Impact of Covid-19 on Six Major Stock Markets and U.S. Sectors

Beatrice Spinola, Rajesh Golla, Suryasivaji Killi, Nihaal Sayala

School of Business

Stevens Institute of Technology

May 14th, 2023

Abstract

This article examines the impact of COVID-19 on the stock markets of six countries with the highest number of confirmed cases. The study employs event study methodology to analyze the most volatile periods during the pandemic and the abnormal returns generated by the markets in relation to the daily rise in cases. Results show that Brazilian indices experienced the highest decline of more than 50%, while Mexican indices saw the lowest fall of around 30%. Additionally, the paper investigates the volatility transmission effects between the US stock market and COVID-19 using a BEKK-multivariate GARCH model. Findings reveal that the US stock market's volatility is positively affected by the death rate but negatively impacted by the recovered rate. Furthermore, the bad news has a more significant impact on the current US stock market than good news, and there are volatility spillover effects. Overall, the study highlights the adverse effects of COVID-19 on stock markets and provides insights into the volatility transmission effects of the virus on the US stock market.

Keywords: Stock Markets, Volatility, BEKK-multivariate GARCH model, Systematic Risk, Principal component factors, Asymmetric volatility, Abnormal returns.

1. Introduction

Globally, COVID-19 cases have crossed the 100 million mark with more than 2 million deaths.

This is one of the deadliest pandemics in recent times, which not only caused loss of human lives but also led to billions of dollars of loss of world economies. Many economies saw historic contractions and disruptions in financial and labor markets. Investors all around the world initially shunned stocks with China's exposure, and with the spread of disease, markets weighed the economic consequences of these crises on firms. The first half of the year 2020 saw one of the most dramatic stock market crashes in history.

This study aims to analyze the effect of covid-19 on six major indices from the most affected countries by the virus. These countries include the Usa, India, Brazil, Mexico, Russia, and Spain.COVID-19 increased volatility and led to panic trading in many major indices.US S&P 500 was halted many times during March 2020 when it fell by 7% or more. Similarly, the Indian indices NIFTY 50 and SENSEX were halted twice in 15 days in March 2020 when the market fell by 10%. Therefore, this study tries to document the overall market response by applying event study methodology in these six affected countries.

We have also selected the 10 Sector ETFs in the Usa, above are the names and symbols of the sectors

Name	Symbol
Energy Sector	XLE
Financial Sector	XLF
Industrial Sector	XLI
Technology Sector	XLK
Health Care Sector	XLV
Materials Sector	XLB
Utilities Sector	XLU
Real Estate Sector	XLRE
Consumer Discretionary Sector	XLY
Consumer Staples Sector	XLP

Based on the change in annualized mean pre and post Covid-19, we divided the above sectors into three groups:

- High Change Sectors: XLE, XLK, XLY

- Mid Change Sectors: XLF, XLB, XLRE, XLU

- Low Change Sectors: XLI, XLV, XLP

2. Literature Review

The global spread of COVID-19 has had a significant impact on the financial markets in almost every country. Its spread created havoc in the market, and risk-averse investors incurred huge financial losses in a short period of time. Similarly, to the research paper "Impact Of Covid-19 Outbreak On The Stock Market: An Evidence From Select Economies" by Irfan Rashid Ganie, Tahir Ahmad Wani, and Miklesh Prasad Yadav, this paper aims to analyze the impact of COVID-19 on stock markets in the top six affected countries based on the total number of cases confirmed. In addition, we'll also analyze the stock market volatility caused by the virus and the abnormal returns generated by the markets during the pandemic. And divide our dataset into pre and post Covid-19 to perform different types of regressions.

The paper found evidence of significant risk transmission from the COVID-19 pandemic to the US stock market. Specifically, they found that shocks to the number of COVID-19 cases and deaths were associated with negative stock returns. The authors noted that the degree of risk transmission varied across different sectors of the economy. For example, the healthcare sector tended to benefit from the pandemic, while the energy and financial sectors were more negatively affected. The paper also found that the transmission of risk was not limited to domestic factors. In particular, the authors noted that shocks to the number of COVID-19 cases in other countries were also associated with negative stock returns in the US. The authors suggested that the findings of the paper could have implications for policymakers and investors. For example, they argued that policymakers should be aware of the potential for risk transmission from health-related events and take steps to mitigate this risk. Similarly, investors may want to consider the potential impact of health-related events on their portfolios.

"Has COVID –19 Infected Indian Stock Market? By Dippi Verma, Praveen Kumar Sinha and Lakshmi S. R , this Research Paper investigates the effect of Covid-19 on Indian stock market returns, it begins with the calculation of return with the time series data of NIFTY50. This paper has empirically investigated the existence of day-of-the-week effect by using closing daily data for Nifty 50, Nifty 50 Midcap, Nifty 100, Nifty100 midcap, Nifty 200, Nifty 100 Small cap. This study used secondary data for all indices over the period 1 April 2005 to 14 May 2020. The impact of Covid-19 cases and deaths is estimated by extending the common GARCH(1,1)model

They concluded that if the volatility of stock markets rises with the increase in total number of COVID-19 cases, then it may affect the investment adversely.

COVID-19’S ADVERSE EFFECTS ON A STOCK MARKET INDEX. Author(s): Kang Hua Caoa, Qiqi Lib, Yun Liuc and Chi-Keung Woob. Journal: Applied Economics Letters, Year: 2021. This article’s goal is to perform a panel data analysis of 14 daily stock market indices during 01/21/2020 – 06/30/2020 to document a stock market index’s negative responsiveness to Covid-19’s spread variations. Double-log Regression with random errors and Stock market index's elasticity is -0.029(p-value < 0.01). The article gives us insights into how the stock market indices were affected and how the markets will recover with time.

3. Data Exploration

For the purposes of this project, we’ve selected the following six countries based on the total number of confirmed cases. Below are the Country Name and R-Symbol chosen for the project:

Country Name	Index
USA	SPY
India	NSEI
Brazil	BVSP
Russia	IMOEX.ME
Spain	IBEX
Mexico	MXX

Below are the first 10 rows of the dataset we used for modeling:

Date	SPY	NSEI	BVSP	IMOEX.ME	IBEX	MXX	XLE	XLF	XLI	XLK	XLV	XLB	XLU	XLRE	XLY	XLP	post_covid
1/10/2017	0	0.00636	0.006977	0.011796731	-0.00431	0.007278	-0.00913	0.002136	0.004141	0.000202	0.003503	0.000593	-0.00312	-0.01258	0.003597	-0.00504	FALSE
1/11/2017	0.002822	0.011044	0.005041	-0.00847378	-0.0046	0.001032	0.010737	0.006382	0.005232	0.004645	-0.01012	0.008064	0.010354	-0.00553	0.002152	0.001942	FALSE
1/12/2017	-0.00251	0.003163	0.023862	-0.00298384	-0.00013	0.002768	-0.00415	-0.00852	-0.00412	-0.00242	0.000988	-0.00196	0.000824	0.003909	-0.00096	-0.00097	FALSE
1/13/2017	0.002293	-0.00082	-0.00473	-0.00762851	0.011015	0.002633	-0.00295	0.005545	0.003804	0.003025	0.000988	-0.00157	-0.00144	-0.00228	0.003699	0.000583	FALSE
1/17/2017	-0.00353	-0.00028	0.010968	-0.00752645	-0.01235	-0.0039	0.006165	-0.02411	-0.00778	-0.00383	-0.00509	-0.00453	0.010866	0.008113	0.001666	0.01407	FALSE
1/18/2017	0.002208	0.00226	-0.00318	-0.00397345	-0.00093	0.007754	-0.00227	0.008245	0.003819	0.002624	-0.00128	0.006299	-0.00122	0.00226	-0.00179	0.003439	FALSE
1/19/2017	-0.00371	0.002148	-0.00311	-0.00361934	-0.00076	-0.00206	-0.00551	-0.00433	0.007279	-0.00141	-0.00641	-0.0063	-0.00882	-0.00972	-0.0031	-0.00363	FALSE
1/20/2017	0.003667	-0.01022	0.008874	-0.00105966	0.000107	0.001433	0.003763	0.004763	0.000157	0.005436	-0.00257	0.008847	0.001852	0.006168	0.002864	0.006487	FALSE
1/23/2017	-0.00261	0.005036	0.018854	-0.00644206	-0.00806	0.016793	-0.01065	-0.0065	-0.00553	0.001805	-0.00416	0.001955	-0.00515	0.00613	0.000714	-0.00038	FALSE

The first column is the date column; the next 6 columns are the log returns of SPY and the five stock market indices; the following 10 columns are the log returns of the ten sector ETFs in the United States; and the post_covid – represents a binary variable; TRUE for rows after 01/10/2020 and FALSE for rows before 01/10/2020 (the start of Covid-19).

We've also selected the ten sector ETFs in the United States. These are the names and symbols of the sectors:

Name	Symbol
Energy Sector	XLE
Financial Sector	XLF
Industrial Sector	XLI
Technology Sector	XLK
Health Care Sector	XLV
Materials Sector	XLB
Utilities Sector	XLU
Real Estate Sector	XLRE
Consumer Discretionary Sector	XLY
Consumer Staples Sector	XLP

Summary of model data over 6 years of historical data:

```
      SPY                  NSEI                 BVSP               IMOEX.ME                IBEX                  MXX                  XLE
Min.   :-0.1005688   Min.   :-0.1390376   Min.   :-0.1599303   Min.   :-0.4046744   Min.   :-1.515e-01   Min.   :-0.0677247   Min.   :-0.2328968
1st Qu.:-0.0040007   1st Qu.:-0.0048051   1st Qu.:-0.0074589   1st Qu.:-0.0056578   1st Qu.:-6.527e-03   1st Qu.:-0.0058011   1st Qu.:-0.0094897
Median : 0.0007964   Median : 0.0012402   Median : 0.0009230   Median : 0.0008221   Median : 1.209e-04   Median :-0.0000870   Median : 0.0003344
Mean   : 0.0004986   Mean   : 0.0006109   Mean   : 0.0004431   Mean   :-0.0000169   Mean   :-6.811e-05   Mean   : 0.0001068   Mean   : 0.0003365
3rd Qu.: 0.0062053   3rd Qu.: 0.0069120   3rd Qu.: 0.0095548   3rd Qu.: 0.0072004   3rd Qu.: 6.568e-03   3rd Qu.: 0.0061606   3rd Qu.: 0.0104743
Max.   : 0.0867310   Max.   : 0.0840030   Max.   : 0.1302228   Max.   : 0.1826195   Max.   : 8.225e-02   Max.   : 0.0604372   Max.   : 0.1487417
      XLF                 XLI                  XLK                  XLV                  XLB                  XLU                  XLRE
Min.   :-0.1144501   Min.   :-0.109258   Min.   :-0.1027027   Min.   :-0.0775900   Min.   :-0.1071682   Min.   :-0.1064023   Min.   :-0.1154493
1st Qu.:-0.0064145   1st Qu.:-0.005565   1st Qu.:-0.0058214   1st Qu.:-0.0045976   1st Qu.:-0.0069416   1st Qu.:-0.0053587   1st Qu.:-0.0057653
Median : 0.0004163   Median : 0.001052   Median : 0.0012483   Median : 0.0008797   Median : 0.0005773   Median : 0.0009104   Median : 0.0008376
Mean   : 0.0004108   Mean   : 0.000451   Mean   : 0.0007839   Mean   : 0.0005648   Mean   : 0.0004576   Mean   : 0.0004554   Mean   : 0.0003045
3rd Qu.: 0.0078519   3rd Qu.: 0.007084   3rd Qu.: 0.0086742   3rd Qu.: 0.0063931   3rd Qu.: 0.0079709   3rd Qu.: 0.0066733   3rd Qu.: 0.0069580
Max.   : 0.1236025   Max.   : 0.119126   Max.   : 0.1109322   Max.   : 0.0742322   Max.   : 0.1111846   Max.   : 0.1021057   Max.   : 0.0841577
      XLY                 XLP               post_covid
Min.   :-0.1065328   Min.   :-0.0986678   Mode :logical
1st Qu.:-0.0053240   1st Qu.:-0.0035892   FALSE:640
Median : 0.0013233   Median : 0.0005828   TRUE :649
Mean   : 0.0004129   Mean   : 0.0004143
3rd Qu.: 0.0071971   3rd Qu.: 0.0053543
Max.   : 0.0947100   Max.   : 0.0816781
```

From the image above, we observe that there are 640 "FALSE" (pre-covid) variables and 649 "TRUE" (post-covid) variables in our dataset.

Minimum (Min.): The minimum is the smallest value in a dataset. It represents the lower bound of the distribution and can be used to identify outliers that fall below this value.

Maximum (Max.): The maximum is the largest value in a dataset. It represents the upper bound of the distribution and can be used to identify outliers that fall above this value.

Median: The median is the middle value of a dataset. It represents the point at which half of the values are above and half are below. The median is a measure of central tendency that is less affected by extreme values or outliers than the mean.

Mean: The mean is the arithmetic average of a dataset. It is calculated by adding up all the values in the dataset and dividing them by the number of values. The mean is a measure of central tendency that is sensitive to extreme values or outliers.

Standard Deviation: The standard deviation is a measure of the spread or dispersion of a dataset. It is calculated by taking the square root of the variance, which is the average of the squared differences from the mean. The standard deviation provides a measure of how much the values in the dataset deviate from the mean.

Quartiles: Quartiles are values that divide a dataset into quarters or four equal parts. The first quartile, or Q1, represents the value below which 25% of the values fall. The second quartile, or Q2, is the same as the median, representing the value below which 50% of the values fall. The third quartile, or Q3, represents the value below which 75% of the values fall. Quartiles can be used to identify the range of values that contain most of the observations and to identify outliers or extreme values.

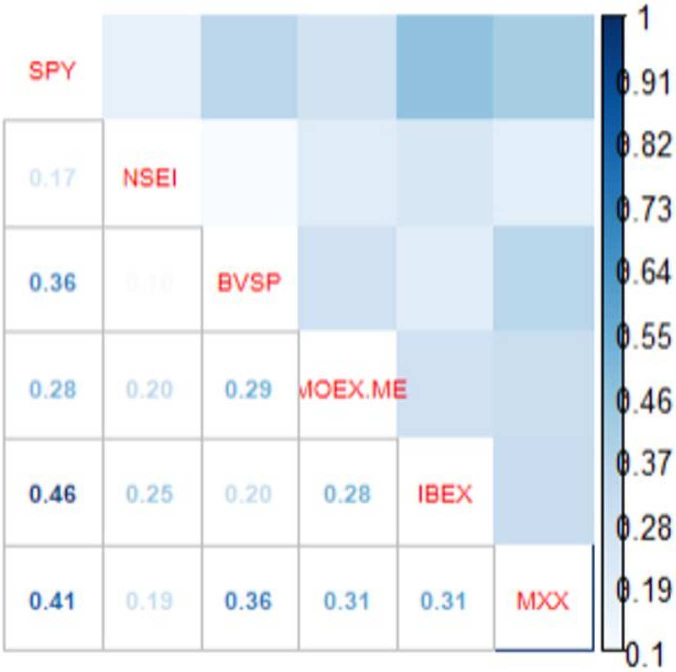
After creating box plots from the summary model data, we found that out of the six stock markets analyzed, Russia had the highest whiskers in comparison to all other countries. On the other hand, Mexico had the smallest whiskers.

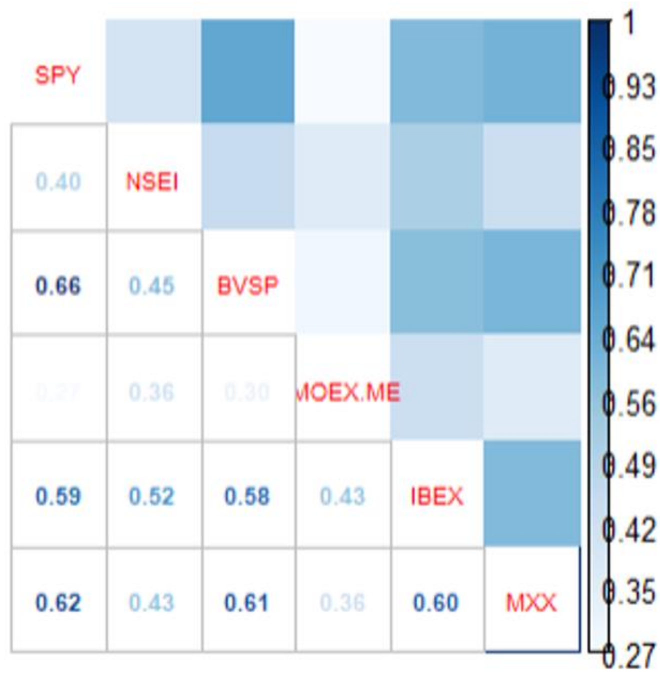
When analyzing the box plots for the pre-COVID period, Brazil had the highest whiskers, while the USA indexes had the lowest whiskers. However, for the post-COVID period, Russia had the highest whiskers according to the box plot analysis, while Mexico had the least whiskers.

Upon analyzing the Sector ETF's box plot for the entire six-year period, we observed that the Energy sector had the highest whiskers while the Health Care sector had the lowest whiskers.

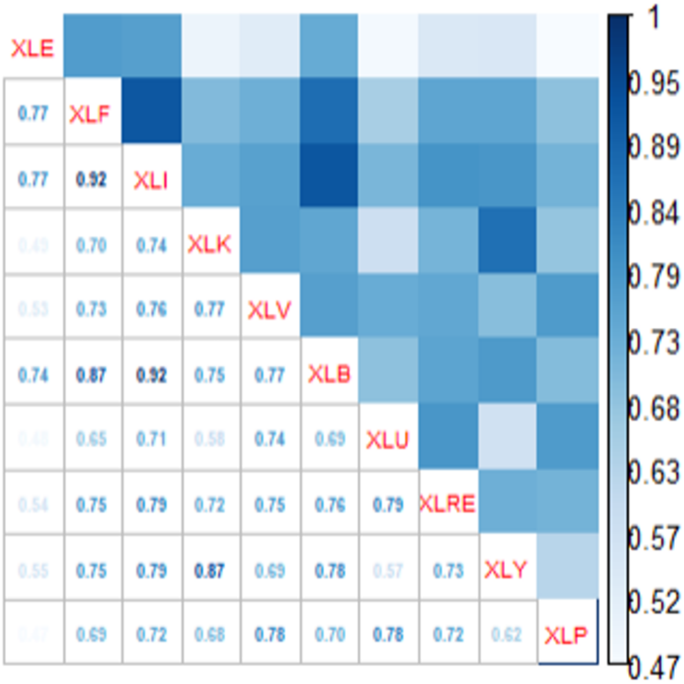
Further, when examining the box plot for the pre-COVID period, we found that the Technology sector in the USA had the highest whiskers, whereas the Utility sector had the lowest whiskers. However, for the post-COVID period in the USA, the Energy sector had the highest whiskers, while the health sector had the least whiskers.

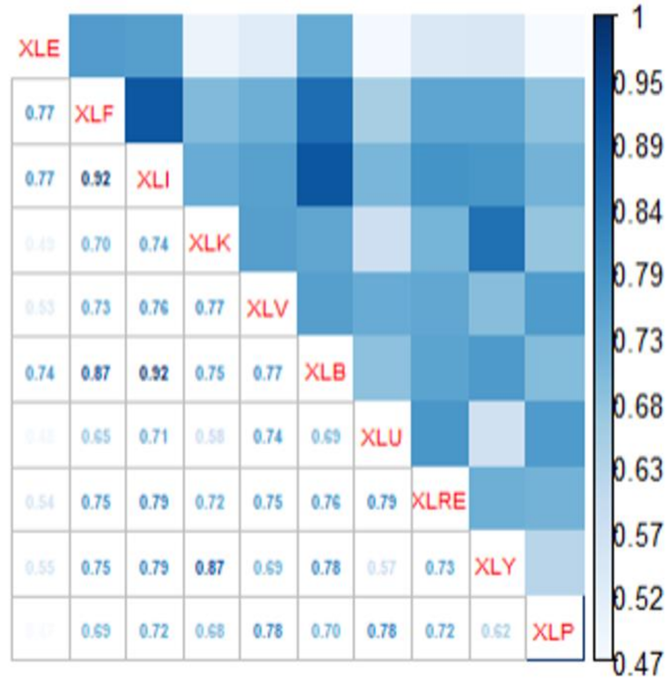
We performed Correlation matrices. It is a representation of the correlation coefficients between multiple variables in a dataset. The correlation coefficient is a measure of the strength and direction of the relationship between two variables. Below are the charts for pre-covid and post-covid stock market indexes:





We then also performed correlation charts for ten sector ETFs for pre-covid and post-covid





Correlation Change:

It indicates how the strength or direction of the correlation between two variables varies under different circumstances. Correlation change can occur due to various factors, such as changes in underlying relationships, shifts in the distribution or range of values of the variables, or the influence of external factors or events.

Below is the Correlation change for both Stock market indexes for six countries and ten sector ETF's:

	SPY	NSEI	BVSP	IMOEX.ME	IBEX	MXX
SPY	NaN	1	1	-1	1	1
NSEI	1	NaN	1	1	1	1
BVSP	1	1	NaN	1	1	1
IMOEX.ME	-1	1	1	NaN	1	1
IBEX	1	1	1	1	NaN	1
MXX	1	1	1	1	1	NaN

	XLE	XLF	XLI	XLK	XLV	XLB	XLU	XLRE	XLY	XLP
XLE	NaN	1	1	1	1	1	1	1	1	1
XLF	1	NaN	1	1	1	1	1	1	1	1
XLI	1	1	NaN	-1	1	1	1	1	1	1
XLK	1	1	-1	NaN	1	1	1	1	1	1
XLV	1	1	1	1	NaN	1	1	1	1	1
XLB	1	1	1	1	1	NaN	1	1	1	1
XLU	1	1	1	1	1	1	NaN	1	1	1
XLRE	1	1	1	1	1	1	1	NaN	1	1
XLY	1	1	1	1	1	1	1	1	NaN	1
XLP	1	1	1	1	1	1	1	1	1	NaN

This matrix shows if the correlation increased or decreased (-1 indicates a decrease and +1 indicates an increase). For the stock markets, the correlation decreased between IMOEX.ME (Russia) and SPY U.S.A). And the correlation decreased also among sector ETFs, XLK (Technology sector), and XLI (Industrial sector).

Annualized Mean Returns:

Below are the Annualized mean returns for pre covid and post covid for both six stock market indexes and ten sector ETFs:

	SPY	NSEI	BVSP	IMOEX.ME	IBEX	MXX	XLE	XLF	XLI	XLK	XLV	XLB	XLU	XLRE	XLY	XLP
Pre	0.166616	0.155223	0.248397	0.1353159	0.003674	-0.00858	-0.03548	0.13176	0.135425	0.27257	0.165339	0.088888	0.151485	0.119979	0.181695	0.108702
Post	0.085266	0.152691	-0.02318	-0.141916	-0.03771	0.061934	0.203401	0.075675	0.092204	0.123566	0.119647	0.141401	0.078543	0.034097	0.027502	0.100161
Change	-0.08135	-0.00253	-0.27158	-0.277232	-0.04139	0.070509	0.238883	-0.05608	-0.04322	-0.149	-0.04569	0.052513	-0.07294	-0.08588	-0.15419	-0.00854

Annualized Mean Volatility:

Below is the Annualized mean volatility for pre covid and post covid for both six stock market indexes and ten sector ETFs:

	SPY	NSEI	BVSP	IMOEX.ME	IBEX	MXX	XLE	XLF	XLI	XLK	XLV	XLB	XLU	XLRE	XLY	XLP
Pre	0.130367	0.131083	0.20802	0.152199	0.143449	0.144572	0.189899	0.169036	0.15871	0.184475	0.145083	0.158925	0.12857	0.138895	0.147915	0.119479
Post	0.257083	0.245964	0.322444	0.40348	0.258901	0.203066	0.483572	0.324036	0.289944	0.322985	0.221282	0.296841	0.268281	0.291678	0.312413	0.202064
Change	0.126716	0.114881	0.114423	0.251281	0.115451	0.058494	0.293673	0.155	0.131234	0.13851	0.076199	0.137916	0.139711	0.152783	0.164498	0.082585

The annualized mean return and annualized mean volatility help us to evaluate the risk-return tradeoff of a particular stock. A higher annualized mean return is desirable, but it must be weighed against the risk of higher annualized mean volatility.

4. Logistic Regression, LDA, QDA, and KNN

We are looking at the impact of Covid-19 on six different stock markets. Our plan is to build a model with the six stock markets and the ten U.S. sector ETFs as the explanatory variables, and the post_covid column (a binary variable) as the response variable. Based on the change in annualized mean pre and post Covid-19, we divided the dataset into three groups:

- High Change Sectors: BVSP, IMOEX.ME, XLE, XLK, XLY

- Mid Change Sectors: SPY, MXX, XLF, XLB, XLRE, XLU

- Low Change Sectors: NSEI, IBEX, XLI, XLV, XLP

To perform Linear Regression, LDA, QDA, and KNN we divided our dataset into train and test sets, as shown below:

```
# Train Set
nrow(filter(train_data, post_covid == 'FALSE'))
nrow(filter(train_data, post_covid == 'TRUE'))

# Test Set
nrow(filter(test_data, post_covid == 'FALSE'))
nrow(filter(test_data, post_covid == 'TRUE'))
```

```
[1] 411
[1] 427
[1] 229
[1] 222
```

The results above can be interpreted as the train set is 65% of the rows (411 pre-covid rows, 427 post-covid rows), and the test set is 35% of the rows (229 pre-covid rows, 222 post-covid rows).

We then performed a Logistic Regression as shown below:

```
# Logistic Regression
# High Change Sectors
glm = glm(post_covid~BVSP+IMOEX.ME+XLE+XLK+XLY, data=train_data, family=binomial)
glm.probs=predict(glm, newdata = test_data, type="response")
glm.pred=rep("FALSE",dim(test_data)[1])
glm.pred[glm.probs>0.5]="TRUE"
table(glm.pred,test_data$post_covid)

# Mid Change Sectors
glm = glm(post_covid~SPY+MXX+XLF+XLB+XLRE+XLU, data=train_data, family=binomial)
glm.probs=predict(glm, newdata = test_data, type="response")
glm.pred=rep("FALSE",dim(test_data)[1])
glm.pred[glm.probs>0.5]="TRUE"
table(glm.pred,test_data$post_covid)

# Low Change Sectors
glm = glm(post_covid~NSEI+IBEX+XLI+XLV+XLP, data=train_data, family=binomial)
glm.probs=predict(glm, newdata = test_data, type="response")
glm.pred=rep("FALSE",dim(test_data)[1])
glm.pred[glm.probs>0.5]="TRUE"
table(glm.pred,test_data$post_covid)
```

```
glm.pred FALSE TRUE
   FALSE    98   95
   TRUE    131  127

glm.pred FALSE TRUE
   FALSE    78   79
   TRUE    151  143

glm.pred FALSE TRUE
   FALSE    25   34
   TRUE    204  188
```

Using the stock index as explanatory variables, we observe that Logistic Regression predicted 127/222 rows correctly for High Change indices; 143/222 rows correctly for Mid Change indices; and 188/222 rows correctly for Low Change indices.

Next, we ran a Linear Discriminant Analysis (LDA), as shown below:

```
# LDA
# High Change Sectors
lda.fit=lda(post_covid~BVSP+IMOEX.ME+XLE+XLK+XLY,data=train_data)
table(predict(lda.fit,test_data)$class, test_data$post_covid)
mean(predict(lda.fit,test_data)$class == test_data$post_covid)

# Mid Change Sectors
lda.fit=lda(post_covid~SPY+MXX+XLF+XLB+XLRE+XLU,data=train_data)
table(predict(lda.fit,test_data)$class, test_data$post_covid)
mean(predict(lda.fit,test_data)$class == test_data$post_covid)

# Low Change Sectors
lda.fit=lda(post_covid~NSEI+IBEX+XLI+XLV+XLP,data=train_data)
table(predict(lda.fit,test_data)$class, test_data$post_covid)
mean(predict(lda.fit,test_data)$class == test_data$post_covid)
```

```
        FALSE TRUE
  FALSE    98   95
  TRUE    131  127
[1] 0.4988914

        FALSE TRUE
  FALSE    78   79
  TRUE    151  143
[1] 0.4900222

        FALSE TRUE
  FALSE    25   34
  TRUE    204  188
[1] 0.4722838
```

Again, using the stock index as explanatory variables, LDA predicted 127/222 rows correctly for High Change indices; 143/222 rows correctly for Mid Change indices; and 188/222 rows correctly for Low Change indices. Note that we got the same results using Logistic Regression and LDA.

We then performed Quadratic Discriminant Analysis (QDA) illustrated below:

```
# QDA
# High Change Sectors
qda.fit=qda(post_covid~BVSP+IMOEX.ME+XLE+XLK+XLY, data=train_data)
table(predict(qda.fit,test_data)$class, test_data$post_covid)
mean(predict(qda.fit,test_data)$class == test_data$post_covid)

# Mid Change Sectors
qda.fit=qda(post_covid~SPY+MXX+XLF+XLB+XLRE+XLU, data=train_data)
table(predict(qda.fit,test_data)$class, test_data$post_covid)
mean(predict(qda.fit,test_data)$class == test_data$post_covid)

# Low Change Sectors
qda.fit=qda(post_covid~NSEI+IBEX+XLI+XLV+XLP, data=train_data)
table(predict(qda.fit,test_data)$class, test_data$post_covid)
mean(predict(qda.fit,test_data)$class == test_data$post_covid)
```

```
         FALSE TRUE
  FALSE    208  115
  TRUE      21  107
[1] 0.6984479

         FALSE TRUE
  FALSE    199  121
  TRUE      30  101
[1] 0.6651885

         FALSE TRUE
  FALSE    191  148
  TRUE      38   74
[1] 0.5875831
```

With stock index as explanatory variables, QDA predicted 107/222 rows correctly for High Change indices; 101/222 rows correctly for Mid Change indices; and 74/222 rows correctly for Low Change indices. Overall, QDA yielded worst results than our previous models.

Lastly, we ran K-Nearest Neighbors (KNN), which is shown below:

```
# KNN
# High Change Sectors
test.x = test_data[,c(3, 4, 7, 10, 15)]
training.x = train_data[,c(3, 4, 7, 10, 15)]
knn.pred=knn(training.x, test.x, train_data$post_covid, k=3)
prediction.knn=cbind(test_data, knn.pred)
table(knn.pred, test_data$post_covid)
mean(knn.pred == test_data$post_covid)

# Mid Change Sector
test.x = test_data[,c(1, 6, 8, 12, 13, 14)]
training.x = train_data[,c(1, 6, 8, 12, 13, 14)]
knn.pred=knn(training.x, test.x, train_data$post_covid, k=3)
prediction.knn=cbind(test_data, knn.pred)
table(knn.pred, test_data$post_covid)
mean(knn.pred == test_data$post_covid)

# Low Change Sectors
test.x = test_data[,c(2, 5, 9, 11, 16)]
training.x = train_data[,c(2, 5, 9, 11, 16)]
knn.pred=knn(training.x, test.x, train_data$post_covid, k=3)
prediction.knn=cbind(test_data, knn.pred)
table(knn.pred, test_data$post_covid)
mean(knn.pred == test_data$post_covid)
```

```
knn.pred FALSE TRUE
   FALSE   168  103
   TRUE     61  119
[1] 0.6363636

knn.pred FALSE TRUE
   FALSE   159   93
   TRUE     70  129
[1] 0.6385809

knn.pred FALSE TRUE
   FALSE   130  118
   TRUE     99  104
[1] 0.518847
```

With stock index as explanatory variables, KNN predicted 119/222 rows correctly for High Change indices; 129/222 rows correctly for Mid Change indices; and 104/222 rows correctly for Low Change indices. Overall, KNN performed better than QDA but worse than Logistic Regression and LDA.

5. Forecasting

Time series forecasting is the process of using historical data to make predictions about future events. It is commonly used in fields such as finance, economics, and weather forecasting. R is a powerful programming language and software environment for statistical computing and graphics that is widely used for time series forecasting.

There are several R packages available for time series forecasting, including:

- “forecast”: This package provides a wide range of methods for time series forecasting, including exponential smoothing, ARIMA, and neural networks.
- “tseries”: This package provides functions for time series analysis and forecasting, including functions for decomposing time series data, and for fitting and forecasting models such as ARIMA

We develop a forecasting model that can predict the future values of the six countries' indexes based on their historical daily data. This will help investors and financial analysts to make informed decisions about buying and selling the stocks in these countries. The six countries' indexes we will be working with are: S&P 500 (USA), NSEI (India), IBEX (Spain), MXX (Mexico), IMOEX.ME (Russia), and BVSP (Brazil).

The dataset consists of daily historical data of the six countries' indexes from 1st January 2017 to 31st December 2022. The data includes the opening price, closing price, highest price, lowest price, and volume. The dataset was obtained from Yahoo Finance.

The output results of a indexes data forecast will depend on the specific forecasting model and techniques used. However, here are some general metrics to analyze our output results.

- Forecasted values: This is the predicted future value of the stock index for a given time period. It can be a single point estimate or a range of values with a certain level of confidence.

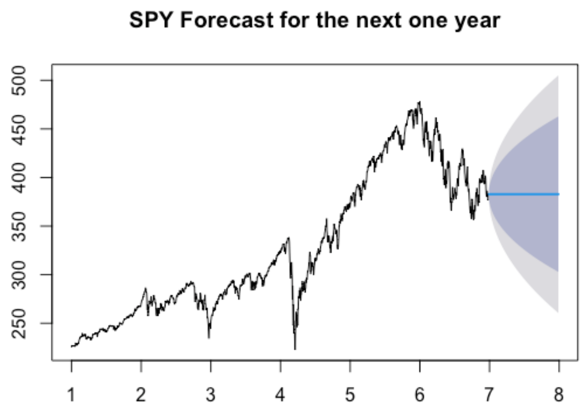
- Confidence intervals: When using a probabilistic forecasting method, you may get confidence intervals along with the point estimates. These intervals show the range of values within which the actual value is expected to fall with a certain level of probability.

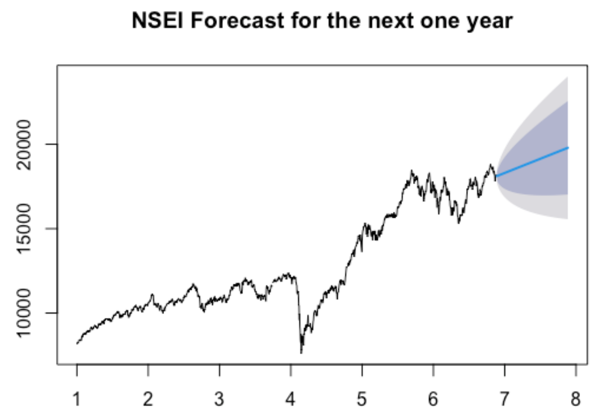
- Residual diagnostics: This output is used to check the assumptions of the forecasting model and the adequacy of the model fit. It typically includes diagnostic plots of the residuals, autocorrelation plots, and tests for normality and stationarity.

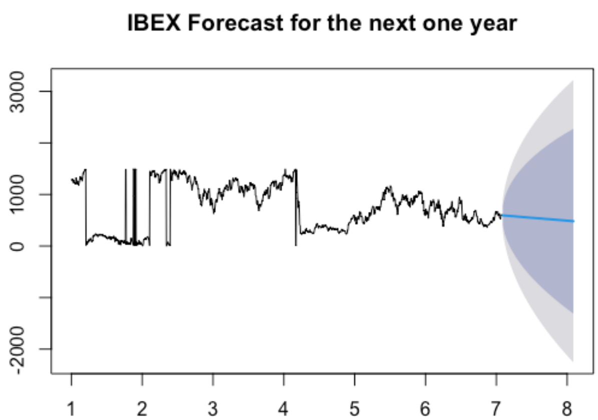
- Model selection criteria: The output may also include model selection criteria such as AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion), which are used to compare different models and select the best one based on goodness-of-fit and complexity.

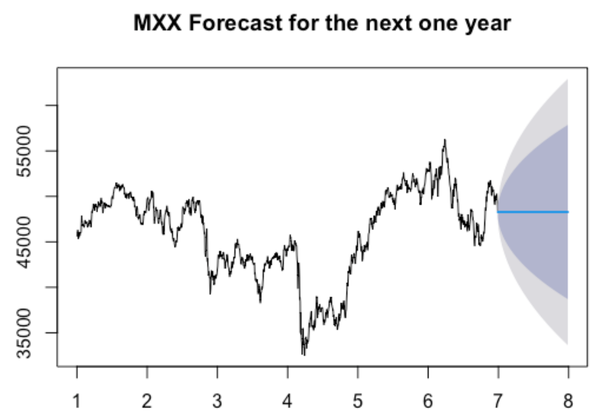
- Performance metrics: These are metrics used to evaluate the accuracy of the forecasting model. Common performance metrics for time series forecasting include mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), and mean absolute percentage error (MAPE).

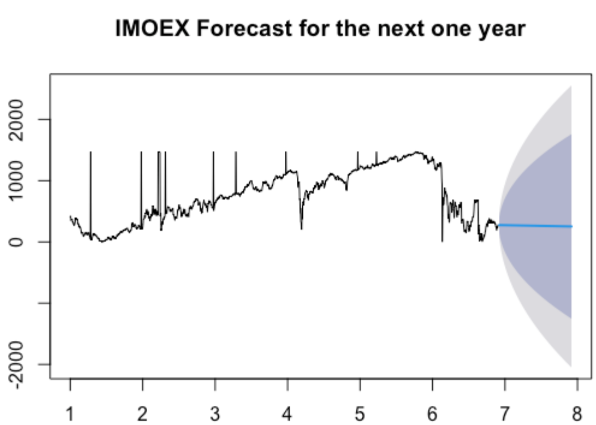
- Visualization of forecast results: Finally, the output may include visualizations of the forecasted values along with the historical data to help interpret the results and identify trends

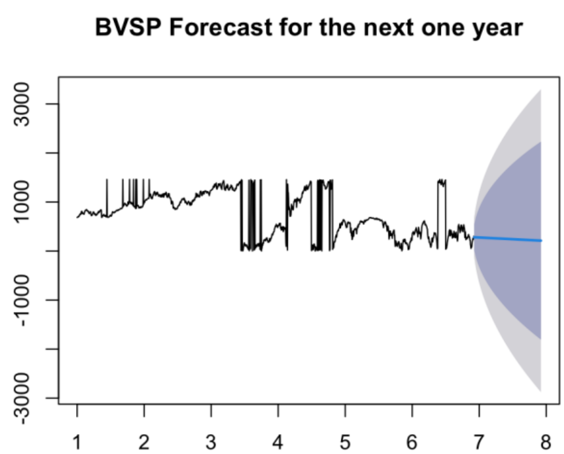












and patterns in the data. Common visualization techniques include line charts, bar charts, and heat maps.

For our forecasting model we have chosen to analysis the visualization of forecast results and and evaluate the results using residual diagnostics and performance metrics.

Model order and Lowest AIC.

SPY.	ARIMA(0,1,2).	AIC - 8524.17
NSEI.	ARIMA(1,1,3) with drift	AIC - 18791.09
MXX.	ARIMA(0,1,1)	AIC - 22586.89
IBEX.	ARIMA(0,1,2)(1,0,0) with drift	AIC - 19504.23
IMOEX	ARIMA(1,1,0) with drift	AIC - 17990.96
BVSP	ARIMA(4,0,2) with drift	AIC - 20123.29

FORECAST PLOTS

Based on the analysis of the stock data for various indices, it can be concluded that the SPY index is likely to experience a stagnate trend over the next 252 trading days, while the NSEI index is expected to have an upward trend during this period.

On the other hand, the IBEX index is expected to show a downward trend, while the MXX index is expected to have a straight stagnate trend. The IMOEX.ME index is also likely to experience a slow downward stagnate trend over the next 252 trading days.

Similarly, the BVSP index is expected to have a slow downward stagnate trend over the next 252 trading days.

It's important to note that stock market predictions are subject to various uncertainties and risks, and the actual trends may differ from the forecasts. Investors should always conduct their own research and analysis before making investment decisions.

6. Conclusion

From the models we built, we can conclude that there is a quantifiable difference in the six international stock markets, as well as the US economy as a whole pre and post Covid-19. We were able to build models with some accuracy, proving that a strong shift was caused by the start of Covid-19. Although Logistic Regression and LDA accurately predicted 84.7% of the rows for Low-Change sectors, QDA and KNN had lower accuracy scores when predicting the same Low-Change sectors. Therefore, we were not able to see any significant differences in model accuracy based on using specific sectors in our model. From our exploratory data analysis, you can definitely see that there were sizeable changes in returns pre and post Covid-19, showing that the pandemic definitely had an impact on those sector stock markets and ETFs. It may just be that because the whole market was affected, we aren’t seeing significant differences between models because even if some stock markets and ETFs were affected at a greater magnitude than others, they were all still affected making them all good predictors at around the same level of importance.

Forecast summary , Stock forecasting involves analyzing historical market data and applying statistical models to predict future trends and movements in stock prices. While stock forecasting can provide valuable insights for investors, it is important to remember that it is not a precise science and comes with inherent risks and uncertainties.

Some of the key factors that can impact stock prices and forecasting include macroeconomic indicators, company performance, geopolitical events, and investor sentiment. It is also important to use a variety of analytical tools and techniques when forecasting, such as technical analysis, fundamental analysis, and sentiment analysis.

Investors should keep in mind that stock forecasting is not a guarantee of future results and should be used as a tool in conjunction with other investment strategies and considerations. Overall, stock forecasting can be a valuable tool for investors looking to make informed investment decisions, but it should be used cautiously and in conjunction with other forms of analysis and research.

7. References:

Ganie, I. R., Wani, T. A., & Yadav, M. P. (2022). Impact of COVID-19 Outbreak on the Stock Market: An Evidence from Select Economies. Business Perspectives and Research, 0(0). https://doi.org/10.1177/22785337211073635

Dippi Verma, Praveen Kumar Sinha and Lakshmi S.R.(2020).Has Covid –19 Infected Indian Stock Market. The IUP Journal of Financial Risk management .

Seungho Baek & Kwan Yong Lee (2021) The risk transmission of COVID-19 in the US stock market from APPLIED ECONOMICS 2021, VOL. 53, NO. 17, 1976–1990 https://doi.org/10.1080/00036846.2020.1854668 https://bozliu.medium.com/financial-data-forecasting-using-r-7a55f2a1599