Maxim Dokukin

017025164

MATH 167R Sec 01

Prof. Gao

03/19/2024

Project Proposal

The debate over whether government officials should participate in stock trading persists. Investment is a key wealth-building strategy, but it could lead to conflicts of interest. Officials are expected to prioritize the public good, which may not always align with profit-maximizing motives. There's also the risk that politicians may exploit non-public information, leading to insider trading.

This project's goal is to examine the trading activities of politicians and determine if their transactions outperform the general market. The S&P 500 index serves as the benchmark for average market returns.

Data for this study was scraped from https://www.capitoltrades.com/trades, yielding a dataset of 38,241 transactions spanning from 2021-03-14 to 2024-03-07. The dataset contains 14 variables: ['Politician Name', 'Party', 'Chamber', 'State', 'Issuer Name', 'Ticker', 'Publication Date', 'Transaction Date', 'Reporting Gap', 'Owner', 'Transaction Type', 'Value Range', 'Price', 'Transaction ID']. Scraping was the chosen method due to the lack of existing comprehensive datasets.

Data Dictionary

Variable Name	Variable Type	Description
Politician Name	String	Name of the politician involved in the transaction
Party	String	Political party affiliation of the politician (Democrat; Republican; Other)
Chamber	String	Legislative chamber (e.g. House; Senate) the politician belongs to
State	String	State the politician represents
Issuer Name	String	Name of the entity issuing the stock or security
Ticker	String	Stock ticker symbol of the security involved in the transaction
Publication Date	Date	Date when the transaction was publicly reported
Transaction Date	Date	Date when the transaction occurred
Reporting Gap	Integer	Number of days between the transaction date and the publication date
Owner	String	Owner of the transaction (Undisclosed; Spouse; Joint; Child; Self)
Transaction Type	String	Type of transaction (SELL; BUY)
Value Range	String	Range of the transaction value (< 1K; 1K–15K; 15K–50K; 50K–100K; 100K–250K; 250K–500K; 500K–1M; 1M–5M; 5M–25M)
Price	Double	Price at which the transaction was made (per share)
Transaction ID	Integer	Unique identifier for the transaction

List of Subquestions

- Is there a correlation between Reporting Gap and trade profitability?

- Is there a difference between Democrat and Republican trades marginality/frequency?

- Who is the most successful investor in the government?

- Is there potential cases of insider trading (selling of stock before its significant drop)?

- Are certain states more active in terms of the number of transactions?

- How does transaction activity vary before and after significant political events (elections)?

- Which ticker symbols appear most frequently in transactions across all politicians?

Outside References

- Yahoo Finance for obtaining missing stock price information.

- Media sites for significant political and public events.

- Methodologies and approaches on estimating the trade volume and profitability

Methodology

Case 1: Single Purchase and Sale

- Compute the estimated quantity of stock acquired by dividing the midpoint of the 'Value Range' by the purchase 'Price.'

- Derive profit by multiplying the quantity of stock by the difference in price between sale and purchase.

Case 2: Single Purchase and Multiple Sales

- Divide the midpoint of the 'Value Range' by the purchase 'Price' to estimate stock quantity.

- Allocate the purchased stock evenly across all sales, calculating profit for each sale by multiplying this allocated quantity by the respective sale 'Price,' and sum the profits.

Case 3: Multiple Purchases and Single Sale

- For each purchase, divide the midpoint of the 'Value Range' by the purchase 'Price' and sum the quantities.

- Calculate profit by multiplying the total estimated quantity of stock by the price difference between sale and aggregate purchase cost.

Case 4: Multiple Purchases and Sales

- Sum the midpoints of the 'Value Ranges' for all purchases, dividing by respective 'Prices' to find total quantity.

- Evenly distribute this total quantity across all sales, calculating profit for each by multiplying the distributed quantity by the sale 'Price,' and sum the profits.

Assumptions:

- The midpoint of the 'Value Range' represents the transaction's value.

- The quantity of stock bought and sold is assumed to be equal.

Considerations:

- If sell dates precede purchase dates, this methodology does not apply.

Alternative Methodology

Only consider the dates of the purchases and sells. Assume that there is always one share bought and sold. Calculate the profit based on the stock price difference on the purchase and sell date.

Questions

- Which methodology should I use, what adjustments do I need?

- Is there anything else I should consider/add?

- Do I need to do all the work with the data from the scraping point in R, or can I use Pandas to preprocess dataset from raw transactions, into trades data: <purchase date><purchase price><sell date><sell price><number of stock traded ><other info>?