Time-Series Forecasting using ARIMA and SARIMAX

Karishma Bhargava, Christian Cassanelli, Ashna Jain, Ayush Rastogi, Benjamin Rice, Markus Zaba

Northeastern University, Boston, MA, USA

Disclaimer: We are not professional financial analysts. Although all recommendations in this report are backed by data and research, you should always perform your own research and consult a professional when making financial decisions.

Abstract

This report presents a data analysis of the performance of stocks in the healthcare, technology, and consumer retail industries from 2018-2023, as well as predicted performance for 2024. Various data analysis techniques, including ARIMA and SARIMAX time series forecasting models, were employed to assess, visualize, and forecast stock performance trends. The primary goal of this project was to use this analysis to predict an optimal investment portfolio for 2024, for both an aggressive and conservative investor. The final product includes insights into the relative performance of the industries, identification of potential investment opportunities and portfolio strategies, and a foundation for further research in the field of financial data as this analysis can be increased in scale. The findings contribute to a better understanding of stock market dynamics and can assist investors in making informed decisions.

Introduction

The primary objective of this project was to utilize predictive models and historical performance of two stocks in each of the Consumer/Retail, Healthcare, and Technology industries to construct investment recommendations for any investor to optimize their portfolio for 2024. We selected the following stocks to perform analysis on: Qualcomm (QCOM) and Nvidia (NVDA) within the technology/software industry, CVS (CVS) and McKesson Corp (MCK) within the healthcare industry, and Starbucks (SBUX) and Walmart (WMT) for the consumer/retail industry. Our main research goal aimed to first determine historically which stocks yielded the highest return, and then predict the returns of those stocks using various time series forecasting models and supplementary analysis to develop investment decisions.

The significance of this project lies in providing valuable insights to those who seek to make informed decisions in structured portfolio investments. It is important to note that while the sample within each industry in our analysis is small, this analysis is scalable and can be applied to cover more companies within the same industry as well as other industries. We chose to examine the healthcare, technology and consumer retail sectors for several reasons. These sectors offer growth, innovation and are extremely important in the global economy. They have shown resilience during challenging times, such as economic downturns or global crises. Healthcare

provides vital services and pharmaceutical advancements, technology enables remote work and digital transformation, and consumer retail has experienced shifts in behavior and e-commerce growth due to the Pandemic.

This analysis can aid in identifying sectors that outperform or underperform the broader market, thus offering guidance for portfolio diversification and allocation strategies. Using five years' worth of daily historical data, our goal was to use time series forecasting to make predictions about the future of these stocks and their respective industries and inform strategic decision making. ARIMA (AutoRegressive Integrated Moving Average), SARIMA (Seasonal Auto Regressive Integrated Moving Average), and SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous factors) models are time series forecasting models that we implemented in our analysis.

Data Sources and Methods

The data for this project was acquired from the Nasdaq website, specifically the historical stock price data section. Each stock was provided as a separate CSV file which was used to generate a multi-table database. After acquiring the data, we cleaned the data using pandas to remove unnecessary columns, such as the open, high, and low. In addition, we added columns to clarify the ticker and industry each stock was in for the sake of comprehensibility. In order to manipulate the data, we assigned the correct data type for each column such as integers for the price columns and datetime for the 'date' column. We also utilized str.replace to remove any additional characters such as "$" in order for the data type casting to work. For our analysis, we utilized matplotlib to generate a visualization of the historical stock prices as well as the statsmodels library to create our time series models (ARIMA, SARIMA, SARIMAX) for our predictions. Furthermore, we created dictionaries to develop additional visualizations.

Analysis

Given our research question of finding the industry that yielded the greatest returns, we analyzed two prominent companies within three different industries (healthcare, semiconductors, and consumer retail) to create a diversified portfolio.

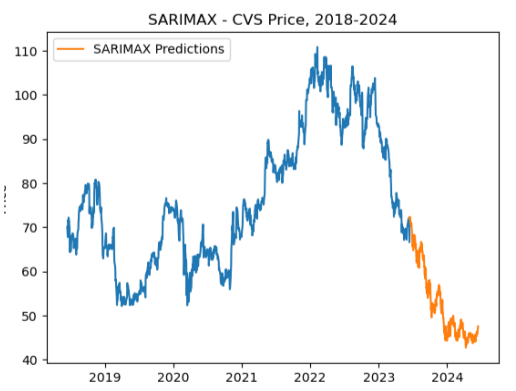

Figure 1: CVS SARIMAX model

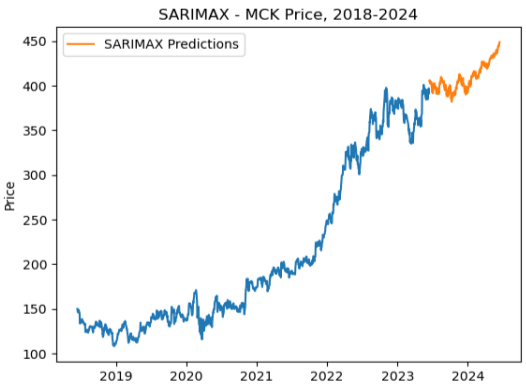

Figure 2: McKesson SARIMAX model

Within the healthcare industry, we analyzed CVS *(Figure 2)* and McKesson (MCK) *(Figure 3)*. CVS is uniquely impacted by consumer-related factors such as the pandemic and inflation rates due to their business model. During the pandemic, CVS supplied COVID tests and vaccines boosting business as seen in *Figure 2* during 2021. In 2022, the graph shows business plateauing coming out of the pandemic. The dip in early 2023 is due to the company lowering the projected stock price range by $0.20 per share because of high acquisition costs (2). This is part of the company's efforts to transition away from a consumer retail company into solely a healthcare provider; due to their transition, the stock could follow patterns of the healthcare industry more closely next year.

Switching over to MCK, they are a leader in pharmaceutical drug sales controlling 37% of the market (4). They also partnered with the US government to help distribute COVID vaccines allowing the company to overcome the market downturn due to the pandemic as people relied on the company as seen in *Figure 3*. The company also took swift measures to overcome supply issues and implement new systems. MCK is expanding into the oncology sector allowing for a boost in clinical trials and drug development shown in the model's predicted company growth in *Figure 3*.

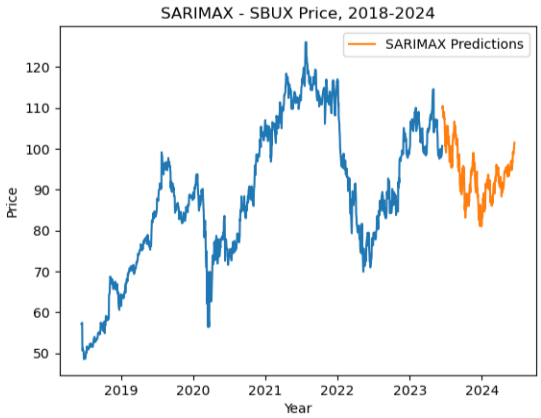

Figure 3: Starbucks SARIMAX model

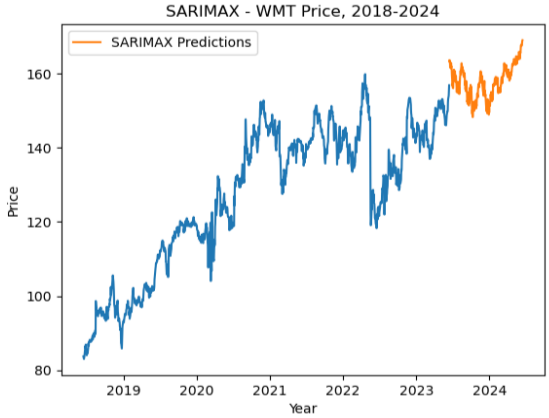

Figure 4: Walmart SARIMAX model

Within the consumer retail industry, Starbucks experienced a sharp decline of 35% in 2022 which is largely attributed to global economic challenges, supply chain disruptions, and changing consumer behavior. However, to mitigate pandemic effects, Starbucks shifted their business model towards increasing their digital presence, which increased their earnings therefore increasing their stock value. Despite their hardships in 2022, Starbucks is benefiting from robust comps and store growth leading to the uptick in stock price in the second half of 2022.

Switching over to Walmart, the company was able to remain relatively stable throughout the challenges of the pandemic. In mid-2022, Walmart revealed that due to increasing inflation levels, the company's operating costs significantly increased leading their stock to drop as seen in *Figure 4*. However, this was a general industry trend. Walmart has been able to outperform the

industry due to expansion of their online presence, new global markets, and an increased focus on grocery delivery and pickup services to cater to shifting consumer preferences.

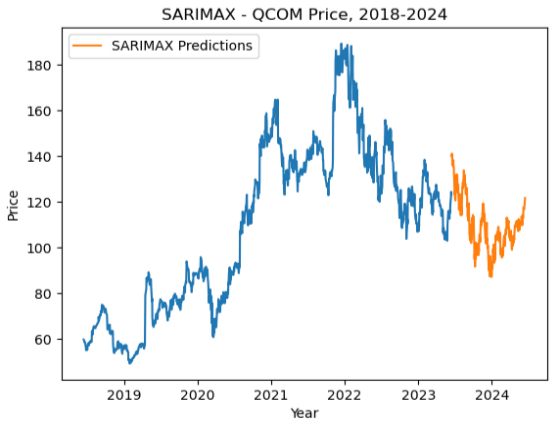

Figure 5: Qualcomm SARIMAX model

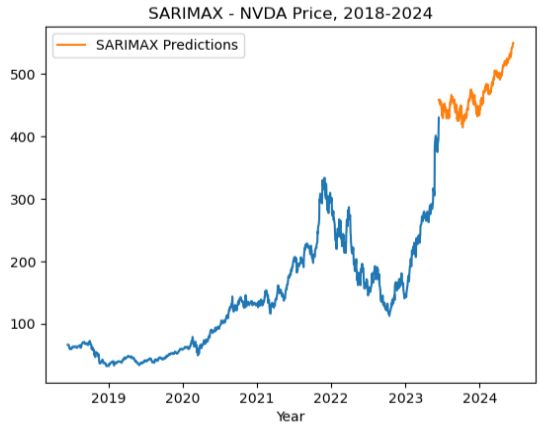

Figure 6: Nvidia SARIMAX model

NVDA and QCOM are both prominent technology companies operating in the semiconductor industry, but as seen in *Figure 5* and *Figure 6* the firms have stark differences in their projected growth rates. QCOM saw massive growth a couple years ago due to its position as the global 5G IP leader; however, with a 14% decline in global cellphone shipments in 2023, Qualcomm's stock value went down due to a resulting 17% decline in sales (7). However, at the turn of the current year, NVDA stock skyrocketed, and the company's success is attributed to its dominance in GPUs and leveraging its GPU technology for AI development, particularly generative AI. Due to the surge in popularity as well as demand of this technology, NVDA earnings skyrocketed and pushed the company towards a $940 billion valuation. Upon reaching one trillion dollars, NVDA will be the first chip manufacturer to reach this mark (8).

It is clear that NVDA is forecasted in a much more positive direction than QCOM, despite being major players in the same industry. NVDA's focus on AI chips for the last decade have deservedly put them in the spotlight, whereas QCOM's reliance on smartphone companies have put the business in a less favorable position. Despite the lopsidedness of the two companies' stocks, it should be noted that the technology industry is very volatile; what may seem like an advantage or disadvantage now could very well change in the near future.

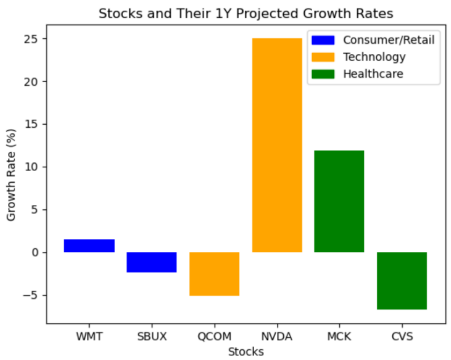

Figure 7. **Projected 1Y Growth Rates**

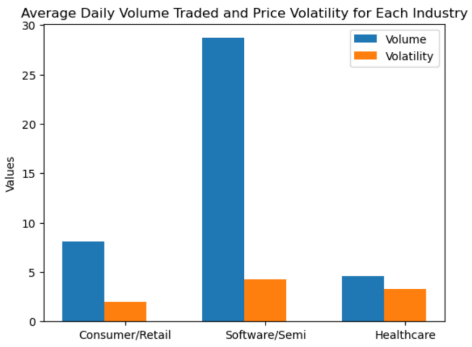

Figure 8: **Average Daily Volume Traded and Price Volatility by Industry**

Tech stocks reported the highest, but most risky, predicted growth from our models. Alongside the highest potential return came the highest spread in predicted growth rates, volume traded, and volatility in any industry we analyzed. Technology is known to be a highly active and constantly changing industry, which is supported by our data, and aggressive investors who enjoy a higher potential return regardless of higher risk should include tech stocks as their primary portfolio investment. Consumer/Retail stocks proved otherwise, being predicted to lose the least value, but also to grow at the slowest rate and show the least price fluctuation. They are a good investment for a risk-averse investor who is "bearish" on the market (believes it will go down), prefers long term stability, and dislikes large fluctuations in their portfolio value. Healthcare stocks tend to be more recession-proof as services like healthcare and health products will always be needed, and can therefore provide some stability to aggressive, tech heavy investors in case of economic downturn. Healthcare is also often disrupted by technological developments in the field and can grow quickly with the discovery of more efficient drugs, tests, or other medical breakthroughs. This speaks to the moderate volatility of the industry, and stocks pouring money into R&D and product development could be profitable for risk-averse investors looking to increase returns without sacrificing too much stability.

Based on this analysis, our final recommendation for risk-tolerant investors would be to invest more heavily in tech stocks, with healthcare as a balancing factor. Risk-tolerant investors won't gain much satisfaction over the next year from consumer and retail stocks based on this analysis. Risk-averse investors should utilize consumer/retail stocks as their foundation if they are bearish on the market, and healthcare if they are bullish. If bearish, they should avoid tech completely; if bullish, it would be smart to add a small amount of tech into the portfolio given technology wins in such a market.

Conclusion

Our initial goal was to use predictive models to gain an understanding of the relationship between past performance and future performance in the stock market, in order to construct an optimal portfolio for 2024. By selecting unique stocks in three very different industries, we hoped to achieve diverse results from our predictive models and gain a greater breadth of understanding of market movements and effects in different sectors.

We did accomplish this initial goal. Each of our 6 stocks had unique forecasted trends from our SARIMAX models- even both stocks from the same industry. Because each stock had a unique predictive model and reasoning for its performance, we had enough strong and diverse evidence from background research on each company to effectively connect this analysis to recommendations for aggressive and conservative investors looking to build a portfolio for 2024.

Tech stocks proved to be the best stocks in an aggressive portfolio, posing the highest potential return but also high volatility, volume traded, and spread between predicted growth rates. Consumer/Retail proved to be a strong conservative investment, with low predicted growth

rates in both a positive and negative direction, as well as the lowest volatility of any sector. Healthcare fell in between in every category and is a good portfolio balancer for both aggressive and conservative investors because of its recession-proof nature.

Although we were able to develop concrete recommendations, our analysis still posed multiple limitations, the first being the extensiveness of our model. Predictive financial models are limited in the amount of time they can accurately forecast because of the dynamic nature of the stock market and the plethora of factors that impact stock price. We were only confident projecting price out to 2024, as macroeconomic changes and major business developments are much more likely to be a disruptive factor after one year. Another limitation was creating a concise and efficient analysis with the understanding that successful portfolios take in mass amounts of stock analytics and companies to develop recommendations. Although we were able to complete a robust analysis, more iterations with other companies and sectors could have helped round out our analysis and provide greater insight. However, our analysis structure is easily scalable, and could incorporate other stocks, indices, and more performance metrics in the future. Regardless of limitations, our research and analysis on the performance of stocks in our chosen industries allowed for tangible conclusions and recommendations for risk-tolerant and risk-averse investors looking to maximize returns in 2024.

Author Contributions

To create this project, Christian did the bulk of the coding analysis and model creation as well as data cleaning. Ben contributed to data cleaning and coding analysis, as well as presentation preparations. Karishma created the database, created the growth rate analysis chart, and helped write the report. Markus developed the research question, created the portfolio recommendations and supplementary volatility/volume graphs, as well as the conclusion for the report. Ashna wrote the analysis on the healthcare and consumer retail sectors. Ayush wrote the analysis on the technology industry. We all contributed to research that was used in selecting the companies as well as understanding the history behind each company's individual market performance.

.

References

1. “Statsmodels.tsa.arima.model.ARIMA — Statsmodels.” *Www.statsmodels.org*.

2. Team, Trefis. “Should You Buy CVS Stock at $70?” *Forbes*.

3. Renauer, Cory. “CVS Health Is a Screaming Buy on the Dip.” *The Motley Fool*, 9 May 2023.

4. Daily, Investor’s Business. “IBD Stock of the Day: Why This Drug Giant Finally Broke Out.” *Investor’s Business Daily*, 16 June 2023.

5. “McKesson Signs up with Warp Speed to Distribute COVID-19 Shots in U.S.”.

6. “Pandemic Lessons Learned on Supply Chain Resilience and Agility - Keelvar.” *Www.keelvar.com*.

7. Leswing, Kif. “Qualcomm Gives Light Forecast, Phone Chip Sales Fall 17%.” *CNBC*, 3 May 2023.

8. “Nvidia Stock Surges Off Huge AI-Focused Earnings Report” *Forbes*.

9. Meyersohn, Nathaniel. “The Real Reasons Stores such as Walmart and Starbucks Are Closing in Big Cities | CNN Business.” *CNN*, 12 May 2023.

10. “WMT | Walmart Inc. Stock Price & News.” *WSJ*.

11. Bowman, Jeremy. “How Walmart Has Changed in the Last 5 Years.” *The Motley Fool*, The Motley Fool, 29 Jan. 2018.

12. Nassauer, Sarah. “Bargain Hunters Turn to Walmart, Boosting Sales.” *Wall Street Journal*, 18 May 2023.

13. “Starbucks (SBUX) Banks on Solid Comps Growth amid Inflation.” *Yahoo Finance*, 22 May 2023.

14. Kalogeropoulos, Demitri. “Why Starbucks Stock Is down 35% so Far in 2022.” *The Motley Fool*, 6 July 2022.

15. Auhl, Miranda. “What Is an ARIMA Model?” *Medium*, 9 Aug. 2021.

16. https://www.facebook.com/jason.brownlee.39. “A Gentle Introduction to SARIMA for Time Series Forecasting in Python.” *Machine Learning Mastery*, 16 Aug. 2018.

17. Blant, Jesse. “Explaining the Components of SARIMAX Using an Example.” *Medium*, 10 Feb. 2021.