Exercise 2

Reading in data to GRETL and Unit Root tests

Name: Dan Schumacher

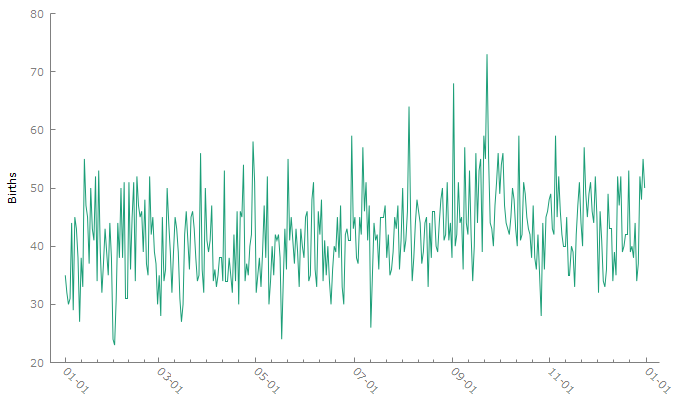
UTSA ID: hdd249

Date: 8/28/2023

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:
   1. **I picked a data set that shows daily female births every day in 1959.** [**Here**](https://machinelearningmastery.com/time-series-datasets-for-machine-learning/) **is the link.**

<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.
   1. 
   2. **I think the dataset might have a small positive correlation.**
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. A graph of a baby

   Description automatically generated with medium confidence
   1. **A typical signiture of a time series data set with no trend (constant mean over time) is flip flopping green bars! ( lag-1 up, lag-2 down, lag-3 up, lag-4 down… more or less)**
   2. **My plot does not look like it has a constant mean across time. All green bars point up!**
4. 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?
      1. **H0: no evidence that there is than one mean (time series is stationary)**
      2. **Halt: data has more than one mean (time series is non-stationary)**
   2. What do you conclude from the KPSS test on your data? Be sure to include the test in your exercise.
      1. A white paper with black text

         Description automatically generated
      2. **If the p is low, the H0 must go. We reject the H0 and support the Halt; the data has more than one mean.**
   3. What is the null and alternative hypothesis for the Augmented Dickey Fuller test?
      1. **H0: data has more than one mean (time series is non-stationary)**
      2. **Halt: no evidence that there is than one mean (time series is stationary)**
   4. What do you conclude from the Augmented Dickey Fuller test on your data?
      1. **With a small p-value, we reject the null hypothesis. This test suggests that the data has no evidence that there is more than one mean.**
      2. A white screen shot of a computer

         Description automatically generated
      3. These results clash with the results of the KPSS test!
5. Select another raw time series data set and repeat steps 2 through 4.
   1. **I will use a dataset that examines the yearly wheat yields in Argentina 1890-1919.**

[**Linked here**](https://users.stat.ufl.edu/~winner/data/argentine_wheat.dat)

* 1. **To my eyes, it has a constant mean over time.**

****

* 1. **My bars are flip-flopping not perfectly, but pretty well.** 
     1. ****
  2. **My KPSS P-Value is greater than 0.10. This means that I fail to reject the null hypotheses. There is no evidence that there is more than one mean.**

**A white background with black text

Description automatically generated**

* 1. **ADF**
     1. **A white text with black text

        Description automatically generated**
     2. **With a p-value of 0.0002536, we fail to reject the null hypothesis. The data has more than 1 mean.**

**Extra credit 7 points**

1. Identify a third time series data set. Read that data set into R, Python, SAS, Stata or name your poison.
   1. **I will be using the electricity production dataset from** [**this Kaggle page.**](https://www.kaggle.com/datasets/shenba/time-series-datasets?resource=download)
2. What are the null and alternative hypotheses for the Phillips-Perron test?
   1. **N0: The data has a unit root (the data is non stationary)**
   2. **Nalt: The data does not have a unit root (the data is stationary)**
3. Apply the Phillips-Perron test to the data set you found in step 1 above. What do you conclude? Be sure to cut and paste a legible copy of the results table that shows the Phillips-Perron test into your exercise.
   1. **with a p value of less than 0.000, we can reject the H0 and conclude that there is strong evidence that the data is weakly stationary.**

A graph showing a line

Description automatically generated with medium confidence

A screenshot of a computer code

Description automatically generated