Lab 2: Market Cap Influences following the Trump Inauguration

DATASCI 203 SEC 02 | https://github.com/mids-w203/lab-2-maroonp005

Amy Steward, Kristen Lin, Chad Adelman, Karim Mattar

April 18, 2025

Introduction

The performance of stocks can shift rapidly in response to presidential administration changes, and since Trump's inauguration on Jan 20th, there has been a highlighted focus on the stock market's performance and concerns about the future outlook. With the stock market as a critical backbone of the U.S. economy, prior research has extensively explored stock market reactions to major events, and financial algorithms have been continuously refined. However, there has been little analysis of how company-specific traits and strategic responses - in the form of corporate actions - are together associated with stock performance during politically uncertain times. This study aims to investigate the combination of company characteristics and actions on market capitalization changes within the Russell 3000 during the initial 60 days of the Trump administration. Examining the first 60 days of the administration allows for an early-term analysis, offering timely insights into influential factors while providing a long enough window to reduce short-term volatility and noise across individual stock performance. To better understand how company-specific factors correlate with early market reactions under this new political leadership, our team is focusing on the following questions:

> *In the first 60 days of the Trump administration, are there any company characteristics that correlate with the 60-day price change of the stock?*
>
> *Do certain characteristics have a higher correlation with price change than others?*

To answer these questions, using the classical linear model, our team will statistically analyze which company characteristics are associated with differences in market capitalization trends during this period. Our outcome variable (Y) is the 60-day net change in adjusted closing prices, and our covariates (Xs) are sector, cap size, employee count, trading volume, and corporate actions. Understanding how company characteristics correlate with stock responses can offer insight into market trends, helping inform strategic decisions within a company, and helping investors react during times of uncertainty or economic turbulence.

Data Source Description

Our analysis uses company-level data from three main sources: the iShares BlackRock list of Russell 3000 constituents, the Yahoo Finance API (via the yfinance Python package), and CompaniesMarketCap.com. Data was collected in April 2025 and includes observations from January 21 to March 21, 2025, covering the first 60 days of the Trump administration.

The Russell 3000 index is a collection of stocks for US-based companies that represent approximately 98% of the American public equity market. The index is created and maintained by FTSE Russell, an organization that builds and manages stock market indexes. The index includes 3,000 of the largest publicly traded U.S. companies, spanning large-cap, mid-cap, and small-cap stocks. This index serves as a comprehensive benchmark for the overall performance of the U.S. stock market and was gathered using iShares BlackRock.

The Yahoo Finance API is a service that provides access to a wide range of financial data, including stock prices, historical market data, company information, and financial news. The API allows users to pull data on stock information and performance in mass. Features that were pulled included company info, such as sector and industry, and daily financial information, such as marketCap.

CompaniesMarketCap.com is a financial data platform that ranks over 10,000 publicly traded companies by market capitalization, offering insights into global corporate valuations. It provides useful financial data such as a company's market capitalization, earning, revenue, and more.

Data Wrangling

We constructed a broad dataset for the Russell 3000 by integrating data from Yahoo Finance API (yfinance package), iShares BlackRock, and CompaniesMarketCap. The BlackRock dataset provided the official list of Russell 3000 tickers and their sector classifications. The Yahoo Finance API was used to pull historical stock data—specifically, the adjusted closing prices at the beginning and end of the 60-day window. These values were used to calculate the net price change. We also retrieved dividend and split activity indicators (using 1 = occurred and 0 = did not occur) during the same window. Employee counts were retrieved from CompaniesMarketCap.com via manual lookup, which were then merged by ticker. Due to resource limitations in DataHub Rstudio, much of the data extraction was done locally. Because stock data is time series-based, we consolidated the measurements into a cross-section of single rows for each stock with standardized features for the same period. The first and final adjusted closing prices were used to calculate the 60-day net change and to indicate if any splits and dividends occurred (using 1 = yes and 0 = no). We then merged the dataset with the CompaniesMarketCap data to assign the number of full-time employees for each company. With the focus being on the first 60 days after Trump's inauguration, we used the Yahoo Finance API to pull the adjusted closing prices, all company actions (splits and dividends), and market capitalizations, which helped us to derive the cap sizes during this period.

The following table displays the adjustments we made to the datasets prior to building our models. The initial dataset from BlackRock contained 2676 stocks of different market cap sizes (small, medium, large). The limitations of using the free Yahoo Finance API resulted in incomplete data capture, in which 121 stocks had missing or unavailable information for all predictors. As such, these 121 samples were removed when joining the two datasets based on stock symbols, which could have introduced bias if these exclusions are non-random. From here, the final dataset was split by a ratio of approximately 30:70 to create a exploratory dataset for model engineering and a confirmatory dataset to evaluate final results. Additionally, we removed 3 stocks that had null values for market cap, as this meant that the predictor variable cap size would also be null, leaving us with a total of 2552 data points.

Sample Size	Filter	Removed Samples	End Sample Size
2676	Unavailable Data from YFin API	121	2555
2555	Missing Values in `market_cap`	3	2552
2552	Subset to Exploratory Dataset	0	765
2552	Subset to Confirmatory Dataset	0	1787

Operationalization

Our null hypothesis for this analysis is that characteristics for companies in the in the Russell 3000, including sector, employee count, dividend and split action, net change in trading volume, and cap size (based on market cap), did not have a statistically significant impact on net change in adjusted close amount (calculated using the difference of adjusted start and end close amount in dollars) during the first 60 days of the Trump Administration. Prior research suggested that firms of varying size and industry respond differently to changes during politically sensitive periods. Employee count serves as a proxy for company scale, and corporate actions often signal information to investors.

We ran a series of single-variable linear regressions to reject the null hypothesis and identify significance (Figure 1). For each regression we checked significance with the `coefest` to look at both t and p-values. Using this method, we found significance in sector, capsize, and natural log of employee count. However, when combining these terms, natural log of employee count no longer had any effect on the overall model. From our testing results, we ended up using Sector and Cap Size as our predictors to create our final linear model and ran an ANOVA against our other models, confirming that the sum of squares was lowest for the linear model with both Sector and Capsize. We then tested the Model Assumptions to confirm if this model meets those expectations and how to derive a better fit .

In the future, we hope to continue segmenting this information to further understand attributes that have the biggest impact on the net change in close price during the Trump administration.

Data Visualization

Figure 1: Single−Variable Linear Regression Exploration

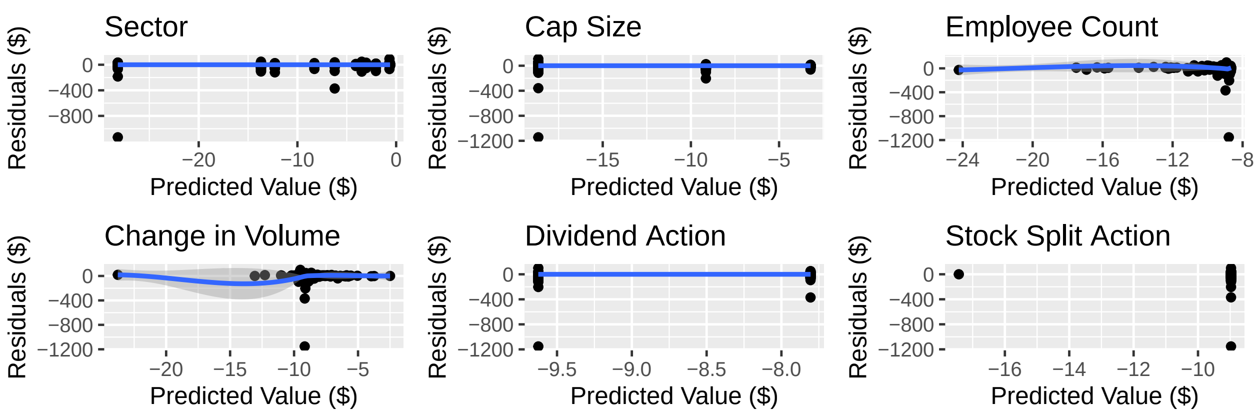
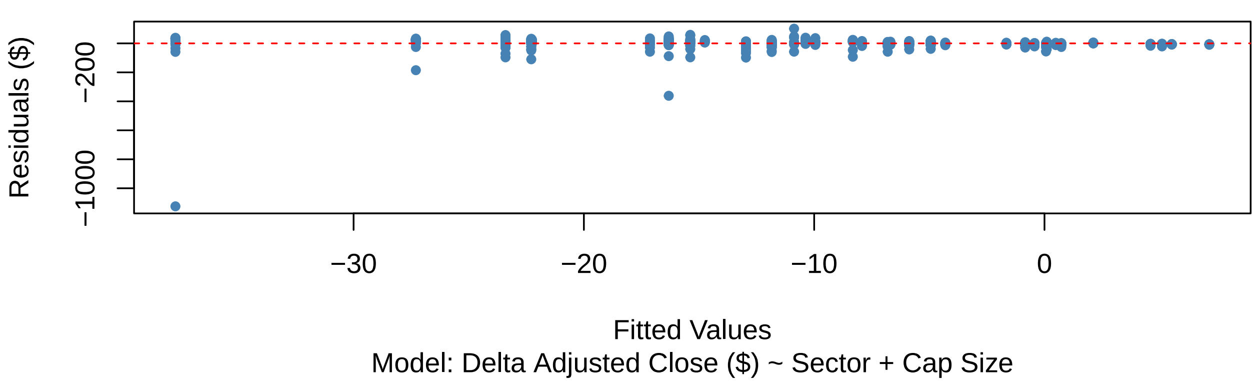


Figure 2: Residuals ($) vs Fitted Values



Model Specifications

We found sector and cap size were the most relevant descriptors of the delta net adjusted close price after looking at linear regressions for each individual attribute. This was confirmed as the residual sum of squares was the lowest for these two characteristics compared to all other individual linear regressions. A review of model assumptions indicated heavy tails on both sides in the QQ plot and a kurtosis of > 400 and skewness of -18, indicating the data is leptokurtic.

While log transformations were considered to reduce skew, they were not feasible due to the nature of stocks and the importance of knowing if the net change in adjusted closing price was negative or positive. Around

3% of data had a Cook's Distance of 4/n and considered influential observations. Of these, we found they were more heavily weighted in the Large Cap sectors of Industrials, IT, and Consumer Discretionary. In a future study, it may be prudent to isolate these groups and continue to investigate why they may act more positively or negatively during this time period. As our data for all models was highly leptokurtic, we opted to use the coeftest with robust standard errors, which would more accurately reflect the data.

Table 2: Confirmatory Model Results

	Dependent variable:	
	Net Change in Adjusted Close	
	(1)	(2)
Sector: Consumer Discretionary	-8.367^{***}	-8.161^{***}
	(3.001)	(3.108)
Sector: Energy	-6.012^{**}	-5.866^{**}
	(2.727)	(2.804)
Sector: Financials	-3.967^{*}	-3.876^{*}
	(2.234)	(2.350)
Sector: Industrials	-14.908^{***}	-14.120^{***}
	(2.683)	(2.754)
Sector: Information Technology	-13.462^{***}	-12.786^{***}
	(2.838)	(2.888)
Sector: Materials	-6.329^{**}	-5.808^{**}
	(2.558)	(2.681)
Cap Size: Medium Cap		4.295^{**}
		(2.181)
Cap Size: Small Cap		8.157^{***}
		(1.986)
Constant	-0.025	-5.503^{*}
	(2.102)	(2.898)
Observations	1,787	1,787
R^2	0.049	0.068
Adjusted R^2	0.044	0.062
Residual Std. Error	22.466 (df = 1776)	22.254 (df = 1774)
F Statistic	9.213^{***} (df = 10; 1776)	10.823^{***} (df = 12; 1774)
Note:		*p<0.1; **p<0.05; ***p<0.01

Model Assumptions

IID: All stocks in the Russell 3000 are impacted by economic factors, decisions and instability. Some stocks may also relate to others, e.g. if you sell your shares for one stock, you may be selling all stocks you own. The residuals have a discernible pattern, suggesting that constant variance is not met. Therefore, it is not IID.

Linear Conditional Expectation: To confirm linearity and conditional expectation, we plotted the residuals against the best-fit line. This confirmed linearity but also identified several significant outliers. We further investigated these outliers in another section below.

No Perfect Collinearity: To confirm no collinearity, we looked at the variance inflation factor and saw that the VIF is <5 and closer to 1, meeting the requirement that the data is not collinear.

Homoskedastic Errors: QQ-Plot, Plot scale-location model, PB test, skewness, and kurtosis were analyzed and indicated heavy tails, which was further confirmed through a kurtosis of > 400 and a skewness of -18. This indicates that our data is heteroskedastic and leptokurtic and does not meet the requirements of the CLM.

Normally Distributed Errors: The QQ Plot showed that errors are mostly equally distributed, however there are still significant outliers.

Results: This provided us with enough information to identify that the CLM may not be the best model due to the outliers. We had several options on how to address this and decided to dive deeper into further investigation at this point.

Model Results and Interpretation

When looking at the 60-day net change in the adjusted closing price, our final confirmatory model rejects the null hypothesis and indicates that both variables of sector and cap size are relevant descriptors of the delta net adjusted close price. However, due to the low R2 value, these descriptors do not fully explain the model. We believe that this is due to not meeting homoskedastic assumptions with leptokurtic data. Cooks Distance confirmed these outliers do impact the model. Of all the models we studied (single variable and multi-variable linear regressions), the model including sector and cap size had the lowest Residual Sum of Squares in the ANOVA test.

In the combined model, some sectors had a negative coefficient, contributing to an increased change in the negative direction when compared to baseline. When looking at the medium cap and small cap, however, both had positive coefficients. This indicates that mid-cap and small-cap stocks typically experienced higher average net gains than large-cap stocks in the first 60 days of President Trump's tenure.

To improve this model, we took the log of the employee count to account for heteroskedasticity, modeled linearity for each variable individually. However, taking the net change in the net delta change omits the sign of the change in close price. We tested applying the log for employee count, however Sector and Cap Size were still more significant. An outlier investigation indicated around 3% of the data had a Cooks Distance $>$ 4/n. These data are comprised disproportionately of Large Cap stocks in the sectors of Industrials and IT.

For next steps we would suggest modeling Large Cap Industrials and IT alone, including a variable for the signage (positive/negative) of the net change in adjusted closing price, then taking the log of the absolute value. Other model types may reflect a more accurate R2 for these data which is worth exploring. Finally, incorporating specific firms or macroeconomic variables may also capture more short-term volatility.

In conclusion, our model suggests that sector and cap size are correlated with the 60-day net change in adjusted closing price, however these descriptor variables are not comprehensive leading to the low R2. Generally, smaller cap firms fared better than large cap firms, and some sectors have better performance than others.

Practical significance of this model suggest that sectors of Industrials (-$14) and IT (-$12) are significantly more negatively correlated by the current political climate. However, we see the small cap stocks actually rose ($8). It is important to recognize that while these variables had a low p-value and were considered very significant, the overall R value of the regression was low meaning that while some industrials and IT did drop and small cap, this is not guaranteed as not all effects are accounted for. However, the limited R2 and multiple outliers indicate that these results do not describe everything that was happening during this 60 day period.

Appendix

A Link to your Data Source:

https://github.com/mids-w203/lab-2-maroonp005/data/external

A List of Model Specifications we Tried:

To start we looked at each variable individually and built a linear model for that variable. This allowed us to clearly see which beta variables may most impact the net delta adjusted close. That let us isolate which variables individually had the largest impact on the net delta change.

Base Models:

$$revenue = \beta_0 + \beta_1 \cdot sector + \epsilon$$

$$revenue = \beta_0 + \beta_1 \cdot cap\ size + \epsilon$$

$$revenue = \beta_0 + \beta_1 \cdot employee\ count + \epsilon$$

$$revenue = \beta_0 + \beta_1 \cdot \ln(employee\ count) + \epsilon$$

$$revenue = \beta_0 + \beta_1 \cdot \Delta\ trading\ volume + \epsilon$$

$$revenue = \beta_0 + \beta_1 \cdot dividend + \epsilon$$

$$revenue = \beta_0 + \beta_1 \cdot split + \epsilon$$

Model 1:

$$revenue = \beta_0 + \beta_1 \cdot sector + \beta_2 \cdot cap\ size + \epsilon$$

Model 2:

$$revenue = \beta_0 + \beta_1 \cdot sector + \beta_2 \cdot cap\ size + \beta_3 \cdot \ln(full\ time\ employees) + \epsilon$$

Final Model:

$$revenue = \beta_0 + \beta_1 \cdot sector + \beta_2 \cdot cap\ size + \epsilon$$

A Residuals-vs-Fitted-values Plot: