CS 7265 : Big Data Analytics –Summer 2023, Instructor: Dr. Dan Lo

Kennesaw state university

COCA COLA sales (stock market) forecasting using econometric modeling

Tamilkumar Subbarayak
Department of Computer Science
Kennesaw State University

Yasaswi Dendukuri
Department of Computer Science
Kennesaw State University

Sunil Kumar Uridi
Department of Computer Science
Kennesaw State University

Dhanusha Sonari
Department of Computer Science
Kennesaw State University

Sai Dinesh Oggu
Department of Computer Science
Kennesaw State University

Abstract

This study analyzes Coca-Cola's competitive landscape, market share, and strategy using the ARIMA model and product analysis to forecast sales. It considers retail marketing, historical sales data, seasonal trends, growth trends, regional differences, and exports. The study also considers macroeconomic factors like spending, as well as the impact of the stock market on consumer behavior and purchasing decisions. The research identifies patterns and trends in Coca-Cola's products and their relationship to sales, addressing changing consumer preferences and healthy behaviors. The study evaluates marketing and advertising activities, including advertising, new products, and promotions, based on changing investor preferences. The research updates sales forecasts based on new information and market conditions, referencing financial reports, market research data, and marketing professionals for accurate forecasting.

Research Statement:

This study intends to investigate the usefulness of adding cutting-edge machine learning methods in Coca-Cola's sales forecasting. The project aims to increase the precision and dependability of sales predictions via the use of big data analytics, artificial intelligence, and predictive modelling, allowing for improved resource allocation within the organization and more informed decision-making.

Conjecture:

It is hypothesized that combining large-scale data analysis with machine learning methods, such as neural networks or random forests, can considerably improve the accuracy of sales estimates for Coca-Cola. The proposed approach has the potential to outperform conventional forecasting techniques and provide more actionable insights for strategic planning and inventory management by effectively capturing complex relationships among various internal and external factors, including economic conditions, consumer preferences, marketing campaigns, and distribution channels.

Literature Review

1. "Sales Forecasting Methods and Techniques: A Comparative Analysis in the Context of Coca-Cola" With an emphasis on how well they work for Coca-Cola's sales forecasting, this evaluation looks at numerous beverage sector sales forecasting methodologies and procedures. In addition to comparing time series analysis, regression models, and machine learning algorithms, it contrasts quantitative and qualitative methods.

2. "Factors Influencing Sales Forecast Accuracy: A Case Study of Coca-Cola" The primary determinants of Coca-Cola's sales prediction accuracy are examined in this study. It investigates the effects of internal (such as marketing tactics, pricing, and product portfolio) and external (such as monetary conditions, consumer trends, and rivalry activities) elements on the accuracy of sales forecasting.

3. "The Role of Big Data Analytics in Coca-Cola's Sales Forecasting: A Comprehensive Review "
   This review examines Coca-Cola's use of big data analytics in sales forecasting. It looks at the aggregation, processing and analysis of large amounts of data from many sources (such as social media, IoT devices,

transaction data, etc.) to improve the accuracy of sales forecasts and provide real-time insights.

4. "Sales Forecasting for New Product Launches: Lessons from Coca-Cola" With respect to Coca-Cola, this assessment focuses on difficulties and Methods associated with sales forecasting for new product releases. It looks at techniques for estimating demand, analyzing market potential, and forecasting sales of new beverages or product line expansions.

5. "Cross-Channel Sales Forecasting: A Case Study of Coca-Cola's Omnichannel Approach" This study examines how Coca-Cola incorporates various sales channels, including food service, e-commerce, and retail, into their sales forecast. It addresses methods of combining data from multiple channels to increase accuracy and looks at the benefits and problems of cross-channel forecasting.

6. "Demand Forecasting and Inventory Management: A Review of Coca-Cola's Supply Chain Practices" The interaction between inventory management and sales forecasting in the supply chain of Coca-Cola is examined in this research. It examines how precise projections of sales aid in inventory optimization, stockout prevention, and operational efficiency.

7. "The Impact of Seasonality on Coca-Cola Sales Forecasting: A Review of Approaches and Models"The interaction between inventory management and sales forecasting in Coca-Cola's supply chain is examined in this research. It examines how accurate estimates of sales support inventory optimization, stockout prevention, and operational efficiency

8. "External Factors in Sales Forecasting: A Comparative Analysis of Coca-Cola's Market-Specific Approaches" This research evaluates how Coca-Cola includes outside variables into estimate of sales across several markets, including monetary situations, legislative changes, and cultural preferences. It emphasizes the significance of market-specific strategies in identifying distinctive elements which influence sales.

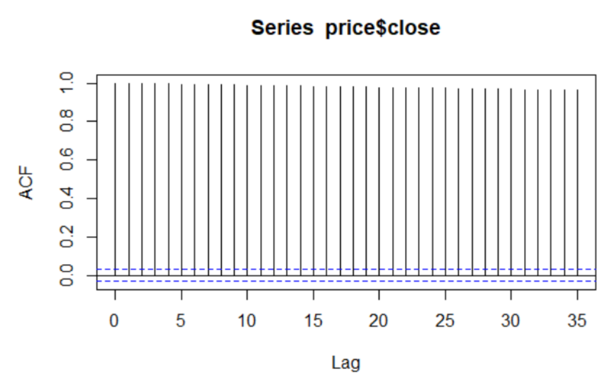
9. "Forecast Accuracy Evaluation Metrics for Coca-Cola Sales Forecasting: A Review" The accuracy of Coca-Cola's sales estimates is examined in this assessment of the evaluation measures. It contrasts frequently employed measures like mean absolute percentage error (MAPE), mean squared error (MSE), and forecast bias, pointing out their advantages and disadvantages in the context of sales forecasting.

Methodology and data

10. Data source - Secondary data taken from yahoo finance,
11. An observation taken from 2008-2023 for analysis
12. We took the help of ARIMA (Auto Regressive integrated Moving Average) model using R studio

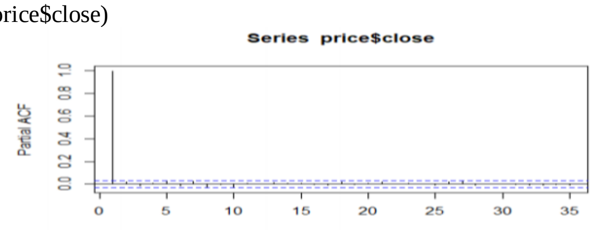
Data analysis With ARIMA model

We can see an upward sloping with a drift.



There is high autocorrelation, as these spikes are going above the blue dotted line.

pacf(price$close)



This diagram shows that there is a Partial auto correlation which has no issues as these spikes are not going above the blue line.

adf.test(price$close)
Augmented Dickey-Fuller Test
data: price$close
Dickey-Fuller = -4.0554, Lag order = 15, p-value = 0.01
alternative hypothesis: stationary
Warning message:
In adf.test(price$close) : p-value smaller than printed p-value
From the Augmented Dicky Fuller test, we can say that the given data is stationary.
pricemodel = auto.arima(price$close, ic="aic", trace = TRUE)
Fitting models using approximations to speed things up...
ARIMA(2,1,2) with drift : 4465.736
ARIMA(0,1,0) with drift : 4474.364
ARIMA(1,1,0) with drift : 4473.301
ARIMA(0,1,1) with drift : 4472.616
ARIMA(0,1,0) : 4474.659
ARIMA(1,1,2) with drift : 4475.476
ARIMA(2,1,1) with drift : 4476.697
ARIMA(3,1,2) with drift : 4458.578
ARIMA(3,1,1) with drift : 4467.409
ARIMA(4,1,2) with drift : 4419.632
ARIMA(4,1,1) with drift : 4417.761
ARIMA(4,1,0) with drift : 4452.879
ARIMA(5,1,1) with drift : 4424.271
ARIMA(3,1,0) with drift : 4471.985
ARIMA(5,1,0) with drift : 4441.818
ARIMA(5,1,2) with drift : 4423.475
ARIMA(4,1,1) : 4418.467
Now re-fitting the best model(s) without approximations...

ARIMA(4,1,1) with drift : 4413.29
Best model: ARIMA(4,1,1) with drift
Price model
Series: price$close
ARIMA(4,1,1) with drift
Coefficients:
ar1 ar2 ar3 ar4 ma1 drift
-0.8326 -0.0072 -0.0258 -0.1017 0.8120 0.0103
s.e. 0.0454 0.0208 0.0208 0.0169 0.0435 0.0063

sigma^2 = 0.1809: log likelihood = -2199.65

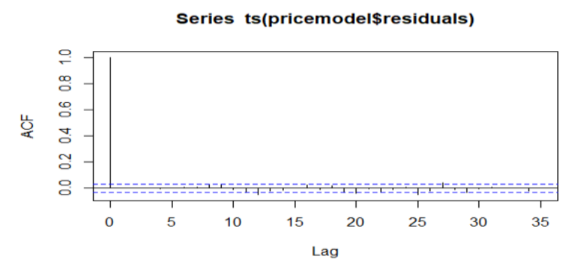
AIC=4413.29 AICc=4413.32 BIC=4457.18

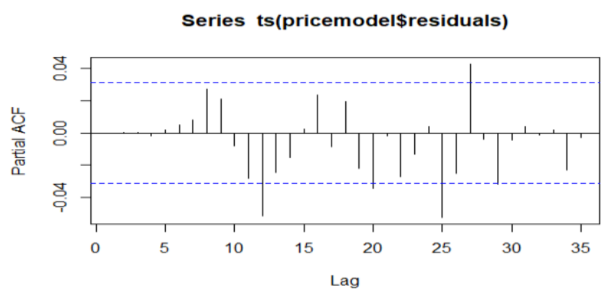
acf(ts(pricemodel$residuals))

From the above result we can inscribe that The result of the ARIMA modeling using the price close data suggests that the best model for forecasting is ARIMA(4,1,1) with drift. This model indicates that the price close variable is influenced by its own lagged values up to four periods ago, and it also includes a moving average component. The drift term accounts for any linear trend or constant term in the data.
The coefficients of the model provide information on the strength and direction of the relationships between the lagged values and the current value of the price close variable. The estimated coefficients are as follows:

13. ar1: -0.8326
14. ar2: -0.0072
15. ar3: -0.0258
16. ar4: -0.1017
17. ma1: 0.8120
18. drift: 0.0103

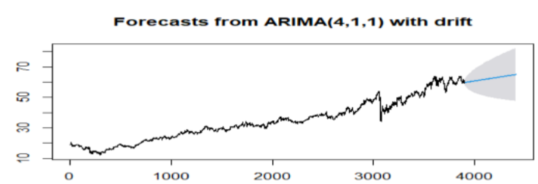
Each coefficient represents the impact of a specific lagged value on the current value of the price close variable. For example, ar1 (-0.8326) represents the effect of the first lagged value, ar2 (-0.0072) represents the effect of the second lagged value, and so on. The ma1 coefficient (0.8120) represents the impact of the moving average component.
The standard errors (s.e.) associated with each coefficient indicate the uncertainty or variability in the estimated coefficients.
In summary, the ARIMA(4,1,1) model with drift suggests that the price close variable is influenced by its own past values up to four periods ago, and it also incorporates a moving average component. The drift term captures any overall linear trend in the data. These coefficients and the model as a whole can be utilized to forecast future values of the price close variable.





priceforecast = forecast(pricemodel, level = c(95), h = 2*253)
priceforecast

plot(price forecast) the formula we use to derive the next two year stock sale forecats for next 2 years.



The above diagram show that the stock market is going to give a increasing trend for next two year that is 2024 and 2025 , and the result is backed by the above result by applying the m formula {(Priceforecast = forecast (price model, level = c(95), h = 2*253)}

Conclusion

This study aims to improve Coca-Cola's stock market forecasting by analyzing its competitive landscape, market share, strategy, historical sales data, seasonality, regional differences, and product output. It also examines the macroeconomic environment, consumer spending patterns, marketing, and advertising activities to develop effective strategies. The study continuously evaluates and updates sales forecasts based on market trends and financial reports, collaborating with marketing professionals. This comprehensive analysis maximizes Coca-Cola's sales performance, enabling informed decision-making for future growth and success.

Future work

Future work can be able to address the research methods to fill the current research gap. by constituting real-time forecasting models using streaming data and providing continuous sales updates. Explore the use of predictive analytics and real-time data to enable faster decision making and effective marketing and sales Strate optimization. Integrating Demand-Side Dynamics: Selling Coca-Cola Examining the Effects of Changing Consumers' Diet, Health Behavior, and Beverage Design to Learn More About Coca. Conduct surveys or focus groups to gain direct customer insight, then incorporate the results into predictive models. Consider issues such as the growing need for alternative health options, environmental concerns, and changing culture to better predict the future of the economy**.**

References

1. Hubbard, C. L., Mohn, N. C., & Reid, J. C. (1979, August). The Backcast: A Sales Performance Evaluation Tool at Coca-Cola. *Interfaces*, *9*(4), 38–45. https://doi.org/10.1287/inte.9.4.38
2. Kamalov, F. (2020, May 4). Forecasting significant stock price changes using neural networks. *Neural Computing and Applications*, *32*(23), 17655–17667. https://doi.org/10.1007/s00521-020-04942-3
3. Schmidt, A., Kabir, M. W. U., & Hoque, M. T. (2022, January 30). Machine Learning Based Restaurant Sales Forecasting. *Machine Learning and Knowledge Extraction*, *4*(1), 105–130. https://doi.org/10.3390/make4010006
4. Tang, L., Li, J., Du, H., Li, L., Wu, J., & Wang, S. (2022, February). Big Data in Forecasting Research: A Literature Review. *Big Data Research*, *27*, 100289. https://doi.org/10.1016/j.bdr.2021.100289
5. Van Calster, T., Baesens, B., &Lemahieu, W. (2017, November). ProfARIMA: A profit-driven order identification algorithm for ARIMA models in sales forecasting. *Applied Soft Computing*, *60*, 775–785. https://doi.org/10.1016/j.asoc.2017.02.011
6. Infosino, W. J. (1986, November). Forecasting New Product Sales from Likelihood of Purchase Ratings. *Marketing Science*, *5*(4), 372–384. https://doi.org/10.1287/mksc.5.4.372
7. Nsomba, G. (2021, September 29). The Coca-Cola Company/Coca-Cola Beverages Africa Merger: Lessons for Robust Regional Competition Enforcement. *The Antitrust Bulletin*, *66*(4), 544–555. https://doi.org/10.1177/0003603x211045436
8. Van Calster, T., Baesens, B., &Lemahieu, W. (2017, November). ProfARIMA: A profit-driven order identification algorithm for ARIMA models in sales forecasting. *Applied Soft Computing*, *60*, 775–785. https://doi.org/10.1016/j.asoc.2017.02.011
9. Miharja, R., Kaltum, U., Primiana, I., &Sarasi, V. (2020, March 23). Evaluation of SME Supply Chain Using Methods Supply Chain Operation Reference (SCOR) (Case Study on Borondong Industry SMEs). *KnE Social Sciences*. https://doi.org/10.18502/kss.v4i6.6659
10. Kumar, S., Teichman, S., &Timpernagel, T. (2012, March). A green supply chain is a requirement for profitability. *International Journal of Production Research*, *50*(5), 1278–1296. https://doi.org/10.1080/00207543.2011.571924
11. Song, X., Wang, B., Yang, Z., &Zuo, H. (2023, April 27). Financial Analysis and Market Analysis of Coca-Cola. *BCP Business & Management*, *44*, 896–904. https://doi.org/10.54691/bcpbm.v44i.4975
12. Kumar, D. M. (2020). Marketing and its Application: Case Study of Ella Drinks Ltd: Comparative Analysis with Coca-Cola. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3595239