ECM and IRF of Vegetable Oils

FRE530 Assignment 2 (12.5 points)
Due in Canvas *before* midnight (11:59pm) on March 30, 2022

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.

You may find loading the following libraries helpful in completing the assignment: `pacman::p_load(here, readr, dplyr, tidyr, janitor, xts, lubridate, urca, forecast, vars, modelsummary)`

Johansen Test (1 point)

Recall that in Assignment 1, we used the Engle-Granger test to determine whether the three pairs of prices were cointegrated or not. To determine if all three price series are simultaneously cointegrated we must use the Johansen test (code is in the Appendix). To confirm the presence of cointegration, this test requires us to establish that the rank of the cointegration matrix is greater than zero (i.e., reject the rank = 0 null)

- Using `readRDS()`, read the `vegoils.RDS` data posted on Canvas. Call this object as `vegoils`.
- Use `xts()` to convert `lnpalm`, `lnsoy`, `lnrapeseed` into a time series object and call it `vegoils_ts`.
- Determine the number of lags to include in the Johansen Test. *Hint: Find the lowest AIC for all three variables simultaneously.*
- Conduct a Johansen Trace Test and interpret the results. (1 point)

Error Correction Model (5 points)

With more than two variables we must use the vector error correction model (VECM) and the Johansen method of estimation. This method is beyond the scope of this class and so we will instead estimate a regular two-variable error correction model (ECM) using one pair of prices. Let's work with the palm oil-rapeseed oil pair.

- Estimate the long run relationship for soybean oil, which is specified as $p^{palm} = \alpha + \beta_1 p^{rapeseed} + \epsilon$.
- Use `resid()` to save the residuals from the long run relationship. Then create a new dataframe called `vegoils_r` that merges `vegoils` with the residuals from the equation you just estimated using the `merge.xts()` function. *Hint: You will have to convert the `vegoils` and residuals into an xts object first before doing the merge.* (1 point)
- Determine the number of lags to include in the error correction model for each variable. (1 point)
- Estimate the general ECM model and interpret your results (1 point)
- Does the speed of adjustment make economic sense? You may refer to FRE501 or other sources (2 points)

Data download and data cleaning (1.5 points)

We now analyze the dynamic relationship between these three vegetable oil prices using VAR and IRF. To run a VAR model, the price series must be stationary, and we must use levels instead of first differences. From the first assignment, we showed that each of the price series is not stationary. One way to address this issue is to deflate prices by an index.

- Download the FAO price index for vegetable oils (CSV) here
- Using `read_excel()`, load the FAO price index data and call it `fao`. *Hint: You can add `skip = 2` to skip the first two lines. You can also use the `clean_names()` function right away to fix the variable names.*

  - If you used `clean_names()`, rename `x1` to `year`, `x2` to `month`, and `oils` to `oil_ppi`
  - You will notice that the year only appears for January of that year. You can use `fill(year)` for R to fill missing values based on the previous entry. Read here for info.
  - Using `mutate()` and `as.Date()`, format the date into a format that R recognizes.
  - Use `filter()` to keep only observations from January 1, 2003 to December 1, 2020.

- Merge `vegoils` and `fao` the two dataframes together by the `date` column. Call this dataframe as `vegoils_fao`.

  - Using `mutate()`, deflate the commodity prices (`palm_oil`, `soybean_oil`, `rapeseed_oil`) by `oil_ppi` and call these columns `palmr`, `soyr`, and `rapeseedr`, respectively.
  - Using `mutate()` take the natural log of each deflated price series. Call these columns `lnpalmr`, `lnsoyr`, and `lnrapeseedr`, respectively.
  - Using `select(-)`, remove the `palm_oil`, `soybean_oil`, and `rapeseed_oil` columns.

- Using `xts()`, convert the dataframe into an xts object. Call this new object `vegoils_real`.

- The first 5 rows of your `vegoils_real` dataframe should match the output below.

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

```
##             lnpalmr   lnsoyr lnrapeseedr
## 2003-01-01 6.627734 6.734201    6.877001
## 2003-02-01 6.639818 6.730898    6.845317
## 2003-03-01 6.640848 6.782405    6.837145
## 2003-04-01 6.611633 6.811253    6.840978
## 2003-05-01 6.610523 6.801358    6.901631
```

VAR and Impulse Response Functions (5 points)

- Perform a test to confirm that the natural log deflated prices of each commodity is stationary. Follow the steps in the Stationary testing flowchart. Make sure you control for autocorrelation and test for the optimal lag length (for all 3 variables simultaneously) to use (0.5 point)
- Estimate a basic VAR model. Make sure you include the optimal number of lags. (0.5 point)
- Perform at least 2 diagnostic tests of your VAR model (0.5 point)
- Perform impulse responses functions and plot palm-soybean, palm-rapeseed, and palm-palm IRF plots nicely (0.5 point)
- Explain the IRF plots for palm-soybean, palm-rapeseed, and palm-palm, first while focusing only on the IRF schedules (i.e., ignoring the confidence intervals) and then accounting for the confidence intervals. (3 points)