12/1/2023

Aditya Nikumbh (Student)

University of dundee

**Stock Exchange Data Analysis**

Effect of correlation between Daily Price Volatility and Traded Volume on Market Dynamics

## **Dataset Selection**

The stock indexes dataset containing data from 1965 to 2021 was selected. This dataset comprises information related to various indexes from different countries, including date, daily opening price, closing price, volume traded, closing price adjustments, and closing price in USD. While only the closing price has been converted to USD, the remaining values are in local currencies of the respective indices. The dataset was obtained from Kaggle and contains additional information about the indexes in a separate file. The table below shows the indexes included in the dataset.

(*https://www.kaggle.com/datasets/mattiuzc/stock-exchange-data*)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sr. no. | Region | Exchange | Index | Currency |
| 1. | United States | New York Stock Exchange | NYA | USD |
| 2. | United States | NASDAQ | IXIC | USD |
| 3. | Hong Kong | Hong Kong Stock Exchange | HSI | HKD |
| 4. | China | Shanghai Stock Exchange | 000001.SS | CNY |
| 5. | Japan | Tokyo Stock Exchange | N225 | JPY |
| 6. | Europe | Euronext | N100 | EUR |
| 7. | China | Shenzhen Stock Exchange | 399001.SZ | CNY |
| 8. | Canada | Toronto Stock Exchange | GSPTSE | CAD |
| 9. | India | National Stock Exchange of India | NSEI | INR |
| 10. | Germany | Frankfurt Stock Exchange | GDAXI | EUR |
| 11. | Korea | Korea Exchange | KS11 | KRW |
| 12. | Switzerland | SIX Swiss Exchange | SSMI | CHF |
| 13. | Taiwan | Taiwan Stock Exchange | TWII | TWD |
| 14. | South Africa | Johannesburg Stock Exchange | J203.JO | ZAR |

As observed from the table, there are fourteen indexes for which the data is recorded in the stock indexes dataset. For the USA alone, two indexes are recorded. This data was collected from Yahoo Finance.

**Research Question:** "How does the daily price volatility interact with trading volume across various indices in the dataset, and is there a noticeable correlation between these key factors influencing market dynamics?"

The dispersion of returns for a security or market index can be measured statistically through volatility. Generally, higher volatility indicates a greater risk (Carr, 2022) for that particular security. The standard deviation or variance between returns from that same security or market index is typically used to calculate volatility. But apart from these two even, range can be applied to measure it (Hayes, 2023).

This study employs the range measure of volatility on daily price to study the relationship between the Daily Price Volatility of Indexes and the volume traded, which is then used to understand the market dynamics.

## **Exploratory Data Analysis**

The dataset was imported, transformed, and explored using Python code written in the Jupyter Notebook and executed in the Anaconda environment. Scatterplots, correlation heatmaps, and time series plots were utilised to analyse the research question.

After importing the dataset and storing it in a data frame, the first five rows were loaded to check it.

A table with numbers and letters

Description automatically generated

**Fig:1** Original dataset for Stock Exchange

In the above data, as we can see, only the Closing price of the Index has been mentioned in US dollars, and all the price columns, i.e., Open, High, Low and Close, are in Local currencies of their Index. So before proceeding further with data exploration, we will change all the necessary columns to USD. Below is the updated data frame.

A table with numbers and letters

Description automatically generated

**Fig:2** Updated Dataset for Stock Exchange

Next, the data frame information was checked to determine the structure of the data set and get an overview. Below are the images of different kinds of information about the dataset.

A screenshot of a computer code

Description automatically generated

**Fig:3.** Dataset information regarding length, non-null count, and data type

As we can observe from the above image, the dataset has 104,224 rows. The index and date columns have an object data type; the rest are float64 data types. We converted the Date column to the datetime64[ns] type to ensure the dataset can be sorted date-wise.

After rechecking the information, we can see the Date column data type has been updated.

A screenshot of a computer

Description automatically generated

Fig:4 Update Date Column

Next, the dataset summary was checked to see the statistical information about the data using the describe function.

A table with numbers and letters

Description automatically generated

Fig:5 Statistical Summary of the data

The above image shows the count, mean, standard deviation, minimum values, maximum values, and quartiles for all the numerical columns of the dataset. Next, as we were interested in the trading volume, we checked the statistical summary of the volume after grouping them by Index.

A group of colorful text

Description automatically generated with medium confidence

**Fig:6** Volume Statistics by Index

The one interesting observation from the above data is that the Index J203.JO has no volume being traded in the whole dataset.

Next, we created a new column of daily volatility by calculating the difference between the High\_USD column and the Low\_USD column. Then, we checked its statistical data using the describe function.

A table with numbers and symbols

Description automatically generated

**Fig: 7** Daily Volatility by Index

Upon observing the index J203.JO again shows a change in its price, meaning it has daily price volatility.

## **Data Visualisation**

To answer the research question, we have utilised scatter plots, correlation heatmaps and time-series plots to understand the market sentiment.

### **Scatter Plot**

A chart of a trading volume

Description automatically generated with medium confidence

**Fig: 8** Scatter Plot of trading volume vs daily price volatility by index

#### Design Justification

A scatterplot is beneficial when observing the relationship between two variables. The colour palette used was Viridis, which makes it easily accessible to people with colour blindness. The plot style is white background because it draws attention towards the cluster rather than the background itself.

#### Interpretation & Analysis 1

As my research question is based on comparing Daily Price Volatility with volume to understand market dynamics, I have employed a scatterplot to know how the two behave for different indices in the stock market.

1. For zero trade volume:

As observed in the above plot, many indices on the legend have volatility in their price when zero volume is traded. One of the data points is seen to have crossed the 1400 mark. This might be due to several reasons. A few of them are as follows:

* Bid-Ask price spread: The bid-ask price spread means the minimum price the seller is willing to sell and the highest price the bidder is willing to pay for a stock. In this case, there is no liquidity.
* News and Events: The news and events in the market can affect the stock prices of an index even when no actual trades occur. As we have not included the time component here, it is difficult to analyse the exact reason for the same.
* The most probable reason for this might be our earlier observation in the case of J203.JO index, which has daily volatility for zero volume traded.

1. For trade volume greater than zero:

In this case, two observations can be made. These observations are based on the two types of shade observed.

* The bluish indices are observed to form a cluster with a positive relationship between daily volatility and volume. There is a higher change in daily price volatility for these indices when there is a slight change in trading volume. This means that there is some liquidity among these indexes.
* The greenish indices also show a positive relationship between daily price volatility and volume, except that in these indices, there is a slight change in daily volatility for significant change in trading volume. These often indicate that there is high liquidity in these stock indexes.

### **Correlation Heatmaps**

A graph with colored squares

Description automatically generated

**Fig: 9** Correlation heatmap of Volume vs Daily Volatility by Index

#### Design Justification

A correlation heatmap was chosen because not only does it identify the correlation between Daily Price Volatility and Traded Volume, but unlike a scatter plot, it also quantitatively determines the strength of the relationship between them. Again, the colour scheme was set to Viridis as it not only helps people with colour blindness to read it but also highlights the most vital relationship with the attention-catching bright colours. Also, the background was set to be white to focus the attention towards the heatmap itself.

#### Interpretation & Analysis

The correlation analysis showed that the index NYA had the highest correlation value of 0.82, followed by IXIC and GSPTSE, with weights of 0.73 and 0.71, respectively. Two of the top three indexes with high correlation values, NYA and IXIC, are from the USA. This is primarily because the USA had multiple indexes in the dataset. The correlation weights were mainly influenced by significant events, such as the dot-com bubble burst between 2001 and 2002, the Great Recession of 2008, and the COVID-19 pandemic in 2020.

Between 1995 and 2000, the NASDAQ index of the US increased from under 1000 to 5000 due to the rise in investment in technology-based companies. This led to a surge in stock prices and volumes, making the index highly volatile. However, the global markets plummeted when the bubble burst, affecting many investors.

This was followed by the Great Recession of 2008, triggered by the US housing market crisis and the COVID-19 pandemic in 2019. These events contribute to the correlation values observed in the NYA, IXIC, and GSPTSE indices. We will further explore this relationship with the help of time series plots for the overall daily price volatility by the top three indexes.

### **Time Series Plots**

A graph showing a number of times

Description automatically generated with medium confidence

**Fig: 10** Time series plot of Daily Price Volatility for NYA

#### A graph of a graph showing the price of a stock market Description automatically generated

**Fig: 11** Time series plot of Daily Price Volatility for IXIC

A graph of a price

Description automatically generated

**Fig: 12** Time series plot of Daily Price Volatility for the GSPTSE Index

#### Design Justification

#### We plotted daily price volatility for the top three indexes using individual time series plots. The graphs had a white background with no borders to draw attention to the data. We included data from 1990 to show events that caused fluctuations in the market.

#### Interpretation & Analysis

NYA index had low volatility until 2008, but it impacted the global market during the Great Recession. The highest volatility of 1292.75 was observed in 2020 due to the pandemic. Upon observation, it appears that IXIC experienced its highest spike in the year 2000 due to the dot-com bubble burst. However, during the Great Recession of 2008 and the COVID-19 pandemic, its spike was less significant than that of NYA. Notably, IXIC is the NASDAQ index and was the primary contributor to the global market crash. GSPTSE index had a spike during the dotcom burst but reached its highest point during the Great Recession with 1332.16 in a day. In COVID-19, it showed high volatility but was still lower than the NYA index.

## **Conclusion**

Based on the research question, it can be concluded that there is a noticeable correlation between daily price volatility and trading volume in the stock market. The scatter plot shows that most indexes positively correlate for the two factors. However, some indexes have a weak or negative correlation, indicating that other factors might influence market dynamics. The correlation heatmap confirms the positive correlation between daily price volatility and trading volume. Additionally, the time series plot shows how the market sentiment changed during different crises in the market. Overall, this study highlights the importance of understanding the relationship between daily price volatility and trading volume to understand how it impacts the market dynamics.

However, there are certain limitations of the dataset that limit the study. That is no volume for the J203.JO index. Another limitation of the study is that it has not calculated the Daily Volatility based on standard deviation or variance, which would have proven an accurate measure. This study can also be further extended to use machine learning models to predict future volatility.

### **References**

Carr, M. (2022) Measure volatility with average true range, Investopedia. Available at: https://www.investopedia.com/articles/trading/08/average-true-range.asp#:~:text=A%20stock’s%20range%20is%20the,small%20ranges%20indicate%20low%20volatility. (Accessed: 01 December 2023).

Hayes, A. (2023) Volatility: Meaning in finance and how it works with stocks, Investopedia. Available at: https://www.investopedia.com/terms/v/volatility.asp#toc-understanding-volatility (Accessed: 01 December 2023).

Inquiry Commission, F.C. (2011) ‘The Financial Crisis Inquiry Report: Final Report of the National Commission on the causes of the financial and economic crisis in the United States’, Choice Reviews Online, 48(12). doi:10.5860/choice.48-7034.