**Regression Report**

All data is stored in Google Drive <https://colab.research.google.com/drive/16Cw1crskZZbLkoSZyRPSwYxkFAB0Nk1c>

1. Data and Variables

Firstly, we download the daily close prices and volume for industrial ETF and S&P 500 components. For ETF, we extract data of 18 ETF from S&P 500 selected of GICS level 1 to level 3 ETF pool on Bloomberg Terminals, setting the period from 2/1/2011 - 9/30/2020 in line with the article. For S&P 500 components, we deploy Yahoo Finance’s API ‘yfinance’ in Python, referring to the component list on Wikipedia retrieved on 06/30/2024. The sample for components also covers the period from 2/1/2011 to 9/30/2020. For components specifically, as there are 5 stocks on 2024 list IPO after 2020, they are removed from the sampel pool as they have fully missing data.

Then, we download the daily 3-month fed treasury yield as our risk free rate, scaling it in percentage as well. With this Rf data, we calculate the feature ‘macro’ by S&P 500 return minus Rf to capture macroeconomic impacts on ETF and stock returns.

After collecting these data, we calculate daily returns in percentage and the corresponding pre-drift. Returns are calculated as:

Meanwhile, the pre\_drift is calculated in Python with function: .rolling(window=3).mean()

**Variable Description**

|  |  |  |
| --- | --- | --- |
| Name | Definition | Observation |
| *nega* | anger+disgust+fear+sadness | 46 |
| *negative emotion average* | anger+disgust+fear | 46 |
| *FFR* | interest change between meetings | 46 |
| *MPU* | Monetary policy uncertainty https://www.policyuncertainty.com/bbd\_monetary.html | 46 |
| *Negative Tone* | From paper <Words that shake traders> | 36 |
| *Hawkishness* | From paper <Words that shake traders> | 36 |
| *stock\_drift\_3d* | The stock price percent change in 3 days before the FOMC press conference, measured in percentage points. | 46 |
| *macro* | Rm-Rf | 46 |

**Statistical Description**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | *nega* | *Negative*  *Emotions Avg* | *Negative Tone* | *Hawkishness* | *EFFR\_CHG* | *MPU\_wrld* | *macro* |
| count | 46 | 46 | 36 | 36 | 46 | 46 | 46 |
| mean | 0.23 | 0.25 | 0.42 | 0.37 | 0.82 | 0.24 | 0.70 |
| std | 0.22 | 0.22 | 0.25 | 0.23 | 0.15 | 0.20 | 0.36 |
| 25% | 0.05 | 0.06 | 0.26 | 0.21 | 0.83 | 0.10 | 0.37 |
| 50% | 0.17 | 0.21 | 0.34 | 0.34 | 0.83 | 0.20 | 0.92 |
| 75% | 0.27 | 0.39 | 0.57 | 0.51 | 0.83 | 0.33 | 0.98 |

After getting all features and target variables, we scale them with the Python function MinMaxScaler() from sklearn.preprocessing package to unify the magnitude for all features and increase the likelihood of significance.

1. Regression: Return

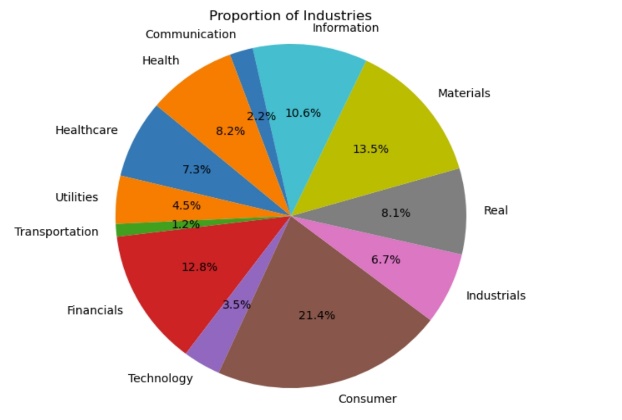
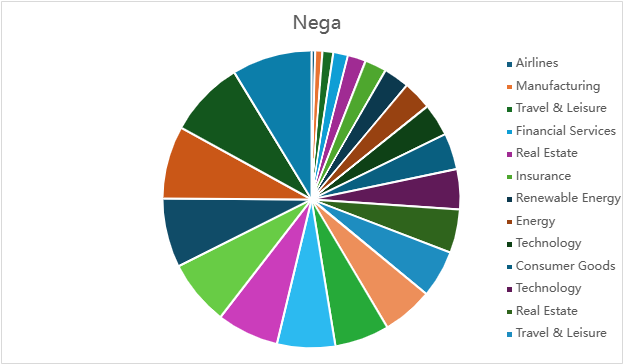
We conducted 4 different empirical combinations with different control variables and fixed effects. The combination of negative tone, hawkishness, quarter and chair fixed effect gives the best significance. With sad in negative emotion, we obtained 23 significant industry with medical (negative) and software (positive) listing the top 2.

1. Regression: Volume
   1. Data Processing

We use np.log to compute the derivative of trading volume for SP500 constituent stocks between the day before and the day after from sp500\_V\_daily.csv excel . After organizing the time series data, we exclude all zero and NA values, align the data with Data.xlsx, and merge based on ‘Date’, retaining control variables: 'Negative Emotions Avg', 'Negative Tone', 'Hawkishness', 'Chair'. Dummy variables are generated based on 'Chair'. The regression code is in the sp500v\_t4.ipynb, and the regression results and summary are in the Regression\_Results\_with\_nega.xlsx. In particular, we computed 'Negative Emotions Avg' using a specific method.

* 1. Regression Data Using OLS

we regress each cleaned SP500 stock separately. We conducted a multiple regression model on a total of 440 stocks, summarizing coefficients, t-values, and p-values respectively, which were then organized into Excel. Under a 90% significance level, we compiled stocks where different control variable coefficients showed significance. We found that negative emotion avg coefficient is much larger than negative emotion avg + sadness avg and the amount of stock that is statistically significant is also larger.



**Industry Distribution with Negative nega** **Industry Distribution with Negative nega**

Negative coefficients (e.g., AAL, CCL, COR, etc.) suggest that an increase in the ‘nega’ is associated with a decrease in the stock volumes. Positive coefficients (e.g., AME, FSLR, LEN, etc.) suggest that an increase in the independent variable is associated with an increase in these stock volumes.