Data Analysis for Global Business

ASSESSMENT 002 COURSEWORK

ID: 21040480

Module Code: FE7066

Tutor: Dr. Lan Jiang

Module Title: Data Analysis for Global Business

School: Guildhall School of Business & Law

Date of Submission: 6-May-2022

Table of Contents

[Introduction 4](#_Toc109348704)

[Question No. 1 4](#_Toc109348705)

[Part. A 4](#_Toc109348706)

[Testing of Hypothesis. 5](#_Toc109348707)

[Level of significance. 5](#_Toc109348708)

[Test statistics for this test. 5](#_Toc109348709)

[Probability value or p-value. 5](#_Toc109348710)

[Decision 6](#_Toc109348711)

[Part. B 6](#_Toc109348712)

[Part. C 7](#_Toc109348713)

[Part. D 8](#_Toc109348714)

[Testing of Hypothesis. 9](#_Toc109348715)

[Level of significance. 9](#_Toc109348716)

[Test statistics for this test. 9](#_Toc109348717)

[Probability value or p-value. 10](#_Toc109348718)

[Decision. 10](#_Toc109348719)

[Part. E 10](#_Toc109348720)

[Testing of Hypothesis. 11](#_Toc109348721)

[Test statistics for DW. 11](#_Toc109348722)

[Part. F 13](#_Toc109348723)

[Part. G 14](#_Toc109348724)

[Part. H 14](#_Toc109348725)

[Part. I 16](#_Toc109348726)

[Question No. 2 16](#_Toc109348727)

[Part. A 17](#_Toc109348728)

[Part. D 20](#_Toc109348729)

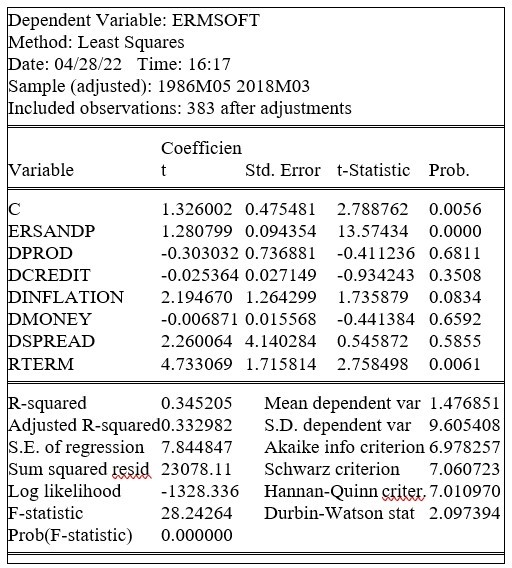
[Conclusion 21](#_Toc109348730)

# Introduction

It was necessary to study regression model that aim to assess if the expected data on Microsoft stock could be described by referring to unforeseen situations in a collection of economic and financial factors in the sense of Arbitrage Pricing Theory (APT). It was also necessary to gather the weekly closing marketplace stock value of Compass Group Plc (CPG) and the FTSE 100 Stock Index (UKX) from Friday 2nd March 2012 to Friday 4th March 2022 in order to run a cointegration test.

# Question No. 1

## Part. A

The regression analysis approach is being used to calculate the variation in the dependent variable as a result of the independent factors. When there is just single linear independent factor, regression is referred to as simple linear regression. Regression research is a quantifiable approach for determining the relationship among at minimum two variables. Typically, the researcher seeks the signiﬁcant impact of one factor upon other, such as the effect of a price rise on demand or the effects of technological in the money stock on the annual increase. A multiple regression analysis is one that comprises several independent variables. The OLS regression outcome and result table are shown below:

The Shown above table illustrates the least square approximation for each factor, whereas the regression results reveal that DPROD, DCREDIT, and DMONEY have such an unfavorable correlation with the dependent variable, whilst remaining variables have a positive relationship.

* The simple regression figure showed that the factors RTERM, DINFLATION, and ERSANDP have quite a significant influence on the predictor variable since their P-value is smaller than alpha, which itself is 10%. The remaining factors all seem to be non-significant.
* The F-statistic for the given model yields a p-value of 0.00, indicating that the combined impact of all factors on the dependent variable is substantial. We say in the null hypothesis of regression

### Testing of Hypothesis.

***Ho: b1=b2=b3…. =0 H1: b1=b2=b3…. ≠0***

Here b1 is coefficient of first variable and b2 is coefficient of second variable and so on.

### Level of significance.

Here, al is the re-specified magnitude of alpha, which would be presumed to be 0.05. This suggests that a 5% margin of error is permissible during analysis.

### Test statistics for this test.

***F- statistics = MST***

***MSE***

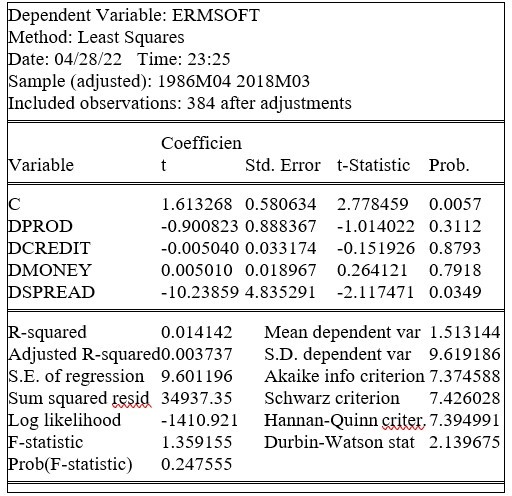
### Probability value or p-value.

According to thumb rule if our P-value is less than alpha which is pre-specified value we reject Ho otherwise don’t reject Ho.

### Decision

In the above scenario, our probability value is 0.00, this is less than 0.05, therefore Ho is discarded, and it is determined that all independent factors have a positive influence on the dependent variables. It informs us about the total combined influence of independent factors on the dependent variable.

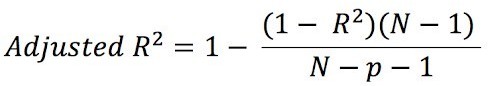
## Part. B



Because the P-value of F-statistics is bigger than just the criterion of importance (5 percent), we do not reject Ho instead infer that all four factors have had no meaningful influence on dependent variables. There are several constraints throughout assessment, such as when using dummy variables. E-Views demonstrate complete collinearity.

## Part. C

* The R square concept recognizes how well our independent factors matched the response variable. If R - squared is 0.6, the connected take into account the needs 60% of the variance in the dependent variable. In a nutshell, R square informs us how much of the variance in dependent variables is explained by variation in a linear regression model. It is used to assess a model's performance. A higher square value indicates that the model is doing better. A model is favored if it has the highest R square value.
* When we add factors to our paradigm to increase the variance of the dependent variable, it is conceivable that unnecessary variables are added to the model, which may cause model specification errors. In the situation of model specification, the number of R - squared would be increased regardless of how relevant these factors are for the dependent variable. For example, adding an unimportant variable to a model should not raise R square conceptually, but it does in practice. To overcome this situation, the notion of modified R square is proposed. It lessens the influence of the extraneous parameter R - squared. Likewise, when numerous correlated factors are included in the model, the value of R - squared is boosted from its current value. To minimize missed specification errors in a model, modified R square is utilized, or an irrelevant variable is omitted. Multi-collinearity and irrelevant factors have no effect on adjusted R square. The updated R square formula is shown below.

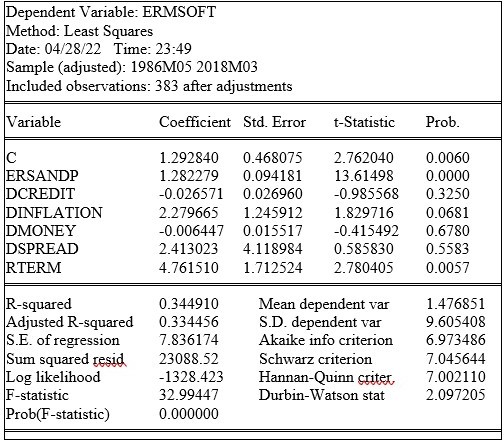


***N= Total sample size,***

***P = Total number of independent variables,***

***R2 = Sample R square.***

* When DPROD variable is dropped from regression model our results are.



The finding suggests that the adjusted R - squared is 0.334, whereas the adjusted R - squared in the initial formulation from component (a) is 0.332. It signifies that the value of the new model has increased somewhat. As a result, there is no need to add the DPROD variable in Model. Because it might lead to misspecification and computation complications.

## Part. D

The White test (1980) is used to determine the heteroscedasticity of a residual term. We assert that there really is no Heteroscedasticity versus Heteroscedasticity inside the null hypothesis. To pass judgment, the test statistics are calculated first, followed by the p-value. For Heteroscedasticity, for example, we may formulate null hypotheses and variant hypotheses as shown below.

### Testing of Hypothesis.

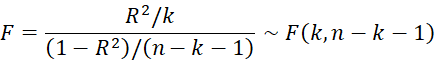
#### Ho: there is no Heteroscedasticity in given data

#### H1: there is Heteroscedasticity in given data

### Level of significance.

Here alpha is pre-specified value which is assumed as 0.05. it means 5% error is acceptable during analysis.

### Test statistics for this test.

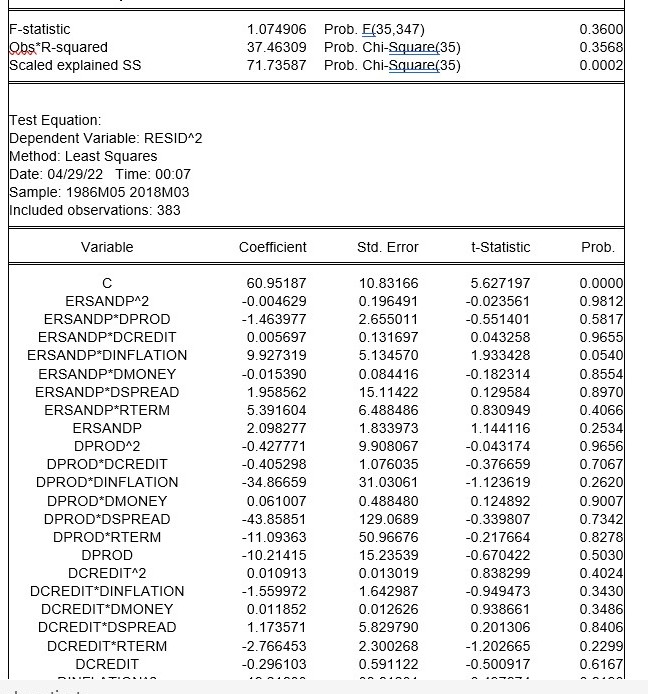


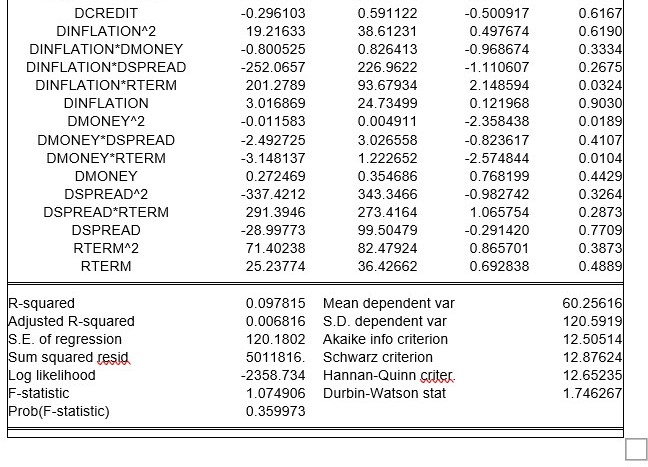
***n= Total sample size,***

***k = Total number of independent variables,***

***R2 = Sample R square.***

Result of white test are given below.





### Probability value or p-value.

If the P-value is smaller than alpha, a pre-specified number, we refuse Ho; alternatively, we do not reject Ho.

### Decision.

In the preceding scenario, our significance level is 0.36, which is more than 0.05, and we conclude that there really is no heteroscedasticity in the supplied data. It signifies that the variance of the error term is constant across all stages and therefore does not fluctuate with time.

## Part. E

The Durbin Watson statistics has been used to assess sequence autocorrelation. It is typically used in regression analysis to assess autocorrelation among standardized residuals. The Durbin-Watson statistic has always had a range of 0 to 4. If it is 2, it indicates that there really is no autocorrelation in the standardized residuals in our sample. Similarly, numbers from 0 to less than 2 points showed promising autocorrelation, whereas values from 2 to 4 indicate negative autocorrelation between series values. Autocorrelation should be about 2. The following are the basic assumptions for using this test: The error term in regression analysis should have a normal distribution with a mean of zero. The second requirement is that these mistakes be stationary.

In Durbin-Watson test our hypothesis testing steps are given below.

### Testing of Hypothesis.

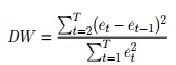
In null hypothesis of regression, we say that,

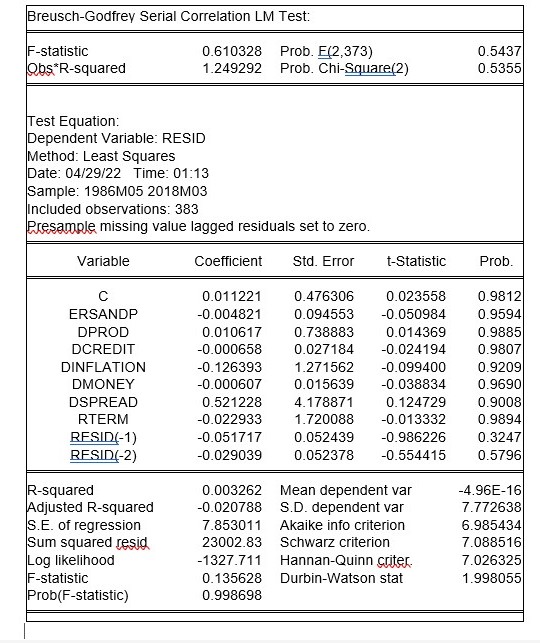
Ho: Residual terms in regression analysis are uncorrelated H1: Residual terms in regression analysis are not uncorrelated

**Level of significance**.

Value for level of significance for above test is fixed which is 5%

### Test statistics for DW.

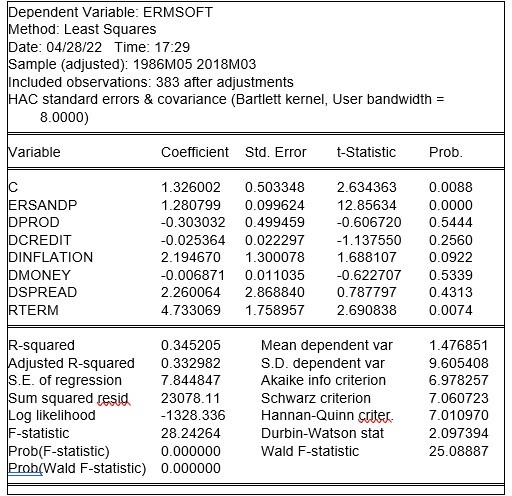




Because the above findings reveal that Durbin Watson =2 as well as the p-value is 0.59, which itself is larger than 0.05, we don't really discard Ho and infer that there really is no autocorrelation in the regression residual variance.

## Part. F

Result for HAC is given below.



According to the aforementioned data, three variables are important. These would be the factors ERSANDP, DINFLATION, and RTERM. These same variables are statistically significant inside the initial formulation, hence there is no change when compared to the original model.

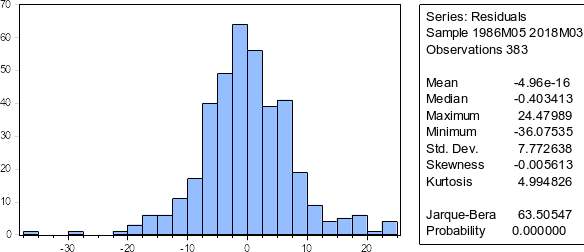
## Part. G

The multicollinearity of independent factors is a connection between variables that is examined using several methods. Variance influence factor (VIF) and correlation matrix are two popular approaches for determining multi-collinearity. To test multi-collinearity, we look for correlation between all independent variables.

Considering 0.80 as a criterion, three regressors are significantly linked with those around. The correlation between RTERM and DCREDIT in the preceding table is 0.93, indicating a substantial link. These variables are highly connected according to the supplied criteria. Similarly, the correlation between RTERM and DMONEY is 0.94, indicating a substantial relationship. Because the correlation between both variables is larger than 0.80, these two variables are highly linked, and multi-collinearity exist here. To eliminate the multi-collinearity, we must delete one of the linked variables. The correlation among RTERM and DSPRED is 0.73, which is significant but not great according to the specified parameters. All other factors' correlation is less than 0.80, indicating that they are not highly connected.

## Part. H

Result of normality test is given below.



The statistics in the histogram is indeed average. Because of its bell form. Because the p-value of the Jarque-Bera test is smaller than zero, we reject Ho and infer that the data is routinely dispersed.

The following code for mock is often used to produce dummy variables in E-Views. For instance, if we really want to create a fake variable for dmoney, we will use the code below.

***Smpl if dmoney > 0***

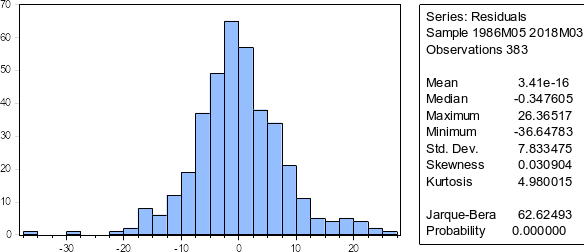
**Series dummy1= 1 Smpl if dmoney < = 0 Series dummy1= 0**

Here,

dmoney is the name of variable. Other dummy variable can be generated using above formula.

**Smpl if variable2 > 0 Series dummy2= 1 Smpl if variable2 < = 0 Series dummy2= 0**

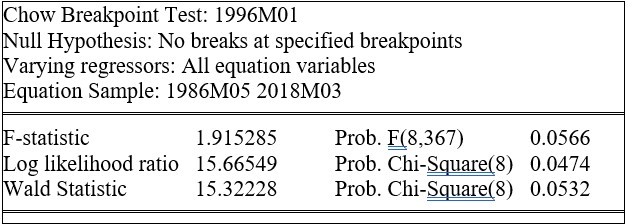
After adding two dummy variables in original model our output is given below.



Result of Jarque-Bera test and histogram shows that data is still normally distributed.

## Part. I

Result for Chow test is given below.



The F-statistics for the chow test are 1.91 and the significant level is 0.056, which is larger than 0.05, therefore we do not reject Ho and infer that there have been no breakdowns at the indicated endpoints.

# Question No. 2

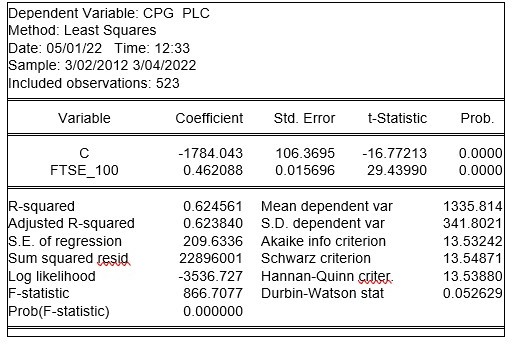
The basic graph of such series is shown below after generating data for weekly closing market share prices of Compass Group Plc (CPG) and the FTSE 100 Stock Index (UKX) from Friday 2nd March 2012 to Friday 4th March 2022.

## Part. A

The graphs above depict the pattern and variability for the both sets of data. Both sets demonstrate an overall rising trend. Both statistics show a fall after 2019, followed by a steady increase.

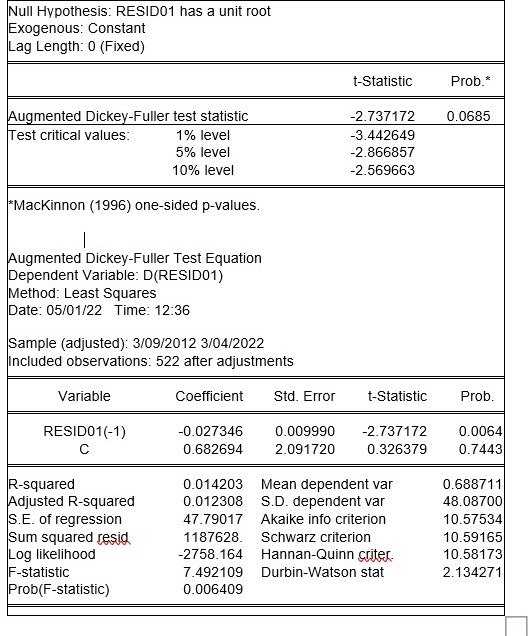
**Part. C**

As we have proven, both the FTSE 100 and the CPG PLC are I (1). So, in the following step, we may run the Engle-Granger co-integration test. For this test, we first ran a regression analysis among CPG PLC and FTSE 100, and the results are shown below.

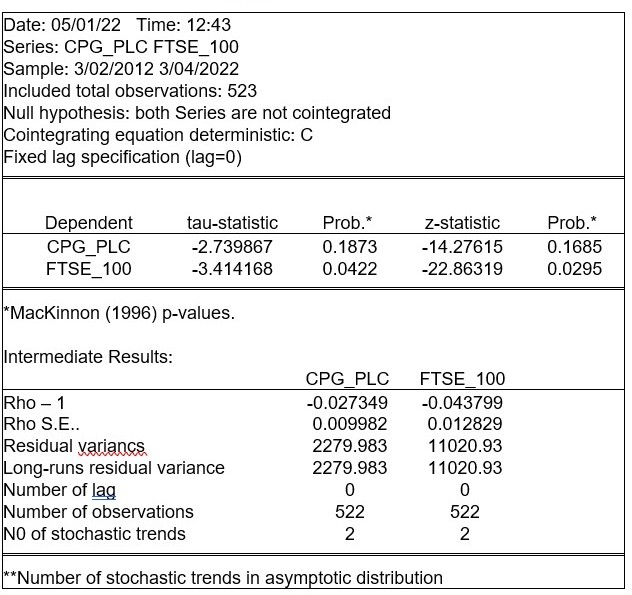


The co-integration research is displayed above in this table, and the result demonstrates that FTSE has a substantial influence on the dependent variable since its p-value is smaller than alpha, which in this case is 0.05. The residue factor "resid01" is generated just after regression analysis. After generating the residue term, the Augmented Dickey-Fuller test is used to determine whether or not data is normal. The remaining for this test is provided below.

The Dickey Fuller test is employed here to determine whether or not our leftover component is stable. Because the calculated t value is -2.868 and the p-value is 0.06, we cannot reject the null hypothesis. Ho comes to the conclusion that there really is no co-integration among these two series.



So, this was just the procedure for testing co-integration through using leftover term. E-Views also offers the Engle-Granger co-integration test with a single click. Click on data to open it in a new window, then click View\ Co-integration Test > Single-Equation Co-integration Test, which yields the findings shown below

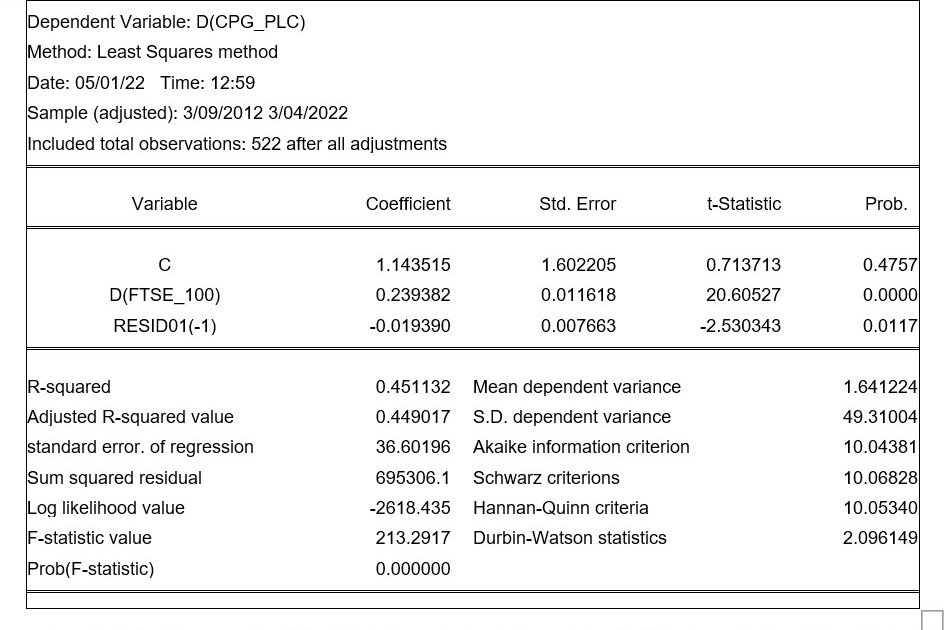


The probability estimates for 2 test statistics are shown. The first is the Engle-Granger tau-statistic, also known as the t-statistic, while the second is the normalized autocorrelation coefficient, often known as the Z-statistic. The outcomes for the both metrics are shown below. When CPG PLC is the dependent variable, the tau-statistic value is -2.73 and the Z-statistic value is -14.2. Both test statistics have probability values less than 0.05, indicating that we cannot reject the null hypothesis of no co-integration. As a result, we may conclude that the evidence from the preceding data strongly suggests that the CPG PLC and FTSE \_100 are not co-integrated.

## Part. D

ECM belongs to a class of temporal arrangement models that are most typically used for information and feature a lengthy universal stochastic drift, commonly referred as cointegration. ECMs are a conceivably technique that may be used to analyze the short- and long-term effects of a one-time configuration on another. The phrase error-correction refers to the fact that the past period's divergence from a lengthy balance, the inaccuracy, has an effect on its short-run components. ECMs may therefore directly assess the rate where a subordinated variable recovers to balance following a change in other variables.

We are now employing Error Correction Models (ECMs), and the results for this model are shown below. Below result are obtained from this model, D (CPG\_PLC) c D (FTSE\_100) RESID01(-1).



The following table depicts the co-integration study, and the corrected R-squared score is 44 percent, indicating that the regression coefficient accounted 44 percent of the variance in the dependent variables. Similarly, both independent quantities have p-values less than 0.05, indicating that both factors have a substantial influence on the dependent variable.

## **Conclusion**

This can be concluded from the above procedure that the independent variable has values less than 0.05 that is five percent. The R squared score and the p-value both have significant impact on the dependent variable.