Homework Assignment #3

The data file DATA02.xlsx contains monthly data for the period Jan-1991 through Mar-2021 (Mar-2021 CPIU is estimated).

There are three data series, RIGS, WTI, and CPIU.

RIGS is the total rig count in the lower 48 states for oil and gas wells.

WTI is the underlying commodity of the New York Mercantile Exchange's oil futures contracts. It is the marker price for oil in the U.S.

CPIU is the consumer price index for all urban consumers.

Rigs are the large apparatus you will find in oil and gas fields that explore for oil and natural gas. They are very expensive and are usually leased on a daily or weekly basis.

Given the equation (1)

1. Estimate , , and by OLS (these are free points -- you should have already done this in Homework 01).
2. Examine the OLS residuals
   1. Calculate , the correlation between and . Construct a correlogram of the OLS residuals, using 24 lags. With evidence from and the correlogram, what is your impression of the validity of the standard errors? How does this impact the t-statistics?
   2. Square the OLS residuals and graph them with Date on the x-axis. Does there appear to be heteroskedasticity?
   3. If there is heteroskedasticity, what is the implication for t-statistics and p-values. Can the p=values be trusted?
3. With high t-statistics and severe autocorrelation, one should reflexively suspect non-stationary data. You should also suspect that RIGS and WTI are cointegrated. Run a cointegration test between RIGS and WTI. Use the weighted version of these variables (RIGS and WTI multiplied by ).
   1. What is your conclusion? (Use the p-value generated by the R function adf.test).
   2. How many lags are needed?
   3. What do these results tell you about the OLS residuals (are they stationary?).
4. Construct and estimate an Error Correction model. The sample used for estimation should start in Jan-1991 and run through Dec-2020. (Because of the lags in the equation below you will lose the first 13 observations, but the R software will automatically do this for you.) The cointegrating equation to be estimated is

The Error Correction equation to be estimated is (no intercept because all variables have a mean of zero):

where are the lagged residuals calculated from the cointegrating equation.

* 1. Calculate , the correlation coefficient between and . Construct a correlogram for . Is there evidence of autocorrelation?
  2. Plot vs time. Is there evidence of heteroskedasticity?
  3. What is your conclusion regarding the validity of the standard errors and the t- statistics now?
  4. Do all the coefficients have the expected signs? Are they all significant at the .05 level? At the .01 level?

1. For January and February of 2021, the values for all the variables are:

|  |  |  |  |
| --- | --- | --- | --- |
| Date | RIGS | WTI | CPIU |
| Jan-21 | 374 | 52.10 | 262.2 |
| Feb-21 | 397 | 59.06 | 262.9 |

1. Using the model developed in #4, forecast for January. (To forecast in January, you need to add your prediction for in January to the actual in December.) Compare this forecast to the forecast made in Homework 01 for the same month. Which one is a better forecast?
2. Use the cointegrating equation developed in #4 to forecast in January. Then subtract this forecast from actual January provided in the table. This difference is for January, which becomes in February. Now, forecast February by first forecasting using the Error Correction equation and adding that forecast to the forecasted January constructed in part a. Compare this February forecast to the February forecast made in Homework 01. Which one is a better forecast?