Ryan J Morrow

Professor Redmond

BAIS:3250 Data Wrangling

9 May 2025

Project Final Report

Do Congress Members’ Trades Outperform the Stock Market

All files relating to this project can be found at:

<https://github.com/ryan-j-morrow/BAIS3250_Final_Project>

# Introduction

The intersection of politics and financial markets has long been scrutinized, with concerns over whether lawmakers leverage privileged information for personal monetary gain. In the United States, members of Congress are legally required to disclose their stock trades under the STOCK Act, enacted in 2012, to promote transparency and deter insider trading. However, despite these regulations, public skepticism remains regarding whether congressional stock trades outperform the broader market due to potential access to non-public information.

This analysis seeks to explore whether the stock trades of U.S. Congress members in 2023 and 2024 have outperformed the overall market. The project will leverage publicly available data from Capitol Trades, which tracks congressional stock transactions, and financial market data sourced from Yahoo Finance. By analyzing the returns of stocks bought and sold by lawmakers and comparing them to market performance, this project aims to determine whether congressional trades yield statistically significant excess returns.

Beyond overall performance, this analysis will examine whether there are notable differences in trading success across political parties and congressional chambers. Specifically, it will test whether members of one party tend to achieve higher returns than the other and whether the House of Representatives or the Senate exhibits stronger market performance.

# Data

## Capitol Trades

To get data on what stocks congress members traded and when, I will scrape the website <https://www.capitoltrades.com/trades>, which contains all reported trades from congress members across the past 3 years. I will then take all trades from 2023 and 2024 to be part of my sample to cover two full periods of the stock market. To ensure an accurate return can be calculated, only tickers that the congress member has bought and sold at least once will be included in the sample. Per [Yahoo Finance](https://finance.yahoo.com/news/average-stock-holding-period-121123957.html?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAAKFkaYec2C7JwyIpiiLiynAwdUJAtZ9jYolIR9q9jPC3h6tyhYPgm7qM_SJ0Iv0JrThMJHFZx4WwhDtDJtGUW8RRYP6BnxdnMDOhttGjjkd1072sqSO7StIz3ncIt-FsDwpH5WbymOyhFgausq-nUkqGzZcDJTwmBALTEc48JNC), the average stock is held for 5.5 months, so the sample will consist of all stocks up to 4 average stock turns.

## GitHub: rreichel3/US-Stock-Symbols

The American stock market contains three leading stock exchanges: NYSE, NASDAQ, and AMEX. A few ticker symbols are shared between the exchanges, which would pose a problem to the method used to get stock prices. Thus, I will use a [data source on GitHub](https://github.com/rreichel3/US-Stock-Symbols), which has all the ticker symbols on the leading three stock exchanges. This will help me remove ticker symbols shared between the exchanges and have a list of all possible ticker symbols. Some ticker symbols scraped from Capitol Trades will be from smaller or private stock markets where prices may not be easily gathered; those will be removed from the sample.

## YFinance

To gather stock prices for both the congress-member traded stocks and the market baseline (S&P 500), the public API from Yahoo Finance will be used via the yfinance Python module. For simplicity of the dataset, the following assumptions will be used:

* Even though somebody can buy and sell the same stock multiple times due to not having data on the investment size, all stocks will be compared based on their percentage return rather than actual returns.
* Any stock in the sample must have been purchased at least once and sold at least once in the sample period of 2023 and 2024. The earliest date bought will be the Buy Date, and the latest sell date will be the Sell Date. If the Sell Date is before the Buy Date, that record will be removed from the sample.
* The S&P 500 Index will be used as the Market Baseline Stock
* The Market Baseline Stock prices will utilize the exact dates as the traded stock
* The open price on the open date will be used for the Buy Price, and the close price on the sell date will be used for the Sell Price.
* If either date falls on a weekend or market holiday, the closest previous date will be used; for example, a date that lands on a Sunday will use the price from the last Friday.

## Data Dictionary

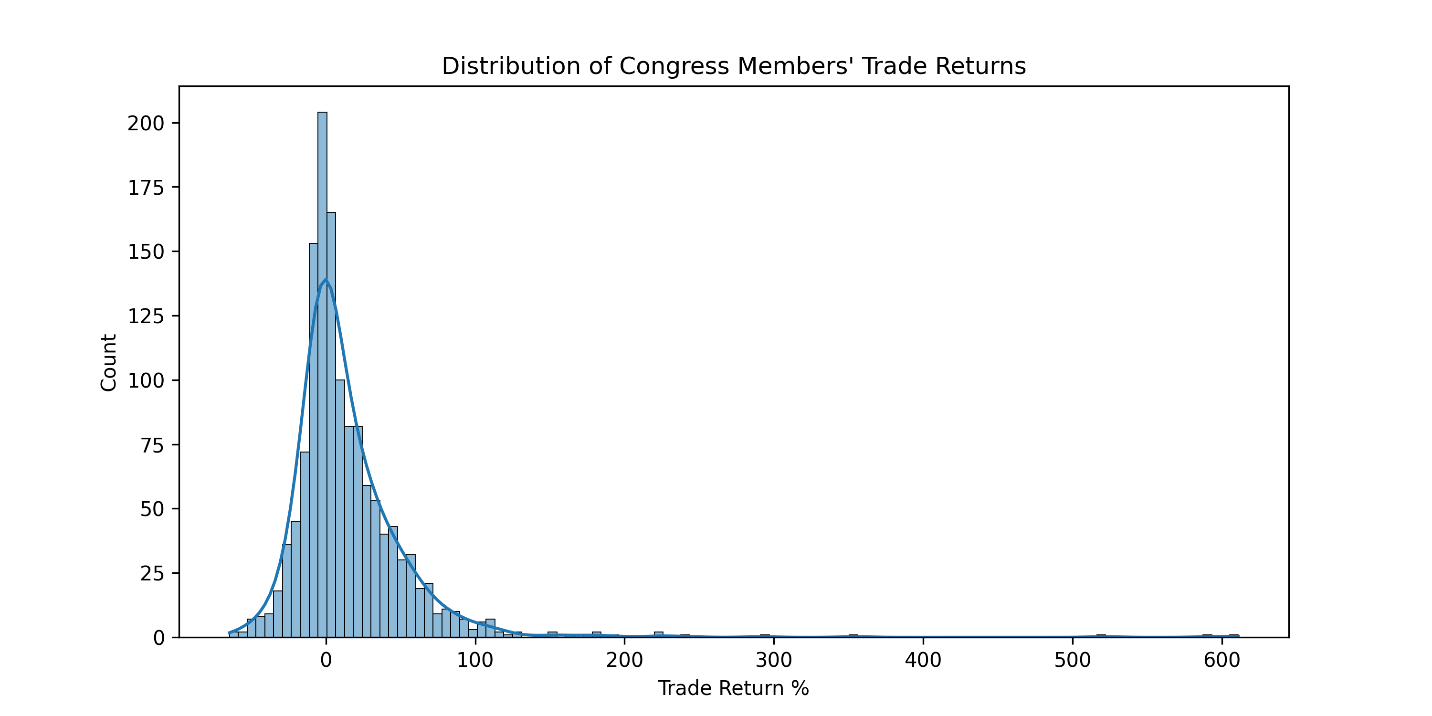
|  |  |  |
| --- | --- | --- |
| Field | Datatype | Description |
| Name | String | Congress member’s name |
| Ticker | String | The stock market ticker of the stock that the congress member traded |
| Party | String | The political party the congress member is affiliated with |
| Chamber | String | The chamber of Congress is where the congress member serves (House or Senate). |
| State | String | The state that the congress member represents |
| Buy Date | String | The earliest date within the sample that the congress member bought shares of the traded stock |
| Sell Date | String | The latest date within the sample that the congress member sold shares of the traded stock |
| Buy Price | Float | The opening price of the traded stock on the Buy Date (or closest valid date). |
| Sell Price | Float | The closing price of the traded stock on the Sell Date (or closest valid date). |
| Trade Return % | Float | The traded stock’s return as a % using:  Return % = 100 \* ((Sell Price – Buy Price)/Buy Price) |
| Market Buy Price | Float | The opening price of the market baseline stock on the Buy Date (or closest valid date). |
| Market Sell Price | Float | The closing price of the market baseline stock on the Sell Date (or closest valid date). |
| Market Return % | Float | The market baseline stock’s return as a % using:  Return % = 100 \* ((Sell Price – Buy Price)/Buy Price) |
| Overperformance % | Float | The target variable represents how much better the traded stock performed than the market baseline stock over the same period, using:  Overperformance % = Trade Return % - Market Return % |
| Overperform | Binary | This target variable represents whether the trade overperformed the market or not, using:  Overperform= Trade Return % > Market Return % |

# Analysis

## Congress Members' Performance against the Market

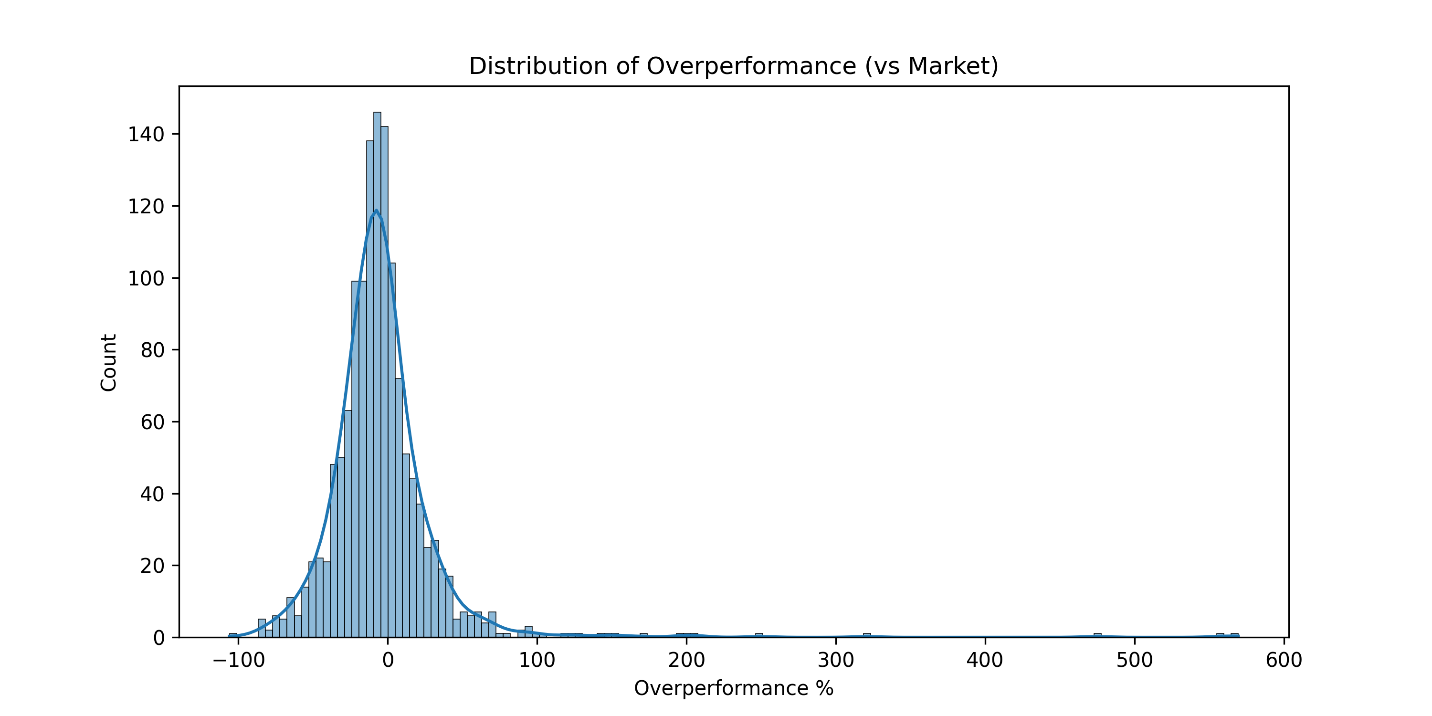
I wanted to determine whether Congress members’ stock trades outperform the broader market and how their performance compares directly to the S&P 500 benchmark. I started by calculating descriptive statistics to understand the distribution of returns from congressional trades. Across 1,356 trades, the average return was 15.38%, with a broad standard deviation of 43.79%. Returns ranged from –64.50% to +610.85%, showing that although some trades had significant positive returns, there was considerable variation overall. The distribution of returns was skewed to the right with a long tail, meaning a few extremely high-return trades pulled the average upward (Figure 1).

**Figure 1: Distribution of Congressional Trade Returns (%)**

****

However, to assess performance relative to the market, I calculated each trade’s overperformance—the percentage return above or below what the S&P 500 achieved over the same time frame. This is a better measure to determine whether Congress members are beating the market, not just generating positive returns. The average overperformance was –4.04%, meaning congressional trades underperformed the market by about 4%. The overperformance distribution appeared more normal than the raw returns but still had a long right tail, indicating a few large outperforming trades (Figure 2).

**Figure 2: Distribution of Congressional Trade Overperformance (%)**

****

I initially thought the few high-return trades might pull up the overall performance enough to beat the market. To test this more formally, I ran three hypothesis tests (Table 1). First, I tested whether Congress members’ trades overperformed the market at a statistically significant level. The p-value was 0.9999, which firmly rejects that idea. Then, I tested whether congressional trades underperformed the market. The p-value here was 0.0001, meaning there is strong evidence that Congress members’ trades do underperform the market. Finally, I tested whether congressional returns simply differed from the market (in either direction). This test had a p-value of 0.0003, showing returns were statistically different—mostly driven by underperformance and high variability.

**Table 1: Hypothesis Test Results – Overperformance of Market**

|  |  |  |
| --- | --- | --- |
| **Test** | **P\_Val** | **Result** |
| Congress members' returns overperform the market at a statistically significant level | 0.999874 | FALSE |
| Congress members' returns underperform the market at a statistically significant level | 0.000126 | TRUE |
| Congress members' returns perform differently than the market at a statistically significant level | 0.000251 | TRUE |

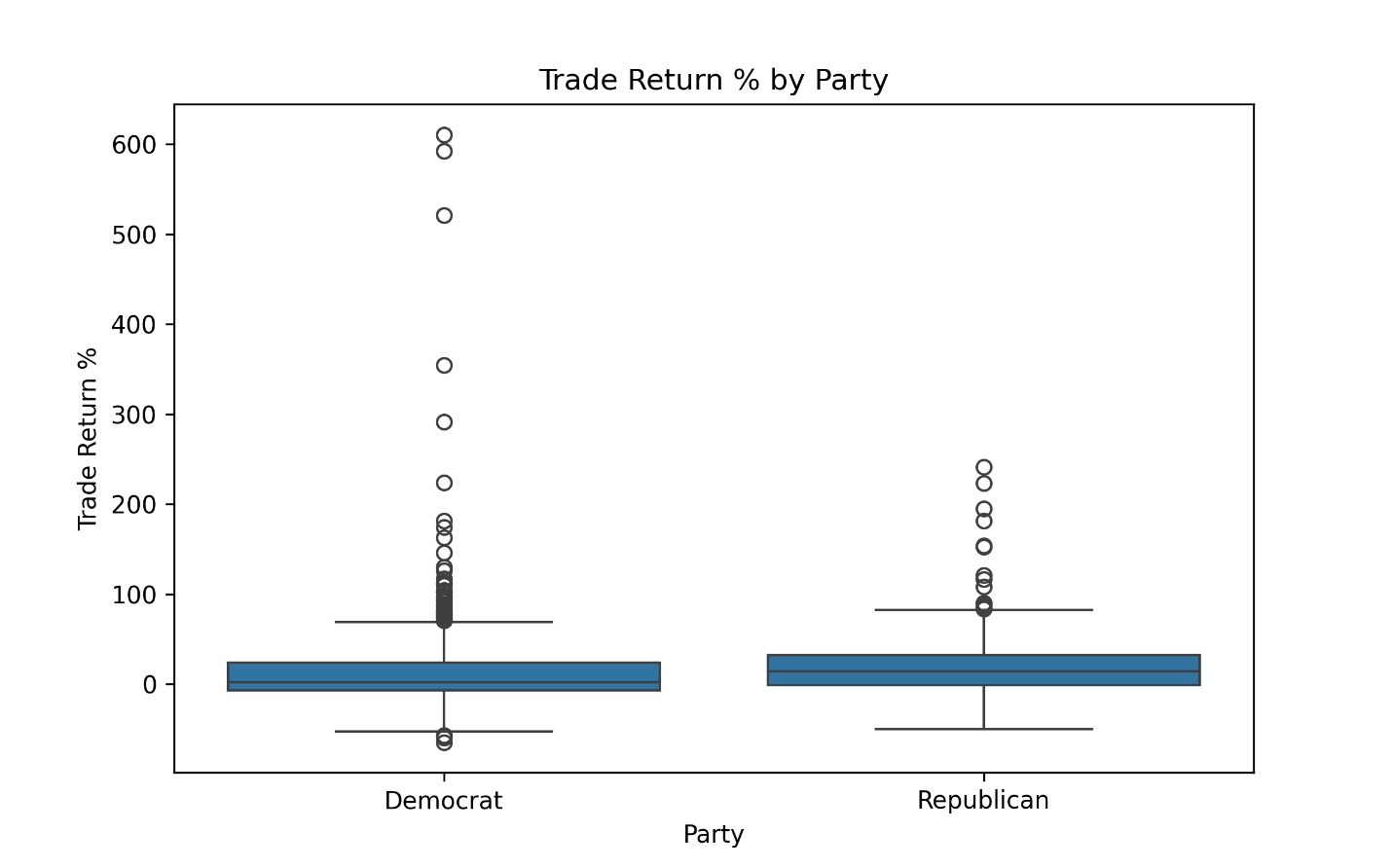
The results show that, despite some public concerns that lawmakers may have an informational edge, their stock trades as a group do not outperform the market. In fact, they underperform significantly. While there were outliers with very high returns, these were not enough to outweigh the broader pattern of losses relative to the S&P 500. This supports the idea that, even with possible access to non-public information, Congress members do not systematically gain an advantage in their investments. These findings contribute to the broader conversation about financial transparency and ethics in government by offering data-driven evidence that challenges assumptions of insider advantage.

## Stock Trading Performance By Party

Beyond overall market performance, I also wanted to investigate whether there were differences in stock trading success between political parties. Using the same method as before, I separated the trades by party affiliation and calculated descriptive statistics for each group. Democrats had 947 trades with an average return of 13.60% and a standard deviation of 47.31%, while Republicans had 409 trades with a higher average return of 19.51% and a lower standard deviation of 33.96%. Both groups had wide ranges in returns, but Democrats exhibited more variability and extreme outliers, with a maximum return of over 610% compared to Republicans’ maximum of about 242%.

A boxplot of returns by party (Figure 3) visually highlights these patterns. While both distributions have long tails and similar spreads in their interquartile ranges, the median return for Republicans is noticeably higher than that for Democrats. Additionally, Democrats had more extreme positive outliers, but their distribution appears more dispersed and less centered around positive returns compared to Republicans.

**Figure 3: Boxplot of Congressional Trade Returns by Party**

****

To formally test whether these differences were statistically significant, I conducted three hypothesis tests (Table 2). First, I tested whether Democrats’ returns outperformed Republicans’ returns. The p-value was 0.9953, which provides no evidence that Democrats outperform Republicans. Next, I tested whether Democrats underperformed Republicans, which produced a p-value of 0.0047. This indicates that Democrats significantly underperformed Republicans in stock returns. Finally, I tested whether returns simply differed between the two parties, regardless of direction. The p-value here was 0.0095, confirming that there is a statistically significant difference in returns between parties.

**Table 2: Hypothesis Test Results – Party Comparison**

|  |  |  |
| --- | --- | --- |
| **Test** | **P\_Val** | **Result** |
| Democrat Congress Members outperform Republican Congress Members in the stock market at a statistically significant level. | 0.995274 | FALSE |
| Democrat Congress Members underperform Republican Congress Members in the stock market at a statistically significant level. | 0.004726 | TRUE |
| Democrat Congress Members's returns are different than Republican Congress Members' returns in the stock market at a statistically significant level. | 0.009452 | TRUE |

These results suggest that, on average, Republican members of Congress had stronger stock market returns than their Democratic counterparts during the 2023–2024 period. While both groups saw highly variable outcomes, Republicans generated higher median and average returns, and this difference is unlikely to be due to chance alone. This finding contributes an additional layer to the analysis by showing that party affiliation may correlate with trading success, though further research would be needed to determine potential causes for this gap.

## Stock Trading Performance By Chamber

In addition to analyzing performance by political party, I also examined whether stock trading outcomes differed between the two chambers of Congress: the House of Representatives and the Senate. The descriptive statistics indicate that both chambers had similar average returns, with House members producing a mean return of 15.22% and Senate members slightly higher at 16.80%. The House had a sample size of 1,215 trades, far larger than the Senate’s 141 trades. Both groups exhibited wide variability in returns. The House had a standard deviation of 44.31%, while the Senate showed a somewhat lower variability with a standard deviation of 39.08%. Both chambers also had substantial ranges, with the House’s maximum return exceeding 610% and the Senate’s maximum return of about 223%.

While the Senate exhibited marginally higher mean and median returns, the differences between the two groups are relatively minor when compared to their large standard deviations. To formally assess whether these apparent differences were statistically meaningful, I conducted two hypothesis tests (Table 3). First, I tested whether Senate members underperformed House members, which yielded a p-value of 0.6728—far above the conventional threshold of significance. Second, I tested whether the returns from the two chambers simply differed (regardless of direction). The resulting p-value was 0.6544, again indicating no statistically significant difference.

**Table 3: Hypothesis Test Results – Chamber Comparison**

|  |  |  |
| --- | --- | --- |
| **Test** | **P\_Val** | **Result** |
| Congress Members who serve in the Senate outperform Congress Members who serve in the House in the stock market at a statistically significant level. | 0.327196 | FALSE |
| Congress Members who serve in the Senate underperform Congress Members who serve in the House in the stock market at a statistically significant level. | 0.672804 | FALSE |
| Congress Members who serve in the Senate's returns are different than Congress Members who serve in the House's returns in the stock market at a statistically significant level. | 0.654392 | FALSE |

These results suggest that, despite slight differences in mean returns, there is no evidence that members of either chamber consistently outperform or underperform their counterparts. The similarity in outcomes implies that institutional differences between the House and Senate—such as term lengths, committee assignments, or access to information—do not appear to manifest in significantly different stock market performance.

## Predicting Trade Overperformance

To assess whether congressional trade overperformance could be predicted based on characteristics of the lawmaker or the trade itself, I created a binary variable, Overperformed, which indicates whether a trade outperformed the market. Features included party affiliation, chamber, state, and the number of days the stock was held. Using an 80/20 train-test split, I tested three classification models: Random Forest, Decision Tree, and Logistic Regression. Model performance metrics are summarized in Table 4.

**Table 4: Model Performance Metrics – Predicting Overperformance**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **Level** | **precision** | **recall** | **f1-score** | **support** |
| Random Forest | 0 | 0.675676 | 0.862069 | 0.757576 | 174 |
| Random Forest | 1 | 0.52 | 0.265306 | 0.351351 | 98 |
| Random Forest | accuracy | **0.647059** | **0.647059** | **0.647059** | **0.647059** |
| Random Forest | macro avg | 0.597838 | 0.563688 | 0.554464 | 272 |
| Random Forest | weighted avg | 0.619587 | 0.647059 | 0.611215 | 272 |
| Decision Tree | 0 | 0.68018 | 0.867816 | 0.762626 | 174 |
| Decision Tree | 1 | 0.54 | 0.27551 | 0.364865 | 98 |
| Decision Tree | accuracy | **0.654412** | **0.654412** | **0.654412** | **0.654412** |
| Decision Tree | macro avg | 0.61009 | 0.571663 | 0.563746 | 272 |
| Decision Tree | weighted avg | 0.629674 | 0.654412 | 0.619315 | 272 |
| Logistic Regression | 0 | 0.644269 | 0.936782 | 0.763466 | 174 |
| Logistic Regression | 1 | 0.421053 | 0.081633 | 0.136752 | 98 |
| Logistic Regression | accuracy | **0.628676** | **0.628676** | **0.628676** | **0.628676** |
| Logistic Regression | macro avg | 0.532661 | 0.509207 | 0.450109 | 272 |
| Logistic Regression | weighted avg | 0.563845 | 0.628676 | 0.537665 | 272 |

Across all models, overall accuracy ranged from 62.9% to 65.4%. However, performance was heavily skewed toward correctly predicting non-overperforming trades (class 0). Precision and recall for overperforming trades (class 1) were much lower, particularly for Logistic Regression, which struggled most with this class. Both the Random Forest and Decision Tree models provided modest improvements in detecting overperforming trades, but none of the models demonstrated strong predictive power overall.

These results suggest that while characteristics like party, chamber, state, and holding period offer some explanatory power, they are insufficient on their own to reliably predict market-beating trades. Incorporating more granular financial data—such as transaction size, sector, or timing relative to policy events—could improve model performance in future studies.

## Forecasting Trade Volume

To explore whether congressional stock trading volume can be forecasted, I aggregated monthly counts of buy-side trades from January 2023 to December 2024. Table 5 shows the number of trades per month, which revealed considerable variability—spikes occurred in January and October 2023 (223 and 260 trades respectively), while activity sharply declined in late 2024.

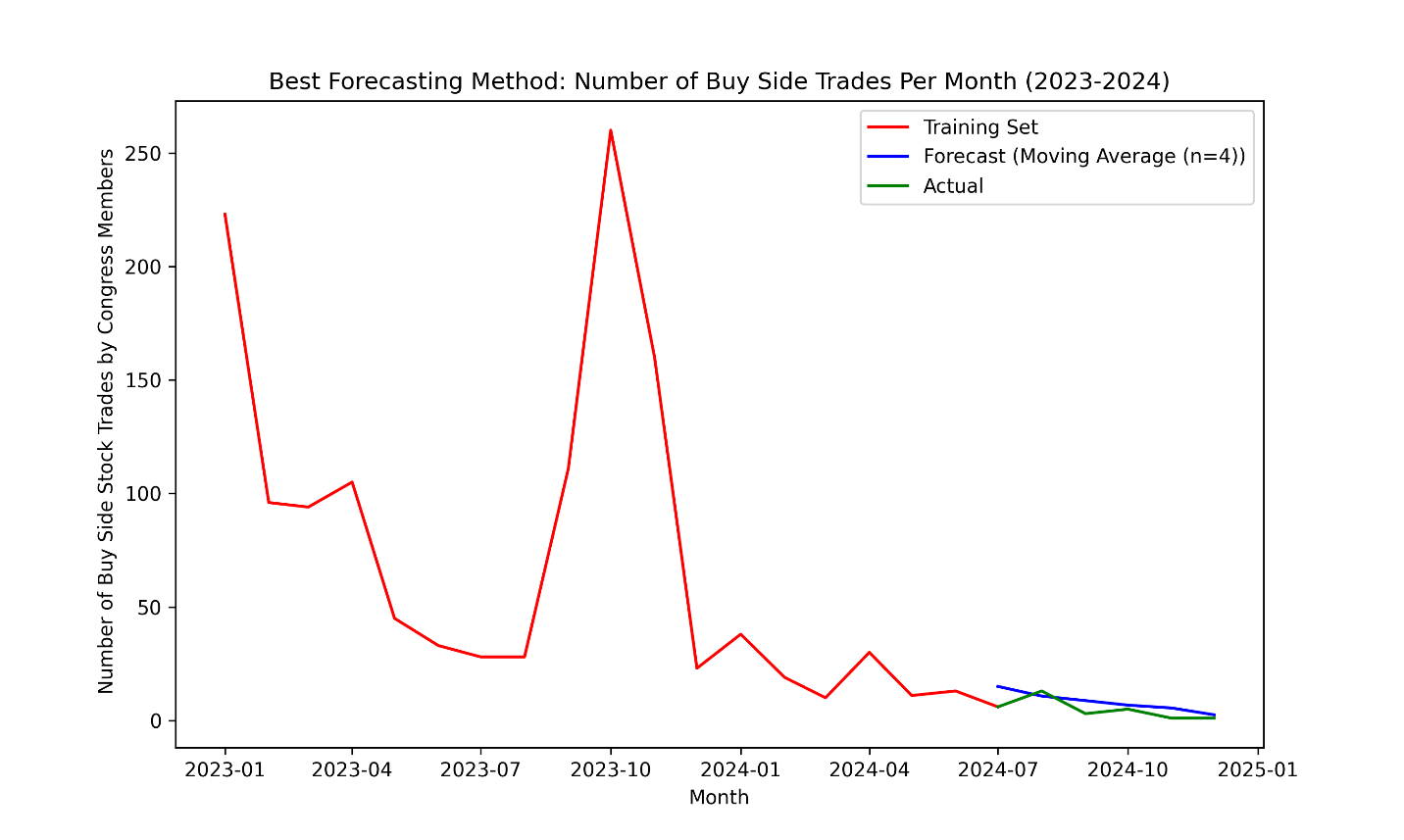
I applied several time series forecasting methods to predict future monthly trade volumes and compared their accuracy using mean absolute error (MAE). The models tested included Moving Average (n=4, n=6), Simple Exponential Smoothing (SES) with two smoothing parameters (α = 0.2 and α = 0.8), Holt-Winters additive and multiplicative methods, and TBATS, a model designed for complex seasonal patterns. Table 6 summarizes the MAE for each technique.

**Table 6: Forecasting Model Performance (Mean Absolute Error)**

|  |  |
| --- | --- |
| **Forecasting Technique** | **MAE** |
| Moving Average (n=4) | 4.125 |
| SES (alpha=0.8) | 4.439151 |
| Moving Average (n=6) | 5.416667 |
| HW Multiplicative | 6.033796 |
| SES (alpha=0.2) | 23.56293 |
| HW Additive | 40.16474 |
| TBATS | 66.38082 |

The Moving Average model with a 4-month window produced the most accurate forecasts, with an MAE of 4.13 trades per month. Figure 4 visualizes this forecast alongside actual trade volumes during the test period (2023–2024). The model captured overall fluctuations reasonably well, though it struggled with abrupt spikes like those in October 2023.

**Figure 4: Best Forecast of Buy-Side Trades**

****

These results suggest that while congressional trading volume shows short-term seasonality or clustering that simple moving averages can capture, more sophisticated models like TBATS or Holt-Winters do not add predictive value for this dataset. Future research could investigate whether external factors (such as legislative calendars or market volatility) improve long-term forecasting accuracy.

# Conclusion

This analysis set out to evaluate whether the stock trades of U.S. Congress members in 2023 and 2024 outperformed the broader market and to assess whether trading success differed meaningfully across political parties and congressional chambers. By leveraging transaction data from Capitol Trades and historical market data via Yahoo Finance, we systematically compared the returns of stocks bought and sold by lawmakers to the performance of the S&P 500 over equivalent holding periods.

The results indicate that, on average, congressional trades underperformed the market at a statistically significant level. Contrary to public skepticism suggesting members of Congress might possess or act upon privileged information to achieve outsized gains, the evidence here suggests their investment outcomes were, in fact, inferior to a passive market benchmark. Additionally, the analysis found that returns varied significantly by political party. Specifically, trades made by Democratic members underperformed those of their Republican counterparts at a statistically significant level, highlighting notable differences in trading success between the two groups.

However, these conclusions must be considered in light of several limitations. First, the study excludes transactions for which complete buy and sell data could not be established, potentially omitting trades that could meaningfully affect the results. Second, because the analysis relied on percentage returns rather than dollar-weighted returns (due to lack of data on trade size or investment amounts), the impact of larger trades versus smaller trades could not be assessed. Additionally, by focusing solely on trades executed in 2023 and 2024, the analysis may not fully capture longer-term trading patterns or behaviors that manifest over a broader time horizon. Finally, although steps were taken to ensure price accuracy (such as adjusting for market holidays and weekends), the absence of intraday transaction data introduces an approximation that may slightly distort realized returns.

Future research could address these limitations by incorporating transaction size where available, expanding the sample period to examine trends across more market cycles, and exploring sector-level trading patterns to determine whether specific industries are disproportionately favored by members of Congress. Moreover, integrating qualitative factors — such as committee assignments or legislative activity — could provide deeper insight into whether access to non-public information influences trading behavior. As transparency and accountability remain central to public trust, continued investigation into congressional financial activity remains both relevant and valuable.