Problem Set 4

Thais Takeuchi

June 15, 2023

Problem 1

1.1

```r
## Question 1.1 - Subset the data to cover only the analyzed period (1995-2020).
# Filter data for the specified date range
start_date <- ymd("1995-01-01")
end_date <- ymd("2020-01-01")
data_brazil <- data_brazil[date >= start_date & date < end_date]
data_usa <- data_usa[date >= start_date & date < end_date]
```

1.2

```r
# Question 1.2 - For each variable X_{k,t}, k \in {1,2,3} in the dataset, define Y_{k,
    t} := 100  [log (X{k,t})     log (X_{k,January1995})]
# X_{1,t} = ipca
# X_{2,t} = exchange rate
# X_{3,t} = cpi

# It creates two new columns, "y_ipca" and "y_exchange_rate", in data_brazil. The
# values in these columns are calculated based on the "ipca" and "exchange_rate"
# columns using the formula above. The formula computes the percentage change
# between the current value and the value at the reference date ("1995-01-01"),
# and multiplies it by 100. The result is assigned to the respective new columns.
base_date <- ymd("1995-01-01")

data_brazil[, `:=` (
  y_ipca = 100 * (log(ipca) - log(ipca[date == base_date])),
  y_exchange_rate = 100 * (log(exchange_rate) - log(exchange_rate[date == base_date]))
)]

# It creates a new column, "y_cpi", in data_usa. The values in this column are
# calculated based on the "cpi" column using the formula above.
# The formula computes the percentage change between the current value and the
# value at the reference date ("1995-01-01"), and multiplies it by 100. The
# result is assigned to the "y_cpi" column.
data_usa[, y_cpi := 100*(log(cpi) - log(cpi[date == base_date]))]

data_merged <- merge(data_brazil, data_usa, by = "date") # Merge data_brazil and data_
    usa
```

1.3

The code in 1.2 was used to define

$$Y_{k,t} := 100 \cdot [\log(X_{k,t}) - \log(X_{k,\text{January1995}})]$$

in the dataset, such that, $X_{k,t}$, where $k \in \{1, 2, 3\}$. According to purchasing power parity empirical example given in class[1], apart from transportation costs, goods should sell for the same effective price in two countries, where:

[1] Lecture 5A: Cointegration using OLS.

- P_t: IPCA;
- P_t^*: CPI;
- E_t: exchange rate in BRL/USD;
- p_t: log of the price level in Brazil;
- p_t^*: log of the price level in USA;
- e_t: log of exchange rate are $I(1)$ processes.

From the definition of the purchasing power parity,

$$P_t = E_t P_t^* \tag{1}$$

If we apply *log* to both sides of equation 1:

$$log(P_t) = log(E_t) + log(P_t^*) \tag{2}$$

$$p_t = e_t + p_t^* \tag{3}$$

Since

$$Y_{k,t} := 100 \cdot [\log(X_{k,t}) - \log(X_{k,\text{January1995}})]$$

We have:

- $X_{1,t} = P_t$, such that,

$$Y_{1,t} := 100 \cdot [\log(X_{1,t}) - \log(X_{1,\text{January1995}})] := 100 \cdot [\log(P_t) - \log(P_{\text{January1995}})$$

- $X_{2,t} = E_t$, such that,

$$Y_{2,t} := 100 \cdot [\log(X_{2,t}) - \log(X_{2,\text{January1995}})] := 100 \cdot [\log(E_t) - \log(E_{\text{January1995}})$$

- $X_{3,t} = P_t^*$, such that,

$$Y_{1,t} := 100 \cdot [\log(X_{3,t}) - \log(X_{3,\text{January1995}})] := 100 \cdot [\log(P_t^*) - \log(P^*_{\text{January1995}})$$

Furthermore,

- $Y_{1,t} = Y_{ipca,t}$
- $Y_{2,t} = Y_{er,t}$, where er=exchange rate
- $Y_{3,t} = Y_{cpi,t}$

Since $Z_t = a'Y_t$, then, $Z_t := p_t - e_t - p_t^*$ is stationary, where $a = (1, -1, -1)'$, according to the slides given in class. Therefore, the order of the variables is $(Y_{1,t}, Y_{2,t}, Y_{3,t}) = (ipca, er, cpi)$.

1.4

```
# Question 1.4 - Define Z_t = a^{\prime}Y_t, where Y_t = (Y_{1,t}; Y_{2,t}; Y_{3,t})^{\prime}
# It adds a new column "z" to the data_merged and assigns it the calculated values of the
# expression y_ipca - y_exchange_rate - y_cpi
data_merged[, z := y_ipca - y_exchange_rate - y_cpi]
```

The code above defines the vector $Z_t = a'Y_t$, where $Y_t = (Y_{1,t}, Y_{2,t}, Y_{3,t})'$

1.5

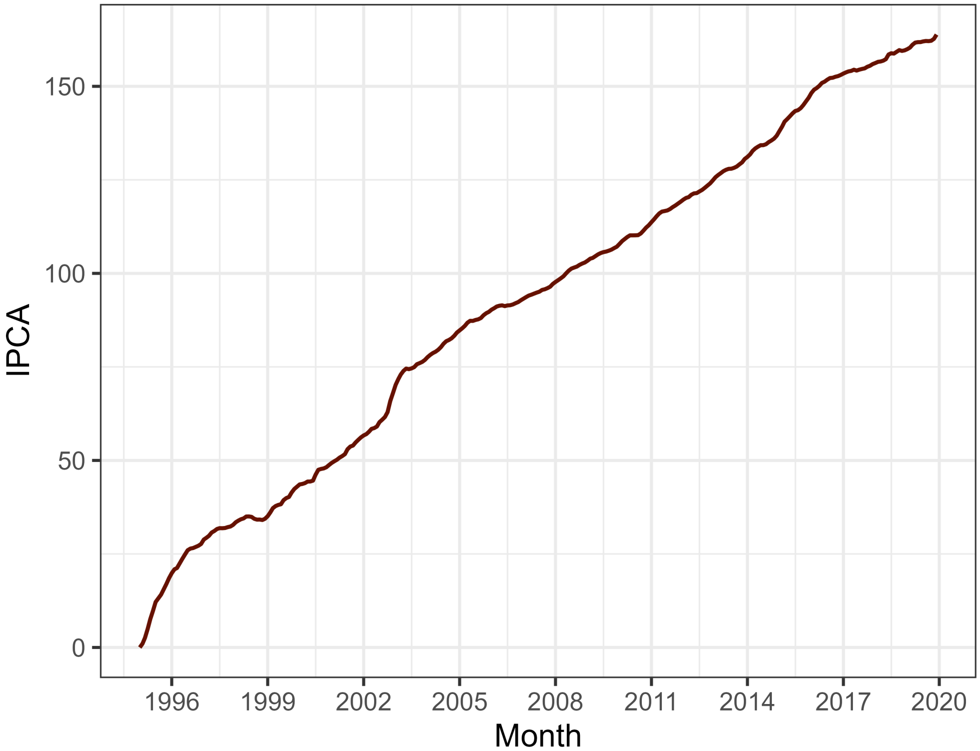

Figure 1: $Y_{1,t}$ through 1995 to 2019

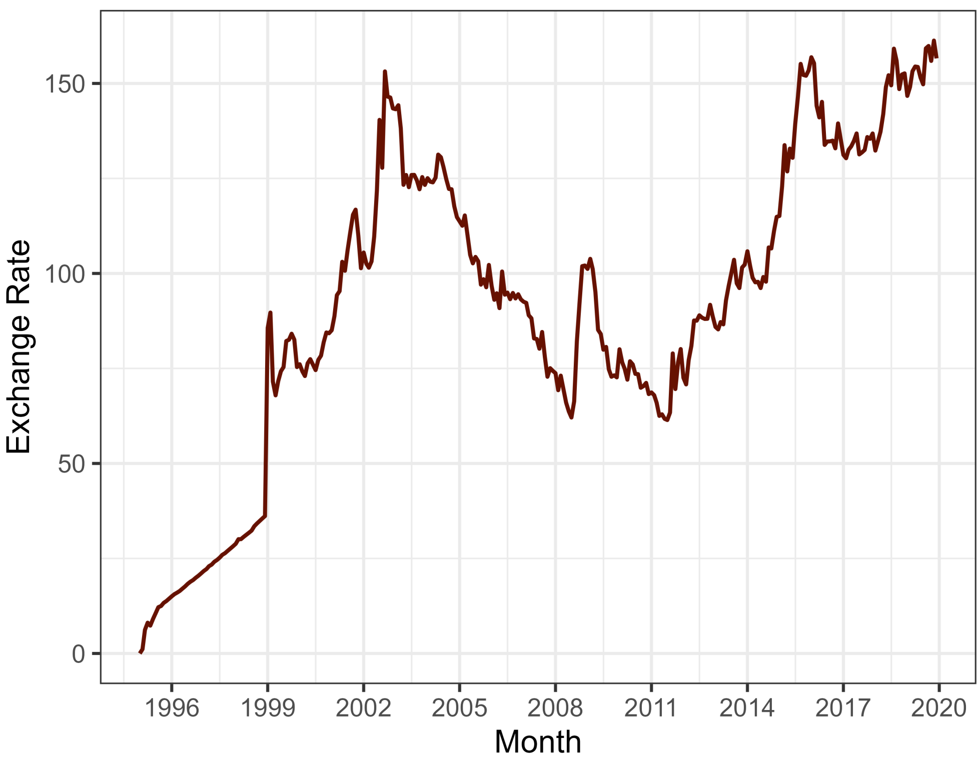

Figure 2: $Y_{2,t}$ through 1995 to 2019

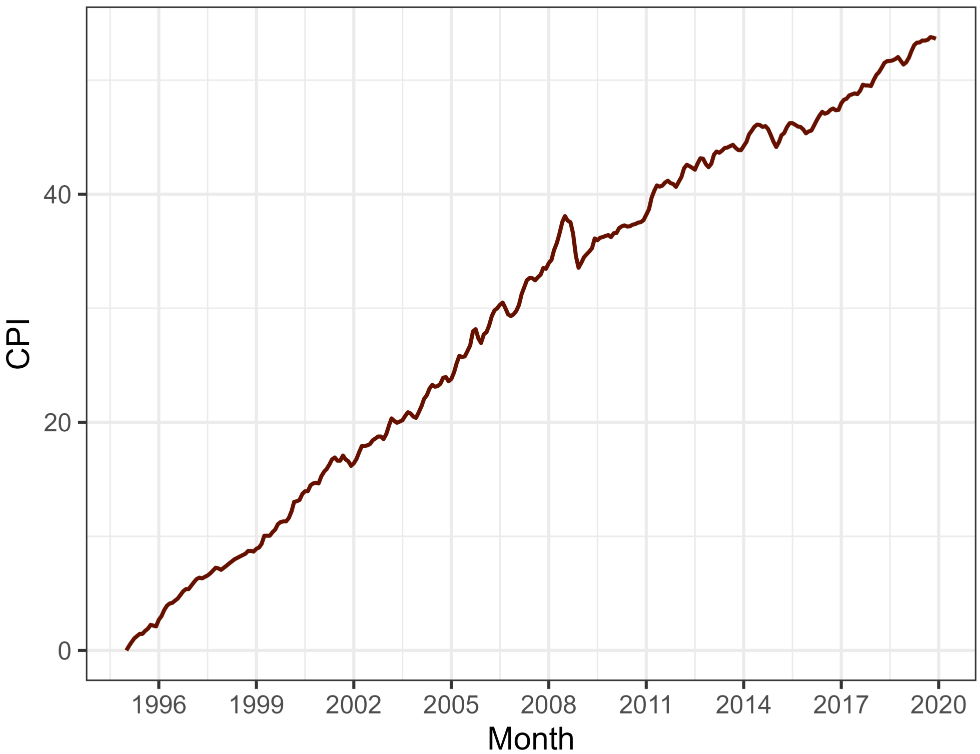

Figure 3: $Y_{3,t}$ through 1995 to 2019

1.6 and 1.8

According to Hamilton's Time Series Analysis[2] and the slides provided by the teacher[3]:
Let $\{Y_t\}$ be a AR(p) stochastic process possibly with a trend term, i.e.,

$$Y_t = \alpha + \delta \cdot t + \phi_1 \cdot Y_{t-1} + \ldots \phi_p \cdot Y_{t-p} + \varepsilon_t$$

We want to test wether $\{Y_t\}$ has one unit root, i.e., if the polynomial function

$$1 - \phi_1 \cdot z - \phi_2 \cdot z^2 - \ldots - \phi_p \cdot z^p \tag{4}$$

has one and only one unit root. To test for the presence of one unit root, we implement the Augment Dickey-Fuller test. It uses the following estimating regressions:

1. No Drift and No Deterministic Time Trend
2. With Drift but No Deterministic Time Trend
3. With Drift and a Deterministic Time Trend

We start by testing the most restrictive model, i.e., the model with drift and a deterministic time trend, illustrated by:

$$\Delta Y_t = \gamma \cdot Y_{t-1} + \delta \cdot t + \alpha + \sum_{i=2}^{p} \beta_i \cdot \Delta Y_{t-i+1} + \varepsilon_t \tag{5}$$

$$\rightarrow (\phi 2) \quad H_0 : \rho = 1 \quad \text{and} \quad \delta = 0 \quad \text{and} \quad \alpha = 0 \tag{6}$$

$$\rightarrow (\phi 3) \quad H_0 : \rho = 1 \quad \text{and} \quad \delta = 0 \tag{7}$$

$$\rightarrow (\tau 3) \quad H_0 : \rho = 1 \tag{8}$$

[2] Chapter 17

[3] Lecture 02C unit root models

such that:
Test (ϕ_2): Presence of unit root, absence of drift and absence time trend under the null.
Test (ϕ_3): Presence of a unit root and absence of time trend under the null.
Test (τ_3): Presence of a unit root under the null.

For all the variables below, the Augmented Dickey Fuller (ADF) test was conducted with 12 lags, selected based on the BIC criteria, assessing the stationarity of the variables. Specifically, in the context of monthly data, using 12 lags means considering the previous 12 months of the variable to account for potential autocorrelation and seasonality patterns.

Monthly data often exhibits patterns or dependencies on past months, and including 12 lags allows the ADF test to capture these dynamics. By examining the test statistic and comparing it to critical values at different significance levels (1%, 5%, and 10%), the ADF test determines whether the null hypothesis of a unit root and non-stationarity should be rejected in favor of the alternative hypothesis of stationarity.

The ADF test is applied to each variable (y_{ipca}, $y_{exchange\_rate}$, y_{cpi} and z_t) in the dataset. The results are summarized in a table that includes the test statistic, critical values at different significance levels, and indicates whether the null hypothesis is rejected or not at each significance level.

Variable	Test	Statistic	CV (1%)	CV (5%)	CV (10%)	Rejection
y_{ipca}	$\tau 3$	-1.83	-3.98	-3.42	-3.13	
y_{ipca}	$\phi 2$	14.84	6.15	4.71	4.05	***
y_{ipca}	$\phi 3$	1.94	8.34	6.30	5.36	

For the IPCA variable we apply a test with drift and time trend. Moreover, the series has a drift and one unit root under the null.

Variable	Test	Statistic	CV (1%)	CV (5%)	CV (10%)	Rejection
$y_{exchange\_rate}$	$\tau 3$	-1.89	-3.98	-3.42	-3.13	
$y_{exchange\_rate}$	$\phi 2$	2.00	6.15	4.71	4.05	
$y_{exchange\_rate}$	$\phi 3$	1.85	8.34	6.30	5.36	

For the exchange rate variable we apply a test with drift and time trend. Moreover, for the 3 tests, we do not reject the null.

Variable	Test	Statistic	CV (1%)	CV (5%)	CV (10%)	Rejection
y_{cpi}	$\tau 3$	-1.81	-3.98	-3.42	-3.13	
y_{cpi}	$\phi 2$	12.03	6.15	4.71	4.05	***
y_{cpi}	$\phi 3$	2.18	8.34	6.30	5.36	

For the CPI variable we again apply a test with drift and time trend. Then, the series has a drift and one unit root under the null.

Variable	Test	Statistic	CV (1%)	CV (5%)	CV (10%)	Rejection
z_t	$\tau 3$	-1.95	-3.98	-3.42	-3.13	
z_t	$\phi 2$	1.39	6.15	4.71	4.05	
z_t	$\phi 3$	1.94	8.34	6.30	5.36	

Finally, for the Z variable we again apply a test with drift and time trend. For the 3 tests, we do not reject the null, i.e., the series has a unit root.

Therefore, $Y_{1,t}$, $Y_{2,t}$, $Y_{3,t}$ and Z_t variables are each individually I(1).

1.7

Figure 4: Z_t through 1995 to 2019

1.9

In the context of testing for cointegration with a known cointegrating vector $a = (1, -1, -1)'$, we need to assess whether each element of the time series $\{Y_t\}$ is individually integrated of order 1 (I(1)). After conducting unit root tests for each series, we found that the null hypothesis of a unit root in each series cannot be rejected. Therefore, we constructed a scalar variable $Z_t := a'Y_t$ to determine whether the combined series $\{Z_t\}$ is integrated of order 1 (I(1)) or of order 0 (I(0)). The null hypothesis for Z_t was not rejected, indicating that $\{Z_t\}$ is integrated of order 1 (I(1)). Consequently, the cointegrating vector a is not valid, suggesting that there is no linear relationship between $Y_{1,t}$, $Y_{2,t}$, and $Y_{3,t}$. This finding contradicts the theory of Purchasing Power Parity (PPP).

Problem 2

The code for question 2 was based on Matheus Junqueira's code.

2.1

According to the Lecture Notes[4], in this exercise we will test for cointegration using the OLS-estimated cointegrating vector, where the H_0 hypothesis is that there is no cointegrating relation amoung our variables Y_t. We will estimate the OLS regression (of an $I(1)$ variable on a set of $I(1)$ variables for which no coefficients produce an $I(0)$ error term) given by,

$$Y_{1,t} = \alpha + \gamma_2 \cdot Y_{2,t} + \cdots + \gamma_n \cdot Y_{n,t} + U_t$$

Then, the Phillips-Ouliaris-Hansen (POH) use the caracteristic of the OLS residuals as a test for cointegration. When conducting the POH test, we can choose an order for the variables based on

[4] Lecture Notes 05A - cointegration using OLS

the economic theory or empirical evidence. The POH test allows us to test for cointegration between variables without the need to specify the cointegrating vector in advance. In this case, we can estimate different models with different variable orders and examine the results of the test.

We can then perform the POH test for each model and examine the results to determine if there is evidence of cointegration. The presence of cointegration would indicate a long-run relationship between the variables, supporting the PPP theory.

It's important to note that the specific order of variables and the results of the POH test may vary depending on the data and the specific econometric methodology used. In this case we use the B9 table from Hamilton's Time Series Analysis[5] regarding case 2, such that, the number of right-hand variables in regression, excluding trend or constant (n-1) = 2.

TABLE B.9
Critical Values for the Phillips Z_t Statistic or the Dickey-Fuller t Statistic When Applied to Residuals from Spurious Cointegrating Regression

Number of right-hand variables in regression, excluding trend or constant $(n-1)$	Sample size (T)	Probability that $(\hat{\rho} - 1)/\hat{\sigma}_{\hat{\rho}}$ is less than entry						
		0.010	0.025	0.050	0.075	0.100	0.125	0.150
		Case 2						
1	500	−3.96	−3.64	−3.37	−3.20	−3.07	−2.96	−2.86
2	500	−4.31	−4.02	−3.77	−3.58	−3.45	−3.35	−3.26
3	500	−4.73	−4.37	−4.11	−3.96	−3.83	−3.73	−3.65
4	500	−5.07	−4.71	−4.45	−4.29	−4.16	−4.05	−3.96
5	500	−5.28	−4.98	−4.71	−4.56	−4.43	−4.33	−4.24

Figure 5: Critical Values for the Phillips Z_t Statistic or the Dickey-Fuller t Statistic When Applied to Residuals from Spurious Cointegrating Regression

Given that the orderings can be: π_t^* (CPI), π_t(IPCA), and e_t (exchange rate); e_t (exchange rate), π_t(IPCA) and π_t^* (CPI) and π_t(IPCA), e_t (exchange rate) and π_t^* (CPI), all 3 orderings are valid for the POH test because π_t(IPCA), π_t^* (CPI) and e_t (exchange rate) cannot have zero as a cointegration coefficient. **Therefore, we choose the order π_t(IPCA), e_t (exchange rate) and π_t^* (CPI) to continue the analysis.**

2.2

Suppose the orderings are:

- 1: π_t(IPCA), e_t (exchange rate) and π_t^* (CPI)
- 2: e_t (exchange rate), π_t(IPCA) and π_t^* (CPI)
- 3: π_t^* (CPI), π_t(IPCA), and e_t (exchange rate)

The OLS regression can be written as:

$$Y_{1,t} = \alpha + \gamma_2 \cdot Y_{2,t} + \gamma_3 \cdot Y_{3,t} + U_t \tag{9}$$

The estimated cointegrating vector for each order is:

[5] Hamilton's Time Series Analysis, page 766.

Order	$\hat{a}_1$	$\hat{a}_2$	$\hat{a}_3$
1	1.000	-0.112	-2.614
2	1.000	-3.250	7.390
3	1.000	-0.3749	0.0364

Table 1: Estimated cointegrating vectors

2.3

As mentioned before, the POH test explore the property of the OLS residuals as a test for cointegration, then, for the POH test it's necessary to estimate Equation 9 and save $\hat{U}_t$, where its H_0 is that $\hat{U}_t$ is a $I(1)$ process i.e., there is no cointegrating relation among our variables. Then, we test whether $\hat{U}_t$ is a $I(1)$ process using the ADF test with no constant term, then:

$$\hat{U}_t = \rho \cdot \hat{U}_{t-1} + \sum_{i=2}^{p} \zeta_{i-1} \cdot \Delta\hat{U}_{t-i+1} + V_t$$

where $H_0 : \rho = 1$

Recall that the critical values of the POH test are different from the standard critical values of the ADF test due to the differences in time trend or not of the stochastic processes $\{Y_t\}$.

To evaluate POH's critical values for model $m1$[6]

```r
# Step 1: Estimating model 1
m1 <- lm(y_ipca ~ y_exchange_rate + y_cpi, data = data_merged)

# Step 2: Computing the cointegration vector
covec_m1 <- c(1, -m1$coefficients[2], -m1$coefficients[3])

# Step 4: Obtaining the residuals vector
resvec_m1 <- residuals(m1)

# Step 5: Performing ADF test on the residuals vector
adf_m1 <- ur.df(
  y = resvec_m1,
  type = "none",
  selectlags = "BIC"
)
```

According to the residuals vector:

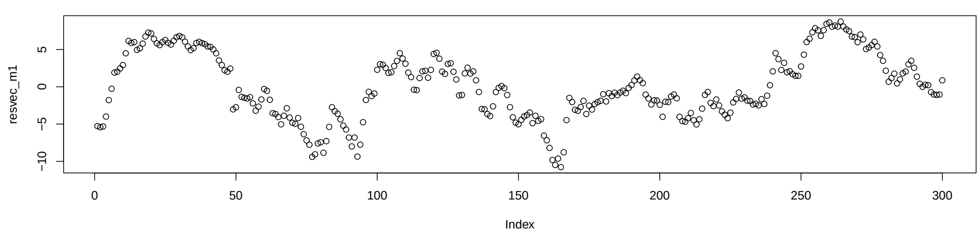

Figure 6: Residuals vector - model 1

The residuals from this regression equation seems to have a zero mean and do not have time trend.[7] Then,

	Statistic	CV (1%)	CV (5%)	CV (10%)
τ_1	-3.089	-4.31	-3.77	-3.45

Table 2: ADF Test: no drift and no time trend for model 1

[6]The ordering is: 1: π_t(IPCA), e_t (exchange rate) and π_t^* (CPI)

[7]Ender's Applied Econometric Time Series Analysis, page 362.

According to the ADF test above for the model 1, since the absolute value of the test statistic is smaller than the critical value of 5%,1% and 10% we cannot reject the null hypothesis, i.e., **the residuals are** $I(1)$.

2.4

When $\hat{U}_t$ is an $I(1)$ process, it means that the variables are not co-integrated, and there is no stable long-term relationship among them. This suggests that any changes or shocks to the variables will have a long-lasting impact, making it difficult to establish a predictable or equilibrium relationship between them, i.e., they cannot be represented by a stationary linear combination.

2.5

The analysis for Model 1 is similar for Models 2 and 3. Recall that Models 2 and 3 have the following orderings:

- 2: e_t (exchange rate), π_t(IPCA) and π_t^* (CPI)
- 3: π_t^* (CPI), π_t(IPCA), and e_t (exchange rate)

To evaluate POH's critical values for model $m2$

```
# Step 1: Estimating model 2
m2 <- lm(formula = y_exchange_rate ~ y_ipca + y_cpi, data = data_merged)

# Step 2: Computing the cointegration vector
covec_m2 <- c(1, -m2$coefficients[2], -m2$coefficients[3])

# Step 4: Obtaining the residuals vector
resvec_m2 <- residuals(m2)

# Step 5: Performing ADF test on the residuals vector
adf_m2 <- ur.df(
  y = resvec_m2,
  type = "none",
  selectlags = "BIC"
)
```

According to the residuals vector:

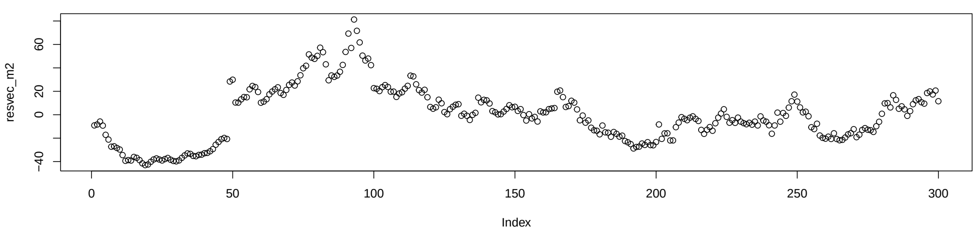

Figure 7: Residuals vector - model 2

The residuals from this regression equation seems to have a zero mean and do not have time trend.[8]. Then,

	Statistic	CV (1%)	CV (5%)	CV (10%)
τ_1	-2.180	-4.31	-3.77	-3.45

Table 3: ADF Test: no drift and no time trend for model 2

[8] Ender's Applied Econometric Time Series Analysis, page 362.

For the model 2, since the absolute value of the test statistic is smaller than the critical value of 5%,1% and 10% we cannot reject the null hypothesis, therefore, **the residuals are** $I(1)$.

Finally, to evaluate POH's critical values for model $m3$

```r
# Step 1: Estimating model 2
m3 <- lm(formula = y_cpi ~ y_ipca + y_exchange_rate, data = data_merged)

# Step 2: Computing the cointegration vector
covec_m3 <- c(1, -m3$coefficients[2], -m3$coefficients[3])

# Step 4: Obtaining the residuals vector
resvec_m3 <- residuals(m3)

# Step 5: Performing ADF test on the residuals vector
adf_m3 <- ur.df(
  y = resvec_m3,
  type = "none",
  selectlags = "BIC"
)
```

According to the residuals vector:

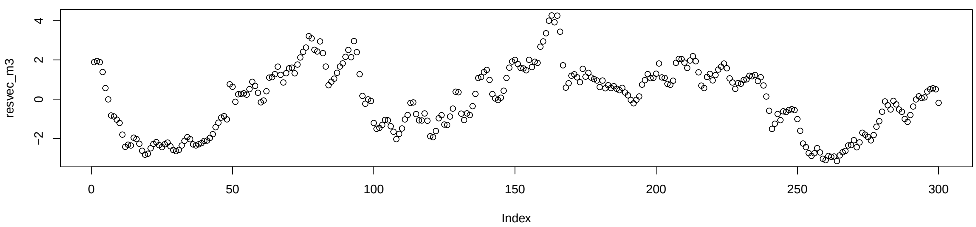

Figure 8: Residuals vector - model 3

The residuals from this regression equation seems to have a zero mean and do not have time trend.[9] Then,

	Statistic	CV (1%)	CV (5%)	CV (10%)
τ_1	-3.118	-4.31	-3.77	-3.45

Table 4: ADF Test no drift and no time trend for model 3

For the model 3, since the absolute value of the test statistic is smaller than the critical value of 5%,1% and 10% we cannot reject the null hypothesis, i.e., **the residuals are** $I(1)$.

Therefore, regardless of the ordering of our variables, the conclusion that we cannot reject the null hypothesis that the residuals are I(1) remains unchanged.

Problem 3

Let's analyse if cointegration exists in this system of equations, i.e., if there exists a linear combination of this non-stationary variables that could potentially be stationary.

3.1

	Test_Statistic	10pct	5pct	1pct
r <= 2 |	2.54	6.50	8.18	11.65
r <= 1 |	6.42	12.91	14.90	19.19
r = 0 |	18.87	18.90	21.07	25.75

[9]Ender's Applied Econometric Time Series Analysis, page 362.

First, we test the H_0 hypothesis of no cointegrating relation (r=0) against the alternative that there is at least one cointegrating relation. Our test statistic is 18.87 and our 5% critical value is 21.07. Hence, we fail to reject the null of no cointegration. Therefore, there is no cointegration relations among our variables of interest.

Furthermore, the Johansen's test was constructed based on:

- `type = "eigen"`;
- `ecdet = "none"`: setting ecedet as "none" implies conducting a unit root test without a common trend. This assumes that the variables have an individual unit root but do not share a common deterministic trend. Hence, we examine whether the variables move together in the long run without assuming a common trend. Therefore, `ecdet = "none"` was chosen due to the PPP model's suggestion that these variables exhibit a stable long-term trend;
- `K = 12`: we are dealing with montly data, hence, we set our model to be a VAR(12);
- `spec = "transitory"`: this is the VECM's specification discussed in class.

```
vecm1 <- ca.jo(
  x = data_merged,
  ecdet = "none",
  type  = "eigen",
  K = 12,
  spec = "transitory")
```

Listing 1: Testing for cointegration

3.1

In this context, the results suggests that there is no evidence of a long-term relationship or cointegration among the variables. This implies that **the purchasing power parity may not hold in this particular case**, indicating that the relative prices between the two countries may not converge in the long run.