or explanatory variables, and the ANOVA and regression results are shown in Table 32.

Table 32. ANOVA and regression results

	df	*F*	*Significance F*	
Regression	8	2.6528	0.0089	
Residual	182			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lpce	0.8610	0.8199	1.0502	0.2950
lindpro	1.4219	0.4900	2.9022	0.0042
loil	-0.0009	0.0526	-0.0174	0.9861
lppi	0.8827	0.6643	1.3288	0.1856
lcpi	-3.2645	1.8212	-1.7925	0.0747
lhyield	-0.0637	0.0565	-1.1274	0.2611
lt10y	-0.0262	0.0735	-0.3565	0.7219
lt1y	0.0476	0.0431	1.1045	0.2708
Intercept	0.0040	0.0049	0.8331	0.4059

In Table 32 the regression coefficient with the least statistically significant (i.e., largest) p-value is the lagged oil (*loil*) independent variable, which is not statistically significant with 95% of statistical confidence. For the model correct specification, we need to perform variable selection by removing the least statistically significant explanatory variables one at a time until all the remaining independent variables are individually and jointly statistically significant (i.e., with their p-value < 0.05) with 95% of statistical confidence. Therefore, we remove *loil* independent variable and the variable selection step 1 results are reported in Table 33.

Table 33. Variable selection (step 1) ANOVA and regression results

ANOVA				
	df	*F*	*Significance F*	
Regression	7	3.0484	0.0047	
Residual	183			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lpce	0.8600	0.8154	1.0546	0.2930
lindpro	1.4218	0.4885	2.9104	0.0041
lppi	0.8776	0.5944	1.4764	0.1416
lcpi	-3.2647	1.8162	-1.7975	0.0739
lhyield	-0.0636	0.0556	-1.1433	0.2544
lt10y	-0.0263	0.0731	-0.3602	0.7191
lt1y	0.0475	0.0429	1.1074	0.2696
Intercept	0.0040	0.0048	0.8385	0.4028

After variable selection step1, the least statistically significant dependent variable with the highest p-value is *lt10y* which is removed from the regression model. The variable selection step 2 results are reported in Table 34, where the least statistically significant dependent variable with the highest p-value is *lpce*, which is removed from the regression model.

Table 34. Variable selection (step 2) ANOVA and regression results

ANOVA				
	df	*F*	*Significance F*	
Regression	6	3.5516	0.0024	
Residual	184			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lpce	0.8686	0.8131	1.0682	0.2868
lindpro	1.4305	0.4868	2.9388	0.0037
lppi	0.8552	0.5898	1.4502	0.1487
lcpi	-3.2524	1.8116	-1.7953	0.0742
lhyield	-0.0633	0.0555	-1.1418	0.2550
lt1y	0.0377	0.0330	1.1425	0.2547
Intercept	0.0040	0.0048	0.8399	0.4021

The variable selection step 3 results are reported in Table 35, where the least statistically significant dependent variable with the highest p-value is *lt1y*, which is removed from the regression model.

Table 35. Variable selection (step 3) ANOVA and regression results

	df	*F*	*Significance F*	
Regression	5	4.0306	0.0017	
Residual	185			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lindpro	1.5642	0.4706	3.3240	0.0011
lppi	0.8531	0.5900	1.4459	0.1499
lcpi	-2.8039	1.7630	-1.5905	0.1134
lhyield	-0.0751	0.0544	-1.3807	0.1690
lt1y	0.0403	0.0329	1.2241	0.2225
Intercept	0.0064	0.0043	1.4821	0.1400

The variable selection step 4 results are reported in Table 36, where the least statistically significant dependent variable with the highest p-value is *lcpi*, which is removed from the regression model.

Table 36. Variable selection (step 4) ANOVA and regression results

ANOVA				
	df	*F*	*Significance F*	
Regression	4	4.6512	0.0013	
Residual	186			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lindpro	1.6696	0.4632	3.6042	0.0004
lppi	0.8849	0.5902	1.4993	0.1355
lcpi	-2.5515	1.7532	-1.4553	0.1473
lhyield	-0.0813	0.0542	-1.5000	0.1353
Intercept	0.0052	0.0042	1.2384	0.2171

The variable selection step 5 results are reported in Table 37, where the least statistically significant dependent variable with the highest p-value is *lppi*, which is removed from the regression model.

Table 37. Variable selection (step 5) ANOVA and regression results

ANOVA				
	df	*F*	*Significance F*	
Regression	3	5.4629	0.0013	
Residual	187			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lindpro	1.6963	0.4643	3.6538	0.0003
lppi	0.2116	0.3676	0.5757	0.5655
lhyield	-0.0760	0.0543	-1.4017	0.1627
Intercept	0.0016	0.0034	0.4752	0.6352

The variable selection step 6 results are reported in Table 38, where the least statistically significant dependent variable with the highest p-value is *hyield*, which is removed from the regression model.

Table 38. Variable selection (step 6) ANOVA and regression results

ANOVA				
	df	*F*	*Significance F*	
Regression	2	8.0574	0.0004	
Residual	188			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lindpro	1.7255	0.4607	3.7458	0.0002
lhyield	-0.0782	0.0540	-1.4481	0.1492
Intercept	0.0020	0.0033	0.6152	0.5392

The variable selection step 7 results are reported in Table 39, where the explanatory variable *lindpro* is statistically significant (i.e., p-value < 0.05) with 95% of statistical confidence.

Table 39. Variable selection (step 7) ANOVA and regression results

ANOVA				
	df	*F*	*Significance F*	
Regression	1	13.9368	0.0003	
Residual	189			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
lindpro	1.7247	0.4620	3.7332	0.0003
Intercept	0.0021	0.0033	0.6191	0.5366

Since after step 7 the regression model consists of only one explanatory variable *lindpro*, the ANOVA F-statistic and the associated p-value (i.e., significance F) is < 0.05, and confirms that the regression model that includes the independent variable *lindpro* is a better fit than the one that only includes the intercept term:

$$stocks_t = \alpha + \beta * indpro_{t-1} + e_t.$$

5.3 Forecasting Correct Specification

The correct functional form for this multiple regression forecasting consists of only one explanatory variable (i.e., *lindpro*) as summarized in Table 39. Hence, in this case we do not need to verify the linear independence or multicollinearity among the independent variables because there is only one explanatory variable.

We need to check the correct functional form through the Ramsey RESET linearity test, by including in the regression model the corresponding squared fitted values from variable selection step 7 model:

$$\widehat{fitted}_{t-1}^2 = (\alpha + \beta * indpro_{t-1})^2,$$

$$\widehat{stocks}_t = \alpha + \beta_1 * indpro_{t-1} + \beta_2 * \widehat{fitted}_{t-1}^2.$$

The results of the Ramsey RESET linearity test are reported in Table 40.

Table 40. Ramsey RESET linearity test results

ANOVA				
	df	F	Significance F	
Regression	2	12.7075	0.0000	
Residual	188			
	Coefficients	Standard Error	t Stat	P-value
fitted(var sel step 7)^2	-25.3533	7.7296	-3.2800	0.0012
lindpro	1.0214	0.4989	2.0471	0.0420
Intercept	0.0072	0.0036	1.9983	0.0471

The p-value of the squared fitted values independent variable is < 0.05, hence, it is statistically significant with 95% of statistical confidence, and we can conclude through the RESET test that there is no linearity. Moreover, the significance F value is < 0.05, thus rejecting the null hypothesis that there is no need of non-linear terms, in favor of the alternative hypothesis that it is necessary to include additional non-linear terms to better capture the relationship between the independent variables and the dependent variable. Therefore, we need a non-linear functional form of the regression model. We consider the quadratic transformation for the independent or explanatory variable *lindpro* (the lagged industrial production index) as shown in the following expression:

$$\widehat{stocks}_t = \alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro_{t-1}^2.$$

The results of the multiple regression on the non-linear model are reported in Table 41.

Table 41. Non-linear functional form regression results

ANOVA				
	df	F	Significance F	
Regression	2	12.7075	0.0000	
Residual	188			
	Coefficients	Standard Error	t Stat	P-value
lindpro^2	-75.4154	22.9922	-3.2800	0.0012
lindpro	0.8408	0.5250	1.6016	0.1109
Intercept	0.0071	0.0036	1.9761	0.0496

The squared lagged *lindpro* explanatory variable (*lindpro*^{2}) has a p-value < 0.05 and it is statistically significant with 95% of statistical confidence. Since the linear lagged *lindpro* independent variable has a p-value > 0.05, then it is not statistically significant. However, it is worth noting that even if the linear variable is not statistically significant and the quadratic form of the explanatory variable is statistically significant, nevertheless, because of the *hierarchy principle*, we can include also the lagged linear variable *lindpro* together with its quadratic transformation *lindpro*^{2}, so that we have the complete quadratic form.

The hierarchy principle in quadratic transformation refers to the idea that when creating a quadratic regression model, the inclusion of predictor variables should be guided by a hierarchical order reflecting the relationship between the variables. In a quadratic regression model, a quadratic term is added to the linear term of the predictor variable. The quadratic term is used to capture the non-linear relationship between the predictor variable and the dependent variable, and the hierarchy principle suggests that the quadratic term should be added to the model after the linear term. Moreover, the ANOVA F-statistic p-value (i.e., significance F) in Table 41 confirms that the linear and quadratic transformation of the explanatory variable are jointly statistically significant.

In order to test the non-linearity through the RESET test, we include in the model the fitted values of the squared non-linear regression represented in Table 41.

$$\widehat{fitted}^2_{t-1} = (\alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro^2_{t-1})^2$$

$$\widehat{stocks}_t = \alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro^2_{t-1} + \beta_3 * \widehat{fitted}^2_{t-1}.$$

The squared fitted values explanatory variable of the non-linear regression of Table 42 has a p-value > 0.05, hence; is not statistically significant with 95% of statistical confidence.

Table 42. Ramsey RESET linear test results

ANOVA				
	df	*F*	*Significance F*	
Regression	3	8.4268	0.0000	
Residual	187			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
fitted(non-lin reg)^2	-0.0714	3.6675	-0.0195	0.9845
lindpro^2	-74.5629	49.5000	-1.5063	0.1337
lindpro	0.8385	0.5398	1.5534	0.1221
Intercept	0.0071	0.0038	1.8779	0.0620

Therefore, we can conclude that there is linearity through the Ramsey RESET linearity test, hence, there is no need to include further functional transformation in the quadratic model represented in Table 41:

$$stocks_t = \alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro^2_{t-1} + e_t.$$

5.4 Forecasting Residuals No Autocorrelation

We consider the non-linear regression model represented in Table 41 and its residuals or forecasting errors. The residuals autocorrelation is evaluated by means of the Breusch-Godfrey autocorrelation test. The fitted residuals or forecasting errors from the non-linear regression model represented in Table 41 are given by:

$$\hat{\varepsilon}_t = stocks_t - (\alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro^2_{t-1}).$$

The dependent variable is given by the fitted residuals, while the independent or explanatory variables are given by the linear and quadratic transformation of the lagged *indpro* variable (*lindpro*) and by the lagged fitted residuals or forecasting errors:

$$\hat{\varepsilon}_t = \alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro^2_{t-1} + \gamma_1 * \hat{\varepsilon}_{t-1}.$$

The results of Breusch-Godfrey autocorrelation test are summarized in Table 43. The evaluation of the Breusch-Godfrey test is performed for 1-lagged explanatory variables and the p-values of the fitted lagged residuals is > 0.05. We can conclude that there is no 1-lagged residuals autocorrelation with 95% of statistical confidence. Furthermore, the significance F value obtained from ANOVA is > 0.05, thus confirming that the null hypothesis cannot be rejected. The null hypothesis of the Breusch-Godfrey autocorrelation test in multiple regression is that there is no autocorrelation in the residuals of the regression model.

Table 43. Breusch-Godfrey no autocorrelation test results

ANOVA				
	df	*F*	*Significance F*	
Regression	3	0.0333	0.9918	
Residual	187			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
resid(non-lin reg)(-1)	0.0233	0.0737	0.3161	0.7523
lindpro^2	0.9163	23.2291	0.0394	0.9686
lindpro	0.0122	0.5276	0.0231	0.9816
Intercept	-0.0001	0.0036	-0.0175	0.9860

5.5 Forecasting Residuals Homoscedasticity

The residuals homoscedasticity has been evaluated through the White(no-cross terms) test. A quadratic transformation of the quadratic independent variable of the non-linear regression model described in Table 41 is applied. The following regression model is an example of the White(no-cross terms) homoscedasticity test with the two independent or explanatory variables of the non-linear regression model in which the current period, previously fitted model squared forecasting errors or residuals (to the power of two) are expressed by the following formula:

$$\hat{\varepsilon}_t^2 = \alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro_{t-1}^2 + \beta_3 * (indpro_{t-1}^2)^2 + e_t.$$

We consider the F-statistic and its corresponding p-value (significance F) reported in Table 44. The results show that the significance F value is > 0.05, therefore we can conclude that the residuals are homoscedastic through the White(no-cross terms) test with 95% of statistical confidence. In fact, if the significance F is > 0.05, we cannot reject the null hypothesis, and the null hypothesis of the White test for heteroscedasticity is that the residuals, or forecasting errors, are homoscedastic, meaning that the variances of the errors are equal across all levels of the independent variables. More specifically, the null hypothesis of the White test affirms that there is no heteroscedasticity that can be explained by including the squared terms of the independent variables in the regression model. The test checks whether the squared residuals from the multiple regression model are linearly related to the independent variables and their interactions. If the test fails to reject the null hypothesis, it suggests that the assumption of homoscedasticity is not violated, and the errors are equally variable across all levels of the independent variables.

Table 44. White(no-cross terms) homoscedasticity test

ANOVA				
	df	*F*	*Significance F*	
Regression	3	1.1658	0.3241	
Residual	187			
	Coefficients	*Standard Error*	*t Stat*	*P-value*
(lindpro^2)^2	-2622.6543	2001.6253	-1.3103	0.1917
lindpro^2	2.5355	3.3074	0.7666	0.4443
lindpro	-0.0525	0.0348	-1.5087	0.1331
Intercept	0.0019	0.0003	7.7145	0.0000

5.6 Forecasting Residuals Normality

The residuals normality is evaluated through the Jarque-Bera normality test. The results are summarized in Table 45. The JB test statistic is < 6, therefore we can conclude that the residuals from the non-linear functional form of the multiple regression have a skewness and excess kurtosis similar to that of a normal probability distribution with 95% of statistical confidence. The p-value > 0.05 confirms that the null hypothesis, i.e., the residuals or forecasting errors are normally distributed, cannot be rejected.

Table 45. Jarque-Bera normality test results

	resid(non-lin reg)
skewness	-0.2789
excess kurtosis	0.2859
jarque-bera	3.1263
P-value (2 df)	0.2095

5.7 Forecasting Accuracy Metrics and Results (in the Testing Range)

Multiple regression forecasting accuracy consists of comparing multiple regression forecasting accuracy within the testing data subset with testing data subset random walk and training data subset arithmetic mean benchmarks, through the metrics of Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE).

- *Mean Absolute Error* consists of scale-dependent measure of forecasting accuracy based on arithmetic mean of absolute value of residuals or forecasting errors:

$$MAE = \frac{1}{n}\sum_{t=1}^{n}|e_t|$$

where the residuals or forecasting errors are expressed as:

$$e_t = y_t - \hat{y}_t$$

and $\hat{y}_t$ are the forecasted or fitted values;

- *Root Mean Squared Error* consists of scale-dependent measure of forecasting accuracy based on square root of arithmetic mean of squared residuals or forecasting errors;

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n} e_t^2}$$

- *Mean Absolute Percentage Error* consists of scale-independent measure of forecasting accuracy based on arithmetic mean of absolute value of residuals or forecasting errors as percentage of actual data:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{e_t}{y_t} * 100\right|.$$

A *scale-dependent* measure of forecasting accuracy, such as Mean Absolute Error (MAE), is sensitive to the units of measurement used for the target variable. For example, if the target variable is measured in dollars, the MAE will also be measured in dollars. This means that the MAE value is difficult to compare between datasets or models that use different scales or units of measurement. To address this issue, *scale-independent* measures of forecasting accuracy, such as the Mean Absolute Percentage Error (MAPE), are often used. These metrics express the forecasting error as a percentage of the actual value, making them comparable across different scales and units of measurement.

The forecasting accuracy metrics are evaluated within the testing range (from Jan 2013 till Feb 2023). The non-linear multiple regression model to estimate the stock market returns in the testing range is expressed by:

$$\widehat{stocks}_t = \alpha + \beta_1 * indpro_{t-1} + \beta_2 * indpro_{t-1}^2,$$

where the regression coefficients are reported in Table 41.

The benchmarks considered to evaluate the multiple regression model (i.e., MODEL) are the random walk (i.e., RANDOM WALK) and the arithmetic mean (i.e., MEAN) in Table 46. The random walk is implemented by using the previous period stocks data plus some randomness [16]. Among the three models, MEAN shows the smallest values for all the metrics, while MODEL has higher MAE and RMSE than RANDOM WALK, but MAPE is smaller than RANDOM WALK. Hence, according to the scale-dependent forecasting metrics, the random walk has a better forecasting accuracy with respect to the non-linear multiple regression model, but according to the scale-independent metric, the multiple regression model shows a better accuracy than the random walk model.

Table 46. Forecasting accuracy (testing range)

	MODEL	RAND WALK	MEAN
Scale Dependent:			
MAE	0.0499	0.0497	0.0328
RMSE	0.1480	0.0673	0.0428
Scale Independent:			
MAPE	232.0575	1212.1982	154.3205

Figure 1 shows the stock market (S&P 500) forecasted returns computed with the non-linear multiple regression fitted model from Jan 2013 till Jan 2020, compared against the actual data and the random walk model.

Figure 1. Testing range: Jan 2013 — Jan 2020

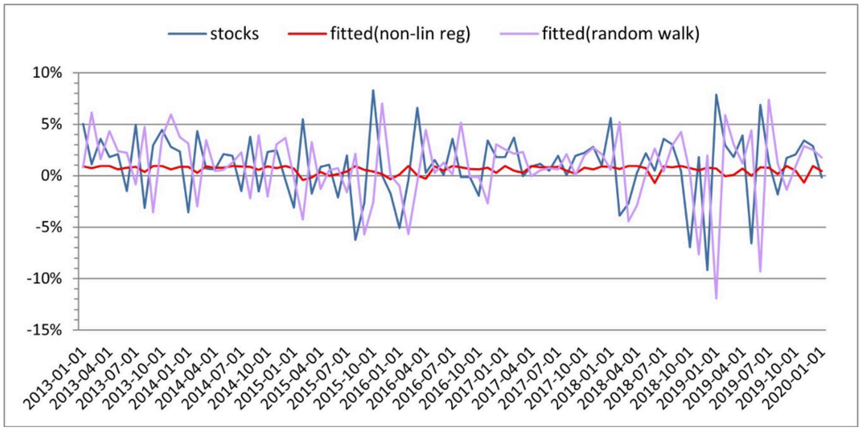
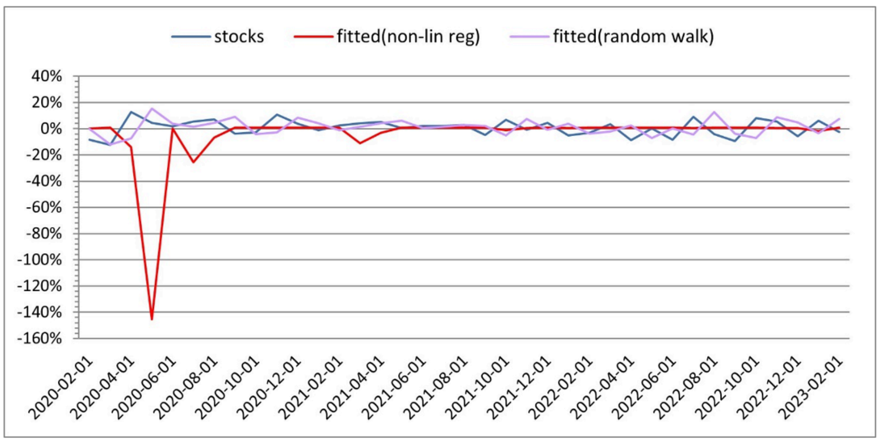


Figure 2 shows the same data and models from Feb 2020 till Feb 2023. The negative spike in the fitted multiple regression model is a consequence of the corresponding large drop in the industrial production index macroeconomic indicator, where in Apr 2020 the *indpro* return plunged by 13.38%.

Figure 2. Testing range: Feb 2020 — Feb 2023



6. Beyond Multiple Regression: Exponential Triple Smoothing

In the previous Sections we have presented a rigorous statistical assessment of multiple regression and have developed a non-linear (i.e., quadratic) model to forecast the stock market behavior (represented by the S&P 500 index) based on the industrial production index explanatory variable. However, multiple regression is not the only approach that can be used to perform predictions.

Exponential Triple Smoothing (ETS) [17][18] is a popular time series forecasting method that uses exponential smoothing to predict future values based on past observations. It is particularly useful for datasets with trend, seasonality, and randomness, making it a valuable technique for a wide range of applications, such as sales forecasting and stock market analysis. Typically, trend, seasonality, and randomness are the three key components of a time series. The trend refers to the overall direction of the data, whether it is increasing, decreasing, or remaining constant over time. The seasonality reflects the regular patterns or cycles in the data, which could be daily, weekly, monthly, or yearly, and it could be influenced by factors such as economic cycles, corporate earnings, fiscal and monetary policies, geopolitical events, and investor behavior. The randomness represents the unpredictable fluctuations that are not explained by trend or seasonality, which could be caused by unexpected news, investor sentiment, market psychology, algorithmic trading, and technical factors like market liquidity, trading volume, bid-ask spreads, or measurement errors.

ETS uses exponential smoothing to estimate the values of the three components and forecast future values based on these estimates. Exponential smoothing is a weighted average method that assigns different weights to the past observations based on their recentness. The closer an observation is to the present, the higher the weight assigned to it, and the farther away it is, the lower the weight. This means that recent observations have more impact on the forecast than older ones, reflecting the concept that more recent data is more relevant to predict the future. ETS is particularly useful when dealing with time series data with a complex seasonal pattern or where the trend is not linear. It can also handle data with missing values or irregular intervals between the observations.

In comparison, multiple regression is a statistical technique that models the relationship between a dependent variable and several independent variables. Unlike ETS, multiple regression does not take into account time series components such as trend and seasonality.

Overall, ETS and multiple regression are two different approaches to forecasting, and the choice between them depends on the specific characteristics of the data being analyzed. If the data has a strong temporal pattern, ETS may be the more appropriate choice. However, if the data has a more complex set of explanatory variables, multiple regression may yield a better solution.

- Pros of Exponential Triple Smoothing:
  - ETS can handle time series data with complex patterns, such as seasonality, trend, and noise (randomness).
  - ETS can handle missing data and irregular time intervals, making it a more flexible tool than (multiple) regression analysis.
- Cons of Exponential Triple Smoothing:
  - ETS assumes that the underlying patterns in the time series data are stable over time, which may not be the case when sudden changes in the underlying time series occur.
  - ETS relies on the most recent observations in the data, which can make the forecasts more volatile and less reliable when the data is highly erratic or there is a lot of noise in the data.
  - ETS does not account for external factors that may influence the time series data, such as economic trends.
  - ETS is primarily a univariate method, meaning it only considers one variable at a time, which may not be sufficient for complex forecasting tasks.
- Pros of Multiple Regression:
  - Multiple regression can handle a wide range of independent variables, making it useful for forecasting complex relationships between variables.
  - Multiple regression can account for external factors that may influence the dependent variable, making it a more robust forecasting tool in some contexts.
  - Multiple regression can handle both continuous and categorical variables, making it a more versatile tool than ETS.
  - Multiple regression can provide estimates of the effect of each independent variable on the dependent variable, which can be useful for interpreting the results.
- Cons of Multiple Regression:
  - Multiple regression assumes a linear relationship between the independent and dependent variables, which may not always be the case.
  - Multiple regression assumes that the independent variables are not highly correlated with each other, which can lead to biased estimates if this assumption is violated.
  - Multiple regression can be sensitive to outliers and missing data.

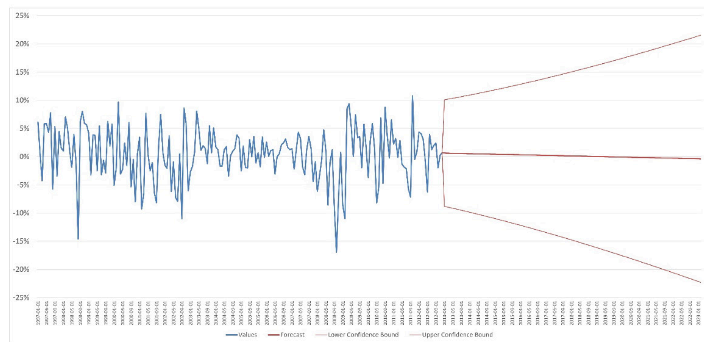
In summary, ETS is a useful method for forecasting time series data with complex patterns, while multiple regression is a more robust method that can handle a wider range of variables and relationships among them.

6.1 Forecasting with ETS

ETS is a method of time series forecasting that does not use external predictors. It is a univariate technique that relies on the historical observations of the time series itself to make predictions about future values. The method uses a combination of smoothing parameters to estimate the current level, trend, and seasonal components of the time series. These estimates are then used to forecast future values of the time series. However, this approach does not take into account any external variables or predictors that may influence the time series.

We have used the same training range for the stocks time series (from Jan 1997 till Dec 2012) to train the ETS model. The results of ETS on the testing range (from Jan 2013 till Feb 2023) are shown in Figure 3, where along with the forecast, also the 95% upper and lower confidence bounds are represented, where there is a 95% probability that the actual value will fall within these confidence bounds.

Figure 3. ETS on stocks (training and testing ranges)



A more detailed representation of ETS forecast in the testing range is shown in Figure 4 and Figure 5. In this case the non-linear multiple regression fitted model provides a better forecast than the ETS model. The ETS model 95% confidence bounds yield a useful reference for the range where the predicted variable could fluctuate.

Figure 4. ETS vs non-linear regression: testing range Jan 2013 – Jan 2020

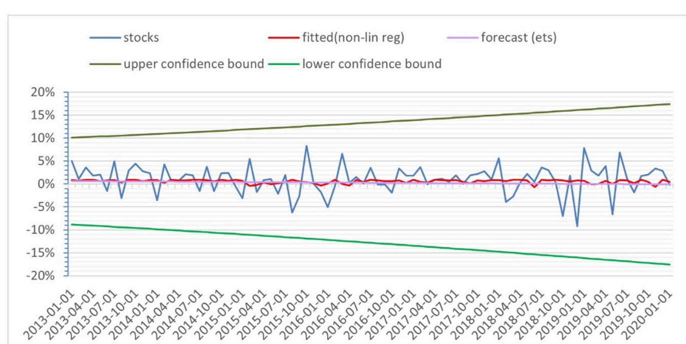
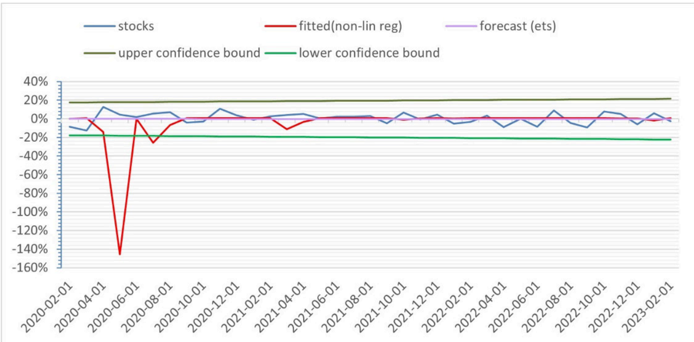


Figure 5. ETS vs non-linear regression: testing range Feb 2020 – Feb 2023



The accuracy metrics values of ETS forecasting are reported in Table 47. Based on these results, multiple regression is a better method than Exponential Triple Smoothing for a qualitative forecasting of the stock market.

Table 47. ETS Forecasting accuracy (testing range)

	ETS
Scale Dependent:	
MAE	0.0334
RMSE	0.0432
Scale Independent:	
MAPE	134.4117

7. Conclusions

Forecasting the stock market is a highly complex and dynamic task requiring a deep understanding of economic and financial factors. Both multiple regression and exponential smoothing are commonly used techniques for forecasting, but the choice of the best model depends on various factors such as data availability, quality, and the specific characteristics of the data being analyzed.

Multiple regression is typically used to identify a relationship between several independent variables and a dependent variable, which can be used to predict future values of the dependent variable. In the context of stock market forecasting, multiple regression can identify key factors that influence stock prices, such as economic indicators, financial data, company-specific information, and market trends, and then predict with some degree of statistical confidence the future movements of the asset prices.

Exponential smoothing is a time series forecasting technique that uses a combination of smoothing and trend analysis to predict future values. This method is particularly useful for data that display seasonality, as it can capture and predict intrinsic seasonal patterns in the data. Ultimately, the choice between multiple regression and exponential smoothing will depend on the specific nature of the data being analyzed. If the data contains multiple independent variables that may influence stock prices, multiple regression is more appropriate. On the other hand, if data display seasonal patterns, exponential smoothing may be more effective.

It is worth noting that predicting the stock market is an inherently uncertain task, and no forecasting model can guarantee accurate predictions. Therefore, it is important to approach any stock market forecasting with caution and always consider a range of factors, including market trends, economic indicators, and risk factors. Moreover, it is mandatory to rigorously assess all the statistical hypotheses in order to understand the limitations and pitfalls of the selected model and to correctly interpret the analysis results.

This work has presented a rigorous evaluation of multiple regression analysis, presenting a detailed assessment of all the tasks that must be performed for an informed and reliable use of multiple regression to forecast the underlying trend in the asset price movements.

Appendix Linear Independence and Multicollinearity

In multiple regression analysis, linear dependence among the explanatory variables and multicollinearity are related concepts, but they are not exactly the same.

Linear dependence among the independent variables means that one or more of the variables can be expressed as a linear combination of the other explanatory variables. This implies that there is a perfect linear relationship among some of the independent variables, which can lead to a lack of identifiability of the regression coefficients and other estimation problems.

On the other hand, multicollinearity refers to a situation where there is a high degree of correlation among the independent variables, but not necessarily perfect linear dependence. In other words, multicollinearity means that the independent variables are highly interrelated, and it may be difficult to determine the individual effect of each independent variable on the dependent variable.

In summary, while linear dependence among the explanatory variables is a more extreme case of multicollinearity, multicollinearity can still occur even when there is no perfect linear relationship among the independent variables.

References

[1] Malkiel, B. G. "*A Random Walk Down Wall Street.*" W. W. Norton & Company, Inc., NY, U.S., 1996.

[2] Kaufman, P. J. "*Trading Systems and Methods.*" 5th edition, John Wiley & Sons, Inc., Hoboken, NJ, U.S., 2013.

[3] Spiga, B. "*Trend Investigation: Operare in Tendenza con le Regressioni Lineari.*" Trading Library, 2022.

[4] Lincoln, T. H. "*Time Series Forecasting: ARMAX*," Technical Analysis of Stocks & Commodities, 1991.

[5] Kleinbaum, D. G., Kupper, L. L., Nizam, A., and Rosenberg, E. S. "*Applied Regression Analysis and Other Multivariable Methods.*" Cengage Learning, Boston, MA, U.S., 2014.

[6] Greene, W. H. "*Econometric Analysis.*" 8th Edition, Pearson Education, 2017.

[7] Wooldridge, J. M. "*Introductory Econometrics: A Modern Approach*." 7th Edition, South-Western Cengage Learning, 2019.

[8] https://finance.yahoo.com/

[9] https://fred.stlouisfed.org/

[10] Ramsey, J. B. "*Tests for Specification Errors in Classical Linear Least Squares Regression Analysis*," Journal of the Royal Statistical Society Series, 1969.

[11] Breusch, T. S. "*Testing for Autocorrelation in Dynamic Linear Models*," Australian Economic Papers, 1978.

[12] Breusch, T. S. and Godfrey, L. G. "*A review of recent work on testing for autocorrelation in dynamic linear models*," Journal of the Royal Statistical Society, 1981.

[13] White, H. A. "*Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity*," Econometrica, 1980.

[14] Breusch, T. S. and Pagan, A. R. "*A simple test for heteroscedasticity and random coefficient variation*," Econometrica, 1979.

[15] Jarque, C. M. and Bera, A. K. "*Efficient tests for normality, homoscedasticity and serial independence of regression residuals*," Economics Letters, 1980.

[16] Ruppert, D. "*Statistics and Data Analysis for Financial Engineering*." Springer, 2011.

[17] Hyndman, R. J. and Athanasopoulos, G. "*Forecasting: principles and practice.*" 3rd Edition, OTexts, 2021.

[18] Winters, P. R. "*Forecasting Sales by Exponentially Weighted Moving Averages*," Management Science, 1960.

Minute-Generated Price Values for Time Series Analysis

Christoph T. Wildensee, Ph.D., MFTA

Nordstr. 7
Seelze-Letter, Niedersachsen
30926 Germany

+49 511 4302128
christoph@wildensee.de

Introduction

Time series of monthly average temperature and precipitation trends at different locations, recorded over longer periods of time up to decades, allow the generation of prediction values for these locations and also bundled for regions with respect to a realistic view on a development of such basic information. The same applies to a large number of other climate-related influencing variables. This is also possible because seasonal fluctuations for each individual variable can be extracted via the models and thus adjusted trends and difference patterns can be derived (ML / AI). An expected value in each case is known. This results in a comprehensible model-based forecast outside of sudden exception restrictions. In the short-term forecast of the weather, forecast models are also used, which represent in detail influencing variables such as wind, precipitation amount and center of gravity, temperature and solar radiation. The quality of these is very high for one day and decreases with increasingly forecast days. The same applies to stable large-scale weather situations, which also have a high forecast quality for the first day, but do not drop as much in quality for the subsequent days as in the case of uncertain weather situations described above. What they all have in common is that the permanent observation and recording of influencing variables takes place at many observation points with pinpoint accuracy and that their significance for the forecast phenomenon is well-known [cf. e.g. also Trömel (2004), p. 27Ff].

A stock price that is displayed minutes later does not primarily depend on the previously emerging stock price trend [see Litz (2019), pp. 15, 19ff, also Schmelzer (2009), pp. 5ff, 10]. Rather, the previous price merely forms a flank or tendency of the new price, around which it fluctuates with a percentage offset. The expectations of market participants determine developments and are in turn supplied by both legitimate and illegitimate information about companies, individuals operating in the corporate environment, governmental / fiscal policy frameworks and other opinion leaders. A mathematically unambiguous description of this speculation is thus already difficult due to very many factors that cannot be operationalized. Otherwise, it could not be explained that a person like Elon Musk could set the price of Bitcoin as well as Twitter into considerable „oscillations“ only by short text statements. The sensitive reaction of market participants to a few hyped opinions and their market actions can certainly convulse the market [see e.g. Businessinsider (2022)].

But if we assume that the pattern of past performance is a significant factor in short-term prediction and that all other factors are so far behind in their effect that a mathematical description of this one factor is sufficient to make a significant forecast, *because such models work extremely well for medium- and long-term trend analysis,* **we can use a time series of past short-term performance to generate a prediction value**. The target should be divided: First, a value for the next time in the context of the choice of the **integration interval in the minute range.** But second, also a model use that targets a time period and later prediction value with such as weekly or monthly basis.

Auto-Regressive-Moving-Average (ARMA) & Co.

Functional Models such as "Auto-Regressive-Moving-Average" (ARMA), ARIMA (integrated), SARIMA (seasonal) and others are available for value prediction [see e.g. Litz (2019), p. 41ff]. Here, we will discuss the use of an **ARMA model.** Such and derived similar models are also supported, for example, in the software systems R and Python with many parameter setting options. If someone would like to integrate this into program code and use thereby **another language** without integration of existing libraries, it is necessary to make a transfer into the language logic. A representation of the Auto Regressive Moving Average ARMA(p, q) is done via...

$$ARMA(p,q) : Y_t = \tau + \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t$$

what means

$$Y_t = \tau + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t$$

where AR(p) is defined as...

$$Y_t = \tau + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t$$

and MA(q) as...

$$Y_t = \tau + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t$$

and can be integrated accordingly. A description for an adequate procedure is presented e.g. in Haile / Olive (2022) as well as Hyndman (2017) and (2018) for ARIMA. It is not necessary at this point to go into detail about the formula, neither about the mathematical derivation, nor about applied optimizations / procedures such as Box-Jenkins, Yule-Walker (cf. e.g. also Schmelzer (2009), p. 8ff), error analyses and others. This is done in an outstanding way at the mentioned place. However, this procedure is **at least codable**.

We integrate the available prices **every three minutes** (nearly 200 points per day). In the following a **with priority setting adapted example code** is used:

```
[...]
// i = > Array index; increase i++ to count_ds (number of records
with new values)
// Example calculation with fitting; YTF = function value
// ArrayNumNV[] = Array with values every 3 minutes from ≈
8:30-18:30 (CET)
AR[i]:=beta1 * ArrayNumNV[i-1] + beta2 * MA[i-1];
MA[i]:=ArrayNumNV[i] - AR[i];
DY[i] :=ArrayNumNV[i] - ArrayNumNV[i-1];
TRY // Attempt to create value
 YTF[i]:=ar1 + ArrayNumNV[i-1] +
 (c1 * (ArrayNumNV[i-1] - ArrayNumNV[i-2])) +
 (c2 * (ArrayNumNV[i-2] - ArrayNumNV[i-3])) +
 (c3 * (ArrayNumNV[i-3] - ArrayNumNV[i-4])) +
 (c4 * (ArrayNumNV[i-4] - ArrayNumNV[i-5])) +
 (c5 * (ArrayNumNV[i-5] - ArrayNumNV[i-6])) +
 (c6 * (ArrayNumNV[i-6] - ArrayNumNV[i-7])) -
 (ma1 * (DY[i-1]));
EXCEPT
// in case of error the value remains as read out
 YTF[i]:=ArrayNumNV[i];
END; // End try
```

This procedure is interesting because the last values are used via the **factors / parameters beta1, beta2 and c1-c6** and a relevance is assumed that cannot be denied, since the last 6-15 values reveal the short-term trend (*up to 15, since the last relevant value is used offset in the case of repeated values*). Values further back are largely irrelevant to the short-term analysis. The factors can of course vary depending on, for example, the market affiliations considered.

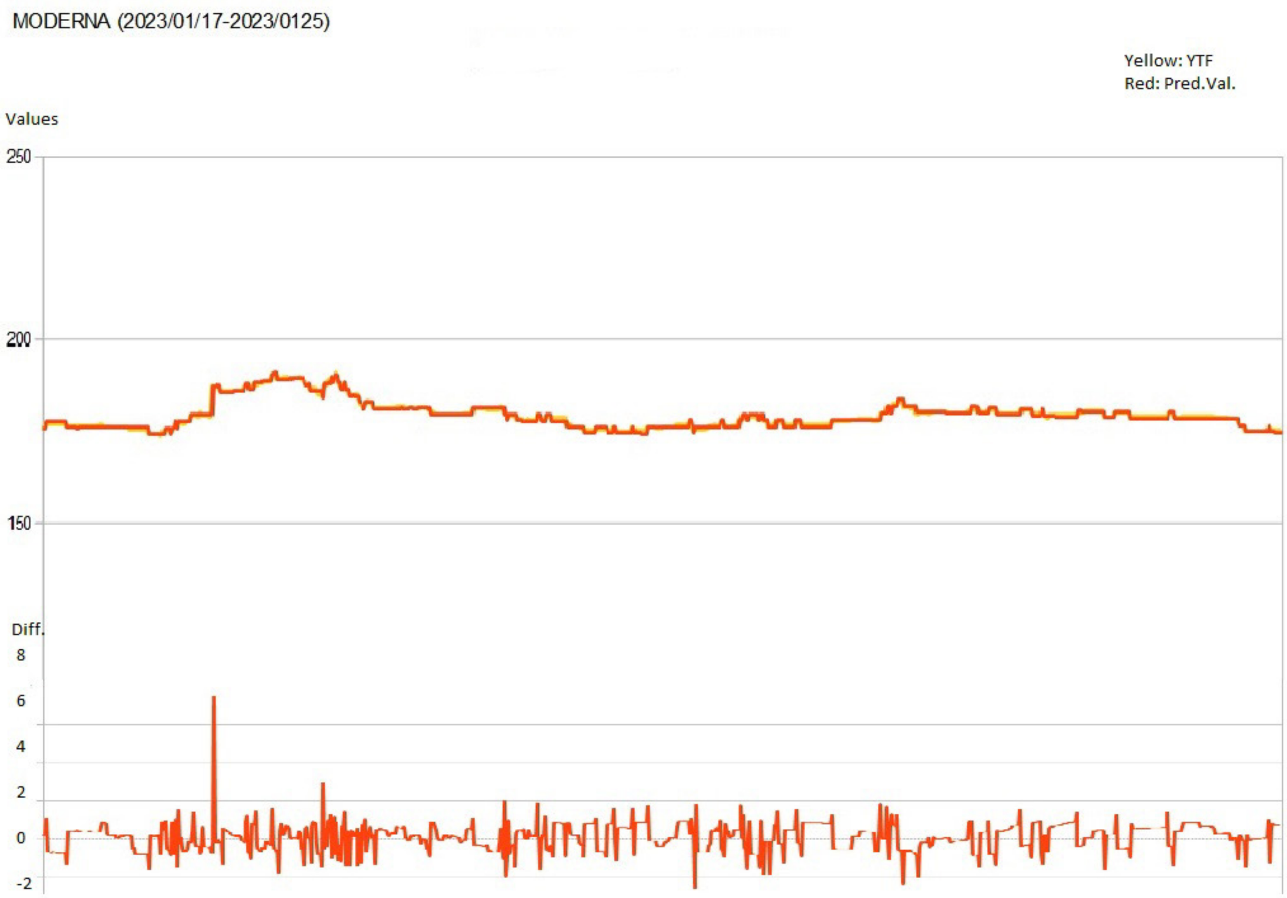
```
ETF[i]:=ArrayNumNV[i] - YTF[i];
// Passing the target prediction value to the variable value_pred
value_pred:=YTF[count_ds];
```

However, corrections of the output value due to different stock price levels are necessary, as the last movements of the price development often determine the direction and have to be included. Basically, it is also true that occasionally omitted values are not detrimental if you do not want to use the prediction value again immediately. In that case, it is necessary to wait for an interval.

Results

The underlying model (values supported by DLL) shows that without using existing library-based integrations specifically designed for this purpose, mapping such a systematics is still useful and possible, and that modeling it can lead to success. The not questioned taking over of supplied functionalities means that if necessary adjustments and mental reinterpretations do not (can) take place any longer. The represented approach of an own programming is arduous, but can offer interesting possibilities. Thus the exemplary representation of the stock development of MODERNA shows a small error rate between determined and prognosticated value (few extreme value outliers: max. 7.49 EUR, min. -2.61 EUR) and only rarely a higher deviation between prognosticated value and next value determination than +- 0.5% .

Figure 1: MODERNA, between 2023-01-17 and 2023-01-25 - value and variance trends.



The analysis of the basic data for this period then also makes it clear that **the model reacts quickly to these in each case for all stocks and adjusts the prediction to them, despite clearly recognizable trends and occasional outliers. Those outliers,** both downward and upward, of course **cannot be predicted** (not model-based determinable; see as a later example the loss of the stock value of near -60% of TUI between 2023-02-24 and 2023-03-31), but they can nevertheless **be identified and displayed. In this way, they are immediately in focus for the user and can be placed in context.**

Figure 2: Between 2023-01-17 and 2023-01-25 and between 2022-10-14 and 2022-12-23 (DAX had reached the level of 13940 points on 2022-12-23) - Percentage distribution of difference outliers across all stocks.

Between 2023-01-17 and 2023-01-25

Diff. NV-PREDVAL	Number of Datasets	
	10310	100,00%
< -1,0%	81	0,79%
< -2,0%	8	0,08%
> 1,0%	57	0,55%
> 2,0%	11	0,11%
-1,0% - 1,0%	10172	98,66%

Between 2022-10-14 and 2022-12-23

Diff. NV-PREDVAL	Number of Datasets	
	47744	100,00%
< -1,0%	1247	2,61%
< -2,0%	255	0,53%
> 1,0%	2174	4,55%
> 2,0%	574	1,20%
-1,0% - 1,0%	44323	92,83%

(< -1% includes values of < -2%; > 1% includes values of > 2%)

Other short-term Interval views

Determining a prediction value for a point in time further in the future - for instance a week or a month here - is more difficult. This view is not dependent on the last value, but must first be seen on a larger scale. As a weekly view, the base is about 1000 points back from the last determined value, in the monthly view about 4000 points away. So, if someone analyzes these **two views with the identical systematics like in the minute-range,** constantly going back in e.g. 200 resp. 800 points (1000 resp. 4000 points by 5 days a week = 200 resp. 800 points per step), **a stable view** on the next period can be determined. For this purpose, the respective interval has to be identified, which is included in the partial view. This means a change of the definition of the relevant evaluation vectors, so that [i-1] to [i-7] is substituted by the - starting from the last point and going back per 200 resp. 800 points per step as a vector - then available reference points for the use in the determination formula. It is therefore pretended that **only these values on average of the sub-ranges** and not the large number of intermediate values have been surveyed.

Seasonal corrections are inappropriate, as there are slight price movements around the time of the dividend payment and at a few other points, but these are not generally significant over the period and these interval views.

Figure 3: Short-term differences 1W/1M between 2023-02-03 and 2023-03-03 respectively 2023-04-03.

PRED_VAL 1 Week Stockname	PRED Fr 03.Feb 23 18:00	Real Fr 10.Feb 23 18:00	Diff.%	PRED Fr 10.Feb 23 18:00	Real Fr 17.Feb 23 18:00	Diff.%	PRED Fr 17.Feb 23 18:00	Real Fr 24.Feb 23 18:00	Diff.%	PRED Fr 24.Feb 23 18:00	Real Fr 03.Mrz 23 18:00	Diff.%
Adidas	156,05 €	138,78 €	-11,07	132,24 €	143,88 €	8,80	137,14 €	136,42 €	-0,53	130,77 €	136,42 €	4,32
NEL ASA	1,65 €	1,66 €	0,61	1,69 €	1,56 €	-7,69	1,62 € *	1,45 €	-10,49	1,48 €	1,45 €	-2,03
Drägerwerk	37,82 €	38,95 €	2,99	37,98 €	37,30 €	-1,79	36,51 €	37,20 €	1,89	36,40 €	37,20 €	2,20
DWS	31,22 €	30,76 €	-1,47	30,48 €	31,04 €	1,84	30,78 €	30,34 €	-1,43	29,90 €	30,34 €	1,47
Moderna	161,59 €	156,80 €	-2,96	155,04 €	154,18 €	-0,55	151,68 €	134,04 €	-11,63	130,53 €	134,04 €	2,69
Biontech	129,97 €	132,45 €	1,91	129,03 €	129,80 €	0,60	126,41 €	125,30 €	-0,88	121,48 €	125,30 €	3,14
SiemensEnergy	19,11 €	18,91 €	-1,05	18,92 €	19,25 €	1,74	19,34 €	18,72 €	-3,21	18,72 €	18,72 €	0,00
Continental	67,99 €	68,98 €	1,46	67,57 €	69,26 €	2,50	67,86 €	66,72 €	-1,68	65,19 €	66,72 €	2,35
LYX05L	21,91 €	21,53 €	-1,73	21,61 €	21,42 €	-0,88	21,50 €	21,20 €	-1,40	21,20 €	21,20 €	0,00

6th calendar week: 1-Week: Adidas with bad performance, still victim of Kanye West (Ye) + german soccer World Cup flop + financial expectations were not achieved.
11th+12th calendar week: ECB increase +0.5%, problems at SVB & other banks; bankruptcy of Silicon Valley Bank and other banking difficulties, and the upheavals in the Swiss banks Credit Suisse with UBS as the stronger part

* : No values on 2023-02-24

PRED_VAL 1 Month Stockname	PRED Fr 03.Feb 23 18:00	Real Fr 03.Mrz 23 18:00	Diff.%	PRED Fr 03.Mrz 23 18:00	Real Mo 03.Apr 23 18:00	Diff.%
Adidas	156,49 €	145,98 €	-6,72	140,90 €	162,80 €	15,54
NEL ASA	1,66 €	1,44 €	-13,25	1,49 €	1,24 €	-16,78
Drägerwerk	37,97 €	36,65 €	-3,48	35,97 €	38,40 €	6,76
DWS	31,16 €	30,58 €	-1,86	30,22 €	28,42 €	-5,96
Moderna	163,00 €	134,98 €	-17,19	132,93 €	144,76 €	8,90
Biontech	131,55 €	122,95 €	-6,54	120,63 €	116,80 €	-3,17
SiemensEnergy	19,09 €	19,89 €	4,19	19,76 €	21,06 €	6,58
Continental	68,14 €	72,32 €	6,13	71,20 €	68,68 €	-3,54
LYX05L	not enough values for 1-Month-prediction value					

The result is (*provided that the price values are collected consistently every weekday during the trading period*) a short-term forecast on a weekly and monthly basis, which can be assessed as model-based, comprehensible and reproducible. (*Documentation via F5 Show Hist. - Button „1W/1M" with files in the application directory.*) However, **as the time horizon increases,** the **validity of these prediction values decreases** and they may **not be** adequately **reflected in the case of large fluctuations. Figure 3 shows** that a forecast for one week and one month **does not approximate actual performance.** However, if there is also such a setback as in the 6th calendar week 2023 (Feb., 10th), this stronger deviation is to be expected. A better and impressive example is the development in the 11th and 12th calendar week with the changes in the ECB's key interest rates by +0.5%, the bankruptcy of Silicon Valley Bank and other banking difficulties, and the upheavals in the Swiss banks Credit Suisse with UBS as the stronger player (emergency rescue deal; stemming financial market panic). Every stock trended downward. These days diverse factors impacted stock performance significantly.

Figure 4: Between 2023-02-02 and 2023-04-06 - Percentage distribution of difference outliers across all stocks - Result during downward trend.

Between 2023-02-02 and 2023-04-06

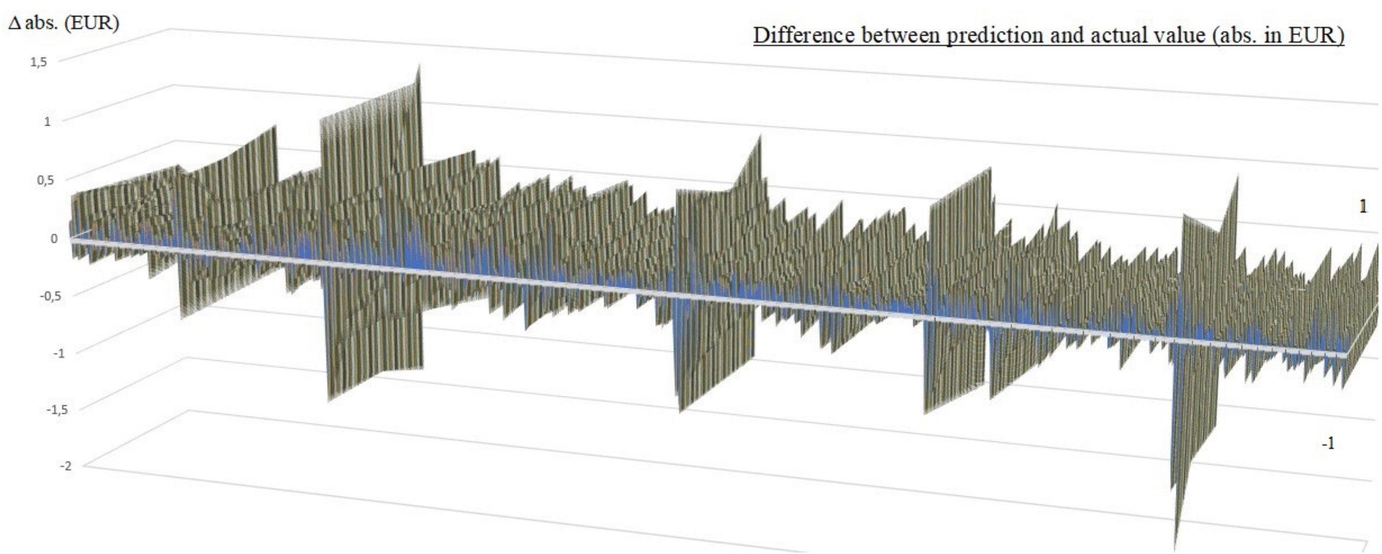
Diff. NV-PREDVAL	Number of Datasets	
	49094	100,00%
< -1,0%	1222	2,49%
< -2,0%	234	0,48%
> 1,0%	2233	4,55%
> 2,0%	450	0,92%
-1,0% - 1,0%	45639	92,96%

Finally, it can be seen that the model can react very quickly to changes in the minute range as shown, however, it can be understood that an appropriate reaction to sudden changes in the market cannot lead to an adjustment of a prediction value in the weekly and monthly range, since these changes take without temporal reference to the model. Then the model can no longer make corrections to the predicted value in a timely manner and becomes uncertain in its consideration of market influence. The consequence is a detachment of the prediction value from a realistic view of the market development due to the various messages from the market players.

Parameter profile

Each choice of parameters in a range between -1 and 1 with step 0.01 leads to an individual profile curve in a simulation, which shows how close the deviations between prediction and actual value are (Figure 5; x = number of values [c], min. 1000; y = abs. deviation in EUR; z = model response per chosen parameter setting between -1 and 1 with step 0.01). These profiles are therefore important for the choice of model parameters. If the curve for one or more particular parameters runs towards 0 for all values, while the other parameter variants show outliers up or down, this chosen parameter is eligible. Frequent strong outliers per subcurve are an indication that the model reacts too slowly with this parameter setting. This must be checked for **all parameters** and **applied in parameter context.**

Figure 5: Example of a Parameter Profil - Continental, c5, -1 (left) to 1 (right) step 0.01.



Conclusion

Even if it is mathematically ambitious at first to use such a **model adjustment for near-minute forecasts when price values are collected in a minute-range, the procedure offers a valid result**. ARMA / ARIMA models and similar ones are just established as linear, time-discrete models for usually short-term economic forecasts. The fast adjustment is immediately included in the next calculations even in case of strong ad-hoc fluctuations and it shows a considerable performance for this portfolio.

The vector-adjusted but identical model usage for other short-term intervals (week, month) is more difficult, as it shows weaknesses especially when strong ad-hoc swings occur, but the low sensitivity for such ad-hoc events is also the goal. For this reason, a weekly view is also rather unsuitable for the model presented here with its parameters. This is too close to get a good approximation when short-term oscillations occur. In this respect, it also shows that a model with parameters for the near-minute forecast is not appropriate for other short periods. They require different model parameter settings. The last period in each case needs a higher priority in the model to generate valid values. Thus, strong upward or downward trends are observed with a considerable delay. More market data is required and must be analysed and conditioned.

As a result, however, **the model can be used to justify** further **variant analyses for test purposes.**

FIBO_VIEW.ZIP provides an application that can be used to observe and experiment with the model over an extended period of time. If you want to get the source code and / or an actual database, send me an email (DLL start must be possible).

(Group of stocks: Adidas, Nel ASA, Drägerwerk, DWS, Moderna, Biontech, Siemens Energy, Continental)

References

Businessinsider (2022): What happens when Elon Musk moves markets with a tweet, Date 2022/05/31, https://www.businessinsider.com/what-happens-when-elon-musk-moves-markets-with-tweet-2022-5 .

Chan / Chen (2011): Subset ARMA selection via the adaptive LASSO, Statistics and its Interface 4, 197-205.

Eshel (2003): The Yule Walker Equations for the AR Coefficients, University of Pennsylvania.

Haile / Olive (2022): Forecast Intervals for Some ARIMA Time Series, Southern Illinois University, September 20, 2022.

Hyndman (2018): Forecasting: principles and practice, 2 ARIMA models, Monash University Business School, https://robjhyndman.com/acemsforecasting2018/2-ARIMA.pdf .

Hyndman (2017): Forecasting: principles and practice, 2.4 Non-seasonal ARIMA models, Monash University Business School, https://robjhyndman.com/uwa2017/2-4-NonseasonalARIMA.pdf .

Litz (2019): Vorhersage von Aktienkursen mithilfe rekurrenter Neuronaler Netze, Masterarbeit, Hochschule Harz.

Schmelzer (2009): Die Volatilität von Finanzmarktdaten - Theoretische Grundlagen und empirische Analysen von stündlichen Renditezeitreihen und Risikomaßen, Diss., Universität zu Köln, 2009.

Stadnik et. al. (2016): Fourier Analysis for Stock Price Forecasting: Assumption and Evidence, Journal of Business Economics and Management, 2016 Volume 17(3): 365–380.

Tang / Röllin (2021): Model identication for ARMA time series through convolutional neural networks, Decision Support Systems, Volume 146, 2021.

Trömel (2004): Statistische Modellierung von Klimazeitreihen, Diss., Universität Frankfurt am Main, 2004.

Wildensee (2022): http://www.wildensee.de/fibo_view.zip . http://www.wildensee.de/laz1a.htm .

Wildensee (2023): Raw Values, http://www.wildensee.de/raw_val.ods .

Christoph T. Wildensee, Ph.D., MFTA, is a Senior Auditor of IT-Risks and a Data / Process Analyst from Germany. He takes care of accounting-related IT-Systems such as SAP, but also Trading Systems.

Validation of a Trading System: Monte Carlo Analysis Applied to the IS & OOS Architecture

Michael Trequattrini, CSTA

Terni (TR), 05100, Italy

+39 3470010554
michael.trequattrini@gmail.com

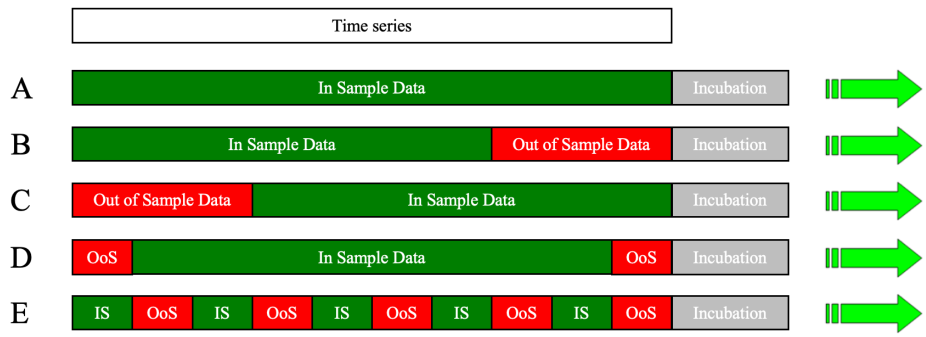
Abstract

This article aims to present a model for analyzing the robustness of a trading strategy, which, applied ex-post to the development and validation chain, will allow the determination of the probability of robustness on future and unknown data. In its various declinations, the model will be applied to a real and already validated strategy, focusing on interpreting the results achieved.

Choice of the Validation Architecture

A crucial step in the development of a trading strategy is the identification of the correct validation architecture. This, in its static and dynamic versions, allows the trader to segment the entire historical series available by obtaining an "In Sample" and an "Out of Sample" period. Regarding the static declination, which originates from Machine Learning, Figure 1 highlights several architectures found in the literature today. Without delving into each of them in detail, which would be beyond the article's focus, it is emphasized that the model presented can be applied to any architecture, making a choice fall based on personal needs and ways of working.

Figure 1: Validation architectures



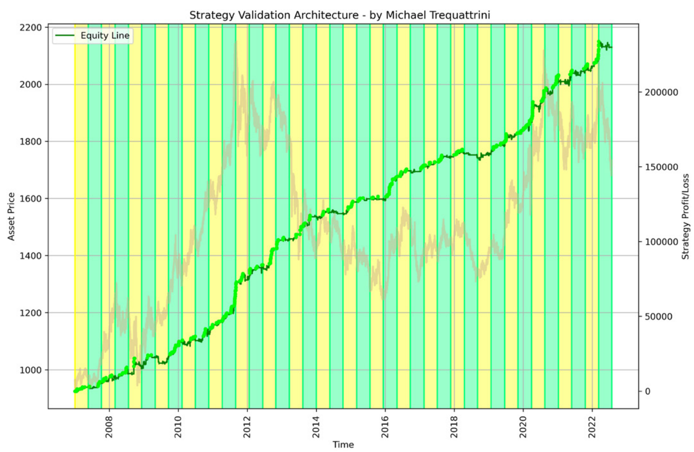
In this article, for the application of the model, will be used a trading strategy coded according to the architecture "E", called GSA ("Gandalf Segmented Architecture") and first presented at the IFTA 2017 event by the R&D group Gandalf Project. This approach makes it possible to eliminate the incubation period and to model and test the system under all trend and volatility conditions. It will also provide a robustness indicator, called GPDR, which will return an objective measure of persistence by comparing the swarm of the trades of the Is Sample and Out of Sample periods. Any value greater than 1.0 will indicate the robustness of the results and subsequent success in the validation process. It can learn more about the GSA architecture and the GPDR indicator in an article in the SIAT Mag #07.

Origin of the Model

The trading strategy to which the model will be applied is a breakout system on the hourly series of the Gold Future. Without analyzing the system's rules, suffice it to say that it has been adequately validated with a GSA in light of a GPDR of 4.5. Figure 2 shows the validation architecture: in yellow, the In Sample training periods, in light green, the Out of Sample testing periods, and finally, in darker green, the equity line of the system superimposed on the Gold Future. The size of each IS, and OOS block is equal to 2300 candles giving rise to a symmetric architecture. This choice is arbitrary and could result from discretionary logic or random assignment. In either case, the choice will affect the entire development chain. Think of optimizing the parameters of a strategy; inevitably, the results will depend on the segmentation chosen. Different splitting of the time series implies performing the optimization, at least in part, on different data and consequently will produce results with some degree of dissimilarity. If the variance of the results produced by the optimization, when varying IS and OOS segmentation, is reduced to the limit with the GSA technique, the same cannot be said for the other architectures. Moreover, whether GPDR or any other analysis is used, the validation process can lead to different results as the fractionation varies. Therefore, the event to be averted is when a system passes the validation due to a

good choice of IS and OOS split. Obviously, during the strategy development, the trader should choose and rely on a specific segmentation. For this reason, the model is only applied ex-post to the trading system's development and validation chain.
Its goal is to test whether the strategy passes the validation process even with different segmentations of the training and testing blocks.

Figure 2: Trading system on Gold Future with a GSA



Model Operation

The first step is creating several N validation architectures with different random segmentations following a Monte Carlo logic. For each of them, the backtest is performed on both the In Sample and Out of Sample portions. The key metrics are computed, and the validation test is performed (in the personal case using the GPDR indicator). At this point, it is counted how many times, out of the total N, the system passed the validation. If we are beyond a certain percentage of validated tests, that percentage can be seen as the probability of future system robustness. Figure 3 shows the model inputs, a thorough understanding of which is required for consistent results.

Figure 3: Model inputs

```
#### INPUTS ###

ASYMMETRICAL_ARCHITECTURE = False  # if True it builds asymmetrical architectures (IS > OOS)


ITERATIONS = 1000              # number of random architectures
MIN_NUMBER_OF_BLOCKS = 8       # min number of blocks (IS + OOS)
MIN_SIZE_SINGLE_BLOCK = 0.03   # min size of a single block --> 0.03 = 3% of the total series

THRESHOLD_GPDR = 1.0           # validation threshold
TOLERANCE_GPDR = 0.1           # tolerance --> 0.1 = 10%
STEP_PERCENTILES = 0.1         # decile


GRAPHS = False                 # if True it shows each GSA architecture

#### INPUTS ###
```

Since the strategy on Gold Future was developed with symmetric GSA, the new architectures must adhere to that pattern. It will suffice to change the specific parameter to obtain random segmentations with the IS portion greater than the OOS one. It is asked to generate 1000 GSA architectures with different segmentations, a minimum number of blocks equal to 8 (4 IS + 4 OOS), and a minimum unit block size of 3% of the whole time series. The GPDR validation threshold is set at 1.0, and the tolerance at 10%. Two of the thousand random architectures generated can be seen in Figures 4 and 5. Note how in the first case, the size of each block is 4208 candles, while in the second of 9213 candles. It can be easily seen how the validation in these two situations could lead to different results.

Figure 4: Random generation #1

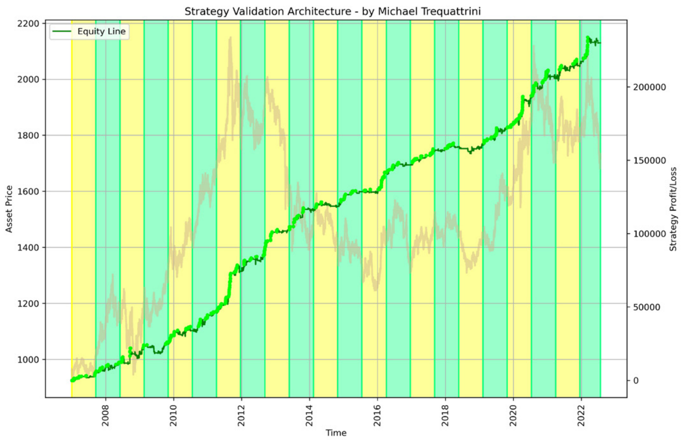
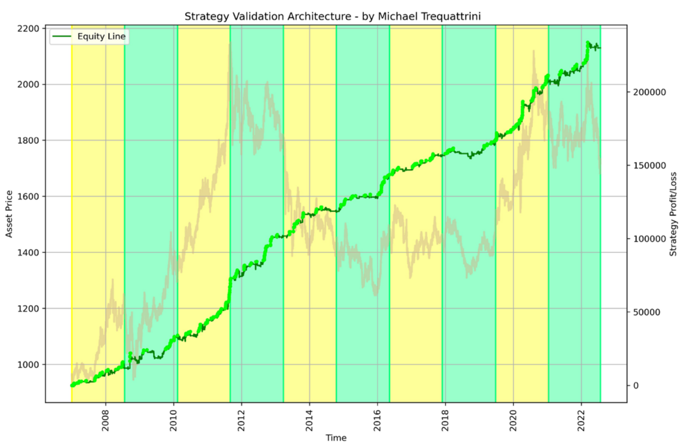


Figure 5: Random generation #2



Results Interpretation

Once the 1000 random architectures have been generated, it is time to evaluate the results. Figure 6 shows the model's output, which provides several indications for an overall assessment of the robustness of the strategy. The key element, the subject of the entire article, is the "Probability of future robustness," which, in this specific case, shows that the trading system passed the validation process 73.4% of the time. As per its name, this same percentage can also be read as the probability of robustness on future, unknown data. Remembering that the model is applied ex-post to a strategy that has already been validated with an initial segmentation, it is sufficient to have a value greater than 50% (if desired, the threshold can be raised, making the model more selective). With the GPDR distribution available, additional information can be obtained by decomposing it into deciles. As in the present case, it is pleasant to find a value greater than 1.0 from the median or previous deciles.

Figure 6: Results

```
Number of shuffles: 1000
Number of validations with GPDR > 1.0 (tolerance: 0.1): 734
Probability of future robustness: 73.4%

Decile 0.0 --> 0.22
Decile 0.1 --> 0.57
Decile 0.2 --> 0.83
Decile 0.3 --> 1.2
Decile 0.4 --> 1.75
Decile 0.5 --> 1.75
Decile 0.6 --> 2.67
Decile 0.7 --> 4.5
Decile 0.8 --> 4.5
Decile 0.9 --> 10.0
Decile 1.0 --> inf
```

Finally, a comparative analysis of the primary metrics produced by the backtest is shown in Figure 7. The mean and standard deviation on the IS and OOS aggregates of all 1000 generated architectures are calculated for each metric. A delta between +-30% ensures acceptable variation.

Figure 7: Comparative analysis

```
In Sample vs Out of Sample Statistics GSA Montecarlo

Average trade (mean):                [IS] 245.73 [OOS] 264.07 -> delta: 7 %
Average trade (std):                 [IS] 26.94 [OOS] 29.65 -> delta: 10 %

Profit factor (mean):                [IS] 1.99 [OOS] 2.0 -> delta: 1 %
Profit factor (std):                 [IS] 0.14 [OOS] 0.14 -> delta: 0 %

Percent Winning Trades (mean):       [IS] 57.08 [OOS] 56.12 -> delta: -2 %
Percent Winning Trades (std):        [IS] 1.5 [OOS] 1.6 -> delta: 7 %

Reward Risk Ratio (mean):            [IS] 1.49 [OOS] 1.56 -> delta: 5 %
Reward Risk Ratio (std):             [IS] 0.08 [OOS] 0.09 -> delta: 12 %

Avg Open Draw Down (mean):           [IS] -1639.59 [OOS] -1701.62 -> delta: 4 %
Avg Open Draw Down (std):            [IS] 343.22 [OOS] 332.81 -> delta: -3 %

Max Open Draw Down (mean):           [IS] -7326.86 [OOS] -7287.71 -> delta: -1 %
Max Open Draw Down (std):            [IS] 903.95 [OOS] 680.48 -> delta: -25 %
```

Noise Addition

As mentioned earlier, the result of the optimizations in the developing phase is conditioned by the initial In Sample and Out of Sample segmentation. The possibility of applying random noise to the strategy parameters was introduced to account for this within the model. In this way, since each segmentation would have led to a diverse optimization, the analysis is more consistent, simulating the likely difference in results. Figure 8 shows the final input panel, with the amount of noise to be applied.

Figure 8: Model inputs

```
#### INPUTS ###

ASYMMETRICAL_ARCHITECTURE = False  # if True it builds asymmetrical architectures (IS > OOS)


ITERATIONS = 1000              # number of random architectures
MIN_NUMBER_OF_BLOCKS = 8       # min number of blocks (IS + OOS)
MIN_SIZE_SINGLE_BLOCK = 0.03  # min size of a single block --> 0.03 = 3% of the total series

THRESHOLD_GPDR = 1.0           # validation threshold
TOLERANCE_GPDR = 0.1           # tolerance --> 0.1 = 10%
STEP_PERCENTILES = 0.1         # decile

NOISE = True                           # if True adds noise to the chosen parameters
PERCENTAGE_NOISE_ADDICTION = 0.2   # random noise to be added to each chosen parameter --> 0.2 = 20% of noise


GRAPHS = False                 # if True it shows each GSA architecture

#### INPUTS ###
```

Figure 9 shows the model's output using the same 1000 segmentations generated earlier with the addition of a 20% random noise added to each strategy parameter. Note how adding different noise to each newly generated segmentation results in a lower validation rate than obtained without noise. It should be noted that no result is more frequent than another, as it depends on how one chooses the parameters during optimization. As an indication, running the model with and without noise is advisable, holding constant the segmentations generated at the first run (using a specific random seed) and verifying that the "Probability of future robustness" is sufficiently high in both cases.

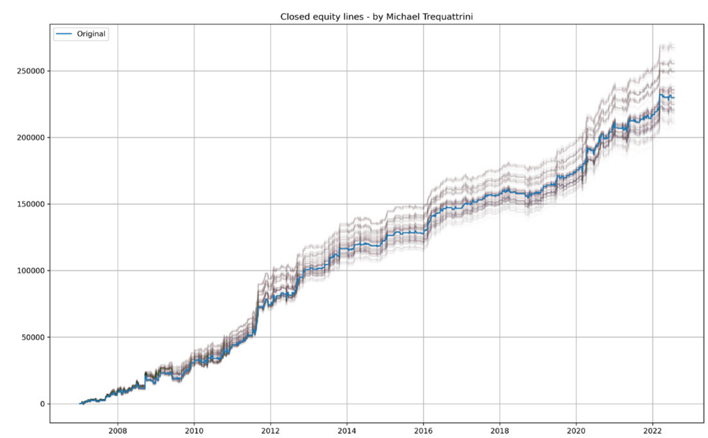
Figure 9: Results with noise

```
Number of shuffles: 1000
Number of validations with GPDR > 1.0 (tolerance: 0.1): 696
Probability of future robustness: 69.6%

Decile 0.0 --> 0.0
Decile 0.1 --> 0.57
Decile 0.2 --> 0.83
Decile 0.3 --> 0.83
Decile 0.4 --> 1.2
Decile 0.5 --> 1.75
Decile 0.6 --> 2.67
Decile 0.7 --> 4.5
Decile 0.8 --> 4.5
Decile 0.9 --> 10.0
Decile 1.0 --> inf
```

Finally, Figure 10 depicts the swarm of equity lines produced by noise addition. The desirable situation is the one with the original equity line located approximately in the middle of the cloud, highlighting a proper optimization process at the developing stage.

Figure 10: Noisy equity lines



Conclusion

The model presented here allows an additional check on the robustness of a strategy already validated with an initial architecture. The possibility of testing the system on different segmentations makes it possible to identify those situations where the validation has been passed due to the choice, voluntary or not, of a specific IS and OOS fractionation. Furthermore, it is fair to point out that although the model proposed here sees only the declination with or without noise, it is possible to imagine many variations on the theme. The goal was to share an idea of analysis that can be customized to help traders evaluate the robustness of a trading system.

The Impact of Machine Learning and Computer Vision to Predict an Opening Range Breakout on E-Mini S&P 500

Edzard Wiener, Ph.D, MFTA, CFTe

Vereinigung Technischer Analysten Deutschlands e.V.
Berlin, Germany

edzard.wiener@gmail.com

Abstract

The motivation for using artificial intelligence to create accurate models for trade optimization and price prediction is increasing and nowadays potentially usable for every trader. Depending on the instrument, the trading hours, and the timeframe, different combinations of technical, financial, macro-economic, and sentiment indicators can be either more effective or less suitable. There is no general consensus about the ideal mix of indicators that are most relevant for predicting price movements. To keep their work simple and clear, traders should focus only on a few. Machine learning is a powerful tool to help traders adapt their indicator setup to the market and improve their strategy. The use of neural networks to trade suitable indicator combinations for specific markets and timeframes has as yet been relatively underexplored. Based on machine learning and "computer vision" approaches, we investigated 19 indicators to identify the optimal combination of indicators for predicting a significant price movement after the opening range of the E-Mini S&P 500 Futures. We also implemented two self-programmed indicators: the intra-bar price movement (IBPM) indicator based on computer vision, and the high-frequency presence (HFP) indicator based on short-term cycle analysis (STCA) using fast Fourier transformation on price and volume data during the opening range. Both indicators are very useful for predicting a significant trending price movement after the opening range period. We argue here that the strategic choice of an indicator combination by an optimized neural network can predict a short-term trend with an accuracy of about 80% during specific market conditions.

Acknowledgements

I could not have embarked on this journey without the staff of the VTAD — the Vereinigung Technischer Analysten Deutschland (the German Association of Technical Analysts) — who generously offered me their knowledge and expertise. I would also be remiss in not mentioning my wonderful wife Berenike and children Marika, Lara, Marius, and Laurenz. Their belief in me has kept my spirits up and motivation high during this process.

Introduction

1.1 The Opening Range

The "opening range" refers to a specific time period at the beginning of a trading session during which the price range of a security is established. It is characterized by a high price volatility and momentum. Opening range can be defined in a time-based way from a few minutes up to two hours, or as pattern-based, in terms of movement up until the first significant price reaction, such as an intra-day micro-trend, a broad or tight channel or a swing. The pattern-based characterization implies a flexible definition that varies each day. The opening range is determined by the highest and lowest prices at which a particular security trades. Traders should pay close attention to the opening range because these can provide valuable information about market sentiment and potential trading opportunities.

The first 30 minutes following the opening of the market is a particularly intense period, as buyers and sellers look for advantageous positions. During this time, incoming orders can lead to considerable price volatility. Such dynamic pricing behavior is often triggered by institutional traders who typically participate in the market with high-volume transactions. Some traders will even wait longer — such as a whole hour — to gauge the balance of power between buyers and sellers. They than develop a trade plan in the direction of a possible breakout.

The idea of analyzing the start of a trading session to gain information about possible price movements during the session was first established by J. Peter Steidlmayer, a trader at the Chicago Board of Trade (CBOT). He developed the concept known as "initial balance" as part of his Market Profile theory in the 1980s, especially in relation to futures trading (Steidlmayer 2002). Based on his observations he noticed that the initial balance can provide important insights into the market's behavior for the rest of the trading day. For example, if the initial balance is skewed to the upside, it might indicate bullish sentiment, while a downside skew might hint at a more bearish sentiment. The high and low of the initial balance can act as support and resistance levels for the rest of the trading day. The size of the initial balance can indicate the potential volatility of the market for the day and if the price moves significantly outside the initial balance range, this could indicate a potential breakout.

Nowadays the definition of the initial balance is interpreted more flexibly. With the recent rapid advances in technology and the growing influence of artificial intelligence, we would like to take a closer look at this specific time period at the beginning of a trading session and analyze it again using a combination of newer indicators and the power of machine learning to optimize trading on a shorter timeframe after initial balance. Motivated by Steidlmayer's approach, the following question arises: is it possible to predict a significant price movement after the initial balance through a suitable selection of indicators and the use of artificial intelligence?

In this study we focused our attention on the very liquid future market, in particular the E-Mini S&P 500 Futures. We used the NYSE opening time of 9:30 a.m., when the market opens and a high volume of trading is coming in. For an intraday trade it is best to wait for the market to stabilize, then determine the trend and try to look for an entry. Opening range breakout patterns can help to find some of the strongest charts for day and swing trading (Calhoun 2015, p 20). In this context we used a time-based definition of the opening range as the first 30 minutes of each trading day, in order to look for successful breakouts and a consequent trending price.

1.2 Technical Indicators

Technical indicators are mathematical formulas used to identify patterns and trends in the market. Frequently they are lagging indicators, but they can help to generate insights into future price movements. Since markets are by their nature never perfectly efficient due to various factors such as information asymmetry, investor irrationality, and transaction costs (Shiller 2003, p 83, Malkiel 1998, p 59), technical analysis and indicators are crucial tools and commonly used, especially for short-term trading strategies. There are many indicators used in technical analysis, such as trend indicators, oscillators, volatility indicators, volume indicators, market breadth, and sentiment indicators (Fang 2014, p 25). Indicators are increasingly being combined over a wide range of timeframes, while a host of instruments and newer approaches are now finding their way into trading, such as social media-based sentiment or time series analysis (TSA), as well as tools for identifying seasonal trends (or seasonality) and shorter-term cyclical patterns (Warren 1984, p 123). Innovative approaches like algorithmic trading based on AI are also increasingly being used by retail traders and can offer several advantages, regardless of the particular trading style preferred.

1.3 Machine Learning — Neural Networks

Neural networks are a subset of machine learning algorithms ostensibly modeled on the workings of the human brain. They have been trained to recognize patterns and interpret data by simulating human cognitive operations (Kumbure 2022, p 21). In the world of trading, neural networks are powerful tools that can be used to predict prices, recognize patterns or optimize portfolios. Building a neural network requires several steps. First, a clear hypothesis needs to be devised and addressed. Then data must be collected and prepared for modeling. This procedure includes data cleaning, primary statistical analyses, normalizing and data splitting into feature and response variables, and subsequently into a training and testing data set. The architecture of the neural network then needs to be designed, based on the hypothesis and the amount and structuring of the data. So-called hyperparameters, such as the number of layers, the number of nodes in each layer, and the activation function also have to be optimized. For the most effective selecting of hyperparameters, the neural network has to be trained and the setup with the lowest validation error should be picked.

During the training phase, in this study the network's predictions were compared to the actual values and the weights and biases adjusted to minimize the difference between predicted and actual values. As this procedure is itself based on the performance of the network, it thus represents an iterative process of optimization. This step requires large amounts of data which are often difficult and expensive to obtain. After this initial training phase, the network is then validated using a smaller test data set. Once the network is up and running, predictions can be made on newer — and as yet unseen — data.

1.4 Computer Vision Algorithm

Computer vision is a method that uses artificial intelligence (AI) to acquire and analyze digital images to extract richly descriptive — or high-dimensional — data from the real world to produce quantifiable, numerical information. The approach attempts to replicate the complexity of human experience (or "vision") with the aim of using its processing of the features of digital images in order to automate tasks that our visual systems typically carry out. In the field of trading, price movements and pattern recognition in real time could be a promising application of this technology. Price movements on each trading platform can be observed and stored like frames from a film. The images are than post-processed to enhance the feature of interest, which can involve resizing the images, converting them to different colors or applying filters. In a next step, structure detection, using segmentation, edge detection or more advanced techniques like convolutional neural networks (CNNs) can also be deployed. Characteristic patterns and typical setups can be extracted and converted into a numerical format. For example, intra-bar price movement (IBPM), defined as the price changes during the formation of a graphically represented candle can be detected and the current rate (or velocity) of price change as well as the current deviation from the opening price can be extracted for further real-time analysis. This process can be complex and may require advanced techniques depending on the quality and complexity of the images (Chen 2016).

Table 1 Technical indicators as feature variables for the neural network

Indicator group	Indicators	Description
Indicators based on COT sentiment (weekly *Commitments of Traders* report)	COT position change: net long [1] or net short [-1	Shows the net long and short positions of commercials and non-commercials. Changes in these positions can indicate shifts in market sentiment
	COT percentage change in relation to previous week	Shows the percentage change in COT positions in relation to the previous week.
Indicators based on bars	30-minute bar of opening range: bullish or bearish	Compares the closing to the opening price of the 30-minute opening range bar: bullish (green bar) or bearish (red bar)
	Bar size of 30-minute opening range bar: large or small	Categorizes the size of the 30-minute opening range bar: large body or small body
	Wick size of 30-minute opening range bar: large or small	Categorizes the size of the wick of the 30-minute opening range bar: large wick or small wick
	Ratio of bullish to bearish bars on one-minute opening rang	Compares the number of one-minute bullish to one-minute bearish bars during opening range.
	Location of 30-minute opening range bar: inside [0], crossing above [1] or crossing below [-1] the previous day's high or low	Indicates whether the 30-minute opening range bar is inside or crossing the previous day's trading range (TR).
Indicators based on volume	Volume profile shape over opening range: D-shape, I-shape, b-shape, P-shape, double distribution, non-significant shape	Indicates the shape of the volume profile over the opening range, which can suggest different market conditions.
	Point of control (POC): above [1] or below [0]	Compare the location of the30-minute opening range bar in relation to the weekly point of control. The POC is the price level with the most trading activity.
	Volume during opening range in relation to a 14-period "lookback" timeframe: over, normal, under average volume	Indicates the trading vol-ume during the opening range in relation to the av-erage volume calculated as the moving average from a 14-period "lookback" timeframe.
Indicators based on VWAP (volume weighted average price)	30-minute bar of opening range in relation to the VWAP line: above [1] or below [-1] VWAP	Compares the location of the 30-minute opening range bar in relation to the VWAP line.

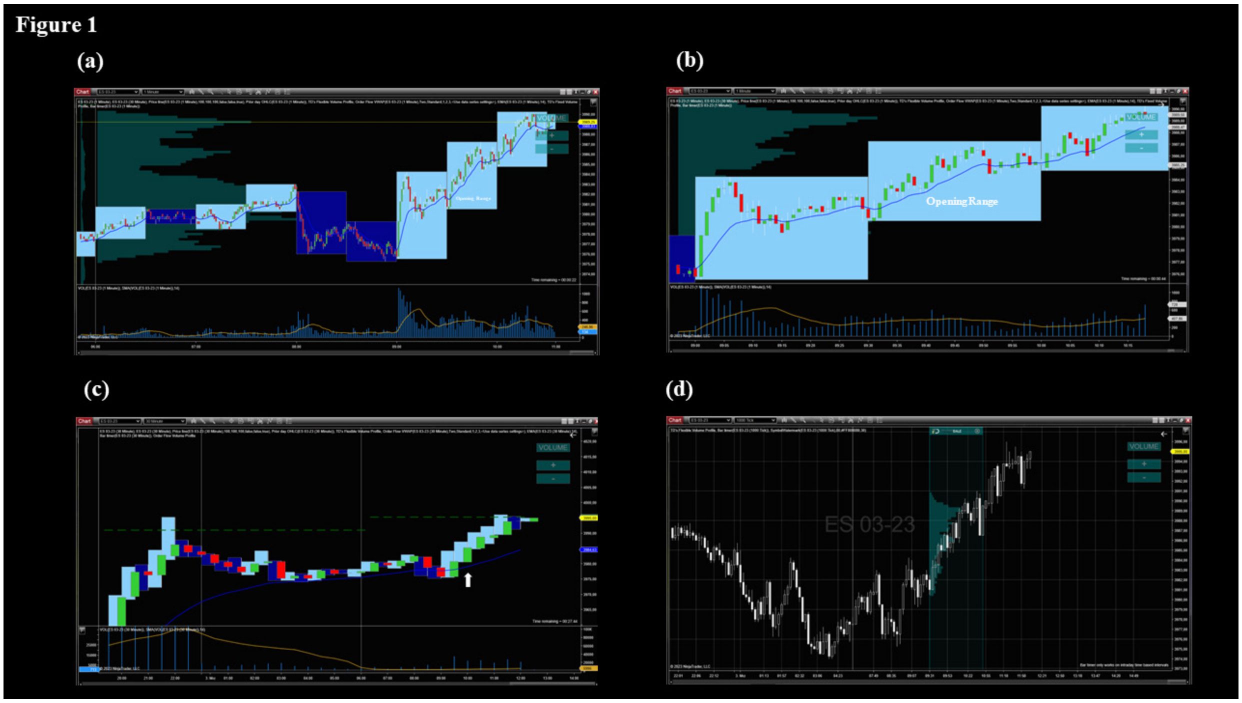
	30-minute bar of opening range in relation to the VWAP band: above [1], below [-1] or inside [0] the first standard deviation VWAP band	Compares the location of the 30-minute opening range bar to the first standard deviation VWAP band.
Indicators based on seasonality	Weekday of opening range [Monday to Friday]	Takes into account the weekday of the opening range based on the assumption of a variety of trading activities.
Indicators based on high-frequency presence (HFP) using fast Fourier transformation on short-term cycles	Presence of high frequency in price data: Spectrogram of price [time *vs.* frequency *vs.* intensity]	Indicates the presence of high frequencies in the price data during the opening range, indicating strong trading activity.
	Presence of high frequency in volume data: Spectrogram of volume [time *vs*. frequency *vs*. intensity]	Indicates the presence of high frequencies in the volume data during the opening range, indicating strong trading activity.
Indicators based on intra-bar price movement (IBPM) using computer vision	Number of high-velocity price movements during formation of a 30-minute bar	Uses computer vision techniques to track the movement of the price line on the chart and to identify impulsive, high-velocity movements.
	Standard deviation of price movements in relation to the opening price of a30-minute bar	Uses computer vision techniques to measure the standard deviation of the actual price line in relation to the opening price of that bar.
Moving average indicators	Price in relation to the EMA 20 line: above [1] or below [-1]	Compares the price to a 20-period exponential moving average (EMA). Prices above the EMA are often seen as bullish, while prices below the EMA are seen as bearish.

Materials and Methods

2.1 Input and Response Variables

The feature variables incorporated into the neural network spanned various types of financial indicator (Table 1) and were plotted on the price or volume chart. For the sake of simplicity, the variables are based either on mathematical calculation and documented as metric variables or visually evaluated on the charts and categorized according to the number of possible values — 1, 0, -1, large or small — mostly in relation to a mean value or to previous values.

Figure 1 Candlesticks Charts from the E-Mini S&P 500 with technical indicators — opening range market movements of the E-Mini S&P 500 Future during the first half-hour of a typical day's trading; Figure 1(a) - (c) shows consecutive bullish candles (in light blue or green), also anticipated by the neural network.



The technical indicators were applied on a 30-minute and one-minute chart (Figure 1). The data were analyzed on a *NinjaTrader* platform (version 8, NinjaTrader Group, LLC) including third party add-ons for flexible volume profile (*Trader Dale*). All data were entered into an Excel spreadsheet for further evaluation. The whole data set consisted of 1260 data cells (21 indicators x 60 days) including a training, testing, and (for prediction) an unseen data set.

The data feed for the futures markets during regular trading hours was obtained by means of *Kinetick* data feed (*Kinetick* market data). The following price values were downloaded: open, high, low, close (often shortened to OHLC), and volume, on a "tick" basis of the E-Mini S&P 500 3-2023 and E-Mini S&P 500 9-2023 future contracts covering the period from January 1 until March 31 (for training and validation), and from the E-Mini S&P 500 September contract from July 1 until July 14 (for prediction). Finally, 19 indicators were used to obtain the input data for the neural network.

Short-term seasonality was considered by documenting the weekday of the daily opening range to consider the impact of recurring price patterns that occur during specific times of the day.

Market sentiment was analyzed using the weekly *Commitment of Traders* report (COT) data released by the U.S. Commodity Futures Trading Commission (CFTC). The report outlines the number of futures contracts held by market participants (hedge funds, banks, producers, and speculators, for example). Two main categories have to be distinguished: commercial market participants deal in the futures market for hedging, and their positions are negatively correlated with the underlying market; and non-commercials, which deal in the future markets for speculative reasons, and whose positions are positively correlated with the underlying market. To indicate shifts in market sentiment, the net positions and the absolute weekly percentage position change of commercials and non-commercials were calculated and entered into the Excel spreadsheet. The net positions were categorized as -1 for a short position and +1 for a long position.

The *trading volume and mean volume* over a 14-period lookback period was calculated and plotted below the price chart (Figure 1). The actual trading volume was categorized in relation to the mean value. If actual trading volume was below average, it was categorized as -1. If actual trading volume was near the mean value, it was categorized as 0 and, if the trading volume was above the mean volume, this was categorized as 1.

The *number of bullish and bearish candles* was counted during the opening range on a one-minute timeframe and the green (bullish) to red (bearish) candle ratio was calculated and documented. This strategy is known as bar counting (Brooks, 2011).

The *relationship between candle body size to wick size* was assessed on a 30-minute timeframe and visually categorized as large, equal or small.

As a short-term trend indicator, *the exponential moving average* with a 20-period lookback period (EMA 20) was used to identify short-term trends. Based on visual inspection of the opening range, the price in relation to the indicator line was documented. If the price was mostly being traded above EMA 20 it was encoded as 1 or, if below, it was entered as -1.

The *weekly volume profile* and its *point of control* (POC) indicating the most traded price over this time period was plotted on the one- and 30-minute charts. The traded price in relation to the weekly POC was evaluated. A price above the POC line was categorized as 1 and below as -1.

In addition, a *flexible volume profile* over the daily opening range was plotted on the one-minute chart and the shape of the volume profile was analyzed visually and classified as either a D-shape, I-shape, b-shape, P-shape, double distribution (DD) or a non-significant shape (non). The shape of the opening range volume profile is useful for gaining market insights and allows a more nuanced grasp of market dynamics, auction processes, and price action, which can all influence the patterns of further daily trading.

The *daily volume-weighted average price* (VWAP) is calculated by multiplying the price of each trade by the number of shares, then summing these values and dividing the result by the total volume. This is a key benchmark for traders, especially for professional traders whose aim is to buy below and sell above. Deviations of the price from VWAP may indicate overbought or oversold conditions. If the price is trading above VWAP the indicator line is represented in green and categorized as 1, and if the price is trading below VWAP it is represented in red and categorized as -1.

Furthermore, the trading price was evaluated in relation to the *first standard deviation band of VWAP*. If the price was trading above the first standard deviation band, it was documented as 1. If the price was trading below the first standard deviation band, it was documented as -1, and in between as 0.

The *previous-day high* (green dashed line) and *previous-day low* (red dashed line) were also plotted on the one-minute and 30-minute chart. If the actual price was above the previous-day high it was notated as 1. If the actual price was below the previous-day low, it was entered as -1 and, if it was in between, this was recorded as 0.

2.2 New Indicators: Intra-Bar Price Movement, High-Frequency Presence

In addition, two proprietary indicators were applied for this study: the intra-bar price movement (IBPM) indicator, deploying computer vision (CV) techniques, and the high-frequency presence (HFP) indicator conducting short-term cycles analysis (STCA) of price and volume data during the opening range. The analysis was carried out within one short timeframe comprising a one-minute chart.

The IBPM indicator is based on computer vision and calculates the speed of price change during candle formation. Therefore, the movement of the price line was recorded in real time from a candlestick chart and refers to the price fluctuations that occur during the predefined period. A price candle represents a specific time period during which the price of a financial instrument may undergo various price movements with different speeds and deviations from the opening price of that candle. The IBPM indicator measures the position of the current trading price relative to the open price. The IBPM could be useful evidence because it provides traders with insights into the conditions underlying very short-term price dynamics and volatility.

The presence of high-speed price movements, defined as a value greater than 250, can be interpreted as an indication of strong trading activity. This suggests the presence of market participants who have the ability to move the price rapidly in a particular direction.

Figure 2 IBPM indicator on price and volume data over opening range (60-minute bar on July 7, 2023)

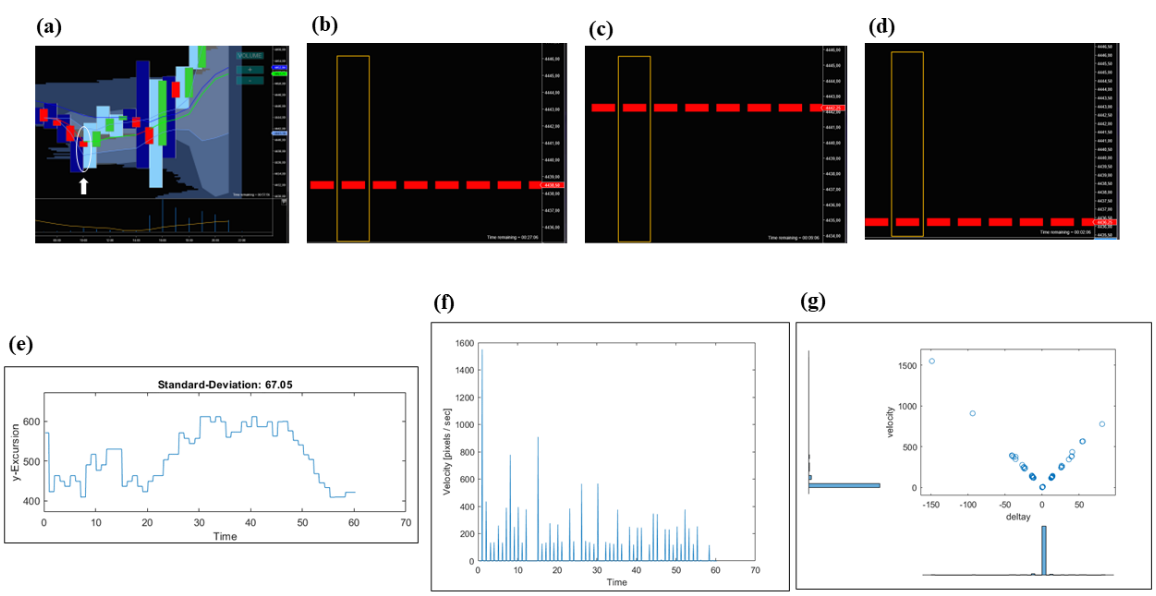
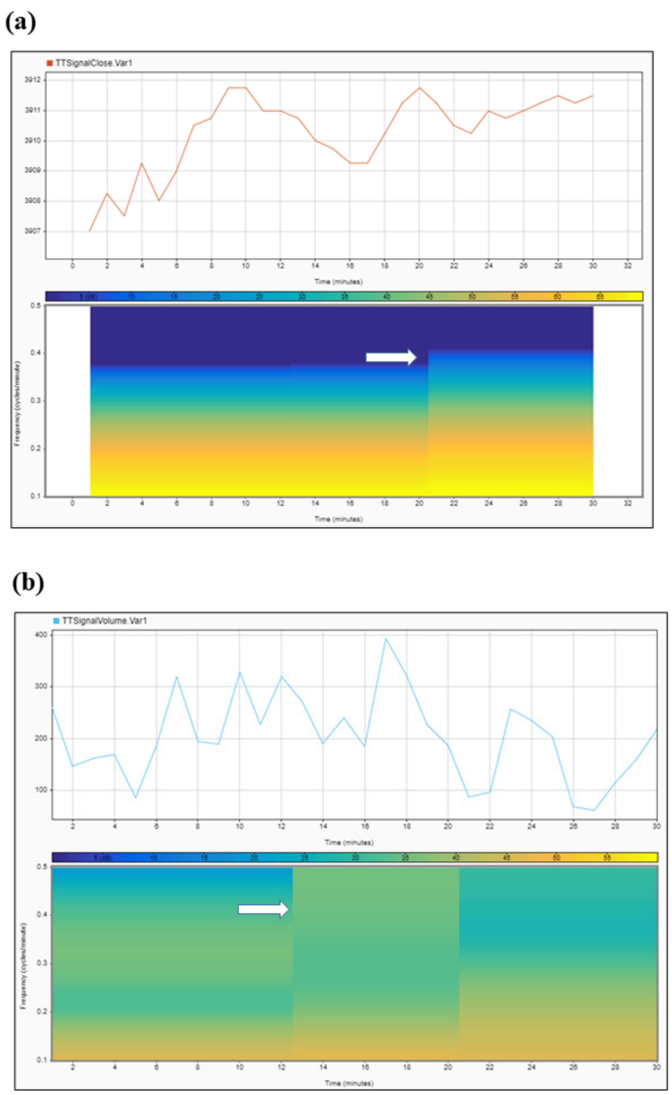


Figure 2(a) depicts a bearish small body candle (red) with a large lower wick. The IBPM indicator identifies the strong trading activity between buyers and seller during the formation of this candle (Figure 2(b-d)). Multiple strong price velocity changes of more than 250 in both directions with large deviations from the opening price (Figure 2(e-f)) can be observed, indicating that buyers try to push the price closer to the opening while the sellers try to stay in control. Figure 2(g) clearly demonstrates that the most frequent location during the candle formation is the opening price corresponding to the zero line. During the strong up and down movements, the zero line is crossed several times until the "closing" of this candle. The candle in this example is a kind of reversed candle with a small body and a large lower wick following three green bullish candles, each with higher closings. This pattern after a reversal or a period of consolidation is also known as the "three white soldiers" and is typically interpreted as a strong bullish signal, indicating that buyers have taken control of the market.

Figure 3 High-frequency presence (HPF) exhibited in spectrograms representing either price (a) or volume (b) data



The basic concept of STCA is to determine whether it is best to perform a trend or a swing trade. STCA over the opening range using fast Fourier transformation (FFT) allows traders to identify the underlying short-term periodicities and cyclical patterns present in price and volume data (Figure 3). FFT reflects the duality between events in the time domain and their representation in the frequency domain. By analyzing the dominant frequencies, traders can identify short-term patterns in the market data. A trader might choose to enter a long position when the dominant frequency suggests an upcoming bullish cycle. By examining the peaks and troughs in the spectrum, traders can determine the frequencies at which the market shows strong activities. For a detailed evaluation, spectrograms of price and volume data were calculated and plotted with time on the horizontal axis, frequency on the vertical axis, and various colors from blue to yellow representing an increasing amplitude or energy of the frequencies.

Figure 3 shows HFP as it appears in the spectrograms, which represent either price data (Figure 3(a)) or volume data (Figure 3(b)) over the opening range. In this spectrogram, the yellow color signifies the presence of high frequencies in the data. This can be interpreted as an increased trading activity, leading to significant price action. During minute 13 to minute 21, high-frequency volume changes are present in the data and, after a slight lag, price action follows during minute 18 to minute 30. This is observable because of the growing number of yellow areas during this time interval.

These indicators serve as input variables — in this case known as feature variables or predictors — and can help the neural network to learn specific patterns to provide valuable information about the financial markets. All indicators have shown predictive value in a similar context. Importantly, while developing the neural network one has to consider multicollinearity, among other indicators.

2.3 Model Training

Initially, the issue under study was whether it would be possible to predict a significant price movement after the start of the regular trading session for a highly liquid futures contract using various technical indicators. All numerical and categorical data, obtained from the technical analysis of both a one-minute and a 30-minute candlestick chart, were documented in an Excel spreadsheet. All data were historical "tick data" — that is, all non-aggregated intra-day prices of a security determined during a trading day — from the March and September 2023 contracts of the E-Mini S&P 500. Overall, 60 opening ranges were analyzed using 19 distinct indicators, as set out in Table 1 above. The data set was then separated into a training and testing subset. The neural network (NN) was created using *MATLAB* Version R2022b, by MathWorks (McCarthy 2018).

The selection of a neural network's architecture is often discretionary and it is not always obvious why it has been designed in that way. In total, 70% of the data are used for training, 10% for validation and 20% for testing. The process of training and validating the neural network constitutes an iterative optimization exercise with varying configurations of hyperparameters. This is conducted under the observation of the network's performance in response to distinct selections of hyperparameters. Afterwards the model's performance is evaluated using the test set.

The training regimen constituted an iterative process with a total of 50 cases. Each case comprised 19 feature variables, which were either of a categorical or a numerical type. Since our interest was to differentiate between trading and trending price, we required only a singular categorical response variable. This variable was classified as 1 to denote a trending price, or 0 to signify a trading price. After training the model, the model's predictive accuracy was assessed by testing it against an additional set of ten cases.

A standard neural network comprises layers of interconnected nodes, often referred to as "neurons". Neural networks can be used for particularly effective handling of high-dimensional and non-linear data and find utility in both regression and classification tasks. A classification task, such as the one examined in the present study, typically involves the prediction of an input's class (in this case, trending or trading price) based on its characteristic features (in this context, the values of various indicators).

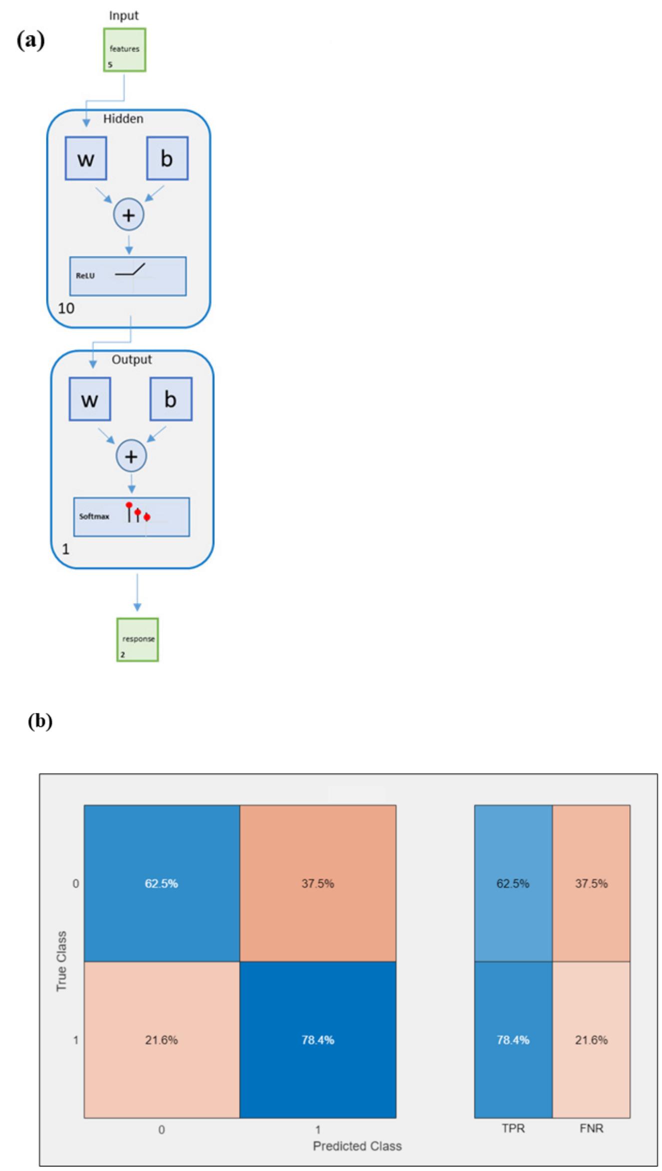
A three-layered neural network typically consists of an input layer, one hidden layer, and an output layer. In the input layer the network takes in the data for processing. The hidden layer transforms the inputs into something that the output layer can use. This is called "hidden" because its inputs and outputs are masked by the activation function and the weights. The output layer produces the final prediction or classification result, whereas a three-layered neural network has an input layer, two hidden layers, and an output layer. The additional hidden layer allows the network to learn more complex representations of the data.

The purpose of an activation function is to introduce non-linearity into the network, allowing it to learn and represent more complex patterns in the data. The activation function for each neuron in the hidden layers is also discretionary. A common choice for this purpose is the rectified linear unit (ReLU) function, which serves to introduce an element of non-linearity. However, when it comes to the neuron in the output layer, to reduce overfitting we used the reserved validation data for control of the training process. If the loss of the validation data stops decreasing, we save the adjusted weights that lead to the lowest loss of the validation data for our model. Classification models are usually evaluated using accuracy, precision, recall, and F1 score. These metrics provide a comprehensive assessment of the model's performance.

Results

3.1 Hyperparameter Setting for the Neural Network

Figure 4 The architecture of the three-layered neural network (a), with its performance percentages plotted on a so-called confusion matrix (b)



In the context of this study, an artificial neural network was assessed for its performance in predicting a trend following a breakout of the opening range, which was defined as the first 30 minutes after the start of the trading session of the E-mini-S&P 500 Future at the New York Stock Exchange opening at 9:30 a.m. Eastern Time. This assessment involved various presets of several hyperparameters, as depicted in Figure 4(a). These hyperparameters included: the number of layers within the network; the size of these layers as determined by the quantity of neurons; a selection of an activation function (sigmoid, hyperbolic tangent, rectified linear unit); the count of feature variables as input into the model; and the strength of regularization, represented by the parameter lambda (β).

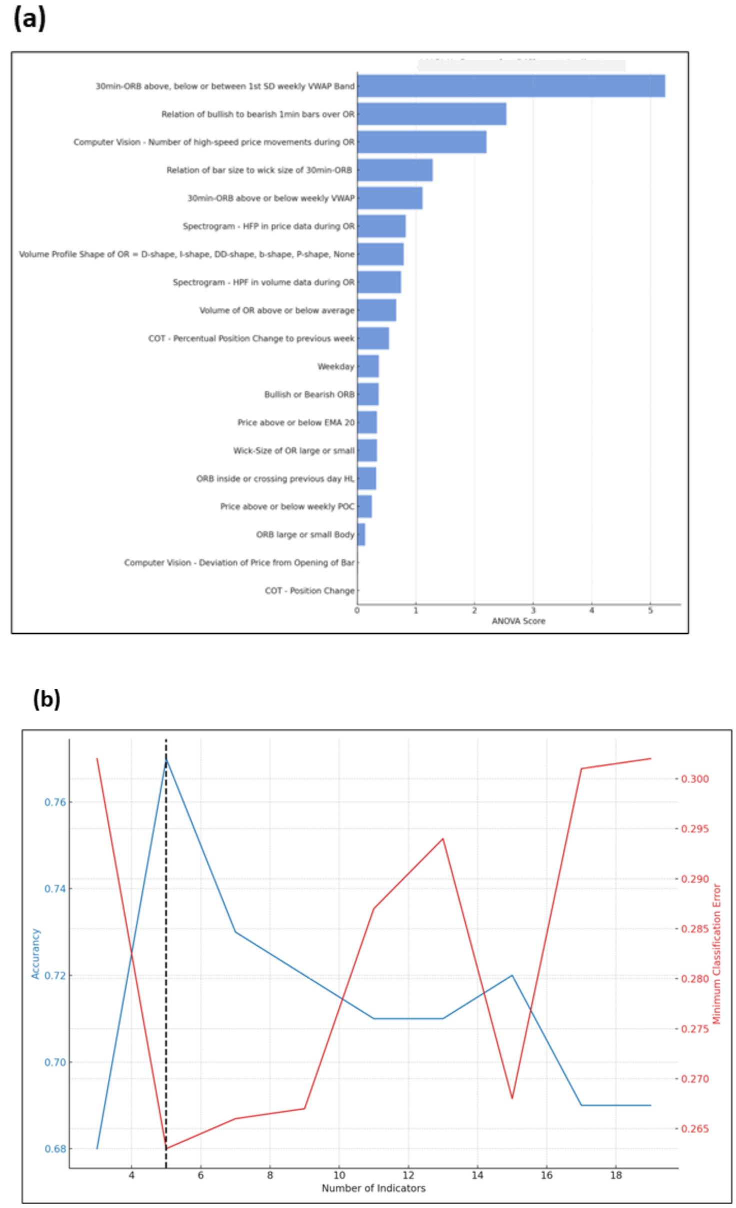
In the context of neural network architecture, especially one featuring an increasing size of layers and nodes, the network will be able to model more complex relationships in the data. Generally, more neurons mean more dimensions into which the training data can be separated. After the dimensions have expanded, the data need to be condensed back into one dimension because the output — in this case, future price — is expressed along a single dimension.

A binary value (0 or 1) was chosen, based on whether the price after the opening range was trending, encoded as 1, or whether there occurred trading in a range (non-trending), in which case this was encoded as 0. However, with an increasing number of layers and nodes a binary value might also be more likely to overfit the model. Overfitting refers to a situation whereby a model has been trained too well and starts to perform poorly, and can occur when the model becomes too complex or when the training data are insufficient. The model starts to memorize noise or specific patterns which do not represent genuine underlying patterns in the data. Signs of overfitting are high accuracy during the training phase of a model but low-quality performance on new data. To avoid overfitting and to simplify the model, it is advisable to increase the amount of data by applying regularization techniques, and to shorten or even curtail the training process when the model's performance is unstable or oscillating.

For our model a three-layered (input, hidden, output) neural network architecture has proven to be the most effective solution, with the input layer consisting of 19 neurons and then the hidden layer of ten neurons. For the output layer, one neuron was found to be the best setup. In a first step, input data were standardized. The best activation function for achieving this was the ReLU function $f(x) = \max(0, x)$ with the input x. If x is positive, the function returns x, and if the input is negative, the function returns 0. $\lambda = 0$ was the best value for the regularization strength. A larger value of λ results in a stronger regularization effect and encourages the model to have smaller weights. The optimal value achieves the best balance between model complexity and generalization and depends on the specific problem and the data set.

The confusion matrix, the true positive rates (TPRs) and the false negative rates (FNRs) of the validated model are presented in Figure 4(b). Each row of the confusion matrix represents the actual class and each column corresponds to the predicted class.

Figure 5 ANOVA scores of different indicators (a) and accuracy and minimum classification error (b) depending on the number of indicators



To further refine the neural network, the number of input variables was evaluated. In a first step, the 19 feature variables listed in Table 1 were used as input. The input variables were then successively reduced by two. The importance of each feature variable was evaluated and ranked using an analysis of variance (ANOVA) statistical technique (Figure 5(a)). To reduce the input variables from 19 to five further improved the performance of the model (Figure 5). The five variables with the greatest input were, in descending order: price in relation to the first standard deviation band of VWAP (ANOVA = 5.25) and in relation to the VWAP line (ANOVA = 1.11); the relationship between bullish and bearish one-minute candles over the opening range (ANOVA = 2.54); the number of high-speed price movements during the opening range analyzed with the intra-bar price movement (IBPM) indicator (ANOVA = 2.21); the relationship between candle size to wick size within the 30-minute opening range "candle" (ANOVA = 1.29); and the high-frequency presence (HFP) in price data during the opening range (ANOVA = 0.83).

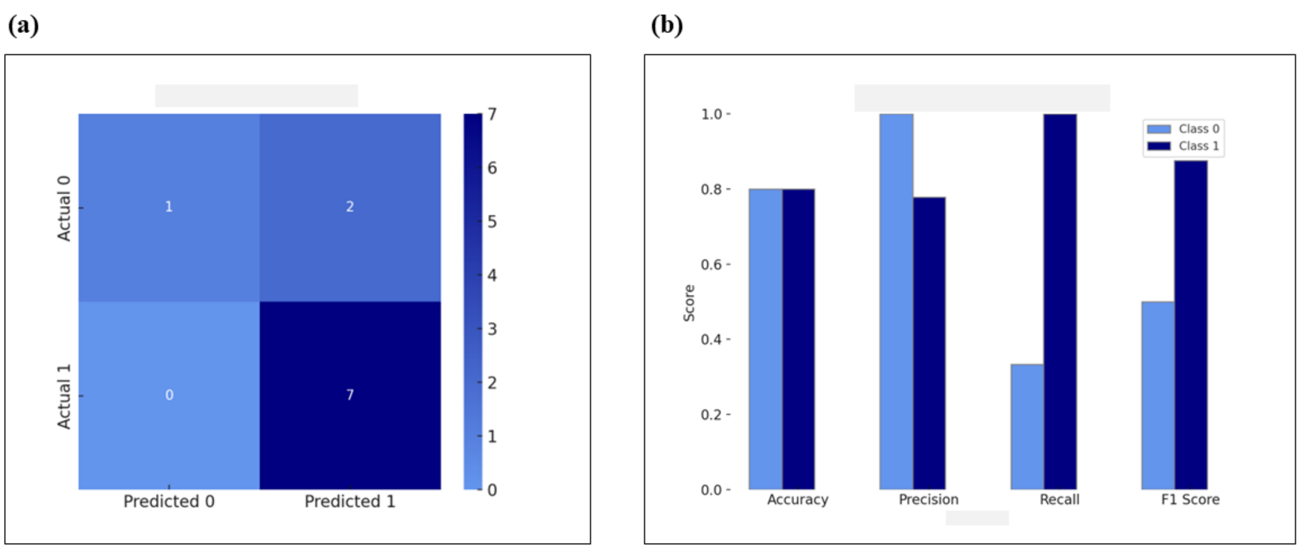
Out of the 19 indicators featured in this study, both indicators we had developed — IBPM and HFP — were included in the final set of the top five indicators for predicting a trend.

The type of indicator best suited for the issue also depends on the amount of data, the trading instrument, timeframe, and trading hours. Figure 5(b) indicates the increase of model performance in relation to the number of indicators. A reduction of indicators leads to a more closely fitting (although not overfitting) model. The number of indicators was reduced successively by two and the accuracy (ACC) and the minimum classification error (MCE) were plotted. As illustrated in Figure 5(b), the selection of five indicators exhibits superior performance with the highest ACC and the lowest MCE. Despite the relatively small amount of data, the accuracy of the model was quite high.

3.2 Prediction of the Model

Once the model had been trained, validated, and finalized, it was used to predict a significant price move of the actual E-mini S&P 500 Future September 2023 contract (see Table 2 in the Appendix). This topic is a typical classification task for supervised learning when it comes to predicting the categorical class labels of new instances based on past observations, with the aim of assigning an input to one of the given classes. The network has now learnt to recognize patterns in the input data that correlate with certain labels. The network does this by iteratively adjusting its internal parameters to minimize the difference between its predictions and the actual labels (a process known as back propagation). For a binary classification problem, where there are only two possible classes — trending (1) or non-trending (0) — a single output neuron can be used that would generate a value of 0 or 1. A *softmax* function can be used to convert the output into a probability distribution over the classes.

Figure 6 Predictions of the finalized neural network, showing confusion matrix with differences between actual and predicted values (a) and the key metrics of the neural network for predictions (b)



The confusion matrix in Figure 6(a) shows the performance of the classification model to predict unseen data. Based on the four values in the confusion matrix (true positive, false negative, false positive, and true negative), various key metrics were calculated to assess the performance of the neural network, such as accuracy, precision, recall (sensitivity), and F1 score. For the non-trending class 0 the precision is 100% indicating that, for all instances classified as non-trending, these are true negative predictions. The recall is approximately 33%, indicating that the model correctly captures 33% of all non-trending price moves. The F1 score is about 50%, representing the harmonic mean of precision and recall. For the trending price move encoded as class 1 the precision is around 78%, the recall is 100%, and the F1 score is about 88%. The overall accuracy of the finalized neural network is approximately 80%, which suggests that it correctly predicted a trend or a trading range after the breakout of the opening range of the E-mini S&P 500 Future in about 80% of cases. The modeled network has a better recall and F1 score for a trend prediction compared to a trading range.

Discussion

4.1 Artificial Intelligence for Optimal Indicator Setup

Technical analysis has a long history of widespread use by market participants with a large body of academic evidence including theoretical support and empirical evidence (Rodolfo 2017, p 115). This study explored the application of a neural network for predicting a price move after the first 30 minutes at the start of the trading session for the E-mini S&P 500 Future. Previous studies have mainly investigated the prediction of stock prices or ways of optimizing portfolios. By evaluating historical data, we aim to analyze the performance of the models in capturing trends and patterns, thereby providing insights into the potential use of such models in making short-term trading decisions. Various steps were involved in the analysis, including defining an indicator combination comprising feature variables, as well as carrying out data preprocessing, model construction, training, evaluation, and visualization.

In today's fast-paced financial markets, intra-day trading has emerged as a popular strategy for those seeking short-term gains. With the rapid advances in technology and the growing influence of artificial intelligence, traders are using the power of machine learning more and more frequently to optimize their trading. In this fast-growing field, continuous studies need to be performed with the aim of identifying useful applications. We have shown that the fusion of sophisticated trading techniques based on technical indicators in combination with the competitive edge provided by machine learning algorithms can be used to find simplified and clear trading strategies. Some may argue that AI is unnecessary for trend identification since any expert can observe a trend. This may be true to an extent but human behaviors like risk aversion during success and risk-seeking during loss often interfere with trading. AI can help to identify potential risks that are difficult for human analysts to be aware of because indicators are mostly selected for the purposes of supporting someone's initial hypothesis. Nowadays the ease of access to information and the increasing use of new technologies and alternative data sources like social media make the selection of pertinent information an ever more complex task — especially for retail traders.

But gaining an edge is paramount for successful trading. This is where AI can help the trader to select the most relevant information and the best indicator setup quickly and efficiently. We have demonstrated how to use a neural network to select the best indicator setup from a large variety of possible indicators for a specific market constellation, instrument, and trading hour. The neural network tested here can be trained to recognize patterns in a large amount of financial data, to run complex mathematical operations, and to help inform buy-or-sell decisions in real time. Furthermore, it can operate 24/7.

But there are also several potential drawbacks like over-optimization, the risk of overtrading, loss of control, and limited learning. If AI ends up informing everyone's trading decisions for them, then the market will drastically change. Further, if an AI service is hacked, directing traders to adopt specific behaviors, it could result in manipulation of the market. As AI is continually evolving while its accuracy will necessarily be less than perfect, traders should be aware of this and always rely on their own judgment when making trading decisions. Neural networks are based on historical data that may or may not hold true in the future. Therefore, predictions generated by AI tools should always be used in combination with other trading and risk management strategies. Trading relies on speculation, diversity, and differences of opinion. So AI should at the moment be viewed as an aid that could make trading more efficient regardless of what trading style is preferred, whether completely discretionary, semi-automated or fully automated. We hope we have demonstrated here that reducing the number of indicators to specific feature variables can enhance a neural network's performance. This observation aligns with the conventional advice to observe the tenet of simplicity when trading. As evidenced by the model's output, the utility of selecting various types of indicators is essential to avoid multicollinearity of indicators. Multicollinearity can lead to unreliable trading signals because the redundant information provided by highly correlated indicators can amplify the effect of noise and lead to an increase in the loss of interpretability. The ranking of both the IBPM and the HFP indicators as among the top-five performers in the ANOVA analysis supports the argument that significant price movements are often driven by big market players and high trading activity, especially after the opening of the trading session.

As previously argued, hyperparameter optimization (HPO) is another important step for enhancing models — even with a relatively scant amount of data. Other suggestions for improving a model's performance are: increasing the amount of training data; adding more input variables; researching more advanced models that can potentially capture more complex patterns in the data; bringing feature engineering expertise to transform the raw data into more complex data by applying mathematical calculations; and resampling the data along different timeframes.

In this study we introduced some methods for helping in trend detection on a one-minute and 30-minute chart from each trading day's data of the E-mini S&P 500 Futures, combined with an appropriate indicator configuration.

One limitation of this study is the small amount of data available. While we achieved still a good accuracy in our model predictions, it is important to consider that a larger dataset might uncover different patterns or exceptions that our current dataset did not capture. Additionally, a small dataset might lead to overfitting. We did not conduct back testing on our neural network model in this study, therefore we cannot accurately measure the real-world performance of the model, especially considering transaction costs, slippage, and other market microstructure effects. Our model utilizes binary feature variables, quantitative feature variables might offer more information and subsequently improve model accuracy and interpretation. On the other hand, binary representation, although simplifying the decision-making process. Financial markets are influenced by a myriad of external factors, ranging from geopolitical events to economic indicators. Our model might not have taken into account all of these externalities, which can significantly influence trading outcomes.

4.2 Practical Implications for Trading

Our research combined three themes — technical indicators, machine learning, and trading strategy. There are numerous trading ideas and algorithms for traders and strategy developers that combine classic indicators with new concepts, and this could represent a veritable goldmine for future trading activities. In today's world, researching and testing fresh strategies, indicators, and trading setups is important for continually improving trading and for adapting to new market structures and dynamics. Simplicity is the key to better real-time performance. Modern testing approaches using machine learning and computer vision represent one possible avenue for identifying and simplifying deployable strategies for traders tailored to particular trading styles. AI trading offers the capacity to process large amounts of data including multiple indicators, news feeds, and historical price data to identify trading opportunities. The ability to process and interpret large data sets quickly could eventually lead to more informed and data-driven trading decisions. For short-term trades, the innovative approach presented in this study offers great potential for enhancing and refining trading strategies. A neural network is a powerful tool for helping to select an optimal indicator setup for a specific market condition and can be highly effective in optimizing new trading strategies for the future.

Summary

5.1 Conclusion and Future Work

In the context of this research study, our focus has been on evaluating the optimal indicator setup for carrying out trades on the very liquid futures markets of the S&P 500 and to predict a short trend after the first 30-minute "candles" of a typical trading day. With our two self-developed indicators and by reducing the 19 initially chosen indicators to the five most important for this special market constellation, we were able to achieve an accuracy of approximately 80% using an optimized neural network. This result demonstrates how effectively classic trading strategies can be combined with AI to gain a competitive edge.

While physical phenomena are often foreseeable and follow recognized formulas, financial data are much more unpredictable but nonetheless amenable to modeling. But, as the British statistician George Box noted 1976, "All models are wrong, but some are useful." Our intention for this work was to combine — and thereby enhance — more traditional trading strategies with the aid of neural networks. In due course the authors hope to build more accurate models based on additional data, alternative indicator configurations, the use of different financial instruments and trading timeframes, and more accurate back-testing. AI can help us select the optimal indicator combination for specific market conditions influenced by the actual economic environment, the vagaries of market sentiment, and broader current events. It is a promising approach to further develop trading because, in trading, personal experience often yields insights that raw data could overlook. Traders should therefore always take into account both personal experience and data-driven models for optimal decision-making.

References

1. Chan, Ernest P. *Algorithmic Trading: Winning Strategies and Their Rationale.* 1st ed., Wiley, 2013.
2. Kumbure, Mahinda Mailagaha, et al. "Machine learning techniques and data for stock market forecasting: A literature review." *Expert Systems With Applications* 197 (2022): 1-42.
3. McCarthy, Edward. *Foundations of Computational Finance with MATLAB.* 1st ed., Wiley, 2018.
4. Steidlmayer, J. Peter, and Steven B. Hawkins. *Steidlmayer on Markets: Trading with Market Profile.* 2nd ed., Wiley, 2002.
5. Warren, Anthony W., and Jack K. Hutson. "Cycle Analysis: Maximum Entropy Optimization." *Technical Analysis of Stocks & Commodities* 2:4 (1984): 123–129.
6. Calhoun, Ken. "Trading First-Hour Breakouts." *Stocks & Commodities* 33:12 (2015): 20–23.
7. Shiller, Robert J. "From Efficient Markets to Behavioral Finance." *Journal of Economic Perspectives* 17:1 (2003): 83–104.
8. Malkiel, Burton G. "The Efficient Market Hypothesis and Its Critics." *Journal of Economic Perspectives* 17:1 (2003a): 59–82.
9. Fang, J, Y Qin, and B Jacobsen. "Technical market indicators: An overview." *Journal of Behavioral and Experimental Finance* 4 (2014): 25–56.
10. Chen, C.H. *Handbook of Pattern Recognition And Computer Vision.* 5th ed., World Scientific, 2016.
11. Brooks, Al. *Trading Price Action Trends: Technical Analysis of Price Charts Bar by Bar for the Serious Trader.* Wiley, 2011.
12. Nazário, Rodolfo Toribio Farias, et al. "A literature review of technical analysis on stock markets." *The Quarterly Review of Economics and Finance* 66 (2017): 115–126.

Endnotes

U.S. Commodity Futures Trading Commission (CFTC) https://www.cftc.gov/MarketReports/CommitmentsofTraders/index.htm

Appendix

Table 2 Technical analysis of opening ranges of the E-mini S&P 500 Future contract for September 2023; data used for prediction of the finalized neural network

COT Re-port: per-centage change to determine market sen-timent [weekly	0.17	0.12	0.29	0.08	0.22	0.44	0.16	0.22	0.16	0.44
Candle opening range: bullish or bearish [30min]	bullish	bullish	bearish	bearish	bullish	bullish	bearish	bullish	bullish	bullish
Body-size candle open-ing range: large body or small body [30min]	small	large	small	small	large	large	large	large	small	large
Relation of body size to wick size of the candle opening range [30min]	-1	1	-1	-1	1	1	1	0	-1	0
Ratio number of bullish to bearish candles over opening range [1min]	1.7	0.94	0.7	1.14	1.14	1.0	0.66	1.3	1.72	1.14
Candle opening range: between [0], crossing above [1] or crossing below [-1] on previous-day high or low [30min]	0	0	-1	-1	0	-1	0	1	0	0
Price trading: above [1] or below [-1] EMA 20	1	1	-1	-1	-1	1	-1	1	1	1
Volume profile shape over opening range: D-shape, I-shape, b-shape, P-shape, double distribution (DD), non-signifi-cant shape (non)	DD	b	D	D	D	b	D	D	D	DD

Candle opening range [30min]: above [1] or below [-1] VWAP line [weekly]	1	1	-1	-1	-1	1	-1	1	1	1
Candle opening range [30min]: above [1] , below [-1] or within [0] one SD VWAP band [weekly]	1	1	0	-1	0	0	0	1	0	0
Candle opening range [30min]: above [1] or below [0] POC [weekly]	1	1	-1	-1	-1	-1	1	-1	-1	-1
Volume opening range in relation to average volume: large [1] = over average; normal [0] = near average; small [-1] = under average	0	0	0	1	0	0	0	1	1	0
FFT short-term cycle high-frequency presence (HFP) in price spectrogram [cycles/min]	0.3	0.3	0.35	0.3	0.35	0.25	0.33	0.3	0.35	0.28
FFT short-term cycle high frequency presence (HFP) in volume spectrogram [cycles/min]	0.4	0.5	0.5	0.3	0.45	0.25	0.5	0.45	0.45	0.35
Computer vision velocity-price tracking of the number of impulsive high-velocity movements [30 min]	14	17	8	9	12	13	17	14	20	13

The 5% Canary
NAAIM Founders

Andrew Thrasher, CMT

Financial Enhancement Group LLC
Noblesville, Indiana USA

athrasher@yourlifeafterwork.com

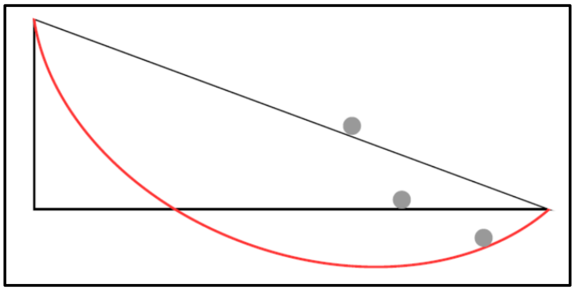
Abstract
This paper will show that the amount of time taken for the S&P 500 or Dow Jones Industrial Average to decline by 5% from a 52-week high illustrates significant insight to the subsequent trend in price. A study on price history for two major equity indices is conducted to show a relationship between duration of an initial decline and the potential for further material weakness or the opportunity for the index to reverse and move higher. With the examination of an 18th century mathematical challenge, a simple lens into the market is shown and evaluated, resulting in a tool that retail and professional investors may apply to markets in the pursuit of capital preservation and appreciation.

Introduction
Buy the dip or prepare for further decline? That's the question presented to investors throughout each trading year as the equity market ebbs and flows, experiencing periods of volatility and persistent trends. Since 1950, the S&P 500 has spent 90% of the time in some degree of drawdown.[1] The proper management of risk is one of the cornerstones of being a successful investor in the long-term, and to do so, a suitable discipline of buy and sell protocols must be in place. This paper focuses on the initial 5% declines from 52-week highs in the price history for the S&P 500 and Dow Jones Industrial Average to determine if the speed at which this decline occurs has predictive value in the ensuing price direction for the equity market. The topic will be divided into two sections: first, the evaluation of the implications of an equity index declining by 5% within a specific period of time; and second, the possible bullish opportunities that may result when these 5% declines occur over a protracted number of trading days. Findings will show that the measure of time it takes the noted equity indices to decline by 5% does have meaningful value in both downside mitigation as well as in upside capture potential.

Literature Review

Early Speed
The initial concept of evaluating the possible risks of a market downturn based on the acceleration of the first signs of market weakness stemmed from this author's research on the brachistochrone curve. In 1696, Johann Bernoulli posed a challenge to the math community that sought to discover the shape of the path an object must travel when being pulled by gravity to travel from point A to point B in the least amount of time. While a straight line may be the *shortest path* between two points, Bernoulli's proposed question and the subsequent answer showed it not to be the quickest.[2] While it took Bernoulli much longer than Isaac Newton, who anonymously submitted his solution to the challenge, Bernoulli applied the findings of Pierre de Fermat who uncovered the principle of least time based on his work calculating refraction angles and the trajectory light takes when traveling between two points.[2] When working on the solution, Bernoulli recognize the most efficient path used the differential equation for a cycloid, which is simply the arc shape that would be traced if one put a pin or writing instrument in the exact middle of a ball or wheel (any circular object) and rolled it on an level surface.[3] In Figure 1, a depiction of the cycloid-shaped curve of the brachistochrone curve (illustrated by the red line), we can see how the dramatic drop and arc of the curved path allows an object to travel at an initial high rate of speed without sacrificing a significant distance, while an identical object traveling on the course of the straight line moves at a slower pace. This application of a cycloid in a brachistochrone curve has been used in many areas outside of academia, including extreme sports like surfing[4] and skiing.[5] The idea of "early speed" in the movement of an object, as adapted from the brachistochrone curve, will be the primary focus of this paper and its application to the equity market. This concept will be drawn upon to help answer the question of whether the initial amount of time it takes price to decline has a relationship with the resulting drawdown experienced by an equity index.

Figure 1: The Brachistochrone Curve



Source: Shi & Celik, 2017

Investor Sensitivity to Losses
Evaluating the movement in price activity of the cited indices as they experience early signs of drawdown, an investor must evaluate if action needs to be taken to protect from further

[1] Based on closing value of the S&P 500 from January 1st, 1950 through September 30, 2022.

[2] See Herrera, 1994

[3] See 3Blue1Brown, 2016

[4] See Henry & Watt, 1998

[5] See Jennings, 2017

declining market prices or if an opportunity is being presented for the index to reverse its decline and begin moving higher. In their paper on Prospect Theory, Kahneman and Tversky showed the experience of loss is twice as great as the pleasure derived from gain.[6] Ultimately this impacts an investor's ability to unemotionally make decisions in regard to avoiding further loss and seeking capital appreciation opportunities.

Because of the overt sensitivity to losses on investors, larger market moves can draw a greater degree of attention as observed by Kaminski and Lo, "in particular, in the event of a significant drop in aggregate stock prices, investors who are generally passive will become motivated to trade because mounting losses will cause them to pay attention when they ordinarily would not."[7] Kaminski and Lo go on to contend that this increase in attention can enhance the irrationality of market participants and exacerbate the downside market move. This focus on attention-driven behavior also impacts buying decisions to a greater degree for retail investors. Barber and Odean note, "individual investors display attention-driven buying behavior. They are net buyers on high-volume days, following both extremely negative and extremely positive one-day returns, and when stocks are in the news. Attention-driven buying is similar for large capitalization stocks and for small stocks."[8]

This paper will address this emotionally driven impediment to buying in Section Two, applying a data dependent lens to buying opportunities presented after protracted periods of price weakness. The emotional emphasis on attention-grabbing data and media is also expanded with the increase in available resources to investors. However, there seems to be a stronger reliance on external resources for broad market viewpoints compared to stock-specific opinions. Peng and Xiong argue that investors put a greater degree of focus on broad market and sector-level information inputs than on individual stock-related data points.[9] This goes to the importance of properly evaluating the price movement of the broad equity market indices in a more pragmatic approach.

Not only do investors over-react to large downside market events, but they also expect them to occur more often than history would suggest. A survey of high-net-worth and institutional investors from 1980 to 2004 found that accredited investors placed a 19% probability of a 1929 or October 1987-type of crash occurring again. Instead, from 1925 through 2015 the market has experienced crash-like declines just 1.7% of the time.[10] While periods of major market weakness are not abnormal, they do not occur nearly as often as many investors believe they will. This irrational belief can be corrected by having a methodology for evaluating market declines early on in their potential developing process. The approach that is discussed in this paper is one such method, which may provide enough advance warning to make proper investment decisions to escape the possible pain from losses Kahneman and Tversky have shown investors are so eager to avoid.

[6]See Kahneman & Tversky, 1979

[7]See Kaminski & Lo, 2014

[8]See Barber & Odean, 2008

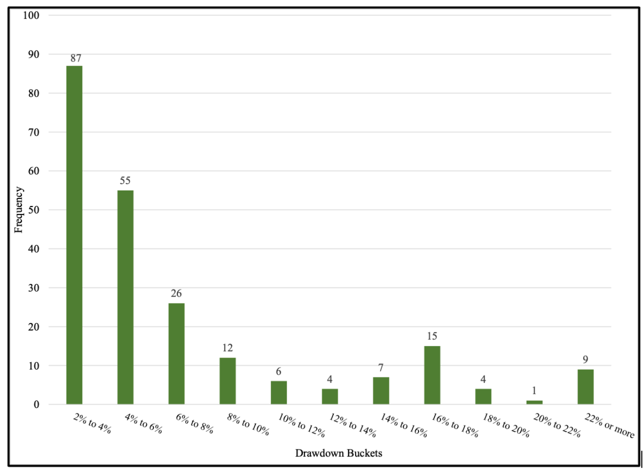
[9]See Peng & Xiong, 2006

[10]See Goetzmann, Dasol & Shiller, 2016

Drawdowns & Market Timing

As previously stated, the S&P 500 has been in some form of decline during 90% of its trading days. While we can only know the final degree of decline after it has occurred, historically most down trends experienced by the S&P 500 have been between 2% and 6%. In Figure 2 we can see that when declines that exceed 2% on a closing basis, 63% of drawdowns in the index are in the noted range.[11] While 2% to 6% may well be within the acceptable risk range for most investors, it's the declines that move into the double-digit category that cause many investors to become uncomfortable and begin to react irrationally. This paper will present a methodology investor can use to evaluate if the more common 5% declines in the S&P 500 and Dow Jones Industrial Average are likely to continue lower and threaten the risk of a more severe drawdown or offer a mean-reversion bullish opportunistic trade.

Figure 2: Number of Declines By Percentage Group in the S&P 500



Percentage declines in price for the S&P 500 based on closing values from January 1950 to September 2022 grouped into eleven buckets. Declines less than 2% were excluded for scaling purposes.

The practice of market timing often involves the use of technical analysis — using the data provided by changes in price and volume to make investment decisions in the active pursuit of capital appreciation and/or capital preservation. Market timing is often viewed as the third rail of the financial industry, drawing strong views from both adamant supporters and hardened critics. One of the most common arguments against market timing is the belief that potentially missing the strongest days of market performance will have a larger cost to the investor than what could be gained by avoiding the largest declines. This belief was disproven in Antoons (2016) with a review of market history from 1961 through 2015, which found the annualized return for missing the 25 best days in the S&P 500 to be 5.74% and avoiding the worst 25 days produced a return of 15.27%. However, if an investor were to miss both the best and worst days, Antoons notes the annualized return would have been 10.94%, still exceeding that of a buy and hold investor of 9.87%.[12]

[11] During the evaluated period, 609 declines of less than 2% occurred and are outside the scope of focus for this paper.

[12]See Antoons, 2016

The perception that markets are entirely random has also been shown to not be accurate.[13] If there is fruit to be harvested by active involvement in financial markets through the use of timing opportune and inopportune periods of investment, then certain methods are likely to be shown to be more qualified than others. Through the extensive testing of tools and lookback periods, technical analysis has been shown to be an appropriate form of market research. Brock, Lakonishok and LeBaron conducted a review of market data from 1897 to 1986, using price history of the Dow Jones Industrial Average. They tested trading signals generated from various moving averages and trading range breakouts, concluding that "overall, our results provide strong support for the technical strategies that we explored. The returns obtained from buy (sell) signals are not likely to be generated by the four popular null models. Consistently, buy (sell) signals generate returns which are higher (or lower) than "normal" returns."[14]

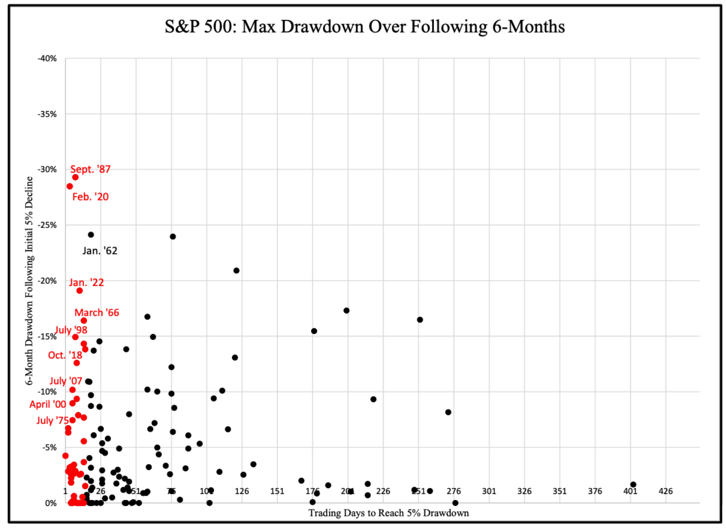
Section One: The Canary Decline

Shown in Figure 2, the majority of the substantive declines experienced in the S&P 500 are between 2% and 6%. In this study, I will be narrowing the focus specifically on closing price declines of 5%. As legendary trader and trend follower, Ed Seykota was known for saying, "The trend is your friend except at the end when it bends." The bend in the trend and how quickly it occurs is the focus of Section One. The initial 5% downward move will be classified as the initial "bend in the trend" to be studied in determining if the trend truly has ended.

While all investors have unique and varying degrees of risk tolerance and time horizons, the data used throughout this paper will be the closing daily values of a stated index obtained from Optuma Software. The two primary indices used in the study will be the S&P 500 and Dow Jones Industrial Average. The S&P 500 is often believed to be the most widely tracked index and viewed as a broad barometer of the U.S. financial market. The Dow Jones Industrial Average, while a smaller sampling of individual securities, has a long history that covers multiple market cycles.

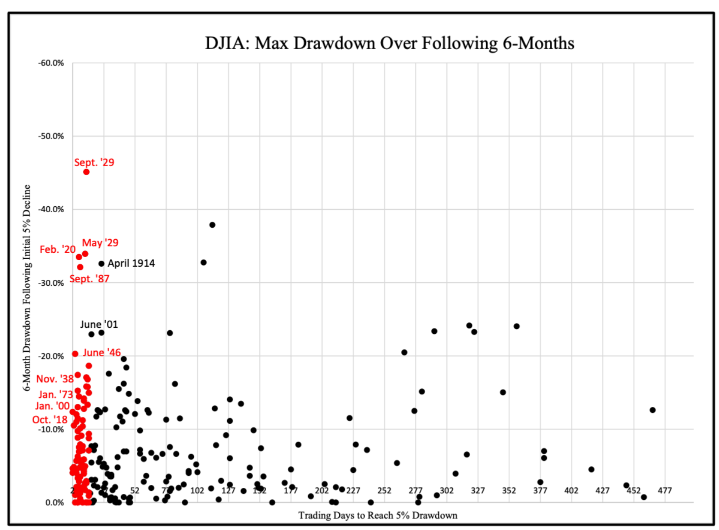
Applying the broad concept of the brachistochrone curve and the insights uncovered by Newton and Bernoulli, which showed that the fastest path between two points involved early speed in an objects decline, Figure 3 and Figure 4 show scatter plots for the S&P 500 and Dow Jones Industrial Average, respectively. Plotted in both figures are the number of trading days it took each index to decline by 5% and the resulting drawdown the index experienced over the following six months after the initial 5% move. The declines that were 15 days or less are plotted in red. Note that the majority of the largest declines often occurred when the index saw the initial decline develop in just three weeks. Included are significant market events such as the Great Depression era of the 1920s and 1930s, the Black Monday market crash of September 1987, the inflation-induced crash of 1946, the bear markets that began in 2000 and 2007, as well as the more recent Covid Crash in 2020 and bear market of 2022. From this look at market history, it appears equity market indices do subscribe to Bernoulli and Newton's concept of the significance early speed has on the ultimate path taken by an object — and in this case — equity prices.

Figure 3: The Relationship Between Time to a 5% Drawdown in the S&P 500 and Resulting Decline



S&P 500 closing values from January 1950 through September 2022 that resulted in a 5% decline from a 52-week high. Vertical axis shows resulting drawdown six months following a 5% decline from a 52-week high.

Figure 4: The Relationship Between Time to a 5% Drawdown in the DJIA and Resulting Decline



Dow Jones Industrial Average closing values from January 1900 through March 2022 that resulted in a 5% decline from a 52-week high. Vertical axis shows resulting drawdown six months following a 5% decline from a 52-week high.

[13] See Lo & MacKinlay, 1988

[14] See Brock, Lakonishok & LeBaron, 1992

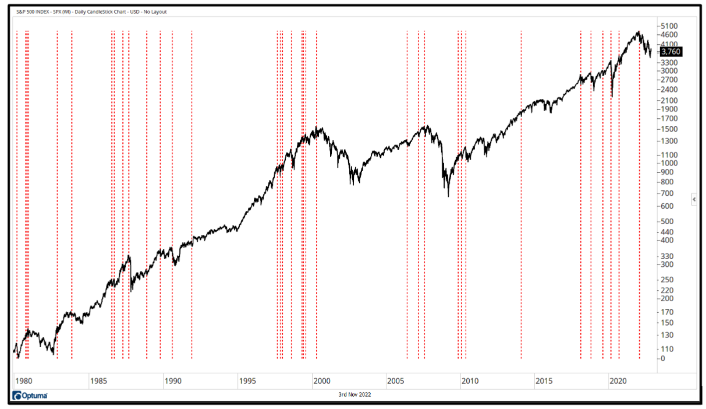
In the early 20th century, canaries and other birds were used in coal mines as early detectors of carbon monoxide. John Haldane, an expert in respiratory physiology, is credited with the initial suggestion of using birds to provide an early warning of the poisonous gas. This was due to their higher metabolic rates compared to humans, which meant they showed symptoms faster, such as before gas levels reached dangerous levels to the coalminers.[15] Canaries in coal mines aren't the only place we find unique "early warning" protocols used. Many vineyards plant rose bushes at the ends of grapevine rows, as both roses and grapes are vulnerable to many of the same diseases. When roses begin to show the negative signs of a fungus, a winemaker can quickly take action to protect the grapes.[16] Just like the canary and the rose bush, the observation of a 5% decline in major equity indices, within a 15-day period, also acts as an "early warning" of what could potentially turn into a material decline in the equity market. With that, we can title these developments "5% Canary" signals.

If seeking an explanation outside of 17th century mathematics for the evident correlation between quick initial declines and material drawdowns, a cornerstone of technical analysis, the concept of momentum is well-suited for the task. Using a classic definition of momentum, a measurement of change over a specified period in order to gauge the strength or weakness of an underlying trend, shows that this accumulation of strong price action in the indices to the downside builds upon itself and draws in additional sellers, historically resulting in larger-than-average drawdowns. However, as this paper will show, momentum alone does not account for the separation of outcomes based on the time for the 5% decline to take place. Momentum is the kindling that is lit with the proverbial match of "early speed" in the 5% Canary signals that result in above-average weakness in the equity indices.

Building on the work cited previously by Barber and Odean, on attention-grabbing data having an impact on the emotionally driven investor behavior, the initial 5% declines in the indices that occur at an accelerated rate likely garner a higher degree of attention and with it, trade activity. The findings of Kaminski and Lo also applies here, that the resulting declines in the equity indices are larger than average, "…in the event of a significant drop in aggregate stock prices, investors who are generally passive will become motivated to trade."[8]

Shown in Figures 3 and 4, a relationship for both the S&P 500 and the Dow Jones Industrial Average has been identified between 5% Canary declines and some of the largest drawdowns that have taken place since 1950 and 1900, respectively. However, not all 5% Canary signals result in a steep and material sell-off. As shown in Figure 5, 5% Canary signals have been produced throughout history, several have shown to be "ignored" by the market and no major decline has come to pass.

Figure 5: 5% Declines From a 52-Week High in The S&P 500



S&P 500 daily line chart from 1980 through November 2nd, 2022. Red lines indicate the first 5% decline from a 52-week high that has taken place within 15 trading days. Repeated instances within 21-days have not been shown.

The next step to be taken in the process of formalizing a methodology for using 5% Canary signals as an early warning protocol for equities is to seek a form of price confirmation. With confirmation, we can begin filtering the 5% Canaries that carry a higher degree of downside risk from those that the market ultimately disregards. Moving Averages are common tools used within technical analysis, with 50-day and 200-day Simple Moving Averages (SMA) being popular lookback periods used when technicians are applying trending identification analysis. Moving averages are arithmetic averages of the price of a security or index over a specified period of time. Paul Tudor Jones also believed the 200-day SMA was a useful tool, "My metric for everything I look at is the 200-day moving average of closing prices. I've seen too many things go to zero, stocks and commodities."[17] Siegel (2014) showed that the 200-day SMA offers a degree of superiority to a passive buy-and-hold investment approach during the 1886 to 2006 period for the Dow Jones Industrial Average. The caveat that Siegal points out is a SMA-based market timing strategy can succumb to whipsawing markets, periods of non-trending price activity.[18] Siegal is correct, when used in isolation, the obstacle to using smoothing mechanisms like moving averages when price volatility is elevated, and a trend is unable to be established. In this paper, the usage of a moving average is applied as a form of signal confirmation rather than initial or standalone signal generation.

Applying the 200-day SMA to the index charts we can begin seeking affirmation of the 5% Canary signal occurrences. The confirmation methodology is as follows:

> When the underlying index declines by 5% within 15 days from a 52-week high, (producing the 5% Canary signal), and closes under the 200-day SMA for two consecutive days.

Going forward, these instances will be called "Confirmed 5% Canary" signals. Like many forms of analysis, whether in academia or financial markets, validation of a hypothesis or a

[15]See Sekhar & Haldane, 2014

[16]See Cameron, 2015

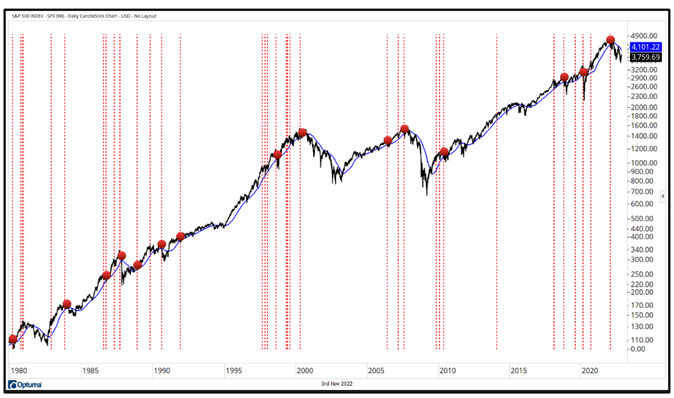
[17]See Robbins, 2016

[18]See Siegel, 2014

price-based signal or indicator provides clarity and verification. For the purposes of this paper, the consecutive closes below the 200-day SMA are sought to occur within two months (42 trading days) of an identification of a 5% Canary signal. This period was selected, and not optimized to avoid curve fitting, to allow ample opportunity for price action of the index to confirm the perceived trend change as noted by the initial signal.

In Figure 6, the same chart of the S&P 500 as used in Figure 5 is shown but with the addition of red dots at the price points when the index has closed below the 200-day SMA for two consecutive days, providing the Confirmed 5% Canary signal. Since 1980, there have been just 15 Confirmed 5% Canary signals in the S&P 500 and 14 Confirmed 5% Canary signals in the Dow Jones Industrial Average (Figure 11 in the Appendix). Included in this sampling are the significant market events that were noted by the 5% Canary signals in Figures 3 and 4. This includes the Great Depression, 1987 Crash, bear market following the Tech Bubble of 2000, Financial Crisis of 2008, Covid Crash of 2020, and the bear market that began in 2022. Of the 15 Confirmed 5% Canary signals on the S&P 500 sampling shown on Figure 6, only four did not occur ahead of what resulted in a double-digit drawdown for the index. Since 1980, only one drawdown of -20% or more for the S&P 500 did not receive a Confirmed 5% Canary signal (the bear market of 1981-1982).

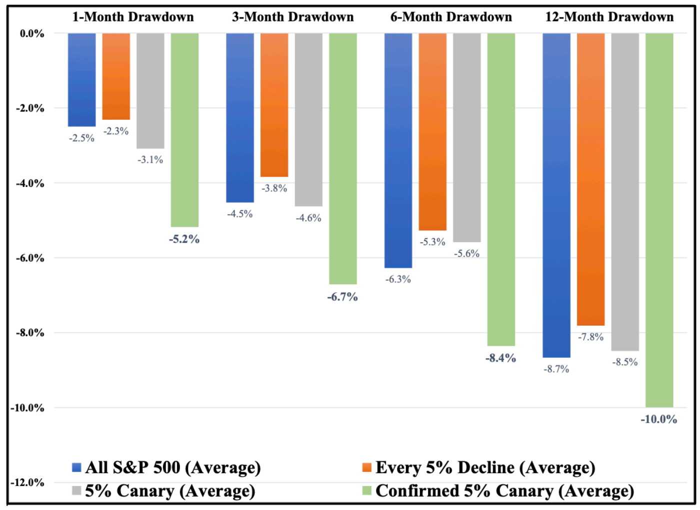
Figure 6: Confirmed 5% Canary Signals



S&P 500 daily line chart from 1980 through November 2nd, 2022 with a 200-day SMA. Red lines indicate the first 5% decline from a 52-week high that has taken place within 15 trading days. Red dots indicate price confirmation of 5% Canary signals based on the index experiencing two consecutive closes below the 200-day SMA within 42 days of the a 5% Canary signal.

Comparing the resulting price declines for the S&P 500 in Table 1 and the Dow Jones Industrial (Table 3 in the Appendix), both average and median drawdowns for 1-month, 3-month, 6-month and 12-month periods are shown. We can see in Figure 7 that more severe price weakness is witnessed following 5% Canary and Confirmed 5% Canary periods compared to the entire sampling of all 5% declines that take place regardless of length of time. For the S&P 500, the median decline 3- and 6-months after Confirmed 5% Canary signals is over twice the size of the Every 5% Decline group.

Figure 7: Comparison of Average Drawdowns Across Various Timeframes



Bar chart of the average drawdowns across multiple timeframes based on S&P 500 data from January 1950 through October 2022.

Table 1: Average & Median Drawdowns For the S&P 500

	1-Month Drawdown	3-Month Drawdown	6-Month Drawdown	12-Month Drawdown
All S&P 500 (Average)	-2.5%	-4.5%	-6.3%	-8.7%
Every 5% Decline (Average)	-2.3%	-3.8%	-5.3%	-7.8%
5% Canary (Average)	-3.1%	-4.6%	-5.6%	-8.5%
Confirmed 5% Canary (Average)	-5.2%	-6.7%	-8.4%	-10.0%
All S&P 500 (Median)	-1.6%	-3.0%	-4.0%	-5.3%
Every 5% Decline (Median)	-1.4%	-2.1%	-3.0%	-4.8%
5% Canary (Median)	-2.0%	-2.6%	-2.8%	-6.7%
Confirmed 5% Canary (Median)	-2.6%	-5.0%	-6.9%	-7.8%

S&P 500 data from January 1950 through October 2022. 5% declines are measured from 252-day closing high on the index.

This decisively shows that the resulting drawdowns in equity indices following an initial 5% decline are not solely the result of price momentum. Instead, it shows the significance of the time duration the decline occurs in having a larger impact on the ultimate price deterioration that has historically followed.

The data for both the S&P 500 and Dow Jones Industrial Average show outsized drawdowns for 1-, 3-, 6-, and 12-month periods when compared to the Every 5% Decline group. Viewed in the context of economic recessions, since 1925 there have been just three of the 16 NBER classified recessions that were not first preceded by a Confirmed 5% Canary signal on either the S&P 500 or Dow Jones Industrial Average (Figure 12 of the Appendix). A closer look at the price history over the last six years is also shown with 5% Canary and Confirmed 5% Canary signals in Figure 13 of the Appendix.

Section Two: Buy The Dip

Throughout the bull market of the 1990s, traders were rewarded for buying nearly each dip in the market as stocks charged higher month after month. This “buy the dip” mentality bred overconfidence and helped add air to the expanding bubble that eventually popped, resulting in the bear market that lasted nearly three years. As human behavior often repeats itself, the “buy the dip” approach grew once again in popularity during the decade following the Financial Crisis and intensified the year after the Covid Crash in 2020. All good things come to an end and the equity market showed the faults of this superficial strategy with outsized declines that wreaked havoc on many investors’ portfolios.

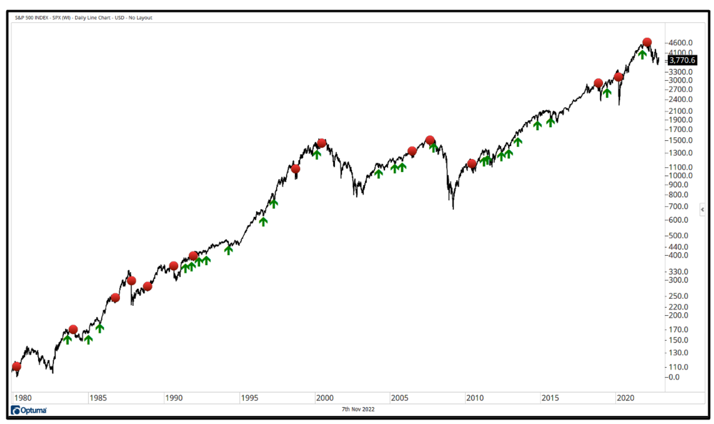
As Section One showed, 5% Canary signals provide early warning to potential major declines in the equity market that then increase in confidence through Confirmed 5% Canary signals. The next question we must ask pertains to the potential significance of 5% declines that take longer than 15 days to develop. One method for validating robustness of a signal is through the review of conditions that do not meet the stated criteria. Section Two of this paper will look to improve upon the “buy the dip” strategy through a more quantitative approach with a focus on when an equity index takes longer than 15 days to experience a 5% decline.

Investors in the periods of the last 10+ years and the 1990s were not alone in their folly of believing the dip can always be bought. Isaac Newton, considered one of the greatest minds in mathematics and physics, who also helped solve the previously discussed Bernoulli math challenge, also was bested by the market due to his dip buying. Newton had been drawn into the speculation of the 1720 South Sea Company bubble and had famously said, “I can calculate the movement of the stars, but not the madness of men” after experiencing a severe loss when the bubble eventually burst. Newton was one of the executors of the Thomas Hall estate, trading in securities on its behalf, including in South Sea Company stock. “The most interesting transactions are the final purchases of South Sea stock, around the middle of September. At that point this security was in a free fall, [...] for a total of 400 stock, at prices of 520, 405, and 395. They [Newton and the other executors of the trust] could only have been motivated by deep conviction that the market’s change of heart about the South Sea Company was just a temporary irrational panic, and that there was real value in that venture.”[19] Appendix Figure 14 shows Newton’s purchases and what we now know were failed attempts to “buy the dip” following the summer peak in South Sea Company shares.

Identifying and respecting the long-term trend of the market is an important first step in determining if a dip presents the proper criteria to warrant an opportunity for mean-reversion in price. Moving averages are popular tools used in trend identification. Similar to the confirmation criteria used in the Confirmed 5% Canary signal of Section One, intermediate and long-term moving averages can be used to identify up and down trends of a respective index. Faber showed that monthly moving averages provide superior returns in active management of portfolios.[19] In a paper published in *The Journal of Wealth Management*, Kilgallen showed that Simple Moving Averages produced 28% less drawdown than a buy-and-hold strategy for commodities, 44% less for equities, and 65% less downside for currencies.[20] Kirkpatrick and Dahlquist also showed that moving averages are useful in identifying trends, “you can see how the moving average tends to follow the trend line fairly well. The moving average then becomes a proxy for the trend line and can be used to determine when a trend is potentially changing direction, just as a trend line can.”[21]

For the purposes of this paper, the 50-day Simple Moving Average (SMA) above the 200-day Simple Moving Average (SMA) will be used for classifying up trends within the index. Only qualifying signals found on a closing basis on daily charts of the index when the 50-day SMA closes above the 200-day SMA will be recognized. Figures 8 and 9 show daily charts of the S&P 500 and Dow Jones Industrial Average with green arrows used to show when the respective index has declined by 5% from a 52-week high that exceeded 15 trading days while the underlying index was in a defined-up trend based on the above-mentioned moving averages. Going forward, these occurrences will be called “Buy The Dip” signals.

Figure 8: Confirmed 5% Canary & Buy The Dip Signals For the S&P 500



S&P 500 daily line chart from 1980 through November 2022 with Confirmed 5% Canary signals shown in red dots and Buy The Dip signals shown in green arrows.

[19]See Faber, 2007

[20]See Kilgallen, 2012

[21]See Kirkpatrick & Dahlquist, 2008

Figure 9: Confirmed 5% Canary & Buy The Dip Signals For the DJIA

Dow Jones Industrial Average daily line chart from 1980 through November 2022 with Confirmed 5% Canary signals shown in red dots and Buy The Dip signals shown in green arrows.

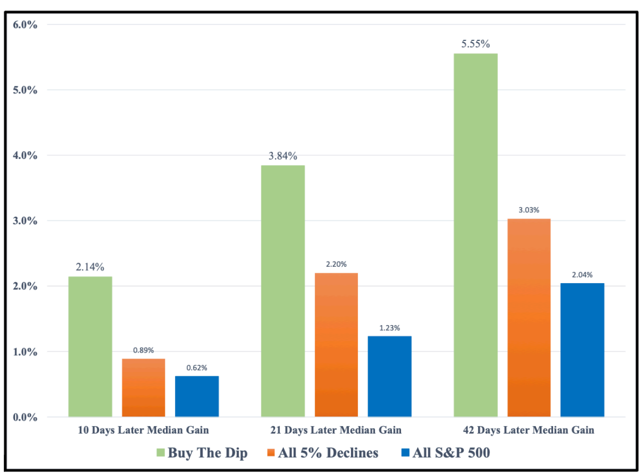
Table 2: Data Results from Buy The Dip and All 5% Declines Over Various Timeframes

	Buy The Dip	All 5% Declines	All S&P 500
10 Days Later % Higher	66.67%	63.01%	60.38%
10 Days Later Median Gain	2.14%	0.89%	0.62%
80th/20th Percentile	3.48%/-1.62%	3.58%/-1.95%	2.47%/-1.63%
21 Days Later % Higher	75.00%	65.90%	63.96%
21 Days Later Median Gain	3.84%	2.20%	1.23%
80th/20th Percentile	5.1%/-0.35%	5.1%/-1.97%	3.8%/-2.14%
42 Days Later % Higher	87.50%	69.36%	67.10%
42 Days Later Median Gain	5.55%	3.03%	2.04%
80th/20th Percentile	8.19%/0.89%	7.03%/-1.77%	5.89%/-2.2%

Values based on closing values of the S&P 500 from January 1980 through June 2022. Buy The Dip based on 5% declines from 52-week highs that take more than 15 days to occur. Both Buy The Dip signals and All 5% Decline instances were included when occurring in up trends based on the 50-day SMA closing above the 200-day SMA.

Since 1980, the S&P 500 was higher 66.7% of the time two weeks after Buy The Dip signals with a median gain of 2.14%, compared to all declines of 5% within an uptrend, which resulted in a median gain of less than 1%. As shown in Table 2, two months after Buy The Dip signals saw the market higher 87.5% of the time with a gain of over 5.5%, compared to less than 70% for the full sample of 5% declines that saw a median gain of 3% during the same period. Alongside the improved upside probability of the Buy the Dip signals, the range of historical outcomes produced a larger spread between the 80th and 20th percentiles compared to the All 5% Decline group, offering a more attractive risk/reward profile based on historical outcomes.

Figure 10: Change in the S&P 500 Following Buy the Dip Signals & All 5% Declines



Values based on closing values of the S&P 500 from January 1980 through June 2022. Buy The Dip signals based on 5% declines from 52-week highs that take more than 15 days to occur. Both Buy The Dip signals and All 5% Decline instances were included when occurring in up trends based on the 50-day SMA closing above the 200-day SMA.

The superior risk/reward profile presented through this nearly 42-year study on the S&P 500 shows the validity of the Confirmed 5% Canary signal as well as the opportunities offered when market declines are not classified as a 5% Canary. As Albert Einstein once said, "Everything should be made as simple as possible, but not simpler." Section Two was not structured on the basis of optimized data, timeframes, or signal thresholds; criteria and structure of the studies presented were intentionally kept simple and broad. While the focus of this paper was on the S&P 500 and Dow Jones Industrial Average, foreign markets were also tested, which showed similar results as the U.S. equity indices. One distinction found with non-U.S. indices was that the volatile nature of several less developed and emerging markets produced less meaningful signals than the larger developed markets. Example charts of both Confirmed 5% Canary and Buy The Dip signals can be found for Germany's DAX Index, Japan's Nikkei 225 Index, France's CAC 40 Index, Australia's ASX Index, and Israel's Tel Aviv 125 Index in Figures 15-19 of the Appendix.

The value of the study stands firmly in the resulting data of the improved risk management of quick initial market declines and in the opportunities presented to investors by index declines that take a protracted period of time.

Conclusion

This paper provides a data-driven argument for using the amount of time it takes an equity index to decline by 5% to serve as either an early warning sign of a material decline in the market or an opportunity to Buy The Dip with the assumption the market will reverse its decline and move higher. Through a study of market history from 1900 on the Dow Jones Industrial Average and 1950 for the S&P 500, Figures 3 and 4 showed the relationship between the resulting drawdown and the initial decline of the two equity indices. A relationship was found in the observation that many of the largest drawdowns in market history began by initially declining by 5% within 15 days. The 5% Canary signal was improved upon with the inclusion of the 200-day Simple Moving Average to assist with price confirmation of the initial signal. With a median drawdown for Confirmed 5% Canary signals of -5% after 3-months compared to -2.1% for all 5% declines for the S&P 500 and -5.3% and -3.3% for the DJIA, respectively, the concept of the 5% Canary proves to be meaningful throughout market history.

The concept of measuring the time to a 5% decline was also shown to have value in a bullish context as well, with the S&P 500 rising over 87% of the time in the two months following a Buy The Dip signal, resulting in a 45% higher median gain compared to all 5% declines in the S&P 500 and a 63% higher median gain than all two month periods for the index. The insight provided by Kaminski and Lo (2014), that investors pay greater attention to significant declines in the market than they typically would, explains the need for a proper lens to view declines such as these and to be better prepared with what action if any should be taken.

Built upon the foundation of technical analysis research, this study of 5% declines from both the viewpoints of bullish and bearish market outcomes offers a unique methodology to the investment community that can be applied by both retail and professional investors. The creation of a complete trading strategy or system is not the objective of this paper. Analysis on other trend classification and/or additional criteria could be studied and tailored to the Confirmed 5% Canary and Buy The Dip signals presented in Sections One and Two to better fit an investor's risk tolerance and trade preferences. Finally, the idea of the "early speed" in an object's path that was initially introduced by the creation of the brachistochrone curve in the 1700s has been shown in this paper to have robust application to financial markets when applied to index price changes and offers investors another tool to be deployed in both downside risk management and upside opportunistic applications.

Resources

3Blue1Brown. (2016, April 1). *The Brachistochrone, with Steven Strogatz* [Video]. YouTube. https://youtu.be/Cld0p3a43fU

Antoons, Wim, Market Timing: Opportunities and Risks (November 2016). Brandes Institute Research Paper No. 2016-06.

Barber, Brad M., and Terrance Odean. 2008. All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. The Review of Financial Studies 21: 785–818.

Brock, William, Josef Lakonishok, & Blake LeBaron. (1992). Simple Technical Trading Rules and Stochastic Properties of Stock Returns. *The Journal of Finance*, 47(5), 1731–1764.

Cameron, Chris. "Ask the Advisor, Roses in the Vineyards." *California Winery Advisor*, Sept. 2015, californiawineryadvisor.com/ask-the-advisor-roses-in-the-vineyards/.

Faber, Meb, A Quantitative Approach to Tactical Asset Allocation (February 1, 2013). The Journal of Wealth Management, Spring 2007

Goetzmann, William N. and Kim, Dasol and Shiller, Robert J., Crash Beliefs from Investor Surveys (March 19, 2016).

Henry, Bruce, and Watt, Simon. "Surfing Brachistochrones." Parabola, vol. 34, no. 3, 1998.

Jennings, Bill. "The Physics and Mathematics of Skiing." The Spokesman Review, 19 Jan. 2017.

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. Econometrica, 47(2), 263–291. https://doi.org/10.2307/1914185

Kathryn M. Kaminski and Andrew Lo, (2014), When do stop-loss rules stop losses?, Journal of Financial Markets, 18, (C), 234-254

Kilgallen, Thomas. "Testing the Simple Moving Average across Commodities, Global Stock Indices, and Currencies." *The Journal of Wealth Management* 15 (2012): 100 - 82.

Kirkpatrick, Charles A., and Julie R. Dahlquist. Technical Analysis: *The Complete Resource for Financial Market Technicians.* Pearson Education, 2008.

Lo, Andrew W, and Craig MacKinlay. "Stock Market Prices Do Not Follow Random Walks: Evidence from a Simple Specification Test." *The Review of Financial Studies*, vol. 1, no. 1, 1988, pp. 41–66., doi:https://www.jstor.org/stable/2962126.

M de Icaza Herrera, Galileo, Bernoulli, Leibniz and Newton around the brachistochrone problem, *Rev. Mexicana Fis.* 40 (3) (1994), 459-475.

Odlyzko, Andrew, Newton's Financial Misadventures in the South Sea Bubble (February 10, 2018).

Peng, Lin, and Wei Xiong. 2006. Investor attention, overconfidence and category learning. Journal of Financial Economics 80: 563–602.

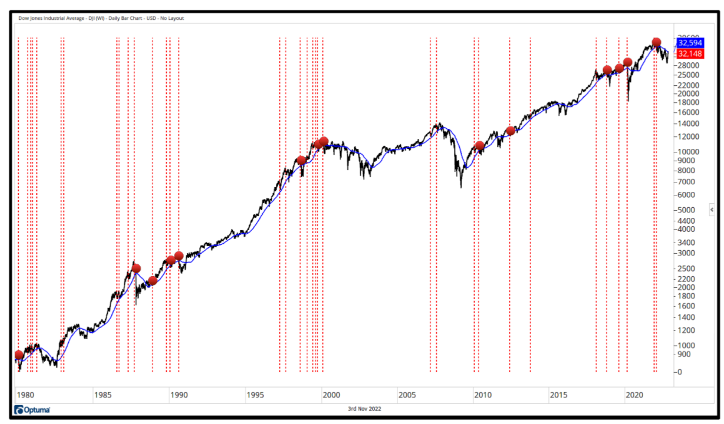
Robbins, Tony. Money: Master the Game. Simon & Schuster LTD, 2016.

Sekhar K, Rao SC. John Scott Haldane: The father of oxygen therapy. Indian J Anaesth. 2014 May;58(3):350-2. doi: 10.4103/0019-5049.135087. PMID: 25024490; PMCID: PMC4091013.

Siegel, Jeremy J. Stocks for the Long Run the Definitive Guide to Financial Market Returns & Long-Term Investment Strategies. McGraw Hill, 2014.

Appendix

Figure 11: 5% Canary & Confirmed 5% Canary Signals For The DJIA



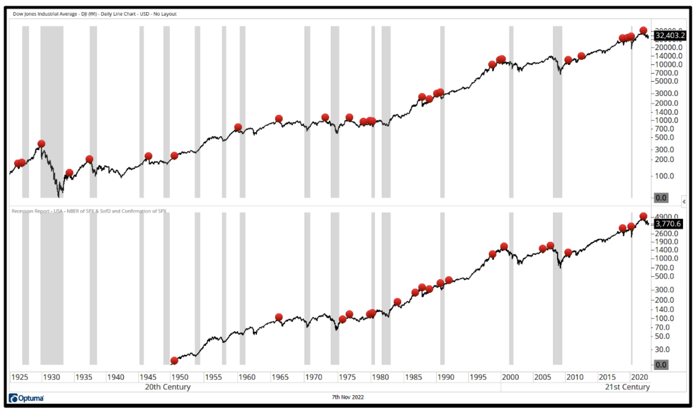
Dow Jones Industrial Average daily line chart from 1980 through November 2nd, 2022, with the 200-day Simple Moving Average. Red vertical lines indicate the first 5% decline from a 52-week high that has taken place within 15 trading days. Red dots indicate price confirmation of 5% Canary signals based on the index experiencing two consecutive closes below the 200-day SMA within 42 days of the initial 5% Canary.

Table 3: Average & Median Drawdowns For the DJIA

	1-Month Drawdown	3-Month Drawdown	6-Month Drawdown	12-Month Drawdown
All DJIA (Average)	-2.9%	-5.4%	-7.7%	-10.8%
Every 5% Decline (Average)	-3.0%	-5.3%	-7.6%	-10.9%
5% Canary (Average)	-3.0%	-5.4%	-7.8%	-11.0%
Confirmed 5% Canary (Average)	-7.7%	-10.0%	-11.5%	-16.0%
All DJIA (Median)	-1.8%	-3.4%	-4.8%	-6.8%
Every 5% Decline (Median)	-1.9%	-3.3%	-5.1%	-8.1%
5% Canary (Median)	-1.8%	-3.9%	-5.1%	-7.9%
Confirmed 5% Canary (Median)	-2.9%	-5.3%	-7.6%	-10.8%

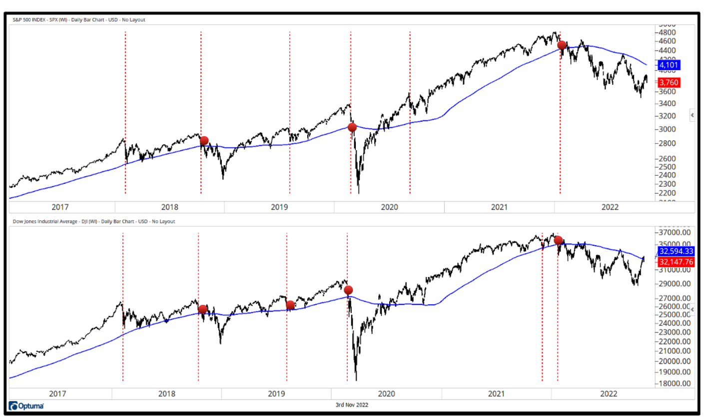
Dow Jones Industrial Average data from January 1900 through October 2022. 5% Declines are measured from 252-day closing high on the index.

Figure 12: Recession Periods Since 1925



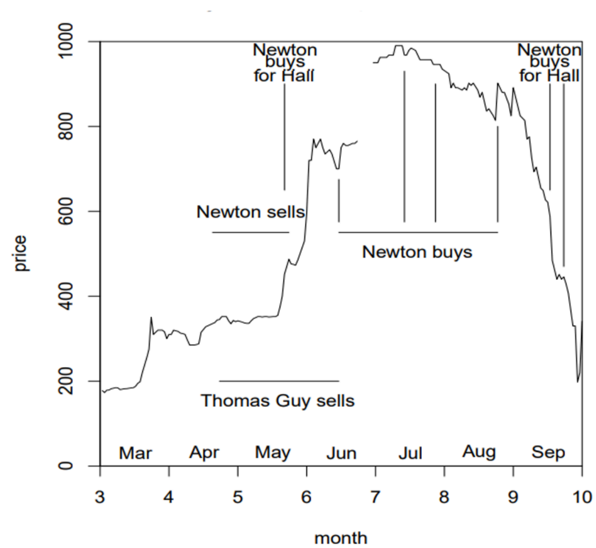
Dow Jones Industrial Average since 1925 and S&P 500 since 1950 with NBER classified recession periods shown in gray and Confirmed 5% Canary signals shown with red dots.

Figure 13: 5% Canary & Confirmed 5% Canary Signals For The DJIA and S&P 500



S&P 500 and Dow Jones Industrial Average daily bar chart from 2017 through October 2022. Price with 200-day Simple Moving Average (blue line), 5% Canary (vertical red lines) and Confirmed 5% Canary Signals (red dots) shown.

Figure 14: Price Chart of The South Sea Company in 1720



Source: Odlyzko, 2018

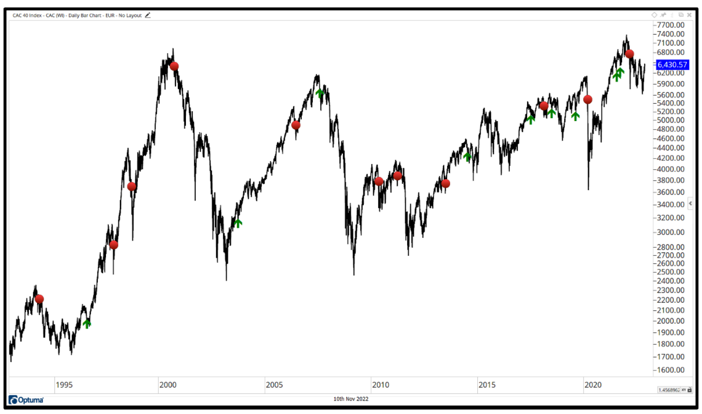
Figure 15: Confirmed 5% Canary and Buy The Dip Signals for Australia's XAO Index

Daily bar chart of XAO Index with Confirmed 5% Canary Signals in red dots and Buy The Dip Signals shown by green arrows.

Figure 16: Confirmed 5% Canary and Buy The Dip Signals for Germany's DAX Index

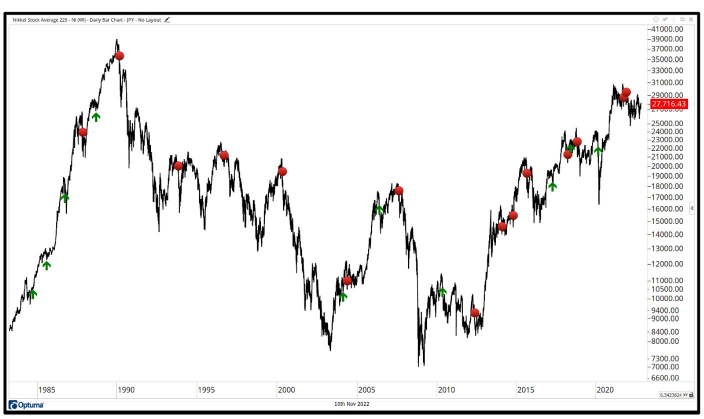
Daily bar chart of DAX Index with Confirmed 5% Canary Signals in red dots and Buy The Dip Signals shown by green arrows.

Figure 17: Confirmed 5% Canary and Buy The Dip Signals for France's CAC40 Index



Daily bar chart of CAC40 Index with Confirmed 5% Canary Signals in red dots and Buy The Dip Signals shown by green arrows.

Figure 18: Confirmed 5% Canary and Buy The Dip Signals for Japan's Nikkei 225 Index



Daily bar chart of Nikkei 225 Index with Confirmed 5% Canary Signals in red dots and Buy The Dip Signals shown by green arrows.

Figure 19: Confirmed 5% Canary and Buy The Dip Signals for Israel's Tel Aviv 125 Index

Daily bar chart of Tel Aviv 125 Index with Confirmed 5% Canary Signals in red dots and Buy The Dip Signals shown by green arrows.

About the NAAIM Founders Award

Launched in 2009, the NAAIM Founders Award is designed to expand awareness of active investment management techniques and the results of active strategies through the solicitation and publication of research on active management. $5,000 is presented annually for the best paper submitted to the competition.

The National Association of Active Investment Managers (NAAIM) is a non-profit trade group of nearly 130 registered investment advisor firms that collectively manage more than $15 billion in assets. NAAIM member firms provide active money management services to their clients to produce favorable risk-adjusted returns as an alternative to passive, buy-and-hold investment strategies. NAAIM publishes the weekly NAAIM Exposure Index and sponsors the annual Uncommon Knowledge conference along with smaller conferences and meetings on topics of interest to its membership. For more information, visit www.naaim.org

Encyclopedia of Chart Patterns *by Thomas N. Bulkowski*

Reviewed by Regina Meani, CFTe

In the time poor world in which we live it is so easy to press a button for most of the things that we want. This pertains to the world of technical analysis. Many technical analysis programs allow us to press a button and many forms of analysis can be applied to the chart. Some may say it is a time saver but are we losing our own individual ability to interpret the charts?

For chart interpretation, I thought that pattern recognition was a good place to start. On my bookshelf, I found my well-worn copy of Thomas Bulkowski's *Encyclopedia of Chart Patterns*. I have the first edition and I believe it has now been expanded with second and third editions, which add extra value and additional features. The third edition offers 10 times the number of samples for statistic reporting.

Delving back into my copy, Bulkowski reminded me that "chart patterns ... are the footprints of the smart money... it pays to know what the market is thinking. It pays to follow the footprints."[1]

In the introduction, Bulkowski explains his methodology, choosing his data base, discussing the market at large, developing an investment style and the use of theoretical sample trades to highlight certain techniques. The body of the book contains detailed descriptions of over 30 patterns. Every chapter has a heading image of the pattern for easy identification followed by a "Results Snapshot." Here a condensed guide serves as a quick reference. The table includes how to identify the pattern, whether it is considered continuation or reversal, the percentage failure rate, the average percentage gain or loss and the trend description. Bulkowskis often adds a "surprising finding" as well as associated patterns.

Each chapter is divided into sections. The "Tour" provides a general description followed by "Identification Guidelines" in text and in a table. Bulkowski then "Focuses on Failures," describing what a failed pattern looks like and why they failed and what to do when they fail.

The next section deals with the statistics and often highlights the trustworthiness of the pattern and the potential returns. The section on "Trading Tactics" is one of the most useful as it sets out a guide that helps maximize profits and minimize losses, identifying where to place stop losses and when to go long or short. Finally, in the "Sample Trade," Bulkowski uses a unique and not without humour story telling style which takes you step by step on how to trade the patterns. In the complex head and shoulders trade, in Chapter 19, he ends by falling off his bicycle.

The third edition adds to the sections in the chapters with "Experience," which lets you share Bulkowski's expertise, sitting at the master's knee, learning the lessons from his trading techniques.

Throughout the pages of this book there is a wealth of examples, which show pattern guidelines and tag the appropriate target price.

As many of us press that "button" in our analysis program, it might be useful to keep a copy of the *Encyclopedia of Chart Patterns* close by as a confirmation/reference tool, and as I am intrigued by the new additions to the third edition, I will be updating my bookshelf.

Notes

[1] Bulkowski, TN, *Encyclopedia of Chart Patterns*, John Wily & Sons, Canada, 2000, pg 3

Author Profiles

Andrea Bolognesi

Andrea Bolognesi is a second-year PPE & Data Science student at University College London (UCL) and an incoming summer analyst at UBS London. He possesses a keen interest in equities, global macro, and quantitative as well as systematic investment strategies. Currently, he is working towards achieving the CFA Level I certification. In addition to his core studies, Andrea has undertaken multiple research projects in the field of fixed income, focusing on the factors that influence recent changes in interest rates and subsequent market's response. Passionate about the integration of new technologies in investment and technical analysis, Andrea welcomes enquiries and opportunities.

Regina Meani, CFTe

Regina Meani covered world markets as a technical analyst and associate director for Deutsche Bank prior to freelancing. She is an author in the area of technical analysis and is a sought after presenter both internationally and locally, lecturing for various financial bodies and universities as well as the Australian Stock Exchange. Regina is a founding member and former president of the Australian Professional Technical Analysts (APTA) and a past journal director for IFTA, carrying the CFTe designation and the Australian AMT (Accredited Market Technician). She has regular columns in the financial press and appears in other media forums. Her freelance work includes market analysis, webinars, and larger seminars; advising and training investors and traders in market psychology; CFD; and share trading and technical analysis. She is also a former director of the Austral-ian Technical Analysts Association (ATAA) and has belonged to the Society of Technical Analysts, UK (STA) for over 30 years.

Davide Pandini, PhD, CMT, MFTA, CFTe, CSTA

Davide Pandini holds a Ph.D. in electrical and computer engineering from Carnegie Mellon University. He was a research intern at Philips Research Labs in Eindhoven, the Netherlands, and at Digital Equipment Corp., Western Research Labs, in Palo Alto, U.S. He joined STMicroelectronics in Agrate Brianza, Italy, in 1995, where he is a technical director and a fellow of ST technical staff.
Davide has authored and co-authored more than 50 papers in international journals and conference proceedings and served on the program committee of several premiere international conferences. He received the STMicroelectronics Corporate STAR Gold Award in 2008 and 2020, and the STRIVE Gold Award in 2022, for R&D excellence. Since June 2015, he has been the chairman of the ST Italy Technical Staff Steering Committee.
In the field of technical analysis, Davide is a Certified Financial Technician and a Master of Financial Technical Analysis of IFTA, holds the Chartered Market Technician (CMT) designation from the CMT Association, and is a professional member of SIAT (Societa' Italiana Analisi Tecnica). In 2021, he was the recipient of the XII SIAT Technical Analyst Award in the Open category. He was a speaker at SIAT Trading Campus in 2022, and at the Investing and Trading Forum in 2022 and 2023. He was the winner of the prestigious IFTA 2021 John Brooks Memorial Award and was a speaker at the IFTA 2023 annual conference in Jakarta, Indonesia.
Davide served as volunteer at the Universal Exhibition Expo2015—*Feeding the Planet, Energy for Life*—in Milano, Italy.

Andrew Thrasher, CMT

Andrew Thrasher, CMT, is the Portfolio Manager for Financial Enhancement Group LLC and founder of Thrasher Analytics LLC. Mr. Thrasher holds a bachelor's de-gree from Purdue University and the Chartered Market Technician (CMT) designa-tion. He is a two-time winner of the Charles H. Dow Award from the CMT Associ-ation. He was also a finalist in the Best Equity Research and Best Commodity Re-search categories for the Technical Analysis Awards in 2022. Andrew resides in Noblesville, Indiana with his wife and daughter. His analysis has been cited by CNBC, Wall Street Journal, MarketWatch, Bloomberg, Fox Business, ValueWalk, Yahoo! Finance, Barron's, U.S. News, TD Ameritrade Network and Opto.

Michael Trequattrini, CSTA

Michael Trequattrini is a quantitative researcher specializing in alpha generation using artificial intelligence algorithms, developing and implementing trading and investment strategies across stocks, futures, and commodities. He holds a B.Sc in Bank, Finance, and Financial Markets with Honours from the University of Pisa and a M.Sc in Finance from the Nova School of Business and Economics. He is a Certified SIAT Technical Analyst (CSTA) and the winner of the SIAT Technical Analyst Award in 2022. He has authored numerous articles and has been featured on various YouTube channels. During his studies in Portugal, he served as a co-portfolio manager for a local bank, actively contributing to portfolio management.

Edzard Wiener, Ph.D, MFTA, CFTe

Member of the Vereinigung Technischer Analysten Deutschlands e.V. (VTAD)
Dr. Edzard Wiener holds a master's in physics and a PhD in medicine, specializing as a Senior Physician in Neuroradiology with a significant focus on artificial intelligence, computer vision and neuroscience. Since 2018, Dr. Wiener has established a specialized position for himself in the field of technical analysis.

He offers a unique mix of academic and practical expertise in physics, neuroscience, and trading. He has investigated various aspects of market behavior, data science, artificial intelligence, and neural networks. His intense involvement in exploring the synergy between data science and trading has not only expanded his professional horizon but also contributed extensively to his success as a private trader over the past five years. He is recognized for merging traditional trading methodologies with cutting-edge AI and data analysis, thereby fostering a blend of human and artificial intelligence in understanding market movements. His innovative indicators offer promising avenues for understanding and predicting significant price movements. In 2022, Dr. Wiener obtained the Certified Financial Technician (CFTe) designation and further strengthened his credentials in 2023 by acquiring the Master of Financial Technical Analysis (MFTA), underscoring his profound expertise and commitment in technical analysis.

Christoph Wildensee, Ph.D., MFTA

Christoph Wildensee has a Ph.D. in business administration. He is a well-known auditor and data/process analyst at enercity AG in Hannover, Germany. Christoph's special focus is on finding errors and optimization potential in IT systems relevant to accounting, including, in particular, the Energy Trade and Risk Management (ETRM) system, which is used to handle all energy trading activities at enercity. He was also a member of the team evaluating the new Pioneer/Hitachi-ABB ETRM system.

IFTA Staff

Executive Director
Linda Bernetich, CAE

Production Manager
Jansen Vera

Managing Editor
Heather Rigby

Director of Accounting Services
Dawn Rosenfeld

IFTA Board of Directors

IFTA HEADQUARTERS
International Federation of Technical Analysts
1300 Piccard Drive, Suite LL 14
Rockville, MD 20850 USA
Phone: +1 (240) 404-6508
Fax: +1 (301) 990-9771
Email: admin@ifta.org | Web: www.ifta.org

President
Wieland Arlt, CFTe (VTAD)
wieland.arlt@ifta.org

Treasurer
Karin Roller, CFTe (VTAD)
karin.roller@vtad.de

Secretary and Membership Director
Warren Murnane, Dip TA, CFTe, ST & I (Wealth Within) (ATAA)
warren.murnane@ataa.asn.au

Vice President Europe, Website Director
David Watts, BSc (Hons) CEng MICE MIWEM, FSTA (STA)
Dwatts360@gmail.com

Vice President Asia-Pacific
Hiwon Yoon, Ph.D., CMTA (NTAA)
yoon@cmdlab.co.jp

Vice President The Americas, Webinar Director
Bruce Fraser (AAPTA)
rdwyckoff@me.com

Vice President Middle East and Africa
Tamer Gamal, CFTe (ESTA)
tgamal@gmail.com

Education Director
Saleh Nasser, CMT, CFTe (ESTA)
snasser72@gmail.com

Examination Director
Gregor Bauer, Ph.D., CFTe, (VTAD)
gregor.bauer@vtad.de

Marketing Director
Eddie Tofpik, MSTA (STA)
Eddie.Tofpik@admisi.com

Quant Director
Giovanni Trombetta, CFTA, Electronic Engineer (SIAT)
giovanni.trombetta@gandalfproject.com

Institutional Director
Alessandro Greppi, Ph.D. (SIAT)
greppi.alessandro01@gmail.com

Non-Institutional Director
Shinji Okada, CMTA, CFTe, MFTA (NTAA)
shinjiokadaresearch@gmail.com

Development Director
Indrawijaya Rangkuti, MBA, CTA (AATI)
Indra.senna@gmail.com

Digital Media Director
Ron William, MSTA, CFTe, MSTA, CFTe (SAMT)
roniwilliam@gmail.com

CFTe
IFTA Certified Financial Technician

Certified Financial Technician (CFTe) Program

IFTA Certified Financial Technician (CFTe) consists of the CFTe I and CFTe II examinations. Successful completion of both examinations culminates in the award of the CFTe, an internationally recognised professional qualification in technical analysis.

Examinations

The **CFTe I** exam is multiple-choice, covering a wide range of technical knowledge and understanding of the principals of technical analysis; it is offered in English, French, German, Italian, Spanish, Arabic, and Chinese; it's available, year-round, at testing centers throughout the world, from IFTA's computer-based testing provider, Pearson VUE.

The **CFTe II** exam incorporates a number of questions that require essaybased, analysis responses. The candidate needs to demonstrate a depth of knowledge and experience in applying various methods of technical analysis. The candidate is provided with current charts covering one specific market (often an equity) to be analysed, as though for a Fund Manager.

The CFTe II is also offered in English, French, German, Italian, Spanish, Arabic, and Chinese, via Zoom, typically in April and October of each year.

Curriculum

The CFTe II program is designed for self-study, however, IFTA will also be happy to assist in finding qualified trainers. Local societies may offer preparatory courses to assist potential candidates. Syllabuses, Study Guides and registration are all available on the IFTA website at **http://www.ifta.org/certifications/registration/**.

To Register

Please visit our website at **http://www.ifta.org/certifications/registration/** for registration details.

Cost

IFTA Member Colleagues	Non-Members
CFTe I $550 US	**CFTe I $850 US**
CFTe II $850* US	**CFTe II $1,150* US**

**Additional Fee (CFTe II only): $100 US Proctor Fee*

MFTA
IFTA Master of Financial Technical Analysis

Master of Financial Technical Analysis (MFTA) Program

IFTA's Master of Financial Technical Analysis (MFTA) represents the highest professional achievement in the technical analysis community, worldwide. Achieving this level of certification requires you to submit an original body of research in the discipline of international technical analysis, which should be of practical application.

Examinations

In order to complete the MFTA and receive your Diploma, you must write a research paper of no less than three thousand, and no more than five thousand, words. Charts, Figures and Tables may be presented in addition.

Your paper must meet the following criteria:

- It must be original
- It must develop a reasoned and logical argument and lead to a sound conclusion, supported by the tests, studies and analysis contained in the paper
- The subject matter should be of practical application
- It should add to the body of knowledge in the discipline of international technical analysis

Timelines & Schedules

There are two MFTA sessions per year, with the following deadlines:

SESSION 1	
"Alternative Path" application deadline	**February 28**
Application, outline and fees deadline	**May 2**
Paper submission deadline	**October 15**

SESSION 2	
"Alternative Path" application deadline	**July 31**
Application, outline and fees deadline	**October 2**
Paper submission deadline	**March 15 (of the following year)**

To Register

Please visit our website at **http://www.ifta.org/certifications/master-of-financial-technical-analysis-mfta-program/** for further details and to register.

Cost

$950 US (IFTA Member Colleagues);
$1,200 US (Non-Members)

IFTA
International Federation of Technical Analysts