Motivation

CONTEXT AND MOTIVATION

We are interested to analyze price sensitivity of sector ETFs to external factors. Economists, traders, and investors are generally interested to understand what variables contribute to the market volatility /movement to inform their decisions on investment, hedging, and others. We have heard economists, traders' and investors' beliefs, experiences, and assumptions on what variables (whether they are specific trends, news, economic reports, earnings, seasons, and many others) actually move the equities market. Some do hold credentials for their claims due to their knowledge and experience. However, many do not show the reliability of their claims and ideas with data other than telling them "it's their experience". Experience is important, however, some things don't come with experience since our brain is not capable of holding large amounts of dynamic information together and analyzing them continuously considering how dynamic the equities market is. We can't depend on data entirely as well, but we need it to complement parts of our decision making process or reevaluate some ideas and assumptions continually. Also, past performance and experience may not be strong indications of the future market performance. You won't know things have changed or shifted unless you have ways to gather the information quickly and monitor them continually.

PROJECT GOALS

Our ultimate goal is to develop an analysis and visualization framework that allow us to fulfil the following:

- Analysing the sensitivity of the market price change, direction, or both to various variables on multiple sector ETFs. This will reveal the relationship between various variables (econometrics, earnings season, special events, ...) and the price movement. And may also lead to findings as to which variables may be the most influential.
- Evaluate traders' assumptions about specific variables that they think will move the equities market.
- Analyse the strength (if any) of various variables across time to see if assumptions still hold.
- Allow us to expand the analysis and visualization to include new variables by abstracting the analysis process. This will help us to quickly evaluate the variables.

Data Sources

Name	Description	Details	Access
Yahoo Finance Ticker Data	The access to the ticker data is facilitated by the finance python module. Dataframes will be created using the module for each sector specific ETFs listed here. Data is structured as a timeseries facts table with open, high, low, and close prices as well as trade volumes, dividends and stock splits. Additionally the CBOE Volatility Index is used which can be found under the "^VIX" ticker name with open, high, low, and close daily measures.	**Size:** 19MB **Shape:** 57K rows **Format:** Pandas DataFrame	YFinance API
Daily Financial News Data	The dataset will be used to bring in newsworthy events that may impact equity prices. The features we will be using are date and headline.	**Size:** 400MB **Shape:** 1.85M rows **Format:** CSV	Kaggle
Observance Dates	Contain multiple dates for each observance or holiday period for each year. For example, Labor Day occurs on every Monday within a specific range of days depending on each specified year. The dates also extends to 3 trading days before and after the event date. Example, the extended days for Christmas will look like this: -3, -2, -1, Christmas, +1, +2, +3. The idea for the extension is derived from a book, "Almanac", by Jeffrey A. Hirsch.	**Size:** 38KB **Shape:** 231 rows **Format:** CSV	Kaggle
Santa Rally	Contain dates for Santa Rally period. Santa Rally occurs 5 days before and 2 days after the end of December.	**Size:** 3KB **Shape:** 238 rows **Format:** CSV	Kaggle
Triple Witching Week	Contain dates for Triple Witching Week. It occurs on the 3rd week of the end of each quarter.	**Size:** 14KB **Shape:** 780± rows **Format:** CSV	Kaggle
Economic Reported Dates	Contain US economic reporting dates for various economic reports such as ISM PMI, Non-farm payroll, FOMC, OPEC, CPI, building permits, and others. The dates was collected from Forexfactory.	**Size:** 99KB **Shape:** 1283 rows **Format:** CSV	Kaggle

Data Manipulation Methods

1. ENRICH THE SECTOR ETF DATA WITH ADDITIONAL METRICS

The manipulation of the Sector ETF Ticker data was focused on creating an additional set of metrics to enable analysis on price movement during the trading day or in between trading days. The metrics below were calculated and added on as new columns to each dataframe.

1.1 PRICE CHANGE

- **price_chg_c2o:** Percent change between previous close and current open.
- **price_chg_o2c:** Percent change between the current day's close and open.
- **price_chg_c2c:** Percent change between previous and current day close price.

1.2 ROLLING VOLUME

- **volume_rollmed:** Rolling median of volume across a 90 day time window.
- **volume_pchg_from_med:** Percent change from the rolling median.
- **volume_diff_to_med:** Difference between current day's volume and the rolling median.

1.3 PRICE CHANGE T-SCORES

For measuring the magnitude of price, a t-score is added for each price change metric **tscore_c2o**, **tscore_o2c**, and **tscore_c2c**. Each represents the amount of standard deviations the current day's measure is away from the rolling mean.

2. ENRICH AND PREPARE THE VOLATILITY INDEX TICKER DATA

For the Volatility Index Ticker data, not all of the measures are relevant or useful. This step focuses on creating a subset of metrics defined above. Also all column names besides the date are prefixed in preparation for joining Volatility Index metrics with Sector Index metrics in the subsequent stage.

2.1 VIX CHANGE

- **vix_chg_c2c:** Same definition and implementation of **price_chg_c2c**.
- **vix_tscore_c2c:** The amount of standard deviations of the current day's close to close metric away from 360 day rolling average.

3 COMBINE ALL SECTOR ETFS AND THE VOLATILITY INDEX

For convenience and efficiency in the next steps of analysis, all of the enriched Ticker data is combined into a single data frame.

3.1 SECTOR ETFS ARE MERGED TOGETHER

For convenience and efficiency in the next steps of analysis, all of the enriched Ticker data was combined into a single data frame.

3.2 ENRICHED VIX DATA IS JOINED TO SECTOR ETF DATA

The join is a left outer join by the date column so Volatility measures are available for each day of each Sector ETF's measures.

Data Manipulation Methods

4. PROCESS STOCK NEWS DATA

The goal of this step is to transform the raw stock news data into a comparable schema as our “additional events of interest” data sources. The raw news data has one row for each news article along with attributes of that article such as headline and publish date. The events of interest data sources have a column per event, and each column holds the list/ series of dates they event occurred.

4.1 DEFINE THE TOPICS AND KEYWORDS

Drawing from a review of the news data, as well as existing knowledge of financial news we defined a set of news topics along with a set of keywords which are commonly associated with the topic. Each topics is to be considered an “event of interest” and the set of dates when there are articles published on that topic will be considered occurrences of the event.

4.2 FOR EACH TOPIC, DETERMINE THE RELATED ARTICLES

The relation of articles to topics is defined by looking for a topic’s key words in the article’s headline. If the article has one or more key word for a topic, then it is considered related to that topic.

4.3 DETERMINE THE DATES WITH ARTICLES RELATED TO EACH TOPIC

Using regex findall on the source data’s headline column, a new Pandas Series is created which contains an array column of all keywords contained in the article’s headline.

A new dataframe is created from the series using Pandas to_list() method so as to have a new column for each occurrence of a key word in a headline. All strings are lowercase at this stage.

Using the same topic/keyword dictionary and Pandas apply and map functions, each key word string is converted back to it’s related topic.

Dates are now joined back into this dataframe through the common index and the dataframe is pivoted

5. COMBINE EVENT DATES DATA

As the individual CSV files with event dates are all of the same format of columns with event names. These are all combined together by loading each CSV file into and using pd.concat to combine them with the axis=”columns” setting.

6. COMBINE TICKER DATA WITH DATE DATA

For each event of interest, a new column is added in the Sector ETF time series data created in the previous phase. The column is boolean as integer, and is set to 1 if there is an occurrence of the event of interest on the date and 0 if there is not. This enables the future step of analyzing impact of events of interest on our measurements as we can now have the intersection of every day’s measure for every equity with every occurrence of each event of interest

Analysis | 1. Identify Influential Variables / Factors

Figure 1.1. Description...

1.1. PRICE MOVEMENT & VOLUME AS A RESULT OF THE EVENT

Figure 1.1. shows that variables like FOMC Press Conference, OPEC meeting, ISM PMI Manufacturing reporting day, and Black Friday and Cyber Monday period seem to play an influential role on specific sector ETFs.

The t-score that we used to measure closing price change after normalising with it standard deviation doesn't take direction into account as we are only interested in the magnitude of price change and not the market sentiment on the outcome whether it's positive or negative. And we only measure the effect on the day the event occurs, not including the day after. Here are some of the findings we have garnered.

First is FOMC Press Conference. The primary objective of the conference is to continually optimise the monetary policies to address the current economic challenges related to employment, inflation, growth, demand, and money supply (APA, ...). We find that all the 11 sectors are quite sensitive to the outcome of the FOMC Conference on that particular day with an average magnitude of 1.0 t-score (one standard deviation) on price change (figure 1.2.). The volume also tends to rise above the median on this day according to figure 1.3. which means more participants tend to involve in the market during the specific event.

Second comes the OPEC meeting. The objective of OPEC is to coordinate petroleum policies among countries that are chosen to be part of OPEC member, to ensure optimal production and supply, hence stable oil price (APA, ...). And, according to our findings, the outcome of the OPEC meeting on that day seem to have an influence on sectors like materials, energy, financial, technology, utilities, and staples with an average magnitude of 0.9 t-score (figure 1.2.). It is not surprising to see the influence of petroleum policies on sectors like materials, energy, and utilities since they are related to oil production, equipment, transportation of materials, shipping, and power supply. However, it would be interesting to investigate sectors like staples, financial, and technology to validate the effect. Similar to FOMC, the volume tends to rise above the median on this day (figure 1.3.).

The ISM PMI manufacturing reporting day also tends to have greater price movement for materials and industrial sectors with an average of 0.9 t-score (figure 1.2). ISM PMI report is a leading indicator that surveys the future prospect related to manufacturing. The effect is less about the reliability of the survey, but more on the response of the market towards the report. It's not difficult to map the

Analysis | 1. Identify Influential Variables / Factors

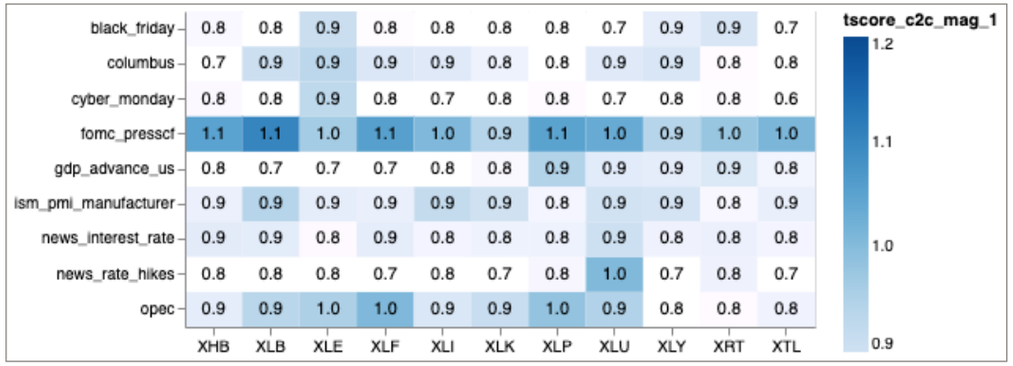

Figure 1.2. Description...

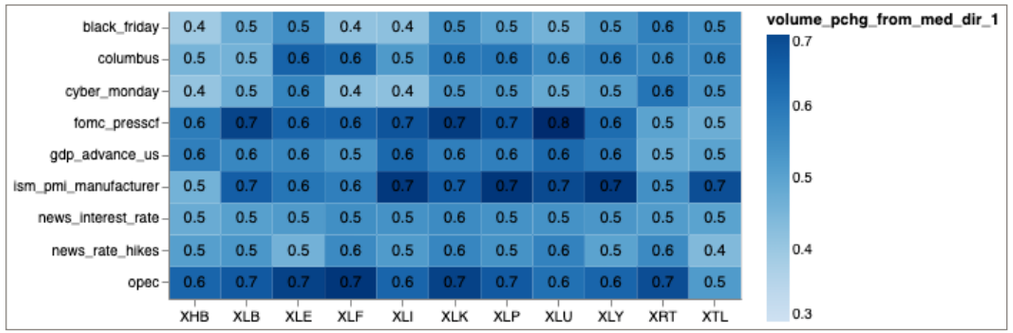

Figure 1.3. Description...

relationship here since materials and industrials sectors are related to manufacturing by nature.

Another interesting variables / factors are Black Friday and Cyber Monday. Our expectation is that events are more likely to affect retail, staples, and discretionary sectors although the findings show an otherwise result. However, this is just a high level analysis using t-score to measure the price movement. Furthermore, energy sector tends to have higher beta or price movement than other sectors. Nonetheless, supply chain and logistics do play a major role during events like these, thus the effect on the energy sector could be due to the demand and supply for oil, plus market reaction on those days.

1.2. IMPLIED VOLATILITY CHANGE AS A RESULT OF THE EVENT

We also validate the price change by measuring the changes in VIX (implied volatility index) to see if the price movement is also supported by a significant change in implied volatility. Based on (figure 1.4.), we find that only FOMC Conference tend to move the VIX by around 1 standard deviation on average (0.9 t-score), followed by OPEC with 0.8 of t-score, while the rest is below 0.8.

Analysis | 1. Identify Influential Variables / Factors

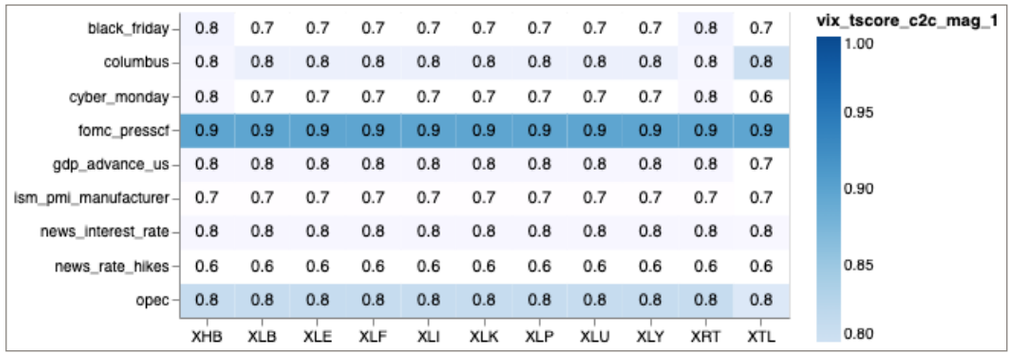

Figure 1.4. Description...

1.3. LIMITATION & CHALLENGES

We are aware, however, that the variables we identified above based on the t-scores are only around 0.2 higher than the rest, which can likely be due to random effect. Likewise for VIX. And, some events have limited sample size since they occur once a year.

Also, since the data can only measure the price change between the previous and current day closing price, we have essentially ignored the narrower timeframe such as price change between the market hours. Price can fall or rise a lot after the result is released or an announcement, and retract after the market has cooled down within hours. Hence, the effect can be watered down. Likewise for VIX.

Furthermore, the analysis doesn't take omitted variables, uncertainty measure, and confounding variables into account. Plus, some variables / factors might allude to an uptrend or downtrend with longer days as a result of the outcome even though the price movement is small, but we are only measuring the price movement on a very short duration.

It is purely a high level scanning and analysis for variables that seem to stand out among the rest for further investigation in the future. Hence we decide to emphasise them.

Analysis | 2. Analyse Influential Variable's Effect Over Time

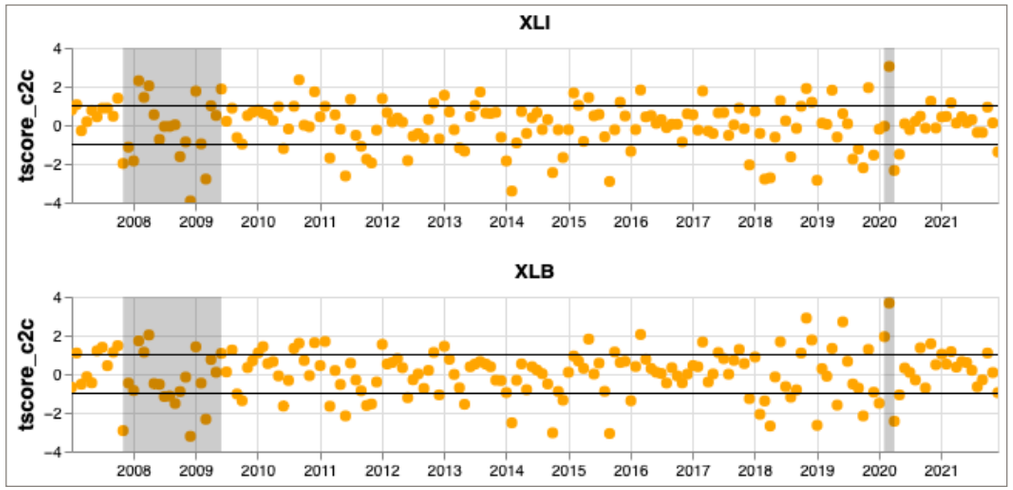

Figure 2.1. Description...

2.1. VARIABLE'S EFFECT OVER TIME

We are interested to observe the effect over time to analyse the sustainability of the effect across time and impact during specific period such as recession after narrowing down our focus to specific variables.

As we observe figure 2.1., we find that the price movement on the ISM PMI Manufacturing reporting day tends to be larger a few years before and during the recession as most of the dots fall outside the t-score range between -1 to 1. We highlight the range with rectangular boxes. This outcome could be due to materials and industrials sectors becoming more reactive to the ISM PMI result (leading indicator for manufacturing) as the economy is falling from the peak or approaching a recession.

The same scenario occurs for Black Friday (figure 2.3.), where the price movement seems to increase during the recession period (refer to the grey box) while the other periods have narrower movement.

Analysis | 3. Analyse Influential Variable's Effect Over Time

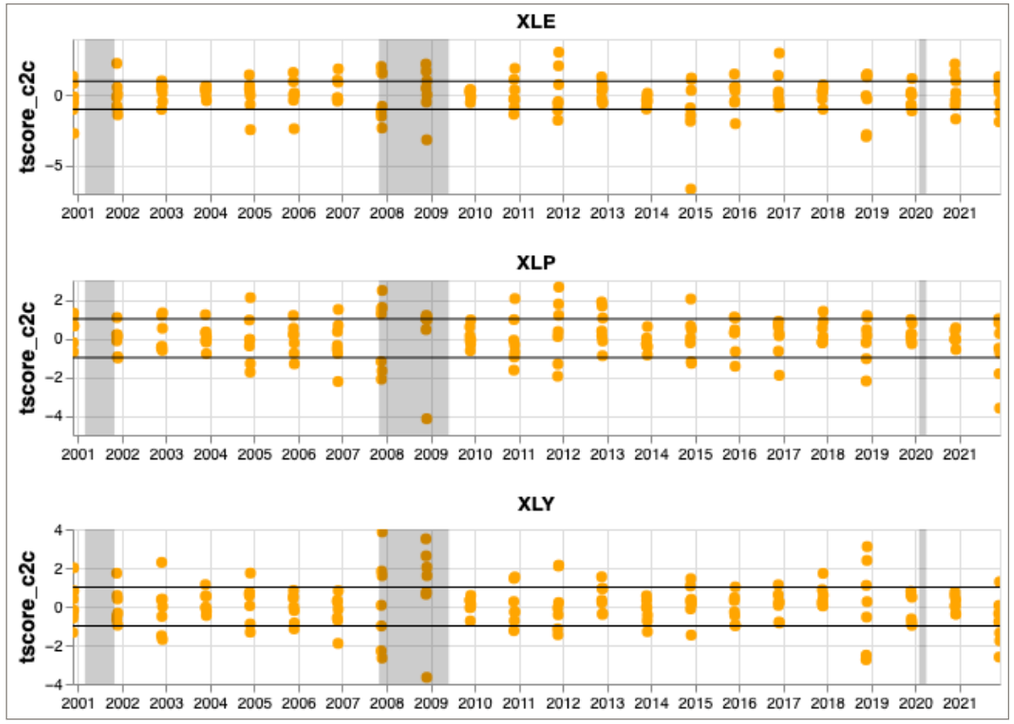

Figure 2.2. Description...

1.2. LIMITATION & CHALLENGES

Since we can only gather data from a very limited time range for both the ETFs and events, we aren't able to compare the price movement on multiple recessionary period to observe the consistency. However, when we compare it with other sectors for ISM PMI, materials and industrial sectors seem to show an increasing price movement during the recessionary period compared to other sectors (figure 2.2.).

Also, the volatility is generally higher during the recessionary period, so the movement is likely to be magnified for other observances or events, which makes effect from ISM PMI and Black Friday less impactful.

Analysis | 3. Summary & Moving Forward

3.1. SUMMARY

We narrow down to FOMC Press Conference, OPEC Meeting, ISM PMI Manufacturing reporting day, and Black Friday and Cyber Monday period after identifying the variables with high average price movement.

As we analyse the effect over time, we find that price movement during the FOMC, OPEC, and Black Friday event day is magnified during the recession. Likewise for ISM PMI. However, materials and industrials sectors seem to become more reactive to ISM PMI with increasing price movement as the economy is falling from the peak or approaching a recession.

3.2. SUGGESTIONS & CONSIDERATIONS

Days like FOMC Conference, OPEC Meeting, ISM PMI Manufacturing reporting day, and Black Friday and Cyber Monday period tend to be more volatile. With that, we have 3 considerations.

- Traders can exit their position days before the event occurs to avoid volatile period if the market reaction and price movement is unpredictable.
- If traders were to hold their position during those period, they may consider hedging to protect profits or mitigate losses until the dust is settled.
- Volatile days can be a profitable also if traders aren't trading the direction but price movement. Meaning prices can either move up or down as long as the movement is large enough to secure an attractive profit. However, if the movement is small, trades might suffer losses due to example like time decay if it's an option trade.

3.3. MOVING FORWARD

We look forward to analyse the hourly price movement for specific variables / factors mentioned above. As mentioned, certain events may cause significant price movement within hours, but retract after the market has cooled down. If the result does present a significant movement, it might offer an opportunity to trade the volatility depending on the context for that day.

To measure the hourly movement, we need to get our hand on hourly market data and compare the movement over multiple years, period, or economic cycle. After identifying the period and average price movement, we can then find out the context to construct the narrative and reasoning.

References

Census Reporter. (2019a). *Alpena, MI*. https://censusreporter.org/profiles/16000US2601740-alpena-mi/

Census Reporter. (2019b). *Lexington-Fayette, KY*. https://censusreporter.org/profiles/16000US2146027-lexington-fayette-ky/

Census Reporter. (2019c). *Household Income in the Past 12 Months (In 2019 Inflation-adjusted Dollars)* [Data set]. Retrieved August 16, 2021, from https://censusreporter.org/data/table/?table=B19001&geo_ids=16000US2146027,16000US2601740&primary_geo_id=16000US2146027

FRED, Federal Reserve Bank of St. Louis. (2021a). *Unemployment Rate in Alpena County, MI* [Data set]. Retrieved August 16, 2021, from https://fred.stlouisfed.org/series/MIALPE7URN

FRED, Federal Reserve Bank of St. Louis. (2021b). *Unemployment Rate in Lexington-Fayette, KY (MSA)* [Data set]. Retrieved August 16, 2021, from https://fred.stlouisfed.org/series/LEXI421URN

Gale Business DemographicsNow. (n.d.-a). *Consumer Expenditure Comparison Report* [Data set]. Retrieved August 16, 2021, from https://https://dnow-gale-com.proxy.lib.umich.edu/dnow/?p=DNOW&u=lom_umichanna

Gale Business DemographicsNow. (n.d.-b). *Consumer Expenditure Restaurant Detail Comparison* [Data set]. Retrieved August 16, 2021, from https://https://dnow-gale-com.proxy.lib.umich.edu/dnow/?p=DNOW&u=lom_umichanna

Google. (n.d.-a). [Google Maps Lexington, KY, USA Restaurants]. Retrieved August 16, 2021, from https://www.google.com/maps/search/Restaurants/@38.0280802,-84.6126788,11z/data=!3m1!4b1?hl=en

Google. (n.d.-b). [Google Maps lexington fayette hockey arena]. Retrieved August 16, 2021, from https://www.google.com/maps/search/lexington+fayette+hockey+arena/@38.0229325,-84.5028275,12z?hl=en

World Population Review. (n.d.-a). *Alpena Population by Year* [Data set]. Retrieved August 16, 2021, from https://worldpopulationreview.com/us-cities/alpena-mi-population

World Population Review. (n.d.-b). *Lexington Population by Year* [Data set]. Retrieved August 16, 2021, from https://worldpopulationreview.com/us-cities/lexington-ky-population

Appendix