Exercise 2

Reading in data to GRETL and Unit Root tests

Name: Collin Real (yhi267)

This exercise will give you the opportunity to find an appropriate time series data set and read it in to GRETL. Once read into GRETL, you can then plot and proceed to apply some unit root tests to the data to see whether or not there is a constant mean across time. There is also an extra credit challenge portion that is worth 7 points if you want to attempt it.

1. Find yourself a nice happy time series data set online. The good news is that unlike market research data sets, free time series data sets are easy to find and download. Here is a starting place for you:

<https://archive.ics.uci.edu/ml/datasets.php?format=&task=&att=&area=&numAtt=10to100&numIns=&type=ts&sort=nameUp&view=table>

1. Once you have found your time series data set then plot the data set (be sure to include the plot in your exercise). Use your Mark I eyeball and tell me if you think the mean is constant across time or not.
2. Run an ACF plot for the data set (be sure to include that plot in your exercise). First, tell me what prominent feature is usually there in an ACF plot if there is a trend or non-constant mean across time? Does your plot look like there is a non-constant mean?
3. Next apply the two unit root tests that test for constant mean across time.
   1. What is the null and alternative hypothesis for the KPSS test?
   2. What do you conclude from the KPSS test on your data? Be sure to include the test in your exercise.
   3. What is the null and alternative hypothesis for the Augmented Dickey Fuller test?
   4. What do you conclude from the Augmented Dickey Fuller test on your data?
4. **Select another raw time series data set and repeat steps 2 through 4d.**

\*\*Answers below\*\*

* 1. – Texas Residential Electricity Prices Dataset
  2. – The mean does not look constant over time.

A green line graph with numbers

Description automatically generated

* 1. – Autocorrelation Function (ACF) measures the correlation between a data point and its lagged values. If ACF values drop off quickly, the series might be stationary. Our ACF plot illustrates a slow decay in ACF values, indicating some correlation between observations and time lags; therefore, we conclude the series as non-stationary.

A graph of a graph of a price

Description automatically generated with medium confidence

1.4a – Null: There is a constant mean. Alternative: There is a non-constant mean.

1-4b – Since the p-value (< 0.01) is less than the significance level (assuming 0.05), we reject the null hypothesis and conclude a non-constant mean. This means we can assume that the time series trend is non-stationary.

A screenshot of a computer

Description automatically generated

1-4c – Null: There is more than one mean. Alternative: There is only one mean.

1-4d – Since the p-value (0.3389) is greater than the significance level (assuming 0.05), we do not reject the null hypothesis and conclude that there is more than one mean.

2.1 – Government Expenditures Dataset

2.2 – The mean does not look constant over time.

A green line graph with numbers

Description automatically generated

2.3 – Autocorrelation Function (ACF) measures the correlation between a data point and its lagged values. If ACF values drop off quickly, the series might be stationary. Our ACF plot illustrates a slow decay in ACF values, indicating some correlation between observations and time lags; therefore, we conclude the series as non-stationary.

A graph of a graph

Description automatically generated with medium confidence

2.4a - Null: There is a constant mean. Alternative: There is a non-constant mean.

2.4b - Since the p-value (< 0.01) is less than the significance level (0.05), we reject the null hypothesis and conclude a non-constant mean. The series is non-stationary.

A screenshot of a computer

Description automatically generated

2.4c - Null: There is more than one mean. Alternative: There is only one mean.

2.4d - Since the p-value (0.9983) is greater than the significance level (assuming 0.05), we do not reject the null hypothesis and conclude that there is more than one mean. (Note: I was getting an error when I tried running with 15 lags. The greatest number of lags GRETL would let me input for this dataset was 9.)

A screenshot of a computer

Description automatically generated

Extra Credit w/ Python:

3.1 – Bitcoin Transactions Dataset

3.2 – The mean does not look constant over time.

A graph showing a line of a bitcoin transaction

Description automatically generated

3.3 – Autocorrelation Function (ACF) measures the correlation between a data point and its lagged values. If ACF values drop off quickly, the series might be stationary. Our ACF plot illustrates a slow decay in ACF values, indicating some correlation between observations and time lags; therefore, we conclude the series as non-stationary.

A graph with blue lines and a white background

Description automatically generated

3.4a - Null: There is a constant mean. Alternative: There is a non-constant mean.

3.4b - Since the p-value (0.032568) is less than the significance level (assuming 0.05), we reject the null hypothesis and conclude a non-constant mean.

A screen shot of a computer

Description automatically generated

3.4c - Null: There is more than one mean. Alternative: There is only one mean.

3.4d - Since the p-value (0.9515) is greater than the significance level (assuming 0.05), we do not reject the null hypothesis and conclude that there is more than one mean. This means we can assume the time series is non-stationary. It does not have a constant variance over time and has some time-dependent structure.

A computer screen with numbers and symbols

Description automatically generated