# Applied Econometrics (Semester V) – Project

OBJECTIVE: *To formulate an econometric model of the impact of economic variables and the COVID-19 outbreak on the stock market.*

TABLE OF AUTHORS (GROUP 3)


| NAME | ROLL NUMBER | INSTITUTION |
|---|---|---|
| Ayush Ghosh Roy | 0518016 | Sri Venkateswara College |
| Aditi Paliwal | 0518077 | Sri Venkateswara College |
| Ashwin Singh | 0518079 | Sri Venkateswara College |


---

**TABLE OF CONTENTS**



---

*SECTION I – PREVIEW*

1. Introduction

Institutions of various forms have been adversely affected with the onset of the COVID – 19 pandemic. The rapid spread of the virus, and its consequent pervasive presence across the globe has caused economies to come to a standstill – supply chains have been disrupted, small scale businesses have been shut down, unemployment has been soaring and growth in industrialized and developing economies alike, has been in negative double digits. An area of interest for economists across the world has been the trends observed in the stock market during this period, with observed trends seemingly contradicting the a priori expectations that theory has established. It, therefore, makes good sense to formulate a model which captures the behaviour of the stock market specifically during this period. This paper details the evolution of a model which can be taken to be an appropriate (but not necessarily perfect) model of the stock market during the pandemic.

2. Sectional Summary

The core content of this paper has been divided into four sections, each section providing information pertaining to a candidate model that fulfils the objective of the paper. It may be beneficial to know at the onset of the paper, that section V consists of the final model, while sections II – IV present the intermediate candidate models, as well as the causes for moving to subsequent models.

Every section consists of the candidate econometric model under examination in the corresponding section, and the theoretical justification for the variables employed in the model. This has been followed by diagnostic tests, which include testing for multicollinearity, autocorrelation, normality of residuals and model specification errors. Information regarding the correction of any errors, if found, has been mentioned in the section itself. The last sub-section of a section consists of reasons as to why we have rejected/not rejected a given model in favour of a subsequent model. The sources used to obtain data for the models have been specified in Section VI of the paper.

---

*SECTION II – Model I*

1. Econometric Model

$$SP500_i = \beta_o + \beta_1 Time_i + \beta_2 MobilityCh_i + \beta_3 WTI_i + \beta_4 NatEmerDec_i + u_i$$

Where;

$SP500_i$ – Value of the S&P500 index as on day *i*

$Time_i$ – The number of business days in the study period, starting from 21st January, 2020

$MobilityCh_i$ – Mobility expressed as a percentage of normal levels , as measured by IHME data, on day *i*

$WTI_i$ – Price of a barrel of West Texas Intermediate Crude Oil as on day *i*

$NatEmerDec_i$ – Dummy Variable with a value of "1" for pandemic related events; "0" otherwise

$u_i$ – Regression error term, on day *i*

*i = 1,2,3,...,n ; n – Total observations*

2. Theoretical Justification for Variables Chosen

Stated below are the justifications for choosing the variables that have been used in the construction of the model:

a) SP500 – This regressand stands for the value of the S&P500 index, which stands for the weighted average index of the largest 500 publicly traded companies traded in the New York Stock Exchange. Due to the efficiency of information, credibility of the stock exchange, and the repercussions of the pandemic in the United States, we found the situation of the US bourses to be an appropriate variable for the behaviour of the stock market. Furthermore, due to the weights given to the companies in the index on the basis of the corresponding industry's size in the economy, we found this specific index to be the best indicator of the behaviour of stock markets in the United States.

b) Time – This trend variable has been added to look at how the market moves in and of itself. By using time as a regressor, we have essentially isolated the long term trend of the market, and how the specific structures of the market and its players influence the movement of the market endogenously. In general, a positive relation a priori is expected to exist between the stock market and time variable.

c) MobilChan – This variable represents Mobility expressed as a percentage of normal levels recorded using cell phone data in the United States with 0% indicating no change, positive values indicating increased mobility and negative values implying declining mobility. Mobility during the pandemic can be considered a proxy for the way people perceive the pandemic (including the restrictive impact of social distancing norms), and are taking economic decisions. These economic decisions reflect on the top and bottom line of companies. For example, if mobility in the economy is low, entertainment and hospitality industries would earn lower revenues and consequently, lower profits. Another way mobility affects the economy is by using it as a proxy for the level of confidence that people have in undertaking economic activities. Lower confidence today does not just translate into lower financial metrics today, but also hurt the future prospects of a business. These two factors together affect how the shares of companies in mobility sensitive industries are traded in the market. In general, a mobility increase, or a positive value of mobility change is considered to be a favourable factor, and thus, a positive relationship between Mobility Change and Stock Market can be expected a priori. We have used IHME data to measure mobility change.

d) WTI – United States is one of the top oil producing economies in the world, in addition to being one of the largest users of oil as well. In any economy, the price of oil has two opposing effects; the first being the effect on the revenue of the producers of oil through their core operating activities in the Statement of Profit and Loss. The second effect is the effect on the industries which use oil as a raw material in their production process, thus affecting gross profits of a firm. Therefore, macroeconomic supply and demand conditions in the economy significantly affect and are affected by the price of oil. During the pandemic, in addition to a decline in the macroeconomic demand for goods and services, a breakdown of the OPEC+ due to Russia's unwillingness to reduce the supply of oil led to a crash in oil prices. This crash had positive as well as negative effects on the stock market. In order to capture the relationship between oil prices and stock market movements, the price of a barrel of West Texas Intermediate Crude Oil has been added to the model. Generally, due to the composition of the US Economy, a positive relationship between the price of WTI and Stock Market movements is expected a priori.

e) NatEmerDec – This dummy variable captures the movements in the stock market related to pandemic related events. Pandemic related events refer to such news, orders, and rules etc. which have significant effects on the functioning of the society. For example, on 11th March the WHO (World health Organisation) declared COVID-19 to be a pandemic; on 13th, March President Donald J. Trump announced a National Emergency as a result of the spread of COVID-19. The variable takes a value of "1" to account for the effect of such announcements, and "0" otherwise. Due to the number and gravitas of negative events and news being greater than that of positive events and news, a negative relationship between this regressor and the regressand is expected a priori.

3. Regression Results

```
      Source |       SS       df       MS              Number of obs =     160
-------------+------------------------------           F(  4,   155) =  339.14
       Model |  11972934.8     4   2993233.7           Prob > F      =  0.0000
    Residual |   1368017.7   155  8825.92066           R-squared     =  0.8975
-------------+------------------------------           Adj R-squared =  0.8948
       Total |  13340952.5   159  83905.3617           Root MSE      =  93.946

------------------------------------------------------------------------------
                      SP500Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------------+------------------------------------------------
                       TimeX1 |   2.701959   .1735035    15.57   0.000     2.359222    3.044695
 MobilityChangechangeagainst |   6.335622   .4811277    13.17   0.000     5.385208    7.286035
NationalEmergencyDeclaration |  -139.4605   37.55201    -3.71   0.000    -213.6402   -65.28071
                WTICrudebarrel |   3.799508   1.054298     3.60   0.000     1.716862    5.882153
                       _cons |   3035.132   57.09998    53.15   0.000     2922.338    3147.927
------------------------------------------------------------------------------
```

In equation form, the model is as follows:

$$\widehat{SP500}_i = 3035.132 + 2.7019 Time_i + 6.3356 MobilityCh_i + 3.7995 WTI_i - 139.4605 NatEmerDec_i$$

From the regression results, following conclusions can be made:

- The explanatory variables are individually significant at the 1% level of significance.
- The regressors are jointly statistically significant as can be seen from the value of the F-statistic.
- The explanatory variables are jointly able to explain 89.75% of the variation in S&P 500.
- The signs of the coefficients of various regressors are consistent with a priori expectations, thus indicating theoretical consistency and data adequacy.

The interpretations of the various values in the estimated regression model are as follows:

- 2.701959 – The mean expected value of the S&P500 index increases by 2.701959 points from one business day to the next, ceteris paribus.
- 6.335622 – The mean expected value of the S&P500 index increases by 6.335622 points when there is an increase in mobility by 100 basis points compared to normal levels, ceteris paribus.
- 3.799508 – The mean expected value of the S&P500 index increases by 3.799508 points when there is an increase of $1 in the price of a barrel of WTI crude oil, ceteris paribus.
- -139.4605 – The mean expected value of the S&P 500 index decreases by 139.4605 points on account of pandemic related news and events, ceteris paribus.


*(Actual vs fitted values of S&P500 using Model I, plotted by observation number)*

4. Diagnostic Testing Results

As mentioned in the Sectional Summary, we moved from Model I to Model II. This was on account of a model misspecification that we discovered in the course of our analysis, the working and results of which are mentioned below. We have used Ramsey's RESET test for ascertaining whether the model has been correctly specified.

*Ramsey's RESET Test (for misspecification error)*

$H_0$: Model is correctly specified

$H_1$: A misspecification error has been committed

Let the level of significance be 1% ($\alpha = 0.01$)

```
Ramsey RESET test using powers of the fitted values of SP500Y
       Ho:  model has no omitted variables
                  F(3, 152) =     19.23
                   Prob > F =      0.0000
```

$$F_{cal} = \frac{(R^2_{new} - R^2_{old})/number\ of\ new\ regressors}{(1-R^2_{new})/(n-number\ of\ parameters\ in\ the\ new\ model)} = 19.23 \sim F_{3,152}\ ;\ F_{critical} = F_{0.01,3,152} \approx 3.95$$

(Note: Three new regressors, viz. $\hat{Y}^2$, $\hat{Y}^3$ and $\hat{Y}^4$ were introduced while running the test in STATA)

Conclusion: Since $F_{cal} > F_{critical}$, we reject the null hypothesis at the 1% level significance. Hence, we can conclude that the old model (Model 1) suffers from a misspecification error. These results lead us to the subsequent LM (Lagrange Multiplier) Test for adding gold prices and ten-year sovereign bond yields as new explanatory variables in the model, which we suspect to be the cause for the misspecification error in Model I. Theoretical justification for adding these variables has been mentioned in Section III of this paper.

*LM Test (for Gold Prices and TenYyear Sovereign Bond Yyields)*

```
      Source |       SS       df       MS              Number of obs =     160
-------------+------------------------------           F(  6,   153) =   19.67
       Model |  595695.105     6  99282.5175            Prob > F      =  0.0000
    Residual |  772322.599   153  5047.86013            R-squared     =  0.4354
-------------+------------------------------           Adj R-squared =  0.4133
       Total |   1368017.7   159  8603.88493            Root MSE      =  71.048

------------------------------------------------------------------------------
           r |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
             TimeX1 |  -.0193318   .4282822    -0.05   0.964     -.865442    .8267783
 MobilityChangechangeagainst |  -3.151478   .5307491    -5.94   0.000    -4.200021   -2.102936
        WTICrudebarrel |  -2.195804    .824096    -2.66   0.009     -3.82388   -.5677277
NationalEmergencyDeclaration |  -68.62955   33.56887    -2.04   0.043    -134.9479   -2.311215
              Gold |   .8524669   .1123627     7.59   0.000     .6304842    1.07445
      YearSovereignBondYield |   464.2461   58.54886     7.93   0.000     348.5775    579.9147
               _cons |  -1941.528   194.2604    -9.99   0.000    -2325.306   -1557.749
------------------------------------------------------------------------------

. scalar Rsq= .4354

. scalar chisqcal= 160*Rsq

. display chisqcal
69.664
```

$H_0$: Restricted Model is correctly specified

$H_1$: A misspecification error has been committed

$\chi^2_{cal} = \text{n.R}^2 = 69.664 \sim \chi^2_2\ ;\ \chi^2_{critical} = \chi^2_{2,0.01} = 9.2103$

Conclusion: Since $\chi^2_{cal} > \chi^2_{critical}$, we reject the null hypothesis, and conclude that a misspecification error has been committed. Therefore, the price of gold and 10-year sovereign bond yields should be added as regressors in the model. This has been done in Section III (Model II) of the paper.

---

*SECTION III – Model II*

1. Econometric Model

$$SP500_i = \beta_o + \beta_1 Time_i + \beta_2 MobilityCh_i + \beta_3 WTI_i + \beta_4 NatEmerDec_i + \beta_5 Gold_i + \beta_6 SovBondYield_i + u_i$$

Where:

$SP500_i$ – Value of the S&P500 index as on day *i*

$Time_i$ – The number of business days in the study period, starting from 21st January, 2020

$MobilityCh_i$ – Mobility expressed as a percentage of normal levels , as measured by IHME data, on day *i*

$WTI_i$ – Price of a barrel of West Texas Intermediate Crude Oil as on day *i*

$NatEmerDec_i$ – Dummy Variable with a value of "1" for pandemic related events; "0" otherwise

$Gold_i$ – Price of one Troy Ounce of gold on day *i* (expressed in US Dollars)

$SovBondYield_i$ – Yield on US 10-Year Sovereign Bonds on day *i*

$u_i$ – Regression error term, on day *i* (measured from the base date)

*i = 1,2,3,...,n ; n – Total observations*

## 2. Theoretical Justification for Changed Introduced

In the previous Section, by running the LM Test, with Gold Prices and 10-Year Sovereign Bond Yields included as the candidate regressors, we came to the conclusion that they should be included in the model. Stated below are the justifications for making changes to the previous model, thus bringing us to Model II:

a) Inclusion of Gold Prices – Theoretically, it makes good sense to include the price of gold as an explanatory variable in the model. Gold is generally considered a safe asset – it is often used by investors to park their funds, when there is an expectation of high volatility or decline in the prices of shares listed on the stock exchange. During adverse Black Swan events such as the ongoing pandemic, a significant amount of funds were shifted to gold on account of the lack of confidence in the market, and subsequent expectations of high volatility and poor performance in the stock market. Using the aforementioned justification, we have an a priori expectation of an inverse correlation between Gold Prices and movements in S&P500 (regressand).

b) Inclusion of 10-Year Sovereign Bond Yields – Also known as Treasury Bonds, a Sovereign Bond refers to a debt instrument issued by the Government. The interest paid on these bonds is known as its yield. Theoretically, the return on these bonds is considered as the risk-free rate of return on investing money in the market. Therefore, an investor will assume the risks associated with investing in stocks only if the rate of return on their investment in shares is more than the 10-Year Sovereign Bond Yields. Since the pandemic has led to a significant increase in volatility and risk in the market, movements in 10-Year Sovereign Bond Yields can be used as an explanatory variable while formulating a model of the stock exchange in the pandemic. Since price of bonds and yields are inversely related and the price of bonds and stock market movements are inversely related, there is an a priori expectation of a direct/positive relationship between this regressor and the regressand.

## 3. Regression Results

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 12568629.9 | 6 | 2094771.65 | Number of obs | 160 |
| Residual | 772322.598 | 153 | 5047.86012 | F( 6, 153) | 414.98 |
| Total | 13340952.5 | 159 | 83905.3617 | Prob > F | 0.0000 |
| | | | | R-squared | 0.9421 |
| | | | | Adj R-squared | 0.9398 |
| | | | | Root MSE | 71.048 |

| SP500Y | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Gold | .8524669 | .1123627 | 7.59 | 0.000 | .6304842 | 1.07445 |
| YearSovereignBondYield | 464.2461 | 58.54886 | 7.93 | 0.000 | 348.5775 | 579.9147 |
| MobilityChangechangeagainst | 3.184143 | .5307491 | 6.00 | 0.000 | 2.1356 | 4.232686 |
| TimeX1 | 2.682627 | .4282822 | 6.26 | 0.000 | 1.836517 | 3.528737 |
| WTICrudebarrel | 1.603704 | .824096 | 1.95 | 0.053 | -.0243725 | 3.23178 |
| NationalEmergencyDeclaration | -208.09 | 33.56887 | -6.20 | 0.000 | -274.4084 | -141.7717 |
| _cons | 1093.605 | 194.2604 | 5.63 | 0.000 | 709.8257 | 1477.384 |

In equation form, the model is as follows:

$$\widehat{SP500}_i = 1093.685 + 2.6826Time_i + 3.1841MobilityCh_i + 1.6037WTI_i - 208.09NatEmerDec_i + 0.8524Gold_i + 464.2461SovBondYield_i$$

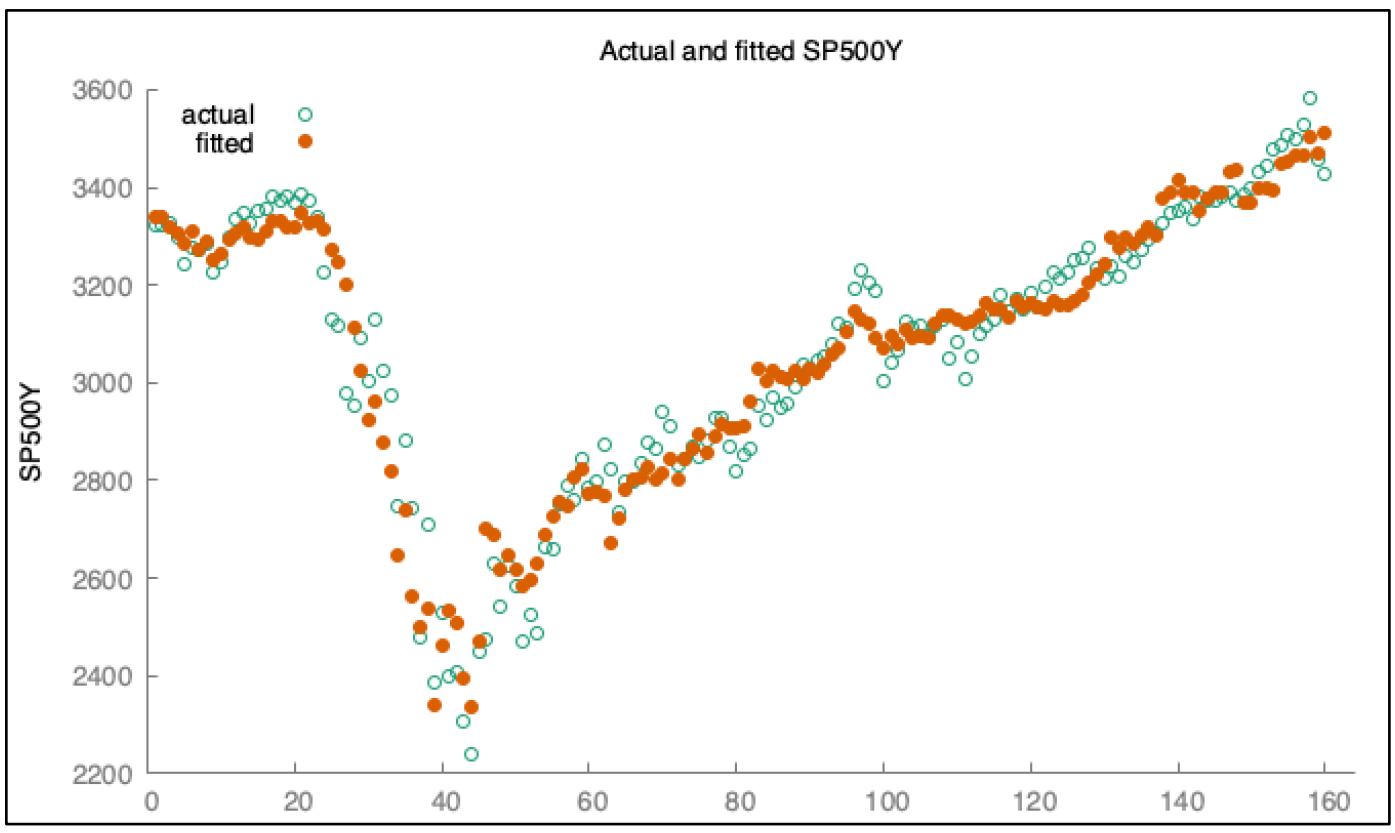
From the regression results, following conclusions can be made:

- All explanatory variables, except WTI Crude Oil Prices (p-value = 0.053), are significant at 1% level of significance.
- The regressors are jointly statistically significant as can be seen from the value of the F-statistic.
- The explanatory variables are jointly able to explain 94.21% of the variation in S&P 500.
- Only the sign of the coefficient of Gold Prices are in contradiction to a priori expectations. Two reasons can be put forth to explain this:
  a) The impact of the event was such that people lost confidence in gold as a safe asset. Therefore, investors removed funds from the stock market as well as from gold, rather than simply shifting to gold from stocks. Investors may have shifted their funds to other assets such as sovereign bonds, corporate bonds, or may simply hold them in liquid form.

b) The relationship between Gold Prices and stock market movements is generally considered to be a long run relationship, on account of portfolio diversification. Therefore, our time period may be too short for the a priori expectations to hold.

The interpretations of the various values in the estimated regression model are as follows:

- 2.682627 – The mean expected value of the S&P500 index increases by 2.682627 points from one business day to the next, ceteris paribus.
- 3.184143 – The mean expected value of the S&P500 index increases by 3.184143 points when there is an increase in mobility by 100 basis points compared to normal levels, ceteris paribus.
- 1.603704 – The mean expected value of the S&P500 index increases by 1.603704 points where there is an increase of $1 in the price of a barrel of WTI crude oil, ceteris paribus.
- -208.09– The mean expected value of the S&P 500 index decreases by 208.09 points subsequently on account of pandemic related news and events, ceteris paribus.
- 0.8524669 – The mean expected value of the S&P index increases by 0.8524669 points where there is an increase of $1 in the price of one troy ounce of gold, ceteris paribus.
- 464.2461 – The mean expected value of the S&P index increases by 464.2461 points when there is an increase in the yield of 10-Year Sovereign Bonds by 100 basis points, ceteris paribus.




*(Actual vs fitted values of S&P500 using Model II, plotted by observation number)*

4. <u>Diagnostic Testing Results (Multicollinearity)</u>

For the purpose of moving from Model II to Model III, we discovered that this model suffers from a serious drawback – high multicollinearity between variables. The results of our multicollinearity diagnosis are shown in this section:

| Variable | VIF | 1/VIF |
|---|---|---|
| YearSovere~d | 13.21 | 0.075680 |
| TimeX1 | 12.40 | 0.080628 |
| MobilityCh~t | 8.37 | 0.119410 |
| Gold | 7.62 | 0.131215 |
| WTICrudeba~l | 3.40 | 0.293799 |
| NationalEm~n | 2.09 | 0.477818 |
| Mean VIF | 7.85 | |

| | SP500Y | Time | Mobili~t | WTICru~l | Nation~n | Gold | YearSo~d |
|---|---|---|---|---|---|---|---|
| SP500Y | 1.0000 | | | | | | |
| Time | 0.3899 | 1.0000 | | | | | |
| MobilityCh~t | 0.8122 | -0.1002 | 1.0000 | | | | |
| WTICrudeba~l | 0.7268 | -0.0070 | 0.8038 | 1.0000 | | | |
| NationalEm~n | -0.5384 | -0.2236 | -0.4324 | -0.2074 | 1.0000 | | |
| Gold | 0.5448 | 0.9148 | 0.0540 | 0.1207 | -0.3498 | 1.0000 | |
| YearSovere~d | 0.2388 | -0.7431 | 0.6491 | 0.5408 | 0.0624 | -0.6114 | 1.0000 |

We come to the following conclusions from the VIF and pair wise correlation matrices:

a) We find that the VIFs of 10-Year Sovereign Bond Yield (13.21) and Time (12.41) are unacceptably high, while those of Mobility Change (8.37) and Gold prices (7.62) are alarmingly close to 10.
b) From the pair wise correlation matrix, we find that the variable Time has an unacceptably high pair wise correlation with Gold (0.9148), while having a high pair wise correlation with 10-Year Sovereign Bond Yields (-0.7431) as well.

Using the conclusions from the analysis of the VIF and pair wise correlation matrices, we believe that the variable time has high multicollinearity with the other variables. In order to confirm this hypothesis, we ran an auxiliary regression of Time on the other regressors.

*Auxiliary Regression*

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 313800.074 | 5 | 62760.0148 | Number of obs | 160 |
| Residual | 27519.9261 | 154 | 178.700819 | F( 5, 154) | 351.20 |
| Total | 341320 | 159 | 2146.66667 | Prob > F | 0.0000 |
| | | | | R-squared | 0.9194 |
| | | | | Adj R-squared | 0.9168 |
| | | | | Root MSE | 13.368 |

| TimeX1 | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Gold | .1795036 | .0154183 | 11.64 | 0.000 | .1490449 | .2099623 |
| YearSovereignBondYield | -88.53505 | 8.393778 | -10.55 | 0.000 | -105.1169 | -71.95325 |
| NationalEmergencyDeclaration | 30.15966 | 5.82976 | 5.17 | 0.000 | 18.64304 | 41.67628 |
| WTICrudebarrel | .2352032 | .1538929 | 1.53 | 0.128 | -.0688103 | .5392167 |
| MobilityChangechangeagainst | .4850831 | .0918934 | 5.28 | 0.000 | .3035487 | .6666174 |
| _cons | -141.5524 | 34.72508 | -4.08 | 0.000 | -210.1513 | -72.95338 |

We than ran a test of overall significance for the auxiliary regression model:

$H_0$: $p^2 = 0$

$H_1$: $p^2 > 0$

$F_{cal} = 351.20$ ; $F_{crtical} = F_{5,154}^{0.01} = 3.17$

Conclusion: Since $F_{cal} > F_{crtical}$, we reject the null hypothesis, stipulating that the model is significant and Time linearly dependent on the other regressors. Therefore, our initial hypothesis of the behaviour of Time being jointly explained by the other regressors cannot be rejected. This can also be witnessed in the extremely high correlation of 0.9194. On the basis of the aforementioned results, we decided to drop Time as a regressor, thus moving to Model III.

---

*SECTION IV – Model III*

1. Econometric Model

$$SP500_i = \beta_o + \beta_3 MobilityCh_i + \beta_4 WTI_i + \beta_5 NatEmerDec_i + \beta_6 Gold_i + \beta_7 SovBondYield_i + u_i$$

Where:

$SP500_i$ – Value of the S&P500 index as on day *i*

$MobilityCh_i$ – Mobility expressed as a percentage of normal levels , as measured by IHME data, on day *i*

$WTI_i$ – Price of a barrel of West Texas Intermediate Crude Oil as on day *i*

$NatEmerDec_i$ – Dummy Variable with a value of "1" for pandemic related events; "0" otherwise

$Gold_i$ – Price of one Troy ounce of gold on day *i*

$SovBondYield_i$– Yield on US 10-Year Sovereign Bond on day *i*

$u_i$ – Regression error term, on day *i* (measured from the base date)

*i = 1,2,3,...,n ; n – Total observations*

2. Regression Results

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| Model | 12370583.1 | 5 | 2474116.62 | Number of obs | 160 |
| Residual | 970369.399 | 154 | 6301.09999 | F( 5, 154) | 392.65 |
| Total | 13340952.5 | 159 | 83905.3617 | Prob > F | 0.0000 |
| | | | | R-squared | 0.9273 |
| | | | | Adj R-squared | 0.9249 |
| | | | | Root MSE | 79.379 |

| SP500Y | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Gold | 1.334008 | .0915549 | 14.57 | 0.000 | 1.153142 | 1.514874 |
| YearSovereignBondYield | 226.7396 | 49.8428 | 4.55 | 0.000 | 128.2757 | 325.2034 |
| MobilityChangechangeagainst | 4.48544 | .545669 | 8.22 | 0.000 | 3.407478 | 5.563403 |
| WTICrudebarrel | 2.234666 | .9138257 | 2.45 | 0.016 | .4294143 | 4.039918 |
| NationalEmergencyDeclaration | -127.1829 | 34.61749 | -3.67 | 0.000 | -195.5693 | -58.79645 |
| _cons | 713.8725 | 206.1998 | 3.46 | 0.001 | 306.5273 | 1121.218 |

In equation form, the model is as follows –

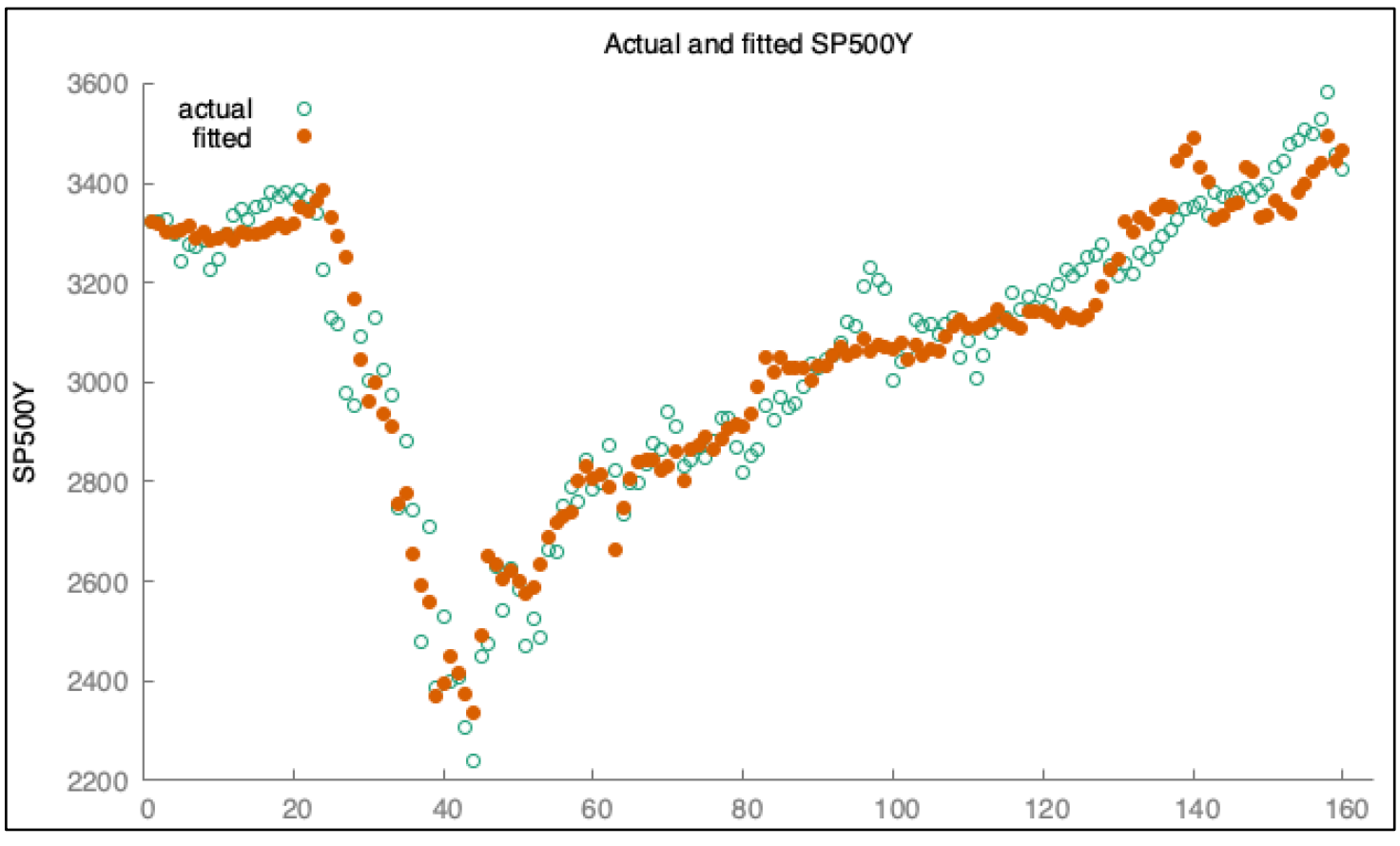
$$\widehat{SP500}_i = 713.8725 + 4.4854 MobilityCh_i + 2.2346 WTI_i - 127.1829 NatEmerDec_i + 1.334 Gold_i + 226.7396 SovBondYield_i$$

From the regression results, following conclusions can be made:

- All explanatory variables, except WTI Crude Oil Prices (t = 2.45, p-value = 0.016), are significant at 1% level of significance.
- The regressors are jointly statistically significant as can be seen from the value of the F-statistic.
- The explanatory variables are jointly able to explain 92.73% of the variation in S&P 500.
- Only the sign of the coefficient of Gold Prices are in contradiction to a priori expectations for reasons mentioned in Section II.

The interpretations of the various values in the estimated regression model are as follows:

- 4.48544 – The mean expected value of the S&P500 index increases by 4.48544 points when there is an increase in mobility by 100 basis points compared to normal levels, ceteris paribus.
- 2.234666 – The mean expected value of the S&P500 index increases by 2.234666 points where there is an increase of $1 in the price of a barrel of WTI Crude Oil, ceteris paribus.
- -127.1829– The mean expected value of the S&P 500 index decreases by 127.1829 points subsequently on account of pandemic related news and events, ceteris paribus.
- 1.33408 – The mean expected value of the S&P 500 index increases by 1.33408 points where there is an increase of $1 in the price of one troy ounce of gold, ceteris paribus.
- 226.7396 – The mean expected value of the S&P index increases by 226.7396 points when there is an increase in the yield of 10-Year Sovereign Bonds by 100 basis points, ceteris paribus.




*(Actual vs fitted values of S&P500 using Model III, plotted by observation number)*

3. Diagnostic Testing Results

Numerous diagnostic tests were run to ascertain how well the model fit the assumptions of the Classical Linear Regression Model. The results of the tests are mentioned below:

*3.a. Multicollinearity*

| Variable | VIF | 1/VIF |
|---|---|---|
| YearSovere~d | 7.67 | 0.130354 |
| MobilityCh~t | 7.09 | 0.141016 |
| Gold | 4.05 | 0.246703 |
| WTICrudeba~l | 3.35 | 0.298255 |
| NationalEm~n | 1.78 | 0.560859 |
| Mean VIF | 4.79 | |

Conclusion: Using Variance Inflation Factor Matrix, we observe that the VIFs of all the regressors are below 10, and therefore, at acceptable levels. This is a welcome result, since in Model II, multicollinearity was a central problem.

*3.b. Autocorrelation: Breusch-Godfrey General Test of Autocorrelation*

In order to determine the appropriate number of lagged error terms for the auxiliary regression, we chose the model with the lowest AIC (Akaike's Information Criterion) and SIC (Schwarz's Information Criterion). The results of our analysis can be summarised using following table –

| Number of Lagged Error Terms | AIC | SIC |
|---|---|---|
| 1 | 1760.017 | 1781.500 |
| 2 | 1746.938 | 1771.439 |
| 3 | 1733.036 | 1760.543 |
| 4 | 1723.402 | 1753.900 |
| 5 | 1712.833 | 1746.311 |
| 6 | 1704.501 | 1740.944 |
| 7 | 1694.173 | 1733.569 |
| 8 | 1675.686 | 1718.020 |
| 9 | 1665.465 | 1710.724 |
| 10 | 1644.103 | 1692.273 |

As can be seen in the table, 10 lags yield the lowest values for AIC and SIC. Hence, we take 10 as the number of lagged error terms for our auxiliary regression:

$$e_i = A_1 + A_2 Gold_i + A_3 MobilityCh_i + A_4 WTI_i + A_5 NatEmerDec_i + A_6 Bonds_i + \rho_1 e_{i-1} + \rho_2 e_{i-2} + \dots \rho_{10} e_{i-10} + v_i$$

$H_0$: No autocorrelation (i.e. $\rho_1 = \rho_2 = \rho_3 = \dots = \rho_{10} = 0$)

$H_1$: Autocorrelation is present

Let the level of significance be 1% ($\alpha = 0.01$)

```
Breusch-Godfrey LM test for autocorrelation
------------------------------------------------------------------------------
    lags(p)  |          chi2              df                 Prob > chi2
-------------+----------------------------------------------------------------
      10     |         91.049             10                   0.0000
------------------------------------------------------------------------------
                        H0: no serial correlation
```

$\chi^2_{cal} = (n-10)R^2_{aux} = 91.049 \sim \chi^2_{10}$ ; $\chi^2_{critical} = \chi^2_{10,0.01} = 23.2093$

Conclusion: Since $\chi^2_{cal} > \chi^2_{critical}$, we reject the null hypothesis at the 1% level of significance and conclude that the model suffers from autocorrelation. To correct the standard errors of the estimated coefficients for autocorrelation, we make use of the HAC (heteroscedasticity and autocorrelation consistent) procedure (developed by Newey and West).

```
Regression with Newey-West standard errors          Number of obs  =       160
maximum lag: 0                                      F(  5,   154)  =    361.52
                                                    Prob > F       =    0.0000

------------------------------------------------------------------------------
                             |             Newey-West
                      SP500Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------------
                        Gold |   1.334008   .0963995    13.84   0.000     1.143572    1.524444
      YearSovereignBondYield |   226.7396   46.33714     4.89   0.000     135.2011     318.278
 MobilityChangechangeagainst |    4.48544   .6515666     6.88   0.000     3.198278    5.772602
                WTICrudebarrel |   2.234666   1.438791     1.55   0.122     -.607649    5.076981
NationalEmergencyDeclaration |  -127.1829   38.59771    -3.30   0.001    -203.4322   -50.93358
                       _cons |   713.8725   206.5349     3.46   0.001     305.8652    1121.88
------------------------------------------------------------------------------
```

From the regression results, following conclusions can be made:

- All explanatory variables, except WTI Crude Oil Prices, are significant at 1% level of significance. This problem has been taken care of, in Model IV of this paper.
- When compared with the OLS standard errors, the above-mentioned HAC standard errors do not differ substantially; therefore despite the evidence of autocorrelation based on the Breusch-Godfrey Test, the autocorrelation problem does not seem to be very serious.

---

*SECTION V – Model IV*

1. Econometric Model

$$SP500_i = \beta_o + \beta_1 MobilityCh_i + \beta_2 BrentCru_i + \beta_3 NatEmerDec_i + \beta_4 Gold_i + \beta_5 SovBondYield_i + u_i$$

Where:

$SP500_i$ – Value of the S&P500 index as on day *i* (measured from the base date)

$MobilityCh_i$ – Mobility expressed as a percentage of normal levels , as measured by IHME data, on day *i*

$BrentCru_i$ – Price of a barrel of Brent Crude Oil as on day *i*

$NatEmerDec_i$ – Dummy Variable with a value of "1" for pandemic related events; "0" otherwise

$u_i$ – Regression error term, on day *i* (measured from the base date)

$Gold_i$ – Price of gold per Troy Ounce in USD from London Bullion Market as on day *i*

$SovBondYield_i$ – US Ten year Sovereign bond yield as on day *i*

*i = 1,2,3,...,n ; n – Total observation*

## 2. Theoretical Justification for substituting WTI with Brent Crude Oil

Since WTI proved to be insignificant in the previous model after using the Newey-West method of correcting autocorrelation, we replace it with Brent Crude. Brent is a major oil benchmark used globally to price two-thirds of the world's crude oil supplies. The demand and supply shock of crude oil due to the pandemic was accompanied by the disagreement between the OPEC (Organization of the Petroleum Exporting Countries) and Russia (which is an OPEC Plus member) which further exacerbated the turbulence in the oil markets. Since Russia and OPEC are some of the major producers of Brent, it would be fitting to include it in the model as an indicator of oil prices. Furthermore, since Brent Crude Oil and WTI move in tandem in the global market, a priori, a positive relationship is expected between movements in the S&P500 and movements in Brent Crude Oil.

## 3. Regression Results

| Source | SS | df | MS | Number of obs | 160 |
|---|---|---|---|---|---|
| Model | 12401067.6 | 5 | 2480213.52 | F( 5, 154) | 406.38 |
| Residual | 939884.896 | 154 | 6103.14867 | Prob > F | 0.0000 |
| Total | 13340952.5 | 159 | 83905.3617 | R-squared | 0.9295 |
| | | | | Adj R-squared | 0.9273 |
| | | | | Root MSE | 78.123 |

| SP500Y | Coef. | Std. Err. | t | P>\|t\| | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Gold | 1.274415 | .0939409 | 13.57 | 0.000 | 1.088836 | 1.459994 |
| YearSovereignBondYield | 194.0646 | 50.96768 | 3.81 | 0.000 | 93.37856 | 294.7507 |
| MobilityChangechangeagainst | 4.037752 | .5728495 | 7.05 | 0.000 | 2.906095 | 5.169409 |
| BrentCrudebarrel | 3.861907 | 1.155576 | 3.34 | 0.001 | 1.57908 | 6.144733 |
| NationalEmergencyDeclaration | -137.5128 | 34.30873 | -4.01 | 0.000 | -205.2893 | -69.73631 |
| _cons | 754.1071 | 203.5918 | 3.70 | 0.000 | 351.914 | 1156.3 |

In equation form, the model is as follows:

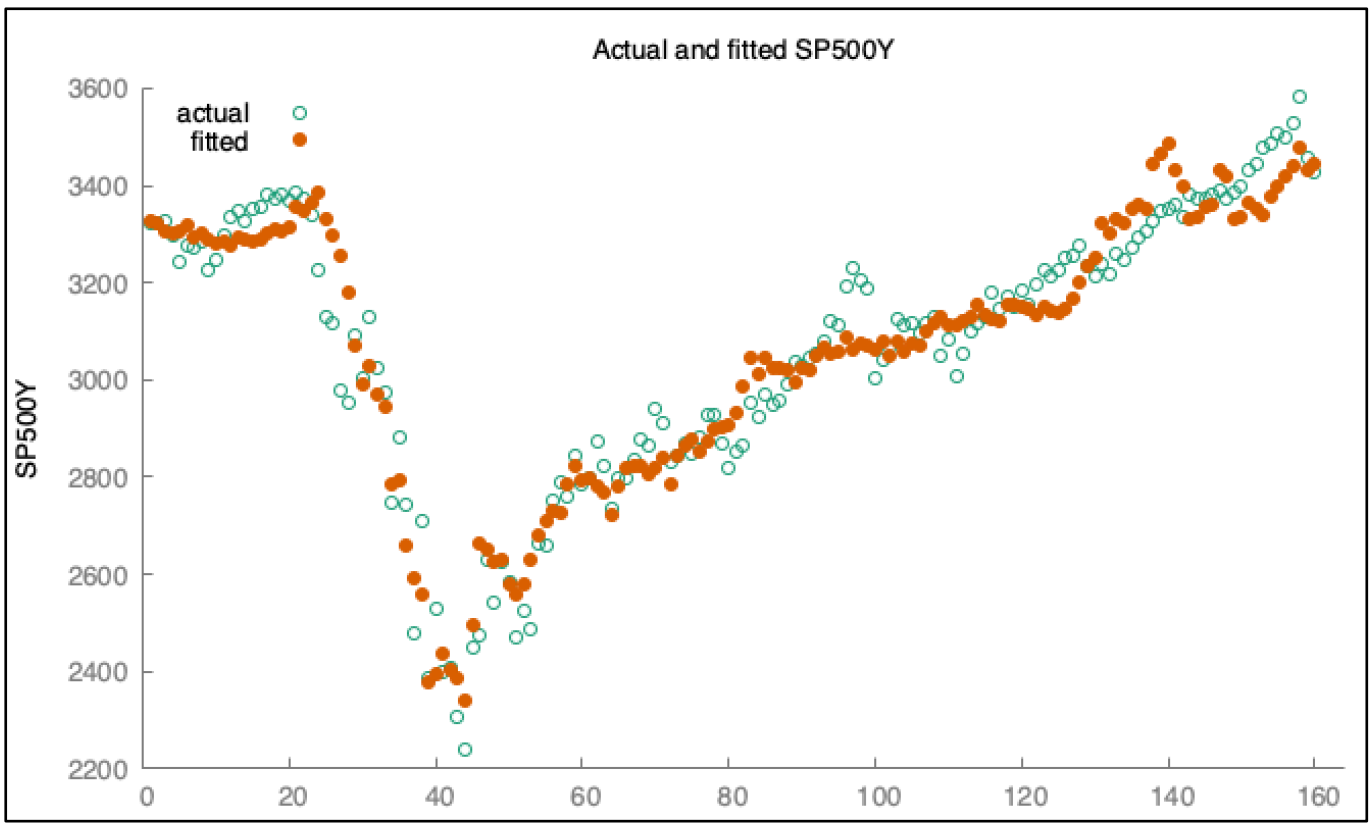
$$\widehat{SP500}_i = 754.1071 + 4.0377MobilityCh_i + 3.8619BrentCru_i - 137.5128NatEmerDec_i + 1.2744Gold_i + 194.0646SovBondYield_i$$

From the regression results, following conclusions can be made:

- All explanatory variables are significant at 1% level of significance.
- The regressors are jointly statistically significant as can be seen from the value of the F-statistic.
- The explanatory variables are jointly able to explain 92.95% of the variation in S&P 500.
- Only the sign of the coefficient of Gold Prices are in contradiction to a priori expectations for reasons mentioned in Section II.

The interpretations of the various values in the estimated regression model are as follows:

- 4.037752 – The mean expected value of the S&P500 index increases by 4.037752 points when there is an increase in mobility by 100 basis points compared to normal levels, ceteris paribus.
- 3.861907 – The mean expected value of the S&P500 index increases by 2.234666 points where there is an increase of $1 in the price of a barrel of Brent crude oil, ceteris paribus.
- -137.5128 – The mean expected value of the S&P 500 index decreases by 137.5128 points subsequently on account of pandemic related news and events, ceteris paribus.
- 1.274415 – The mean expected value of the S&P index increases by 1.33408 points where there is an increase of $1 in the price of one troy ounce of gold, ceteris paribus.
- 194.0646 – The mean expected value of the S&P index increases by 226.7396 points when there is an increase in the yield of 10-Year Sovereign Bonds by 100 basis points, ceteris paribus.




*(Actual vs fitted values of S&P500 using Model IV, plotted by observation number)*

4. Diagnostic Testing Results

Being the final model in our paper, extensive diagnostic tests were run to test the credibility of the model, by checking the adherence of the model to the assumptions of the Classical Linear Regression Model. Results of our diagnostic tests are as follows;

*4.a. Multicollinearity*

| Variable | VIF | 1/VIF |
|---|---|---|
| YearSovere~d | 8.28 | 0.120747 |
| MobilityCh~t | 8.07 | 0.123932 |
| BrentCrude~l | 5.69 | 0.175751 |
| Gold | 4.41 | 0.226968 |
| NationalEm~n | 1.81 | 0.553062 |
| Mean VIF | 5.65 | |

Conclusion: From the VIF matrix, we see that none of the VIFs exceed 10. Therefore, the Variance Inflation Factors are at acceptable levels (and so is multicollinearity).

*4.b. Autocorrelation - Breusch-Godfrey General Test of Autocorrelation & Graphical Observations*

In order to determine the appropriate number of lagged error terms for the auxiliary regression, we chose the model with the lowest AIC (Akaike's Information Criterion) and SIC (Schwarz's Information Criterion). The results of our analysis can be summarised using the following table –

| Number of Lagged Error Terms | AIC | SIC |
|---|---|---|
| 1 | 1749.742 | 1771.225 |
| 2 | 1732.928 | 1757.428 |
| 3 | 1719.750 | 1747.256 |
| 4 | 1709.348 | 1739.847 |
| 5 | 1696.788 | 1730.625 |
| 6 | 1688.579 | 1725.023 |
| 7 | 1678.840 | 1718.236 |
| 8 | 1660.183 | 1702.518 |
| 9 | 1649.854 | 1695.113 |
| 10 | 1631.026 | 1679.197 |

As can be seen in the table, a lag length of 10 yields the lowest values for AIC and SIC. Hence, we take 10 as the number of lagged error terms for our auxiliary regression:

$$e_i = A_1 + A_2 Gold_i + A_3 MobilityCh_i + A_4 BrentCru_i + A_5 NatEmerDec_i + A_6 Bonds_i + \rho_1 e_{i-1} + \rho_2 e_{i-2} + \dots \rho_{10} e_{i-10} + v_i$$

$H_0$: No autocorrelation (i.e., $\rho_1 = \rho_2 = \rho_3 = \ldots = \rho_{10} = 0$)

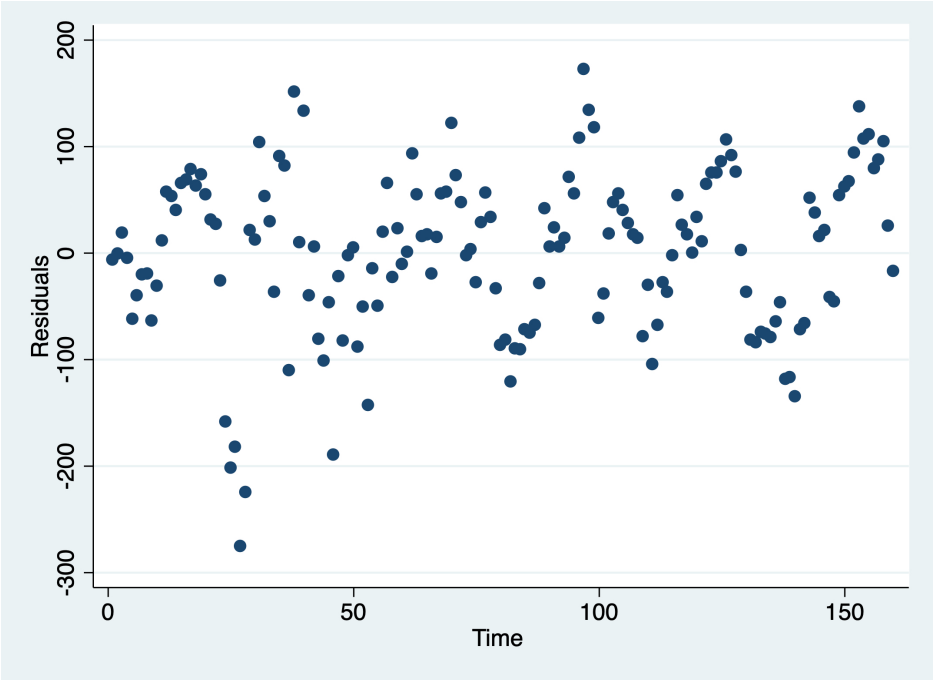
$H_1$: Autocorrelation is present

Let the level of significance be 1% ($\alpha = 0.01$)

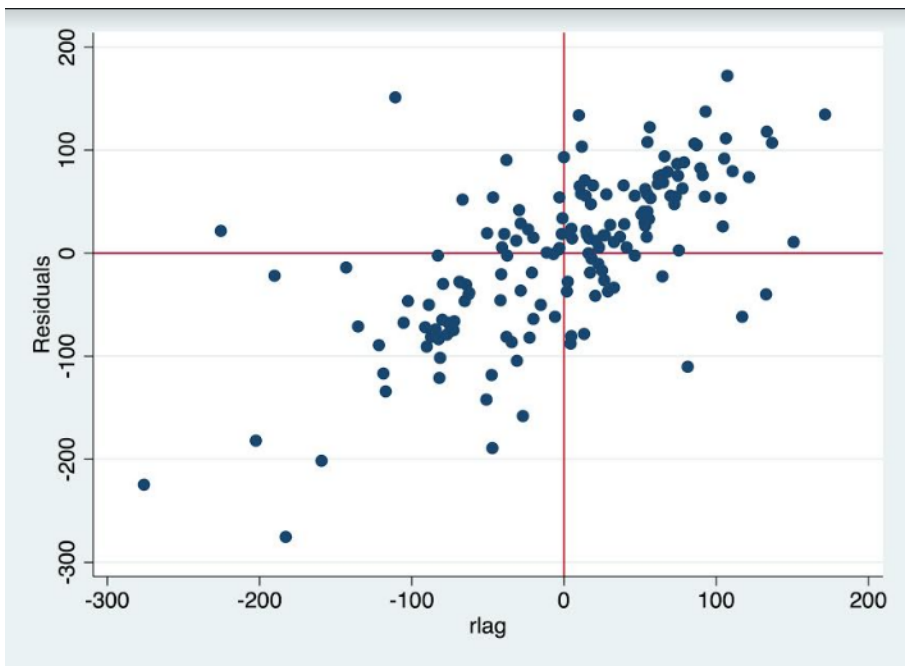
```
Breusch-Godfrey LM test for autocorrelation
-------------------------------------------------------------------------------
    lags(p)  |          chi2              df                 Prob > chi2
-------------+-----------------------------------------------------------------
      10     |         94.381             10                   0.0000
-------------------------------------------------------------------------------
                        H0: no serial correlation
```

$\chi^2_{cal} =$ (n-10). $R^2_{aux} = 94.381 \sim \chi^2_{10}$ ; $\chi^2_{critical} = \chi^2_{10,0.01} = 23.2093$

Conclusion: $\chi^2_{cal} > \chi^2_{critical}$ , thus we reject the null hypothesis at 1% level of significance.




*(Scatter plot of $\hat{u}_i$ vs Time)*




(*Scatter plot of $\hat{u}_i$ vs $\hat{u}_{i-1}$)*

Using the evidence mentioned above, we conclude that the model suffers from autocorrelation. To correct the standard errors of the estimated coefficients for autocorrelation, we make use of the HAC (heteroscedasticity and autocorrelation consistent) procedure (developed by Newey and West).

```
Regression with Newey-West standard errors          Number of obs  =       160
maximum lag: 0                                      F(  5,   154)  =    398.88
                                                    Prob > F       =    0.0000

------------------------------------------------------------------------------
                             |             Newey-West
                      SP500Y |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------------------+------------------------------------------------------------------
                        Gold |   1.274415   .0938575    13.58   0.000      1.089001    1.459829
     YearSovereignBondYield |   194.0646   45.67338     4.25   0.000      103.8374    284.2918
 MobilityChangechangeagainst |   4.037752   .6058077     6.67   0.000      2.840986    5.234518
            BrentCrudebarrel |   3.861907   .9036951     4.27   0.000      2.076668    5.647145
NationalEmergencyDeclaration |  -137.5128   38.24363    -3.60   0.000     -213.0626   -61.96295
                       _cons |   754.1071   207.5103     3.63   0.000      344.1729    1164.041
------------------------------------------------------------------------------
```

From the regression results, following conclusions can be made;

- All explanatory variables are significant at 1% level of significance using HAC estimates.
- Significant changes to the Standard Errors have not been observed. We have thus, corrected for Autocorrelation in the model.

*4.c. Ramsey's RESET Test (for misspecification error)*

$H_0$: Model is correctly specified

$H_1$: A misspecification error has been committed

Let the level of significance be 1% ($\alpha = 0.01$)

```
Ramsey RESET test using powers of the fitted values of SP500Y
       Ho:  model has no omitted variables
                  F(3, 151) =      3.67
                   Prob > F =      0.0138
```

$$F_{cal} = \frac{(R^2_{new} - R^2_{old})/number\ of\ new\ regressors}{(1-R^2_{new})/(n-number\ of\ parameters\ in\ the\ new\ model)} = 3.67 \sim F_{3,151}\ ;\ F_{critical} = F^{0.01}_{3,151} \approx 3.95$$

(Note: Three new regressors, viz. $\hat{Y}^2$, $\hat{Y}^3$ and $\hat{Y}^4$ were introduced while running the test in STATA)

Conclusion: Since $F_{cal} < F_{critical}$, we do not reject the null hypothesis at 1% level of significance, and stipulate that no misspecification error has been committed.

*4.d.i. Jarque-Bera Test for Normality of Residuals*

Using STATA, we obtain the following results for Model IV:

```
                            Residuals
-------------------------------------------------------------
      Percentiles      Smallest
 1%    -225.0504      -275.7295
 5%    -128.2631      -225.0504
10%    -89.18578      -202.2785       Obs                 160
25%    -46.92388       -189.748       Sum of Wgt.         160

50%     11.00256                      Mean           7.79e-08
                        Largest       Std. Dev.       76.8845
75%     54.98725       133.7314
90%      90.7953       137.1339       Variance        5911.226
95%     109.1763       151.0332       Skewness       -.6078037
99%     151.0332       171.9869       Kurtosis        3.662651
```

From the regression results we find that the residuals exhibit the following characteristics-

i) Skewness (S) = -0.6078037
ii) Kurtosis (K) = 3.662651

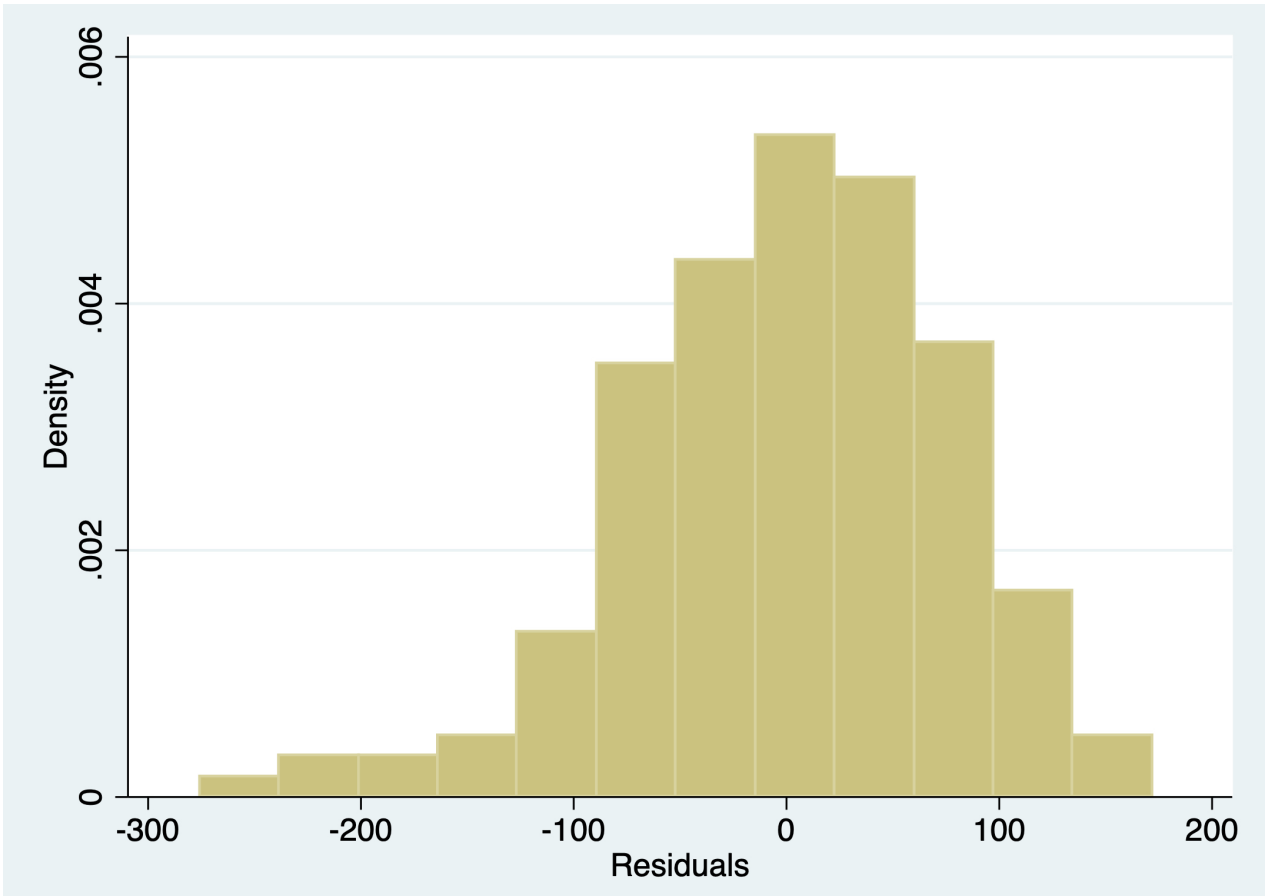
$H_0$: Residuals are normally distributed (S=0, K=3)

$H_1$: Residuals are not normally distributed

Let the level of significance be 1% ($\alpha = 0.01$)

$\chi^2_{cal} = \frac{n}{6}( S^2 + \frac{(K-3)^2}{4} ) = 12.7787 \sim \chi^2_2$ ; $\chi^2_{critical} = \chi^2_{2,0.01} = 9.2103$

Conclusion: Since $\chi^2_{cal} > \chi^2_{critical}$, we reject the null hypothesis at the 1% level of significance and conclude that the residuals are not normally distributed. At the 0.2% level of significance, however, we do not reject the null hypothesis ( $P(\chi^2_2 > 12.7787) = 0.002$ ). The negative skew can be seen in the graphical representation of residual terms-




*(Histogram plot of residuals)*

*4.d.ii. Jarque-Bera Test for Normality of Residuals (using $log_{10}$ S&P500 as the regressand)*

In an attempt to address the problem of non-normality of residuals, we tried to experiment with a different functional form for our model where we took log (with base 10) values of S&P 500 as the dependent variable. We let the corresponding residuals of the new model be denoted by $v_i$.

Using STATA, we obtain the following results:

| Residuals | | | | |
|---|---|---|---|---|
| | Percentiles | Smallest | | |
| 1% | -.0317668 | -.0384698 | | |
| 5% | -.0207435 | -.0317668 | | |
| 10% | -.0147351 | -.0316463 | Obs | 160 |
| 25% | -.0067759 | -.0280889 | Sum of Wgt. | 160 |
| 50% | .0015754 | | Mean | 3.53e-11 |
| | | Largest | Std. Dev. | .0114496 |
| 75% | .007984 | .0194062 | | |
| 90% | .0126332 | .0220527 | Variance | .0001311 |
| 95% | .0166908 | .0248764 | Skewness | -.5716162 |
| 99% | .0248764 | .028912 | Kurtosis | 3.657857 |

From the regression results we find that the residuals exhibit the following characteristics-

i) Skewness (S) = -0.5716162
ii) Kurtosis (K) = 3.657857

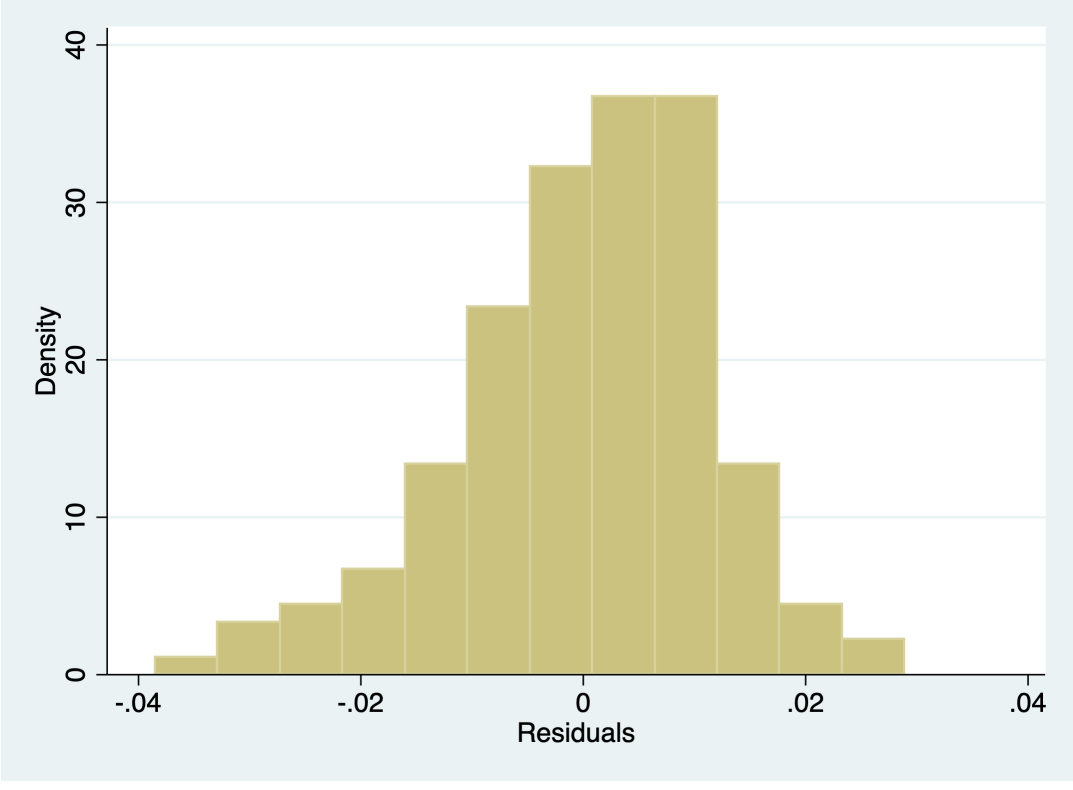
$H_0$: Residuals are normally distributed (S=0, K=3)

$H_1$: Residuals are not normally distributed

Let the level of significance be 1% ($\alpha = 0.01$)

$\chi^2_{cal} = \frac{n}{6}( S^2 + \frac{(K-3)^2}{4} ) = 11.5983 \sim \chi^2_2$ ; $\chi^2_{critical} = \chi^2_{2,0.01} = 9.2103$

Conclusion: Since $\chi^2_{cal} > \chi^2_{critical}$, we reject the null hypothesis at the 1% level of significance and conclude that the residuals are not normally distributed. At the 0.3% level of significance, however, we do not reject the null hypothesis ( $P(\chi^2_2 > 11.5983) = 0.003$). We conclude that our final model suffers from non-normal distribution of residual terms.




*(Histogram plot of residuals)*

As can be seen from the histogram plot, the logarithmic transformation of the dependent variable does improve the skewness and kurtosis of the residuals; this improvement, however, is not significant enough to not reject the null hypothesis of normality. Therefore, despite a different functional form for the regressand, the problem of non-normality of residuals still persists.

---

*SECTION VI – Sources*

1. https://covid19.healthdata.org/united-states-of-america?view=social-distancing&tab=trend
2. https://fred.stlouisfed.org/series/GOLDAMGBD228NLBM
3. https://in.investing.com/rates-bonds/u.s.-10-year-bond-yield-historical-data
4. https://fred.stlouisfed.org/series/DCOILWTICO
5. https://fred.stlouisfed.org/series/DCOILBRENTEU
6. https://finance.yahoo.com/quote/%5EGSPC/history?period1=1547942400&period2=1599955200&interval=1d&filter=history&frequency=1d&guccounter=1

---

*SECTION VII – Back Questions*

*Question 17.26*

Summary Table -

| Direction of Causality | Number of Lags | Calculated F value | Critical F Value (5%) | Decision |
|---|---|---|---|---|
| S -> P | 1 | 20.27 | 4.38 | Reject |
| S -> P | 2 | 10.99 | 3.63 | Reject |
| S -> P | 3 | 6.14 | 3.42 | Reject |
| S -> P | 4 | 3.36 | 3.48 | Do not reject |
| S -> P | 5 | 1.82 | 3.97 | Do not reject |
| S -> P | 6 | 1.16 | 6.16 | Do not reject |
| P -> S | 1 | 56.45 | 4.38 | Reject |
| P -> S | 2 | 34.26 | 3.63 | Reject |
| P -> S | 3 | 24.71 | 3.42 | Reject |
| P -> S | 4 | 13.42 | 3.48 | Reject |
| P -> S | 5 | 7.58 | 3.97 | Reject |
| P -> S | 6 | 6.03 | 6.16 | Do not reject |

$H_o$: Sales does not Granger cause Planned Expenditure and vice-versa.

Up to 3 lags, there is bilateral causality between Sales and Planned Expenditure.

Up to 5 lags, Planned Expenditure Granger causes Sales.

Beyond 3 lags Sales does not Granger cause Planned Expenditure.