Descriptive

The data variables used in the study are;

IR: Inflation rate of the United States form January 1980 to December 2021

GDP: Gross Domestic product of the United States form January 1980 to December 2021

CPI: The Consumer Price Index

UR: Unemployment rate

SP: S&P500 close price

MS: Money supply

The summary statistics were as shown below;

	Mean	Median	Minimum	Maximum
IR	3.2181	2.8000	-2.1000	14.800
GDP	99.910	100.00	91.705	101.85
l_MS	7.2176	7.0473	5.9501	9.9242
l_UR	1.7910	1.7579	1.2528	2.6878
l_CPI	5.0653	4.8963	4.3541	5.6305
l_SP	6.6325	6.9602	4.6259	8.4350

	Std. Dev.	C.V.	Skewness	Ex. kurtosis
IR	2.4612	0.76481	2.2968	7.1366
GDP	1.1886	0.011897	-1.9618	8.0772
l_MS	0.79874	0.11067	1.2357	2.3138
l_UR	0.27155	0.15162	0.25812	-0.51887
l_CPI	0.34477	0.068066	0.016724	-1.3794
l_SP	0.97663	0.14725	-0.37944	-0.88546

	5% perc.	95% perc.	IQ range	Missing obs.
IR	0.62500	8.2000	2.0750	0
GDP	97.740	101.41	1.1064	0
l_MS	6.0915	8.2900	0.89883	0
l_UR	1.3610	2.2772	0.39864	0
l_CPI	4.5470	5.5488	0.57264	0
l_SP	4.8667	7.9995	1.4193	0

From the descriptive table above, consumer price index recorded the least range of 1.2764 (max-min) showing that CPI was stable for that time period while inflation rate recorded the highest range showing that it had great fluctuations in the period.

The standard deviation which is the distance from the mean was highest for inflation rate and least for unemployment rate

Skewness shows the direction of the outliers. For positive skew, the tail of the distribution curve is longer on the right meaning that the outliers are further out towards the right and closer to the mean on the left. The same is opposite for negative skewness. From the analysis, inflation rate had the highest positive skewness while GDP has the highest negative skewness.

Kurtosis measures whether the data are heavy tailed or light tailed relative to the normal distribution. That is, the variables with high kurtosis tend to have outliers (IR , GDP, MS)while those with low kurtosis have no outliers (UR, CPI , SP)

The dataset recorded no missing observations.

Correlation

The correlation analysis was conducted to establish association between each pair of variables, the summary was as shown;

```
Correlation Coefficients, using the observations 1980:02 - 2021:12
5% critical value (two-tailed) = 0.0874 for n = 503

          IR          GDP          l_UR        d_l_MS         d_l_SP
      1.0000       0.1517        0.1303       -0.0701        -0.0616  IR
                   1.0000       -0.5992       -0.3632        -0.1055  GDP
                                 1.0000        0.1827         0.0859  l_UR
                                               1.0000         0.0348  d_l_MS
                                                              1.0000  d_l_SP

     d_l_CPI
      0.1041  IR
     -0.0234  GDP
      0.0113  l_UR
     -0.0137  d_l_MS
     -0.0388  d_l_SP
      1.0000  d_l_CPI
```

The study established a regression coefficient of .1041 between the variable d_I_CPI and IR, it follows that there exist a weak positive association between the consumer price index and the inflation rate. The variables UR also had a direct correlation with d_I_CPI.

There was a weak direct correlation between the variables GDP and IR, I_UR and IR, d_I_MS and I_UR then d_I_SP and d_I_MS.

The study established an inverse association between the variables; d_I_CPI and GDP, d_I_MS, d_I_SP. This was also the case for the variables d_I_SP and IR and correlation between d_I_SP and GDP.

In summary, the study established that there was at least an association between each of the pairs of variables in the study.

Stationarity

Since the data is a time series data, it would be necessary to test for stationarity, the study used the Augmented Dickey Fuller test to test this and the results were as follows;

```
Augmented Dickey-Fuller test for d_l_CPI
testing down from 17 lags, criterion AIC
sample size 502
unit-root null hypothesis: a = 1

  test with constant
  including 0 lags of (1-L)d_l_CPI
  model: (1-L)y = b0 + (a-1)*y(-1) + e
  estimated value of (a - 1): -0.953376
  test statistic: tau_c(1) = -21.355
  asymptotic p-value 1.562e-49
  1st-order autocorrelation coeff. for e: -0.002
Augmented Dickey-Fuller test for d_l_MS
testing down from 17 lags, criterion AIC
sample size 502
unit-root null hypothesis: a = 1

  test with constant
  including 0 lags of (1-L)d_l_MS
  model: (1-L)y = b0 + (a-1)*y(-1) + e
  estimated value of (a - 1): -0.896199
  test statistic: tau_c(1) = -20.1485
  asymptotic p-value 1.183e-47
  1st-order autocorrelation coeff. for e: -0.004
Augmented Dickey-Fuller test for d_l_SP
testing down from 17 lags, criterion AIC
sample size 502
unit-root null hypothesis: a = 1

  test with constant
  including 0 lags of (1-L)d_l_SP
  model: (1-L)y = b0 + (a-1)*y(-1) + e
  estimated value of (a - 1): -0.961662
  test statistic: tau_c(1) = -21.5235
  asymptotic p-value 9.296e-50
  1st-order autocorrelation coeff. for e: -0.000
Augmented Dickey-Fuller test for IR
testing down from 17 lags, criterion AIC
sample size 488
unit-root null hypothesis: a = 1

  test with constant
  including 15 lags of (1-L)IR
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0339142
  test statistic: tau_c(1) = -4.34867
  asymptotic p-value 0.0003603
  1st-order autocorrelation coeff. for e: -0.003
  lagged differences: F(15, 471) = 27.404 [0.0000]
```

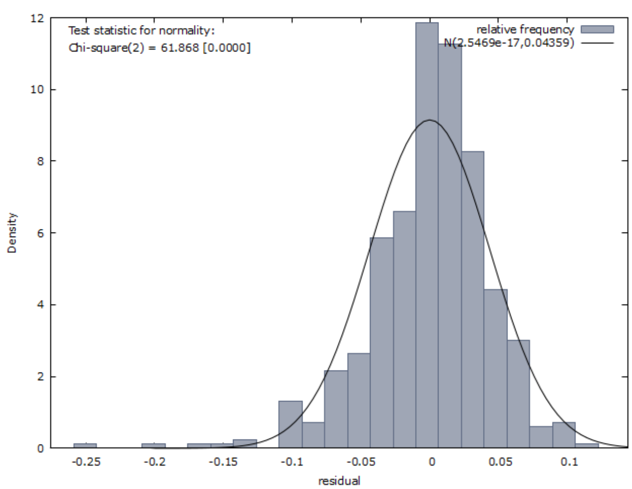
```
Augmented Dickey-Fuller test for GDP
testing down from 17 lags, criterion AIC
sample size 499
unit-root null hypothesis: a = 1

  test with constant
  including 4 lags of (1-L)GDP
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0451051
  test statistic: tau_c(1) = -5.05026
  asymptotic p-value 1.609e-05
  1st-order autocorrelation coeff. for e: 0.009
  lagged differences: F(4, 493) = 137.508 [0.0000]
Augmented Dickey-Fuller test for l_UR
testing down from 17 lags, criterion AIC
sample size 502
unit-root null hypothesis: a = 1

  test with constant
  including one lag of (1-L)l_UR
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0286527
  test statistic: tau_c(1) = -2.76582
  asymptotic p-value 0.06327
  1st-order autocorrelation coeff. for e: 0.005
```

The study established that all the variables were stationary, and those that were not stationary were differenced and became stationary on first differencing, the variables differenced were; I_MS, I-CPI, and I_SP.

Assumptions

Normality

The histogram below was used to test for the normality assumption.



The graph implied that the data is symetric, this indicates that the data is normally distributed. The assumption of normality has been met.

Autocorrelation

The study tested for the autocorrelation using the Durbin Watson test as shown;

```
Durbin-Watson statistic = 1.94844

H1: positive autocorrelation
   p-value = 0.237913
H1: negative autocorrelation
   p-value = 0.762087
```

A Durbin Watson test statistic value of 1.94844 being less than 2.0 indicates that there exist a positive autocorrelation, implying that the assumption has been met.

Autocorrelation function results

While some variables needed lagging to attain stationarity, before this, some variables like the CPI, Money Supply and the Unemployment rate were logged given they contained extreme values that may distort results of the OLS regression, this is the summarizing function of number of lags that was required for each variable to attain stationarity;

Variable:	**Number of Lags**
Inflation	0
Ln Unemployment Rate	0
GDP	0
Ln S&P500	1
Ln_Money Supply	1
Ln CPI	1

After lagging was done, Breusch-Godfreys test was used to assess whether there exist a serial correlation and the results were as shown below;

```
Breusch-Godfrey test for first-order autocorrelation
OLS, using observations 1980:02-2021:12 (T = 503)
Dependent variable: uhat

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const       0.00567158    0.235318       0.02410   0.9808
  IR         -1.34907e-05   0.000854636   -0.01579   0.9874
  GDP        -5.85871e-05   0.00225562    -0.02597   0.9793
  d_l_CPI     0.00415019    0.136543       0.03039   0.9758
  d_l_MS     -0.00211860    0.0380918     -0.05562   0.9557
  l_UR        0.000129072   0.00933544     0.01383   0.9890
  uhat_1      0.0261614     0.0452351      0.5783    0.5633

  Unadjusted R-squared = 0.000674

Test statistic: LMF = 0.334480,
with p-value = P(F(1,496) > 0.33448) = 0.563

Alternative statistic: TR^2 = 0.338972,
with p-value = P(Chi-square(1) > 0.338972) = 0.56

Ljung-Box Q' = 0.335757,
with p-value = P(Chi-square(1) > 0.335757) = 0.562
```

Breusch-Godfrey tests assess the null hypothesis that the there is no serial correlation of any order up to p, since the p-value of the test is .563, a value greater than the 1% level of significance, we fail to reject the null hypothesis and conclude that there does not exist a serial correlation up to order p, this implies that the problem of autocorrelation does not exist.

Therefore, the general autocorrelation model is;

$$S\&P500_t = \alpha + \beta_1 IR + \beta_2 GDP + \beta_3 UR + \beta_4 CPI_t + \beta_5 MS_t + \epsilon$$

Linearity

This assumption is tested to ensure that the variables are not linearly related, the variance inflation factor was used to test this assumption and the resullt were as follows;

```
Variance Inflation Factors
Minimum possible value = 1.0
Values > 10.0 may indicate a collinearity problem

      d_1_MS    1.155
     d_1_CPI    1.014
         GDP    1.889
        l_UR    1.698
          IR    1.125

VIF(j) = 1/(1 - R(j)^2), where R(j) is the multiple correlation coefficient
between variable j and the other independent variables

Belsley-Kuh-Welsch collinearity diagnostics:

  variance proportions

  lambda    cond   const  d_1_MS d_1_CPI     GDP    l_UR      IR
   3.767   1.000   0.000   0.002   0.004   0.000   0.001   0.017
   0.996   1.944   0.000   0.599   0.275   0.000   0.000   0.003
   0.939   2.003   0.000   0.254   0.706   0.000   0.000   0.001
   0.283   3.651   0.000   0.017   0.011   0.000   0.003   0.896
   0.015  15.758   0.001   0.035   0.000   0.001   0.590   0.006
   0.000 324.900   0.999   0.093   0.003   0.999   0.406   0.077

  lambda = eigenvalues of inverse covariance matrix (smallest is 3.56843e-005)
  cond   = condition index
  note: variance proportions columns sum to 1.0

According to BKW, cond >= 30 indicates "strong" near linear dependence,
and cond between 10 and 30 "moderately strong".  Parameter estimates whose
variance is mostly associated with problematic cond values may themselves
be considered problematic.

Count of condition indices >= 30: 1
Variance proportions >= 0.5 associated with cond >= 30:

   const     GDP
   0.999   0.999

Count of condition indices >= 10: 2
Variance proportions >= 0.5 associated with cond >= 10:

   const     GDP    l_UR
   1.000   1.000   0.996
```

Since the VIF values are all below 10, it indicates that the problem of linearity does not exist therefore the assumption has been met.

Heteroscedasticity

The problem of constant variance is another that needs to be addressed, the study used Breusch-Pagan test to assess this assumption and the result was as follows;

```
Model 4: OLS, using observations 1980:02-2021:12 (T = 503)
Dependent variable: d_l_SP

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const        0.241828      0.234956       1.029     0.3039
  d_l_MS      -0.00298904    0.0378899     -0.07889   0.9372
  d_l_CPI     -0.107375      0.136263      -0.7880    0.4311
  GDP         -0.00246820    0.00225184    -1.096     0.2736
  l_UR         0.00869195    0.00932652     0.9320    0.3518
  IR          -0.000996875   0.000853745   -1.168     0.2435

Mean dependent var   0.007322   S.D. dependent var   0.043731
Sum squared resid    0.944341   S.E. of regression   0.043590
R-squared            0.016350   Adjusted R-squared   0.006454
F(5, 497)            1.652210   P-value(F)           0.144633
Log-likelihood       865.1552   Akaike criterion    -1718.310
Schwarz criterion   -1692.987   Hannan-Quinn        -1708.376
rho                  0.025760   Durbin-Watson        1.948441

Excluding the constant, p-value was highest for variable 17 (d_l_MS)

Breusch-Pagan test for heteroskedasticity -
  Null hypothesis: heteroskedasticity not present
  Test statistic: LM = 12.398
  with p-value = P(Chi-square(5) > 12.398) = 0.0297225
```

The p-value of the test was established to be .0297, a value greater than the 1% level of significance, we therefore fail to reject the null hypothesis that the data is homoscedastic and conclude that the assumption has been met.

Cointegration test

This test identifies scenarios where two or more non-stationary time series variables are integrated together in a way they cannot deviate from the equilibrium, it is used to identify the degree of sensitivity.

This test carried the following stationarity tests using the Augmented Dickey-Fuller test to assess stationarity of the variables and the result was as follows;

```
Step 1: testing for a unit root in IR

Augmented Dickey-Fuller test for IR
including 7 lags of (1-L)IR
sample size 495
unit-root null hypothesis: a = 1

  test with constant
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0357674
  test statistic: tau_c(1) = -4.58746
  asymptotic p-value 0.0001
  1st-order autocorrelation coeff. for e: 0.000
  lagged differences: F(7, 486) = 18.082 [0.0000]

Step 2: testing for a unit root in GDP

Augmented Dickey-Fuller test for GDP
including 7 lags of (1-L)GDP
sample size 495
unit-root null hypothesis: a = 1

  test with constant
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0456867
  test statistic: tau_c(1) = -4.73703
  asymptotic p-value 6.829e-05
  1st-order autocorrelation coeff. for e: -0.001
  lagged differences: F(7, 486) = 77.424 [0.0000]

Step 3: testing for a unit root in l_UR

Augmented Dickey-Fuller test for l_UR
including 7 lags of (1-L)l_UR
sample size 495
unit-root null hypothesis: a = 1

  test with constant
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.0264315
  test statistic: tau_c(1) = -2.42026
  asymptotic p-value 0.1361
  1st-order autocorrelation coeff. for e: 0.000
  lagged differences: F(7, 486) = 1.829 [0.0797]
```

```
  asymptotic p-value 0.1361
  1st-order autocorrelation coeff. for e: 0.000
  lagged differences: F(7, 486) = 1.829 [0.0797]

Step 4: testing for a unit root in d_l_MS

Augmented Dickey-Fuller test for d_l_MS
including 7 lags of (1-L)d_l_MS
sample size 495
unit-root null hypothesis: a = 1

  test with constant
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.826381
  test statistic: tau_c(1) = -7.35609
  asymptotic p-value 3.484e-11
  1st-order autocorrelation coeff. for e: -0.000
  lagged differences: F(7, 486) = 0.164 [0.9919]

Step 5: testing for a unit root in d_l_SP

Augmented Dickey-Fuller test for d_l_SP
including 7 lags of (1-L)d_l_SP
sample size 495
unit-root null hypothesis: a = 1

  test with constant
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.921441
  test statistic: tau_c(1) = -7.46722
  asymptotic p-value 1.7e-11
  1st-order autocorrelation coeff. for e: -0.001
  lagged differences: F(7, 486) = 0.912 [0.4964]

Step 6: testing for a unit root in d_l_CPI

Augmented Dickey-Fuller test for d_l_CPI
including 7 lags of (1-L)d_l_CPI
sample size 495
unit-root null hypothesis: a = 1

  test with constant
  model: (1-L)y = b0 + (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.992424
  test statistic: tau_c(1) = -7.96076
  asymptotic p-value 6.512e-13
  1st-order autocorrelation coeff. for e: -0.000
  lagged differences: F(7, 486) = 0.070 [0.9995]
```

Since all the p-values were less than 1% level of significance, the study can conclude that all the variables were stationary.

Cointegrating regression analysis

The summarized regression analysis resulting from cointegration was as follows;

```
Step 7: cointegrating regression

Cointegrating regression -
OLS, using observations 1980:02-2021:12 (T = 503)
Dependent variable: d_l_SP

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const       0.241828      0.234956       1.029     0.3039
  IR         -0.000996875   0.000853745   -1.168     0.2435
  GDP        -0.00246820    0.00225184    -1.096     0.2736
  l_UR        0.00869195    0.00932652     0.9320    0.3518
  d_l_CPI    -0.107375      0.136263      -0.7880    0.4311
  d_l_MS     -0.00298904    0.0378899     -0.07889   0.9372

Mean dependent var   0.007322   S.D. dependent var   0.043731
Sum squared resid    0.944341   S.E. of regression   0.043590
R-squared            0.016350   Adjusted R-squared   0.006454
Log-likelihood       865.1552   Akaike criterion    -1718.310
Schwarz criterion   -1692.987   Hannan-Quinn        -1708.376
rho                  0.025760   Durbin-Watson        1.948441

Step 8: testing for a unit root in uhat

Augmented Dickey-Fuller test for uhat
including 12 lags of (1-L)uhat
sample size 490
unit-root null hypothesis: a = 1

  test without constant
  model: (1-L)y = (a-1)*y(-1) + ... + e
  estimated value of (a - 1): -0.932803
  test statistic: tau_c(6) = -5.91614
  asymptotic p-value 0.0008026
  1st-order autocorrelation coeff. for e: -0.001
  lagged differences: F(12, 477) = 0.789 [0.6618]

There is evidence for a cointegrating relationship if:
(a) The unit-root hypothesis is not rejected for the individual variables, and
(b) the unit-root hypothesis is rejected for the residuals (uhat) from the
    cointegrating regression.
```

From the cointegretion regression findings above, the integration coefficient of S&P500 and inflation rate is -0.000996875 indicating a negative association between the variables. Hence a unit change in inflation rate would lead to a decrease in S&P500 by 0.000996875 units.

The integration coefficient of S&P500 and GDP is -0.0024682 indicating a negative association between the variables. Hence a unit change in GDP would lead to a decrease in S&P500 by 0.0024682 units.

The integration coefficient of S&P500 and unemployment rate is 0.00869195 indicating a positive association between the variables. Hence a unit change in unemployment rate would lead to an increase in S&P500 by 0.00869195 units.

The integration coefficient of S&P500 and consumer price index is -0.107375 indicating a negative association between the variables. Hence a unit change in consumer price index would lead to a decrease in S&P500 by 0.107375 units.

The integration coefficient of S&P500 and money supply is -0.00298904 indicating a negative association between the variables. Hence a unit change in money supply would lead to a decrease in S&P500 by 0.00298904 units.

The regression constant is 0.241828 indicating that S&P500 would be at 24.1828% if all the independent variables were zero.

The value of the adjusted R^2 (0.006454) indicates that the model explains 0.65% influence on S&P500.

The regression equation is,

$$S\&P500_t = 0.2418 - 0.000997IR_t - 0.00247GDP_t + 0.00869UR_t - 0.1074CPI_t \\ - 0.002989MS_t + \epsilon$$

Conclusion

Evidence of cointegration

H_0: Unit root is present, indicating residuals are non-stationary

H_1: No unit root present, indicating residuals are stationary.

From the above table: for all the independent variables t-statistic is less than the p-values, we reject H_0 and conclude that there isn't enough evidence to show that unit root is present.

Since there is proof of the residuals being stationary, relationship between the variables in the model is evident in the long run.

Error correction model results

```
Model 1: OLS, using observations 1980:02-2021:12 (T = 503)
Dependent variable: d_l_SP

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const       0.241828      0.234956       1.029     0.3039
  IR         -0.000996875   0.000853745   -1.168     0.2435
  GDP        -0.00246820    0.00225184    -1.096     0.2736
  l_UR        0.00869195    0.00932652     0.9320    0.3518
  d_l_MS     -0.00298904    0.0378899     -0.07889   0.9372
  d_l_CPI    -0.107375      0.136263      -0.7880    0.4311

Mean dependent var   0.007322   S.D. dependent var   0.043731
Sum squared resid    0.944341   S.E. of regression   0.043590
R-squared            0.016350   Adjusted R-squared   0.006454
F(5, 497)            1.652210   P-value(F)           0.144633
Log-likelihood       865.1552   Akaike criterion    -1718.310
Schwarz criterion   -1692.987   Hannan-Quinn        -1708.376
rho                  0.025760   Durbin-Watson        1.948441

Excluding the constant, p-value was highest for variable 14 (d_l_MS)
```

From the table above, the correlation coefficient of inflation rate and S&P500 is -0.000997 indicating a very week negative correlation between the two variables. Sam case applies to GDP and the dependent variable having a correlation coefficient of -0.00246820

Unemployment rate however has a positive correlation with S&P500, hence a change in unemployment rate affects S&P500 by 0.869%

Narrow money supply and consumer price index have negative correlations with the dependent variables of -0.00298904 and -0.107375. This shows that narrow money supply and S&P500 move in different directions as well as consumer price index and S&P500.

In conclusion, since the correlation coefficients are more close to 0 than one, we notice a very week association between each of the independent variables and S&P500.

The value of the adjusted R^2 (0.006454) indicates that, in a general contest, S&P500 is affected by only 0.65% of the said independent variables. We can conclude therefore that there is no significant relationship between S&P500 and each of the independent variables. The final regression model is;

$$S\&P500_t = 0.2418 - 0.000997IR_t - 0.00247GDP_t + 0.00869UR_t - 0.1074CPI_t - 0.002989MS_t + \epsilon$$

Ramsey RESET test

```
Auxiliary regression for RESET specification test
OLS, using observations 1980:02-2021:12 (T = 503)
Dependent variable: d_l_SP
```

	coefficient	std. error	t-ratio	p-value
const	-0.123515	0.335398	-0.3683	0.7128
d_l_MS	-0.0273962	0.0410872	-0.6668	0.5052
d_l_CPI	-0.149508	0.138860	-1.077	0.2821
GDP	0.00126321	0.00332393	0.3800	0.7041
l_UR	0.000416443	0.0107806	0.03863	0.9692
IR	-0.000242019	0.000985967	-0.2455	0.8062
yhat^2	61.9737	40.6528	1.524	0.1280

```
Test statistic: F = 2.323983,
with p-value = P(F(1,496) > 2.32398) = 0.128
```

The p-value of our F-statistic is 0.128 (12.8%). At 5% level of significance, we fail to reject the null hypothesis and conclude that the functional form is correct and our model does not suffer from omitted variables.

It is therefore clear that the model is reliable.