Autoregressive Stock Prediction Modeling

CSC-475: Seminar in Computer Science
Fall 2023

Emil Westling
Department of Computer Science
Furman University
Emil.Westling@furman.edu

Abstract

This report explores the feasibility of predicting stock prices using autoregressive models, focusing on two major companies, Apple and Amazon. The purpose of this study is to serve as an introduction to stock price prediction using machine learning and time series analysis. Utilizing data from Yahoo Finance, the study involves preprocessing, including sentiment analysis of related tweets, and visualizing correlations through various plots and heatmaps. Initial findings suggest a strong correlation between stock prices and the S&P 500 index, with a lower correlation between stock prices and sentiment scores and trading volumes. The project employs the AutoReg library for model construction, experimenting with different lags and exogenous variables. Results indicate that lower lags yield better predictions, and the models' performance varies between the two companies. Surprisingly, the inclusion of exogenous variables does not significantly improve model accuracy. The report acknowledges the limitations of its short-term prediction scope and suggests future work focusing on long-term forecasting, integrating different models, and identifying optimal exogenous variables.

1 Introduction

Wall Street is a long-held dream destination for many. The barrier to entry into the world's financial capital is exceptionally high, requiring two essential elements: connections and technical skills. The technical skills required depend on the type of job one seeks. A prominent role in many major financial firms is that of a quantitative analyst, commonly known as "quants". Quants employ mathematics and statistics to enhance companies' decision-making processes, particularly in areas like trading. Quantitative traders, for example, deploy machine learning models to optimize the profitability of their trading business. This project serves as a good learning experience and entry point for anyone with an interest in computational finance.

Predicting the stock market in the hopes of accumulating wealth has been a problem people have been hoping to solve ever since its creation. The closest anyone has come to solving this challenge is the mathematician Jim Simons. He earned a PhD in Mathematics from the University of California, Berkeley, and served as a professor at Stony Brook University before starting his hedge fund Renaissance Technologies. Renaissance Technologies' Medallion fund earned a net average return of 39.1% between 1988-2018, making Jim Simons the most successful investor in the history of Wall Street. To put his performance in perspective, Warren Buffet, widely regarded as the greatest investor of all time, averaged 20.5% 1965-2018. Simons and his colleagues employed groundbreaking trading techniques, using mathematical models to determine which set of states fit the pricing data the best. The hyper-successful model used this idea to make its pricing predictions [1]. Today, algorithmic trading, typically reliant on machine learning, makes up roughly 70% of intra-day trading volume in the USA [2].

This report tackles the problem of stock prediction using autoregressive models. The ultimate goal is not to make money off this project, but to learn as much as possible about stock prediction and time series analysis. Autoregression is commonly used for time series modeling making it a good starting point for a newcomer. Stock price data inherently represents time series data, characterized by its sequential ordering based on time [3]. While this report explores a general Autoregressive model, an interesting progression of this project would be to explore other models, including other autoregressive models such as ARIMA or VAR as they are both commonly used in the world of computational finance.

This report concentrates on two of the world's largest companies, Apple and Amazon, which constitute 7.29% and 3.51% of the S&P 500 index, respectively [4]. Furthermore, their substantial daily trading volumes make them particularly interesting stocks for analysis. Comparing the results between different companies is a good way to avoid biases and arrive at more general results. Additionally, the distinct nature of the businesses—Apple being a tech giant with a focus on consumer electronics and software, and Amazon a leader in e-commerce and cloud computing—provides a diverse set of data. This diversity is essential in understanding how different sectors react to market trends and external factors. Moreover, examining the performance of these companies' stocks will offer insights into the broader market dynamics. It's well acknowledged that tech companies, especially ones as influential as Apple and Amazon, often serve as market leaders for not only the tech industry, but the stock market as a whole. Their performance can influence investor sentiment and market trends. Understanding their stock movements could provide valuable lessons in market psychology and trend analysis.

With this in mind, this project will explore models focusing not only on past pricing data but also incorporating a variety of exogenous variables to enhance the predictive accuracy. The selected exogenous variables include S&P 500 pricing data, sentiment analysis, and daily trading volumes. We will experiment with different lags (10, 15, or 20 days) to forecast the next day's stock price. Interestingly, our initial findings indicate that the inclusion of these exogenous variables doesn't significantly enhance the performance of our models. A possible explanation for this could be the models' short-term prediction horizon, which is limited to just one day ahead. This short-term forecasting might restrict the impact of external variables, suggesting that their effects could be more pronounced over longer prediction periods.

2 Related Work

As the field of AI keeps evolving and the models become more versatile and powerful, the financial industry seems to find more and more applications for it. The intersection of finance and technology is called FinTech and its growing at a rapid pace. AI has caused numerous subfields of FinTech to evolve. Some of these include BankingTech, WealthTech, TradeTech and PayTech. Today, Data science and artificial intelligence plays a crucial role in the economy and its continued growth. Despite the substantial financial investment made in extensive research over the years, the field continues to remain open for research [5]. Furthermore, the question of whether accurate predictions of stock prices can be reliably achieved over time, due to the inherent noise and unpredictability present in capital markets, has been a hot debate for decades. As early as 1973, an economist at Princeton University published a best-selling book *A Random Walk Down Wall Street* arguing that it is impossible to consistently beat average market returns due to an underlying randomness in financial markets [6].

Nonetheless, financial institutions annually spend billions of dollars on market research, leveraging machine learning techniques to boost their revenue. When attempting to forecast market trends and future stock prices, analysts utilize a mix of methods, focusing on fundamental, technical, and sentiment data related to the company, industry, and overall market. Fundamental analysis involves a thorough evaluation of a company's financial health, examining its financial statements, management effectiveness, industry standing, and economic factors that could influence its performance. Technical analysis, on the other hand, looks at statistical trends from market activities, such as past prices and volumes, using charts and indicators to identify patterns predicting future market behavior. Lastly, sentiment analysis offers insights into market psychology by analyzing news articles, social media, and other textual sources, acknowledging the significant impact of public opinion and perceptions on market movements. These approaches together provide analysts with a holistic view of potential stock behaviors in the constantly evolving financial markets [7].

When it comes to algorithmic trading, there are various techniques that data scientists have employed historically. Machine learning techniques are commonly combined with supervised algorithms to build complex models such as the support vector machine (SVM). SVM is an algorithm developed in 1999 that searches for a hyper plane in higher dimension to separate classes. This characteristic makes it especially useful in algorithmic trading where it can be employed to forecast the direction of daily stock price movements by analyzing specific technical attributes, such as price momentum, moving averages, or trading volume. Additionally, SVM's robustness against overfitting and its ability to handle high-dimensional data make it a preferred choice in financial markets, where data is often complex and multidimensional [8].

Time series analysis is a vital area in algorithmic trading, with ARIMA and GARCH being two of the most researched and implemented algorithms. ARIMA (Autoregressive Integrated Moving Average) is particularly effective for analyzing and forecasting time series data when there is a presence of a clear trend or seasonal pattern. It combines autoregressive features with moving average capabilities, making it versatile for different market conditions. On the other hand, GARCH (Generalized Autoregressive Conditional Heteroskedasticity) is extensively used for modeling the volatility of financial time series, which is crucial in assessing risk and making informed trading decisions. GARCH models are good at capturing the 'volatility clustering' phenomenon often observed in financial markets, where high-volatility events tend to cluster together [9].

Although the use of time series algorithms have proved useful and is still one of the most commonly used trading algorithms, recent research has explored Artificial Neural Networks (ANN) and Hidden Markov Models (HMM) for market analysis since they are more effective in capturing heterogeneous variables and coupling relationships. This can be useful when traders are particularly interested in including information from outside the capital markets in their models. History has repeatedly shown that there is a strong relationship between different markets nationally and internationally. For example, the global financial crisis in 2008 originated in the United States, but spread to all sorts of markets across the globe. An example of a model developed based on this idea is the Multi-layer Coupled Hidden Markov Model (MCHMM) which was developed to specifically capture relationships between intra-country markets [10].

Another area researchers have focused their attention on in recent years is deep learning. Deep learning technology has proven useful in predicting price fluctuations in capital markets. The performance of trading strategies established by applying deep learning to analyze and predict financial data is significantly better than traditional machine learning models previously discussed [11]. However, deep learning models have their downsides. Financial data tends to be significantly volatile and includes unpredictable noise which makes their patterns unclear. These characteristics often causes overfitting among deep learning models and consequently they tend to fall into local minimums [12].

Although there are many deep learning models available, the one that tends to be used in the context of algorithmic trading is deep reinforcement learning. The deep learning part perceives the current market environment for feature learning, whereas the reinforcement learning part constructs the interaction with deep characterization and makes the trading decisions. Deep reinforcement learning combines the perception ability of deep learning and the decision-making ability of reinforcement learning to create a high-performing model [13, 14].

Recognizing the extensive body of advanced research existing in the realm of algorithmic trading, it is important to acknowledge that this project may not lead to groundbreaking contributions in the field. Nevertheless, it holds immense personal value as a unique learning experience. This project serves as an introduction to the world of algorithmic trading, providing a platform to cultivate practical skills and a deeper understanding of the underlying principles. These skills are anticipated to extend beyond trading and into broader areas of finance and technology.

3 Data

The stock pricing data obtained for this project comes from Yahoo Finance. Yahoo Finance is part of Yahoo and it provides live financial data, news, and commentary about individual companies, economies across the globe and anything else related to finance. Yahoo Finance offers a free API, yfinance, that allows individuals to download market data for any security over a specified period of time. The API also allows for download of financial news data, financial statements, and other forms of data related to the company and its stock. The API is publicly available [15].

The fetched data includes pricing data for Apple and Amazon, and the S&P 500 index from 2021-09-30 to 2022-09-30 resulting in 242 observations (the stock market is closed on weekends and during holidays). The initial API calls yield data frames containing multiple columns including opening price, closing price, daily high, daily low, closing price, trading volume, stock splits, and dividends for each date. The columns containing the stock splits, and dividends were omitted as they will not be used for forecasting. The open, close, high, and low were not used for forecasting either, however, these columns were used to generate Figure 1. Additionally, three additional columns were generated. One representing tomorrow's closing price and one indicates whether tomorrow's stock price is higher than today's closing price. This column takes on a binary value (1 or -1). The third column includes the daily sentiment score of the company accounting for all tweets posted related to the company that day.

These tweets are fetched from a dataset on Kaggle containing all tweets related to a company from 2021-09-29 to 2022-09-29 [16]. Hence why the pricing data is limited to these dates. This dataset did not contain a column for the sentiment of the tweet. Therefore, this column had to be generated. This was done using a natural language processing algorithm. Depending on the sentiment of the tweet, it was awarded a sentiment score. A 1 for a positive tweet, a 0 for a neutral tweet, and a -1 for a negative tweet. Since there were multiple observations each day, the sentiments for each day were summed up. If something negative happened to one of the companies, for example a bad earnings report, resulting in a lot of negative tweets regarding the company, then it would receive a large negative sentiment score for that day. Conversely, if a positive event took place and consequently many positive tweets were published related to a company, it would receive a high sentiment score for the day.

The first 6 rows of the resulting Apple data frame are displayed in Table 1. Similarly, Table 2 shows the first 6 rows of Amazon's data frame. Note, the columns for the opening price, daily high, and daily low are not displayed as they are not used in the forecasting. In addition to these three data frames, a separate data frame containing S&P 500 closing prices for the same time period is stored.

Date	**Close**	**Volume**	**Tomorrow**	**Direction**	**Sentiment**
2021-09-30	139.697632	89056700	140.832993	1	4
2021-10-01	140.832993	94639600	137.367691	-1	3
2021-10-04	137.367691	98322000	139.312592	1	4
2021-10-05	139.312592	80861100	140.191269	1	0
2021-10-06	140.191269	83221100	141.464844	1	2

Table 1: First 6 rows, one per day, of the data frame containing the Apple data. The table includes a closing price, trading volume, tomorrow's closing price, the direction (whether tomorrow's price is higher than today's price), and the sentiment score (sum of number of positive/negative tweets that day).

Date	**Close**	**Volume**	**Tomorrow**	**Direction**	**Sentiment**
2021-09-30	164.251999	56848000	164.162994	-1	5
2021-10-01	164.162994	56712000	159.488998	-1	2
2021-10-04	159.488998	90462000	161.050003	1	4
2021-10-05	161.050003	65384000	163.100494	1	3
2021-10-06	163.100494	50660000	165.121506	1	5

Table 2: First 6 rows, one per day, of the data frame containing Amazon data. The table includes a closing price, trading volume, tomorrow's closing price, the direction (whether tomorrow's price is higher than today's price), and the sentiment score (sum of number of positive/negative tweets that day).

3.1 Data Visualization

Figure 1 displays a classic candle stick plot of all three companies' pricing data. The trading volume is also displayed in the plot. It appears as if Apple had large up- and downs throughout the year whereas Amazon trended downwards. Additionally, Apple's average trading volume appears to be slightly larger than Amazon's.

Figure 1: Candle stick plot of Apple (top) and Amazon (bottom) pricing data. Note large up- and downs for Apple's stock and a consistent downwards trend for Amazon's stock. The trading volume is also shown in the plot showing a slightly higher average trading volume for Apple.

It seems logical to assume a high correlation between the trading volume and an extreme sentiment score. If a major event were to occur, there would be an unusual amount of people tweeting about the company as well as trading the security. Similarly, if a lot of people are tweeting negative/positive things about a company, then it would make sense for the stock price to decrease/increase. Hence, a high correlation between the closing price and the sentiment can be expected. Lastly, since both Apple and Amazon make up significant parts of the S&P 500 index, one could imagine there being a large correlation between the overall market index and the individual stock price. Figure 2 displays an initial visualization of the relationship between the predictors and response variables. Observing the line plots, there appears to be a large correlation between the S&P 500 and the stock price of both companies. Additionally, it appears as if the closing price tends to rise when the sentiment is large and fall when the sentiment is particularly low. This effect is particularly evident in Apple's plot. Lastly, it appears as if the volume and sentiment are moving together confirming our earlier reasoning.

The heatmaps in Figure 3 also gives a visual of the correlations between the predictor and response variables. As expected, tomorrow's closing price has a large correlation to today's closing price implying that the price does not change much in one day. It is also evident that the stock price is highly correlated to the S&P 500 price. This relationship is particularly strong for Amazon's stock. Another noteworthy relationship exists between the sentiment score and the trading volume, again, particularly for Amazon. Surprisingly, perhaps, the closing price has low correlations to the trading volume and the sentiment score. One potential explanation for this could be that the sentiment score and the trading volume only influence the closing price when the metrics are taking on extreme values.

Looking at Figure 4 we make an interesting observation. The Apple closing price is, although slightly left-skewed, Gaussian distributed whereas Amazon's closing price is distributed evenly. This may be explained by looking at the line plots 2. Apple's stock price seems to have fluctuated up and down

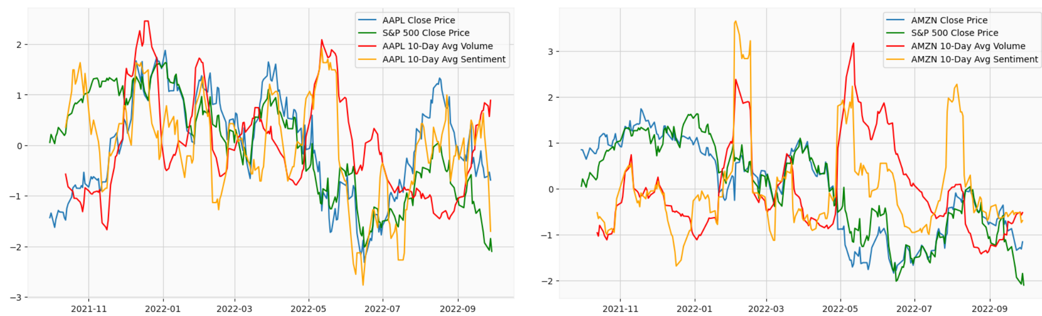

Figure 2: Normalized line plots of Apple's (left) and Amazon's (right) closing price related to their predictor variables. Note how the S&P 500 index and the closing price are moving together. The 10-day rolling mean trading volume and sentiment scores also appear to be moving alongside each others.

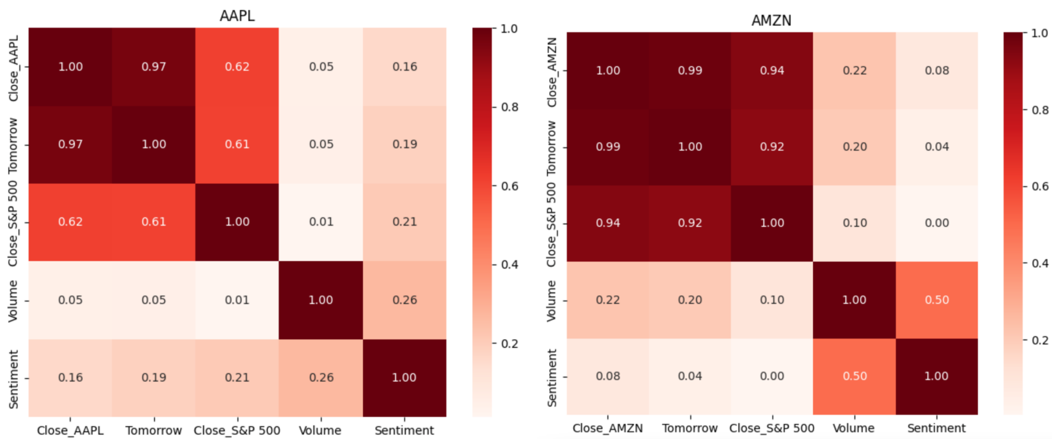

Figure 3: Heatmap visualizing the correlation between the various variables with Apple (left) and Amazon (right). Note the clustering and how the closing price has a high correlation with tomorrow's price and the S&P 500 index price. Also note the correlation between the trading volume and the sentiment score.

without a clear upwards or downwards trend. Amazon's stock price, however, seems to have had a downwards trend throughout the year yielding a more even distribution.

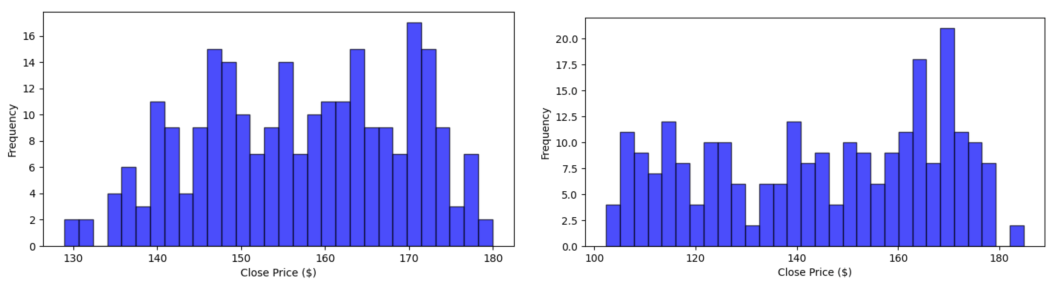

Figure 4: Histogram visualizing the closing price distribution of Apple (left) and Amazon (right). Apple is taking on a more Gaussian distribution whereas Amazon is more evenly distributed.

4 Methods

Fortunately, the acquisition of stock pricing data for this project was straightforward, facilitated by readily accessible API calls. However, gathering sentiment data was more challenging due to

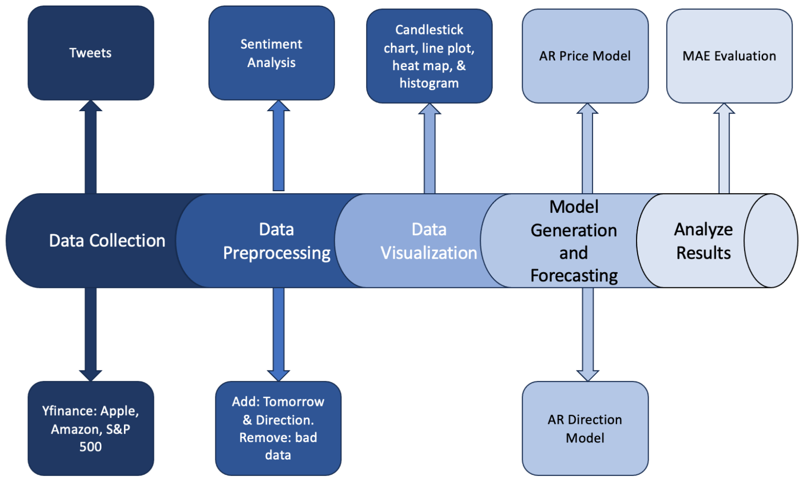

Figure 5: Pipeline diagram of the project. Visualizes all steps of the research inclduing data collection, processing, visualization, model generation and forecasting, and lastly the analysis of the results.

a scarcity of publicly available time series news sentiment datasets. Despite these challenges, a suitable sentiment dataset, apt for stock prediction, was eventually sourced. The next phase involved data processing, where the existing datasets were enhanced. Columns for tomorrow's price and the direction of tomorrow's price were calculated and added to the datasets for Apple and Amazon. Furthermore, the dataset containing tweets originally lacked a sentiment column, which then had to be created. This sentiment was calculated using the sentiment analysis library eng_spacysentiment provided by spaCy [17]. The final step entailed the exclusion of data that was not relevant for forecasting, along with the removal of records containing null values.

After all data was collected, cleaned, and processed, the focus shifted to data visualization. This included generating tables and various figures. A distinctive feature in financial analysis is the candlestick plot, which condenses substantial information into a single graphic. Consequently, a candlestick plot was the natural choice for initially presenting the Apple and Amazon data. Heatmaps and line plots were then generated to highlight relationships among data while tables provided a clear understanding of the data structure. These visualizations were constructed using Matplotlib and Bokeh [18, 19]. An interactive Bokeh line plot can be accessed at `www.emilwestling.com/AlgorithmTrading` whereas most plots in this paper were generated using the library Matplotlib. Figure 3 was constructed using another visualization library, seaborn [20].

Once the data visualization was completed and a thorough understanding of the data had been obtained, the next phase involved developing the initial autoregressive model. This stage required the creation of forecasting and backtesting functions, employing the AutoReg library from the statsmodels package [21]. The model iterates through the data frame making predictions on each data point using the most recent 10, 15, or 20 observations, yielding 151 predictions. Adjusting the inputs, the number of lags, and the exogenous variables within these functions allows for varied predictions. The second model is identical aside from the fact that it is only concerned with the direction of the price, in other words, whether tomorrow's price is higher or lower than today's price.

$$y_t = \delta_0 + \delta_1 t + \phi_1 y_{t-1} + \cdots + \phi_p y_{t-p} + \sum_{j=1}^{m} \kappa_j x_{t,j} + \varepsilon_t. \tag{1}$$

The equation above illustrates the autoregressive equation used by Statsmodel's AutoReg library where y_t is the response variable. In the context of this paper, it is tomorrow's closing price of a stock. The prediction is based on several variables. Specifically, δ_0 serves as the intercept, indicating the starting level of y_t when all other factors are zero. The term $\delta_1 t$ captures any linear trend in the series, signifying that y_t is expected to change over time at a rate proportional to the coefficient

δ_1. The autoregressive components, $\phi_1 y_{t-1} + \cdots + \phi_p y_{t-p}$, reflect the influence of p previous time periods (lag p) on the current value, suggesting that past values have a lingering effect. The model also includes m exogenous variables, $x_{t,j}$, each with an associated coefficient κ_j, representing the influence of additional factors that may affect y_t. Lastly, ϵ_t is the error term, accounting for random fluctuations and unexplained variability in y_t.

Lastly, the models were used for forecasting varying the hyper parameters, number of lags and exogenous variables, among both companies. To assess the accuracy of these predictions, the mean absolute error (MAE) was utilized as the evaluation metric, a widely recognized standard in gauging the performance of regression models. The MAE scores were summarized in a table, sorted by the hyper parameters allowing for comparison and evaluation of the various models. Additionally, to provide a more intuitive understanding of the pricing model's performance, the predicted price data were plotted alongside the actual pricing data.

Note, all code for this project can be accessed at the link: `https://github.com/emilwestling/AlgorithmTrading`.

5 Results

The results for the price prediction model is summarized in Table 3. In general, it appears as the lower the lag, the better performance for both companies. This make intuitive sense due to the nature of the model. Fitting a line through more recent points will likely yield smaller errors on average. The lower MAE scores for Apple indicates that the model is better at predicting Apple's stock price than Amazon's. Although uncertain, a possible explanation to this might be that Apple has slightly smaller price fluctuations from day to day compared to Amazon. The mean of the absolute difference between each day's closing price and the consecutive day's closing price is \$2.417 for Apple as opposed to \$2.900 for Amazon.

Introducing S&P 500 as an exogenous variable into the models yield interesting results. To start, it does a great job of predicting the direction for both stocks. This holds true regardless of the lag for Amazon while increasing the lag worsen the Apple predictions. However, the price predicting model does exceptionally bad making Amazon price predictions using the S&P 500 variable. Conversely, the model does well predicting Apple's stock price with the S&P 500 variable and in particular when using longer lags. Looking back at the heatmap in Figure 3, we noted a high correspondence between the Amazon stock price and the S&P 500 price and a smaller, while still significant, correspondence between the Apple stock price and S&P 500 price. This inidicates that the model is affected too largely by the S&P 500 variable when making predictions for Amazon. Figure 6 confirms this notion as the price predictions using the S&P 500 variable are consistently too big.

Observing the MAE scores for the price predicting model for Amazon, a clear pattern is discovered. Generally, the lower the lag, the better the prediction. Additionally, the model performs the best without the inclusion of exogenous variables, although, its performance does not significantly worsen using either of the sentiment or volume variables. This pattern is not present in the Apple predictions as the inclusion of exogenous variables does seem to be successful. However, it does not improve the performance by much. The model performance is similar regardless of the lag and exogenous variables. To improve the interpretability of Table 3, the scores of the best performing model for each company are bolded.

All predictions are plotted in Figure 6 along with the actual pricing data. The most noticeable things to point out from the pricing predictions are the Amazon predictions using S&P 500. Beginning around May 2022, the predictions start to become too high explaining the high MAE scores. Outside of those predictions, all models make extremely similar predictions elucidating their similar MAE scores. Taking a broader look at the plots, it appears as if the predictions are not particularly far off. They seem to be following the trend of the actual pricing data, however, delayed by a few days. This is an indication that the model relies too heavily on the past pricing data to make predictions. Therefore, the models will never be able to predict sudden stock price changes. Another consequence of this, visible in the plots, is that the predictions take on higher highs and deeper lows. The exogenous variables were introduced in the hopes of being able to predict sudden price changes. However, none of the variables seem to have done this successfully.

Lag	Company / Exog	Amazon MAE_Direction	Amazon MAE_Price	Apple MAE_Direction	Apple MAE_Price
10	NA	1.019868	**3.346440**	0.953642	2.762461
	S&P500	**0.993377**	4.859555	**0.940397**	2.795329
	Sentiment	1.033113	3.356564	0.953642	**2.744114**
	Volume	1.006623	3.360806	0.953642	2.755890
15	NA	1.033113	3.411418	0.966887	2.839545
	S&P500	**0.993377**	5.569160	1.059603	2.787392
	Sentiment	1.006623	3.413289	1.006623	2.829494
	Volume	1.006623	3.443118	0.980132	2.835158
20	NA	1.086093	3.424944	0.993377	2.853466
	S&P500	**0.993377**	5.841501	0.980132	2.768990
	Sentiment	1.033113	3.474168	0.953642	2.837408
	Volume	1.086093	3.447584	0.980132	2.851404

Table 3: Table summarizing the MAE scores for every model. The scores for the best performing models are bolded. Note how lower lag models tend to do make more accurate predictions and how certain exogenous variables are better at predicting direction and worse at predicting price.

6 Future Work

The exploration of autoregressive stock prediction modeling in this research, focusing on Apple and Amazon, opens several avenues for future investigation. The potential expansions of this study are manifold and can significantly enhance the understanding and application of stock prediction models.One area of expansion is the exploration of more advanced predictive models. Further work could benefit from incorporating sophisticated time series models like ARIMA and VAR which might offer improved accuracy in stock price predictions. Moreover, integrating deep learning techniques such as neural networks and reinforcement learning could prove beneficial in recognizing complex patterns in the data.

Another promising direction for future work is the integration of a broader range of exogenous variables. This research primarily focused on the S&P 500 index, sentiment analysis, and trading volumes. Future studies could enhance prediction models by including diverse economic indicators, such as interest rates, inflation rates, and GDP growth. Furthermore, leveraging data from news articles, financial reports, as well as social media for sentiment analysis could provide a more comprehensive view of the general attitude towards the stock.

Extending the research to encompass long-term forecasting presents another valuable opportunity. Shifting focus from short-term to medium and long-term predictions would cater to a wider range of investment strategies and test the models' effectiveness over varying time frames. It could also provide insights into the models' adaptability and resilience to different market conditions.

A comparative analysis across various sectors could also enrich the research. Examining companies from different industries would shed light on sector-specific trends and how various types of stocks react to market fluctuations. This approach would broaden the applicability of the findings and deepen the understanding of market dynamics.

In summary, the field of algorithmic trading and stock prediction is ripe for further research. The opportunities for exploring more complex models, diverse datasets, and extended forecasting horizons are vast. Advancements in this area promise not only enhanced predictive accuracy but also a deeper understanding of the intricate and dynamic nature of financial markets.

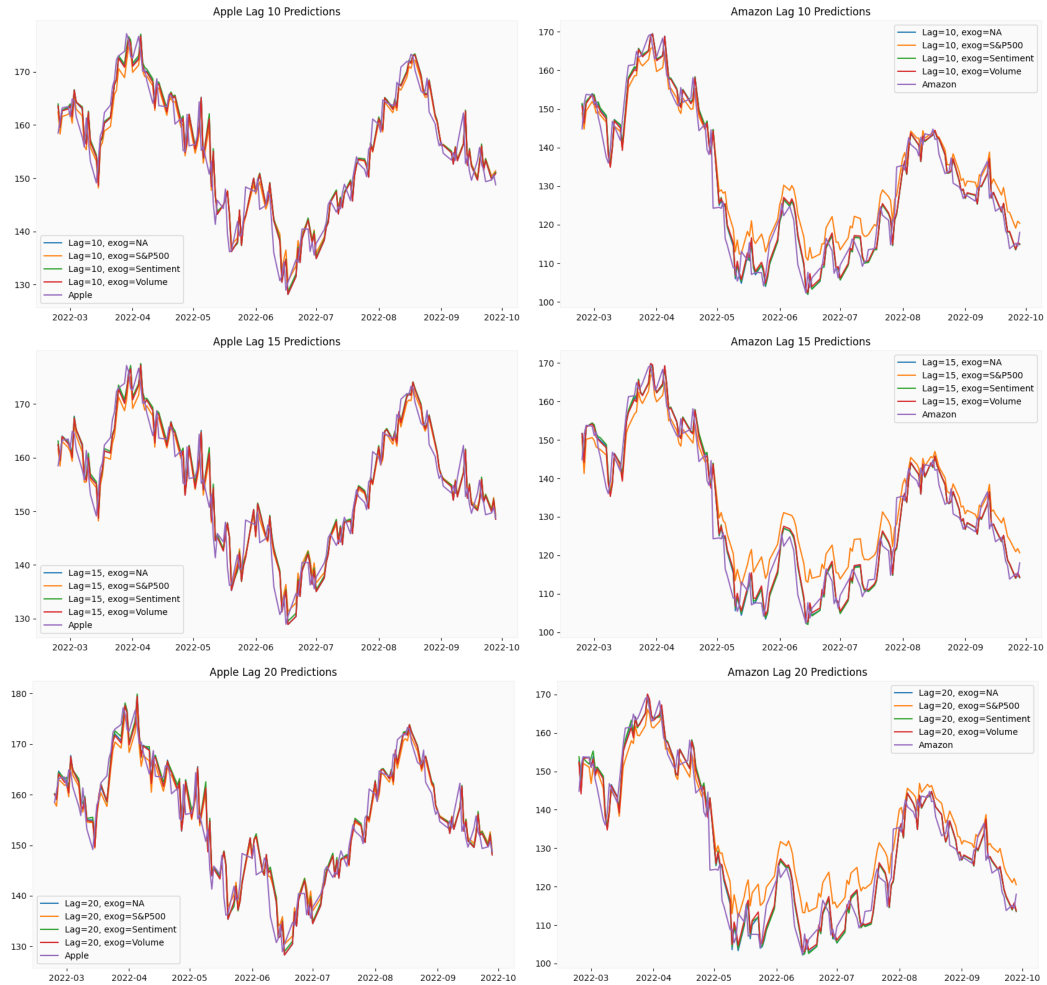

Figure 6: Line plots, Apple (left) and Amazon(right), displaying the predictions for a select few model inputs vs the actual stock price data. In general, most predictions are similar. Additionally, predictions for Apple seem to be more accurate slightly more accurate than Amazon.

References

[1] Gregory Zuckerman. *The Man Who Solved The Market: How Jim Simons Launched the Quant Revolution*. Penguin Random House, 2019.

[2] Gil Cohen. Algorithmic trading and financial forecasting using advanced artificial intelligence methoddologies. *Mathematics*, 2022.

[3] James D. Hamilton. *Time Series Analysis*. Princeton University Press, 2020.

[4] `https://www.slickcharts.com/sp500`. Accessed: 2023-11-29.

[5] Philip S. Yu Longbing Cao, Qiang Yang. Data science and ai in fintech: an overview. *Springer Link*, 2021.

[6] Burton Malkiel. *A Random Walk Down Wall Street*. W. W. Norton Company, Inc., 1973.

[7] Benjamin Asubam Weyori Isaac Kofi Nti, Adebayo Felix Adekoya. A systematic review of fundamental and technical analysis of stock market predictions. *Springer Link*, 2020.

[8] Amin Shoukry Rasha AbdelKawy, Walid M. Abdelmoez. A synchronous deep reinforcement learning model for automated multi-stock trading. *Springer Link*, 2021.

[9] Thomas Mikosch Torben G. Andersen, Richard A. Davis Jens-Peter Kreiß. *Handbook of Financial Time Series*. Springer, 2009.

[10] Yves Demazeau Weidong Zhu. Multi-layer coupled hidden markov model for cross-market behavior analysis and trend forecasting. *IEEE Access*, 2019.

[11] Guofeng Yan Yongfeng Wang. Survey on the application of deep learning in algorithmic trading. *AIMS*, 2021.

[12] Ha Young Kim Gyeeun Jeong. Improving financial trading decisions using deep q-learning: Predicting the number of shares, action strategies, and transfer learning. *Elsevier*, 2019.

[13] Victor Chang Yuming Li, Pin Ni. Application of deep reinforcement learning in stock trading strategies and stock forecasting. *Springer Link*, 2020.

[14] Yu Li Min Yang Ying Shen Kai Lei, Bing Zhang. Time-driven feature-aware jointly deep reinforcement learning for financial signal representation and algorithimic trading. *Elsevier*, 2020.

[15] Yahoo finance api. `https://pypi.org/project/yfinance/`. Accessed: 2023-11-20.

[16] Stock tweets for sentiment analysis and prediction. `https://www.kaggle.com/datasets/equinxx/stock-tweets-for-sentiment-analysis-and-prediction?select=stock_tweets.csv`. Accessed: 2023-11-20.

[17] `https://spacy.io/universe/project/eng_spacysentiment`. Accessed: 2023-11-20.

[18] `https://bokeh.org`. Accessed: 2023-11-20.

[19] `https://matplotlib.org`. Accessed: 2023-11-20.

[20] `https://seaborn.pydata.org`. Accessed: 2023-11-20.

[21] `https://www.statsmodels.org/stable/generated/statsmodels.tsa.ar_model.AutoReg.html`. Accessed: 2023-11-20.