MGT_6203 - Project Proposal

Neill Killgore, Rajiv Kamal, Eissa Qader, Kha Tran, Bryce Meyering

2022-02-19

Group Member Bios

Neill Killgore (Louisiana, USA) I work for Louisiana State University managing a team in the IT department. My background is mostly in various different roles in IT. My interest in analytics started with a project for the US Probation and Pretrial offices. The goal was to optimize treatment and rehabilitation programs to determine the best allocation of budget dollars with the goal of reducing recidivism (someone re-offending and ending up back in the system). Since then, most of the data analysis I have done has largely been time-series data - logs from things like firewalls and servers or raw network packet streams. The most interesting project lately was developing a method to detect and automatically mitigate a recurrent DDOS (distributed denial of service) attack before it could impact the campus.

Rajiv Kamal (Delaware, USA) I am working for Citibank NA as portfolio risk manager, looking after the small business cards portfolios. I am responsible for all machine learning based credit risk scoring models (both in-house and vendor) for Costco and American Airlines cards. These models are acquisition and behavior score cards which are based upon gradient boosting machines. On the educational front, I am a 2009 MBA from Vinod Gupta School of Management, IIT Kharagpur specializing in finance which is further supplemented by FRM®-I, II certifications from GARP.

Eissa Qader (Virginia, USA) I am a current master's student at Georgia Tech studying analytics. I come from a business tech background and work at Deloitte as a tech consultant. I specialize in data cleaning, modeling, and analysis.

Kha Tran (Georgia, USA) I am currently working for Norfolk Southern (NS) as an IS Auditor. I get to interact with both IT and business department within Norfolk. My job is to gather all the data and evidences within our environment and analyze them to identify problems and suggest improvements to NS's processes and systems.

Bryce (Bo) Meyering (Florida, USA) I am currently a lab manager and agricultural researcher at the University of Florida in the Horticultural Sciences Department. My research focuses mainly on citrus tree root system architecture, physiology, phenotyping and rootstock-scion interactions. This is my fourth class in the OMSA program. I do a lot of root system image analysis (both on microscopic and large scale whole tree root system images) and would love to work for a large agribusiness company in the future, particularly with remote sensing data or accelerated phenotyping systems.

Research Questions

Recently, news of insider trading and violations of the STOCK act by congressional members has made many headlines in US national news [1]. This type of behavior is troubling due to the timely and, many times, highly sensitive information they are exposed to on a regular basis. Though the purpose of the STOCK act is to prevent our representatives from acting on this information and gaining a distinct advantage in the market, the penalties for not complying are minimal and ineffective, thus insider trading continues. While proving the intent behind a congressional member's trades is a difficult task since we do not know what they know, we thought it would be worthwhile to analyze representatives' trading behavior and transaction timing before major moves in either the market or individual securities they hold.

1. Thus, our first and main research goal is to leverage publicly available daily stock price data and Congressional stock transaction disclosures to identify 'suspicious' trades within Congress. We plan to investigate any

---

[1] see (https://www.theatlantic.com/politics/archive/2022/01/congress-stock-trading-ban/621402/) for more reading

correlations between congressional members' trades and price movement in the market following the trades that congressional members make.

2. The second goal of this project is to identify the best and worst traders within each respective house of congress or political party by computing their individual returns compared to broader market indices.

3. Our final goal is to identify any distinct clusters of congressional members based on similar trading patterns, positions held, cumulative returns, etc.. We will try to find the variables that are most associated with each cluster.

Data Sources

We have identified two separate datasets that we plan to leverage in addressing our research questions.

- The first dataset is a live database containing all of the disclosed stock purchases/sales of all U.S. Congress members. Under the STOCK (Stop Trading on Congressional Knowledge) act of 2012, senior US government officials have to report any stock, bond, or commodities transaction within 45 days of the transaction, provided that the total transaction amount is greater than $1000. We have accessed this database through a Python API at Quiver Quant for a small fee. Pursuant to the Code of Federal Regulations, government officials are only required to report the value of a sale or purchase within 11 distinct bins and what investment they sold/purchased, but not the quantity or exact price. This leaves some ambiguity in the numerical side of the data that we will have to address by making some broad, less-than-ideal assumptions. Our dataset currently has coverage from 2016 until present for all congressional members who disclosed any transactions.

- In order to supplement the above dataset, we have also pulled daily stock data for all of the unique ticker symbols found in our congressional trading data. This data was pulled from Yahoo Finance using the R package `tidyquant` starting from January 1st, 2016 until present date. We plan to focus most on the `adjusted_close` variable in order to calculate returns.

Dataset Description

Both datasets are in tidy formats with self explanatory variable names. However, I have listed them below

Congressional Trading Data

- **Report Date:** Date that the transaction was disclosed under the STOCK act.
- **Transaction Date:** Date of the transaction
- **Ticker:** Ticker symbol of the transaction
- **Representative:** Name of the Congress member
- **Transaction:** Type of transaction, 'Sale' or 'Purchase'
- **Amount:** Lower bound of the transaction range
- **House:** Congressional body that the representative is a member of, 'Representatives' or 'Senate'
- **Range:** The range of the transaction value

Here is a glimpse at a few random rows of the congressional trading dataset (only select columns for brevity's sake):

TransactionDate	Ticker	Representative	Transaction	Range
2019-02-04 00:00:00	HON	John F. Reed	Sale	$1,001 - $15,000
2020-04-02 00:00:00	BP	Daniel Meuser	Sale	$1,001-$15,000
2018-07-30 00:00:00	JPM	Pat Roberts	Sale	$1,001 - $15,000
2021-07-19 00:00:00	AAPL	Patrick Fallon	Sale	$250,001-$1,000,000

Historical Market Data

- **Symbol:** Stock ticker symbol
- **Date:** Date of the observation
- **Open:** Opening price
- **High:** High price
- **Low:** Low price

- **Close:** Closing price
- **Volume:** The number of shares traded
- **Adjusted:** Closing price after accounting for major corporate actions

symbol	date	open	high	low	close	volume	adjusted
AAPL	2020-05-04	72.2925	73.4225	71.580	73.2900	133568000	72.29656
AAPL	2021-04-14	134.9400	135.0000	131.660	132.0300	87222800	131.24994
AAPL	2020-10-29	112.3700	116.9300	112.200	115.3200	146129200	114.27045
AAPL	2016-05-11	23.3700	23.3925	23.115	23.1275	114876400	21.51419

Modeling Plan

As with any large project, our first step in the modeling process will be to clean the data if needed, and perform exploratory data analysis by plotting changes in market indices over the data time periods. We the plan to track the positions held by each representative/senator and mark the dates that they either close/open a position. We will then analyze the trends in the daily percentage price change for that stock after detrending to look for large fluctuations in price. We plan to try out many different detrending techniques and lag periods with the data in order to come up with parameters that are optimized for the majority of cases. If we can successfully identify transactions preceding large stock movements that work to the traders' benefit, then we will classify those as 'suspicious' trades. Our plan is to try CUSUM or other time series models to detect these changes. We realize that there are many rational and perfectly legal reasons that a congressman might buy or sell before market movements, and so will make our classifications as judiciously as possible.

The second goal of the project is to determine who is the most effective trader in Congress. While on the surface, this seems to be fairly straightforward, the data from the congressional trades does not include the number of shares sold, only a value range in dollars. Though it isn't possible to calculate the *precise* returns since we don't have access to their individual portfolios, we can make a series of assumptions about the data (i.e. always taking the mean value, minimum value, or a random value according to a sampling distribution for any given range of transaction amounts) and then calculate cumulative returns from these values.

Finally, we want to try out several different clustering algorithms, such as k-means or DBscan, to find distinct groups of traders with Congress. Since this is a tangential goal to our first two questions, the data that we use for clustering will vary depending on our initial findings.

Anticipated Results

We anticipate finding some correlation between the congressional stock transactions and stock price movements and expect their return bets to surpass market returns. We anticipate finding some officials who trade well, but also expect that there are those who trade poorly. In addition, we expect to see that there are distinct types of congressional traders based on their trading habits and discovering some unique trading networks/groupings after our clustering analysis between congressional houses, parties, and/or congressional committees.