Cointegrated Processes

Financial Econometrics

Saransh Gupta   Harsh Gupta   Yash Gupta

Contents

1 Abstract **2**

2 Introduction **2**
2.1 Examples . . . 2

3 Applying Cointegration **3**
3.1 Preparing for Cointegration . . . 3
3.2 ADF Test . . . 4
3.3 The Error Correction Model (ECM) . . . 4

4 Data Analysis **5**
4.1 Data Description . . . 5
4.2 Data Preprocessing (Gold and Silver) . . . 5
4.3 ADF Test on Gold Series (with Trend and Drift) . . . 6
4.4 ADF Test on First Difference of Gold Series . . . 7
4.5 ADF Test on Silver Series (with Trend and Drift) . . . 7
4.6 ADF Test on First Difference of Silver Series . . . 8
4.7 Plotting Residuals . . . 9
4.8 Testing stastionarity of residuals . . . 10
4.9 ECM Model Evaluation . . . 11

5 Data Analysis : Gold and Copper **11**
5.1 Data Preprocessing . . . 11
5.2 ADF Test on Copper Series (with Trend and Drift) . . . 12
5.3 ADF Test on First Difference of Copper Series . . . 13
5.4 Plotting Residuals . . . 13
5.5 Testing stastionarity of residuals . . . 14

6 References **15**

1 Abstract

Most macroeconomic and many financial variables are non-stationary. To study long term linear equilibrium between two or more time series, we use the concept of cointegration. It is exhibited in various fields like finance, economics, medicine, etc., where a linear combination of unit root processes is rendered stationary. In our work, we demonstrate the application of cointegrated processes between bivariate series (using Johansen's Approach) using daily closing price of metals: Gold, Silver and Copper.

2 Introduction

Most macroeconomic and many financial variables are non-stationary. They drift upwards over time and often exhibit characteristics, which suggest that they have a stochastic trend. Economic theory suggests that many of these variables will move together, fluctuating around a long-run equilibrium. In econometrics and statistics, this long-run equilibrium is tested and measured using the concept of cointegration.

Cointegration occurs when two or more nonstationary time series exhibit a long run equilibrium and move together in such a way that their linear combination results in a stationary time series. This signifies that they share an underlying common stochastic trend.

An $(n \times 1)$ vector time series $\boldsymbol{y}_t$, is said to be cointegrated if each of the series $y_{i,t}$ taken individually is $I(1)$, that is, non-stationary with a unit root, while some linear combination of the series $\boldsymbol{a}^T \boldsymbol{y}_t$, is stationary, or $I(0)$, for some nonzero $n \times 1$ vector $\boldsymbol{a}$. In the context of cointegration, $\boldsymbol{a}$ is called the cointegrating vector dictates how cointegrating series are combined, and it does not have to unique.

Cointegration means that although many developments can cause permanent changes in the individual elements of y„ there is some long-run equilibrium relation tying the individual components together, represented by the linear combination $\boldsymbol{a}^T \boldsymbol{y}_t$. If $\boldsymbol{y}_t$ is cointegrated, , then it is not correct to fit a vector auto-regression to the differenced data.

2.1 Examples

An example of such a system is the model of consumption spending proposed by Davidson, Hendry, Srba, and Yeo (1978). Their results suggest that although both consumption and income exhibit a unit root, over the long run consumption tends to be a roughly constant proportion of income, so that the difference between the log of consumption and the log of income appears to be a stationary process.

Another example of an economic hypothesis that lends itself naturally to a cointegration interpretation is the theory of purchasing power parity. This theory holds that, apart from transportation costs, goods should sell for the same effective price in two countries. Let P, denote an index of the price level in the United States in dollars per good), P; a price index for Italy (in lire per good), and S, the rate of exchange between the currencies (in dollars per lira). Then purchasing power parity holds that $P_t = S_t P_t^*$ In practice, errors in measuring prices, transportation costs, and differences in quality prevent purchasing · power parity from holding exactly at every date t.

What is an Example of Cointegrated Time Series?

Field	Supporting Theory	Time Series
Economics	The permanent income hypothesis describes how agents spread their consumption out over their lifetime based on their expected income.	Consumption and income.
Economics	Purchasing power parity is a theory that relates the prices of a basket of goods across different countries.	Nominal exchange rates and domestic and foreign prices.
Finance	The present value model of stock prices implies a long-run relationship between stock prices and their dividends or earnings.	Stock prices and stock dividends/earnings.
Epidemiology	Joint mortality models imply a long-run relationship between mortality rates across different demographics.	Male and female mortality rates.
Medicine	Time series methodologies have been used to examine comorbidities of different types of cancers and trends in medical welfare.	Occurrence rates of different types of cancer.

3 Applying Cointegration

3.1 Preparing for Cointegration

Remember the statistical adage, correlation doesn't imply causation! Similarly, before proceeding with verification of existence of cointegration relation, we should have a theoretic support or conjecture that such a relation indeed makes sense. Even a small theoretic support helps strengthen our belief that we are indeed deciphering **useful** relation between two or more variables.

We then proceed to visualize the time series as it is always good to get a feel of the data before apply any of the procedures mentioned in the book. We verify that no structural breaks exist in our time series. Structural breaks occur when a time series suddenly changes its mean. This could mean the simple procedures that we would demonstrate in our current work would be invalid. Structural breaks require more complicated models.

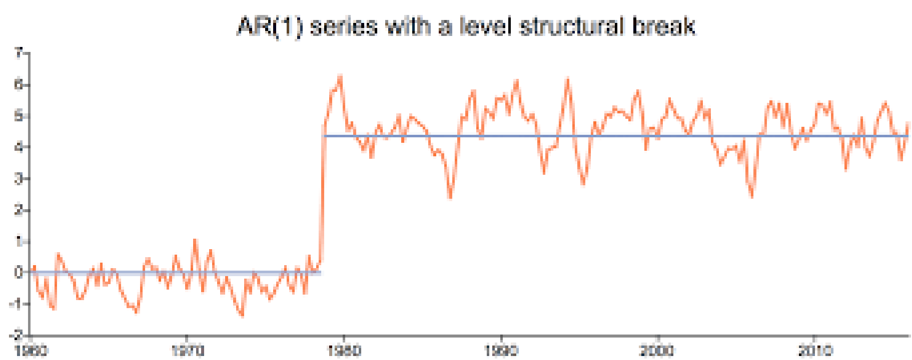

Figure 1: Structural Breaks

At last we require Unit Root tests for repeated application on the original and the differenced time series. We would be using the **Augmented Dickey-Fuller (ADF)** Test for testing stationarity. We detail Johansen's Approach for a bivariate time series where we test each component to be $I(1)$

and then using linear regression to proceed with testing stationarity of residuals.

3.2 ADF Test

ADF test have null hypothesis that a unit root is present in a time series sample. The alternative hypothesis is different depending on which version of the test is used, but is usually stationarity or trend-stationarity. It is an augmented version of the Dickey–Fuller test for a larger and more complicated set of time series models.

$$H_0 : \text{Time Series has a Unit Root}$$
$$H_1 : \text{Time Series does not have Unit Root}$$

3.3 The Error Correction Model (ECM)

- A cointegration model may make use of the term equilibrium which refers to the existence of a long-run relationship.
- This can only occur if there is a common stochastic trend amongst the variables i.e. two variables share a common equilibrium path
- Co-integrating variables will periodically move away from the equilibrium path, but the effect of this will not be permanent (i.e. the errors are stationary)
- The variables return towards the equilibrium path over time
- The residuals from the cointegrated model are then described as equilibrium errors
- In a cointegrated model variables return to the equilibrium value which signifies cointegrated system has an error correction representation which we model using Error COrrection Model
- Assume bivariate system $\boldsymbol{y}_t = (y_{1,t}, y_{2,t})$ with co-integrating vector $\boldsymbol{\alpha} = (1, -\beta_2)$
- Then ECM Model is expressed using these two expressions:

$$\Delta y_{1,t} = c_1 + \delta_1(y_{1,t} - \beta_2 y_{2,t}) + \sum_j \psi_{11}^j \Delta y_{1,t-j} + \sum_j \psi_{12}^j \Delta y_{2,t-j} + \epsilon_{1,t}$$
$$\Delta y_{2,t} = c_2 + \delta_2(y_{1,t} - \beta_2 y_{2,t}) + \sum_j \psi_{21}^j \Delta y_{1,t-j} + \sum_j \psi_{22}^j \Delta y_{2,t-j} + \epsilon_{2,t}$$

- The parameters δ_1 and δ_2 are speed of adjustment parameters that describe how changes to share prices react to past deviations from the equilibrium path in the respective share prices.
- Insignificant values of δ_i imply a relatively unresponsive relationship, where it would take a long time to return to equilibrium.
- If both δ_1 and δ_2 equal zero there are no equilibrium relationship, no error-correction and no co-integration.

4 Data Analysis

4.1 Data Description

We will be working with commodity prices data.[1]

It is a historic time series data, providing information on precious metals futures, like Gold, Silver, Copper, Palladium etc. This dataset offers detailed, up-to-date information on precious metals futures. Futures are financial contracts obligating the buyer to purchase, and the seller to sell, a particular precious metal (such as gold, silver, platinum, etc.) at a predetermined future date and price.

Column Descriptions:

Date: The date the data was recorded. Format YYYY-MM-DD.

Open: Market opening price.

High: Highest price during the trading day.

Low: Lowest price during the trading day.

Close: Market closing price.

Volume: Number of contracts traded during the day.

Ticker: Market quotation symbol for the future.

Commodity: Name of the precious metal the future refers to.

However, **we will only be studying the closing prices of Gold, Silver and Copper.**

4.2 Data Preprocessing (Gold and Silver)

- Loading gold and silver data

```
gold_data = read.csv("Commodity_Data/Gold_data.csv")
silver_data = read.csv("Commodity_Data/Silver_data.csv")
```

- Storing the data as time Series objects.

```
gold_ts <- ts(log(gold_data$close), start = c(2000, 08), frequency = 12)
silver_ts <- ts(log(silver_data$close), start = c(2000, 08), frequency = 12)
```

- Visualizing the closing prices of the commodities Gold and Silver.

```
par(mfrow = c(1, 1), mar = c(2.2, 2.2, 1, 2.2), cex = 0.8)
plot.ts(cbind(gold_ts, silver_ts), plot.type = "single", ylab = "",
    col = 4:3)
legend("topleft", legend = c("gold", "silver"), col = 4:3,
    lty = 1, bty = "n")
```

We observe that as gold price increases or decreases, also does the silver price.

Therefore, we suspect if the two series are co-integrated.

To validate our suspicion, we will perform some tests on both Gold and Silver data as follows.

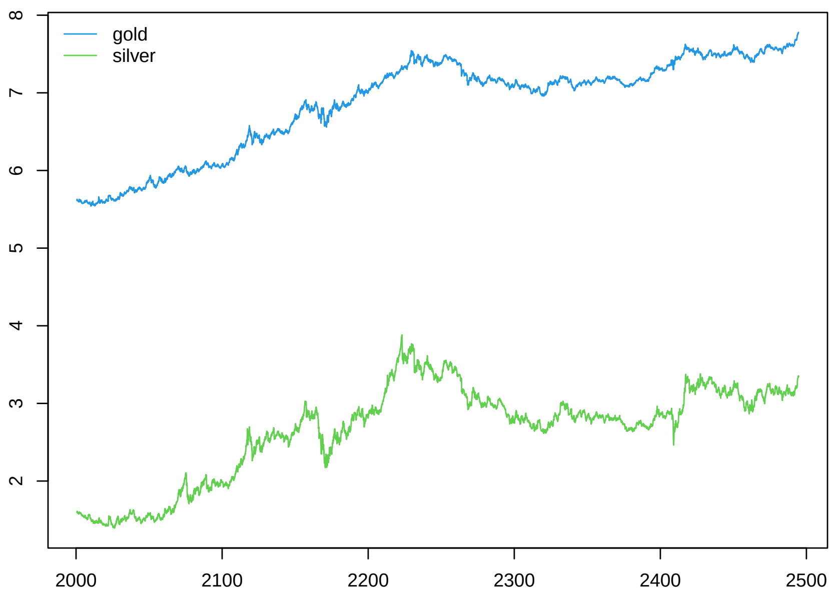

Figure 2: Gold and Silver Time Series

4.3 ADF Test on Gold Series (with Trend and Drift)

ADF test tests for presence of unit root in the time series data. We have applied the test on Gold time series (gold_ts) with a trend and drift component.

```
adfg1 <- ur.df(gold_ts, type = "trend", selectlags = c("BIC"))
summary(adfg1)
```

```
Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.753e-03  3.106e-03   1.852   0.0641 .
z.lag.1     -8.866e-04  5.250e-04  -1.689   0.0913 .
tt           2.353e-07  1.905e-07   1.235   0.2169
z.diff.lag  -1.084e-02  1.299e-02  -0.835   0.4040
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01089 on 5923 degrees of freedom
Multiple R-squared:  0.0006793, Adjusted R-squared:  0.0001731
F-statistic: 1.342 on 3 and 5923 DF,  p-value: 0.2588


Value of test-statistic is: -1.6888 3.3345 1.6438

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Result

- Value of test statistics : -1.69
- 5% significance level : -3.41
- Corresponding p-value : 0.75

Conclusion

- Falied to reject the Null hypothesis at 5% level of significance.
- Gold series has unit root and is non-stationary.

4.4 ADF Test on First Difference of Gold Series

Now, we will test stationarity of gold first difference series diff(gold_ts) using ADF test.

```
adfg2 <- ur.df(diff(gold_ts), selectlags = c("BIC"))
summary(adfg2)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1    -1.014618   0.018465 -54.949   <2e-16 ***
z.diff.lag  0.004642   0.012991   0.357    0.721
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.0109 on 5924 degrees of freedom
Multiple R-squared:  0.505,     Adjusted R-squared:  0.5048
F-statistic:  3022 on 2 and 5924 DF,  p-value: < 2.2e-16


Value of test-statistic is: -54.9493

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Result

- Value of test statistics : -54.94
- 5% significance level : -1.95
- Corresponding p-value : less than 0.01

Conclusion

- Rejected the Null hypothesis at 5% level of significance.
- First difference gold series is stationary.
- Since gold series has unit root but first difference gold series has no unit root, we came to conclusion that gold series is an I(1) process.

4.5 ADF Test on Silver Series (with Trend and Drift)

As we did for gold series, similarly we will apply ADF test for silver series (silver_ts) with trend and drift component.

```
adfs1 <- ur.df(gold_ts, type = "trend", selectlags = c("BIC"))
summary(adfs1)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.830e-03  1.318e-03   2.147   0.0319 *
z.lag.1     -1.246e-03  6.457e-04  -1.930   0.0537 .
tt           2.641e-07  2.260e-07   1.169   0.2426
z.diff.lag  -2.085e-02  1.299e-02  -1.605   0.1085
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01971 on 5923 degrees of freedom
Multiple R-squared:  0.001117,  Adjusted R-squared:  0.0006108
F-statistic: 2.207 on 3 and 5923 DF,  p-value: 0.0851


Value of test-statistic is: -1.93 1.757 1.9502

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Result

- Value of test statistics : -1.93
- 5% significance level : -3.41
- Corresponding p-value : 0.63

Conclusion

- Falied to reject the Null hypothesis at 5% level of significance.
- Silver series has unit root and is non-stationary.

4.6 ADF Test on First Difference of Silver Series

Now, we will test stationarity of silver first difference series diff(silver_ts) using ADF test.

```r
adfs2 <- ur.df(diff(gold_ts), selectlags = c("BIC"))
summary(adfs2)
```

```
Coefficients:
           Estimate Std. Error t value Pr(>|t|)
z.lag.1    -0.99937    0.01856 -53.836   <2e-16 ***
z.diff.lag -0.02137    0.01299  -1.645   0.0999 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01971 on 5924 degrees of freedom
Multiple R-squared:  0.5108,    Adjusted R-squared:  0.5107
F-statistic:  3093 on 2 and 5924 DF,  p-value: < 2.2e-16


Value of test-statistic is: -53.8359

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Result

- Value of test statistics : -53.84

- 5% significance level : -1.95
- Corresponding p-value : less than 0.01

Conclusion

- Rejected the Null hypothesis at 5% level of significance.
- First difference silver series is stationary.
- Since silver series has unit root but first difference silver series has no unit root, we came to conclusion that silver series is also an I(1) process.
- **Both Gold and Silver series are I(1) process.**

4.7 Plotting Residuals

Now that we know the gold series and silver series are $I(1)$ processes, we would like to see the residual corresponding to the co-integrating vector (if exists) forms an $I(0)$ process or not.

Since we do not know the co-integrating vector, we will regress gold_ts over silver_ts to obtain the co-integrating vector.

```
data <- ts.union(gold_ts, silver_ts)
gold.silver.eq <- lm(gold_ts ~ silver_ts, data = data)
```

Now plotting residuals of the fitted linear model.

```
plot.ts(gold.silver.eq$residuals)
```

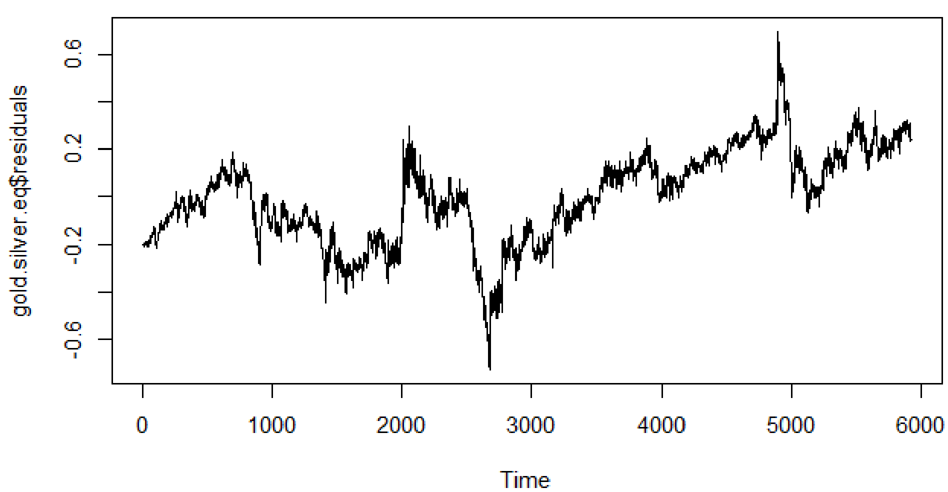

Figure 3: Residuals from regressing Gold on Silver

Visually the residuals appear to be stationary, but let's check it using a formal test.

4.8 Testing stastionarity of residuals

We will test the stationarity with ADF test.

```r
error.gold.silver <- ur.df(gold.silver.eq$residuals, lags = 1, type = "none")
summary(error.gold.silver)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
z.lag.1     -0.99939    0.01857 -53.831   <2e-16 ***
z.diff.lag -0.02139    0.01299  -1.646   0.0997 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01971 on 5923 degrees of freedom
Multiple R-squared:  0.5108,    Adjusted R-squared:  0.5107
F-statistic:  3093 on 2 and 5923 DF,  p-value: < 2.2e-16


Value of test-statistic is: -53.8314

Critical values for test statistics:
      1pct  5pct 10pct
tau1 -2.58 -1.95 -1.62
```

Result

- Value of test statistics : -53.84
- 5% significance level : -1.95
- Corresponding p-value : less than 0.01

Conclusion

- Rejected the Null hypothesis at 5% level of significance. (and at 1% as well)
- Residual series is stationary.
- The process is co-integrated.

4.9 ECM Model Evaluation

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -9.468e-05  1.990e-04  -0.476   0.6343
error.ecm1  -2.661e-03  1.036e-03  -2.568   0.0102 *
gold.d1      1.149e+00  1.843e-02  62.350   <2e-16 ***
silver.d1   -1.070e-01  1.018e-02 -10.504   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01531 on 5923 degrees of freedom
Multiple R-squared:  0.3971,    Adjusted R-squared:  0.3967
F-statistic:  1300 on 3 and 5923 DF,  p-value: < 2.2e-16
```

Result

- Coefficient of δ_2 from ECM model is significant.
- Also, average time take to return to a long term equilibrium is approximately equal to $\frac{1}{\delta_2}$. Thus,

$$\text{Long term equilibrium established in } \frac{1}{0.002661} \approx 376 \text{ trading days}$$

Conclusion

- Significance of δ_2 indicates the presence of cointegration and validates our result.
- This period is quite long (~ 1 year), which indicates equilibrium relationship, though exists, is not very reactive.

5 Data Analysis : Gold and Copper

5.1 Data Preprocessing

- Loading the Gold and Copper data: 5933 observations
- Preprocessing the data.
- Storing the data as Time Series objects.
- Visualizing the closing prices of the commodities Gold and Copper.

```
plot.ts(cbind(gold_ts, copper_ts), plot.type = "single", ylab = "",
    col = 4:3)
legend("topleft", legend = c("gold", "copper"), col = 4:3,
    lty = 1, bty = "n")
```

We observe that as gold price increases or decreases, also does the silver price.

Therefore, we suspect if the two series are co-integrated.

To validate our suspicion, we will perform some tests on copper and gold data as follows.

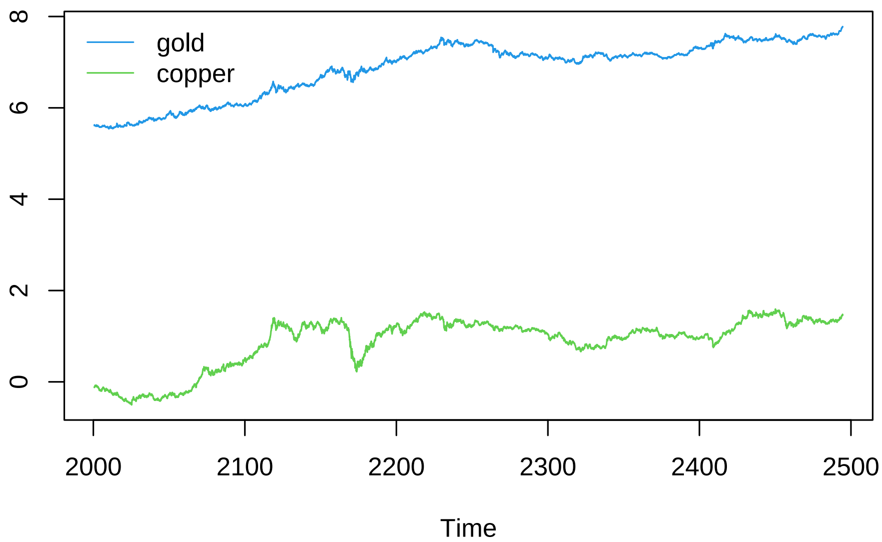

Figure 4: Gold and Copper Time Series

5.2 ADF Test on Copper Series (with Trend and Drift)

Now, we will apply ADF test (with trend and drift) for copper series (copper_ts) with trend and drift component as we have done for gold and silver series previously.

```
adfc1 <- ur.df(copper_ts, type = "trend", selectlags = c("BIC"))
summary(adfc1)
```

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.905e-04  4.512e-04   1.530    0.126
z.lag.1     -9.369e-04  5.627e-04  -1.665    0.096 .
tt           1.409e-07  1.814e-07   0.777    0.437
z.diff.lag  -6.623e-02  1.297e-02  -5.108 3.35e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.01682 on 5923 degrees of freedom
Multiple R-squared:  0.004959,  Adjusted R-squared:  0.004455
F-statistic: 9.839 on 3 and 5923 DF,  p-value: 1.804e-06


Value of test-statistic is: -1.665 1.6094 1.5518

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Result

- Value of test statistics : -1.67

- 5% significance level : -3.41
- Corresponding p-value : 0.65

Conclusion

- Falied to reject the Null hypothesis at 5% level of significance.
- Copper series has unit root and is non-stationary.

5.3 ADF Test on First Difference of Copper Series

Now, we will test stationarity of copper first difference series diff(copper_ts) using ADF test.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.095e-04  4.374e-04   1.165    0.244
z.lag.1     -1.082e+00  1.898e-02 -57.027   <2e-16 ***
tt          -7.390e-08  1.278e-07  -0.578    0.563
z.diff.lag   1.463e-02  1.299e-02   1.126    0.260
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01682 on 5922 degrees of freedom
Multiple R-squared:  0.5334,    Adjusted R-squared:  0.5332
F-statistic:  2257 on 3 and 5922 DF,  p-value: < 2.2e-16


Value of test-statistic is: -57.0267 1084.015 1626.022

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Result

- Value of test statistics : -57.02
- 5% significance level : -3.41
- Corresponding p-value : less than 0.01

Conclusion

- Rejected the Null hypothesis at 5% level of significance.
- First difference copper series is stationary.
- Since copper series has unit root but first difference copper series has no unit root, we came to conclusion that copper series is also an I(1) process.

5.4 Plotting Residuals

Now that we know the gold series and copper series are $I(1)$ processes, we would like to see the residual corresponding to the co-integrating vector (if exists) forms an $I(0)$ process or not.

Since we do not know the co-integrating vector, we will regress gold_ts over copper_ts to obtain the co-integrating vector.

```r
data <- ts.union(gold_ts, copper_ts)
gold.copper.eq <- lm(gold_ts ~ copper_ts, data = data)
plot.ts(gold.copper.eq$residuals)
```

Now plotting residuals of the fitted linear model.

```r
plot.ts(gold.copper.eq$residuals)
```

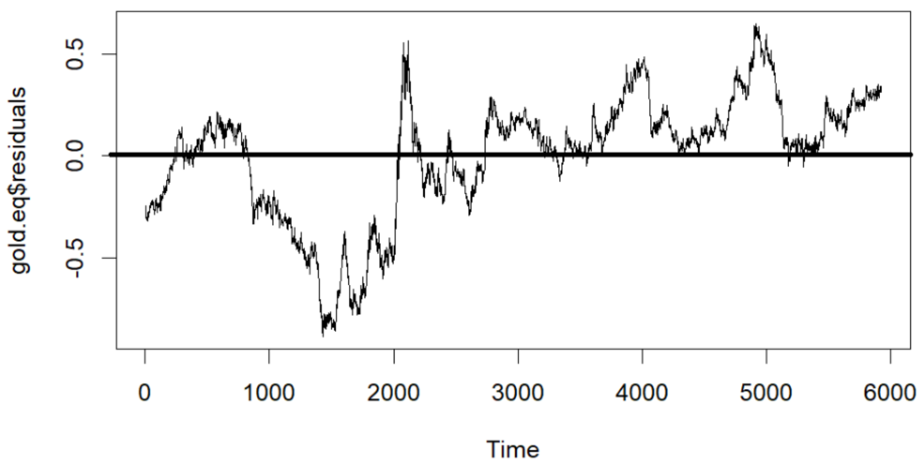

Figure 5: Residuals from regressing Gold on Copper

Visually the residuals appear to be stationary, but let's just confirm it with a formal test.

5.5 Testing stastionarity of residuals

We will test the stationarity with ADF test.

```
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -6.673e-04  5.479e-04  -1.218   0.2233
z.lag.1     -2.436e-03  9.567e-04  -2.547   0.0109 *
tt           2.617e-07  1.672e-07   1.566   0.1175
z.diff.lag  -1.138e-01  1.291e-02  -8.819   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.01796 on 5923 degrees of freedom
Multiple R-squared:  0.01432,   Adjusted R-squared:  0.01382
F-statistic: 28.68 on 3 and 5923 DF,  p-value: < 2.2e-16


Value of test-statistic is: -2.5465 2.2385 3.2489

Critical values for test statistics:
      1pct  5pct 10pct
tau3 -3.96 -3.41 -3.12
phi2  6.09  4.68  4.03
phi3  8.27  6.25  5.34
```

Result

- Value of test statistics : -2.546
- 5% significance level : -3.41
- Corresponding p-value : more than 0.05

Conclusion

- Failed to reject the Null hypothesis at 5% level of significance.
- Residuals are not stationary.
- The process is not co-integrated.

6 References

[1] Data Source: https://www.kaggle.com/datasets/guillemservera/precious-metals-data

[2] https://kevinkotze.github.io/ts-10-tut/

[3] https://www.aptech.com/blog/a-guide-to-conducting-cointegration-tests/

[4] https://faculty.washington.edu/ezivot/econ584/notes/cointegration.pdf

[5] https://kevinkotze.github.io/ts-10-cointegration/