Stationarity and Cointegration of Vegetable Oils

FRE530 Assignment 1 (12.5 points)
Due in Canvas *before* midnight (11:59pm) on xx

1 Background

This assignment has three main objectives: (1) reinforce the time series topics covered in class, (2) build your intuition about time series in economics within the FRE sector, and (3) build your R toolkit.

In the labs, we have been exploring the linkage between diesel prices and soybean oil. In this assignment, you will explore the linkage between 3 vegetable oils - palm oil, soybean oil, and rapeseed oil. You will submit the rendered PDF or html file that includes your codes, output, and answers.

2 Data Download and Cleaning (1 point)

- Download the World Bank Commodity Markets "Pink Sheet" Data (Monthly Prices, XLS) here
- You may find the following packages useful to conduct the cleaning and analysis: `pacman::p_load(here, readxl, dplyr, janitor, xts, lubridate, urca, forecast)`.
- Using `read_excel()`, load the Commodity Price Data and call this object `cmo`. *Hint: You can add `skip = 4, na = ".."` to skip the first four lines and ask R to read ".." as missing. You can also use the `clean_names()` function right away to fix the variable names.*
  - If you used `clean_names()`, rename `x1` to `date`
  - Use the `dplyr::select()` function to select the following columns only: `date, palm_oil, soybean_oil, rapeseed_oil`.
  - Use `filter()` to keep only the non-missing data for `date` and `rapeseed_oil`
  - You will notice that the date is stored as YYYYM01 format. R needs "a day", as in 1 for January 1, for a year-month date to be recognize as a date. Use `mutate(date = as.Date(paste0(date, "01"), "%YM%m%d"))` to recode the current `date` column into a date format that R can recognize.
  - Use `filter()` to keep only observations from January 1, 2003 to December 1, 2020, inclusive.
  - Use `mutate()` to take the natural log of each price series. Name these columns `lnpalm`, `lnsoy`, and `lnrapeseed`, respectively.
- Use `xts()` to create a new time series object called `vegoils`. This object should only contain `lnpalm`, `lnsoy`, and `lnrapeseed`.
- The first 5 rows of your `vegoils` dataframe should match the output below.

```
##              lnpalm    lnsoy lnrapeseed
## 2003-01-01 6.186435 6.292902   6.435702
## 2003-02-01 6.167412 6.258491   6.372910
## 2003-03-01 6.117106 6.258663   6.313403
## 2003-04-01 6.093615 6.293234   6.322960
## 2003-05-01 6.117701 6.308536   6.408809
```

```
## [1] "xts" "zoo"
```

3 Testing for Stationarity (8 points)

a) Perform a Dickey-Fuller (DF) test (i.e., no constant, no trend, no lags) to determine whether each price series is stationary or not. Be clear and concise in explaining how you concluded whether the price series is stationary or not. To be clear, you will do 3 separate tests; all tests are done with log prices. (2 points)

b) In recent years, an Augmented Dickey Fuller (ADF) test has been developed to account for potential autocorrelation in the residuals.

- Use a partial autocorrelation plot and the `selectlags()` function to explain how many lag terms you would include your ADF test **for log soybean oil prices only**. (1 point).
- Follow the Stationarity Testing flow chart to test for the stationarity of **log soybean oil prices only**. Use the same number of lags in these tests based on your previous answer. Make sure you explain what the different test statistics mean per specification when you discuss your results. (2 points)
- In 2-3 sentences, explain why it is important to go through the flow chart and use different specifications of the ADF test when testing for stationarity. (1 point)

c) Perform an ADF test whether the first difference of each log price series is stationary or not. For each ADF test, control for autocorrelation by using 4 lags for palm, 1 lag for soybean oil, and 1 lag for rapeseed oil. Discuss your ADF test results. In your answers, briefly explain why it would have been acceptable to begin the flow chart at stage 3 rather than stage 1 (i.e., omit testing for a unit root in the presence of a time trend). *Hint: Think about what happens to a time trend when you take a first difference. Also, compare plots of your data with and without differencing.* (1 point)

d) In 1-2 sentences, explain why we took the first difference of each price series and conducted an ADF test on the first difference? (1 point)

4 Testing for Cointegration (3.5 points)

- Use the 2-step Engle-Granger test to determine if palm-soybean, palm-rapeseed, soybean-rapeseed are cointegrated or not. As we have done in the labs, do not add trend or drift terms to the second stage of the ADF test on residuals. For this part, specify `lags = 0` for the second stage ADF test on residuals. Use the `englegranger()` function we used in the lab to determine the appropriate critical values for the ADF test on residuals. Discuss your results. (1 points)
- Do the Step 2 of Engle-Granger test again but now add lags to your ADF test specification. Specifically, add 1 lag to palm-soybean, 2 lags to soybean-rapeseed, and 2 lags to soybean-rapeseed. Discuss your results and the importance of lag length selection in conducting the ADF test on residuals. (1.5 points)
- In a few sentences, explain why certain countries are interested in determining whether there exists a cointegrating relationship among vegetable oils, and why? (1 point)