Stellenbosch
UNIVERSITY
IYUNIVESITHI
UNIVERSITEIT

Analysis to indicate the impact Hindsight Bias have on the outcome when forecasting of stock in the South African equity market

by
Anton Lafrass Heyneke

Thesis presented in fulfilment of the requirements for the degree of Master of Commerce in Financial Risk Management in the Faculty of Economic and Management sciences at Stellenbosch University

Supervisor: Emeritus Prof. W. J. Conradie
Co-supervisor: Dr. Mesias Alfeus

December 2023

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the sole author thereof (save to the extent explicitly otherwise stated), that reproduction and publication thereof by Stellenbosch University will not infringe any third party rights and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

December 2023

Acknowledgements

I would like to thank my parents for their continuous support throughout my academic career. They have not done anything but to show me endless love and support.

I would show my greatest gratitude to my Lord and Saviour, that provided me with the necessary gifts and courage to finish my master’s degree and go onto working. I am grateful to have been given the opportunity to prove myself and I dedicate all my accomplishments to Him.

I would finally like to thank Professor Conradie and Dr Alfeus for supporting me throughout the journey and showing me tons of patience and help.

Abstract

A novel Artificial Neural Network (ANN) framework presented in this study has the ability to mimic the effect that cognitive biases, specifically hindsight bias has on the financial market. This study investigates how hindsight bias influences models and their outcomes. During this study the hindsight bias effect will be measured within a South African context.

The decisions that people make when faced with uncertainty are characterized by heuristic judgments and cognitive biases. If these characteristics are systematic and confirmed through research and literature related to this topic, it would form a quintessential part to the explanation of the behaviour of financial markets. This research presents a methodology that could be used to model the impact of cognitive biases on the financial markets. In this study, an ANN will be used as a stand-in for the decision-making process of an investor. It is important to note that the selection of the companies, on which the ANN will be trained, validated and tested, demonstrated cognitive bias during the study's preparation. Though there are many cognitive biases that have been identified in the literature on behavioural finance, this study will concentrate solely on the impact of hindsight bias. On financial markets, hindsight bias manifests when outcomes seem more predictable after they have already happened. This study attempts and succeeds – to some degree - to replicate the return characteristics of the ten chosen companies for the assessment period from 2010 to 2021.

The study described here may still be subject to various cognitive biases and systemic behavioural errors in addition to the hindsight bias. The further application of this technique will stimulate further research with respect to the influence of investor behaviour on financial markets.

Opsomming

'n Nuwe Kunsmatige Neurale Netwerk (ANN) raamwerk wat in hierdie studie aangebied word, het die vermoë om die uitwerking van kognitiewe afwykings, spesifiek terugskouende sydigheid, op die finansiële mark na te boots. Hierdie studie ondersoek hoe terugskouende sydigheid, modelle en hulle uitkomste beïnvloed. In hierdie studie word die terugskouende sydigheid-uitwerking binne 'n Suid-Afrikaanse konteks gemeet.

Die besluite wat mense maak wanneer hulle met onsekerheid gekonfronteer word, word gekenmerk deur heuristiese oordele en kognitiewe afwykings. As hierdie eienskappe aangaande hierdie onderwerp sistematies en bevestig is deur navorsing en literatuur, dan sou dit 'n noodsaaklike deel vorm van die verklaring van die gedrag van finansiële markte. In hierdie studie word 'n metodologie wat gebruik kan word om die impak van kognitiewe afwykings op die finansiële markte te modelleer, aangebeied. 'n ANN as 'n vervanging vir die besluitnemingsproses van 'n belegger word gebruik. Dit is belangrik om daarop te let dat die keuse van die maatskappye, waarop die ANN geleer, gevalideer en getoets sal word, kognitiewe afwyking gedemonstreer het gedurende die voorbereiding van die studie. Alhoewel daar baie kognitiewe afwykings is wat in die literatuur oor gedragsfinansies geïdentifiseer is, sal hierdie studie slegs konsentreer op die impak van terugskouende sydigheid. Op finansiële markte manifesteer terugskouende sydigheid wanneer uitkomste ná hulle reeds plaasgevind het, meer voorspelbaar lyk. Hierdie studie poog en slaag – tot 'n mate – om die obrengs-eienskappe van die tien gekose maatskappye vir die beoordelingsperiode vanaf 2010 tot 2021 te repliseer.

Hierdie studie mag steeds onderhewig wees aan verskeie kognitiewe afwykings en sistemiese gedragsfoute, naas die terugskouende sydigheid. Die verdere toepassing van hierdie tegniek sal verdere navorsing ten opsigte van die invloed van beleggingsgedrag op finansiële markte stimuleer.

Table of Contents

List of Figures

List of Tables

Chapter 1:

Introduction

1.1 Introduction

Stock price forecasting has been a popular discussion for the past few decades. Numerous statistical methods and philosophies have been established to increase and improve the forecasting accuracy of stock prices. The desired objective for this study is to enable readers to be cognisant of the fact that behavioural factors play a much larger, more important role than most people anticipate. Some would argue that cognitive biases have a better capability to forecast stock prices than trends, technical and fundamental indicators. Several researchers have attempted to improve model's accuracy when considering these indicators, without considering behavioural factors and struggle to develop more robust models.

Behavioural finance has become an interesting topic over recent years in the field of finance. The most commonly known theory that would challenge this topic is the Efficient Market Hypothesis (EMH) that was published in the 1960s. Briefly, efficient markets can be explained as the theory which states that rational investors, with full access to complete information, can accurately forecast stock prices, (Fama, 1965).

Behavioural finance is defined as "the study of the impact of psychology on the behaviour of financial professionals and the ensuing outcome on markets", (Sewell, 2007). Behavioural finance theory is supported and is derived from inefficiencies occurring in the market and investor's irrationality when it comes to decision making, (Chaudhary, 2013). The recent curiosity shown surrounding behavioural finance is caused by a rising degree of uncertainty that originate among investors. Human behaviour cannot be easily quantified, and a single theory cannot necessarily understand certain actions and what investors' expectations would be.

The core principles for behavioural finance state investors, by assumption, do not act rationally at all times, but rather make choices based on their emotions and feelings. Investors act naïve to the reality that they show behavioural patterns during trading, portfolio selection, market timing etc.

There have been various popular statistical methods using various sort of techniques, to predict the stock market. Researchers, investors, money managers, etc., use these techniques to try and constantly outperform the market and it has been shown that none of these methods can consistently outperform the market. However, an Artificial Neural Network (ANN) model holds characteristics that provide improved accuracy and results, when compared to other statistical methods available. One of the most desired models to create, would be a model that can analytically incorporate emotional information into a quantitative model to potentially improve the prediction accuracy. The objective of a behavioural model is not always to improve accuracy, but to help the user interpret and explain why accuracy can be improved when incorporating specific behavioural data. If a model can efficiently combine public stock market data with emotional data, what will the outcome be? Can this proposed behavioural model outperform a traditional model?

1.2 Artificial neural network's contribution

An ANN model is implemented in this study. With an ANN, it tries to replicate the process of the human brain, where an ANN is a computer program that is created to learn in a parallel manner to the human brain. The brain continues to learn and evolve by acquiring more knowledge and information through the network by implementing a learning process and would consequently store the acquired knowledge based on interneuron connection strengths, (Li, Ma, 2010). The idea of an ANN model can be explained as a computational network which aim and attempt to emulate nerve cells or neurons of the biological nervous system of humans, (Malekian, Chitsaz, 2021). This study can be enabled by the use of ANN to incorporate input variables in conjunction with the ability to handle large datasets swiftly. Many different models such as linear regression models can be incorporated using an ANN framework by simply adjusting the activation functions or the ANN's architecture.

The framework of an ANN allows it to be applied to stock market data, due to their strong ability to understand existing non-linear mappings between the inputs and output. The practical use of an ANN to analyse portfolio management, credit and loan analysis, predicting stock prices and simulating market behaviour, proves that it is a popular and useful method.

1.3 Objective of the study & motivation

It is important to state that the desired outcome is not to only obtain an efficient and highly accurate model, but to also provide an outcome which can be used to explain the behavioural element that exist in stock prices. It is a challenge to understand investor decision-making and it is also difficult to assess how it influences future stock prices, since most investors –

realistically – do not act rational. This study is devoted to show that investors are vulnerable to cognitive biases, as it plays a big role in the individual's decision making.

Most theory surrounding traditional finance is based on the key assumption that individual investors act rationally and will incessantly consider all the existing information when making an informed investment decision, as they are confident of the outcome and consequently become overconfident.

Investors are not prone to making systematic errors, which will be referred to as cognitive biases. Investors experience cognitive biases which leads to less rational behaviour in investment decisions and overall decision-making. Cognitive biases have been proven to be the root cause to numerous occurrences existing in the financial market.

The current challenge investors in South Africa experience, are that they struggle to make an informed decision, due to the information they need might not always be readily available, transparent, or timely information. Therefore, incorporating cognitive psychology into decision-making may provide investors with the ability to make improved decisions for themselves. "Decisions are non-programmed or heuristics to the extent that they are novel, unstructured, and consequential", (Noor, Hossain, 2005).

Financial markets and cognitive decisions are highly correlated. The purpose of this study is to investigate the effects behavioural finance would have on the outcome of a few selected South African companies' stock prices, by introducing an ANN model that would incorporate the behavioural finance element in its architecture. Many studies have proven that behavioural factors have a tremendous effect on the movement in stock prices.

It is worth noting that it is not necessarily true that a model with hindsight bias would perform better than a standard ANN model. Hindsight bias can lead to overconfidence in predictions and misinformed decision making, which may result in suboptimal performance in the financial market. The performance of a model depends on various factors such as the quality of the data, the complexity of the model, and the evaluation criteria used. A standard ANN model, free of hindsight bias, can also perform well if it is well-designed and trained with appropriate data. It is important to consider the potential biases and limitations of a model when evaluating its performance.

It is essential to highlight the importance of incorporating behavioural finance into financial models, as it can offer a more accurate representation of the decision-making process of investors. The uniqueness of the study is promoted by focusing on the fact that it is one of the first to apply an ANN framework to assess the impact of hindsight bias in a South African context, filling a gap in the current literature. Additionally, highlighting the results of the study and showing how the proposed framework has demonstrated some success in replicating the return characteristics of the ten selected companies could increase the appeal of the study.

1.4 Contribution

South Africa, as an emerging economy, exhibits unique characteristics in its financial markets, including greater volatility and sensitivity to global and domestic events. Analysing behavioural biases in this context can provide insights into how investor sentiment influences stock price movements, offering valuable information to market participants, policymakers, and investors in similar economies.

The study's primary contribution is assessing hindsight bias using an ANN framework and the provision of a novel framework for mimicking the effect of cognitive biases, particularly hindsight bias, on financial markets. By using an ANN as a stand-in for the decision-making process of an investor, the study presents a methodology for modelling the impact of cognitive biases on financial markets. The study's concentration on the impact of hindsight bias and its attempt to replicate the return characteristics of a set of companies provides valuable insights into how investor behaviour affects financial markets. The results of this study will contribute to further research in the field and a deeper understanding of the influence of biases in financial markets.

The key defining, value-adding element of this study is the development of a novel ANN framework that can mimic the effect of cognitive biases, specifically hindsight bias, on the financial market. By doing so, this study offers a new and innovative approach to understanding how cognitive biases influence models and their outcomes in a South African context. This research endeavour seeks to render quantitative the impact of a qualitative factor, such as cognitive bias, and to gauge its potential numerical ramifications on equity prices within the South African market. As a result, the study melds established methodologies with financial market data, thus evaluating the quantitative effects of cognitive bias on projected stock price outputs.

1.5 Research objectives

In this section, different objectives are introduced.

1.5.1 Primary objective

The primary objective is to assess how the study can incorporate behavioural finance and eventually improve the accuracy of an ANN model. The model will be developed using market data from a selected list of Johannesburg Stock Exchange (JSE) - listed companies. The model would test whether the behavioural element added to the model, would provide more accurate results. To assess the suitability of this model, two models are developed; one which includes the behavioural element and one which excludes this element. Both models will be compared to the actual result. Incorporating behavioural finance into financial models is important because it recognizes that human emotions, biases, and heuristics can affect financial decision-making and market outcomes. By including these behavioural elements in the model, the study aims to improve the accuracy of stock forecasting and better reflect the reality of the stock market.

1.5.2 Secondary objective

The secondary objective is to provide a detailed literature review on how an ANN model can be used to incorporate and explain behavioural factors that exist in the stock prices of listed companies. This literature review will help to inform the design and implementation of the study and contribute to a deeper understanding of how behavioural finance can be incorporated into ANN models for stock forecasting.

1.5.3 Conclusion

The primary objective of the study is to assess the potential improvement in the accuracy of stock forecasting by incorporating behavioural finance into an ANN model, while the secondary objective is to provide a comprehensive literature review on the use of ANN models to incorporate behavioural factors in stock prices.

The study is unique and provide valuable context to the South African financial market, especially for the following reasons:

1. Relevance: Hindsight bias is a common behavioural finance phenomenon that can impact financial forecasts and decision-making.
2. Novelty: It explores the impact of hindsight bias on stock market forecasting using ANNs, an area that has received limited attention in the existing literature.

3. Importance: Understanding the impact of hindsight bias on financial forecasts, as it can provide valuable insights for practitioners and inform future research in the field.
4. Methodology: The robustness and rigor of the methodology, and the planning to compare the performance of the two models to assess the impact of hindsight bias on stock market forecasting.
5. Contribution: The contribution that the study will make to the existing body of knowledge, regardless of the results, as it will add to the understanding of how behavioural factors such as hindsight bias can impact financial forecasting models in an emerging market such as South Africa. Finally, the input data used to obtain the predicted output uses stock prices instead of stock returns.

1.6 Chapter outline

This study consists of five chapters. The assignment will be laid out as follows: In Chapter 1, the introduction of the assignment will be outlined, along with the motivation of why this study is done and some key points and concepts are briefly introduced. In Chapter 2 a literature review is provided, where some of the major behavioural finance theories are introduced. In the chapter theory will be introduced which includes the explanation of traditional finance and how this study disagrees with it. In Chapter 3 the methodology for the ANN model is introduced. The theory and why an ANN model is beneficial when compared to other regression methods are introduced. The ANN general framework and architecture will be explained. The parameters used for the learning process are also defined. In Chapter 4 the research design and methodology implemented to the data used for this study are introduced. The empirical evidence pertaining to the behaviour and results found in the stock prices from the selected companies is discussed. Also, a detailed review of the empirical evidence is found as well as a discussion on which models performed well, and which did not. The models' results are assessed with the help of descriptive and inferential statistics to draw conclusions. The chapter presents the findings of this assignment using behavioural finance theory and consequently potentially suggest a more radical approach to the modelling of companies' share prices in a South African context. In Chapter 5 the study and the discussion on how well the models performed are concluded. The findings are summarised, along with the study's implications and limitations and will conclude after mentioning the main contributions to the study and finally suggest any further research possible.

Chapter 2:

Literature Review

2.1 Introduction and Framework

The 21st century has entered a stage where data has become the new gold standard or currency, others even call it the King of the Century. Due to the rapid development and improvement of the internet and information being available by the click of a button, leads to countless, eager investors searching the internet for information before making a decision. This has enabled sentiment analysis to become a viable feature when looking at different investment or stock options available. Investors regards information that is available on news channels; social media platforms and financial markets as valuable, reliable information, which mean that they are exposed to information that was not easily attainable a couple of decades ago.

It is a major concern that prices observed in the market deviate from its fundamental value, largely caused by irrational, noisy investors in the market. Understanding how the 'real-world' works and how people act to information is often neglected when it comes to financial market related decision making. Without even taking into account what the true value of the company is, investors make their own ideas and carry out their own valuations of companies. Understanding behavioural models that investigate this paradigm, is key to this study. Understanding investor behaviour and preferences is a key factor in making investment decisions and is frequently very challenging to quantify.

2.2 Efficient Market Hypothesis

The efficient market hypothesis (EMH) is a traditional economic theory that, at its core, assumes that stock market data simply follows a random walk and that investors make fully informed decisions that take into account all relevant information, (Kumar, 2016). As a result, the stock price would incline toward its "fundamental value.". The fundamental indicators are confirmed to be the main driving force for stock price movements. The EMH states that once new, verified information is available to investors, they would rapidly respond to the change and act accordingly by indicating no form of bias or preference. Therefore, the theory implies it is near impossible for the investors to outperform and earn exceptional excess returns, since the theory clearly states all new information is always incorporated into prices in an orderly manner. The EMH, assume that all investors all have equal access to information and are able to obtain information equally fast, (Ang et al., 2011).

This assumption is incorrect since, for instance, some market participants can learn information more quickly thanks to insider information.

The EMH can be fragmented into three different components, these components were formally introduced by Fama (1970). Fama identified that market efficiency can be defined in three different information sets.

1. **Weak Form** claims that stock prices reflect prior prices and volatility. Information can be adequately conveyed by stock prices and volume alone because trends and historical analyses are weak indications of future prices.
2. **Semi-strong Form:** This is demonstrated by extra sources of public information, where data on top management change/turnover, share buybacks, and earnings announcements play a significant role in market efficiency.
3. **Strong Form:** According to the information, prices would reflect all publicly and privately available information. It indicates that since investors lack "monopolistic access to some information," and are unable to generate positive excess returns, (Fama, 1970).

This study disagrees with the EMH, and results obtained by the study will closely support the first two information sets, to show that there are more factors available to forecast stock prices.

Accuracy in stock price forecasting could be increased by taking into account historical trends and the potential that investors will make bad judgments based on incomplete information. This argument runs counter to Eugene Fama's notion. As many researchers and papers have explained, they disagree with the implementation of the EMH, as researchers are more aware of other factors such as cognitive, psychological factors that could possibly also affect stock prices.

Having said this and introducing the possibility of including behavioural factors into a quantitative model would imply that asset pricing methods would not consider measures such as the Sharpe Ratios and Capital Asset Pricing Model (CAPM). These asset pricing methodologies and techniques depend on the effectiveness of the market and on investors acting rationally and consistently avoiding risk. The CAPM is a very well-liked model that offers many benefits, including the ability to absorb systematic risk, but it is predicated on assumptions that can be quite constrictive. It also assumes that investors would hold a diversified portfolio and would essentially eliminate specific, unsystematic risk. (Zucchi, 2021). The CAPM model has been used in research, particularly in attempts to comprehend potential and probable stock market behaviour. These algorithms have a limited ability to predict stock price changes effectively, though.

A neural network is one of several other robust and adaptive models that might be used when attempting to forecast stock prices. The results of the CAPM model of financial market behaviour will be insufficient, (Mullins Jr., 1982).

It has been shown that the market is in fact inefficient. Jensen demonstrated his point by claiming that aberrant returns are not just caused by the flaws in today's traditional pricing methods, but also by market inefficiencies, (Jensen, 1978).

White used a straightforward neural network to investigate IBM's daily stock return and discovered non-linear regularities in the data, (White, 1988). White proposed to broaden the study's scope and attempt to consider aspects other than the EMH theory, including volume, market indexes, macroeconomic data, and qualitative information like news events, company announcements, and political events. According to the study's findings, while neural network models are not always thought of as "money-making machines" (Qi, 1996), they are successful in helping us comprehend the dynamic behaviour of stock prices.

A neural network is a great substitute for a linear regression model in a dynamic multi-factor model for stock returns, (Refens et al.,1994).

As previously said, many have disregarded the EMH since there is not enough evidence to support it and because it cannot be used to reliably predict stock prices. Given that neural networks still significantly rely on their architecture and design, this does not imply that neural networks are the most likely answer to the problem.

2.3 Stock Market and Anomalies

The term "anomalies" has become well-known as a result of numerous empirical research examining stock market investor behaviour. The window where above-average results are seen is represented by anomalies. When assuming that investors behave rationally inside an efficient market, but outliers would still exist, the following phrases exist.

2.3.1 Long-term reversals:

Investors often assume that when a company consistently reports bad news over a long period of time and performs poorly, their investment will stay a terrible investment. However, research led by De Bondt and Thaler has demonstrated that this is not necessarily the case, (De Bondt and Thaler, 1985). Their research demonstrated that once the business has fully recovered, these "extreme losers" frequently generate higher profits and that one of the most important aspects of behavioural finance that affects investors' decision-making is the market's tendency for investors to overreact to dramatic media news events, particularly in developing, emerging markets where information may be incomplete or inaccurate. Investors tend to overreact to corporate statements that could cause a negative feeling or reaction.

2.3.2 Price Earnings Anomaly

Given that stock prices are skewed, many investors think that the price-to-earnings (P/E) ratio is an indicator of bias. Additionally, data suggests that stock returns with a low P/E ratio would typically produce superior returns when taking into account all underlying risk; nonetheless, this assertion would be at odds with the EMH, (Basu, 1977). The analysis came to the conclusion that the publicly disclosed P/E ratio has some "information substance" and that there appears to be some lag.

2.3.3 Other factors considered

Size, momentum, low-volatility, and value are further considerations made by (Nielson, Nielson & Barnes, 2016:2). These elements may assist in offering a helpful explanation for current anomalies or abnormal returns seen in the market.

2.3.3.1 Size

This factor is based on the idea that companies with smaller market capitalizations may perform better over time than those with greater market capitalizations, (Nielson et al., 2016:3).

2.3.3.2 Low volatility

The EMH is refuted by the long-term performance of low-risk stocks, which beat high-risk ones. Investors are not compensated for taking more risk. The size and sector anomaly, not necessarily the volatility differential, may be the cause of low risk, low volatility stocks outperforming, (Nielson et al., 2016:3).

2.3.3.3 Value factor

When evaluating components like Book-to-Market (B/M), P/E ratio, and high dividend yield, the value factor is technically defined as when an investor purchases a stock that is valued below its fundamental value, (Haitsma et al., 2016).

2.4 Behavioural Finance

"Behavioural finance postulates that psychological factors and biases influence investors' financial behaviours. Furthermore, understanding and predicting significant rises or declines in stock prices can be done using influences and biases as the basis for explaining all different kinds of stock market anomalies, (Hayes, 2021).

The paradoxical version of the conventional economic framework, such as the EMH, is frequently referred to as behavioural finance. Researchers and psychologists have demonstrated that investors are prone to a variety of cognitive errors brought on by heuristics. The overall principle or method used to guide wise decision-making is known as a heuristic. Heuristics are ways for investors to make decisions with little mental effort. In general, heuristics are referred to as "mental shortcuts", (Kahneman, 2011).

This section discusses how investors may respond to market conditions and how they base their decisions on informational judgments. Modern financial theory has been persuaded by behavioural finance theory that investors are affected by emotions like greed and fear and do not always act impartially or rationally. Due to mounting evidence that behavioural patterns do in fact exist in stock trading and their price, behavioural finance has recently attracted a lot of interest.

By taking into account unconventional methodologies, behavioural finance permits studies to adopt a fresh, non-traditional approach to stock price projections, which would in turn produce some valuable contributions to society. One prevalent finding in society is that investors are naive and oblivious of the behavioural patterns of uncertainty they exhibit when trading, choosing a portfolio, timing the market, etc.

In behavioural finance, the concept of uncertainty is a hotly debated subject since it causes investors to make biased decisions when their time and resources are restricted. Eventually, this would raise the possibility that investors will use representativeness that is similar to a heuristic, (Hirshleifer, 2001).

2.4.1 Behavioural biases that exist among investors

When making financial decisions, investors are known to engage in a variety of cognitive biases. The various behavioural biases that can be found in the financial market environment are described in this section.

2.4.1.1 Hindsight bias

When researching behavioural finance, this may be one of the most significant biases to investigate, if not the most significant one. A psychological phenomenon known as hindsight bias enables people to delude themselves into believing that they were able to foresee an outcome properly or mistakenly before it actually happened.

People may believe they can precisely forecast the fate of similar events as a result of this, (Chen, 2022). In the world of finance, this occurs when a shareholder identified—but refrained from purchasing—a stock that was underperforming and used hindsight to determine whether they should have made the purchase. They have no doubt that they could prevent any other losses in the future in addition to this particular one. Even when considering an overperforming stock, the investor would still be biased toward the past. The investor avoided huge profits and used hindsight by realizing that he or she may lose out on potential future gains if, for some reason, they did not purchase the overperforming stock after identifying it.

It is known that markets do not perform efficient nor rational. In 2018, the bitcoin bubble implosion was foreseen by the log-periodic power law model eight days beforehand. However, this does not imply that it will be able to anticipate when future bitcoin bubbles will burst, but it is still an intriguing finding in hindsight.

2.4.1.2 Representativeness bias

The representativeness bias is the tendency for people to overestimate current trends and events in datasets while undervaluing population traits, (Kahneman and Tversky 1982). It causes people to overreact and underreact, which would lead to market abnormalities. Corporate announcements, newsworthy events, and investment decisions made in haste or under the influence of emotion can also result in anomalies.

Where a pattern may not always exist, representativeness looks for it. Investors rely on this feeling and predict that recent news and events will have an impact on the company's stock price or even industry pricing. The possibility for long-reversal and overreaction eventually results in poor decision-making that is based on incomplete information. When faced with uncertainty and a lack of time, information, or resources, investors tend to make biased decisions. Investors are compelled to use heuristics like representativeness as a result.

Incorporating a sufficiently large sample size and taking into account the possibility that the information at hand may not always represent the whole truth or provide the most accurate forecasts are two common problems that people encounter, (Kahneman, Tversky, 1974). The lack of sufficient data or a small enough sample size may frequently be caused by the inadequate resources available and might lead investors to quickly act on new information, when received.

Mullainathan, (2001) presupposes in his paper on Behavioural Economics that investors can be divided into Bayesian and non-Bayesian subcategories. "Investors are not Bayesian, that they think in categories, and that they assume the most typical scenario, which is ignoring(underweighting) realistic states of the world".

Mullainathan groups investors into the following:

1. News viewers, who will represent non-Bayesian people, tend to underreact to new information.

2. Traders who use momentum extrapolate prior trends and performances of price changes, which may or may not result in overreaction.

2.4.1.3 Confirmation Bias

Confirmation bias is the tendency to focus on information that would only confirm or support preconceived assumptions rather than actual information that would offer an alternative interpretation.

2.4.1.4 Look-ahead Bias

This is referred to as the bias that occurs when the study's dataset was unknown during the time it was being analysed, which means the data were not easily accessible at the time they were used, (Kenton, 2020). When using the actual future information in the training data set, there is a look-ahead bias. It is important to carefully choose the model framework for this investigation in order to prevent the underlying look-ahead bias. Look-ahead bias refers to the situation where the model uses data that won't be easily accessible at the time of assessment, thereby producing unreliable results.

2.4.1.5 Herding Bias

Herding occurs when investors make investments in the same way that other investors do (Graham, 1999). Investors frequently mimic the actions of their fellow investors or trusted portfolio managers. Investors occasionally chose to disregard their personal information and rely only on the consensus. Some traders may even emulate the actions and behaviours of the most profitable deals made within a particular time period. Investors that use this method base their decisions on the success of the prior period. Investors would then anticipate comparable outcomes for the subsequent period.

2.4.1.5 Recency Bias

Recency bias is the propensity to overemphasize recent experiences, even when they are not the most accurate or pertinent, (Aguilar, 2021). When investors focus their decisions on the most recent occurrences and anticipate that a similar pattern will continue, this is known as recency bias. This might be anything from sharing information on your neighbour's investment to following a social media page offering free investing advice.

2.4.1.6 Self-Attribution Bias

Self-attribution bias is the propensity for people to credit their own abilities for accomplishments while attributing outside forces for failures, (Hoffman 2014). This bias describes those who, in defiance of all preceding information or research, back their own opinion on a certain stock selection.

This notion and overconfidence go hand in hand. Investors' confidence might rise if they learn new information that supports their personal beliefs or signals.

2.4.1.7 Anchoring

When making judgments about a particular stock in the future, people often unconsciously utilize irrelevant information, such the stock price, as a fixed reference point (or anchor), (Hayes, 2021). This practice is known as anchoring.

According to evidence, investors would 'anchor' their decision to their original attitude rather than changing it accordingly. Anchoring is based on the premise that investors create initial sentiments and guesstimates about a particular investment. This would mean, in a financial sense, that investors would expect a company's earnings to continue performing effectively at their anticipated level without any stagnation. The investor also anticipates that the earnings will follow the noted historical trends. Because of their current anchored positions, investors frequently underreact to large earnings reports, corporate statements, or any news about the company that is alarming.

2.4.1.8 Overconfidence

There is a lot of this bias among investors. Investors would attempt to reach their own conclusions and develop their own "private" buy or sell signals in a market like South Africa where information might not always be reliable or timely. However, these investors take a risk when new information surfaces, and they occasionally hesitate to act on it since their professional interests trump their personal opinions. This is closely connected to anchoring.

2.4.1.9 Overreaction or Underreaction

Barberis et. al., (1998) used a case study to illustrate the concepts of overreaction and underreaction. In order to control investors' expectations when given information, the article provided a model of investor sentiment. “People pay too much attention to the strength of the information given and too little attention to its statistical weight," the article concluded. This is especially true for elements like corporate announcements, which despite their modest strength, significantly influence the final result, (Barberis et al., 1998).

Some basic news events, such as earnings announcements, tends to get a subpar response from investors. Prices frequently increase when specific news is released, especially when it is positive. Long-term reversal and several accounting ratios are directly tied to overreaction because companies with high, positive performance and consistent excellent news eventually revert to the mean.

In 1974, Kahneman and Tversky established the underreaction theory. Investors exhibit underreaction when they do not properly factor all available information into the valuation of stocks and do not respond appropriately to newly received information (Hong & Stein, 1999). In the sense that underreaction can lead to delayed overreaction, overreaction and underreaction can also be effects of one another, (Chan et al., 1996).

Pederson (2015) defines a trend cycle, comprising of distinct stages that elucidate investors' behaviours in response to evolving market conditions:

1. Commencement of the trend: Investors display a delayed reaction at this point, which causes them to make slow, gradual adjustments to their investing decisions.
2. Middle part of the trend: Investors exhibit a protracted response as they gradually adjust to the developing market conditions, causing long-lasting fluctuations in stock values or other assets.
3. End of the trend - Reversion to fundamental value: When trends emerge, the market has a tendency to revert back to its fundamental value, which can signal a correction in asset prices with respect to their underlying fundamentals.

2.4.1.10 Earnings Surprises

Earnings surprise is a general term for announcements and shocks made after earnings. It can be characterized as a time when a company's stock is enjoying unusual returns after a recent announcement, (Bernard, 1992). When a company announces surprisingly large or low earnings, it is known as an earnings surprise, (Chorida & Shivakumar, 2006: 628). These earnings surprises are frequently represented as a subset of the market's underreaction to information already available related to earnings releases, which may in turn cause people to underreact to earlier data, (Bernard & Thomas, 1989, 1990).

2.5 Technical Analysis

Since technical and fundamental analysis are simply based on historical stock price movements, utilizing these approaches to forecast stock market prices alone does not ensure an accurate result, (Dami, 2021).

Technical analysis takes market prices, fluctuations, and volumes into account in order to comprehend and forecast the dynamics of stock values. It places more emphasis on movements than on profits and market share. (It may be interesting to investigate how effective volume pattern heuristics are). The recent progression in technology development that has allowed for the creation of more reliable and complex models has made this endeavour possible. Neural networks are one of the most recent machine learning methodologies and approaches developed and utilized.

Technical analysis serves as a function that effectively combines personal, political, and economic events into a well-functioning model to forecast stock values, making it a very helpful and well-liked instrument to use, (Achelist, 1995). Technical analysis takes use of observable stock market trends and data, and it is essential to have access to historical price data since it assumes that the past will repeat itself.

There are numerous models to use that might be able to forecast changes in pricing. The goal of a prediction model is to merely attempt to highlight some market inefficiencies. Theoretically, these models wouldn't work in a market that is 'fully efficient'. There are various methods that can also be introduced, such as sentiment analysis that analyse earnings announcements and Tweets sent by Political leaders or economists.

The EMH contends that stock prices and returns ought to represent all information as soon as it becomes available. Any information that can potentially have an impact on stock prices is therefore already factored into the pricing. Technical analysis runs counter to the EMH since trends are set off by investors' ongoing shifts in attitude. Because technical analysis is subjective and market players interpret trends differently, it is both widely used and highly contentious.

2.6 Time Series Forecasting

The South African stock market is inefficient and vulnerable to many factors due to its complex, random, non-linear structure and this makes stock price forecasting quite challenging.

When predicting the stock market, there are two viable strategies: qualitative and quantitative. The qualitative approach makes use of the effective forecasting techniques employed by market participants and industry experts. The analytical quantitative method simulates many scenarios and results using computational and algorithmic tools, (Zhuge, Xu and Zhang, 2017).

Time series forecasting aids in decision-making and may assist investors choose less risky positions. Utilizing past data, time series forecasting can be used to identify prospective trends in datasets and, as a result, anticipate future stock prices.

There are many different kinds of time series models, some of which are applicable to finance. Regression techniques, such as neural networks for example, have been utilized and compared to other. Overall, time series forecasting offers fair accuracy for short-term predictions, but as the length of the prediction lengthens, the accuracy can drop down significantly. (Lawrence, 1997). An article by Khoa, Sakakibara, and Nishikawa, focused on using back propagation neural networks to predict stock values.

According to their analysis, a stock price depends on a number of elements that may be known or unknown, (Khoa, Sakakibara and Nishikawa, 2006).

Elaal et. al., (2013) presented a method for forecasting multivariate-factors fuzzy time series that is based on the idea that fuzzy clustering will be able to handle real-world Multivariate forecasting issues. Because stock market information is erratic and often non-stationary in nature, this approach failed to handle it adequately.

Shen, Guo, and Wu, (2010) used a model of radial basis function Neural Network to try and predict stock market indices in the SSE (Shanghai Stock Exchange). In a nutshell, a radial basis function uses linear term combinations based on a single univariate function to approximate multivariate functions, (Buhmann, 2010). They discovered during their research that there was a significant relationship between certain technical indicators and stock indexes. The study came to the conclusion that technical indicators are useful for analysing and forecasting stock index movement. But they also pointed out that non-quantitative elements like political announcements, social media, and any other psychological elements might have a significant impact on stock price changes.

2.7 Conclusion

It is beneficial to this study if key concepts and principles are discussed and defined. Without the aid of investor or general public questionnaires, this study will attempt to include behavioural finance theory in an alternative way. Instead, the study will model and anticipate how investors respond and how these results might help to reliably forecast the fate of a company's stock price using statistical theory, such as an ANN. The data set will be altered to reveal the fact that it contains inherent hindsight bias and so demonstrate that most stock market data, and particularly that which is available on Yahoo Finance, contain embedded hindsight bias.

Behavioural finance tries to overcome some of the limitations in the financial market, since it models investors as “less than fully rational” and therefore the study uses systematic mental biases to try and incorporate this existing irrationality. Although this study will attempt to represent investor irrationality, the framework as a whole is still based on the idea of rationality, which poses a significant limitation in this literature. As a result, the study will continue to employ a very basic model of investor behaviour.

Chapter 3:

Artificial Neural Network – Application and Theory

3.1 Introduction

Over the past ten years, there has been an increase in research on artificial intelligence. In the discipline of artificial intelligence, computer programs are created to resemble or replicate human intelligence. An effective, cutting-edge computer program called a neural network makes intelligent decisions by taking into account several related variables. A neural network model is capable of capturing intricate connections between several variables. A neural network can be used in a wide variety of applications. It gains knowledge from historical stock price data and seeks to identify factors that could improve or worsen accuracy when adding more data to the model. In order to learn from the model itself and to understand what works and what doesn't, performance is generally enhanced by feeding the model new data and by studying historical stock price patterns.

A neural network's architecture can be compared to the structure of the human brain, where neurons are also arranged in layers. Complex connections are used by neurons in the human brain to communicate. Weights associated with these "connections" might be both good and harmful. In this regard, a neural network is extremely comparable since it will attempt to mimic the learning process of the human brain using sophisticated algorithms, (Han et al., 2015).

One of the most modern artificial intelligence techniques, the artificial neural network (ANN), will be used in this research. An ANN is best described as a mathematical model which has the capability to mimic the processing capability of the human neural network system and it has also recently been regarded as one of the most effective techniques for pattern classification, (Zhuge, Xu and Zhang, 2017).

A set of input variables will be used in the conventional method of modelling an ANN in order to forecast a selected target variable(s). The selected variables are capable of capturing impacts that can accurately forecast the target variable.

Contrary to common opinion, stock market data is non-linear and extremely complex in nature. As a result, adopting a linear model for forecasting may not be a reliable option because a linear model cannot recognize the non-linear relationships present in historical financial data. As a result, ANN is an appropriate modelling technique for non-linear dynamic systems like the stock market.

It also has the ability to extract meaningful information from ambiguous and complex data, as well as to identify patterns and trends that are too complex for people or other conventional statistical methods to recognize.

According to Zahedi, ANNs have improved qualitative approaches for financial economic systems more than standard quantitative methods, which are unable to effectively quantify stock market data due to its complexity, (Zahedi, 1993). A huge volume of non-linear, non-parametric data can be processed by an ANN model, which can also learn from experience and use that knowledge to accurately detect patterns of behaviour in the incoming data. An ANN's remarkable ability to "learn" from the input data and evolve over time is made feasible by a lengthy sequence of trials that enable the optimization of the weights connecting inputs to outputs through intermediate layer neurons, (Parisi et al., 1990).

An ANN can maximize the stock market data provided because it allows the data to decide the structure and parameters of the ANN itself, without making any "restricting" assumptions or imposing any limits. Assumptions in the stock market are extremely subjective, (Qi, 1996).

The idea can be explained with the following illustration:

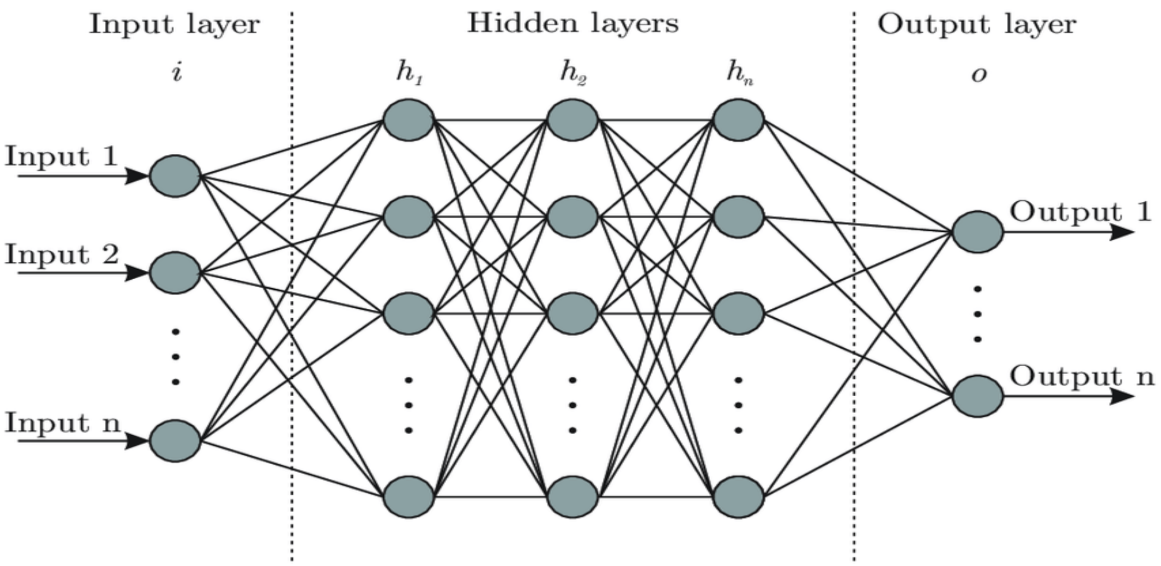

Figure 3.1: Artificial Neural Network Architecture

Source: (Bre et al., 2018)

The input layer, hidden layers, and output layer are the three main elements of an ANN architecture that are graphically represented in Figure 3.1. The input layer serves as the initial point of data entry, receiving input features such as historical stock prices, or any other relevant financial indicators. The hidden layers, which are intermediate layers between the input and output layers, contain numerous interconnected neurons. The neurons carry out complicated calculations and data transformations on the input, enabling the ANN to discover intricate patterns and connections in the financial data. The forecasted output is then generated by the output layer using the learnt patterns from the hidden layers.

Throughout this paper, the ANN architecture will be discussed comprehensively, including the specific activation function used in the neurons, the number of hidden layers and neurons, and the training algorithms employed to optimize the model's performance. Additionally, how the ANN incorporates the concept of hindsight bias will be explored, and its impact on forecasting accuracy will be examined.

3.2 LSTM (Long- Short Term Memory) as an ANN-variant

An efficient technique for time series and non-stationary data sets is an LSTM neural network. The LSTM's architecture can be utilized to predict the return for the upcoming period. The neural network is given the capacity to memorize information by the LSTM. For sequential datasets, this capability offers useful results. As a superior substitute for the Autoregressive integrated moving average (ARIMA) model, LSTM is well-liked in the financial sector. Long-term reliance can be found in stock market data, and because the LSTM – a recurrent neural network - has the ability to learn the long-term dependence, this model can provide predictions that are more accurate. A recurrent neural network (RNN) incorporates past predictions and outcomes into their current decision-making process, mimicking how individuals change their beliefs based on new information.

However, this study would follow a proposed DNN (Deep Neural Network)-ANN approach as DNN models, including feedforward networks, provide insights into how different features influence the output, and would allow for some level of interpretability. This could be crucial in financial analysis to understand which features drive the predictions. LSTMs, due to their complexity, might be more challenging to interpret in this context and can be suggested to be implemented in future research endeavours.

This research assignment aims to understand the impact of hindsight bias on financial markets and involves using historical data to simulate this bias, a DNN could be a more appropriate choice. DNNs can capture nonlinear relationships in the data, which is crucial when modelling behavioural biases that might not have a simple linear effect. DNNs offer a balanced trade-off between model complexity and capturing meaningful patterns in this specific context.

3.3 Deep Neural Network (DNN) as a proposed ANN approach

DNN or also commonly known as a Multi-Layer Perceptron (MLP) is a form of ANN that consists of multiple hidden layers of interconnected neurons or nodes.

These hidden layers, along with the input and output layers, form a deep architecture. The term "deep" in DNN refers to the depth of the network, meaning it has many layers. The DNN model is suitable for handling tasks with high-dimensional and non-linear relationships.

The neural network used in this study can be beneficial for capturing hindsight bias within stock market data, since DNNs can automatically learn relevant features and representations from the input data during the training process. This capability allows the model to identify key factors and behavioural elements that might contribute to hindsight bias in stock market predictions. Furthermore, the DNN's architecture, with its multiple hidden layers, can assist in capturing temporal dependencies in time series data like stock prices.

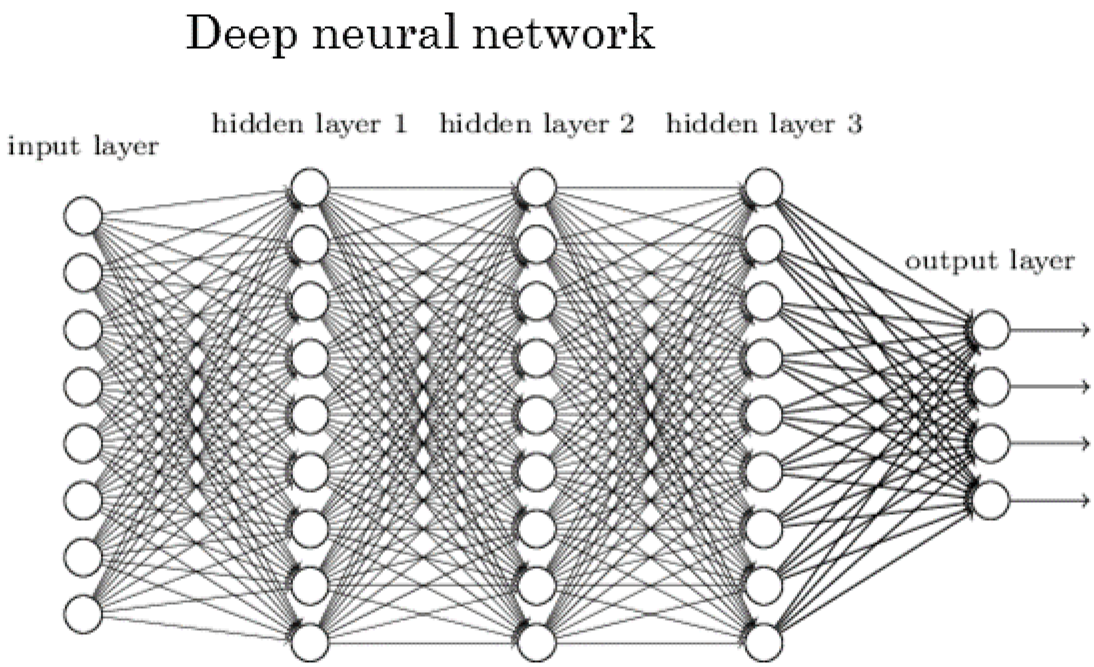

Figure 3.2: Deep Neural Network with more than one hidden layer

Source: (Yang, 2018)

Figure 3.2 can be explained as follows, each neuron in the hidden layers does a weighted sum of the previous layer's inputs. Following the weighted sum is an activation function, which introduces non-linearity into the model. Each neuron's output is then passed on to the neurons in the following layer.

The lines in the figure represent the connections between neurons in adjacent layers. Each connection has a corresponding weight, which defines the connection's relative strength. During the training process, the DNN adjusts these weights to minimize the loss function and generate accurate forecasts. The input data travels through the hidden layers to the output layer during forward propagation, and the model produces predictions based on the learnt weights and biases. Backpropagation is used by the DNN during training to change the weights depending on the error between the predicted outputs and the true labels. This approach refines the model's parameters iteratively to improve its performance.

The activation functions add nonlinearity to the model, allowing it to capture complicated data interactions. The DNN would collapse to a linear model if activation functions were not present.

3.4 Potential risks considered

Overfitting an ANN model is a potential risk. Testing the trained ANN using a hold-out sample or evaluating the trained ANN using data that was not used in training it, is crucial to avoiding overfit. The out-of-sample performance is the main factor that determines whether the chosen ANN model is useful, (Qi, 1996). After mentioning the potential of overfitting, the risk will be addressed and eliminated during the presentation of the study's findings. Artificial Neural Networks (ANNs) have a high capacity to memorize training data, but struggle to generalize to new, unseen data. To prevent overfitting, techniques like early stopping or regularization can be used.

The likelihood that the data set used for this study may be partial or contain some missing data, which may cause noise owing to market inefficiency that exists in the South African market, is another extremely significant risk to take into account. To get the output, one might need to choose the data frequency. An ANN is more robust when coping with missing and erroneous, partial data since it is free of statistical assumptions, (Li, Ma, 2010). Therefore, the risk that the data might be missing, or partial is minimal.

Many would regard ANNs as "black boxes" because of their sophistication and difficulty in interpreting the results. The nature of the model’s complexity might also pose as a significant risk. However, if the neural network's relationship between its inputs, weights, and outputs are clearly defined despite its complexity, the risk will be mitigated. This enables the study into this "black box" to proceed, (Li, Ma, 2010).

3.5 ANN Benefits

Large amounts of noisy, partial, or incomplete data that are non-linear in nature can be handled effectively by an ANN. This model is adaptable and reliable, which makes it valuable in the financial sector where the stock market is erratic and volatile.

The training procedure is really simple. It is thought to be a simpler model to use than the majority of traditional statistical techniques. Setting up and solving a problem using an ANN is simple due to the methodology's simplicity.

As opposed to other sophisticated computer programs like expert systems, ANNs can extract rules without needing to explicitly define them. This gives ANNs an edge. A sophisticated computer system that employs artificial intelligence to replicate human behaviour and solve complicated problems using both facts and heuristics is known as an expert system. Expert systems are specialized and extremely adaptable, whereas ANN models are learned.

An ANN learns on unstructured data and produces an outcome, unlike expert systems, which depend on facts and rules to do so. This is a critical element in the stock market, a highly chaotic environment. It is challenging to take data from expert systems and define it in a way that expert systems can use it. Expert systems perform best when the information is thorough and accurate, and within their specific field of expertise. Neural networks are better at handling changing input, generalising, and making "informed guesses." In light of this, expert systems are less suitable to the stock market setting than neural networks, (Banger, 2022).

An ANN model is endowed with a limited number of hidden units that can accurately approximate any continuous function. ANN routinely outperform linear regression models in terms of forecasting accuracy, (Lee and Ho, 2001). Due to their inability to successfully incorporate innate noise patterns, fat tails, random walks, and the non-linearity of financial data, conventional time series analysis based on stationary process - regression methods, and ARIMA models do not produce results that are as accurate as ANN's when applied to stock market data. ARIMA models are often used to identify the linear structure of data, where the patterns of the ARIMA models are directional, (Ayub, 2020).

3.6 ANN Limitations

While providing all these benefits, ANN also has some disadvantages. The disadvantages are best described as follows. An ANN is adaptable, but because connection weights are not always straightforward to interpret, it is unable to explain the outcomes. It is especially true in financial markets where trends can develop. Large datasets would be needed for the model, so training might take a long time. It is challenging to evaluate the significance of input data and how the ANN obtains its outputs from it, (Lawrence, 1997). This is a significant restriction. Another disadvantage is the paucity of research on determining the optimal ANN structure.

3.7 General Artificial Neural Network Structure

A Deep Neural Network (DNN) will be used for this investigation. An ANN with several layers between the input and output layers is called a DNN. Only feed-forward neural networks are allowed in this model. There must be two or more hidden layers for it to qualify as a DNN. The choice of the number of hidden layers and the number of neurons within each hidden layer define an ANN. There are two hidden layers in the model being employed for this investigation. This model is also considered to be a time series model because it will take sequential data into account.

3.7.1 Layers and Neurons

A network of interconnected processors called neurons constitute an ANN model, which performs a summing operation. The weights on the connectors hold information, (Tan, 1997). It is crucial to stress the importance of simplicity when presenting this model's design in order to avoid making learning too difficult. Input, hidden, and output layers are present in all ANNs' layouts and structural components. Due to the connections between the neurons in each layer and those in the next layers, information is processed in parallel. The layers can be described as follows; the dataset's information is represented by the input layer. An output layer is the model's final step after a hidden layer has detected and projected the relationships within the dataset. The output layer is also known as the target variable.

According to research, it is sufficient to predict recurrent non-linear functions with just one hidden layer. The main restriction on this principle is that a large number of neurons must be used, which would slow down learning. It may be advised to employ a 2-layer DNN for training to remember the long short-term memory if the dataset exhibits high dimensionality and frequency.

According to this principle, if too many neurons are added to the single hidden layer when determining how many should be assigned to it, the model may be able to predict historical data but may not be able to predict output accurately because the input data's influence on the model has been diminished. The model might not be able to learn effectively in the opposite scenario, where there are few to no neurons. The number of neurons inside the hidden layers will continue changing depending on the ANN model.

The input data are either outputs from earlier neurons or come from the original dataset. Each input variable corresponds to a single output from a prior neuron or a variable from the dataset. The input variable consists of 30 observations, which represent a look back period of 30 days, i.e., the 30 most recent days' stock closing prices.

The input variable for the study is therefore the next 30 days of stock prices for each of the selected companies. Once the output is obtained, stock returns are computed. The stock returns will be subjected to statistical analysis shown in Chapter 4. Stock returns, which can be stated as a percentage or ratio, are a measure of the change in the value of a stock over a specific time period.

When using stock prices as input for the ANN model, the model can analyse the relationship between the stock prices of the selected companies and the future stock prices. By analysing this relationship, the study can assess the impact of hindsight bias on stock forecasting in the South African equity market.

The choice of using the next 30 days of stock prices as input is based on the idea that past stock prices can provide valuable information about future stock prices. By analysing this information, the ANN model can identify patterns and relationships in the data, allowing the study to make more accurate predictions about the future stock prices of the selected companies.

Up to the point at which the design is introduced, explained, and optimized, the sequential model will continue to build layers. On the top two layers, the ReLu activation function will be applied, (Brownlee, 2022). The output layer will only contain one node, with the first hidden layer having 30 nodes, the second layer having 8 nodes. On the training dataset, the ADAM optimizer will be used to optimize the training for 50 epochs. The ADAM optimizer will be discussed in Section 3.7.4. An Epoch is defined as the number of passes taken through all the rows within training dataset to train the model. The batch size is set to 10 where batch size refers to the number of training examples that would exist in a single pass, before weights get updated. After training all the parameters, the study simply uses standard simulations methodology that yields a long-term forecast by using the output from one simulation as an input into the next.

Last but not least, it should be highlighted that the study never encountered any significant problems with the ANN model's run time (i.e., the time required to complete "one forward pass") or the time required to forecast the future period stock price and assess its accuracy.

3.7.2 Loss Function

A loss function is a mathematical function that is used to measure the difference between the predicted output of an artificial neural network (ANN) and the actual output. The goal of training an ANN is to minimize the value of the loss function, so that the predictions of the network are as close as possible to the actual output.

When building an ANN architecture, the loss function is important since some functions may be more suitable for use with financial data than others. The error is often referred to as the loss. The error shows how effective the model is. Therefore, the model's performance improves as the error decreases. Since the predicted output is minimised and the observed output will be repeated iteratively across the training set, a loss function or error function will be produced. Therefore, the more you iterate, the more you train over your data set, the more you optimize your parameters and the lower your error rate.

The choice of loss function depends on the specific task and properties of the data. For example, if the task is a regression problem, mean squared error (MSE) is a good choice, if the task is a classification problem, cross-entropy is a good choice, and if the task is a multi-class classification problem, categorical cross-entropy is a good choice.

There are several commonly used loss functions in ANNs, each with their own properties and use cases. For this study, (mean absolute error) MAE and (root mean squared error) RMSE are introduced and defined.

3.7.2.1 Mean Absolute Error (MAE)

Without taking into account their direction, MAE calculates the average size of errors in a set of forecasts. All individual differences are equally weighted in the test sample's average of the absolute disparities between prediction and observation. Where, y_i denotes the realized value of the process, while $\hat{y}_i$ its forecasted value,

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |\hat{y}_i - y_i| \tag{3.1}$$

3.7.2.2 Root mean squared error (RMSE)

A quadratic scoring rule called RMSE also calculates the average error magnitude. It is the average of the squared discrepancies between predicted results and actual observations. RMSE is a helpful statistic since it provides a lot of weight to huge errors encountered; as a result, it becomes useful when errors turn out to be undesirable. The RMSE metric is a very common loss function for regression problems.

The fact that both MAE and RMSE reflect average model prediction error in units of the relevant variable highlights their similarities. Both measures are unaffected by the direction of mistakes and have a range of 0 to ∞. If RMSE = 0, the realized and forecasted values are identical. If it is larger than zero, the model exerts some error between the two values.

Mathematically, it can be represented as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}{n}} \tag{3.2}$$

$$RMSE \in [0, \infty),$$

n = number of observations in time series.

The models are compared to the actual results using the values derived from these metrics.

The two measures, MAE and RMSE are useful measures for assessing model performance and evaluating predictions. Only when the results are shown in the same units can these metrics be applied. One should also take into account if the data has been overfitted when evaluating the reliability of the error's metrics used in the training procedure. The model starts to overtrain when the validation error starts to rise. This never occurred by any of the models that were shown, hence no model was overtrained.

The best way to establish whether a model has the capacity to predict the future is to evaluate its performance through validation. Because of the variations in sample size and splits, validation results might not always be accurate. The trained ANN is validated similarly to statistical regression, where the 'goodness' of the ANN is determined by how well it forecasts the test set using data that it has never seen before. The study created the test set to reflect 10% of the entire dataset across all six models in order to compare the six distinct models.

The suggested model's goal is to produce and acquire better and improved results by improving the loss values, RMSE, and MAE when hindsight bias is included in the training set.

3.7.3 Transfer Function

The transfer function, which specifies the output from a summation of the weighted inputs of a neuron, is typically non-linear and also contributes to the network's non-linearities. It is used to determine the output of a neuron given its input. Arguably, any mathematical function can be used as a transfer function. Most commonly a sigmoid function can be used, along with a logistic and hyperbolic tangent function, which are expressed as:

Logistic Function:

$$f(x) = \frac{1}{1 + e^{-\lambda x}} \tag{3.3}$$

- Restricted to (0:1)

Hyperbolic Tangent Function:

$$f(x) = \frac{2}{1 - e^{-2\lambda x}} - 1 \tag{3.4}$$

- Restricted to (-1:1)

The Rectified Linear Unit (ReLu) function is thought to be the most appropriate due to the non-linear nature of the study's input data. It is extensively used in multi-layer neural networks as a non-linear activation function and is referred to as the default option for neural networks. Another reason for opting to use the Rectified Linear Unit activation function for this regression problem is that the result variable is numerical rather than categorical.

The ReLu function can be expressed as:

$$f(x) = \max(0, x), \tag{3.5}$$

where the derivative of the ReLu is,

$$f'(x) = \begin{cases} 1, & x > 0 \\ 0, & otherwise. \end{cases} \tag{3.6}$$

Equation 3.6 shows that its derivates are both monotonic, which means that if the function receives a negative input, it will return zero, and if it receives a positive input, it will return the value. When using this function in a neural network, x represents the input of the model and f(x) represents the value that is passed through to the final output or to the input of the next layer. The ReLU function is computationally efficient and easy to calculate. It also helps to introduce non-linearity in the network, which allows the network to learn more complex patterns in the data.

The function may run into the vanishing gradient problem when employing Sigmoid or Tanh in the hidden layers, which prevents learning from earlier layers if the ANN is backpropagating. Equations 3.5 and 3.6 show that the ReLu function is simple and will not require a lot of computation power to incorporate into the model, so the model will take less time to train.

ReLU is commonly used as a transfer function in the hidden layers of an ANN, but it has some drawbacks as well. The main problem is the "dying ReLU" issue, which occurs when the input to the neuron is always negative, causing the output to be always 0. This can cause the neuron to become inactive and stop learning, which is not desired. To mitigate this issue, the Leaky ReLU function is introduced, which is similar to the ReLU function, but with a small negative slope for negative input values. This was not included, to favour the computational simplicity of the model.

3.7.4 Optimization

Optimization is used in neural networks to adjust the weights and biases of neural networks in order to minimize error in the model's predictions. The process is carried out on the model's training data set, as the model uses this data to learn the most optimal weight values. Once trained, the model can be applied to a test set to evaluate its performance, (Hurra, 2020).

The effective stochastic gradient descent (SGD) algorithm known as ADAM is what is referred to as the algorithm optimizer (Adaptive Moment Estimation). This algorithm is tailormade for deep neural network training as it can train a model as quickly and effective as possible. It can be regarded as an extension to SGD.

The ADAM optimizer is a machine learning algorithm that uses weight decay for regularization. Smoother, simpler functions that generalize better than spiky and noisy ones are critical for neural networks to learn. Weight decay is a simple and effective regulariser that uses a small amount of force to gradually push the weights toward zero, (Rita, 2022).

This stochastic optimization technique only requires first-order gradients with little memory requirement, (Diederik and Ba, 2015). Stochastic optimization is characterised as a function with randomness present to finally optimize the objective. Firstly, SGD is suitable for a model with a lot of data and parameters and can estimate the gradient using a random subset of that data, (Alabdullatef, 2020).

The ADAM optimizer utilizes the concept of adaptive learning rates and momentum to efficiently update the model's parameters during the training process. In the algorithm below, the parameters are defined as:

- β_1 and β_2 are denoted to the power t, β_1^t and β_2^t and are hyperparameters for the exponential decay rates of the moment estimates (set to values close to 1, i.e., 0.9 and 0.999, respectively).
- ϵ is a small constant (a very small value, 1e-8) added to the denominator to prevent division by zero.
- g_t represents the gradients of the loss function with respect to the parameters at time step t.
- m_t and v_t are the first and second moment estimates, respectively, accumulated over time.
- α is the learning rate, which determines the step size for parameter updates.

The algorithm for ADAM can be explained as follows:

1. Initialize the time step t = 0 with the initial parameter vector θ_0, the first and second moment vectors, $m_0 \ and \ v_0$.
2. Each time step t is updated incrementally with $t = t + 1$, the gradients of the loss function is calculated with respect to the parameters, denoted as g_t;

$$g_t = grad(\theta_{t-1}). \tag{3.7}$$

3. Update equations for the parameters using ADAM:

   The first moment estimate (mean of the gradients) with exponential decay rate β_1 is computed as:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_{t.} \tag{3.8}$$

The second moment estimate (uncentered variance of the gradients) with exponential decay rate β_2 is computed as:

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2. \tag{3.9}$$

4. The bias of the first and second moment estimates are corrected:

$$\widehat{m}_t = \frac{m_t}{(1 - \beta_1^t)}, \tag{3.10}$$

$$\hat{v}_t = \frac{v_t}{(1 - \beta_2^t)}. \tag{3.11}$$

5. Update the parameters using the adaptive learning rate α:

$$\theta_t = \theta_{t-1} - \frac{\alpha}{\sqrt{\hat{v}_t} + \epsilon} \widehat{m}_t. \tag{3.12}$$

By adaptively adjusting the learning rates for each parameter based on the historical gradients (first moment) and the squared gradients (second moment), ADAM can converge faster and efficiently navigate complex loss landscapes, making it a popular choice for training deep learning models.

The optimiser is incorporated into the model to alter an ANN's attributes, in this case, the weights, in an effort to reduce error and improve the model. The ADAM optimiser is a powerful optimisation technique for large problems with a large amount of data and parameters that necessitate extensive data processing. An exponentially decaying average of previous gradients is used by the ADAM optimizer. The model can get an adaptable learning rate because of this. As a result, the model would be able to handle sparse data.

3.8 Back-propagation

Determining appropriate weight values is crucial for DNN effectiveness. To achieve this, loss optimization aims to discover a set of weights, that minimizes the computed loss function. For models with a single weight component, charting the weight and loss on a 2-dimensional graph is possible; however, the majority of DNNs include several weight components, making the visualization of n-dimensional graphs difficult.

To avoid this, the direction of maximum ascent will be found by taking the loss function's derivative with respect to all weights. Understanding this direction, the model descends the gradient in search of a point of convergence at a local minimum. After completing this descent, the DNN obtains and employs a set of ideal weights. In an effort to enhance the model's performance, a hidden layer employs backpropagation to optimize the weights of the model's inputs, (Grogan, 2020).

The process of computing the derivative is known as backpropagation, which fundamentally utilizes the chain rule from calculus. The effect that a slight modification to the initial set of weights has on the overall loss using a neural network with multiple layers will be investigated. This information is captured by the derivative, also known as the gradient, (Joseph, 2020).

The first set of weights, however, is input into a hidden layer, which then influences the anticipated output and the loss through a different set of weights. Therefore, seeing how weight changes affect the hidden layer, need to be taken into account.

3.9 Model Framework

Determining the best learning strategy for this model is crucial, particularly when determining how quickly the ANN should change the weights of the neurons based on the importance of the error. Remember that the goal of an ANN model is to derive an estimating function from a given input x that, in this example, will forecast the stock return for the upcoming period, the target variable.

The input variables described in Section 3.7 will be used by the model. The target variable equation is represented as:

$$Y = \phi(x_1, x_2, \ldots, x_n) + \epsilon \tag{3.13}$$

where:

Y represents the target variable,

The non-linear function ϕ of the input variables is seen as the conditional expectation of

Y given $x_1, x_2, \ldots, x_n$, i.e,

$$E(Y|x_1, x_2, \ldots, x_n) = \phi(x_1, x_2, \ldots, x_n) \tag{3.14}$$

with x_1, x_2, x_n the explanatory variables of Y, and ϵ represents the 'noise' element of y that cannot be explained by the explanatory variables. The noise element, also known as the error term, residual or disturbance, represents the unexplained variance in the response variable in a statistical model. When using explanatory variables to model the relationship between a response variable and predictor variables, the noise element captures any factors that influence the response variable that are not included in the model. The noise element is typically assumed to be a random variable with zero mean, which represents the random fluctuations in the response variable that cannot be explained by the explanatory variables.

$$E(\epsilon | x_1, x_2, \dots, x_n) = 0 \tag{3.15}$$

Equation 3.13 is a traditional representation of a forecasting scenario, where Y represents the next period's price of a specific stock and $x_1, \dots, x_n$ are variables that have an influence in the predictability of Y.

A choice will be made regarding the precise target variable for this study. The most logical option might be to set the target variable as the adjusted closing price for the next day. The value after corporate actions have been taken into account is reflected in the adjusted closing price, which will be factored into the price. However, the model may be susceptible to look-ahead bias given the adjusted closing price.

Stock prices can show long-term trends and patterns, such as growth phases, stabilization periods, and potential resistance levels. These trends could be useful in understanding a stock's historical performance. Observing unexpected price swings might also suggest changes in market mood, investor reactions to news, or changes in market dynamics. Price changes can be used to evaluate market mood. Price movement analysis can give light on investor behaviour and decision-making tendencies. Price charts might show repeated patterns due to behavioural biases and recognizing these trends can assist in forecasting short-term price fluctuations.

The target variable O_j (Adjusted closing price for the next day) are explained as the inputs being connected to the j^{th} neuron with its associated weights $(w_{1j}, w_{2j}, \dots, w_{ij})$ at each connection. The signals that the neurons receive are added together, and they are then multiplied by their weights. The output denoted as h_j is passed through the transfer function, $g(h)$ which is non-linear to give output O_j. The output for a single *neuron* with *weights* $w_1, \dots, w_n \in \mathbb{R}$, bias $b \in \mathbb{R}$ and transfer function $f_j: \mathbb{R} \to \mathbb{R}$ can be displayed through the following function $f: \mathbb{R}^n \to \mathbb{R}$:

$$O_j = f(x_1, \dots, x_n) = f_j\left(\sum_{i=1}^{m} w_{ij} x_i + \boldsymbol{b}\right) = f(\boldsymbol{w^t x} + \boldsymbol{b}) \tag{3.16}$$

where w_{ij} represents the weight vector:

$$\boldsymbol{w_{ij}} = [w_1, \dots, w_m]^t, \tag{3.17}$$

and x represents the input vector,

$$\boldsymbol{x_i} = [x_1, \ldots, x_m]^t. \tag{3.18}$$

Expression $f(\boldsymbol{w^t x} + \boldsymbol{b})$ is the function for the transfer or activation function. The bias has been introduced as an intercept to provide some freedom for the perfect fit, and the function now approximates the actual value (Y). This will make it easier to use loss functions to get the predicted values. As a result, the bias enables the transfer function's curve to move along the axis.

Following that, each point will be transmitted as activations through the neurons, as seen in Figure 3.3 of a single artificial neuron:

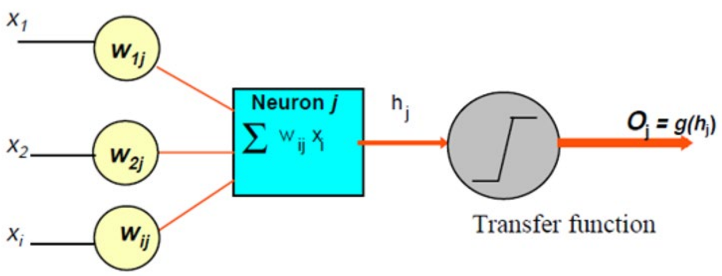

Figure 3.3: Transfer Function illustrated

Source: (Tan, 1997)

An ANN model is trained using the input data, and when doing so, the three key considerations are: selecting the right input data, parameters, and algorithm for training the network, (Lawrence, 1997). It is also crucial to remember that an ANN will start working once there is enough data available to train it. The model divides the observed data into three data sets, as is customary for machine learning with a supervised learning algorithm, in order to train, validate and test the ANN. The training set is utilized to establish the model's parameters.

The form of Equation 3.16 represents a single layer of a deep learning model, or simply put an artificial neuron. In this case the vector inputs $x \epsilon \mathbb{R}^p$ are transformed to a scalar output $\hat{y} \epsilon \mathbb{R}$. Generally, the deep learning model allow for vector inputs.

3.10 Data preparation

The success of the study depends largely on the pre-processing of the data. It's crucial to limit input data types to numbers. If not, the data will be normalized, scaled, and transformed into a numerical form that the model will accept. For this investigation, the data set has been cleansed and pre-processed. For a neural network, it is crucial that the data be normalized so that it can be interpreted. Keep in mind that scaling only occurs once the data have been divided into distinct datasets.

Most of the effort consumed for data analysis goes toward setting up and cleaning the dataset because doing so will cut down on any needless or excessive noise. ANNs are sensitive to noise in the data, which can occur due to measurement errors, errors in data collection, or other sources. Noise can cause the network to learn patterns that do not exist in the data, resulting in poor performance. Another risk is data leakage which occurs when information from the test data is used to train the model. It can cause the model to be overfitted and perform poorly on new, unseen data.

Finding advanced filtering methods or using some domain knowledge, personal recommendation or opinions are all feasible solutions if it is properly introduced and explained. The model would eliminate this risk by giving little to no weight values to these specific connections between the input nodes of those variables due to the nature of an ANN and its capacity to recognize and neutralize inputs that are inconsequential. A large ANN model would result in an unfavourable trade-off thanks to increased computing time. However, due to recent technical developments, including the creation of and access to supercomputers, computation time is no longer seen as a significant barrier.

It is mentioned that there is a chance that the data may be unbalanced. The market could, on average, be slightly more bullish than bearish during the study period, but in this particular case it will be more bearish due to the recent COVID pandemic, and the announcement made before COVID that South Africa had entered junk status and that the country's economy had entered a recessionary stage. Therefore, it's crucial to recognize that the ANN's accuracy in predicting directions may skew.

It's important to keep in mind that data preparation and data cleaning is a crucial step in the machine learning pipeline, and it's essential to invest time and resources to ensure that the data is of high quality and that the ANN is trained on the most relevant data.

3.11 Training, Validation, Test split

This project will analyse a sizable financial time series data collection, which will need a sizable amount of training time. Thus, dividing the data set into training, validation, and test data sets is crucial.

This section outlines the key distinctions between the study’s proposed model and the accepted standard approach.

When hindsight bias is introduced into the study model, it allows the validation set to become a strict subset of the training set, which would essentially mean that the model's validation data set is part of the training dataset. The standard model, with its validation set and training set, will remain mutually exclusive, (Case & Clements, 2021). As a result, it is best to think of the validation dataset as a dependant (subset) of the training dataset to summarize the study's main point.

This can be represented as follows:

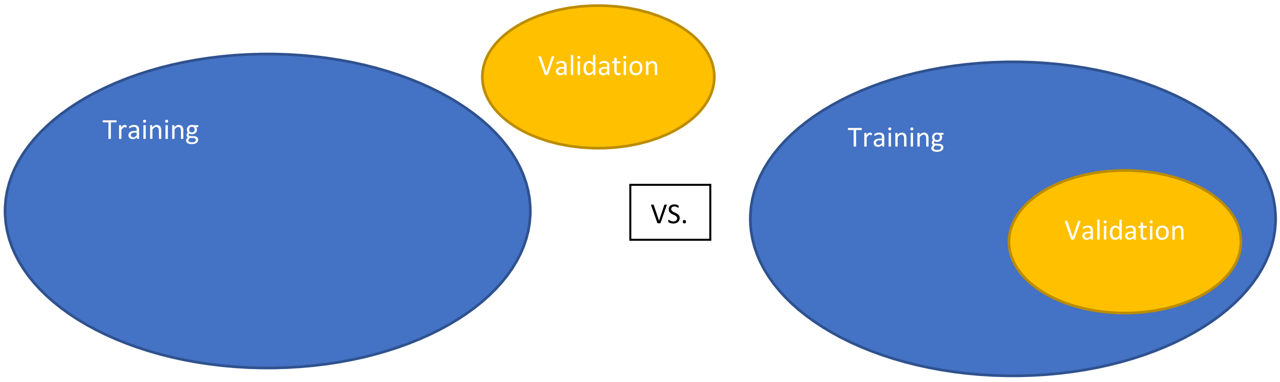

Figure 3.4: Standard Practise Model vs. Hindsight Bias Model

Source: (Case & Clements, 2021)

The standard practice model is a representation of the standard modelling methodology, which divides the entire dataset into three data sets for training, validation, and testing. The data sets are entirely distinct and do not overlap. The Hindsight Bias Model (also known as the Proposed Model) is divided into three datasets, but because the validation dataset depends on the training set, the dataset will be divided into training and test for the entire set of data, because the validation dataset is "absorbed" by the training dataset.

The training, validation, and prediction split will be a key factor in whether this study is successful. The investigation will generate twelve distinct models with twelve distinct results. The assumption of hindsight bias will be used in six of the models, whereas it will not be used in the other six models. The training, validation, and test data splits within each set of models will change to assess the effects of having a smaller prediction set or training set, as well as including or excluding hindsight bias. The validation dataset will strictly become a subset of the training dataset if hindsight bias is introduced during the training of an ANN, (Case & Clements, 2021).

This study will change the training-to-validation ratios, therefore the lower the ratio and the more data is contributed to the training set, the harder it is for an ANN to perform, but the more robust and significant the learning is.

The learning process makes use of the training set. When testing the predictiveness of the model and determining when to end training, the model observes the validation set which is the dataset that was not exposed to the ANN during the learning process.

3.12 Learning Process

The learning process and approach of the ANN can be explained as follows:

1. Define and bestow the data to the ANN model to find and identify a pattern in the input data and determine the forecasting target variable. Introducing and understanding the structure (i.e., the inputs, layers and outputs) is important.
2. Data pre-processing
    i. One of the problems arising in stock market data with deep models is that there is not enough data that is correctly formatted and processed. It is difficult to train a large model when the study has a lack of data without overfitting.
3. Initialization: The custom architecture and parameters necessary to create the connection weights, which are initially assigned at random, between neurons will be used to activate the network learning process.
4. Starting the training process, entails computing the output using the supplied data and associated weights. During this stage, an ANN model will make mistakes and learn from them, adjusting the weights accordingly. Forward propagation is the process by which the input data is transferred through the neural network and the output of each neuron is generated by applying an activation function to the weighted sum of the inputs.
5. The loss calculation: The difference between the predicted output and the actual output is calculated using a loss function.
6. The weights and biases of the ANN are modified to reduce the error after it has been propagated backwards through the system. Backpropagation is the process step that is carried out utilizing the optimization algorithm ADAM.
7. Repeat steps 3-5: The process of forward propagation, loss calculation and backpropagation will be repeated multiple times (epochs) using different parts of the training data set.
8. Once the optimal parameters are obtained, the model will continue by using the validation set, which represents an out-of-sample test on novel data. The model requires parameters that generalise to novel data, data that is not within your data set. If the model does not generalise, it will become overfit and ineffective in terms of

making predictions because it cannot apply to data outside of the dataset that it was trained on.

9. Evaluation: The performance of the network is evaluated using a set of test data that the neural network has not seen before. Out-of-Sample Testing is a method for assessing how well the network performs with new, untested data. The network is tested on a new, independent set of data after being trained on a subset of the data. The network's performance is then assessed using metrics like MAE and RMSE.
10. Repeat steps 3-7: The process is repeated with different hyper-parameters or architecture, in order to optimize the results of the model.
11. The process needs to observe a generalised outcome to determine if the model was able to learn effectively and produce an accurate result for the stock price forecasting. The test data set is an extract of the data set to compare the accuracy of the model outcome against the actual result.
12. The desired outcome of this model is found when the difference between the actual result and the output, i.e., the delta is minimised. The delta is minimised through the ANN model changing the weights. It's crucial to remember that since the stock market is extremely dynamic and the future is difficult to anticipate, it's difficult to have a strong performance for stock price prediction. Additionally, it's critical to remember that even with a strong performance, it might be challenging to determine whether the performance is reliable or the result of chance alone.

3.13 Summary

Artificial Neural Networks (ANNs) can be used to complement a study that focuses on hindsight bias in stock price forecasting in South Africa. ANNs are a type of machine learning model that are well suited for modelling complex, non-linear relationships between input variables and a response variable. In the context of stock price forecasting, ANNs can be used to capture the intricate relationships between various economic and financial variables, including past stock prices, and predict future stock prices.

In a study on hindsight bias, ANNs can be used to control for this bias by allowing the model to be trained on historical data with or without incorporating the knowledge of future stock prices. This can help to control the effect of hindsight bias, which occurs when researchers make predictions about past events using the knowledge of what actually happened. Overall, ANNs can be a valuable tool for complementing a study on hindsight bias in stock price forecasting, helping researchers to better understand the factors that influence stock prices and make more accurate predictions.

The next chapter of this study will focus on the results and summary statistics obtained from the ANN model. In this chapter, the results of the model's predictions for stock prices in South Africa are presented, as well as various statistical measures of the model's performance.

These results will provide a deeper understanding of the effectiveness of the ANN model in capturing the complex relationships between the input variable and the predictability of the output. The chapter will also provide insights into the strengths and limitations of the model, and the impact of hindsight bias on stock price forecasting.

Chapter 4:

Model Layout: Data; Description and Inputs

4.1 Introduction

The information used in the study is not primary data, but secondary data that has been processed and collected from a Google-based website called Yahoo Finance. For this study, a sample of ten companies' share prices has been collected from 2010/01/02 to 2021/12/31. This study prefers selecting companies with data available over the chosen time period.

This study used a technique known as DNN/ANN-based prediction modelling. The dataset will be historical financial stock data. Unstructured data may be included in the ANN model used. Most of the stock market data that is gathered will be non-linear, noisy, and long-term dependent.

4.2 Similar studies

In an effort to accurately anticipate the next 500 closing prices of the IBM stock, a neural network model was examined using 1000 closing values of the stock, (White, 1988). White thought about determining the best fit for the previous three days in order to subsequently forecast the closing price for the next day. This did not result in a successful trading approach.

Blume et al., (1994) discovered that incorporating volume together with price can provide useful information and claimed that this might just be enough to interpret the outcomes achieved from a competitive standpoint.

4.3 Companies list for Portfolio

The dataset for the ANN model on stock price forecasting consists of historical data related to the stock prices of a portfolio of companies. This data is used as input to the model to predict future stock prices. The companies chosen for the portfolio are equally weighted. The following South African companies' stock prices have been collected from 2010/01/02 – 2021/12/31:

1. Clicks Group Ltd (CLS) – Health and Wellness Retailer
2. Bidvest Group Ltd (BVT) – Services, Trading and Distribution
3. Shoprite Holdings Ltd (SHP) – Supermarket Retailer
4. Anheuser-Busch Inbev SA (ANH) – Drink and Brewing
5. Sasol Ltd (SOL) – Energy
6. Absa Group Ltd (ABG) - Bank and Financial Services
7. Standard Bank Group Ltd (SBK) – Bank and Financial Services

8. Naspers Limited (NPN) – Technology & Multimedia
9. Anglo American Plc (AGL) - Mining
10. Compagnie Financiere Richemont (CFR) – Luxury Goods

Eight of the ten companies chosen for the portfolio were in the Top40 JSE. Anheuser-Busch Inbev SA (ANH) and Compagnie Financiere Richemont (CFR) are not part of the JSE Top40 as of 01/01/2022. The companies listed in the JSE Top40 index are typically considered a representative of the South African economy and can be used as a benchmark for the performance of financial assets, such as stocks. Using these companies in an assessment of a South African artificial neural network (ANN) model can provide a clearer understanding of the model's performance and its ability to accurately predict stock prices.

Additionally, these companies are likely to have a significant impact on the South African stock market, given their size and market capitalization, and can provide a good representation of market trends and movements. Moreover, by selecting companies from different sectors of the economy and with varying levels of market capitalization, the study can provide a more holistic assessment of the impact of hindsight bias. This can help to identify any sector-specific or size-specific effects, providing a more nuanced understanding of the topic.

By evaluating the performance of the ANN model on these top companies, a better understanding of its overall accuracy and potential for use in real-world applications can be identified. It can also provide insight into the model's strengths and weaknesses, allowing for further improvements and refinements.

Using the companies listed to form the portfolio for this study is beneficial for several reasons:

1. Diversification: By including companies from different sectors, such as health and wellness retail, services, banking, energy, and luxury goods, the portfolio is well diversified and less vulnerable to sector-specific risks. Additionally, it provides a portfolio that is better equipped to withstand market fluctuations.
2. Representativeness: The companies chosen to represent a broad cross-section of the South African economy, including retail, services, finance, energy, and mining. This provides a comprehensive and representative sample of the stock market, providing valuable insights into the performance of the South African economy. These sectors represent some of the key economic drivers in South Africa and are therefore essential to understanding the overall state of the finance market in South Africa.
3. Performance: The companies selected have a strong track record of performance, which makes them ideal candidates for inclusion in the portfolio. This ensures that the portfolio returns provide a good representation of the performance of the South African stock market as a whole.

Overall, the use of the chosen companies as a portfolio provides a balanced and representative sample of the South African stock market and provides a good basis for generating insightful results in the study.

The input data of the model forms a critical component of the study and is therefore a critical factor in ensuring the validity and reliability of the results. The study can contribute to the development of more accurate and effective stock forecasting models, reducing the impact of hindsight bias, and improving investment decision-making in the South African equity market.

The input data for the study, which consists of the selected companies' prices, is used as input for the artificial neural network (ANN) model. The purpose of using an ANN model in the study is to analyse the relationship between the input data (the selected companies stock prices) and the output data (the future stock prices of the companies).

By using an ANN model, the study can identify patterns and relationships in the data that may not be easily noticeable using traditional statistical methods. The model can learn from the input data and make predictions about the output data, allowing the study to assess the impact of hindsight bias on stock forecasting in the South African equity market.

A market portfolio return (the selected companies' returns) will be determined by calculating the returns for each company, whereas thereafter a portfolio return is calculated by taking an equally weighted average of the returns of each company in the portfolio. This means that each company's contribution to the portfolio return is proportional to its weight in the portfolio.

The adjusted closing price over the open price was selected because, according to studies, a certain stock price's closing value on one day may fluctuate slightly from its opening value on the next day. The recent implementation of after-hours trading between institutions' own exchanges is largely to blame for the difference, (O'Connor, Madden, 2006).

In conclusion, using the companies in the JSE Top40 index in the assessment of a South African ANN model provides a practical and relevant benchmark for the model's performance, allowing for a more accurate evaluation of its potential for use in the real world. Due to the time-consuming nature of this approach, the dataset used in this study does not include any individual opinions or investor sentiment. Finally, hindsight bias will be the only behavioural element that will be included in this study's aspect of behavioural finance, which will result in insightful findings.

4.4 Finding and Results

The goal is to ascertain whether the prediction error term - that is, the difference between the actual and anticipated Y (equation 3.13 – target variable) - improves when hindsight bias is incorporated into the model. The first method will involve forecasting the closing price for the following period using the closing price of the prior trading day data. The second method will make use of closing price and incorporate the assumption of hindsight bias in order to predict the closing price of the following period.

Artificial neural networks (ANNs) can detect hindsight bias in stock price forecasting by comparing the network's performance on different time periods of historical data. Hindsight bias is a cognitive bias that occurs when people believe they knew the outcome of an event before it happened, simply because they know the outcome now. To detect this bias, the following steps are taken:

1. Divide the historical data into three parts: training data, validation data, and test data.
2. Train the ANN on the training data and evaluate its performance on the validation data using metrics like mean absolute error (MAE) or root mean squared error (RMSE).
3. Repeat steps 1 and 2 but use different time periods for the training and validation data. For example, use the data from the last 3 years as training data and the data from the last 2 years as validation data.
4. Compare the performance of the ANN on different validation periods. For example, when the performance drops significantly for validation data from a more recent longer period (e.g., the last 2 remaining years), it is likely that the ANN is suffering from hindsight bias. This drop in performance is caused by the ANN model unintentionally "learning" from future data, making it appear better than it is at predicting the stock prices during that time period. This is the essence of hindsight bias - believing it could have predicted an event accurately because it now knows what the actual outcome is.

To prevent this bias, techniques like cross-validation and out-of-sample testing can be used to evaluate the network's performance on unseen data.

It's important to note that this is a simplified explanation of the process, and the specifics of the process would vary depending on the complexity of the model and the specific task. Additionally, it is important to note that, even if a model is not suffering from hindsight bias, it is hard to know if its performance will be consistent in the future.

The bracket values for the different tests represent the validation dataset which forms part as a subset of the training dataset. For this study, six tests were performed. The data split is defined as:

- The different tests to compare if the standard model performed better or worse than the hindsight bias model, ANN is used to predict the returns against the actual returns observed in the market. The ANN model will follow the learning process defined in Section 3.12. The proposed model will include the validation set as a strict subset in the training data set, as illustrated in Figure 3.4.
- The first two tests (Test 1 and Test 2) for the Proposed Model possess longer time periods than the Standard model under the training data set. This may also indicate that hindsight bias may improve performance.
- Test 5 and Test 6 would ignore the first and second years of the ten-year assessment period respectively. This might attribute more value to more recent data and ignore older data.

The different tests can be described as follows:

1. **Test 1:** The split between training, validation and test set is:
    a. Standard Model - 80:10:10
    b. Proposed Model - 90:(10):10
2. **Test 2:** The split between training, validation and test set is:
    a. Standard Model - 70:20:10
    b. Proposed Model - 90:(20):10
3. **Test 3:** The split between training, validation and test set is:
    a. Standard Model - 80:10:10
    b. Proposed Model - 80:(10):10
4. **Test 4:** The split between training, validation and test set is:
    a. Standard Model - 70:20:10
    b. Proposed Model - 70:(20):10
5. **Test 5:** The split between training, validation and test set is:
    a. Standard Model - 70:20:10
    b. Proposed Model - 70:(20):10
        i. Whereby the first two years are not assessed.
6. **Test 6:** The split between training, validation and test set is:
    a. Standard Model - 80:10:10
    b. Proposed Model - 80:(10):10
        i. Whereby the first year is not assessed.

The layout of the standard model and proposed model can be respectively illustrated as follows:

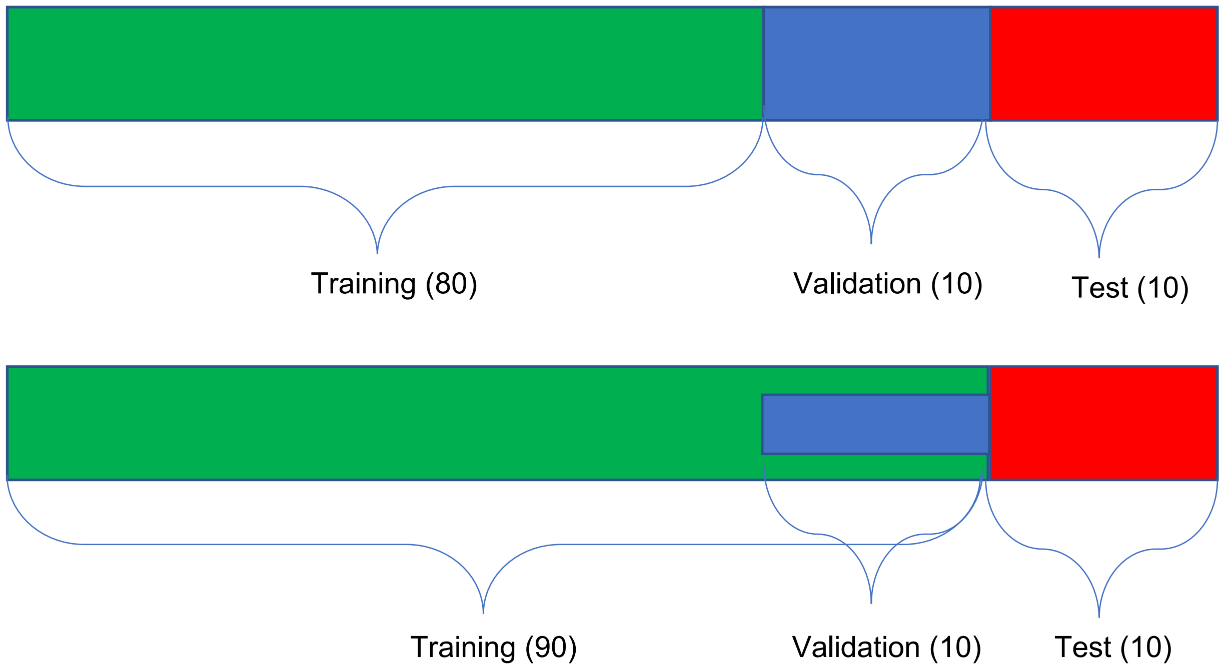

Figure 4.1: Test 1 Illustrative Example

4.5 Inferential Statistics & Prediction Accuracy Measures

Descriptive statistics are used to describe a dataset whereas inferential statistics enable predictions to be made from that data. It provides a framework with which to judge the significance of results. There are and be plenty of measures that can be used to assess the different models' accuracy and efficiency:

- Measuring the errors during the training period: RMSE; MAE.
- Errors during the validation period: RMSE; MAE.
- Residual Diagnostics & Goodness-of-fit Tests:
  - Plots of Actual vs Predicted.
- Qualitative considerations:
  - Intuitive reasonableness of the model, simplicity of the model.

To compare two models against actual results using summary statistics the following will be compared, calculated, and assessed:

- Calculate the mean; standard deviation; standard error; median; kurtosis; skewness and range for each model, which represent the difference between the actual results and the models' predicted values.

- Calculate the root mean squared error (RMSE) and mean absolute error (MAE) for each model, which are measures of the magnitude of the errors in the model's predictions.
- Calculate the R-squared value for each model, which represents the proportion of the total variability in the actual results that is explained by the model's predictions. A higher R-squared value indicates that the model fits the data better.
- Compare the results of the two models in terms of the mean, standard deviation, RMSE, MAE, and R-squared values. A lower mean and standard deviation of the residuals, as well as lower RMSE and MAE values, indicate that the model has a better fit to the data. A higher R-squared value also indicates a better fit.
- Visualize the residuals for each model to check if they are normally distributed and to check for outliers. A normally distributed residuals plot with no outliers is a good indication that the model is a good fit for the data.
- Visualize the time series plot for each model and compare it against the actual result.
- The two models' distributions are compared by performing the Mann-Whitney U test and the Diebold-Mariano test.

4.5.1 Summary Statistics

Summary statistics such as mean, standard deviation, and other relevant measures of central tendency and dispersion can provide an overall understanding of the data distribution within a model. For example, if the mean of the actual returns and the predicted returns from the model are similar, it may indicate that the model is capturing the central tendency of the data well. Similarly, if the standard deviation of the predicted returns is similar to that of the actual returns, it may indicate that the model is capturing the dispersion of the data well.

Kurtosis is defined as a measure of the peakedness or flatness of a distribution compared to a normal distribution. A lower kurtosis indicates a flatter distribution, which is generally considered to be a better fit to the actual data, whereas a kurtosis of three, implies that its distribution has the same level of peakedness as a mesokurtic distribution. Skewness is defined as a measure of the asymmetry of a distribution. A skewness close to zero indicates a symmetric distribution, which is generally considered to be a better fit to the actual data.

Furthermore, standard error can be explained as a measure of the average distance that the sample mean of the data is from the true population mean. The standard error of a model can be compared between two models to determine which model has a better fit. A smaller standard error indicates a better fit as it shows that the model's predictions are closer to the actual values. The standard errors can be directly compared by calculating the ratio of the standard errors and comparing it to 1. A ratio close to 1 indicates that the standard errors are similar, while a ratio significantly less than 1 indicates that one standard error is smaller than the other. The ratio will be calculated for the means too. Assume the standard errors (S.E.) for Model 1 and Model 2 are SE1 and SE2 respectively. Then, the ratio of their standard errors can be calculated as follows:

$$SE_{ratio} = \frac{SE_1}{SE_2}. \tag{4.1}$$

The summary statistics presented in the tables below depict the results of the test set. The test set represents a 252-day projection of the stock price and corresponding returns for the portfolio. These summary statistics serve to aid in determining the superiority of the two models.

Test 1

Table 4.1: Test 1 Summary Statistics

	Without_bias	***With_bias***	***Actual_result***
Mean	-0.00028	-0.00030	-0.00017
Mean Ratio	1.647	1.765	1
Standard Error	0.00106	0.00102	0.00135
S.E. Ratio	0.785	0.756	1
Median	0.00038	0.00071	0.00223
Standard Deviation	0.01687	0.01625	0.02149
Kurtosis	69.6341	91.44708	22.89861
Skewness	-6.08786	-7.44521	-2.63575
Range	0.23216	0.24125	0.26967

S.E. Ratio for the hindsight bias model is smaller, which means that this model has a lower standard error compared to the actual result and is therefore more accurate. With the means ratio for the hindsight bias model being larger, it would imply that this model has a higher mean compared to the actual result and is therefore less accurate.

The standard model has a lower kurtosis which indicates a flatter distribution. The skewness of this model is closer to zero which indicates a more symmetric distribution, which is generally considered to be a better fit to the actual data. Furthermore, the range observed for the standard model is closer to the actual returns, when compared to the hindsight bias model. Therefore, the standard model is considered to be a better fit to the actual data.

Test 2

Table 4.2: Test 2 Summary Statistics

	Without_bias	*With_bias*	*Actual_result*
Mean	-0.00024	-0.00027	-0.00017
Mean Ratio	1.411	1.588	1
Standard Error	0.00102	0.00118	0.00135
S.E. Ratio	0.756	0.874	1
Median	0.00074	0.00063	0.00223
Standard Deviation	0.01634	0.01888	0.02149
Kurtosis	90.15839	46.59502	22.89861
Skewness	-7.36961	-4.42463	-2.63575
Range	0.23827	0.258037	0.26967

S.E. Ratio for the standard model is smaller, which means that this model has a lower standard error compared to the actual result and is therefore more accurate. The means ratio for the hindsight bias model is larger, which would mean that this model has a higher mean compared to the actual result and is therefore less accurate.

The hindsight bias model has a lower kurtosis which indicates a flatter distribution. The skewness of this model is closer to zero which indicates a more symmetric distribution, which is generally considered to be a better fit to the actual data. Furthermore, the range observed for the hindsight bias model is closer to the actual returns, when compared to the standard model.

Therefore, the results obtained are inconclusive as to which model is considered to be a better fit to the actual data. Although, the S.E. ratio proves to be a valuable metric, as it provides information on the variability of the model's predictions. Therefore, the standard model would be suggested as a slightly better fit to the actual data.

Test 3

Table 4.3: Test 3 Summary Statistics

	Without_bias	*With_bias*	*Actual_result*
Mean	-0.00032	-0.00021	-0.00017
Mean Ratio	1.882	1.235	1
Standard Error	0.0010	0.00109	0.00135
S.E. Ratio	0.741	0.807	1
Median	0.00076	0.00013	0.00223
Standard Deviation	0.01715	0.01746	0.02149
Kurtosis	73.63773	60.57649	22.89861
Skewness	-6.45412	-5.33112	-2.63575
Range	0.23349	0.24708	0.26967

S.E. Ratio for the standard model is smaller, which means that this model has a lower standard error compared to the actual result and is therefore more accurate. The means ratio for the hindsight bias model is smaller, which would mean that this model has a lower mean compared to the actual result and is therefore more accurate.

The hindsight bias model has a lower kurtosis which indicates a flatter distribution. The skewness of this model is closer to zero which indicates a more symmetric distribution, which is generally considered to be a better fit to the actual data. Furthermore, the range observed for the hindsight bias model is closer to the actual returns, when compared to the standard model. Therefore, the hindsight bias model is considered to be a better fit to the actual data.

Test 4

Table 4.4: Test 4 Summary Statistics

	Without_bias	***With_bias***	***Actual_result***
Mean	-0.00024	-0.00029	-0.00017
Mean Ratio	1.411	1.706	1
Standard Error	0.00104	0.00110	0.00135
S.E. Ratio	0.770	0.815	1
Median	-0.000015	0.00011	0.00223
Standard Deviation	0.01655	0.01750	0.02149
Kurtosis	84.12707	72.19329	22.89861
Skewness	-6.95218	-6.14715	-2.63575
Range	0.23997	0.25656	0.26967

S.E. Ratio for the standard model is smaller, which means that this model has a lower standard error compared to the actual result and is therefore more accurate. The means ratio for the standard model is smaller, which would mean that this model has a lower mean compared to the actual result and is therefore more accurate.

The hindsight bias model has a lower kurtosis which indicates a flatter distribution. The skewness of this model is closer to zero which indicates a more symmetric distribution, which is generally considered to be a better fit to the actual data. Furthermore, the range observed for the hindsight bias model is closer to the actual returns, when compared to the standard model. Therefore, the hindsight bias model is considered to be a better fit to the actual data.

Therefore, the results obtained are inconclusive as to which model is considered to be a better fit to the actual data. Although, the S.E. ratio proves to be a valuable metric, as it provides information on the variability of the model's predictions. Therefore, the standard model would be suggested as a better fit to the actual data.

Test 5

Table 4.5: Test 5 Summary Statistics

	Without_bias	*With_bias*	*Actual_result*
Mean	0.00034	-0.00013	-0.00017
Mean Ratio	2.000	0.765	1
Standard Error	0.00069	0.001109	0.001351
S.E. Ratio	0.911	0.822	1
Median	0.001313	0.000244	0.002234
Standard Deviation	0.010981	0.017633	0.021495
Kurtosis	67.4360	48.643	22.89861
Skewness	-6.47365	-4.53462	-2.63576
Range	0.21244	0.230661	0.269674

S.E. Ratio for the hindsight bias is smaller, which means that this model has a lower standard error compared to the actual result and is therefore more accurate. The means ratio for the hindsight model is smaller, which would mean that this model has a lower mean compared to the actual result and is therefore more accurate.

The hindsight bias model has a lower kurtosis which indicates a flatter distribution. The skewness of this model is closer to zero which indicates a more symmetric distribution, which is generally considered to be a better fit to the actual data. Furthermore, the range observed for the hindsight bias model is closer to the actual returns, when compared to the standard model. Therefore, the hindsight bias model is considered to be a better fit to the actual data.

Test 6

Table 4.6: Test 6 Summary Statistics

	Without_bias	***With_bias***	***Actual_result***
Mean	-0.00023	-0.00022	-0.00018
Mean Ratio	1.352	1.294	1
Standard Error	0.001034	0.001052	0.001314
S.E. Ratio	0.766	0.780	1
Median	0.000127	5.71E-05	0.002
Standard Deviation	0.016449	0.01673	0.02090
Kurtosis	82.8817	66.60309	23.09693
Skewness	-6.88204	-5.81393	-2.70824
Range	0.23809	0.23333	0.256176

S.E. Ratio for the standard is slightly smaller, which means that this model has a lower standard error compared to the actual result and is therefore more accurate. The means ratio for the hindsight model is smaller, which would mean that this model has a lower mean compared to the actual result and is therefore more accurate.

The hindsight bias model has a lower kurtosis which indicates a flatter distribution. The skewness of this model is closer to zero which indicates a more symmetric distribution, which is generally considered to be a better fit to the actual data. Furthermore, the range observed for the hindsight bias model is closer to the actual returns, when compared to the standard model. Therefore, the hindsight bias model is considered to be a better fit to the actual data.

4.5.1.1 Conclusion

When observing the mean returns, all six tests for both models produce a mean fairly similar to the actual values obtained.

The standard deviation is quite low when compared to the actual standard deviation, which is based on the return of a portfolio made up of 10 stocks and is shown as a measure of risk. However, it is important to note that volatility can be significant because stocks are highly risky assets to keep. In comparison to the standard practice model, the standard deviation results produced by the hindsight bias model are generally closer to the actual values.

An important assumption to confirm when assessing the results will be to assess if the returns are normally distributed. One way is to look at the histograms of the data and visually inspect the shape of the distribution. A histogram that is roughly bell-shaped and symmetrical with a single peak can suggest that the data is approximately normally distributed, A normal or Gaussian distribution is characterized by a symmetrical bell-shaped curve, while a skewed distribution may indicate the presence of outliers or a bias in the data. Another way to assess normality is to calculate the skewness and kurtosis statistics of the data, which describe the shape and distribution of the data. However, it's important to keep in mind that these methods are not fool proof, and it's possible to have a histogram that appears normal but fails normality tests, or vice versa. Ultimately, the best way to assess normality is to use a combination of these methods and make a judgement based on multiple lines of evidence.

The skewness results for the model that includes the hindsight bias assumptions, arguably have a skewness closer to the actual values when compared to the model that does not include this assumption. If the skewness is positive, then assume the distribution is not symmetrical and vice versa. This finding holds for all tests except Test 1.

For kurtosis the same conclusion can be drawn; the hindsight bias assumption has improved model results and kurtosis value are closer to the actual values. Although the kurtosis is different from zero, it simply implies that the distribution is significantly different from a normal distribution. All results have improved, except for Test 1.

Based on the summary statistics for the various tests, it appears that there have been overall improvements. However, these improvements are not substantial enough to be considered definitively convincing. Further analysis and testing may be necessary to confirm these results and draw stronger conclusions. In conclusion, the results of the summary statistics emphasize the importance of continuous evaluation and improvement of the models in order to achieve optimal performance.

4.5.2 R-squared (Coefficient of determination)

R-squared is a statistical measure of the proportion of the variance in the dependent variable that is predictable from the independent variable(s). In the context of comparing two regression models, a higher R-squared value indicates that the model explains more of the variation in the dependent variable, and therefore is a better fit to the data.

The suggested hindsight model's goodness of fit will also be evaluated statistically using the R-squared value. In order to determine if the suggested model is a better fit or not against the actual values, it is crucial to determine whether the model will show a greater or lower R-squared number. Values closer to 1 will indicate a good fit and explains more of the variance in the dependent variable.

However, this is not always the case. Other factors such as the presence of outliers, multicollinearity, or omitted important variables can also impact the R-squared value. It is also important to consider other evaluation metrics such as mean absolute error (MAE), root mean squared error (RMSE), or adjusted R-squared, to get a more comprehensive view of the performance of the models.

The coefficient of determination equation is given formulated as:

$$R^2 \in (\infty; 1] \tag{4.2}$$

$$R^2 = 1 - \frac{SSE}{SST} \tag{4.3}$$

Where:

$SSE = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2$ (4.4)

$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2, \bar{y} = \sum_{i=1}^{n} y_i$ and n is the number of data points. (4.5)

Table 4.7: R-squared results for different test cases

R-squared	**Without_bias**	**With_bias**
Test1	0.882782	**0.904716**
Test2	0.827726	**0.893515**
Test3	**0.905492**	0.874668
Test4	**0.881322**	0.879479
Test5	0.734363	**0.814563**
Test6	0.875415	**0.883874**

Based on the R-squared values from the various test, it appears that the R-squared values tend to improve (four out of six tests) when including hindsight bias into the model and compared to the actual values. The magnitude of a substantial improvement in R-squared values between two models is generally, an improvement of 0.1 or more in R-squared values.

Although some tests show significant improvement in R-squared, others show only marginal improvement or no improvement at all. These results suggest that further analysis is needed to determine the effectiveness of the models and to identify potential sources of variation.

Conversely, a large improvement in R-squared values may not be considered substantial if it does not result in a noticeable improvement in the accuracy of the model when assessing the RMSE, MAE and Time Series results.

4.5.3 Histograms

The histograms of the actual and predicted returns can provide a visual representation of the data distribution. Histograms can help confirm the results of the summary statistics by showing the distribution of the data in a graphical form. For example, if the histogram of the predicted returns closely matches the histogram of the actual returns, it may indicate that the model is accurately capturing the distribution of the returns.

Therefore, by comparing the summary statistics and histograms of the actual and predicted returns, it is possible to confirm the validity of the model and understand how well it is capturing the underlying data.

When comparing the histograms of returns for two models, the interpretation of the distributions depends on the specific details of the models. Some general considerations when interpreting histograms of returns include:

1. Shape of the distribution: The shape of the forecasted distribution of a model can also give insight into its accuracy.
2. Mean and median: The mean and median of the distribution can provide an overall sense of the central tendency of the data, and the difference between the two can provide insight into the skewness of the distribution.
3. Spread: The spread of the distribution, as indicated by the standard deviation or interquartile range, can provide information about the degree of variability in the data. The spread (standard deviation, range) of the forecasted distribution can also give insight into the accuracy of the model. A model with a forecasted distribution that has a similar spread to the actual distribution is likely to be a better model.
4. Skewness: A histogram can show whether the data is skewed to the left or right, or if it is symmetrical.
5. Outliers: Outliers can be easily identified in a histogram as data points that fall outside the general distribution.

In comparing the histograms of returns for the Standard practice model and the ANN model with hindsight bias, it is important to consider these factors and to look for any notable differences in the shape, central tendency, and spread of the distributions. If the histograms are significantly different, it may indicate that the models are providing different results and that one of the models may be more effective in capturing the relevant patterns in the data.

Immediately when observing the first set of results, it is clear that the hindsight bias study model produces better distributions when compared to a standard model. The histograms indicate that the hindsight bias model is able to capture the volatility of the market. The histogram is ultimately used to display the distributions of the actual returns against the hindsight model and the standard model.

Ideally, the predicted returns should be normally distributed with a mean that is close to the true returns and a small standard deviation. This indicates that the model is making accurate predictions and that the errors are random and small.

In the histogram, if one model has a higher peak at the true return value, meaning the predicted returns are closer to the true returns, it would be considered better than the model with a higher peak at larger or lower returns. Additionally, if the first model has a narrower distribution, meaning the predicted returns are more consistent, it would be considered better than the second model with a wider distribution of predicted returns.

The x-axis of the histogram represents the bin axis, displaying the errors of the predicted return for both the standard model and the hindsight bias model. The values on the x-axis are grouped and rendered based on their respective bin values, allowing for a clear visualization of the distribution and comparison of prediction errors between the two models. The standard practice model and hindsight bias model are compared against the actual results observed over 252-days. The figures below illustrate the distributions for the six different tests:

Test 1

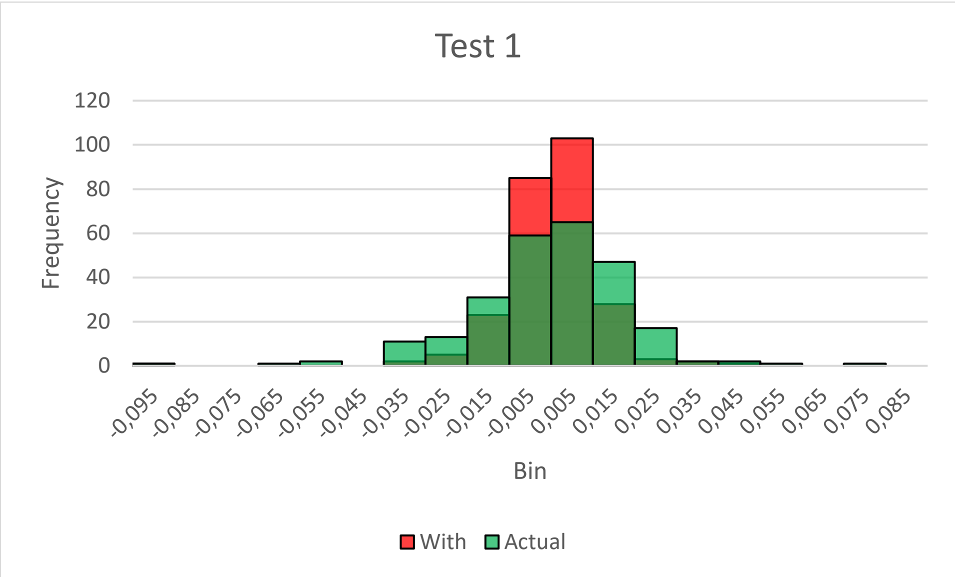

*Figure 4.2: Test 1 Results **with** Hindsight Bias assumption*

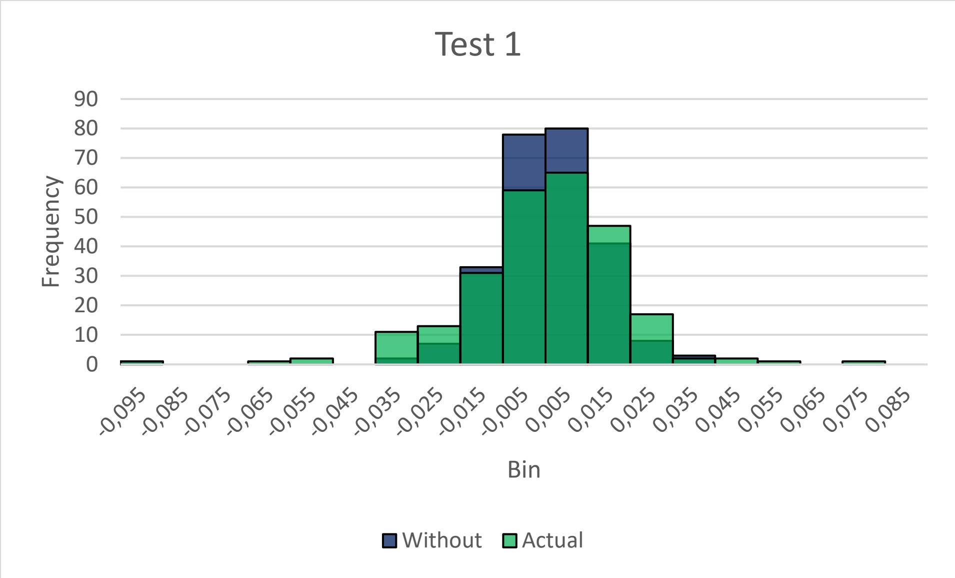

*Figure 4.3: Test 1 Results **without** Hindsight Bias assumption*

The hindsight model has a higher peak at the true return value, meaning the predicted returns are closer to the true returns, Furthermore the hindsight model has a narrower distribution, meaning the predicted returns are more consistent. The hindsight model closely mimics the shape of the actual returns. The range observed for the model with hindsight bias is closer to the actual returns, when compared to the standard practise model.

Therefore, the hindsight bias model is considered to be a better fit to the actual data.

Test 2

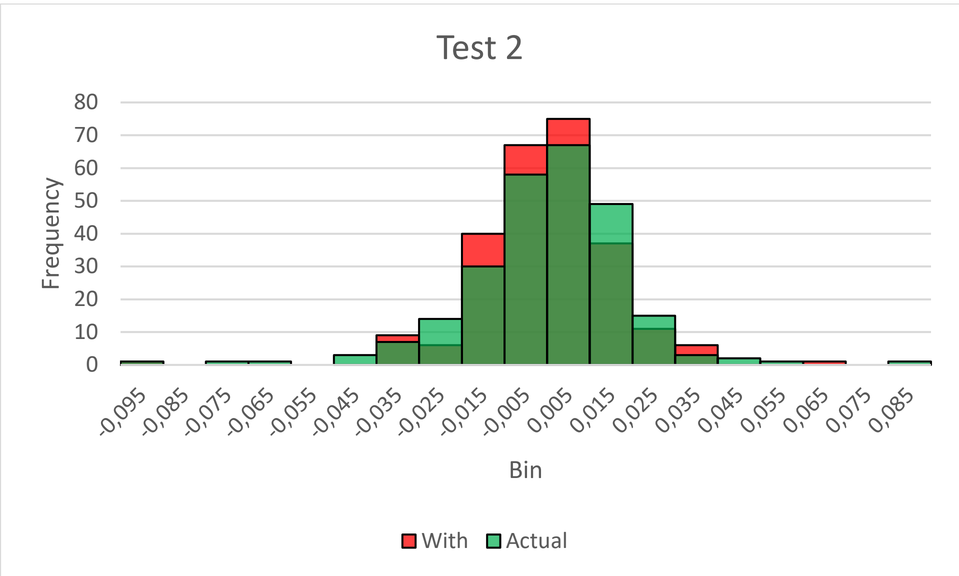

*Figure 4.4: Test 2 Results **with** Hindsight Bias assumption*

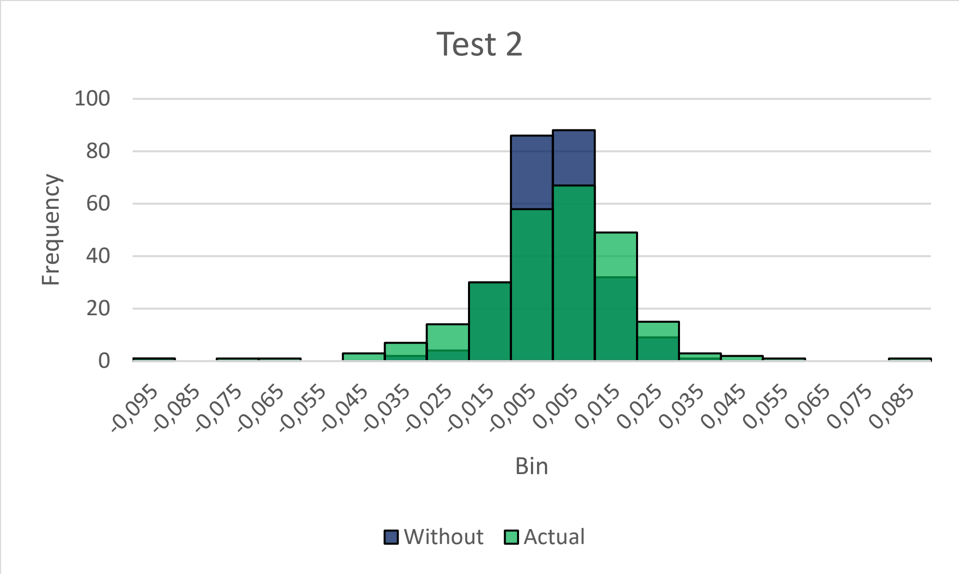

*Figure 4.5: Test 2 Results **without** Hindsight Bias assumption*

The standard model is showing a higher peak at the true return value, and is indicating a clear narrower distribution, meaning the predicted returns are more consistent for the standard practise model. The range observed for the hindsight model is closer to the actual returns, when compared to the standard practise model.

Therefore, the hindsight bias model is considered to be a better fit to the actual data.

Test 3

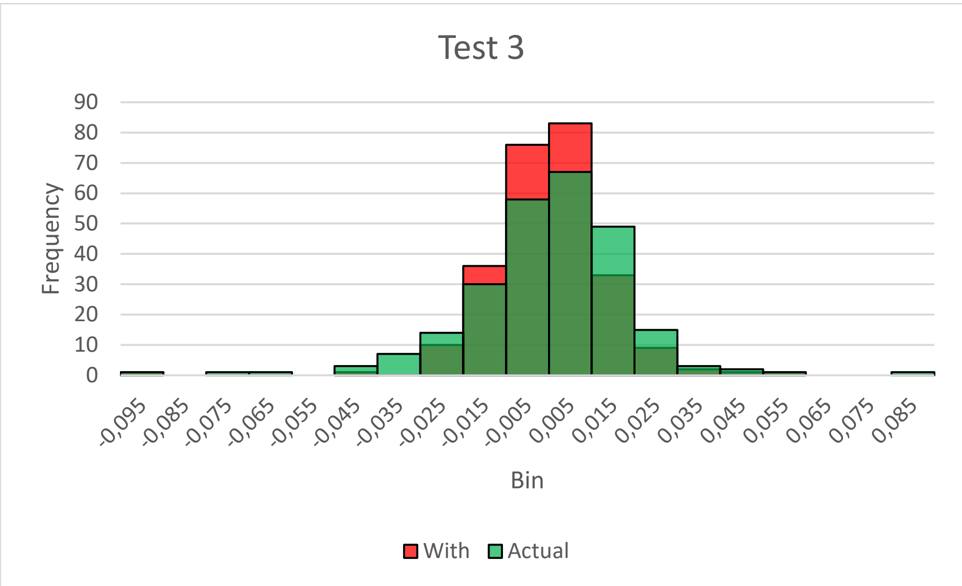

*Figure 4.6: Test 3 Results **with** Hindsight Bias assumption*

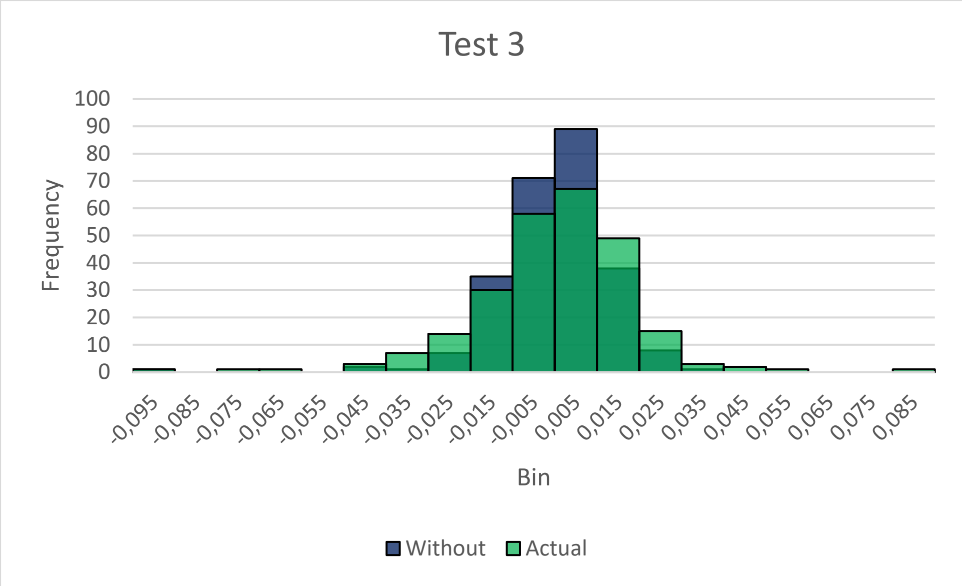

*Figure 4.7: Test 3 Results **without** Hindsight Bias assumption*

The standard model has a higher peak at the true return value, which indicate values to be closer to the actual results, which also confirms that the standard practise model has a narrower distribution, meaning the predicted returns are more consistent. However, the range observed for the model with hindsight bias is closer to the actual returns, when compared to the standard practise model.

Therefore, the standard model is considered to be a slightly better fit to the actual data.

Test 4

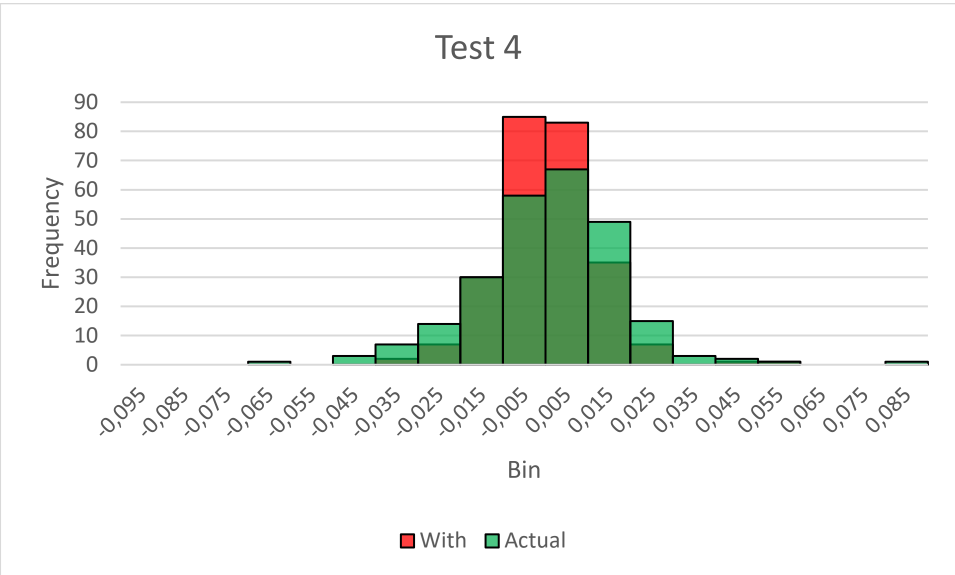

*Figure 4.8: Test 4 Results **with** Hindsight Bias assumption*

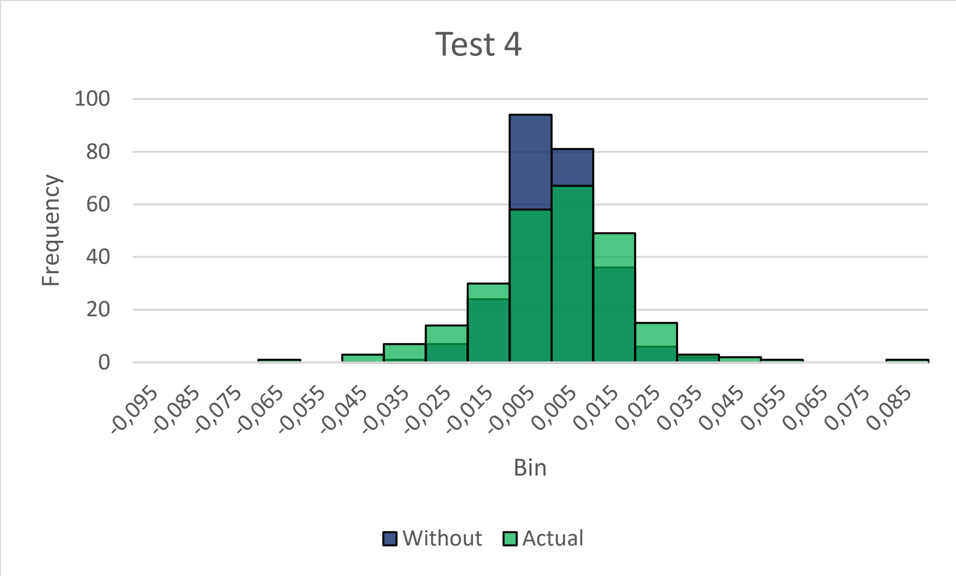

*Figure 4.9: Test 4 Results **without** Hindsight Bias assumption*

The standard model has a higher peak at the true return value, which indicate values to be closer to the actual results, which also confirms that the standard practise model has a narrower distribution, meaning the predicted returns are more consistent. However, the range observed for the model with hindsight bias is closer to the actual returns, when compared to the standard practise model.

Therefore, the standard model is considered to be a slightly better fit to the actual data.

Test 5

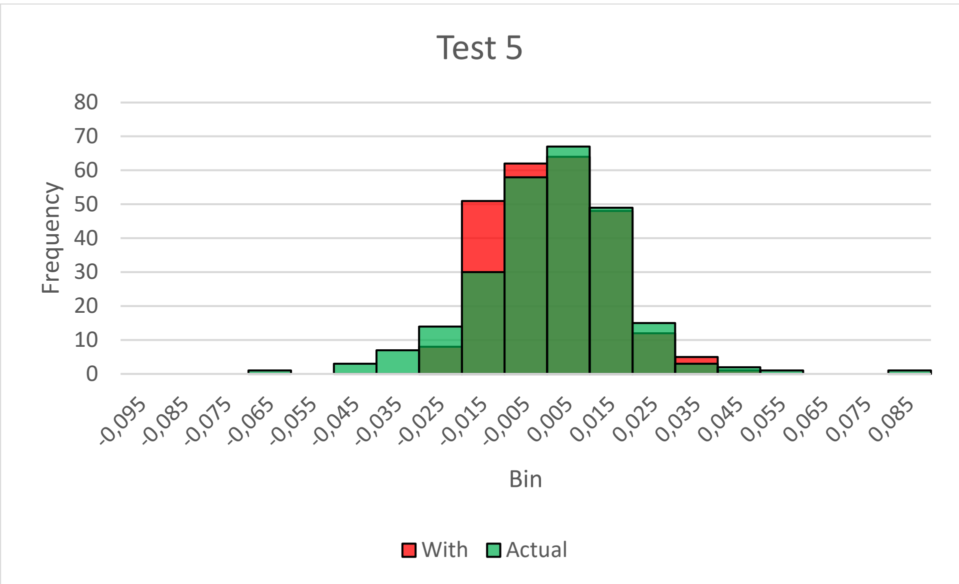

*Figure 4.10: Test 5 Results **with** Hindsight Bias assumption*

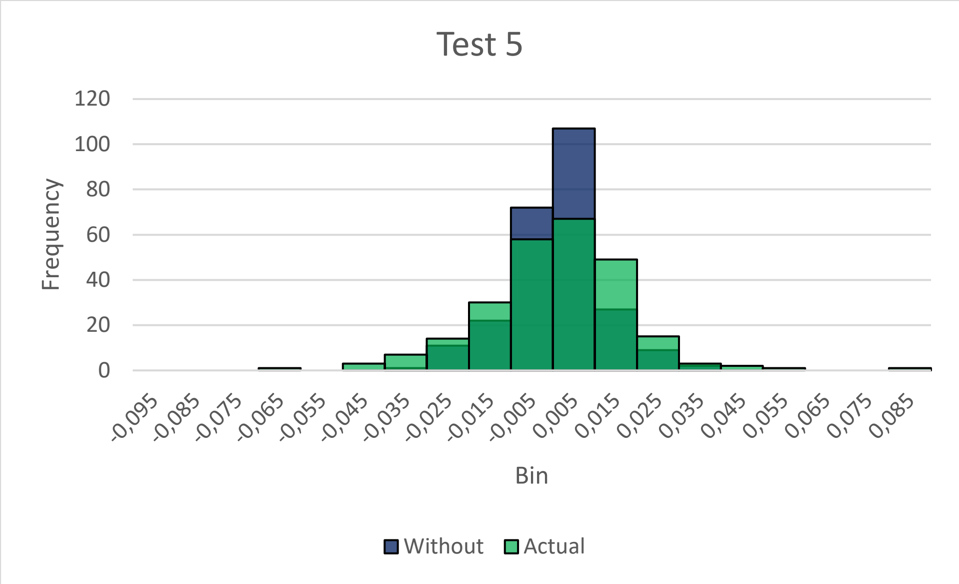

*Figure 4.11: Test 5 Results **without** Hindsight Bias assumption*

The hindsight bias model can effectively mimic the distribution of the actual results, whereas the standard practise model is unable to do so. Additionally, the hindsight bias model has a narrower distribution, meaning the predicted returns are more consistent, and would be considered a better fit than the standard practise model with a wider distribution of predicted returns. The range observed for the model with hindsight bias is also closer to the actual returns, when compared to the standard practise model.

Therefore, the hindsight bias model is considered to be a better fit to the actual data.

Test 6

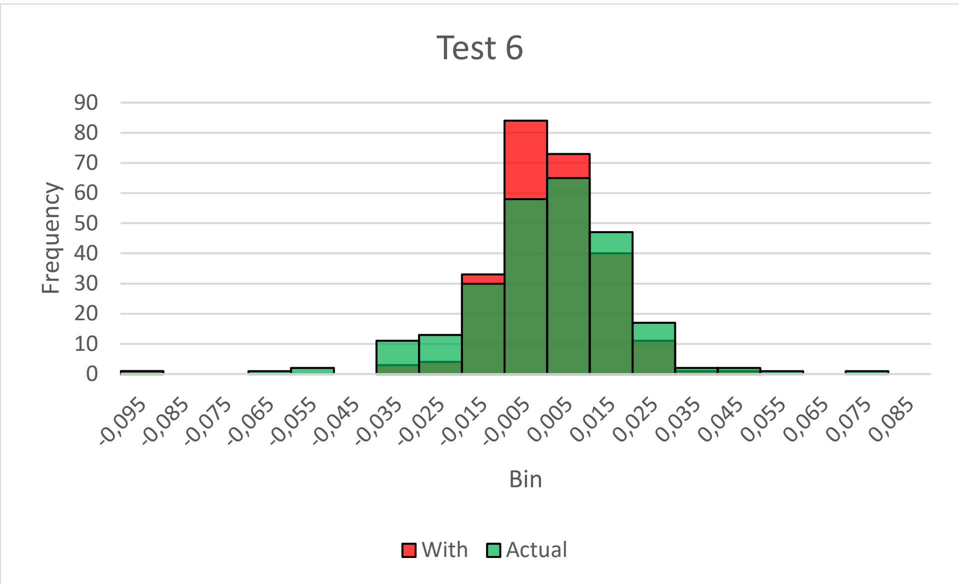

*Figure 4.12: Test 6 Results **with** Hindsight Bias assumption*

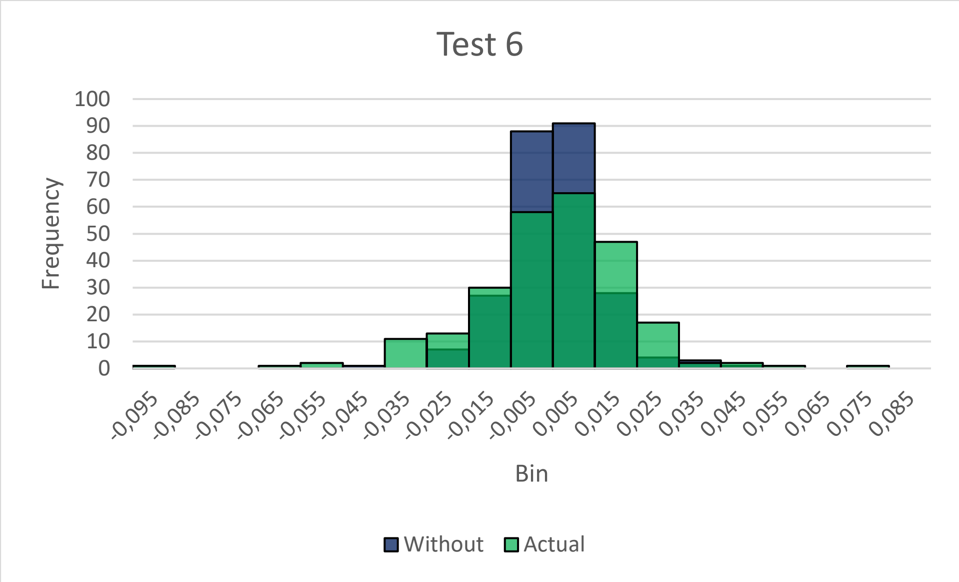

*Figure 4.13: Test 6 Results **without** Hindsight Bias assumption*

The standard practise model has a higher peak at the true return value, meaning the predicted returns are closer to the true returns. Additionally, the hindsight bias model has a narrower distribution, meaning the predicted returns are more consistent, it would be considered better than the standard practise model with a wider distribution of predicted returns. The range observed for the standard model is closer to the actual returns, when compared to the standard practise model.

Therefore, the hindsight bias is considered to be a slightly better fit to the actual data.

4.5.4 Summary

In Section 4.5.1 the returns for all six tests have indicated a mean around zero. The hindsight bias models have a mean close to the true returns and observes a small standard deviation. The hindsight bias model also shows higher peaks at the true returns value, which indicate the model to be more consistent as they have a narrower distribution than the standard model.

The results of comparing the standard practice model and the hindsight bias model against actual returns were inconclusive, as neither model showed a clear advantage over the other. This suggests that further analysis is needed to determine the most effective approach for stock price forecasting in this particular context.

Assessing the histograms of two models against the actual returns can provide some insight into the performance of the models, but it does not provide a convincing conclusion to determine the best fit model. A histogram can show the distribution of returns for a particular model, but it does not take into account the accuracy of the model in predicting future stock prices.

Furthermore, a histogram comparison between models and actual returns may not fully capture the differences in the models' performance, as it does not account for other relevant factors such as the size of the errors or the consistency of the predictions over time.

In conclusion, the hindsight bias model in the context of this study has the ability to capture volatility in the market and the actual results significantly better than the standard model.

4.6 Time Series Plots

Interpreting the actual results against the proposed results of a predicted time series plot involves comparing the predicted values (from the standard model and hindsight bias model) against the actual values. The plot shows how well the models can replicate the actual results over time.

The closer the predicted values are to the actual values, the better the model is at replicating the actual results. If the predicted values deviate significantly from the actual values, this indicates that the model is not performing well. Additionally, the variability of the predicted values should also be considered. If the predicted values are highly variable, it indicates that the model is not stable and may not be reliable. On the other hand, if the predicted values are relatively stable, it indicates that the model is more reliable.

It is worth noting that a single plot may not be sufficient to fully evaluate the performance of a model.

When comparing the time series of a model's forecasted return against the actual return, some key observations that should be noted are:

- Accuracy: How closely does the forecasted return match the actual return over time? This can be measured using metrics like Mean Absolute Error (MAE) or Mean Absolute Error (MAPE), this will be discussed in section 4.7.
- Consistency: Is the model's forecast consistently accurate, or does its accuracy fluctuate over time?
- Trend: Does the forecasted return follow the trend of the actual return? If not, it could indicate that the model is failing to capture important underlying patterns.
- Seasonality: Does the forecasted return account for seasonal patterns in the actual return? If not, it could result in over or under forecasting during certain times of the year.

The Figures below present a comparison of the forecasted time series plots for the standard practice model and the hindsight bias model against the actual result over a forecasted 252-day time period.

Test 1

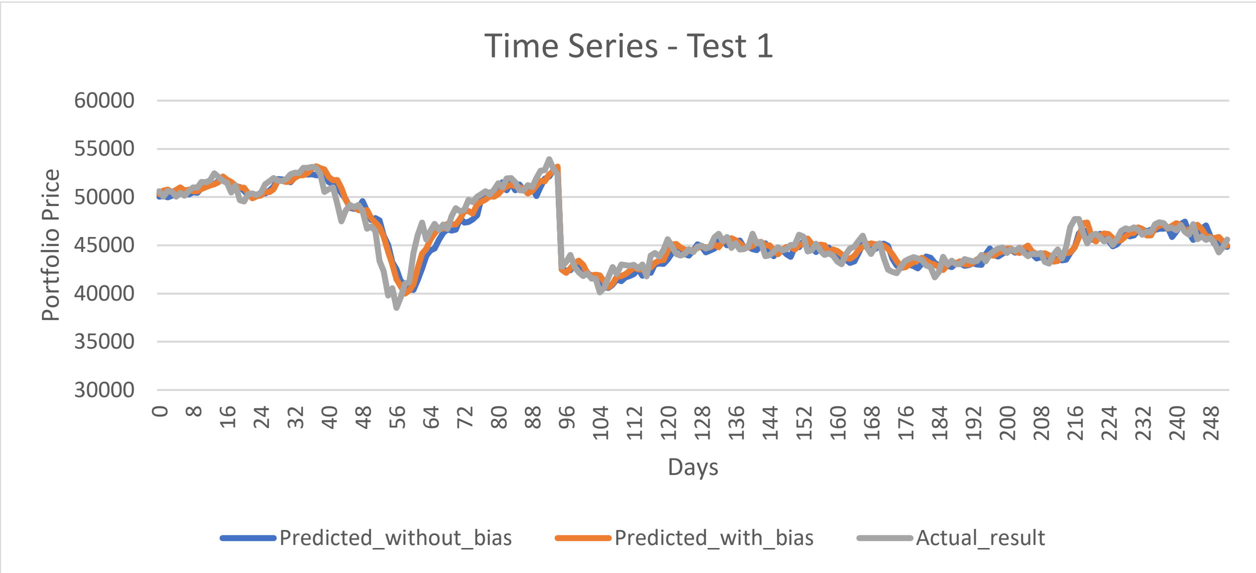

Figure 4.14: Test 1 - Comparing results for the three different predictions

It is evident that the model with hindsight bias is able closely follow the actual results, especially when there are significant drops or rises occurring. However, the standard model is also able to effectively mimic the trend and seasonality of the actual result.

Test 2

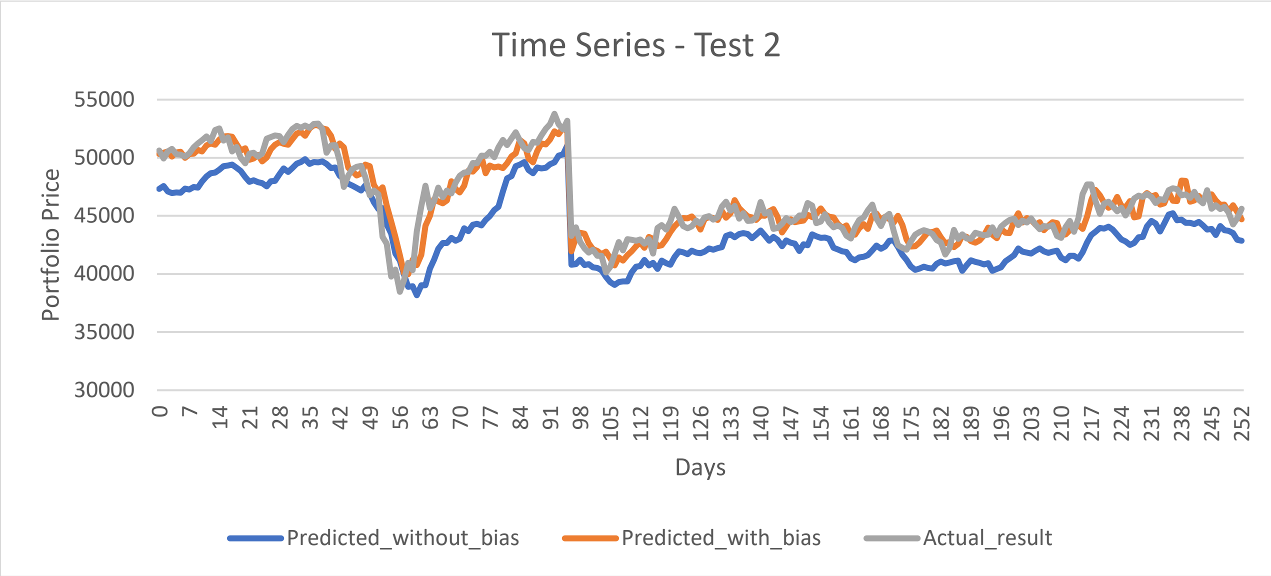

Figure 4.15: Test 2 - Comparing results for the three different predictions

It is very clear that the model with hindsight bias is able closely follow the actual results, especially when there are significant drops or rises occurring, whereas the standard practise model fails to do so and clearly understates the observed returns over time. The model does however mimic the trend and seasonality effectively.

Test 3

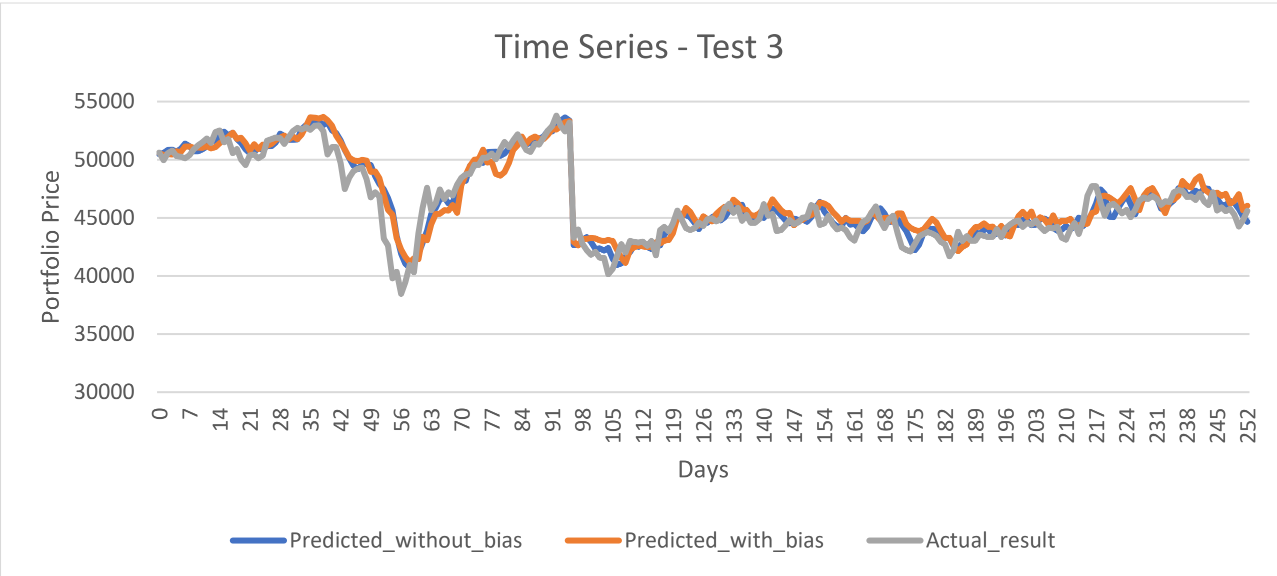

Figure 4.16: Test 3 - Comparing results for the three different predictions

It is evident that the model with hindsight bias is able closely follow the actual results, especially when there are significant drops or rises occurring. The standard model does however also effectively mimic the trend and seasonality of the actual result.

Test 4

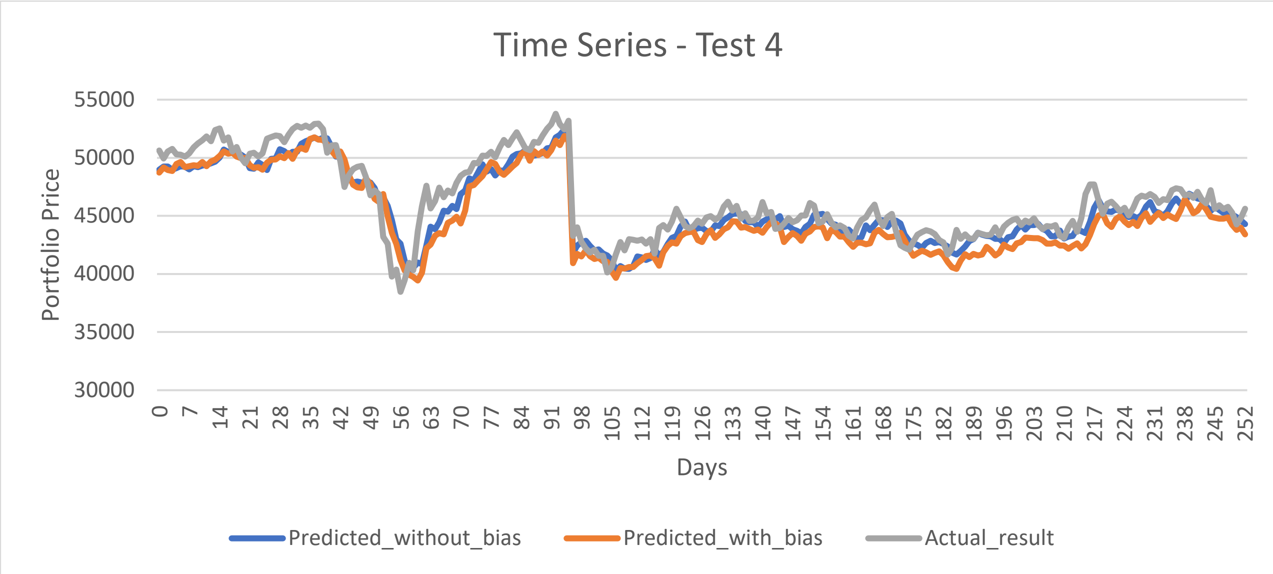

Figure 4.17: Test 4 - Comparing results for the three different predictions

It can be seen that both models are equally effective in following the actual results. The standard model does however prove to be more accurate and is able to follow the actual results over time and both models are able to effectively mimic the trend and seasonality of the actual result.

Test 5

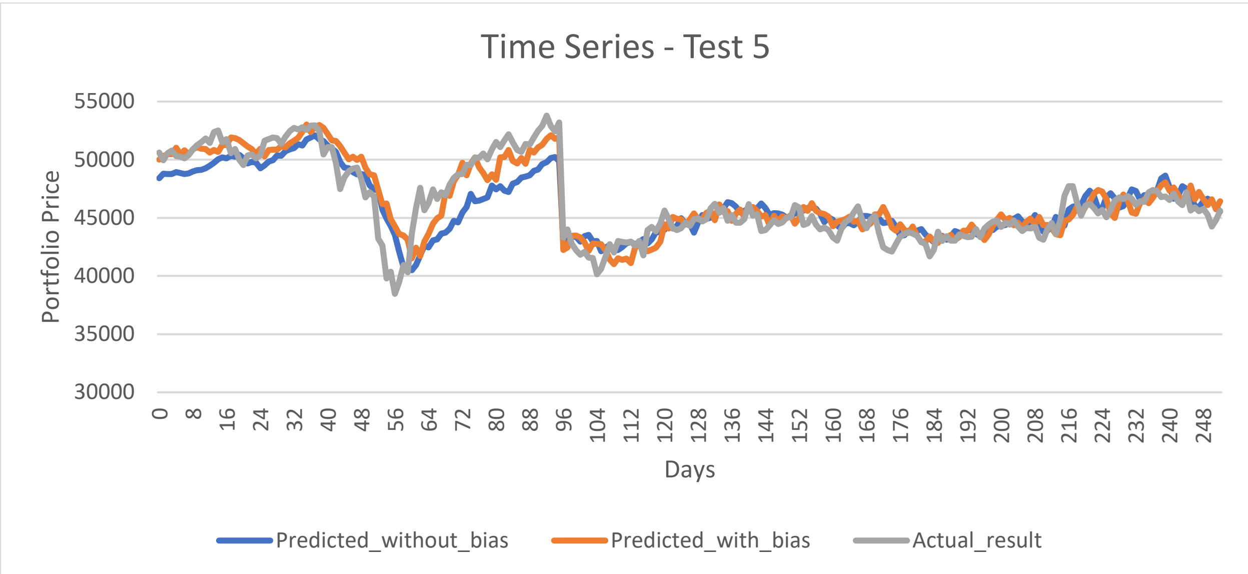

Figure 4.18: Test 5 - Comparing results for the three different predictions

It is very clear that the model with hindsight bias is able closely follow the actual results, especially when there are significant drops or rises occurring, whereas the standard practise model fails to do so and clearly understates the observed returns over time. The standard model does relatively follow the trend and seasonality of the actual results.

Test 6

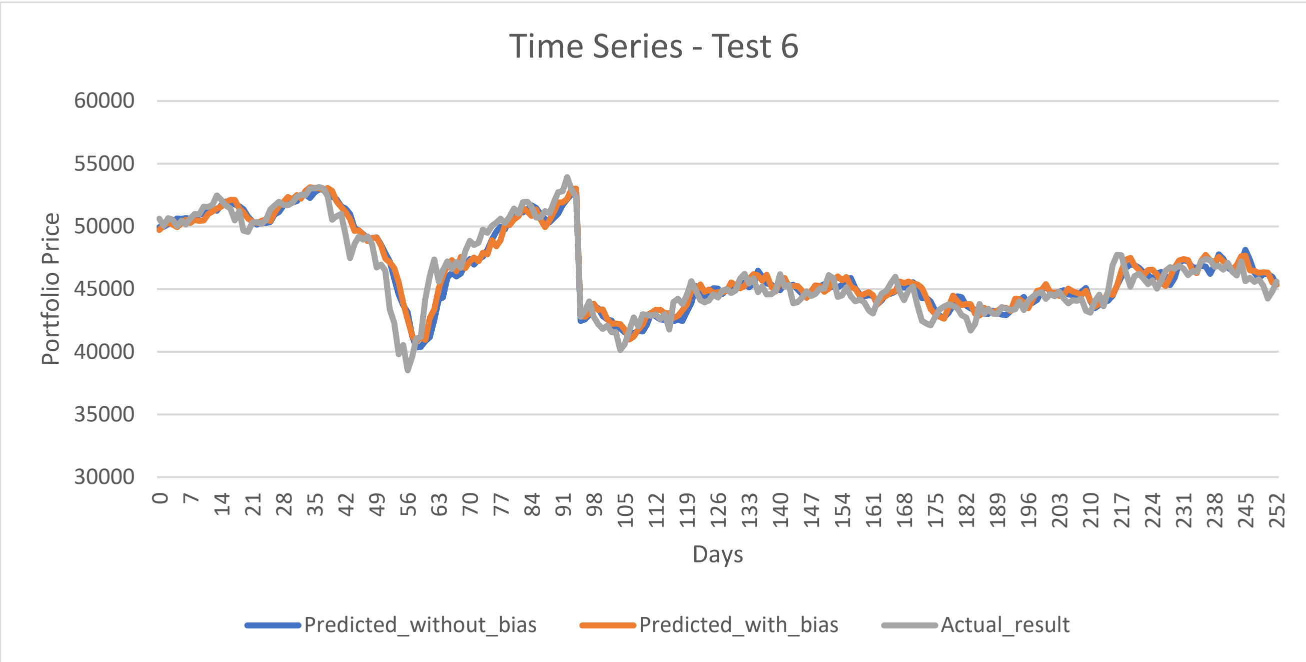

Figure 4.19: Test 6 - Comparing results for the three different predictions

It can be seen that both models are equally effective in following the actual results. Both models are able to effectively mimic the trend and seasonality of the actual result.

4.6.1 Conclusion

As shown by the plots, the proposed model, which takes into account hindsight bias, is able to more accurately replicate actual market returns. The high volatility seen in the plots can be attributed to the fact that investors may develop a mental or statistical model that fits well with historical data but has not been tested on unseen data. As a result, when the investor uses this model to make predictions, they are basing their decisions on a limited set of information, leading to more errors in their predictions and ultimately higher volatility in the financial markets.

The hindsight bias model may replicate market returns better than the standard model because it incorporates additional information or factors that the standard model does not consider. By incorporating this bias into the model, the proposed model may be able to more accurately replicate market returns as it is able to account for the way investors may have acted based on their perception of past events.

4.7 Inferential Statistics

Tables 4.8 – 4.13 below indicate the loss, RMSE, MAE of the training and validation data sets. The RMSE and MAE of the models were calculated and the MSE Comparison Tables were produced.

Row name: ***Improve*** simply indicates if the model's values in the table have improved when introducing hindsight bias into the model. The first three columns in the table represent the values obtained for the training set and the following three columns represent the validation dataset values. The values in tables are observed after 50 Epochs. The expectation will be that the loss will decrease with each epoch iteration, therefore the model will effectively predict the output more accurately as the model continues to train. If the loss decreases exponentially, it suggests that the model gains accuracy as the model is trained, and epochs are increased. The proposed hindsight bias model has indicated improved performance in stock price forecasting for Test 1; Test 2; Test 4 and Test 5.

To compare the performance of two models based on RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error) metrics, smaller values for both RMSE and MAE are beneficial. A smaller value indicates that the model is making fewer errors and making predictions that are closer to the actual values. If both RMSE and MAE are smaller for one model than the other, then it can be concluded that this model is performing better. However, it is important to keep in mind that these metrics should be compared between the training and validation datasets, and not directly between two models.

If the root mean squared error (RMSE) of a model is lower for the training data set than the validation data set, it means that the model is overfitting. This means that the model is too closely fitting the training data set and is not generalizing well to new, unseen data. A model with a lower RMSE for the validation data set is preferred as it suggests that the model is making accurate predictions on unseen data, which is a more important metric for evaluating the performance of a model.

Test 1

Table 4.8: Test 1 - Training and Validation results for the Standard practise model and Hindsight Bias model

	train_loss	**train_rmse**	**train_mae**	**val_loss**	**val_rmse**	**val_mae**
Without	3 883 188	1 970.58	875.19	12 691 710	3 562.54	1 616.40
With	2 550 582	1 597.05	759.97	5 215 658	2 283.78	1 138.46
Improve	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

The loss, RMSE and MAE values have improved when using the hindsight bias model when assessing the training data, however both models' validation data results, show worse values for the loss, RMSE and MAE and therefore implying that the models have overfitted.

Test 2

Table 4.9: Test 2 - Training and Validation results for the Standard practise model and Hindsight Bias model

	train_loss	train_rmse	train_mae	val_loss	val_rmse	val_mae
Without	8 531 153	2 920.81	1 437.11	44 805 824	6 693.72	3 193.74
With	2 952 360	1 718.24	782.66	5 380 167	2 319.52	1 129.66
Improve	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

The loss, RMSE and MAE values have improved when using the hindsight bias model when assessing the training data, however both models' validation data results, show worse values for the loss, RMSE and MAE and therefore implying that the models have overfitted. The standard model's results have significantly worsened when assessing the validation results, as this model has significantly overfitted on the validation data set.

Test 3

Table 4.10: Test 3 - Training and Validation results for the Standard practise model and Hindsight Bias model

	train_loss	train_rmse	train_mae	val_loss	val_rmse	val_mae
Without	1 902 551	1 379.33	681.23	6 584 461	2 566.02	1 260.58
With	3 068 173	1 751.62	790.04	9 456 200	3 075.09	1 372.35
Improve	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

The loss, RMSE and MAE values have improved when using the standard model when assessing the training data, however both models' validation data results, show worse values for the loss, RMSE and MAE and therefore strongly implying that the models has overfitted.

Test 4

Table 4.11: Test 4 - Training and Validation results for the Standard practise model and Hindsight Bias model

	train_loss	train_rmse	train_mae	val_loss	val_rmse	val_mae
Without	18 556 543	4 307.73	1 982.96	44 194 193	6 647.87	3 178.96
With	26 809 882	5 177.83	2 248.85	34 790 738	5 898.37	2 669.58
Improve	FALSE	FALSE	FALSE	TRUE	TRUE	TRUE

The loss, RMSE and MAE values have improved when using the hindsight bias model when assessing the training data, however both models' validation data results, show worse values for the loss, RMSE and MAE and therefore strongly implying that the models has overfitted.

The hindsight bias model has significantly overfitted when compared against the standard model.

Test 5

Table 4.12: Test 5 - Training and Validation results for the Standard practise model and Hindsight Bias model

	train_loss	train_rmse	train_mae	val_loss	val_rmse	val_mae
Without	2 172 462 438	46 609.68	14 978.95	1.23E+09	35 071.36	11 851.68
With	5 878 629	2 424.59	1 116.07	10 641 967	3 262.20	1 517.74
Improve	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE

The loss, RMSE and MAE values have improved when using the hindsight bias model when assessing the training data, however both models' validation data results, show worse values for the loss, RMSE and MAE and therefore strongly implying that the models has overfitted.

Test 6

Table 4.13: Test 6 - Training and Validation results for the Standard practise model and Hindsight Bias model

	train_loss	train_rmse	train_mae	val_loss	val_rmse	val_mae
Without	3 883 530	1 970.67	872.37	11 437 602	3 381.95	1 481.50
With	6 177 877	2 485.53	1 041.84	13 909 607	3 729.56	1 525.18
Improve	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

The loss, RMSE and MAE values have improved when using the standard bias model when assessing the training data, however both models' validation data results, show worse values for the loss, RMSE and MAE and therefore implying that the models has overfitted. The hindsight bias model has significantly overfitted when compared against the standard model.

4.7.1 Summary

The key takeaway from tables 4.8 – 4.13 is to find that the model with hindsight bias performs better than the standard model without hindsight bias.

A standard Artificial Neural Network (ANN) model may show a worse loss function for validation data compared to training data due to overfitting. Overfitting occurs when the model has learned the patterns in the training data too well and is unable to generalize to new, unseen data. As a result, the model may perform well on the training data but poorly on the validation data.

The main reasons for overfitting in ANN models include having too many parameters, a lack of regularization, or having too small of a training data set. When the model has too many parameters, it can learn the noise in the training data, leading to poor generalization performance. Regularization helps to prevent overfitting by adding a penalty term to the loss function, discouraging the model from learning the noise in the data. The regularisation techniques used in this model did not reduce the existing overfitting.

Another suggestion is to find the optimal epoch, where the validations results was at its lowest. When increasing the number of epochs during the training of an Artificial Neural Network (ANN) model, several things can potentially happen. Overfitting occurs where there is an increase in the number of epochs, the model may continue to learn the patterns in the training data, leading to overfitting, where the model performs well on the training data but poorly on unseen data. To prevent overfitting, it is common to use early stopping, where the training is stopped when the performance on the validation data plateaus or starts to decline.

4.8 Nonparametric tests

When dealing with non-normal data, non-parametric tests can provide more reliable and robust results for hypothesis testing and model comparison. Financial data, such as stock returns, often exhibit non-normal distributions and the models' performance may not be normally distributed.

4.8.1 Mann-Whitney U Test

The Wilcoxon Rank Sum Test, also commonly known as the Mann-Whitney U Test is a non-parametric alternative to the t-test and can be used when data is not normally distributed. This test is used to compare the central tendencies (median) of two independent populations and determine if there is a statistically significant difference between two models. It performs a similar function as the two-sample independent t-test except that, unlike in the two-sample independent t-test, it does not require the normality of the population.

When comparing two models (Standard ANN model and Hindsight bias ANN model) against the actual results, the Mann-Whitney U Test would be a suitable test, when there is no assumption made on the specific underlying distribution of the data.

The test statistics for the Wilcoxon Rank Sum Test (Mann Whitney U Test) is denoted by U and is the smaller U_1 and U_2 defined by:

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1 \tag{4.6}$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2, \tag{4.7}$$

where n_1 and n_2 are the sample sizes of the two models, and R_1 and R_2 are the sums of the ranks.

The test is described as:

- Null Hypothesis (H_0): $\eta_1 = \eta_2$The medians of the two samples are equal (there is no difference between the groups).
- Alternative Hypothesis (H_1): $\eta_1 \neq \eta_2$The medians of the two samples are not equal.
- The significant level is $\alpha = 0.05$.

The U statistic is calculated for one of the samples and represents the sum of the ranks for that sample. The other sample's U statistic is then determined by subtracting the first U from the total sum of ranks. The U statistic is used to calculate the p-value. The Mann-Whitney test uses a normal approximation method to determine the p-value of the test,

$$Z = \frac{U - \frac{n_1 . n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}, \tag{4.8}$$

this can be used to obtain the two-sided p-value.

Table 4.14 Mann-Whitney U Test Results

Parameter	**Standard model**		**Proposed Model**	
	U Statistic	**p-value**	**U Statistic**	**p-value**
Test 1	31 025.5	0.9577	30 971.5	**0.9554**
Test 2	30 785.5	**0.9474**	31 128.5	0.9621
Test 3	31 160.5	0.9635	30 752.5	**0.9459**
Test 4	30 732.5	**0.9451**	30 958.5	0.9548
Test 5	31 174.5	0.9641	31 001.5	**0.9567**
Test 6	30 909.5	0.9527	30891.5	**0.9519**

When the p-value is less than the significance level (e.g., 0.05), the null hypothesis can be rejected, and the conclusion that there is a significant difference between the medians of the two samples can be reached. When the p-value exceeds the significance level, the null hypothesis fails to reject, indicating that there is insufficient evidence to determine a statistically significant difference between the medians of the two samples. This could imply that the two models are equally incapable of forecasting actual results.

When assessing the p-value from the Mann-Whitney U Test, it can be deduced from the table that the model with hindsight bias performs better for four of the six tests. When one p-value is significantly lower than the other, it implies that there is a statistically significant difference in performance between the two models in terms of accuracy or efficacy in capturing actual returns. In such instances, the test can infer that the model with the lower p-value is statistically significantly better than the model with the higher p-value at forecasting or mimicking actual returns. However, because the p-values for the two models are remarkably similar, determining which model would perform better in general for various datasets would be challenging.

4.8.2 Diebold-Mariano Test

The Diebold-Mariano Test is a statistical test that compares the predicting accuracy of two competing models using a rigorous statistical framework. It is especially useful in financial and economic forecasting since it allows researchers to determine whether one forecasting model significantly outperforms another.

Accurate forecasts are critical in financial markets for making informed investment decisions. Researchers can use the Diebold-Mariano Test to determine if a proposed model, such as the Proposed Hindsight Bias model, has a statistically significant advantage over the standard model. This is accomplished by comparing the accuracy of their forecasts to the actual results.

The Diebold-Mariano Test can be used to quantify the performance of the two models and assess whether any detected differences are statistically significant.

The Diebold-Mariano Test is used to examine whether the forecasts of two models are significantly different. $f_1, \dots, f_n$ and $g_1, \dots, g_n$ represent two sets of forecasted returns and $y_1, \dots, y_n$ represent the actual time series data. Let e_i and r_i be the residuals of the two forecasted models:

$$e_i = y_i - f_i \qquad r_i = y_i - g_i \tag{4.9}$$

d_i can be defined as:

$$d_i = e_i^2 - r_i^2. \tag{4.10}$$

The time series d_i is called the loss-differential. Define:

$$\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \quad \mu = E[d_i], \tag{4.11}$$

for $n > k \geq 1$, define:

$$\gamma_k = \frac{1}{n}\sum_{i=k+1}^{n} (d_i - \bar{d})(d_{i-k} - \bar{d}), \tag{4.12}$$

where γ_k is the autocovariance at lag k.

For $h \geq 1$, the Diebold-Mariano statistic can be represented as:

$$DM = \frac{\bar{d}}{\sqrt{\frac{\left[\gamma_0 + 2\sum_{k=1}^{h-1}\gamma_k\right]}{n}}}, \tag{4.13}$$

where it would generally suffice to use $h = n^{\frac{1}{3}} + 1$.

Assume the null hypothesis: $\mu = 0$, where $DM \sim N(0,1)$.

Thus, there is a significant different between the forecasts if $|DM| > z_{\alpha/2}$ where $z_{\alpha/2}$ is the two-tailed critical value for the standard normal distribution. The key assumption for using the Diebold-Mariano test is that the loss differential time series d_i is stationary, (Zaiontz, 2022).

Table 4.15: Diebold-Mariano Test Results

Test Case	DM Statistic	p-value	Result
Test 1	1.2941	0.1956	Do not reject
Test 2	1.4350	0.1512	Do not reject
Test 3	1.2220	0.2216	Do not reject
Test 4	1.1971	0.2312	Do not reject
Test 5	1.5936	0.1110	Do not reject
Test 6	1.7169	0.0859	Do not reject

Most of the p-values obtained in Table 4.15 are high and do not reject the null-hypothesis. A high p-value from the Diebold-Mariano test indicates that there is no statistically significant difference in forecasting accuracy between the standard model and the hindsight bias model when compared to actual returns. In other words, both models forecast real returns identically, and any variations found can potentially be attributed to random fluctuations. Finally, the results indicate a lack of significant evidence to claim that the one model's forecasts are consistently better than the other models across a range of different dataset layouts.

While the p-values show that one model has no statistically significant advantage over the other, this result is meaningful in and of itself. This implies that, despite its innovative approach, the hindsight bias model may not always outperform the traditional model in terms of forecasting accuracy.

4.9 Conclusion

One might say that the suggested model can "learn" the law of this time series from the inferential statistics and patterns detected within the time series. The findings offer valuable insights for following studies and considerations. The figures show that the hindsight model is capable of closely predicting and identifying a trend in the time series. It should also be noted that increasing the number of epochs, time series data, and DNN layers can all enhance accuracy.

The loss values improve when incorporating hindsight bias and it is clear that the standard practise model have weaker loss values. The risk is that the optimal parameter set might not be identified as the model is training and validating on the same data and therefore the model might not be able to generalise to novel data. This essentially describes how hindsight bias exist in real word data – investors might over train their mental models on historical data without testing them out of sample. An artificial neural network (ANN) model that has been trained on specific historical data may perform well on that data, but struggle to generalize to new, unseen data.

Although the model overfits on validation data, it is evident from the Time Series Plots in section 4.6 that the hindsight bias model generally performs better against the actual prices than the standard practise model. This would imply that the hindsight model is able to effectively forecast stock prices, even when the validation data is overfitted. It is important to note that overfitting would obviously be seen in the hindsight bias model, since the validation data forms a subset of the training data set and therefore, the model is using seen data as its validation data set.

Forming the validation data set as a strict subset of the training data set would eventually lead to overfitting. This occurs because the model has memorized the training data instead of learning the underlying patterns and relationships.

The model has learned patterns that are specific to the training data but may not be relevant in the future. This can lead to a higher accuracy on the training data, but lower accuracy on unseen data, especially when the training data is more recent. Additionally, having hindsight bias can also lead to a lack of robustness, meaning that the model may not perform well in different scenarios, with different data distributions and splits. This can lead to a model that performs well when the terms are defined and split differently.

Although the validation data is overfitted, the best way to determine if a model is overfitting is to evaluate its performance on a separate, independent test data set that has not been used during training or validation, which was tested for this study. The test set showed improved results for the hindsight bias model, specifically when observing the time series plots.

Even in the absence of statistically significant difference, it is important noting the originality and value of this research. When studying and comparing the standard model with the hindsight bias model, rigorous statistical tests were applied, adding to the body of knowledge in the field of financial forecasting. This is especially true if there has been little research on the subject in a South African context specifically. These important insights help to understand the potential implications of hindsight bias on forecasting accuracy.

Chapter 5:

Summary, Conclusions and Recommendations

5.1 Introduction

Incorporating exchange data along with emotional biases and using behavioural finance assumptions about the dataset and its inputs would result in better results. The model forecasted the adjusted closing stock price of the listed companies using a model called the DNN (a variant of ANN). ANN's have been used by researchers to question the validity of the EMH.

The two models introduced were analysed and assessed to determine how well each model performed; what their prediction accuracies were and finally the results from each model were compared and measured.

5.2 Key takeaways

No assumptions were made regarding the rationality of the investor and this differs from just about every other modelling methodology in behavioural finance asset pricing and instead the investor is modelled with just the perception of the market data and that perception can either be a rational perception where investors have tested their mental model out of sample or as the study have seen with the hindsight bias, there are existing implicit cognitive bias and this allows the study to model that within ANN methodology.

In the context of a machine learning model, incorporating hindsight bias would involve using knowledge of the outcomes of a financial market to make predictions about its future. This can result in the model memorizing the training data and not generalizing well to new data. The notion that a model that includes hindsight bias would be more accurate is not universally accepted in the field of machine learning and artificial intelligence. In fact, including hindsight bias in a model can often lead to overfitting and reduced accuracy.

It is important to note that, financial markets are often characterized by uncertainty, and including hindsight bias in a model may not accurately reflect the true nature of the market.

In conclusion, while the notion of a model that includes hindsight bias may seem attractive, it may not always be a reliable approach and may lead to reduced accuracy and overfitting. It is generally better to use models that are based on well-established machine learning algorithms and principles, and to avoid incorporating any biases or preconceived notions into the model.

Even if the study finds that incorporating hindsight bias only results in a slight improvement in the accuracy of the ANN model, the study's contribution remains unique and valuable. This is because the study is filling a gap in the existing body of knowledge by exploring the impact of cognitive biases (specifically hindsight bias in this study) on stock market forecasting using ANNs. The study's contribution to the field lies in its empirical exploration of the topic and its results, whether positive or negative, adds to the understanding of how behavioural factors such as hindsight bias can impact financial forecasting models. The results of the study also provide a foundation for further research in the field and offer practical implications for practitioners in finance.

5.3 Future Research and Considerations

There are several ways to continue the research direction initiated by this work. Implementing other machine learning algorithms to see if approaches like regression trees, SVN, and nearest neighbour can improve outcomes, is one of the straightforward expansions that would be taken into consideration.

Modelling cognitive biases in financial markets using artificial neural networks (ANNs) can be done by incorporating behavioural factors into the model. This can be achieved by combining traditional financial indicators, such as stock prices and trading volume, with behavioural indicators, such as sentiment analysis of news articles and social media posts. The ANN can then be trained to predict market movements based on this combined dataset, taking into account both traditional financial factors and behavioural factors that may influence investor decision-making. Another way to do this is to use traditional financial indicators, such as stock prices and trading volume, with behavioural indicators, such as self-report measures of confidence in predictions before an outcome was revealed, as input features in the ANN and train the ANN to predict market movements based on this combined dataset, taking into account both traditional financial factors and behavioural factors that may influence investor decision-making.

An alternative method would be to use a generative model, such as a generative adversarial network (GAN), to create synthetic market situations that incorporate hindsight bias. These scenarios can then be used to simulate the effect of hindsight bias on market movements and to identify specific scenarios that may be particularly susceptible to the bias. Another approach would be to train an ANN on historical market data and use the model to generate synthetic market situations, which can be used to simulate the effect of irrational investor behaviour on market movements. In this case, the ANN's predictions can be compared to actual market data to evaluate the accuracy of the model and identify specific behavioural factors that may be driving irrational decision-making. It's worth mentioning that this field is an active area of research, and there's no definitive way of modelling it.

5.3.1 Survey-based data

The study's major drawback was its failure to collect survey-based data from investors. It was simply not a viable option to create and distribute surveys to the general population due to time and resource limitations. If investors had contributed by outlining their thought processes and decisions, the data could have been more valuable. It would have been useful to analyse the reasons why investors might steer clear of particular equities in the market amid some business announcements. This type of knowledge could have provided an explanation for why investors behave in a particular way, what motivates them to act on instinct, and why emotion rather than the information at their disposal is more important.

5.3.2 Sentiment Analysis

Over the past few years, researchers have tried to prove that market sentiment, formally defined by Malcolm Baker and Jeffrey Wurgler as *“a belief about future cash flows and investment risks that is not justified by the facts at hand”* could affect stock prices, (Baker and Wurgler, 2007). Zhang clearly stated that there is no widely accepted definition for investor sentiment and many researchers have different opinions on this matter, (Zhang, 2008).

A study by Lavrenko, (2000) presented an approach to predicting stock exchange movements, where the study attempted to forecast future stock prices by analysing news articles related to the specific company's stock. The study found that it is possible that a single article can affect trends found in stock prices. Within this study, a language model was designed to find which words were used that relate to the trends observed in the market.

Unstructured data can be obtained through social media, web searches, etc., and it can now be incorporated with the publicly available information mentioned by Eugene Fama. It is now possible to quantify the value of including Tweets, news stories, and other content, which was not always straightforward a few years ago. Computers can read and "understand" the context of a written document using a method known as natural language processing in the field of artificial intelligence. The assessment would then determine if the written document was favourable, unfavourable about the stock of a particular company.

5.3.3 Google Trends

Additionally, Google Trends has been used to look into the relationship between results and the quantity of searches for a specific company's name. Google trends and search data can be used to anticipate a company's trading volume, (Da et al., 2011). A crucial finding was the correlation between a negative mood or term and a drop in investor confidence.

5.4 Conclusion

This assignment has analysed the impact of hindsight bias on stock price forecasting in the South African equity market. By comparing the standard practice model and the hindsight bias model, this study aimed to provide insights into the effect of incorporating past stock price information into the validation process and consequently, the forecasting process. The results showed that, while the models had mixed results in terms of summary statistics, R-squared values, and histograms, it is clear that the impact of hindsight bias on stock price forecasting is not necessarily convincing. Further research might be required to fully understand the effects of hindsight bias on stock price forecasting and to develop more effective forecasting models. However, this study serves as an important step towards understanding the complex relationships between hindsight bias, stock price information, and stock price forecasting in the South African equity market.

To summarise, the aim of this study was to investigate the impact of incorporating the behavioural element of hindsight bias in the forecasting of stock prices in the South African equity market. The validation dataset was merged with the training dataset to form a subset, which served as a crucial component in determining the effectiveness of hindsight bias. The literature review provided background information on behavioural finance and its impact on investor decision making in developing markets. The results of the study showed that adding hindsight bias to an DNN-ANN model can improve its accuracy, challenging the conventional modelling methodologies. Investors are encouraged to consider the role of behavioural finance in explaining market anomalies and consider incorporating hindsight bias in their forecasting models.

Disclosure / ethical consideration

The data collected for this study does not include any personal data, but only publicly available secondary data on YahooFinance and no consent was required in the acquisition of data. As a result, the study made sure it is private and of the utmost integrity. The study does not put anything or anyone at risk of any harm.

Applying the techniques and philosophies discussed in this study may cause other people to lose money because the work done by individuals may be inaccurate and defective. It is worthy to clarify that the findings and work done should not be interpreted as financial advice.

In addition, the study used the Harvard referencing style to cite content written by other authors that was used in this study.

Program Code

This appendix provides an overview of all the code used in the project, including the implementation of the proposed model and the standard model. The code is written in Python and makes use of popular libraries such as pandas, numpy, matplotlib, and yfinance. The implementation of the proposed model includes the use of advanced techniques such as implementing a DNN model and regularization methods. The standard model is also implemented using common techniques such as feedforward neural networks.

The code is provided in a clear and well-organized manner, with detailed comments that explain the different steps of the process. It is accompanied by the step that extracted input data files from YahooFinance, in case someone wants to replicate the results. The code is also open-source and available for anyone to use, modify, or build upon.

It's worth noting that this appendix is intended for readers who have a basic understanding of programming and machine learning concepts. However, even if not, the comments in the code and the explanations in the main text should be enough to understand the main steps of the project.

The script starts off by extracting data information from YahooFinance for the chosen companies listed in Section 4.2. Once the data has been extracted, the code will split the input data into the three categories. Once split there will be two different model. The First model will resemble a standard neural network model and the second model will represent the Hindsight Bias Model. Once both models are set up, the predicted output will be produced.

The script used to model the Artificial Neural Network (ANN) model is presented below for reference and includes the steps and algorithms necessary to train and test the model for Test 1.

```
# Importing the needed modules
import pandas as pd
import numpy as np
from matplotlib import pyplot as plt

import yfinance as yf

# %%
# Creating a function to split the data
def convert2matrix(data_arr, look_back):
    X, Y = [], []
    for i in range(len(data_arr) - look_back):
        d = i + look_back
        X.append(data_arr[i:d, 0])
        Y.append(data_arr[d, 0])
    return np.array(X), np.array(Y)
```

```
# %%
# Defining Neural Network
from keras.models import Sequential
from keras.layers import Dense

# model function

def model_dnn(look_back):
    model = Sequential()
    model.add(Dense(units=30, input_dim=look_back, activation='relu'))
    model.add(Dense(8, activation='relu'))
    model.add(Dense(1))
    model.compile(loss='mean_squared_error', optimizer='adam',
metrics=['mse', 'mae'])
    return model

# Model Evaluation function

def model_loss(history, label, entity):
    plt.figure()
    plt.plot(history.history['loss'], label='Train Loss')
    plt.plot(history.history['val_loss'], label='Test Loss')
    plt.title(f'model loss - {label}')
    plt.ylabel('loss')
    plt.xlabel('epochs')
    plt.legend(loc='upper right')
    plt.show()
    plt.savefig(f'{entity} - {label}.jpg')

# %%
# Downloading the data and saving as a csv

# Declaring Environment Variables
start_date = '2010-01-02'
end_date = '2020-12-31'

# list_of_entities = ['AAPL', '^GSPC', 'AMZN', 'META', 'MSFT', 'NVDA',
'ORCL', 'QCOM', 'SBUX', 'TSLA', 'TWTR', 'VZ']

list_of_entities = ['AGL', 'NPN', 'SBK', 'ABG', 'SOL', 'ANH', 'SHP', 'BVT',
'CLS', 'CFR']

for entity in list_of_entities:

    # entity_df = yf.download(f'{entity}.JO', start=start_date,
end=end_date)
    #
    # entity_df.to_csv(f'{entity}.csv')

    # %%
    # Reading in the csv

    entity_df = pd.read_csv(f'{entity}.csv')

    entity_df = entity_df[['Date', 'Close']]

    entity_df = entity_df.set_index('Date')

    # %%
    epochs=50
    batch_size=10
```

```
    verbose=0
    steps_per_epoch=10

    train_size = 80
    test_size = 10
    predict_size = 10

    # Split data set into testing dataset and train dataset for without
bias
    train_size_without = round(train_size * len(entity_df.Close) / 100)
    test_size_without = round(test_size * len(entity_df.Close) / 100) +
train_size_without
    predict_size_without = round(predict_size * len(entity_df.Close) / 100)
+ test_size_without

    train_without, test_without, predict_without =
entity_df.values[0:train_size_without, :], entity_df.values[

train_size_without:test_size_without,

:], entity_df.values[

test_size_without:predict_size_without,

:]
    print(f'For without: train_size_without = {train_without},
test_size_without = {test_without}, predict_size_without =
{predict_without}')
    # Split data set into testing dataset and train dataset for with bias
    train_size_with = round(train_size * len(entity_df.Close) / 100)
    test_size_with = round(test_size * len(entity_df.Close) / 100) +
train_size_with
    predict_size_with = round(predict_size * len(entity_df.Close) / 100) +
test_size_with

    train_with, test_with, predict_with =
entity_df.values[0:test_size_with, :],
entity_df.values[train_size_with:test_size_with,

:], entity_df.values[

test_size_with:predict_size_with, :]
    print(f'For with: train_size_with = {train_with}, test_size_with =
{test_with}, predict_size_with = {predict_with}')
    # setup look_back window
    look_back = 30
    # convert dataset into right shape in order to input into the DNN
without bias
    trainX_without, trainY_without = convert2matrix(train_without,
look_back)
    testX_without, testY_without = convert2matrix(test_without, look_back)
    predictX_without, predictY_without = convert2matrix(predict_without,
look_back)
    # convert dataset into right shape in order to input into the DNN with
bias
    trainX_with, trainY_with = convert2matrix(train_with, look_back)
    testX_with, testY_with = convert2matrix(test_with, look_back)
    predictX_with, predictY_with = convert2matrix(predict_with, look_back)


    # %%
```

```
    # Fit the model without bias
    model_without = model_dnn(look_back)

    history_without = model_without.fit(trainX_without, trainY_without,
validation_data=(testX_without, testY_without), epochs=epochs,
verbose=verbose)
                                # batch_size=batch_size,
                                # steps_per_epoch=steps_per_epoch,
                                # shuffle=False)


    history_without_df = pd.DataFrame(history_without.history)
    history_without_df.to_csv(f'{entity}_history_without.csv')

    # Model scores without bias

    train_score_without = model_without.evaluate(trainX_without,
trainY_without, verbose=0)
    print(
        f'Train_without Root Mean Squared Error(RMSE):
{np.sqrt(train_score_without[1])}; '
        f'Train_without Mean Absolute Error(MAE) :
{train_score_without[2]}')
    test_score_without = model_without.evaluate(testX_without,
testY_without, verbose=0)
    print(f'Test_without Root Mean Squared Error(RMSE):
{np.sqrt(test_score_without[1])}; '
          f'Test_without Mean Absolute Error(MAE) :
{test_score_without[2]}')
    model_loss(history_without, 'without bias', entity)

    results_without = model_without.predict(predictX_without)

    # %%
    # Fit the model with bias
    model_with = model_dnn(look_back)

    history_with = model_with.fit(trainX_with, trainY_with,
validation_data=(testX_with, testY_with), epochs=epochs, verbose=verbose)
                                  # , batch_size=batch_size,
verbose=verbose,
                                  # steps_per_epoch=steps_per_epoch,
                                  # shuffle=False)

    history_with_df = pd.DataFrame(history_with.history)
    history_with_df.to_csv(f'{entity}_history_with.csv')

    # Model scores with bias

    train_score_with = model_with.evaluate(trainX_with, trainY_with,
verbose=0)
    print(
        f'Train_with Root Mean Squared Error(RMSE):
{np.sqrt(train_score_with[1])}; '
        f'Train_with Mean Absolute Error(MAE) : {train_score_with[2]}')
    test_score_with = model_with.evaluate(testX_with, testY_with,
verbose=0)
    print(f'Test_with Root Mean Squared Error(RMSE):
{np.sqrt(test_score_with[1])}; '
            f'Test_with Mean Absolute Error(MAE) : {test_score_with[2]}')
```

```
    model_loss(history_with, 'with bias', entity)

    results_with = model_with.predict(predictX_with)

    # %%
    # Tables to csvs
    results_without = pd.DataFrame(results_without)

    results_with = pd.DataFrame(results_with)

    actual_results = pd.DataFrame(predictY_without)

    final_results = pd.merge(results_without, results_with,
left_index=True, right_index=True)
    final_results = pd.merge(final_results, actual_results,
left_index=True, right_index=True)
    final_results.columns = ['Predicted_without_bias',
'Predicted_with_bias', 'Actual_result']
    final_results.to_csv(f'{entity}_final_results.csv')

    # %%
    print(((final_results['Actual_result'] -
final_results['Predicted_without_bias'])**2).sum())
    print(((final_results['Actual_result'] -
final_results['Predicted_with_bias'])**2).sum())
#%%
```

References

- Achelist, S. B. (1995). Technical analysis from A to Z: covers every trading tool from the absolute breadth index to the zig zag. Chicago, IL: Probus Publisher.
- Aguilar, O. (2021). The behavioral advantage, Schwab Brokerage. Available at: https://www.schwabassetmanagement.com/content/behavioral-advantage (Accessed: February 4, 2023).
- Ayub Shafaq, Jafri Yasmin Zahra (2020). Comparative Study of an ANN-ARIMA Hybrid Model for Predicting Karachi Stock Price, American Journal of Mathematics and Statistics, Vol. 10 No. 1, 2020, pp. 1-9. doi: 10.5923/j.ajms.20201001.01.
- Baker, M and Wurgler, J (2007). Investor Sentiment in the Stock Market. In: Journal of Economic Perspectives 21.2 (2007), pp. 129-152.
- Banger, E.R.S. (2022). Advantages and disadvantages of expert system: Characteristics in ai, DigitalThinkerHelp. Available at: https://digitalthinkerhelp.com/advantages-disadvantages-characteristics-of-expert-system-in-artificial-intelligence/ (Accessed: February 4, 2023).
- Bernard, V.L. (1992). Stock price reactions to earnings announcements: A summary of recent anomalous evidence and possible explanations. Working Paper.
- Bernard, V.L. and Thomas, J.K. (1989). Post-earnings-announcement drift: delayed price response or risk premium? Journal of Accounting research. 27:1–36.
- Bernard, V.L. and Thomas, J.K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. Journal of accounting and economics. 13(4):305–340.
- Blume, L., Easley, D. and O'Hara, M. (1994). Market Statistics and Technical Analysis: The Role of Volume. The Journal of Finance, 49(1), 153-181. https://doi.org/10.1111/j.1540-6261.1994.tb04424.x
- Bre, Juan M Facundo, Víctor D Gimenez, Fachinotti (2018). Prediction of wind pressure coefficients on building surfaces using artificial neural networks; Energy and Buildings, volume 158, p. 1429 – 1441.
- Brownlee, J. (2022). Your first deep learning project in python with keras step-by-step, MachineLearningMastery.com. Available at: https://machinelearningmastery.com/tutorial-first-neural-network-python-keras/ (Accessed: February 4, 2023).
- Buhmann, M. (2010). Radial basis function, Scholarpedia. Available at: http://www.scholarpedia.org/article/Radial_basis_function#:~:text=Radial%20basis%20functions%20are%20means%20to%20approximate%20multivariable,can%20be%

20used%20in%20more%20than%20one%20dimension. (Accessed: February 4, 2023).
- Chan, L.K.C. Jegadeesh, N. and Lakonishok, J. (1996). Momentum Strategies. The Journal of finance (New York). 51(5):1681-1713
- Chen, J. (2022). What is hindsight bias?, Investopedia. Investopedia. Available at: https://www.investopedia.com/terms/h/hindsight-bias.asp (Accessed: February 4, 2023).
- Chordia, T. and Shivakumar, L. (2006). Earnings and price momentum. Journal of financial economics. 80(3):627–656.
- Da, Z. Engelberg, J. Gao, P. (2011). In search of attention, The Journal of Finance, 66, pp. 1461-1499
- De Bondt, W. F. M., and Thaler, R. H. (1985). Does the stock market overreact. Journal of finance, 40, 793-808.
- Diederik P. Kingma, Jimmy Ba, Adam (2015). A Method for Stochastic Optimization (2015).
- Duncan, Brett and Yanqing Zhang (2015). "Neural networks for sentiment analysis on Twitter." 2015 IEEE 14th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC): 275-278.
- Eugene Fama (1970). Efficient Capital Markets: A Review of Theory and Empirical Work." In: The Journal of Finance 25.2, pp. 382-417.
- Felicity (2022). Exploring deep learning & CNNS, RSIP Vision. Available at: https://www.rsipvision.com/exploring-deep-learning/ (Accessed: February 4, 2023).
- Graham, J.R. (1999). Herding among Investment Newsletters: Theory and Evidence. Journal of Finance, 54, 237-268. https://doi.org/10.1111/0022-1082.00103
- Grogan, M. (2020). Regression-based neural networks with tensorflow v2.0: Predicting average daily rates, Medium. Towards Data Science. Available at: https://towardsdatascience.com/regression-based-neural-networks-with-tensorflow-v2-0-predicting-average-daily-rates-e20fffa7ac9a (Accessed: February 4, 2023).
- Haitsma, R., Unalmis, D. and de Haan, J. (2016). The impact of ECB's conventional and unconventional monetary policies on stock markets. Journal of Macroeconomics. 48:101-116.
- Han, S. et al. (1970). Learning both weights and connections for efficient neural network, Advances in Neural Information Processing Systems. Available at: https://papers.nips.cc/paper/2015/hash/ae0eb3eed39d2bcef4622b2499a05fe6-Abstract.html (Accessed: February 4, 2023).

- Hayes, A. (2022). Behavioral finance: Biases, emotions and financial behavior, Investopedia. Investopedia. Available at: https://www.investopedia.com/terms/b/behavioralfinance.asp (Accessed: February 4, 2023).
- Hirshleifer, David A. (2001). Investor Psychology and Asset Pricing. AFA 2001 New Orleans Meetings, Available at SSRN: https://ssrn.com/abstract=265132 or http://dx.doi.org/10.2139/ssrn.265132.
- Hoffman, A.O.I. and Post, T. (2014). Self-Attribution Bias in Consumer Financial Decision-Making: How Investment Returns Affect Individuals' Belief in Skill. Journal of Behavioral and Experimental Economics, 52, 23-28. https://doi.org/10.1016/j.socec.2014.05.005
- Hong, H and Stein, J.C. (1999). A Unified Theory of Underreaction, Momentum Trading, and Overreaction in Asset Markets. The journal of finance (New York). 54(6):2143-2184.
- Hong, D., Lee, C. and Swaminathan, B. (2003). Earnings momentum in international markets. Available at SSRN 390107. 1:1–39.
- Hurra, T. (2022). 9 steps in model training process that demands a sequence, Medium. Towards Data Science. Available at: https://towardsdatascience.com/9-steps-in-model-training-process-that-demands-a-sequence-7672a95fcf79 (Accessed: February 4, 2023).
- January, V. (2021). Adam Optimizer explained, Machine learning journey. Available at: https://machinelearningjourney.com/index.php/2021/01/09/adam-optimizer/ (Accessed: February 4, 2023).
- Joseph, T. (2020). The mathematics behind Deep Learning, Medium. Available at: https://towardsdatascience.com/the-mathematics-behind-deep-learning-f6c35a0fe077 (Accessed: 03 August 2023).
- Kahneman, D., Slovic, P. and Tversky, A. (1982). Judgment under Uncertainty: Heuristics and Biases. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511809477
- Kahneman, D. (2011). Thinking, fast and slow. Farrar, Straus and Giroux.
- Kenton, W. (2022). Look-ahead bias, Investopedia. Investopedia. Available at: https://www.investopedia.com/terms/l/lookaheadbias.asp (Accessed: February 4, 2023).
- Kumar, R. (2016). Valuation: Theories and Concepts. Pages 73-91.
- Lavrenko, V. and Schmill, (2000). M., Lawrie, D., Ogilvie, P., Jensen, D., Allan, J.: Language models for financial news recommendation. In: Proceedings of the Ninth International Conference on Information and Knowledge Management, pp. 389–396.

- Lee, T.S. and Ho, C.K. (2001). "A comparison of neural network and linear regression models for forecasting stock prices", Expert Systems with Applications, vol. 20, pp. 91-102.
- Malekian, A., and Chitsaz, N. (2021). Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. In Advances in Streamflow Forecasting.
- Mullainathan., S, and Thaler, R. (2001). Behavioral Economics. In N. Smelser, & P. Baltes (Eds.), International Encyclopedia of the Social and Behavioral Sciences (Volume 20, pp. 1094-1100). Oxford: Oxford University Press. http://dx.doi.org/10.1016/B0-08-043076-7/02247-6.
- Nielson, D., Nielsen F. and Barnes, B. (2016). An overview of Factor Investing. Leadereship Series. (1):1-8. [Online], Available: https://www.fidelity.com/bin-public/060_www_fidelity_com/documents/brokerage/overview-factor-investing.pdf
- Parisi D., Cecconi F. and Nolfi S. (1990). Econets: Neural networks that learn in an environment, Network, 2, 149–168.
- Refenes, A.N., Zapranis, A., Francis, G. (1994). Stock performance modeling using neural networks: a comparative study with regression models. Neural Netw. 7(2), 375–388. ISSN 0893-6080.
- Rita (2022). Adam: An optimization algorithm for neural networks, Adam: An Optimization Algorithm For Neural Networks. Available at: https://www.surfactants.net/adam-an-optimization-algorithm-for-neural-networks/ (Accessed: February 4, 2023).
- Shen, X. Guo, C. Wu and D. Wu, (2011). “Forecasting Stock Indices Using Radial Basis Function Neural Networks Optimized by Artificial Fish Swarm Algorithm,” Knowledge-Based Systems, Vol. 24, No. 3, 2011, pp. 378-385.
- Tversky, A., and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1130. doi:10.1126/science.185.4157.1124.
- White, Halbert L. (1988). “Economic prediction using neural networks: the case of IBM daily stock returns.” IEEE 1988 International Conference on Neural Networks: 451-458 vol.2.
- Yang, R. (2018). Classifying Hand Written Digits with Deep Learning. Intelligent Information Management, 10, 69-78. doi: 10.4236/iim.2018.102005.
- Zahedi, F. (1993). Intelligent Systems for Business: Expert Systems with Neural Networks, Wadsworth Publishing Company, Belmont, USA, pp. 10-11, 1993.
- Zaiontz, C. (2022). *Diebold-Mariano Test*, *Real Statistics Using Excel*. Available at: https://real-statistics.com/time-series-analysis/forecasting-accuracy/diebold-mariano-test/ (Accessed: 03 August 2023).

-

- Zhang, C. (2008). Modelling, and Measuring Investor Sentiment. The-sis. University of California, Berkeley.
- Zucchi, K. (2022). CAPM model: Advantages and disadvantages, Investopedia. Investopedia. Available at: https://www.investopedia.com/articles/investing/021015/advantages-and-disadvantages-capm-model.asp (Accessed: February 4, 2023).