Tracking Inflation's Impact on the S&P1500

Kevin Sy | USC ID: 9361-4024-01 | GitHub Username: KevS-777 (Link) | Google Drive

Introduction

In this post-pandemic economy, stock market volatility and rising inflationary pressures have created a highly unpredictable market environment. This case study's objective is to map out potential insights, trends, and anomalies that can illustrate the interplay between these economic forces.

How does inflationary pressure correlate with stock market performance indicators? Supplementally, across different sectors and market capitalizations of publically traded companies, which are considered high or low performers in this macroeconomic environment?

Data Collection

STOCK DATA: S&P1500 Composite for past 2 years - 4 Sources

Polygon.io API | Raw stock data [5,308,331 rows x 9 cols] | Enriched [743,971 rows x 22 cols]

This US stock market exchange data from Polygon.io API "Grouped Daily (Bars)" is the foundation of this data research as it provides daily historical information on stock data regarding individual tickers, stock price (Volume Weight Adjusted Price, VWAP), trade volume, and number of transaction made over the past two years (12/19/2022 to 12/13/2024).

Wikipedia API | S&P500 [503rows x 8col]; S&P400 [401rows x 8col]; S&P600 [602rows x 8col]

I collected the HTML contents of the Wikipedia pages that would be parsed into data tables for these three S&P500/400/600 representing chosen large/mid/small cap companies. This also has vital information on which GICS industry sector and sub-industry each company is classified under. As a collective, all 3 S&P lists make up the S&P1500 composite.

INFLATION DATA: Consumer Price Index (CPI) - 2 Sources USBureauofLaborStatisticsAPI

inflation_all data [239 rows x 7 col] | inflation_less_food&energy [239 rows x 7 columns]

The Consumer Price Index (CPI) is the average change in price for a basket of goods in comparison to the years between 1982-84 which benchmarks at an index of 100. I use both the CPI for all items as well as Core CPI which excludes volatile categories of food and energy.

One example of a crucial knowledge gaps from the proposal to this study was understanding how CPI inflation metrics are calculated in order to cross-analyze with stock data time series. While data collectors may gather gather price data during particular weeks in the year, the aggregate is what makes up the index. This along with learning technical stock metrics like VWAP required subject matter expertise research online.

Since the proposal, the high number of proposed data sources have been removed to focus on more salient datasets. The Fed fund interest rates, Treasury yield, and PPI inflation data were not included in this particular research.

Data Cleaning

In its original form, Wikipedia API requests provide the HTML code on a txt file. I parsed the relevant data tables using pandas.read_html and downloaded them in a csv file. The Polygon stock data API requests came in the form of a JSON that I eventually stored as a csv file. During intermediary data cleaning, analyzing, and visualization, I relied on pandas DataFrames for data manipulation techniques. I chose this avenue for data processing because tabular data format is easier to manipulate and could also be easily exported for future analysis rather than just relying on terminal output logs. The csv file also made it possible for different .py files to run and read the data as it went through different stages of data processing.

To prepare the stock dataset, I needed to merge the US stock market exchange data with the S&P classification information. The S&P500/400/600 datasets were concatenated into a single S&P1500 dataset and then merged with the stock market data. I dropped any rows that weren't a part of the S&P1500 for this study. [743971 rows x 12 columns]. I changed the format of the timestamp from milliseconds from Epoch to a datetime object so it would be comparable to the inflation data at a later point. With this original subset of columns, I feature engineered additional calculation columns. I added metrics to capture price volatility, magnitude of trading volume, and rolling moving averages.

The CPI inflation data came in the form of a JSON which would be converted to a pandas DataFrame. One of the trickier parts of handling this data was that there was a nested dictionary in one of the calculated columns provided. This calculations column contained the various 1/3/6/12 month percent change CPI. CPI inflation data is also collected on a monthly basis. So when comparing CPI data with stock data, I downsampled and aggregated the stock data to be consistent with CPI time series. This left me with 24 observations (24months) for the merged data.

Data Analysis

Now that the stock data has been cleaned and enriched, I looked at overview statistics to get a sense for how industries and S&P categories differed from one another. I used a combination of .describe(), groupby().agg(), and pivot functions for the entire stock data and subsets (S&P categories and GICS Sectors). This [spreadsheet] has the full statistics available.

Overview S&P1500: For the S&P which accounts for around 90% of the market capitalization of the US stock market, the across-the-board average VolumeWeightAveragePrice is $107.81, AvgVolumePerTransaction is 72.61 shares, AvgPricePerTransaction is $4,972.94. This is a good baseline to benchmark against the other sub-categories. The OpenCloseDelta min-max ranges from 49.93% to 202.64% which shows that day trading ROI returns can result in a loss of half to a doubling of investment.

Top5 avg_daily_volume: NVDA, TSLA, PLTR, AAPL, AMD. This indicates that these companies have substantial trading activity and all 5 are in booming AI and tech driven industries.

Pulling the initial insights from a "Corr_Matrix_Stock_CPI_all(Heatmap).png" and "Corr_Matrix_Stock_CPI_Core(Heatmap).png" = CPIvsVWAP: +0.903,+0.923; CPIvsTransactions: +0.753,+0.766; CPIvsVolume: +0.024,+0.031. There is a extremely strong positive correlation between both CPI's and VWAP transactions of shares. There is markedly no correlation between CPI's and the volume of shares that are being traded.

To understand the directional relationship between CPI (All and Core) with the Stock Market (VWAP, Volume, Transactions), I used Granger causality F-test and chi2 statistical tests at a max-lag of 6, to identify the predictive capabilities of one time series to another. It is important to note that this is not a causal relationship and is only indicative of the temporal relationship between variables.

Critical Values: To reject the null hypothesis at df=1 df_denom=20: (Upper one-sided significance value: $p<0.05$, critical value >4.35; $p<0.025$, critical value > 5.87). At df=3 df_denom=14: $p<0.5$, critical value 3.34; $p<0.025$; critical value: 4.24

VWAP → Core&All CPI: The null hypothesis is strongly rejected, which means that VWAP Granger causes Core&All CPI. Stock prices is a strong predictor of CPI inflation in the short term 1 month lag, which could suggest that increasing stock market prices signal predicted inflation faster than CPI can be measured. This could reveal that stock assets are reflectionary of inflationary expectations that is yet to come.

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=11.0514 , p=0.0034  , df_denom=20, df_num=1
ssr based chi2 test:   chi2=12.7091 , p=0.0004  , df=1
likelihood ratio test: chi2=10.1180 , p=0.0015  , df=1
parameter F test:         F=11.0514 , p=0.0034  , df_denom=20, df_num=1
```

```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=12.1888 , p=0.0023  , df_denom=20, df_num=1
ssr based chi2 test:   chi2=14.0171 , p=0.0002  , df=1
likelihood ratio test: chi2=10.9454 , p=0.0009  , df=1
parameter F test:         F=12.1888 , p=0.0023  , df_denom=20, df_num=1
```

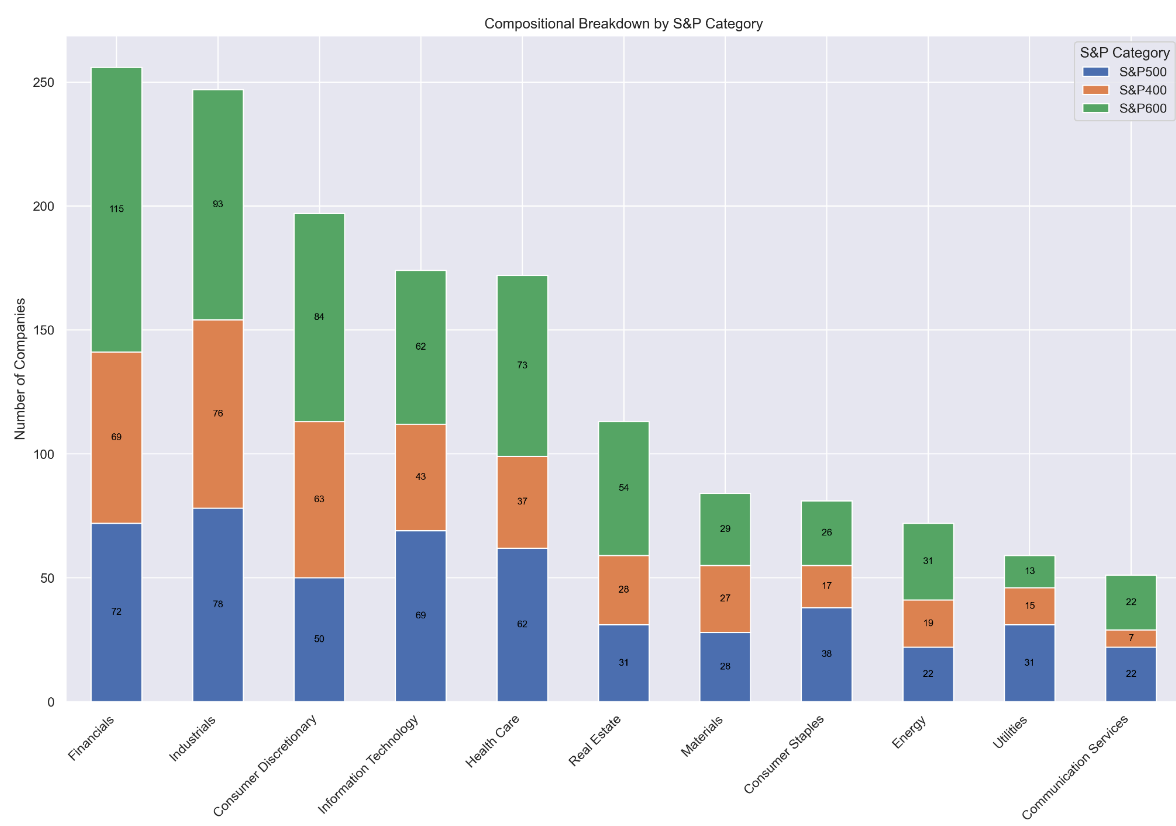
Transactions → Core&All CPI: The null hypothesis is strongly rejected for lags of 1-3, which means that Transactions Granger causes Core&All CPI. Stock trading activity is a strong predictor of CPI inflation across medium term time windows between 1 to 3 months and indicates that trading activity rises and falls in anticipation of near term inflationary predictions.

```
number of lags (no zero) 1
ssr based F test:         F=12.0435 , p=0.0024  , df_denom=20, df_num=1
ssr based chi2 test:   chi2=13.8501 , p=0.0002  , df=1
likelihood ratio test: chi2=10.8414 , p=0.0010  , df=1
parameter F test:         F=12.0435 , p=0.0024  , df_denom=20, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=9.7286  , p=0.0015  , df_denom=17, df_num=2
ssr based chi2 test:   chi2=25.1799 , p=0.0000  , df=2
likelihood ratio test: chi2=16.7844 , p=0.0002  , df=2
parameter F test:         F=9.7286  , p=0.0015  , df_denom=17, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=9.5803  , p=0.0011  , df_denom=14, df_num=3
ssr based chi2 test:   chi2=43.1115 , p=0.0000  , df=3
likelihood ratio test: chi2=23.4381 , p=0.0000  , df=3
parameter F test:         F=9.5803  , p=0.0011  , df_denom=14, df_num=3
```

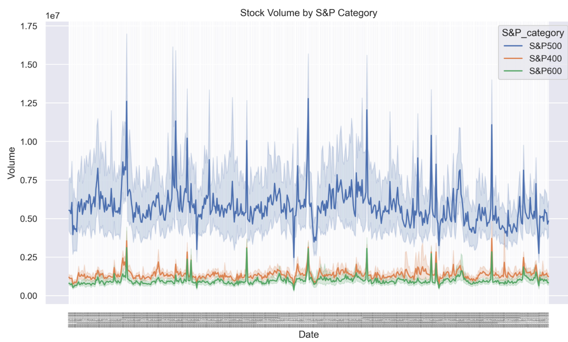
```
Granger Causality
number of lags (no zero) 1
ssr based F test:         F=10.0980 , p=0.0047  , df_denom=20, df_num=1
ssr based chi2 test:   chi2=11.6127 , p=0.0007  , df=1
likelihood ratio test: chi2=9.4007  , p=0.0022  , df=1
parameter F test:         F=10.0980 , p=0.0047  , df_denom=20, df_num=1

Granger Causality
number of lags (no zero) 2
ssr based F test:         F=4.9751  , p=0.0199  , df_denom=17, df_num=2
ssr based chi2 test:   chi2=12.8768 , p=0.0016  , df=2
likelihood ratio test: chi2=10.1372 , p=0.0063  , df=2
parameter F test:         F=4.9751  , p=0.0199  , df_denom=17, df_num=2

Granger Causality
number of lags (no zero) 3
ssr based F test:         F=4.4542  , p=0.0214  , df_denom=14, df_num=3
ssr based chi2 test:   chi2=20.0437 , p=0.0002  , df=3
likelihood ratio test: chi2=14.0724 , p=0.0028  , df=3
parameter F test:         F=4.4542  , p=0.0214  , df_denom=14, df_num=3
```

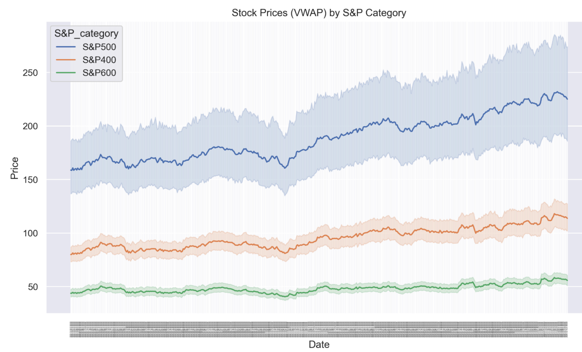
<u>Data Visualizations</u>

First delving into the US stock market stacked bar chart, we see that Financial (256), Industrial (247), and Consumer Discretionary (197) sector lead with the highest total number of companies within the S&P1500. Surprisingly, the aggregate number of Information Technology companies is in the middle of the group despite this sector getting the most buzz on the news reportings. The Information Technology sector ranks third in S&P500 representation, with 39.7% of its companies classified as large-cap, which is above the average. The Real estate sector has the highest percentage of its companies in the S&P600 small cap at 47.8%. [Stats]

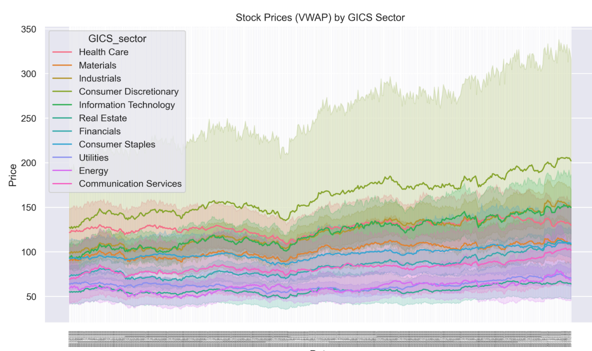


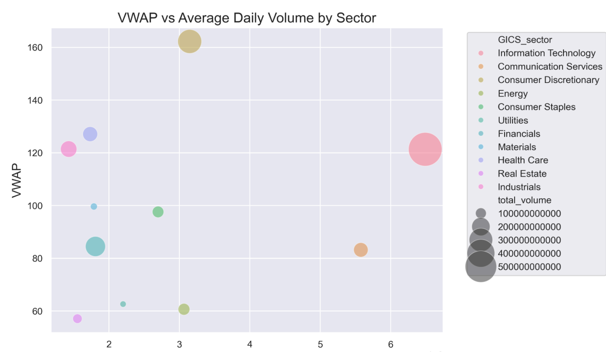
To give a scale of magnitude between the S&P categories, the S&P500 average daily trading volume per company is 5.908 times greater than the S&P600 and the average daily transactions per company is 4.886 times greater. The lineplot also displays the range in distribution for the volume and price at a lower opacity. It is evident the S&P500 large cap companies have a wider distribution of volatility comparative to the S&P600 small cap companies.



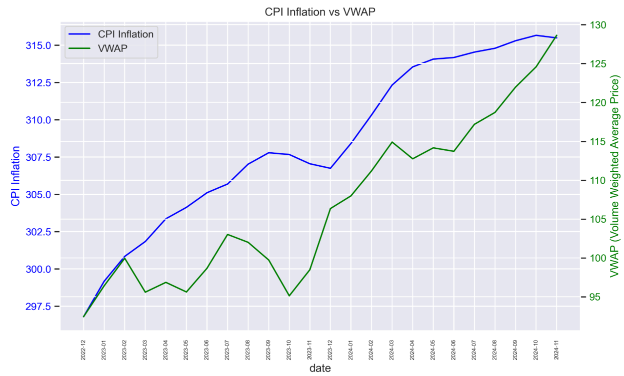


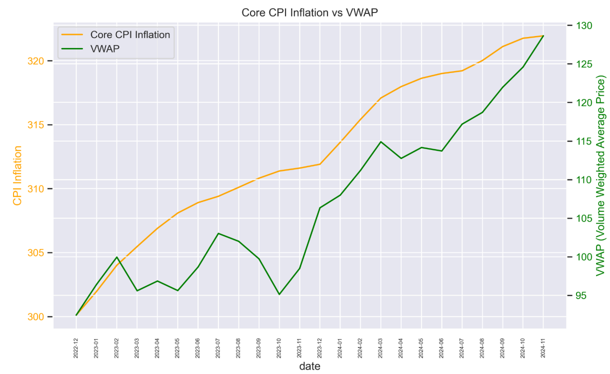
The S&P1500 broken up into GICS sectors reveal that Information Technology is a hotly traded sector with 564,335,625,201 total volume, an average of 6,492,257 average trading volume, and 43,024 transactions. The lowest trading activity occurs within the Industrial sector with 168,684,689,483 total volume, an average of 1,427,282 trading volume and 17,003 transactions. Over the past 2 years, the average stock price has trended upwards across all industry sectors with Consumer Discretionary holding the highest average price to buy in at $162.26 per share.





The dual axis lineplots showcase the visual relationship between both VWAP and Inflation over 2 years and are denoted in monthly data points (left: CPI; right: Core CPI). For recent Nov2024, the CPI index is 315.5 and Core CPI is 321.9. You can begin to see the subtle lag effect where the stock market price sharply drops around Oct 2023 and CPI follows with a decrease around 2 months later Dec 2023. This would coincide with Oct 13, 2023 when the stock market reacted poorly to the fresh September inflation report raising concerns of increasing inflationary pressures (source). This chart visually reinforces what was found in the Granger causality tests that the stock market volatility and reactions are a strong predictor of CPI numbers.





Conclusion

My initial hypothesis in my proposal was that inflationary pressure measured by the CPI index would be a predictor for stock market price and activity. In actuality, I found that it was the other way around. CPI inflation numbers lag behind aspects of stock market activity by around 1 to 3 months. The implication is that the stock market trading is a strong predictor for CPI inflation. My theory is that the aggregate stock market trading activity happens at an extremely volatile and reactionary manner to headwinds of predicted macroeconomic performance. So, if there is news or an expectation that inflation is fluctuating, individual traders pick up on the trends and fluctuations faster than CPI index can be calculated via data collectors. Stock market trades are profit driven to capitalize on the margins between when a share is bought and when it is sold. This incentivizes fast-natured speculative valuations of companies which also take into account how inflation will impact how much the US dollar will stretch in the future.

Future Work

I would continue utilizing Polygon.io for stock market research as stock data collection is easily accessible and scalable. The free plan limited historical data to 2 years, which meant the monthly downsampled dataset I used for granger causality tests was smaller scoped. There is dditional historical data available with a paid membership to increase up to 5-20+ years. This allows for longer duration cross analysis between stocks and CPI inflation. The US Bureau of Labor Statistics also has expansive metrics on PPI inflation, productivity, and employment.

Given more time, I would expand on these fields of study: Macroeconomic events, Government policies, and individual stock analysis. Fed fund interest is one of the main levers used to curb hyperinflation. Fed chair Jerome Powell has projected that in 2025, the Federal Reserve plans to decrease the number of rate cuts from four to two since inflationary percentages have been resistant to cool offs, utilizing Core CPI as the main index, while the labor market has remained strong (source). The cascading impacts of expectations to reduce rate cuts lead to an increase Treasury yields percentage, which makes stock market assets less attractive (source, source). President elect Trump also claims to implement an international tariff of 25% on all imported goods from China, Mexico, and Canada. This was dramatically alter the cost of the supply chain of goods and eventually impact the individual consumers' wallets across multiple sectors (source). From the interplay between macroeconomic events, government policies, and stock market reactions, deeper dives into individual companies can help forecast which companies are outliers or are undervalued. Ticker details can be pulled from Polygon.io for individual company stats like market cap, employee numbers, number of shares to further pull the string on invidualized portfolio analysis.