The Influence of News Sentiment on Stock Price Ratio Movement and its Utilisation for Pairs Trading Strategies

W.F.D. Lee

Faculty of Behavioural and Management Sciences (BMS)

Master Business Administration

University of Twente

Supervisors:

Dr. M. R. Machado

Dr F.S. Bernard

Dr. I. Skute

10-12-2024

Abstract

This study investigated the predictive relationship between sentiment extracted from Google News headlines and the stock price ratios of highly correlated stocks. The aim of the study was to find ways to harness sentiment information to increase the profits of pairs trading strategies. Little research has been conducted on the utilisation of sentiment analysis for pairs trading, and little consensus was found on the temporal nature of the effects of sentiment on stock price or stock price ratio movement; particularly regarding the optimal time periods to sample sentiment, how long the lag between sentiment and effect on close price (or ratio) is, and how long the effect lasts. Furthermore, although many studies confirm that the LSTM models augmented with sentiment can accurately predict stock prices, neither the prediction of the stock price ratio or the subsequent application to pairs trading has seen any attention in the literature. Thus, this study investigated the temporal properties of the relationship between sentiment and stock price movement, and trialled three novel pairs trading strategies assisted by predictions made using statistical models based on sentiment data.

Eight highly correlated stocks forming four stock pairs from different industry classes were selected from the S&P 500, after which close price data and Google News headlines pertaining to the selected stocks were collected from online sources via webscraping. Linear univariate and multivariate ordinary least squares regressions were conducted to determine the explanatory power of sentiment for individual stocks and stock price ratios respectively, experimenting with both the variance and average values of both sentiment and close price (ratio) for the sampling periods. This was followed by the development and testing of three novel pairs trading strategies that sought to utilise the regression relationships. Further, an LSTM model was created which predicted the minimum and maximum values of the stock price ratio for a future period based on historical close prices and sentiment.

The study found that on average, long sentiment sampling periods (28 days) had moderate effects for both individual stocks, and stock pairs, however, for 75% of the individual stocks the strongest relationships were found to occur when the total sentiment sampling, lag, and close sampling time was less than or equal to four days, whereas for 75% of the stock pairs, the strongest relationships (albeit 43% weaker) were found for both sentiment and lag times of 28 days combined with a close price ratio sampling time of 1 hour. Despite the statistically significant results of the regression analyses, none of the novel pairs trading strategies outperformed a standard Bollinger Bands based approach over the testing period. Sentiment was shown to increase the predictive accuracy of the LSTM model's predictions for the minimum and maximum stock price ratio.

Keywords: stock trading, sentiment analysis, pairs trading, LSTM, artificial intelligence.

Acknowledgements

I would like to extend my gratitude firstly to my primary supervisor Dr. M. R. Machado, for investing his time to read and provide helpful feedback on my drafts, for taking the time to meet with me and give me advice, for his flexibility regarding my time scheduling whilst writing my other thesis, and not least, for agreeing to take over the supervision of this thesis part way through. I also highly appreciate the time invested by my additional supervisors Dr F.S. Bernard and Dr. I. Skute to read, discuss, and give feedback on these works.

Further, I would not have enjoyed my stay in the Netherlands nearly as much if it was not for the fantastic, caring, and wild friends that I made whilst living at Campuslaan 37. Thank you for all the hilarity, gezelliheid, and support over the last 4.5 years.

Thank you to the Van den Heuvel family for being my Dutch stand-in extended family, and for always being there for me.

Thank you to the Wellers family, for caring for me, and for doing so much for me over the last 10 years.

To Luisa, my warmest gratitude for your patience and support, for believing in me, and for all that you do that makes my life better.

To my friends and family back in New Zealand, thank you for understanding my decision to study so far away from you all, and for making me feel as if I never left whenever I come home.

Finally, thank you Mum, Dad, Sophie, and James for laying the foundations for this all to be possible, and for your ongoing support and love.

Kaikoura 22-01-2025

Finn Lee

1 Contents

Appendices

List of figures

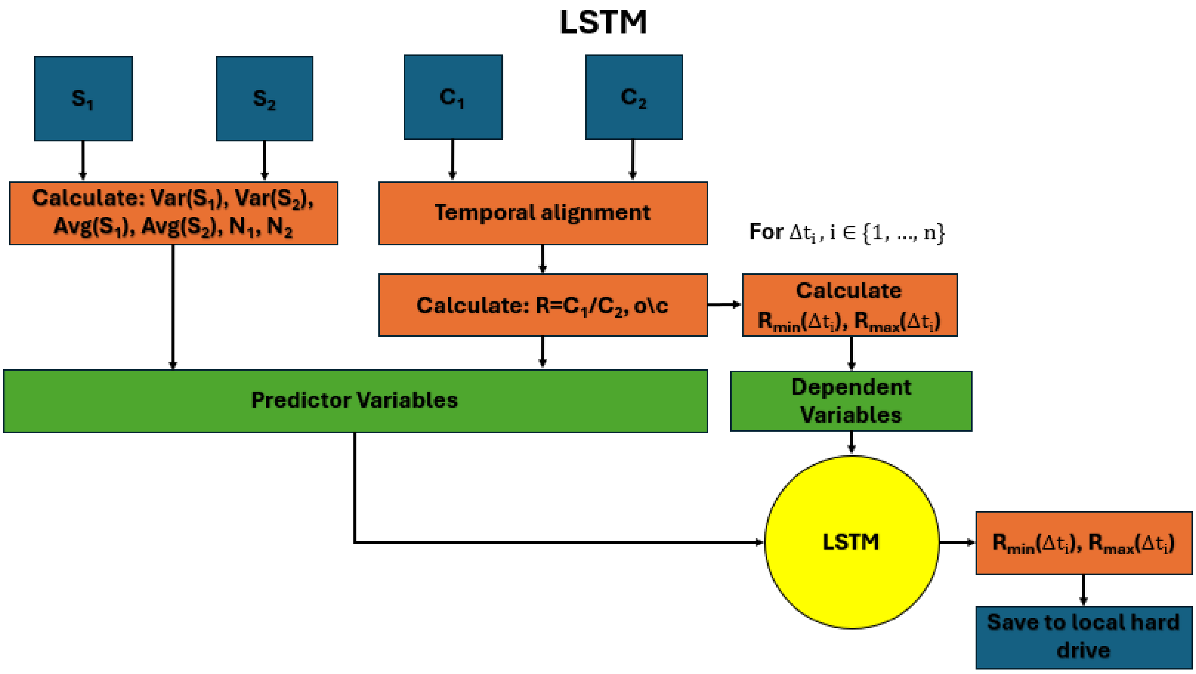

Figure *7: LSTM model implementation architecture*

Error! Reference source not found.

List of tables

List of abbreviations

AA	Additive Attention
AI	Artificial Intelligence
AMAPE	Adjusted Mean Average Percentage Error
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average (ARIMA)
AR-MNN	Auto-Regressive Moving Reference Neural Network
ATR	Average True Range
Avg	Average
A-LSTM	Attention Based LSTM
CAGR	Compound Annual Growth Rate
CCI	Consumer Confidence Index
C_i	Close Price of Stock *i*
CNN	Convolutional Neural Network
CSV	Comma Separated Values
DDPG	Aware Deep Deterministic Policy Gradient
DL	Deep Learning
DMLP	Deep Multilayer Perceptron
DRL	Deep Reinforcement Learning
GB	Gigabyte
GARCH	Generalised Autoregressive Conditional Heteroskedasticity
GIC	Global Industry Classification
Gn_0_Sent	Google News Sentiment Scores Including Neutral Sentiment
Gn_Sent	Google News Sentiment Scores Without Neutral Sentiment
GRU	Gated Recurrent Unit
HHISSO	Harris Hawks Induced Sparrow Search Optimisation
IPO	Initial Public Offering
k-NN	K-Nearest Neighbours
LDA	Latent Dirichlet Allocation
LSTM	Long Short-Term Memory

MAE	Mean Absolute Error
MAPE	Mean-Average Percentage Error
ML	Machine Learning
MLP	Multilayer Perceptron
MV	Mean Variance
NaN	Not a Number
NISI	New Investor Sentiment Index
NLP	Natural Language Processing
NN	Nearest Neighbour
N_{trades}	Number of Trades
OLS	Ordinary Least Squares
PE	Price Earnings
ReLU	Rectified Linear Unit
RF	Random Forest
RL	Reinforcement Learning
RNN	Regressive Neural Network
RNN	Regressive Neural Network
ROE	Return on Equity
R^2	R Squared
R	Close Price Ratio
SA	Self Attention
SMC	Simple Moving Covariance
SMA	Simple Moving Average
SMV	Simple Moving Variance
SPLS	Scaled Partial Least Squares
SPCA	Scaled Principal Component Analysis
sntmt	Sentiment
SVM	Support Vector Machine
SVR	Support Vector Regression
S_i	Sentiment X

T2V_TF Time2Vec and Transformer Technologies

Var Variance

VADER Valence Aware Dictionary and Sentiment Reasoner

Δt_i Sampling Period

1. Introduction

If it was possible to predict the future, the stock market would be a very different place. If investors could know with certainty which stocks would increase in value, much of the risk associated with trading would be obsolete. Trading stocks would be a much safer way to make money, comparable to buying treasury bonds which have an almost certain payout. However, because there is a strong relationship between risk and return in most investments, if stocks were completely predictable the potential returns would likely also shrink. Again, treasury bonds offer modest returns and are generally considered a low risk but not particularly high-return investments. The highest potential for profit for the individual trader is when they alone possess an accurate strategy for predicting the stock market, giving them an edge on all other traders who still perceive the stock market to be uncertain.

Although many different methods have been trialled, only variable levels of success have been reported and the movement of the stock market retains non-negligible levels of uncertainty. The difficulty of predicting stock price movements arises from the immense number of factors which influence the stock market. Stock price prediction methods can be divided into two aggregate categories: fundamental analysis, and technical analysis. Fundamental analysis assumes that the real value of a stock may differ to its current price and can be calculated by analysing all of the fundamental variables which effect the price, such as company growth, revenue, assets, etc. Conversely, technical analysis assumes that the stock price already accounts for all of the fundamental variables, and its movement can be predicted using analytical tools such as statistics and other pattern identification approaches (Thompson, Anderson, et al., 2024). A major shortcoming of both methods is that often all fundamental factors are not accurately represented in the stock price. This leads to inaccurate valuations (Zhou et al., 2023), often as a result of over-confident or fearful investors – which is where sentiment analysis, which is a means of quantifying the emotional content of text, can sometimes provide the missing explanatory factors.

One approach that has seen increasing attention in the past years is the analysis of market sentiment, and the effect that it has on the movement of stock prices. In the past, this could have been gathered via public opinion surveys to assess the sentiment concerning economic or industrial conditions, interviews, and careful reading and analysis of newspaper articles. In addition to this, stockbrokers paying attention to the “word on the street” (rumours) or the mood of others around them has also always had a tangible effect on the stock market (Smith & Rhinehart, 2023). The advent of computers and the world wide web has drastically changed this. They have provided the ability to share sentiment almost instantly via news or social media channels, and correspondingly, for computers around the globe to interpret and act based on this sentiment in addition to other factors (Bharathi & Geetha, 2017; Dahal et al., 2023; Liapis et al., 2023).

A 2019 report from JPMorgan Chase (Saikat, 2019) found that between 70 and 80 percent of all trades on the U.S. stock market were initiated automatically by computers. Five years on, it can be assumed that this share has only increased, especially with the recent advances in Artificial Intelligence (AI) technology. Today, trading algorithms use vast amounts of data to predict stock price movements and execute trades and may be based on anything from simple statistical models (Deveikyte et al., 2022; Lin et al., 2021) to advanced artificial neural networks (Agarwal & Muppalaneni, 2022; Ma et al., 2021; Rather, 2012). Furthermore, computational interpretation of sentiment, known as sentiment analysis, and its application for stock price prediction has seen a large amount of attention, and has been proven to be a relevant variable for stock traders (Ayyappa & Siva Kumar, 2022; Li, 2022; Owen & Oktariani, 2020; Shastri et al., 2018; Tirea & Negru, 2013). Methods of applying sentiment analysis to predict stock prices using machine learning and neural networks have been successful, yet there are still applications that have not yet been tested (Du, 2022; Jain et al., 2022; Ma et al., 2021; Theodorou et al., 2021; Yao et al., 2022).

At time of writing, the application of sentiment analysis to a specific trading strategy by the name of pairs trading had not received any academic attention. To the best of the author's knowledge there were no publications that directly addressed the applicability of sentiment analysis for pairs trading. Pairs trading is a trading strategy which requires the trader to open a position on two stocks simultaneously, namely a long position, wherein the trader buys the stock and hopes that its price will increase, and a short position, where the trader borrows the stock and sells it with the intention of buying it back later because they anticipate that its price will decrease (James & Scott, 2021). The most common pairs trading strategy uses so-called Bollinger Bands, which are momentum indicators which represent the recent variability of a stock price or ratio. The inner bands are calculated by adding or subtracting one standard deviation from the price's simple moving average, and the outer bands are calculated by adding or subtracting two standard deviations from the simple moving average (Thompson, Potters, et al., 2024). Pairs trading is a market neutral approach, i.e. it is a strategy which can be profitable in both bullish (rising) and bearish (falling) markets, because it relies on the relative movement of the two stocks. To implement a pairs trading strategy, two stocks whose past movements show high correlation must be selected. Often these stocks will be from similar companies from the same industry who are affected by the same factors. The principle of pairs trading is that for two highly correlated stocks, the ratio of their stock prices should remain approximately constant, and if a deviation occurs it is expected to exist only temporarily before the price ratio reverts back to its mean value (Chen & Scott, 2023). This is known as "reversion to mean" theory, and pairs traders use it to their advantage.

This study investigates the influence of sentiment derived from published news articles on the stock price ratio of two stocks whose movement shows high historical correlation for pairs trading. In particular, the study focusses on the relationships between sentiment and the value of the stock price ratio, in addition to its variance, because it is hypothesized that a pairs trading strategy should be more profitable when the stock price ratio is more variable (volatile). Subsequently, predictions will be generated using both statistical and AI-based methods, and novel pairs trading strategies that harness the power of sentiment analysis will be developed. This yields the following two-part research question:

1.1 Research Questions

Firstly, research question 1:

How does sentiment data derived from news headlines affect the prices of two stocks selected for pairs trading?

Which can be refined to research question 1.1:

How does the moving average and variance of news sentiment affect the moving average and variance of the stock price ratio of a selected stock pair?

And secondly, research question 2:

Can sentiment-based predictions be used to improve the returns of a pairs trading strategy?

Research question 2 can be divided into two sub-questions.

Research question 2.1:

How *can the inclusion of sentiment data improve the returns of a statistically driven pairs trading strategy?*

Research question 2.2:

How can the accuracy of an AI model for predicting the stock price ratio of a selected pair of stocks be improved when sentiment data is included?

This study will seek to answer these questions in detail, to determine whether news sentiment can be used to improve the accuracy of stock price prediction in ways that can be used to improve a pairs trading strategy. Determining approaches to utilising news sentiment for the prediction of stocks used in pairs trading could result in lower risks and higher profits for traders using the pairs trading strategy and is therefore worthy of investigation. The specific hypotheses that will be tested and the corresponding approaches are presented in the methodology section.

This remainder of this thesis is outlined as follows. The following section is a review of the literature exploring existing applications of AI for the prediction of stock prices and their subsequent applications in finance. Portfolio management is studied because similarly to pairs trading, it is dependent on the movement of multiple stocks, and because of the lack of literature specific to pairs trading. Particular attention is paid to the long-short term memory (LSTM) network, which emerged as the most successful type of AI model for stock price prediction. Additionally, the integration of sentiment analysis into stock price prediction and portfolio management was studied to ascertain the strategies which have been successful and identify research gaps. The literature review is concluded by a survey of the literature which investigated volatility and pairs trading. In the proceeding section, the method for the experiments is outlined, including the collection of data and its analysis, the statistical tests, the structure of the LSTM model, and the assessment criteria. The thesis will be concluded with a discussion section, recommendations for future research, and the conclusions to the respective hypotheses.

2 Literature Study

The world's first stock market opened at the beginning of the 17th century in Amsterdam with the Dutch East India Company being the first company to be publicly traded on a stock market (Hwang, 2024). For several years the East India Company remained the sole company listed on the stock exchange, with other companies gradually joining the exchange, thus progressively developing the market until it became the present-day Euronext exchange. Immediately after the first East India Company Stocks were listed, investors sought to profit from buying stocks in the hope that their value would increase or that dividends would be paid (Hwang, 2024). Correspondingly, the practice of attempting to predict the movement of stocks and other financial derivatives is a practice that has busied investors for the last 400 years.

In the present day, investors can access vast amounts of information in milliseconds. Investors employ a wide range of different approaches for predicting stock movements; however, the application of AI is a novel method with ample potential for innovation. Therefore, this literature study focussed on the application of AI to four areas of interest: portfolio management, sentiment analysis for trading applications, sentiment analysis and volatility prediction, and the effects of volatility on pairs trading strategies. These strategies for forecasting or exploiting predictive modelling in the financial market all belong to the category of technical analysis. Although statistical methods and the mathematical calculation of technical indicators have dominated the field of predicting financial markets for a long time, emerging artificial intelligence (AI) technologies have begun to demonstrate success in a variety of applications including finance. These AI technologies may employ (but are not limited to) models based on machine learning (ML) and its various sub-categories such as Deep Learning (DL) (IBM, n.d.) possibly combined with other novel concepts such as fuzzy logic (Mathworks, 2024). ML methods have aided traders in extracting information from both structured and unstructured data sets by performing high-dimensional transformations, allowing for the extraction of highly non-linear, non-stationary patterns. This often delivers much better out-of-sample predictive performance compared to established prediction methods such as ordinary least squares (OLS), autoregressive conditional heteroskedasticity (ARCH), autoregressive integrated moving average (ARIMA), and other comparable methods (Du, 2022; Yao et al., 2022). Hence, this literature study focusses on the state-of-the-art application of AI in finance. The papers reviewed as part of this study focus on AI assisted stock portfolio optimisation, sentiment analysis, the connection between sentiment and stock volatility, and the connection between stock volatility and pairs trading. The surveyed literature from the respective focus groups is summarised in tables 1, 2, 3, and 4.

2.1 AI and Portfolio Optimisation

Portfolio optimisation is the process of selecting, purchasing, and then managing a collection of stocks or other assets such as to generate a desired return at an accepted level of risk (Hayes et al., 2024). Portfolio optimisation is similar to pairs trading due to its capitalisation on the interdependence of stocks. Additionally, the literature on portfolio optimisation is far more extensive than that of pairs trading, especially in connection with AI technology, hence the inclusion of this literature in the study. The mean variance (MV) approach is the most basic and well-known portfolio management strategy (Chen et al., 2021), however, the implementation of AI in portfolio optimisation in the last decade has often provided superior returns as has been shown by the studies summarised in Table 1.

The findings of these papers demonstrate that AI models, in particular Random Forest (Du, 2022; Ma et al., 2021), LSTM (Cao et al., 2020; Du, 2022; Ma et al., 2021; Sen et al., 2021; Singh et al., 2021; Touzani et al., 2019; Yao et al., 2022), and reinforcement learning have shown strong predictive power for applications where the movement of multiple stocks need to be considered. Furthermore, AI methods like Attention-based LSTM and Deep Reinforcement Learning have demonstrated superior performance in portfolio optimization (Luthfianti et al., 2022; Shen et al., 2021), achieving high Sharpe ratios and outperforming traditional approaches such as the MV approach.

Table 1: AI & Portfolio Optimisation

Author(s)	**Purpose**	**Models**	**Summary of main points**
(Ma et al., 2021)	Improve portfolio return prediction using ML.	RF[1], SVR, LSTM, DMLP.	Random forest (RF), support vector regression (SVR), long short-term memory (LSTM), and deep multilayer perceptron (DMLP) models were trialled for predicting the next day returns and then selecting stocks for mean variance (MV) stock portfolios. The RF model produced the greatest returns, followed by SVR, and then LSTM.
(Rather, 2012)	Optimisation of a stock portfolio using NN.	AR-MNN	An auto-regressive moving reference neural network (AR-MRNN) was employed to optimise a MV stock portfolio. Instead of using historical mean returns and variances, the returns predicted by the AR-MRNN and the prediction errors were used respectively. This technique had a relatively high mean absolute percentage error compared to other newer models.
(Agarwal & Muppalaneni, 2022)	Optimisation of a stock portfolio using a time series forecasting algorithm.	Facebook-Profit.	Well performing stocks were first selected using technical indicators such as price/earnings ratio (PE), return on equity (ROE), alpha, beta, and compound annual growth rate (CAGR). The Facebook profit algorithm was then implemented to predict the returns of the stocks, which were subsequently used to optimise the stock portfolio via a MV type strategy. The strategy generated an annual return of 46.505%.

(Du, 2022)	Used ML models to predict stock prices for pairs trading in addition to MV portfolio optimisation.	SVM, RF, A-LSTM[1].	Compared support vector machine (SVM), random forest (RF), and attention-based long short-term memory (A-LSTM) for predicting stock returns for optimising stationary pairs trading, and MV type portfolios. The models included a large number of different variables, including historical data and technical indicators. A-LSTM provided the most accurate predictions.
(Sen et al., 2021)	Investigated the predictive power of an DL model for portfolio returns.	LSTM.	An LSTM model was employed to predict next day returns for MV type portfolios. The model was trained on daily stock close prices and showed high accuracy of prediction (9.51% predicted return versus 9.30% actual return) over the five-month trial period.
(Cao et al., 2020)	Compared different DL models for creating optimal stock portfolios.	ResNet, LSTM, GRU, SA, AA, and combinations: SA+LSTM, SA+GRU, AA+LSTM, and AA+GRU[1].	Residual Networks (ResNet), LSTM, Gated Recurrent Unit (GRU), Self Attention (SA), Additive Attention (AA), and various combinations of models (SA+LSTM, SA+GRU, AA+LSTM, and AA+GRU) were trialled for creating maximised Sharpe ratio portfolios. Training data included market volume, stock prices, and returns. The two most promising models (AA + GRU and SA + LSTM) achieved respective Sharpe ratios of 1.1056 and 1.0206.
(Luthfianti et al., 2022)	Trialled a DL technique to create an optimised stock portfolio.	DRL.	Portfolios optimised with respect to returns and Sharpe ratios were created using deep reinforcement learning (DRL), MV, and equal weights. The DRL method performed best for 3-, 5-, and 7-stock portfolios, while MV performed best for a 42-stock portfolio.
(Shen et al., 2021)	Investigated a DL technique to recommend stocks for portfolios.	RL.	The added benefit of using reinforcement learning (RL) to recommend stocks for portfolios was demonstrated. Relational graphs were created between stocks and different industries, which provided the basis for stock selection for mean and minimum variance strategies. Both portfolios outperformed the S&P 500 over a 12-week period.
(Yao et al., 2022)	Used DL to predict stock prices and subsequently rebalance a portfolio.	LSTM.	An LSTM model was employed with Fama-French's asset pricing model to predict next-day stock prices. The study showed that the Fama-French model is also applicable to the Chinese stock market and trialled three different portfolio management strategies based on LSTM predicted next day prices, and different confidence levels. The strategy that relied on a

			confidence level of 65% outperformed 60% and 70% confidence level portfolios.
(Touzani et al., 2019)	Prediction of stock price uptrends using DL.	NN-LSTM.	A combine nearest-neighbour (NN) and LSTM model was used to predict uptrends for stock prices. The uptrend prediction accuracy was 71.3% over a 20-day period.
(Singh et al., 2021)	Investigation of price prediction and portfolio allocation techniques.	LSTM, K-means clustering.	Statistical and ML based models were applied for stock price prediction and portfolio allocation. Beta values were calculated via OLS regression to aid in portfolio allocation, LSTM was used for predicting next-day stock prices, and K-means clustering was mentioned as a means of grouping similar stocks.

2.2 Sentiment Analysis

Sentiment describes a person or group of people's view or opinion towards a certain thing and can be described by a continuous qualitative variable ranging from very negative, to neutral, to very positive (*Sentiment*, n.d.). With the rapid progress made in AI technologies over the last decade and an explosion in digitally data available for training, automated sentiment analysis methods have also greatly improved. Similarly to the portfolio optimisation strategies presented in Table 1, sentiment analysis tools are also often based on AI technologies such as machine learning.

Sentiment scores are a promising valuation metric for stocks because investor sentiment has been shown to have a large impact on the short-term movement of stock prices, meaning that for the prediction of short term stock price movements, the historical stock price movements are less relevant (Ma et al., 2021). Typically, if investor sentiments towards a company and its future are positive, then investors will be more inclined to buy shares, and vice versa. In many cases, it is possible to extract investor sentiment from online sources such as social media platforms and news websites. A well-known example of the effects of online sentiment on stock prices is the relationship between Elon Musk's X (formerly Twitter) posts and the share price of his company Tesla INC (TSLA). One study confirmed that if Musk tweeted positive information, Tesla stock prices tended to increase, and if negative or irrelevant information was tweeted both returns, and volatility tended to decrease. Strong sentiment was also linked to increases in trading volume. However, it was also found that the more frequently Musk tweeted over the course of a day, the less the stock price moved with each tweet (Dam, 2023).

Table 2 provides summaries of recent studies that utilised sentiment analysis for stock market prediction. The key takeaways are that the addition of sentiment data has consistently been shown to improve the accuracy of stock price predictions using ML or DL based models, with the LSTM model consistently providing the most accurate predictions (Ayyappa & Siva Kumar, 2022; Dutta et al., 2021; Gehlor & Singh, 2022; Li, 2022; Muthivhi & van Zyl, 2022). Hybrid approaches, such as LSTM with CNN also showed promise (Owen & Oktariani, 2020), and the VADER sentiment analysis tool has proven to be an effective way of quantifying sentiment (Dutta et al., 2021; Gehlor & Singh, 2022; Koratamaddi et al., 2021; Muthivhi & van Zyl, 2022). A recurrent challenge across the literature is the inconsistency of the effects of sentiment, and the effect of the lag time between sentiment publishing and stock price movement.

Table 2: Sentiment Analysis & Stock Market Prediction

Author(s)	**Purpose**	**Models**	**Summary of main points**
(Li, 2022)	Summary of big data applications for finance.	ARIMA, Facebook-Prophet, LSTM[1].	The stock market is a complex environment influenced by factors including governmental actions, online information, and human behaviour. Sentiment analysis can aid in identifying trends, but online sentiment can be unreliable. It is recommendable to employ data reduction techniques. The study compared ARIMA, Facebook Prophet, and LSTM for predicting next day stock prices. LSTM offered significantly better performance.
(Ayyappa & Siva Kumar, 2022)	Predict stock prices via ML using both technical indicators and news sentiment.	LSTM-HHISSO[1], LSTM, RF, SVM, RNN.	Used a combination of stock data (technical indicators such as exponential rolling average, average true range (ATR) and true range) with sentiment scores extracted from news data to predict stock prices using a LSTM network combined with a Harris Hawks Induced Sparrow Search Optimisation (HHISSO) algorithm, in addition to RF, SVM, and regressive neural network (RNN) models. News features included term frequency-inverse document frequency, word occurrence count, n-Gram, and improved cosine similarity.
(Zhou et al., 2023)	Created a novel DL model to predict stock prices.	T2V_TF[1], MLP, SVM, GBDT, LSTM, A-LSTM, Transformer.	The study proposed a novel deep learning model based on Time2Vec and Transformer technologies (T2V_TF) to predict stock prices to build inter-day portfolios. The model used trading data, time frequency features, Alpha 101 and 191 indicators, and sentiment scores as inputs and reportedly outperformed all other models compared in the study. News sentiment data resulted in the greatest increase in prediction accuracy when paired with historical data only.

(Theodorou et al., 2021)	Created a stock management platform utilising ML and SA.	Not specified.	An online stock management platform based on machine learning and sentiment analysis was created. The ML model took 60 financial indices derived from financial and news data as input to predict stock returns and then make recommendations.
(C. Zhang et al., 2022)	Review of decision fusion literature.	Not applicable.	Two main categories identified: classification and regression. Fusion of different information sources invariably improves results. For classification, voting and tree-based information fusion methods are most common, and for regression, averaging. The study noted the potential of sentiment as an information source.
(Owen & Oktariani, 2020)	ML based model using historical data and sentiment for stock price prediction.	Ensemble of: LSTM, MLP, & CNN.	An ensemble model which combined LSTM, MLP, and a convolutional neural network (CNN) predicted next day stock prices using sentiment data extracted from the Stocktwits platform, and financial data from Yahoo! Finance. The model also proposed an adjusted mean average percentage error (AMAPE), which improved the training of the model by doubling the error if the predicted movement was in the wrong direction. The study found that including sentiment scores decreased the AMAPE from 1.188% to 0.89%.
(Shastri et al., 2018)		MLP.	A MLP neural network was trained on sentiment (classified using a Naïve Bayes tool) and historical data such as opening price, highest daily value, lowest daily value, and daily share volume. The model predicted whether the current stock price trend was bearish or bullish. The study found that the model performed better when trained on one year of data as opposed to three (MAPE 1.5830% versus 8.2148%. The Naïve Bayes classifier was seldom mentioned as a sentiment analysis tool in more recent literature.
(Tirea & Negru, 2013)	Investigation of a portfolio optimisation system based on news sentiment, historical prices, and other data.	Not specified.	A stock portfolio optimisation model based on historical prices, the effect of news articles, trader behaviour, confidence levels, and risk evaluation. Stock price, volume, and number of transactions in addition to financial information about the relevant companies extracted from shareholder reports was also considered. The model predicted next day stock prices and trend directions.
(Koratamaddi et al., 2021)	Create a virtual DL-based stock	DDPG.	An automatic DL-based stock trader with historical price data and market sentiment inputs was created. The adaptive sentiment

	trader with market sentiment and historical price data as input.		aware deep deterministic policy gradient (DDPG) algorithm was used to optimise portfolios, and it was found that this method provided better returns (22.05% versus 15.86%) MV strategy, although the inclusion of sentiment data did increase portfolio volatility (Sharpe value 2.07 versus 1.25). Sentiment was scraped from Twitter and Google News, and the study found that news headlines contained sufficient information for accurate sentiment data generation. Sentiment scores were calculated using VADER.
(Gehlor & Singh, 2022)	DL-based stock price and trend prediction using sentiment and historical stock price data.	XGBoost[1], logistic regression, k-NN, decision tree, Gaussian Naïve Bayes, SVM, RF. LSTM[1], Auto ARIMA, Facebook-Prophet.	Stock market data and sentiment scores derived via VADER were used for trend prediction (94.5% accurately), in which the XGBoost classifier performed best, and stock price prediction, wherein the LSTM model performed best (MAPE 4.164%). A sentiment subjectivity metric was shown not to effect prediction accuracy; however, its inclusion did reduce training time.
(Dutta et al., 2021)	ML stock price prediction using closing prices and sentiment.	LRGS, LRGS +VADER, MAVG, MAVG +VADER, KNBR, KNBR +VADER, ARM, ARM +VADER, LSTM, LSTM+NBY, LSTM +SVM, LSTM +VADER[1].	The VADER Lexicon was used to assign sentiment scores to online news articles which were then combined with closing prices from the previous seven days in a LSTM model to predict the next day closing prices. The study compared 11 different models and found that the inclusion of sentiment increased prediction accuracy for all models. The LSTM-VADER model achieved a prediction accuracy of 77.496%.
(Muthivhi & van Zyl, 2022)	Created a sentiment aware DL model to predict stock	LSTM, LSTM +VADER[1].	Different models were compared for augmenting a MV type portfolio strategy. The sentiment aware LSTM stock prediction model produced the highest returns in both bullish and bearish markets. Additionally, this study

	prices for portfolio optimisation.		observed significant differences in sentiment classification accuracy. An accuracy of 43% was observed for sentiment about Microsoft, versus 60% for Disney. This relatively low sentiment accuracy is also indicative of the care that should be taken when implementing sentiment in prediction models. The study also found that VADER is most likely to falsely classify data as neutral and is slightly more likely to falsely classify negative sentiment. Furthermore, the study found that the inclusion of neutral sentiment did not affect prediction accuracy, and that there is an observable lag between sentiment publishing time and stock price movement.

2.3 Sentiment and Volatility

A large amount of research has been undertaken to determine the influence of sentiment on predicting future stock prices and returns; a search of the Scopus data base using the search string "stock AND prediction AND sentiment" returned 1378 results, however, less research has been conducted specifically regarding the influence of sentiment on stock price volatility; a search of the Scopus database for "volatility AND prediction AND sentiment" returned 264 results. The following section will explore the existing knowledge concerning sentiment analysis and its application for both stock price and volatility prediction.

Intuitively, a connection between the mood of company stakeholders and the volatility of the stock market seems logical. Indeed, the surveyed literature in Table 3 shows that a relationship exists between sentiment metrics and stock market volatility metrics. The studies have shown that both the strength, polarity, and volume of the sentiment all influence the stock price volatility, however, care must be taken when applying these findings because the correlations vary between asset classes, sectors, and economies (Alomari et al., 2021; Muguto et al., 2022), because of the lag time between the publishing of the sentiment and its effect (Deveikyte et al., 2022). Furthermore, the application of LSTM models were also shown to improve the prediction of volatility (Jain et al., 2022; W. Zhang et al., 2021).

Table 3: Sentiment and Stock Market Volatility

Author(s)	**Purpose**	**Models**	**Summary of main points**
(Song et al., 2023)	Predict stock price volatility as a function of sentiment	SPCA, SPLS.	Six known sentiment proxies (closed-end fund discount rate, first-day returns of IPOs, number of IPOs, share turnover, number of newly opened individual investor accounts, and the consumer confidence index) were used to

	using statistical methods.		predict stock price volatility using scaled principal component analysis (SPCA), and scaled partial least squares (SPLS). Both techniques offered high predictive power during non-crisis periods, while SPCA was slightly better during the Corona pandemic. The predictors were found to be effective across all industrial stock categories, however, SPCA was found to work especially well for financial stocks.
(Deveikyte et al., 2022)	Prediction of volatility based on sentiment.	Pearson's r test, Granger's causality test, latent Dirichlet allocation (LDA).	Predicted next day volatility trend direction with 63% accuracy based on sentiment. The study investigated the relationship between a single day of positive, neutral, negative, and average sentiment scores from either financial news or Twitter with either zero lag, or a single day's lag. News derived sentiment was the best predictor for same-day returns, with the strongest correlation (r=-0.45) existing for negative sentiment. Neutral sentiment had r=0.291, average sentiment had r=0.367, and a statistically significant relationship between positive sentiment and same day returns was not observed. News based sentiment was a relatively weak indicator for same day and next day volatility (r ≈ 0.25). In contrast, Twitter derived sentiment did not have a statistically significant relationship with stock returns, however, positive, neutral, and average Twitter-based sentiment showed strong correlations with same day volatility (r=-0.698, r=0.754, and r=-0.487 respectively) and next day volatility (r=-0.70, r=0.746, and r=-0.492 respectively).
(Groß-Klußmann & Hautsch, 2011)	Investigated the influence of news on stock market volatility.	Not applicable.	Predominantly highly relevant news stories induced increases in return volatilities, with negative news having the strongest influence.
(Caporin & Poli, 2017)	Compared the influence of news versus social media sentiment on stock volatility.	Not applicable.	Specific news topics such as earning announcements, and up/downgrades had the greatest influence over return volatilities while high social media activity surrounding a certain company predicts a significant increase in volatility, whilst attention from reputable press outlets (such as the Wallstreet Journal) has the opposite effect.
(Frino et al., 2022)	Examined the causality	Not specified.	The study found that tweets had a stronger influence on stock realised volatility as

	between tweets and stock realised/implied volatility.		opposed to option implied volatility. Additionally, the correlation between tweet volume and stock realised volatility was positive, but negative for average sentiment.
(Gong et al., 2022)	Development of a new investor sentiment index.	PLS.	The new investor sentiment index (NISI) developed in this study via PLS aggregation outperformed existing investor sentiment proxies. The NISI was based on the following sentiment proxies: closed end fund discount, stock market turnover, number of new investor accounts, consumer confidence index (CCI), Number of IPOs, average first day returns of IPOs, Advance-Decline Line, the ratio of Advance-Decline.
(Muguto et al., 2022)	Investigation of the effect of sentiment on returns and volatility.	Not specified.	Prevailing sentiment was shown to have a negative relationship with market returns, yet a positive return with market volatilities. The study also found that less experienced traders were more susceptible to sentiment driven biases. Sentiment was found to have a mean reverting effect, but the overall influence of sentiment was different per sector.
(Niu et al., 2022)	Observed the influence of sentiment on gold market forecasting.	Not specified.	Sentiment was found to improve gold market forecast accuracy, particularly in the short term, whereas stock market sentiment improved longer term forecast accuracy. The study found that portfolio performance could be improved by including sentiment forecasted volatility.
(Jain et al., 2022)	Prediction of stock price and volatility from sentiment via DL.	Various RNNs and CNNs including LSTM[1].	News sentiment was used as a predictor for stock price and volatility using several different regressive and convolutional neural networks (RNNs and CNNs respectfully). LSTM performed best for stock price prediction, while all models performed equally for volatility prediction, providing marginal prediction accuracy.
(W. Zhang et al., 2021)	Used sentiment to predict stock volatility with DL and then utilised this information for portfolio allocation.	LSTM.	Four sentiment indices were created and then used to analyse investor sentiment in text. Sentiment was found to have a significant negative correlation with short term market volatility. In developing markets, the relationship between sentiment and market volatility was found to be non-linear, and the inclusion of sentiment from published financial articles in an LSTM model was found to increase volatility prediction accuracy. Including sentiment in a forward-looking portfolio allocation strategy resulted in RMSE

			decreases of 3.67%, MAE decreases of 5.88%, and out-of-sample R^2 increases of 7.19% compared to a model based on historical data alone.
(Alomari et al., 2021)	Investigation of news and social media sentiment's effect on the stock and bond market's volatilities and dynamic return correlations.	Not specified.	The effect of news and social media sentiment on both the stock and bond market was investigated. News-based sentiment was found to be a stronger predictor for volatility and returns, while social media sentiment was found to be a superior predictor for the correlation between stock and bond prices. This correlation tends to exist because when stock market volatility is high, more investors seek more secure investment opportunities such as bonds. Additionally, the study found that the effect of sentiment varies per asset class, that sentiment can prolong volatility persistence, and that high news and social media coverage tends to decrease market volatility and turnovers.
(Hsu et al., 2021)	Investigated the effect of news sentiment on stock volatility.	GARCH	The Generalised Autoregressive Conditional Heteroskedasticity (GARCH) statistical model was used to investigate the effect of news sentiment on stock market volatility. The results indicated that both current and lagged news sentiment affected market volatility.
(Jiang & Jin, 2021)	Investigation of investor sentiment on stock return volatility.	Not specified. [1]	The effects of investor sentiment on stock returns were investigated. A key finding of the study was that sentiment pertaining to one stock also affects stocks in the same geographic region or economic sector. The authors recommend considering these factors when diversifying a portfolio, however, economic distance is more significant than geographic distance.

2.4 Pairs Trading Strategies and Volatility

Pairs trading is a unique arbitrage trading strategy that proceeds by identifying two cointegrated stocks whose price movements are usually in the same directon, recognising when their prices diverge, buying the underperforming stock, (borrowing and) short-selling the overperforming stock, then doing the opposite when the prices re-converge in the hope of generating a profit (James & Scott, 2021). Because pairs trading is a strategy that relies on relative price movements between two correlated assets, increased market volatility often leads to greater price discrepancies and divergence between these assets, thus presenting

[1] Best performing model in study.

more trading opportunities for pairs traders (James & Scott, 2021). Higher volatility can create wider spreads, enhancing the potential for profit and it is therefore logical to investigate whether sentiment data can be used to predict volatility, and if these predictions can be used to augment an enhanced pairs trading strategy. The following papers summarised in Table 4 are a collection of recent papers which discussed the use of volatility in pairs trading strategies. Volatility was primarilly used to manage the risk levels (Göncü & Akyıldırım, 2016; Ramos-Requena et al., 2021), and to predict the returns of stock pairs (Lin et al., 2021).

Table 4: Volatility and Pairs Trading

Author(s)	**Purpose**	**Models**	**Summary of main points**
(Ramos-Requena et al., 2021)	Developed a pairs trading strategy using stocks with the lowest volatility.	Hurst exponent.	The study presented a pairs trading strategy which reduced the stock universe to the least volatile stocks and then selected stock pairs based on the Hurst exponent (a divergence indicator). The strategy was not completely market neutral and performed best during bullish periods with low market volatility.
(Lin et al., 2021)	Predicted risk adjusted returns for pairs trading strategies using volatility data.	Not specified.	The study outlined five main categories of pairs trading strategies: distance methods, cointegration methods, time series methods, stochastic control methods, and other methods. The study created a statistical learning approach with a model trained on volatility data, then used it to predict risk-adjusted returns for pairs trading strategies.
(Göncü & Akyıldırım, 2016)	Identification of the optimal market conditions for a specific pairs trading strategy and the optimal stocks to select.	Not applicable.	The objective of the research was to identify the optimal conditions for implementing such a mean reversion pairs trading strategy and to calculate the pairs with the highest likelihood of statistical arbitrage within a specific time frame. The strategy incorporated a random noise variable; however, this is rather arbitrary and could be an area in which sentiment predicted volatility could generate some improvement. The introduction of noise into the reversion equation did not impact whether the spread reverted to the mean, but it did influence the time it took for mean reversion to occur. It was also noted that pairs traders can improve their market neutrality based on the betas of the selected stocks, and that pairs trading relies on high-speed trading to capitalise on short-lived statistical anomalies.

2.5 Proposed Research Direction

It is evident from the literature that AI, especially deep learning, is a powerful technology that can be applied in a broad array of trading applications. Different techniques have been demonstrated to be successful in stock forecasting applications; however, LSTM networks were repeatedly acclaimed in the literature for providing the most accurate predictions (Ayyappa & Siva Kumar, 2022; Dutta et al., 2021; Jain et al., 2022; W. Zhang et al., 2021), not least in applications involving the interrelation of multiple assets such as portfolio optimisation (Cao et al., 2020; Du, 2022; Ma et al., 2021; Sen et al., 2021; Singh et al., 2021; Touzani et al., 2019; Yao et al., 2022). Sentiment data sourced from both news articles and social media was shown by all of the reviewed studies to be a predictor for stock price movement, and improved all stock price prediction models compared to when only historical price data was used (Ayyappa & Siva Kumar, 2022; Dutta et al., 2021; Gehlor & Singh, 2022; Koratamaddi et al., 2021; Li, 2022; Muthivhi & van Zyl, 2022; Owen & Oktariani, 2020; Shastri et al., 2018; Theodorou et al., 2021; Tirea & Negru, 2013; C. Zhang et al., 2022; Zhou et al., 2023). The effect of sentiment on volatility was also evidenced in the literature (Caporin & Poli, 2017; Deveikyte et al., 2022; Frino et al., 2022; Gong et al., 2022; Groß-Klußmann & Hautsch, 2011; Jain et al., 2022; Muguto et al., 2022; Niu et al., 2022; Song et al., 2023; W. Zhang et al., 2021), however, some results were contradictory, in particular in regard to the best sentiment metric for prediction, and the sentiment source (social media or news). Furthermore, the strength of the reported correlations varied. The temporal aspect of sentiment on stock price and volatility also yielded differing results across studies, however, most studies observed a stronger effect on the near-term behaviour of the stocks. Only three papers (Göncü & Akyıldırım, 2016; Lin et al., 2021; Ramos-Requena et al., 2021) were found which attempted to augment pairs trading strategies with volatility data, and subsequently it can be concluded that this research direction is largely unexplored. None of these studies implemented ML models or sentiment data, and direct prediction of the stock price ratio also was not studied.

Therefore, this study will focus on investigating the relationship between news sentiment and stock price movement, investigating both the average and variance (volatility) of stock prices within different time windows. Because a consensus of the temporal validity of sentiment on stock price and volatility could not be found in the literature, combinations of sentiment sampling period, lag time, and time window for the stock price average and variance will be trialled. This will be investigated using both statistical methods and a LSTM network.

Finally, the strongest predictor methods will be implemented in a number of novel pairs trading strategies to determine whether news sentiment and predicted volatility can improve the returns of pairs trading.

3 Materials and Methods

Based on the research questions given in Section 1.1, the following hypotheses were formulated and subsequently tested to provide answers to the two primary research questions and their respective sub-questions.

3.1 Hypotheses:

1. Sentiment collected from Google News [2] headlines is a predictor for the price ratio of highly correlated stocks.
    a. The average sentiment score of two correlated stocks is a predictor for the average stock price ratio of the two stocks.
    b. The average sentiment score of two correlated stocks is a predictor for the variance of the stock price ratio of the two stocks.
    c. The variance of the sentiment score of two correlated stocks is a predictor for the average stock price ratio of the two stocks.
    d. The variance of the sentiment score of two correlated stocks is a predictor for the variance of the stock price ratio of the two stocks.
2. Google News derived sentiment can be implemented in combination with a linear regression to improve a modified pairs trading strategy to yield increased returns compared to a basic Bollinger Bands-based strategy.
3. Google News derived sentiment can improve the predictive accuracy of a LSTM model which predicts the stock price ratio of a pair of highly correlated stocks.

The three hypotheses were tested as follows: For hypothesis 1, the averages and variances of both the news sentiment and the closing prices for the selected stocks were calculated for a range of sample periods before conducting linear regressions to determine if there was a linear relationship between these values. Hypothesis 2 was tested by creating a modified Bollinger-Bands pairs-based trading strategy which utilised the statistically significant predictors discovered during the investigation of hypothesis 1. The profit of the modified strategy was compared to that generated by a standard Bollinger-Bands based strategy.

[2] Google News is Google LLC's news platform.

Hypothesis 3 was tested by creating and testing two LSTM models which forecasted the stock price ratio, one which utilised news sentiment, and another which utilised random noise instead of sentiment. The accuracy of the two models and thus the hypothesis was evaluated based on the forecast error.

Empirical Approach

The aim of this study was to determine whether sentiment extracted from online news headlines could be used to predict the behaviour of a pair of highly correlated stocks, and subsequently create an improved pairs trading strategy relative to a basic Bollinger bands-based approach. The improved pairs trading method was inspired by the pairs-based trading approach presented by Göncü & Akyıldırım, who used random noise to improve their model (Göncü & Akyıldırım, 2016), however, this study will augment the pairs trading strategies with sentiment data as opposed to random noise.

The structure of the research method is illustrated by the flow chart in Figure 1. First, a total of eight companies whose stock prices exhibited high historical covariance were selected. Next, the historical close price datasets from the eight stocks were downloaded in addition to the most relevant news headlines that mentioned the companies during the same period. The sentiment scores were then calculated for each of the headlines and stored on a hard drive with the stock prices. During the first analysis phase, different sampling windows were used to investigate the relationships between the sentiment about a company, and the subsequent stock price's variance within a finite period (volatility) in addition to the stock price's average value for the same period using statistical models. The motivation for the use of many different sampling windows was to reveal the temporal relationship between news sentiment and stock behaviour. During the second analysis phase, the value of adding sentiment to a deep learning based predictive model (LSTM) was investigated. Finally, the statistical results found in the first analysis phase were used to augment different pairs trading models, in an attempt to create a pairs trading model that provides superior returns to a standard Bollinger Bands based model.

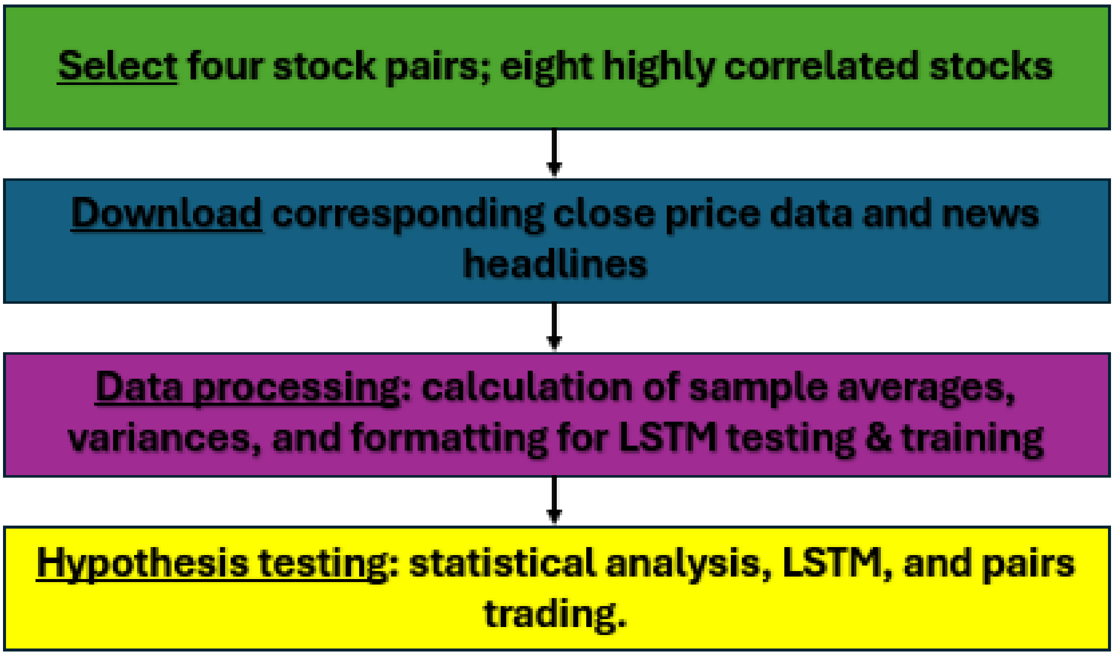

Figure 1: Methodology overview

The study was implemented in python[3] and included webscraping[4] elements which automatically extracted news headlines and stock price data from online sources, sentiment analysis to autonomously interpret the news headlines, data cleaning and handling, statistical analysis, deep learning (LSTM), and during the final stages, graphical displays and data interpretation methods.

3.2 Selected Participant Companies

A total of four stock pairs comprised of eight unique stocks were selected from Standard and Poor's (S&P) 500 for this study. Four was deemed to be a sufficient quantity of stock pairs to reduce the probability of accidentally drawing conclusions based on false correlations. Furthermore, four stock pairs were assumed to be sufficient to demonstrate the generalisability of the results, and thus it was not deemed necessary to include more stock pairs. The S&P 500 is a stock index which lists 500 of the largest companies on the American stock market, correspondingly, it was assumed that any company large enough to be listed should be an industry leading company mentioned sufficiently frequently in the news such as to generate ample sentiment data for the study, with plenty of pre-existing history available

---

[3] *Python* is an open-source, interpreted, object-oriented, high-level programming language created in the 1990s by Dutch programmer Guido Van Rossum that has since been used by over 8 million programmers for applications ranging from stock trading to space flight (Van Deusen, 2023).
[4] *Webscraping* is the process of automatically (often programmatically) extracting data from the world wide web.

for further investigation. Furthermore, most relevant news articles about these companies were published in English, rendering them compatible with the VADER[5] sentiment analysis tool. It was also assumed that companies large enough to be listed on the S&P 500 should also be stable enough to provide reliable data over the duration of this study and provide relatively generalisable results that should be applicable to the stocks of other comparable companies. Additionally, each of the four stock pairs are from different Global Industry Classifications (GICs), which further improved the generalisability of the study's results by revealing the sentiment's influence on the stock price's co-dependency on the respective industry of the stock, whilst improving generalisability within each GIC due to the selected companies being industry leaders. The final, and most critical selection criteria for the pairs trading strategy was the correlation of the stock prices. Pairs trading relies on the reversion of the pair's stock price ratio to its mean value after an observed deviation, and thus, pairs of highly correlated stocks were required. In this study, a historical 3-month Pearson correlation coefficient with magnitude greater than 0.8 at the time of selection (July 2023) was used as the selection criteria. The selected stock pairs, correlation coefficients, and GICs are given in Table 5:

Table 5: Stock pairs with their 3-month correlation coefficients[6] and GIC

Stock names, [Tickers]	**Correlation Coefficient**	**GIC**
HP Inc, Dell Technologies [HPQ,DELL]	-0.83	Information technology
United Airlines, American Airlines [UAL,AAL]	0.93	Industrials
(The) Coca-Cola Company, PepsiCo [KO,PEP]	0.95	Consumer Staples
Mastercard, Visa Inc [MA,V]	0.97	Financials

3.3 Measurement

The sentiment analysis tool VADER (Hutto & Gilbert, 2014), was selected due to its popularity, and because it had the highest number of recommendations compared to other sentiment analysis methods (Koratamaddi et al., 2021), (Gehlor & Singh, 2022), (Dutta et al., 2021), (Muthivhi & van Zyl, 2022). Furthermore, it is an open-source python module

[5] *VADER* is an acronym for Valence Aware Dictionary and Sentiment Reasoner, and is an open source tool created by Hutto and Gilbert (Hutto & Gilbert, 2014) that can be used to quantify the sentiment in a word or passage of text. https://github.com/cjhutto/vaderSentiment

[6] Pearson's correlation coefficients calculated from stock price data extracted from Yahoo! Finance prior to the collection of the experimental dataset. https://finance.yahoo.com/

rendering it a free, ethical, and up-to-date software choice. Despite being an extremely effective tool based on complex algorithms and an expansive external library, VADER is simple to implement and can generate a sentiment score for a given sentence in a single line of code. This study will use the “compound” sentiment score, which is the sum of the negative, neutral, and positive sentiment scores of the sentence normalised between negative -1 and 1, which reduces the number of variables in the investigation whilst retaining sufficient sentiment assessment information (Gehlor & Singh, 2022).

VADER’s compound score represents the emotional polarity (positive or negative) in addition to the emotional intensity; -1 being extremely negative, and 1 being extremely positive. VADER can interpret and aggregate the sentiment of individual words within a sentence, in addition to being able to interpret context (e.g. negation), interpret punctuation, and account for emphases such as words written in capital letters (Hutto & Gilbert, 2014).

In this study, a python function was created which took a pythonic dataframe object containing headlines scraped from Google News as input, and then returned the compound sentiment score for each headline, appended to the original dataframe in a new column. Six recent headlines that mentioned “Pepsi” and their respective sentiment scores are provided in Table 6. These dataframes were then saved in python’s ultralightweight feather file format, which is a binary format that saves and loads up to 150 times faster than the conventionally used comma separated value (csv) format and can be used to write and read Pandas dataframes directly to and from file. The utilisation of feather type files in this study greatly increased computational speed and thus the overall efficiency of this research.

Table 6:Example headlines and sentiment scores

Example headline mentioning “Pepsi”	**VADER Sentiment score**
“Pepsi pushes new, ‘festive’ flavour to replace pumpkin spice — here’s where to get the ‘holiday in a can’” (Steinberg, 2024)	0.6908
“Agriculture, Value Chain & Choices: PepsiCo’s pep+ Strategy” (King, 2024)	0.3400
“SL Green to buy former Pepsi headquarters building on Manhattan’s Park Avenue” (Cheng, 2024)	0.0000
“PepsiCo is closing 4 bottling plants and cutting nearly 400 jobs as it streamlines operations” (Durbin, 2024)	-0.1280

"PepsiCo beats New York state's lawsuit over plastics pollution" (Stempel, 2024)	-0.2263
"Coca-Cola and Pepsi, two of the world's top plastic polluters, just got slapped with a lawsuit alleging exaggerated recycling claims and downplayed health effects" (Ding, 2024)	-0.5423

3.4 Data: Sources and Extraction

This study utilised two sources of data: the historical stock close prices of each of the selected company's stock, and online news headlines that mentioned the company. Both data sources are derived from online sources in the public domain. Both datasets were collected via webscraping modules implemented in python, which are open-source code libraries with functions that can automatically extract user specified data from the internet. The scraped data was subsequently saved to a local hard drive. A visualisation of the data collection software architecture is presented in Figure 2.

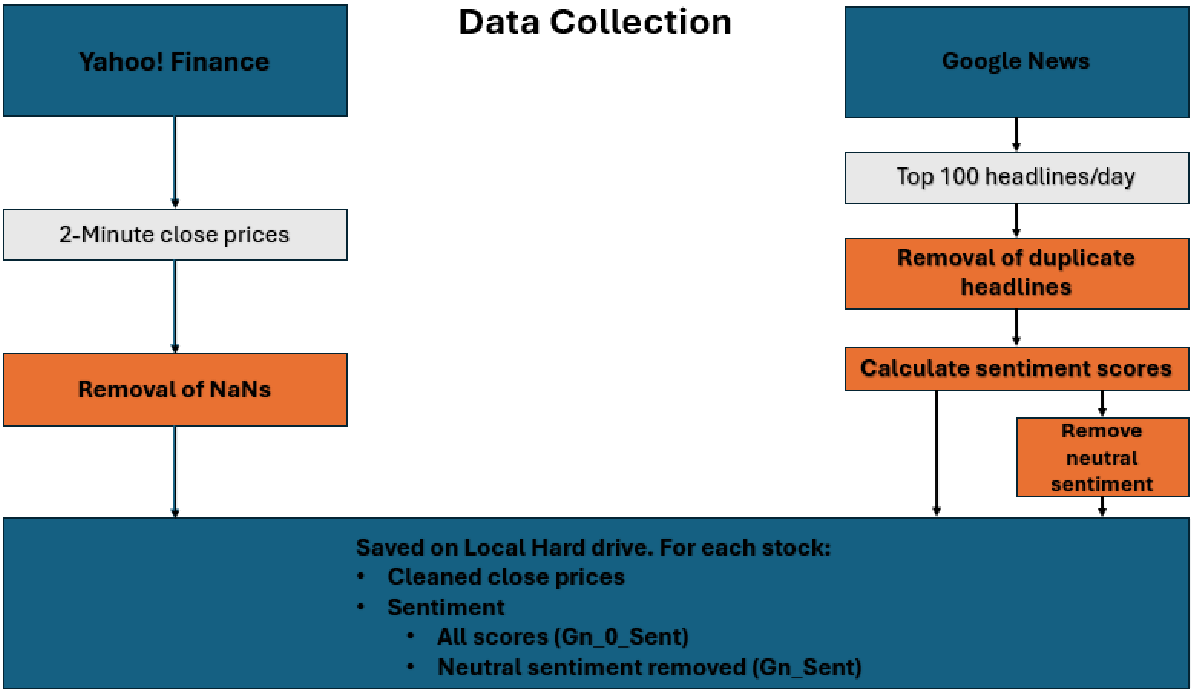

Figure 2: Data collection software structure

3.4.1 Google News

Online news headlines was automatically extracted from Google News using scripts written in python with the assistance of the open-source pygooglenews[7] module (Bugara, 2020). The pygooglenews module allows the user to specify the language of the news articles to search for, and the region. By default, these values are set to English and the United States of America, and for this research these parameters were not changed. This is because all of the companies listed on the S&P 500 are incorporated in the United States of America, and thus the most relevant news was likely published for American readers. Pygooglenews functions by taking a search query (string) comprised of a word or a series of words which should appear in the article or the article headline in addition to a publishing time, and then returns information about the news article, including the headline text and the publishing datetime. This research project used the company name as the search string and extracted the headline text and the publishing datetime.

An example search string call for the Coca-Cola company would be: "search_results = gn.search('Coca-Cola', from_='14-04-2023', to_='15-04-2023')". The pygooglenews scraper is limited to 100 articles per day, however, these articles are the highest ranked articles for each day according to Google and are thus assumed to be most likely the most influential. For this reason, when collecting data for longer periods of time the Google News scraper was implemented in a loop which then gathered the top 100 articles of each day. Additionally, duplicate headlines were removed because separate websites often post popular articles from the same source. The search strings used for the eight companies investigated in this study are provided in Table 7.

Table 7: Companies and their respective search strings

Company	Search String
HP Inc	'Hewlett Packard'
Dell Technologies	'Dell'
United Airlines	'United Airlines'
American Airlines	'American Airlines'
PepsiCo	'Pepsi'
(the) Coca-Cola Company	'Coca-Cola'
Mastercard	'Mastercard'
Visa Inc	'Visa'

[7] *pygooglenews* is an open-source library for python created by A. Bugara that provides functions that allow users to automatically extract news articles from Google News (Bugara, 2020). https://github.com/kotartemiy/pygooglenews

3.4.2 Close Prices

This study employed the open-source yfinance[8] python module (Aroussi, 2019) to scrape adjusted close price data from Yahoo! Finance[6]. A python script was written which takes the desired time span, the period between the data points (the frequency), and the stock tickers as inputs, and then returns a pythonic "dictionary" object containing all of the corresponding closing prices and the corresponding date-time values. This dictionary was subsequently transformed into a dataframe object, allowing it to be saved and loaded for further processing as a feather type file on a local hard drive. An example call in python to retrieve 60 days of 2-minute close price data with the yfinance module for the Coca-Cola Company (KO) is: stock_x = yfinance.Ticker('KO'), stock_data_frame = stock_x.history(period = '60d', interval = '2m'). In this study, closing prices at a frequency of 2 minutes were collected for a total period of 105 days. Because Yahoo! Finance does not provide data with a 2-minute frequency from more than 60 days in the past, it was necessary to scrape the data on more than one occasion and then merge the subsequent dataframes programmatically.

3.4.3 Data Cleaning

The headlines scraped from Google News were partially cleaned during the collection process, when duplicate headlines were removed. After the sentiment score of each headline was calculated, a copy of the dataset was created for which all instances with sentiment score magnitudes less than 0.1 were removed, resulting in one dataset without neutral sentiment, and one with. This was done because some studies (Muthivhi & van Zyl, 2022) found that neutral sentiment did not affect stock prices, however, other studies (Deveikyte et al., 2022) found results to the contrary.

The closing price data was first filtered for any NaN (empty) values. During stages of the data analysis where both of the stocks in a pair were simultaneously analysed, the close prices were also chronologically aligned to ensure that the dataset only consisted of close prices from times when both stocks had reported a close price.

Because news headlines are not always released at the same times, it was necessary to use sampling windows and create aggregate values for the sentiment, and corresponding close price data. A range of sampling windows (i.e. time periods) were used to partition both the sentiment and close price data, and then calculate averages, variances, and covariances of the

[8] *yfinance* is an open-source library for python created by R. Aroussi that provides functions that allow users to automatically download (webscrape) stock price data from Yahoo! Finance.
https://github.com/ranaroussi/yfinance

respective values. Consequently, the models are designed to generate predictions for aggregate future close prices or ratios for a specified period based on aggregate sentiment values from a specific sample period, as opposed to a direct prediction between a single news article's sentiment score and a single close price or instantaneous price ratio.

3.4.4 Validation Using Random Data

Additional vectors filled with normally distributed random values ranging from -1 to 1 were also created to validate the predictive effect of the sentiment data. A normal distribution was chosen because some of the statistical methods used in the analysis, such as Pearson's R, operate under the assumption of a normal distribution. This random data was then used as an input to both the statistical and AI based (LSTM) models to facilitate a comparison with the effects of the sentiment data.

3.5 Data Analysis

The entire data analysis was implemented in python, primarily making use of the statistical models available via the open source scipy[9] (Virtanen et al., 2020), and statsmodels.api[10] (Seabold & Perktold, 2010) python modules. Statistical analysis is required to address hypothesis one and its sub-parts. The statistical analysis methods aim to determine if the effect of sentiment on stock price volatility is visible over the short-term. Namely, whether the simple moving average (SMA), simple moving variance (SMV), or simple moving covariance (SMC) of the sentiment scores have a statistically significant relationship with the SMA or SMV of the close prices, or the stock price ratio. This investigation was carried out via the following steps: chronologically aligning the close prices, creation of samples of close prices, sentiment, and random values, followed by both univariate and multivariate statistical analysis of the samples, and then finally statistics-based filtering of the results. This process is outlined by the diagram displayed in Figure 3, note "Gn_sent" is an abbreviation for sentiment from Google News with neutral scores removed, while "Gn_0_sent" is an abbreviation for the Google News sentiment data which still contains the neutral (zero) scores.

---

[9] scipy.py is an open source python module created by Virtanen et al (Virtanen et al., 2020) that provides functions for programmatic statistical analysis.
https://github.com/scipy/scipy

[10] statsmodel.api is a python library providing open-source statistical analysis tools to enable programmatic statistical analysis, created by S. Seabold and J. Perktold in 2010 (Seabold & Perktold, 2010).
https://github.com/statsmodels/statsmodels

Chronologically aligning the close prices was important because it reduced the influence of any anomalous missing data values. When conducting a multivariate regression, all sample sizes must be equal for both the regressors and the response variable. Additionally, the stock price ratio is defined as the ratio of the (on average) higher close price to the lower close price. This ratio can only be calculated at moments when both close prices are known, and furthermore, requires that both close prices correspond to the same time. The alignment was time based and utilised the associated datetime indices of the close prices with a resolution in minutes to ensure that each row of close prices in each respective dataframe column corresponded to the same date-time instance. During alignment, any close prices corresponding to time instances for which both close prices were not reported were removed.

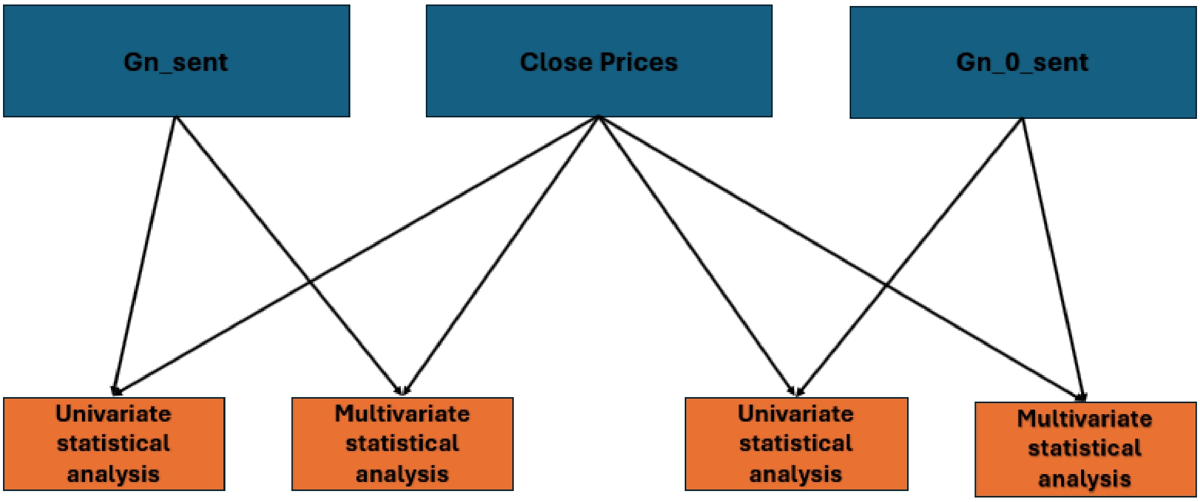

Figure 3: Independent variables and statistical analysis methods.

During the analysis phase, numerous sampling periods (Δt_i) were trialled to determine the temporal range over which the sentiment held influence because this was one of the research gaps identified in the literature study. This was also necessary to determine which sampling periods would yield the most accurate predictions to use in the pairs trading models. These sampling periods belonged to three separate categories, namely: sentiment sampling times, lag times, and close price sampling times. The lag times that were tested were: 2 minutes, 30 minutes, 1 hour, 6 hours, 1 day, 3 days, 1 week, 2 weeks, and 4 weeks (1 month). The same time periods were used for both sentiment and close price sampling. This study iterated through the datetime indexed sentiment and close price data taking samples of sentiment for each corresponding sampling period, in addition to samples of closing prices which corresponded to the different close price sampling periods lagged by the corresponding lag times. These samples were then used for the statistical analyses.

The statistical analyses carried out within this study were limited to linear regressions that investigated the predictive relationship between sentiment and stock price or stock price ratio movement. All of the regressions used a sample size of 100 to ensure an adequate statistical power. Because of the finite size of the data set, the possible time combinations were limited to those for which 100 samples were possible within the span of the data set. 80% of the data set was used to fit the regression model, leaving 20% for testing the sentiment assisted pairs trading strategies. For the 105-day data set, data spanning 84 days was used for the regression models which meant that 384 different time combinations were possible, as illustrated by the 3D graph presented in Figure 4.

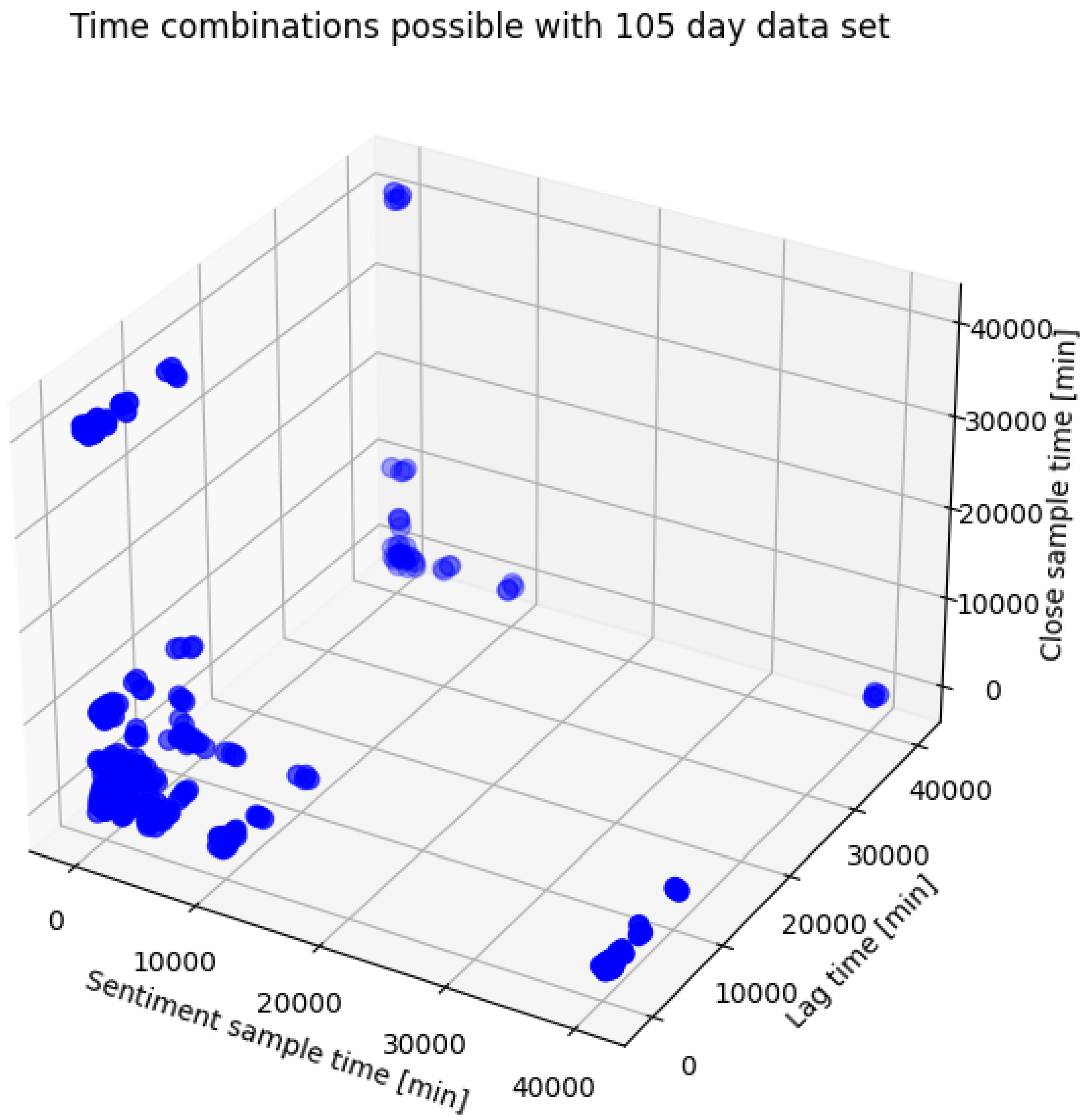

Figure 4: All possible sample and lag time combinations (with slight jitter to aid visibility)

The univariate statistical model that was employed conducted the Pearson's R test and an ordinary least squares regression using the scipy.stats python module to determine the strength of the relationship between both the SMA and SMV of the sentiment and close prices of individual stocks. The architecture of the univariate model is displayed in Figure 5. Four different investigations were conducted for each stock, namely: SMV of sample sentiment as the independent variable with SMV of sample close price as the dependent variable, SMV of sample sentiment as the independent variable with SMA sample close price as the dependent variable, SMA of sample sentiment as the independent variable with SMA sample close price as the dependent variable, and SMA of sample sentiment as the independent variable with SMV of sample close price as the dependent variable. A maximum p-value of 0.05 was imposed upon the results to limit the results to those with a statistical significance of 95% or higher. The remaining Pearson's R coefficients were then used in the Modified Bollinger Bands strategies detailed in section 3.7.

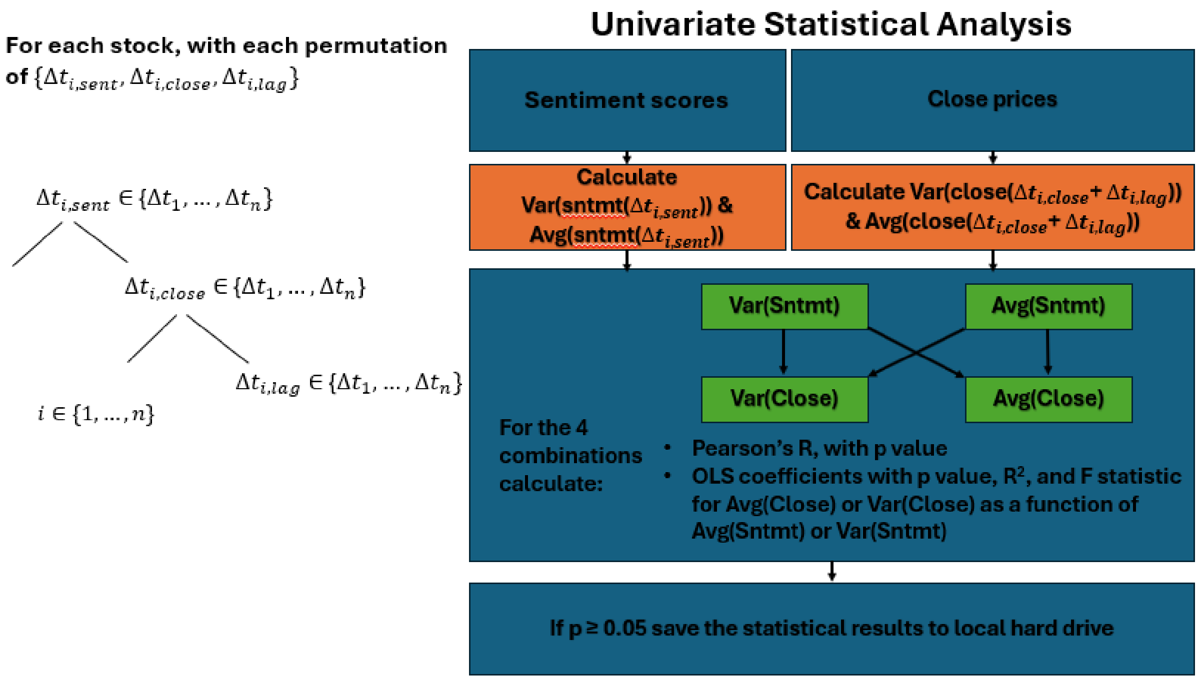

Figure 5: Univariate statistical analysis architecture

The remainder of the statistical analysis was multivariate, and employed an ordinary least squares (OLS) type regression (Singh et al., 2021; Yao et al., 2022) using the statsmodels.api python module (Seabold & Perktold, 2010). An overview of the architecture used for the multivariate statistical analysis can be found in

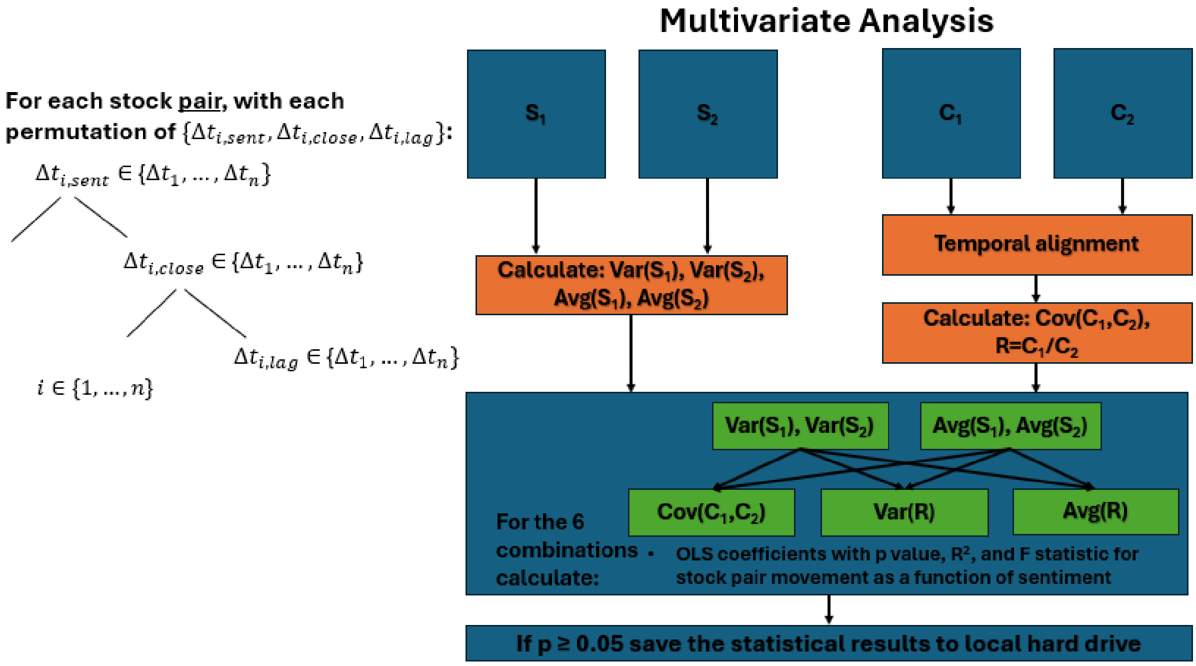

Figure 6. Note that in

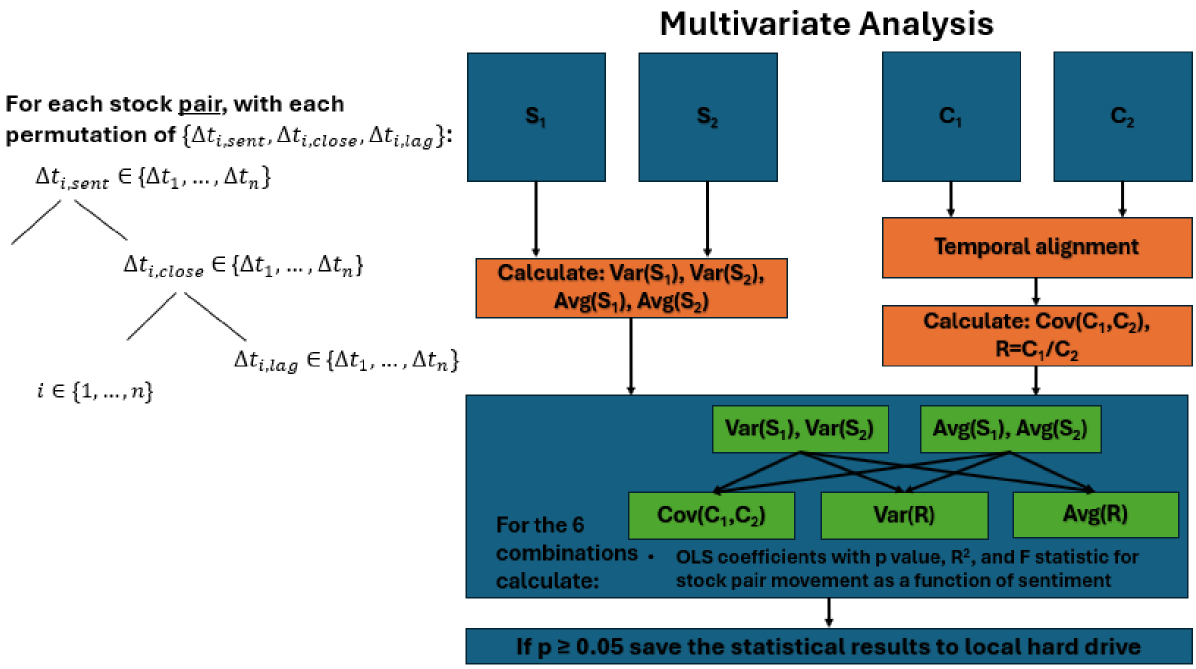

Figure 6 the close price and sentiment score datasets of the two stocks have been abbreviated to C_1 and C_2, and S_1 and S_2 respectively. Multivariate linear regression (MLR) was used to determine the relationship between the SMA or SMV of the sample sentiment for both stocks (as the independent variable), and the SMC of their closing prices (as the dependent variable) for the aforementioned sample and lag times. Additionally, MLR was conducted to test hypotheses 1a, 1b, 1c, and 1d, regarding the relationship between the SMA or SMV (as the independent variable) of both sentiment samples and the SMA or SMV of the stock price ratio (as the dependent variable) for the different sample and lag times. The resulting correlation coefficients were then filtered based on both the F-statistic of overall significance, which shows that a significant amount of the variance of the dependent variable is explained by the independent variables, and the p-value. Again, a maximum p-value of 0.05 was used, with a corresponding minimum F-statistic of 3.165 (independent variables with an F-statistic lower than 3.165 are responsible for an insignificant amount of variance in the dependent variable). All sample and lag time combinations which did not yield significant results based on these criteria were removed. Additionally, the adjusted R^2 value of each MLR model was returned, which revealed the models which provided the best fit.

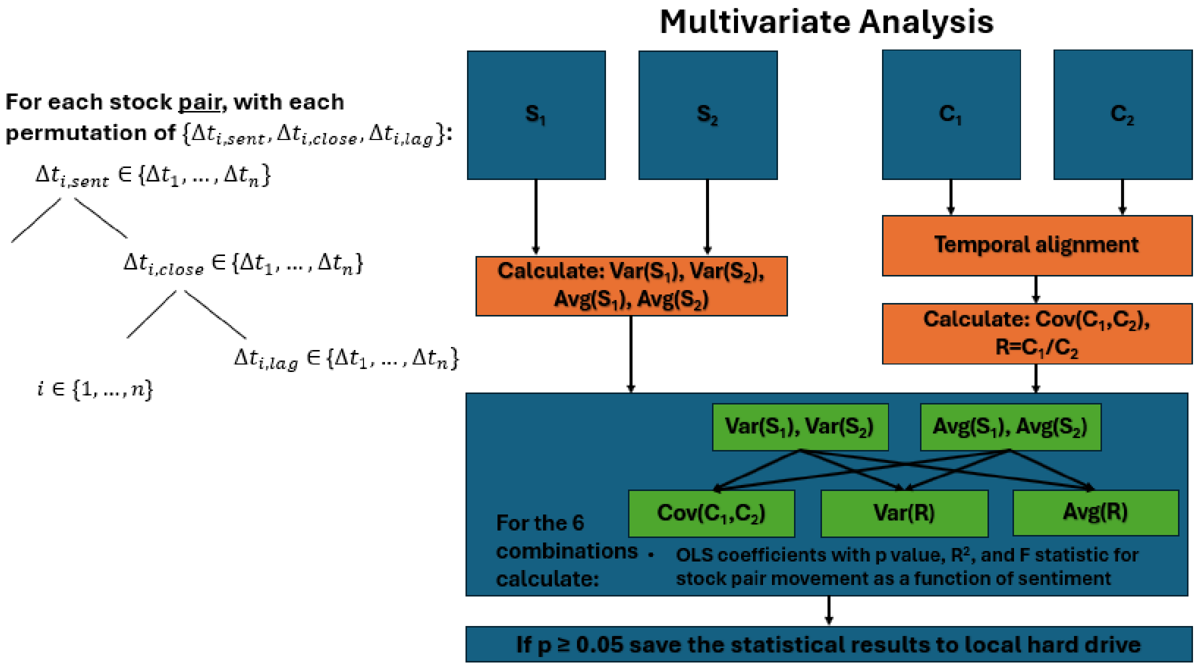

Figure 6: Multivariate statistical analysis architecture

3.6 LSTM Model

This section sought to evaluate hypothesis 3: *Google News derived sentiment can improve the predictive accuracy of an LSTM model which predicts the stock price ratio of a pair of highly correlated* stocks. To capture the patterns hidden in the longer-term stock pairs data a deep learning model was used because a substantial amount of research (Du, 2022; Li, 2022) has shown that purely statistics-based methods often struggle to capture the complex non-linear behaviour of the stock market. An LSTM model is deemed to be the most suitable model for this task based on recent research (Cao et al., 2020; Du, 2022; Sen et al., 2021; Singh et al., 2021; Touzani et al., 2019; Yao et al., 2021), in particular due to its ability to utilise sentiment data for stock market prediction (Ayyappa & Siva Kumar, 2022; Dutta et al., 2021; Gehlor & Singh, 2022; Jain et al., 2022; Li, 2022; Ma et al., 2021; Muthivhi & van Zyl, 2022;

Owen & Oktariani, 2020; W. Zhang et al., 2021; Zhou et al., 2023). The subsequent LSTM model was coded, implemented and evaluated in python using the keras and sklearn libraries respectively (Keras-Team, 2024), (Pedregosa et al., 2011). The LSTM model can take multiple independent variables as predictors, and then utilise these input (X) variables to predict one or more dependent (Y) variables. The model implemented in this study will use a variety of different independent variables to predict the minimum and maximum stock price ratio for a range of different future periods. The architecture of the code which was used to implement the LSTM model is depicted in

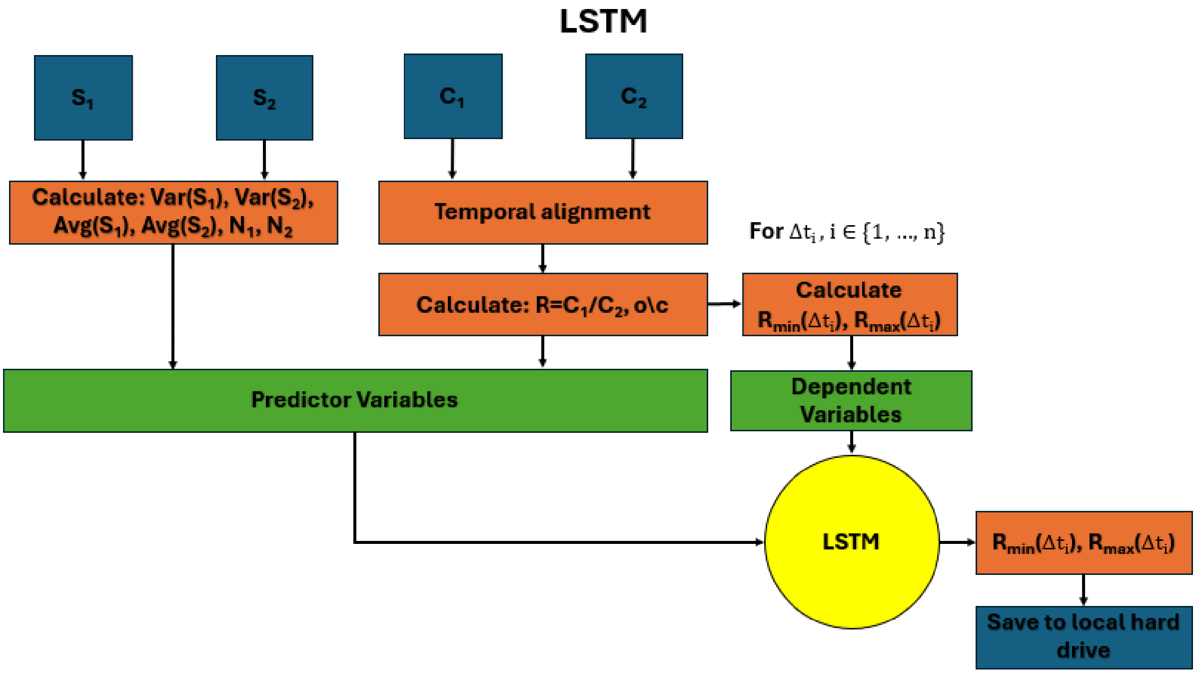

Figure 7. Note that the same abbreviations for close price and sentiment hold as in

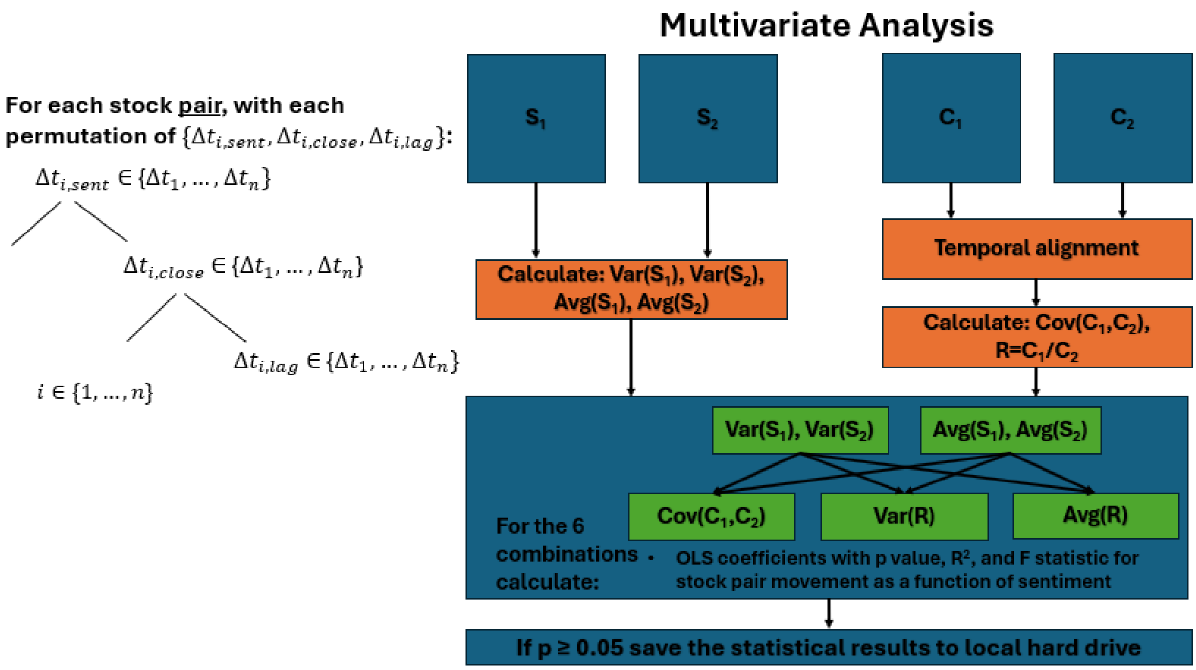

Figure 6, with the addition of the abbreviation o/c for the dichotomous variable that indicated whether the stock market was open or not, and the abbreviations N_1 and N_2 representing the number of news articles about the respective companies that were published within the time period. The resulting predictions were used to improve a pairs trading model. The accuracy of the LSTM model was evaluated via the root mean square error (RMSE), which is a commonly accepted evaluation method for LSTM models (Li, 2022; Singh et al., 2021; W. Zhang et al., 2021).

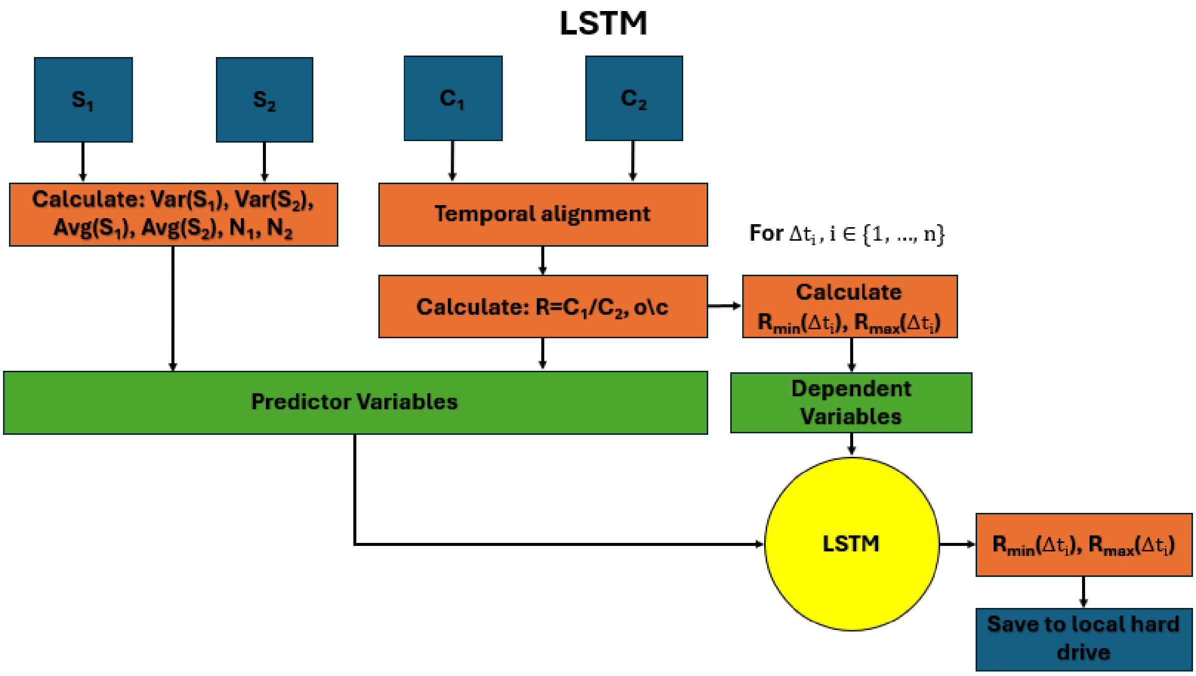

Figure 7: LSTM model implementation architecture

In this study, the variables were provided to the model in a time series and were aligned such that it was ensured that there was close price data from both stocks in the pair at every time instance. The LSTM model can take multiple independent variables as predictors, and then utilise these input (X) variables to predict one or more dependent (Y) variables. These variables were provided as a time series with a frequency of one entry per two minutes, because this was the frequency of the close price data collected from Yahoo! Finance. All sentiment data points (SMA and SMV) were calculated over the 2-minute period before each close price. If no news articles were published in this period, the sentiment score was set to zero, in addition to the *number of articles* variable. This study utilised an LSTM model with eight predictor variables, namely: a dichotomous variable to indicate whether the stock market was open or closed, the stock price ratio, the number of news articles mentioning each respective stock, and the SMA and the SMV of sentiment for each stock (calculated since the last reported close price). The dependent variables were the maximum and minimum stock price ratio (R_{min} and R_{max}) for the next *t* minutes, where *t* was a time period variable. Because the LSTM assisted pairs trading is a novel approach, the optimal time period for this prediction was not known, and it was speculated that the ideal forecasting window would need to strike a balance between being short enough to enable accurate forecasting, but long enough to allow the algorithm to see if large changes in the stock price ratio are imminent. Therefore, the following values of *t* were trialled in a parameter sweep to ascertain the ideal time period to use for the pairs trading strategy: t = [range(2,11,2),range(16,33,4),range(40,61,10), range(90,151,30), range(180,1441,60)]. The LSTM model was trained using the first 84 days of data, and then tested using the last 21 days of data. The model was then further validated by comparing to the results of a model which used the random sentiment dataset instead of the actual sentiment scores.

LSTM models have different settings which can be adjusted and optimised depending on the application. This study set the maximum number of epochs for training the model to 200 with a patience value of 30 epochs – after which point if the model had not improved, the training would stop. This was deemed sufficient after observing a selection of trial models. The batch size for each prediction was set to 15 (30 minutes), the rectified linear unit (ReLU) activation function was used (Ma et al., 2021), the Adam optimisation algorithm was used (Cao et al., 2020; Owen & Oktariani, 2020), and the mean absolute error (MAE) loss function was used (Du, 2022; W. Zhang et al., 2021). The accuracy of the LSTM model's predictions was assessed via the root mean square error (RMSE) (Li, 2022; Singh et al., 2021; W. Zhang et al., 2021)..

3.6.1 LSTM Data Preparation

All of the data for the LSTM model required preparation prior to the training and testing of the model. For each stock pair, this entailed reformatting the sentiment data into the SMA and SMV for each 2-minute period, creating a news article counter variable for each 2-minute period, and partitioning the data into a training set, and a test set. Additionally, the minimum and maximum values of the stock price ratio (the dependent variables) were calculated over the data set for the range of forecasting windows. To maintain comparability with the statistics-based strategies, the same partition of 84 days for training and 21 days for testing was used. Following the removal of any empty values from the close price data set, the data sets of both stocks in the pair were chronologically aligned to ensure that there was always a close price value for both stocks at the same time instance. The remaining eight independent variables and two dependent variables were subsequently calculated for each time instance and then saved in a feather type file for future use. The randomly generated datasets were also used as inputs to the LSTM model in separate trials to allow a comparison with the effects of the sentiment data.

3.7 Pairs Trading Strategies

Hypotheses 2 requires the implementation and subsequent evaluation of a variety of different pairs trading strategies. Hypothesis 2 (*Google News derived sentiment can be implemented in combination with linear regression to improve a modified pairs trading strategy to yield increased returns compared to a basic Bollinger Bands-based strategy*) necessitated the implementation of a basic Bollinger Bands-based pairs trading strategy, in addition to three variations of a pairs trading strategy that utilised information extracted from the sentiment-close price statistical analysis. The first variation of the Bollinger Bands-based strategy incorporates a method of weighting the bands to either widen or contract them according to the level of variance to be expected based on previous statistical analysis such as to capitalise on expected large stock movements before opening or closing a position. The second and third variations attempt to directly predict the standard deviation of the stock price ratio, which is one of the key components used to calculate the Bollinger Bands. Because this cannot be achieved via analytical mathematics a numerical approximation by means of a second order Taylor series approximation as given in equation 1 was used. Note that the standard deviation is the square root of the variance and is represented in equation 1 with the Greek letter σ.

$$\sigma\left[\frac{C_1}{C_2}\right] \approx \sqrt{\frac{var[C_1]}{E[C_2]^2} - \frac{2E[C_1]}{E[C_2]^3}cov[C_1, C_2] + \frac{E[C_1]^2}{E[C_2]^4}var[C_2]}$$

Equation 1: Taylor series approximation for the variance of a ratio (van Kempen & van Vliet, 1995).

Where C_1 and C_2 are the close prices of the two stocks of the stock pair. The values used for the expected values, variances, and covariances differ between the two strategies.

Note that in all of the proceeding strategies a hedging ratio derived via OLS was employed, where the hedging ratio was calculated as the first regression coefficient of the OLS regression used to find the (on average) more expensive stock price as a function of the (on average) less expensive stock price. This is a commonly used approach to ensure that the relative movements of both stocks in the pair affect the returns equally by ensuring that during each trade, the amount of money invested in each stock in approximately equal (Longmore, 2019). Additionally, the initial capital invested in the first open position was 100 USD for each trial, the following open positions reinvested the total amount returned by the previous close. The final profit returned was thus the balance following the final closure of the positions, minus 100 USD.

3.7.1 Basic Bollinger Bands

Pairs trading via a fundamental Bollinger Bands-based strategy proceeds as follows. The price ratio of the two stocks is monitored, in addition to four so-called Bollinger Bands. These bands represent distances from the SMA of the stock price ratio, which were calculated using the 20-day SMA of the stock price ratio and the 20-day standard deviation of the stock price ratio at each 2-minute close instance. The lowest (first) Bollinger Band was calculated by subtracting two standard deviations from the SMA, the second Bollinger Band was the SMA minus one standard deviation, the third Bollinger Band was the SMA plus one standard deviation, and the fourth Bollinger Band was the SMA plus two standard deviations. If at any instance the stock price ratio exceeded the uppermost (fourth) Bollinger Band, a pairs trading position was opened with a short position on the (likely over valued) more expensive stock, and a long position on the (likely undervalued) less expensive stock. This position was held until the price ratio fell below the third Bollinger Band, at which point the respective long and short positions were closed. This logic is based on reversion to mean theory, which postulates that for a pair of highly correlated stocks, short-term deviations from the long-term average price ratio are highly likely to revert (Göncü & Akyıldırım, 2016; Ramos-Requena et al., 2021). Inversely, when the stock price ratio dropped beneath the first Bollinger-Band, a long position was opened on the more expensive stock, and short on the less expensive stock, and then subsequently closed when the stock price ratio rose again above the second Bollinger Band.

3.7.2 Modified Bollinger Bands: Weighted R

This is a novel pairs trading method, wherein the strength of statistical relationships (Pearson's R) between sentiment and price ratio volatility were harnessed to "weight" the Bollinger Bands and thus move them further or closer to the SMA, as a function of the predicted volatility of the price ratio. To the best of the author's knowledge, this strategy has not been used in any other works and is proposed here for the first time. The weight multiplier was calculated at each 2-minute instance, corresponding to the available close price points. If high volatility was expected, the bands were shifted further away in an attempt to capitalise on large price fluctuations, conversely, if low volatility was expected the bands were shifted closer to the SMA. This strategy utilised the Pearson's R values calculated during the statistical analysis phase that described the strength of the relationship between either the average or the variance of the news sentiment, and the variance of the close price for each individual stock for specific combinations of sentiment sample, lag, and close price sample times. Four different independent variables were trialled, namely: average sentiment, variance of sentiment, average sentiment with neutral sentiment removed, and variance of sentiment with neutral sentiment removed.

The weights were created by first determining the expected contribution to the volatility of the price ratio from each stock, and then taking the sum of the respective contributions multiplied by a user specified "trust weight" (the optimum value of which was determined via a parameter sweep). This proceeded as follows.

For each stock:

- Determine the statistically significant combinations of sentiment sample time, lag time, and close sample time.
- Transform the remaining corresponding Pearson's R values such that the sum of these transformed values is equal to one.
- Sample the sentiment data at the current time point using the corresponding lag and sample times, then multiply the average or variance of the sample sentiment by the corresponding transformed Pearson's R value.
- Take the sum of all of these sentiment-Pearson's R value products and divide by either the average or the variance of the sentiment from the previous 20 days.

Once this weight has been determined for both stocks, the average is taken, after which the modified Bollinger bands are determined by multiplying the 20-day standard deviation by the calculated weight, and the user specified "trust weight".

3.7.3 Modified Bollinger Bands: Semi-Predicted Volatility

The second variation of the Bollinger Bands-based strategy sought to predict the standard deviation of the price ratio, and subsequently use the predicted value to compute the Bollinger bands as opposed to computing a value based solely on the past 20 days of movement, because historical data is not always representative of future movements (Ma et al., 2021). This was achieved via equation 1, which predicts the standard deviation of the stock price ratio as a function of the expected close prices, the variance of each close price, and the covariance of the close prices. In this semi-predicted volatility strategy, the variances of the two stocks were predicted based on statistical correlation with sentiment. The expected values of the close prices and their covariance were taken as the 20-day SMA and SMC respectively. The predicted variance for each close price was calculated as the weighted average of all of the predicted variances calculated via the statistically significant sample and lag time combinations of the OLS MLR, wherein the weights were the min-max normalised respective F-statistics. The subsequent semi-predicted standard deviation was then used in the formation of the Bollinger Bands analogously to the Basic Bollinger Bands-based strategy.

3.7.4 Modified Bollinger Bands: Fully Predicted Volatility

The third variation of the Bollinger Bands-based strategy also sought to predict the standard deviation of the price ratio and then utilise it to compute the Bollinger Bands, however, in addition to utilising predicted variance, this strategy utilised predicted covariance, and predicted expected close price, which were predicted analogously to the predicted variance in *modified Bollinger Bands-based strategy 2*. The subsequent fully predicted standard deviation was then used in the formation of the Bollinger Bands analogously to Basic Bollinger Bands-based strategy.

The different pairs trading model variants were evaluated based on their net returns over the trial period, with consideration for the total number of trades executed. This is an important consideration because excessive trading can erode returns due to transaction costs in a real stock market.

3.8 Hardware and Software

All code was written in the python 3.11.2 using Microsoft Visual Studio code editor. The subsequent models were run on a Lenovo G580 laptop with 8GB of RAM, an Intel i5-3210M CPU, with Microsoft Windows 10 installed. All code and saved sentiment, close price, statistical, and model data was saved on the internal local hard drive with periodic back-ups made in Microsoft OneDrive.

4 Results

This chapter provides the descriptive statistics of the datasets which were used to investigate hypotheses one through four, followed by an explanation of the working principals of the statistical methods that were used to assess the relationships between news sentiment and stock price movement. This is followed by the results of the tests for hypotheses one through four, which illustrate the strength of the relationship between news sentiment and stock price movement for both individual stocks and the stock price ratio of a pair of stocks, how this information can be used to improve a pairs trading model, and then finally, how news sentiment can be used to improve the predictive accuracy of a LSTM model which can be used for pairs trading.

4.1 Descriptive Statistics

The descriptive statistics for the Google News sentiment scores (with and without neutral sentiment), and the close price data for each stock are presented in tables 8, 9 and 10. Note that the p values for all of the Shapiro-Wilk statistics were in the order of 10^{-19} or less and were therefore considered to be effectively zero.

The descriptive statistics of the original sentiment score data (including neutral scores) for each of the eight stocks are presented in Table 8.

Table 8: Descriptive Statistics: Google News sentiment (with neutral scores)

	PEP	KO	DELL	HPQ	UAL	AAL	MA	V
Total	2493	2645	2719	2686	2576	2475	2615	2641
Mean	0.099	0.102	0.075	0.145	0.097	0.045	0.135	0.040
Std. dev.	0.300	0.296	0.253	0.252	0.326	0.317	0.287	0.312
Variance	0.090	0.088	0.064	0.064	0.107	0.100	0.082	0.097
Minimum	-0.914	-0.869	-0.926	-0.765	-0.875	-0.959	-0.920	-0.891
Maximum	0.920	0.917	0.957	0.908	0.912	0.904	0.926	0.896
Shapiro-Wilk, p	0.868, 0.0	0.900, 0.0	0.748, 0.0	0.817, 0.0	0.930, 0.0	0.897, 0.0	0.847, 0.0	0.897, 0.0

The descriptive statistics of the sentiment score data without neutral scores for each of the eight stocks are presented in Table 9.

Table 9: Descriptive Statistics: Google News sentiment (neutral scores removed)

	PEP	KO	DELL	HPQ	UAL	AAL	MA	V
Total	1067	1158	810	1103	1383	1135	1121	1205
Mean	0.231	0.232	0.251	0.352	0.179	0.096	0.314	0.088
Std. dev.	0.424	0.412	0.413	0.287	0.429	0.463	0.369	0.457
Variance	0.180	0.170	0.170	0.083	0.184	0.214	0.136	0.209

Minimum	-0.914	-0.869	-0.926	-0.765	-0.875	-0.959	-0.920	-0.891
Maximum	0.920	0.917	0.957	0.908	0.912	0.904	0.926	0.896
Shapiro-Wilk, p	0.901, 0.0	0.900, 0.0	0.882, 0.0	0.887, 0.0	0.902, 0.0	-0.959, 0.0	0.910, 0.0	0.916, 0.0

The descriptive statistics of the close prices for each of the eight stocks and the corresponding stock price ratios are presented in Table 10.

Table 10: Descriptive Statistics: close prices of the eight selected stocks

	PEP	**KO**	**DELL**	**HPQ**	**UAL**	**AAL**	**MA**	**V**
Total	13542	13545	13543	13545	13545	13540	13539	13543
Mean	180.818	59.623	59.791	30.261	50.212	15.449	399.663	239.008
Std. dev.	7.058	2.256	7.164	2.491	4.828	1.961	9.782	5.559
Variance	49.810	5.088	51.323	6.204	23.306	3.844	95.693	30.905
Minimum	156.090	54.565	50.630	25.295	40.365	12.260	373.267	226.200
Maximum	192.345	63.230	72.790	33.860	58.220	19.032	418.310	249.970
Shapiro-Wilk, p	0.948, 0.0	0.913, 0.0	0.796, 0.0	0.899, 0.0	0.927, 0.0	0.936, 0.0	0.951, 0.0	0.963, 0.0
Total(R)	13542		13543		13540		13539	
Mean(R)	3.03		2.01		3.26		1.67	
Variance(R)	0.002		0.157		0.020		0.001	
Pair Correlation	0.93		-0.77		0.94		0.81	

The random datasets that were used to compare the predictive power of sentiment compared to a random input were normally distributed between -1 and 1, and therefore had a mean of 0, with a standard deviation of 0.34, and a variance of 0.116. The datasets were generated such that there was an artificial sentiment value for each stock at every moment that a close price was reported.

4.2 Linear Regression Results

The relationships between the sentiment and the close prices of the individual stocks in each pair were investigated by calculating Pearson's R, performing a linear (OLS) regression, and then evaluating the results based on the corresponding p values, adjusted R^2 values, and F statistics.

Pearson's R is a statistic which describes the strength and direction of a linear relationship between two variables. It is the ratio of the covariance of the two variables divided by the product of their standard deviations, i.e. the ratio of the amount of variance in one dataset that can be explained by the variance in the corresponding predictor dataset. Pearson's R varies from -1, indicating a perfect negative relationship, to 1, indicating a perfect positive relationship. Because Pearson's R is sensitive to outliers, it was important to ensure that the dataset did not contain any outliers. Pearson's R also assumes that the data is normally distributed, which necessitated the use of the Shapiro-Wilk normality test on the respective data sets. The null hypothesis would be accepted if a Pearson's R of zero was found.

OLS shows the strength and direction of the relationship between an independent and a dependent variable, and the regression coefficients can subsequently be used to directly predict the dependent variable as a product of the first regression coefficient and the independent variable, sometimes summed together with an additional constant if required. The principle of an OLS regression is to fit a line between all of the data points which minimises the squares of the normal Euclidean distances from each of the data points to the line. The gradient of this line is given by the regression coefficient and the possible additional constant represents the y intercept of the line. OLS can also be applied to datasets with higher dimensionality, and a regression coefficient for each independent variable will be calculated. The main shortcoming of the OLS method is that it is only able to capture linear (straight line) relationships.

The R^2 value provides an indication of the proportion of the variance in the dependent variable that is explained by the independent variable in the model. The adjusted R^2 provides an analogous indication, but with an adjustment to account for the effect of the sample size. A model which can explain all of the variance in the data would have an R^2 value of 1, whereas a model which cannot explain any of the variance (but may still provide a reasonable fit, for example the mean value of the data) would have an R^2 value of 0. The disadvantage of R^2 is that small sample sizes can yield artificially high R^2 values, and that it is also only suitable for assessing linear relationships. Hence, the adjusted R^2 was used in this study to account for smaller and varying sample sizes. The adjusted R^2 extends the range of possible scores downwards to negative infinity, which is indicative of a model which provides a worse fit than a horizontal line.

The F statistic is a measure of the model's statistical significance and is the ratio of the variance explained by the model to the unexplained variance in the dataset. In the context of regression, the F statistic indicates whether at least one of the regression coefficients predicts a statistically significant amount of the variance in the dependent variable. Small values of the F statistic, i.e. below 3.165 are indicators that the null hypothesis should be accepted.

The p value is an important concept in statistics, because it represents the probability that the observed data would have occurred completely randomly, when the null hypothesis is correct. The p value ranges from 0, representing complete statistical confidence, to 1, representing no confidence. In this study 95% statistical significance was required, which corresponded to p values no larger than 0.05. When regression coefficients were found with a p value no larger than 0.05, the null hypothesis was rejected, and it was concluded that a statistically significant relationship was present. The disadvantage of the p value is that it is not indicative of the strength of the effect, hence it is important to also inspect other values such as the R^2 value and the correlation coefficients. Furthermore, the p value is sensitive to the sample size. In large samples very weak yet statistically significant effects may be detected, whereas in small samples stronger effects may not be classed as being statistically significant, which further justifies the use of the adjusted R^2 value in this study.

4.3 Hypothesis 1

Hypothesis 1 proposes that sentiment collected from Google News headlines is a predictor for the price ratio of highly correlated stocks. The hypothesis is split into four sub-parts:

a. The average sentiment score of two correlated stocks is a predictor for the average stock price ratio of the two stocks.
b. The average sentiment score of two correlated stocks is a predictor for the variance of the stock price ratio of the two stocks.
c. The variance of the sentiment score of two correlated stocks is a predictor for the average stock price ratio of the two stocks.
d. The variance of the sentiment score of two correlated stocks is a predictor for the variance of the stock price ratio of the two stocks.

Hypothesis 1 was tested by calculating the ordinary least squares (OLS) correlation coefficients for each of the four combinations to more specifically test each of the four sub-hypotheses. This process involved a univariate analysis, in which the relationships between the news sentiment about each individual stock and its respective stock price movement was analysed, and a multivariate analysis, where the relationship between the news sentiment about both stocks in a pair and the movement of their stock price ratio was analysed.

4.3.1 Univariate Analysis

Figure 8 illustrates the results of the univariate analysis for Pepsi. The results for the other seven companies are provided in appendix A. The three-dimensional axes represent the different sentiment sample times, lag times between sentiment and close price, and close prices sample times, and each data point is located in three-dimensional space on the plot according to the combination of times used for the respective OLS models. For each stock, the magnitude of the adjusted R^2 value (indicating the explanatory power of the model) for the OLS regressions of sentiment variance or average with the variance or average of the close price is represented by the colour of the marker; warmer tones indicating larger adjusted R^2 values and thus better explanatory power. Note that only statistically significant ($p < 0.05$, $F > 3.165$) regressions with an adjusted R^2 value greater than zero were included in the graphs. The different sentiment types (sentiment including neutral scores, "gn_0", sentiment without neutral scores, "gn", and the randomly generated data, "rand") with the four different data types result in a total of 12 regression models for each stock, represented by the 12 different marker types respectively. The total number of statistically significant regressions for each combination is included in the legend of each figure, in addition to the average OLS coefficient. The legend entries are formatted as follows:

"stock ticker" close_variance/average" _" sentiment_type" _" sentiment variance/average" + number of positive statistically significant adjusted R^2 values, average statistically significant OLS coefficient.

The legends for Figure 8 and the additional figures in appendix A are formatted such that the legend entries are in ascending order when read from top to bottom, from lowest to highest average adjusted R^2 score. Thus, the legend also provides the order of strongest to weakest relationships for the stock.

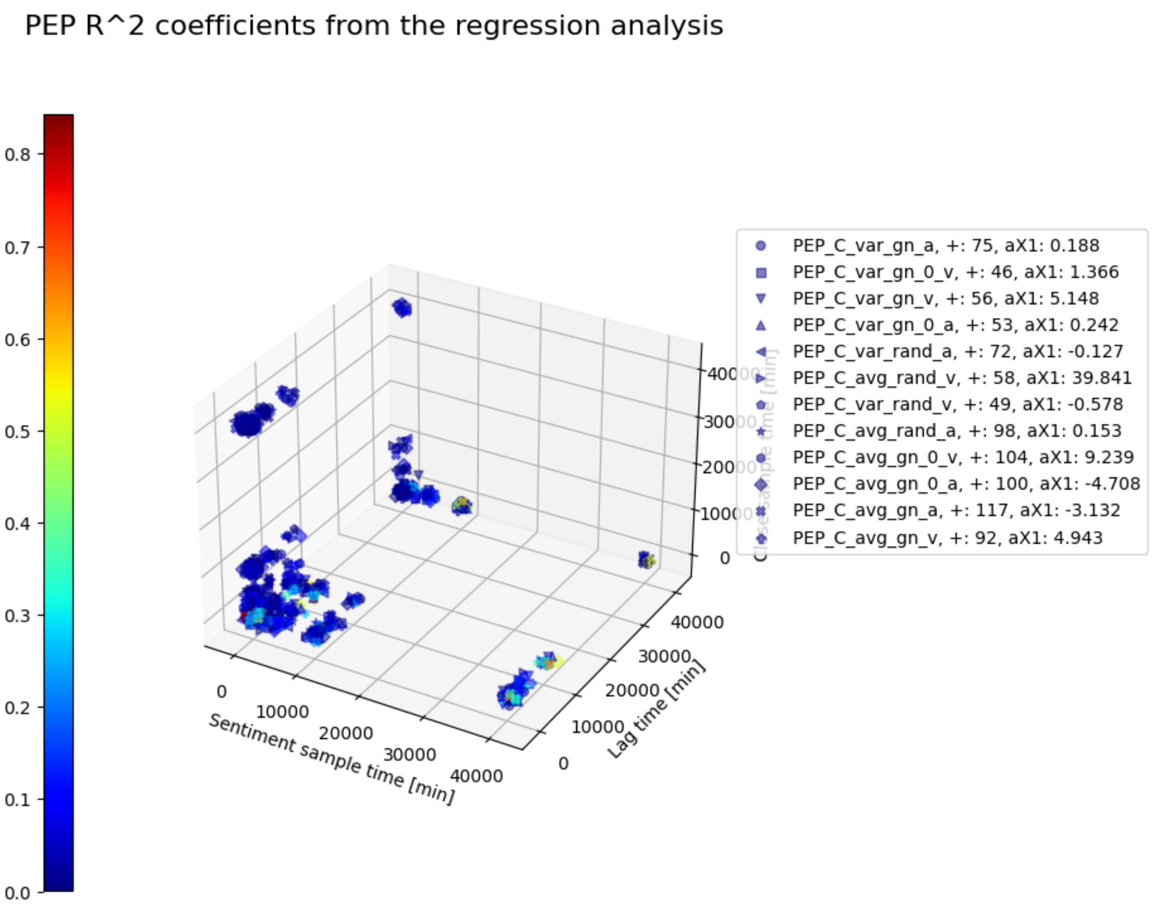

Figure 8: Adjusted R^2 values for Pepsi

It can be seen from Figure 8 (and the figures in appendix A) that for all 12 variable combinations there were statistically significant relationships present at a 95% significance level ($p<0.05$, $F>3.165$). Despite being statistically significant, the predominantly blue tone of the clusters indicates that most of these relationships had a weak explanatory power (adjusted $R^2<0.3$). However, a small number of stronger relationships were found. For Pepsi, a small number of markers with a yellow tone can be seen in different regions, indicating the presence of several regressions with moderate explanatory power. For the remaining stocks, there was at least one strong relationship identified, with the maximum adjusted R^2 values of the regressions for each stock ranging from 0.7821 to 0.9693. This can be confirmed by inspecting the summary data presented in Table 11, and the maximum adjusted R^2 values for each stock given in Table 12. The model types are given in Tables 11 and 12 in the format[10]: close_ {variance or average} _sentiment type_ {variance or average}. It is evident from both the graphs and the summary data that the variance of the close price (C_Var) is less explainable with any of the types of news sentiment. Additionally, all of the average adjusted R^2 values of the close price average prediction models are greater than the average adjusted

R^2 values of the close price variance prediction models, and all of the maximum adjusted R^2 values for each stock come from the average close price prediction models. Interestingly, for close price variance prediction the random sentiment model's adjusted R^2 values were higher on average than those models which used real sentiment, and for average close price prediction the variance of random sentiment was the second-best predictor. 87.5% of the maximum adjusted R^2 values came from models that used real news sentiment. Because all of the average adjusted R^2 values have a magnitude of approximately 0.1 or less it can be concluded that although there are a relatively high number of statistically significant regression models, most of them provide relatively poor explanatory power, which can be confirmed by the observation that randomly generated "sentiment" performs similarly well on average. However, there are a small number of combinations which work well, demonstrating the importance of selecting suitable sampling period sizes.

Table 11: Average adjusted R^2 values for the univariate models

Model Type[11]	Average Adjusted R^2	Average OLS Coefficient
C_var_gn_a	0.0244	-0.118
C_var_gn_0_v	0.0309	5.685
C_var_gn_0_a	0.0312	-1.214
C_var_gn_v	0.0420	10.320
C_var_rand_a	0.0426	-0.343
C_var_rand_v	0.0518	0.100
C_avg_rand_a	0.0880	-13.441
C_avg_gn_0_v	0.0971	2.949
C_avg_gn_a	0.101	-1.369
C_avg_gn_0_a	0.102	-23.350
C_avg_rand_v	0.104	-10.422
C_avg_gn_v	0.109	17.502

Table 12: Maximum adjusted R^2 values for each stock from the univariate regressions

Stock	Model Type[10]	Adjusted R^2_{max}	$t_{sentiment}$, t_{lag}, t_{close}	OLS Intercept	OLS Coefficient

[11] The model type code for the univariate analysis can be understood accordingly: The first letter C, stands for "close price", the second term is the aggregation method (avg = average, var = variance), the third term is the sentiment type (rand = the random data set, gn_0 = Google News sentiment with neutral sentiment, gn = Google News sentiment without neutral sentiment, and the final letter indicates the aggregation method used for the sentiment (a = average, v = variance).

PEP	C_avg_gn_0_v	0.8423	40320, 10080, 60	160.5841	210.1236
KO	C_avg_gn_v	0.7821	360, 10080, 2	60.3137	0.9257
DELL	C_avg_gn_a	0.9688	360, 60, 2	52.5374	-2.4169
HPQ	C_avg_gn_v	0.9459	360, 4320, 2	29.7139	5195.9160
UAL	C_avg_rand_a	0.9056	1440, 4320, 2	54.2361	25.7019
AAL	C_avg_gn_v	0.9693	30, 60, 2	18.4404	-119.8032
MA	C_avg_gn_0_v	0.8143	4320, 360, 2	382.9347	-99.9299
V	C_avg_gn_0_v	0.9472	360, 4320, 2	226.7638	62.5955

4.3.2 Time Dependence

This study trialled 384 different sentiment sample time, lag time, and close price sample time combinations due to the disagreement among the surveyed literature regarding the temporal persistence of the effect of sentiment on stock price behaviour. The average adjusted R^2 values obtained for each combination with Google News sentiment with all sentiment scores are plotted in Figure 9. The respective plots for Google News sentiment without neutral scores, and for the random data set are provided in appendix B. The average statistically significant adjusted R^2 values from all of the models for each of the time combinations are plotted, with the colour indicating the magnitude of the adjusted R^2 value. On average, Google News with neutral sentiment removed has the best explanatory power (average adjusted R^2 = 0.072) followed by Google News with all sentiment (average adjusted R^2 = 0.070), and then random sentiment (average adjusted R^2 = 0.069). Figure 9 illustrates that the highest adjusted R^2 values tend to occur for long sentiment sampling times such as 40320 minutes (28 days) combined with relatively short lag times, such as 1440 minutes (1 day) to 360 minutes (6 hours), and very short close price sampling times (2 – 60 minutes) suggesting that the sentiment of the last month tends to have a strong influence on the immediate behaviour of a single stock. Further, all of the maximum adjusted R^2 values (except for Pepsi and Coca-Cola) displayed in Table 12 occur when shorter sentiment sampling, lag, and close price sampling times are used (in combination less than four days), indicating that perhaps the strongest effects of the sentiment occurs within this four day period. It should also be noted that a higher volume of significant, but smaller adjusted R^2 values occur for the shorter respective time span combinations. However, this is largely because the time span grid contained a larger number of smaller values due to the limited size of the data set, and due to the finer grid at the lower end of each time dimension. These plentiful yet weak adjusted R^2 values for shorter sentiment sampling periods suggest that the immediate effects of news sentiment are variable, and if a usable relationship exists, it is not a linear one, and OLS regression results should be used with caution for forecasting of stock prices based on shorter sentiment sampling windows.

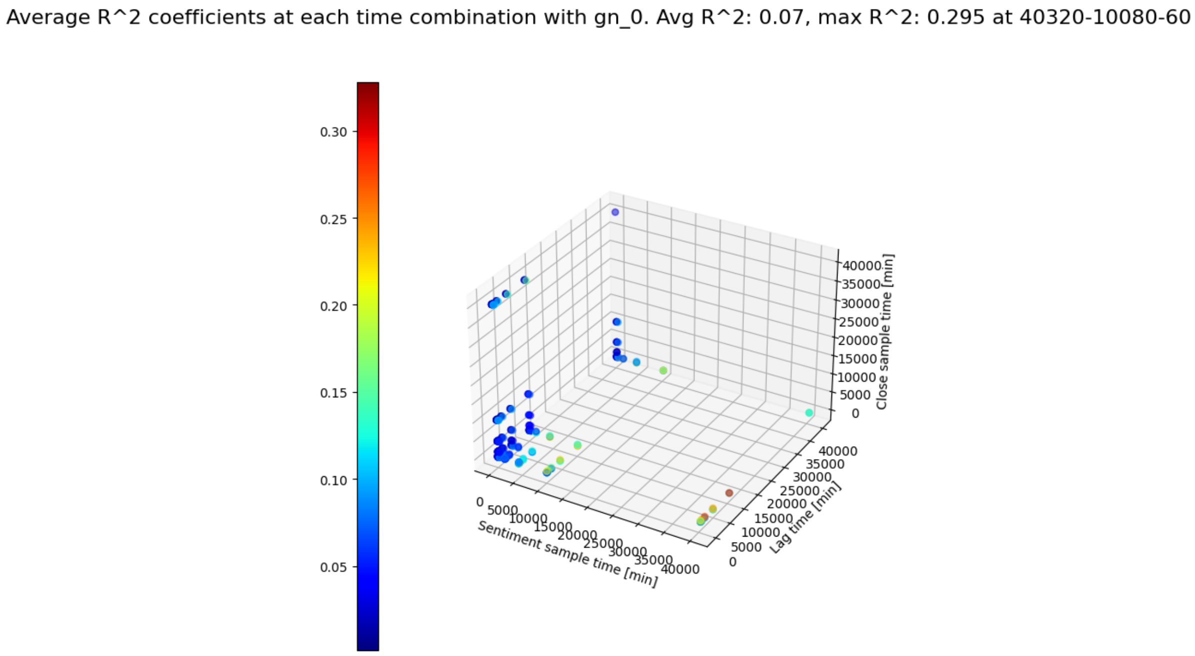

Figure 9: Univariate regression time dependence for Google News sentiment including all scores

4.3.3 Multivariate Analysis

The multivariate regression results for Pepsi and the Coca-Cola company are displayed in Figure 10, with the figures displaying the multivariate analysis results for the remaining pairs provided in appendix C. The figures can be interpreted analogously to the univariate results, with minor changes being that the adjusted R^2 values show the explanatory power between the sentiment pertaining to both stocks in the pair, and either the covariance (cov) of their stock prices, or the average or variance of their stock price ratio (R). Note that some models did not produce statistically significant results and are therefore absent from the plot. The average correlation coefficients, ax1 for stock 1, and ax2 for stock 2, are also provided in the legend with the value for models which failed to yield a statistically significant relationship showing “nan”.

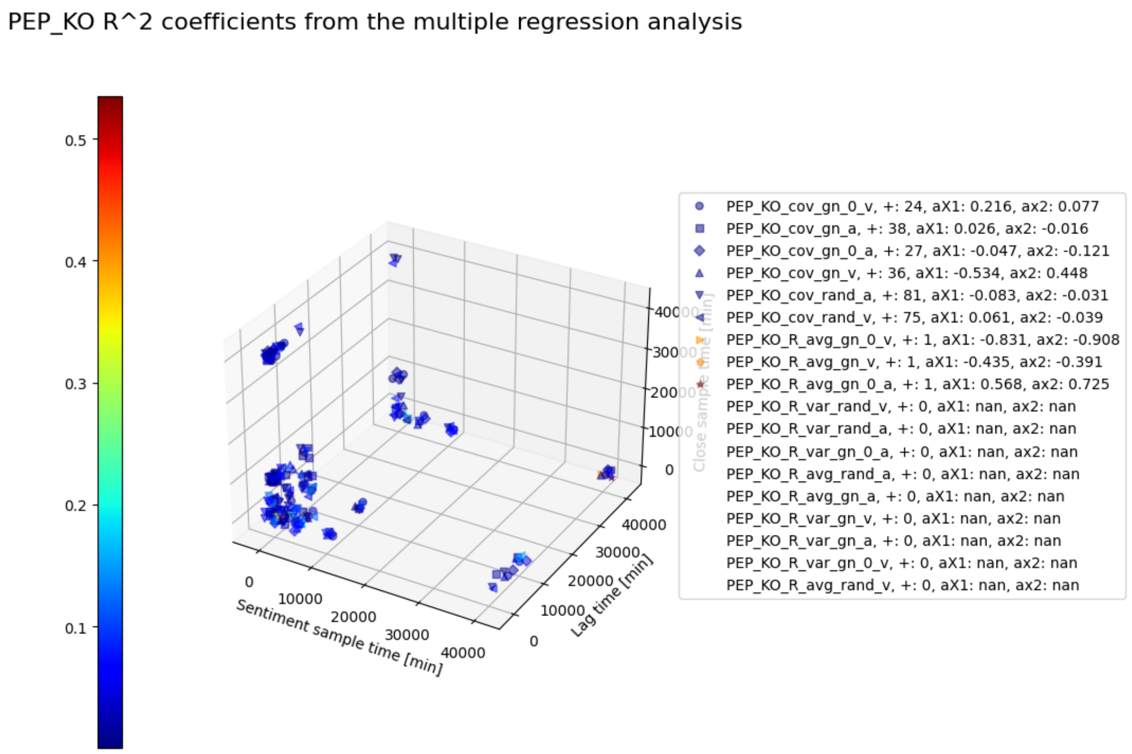

Figure 10: Multivariate regression adjusted R^2 values for PEP_KO

It is observable that in general, there is a stronger relationship between the sentiment pertaining to an individual stock and its own stock price movement than for the sentiment of two stocks and their respective stock price ratio when comparing the maximum adjusted R^2 values of all models. The average adjusted R^2 values of the multivariate regressions are higher, but this is because there are a smaller number of significant adjusted R^2 values with a moderate magnitude, whereas for the univariate case these values were present in addition to many small, yet statistically significant R^2 values. This is likely to be attributable to the increase in complexity of the regression relationship due to the combination of both individual and common behaviour of the two stocks in each pair. For the multivariate case, the models which predicted the stock price covariance with both sentiment metrics consistently returned statistically significant results for all pairs, but the models which predicted the average of the stock price ratio had the highest adjusted R^2 values on average in addition to accounting for 75% of the maximum adjusted R^2 values across all models. Table 13 presents a summary of the results for the multivariate statistical investigation. The temporal distribution of the statistically significant models is similar to the univariate case; and is concentrated in the shorter sentiment sample/ lag/ close time area of the graph, although for 75% of the pairs the maximum R^2 values (displayed in Table 14) occurred for the 40320, 40320, 60 time combination, indicating that for the multivariate case, the effect of sentiment requires longer to manifest.

Table 13: Average adjusted R^2 values and OLS coefficients for the multivariate models

Model Type[12]	**Average Adjusted R^2**	**Average OLS Coefficient 1**	**Average OLS Coefficient 2**
R_var_gn_v	0.0089	0.0003	-0.0005
R_var_gn_0_v	0.0138	-0.0011	0.0005
R_var_rand_a	0.0228	0.0016	-0.0019
R_var_gn_a	0.0278	-0.0004	0.0000
Cov_gn_0_a	0.0313	-1.1505	-0.1677
Cov_gn_a	0.0330	-0.0830	0.0603
Cov_gn_v	0.0441	2.6100	2.7048
Cov_rand_a	0.0462	0.4685	0.3848
Cov_gn_0_v	0.0562	-2.1457	0.9064
R_var_rand_v	0.0626	0.0080	-0.0013
R_var_gn_0_a	0.0632	-0.0012	0.0006
Cov_rand_v	0.0650	-0.1688	-0.3562
R_avg_gn_0_v	0.3153	-1.3778	-2.0618
R_avg_rand_a	0.3472	10.6206	26.7528
R_avg_gn_a	0.3891	-1.6914	0.2691
R_avg_gn_v	0.3918	1.5765	-2.1450
R_avg_rand_v	0.4487	10.0943	12.3986
R_avg_gn_0_a	0.5416	-3.4303	1.7133

[12] The model type code for the multivariate analysis can be understood accordingly: The first term indicates whether the regression focused on R, which stands for stock price ratio, or Cov, which stands for the covariance of the stock prices of the two stocks in the pair. The second term is the aggregation method for the price ratio or covariance (avg = average, var = variance), the third term is the sentiment type (rand = the random data set, gn_0 = Google News sentiment with neutral sentiment, gn = Google News sentiment without neutral sentiment, and the final letter indicates the aggregation method used for the sentiment (a = average, v = variance).

Table 14: Best performing regression models for each stock pair and average adjusted R^2 values for each pair.

Stock Pair	**Model Type**[11]	**Adjusted R^2_{max}**	**t_{sent}, t_{lag}, t_{close}**	**OLS Intcpt**	**OLS Coeff. 1**	**OLS Coeff. 2**	**Adjusted R^2_{mean} (All models)**
PEP_KO	R_avg_gn_0_a	0.5348	40320, 40320, 60	2.8213	0.5678	0.7249	0.1800
DELL_HPQ	R_avg_gn_0_a	0.7010	40320, 40320, 60	3.0824	-11.1632	1.4668	0.1917
UAL_AAL	R_avg_gn_v	0.4194	40320, 40320, 60	4.1070	-1.4451	-2.0710	0.1332
MA_V	COV_rand_v	0.3930	360, 10080, 30	0.2305	-0.4871	-0.2211	0.0676

4.3.4 Hypothesis 1 Assessment

The multivariate results indicate that in general the null hypothesis for hypotheses 1a, 1b, 1c, and 1d for cannot be accepted, as there is statistically significant evidence for an explanatory relationship between both the variance and average values of sentiment of two highly correlated stocks, and the variance and average of their stock price ratio for at least one pair. For both individual stocks and stock price ratios, variances and covariances proved to be less predictable than mean values. This is likely because stock price variance and covariance is a result of uncertainty or unforeseen events, which are very unpredictable, and do not appear to be largely influenced by news sentiment. When each pair was examined individually many of the results differed, proving that the effect of news sentiment is not variable across stock pairs. The results for hypothesis 1 for each pair are displayed in Table 15 and Table 16, showing the results for sentiment with and without neutral scores respectively. It is noteworthy that there was no stock pair for which null hypotheses 1a, 1b, 1c, and 1d could be rejected. Ultimately, most of the observed relationships tend to be weak, and the difference in linear predictive power of real sentiment and randomly generated sentiment is small. The fact that the model which used the variance of randomly generated "sentiment" to predict the average stock price ratio has the second highest average adjusted R^2 value in addition to relatively high OLS coefficients is a good indication that despite the statistical significance of the results at a 95% level, these results should be considered with caution.

Table 15: Hypotheses 1(a-d) assessments for each stock pair using all sentiment

Hypothesis	**PEP_KO**	**DELL_HPQ**	**UAL_AAL**	**MA_V**
1a: R_avg_gn_0_a	Reject null hypothesis	Reject null hypothesis	Reject null hypothesis	Accept null hypothesis
1b: R_var_gn_0_a	Accept null hypothesis	Reject null hypothesis	Accept null hypothesis	Accept null hypothesis
1c: R_avg_gn_0_v	Reject null hypothesis	Reject null hypothesis	Reject null hypothesis	Accept null hypothesis
1d: R_var_gn_0_v	Accept null hypothesis	Reject null hypothesis	Reject null hypothesis	Accept null hypothesis

Table 16: Hypotheses 1(a-d) assessments for each stock pair excluding neutral sentiment

Hypothesis	**PEP_KO**	**DELL_HPQ**	**UAL_AAL**	**MA_V**
1a: R_avg_gn_a	Accept null hypothesis	Reject null hypothesis	Reject null hypothesis	Reject null hypothesis
1b: R_var_gn_a	Accept null hypothesis	Reject null hypothesis	Reject null hypothesis	Accept null hypothesis
1c: R_avg_gn_v	Reject null hypothesis	Reject null hypothesis	Accept null hypothesis	Accept null hypothesis
1d: R_var_gn_v	Accept null hypothesis	Reject null hypothesis	Reject null hypothesis	Accept null hypothesis

4.4 Hypothesis 2

Hypothesis 2 proposed that Google News derived sentiment can be utilised in combination with a linear regression to improve a modified pairs trading strategy to yield increased returns compared to a basic Bollinger Bands-based strategy, contingent to the identification of statistically significant relationships being identified whilst investigating hypothesis 1. Null hypotheses 1a, 1b, 1c, and 1d could not be comprehensively rejected and statistically significant correlations were observed between news sentiment of individual stocks and their respective close prices. Therefore, the effects of using sentiment to improve pairs trading were tested for each of the four stock pairs, and the results of the different strategies are presented as histograms in Figures 12 to 16. The histograms display the profit made by each model over the 21 day trial period with each type of sentiment that was used as an input to the model denoted by the colour of the bar.

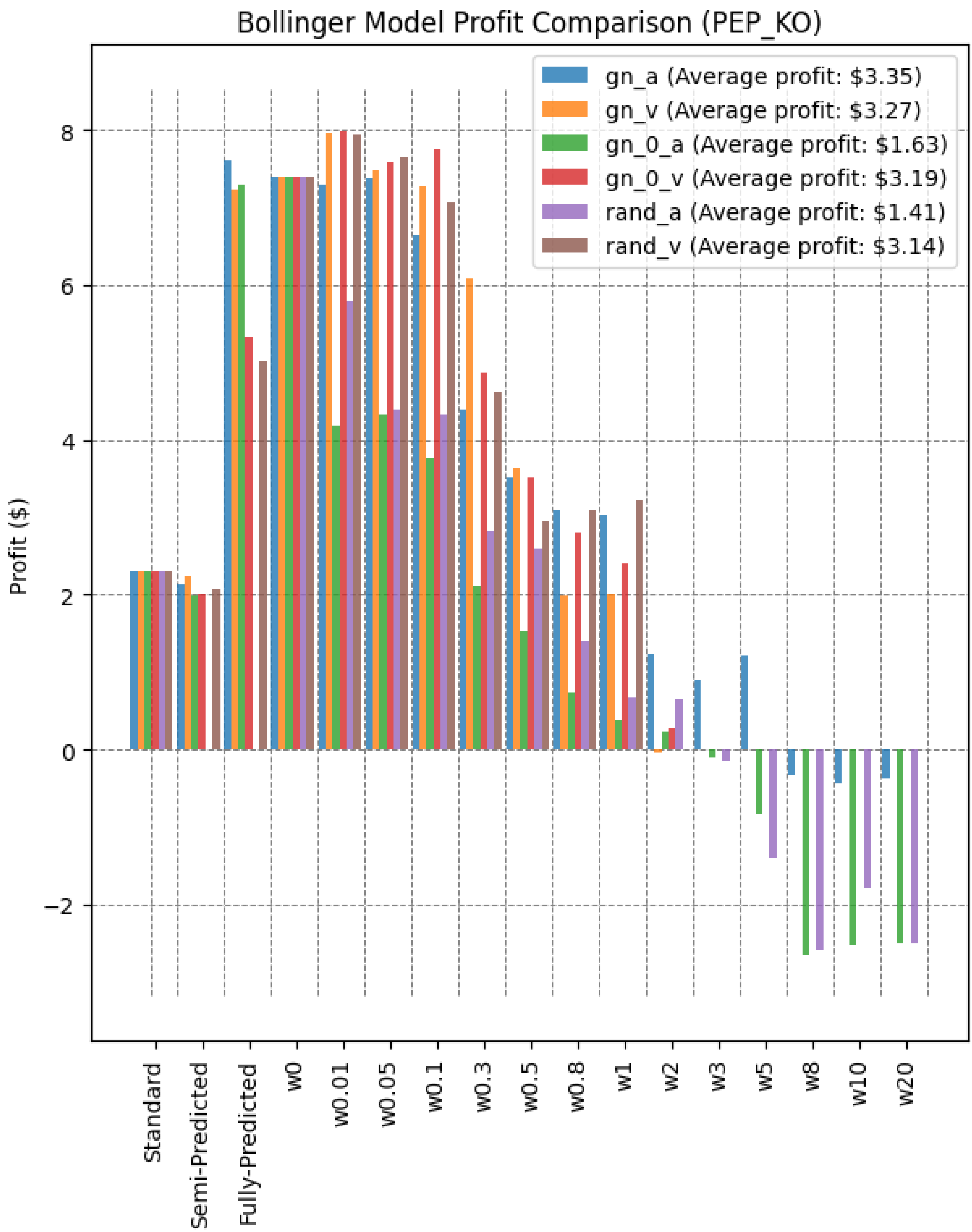

Figure 11: Pairs trading profit comparison PEP_KO

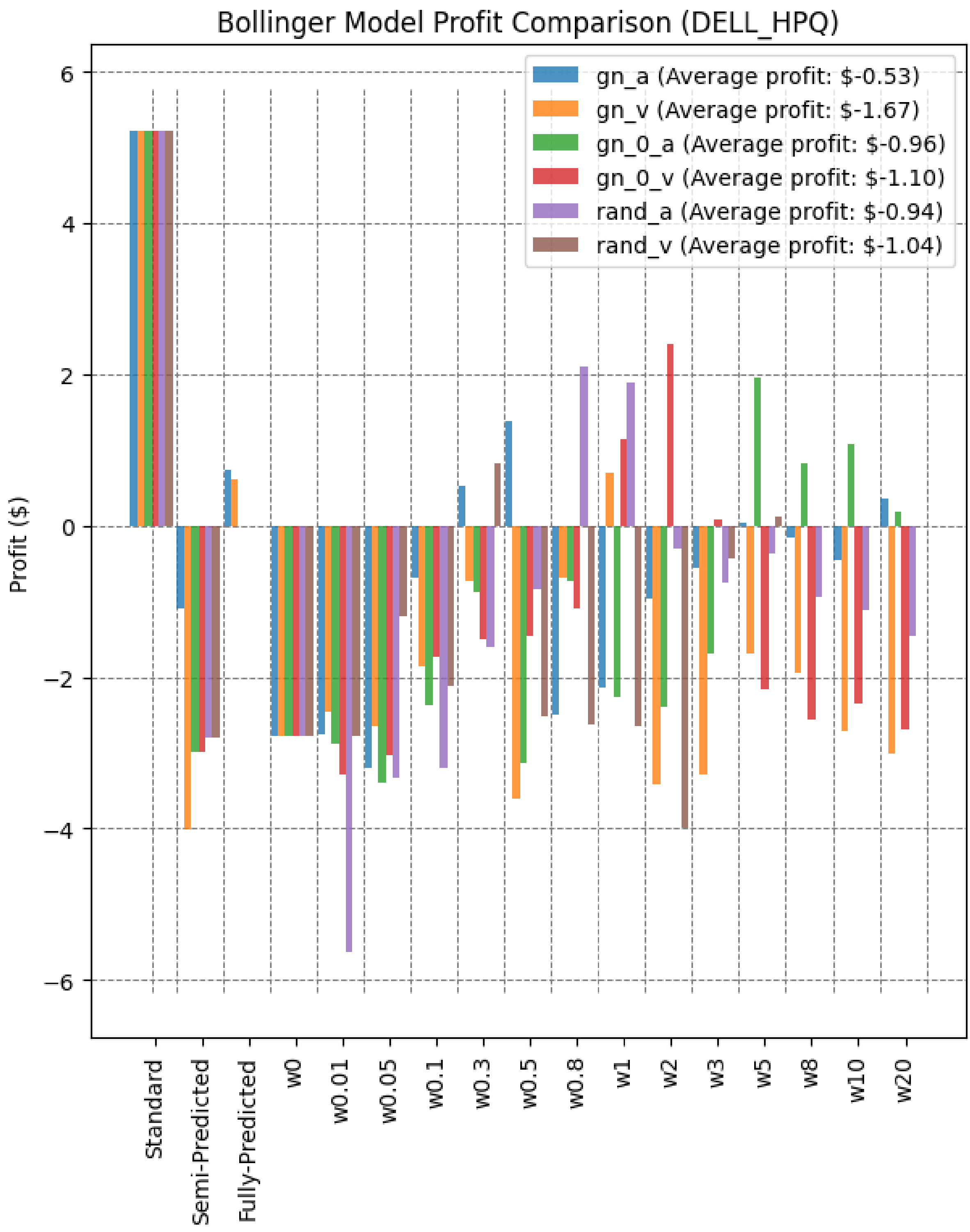

Figure 12: Pairs trading profit comparison DELL_HPQ

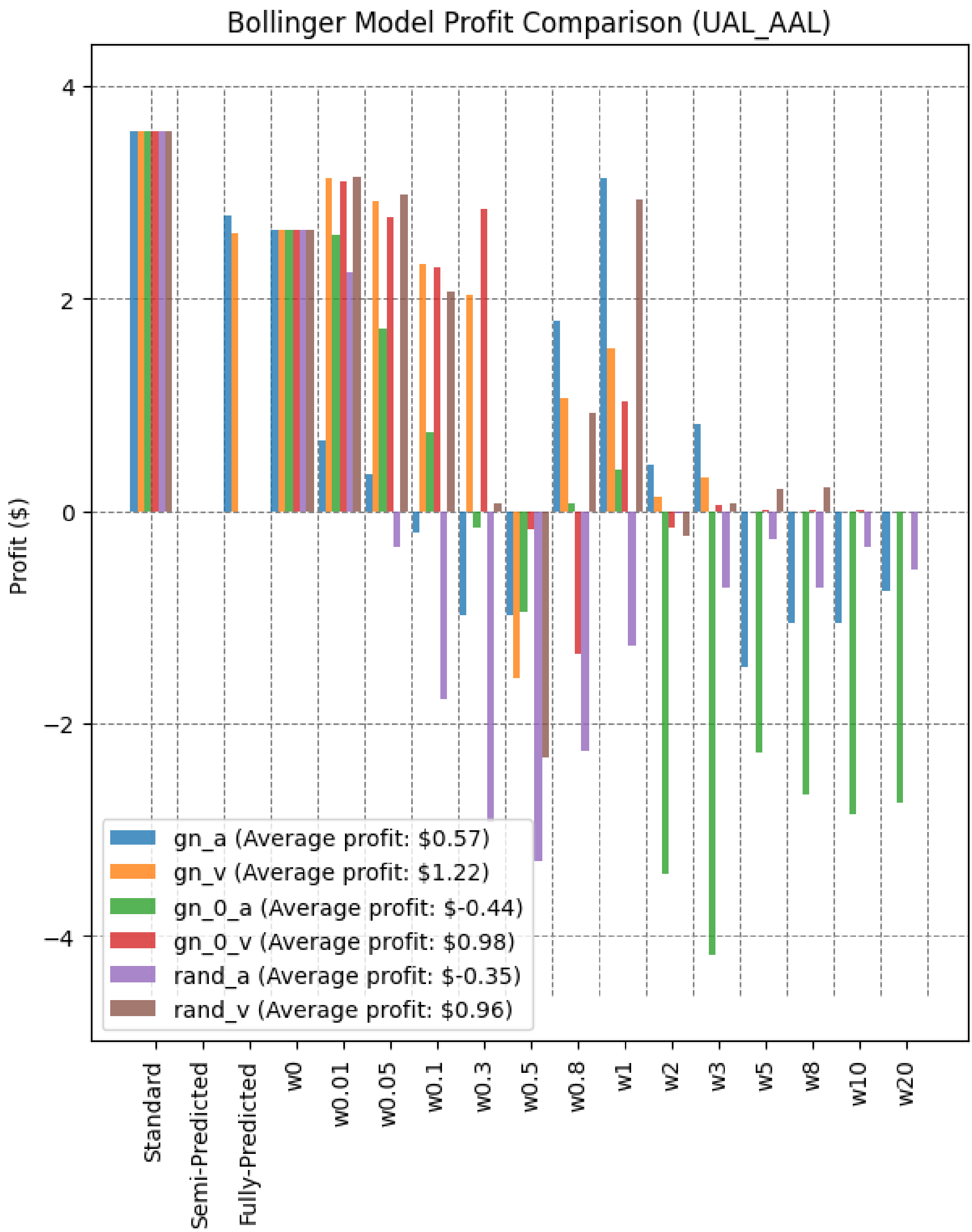

Figure 13: Pairs trading profit comparison UAL_AAL

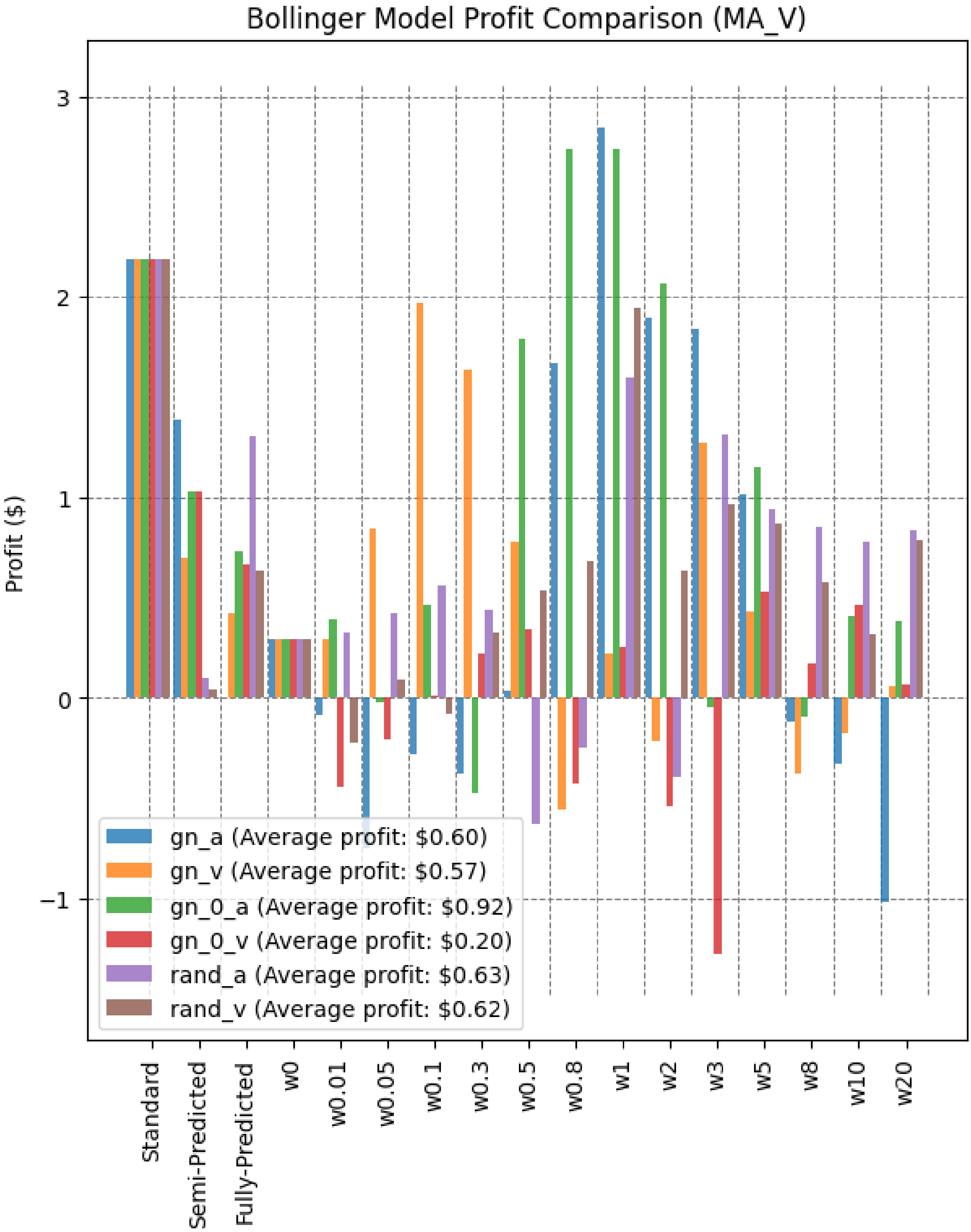

Figure 14: Pairs trading profit comparison MA_V

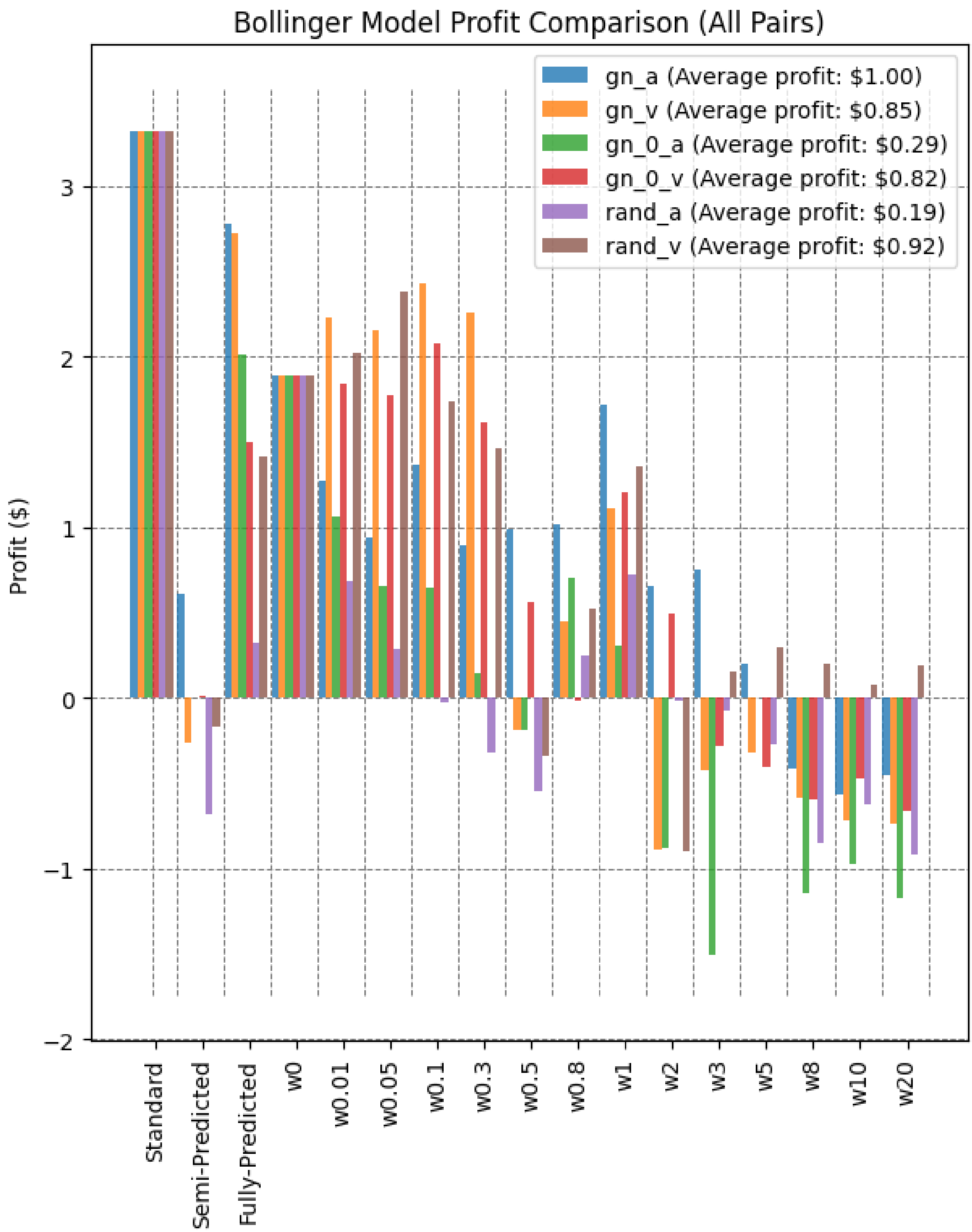

Figure 15: Average profits for the different pairs trading strategies across all pairs

The profit histogram in Figure 15 (average profits) show that the standard Bollinger bands method which does not account for any sentiment effects yields the highest average profits, yielding an average return of $3.33 across the four stock pairs over the 21 day period. Furthermore, it is one of two methods which does not return a loss for any of the four stocks over the 21 day period, in addition to the fully predicted method. The second best returns on average were provided by the semi-predicted model using average Google News sentiment without neutral sentiment scores which provided a $2.78 average return over the 21 day trial period, as displayed in Figure 15.

On average, Google News sentiment without neutral scores generated the highest profits for all of the models which utilised sentiment. Google News sentiment without neutral scores yielded an average return of $1.00 across the 21d period, 8% higher than the next best sentiment type which is the variance of randomly generated sentiment. The average returns of the different sentiment types are included in the legend in Figure 15. On average, Google News sentiment without neutral sentiment scores performed better than Google News sentiment with neutral sentiment scores for both average and variance based methods. But when neutral sentiment was removed, on average, average sentiment outperformed sentiment variance, whereas the opposite was observed when all sentiment was included. These results are aggregated in Table 17 with the sentiment types ordered from most profitable to least profitable for each stock, and the average across all stocks in the right most column. The highest returns overall were observed when the variance of Google News sentiment (with all sentiment scores, gn_0_v) was used for PEP_KO with the weighted model with a weighting of 0.01, generating a profit of $7.99 over the 21 day period with a total of 88 trades. The trades that were made for this particular model are shown in Figure 16 in addition to the stock price ratio and the calculated weighted Bollinger Bands.

An additional important result is that the novel trading strategies yielded very different results for the four pairs. While for PEP_KO the novel methods greatly outperformed the standard Bollinger bands based approach (by as much as 333%, see Figure 11) until a weighting of approximately 1 was used, for DELL_HPQ the standard Bollinger Bands based approach returned a respectable $5.22, while 95% of the novel models generated a loss, and the best performing novel strategy performed approximately half as well as the standard Bollinger Bands based strategy as is evident in Figure 12. Figure 13 and 15 illustrate that for UAL_AAL and MA_V respectively further variation was seen in performance of the novel strategies. From these results, it cannot be concluded with certainty whether any of the novel strategies that returned positive returns on this data set would do this more than 50% of the time on a larger dataset. Therefore, further investigation is recommended.

Table 17: Average returns for each pair with each sentiment type

AVG Return PEP_KO		AVG Return DELL_HPQ		AVG Return UAL_AAL		AVG Return MA_V		AVG Return ALL	
gn_a	$3.35	gn_a	-$0.53	gn_v	$1.22	gn_0_a	$0.92	**gn_a**	**$1.00**
gn_v	$3.27	rand_a	-$0.94	gn_0_v	$0.98	rand_a	$0.63	**rand_v**	**$0.92**
gn_0_v	$3.19	gn_0_a	-$0.96	rand_v	$0.96	rand_v	$0.62	**gn_v**	**$0.85**
rand_v	$3.14	rand_v	-$1.04	gn_a	$0.57	gn_a	$0.60	**gn_0_v**	**$0.82**
gn_0_a	$1.63	gn_0_v	-$1.10	rand_a	-$0.35	gn_v	$0.57	**gn_0_a**	**$0.29**
rand_a	$1.41	gn_v	-$1.67	gn_0_a	-$0.44	gn_0_v	$0.20	**rand_a**	**$0.19**

Table 18: Best performing model and the corresponding profits for each pair

PEP_KO		DELL_HPQ		UAL_AAL		MA_V	
Max. Profit	**Model**	**Max. Profit**	**Model**	**Max. Profit**	**Model**	**Max. Profit**	**Model**
$7.99	gn_0_v, weighted, weight= 0.01	$5.22	Standard Bollinger	$3.58	Standard Bollinger	$2.84	gn_a, weighted, weight=1

Table 19 shows the average number of trades made by the models using each of the different sentiment types as an input. This is an important parameter because overly frequent trading on a real market erodes profit due to transaction costs.

Table 19: Average number of trades using different types of sentiment

Sentiment type	gn_0_a	gn_0_v	gn_a	gn_v	rand_a	rand_v
Average number of trades	154	64	125	109	96	110

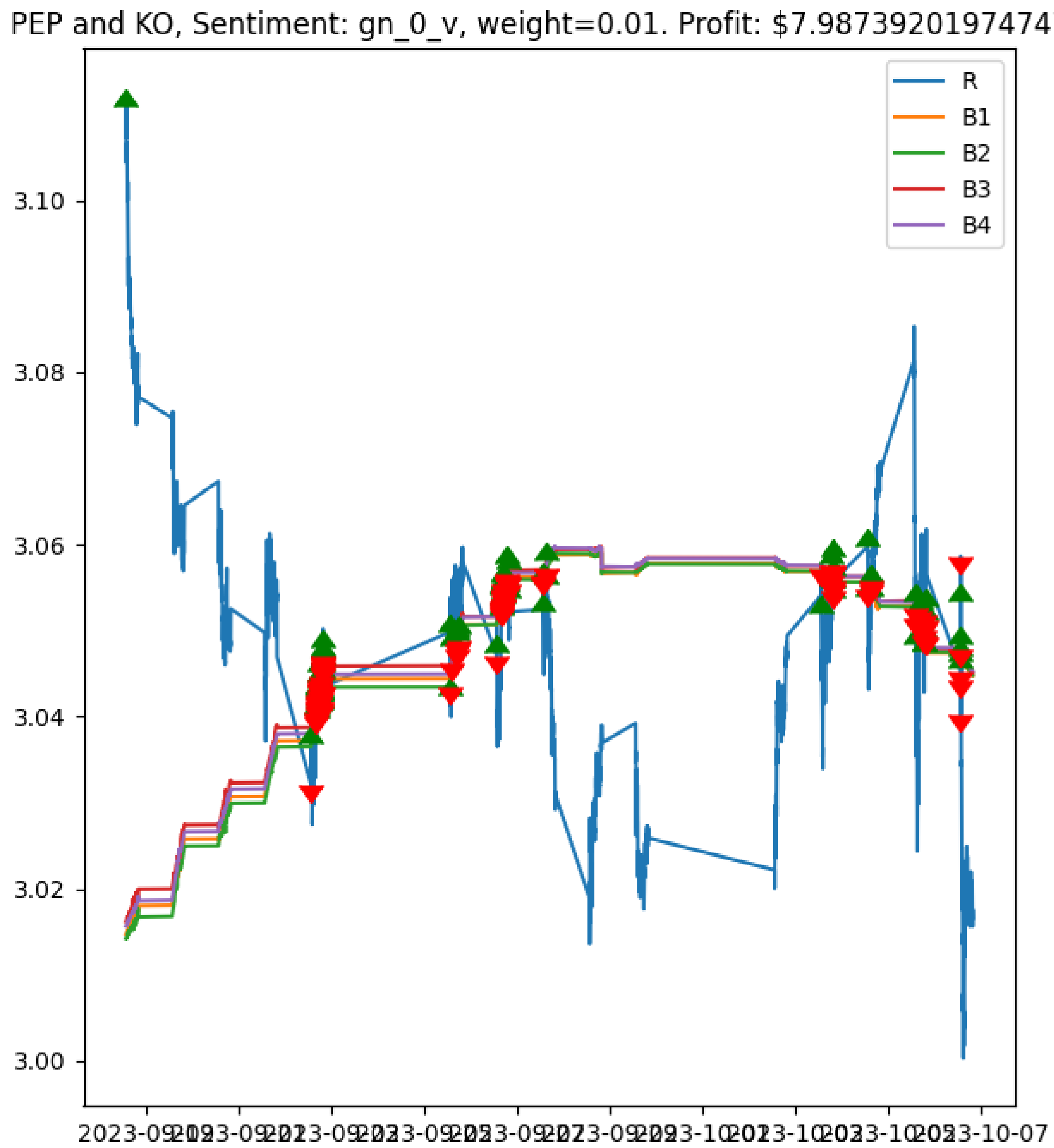

Figure 16: Trades made by the most profitable model

4.4.1 Hypothesis 2 Assessment

These results show that the average and variance of google news sentiment can likely be used in some cases to improve the outcome of a pairs trading strategy, however, the strategies trialled in this study do not all yield improved results compared to a standard Bollinger bands based approach, and on average the standard Bollinger bands approach provides the best returns with the least risk. Additionally, the results show that the models can provide variable profits for each stock pair, and as such each model should be carefully developed for each pair. In spite of this, the highest profit overall was generated by one of the novel strategies (PEP_KO, gn_0_v, $7.99), which suggests that there is potential to improve pairs trading strategies by augmenting them with linear regression models based on news sentiment. However, hypothesis 2 cannot be comprehensively accepted based on these results because the sentiment integrated models did not consistently outperform the standard Bollinger Bands based approach. Furthermore, additional research is required to ascertain whether the different strategies would perform similarly on a data set from a different time period.

4.5 Hypothesis 3

Hypothesis 3 proposed that Google News derived sentiment can improve the predictive accuracy of a LSTM model which predicts the minimum and maximum stock price ratio of a pair of highly correlated stocks for a future time period. This was tested by comparing the average RMSEs of three models for all prediction periods, one which used Google News sentiment, a second which used Google News sentiment with the neutral scores removed, and a third which used the random dataset. The results of the comparisons for each stock pair are presented in Table 20.

Table 20: Average RMSE (absolute value and % of average stock price ratio) of LSTM models for all prediction periods with different sentiment

Sentiment type	**RMSE & %R PEP_KO**	**RMSE & %R DELL_HPQ**	**RMSE & %R UAL_AAL**	**RMSE MA_V & %R**	**RMSE & %R Average**
gn_0	0.02868 (0.95%)	0.01725 (0.85%)	0.07581 (2.33%)	0.02682 (1.61%)	**0.03714 (1.45%)**
gn	0.02512 (0.83%)	0.01929 (0.96%)	0.0808 (2.48%)	0.02587 (1.55%)	**0.03778 (1.52%)**
rand	0.02920 (0.96%)	0.02211 (1.10%)	0.09942 (3.05%)	0.02661 (1.59%)	**0.04433 (1.78%)**

The lowest average RMSE and therefore the most accurate predictions were achieved by the model that included Google News with all sentiment, followed by Google News without neutral sentiment, followed by the random dataset. The random dataset performed worst for every stock pair except for MA_V, for which it produced the second-best results. For DELL_HPQ and UAL_AAL Google News sentiment with all sentiment scores provided the most accurate predictions, while for PEP_KO and MA_V Google News sentiment without neutral sentiment provided the most accurate predictions.

Figure 17 displays the RMSE for the DELL_HPQ LSTM models as a function of the time period for which the predicted minimum and maximum stock price ratio should occur within (the figures for the remaining pairs are located in appendix D). The RMSE tends to increase as the time period increases, and for all pairs except MA_V, random sentiment performs worst. For all of the models the addition of sentiment as a predictor variable increased the LSTM model's predictive accuracy.

Figure 17 and the graphs in appendix D show that for all of the stock pairs, RMSE tends to increase as the time period for the minimum and maximum predictions increases, indicating that the LSTM model's predictive accuracy decreases as the future timespan increases. For PEP_KO this effect was present but less noticeable due to the high variance of the RMSE. DELL_HPQ however, showed clear decreases in model accuracy at two different levels. There is a clear elbow at 1140 minutes (19 hours) at which the RMSE drastically increases, indicating that the model is no longer able to accurately predict the range of the stock price ratio for periods upwards of 1140 minutes. The RMSE graph for UAL and AAL in appendix D shows that the actual Google News sentiment consistently outperformed the randomly generated sentiment for UAL_AAL, however, the best average RMSE for UAL_AAL was 204% higher than the best average RMSE for all stocks, indicating that the LSTM model had the lowest accuracy for UAL_AAL.

4.5.1 Hypothesis 3 Assessment

These results illustrate that including sentiment in an LSTM model for stock price ratio prediction results in a clear decrease in the RMSE which is an indicator of model accuracy, although the strength of this effect varied per observed stock pair. Ultimately, real news sentiment outperformed randomly generated sentiment for all stock pairs, showing that news sentiment does hold predictive information for stock price ratios, and thus hypothesis 3 is accepted.

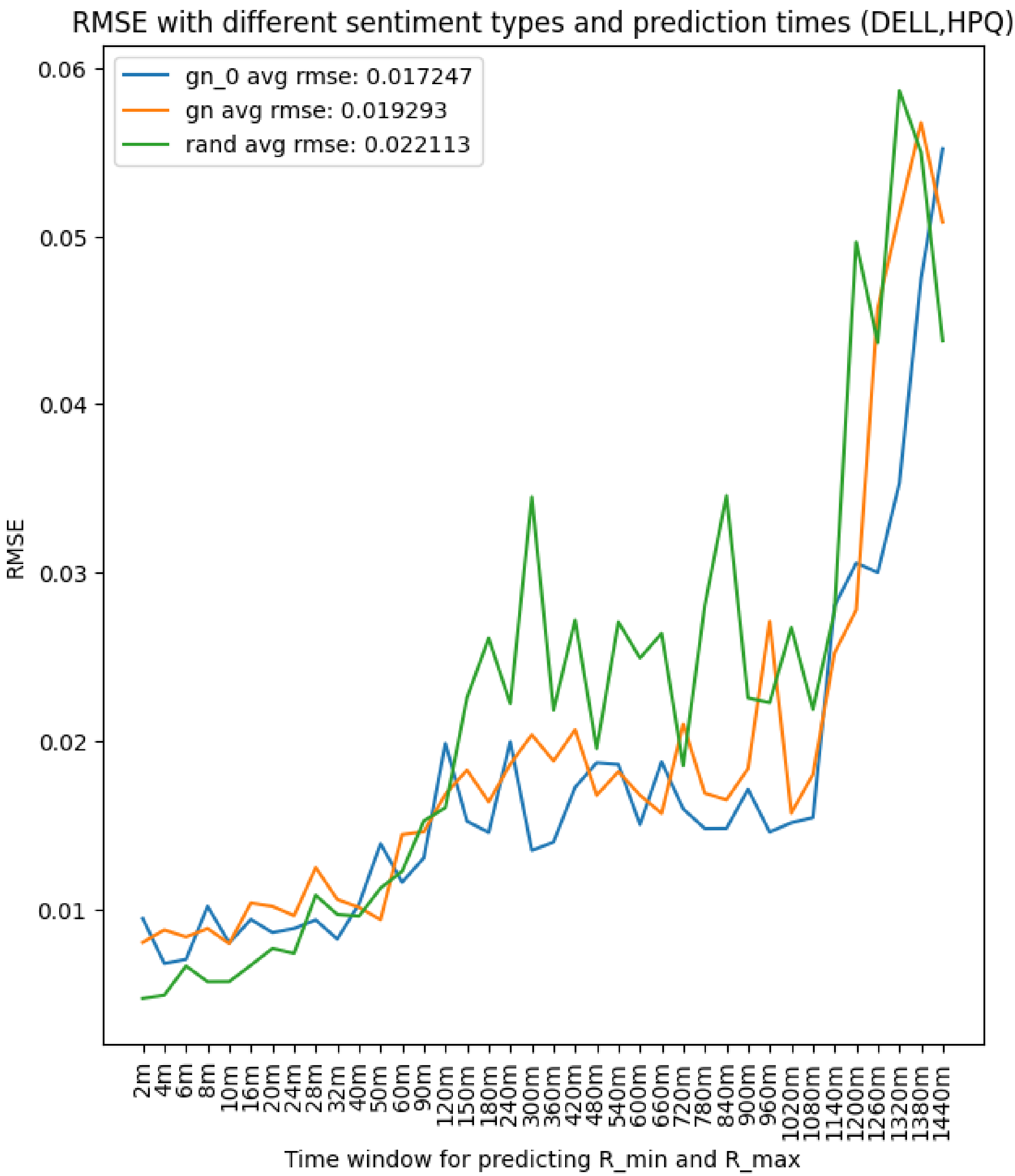

Figure 17: RMSE for the LSTM models with different sentiment types and prediction periods for DELL_HPQ

4.6 Summary

These results show that for individual stocks there is an explanatory relationship between Google News articles that mention the company name in the headline, and the movement of the stock price. The predictive relationship is on average much stronger (approximately five times) between sentiment and the average close price, as opposed to the variance of the close price. Similarly, the relationship between the news sentiment that mentioned either of the stocks in a pair was found to be a much better predictor for the average of the stock price ratio as opposed to the covariance of the two stocks, or the variance of the stock price ratio. There was not a clear pattern observable across all stocks as to whether the sentiment average or the variance was a better predictor, or whether Google News sentiment with or without neutral sentiment was a better predictor. The proposed pairs trading strategies produced variable results across the four different pairs over the 21-day trial period. Some of the models produced respectable profits, of up to $7.99, however, others resulted in significant losses, as low as -$5.60. Consequently, the novel strategy with the best results across the four stock pairs was the semi-predicted method using average Google News sentiment without neutral sentiment, with a profit of $2.78, while the standard Bollinger Bands based approach achieved an average profit of $3.33 across the four stock pairs. The LSTM model's predictive power was improved with the inclusion of Google News sentiment, however, for half of the stock pairs Google News with all sentiment scores resulted in the most accurate predictions, whereas for the other half Google News without neutral sentiment resulted in the most accurate predictions. Ultimately, these results show that news sentiment does contain valuable predictive information about future stock price ratio movements. However, this study was unable to find a way to improve a pairs trading strategy by utilising sentiment data, and further investigation is required to discover how this can be achieved.

5 Discussion

5.1 Theoretical Contributions

The main goals of this study were outlined by hypotheses one through three and focused on studying the relationship between news sentiment and stock price movement and investigating methods of utilising these relationships for pairs trading.

To answer the research question 1 and sub question 1.1, this study has shown that there is a relationship between news sentiment and the stock price (ratio) of two stocks used for pairs trading. The results indicated that both the mean and the variance of the news sentiment can be used as a predictor for the mean stock prices or stock price ratio, and the sentiment sampling time, lag time, and close price sampling time are very influential on the strength of these relationships. Both positive and negative relationships were observed, and the strengths and polarities of the observed relationships varied per stock and stock pair.

To answer research sub-question 2.1, this study did not find a reliable method of using sentiment data to consistently improve the returns of a pairs trading strategy. Although some of the novel pairs trading strategies that used statistical predictions based on sentiment data did return higher profits than the Bollinger Bands based approaches, many of the approaches produced lower returns and there was no clear pattern observed regarding configurations that consistently outperformed the Bollinger Bands based approaches.

To answer research sub-question 2.2, an LSTM model which was developed to predict the minimum and maximum stock price ratio values in a specified future time period did benefit from the addition of sentiment data, and consistently produced more accurate predictions (lower RMSE) than the control model which used a random data set instead of sentiment. However, this study did not find a way to utilise these predictions such as to improve the returns of a pairs trading strategy relative to the Bollinger Bands based approach.

The results of this study shows that sentiment is a predictor for the stock price ratio, however, further research is required to determine how this information can be utilised to increase the profitability of pairs trading strategies, or at the least, lower the risk.

5.1.1 H1: Sentiment and Price Ratio

Conducting linear regressions between Google News sentiment and stock close price or price ratios revealed that there are linear relationships present, however, these relationships vary considerably depending on the time periods used for sampling and lag and for the different stocks. The strongest relationships were found for predicting average close price or price ratios, and close price or price ratio variance were much less explainable via the regression models. Specifically, the average adjusted R^2 magnitudes for the variance prediction regressions were all less than half those of average prediction regressions.

5.1.1.1 H1(a): Average Sentiment – Average Price Ratio

For individual stocks, the OLS regression between the average of Google News sentiment with and without neutral sentiment and the close prices of individual stocks returned average adjusted R^2 values of 0.102 and 0.101 respectively, with the maximum adjusted R^2 value returned for DELL, at 0.9688, using the data set that excluded neutral sentiment. The average OLS coefficients for both types of sentiment models were negative, indicating that on average, stock price tends to decrease slightly with positive sentiment, which is in agreement with Frino et al, (Frino et al., 2022), but in disagreement with Deveikyte et al (Deveikyte et al., 2022). However, both positive and negative OLS coefficients were observed for the models which used average sentiment as a predictor for average stock price, validating both the results of Frino et al (Frino et al., 2022) and Deveikyte et al (Deveikyte et al., 2022).

Furthermore, these results demonstrate that the correlation between sentiment and stock price is also highly dependent on the stock, the sentiment sampling time, lag time, and close price sampling time.

For the models that predicted the average stock price ratio as a function of the average sentiment of two stocks, the OLS regression models that used average Google News sentiment with and without neutral sentiment returned average adjusted R^2 values of 0.5416 and 0.3891 respectively, with the highest adjusted R^2 value returned for DELL_HPQ, at 0.7010 using the data set that included neutral sentiment. There was a negative and positive average OLS coefficient for both models, however, this was not always the case for the other pairs, highlighting again the dependence of the model on the sampling periods, the lag time, and the specific stocks.

This study has shown that in addition to being a useful predictor for individual close prices, the sentiment from two stocks can also be used to predict the stock price ratio, however, the linear relationships are often relatively weak, and results vary across stocks.

5.1.1.2 H1(b): Average Sentiment – Variance of Price Ratio

For the models that predicted the variance of the stock close price as a function of the average sentiment of the stock, the OLS regression models that used average Google News sentiment with and without neutral sentiment returned average adjusted R^2 values of 0.0312 and 0.0244 respectively, with the highest adjusted R^2 value returned for HPQ, at 0.1774, with neutral sentiment included.

For the models that predicted the variance of the stock price ratio as a function of the average sentiment of two stocks, the OLS regression models that used average Google News sentiment with and without neutral sentiment returned average adjusted R^2 values of 0.0632 and 0.0278 respectively, with the highest adjusted R^2 value returned for DELL_HPQ, at 0.0632 with neutral sentiment included.

The small magnitudes of the adjusted R^2 values illustrate that average sentiment score is a poor predictor of close price and stock price ratio, at least when linear methods are used. Zhang et al found that sentiment is a predictor of market volatility for developing markets, however, the relationship was found to be non-linear, which may also be the case even a developed market such as the American one, where the S&P 500 is located. This would explain the failure to observe a strong relationship between sentiment and close price or stock ratio via a linear OLS regression.

5.1.1.3 H1(c): Variance of Sentiment – Average Price Ratio

For the models that predicted the average stock close price as a function of the variance of the sentiment of the stock, the OLS regression models that used the variance of Google News sentiment with and without neutral sentiment returned average adjusted R^2 values of 0.0971 and 0.109 respectively, with the highest adjusted R^2 value returned for AAL, at 0.9693, using the data set that excluded neutral sentiment.

For the models that predicted the average stock price ratio as a function of the variance of the sentiment of two stocks, the OLS regression models that used the variance of Google News sentiment with and without neutral sentiment returned average adjusted R^2 values of 0.3153 and 0.3918 respectively, with the highest adjusted R^2 value returned for UAL_AAL, at 0.4194 excluding neutral sentiment.

These findings illustrate that for some time combinations and stocks, the variance of the sentiment is a useful predictor for the close price or stock price ratio. These findings are a key contribution of this work, as no other study was found that examined the relationship between the variance of the sentiment's effect on stock price, or stock price ratio. However, again, substantial variance was observed depending on the sampling and lag times used, and across the different stocks.

5.1.1.4 H1(d): Variance of Sentiment – Variance of Price Ratio

For the models that predicted the variance of the stock close price as a function of the variance of the sentiment of the stock, the OLS regression models that used the variance of Google News sentiment with and without neutral sentiment returned average adjusted R^2 values of 0.0309 and 0.0420 respectively, with the highest adjusted R^2 value returned for DELL, at 0.4861, excluding neutral sentiment.

For the models that predicted the variance of the stock price ratio as a function of the variance of the sentiment of two stocks, the OLS regression that used the variance of Google News sentiment with and without neutral sentiment returned average adjusted R^2 values of 0.0138 and 0.0089 respectively, with the highest adjusted R^2 value returned for DELL_HPQ, at 0.02466 with neutral sentiment included.

These findings align with the findings of hypothesis 1b, namely that close price or stock ratio variance is more difficult to predict than the respective average values.

5.1.1.5 H1 Summary

Overall, the key contributions of hypothesis 1 are the demonstration that the relationship between sentiment and close price or stock ratio are very variable depending on the stock, and the sentiment sample time, the lag time, and the close price or stock ratio sample time. This aligns with the findings of Hsu et al, which were that both current and lagged sentiment effect the market volatility (Hsu et al., 2021). This study adds that these effects are very variable depending on the lag time and the individual stock, and that the linear relationship between sentiment and stock variance (volatility) is very weak, suggesting that if a relationship is present, it is likely to be non-linear as per Zhang et al (W. Zhang et al., 2021).

The study could not find a clear difference in linear performance of sentiment with and without neutral scores across all stocks, and clear differences were observed between different stocks (as reported by Muguto et al (Muguto et al., 2022)) and stock pairs.

Ultimately, both average and variance of sentiment were found to be moderate linear predictors of average close price or average stock price ratio, but not variance or covariance. The optimal sentiment sampling time, lag time, and close price or stock price ratio sampling time varied between each stock or stock pair, but a trend was observed that longer sentiment sampling periods (28 days) paired with shorter close price (ratio) sampling times (1 hour for individual stocks, 6 hours for stock pairs) had higher adjusted R^2 values on average. However, for 75% of the individual stocks, the strongest effects were observed when the total time for the combined sentiment sampling, lag, and close sampling time was less than four days. Whereas for the stock price ratio for 75% of the pairs the strongest effects were observed for sentiment sampling and lag times of 28 days combined with a six-hour close price sampling period. The maximum adjusted R^2 values for the stock price ratio regressions were on average 43% lower than those for the individual stocks.

The observed variance across the pairs is likely due to different market dynamics of the four GICs that the four stock pairs belong to. This is consistent with the findings of Muguto et al, who found that the strength of sentiment effects differed across different sectors (Muguto et al., 2022), and Muthivhi and van Zyl, who found that the VADER model's ability to accurately classify sentiment varied for different companies (Muthivhi & van Zyl, 2022). Interestingly, the highest R^2 values were obtained for DELL_HPQ, which had the lowest historical correlation coefficient magnitude.

Attention should be given to some of the high adjusted R^2 values found using the random data set. This is unexpected, and the spurious correlations could be caused by several reasons, including non-stationarity of the data, which this study did not test for. Alternatively, it could be due to instances of insufficient sample size and the model achieving a fit with the random data. This is an issue which requires further attention in future studies.

5.1.2 H2: Sentiment Assisted Pairs Trading

Varying levels of success were observed for the three different sentiment integrated pairs trading strategies that were trialled in this study. The standard Bollinger Bands-based strategy returned the most consistent results and generated the highest average profit at $3.33 across the 21-day trial period and yielded no overall losses for each pair. The overall highest returns were observed when the variance of Google News sentiment with all sentiment scores was used for PEP_KO with the weighted model with a weighting of 0.01, which yielded a profit of $7.99 over the 21 day period with a total of 88 trades. However, these results were not reflected across the other stocks pairs, where the standard Bollinger Bands based model consistently outperformed the aforementioned model. In 50% of the stock pairs investigated (PEP_KO & MA_V) one of the weighted strategies returned the best profit overall, whereas for the other 50% (DELL_HPQ & UAL_AAL) the Bollinger Bands based strategy returned the highest profit. This echoes the findings of Koratamaddi et al, who found that inclusion of sentiment increased portfolio returns at the cost of increassed volatility. (Koratamaddi et al., 2021). Furthermore, the inconsistency of the proposed model's returns could be a demonstration of the varying strength of the relationship between sentiment and stock price movement for different stocks as identified by Muguto et al (Muguto et al., 2022). However, it is more likely to be because the novel strategies are not well optimised for pairs trading, because the stocks which showed the highest average and maximum R^2 values (DELL_HPQ & PEP_KO) are the stock pairs which performed best with the new models. This suggests that if the predictive power of the sentiment is more accurate, the novel strategies are more profitable. The sentiment augmented pairs trading models are novel strategies, and in this study the novel strategies did not outperform the standard Bollinger Bands based method. However, the results suggest that further investigation is warranted to determine more effective methods of using sentiment for pairs trading.

5.1.3 H3: Sentiment assisted LSTM Prediction

Ultimately, the inclusion of Google News sentiment in the LSTM model increased the accuracy of the prediction of the stock price ratio. On average, the data set that included neutral sentiment yielded the greatest increase in accuracy, providing an average RMSE of 1.45% as opposed to 1.52% when neutral sentiment was removed, and 1.78% when random “sentiment” was used as a control input. Although Google News with neutral sentiment included performed best on average, for half of the pairs, Google News sentiment without neutral sentiment provided the best performance. This is contradictory to the findings of Muthivhi and van Zyl, who found that the inclusion of neutral sentiment had no effect, (Muthivhi & van Zyl, 2022). To the contrary, all of the LSTM models with and without neutral sentiment displayed small differences in accuracy, with the average absolute difference being 0.11%. A notable case was the results of the LSTM model predicting the ratio of UAL_AAL. The RMSEs for this model were the highest, indicating that the model struggled to predict the stock price ratio, however, the inclusion of sentiment also provided the greatest improvement, with Google News sentiment with neutral scores included providing a 0.72% decrease in RMSE compared to the control model. Although not as large as the 3.67% improvement reported by Zhang et al (W. Zhang et al., 2021), a 0.72% improvement is still significant for arbitrage trading applications such as pairs trading. These results suggests that the other predictors are likely stronger than sentiment, but that including sentiments yields the greatest improvements when the stock price ratio prediction becomes more difficult. The RMSE showed a positive trend with increasing time window, indicating that the LSTM model performs better at predicting stock price ratios closer in the future, and suggesting that the predictive power of sentiment data is most valid immediately after and on the same day that the news article is published.

These investigations confirmed the findings that LSTM is an effective tool for stock price prediction (Cao et al., 2020; Du, 2022; Sen et al., 2021; Singh et al., 2021; Touzani et al., 2019; Yao et al., 2022), and extended the findings that sentiment can improve the prediction of stock price (Dutta et al., 2021; Jain et al., 2022; Li, 2022; Owen & Oktariani, 2020; Sen et al., 2021; Zhou et al., 2023) to the prediction of stock price ratio. Another important contribution is that for half of the pairs, the model accuracy was improved by removing neutral sentiment, while for the other half this was not the case.

5.2 Managerial Implications

These findings are relevant for all entities with an interest in exploiting the market neutral arbitrage strategy of pairs trading. A key finding of this study is that the effects of news sentiment, gathered by assessing the sentiment of Google News headlines, are highly variable between companies, but also temporally, with the sentiment sampling time span, the lag time, and the time span of the effect on the stock prices also varying greatly. Thus, prior to utilising sentiment data for pairs or other types of trading, a thorough assessment of the relationship between news sentiment and close price or stock ratio is required for each stock or stock pair for different sampling periods and lag times.

It should also be noted that some stock predictions improved with the removal of neutral sentiment, while others did not, but ultimately, the difference between models that included neutral sentiment and models which did not was small. Again, to capitalise on this factor an investigation for each stock or stock pair is required, and when trading large volumes of stock, the increased arbitrage potential may justify the investigation.

Statistically significant linear relationships between sentiment and stock price or stock price ratio were observed, however, the explanatory power of linear models such as OLS were generally weak. The strongest relationships occurred for 75% of the individual stocks when the total sentiment sampling, lag, and close price sampling time was four days or less, whereas for stock price ratios, the strongest relationships (albeit 43% weaker) were observed for 75% of the pairs when both the sentiment sampling and lag time was 28 days, and the close price ratio sampling period was one hour. This highlights the importance of using different models to predict a stock price ratio as opposed to individual stock prices.

The strongest regression relationships used either sentiment variance or sentiment average as the predictor for the mean stock price or stock price ratio. The variance of the stock price or stock price ratio, or the covariance of the two stocks in a pair was found to be difficult to predict using sentiment. For 75% of the individual stocks, the variance of the sentiment was responsible for the strongest respective relationships, whereas for the stock price ratios half of the strongest relationships used the variance of the sentiment while the other half used the average sentiment. Again, highlighting the need for individual investigations.

An additional important finding is that although it is difficult to capture the relationship between sentiment and stock price or stock price ratio using linear statistical methods, Google News sentiment consistently improved the predictive accuracy of a LSTM network tasked with predicting the minimum and maximum stock price ratio in a prescribed future time span, but again, the time period for which these predictions remained accurate varied for each stock pair. These improvements demonstrated the potential deep learning models such as LSTM have for predicting stock price ratios and suggest that there is potential for increased arbitrage. The predictive accuracy of the model could likely be improved if a complete optimisation study is conducted using one of the prominent open source optimisation modules such as Optuna (Akiba et al., 2019).

It should be noted that this study has been conducted without consulting the laws and regulations surrounding the use of AI technologies in the EU, or other parts of the world. Subsequently, it is recommended to ensure all compliances with AI laws are met before using the findings of this study to trade on any real stock markets. Further, the terms of use for the open-source python packages should be examined to ensure compliance, however, for most commercial applications this is unlikely to be problematic as python is also free to use for commercial applications.

Although none of the novel pairs trading strategies proposed in this study outperformed the standard Bollinger Bands based approach, the discovery of the presence of predictive relationships between news sentiment and stock price ratio suggest that there is potential to increase the profitability of pairs trading strategies, or at a minimum, reduce risk. The findings of this study imply that further investigation of augmenting pairs trading strategies with sentiment data could lead to higher profitability, especially for the first traders to discover a functioning strategy. Therefore, investing in researching methods of implementing sentiment data or sentiment-based predictions into pairs trading strategies is recommended. Experimenting with non-linear regressions or using stricter filtering of regression results such as a lower p value, or only making predictions with regressions with an adjusted R^2 value above a certain threshold is suggested for the statistics-based approaches, and an optimisation study for the LSTM model would also likely improve the performance of the corresponding trading strategy. Experimenting with tuning the parameters of the models presented in this study could lead to the discovery of a superior pairs trading model. Alternatively, additional novel pairs trading strategies to implement sentiment aided predictions could be developed.

5.3 Limitations

Because the number of combinations of time variables, stocks, sentiment metrics, close price metrics, trading models, and other variables is essentially infinite, it was impossible to study every possible combination. Undoubtedly, there are additional time combinations that this study did not trial that could yield positive results for different stocks. The statistical analysis was limited to linear regressions; however, it appears likely that a portion of the effect of sentiment on stock price and stock price ratio may be non-linear. Further, the results of the statistical analysis may be improvable by decreasing the acceptable p value to 0.01. Using a

99% significance level could help to further distinguish the effects of real sentiment and random sentiment.

This was a quantitative study that used a dataset spanning 105 days, with 84 days used for the regression and the LSTM model training. The size of the dataset restricted the investigation to shorter sampling time periods and lag times and meant that the trial period for the different pairs trading strategies was 21 days, which limits the generalisability of these results. The LSTM model may have performed better when trained on a larger data set, for example a data set spanning an entire year (Shastri et al., 2018).

Another factor which restricts the generalisability of these results is that a total of eight stocks and four pairs were used, all of which are large market capitalization stocks that are listed on the S&P 500. Furthermore, the company name of each stock was the sole keyword used to query Google News and extract sentiment, and it is likely that there are more keywords that would provide useful sentiment for each stock or stock pair.

5.4 Further Research

From these results it is evident that sentiment analysis has the potential to improve pairs trading returns, with proven predictive power. However, further research is required to improve these predictive models, and to develop effective sentiment aided pairs trading models. To improve the accuracy of statistical models, using a smaller p value, for example 0.01, may help to filter out unwanted coincidental fits, in addition to experimenting with non-linear regression models (W. Zhang et al., 2021), and refining the time dimension grid around the values which showed the best fits. Separating sentiment into positive and negative datasets has also been shown in some studies to be beneficial (Deveikyte et al., 2022). Additionally, the close price data sets should be tested for stationarity using tests such as the Augmented Dickey-Fuller (ADF) or KPSS test to determine if non-stationarity is to blame for the statistically significant relationships observed between some of the random data sets and the close prices. In the case that non-stationarity is present in the data set, techniques such as differencing are recommended for preprocessing the data, or a statistical method such as ARIMA which is not sensitive to non-stationarity.

A qualitative study with a data set spanning several years is also recommended to better understand the relationship between news sentiment and stock pairs and improve generalisability. A larger data set may also help prevent spurious correlations based on random data from occurring due to insufficient sample size. A qualitative study would, however, require more time than is available for a master's thesis study such as this one. Generalisability could also be improved by investigating a larger number of stock pairs; however, the number required for a qualitative study is likely to be several orders of magnitude higher than the number of stock pairs investigated in this quantitative study. Furthermore, it could be beneficial to focus on investigating whether sentiment has a similar effect on stock pairs from companies within the same GIC, or the effects of sentiment pertaining to lessor known companies than those listed on the S&P 500, or companies in different geographical regions. Expanding the key words used when searching for news headlines that are relevant to each stock pair could also be beneficial, in addition to determining which additional key words, topics, or events yield influential sentiment for the models.

For the LSTM model predictions, further improvements in accuracy can be made by further tuning the hyperparameters and additional model settings using an optimisation tool such as Optuna (Akiba et al., 2019).

For the novel pairs trading strategies, it may be beneficial to exclude some of the weaker regression models when calculating the average predictions or the weights when using the weighted approaches, as these weak yet numerous relationships may be diluting the predictive accuracy of the other stronger regression models. The low average R^2 values compared to the maximum adjusted R^2 values that were observed in the regression analysis indicate that this may be the case.

6 References

Agarwal, S., & Muppalaneni, N. B. (2022). Portfolio optimization in stocks using mean–variance optimization and the efficient frontier. *International Journal of Information Technology*, *14*(6), 2917–2926. https://doi.org/10.1007/s41870-022-01052-2

Akiba, T., Sano, S., Yanase, T., Ohta, T., & Koyama, M. (2019). Optuna: A next-generation hyperparameter optimization framework. *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/https://doi.org/10.1145/3292500.3330701

Alomari, M., Al Rababa'a, A. R., El-Nader, G., Alkhataybeh, A., & Ur Rehman, M. (2021). Examining the effects of news and media sentiments on volatility and correlation: Evidence from the UK. *The Quarterly Review of Economics and Finance*, *82*, 280–297. https://doi.org/10.1016/j.qref.2021.09.013

Aroussi, R. (2019). *yfinance*. GitHub. https://github.com/ranaroussi/yfinance

Ayyappa, Y., & Siva Kumar, A. P. (2022). Optimized long short-term memory-based stock price prediction with sentiment score. *Social Network Analysis and Mining*, *13*(1), 13. https://doi.org/10.1007/s13278-022-01004-5

Bharathi, S., & Geetha, A. (2017). Sentiment analysis for effective stock market prediction. *International Journal of Intelligent Engineering and Systems*, *10*(3), 146–154. https://doi.org/10.22266/ijies2017.0630.16

Bugara, A. (2020). *pygooglenews*. GitHub. https://github.com/kotartemiy/pygooglenews

Cao, H. K., Cao, H. K., & Nguyen, B. T. (2020). DELAFO: An Efficient Portfolio Optimization Using Deep Neural Networks. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12084 LNAI*, 623–635. https://doi.org/10.1007/978-3-030-47426-3_48

Caporin, M., & Poli, F. (2017). Building News Measures from Textual Data and an Application to Volatility Forecasting. *Econometrics*, *5*(3), 35. https://doi.org/10.3390/econometrics5030035

Chen, J., James, M., & Li, T. (2021). *Mean-Variance Analysis: Definition, Example, and Calculation*. Investopedia. https://www.investopedia.com/terms/m/meanvarianceanalysis.asp

Chen, J., & Scott, G. (2023). *What Is Mean Reversion, and How Do Investors Use It?* Investopedia. https://www.investopedia.com/terms/m/meanreversion.asp

Cheng, A. (2024, November 26). SL Green to buy former Pepsi headquarters building on Manhattan's Park Avenue. *Costar News*. https://www.costar.com/article/41006820/sl-green-to-buy-former-pepsi-headquarters-building-on-manhattans-park-avenue

Dahal, K. R., Pokhrel, N. R., Gaire, S., Mahatara, S., Joshi, R. P., Gupta, A., Banjade, H. R., & Joshi, J. (2023). A comparative study on effect of news sentiment on stock price prediction with deep learning architecture. *PLoS ONE, 18*(4 April), 1–19. https://doi.org/10.1371/journal.pone.0284695

Dam, J. G. (2023). *CEO's tweets and firm stock returns: A case study of Elon Musk and Tesla*. https://digitalcommons.georgiasouthern.edu/honors-theses/850/

Deveikyte, J., Geman, H., Piccari, C., & Provetti, A. (2022). A sentiment analysis approach to the prediction of market volatility. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.836809

Ding, J. (2024, November 1). Coca-Cola and Pepsi, two of the world's top plastic polluters, just got slapped with a lawsuit alleging exaggerated recycling claims and downplayed health effects. *Associated Press*. https://fortune.com/2024/11/01/coca-cola-pepsi-plastic-pollution-lawsuit-los-angeles-county/

Du, J. (2022). Mean–variance portfolio optimization with deep learning based-forecasts for cointegrated stocks. *Expert Systems with Applications*, *201*, 117005. https://doi.org/10.1016/j.eswa.2022.117005

Durbin, D.-A. (2024, October 30). PepsiCo is closing 4 bottling plants and cutting nearly 400 jobs as it streamlines operations. *Associated Press*. https://apnews.com/article/pepsi-plant-closings-layoff-b9240378d4e837edaeda3941b3562bc2

Dutta, A., Pooja, G., Jain, N., Panda, R. R., & Nagwani, N. K. (2021). A Hybrid Deep Learning Approach for Stock Price Prediction. *Lecture Notes in Networks and Systems*, *141*, 1–10. https://doi.org/10.1007/978-981-15-7106-0_1

Frino, A., Xu, C., & Zhou, Z. I. (2022). Are option traders more informed than Twitter users? A PVAR analysis. *Journal of Futures Markets*, *42*(9), 1755–1771.

https://doi.org/10.1002/fut.22303

Gehlor, A., & Singh, R. (2022). Estimation of Sentiment Analysis Base Stock Market Crisis. *2022 IEEE 3rd Global Conference for Advancement in Technology (GCAT)*, 1–5. https://doi.org/10.1109/GCAT55367.2022.9972059

Göncü, A., & Akyıldırım, E. (2016). Statistical Arbitrage with Pairs Trading. *International Review of Finance*, *16*(2), 307–319. https://doi.org/10.1111/irfi.12074

Gong, X., Zhang, W., Wang, J., & Wang, C. (2022). Investor sentiment and stock volatility: New evidence. *International Review of Financial Analysis, 80*, 102028. https://doi.org/10.1016/j.irfa.2022.102028

Groß-Klußmann, A., & Hautsch, N. (2011). When machines read the news: Using automated text analytics to quantify high frequency news-implied market reactions. *Journal of Empirical Finance*, *18*(2), 321–340. https://doi.org/10.1016/j.jempfin.2010.11.009

Hayes, A., Brown, J. R., & Perez, Y. (2024). *Portfolio Management: Definition, Types, and Strategies*. Investopedia. https://www.investopedia.com/terms/p/portfoliomanagement.asp

Hsu, Y.-J., Lu, Y.-C., & Yang, J. J. (2021). News sentiment and stock market volatility. *Review of Quantitative Finance and Accounting*, *57*(3), 1093–1122. https://doi.org/10.1007/s11156-021-00971-8

Hutto, C. J., & Gilbert, E. (2014). VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text. *Eighth International AAAI Conference on Weblogs and Social Media*, 216–225. https://ojs.aaai.org/index.php/ICWSM/article/view/14550

Hwang, I. (2024). *A Brief History of the Stock Market*. Sofi. https://www.sofi.com/learn/content/history-of-the-stock-market/

IBM. (n.d.). *What is machine learning (ML)?* IBM. Retrieved November 30, 2024, from https://www.ibm.com/topics/machine-learning

Jain, N., Motiani, M., & Kaur, P. (2022). *Stock Direction Prediction Using Sentiment Analysis of News Articles* (pp. 195–206). https://doi.org/10.1007/978-981-16-2164-2_16

James, C., & Scott, G. (2021). *Pairs Trade: Definition, How Strategy Works, and Example*. Investopedia. https://www.investopedia.com/terms/p/pairstrade.asp

Jiang, S., & Jin, X. (2021). Effects of investor sentiment on stock return volatility: A spatio-temporal dynamic panel model. *Economic Modelling*, *97*, 298–306. https://doi.org/10.1016/j.econmod.2020.04.002

Keras-Team. (2024). *Keras*. GitHub. https://github.com/keras-team/keras

King, C. (2024, December 2). Agriculture, Value Chain & Choices: PepsiCo's pep+ Strategy. *Sustainability Magazine*. https://sustainabilitymag.com/articles/agriculture-value-chain-choices-pepsicos-pep-strategy

Koratamaddi, P., Wadhwani, K., Gupta, M., & Sanjeevi, S. G. (2021). Market sentiment-aware deep reinforcement learning approach for stock portfolio allocation. *Engineering Science and Technology, an International Journal*, *24*(4), 848–859. https://doi.org/10.1016/j.jestch.2021.01.007

Li, J. (2022). Preliminarily Explore the Steps of Financial Big Data Analytics. *Procedia Computer Science*, *208*, 471–477. https://doi.org/10.1016/j.procs.2022.10.065

Liapis, C. M., Karanikola, A., & Kotsiantis, S. (2023). Investigating Deep Stock Market Forecasting with Sentiment Analysis. *Entropy*, *25*(2). https://doi.org/10.3390/e25020219

Lin, T.-Y., Chen, C. W. S., & Syu, F.-Y. (2021). Multi-asset pair-trading strategy: A statistical learning approach. *The North American Journal of Economics and Finance*, *55*, 101295. https://doi.org/10.1016/j.najef.2020.101295

Longmore, K. (2019). *Practical Pairs Trading*. Robot Wealth. https://robotwealth.com/practical-pairs-trading/

Luthfianti, B. A., Saepudin, D., & Ihsan, A. F. (2022). Portfolio Allocation of Stocks in Index LQ45 using Deep Reinforcement Learning. *2022 10th International Conference on Information and Communication Technology (ICoICT)*, 205–210. https://doi.org/10.1109/ICoICT55009.2022.9914892

Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, *165*, 113973. https://doi.org/10.1016/j.eswa.2020.113973

Mathworks. (2024). *What Is Fuzzy Logic?* Mathworks. https://www.mathworks.com/help/fuzzy/what-is-fuzzy-logic.html

Muguto, H. T., Muguto, L., Bhayat, A., Ncalane, H., Jack, K. J., Abdullah, S., Nkosi, T. S., & Muzindutsi, P.-F. (2022). The impact of investor sentiment on sectoral returns and volatility: Evidence from the Johannesburg stock exchange. *Cogent Economics & Finance, 10*(1). https://doi.org/10.1080/23322039.2022.2158007

Muthivhi, M., & van Zyl, T. L. (2022). Fusion of Sentiment and Asset Price Predictions for Portfolio Optimization. *2022 25th International Conference on Information Fusion (FUSION)*, 1–8. https://doi.org/10.23919/FUSION49751.2022.9841261

Niu, Z., Demirer, R., Suleman, M. T., & Zhang, H. (2022). Speculation, Cross-Market Sentiment and the Predictability of Gold Market Volatility. *Journal of Behavioral Finance*, 1–18. https://doi.org/10.1080/15427560.2022.2109639

Owen, L., & Oktariani, F. (2020). SENN: Stock Ensemble-based Neural Network for Stock Market Prediction using Historical Stock Data and Sentiment Analysis. *2020 International Conference on Data Science and Its Applications (ICoDSA)*, 1–7. https://doi.org/10.1109/ICoDSA50139.2020.9212982

Pedregosa, F., Varoquax, G., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Ramos-Requena, J. P., López-García, M. N., Sánchez-Granero, M. A., & Trinidad-Segovia, J. E. (2021). A Cooperative Dynamic Approach to Pairs Trading. *Complexity*, *2021*, 1–8. https://doi.org/10.1155/2021/7152846

Rather, A. M. (2012). Optimization of Predicted Portfolio Using Various Autoregressive Neural Networks. *2012 International Conference on Communication Systems and Network Technologies*, 265–269. https://doi.org/10.1109/CSNT.2012.65

Saikat, C. (2019). *How to train your machine: JPMorgan FX algos learn to trade better*. Reuters. https://www.reuters.com/article/us-jpm-trading-machines/how-to-train-your-machine-jpmorgan-fx-algos-learn-to-trade-better-idUSKCN1S61JG/

Seabold, S., & Perktold, J. (2010). statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*. https://github.com/statsmodels/statsmodels

Sen, J., Dutta, A., & Mehtab, S. (2021). Stock Portfolio Optimization Using a Deep Learning

LSTM Model. *2021 IEEE Mysore Sub Section International Conference, MysuruCon 2021*, 263–271. https://doi.org/10.1109/MysuruCon52639.2021.9641662

Sentiment. (n.d.). Oxford Learner's Dictionaries. Retrieved December 2, 2024, from https://www.oxfordlearnersdictionaries.com/definition/english/sentiment

Shastri, M., Roy, S., & Mittal, M. (2018). Stock Price Prediction using Artificial Neural Model: An Application of Big Data. *ICST Transactions on Scalable Information Systems*, 156085. https://doi.org/10.4108/eai.19-12-2018.156085

Shen, Y., Liu, T., Liu, W., Xu, R., Li, Z., & Wang, J. (2021). Deep Reinforcement Learning for Stock Recommendation. *Journal of Physics: Conference Series*, *2050*(1), 012012. https://doi.org/10.1088/1742-6596/2050/1/012012

Singh, A., Gupta, P., & Thakur, N. (2021). An Empirical Research and Comprehensive Analysis of Stock Market Prediction using Machine Learning and Deep Learning techniques. *IOP Conference Series: Materials Science and Engineering*, *1022*(1), 012098. https://doi.org/10.1088/1757-899X/1022/1/012098

Smith, T., & Rhinehart, C. (2023). *What Is Market Sentiment? Definition, Indicator Types, and Example*. Investopedia. https://www.investopedia.com/terms/m/marketsentiment.asp

Song, Z., Gong, X., Zhang, C., & Yu, C. (2023). Investor sentiment based on scaled PCA method: A powerful predictor of realized volatility in the Chinese stock market. *International Review of Economics & Finance*, *83*, 528–545. https://doi.org/10.1016/j.iref.2022.10.007

Steinberg, B. (2024, November 27). Pepsi pushes new, 'festive' flavor to replace pumpkin spice — here's where to get the 'holiday in a can.' *New York Post*. https://nypost.com/2024/11/27/lifestyle/pepsi-pushes-new-festive-flavor-to-replace-pumpkin-spice-heres-where-to-get-the-holiday-in-a-can/

Stempel, J. (2024, November 1). PepsiCo beats New York state's lawsuit over plastics pollution. *Reuters*. https://www.reuters.com/legal/pepsico-beats-new-york-states-predatory-lawsuit-over-plastics-pollution-2024-11-01/

Theodorou, T.-I., Zamichos, A., Skoumperdis, M., Kougioumtzidou, A., Tsolaki, K., Papadopoulos, D., Patsios, T., Papanikolaou, G., Konstantinidis, A., Drosou, A., &

Tzovaras, D. (2021). An AI-Enabled Stock Prediction Platform Combining News and Social Sensing with Financial Statements. *Future Internet*, *13*(6), 138. https://doi.org/10.3390/fi13060138

Thompson, C., Anderson, S., & Kvilhaug, S. (2024). *Fundamental vs. Technical Analysis: What's the Difference?* Investopedia. https://www.investopedia.com/ask/answers/difference-between-fundamental-and-technical-analysis/

Thompson, C., Potters, C., & Jackson, A. (2024). *Bollinger Bands: What They Are, and What They Tell Investors*. Investopedia. https://www.investopedia.com/terms/b/bollingerbands.asp

Tirea, M., & Negru, V. (2013). Investment portfolio optimization based on risk and trust management. *2013 IEEE 11th International Symposium on Intelligent Systems and Informatics (SISY)*, 369–374. https://doi.org/10.1109/SISY.2013.6662604

Touzani, Y., Douzi, K., & Khoukhi, F. (2019). Stock price forecasting: Improved long short term memory model for uptrend detecting. In *Lecture Notes in Networks and Systems* (Vol. 66, pp. 214–219). Springer. https://doi.org/10.1007/978-3-030-11914-0_23

Van Deusen, A. (2023). *Python Popularity: The Rise of A Global Programming Language*. FLATIRON SCHOOL. https://flatironschool.com/blog/python-popularity-the-rise-of-a-global-programming-language/

van Kempen, G. M. P., & van Vliet, L. J. (1995). Mean and variance of ratio estimators used in fluorescence ratio imaging. *Journal of Quantitative Cell Science*, *39*(4), 300–305. https://doi.org/https://doi.org/10.1002/(SICI)1097-0320(20000401)39:4%3C300::AID-CYTO8%3E3.0.CO;2-O

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, Pearu Weckesser, W., Bright, J., van der Walt, Stefan J. Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, I., … Contributors, S. 1. . (2020). Scipy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272.

Yao, H., Xia, S., & Liu, H. (2022). Six-factor asset pricing and portfolio investment via deep learning: Evidence from Chinese stock market. *Pacific-Basin Finance Journal*, *76*, 101886. https://doi.org/10.1016/j.pacfin.2022.101886

Zhang, C., Sjarif, N. N. A., & Ibrahim, R. B. (2022). Decision Fusion for Stock Market Prediction: A Systematic Review. *IEEE Access*, *10*, 81364–81379. https://doi.org/10.1109/ACCESS.2022.3195942

Zhang, W., Gong, X., Wang, C., & Ye, X. (2021). Predicting stock market volatility based on textual sentiment: A nonlinear analysis. *Journal of Forecasting*, *40*(8), 1479–1500. https://doi.org/10.1002/for.2777

Zhou, F., Zhang, Q., Zhu, Y., & Li, T. (2023). T2V_TF: An adaptive timing encoding mechanism based Transformer with multi-source heterogeneous information fusion for portfolio management: A case of the Chinese A50 stocks. *Expert Systems with Applications*, *213*, 119020. https://doi.org/10.1016/j.eswa.2022.119020

7 Appendices

7.1 Appendix A: *Univariate adjusted R^2 graphs*

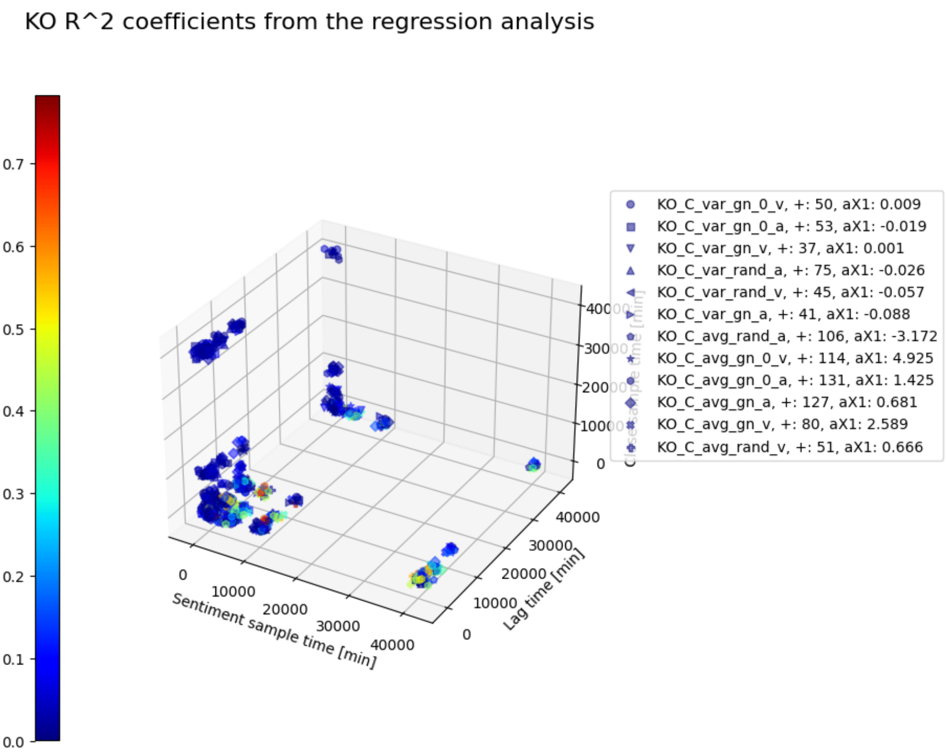

Figure 18: Adjusted R^2 values for the Coca-Cola company

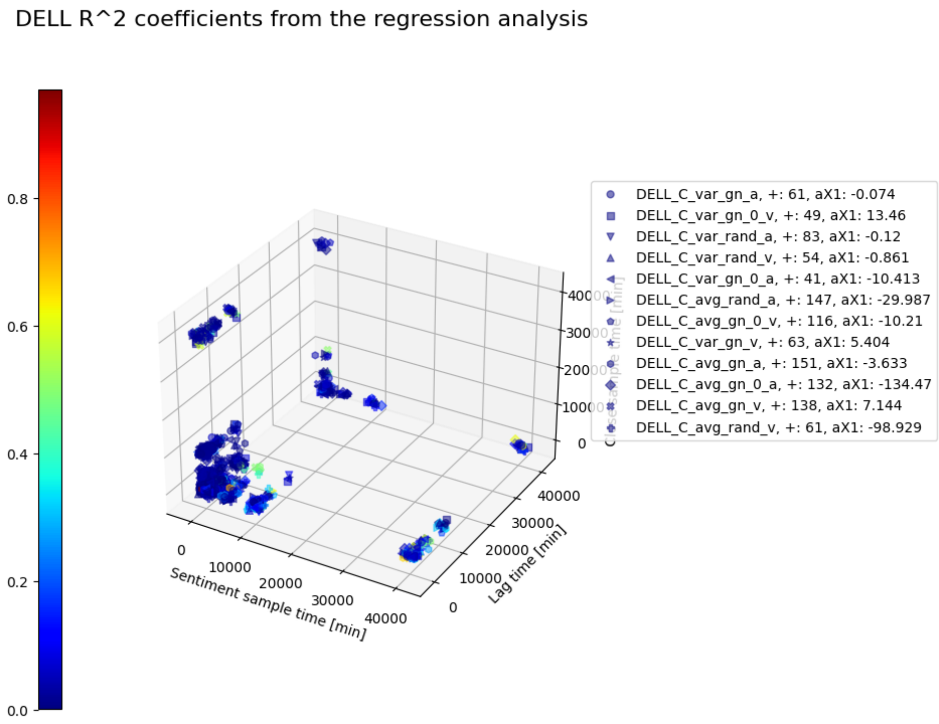

Figure 19: Adjusted R^2 values for Dell

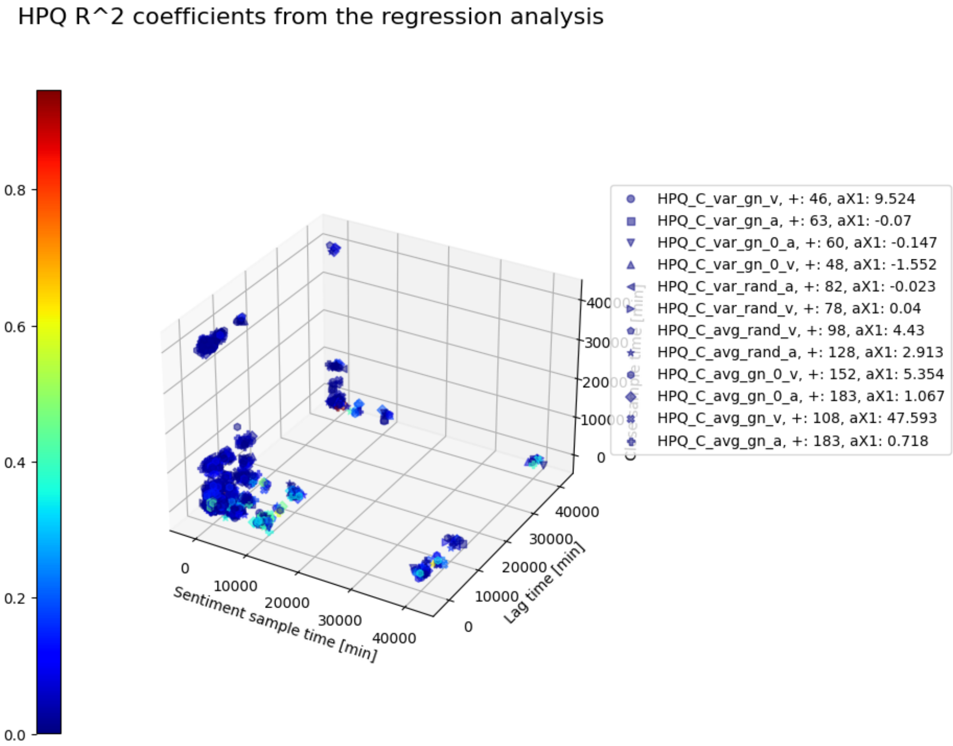

Figure 20: Adjusted R^2 values for Hewlett Packard

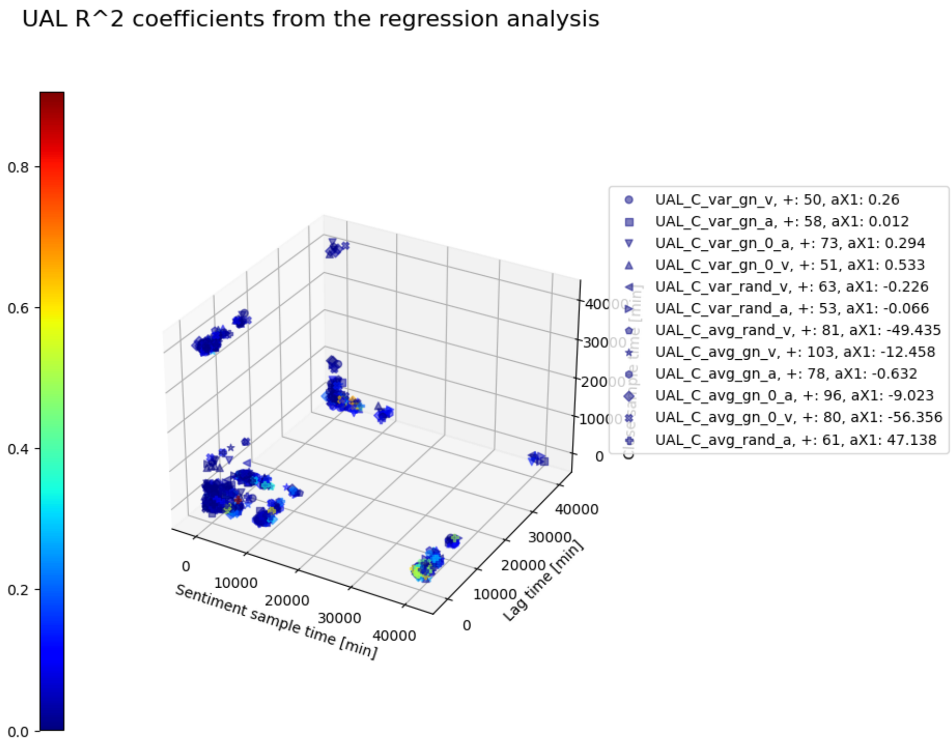

Figure 21: Adjusted R^2 values for United Airlines

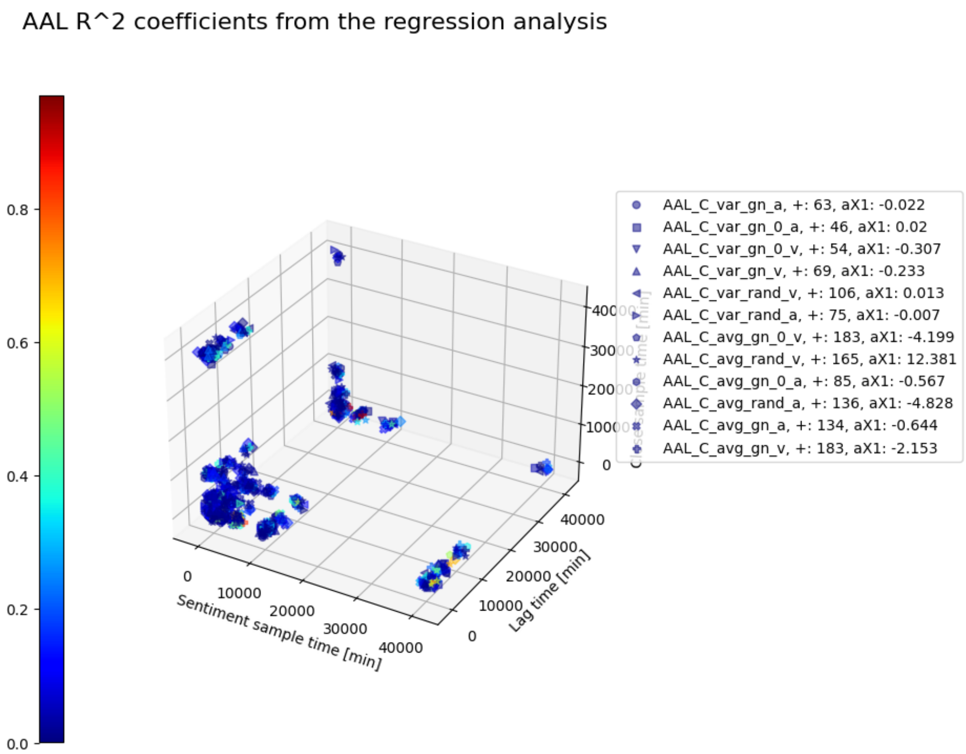

Figure 22: Adjusted R^2 values for American Airlines

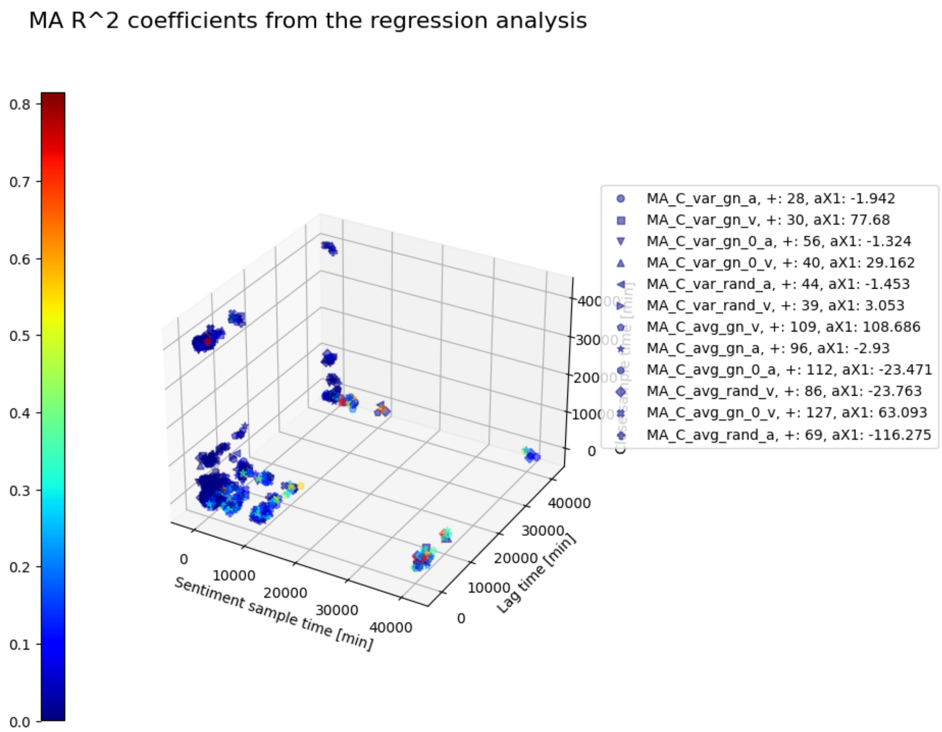

Figure 23: Adjusted R^2 values for Mastercard

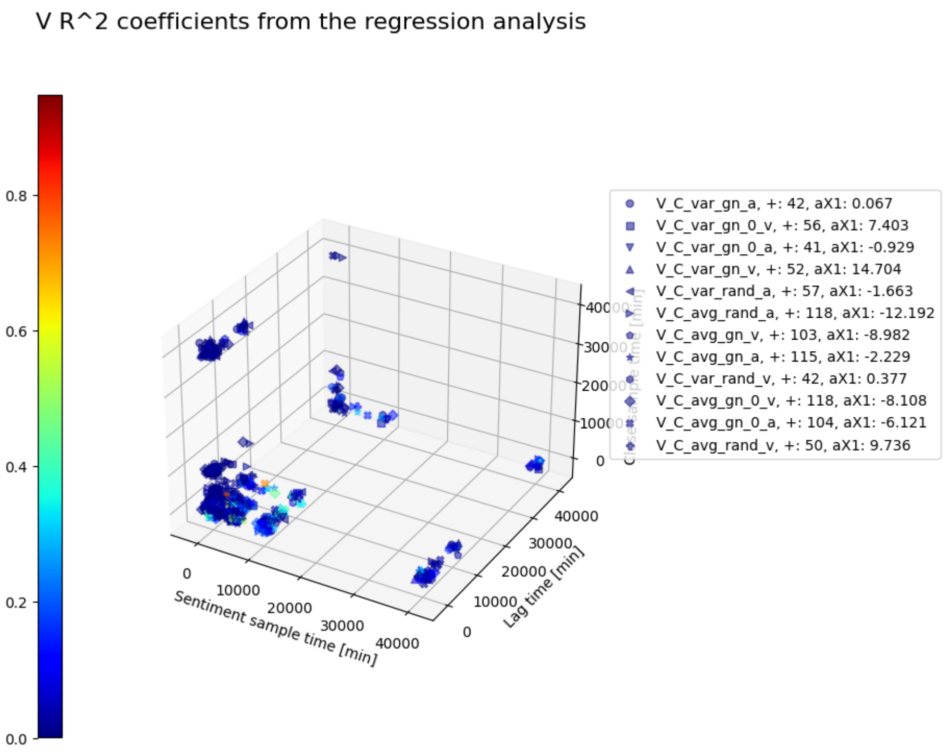

Figure 24: Adjusted R^2 values for Visa

7.2 Appendix B: *Time dependence graphs*

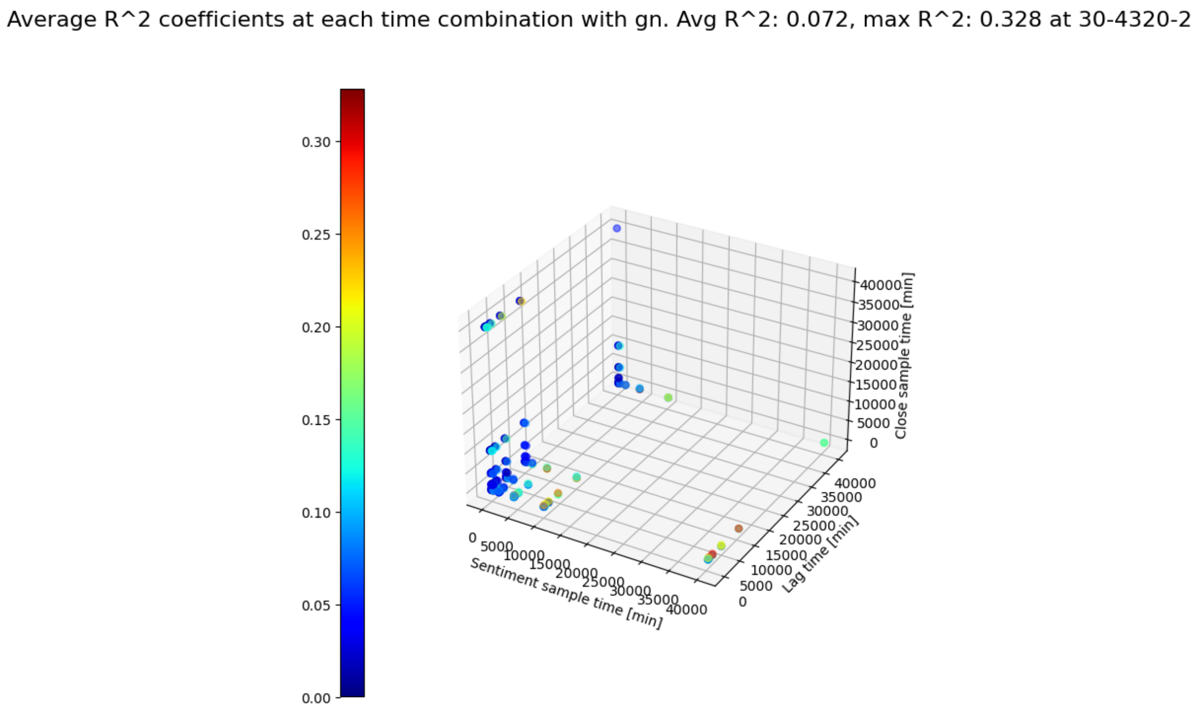

Figure 25: Average adjusted R^2 values across all stocks using Google News sentiment without neutral scores

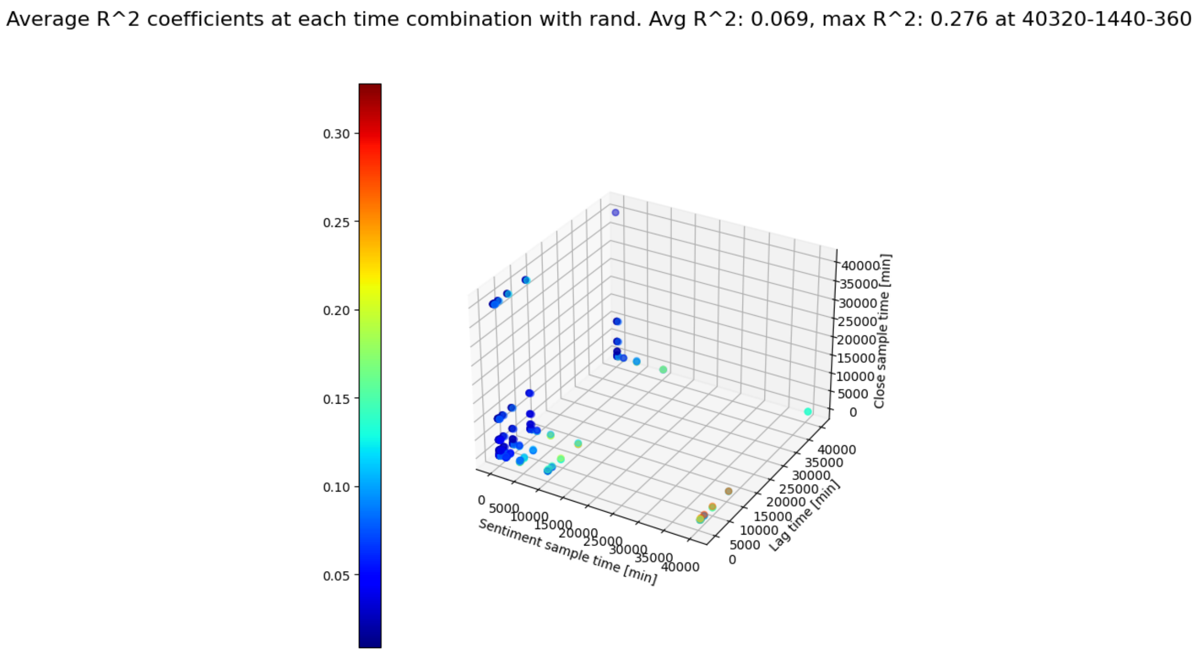

Figure 26: Average adjusted R^2 values across all stocks using randomly generated sentiment

7.3 Appendix C: *Multivariate adjusted R^2 graphs*

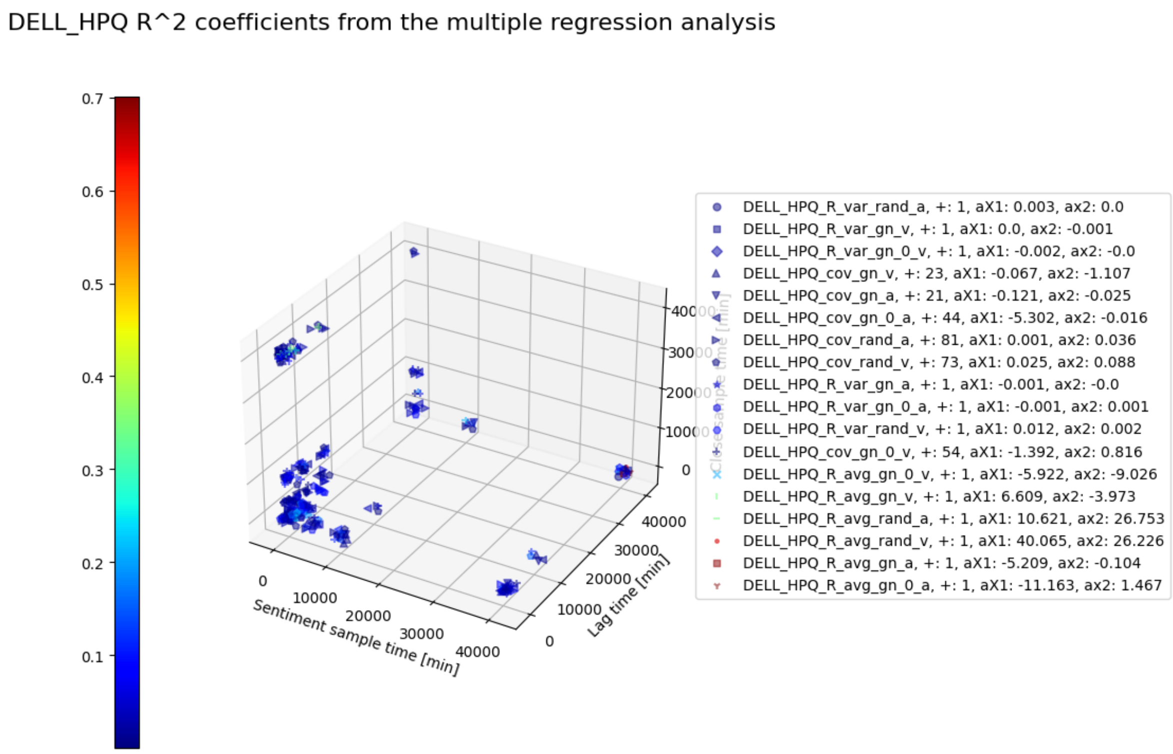

Figure 27: Adjusted R^2 values for DELL_HPQ

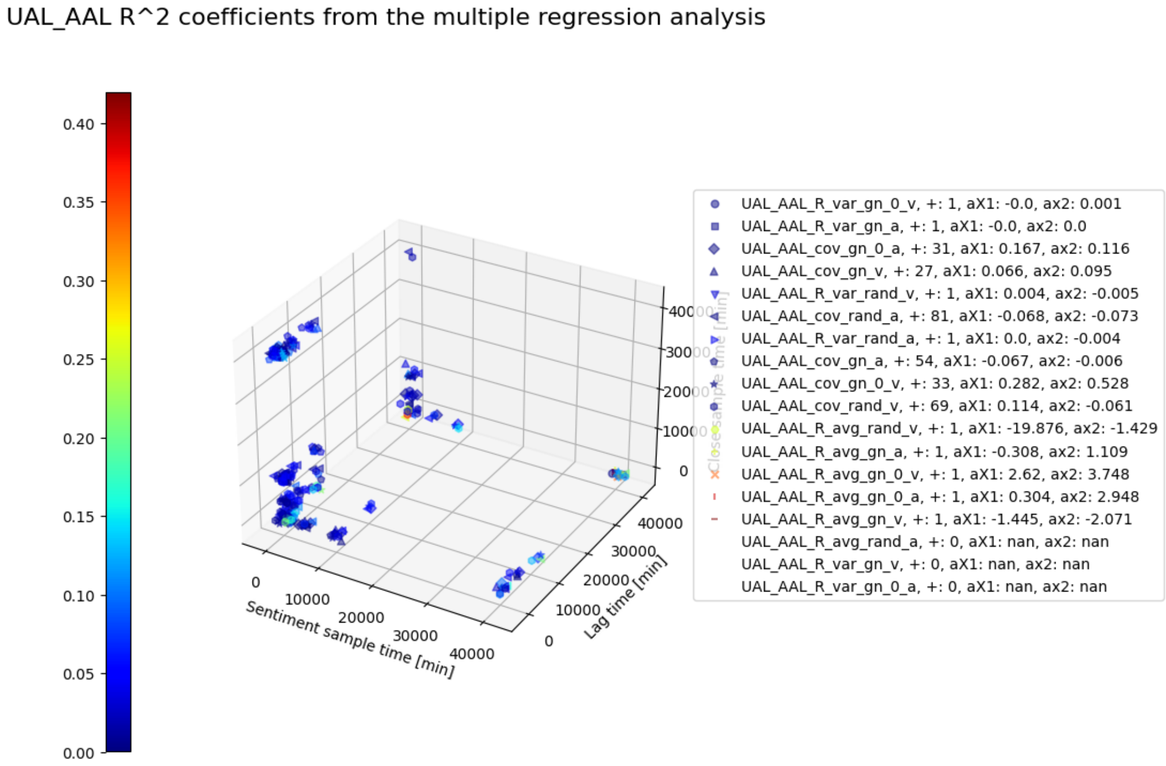

Figure 28: Adjusted R^2 values for UAL_AAL

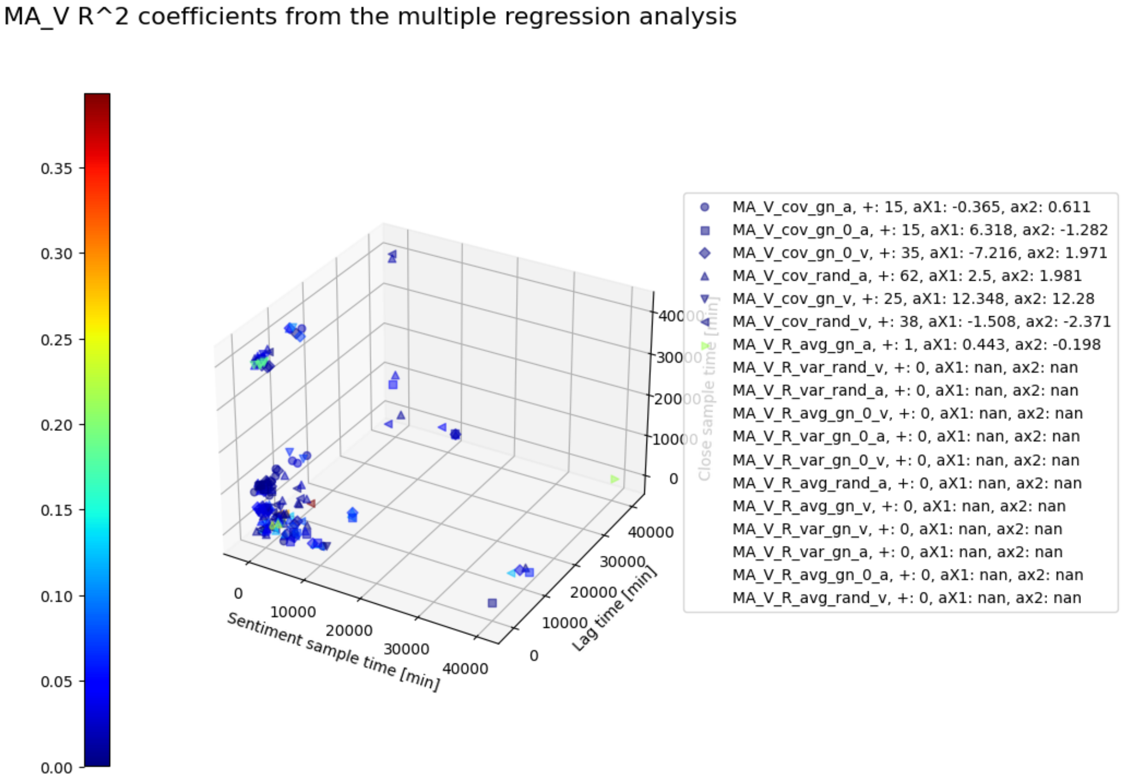

Figure 29: Adjusted R^2 values for MA_V

7.4 Appendix D: Influence of sentiment of LSTM prediction accuracy

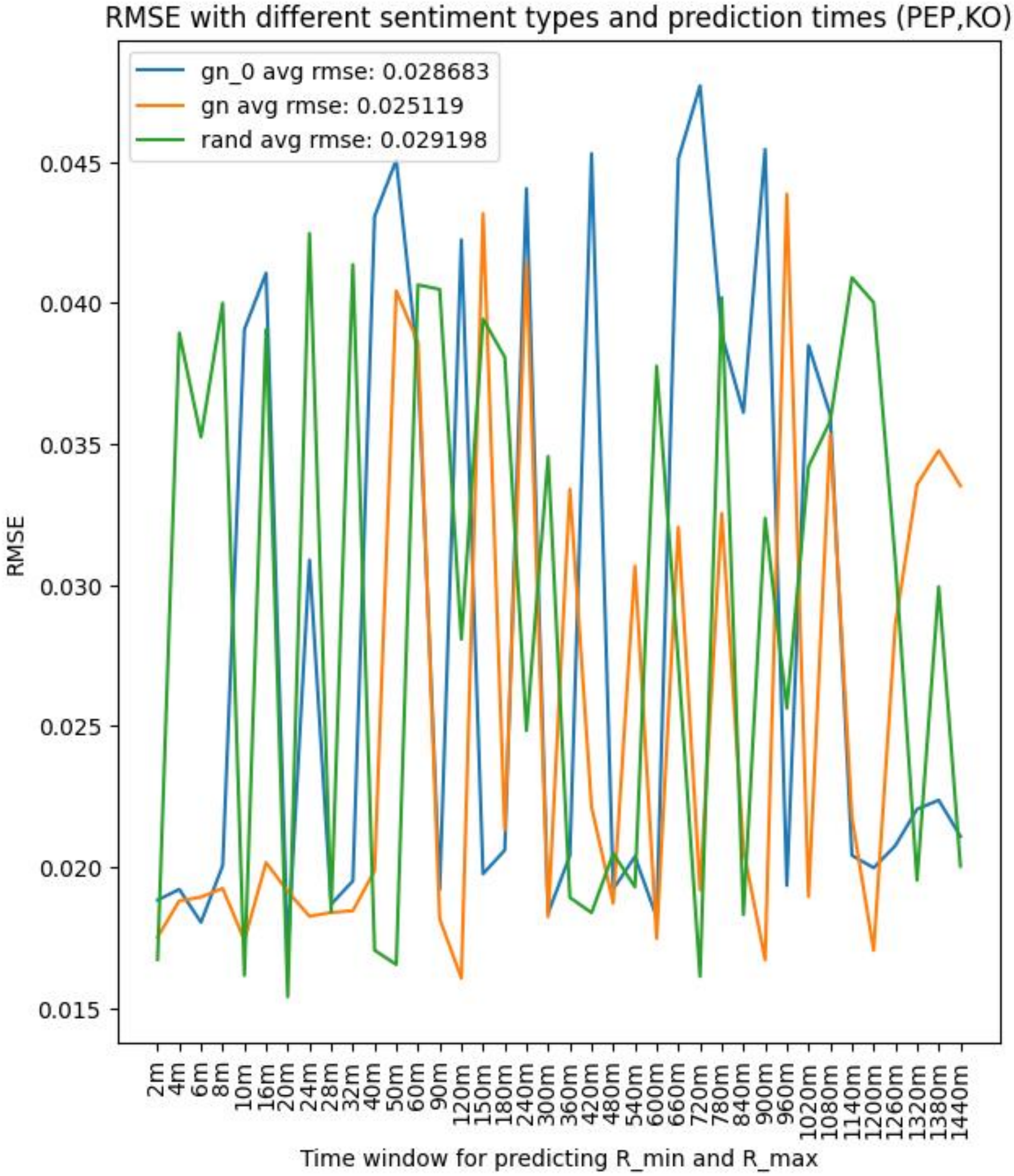

Figure 30: RMSE from LSTM models with different prediction periods and sentiment types for PEP_KO

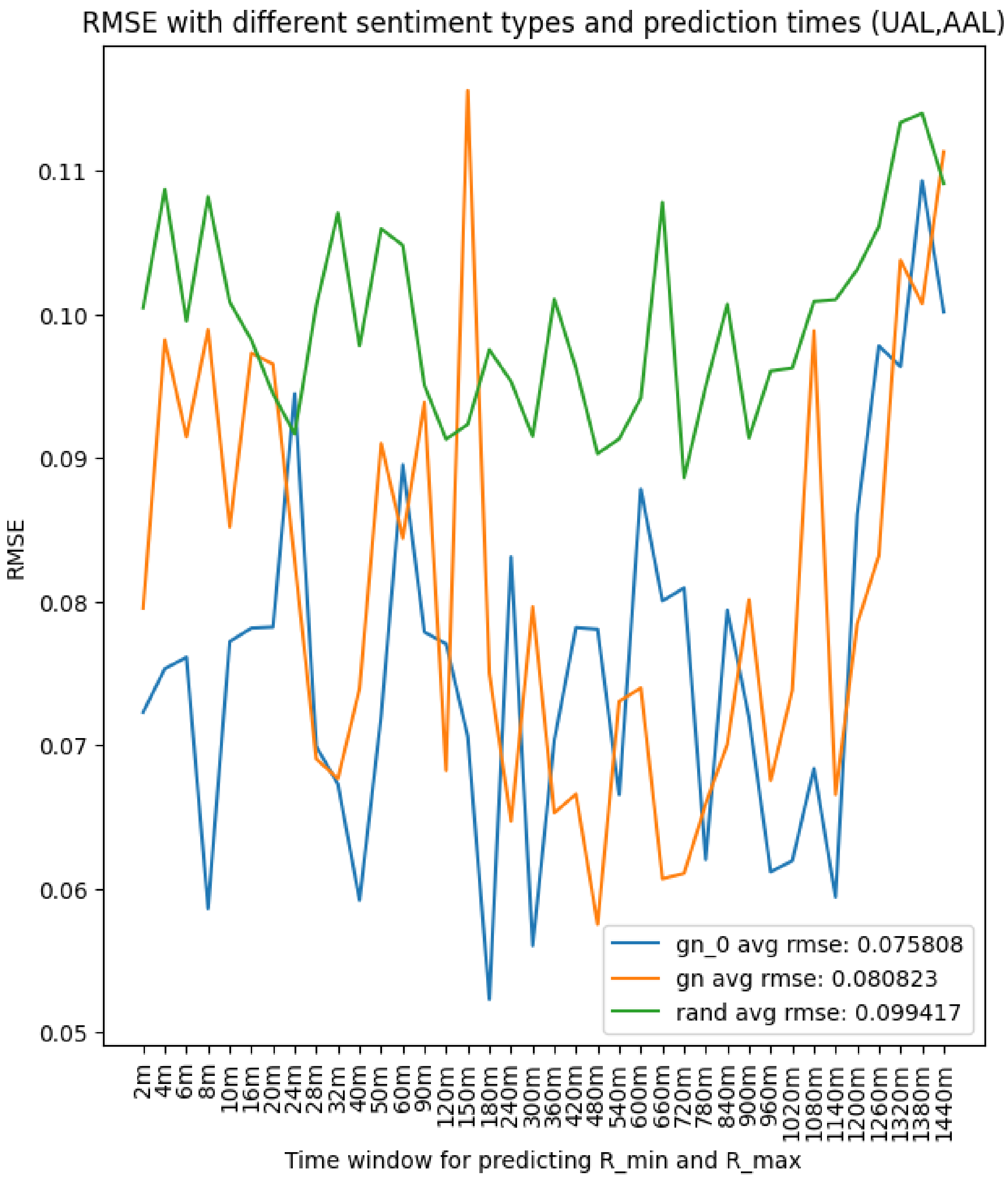

Figure 31: RMSE from LSTM models with different prediction periods and sentiment types for UAL_AAL

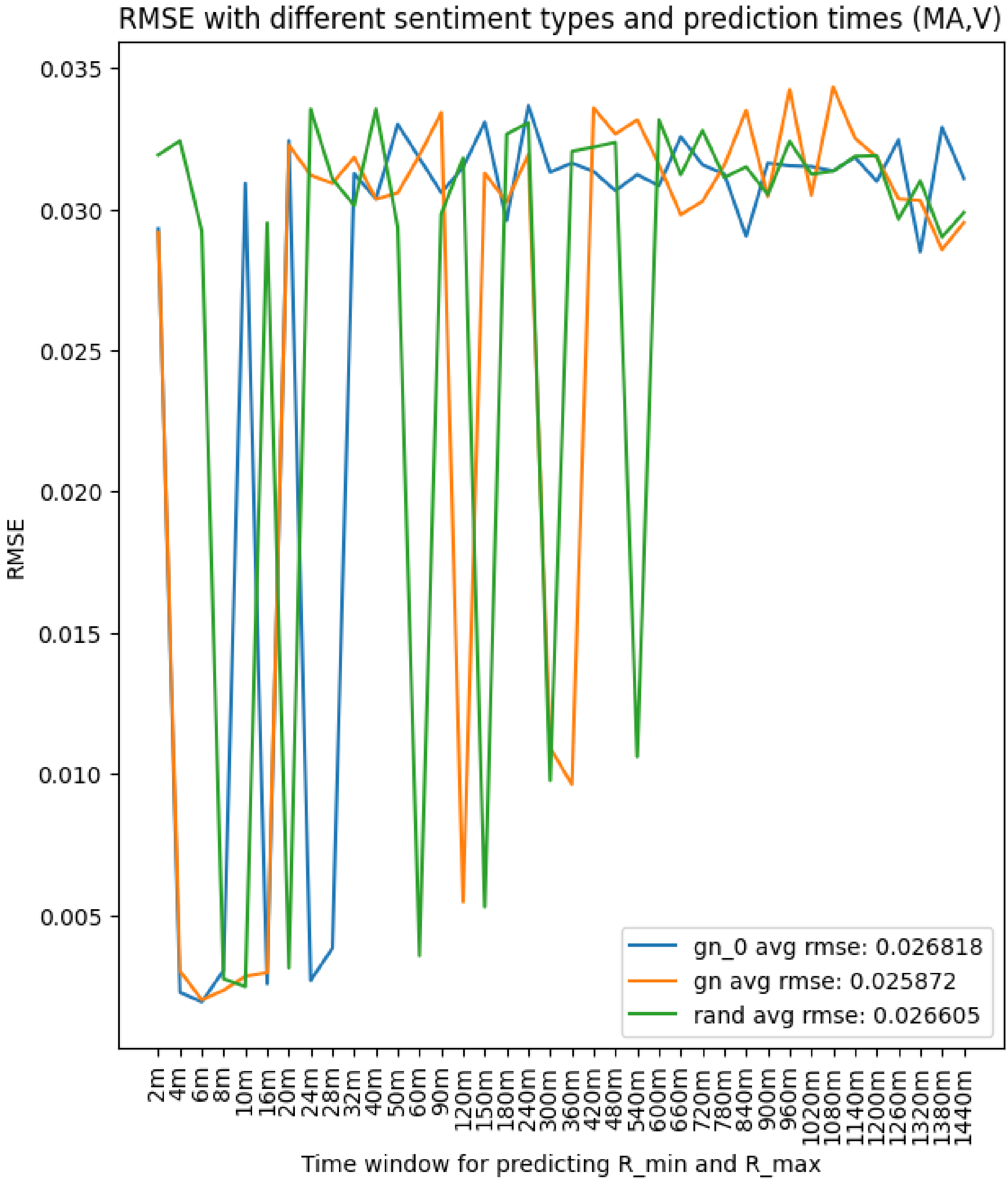

Figure 32: RMSE from LSTM models with different prediction periods and sentiment types for MA_V