Will Dividends Go Up?

Predicting S&P500 Payout with Microeconomic and Macroeconomic Data

Ariel Khait| Khaitari | ariel.khait@mail.utoronto.ca
Aleksander Dobak | dobakale | alex.dobak@mail.utoronto.ca
Paul Esseyin | esseyinp | p.esseyin@mail.utoronto.ca

Table of Contents

INTRODUCTION 1
MOTIVATION 1
RESEARCH QUESTIONS AND PREVIEW 2
LITERATURE REVIEW AND CONTRIBUTION 2
DATA AND METHODOLOGY 3
DATA 3
METHODOLOGY 6
RESULTS 7
CONCLUSION 10
BIBLIOGRAPHY 11

Introduction

Motivation

Dividends are the cornerstone of stock valuation, serving as a signal of a company's strength and commitment to generating long-term shareholder value. Investors have long recognized dividends as a key component of returns and profitability. With the emergence of modern finance in the 20th century, this understanding was formalized through the dividend discount model, which values a stock as the sum of its discounted future dividend payments.[1]

The data underscore the fundamental significance of dividends. From 1987 to 2023, reinvested dividends accounted for 55% of total market returns of the All-Country World Index,[2] showcasing their significance in driving long-term gains. Moreover, companies that increased or initiated dividends have consistently delivered the highest returns among S&P 500 stocks. This is best illustrated by a $100 investment in 1973: by the end of 2023, companies that grew or initiated dividends yielded a remarkable $14,118, compared to $8,756 for companies that paid dividends but did not actively grow them. By comparison, the equal-weighted S&P 500 index returned $4,439, whereas non-dividend-paying companies yielded a meagre $843.[3]

This paper aims to predict whether dividends will go up among S&P 500 companies. Predicting dividend changes requires analysing both firm-specific factors and broader economic indicators—a complex task suited for machine learning models capable of discerning non-linear patterns. Ultimately, in constructing these models, this project grants investors a forward-looking instrument to optimize their portfolio strategies, allowing them to seek-out dividend growers.

---

[1] John Burr Williams, *The Theory of Investment Value*, (Cambridge, MA: Harvard University Press), 1938.
[2] J.P. Morgan Asset Management, "The Era of Accelerated Dividend Growth Is Upon Us", accessed November 26, 2024.
[3] Hartford Funds, "The Power of Dividends: Past, Present, and Future", whitepaper, 2021, accessed November 26, 2024.

Research Questions and Preview

The main research question is: Using previously available data on firm-specific and macroeconomic factors, can we predict an increase in a firm's total dividends over the next calendar year? We find that with a boosting tree, our best model, we achieve an 86.85% accuracy rate with a 92.97% recall rate. That is, we effectively leverage micro and macro factors to predict increases in a firm's total dividends. This raises subsidiary questions. Are microeconomic or macroeconomic features more powerful at predicting the outcome? Which specific features are most effective at predicting a dividend increase? What is the predictive power of previous dividend issuance for classifying future increases? As a preview, we find that micro features are more effective, with momentum (previous dividend issuance) being the most important feature; however, we will conclude that macro features have a role to play.

Literature Review and Contribution

During the initial stages of our research, we explored various papers in the field of finance to refine our focus. Among these, Alexiei Dingli and Karl Fournier's paper, *"Financial Time Series Forecasting – A Machine Learning Approach,"*[4] proved particularly influential in shaping our topic. This paper employed classification and regression techniques to estimate next-period stock direction and price changes, achieving an 81% accuracy in predicting future stock direction when using artificial neural networks. This inspired us to explore another dimension of firm 'strength'—dividends—focusing on the classification of dividend movements.

---

[4] Alexiei Dingli and Karl Sant Fournier, "Financial Time Series Forecasting – A Machine Learning Approach", *Machine Learning and Applications: An International Journal* 7, no. 5 (October 2017): 11-27.

To build on this idea, we examined two machine learning studies addressing dividend prediction: Ana Catarina Moreira's *"Prediction of Dividend Yields"*[5] and *"A Knowledge Integration Model for the Prediction of Corporate Dividends"* by Jinhwa Kim, Chaehwan Won, and Jae Kwon Bae[6]. Both papers focused on regression, employing models such as CART, random forest, KNN, and neural networks, which helped to inform our own selection of ML algorithms. However, most importantly, we noted two distinctive similarities shared by all three papers in our literature review. They all relied on market-derived data (volatility, momentum, etc.) rather than incorporating firm fundamentals, and they did not use macroeconomic features.

Building on these findings, we identified the two key directions for our contribution to this project. First, we decided to incorporate financial statement data. This approach would provide a granular and transparent examination of how specific financial statement features affect dividend movements. Second, and most significantly, we would include macro variables. Our hypothesis was that macro factors play a meaningful role in dividend decisions; for example, we believed that firms would be less likely to increase dividends during a deep recession. Overall, our project aimed to shift the focus from market features such as seasonality indicators, time distance to dividend announcements, and volatility, redirecting attention to features like cost of goods sold, revenue, GDP, CPI, and other macroeconomic and financial statement variables

Data and Methodology

Data

In this study, we utilized a large synthetic dataset consisting of 13,164 data points spanning 33 years, from 1990 to 2023. We retrieved 15 macroeconomic variables from the

---

[5] Ana Catarina Fernandes Moreira, "Prediction of Dividend Yields", master's thesis, Universidade do Porto, 2022.
[6] Jinhwa Kim, Chaehwan Won, and Jae Kwon Bae. "A knowledge integration model for the prediction of corporate dividends." *Expert Systems with Applications* 37, no. 2. (March 2010): 1344-1350.

Federal Reserve Economic Data (FRED)[7] and microeconomic variables from Wharton Research Data Services (WRDS)[8]. Since most of these variables are continuous, we aggregated the data by calculating yearly averages and totals to create a representative observation for each company for each year. Some observations were excluded due to insufficient information, particularly for companies that had their IPOs after 1990, resulting in the removal of 3,435 observations and leaving a final dataset of 444 unique S&P 500 companies over the 33-year period.

After aggregating data from different sources, we introduced industry dummy variables in order to assess the impact of being in a particular industry on dividend increases. We then created our dependent variable, representing dividend increases. A dividend increase was defined as the sum of dividends in period t being greater than the sum in period $t-1$. This methodology was designed to include companies paying both yearly and quarterly dividends. Drawing inspiration from the literature, where lagged variables demonstrated strong predictive power for stock prices and dividends, we applied this approach to our data.[9,10,11] Specifically, we created lagged variables for dividends, dividend increases, and industry average increases. Overall, this resulted in one dependent variable and 41 independent variables.

Analyzing the data in Figure 1, we observe a consistent upward trend in average dividends, growing faster than nominal GDP. Between 1990 and 2023, the S&P 500 experienced an annual growth rate of 10.3%, while the average dividend per company grew at a close 10.24% annually. This near-parallel growth highlights dividends as an excellent income source for investors. Even after adjusting for inflation, average dividend growth outpaces nominal GDP and

---

[7] Federal Reserve Bank of St. Louis, "Federal Reserve Economic Data (FRED)", accessed 26th October 2024. https://fred.stlouisfed.org.
[8] Wharton Research Data Services (WRDS), accessed throughout Fall of 2024.
[9] Moreira, ”Prediction of Dividend Yields”, 10.
[10] Dingli and Fournier, ”Financial Time Series Forecasting”, 14.
[11] Kim, Won, and Bae. “A knowledge integration model for the prediction of corporate dividends”, 1347.

remains comparable to market growth. This indicates that dividends could serve as an effective way for investors to diversify their portfolios or generate income distinct from capital gains.

Since dividends are non-negative, we observe a lognormal distribution in Figure 2, with an average annual dividend payout of approximately $602 million. Notably, the distribution exhibits a skewness of 6.5 and a kurtosis of 78, reflecting the presence of extreme values such as Microsoft's $32 billion special dividend in 2005. Additionally, Figure 3 highlights that over 55% of current S&P 500 companies belong to the Technology, Industrials, Healthcare, or Financial Services sectors.

Finally, Figure 4 reveals a clear inverse correlation between the percentage change in unemployment and dividends. This relationship likely reflects the broader economic environment: a fall in unemployment is usually accompanied by higher economic growth boosting revenues and profitability, which in turn fuels higher dividend payouts. Conversely, higher unemployment signals weaker economic conditions, which can coincide with lower corporate earnings and reduced dividends. Figure 4 illustrates this correlation between labor market dynamics and dividend behavior, highlighting how dividends may align with broader macroeconomic shifts.

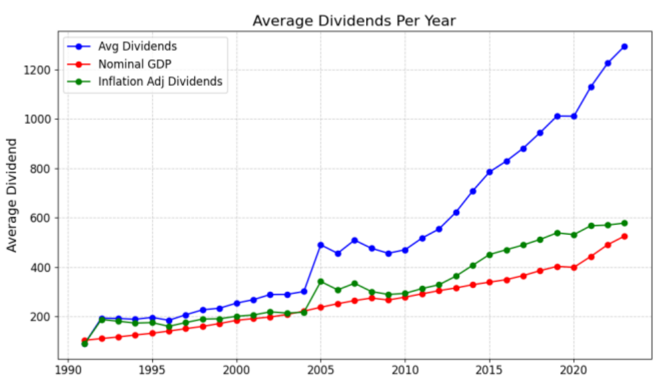

Figure 1: Average Dividends vs. Nominal GDP

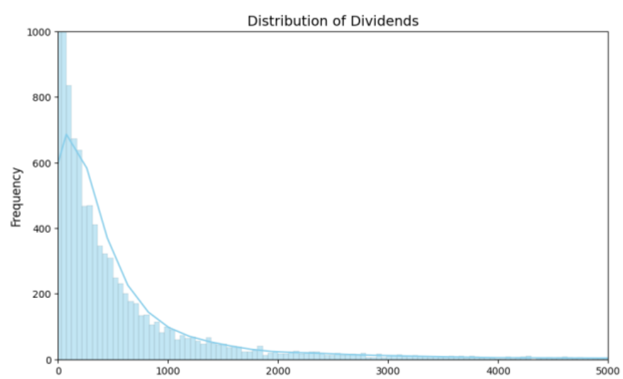

Figure 2: Distribution of Dividends

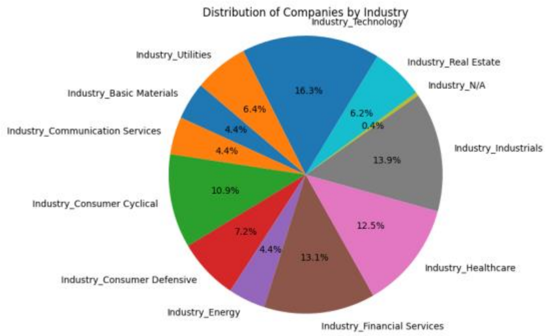

Figure 3: Distribution of Industries

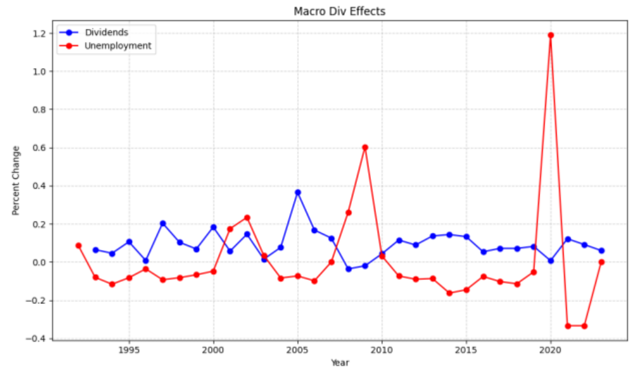

Figure 4: % Change Div vs. % Change Unemployment

Methodology

To evaluate our models, we chose to use accuracy, false positive rate, and recall rate as key metrics for comparison. The false positive rate represents the proportion of negative events that are incorrectly classified as positive. In the context of this model, it specifically refers to the percentage of non-increases in dividends that are misclassified as increases. On the other hand, the recall rate measures the proportion of actual dividend increases that are correctly predicted, relative to the total number of dividend increases.

We first performed an Augmented Dickey-Fuller (ADF) test for stationarity on all 41 features. For any non-stationary data, we transformed them by calculating the percent change. This step is crucial because it helps mitigate bias between steadily increasing data from distinct years, such as comparing 1992 and 2021. That is, by differencing our data, we obtain a consistent mean and variance, preventing the bias of higher GDP values in 2021 compared to those in 1992.

In this study, we optimized the predictive accuracy of each model using a grid search algorithm with a minimum of five-fold cross-validation. To begin our analysis, we applied a Logistic Regression (Logit) model, both with and without regularization. Given its simplicity and consistent estimators compared to other classification models, it was a good starting point, although it produced unsatisfactory results.

Next, we explored a Naïve Bayes approach, although we hypothesized limited success due to its strong assumption of independent draws, rarely observed in financial data. To address this, we applied a K-Nearest Neighbors (KNN) model, which groups and predicts based on similarities between data points, capitalizing on correlations that the Naïve Bayes model overlooks.

We then explored tree-based models, including classification trees, random forests, and gradient boosting trees. These models were implemented to address the sub-research question: Which feature is the most important? Tree-based models are highly interpretable, particularly in identifying the relative importance of features. By examining the structure of these models, we identified the features appearing at the top of the trees as the "most important". Additionally, we compared feature importance results with and without the inclusion of macroeconomic variables.

Results

The non-linear models, particularly XGBoost, demonstrated an impressive predictive accuracy of nearly 87% in forecasting whether dividends would increase in the next calendar year. Our analysis also highlighted a notable improvement in predictability when macroeconomic features were incorporated.

Basic models such as logit and KNN performed reasonably well, achieving an average accuracy of 70% with a variance of about 2%, depending on the data split. However, the naïve bayes model underperformed significantly, with an accuracy of only 65%, making it relatively weak compared to the other basic models. When evaluated against tree-based models, which consistently achieved around 80% accuracy, these basic models performed significantly worse.

Delving deeper into the performance of the tree-based models revealed intriguing insights. By running three tree-based models on datasets both with and without macroeconomic variables, we were able to deduce the substantial impact of these features on model performance. When evaluating feature importance, we observed that lagged variables played a critical role. Specifically, the sum of dividends issued by a company in the previous calendar year emerged as the most significant predictor of future dividend increases. This variable appeared as the top node in the classification tree (Figures 5 and 6) and was consistently identified as the most important feature in both the random forest and boosting tree models (Figures 7–10). Additionally, as hypothesized, macroeconomic variables, such as the percent change in unemployment, were found to be the second most important predictors across all three models. In the classification tree, macroeconomic variables appear in both the second and third level decision nodes, with percent change of unemployment's Gini value only 0.08 lower than the top node. In the gradient boosting model, three out of the five most important features included macroeconomic variables, while in the random Forest model, two out of the top five features were macroeconomic.

A comparison of Figure 11 and Figure 12 illustrates the impact of including Macroeconomic variables, resulting in a 5.43 percentage point increase in accuracy. More importantly, we observed a remarkable 17.45 percentage point reduction in the false positive rate across all three models with the inclusion of these variables. This indicates that while macroeconomic variables significantly improved accuracy, their primary value lay in reducing the number of false positives.

Overall, the gradient boosting model delivered the best results, achieving an accuracy of 86.85%, a false positive rate of 21.26%, and a recall rate of 92.97%.

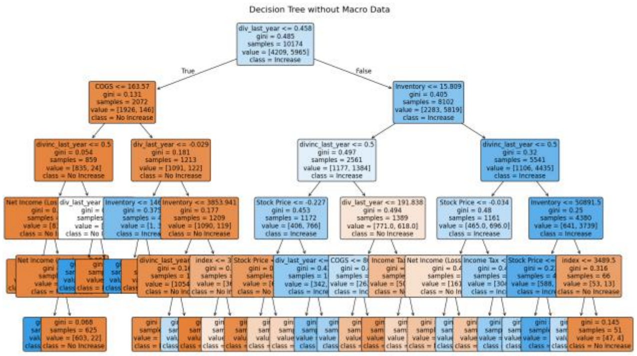

Figure 5: Classification Tree (No Macro)

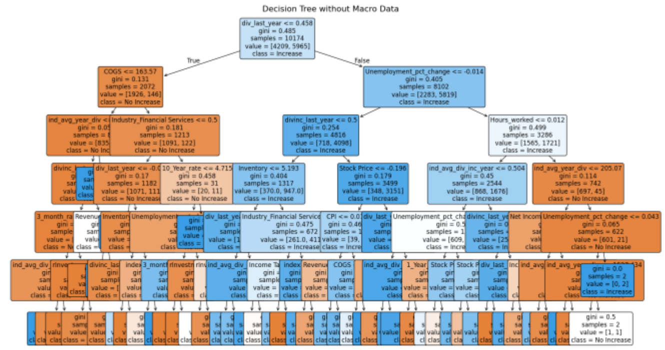

Figure 6: Classification Tree (With Macro)

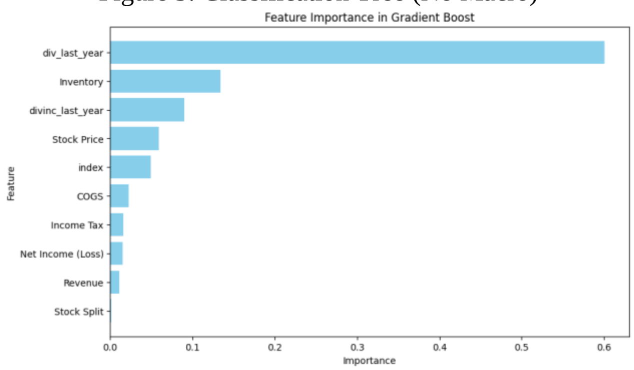

Figure 7: Feature Importance in Gradient Boost (No Macro)

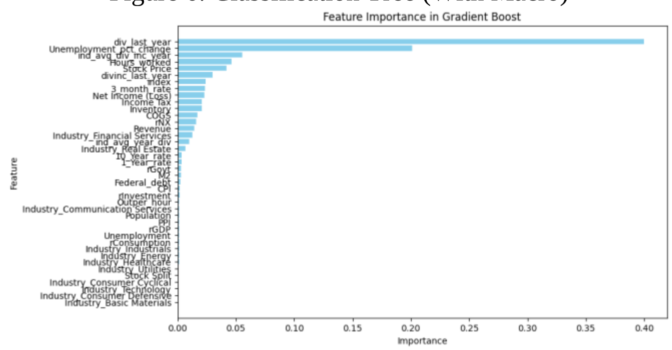

Figure 8: Feature Importance in Gradient Boost (With Macro)

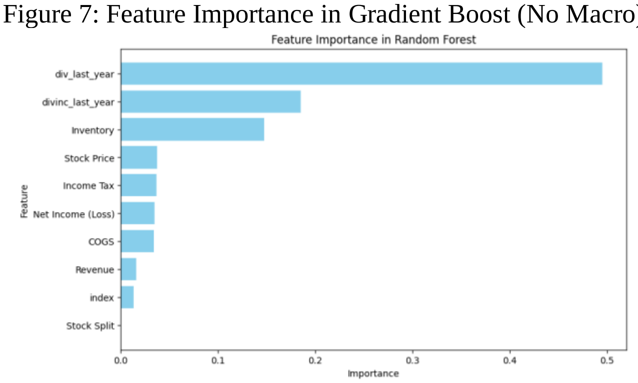

Figure 9: Feature Importance in Random Forest (no Macro)

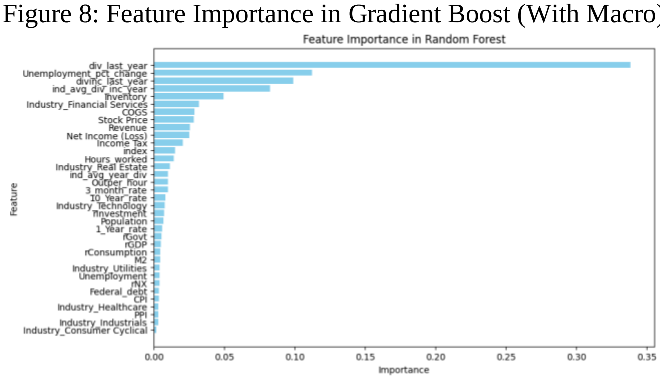

Figure 10: Feature Importance in Random Forest (with Macro)

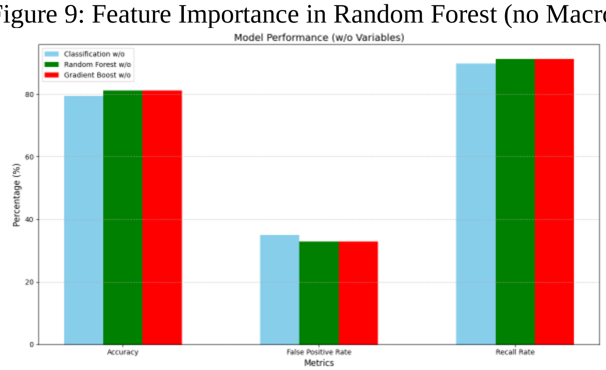

Figure 11: Model Performance without Macro Variables

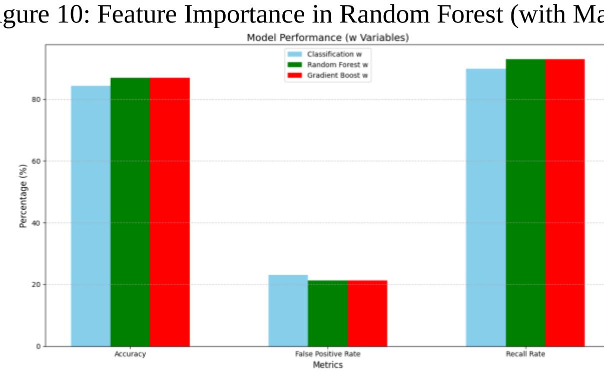

Figure 12: Model Performance with Macro Variables

Conclusion

In conclusion, to address the research questions posed at the beginning, we can indeed predict future dividend increases using macro and firm specific factors. We find that overall firm-specific features are more powerful predictors than macro variables. Dividend issuance last year was twice as important as the second-best feature in our most accurate model (gradient boost with macro variables), demonstrating the power of momentum. Nevertheless, macro features do improve predictive capacity significantly. Their most important contribution being the reduction in the false positive rate. Percent change in unemployment was the most relevant macro feature. In fact, it was the aforementioned second-best feature and was four times as important as the third-best feature in our best model. This is not entirely surprising as unemployment is a good index for the general level of health in an economy, capturing trends in growth, demand, and overall economic stability.

A potential area for improvement is incorporating sentiment analysis into our model, which could enhance our understanding of dividend increases. By analyzing news articles related to the companies and reviewing annual and quarterly filings, such as 10-K reports, we could extract sentiment indicators that shed light onto corporate strategies and market expectations. Incorporating these elements could add a new dimension to our analysis, enhancing the accuracy and robustness of our predictions. That being said, the undeniable fact remains that this research equips investors with the ability to identify and capitalize on dividend growth opportunities, enhancing their strategic decision-making.

Bibliography

Dingli, Alexiei and Karl Sant Fournier. ”Financial Time Series Forecasting – A Machine Learning Approach.” *Machine Learning and Applications: An International Journal* 7, no. 5 (October 2017): 11-27.

Federal Reserve Bank of St. Louis. "Federal Reserve Economic Data (FRED)." Accessed 26th October 2026. https://fred.stlouisfed.org.

Hartford Funds. "The Power of Dividends: Past, Present, and Future." Whitepaper, Hartford Funds, 2021. Accessed November 26, 2024. https://www.hartfordfunds.com/dam/en/docs/pub/whitepapers/WP106.pdf.

J.P. Morgan Asset Management. "The Era of Accelerated Dividend Growth Is Upon Us." Accessed November 26, 2024. https://am.jpmorgan.com/gb/en/asset-management/adv/insights/portfolio-insights/pm-perspectives/equities/the-era-of-accelerated-dividend-growth-is-upon-us/.

Kim, Jinhwa, Chaehwan Won, and Jae Kwon Bae. “A knowledge integration model for the prediction of corporate dividends.” *Expert Systems with Applications* 37, no. 2. (March 2010): 1344-1350.

Moreira, Ana Catarina Fernandes. ”Prediction of Dividend Yields.” Master's thesis, Universidade do Porto, 2022.

Wharton Research Data Services (WRDS). Accessed throughout fall of 2024.

Williams, John Burr. *The Theory of Investment Value.* Cambridge, MA: Harvard University Press, 1938.