Problem Set 4

Pedro Scatimburgo

23/06/2022

Preamble

Here I load the main packages and the data. I also rename the columns.

```
rm(list=ls())

library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(urca)
library(latex2exp)
```

```
## Warning: package 'latex2exp' was built under R version 4.0.5
```

```
brazil_data_raw <- readr::read_csv(paste0("data\\brazil_data.csv"))
```

```
## Warning: Missing column names filled in: 'X4' [4]
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   Data = col_double(),
##   `Taxa de câmbio - R$ / US$ - comercial - venda - fim período - R$ - Banco Central do Brasil- Boleti
##   `IPCA - geral - índice (dez. 1993 = 100) - - - Instituto Brasileiro de Geografia e Estatística- Si
##   X4 = col_logical()
## )
```

```
brazil_data_raw <- brazil_data_raw %>%
  mutate(
    date = brazil_data_raw$Data,
    exrate = brazil_data_raw$`Taxa de câmbio - R$ / US$ - comercial - venda - fim período - R$ - Banco C
    ipca = brazil_data_raw$`IPCA - geral - índice (dez. 1993 = 100) - - - Instituto Brasileiro de Geogra
  ) %>%
  select(date, exrate, ipca)

usa_data_raw <- readr::read_csv(paste0("data\\usa_data.csv"))
```

```
##
## -- Column specification --------------------------------------------------------
## cols(
##   DATE = col_date(format = ""),
##   USACPIALLMINMEI = col_double()
## )
```

```
usa_data_raw <- usa_data_raw %>%
  mutate(
    date = usa_data_raw$DATE,
    cpi = usa_data_raw$USACPIALLMINMEI
  ) %>%
  select(date, cpi)
```

Question 1 (Testing for Cointegration when the Cointegrating Vector is Known - 150 points)

Subset your data to cover only the analyzed period. (10 points)

Although this is pretty straightforward using base R, I prefer to use the `tidyverse` collection. `lubridate::ym` automatically converts a object into the proper '`date` format, but it cannot recognize `yyyy.mm` as a date. Because of that, first I replace `.` for `:` using `stringr::str_replace` and only then I use `lubridate::ym`.

```
brazil_data <- brazil_data_raw %>%
  mutate(
    date = lubridate::ym(str_replace(date, pattern = "\\.", replacement = "\\:"))
    ) %>%
  filter(date >= "1995-01-01" & date <= "2019-12-01") %>%
  select(date, exrate, ipca)

usa_data <- usa_data_raw %>%
  filter(date >= "1995-01-01" & date <= "2019-12-01")

data <- inner_join(brazil_data, usa_data, by="date")
```

For each variable $X_{k,t} \in \{1, 2, 3\}$ in your dataset, define $Y_{k,t} := 100[\log(X_{k,t}) - \log(X_{k,January1995})$. (10 points)

```
data <- data %>%
  mutate(
    log_exrate = 100*(log(exrate)-log(data$exrate[1])),
    log_ipca = 100*(log(ipca)-log(data$ipca[1])),
    log_cpi = 100*(log(cpi)-log(data$cpi[1]))
  )
```

According to the purchasing power parity, what is the value of the cointegrating vector a? To answer this question, you must be clear about the ordering of your variables and careful about measurement units.

The purchasing power parity states that the variation in the exchange rate, measured as the domestic price of the foreing currency, should be equal to the inflation spread between the two countries:

$$\Delta\epsilon = \pi - \pi^* \implies \pi - \pi^* - \Delta\varepsilon = 0$$

The weaker version of the PPP states that $Z_t := \pi - \Delta\varepsilon - \pi^*$ should be a stationary process.

The exchanged rate is already in the desired format $\frac{R\$}{US\$}$. Also, given the definition of Z_t above, the ordering of our variables should be: `log_ipca`, `log_exrate` and `log_cpi`. Then the cointegrating vector a should be:

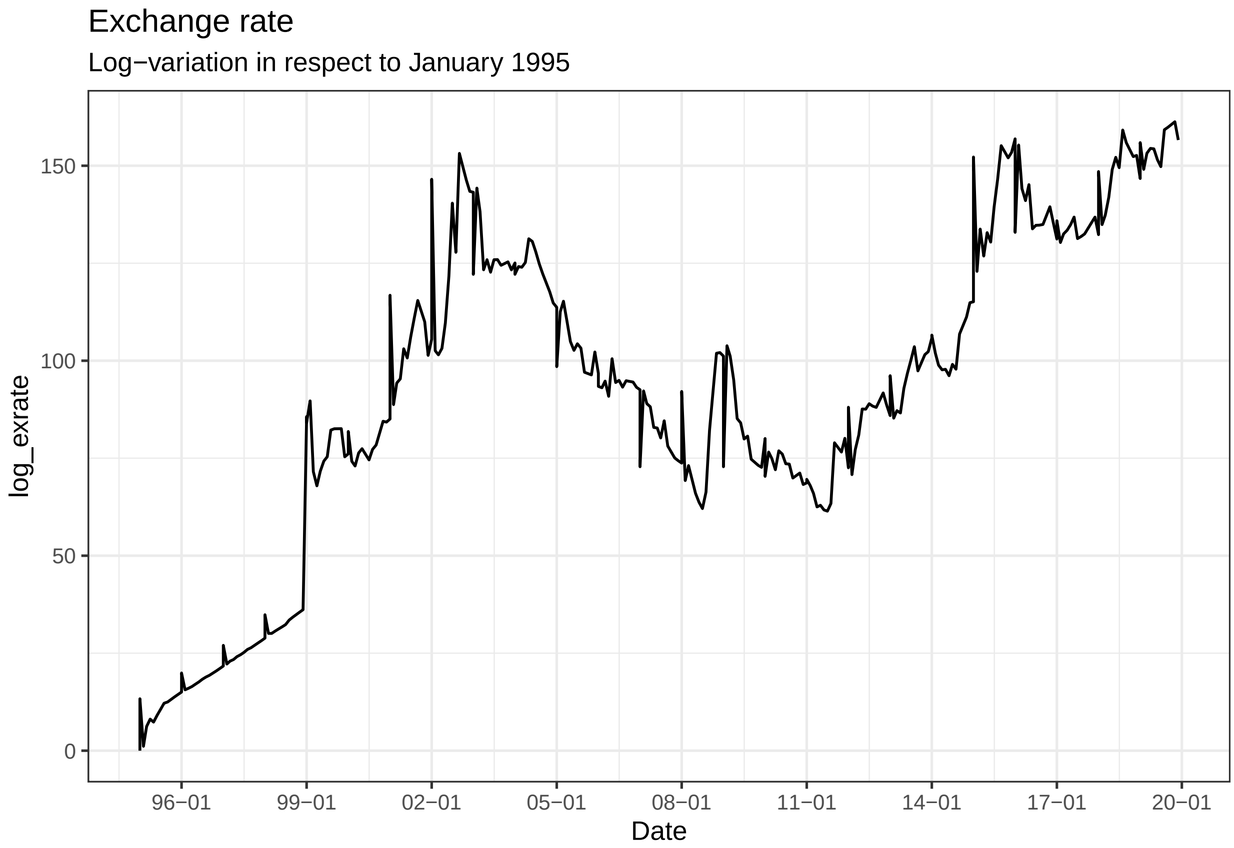
$$a = (1, -1, -1)'$$

Define $Z_t = a'Y_t$, where $Y_t = (Y_{1,t}, Y_{2,t}, Y_{3,t})'$.

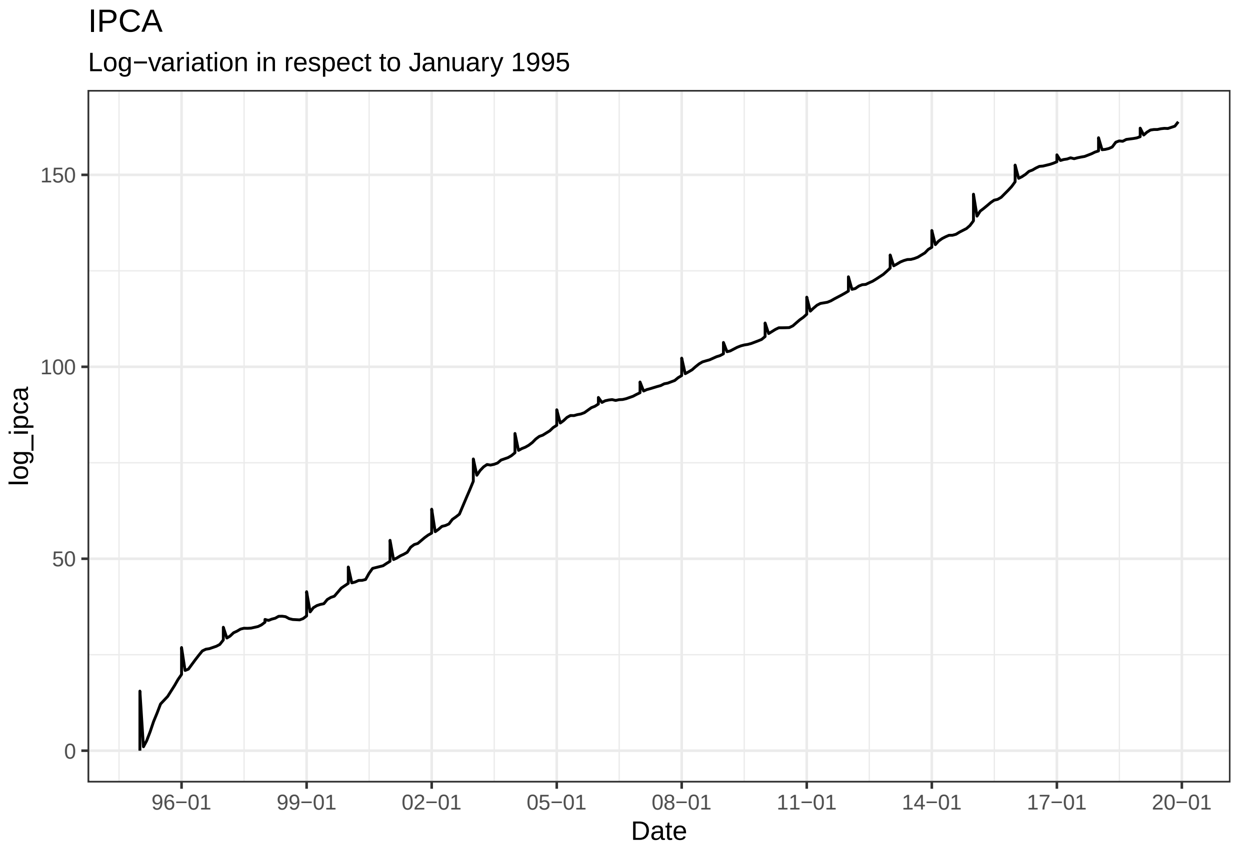
```
Zt = data$log_ipca - data$log_exrate - data$log_cpi
```

Plot the data for $Y_{t,k}$, $k \in \{1, 2, 3\}$.

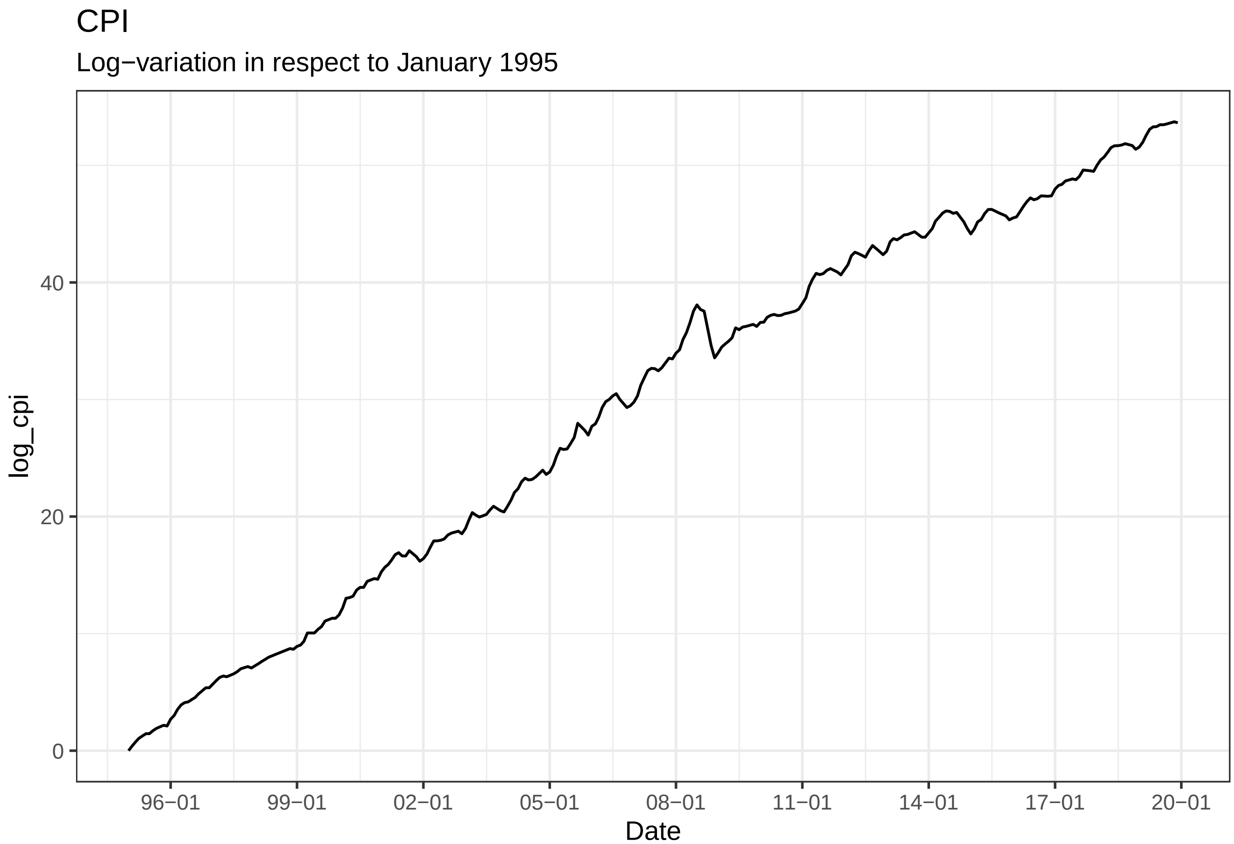
```
log_exrate_plot <- data %>%
  ggplot() +
  geom_line(aes(x = date, y = log_exrate)) +
  scale_x_date(name = "Date", date_breaks = "3 years", date_labels = "%y-%m") +
  labs(title = "Exchange rate", subtitle = "Log-variation in respect to January 1995") +
  theme_bw(base_size = 10)

log_ipca_plot <- data %>%
  ggplot() +
  geom_line(aes(x = date, y = log_ipca)) +
  scale_x_date(name = "Date", date_breaks = "3 years", date_labels = "%y-%m") +
  labs(title = "IPCA", subtitle = "Log-variation in respect to January 1995") +
  theme_bw(base_size = 10)

log_cpi_plot <- data %>%
  ggplot() +
  geom_line(aes(x = date, y = log_cpi)) +
  scale_x_date(name = "Date", date_breaks = "3 years", date_labels = "%y-%m") +
  labs(title = "CPI", subtitle = "Log-variation in respect to January 1995") +
  theme_bw(base_size = 10)

plot(log_exrate_plot)
```




```
plot(log_ipca_plot)
```




```r
plot(log_cpi_plot)
```



Using the Augmented Dickey-Fuller test, test whether your $Y_{k,t}$ variables are each individually $I(1)$. Be clear about the specification of your Augmented Dickey-Fuller test and about your null hypothesis, explaining how you choose the number of lags and your null hypothesis.

Looking at the plots, it seems reasonable to assume that all variables have a time trend. This is very clear for `log_ipca` and `log_cpi`, but less so for `log_exrate`. Still, you could argue there is a trend and a structural break for the latter. Therefore, the null hypothesis is simply $H_0 : \rho = 1, \delta = 0$, which means we will test using ‘`type = "trend"`.

For the lag selection, I used the `selectlags = "BIC"` option. However, since in Possebom's latest lecture, he advocated for a more theory-oriented testing, I decided to set a maximum number of lags of 12, which seems appropriate since we are dealing with monthly data.

```
data_testing <- cbind(data$log_exrate, data$log_ipca, data$log_cpi)
var_names <- c("Exchange Rate","IPCA","CPI")

for(p in 1:3){

  test <- ur.df(y = data_testing[,p],
               type = "trend",
               selectlags = "BIC",
               lags = 12)

 sum_test <- summary(test)

 cat(paste0("Results of the ADF Test for the ",var_names[p]),"\n")
 print(sum_test@teststat)
```

```r
 cat("","\n")
 cat("","\n")

}
```

```
## Results of the ADF Test for the Exchange Rate
##                 tau3     phi2     phi3
## statistic -1.888891 1.999475 1.850617
##
##
## Results of the ADF Test for the IPCA
##                 tau3    phi2     phi3
## statistic -1.832826 14.8383 1.940368
##
##
## Results of the ADF Test for the CPI
##                 tau3     phi2    phi3
## statistic -1.390849 6.164567 1.70619
##
##
```

```r
cat("The critical values for the tests are: ","\n")
```

```
## The critical values for the tests are:
```

```r
print(sum_test@cval)
```

```
##       1pct  5pct 10pct
## tau3 -3.98 -3.42 -3.13
## phi2  6.15  4.71  4.05
## phi3  8.34  6.30  5.36
```

Remember that we have the following convention:

($\phi 2$) $H_0 : \rho = 1$ and $\delta = 0$ and $\alpha = 0$

($\phi 3$) $H_0 : \rho = 1$ and $\delta = 0$

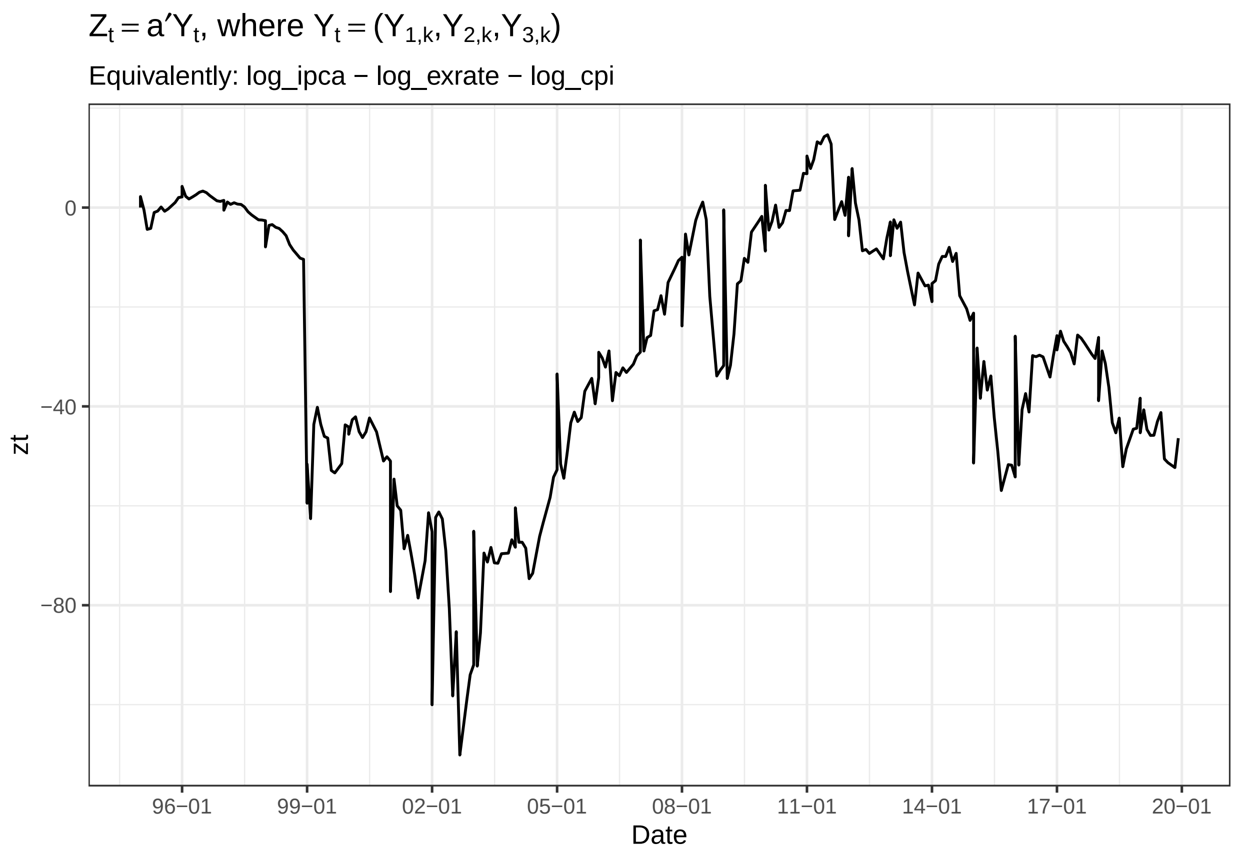
($\tau 3$) $H_0 : \rho = 1$

Because $1.999 < 4.05$, we do not reject the null hypothesis ($\phi 2$) for the exchange rate at the 10% level: we can say that the log-variation of the exchange rate has a unit root, a time trend and a drift. Because $14.8383 > 6.15$, we reject the null hypothesis ($\phi 2$) for the IPCA at the 1% level; but we cannot reject the null hypothesis ($\phi 3$): we can say that the log-variation of the IPCA has a unit root and a time trend, but no drift. Similarly, because $6.16 > 6.15$, we reject the null hypothesis ($\phi 2$) for the CPI, but we cannot reject the null hypothesis ($\phi 3$): we can say that the log-variation of the CPI also has a unit root and a time trend, but no drift as well.

Plot the data for Zt.

```r
Zt_df <- data.frame(
  "date"=data$date,
  "zt"=Zt
)

zt_plot <- Zt_df %>%
  ggplot() +
  geom_line(aes(x = date, y = zt)) +
```

```r
  scale_x_date(name = "Date", date_breaks = "3 years", date_labels = "%y-%m") +
  labs(title = unname(TeX("$Z_t = a'Y_t$, where $Y_t = (Y_{1,k},Y_{2,k},Y_{3,k})$")),
       subtitle = "Equivalently: log_ipca - log_exrate - log_cpi") +
  theme_bw(base_size = 10)

plot(zt_plot)
```



Using the Augmented Dickey-Fuller test, test whether Z_t is $I(1)$. Be clear about the

specification of your Augmented Dickey-Fuller test and about your null hypothesis, explaining how you choose the number of lags and your null hypothesis.

This time it's much harder to argue that the series has a time trend, so I will use `type = "drift"`. The remaining arguments for the test specification remains the same as before: I use `selectlags = "BIC"` for automatic lag selection using the BIC and set a maximum lag of 12 since we have monthly data.

```r
zt_test <- ur.df(y = Zt,
               type = "drift",
               selectlags = "BIC",
               lags = 12)

sum_zt_test <- summary(zt_test)

cat(paste0("Results of the ADF Test for the Zt object:","\n"))
```

```
## Results of the ADF Test for the Zt object:
```

```r
print(sum_zt_test@teststat)
```

```
##                 tau2     phi1
## statistic -1.876845 1.908005
```

```r
cat("","\n")
```

```r
cat("","\n")
```

```r
cat("The critical values for the test is: ","\n")
```

```
## The critical values for the test is:
```

```r
print(sum_zt_test@cval)
```

```
##       1pct  5pct 10pct
## tau2 -3.44 -2.87 -2.57
## phi1  6.47  4.61  3.79
```

Since $-1.876845 > -2.57$, we cannot reject the null hypothesis (τ_2). This means that Z_t has a unit root under the null, and is not stationary.

Based on your analysis, do you believe that the purchasing power parity holds in this context? Explain.

The testable conclusion of the PPP is that $Z_t := \pi - \Delta\varepsilon - \pi^*$ is a stationary process. Remember that the PPP is a generalization of the Law of One Price for bundles of goods: two equal bundles must have the same price, when priced in the same currency. The weaker version allows for the fact that there might some variation in the price, but this variation will be stationary. When we reject the hypothesis that $\pi - \Delta\varepsilon - \pi^*$ is a stationary process, we are rejecting the testable conclusion of the PPP.

Based solely on the ADF test and the data that we have collected, we cannot say that the PPP holds in this specific context.