Evaluating the Impact on Election Betting on the Stock Market in an Election Year

By: Dennis Goldenberg, Samaa Nadkarni, Kevin Shao

Project Impetus

Politics and Economics have a strong relationship, and this year was no different than the other. The American presidential election occurred on November 5th, 2025, with Donald Trump declared the winner the following day. The stock market responded, with the Tech sector (NASDAQ 100 Tech index) immediately spiking 500 points in 2 days and being up over 8.66% over the following month. The Healthcare Sector (S&P Healthcare Index) went in the other direction, dropping by 75 points over the following two weeks, a 4.28% drop. The tech sector boomed likely due to Trump's promised 'sizable deregulatory efforts', while the healthcare sector may have dropped due to his appointment of Robert F. Kennedy Jr. to reorganize the public health agencies, most notably to end the 'corporate capture' of their decision making.

This led our group to wonder: **do political headwinds impact stocks *before election day* as much as they do after?** This election contained a new wrinkle: the ability to bet on elections. On the website Polymarket, non-US bettors could wager money on who they believed would win the election by essentially "buying shares" related to a particular outcome, and the volume betted on each candidate was used to calculate implied probability of victory for said candidate. On Polymarket's website, there exists 11 months of implied probability odds of victory (from January 5th to November 5th), which we utilized to examine this question.

Data Sources and Exploration

Our response variables were the NASDAQ 100 Tech Index and S&P Healthcare Index. The Tech Index data was pulled from *MSN Money* while the Healthcare Index was pulled from *MarketWatch.com*. For our stand-in for political sentiment, we used Donald Trump's implied odds of victory from Polymarket, since the democratic nominee switched several months before the election. We wanted to control for the impact of inflation and general market volatility, so we utilized the Break-Even Interest Rate and the VIX index as measures of these two. We considered other variables, such as the 2 and 10 year market yield of Treasury securities and the dollar index, but dropped them due to lack of predictive power or multicollinearity concerns.

We explored the correlation between Trump's odds and both indices and the cross-correlograms of our exogenous variables and the response variables. Trump's odds and VIX seemed to have a relationship with the Tech Index, while Trump's odds and BEIR had a correlation with the Healthcare Index. These relationships could be spurious due to non-stationarity of the response variables. We conducted Augmented Dickey-Fuller tests as a test stationarity; the result returned that the healthcare index was NOT stationary. However, our cointegration tests showed that the non-stationary time series were co-integrated, which led us to conclude that differencing wasn't required before running time series modeling.

Training Our Models

To get the data ready for modeling, we split our data into train and test sets 80%-20%. We standardized the variables (exogenous and response) for more stable results, using the training moments to avoid look-ahead bias. Finally, we generated lagged versions of our exogenous variables; we used these for autoregression so that all predictor variables were in the past and an investor could potentially make decisions at time t.

ARIMA model

As a baseline, we decided on an Autoregressive Integrated Moving Average (ARIMA) model with the exogenous variables defined above. Note that, as ARIMA can only model one response model at a time, we had 2 separate models. We used a grid search to determine the optimal p (AR terms) and the optimal q (MA terms). Our criterion for optimality was 'test' MSE - the quotations belie the fact that we used the test set as more of a validation set due to our lack of data. For the tech index, our grid search decided on ARIMA(0,0,2) and for the healthcare index, it was ARIMA(2,0,1) - note that we did not allow the model to difference the response variables due to fear of differencing away the trend we were trying to model.

The tech index model showed that the **'trump_odds' was statistically significant at the 1% level**, and positive at that. The healthcare index model did not show a statistically significant result for 'trump_odds', and further, the optimal model turned out to effectively be a random walk. Neither model was particularly effective in capturing the underlying trend according to their respective R-squared values of -.498 and .477. This is likely because the underlying relationships between our variables and the indices are non-linear and changing, and there is some evidence that we should have differenced them.

RNN model

To capture the non-linear relationships better, we turned to a One-Layer Recurrent Neural Network. RNN's can model non-linear relationships and can model more than one output. We used one layer to avoid overfitting due to too many parameters for too little data. Our RNN structure was simple: an input layer, a simple RNN layer with 8 activations, and a dense output layer with 2 outputs (the Tech index and Healthcare Index at time *t*, respectively). Our inputs were the lagged 1 version of the Tech index, the Healthcare Index, and each of our Exogenous variables. The model had 130 parameters, and we gave it a learning rate of 0.001 to learn quicker given less data. We gave the model 500 epochs to learn, but used the stopping criterion of loss minimization to ensure the best final model was selected.

This model performed much better on the test data, with R-squared values of .636 on the Tech Index and .877 on the Healthcare Index. This gives credence to the idea that the relationships between our variables and the response was non-linear. It is notable that **predictions spiked in mid-October for predictions of the Tech Index, which corresponded to a period in which trump_odds rose**.

Conclusion and Limitations

Both the ARIMA model and the RNN model showed that the implied probability of Donald Trump winning the election based on implied betting odds had a positive relationship with the value of the NASDAQ 100 Tech index, which mirrors the post-election surge. However, the trump_odds variable did not appear to have a significant negative impact on the Healthcare index in our models, even potentially being slightly positive, which did not match the reaction post-election. Our models were not perfect, and there is a large amount of hyperparameter tuning that could have been done to improve performance. Also, given more time, we aimed to use interpretable machine learning techniques such as partial dependence plots to confirm the predictive power of the 'trump odds' variable in the RNN. Still, **there is evidence that the public's political sentiments during an election year had influence on the Stock Market**.

There are several limitations and asterisks to our results. Most notably, we only had 11 months of betting odds data from which to go off of (so about 300 data points). Models for the stock market typically use data over a much longer time frame. Furthermore, our data only spanned one election cycle - it is very possible that there were some Trump or Harris-specific movements, and that future campaigns with different candidates and policy platforms may impact financial markets in a different manner. Future analyses, with more elections on which public betting data is available, would be able to confirm or deny this.

Another difficulty was our feeding the RNN the entire test set to predict upon, rather than letting it extrapolate on its own predictions. This created a self-correcting mechanism by which, even if the RNN predictions were inaccurate for a specific day, the RNN would be fed the correct values for the next day, and correct itself. This created a lagged effect where our model often picked up on movements the day after they happened, which calls into question the model's efficacy. Future analyses should let the RNN model predict using its own predictions for the indices in the previous day, rather than the actual test data.

Sources

Cuomo, A. (2024, October 30). YouTube. https://www.youtube.com/watch?v=sDn1yKCIMfA&ab_channel=NewsNation

Goldbeck, D. (2024, August 22). *Harris vs. Trump: Assessing the potential regulatory policy paths in the 2024 election*. American Action Forum. https://www.americanactionforum.org/insight/harris-vs-trump-assessing-the-potential-regulatory-policy-paths-in-the-2024-election/

NASDAQ-100 Technology Sector Index (NDXT) Watchlist. MSN. (n.d.). https://www.msn.com/en-us/money/chart?id=a3yzqh&timeFrame=1M&chartType=baseline&projection=false

Morrow, A. (2024, November 8). *How prediction markets saw something the polls and pundits didn't | CNN business*. CNN. https://www.cnn.com/2024/11/08/business/polymarket-election-trump-nightcap/index.html

Polymarket. *Presidential election winner 2024*. Polymarket. https://polymarket.com/event/presidential-election-winner-2024?tid=1733092143511

Standard & Poors. *S&P 500 Health Care Sector index price data*. MarketWatch. https://www.marketwatch.com/investing/index/sp500.35/download-data?countrycode=xx