**Exploratory Data Analysis**

**Project 1**

**Sector performance during and immediately after a recession.**

**By: Ash Tao, Chris Burley,**

**Maica Tran and Sabrina Tsan**

Table of Contents

[1. INTRODUCTION 3](#_Toc99289637)

[Aim 3](#_Toc99289638)

[2. STEP BY STEP APPROACH 3](#_Toc99289639)

[Data Fields and Variables – 3](#_Toc99289640)

[Approach to breakdown the problem 4](#_Toc99289641)

[**What we already know (Baseline knowledge, research, and studies):** 4](#_Toc99289642)

[**Cleansing Data** 4](#_Toc99289643)

[**Analysis** 4](#_Toc99289644)

[**Time-based analysis** 4](#_Toc99289645)

[Results - Analysis 5](#_Toc99289646)

[**1. We can tell some trends from the cause of a recession** 5](#_Toc99289647)

[**2. Trade volume makes the performance differences:** 6](#_Toc99289648)

[**3. the previous price can help to make a judgment of buying or selling:** 11](#_Toc99289649)

[**4. The correlation between sectors become stronger than ever:** 13](#_Toc99289650)

[**5. Continue holding or buying energy sector doesn't necessarily mean dangerous:** 15](#_Toc99289651)

**1 -** **INTRODUCTION**

**Aim:**

The aim of this project is to analyse historical stock market data and attempt to draw insights relating the ways sectors within the S&P500 behave during, and immediately after, a recession. We will try to use those insights to predict future market behaviour. In addition we will seek to determine the factors which impact the length of a sectors recovery, post recession.

**S&P 500:**

The S&P 500 is a group comprising the largest 500 publicly listed companies based on market capitalisation.

The S&P 500 is widely regarded as one of the best indicators of large company performance and of the performance of the stock market more broadly.

**S&P 500: Sectors**

The companies listed on the S&P 500 can be categorised using the following eleven sectors:

|  |  |  |  |
| --- | --- | --- | --- |
| 1. Energy | (2) Material | (3) Industrials | (4) Consumer (Discretionary) |
| (5) Consumer (Staples) | (6) Health Care | (7) Financial | (8) Technology |
| (9) Communications | (10) Utilities | (11) Real Estate |  |

2 - DATA ACQUISITION

Our stock price data and sector performance data was obtained through Yahoo Finance.

## 

**Data Fields and Variables:**

**yfinance**: A python package that downloads market data from the Yahoo! Finance's API.

**yf.download()**: A function which can download the market data based on the selected time period.

With the adjustment of the start and end dates below, we were able to send a request to the Yahoo Finance website and acquire the data we needed for the relevant timeframe.

Scatter chart

Description automatically generated with medium confidence

2 - DATA ACQUISITION (continued)

Once downloaded, the SECTOR data that the yfinance package provides includes:

|  |  |
| --- | --- |
| **Fields** | General field information. |
| **Date** | The date for this record. |
| **Open** | The opening price on that date. |
| **High** | The highest price on that date. |
| **Low** | The lowest price on that data. |
| **Close** | The closing date on that date. |
| **Adj Close** | Amended stock closing price after any corporate actions. |
| **Volume** | Traded volume on that date. |

Once downloaded, the INDIVIDUAL COMPANY data that the yfinance package provides includes:

|  |  |
| --- | --- |
| **Fields** | General field information. |
| **Symbol** | ID for the stock. |
| **Shortname** | Name of the stock. |
| **Longname** | Name of the stock. |
| **Sector** | Sector the stock belongs to. |
| **Industry** | Industry the stock belongs to. |
| **Currentprice** | The latest price of the stock. |
| **Marketcap** | The combined value of all shares for the company. |
| **Ebitda** | Earnings before interest, taxes, depreciation, and amortisation. |
| **Revenuegrowth** | The amount of money the company makes over a pre-determined time period compared to the previous, identical amount of time. |
| **City** | The register city of the stock. |
| **State** | The state of the stock. |
| **Country** | The city of the stock. |
| **Fulltimeemployees** | Number of full-time employees |
| **Longbusinesssummary** | Business summary of the company. |
| **Weight** | The percentage value of each stock in the **portfolio.** |

**Feature Table: Feature Table Trying to calculate the weight of each sector**

|  |
| --- |
| **Fields** |
| **Sector** |
| **Marketcap** |

3 – IMPORTANT DEFINITIONS and BASELINE ASSUMPTIONS

During a recession the total market index will decrease.

History tells us that the market will **ALWAYS** recover from a recession. We know this because it hit a record high in January 2022.

In addition, we know that **ALL** sectors recover from a recession, albeit at different rates. This can be seen in the figure below showing the market ‘bounce-back’ after the most recent recession brought about by COVID 19:

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, line chart

Description automatically generated

Later in the report, we will feature a deeper dive into the post COVID-19 recession recovery.

3 – DEFINITIONS and ASSUMPTIONS (continued)

For the purposes of this project, please use the following definitions:

|  |  |
| --- | --- |
| **Daily Return:** | A metric for analysing index performance.  It represents the daily change in price as a percentage of the opening price. |
| **Moving Average:** | A [technical analysis](https://www.investopedia.com/terms/t/technicalanalysis.asp) tool that smooths out price data (removing the sudden undulations). It does so by creating a constantly updated [average price](https://www.investopedia.com/terms/a/averageprice.asp).  The average is taken over a pre-defined period, like days, minutes, weeks.  The time period can be updated to suit the needs of the analyst. |
| **Daily Trade Volume:** | The total number of a particular share that are traded on a given day. |
| **Expected Return & Risk:** | **Can be displayed on a scatter plot and is based on the mean and standard deviation of daily return.** |
| **Quotient of Recovery (QRec):** | This quotient is a metric which has been designed by the group. It is a measure of how long recovery takes relative to the length of the recession. It allows for the comparison of sector recoveries after recessions of DIFFERENT LENGTHS. |

**Calculating the Quotient of Recovery:**

**Using the Quotient of Recovery:**

QRec = 1.5 - Means the recover took 1.5 times longer than the length of the recession.

QRec = 0.75 - Means the recover took 75% of the length of the recession.

**Why use a Quotient of Recovery instead of ‘days’?**

Recovery can indeed be measured in days, and when comparing sector performance after a particular recession it serves us well. The problem we encounter, however, is when we wish to compare sector performances across multiple different recessions. Our group feels that a recovery will likely take longer if the recession, itself, was longer. Thus, having a measure such as QRec, which takes into account the recession length, was seen by the team as beneficial when attempting to compare recovery data for the sectors over time.

3 – DEFINITIONS and ASSUMPTIONS (continued)

**A demonstration of how you might identify the quantities needed to calculate a QRec value.**

|  |  |  |  |
| --- | --- | --- | --- |
| (1) | Chart, line chart  Description automatically generated | (2) | Chart  Description automatically generated |
| (3) | Chart  Description automatically generated | (4) | Chart  Description automatically generated |
| (5) | Chart  Description automatically generated | (6) | Chart  Description automatically generated |
| (7) | Chart  Description automatically generated | (8) | Chart  Description automatically generated |

4 – CLEANING THE DATA

The data retrieved from Yahoo finance does not need to be cleaned. All values of interest are present.

There is no missing data in either the “Sector” or “Marketcap”, CSV files.

5 – RECESSIONS IN FOCUS

Throughout this project the group will perform an exploratory data analysis (EDA) focusing on the first 3 recession periods in the table below.

All 6 recessions in the table will be used in the final section of the project where we seek to find a correlation between the ‘GDP decrease’ observed during a recession and the time since the last recession.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Name** | **Start** | **End** | **Duration**  **(Months)** | **Last Recession**  **(Months)** | **Unemployment**  **(Peak)** | **GDP decline**  **(Peak)** |
| **COVID 19** | Feb 2020 | April 2020 | 2 | 128 | 14.7% | −19.2% |
| **Global Financial Crisis** | Dec 2007 | June 2009 | 18 | 73 | 10.0% | −5.1% |
| **Dot Com**  **(9-11 attack)** | Mar 2001 | Nov 2001 | 8 | 120 | 6.3% | −0.3% |
| **Early 1990s** | July 1990 | Mar 1991 | 8 | 92 | 7.8% | −1.4% |
| **Early 1980s**  **(Part 1)** | July 1981 | Nov 1982 | 16 | 12 | 10.8% | −2.7% |
| **Early 1980s**  **(Part 2)** | Jan 1980 | July 1980 | 6 | 58 | 7.8% | −2.2% |

Source: Wikipedia

6 – OUR KEY QUESTIONS

|  |  |
| --- | --- |
| 1 | Do correlations exist in sector performance between the following metrics:   * closing price * price return * daily trade volume * moving average |
| 2 | How long does it take the market and its eleven sectors to recover after a recession? |
| 3 | Do all sectors take the same amount of time to recover after a recession? |
| 4 | Is there a link between the number of days a recession lasts for and the number of days it takes for the market to recover? |
| 5 | How has each sector fared in previous recessions? |
| 6 | Can we use previous recession data to predict how long recovery will take? |
| 7 | Can we identify the cause of a recession? |
| 8 | Are there factors that impact the length and severity of a recession? |
| 9 | What makes one sector perform differently to the others? |
| 10 | Can the previous price help to make a judgment of buying or selling? |
|  |  |

**7 - ANALYSIS -** COVID-19 RECESSION (Feb 2020 – April 2020)

**Market and Sector Performance during the COVID-19 recession:**

|  |  |
| --- | --- |
| Graphical user interface  Description automatically generated | Chart  Description automatically generated |

**Data frame showing recovery information for market and sectors:**

Table

Description automatically generated

**Bar charts showing sector QRecs during COVID 19 Recession:**

|  |  |
| --- | --- |
| Chart  Description automatically generated | Chart, line chart  Description automatically generated |

**7 - ANALYSIS -** COVID-19 RECESSION (continued)

|  |  |
| --- | --- |
| **Daily Return during the COVID 19 recession:**  **(By Sector)**  Background pattern  Description automatically generated with low confidence | **Daily Stock Volume during the COVID 19 recession:**  **(By Sector)**  A picture containing text  Description automatically generated |

**Daily Sector Return and Daily Stock Volume during COVID recession.**

**OVERLAID FOR COMPARISON:**

|  |  |  |
| --- | --- | --- |
| Chart  Description automatically generated | Chart  Description automatically generated | Chart  Description automatically generated |
| Chart  Description automatically generated | Chart  Description automatically generated | Chart, line chart  Description automatically generated |
| Chart  Description automatically generated | **Chart  Description automatically generated** | Chart, line chart  Description automatically generated |
| Chart  Description automatically generated | Chart  Description automatically generated |  |

**7 - ANALYSIS -** COVID-19 RECESSION (continued)

**10, 20 and 50 day Moving Averages for each sector during the COVID 19 recession.**

Diagram

Description automatically generated

**Correlation heat-map between sectors during the COVID-19 recession:**

Chart, treemap chart

Description automatically generated

**7 - ANALYSIS -** COVID-19 RECESSION (continued)

**Insights:**

* All sectors were negatively impacted by the recession, but the severity of that impact varied. This is demonstrated by the fact that:

1. All sectors recorded negative returns.
2. All sectors took at least some time to recover. Consequently, they each have a blue bar on the QRec charts above. The size of each blue bars indicates the duration of the sector’s recovery.

* The market took 1.11 times the length of the recession to recover.
* Real Estate, Utilities, Financial, Industrial, Consumer (Discretionary) and Energy all took considerably longer than the market average to recover.
* Materials, Consumer Discretionary, Health Care, Technology and Communications took considerably less time than the market average to recover despite all taking a hit of some degree.
* The slowest recover was Utilities which took 8.25 times the length of the recession to recover. Energy wasn’t too far behind, taking 6.74 times the length of the recession to recover.
* The fastest sector to recover was Health Care. It only took approximately 26% of the recession duration to recover.
* Trade Volume vs Daily Return insight needed.
* Volume measures the number of shares traded in a sector. Volume can indicate market strength. From the plot we generated, we found that “Daily Trade Volume" and "Daily Return" seem to have similar patterns through observation.
* Moving Averages insight needed. 50 days line missing.
* As can be seen from the plot above, the ‘Adjusted Closing Price’ oscillates from above to below the average line, making regular crossovers. The moving average lines can help to identify trend direction and to determine support and resistance levels. The short-term average is trending closer to price. NO GREEN LINE ON GRAPHS
* Correlation Heatmap insight needed.

**Explanations:**

International incidents can lead to stock market recessions. Different incidents can have vastly different impacts on sectors within the same market.

Sectors which are directly related to the incident can incur the greatest losses. The financial sector, for instance, experienced the largest losses in the 2008 Global Financial Crisis due to this incident centring around banks and other lending institutions. Similarly, the Communications sector was decimated during the Dot-Com recession due to the incident relating to a sudden decline in the price of internet-based stocks.

Other sectors may be virtually unaffected during a recession. The Health Care sector during the COVID-19 recession, for instance, experienced only a very minor loss.

Unfortunately for many investors, the cause of a market crash is only discovered after the huge decline in stock prices.

**8 - ANALYSIS -** GFC RECESSION (Dec 2007 – June 2009)

**Market and Sector Performance during the GFC recession:**

|  |  |
| --- | --- |
|  |  |

**Data frame showing recovery information for market and sectors:**

Table, Excel

Description automatically generated

**Bar charts showing sector QRecs during GFC Recession:**

|  |  |
| --- | --- |
| Chart  Description automatically generated | Chart  Description automatically generated |

**8 - ANALYSIS -** GFC RECESSION (continued)

|  |  |
| --- | --- |
| **Daily Return during the GFC recession:**  **(By Sector)** | **Daily Stock Volume during the GFC recession:**  **(By Sector)** |

**Daily Sector Return and Daily Stock Volume during GFC recession.**

**OVERLAID FOR COMPARISON:**

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

**8 - ANALYSIS -** GFC RECESSION (continued)

**10, 20 and 50 day Moving Averages for each sector during the GFC recession.**

Chart, map

Description automatically generated

**Correlation heat-map between sectors during the GFC recession:**

Chart

Description automatically generated

**8 - ANALYSIS -** GFC RECESSION (continued)

**Insights:**

* The Financial and Real Estate sectors incurred the biggest percentage losses in value throughout this particular recession. Of the remaining sectors, Consumer (Discretionary) fared best on this metric.
* All sectors were negatively impacted by the recession, but again the severity of that impact varied. As compared to the COVID-19 recession, each sector took a greater amount of time to recover relative to the length of the recession. The fastest sector to recover after this recession was Consumer (Discretionary). It took 58% of the recession length to bound back to pre-recession levels. In the COVID recession the fastest recovery time belonged to the Health Care sector at 26% of *its* recession length. Consumer Staples was the second fastest to recover with a QRec of 1.03.
* The Financial and Real Estate sectors incurred the biggest percentage losses in value throughout this particular recession.
* The market took 2.42 times the length of the recession to recover.
* The variation in recovery times across the sectors is less for the GFC than it was during the COVID-19 recession. The range in QRecs is 4.6 as compared to 8. Sectors had a more similarity negative experience throughout the GFC.
* Except for Consumer (Discretionary), all sectors took longer than the duration of the recession to recover.
* Only Financial and Utilities underperformed by way of recovery as compared to the market average. All others matched or outperformed the market average.
* The peaks and deeps on Return & Volume always seem to appear in pairs. The turning point of volume (the direction change of acceleration) often indicates the turning point of return.
* Need a moving averages insight.
* Need a correlation heatmap insight.

**9 - ANALYSIS –** DOT-COM RECESSION (Dec 2007 – June 2009)

**Market and Sector Performance during the DOT-COM recession:**

|  |  |
| --- | --- |
|  |  |

**Data frame showing recovery information for market and sectors:**

Table

Description automatically generated

**Bar charts showing sector QRecs during DOT-COM Recession:**

|  |  |
| --- | --- |
|  |  |

**9 - ANALYSIS –** DOT-COM (continued)

**Daily Return during the DOT-COM recession:**

**(By Sector)**

|  |  |
| --- | --- |
|  |  |

**10, 20 and 50 day Moving Averages for each sector during the GFC recession.**

Chart

Description automatically generated

**Correlation heat-map between sectors during the GFC recession:**

Chart

Description automatically generated

**9 - ANALYSIS –** DOT-COM (continued)

**Insights:**

* All sectors fell in value. The Utilities and Technology sectors fell the most, and the Materials sector the least.
* The Materials sector took only 6 days to recover its losses from this recession. It’s QRec score of 0.02 suggests it recovers in 2% of the time the recession lasted.
* The market took 5.55 times the length of the recession to recover. This Market QRec score is the highest of the 3 recessions analysed, by some distance. (GFC: 2.42 and COVID-19: 1.11)
* 5 of the 10 sectors were nearly unaffected by this recession, with Communications, Technology Utilities, Health Care and Industrials taking the brunt of the downturn.
* The Communications sector recorded a QRec score of 27.93, meaning that it took 6872 days to recover. Relative to other sectors in the same recession, no sector has taken a bigger hit. The second highest QRec score for the Dot-Com recession belonged to Technology at 8.45. This score represents only 30% of the Communications QRec score.

In the GFC the second highest QRec score was 63% of the highest.

In the COVID 19 recession the second highest QRec score was 82% of the highest.

* Need a moving averages insight.

The correlations in sector behaviour throughout a recession have become progressively stronger over time. This is evidenced by the change in colour scheme observed in the respective heat maps. The colours within the plots are much cooler during the Dot-Com recession as compared to the most recent COVID-19 recession. This can possibly be explained by the increased interconnectivity of the sectors in their daily operations. Take Apple as an example. If demand is high for their products then all of the companies in Apple’s supply chain will have improved business outcomes as a result. Their stock price will, in some part, be linked to Apple’s stock price. Many of these businesses will be in a different sector to Apple and so there will exist a link between sectors, too.

*Please note that the Real Estate sector data is not available for the Dot-Com recession or any other recession prior.*

**10 - ANALYSIS –** ALL 3 Recessions (continued)

**QRec scores for each sector in all three recessions:**

Chart, bar chart, box and whisker chart

Description automatically generated

**Average QRec scores for each sector across all three recessions:**

Chart

Description automatically generated

**Insights:**

* As we can throughout the past three recessions, Consumer related sectors recovered at the fastest rate, followed by the Materials sector. That means a recession has the capacity to bolster the strength of these two sectors relative to the other nine.
* The Communications sector has been the slowest to recover in the past three recessions, primarily due to being hit so hard, relative to the rest of the market, throughout the Dot-Com recession. It might also be explained by the fact that consumers prioritise the purchase of food and essential services during recessions and spend less of their income on products and services supplied by other sectors.

**11 - ANALYSIS –** Post COVID-19 Recession (A deeper dive)

As discussed in section 3 of this report, the market and all eleven of its sectors have recovered from the COVID-19 recession. Each has hit record heights in terms of price in recent months.

We will now investigate this ‘bounce-back’ in greater detail. The following charts are based on stock market data from the end of the recession until 27/3/2022.

|  |  |
| --- | --- |
| **Adjusted S&P500 Close Price since the end of the COVID-19 recession:**  Chart, line chart  Description automatically generated | **Revenue Growth by Sector since the end of the COVID-19 recession:**  Chart, pie chart  Description automatically generated |
| **Weight by Sector since the end of the COVID-19 recession:**  Chart, pie chart  Description automatically generated |  |

Would be great to have a Weight by Sector for the day after the recession to compare this one to.

**11 - ANALYSIS –** Post COVID-19 Recession (continued)

**Insights:**

* We can see from the Adjusted S&P500 Close Pricechart above that the S&P 500 market has increased in overall value by more than 90%.
* All sectors have experienced positive growth, led by Consumer (discretionary) with 34.28% growth since the end of the recession. It is likely that Tesla has contributed significantly to that result, growing a whopping 12% in the 9 days leading up to the 27th of March, 2022.
* Consumer (discretionary) is a sector classified as providing non-essential goods and services. Consumers tend to spend more on Consumer (discretionary) products in economic growth phases, when individuals have more disposable income.
* The Technology sector makes up almost a quarter (23.68%) of the S&P 500 total weighting.
* The largest U.S. companies, based on Market Capital, are Apple, Microsoft, Alphabet, Amazon and Facebook. They currently account for 17.5% of the S&P 500. That means that anyone who invests in the SPDR S&P 500 ETF is simultaneously investing heavily, and perhaps unknowingly, in these companies too.

**12 - ANALYSIS – EXPLORING POSSIBLE CORRELATIONS**

**Potential Correlation 1:**

Are recessions becoming progressively shorter in duration?

To investigate this potential correlation, we constructed a bar chat showing the duration of each recession since the beginning of the 1980s. The x-axis has been ordered in such a way that the recessions appear in chronological order.

Chart, bar chart

Description automatically generated

**Insights:**

* There is no apparent trend in this data. It is unlikely that a correlation exists between the recency of a recession and its length.

**Potential Correlation 2:**

Is there a correlation between GDP decline and Unemployment?

Could we possibly reverse the order of the x axis values?

Chart, waterfall chart

Description automatically generated

Could we possibly order the blue lines in ascending height order?

This will better show the decline in orange lines.

**Insights:**

* Based on the data associated with the past 6 recessions, there appears to be a correlation between GDP declines and unemployment. As GDP plumets, unemployment rates rise.

**12 - ANALYSIS – EXPLORING POSSIBLE CORRELATIONS**

**Potential Correlation 3:**

Is there a correlation between GDP decline in a recession and the time since the last recession?

Chart, scatter chart

Description automatically generated

**Insights:**

* No correlation.
* What is the r-squared value?

**Potential Correlation 4:**

Is there a correlation between GDP growth in a calendar year and S&P500 returns?

Chart, scatter chart

Description automatically generated

**Insights:**

* No correlation.
* What is the r-squared value?

**2.3 Daily Return and Daily Volume can help predict the trend of the market.**

Daily Return and Daily Volume are closely related courtesy of the way they are calculated.

From a market perspective, Trading Volume can help investors identify market trends.

If the Trade Volume keeps moving on one way, the price usually does not change its direction. In other words, if the volume trend remains unchanged (continuously increasing or decreasing), the price trend does not change. At this point, it is safe to keep the strategy you have.

If the volume trend changes, from increasing to decreasing (or a decrease becomes an increase), which means that the price is facing a reversal point. At this point, you may need to consider changing your strategy.

**2.4 Limitation**

However, in theory, the above statement is correct. But in the real world, this theory might not work due to various reasons. For example, volume changes every second, and we can only know the change of trend after it crosses the extreme value. Therefore, the relationship between Volume and Return is not enough to help investors make a clear judgment. But there is indeed a close correlation between them, and volume can reflect investors' confidence and interest in the market.

**3.2 The moving average (MA) can help predict the trend**

Crossovers are one of the main moving average strategies.

First, we can judge by the crossover on the closing price and the moving average line. When the price crosses above or below a moving average, there will be a signal for change in trend.

Chart, map

Description automatically generated

Second, we can judge by the crossover on the short-term and long-term moving average lines. When the shorter-term MA crosses above the longer-term MA, it's a buy signal, as it indicates that the trend is shifting up. This is known as a golden cross. Meanwhile, when the shorter-term MA crosses below the longer-term MA, it's a sell signal, as it indicates that the trend is shifting down.

Chart, line chart

Description automatically generated

**3.3 Limitation**

Moving averages are calculated based on historical data and nothing about the calculation is predictive in the real world. if the price becomes choppy, the price may swing back and forth, generating multiple trend reversals or trade signals. When this occurs, we need another indicator to help clarify the trend.

If prices start fluctuating, sometimes the market won't respect MA support/resistance trade signals. Moving averages work well in strong trending conditions but poorly in ranging conditions.

Chart, pie chart

Description automatically generated