Assignment #6: Preliminary Data Analysis

Please answer the following questions regarding the data you intend to use for your research project. For each of the questions below, you should include a typed answer, graph, or table.

Rough Draft: Include typed answers to the questions below.
Final Draft: Incorporate the preliminary data analysis into your research paper (see comments on next page)

Documentation

1. Will your regression involve cross section or time series data? If your study involves cross section data, identify what the individual observations represent (i.e., individual students, individual schools, states, countries, etc.) If your study involves time series data, identify the frequency of the data (quarterly, monthly, annual, etc.) and the sample period(s) you intend to use.

2. What is/are the dependent variable(s) you will use in your regression?

3. What are the independent (explanatory) variables you will use in your regression?

4. For EACH of the variables listed in questions 2 and 3 above, document the following:
   - Primary source of the data and, if applicable, the electronic database/secondary source from data.
   - Units of measurement (billions of dollars, index, real 2000 dollars, years of age, etc.)
   - Any adjustments that have been made to the data (seasonal adjustments, moving average, etc.)

   Create a data appendix that includes relevant information about your data. You will want to list each data series with the information above included. If you are using several series from one or two sources, it may make sense to group the data according to source. This appendix can be added to the end of your paper (before the Bibliography/Works Cited pages) and labeled Data Appendix.

Gathering and Preparing Data

You will need to put each of the variables listed in questions above into Excel. You should carefully format the data – you will eventually need to import this information into EViews. I recommend giving yourself plenty of time to do this so that I can help you if you have difficulty.

The instructions below are designed to help you get your data into Excel and EViews. Once in Excel, it will be easy to import the data into EViews.

- Download the data file. If it a file that can not be opened in Excel, you may need to get help from me so that you can get the data ready to use. Give yourself plenty of time to do this! Before you get my help, you need to know how the file is laid out (which variables are 1st, 2nd, 3rd, and how they are formatted), which variables you want in Excel, and which observations you want to select from the entire data file. Common formats for data include text (.txt) or comma separated values (.csv).
- Once in an Excel spreadsheet, import the data into an EViews workfile. Remember, you need to indicate the first cell with data (usually A2 or B2 depending on how your data is laid out) and how many series are in your data file.
- Check that the data was imported correctly. There are two quick ways to do this. Look at the summary statistics and histogram. This will immediately alert you to any excessively high or low values that could indicate the data were imported incorrectly. Also, view a graph of the data – this will indicate any missing values, outliers, or glaring mistakes in importing the data.

Preliminary statistics

Note, if you are using time series data, you may need to transform your data. Carefully study the papers from your literature review to become familiar with which transformations are most common in your research topic. Also, think ahead to how your data and economic model are connected.

5. For your dependent variable(s), compute the following summary statistics. These should be included in a table.
    - Mean
    - Median
    - Standard deviation

6. To get a sense of what your data look like, do one of the following:
   Cross section data: Plot a histogram of your dependent variable
   Time series data: Plot your dependent variable over the sample period(s) you intend to study.
   ***Note, you may need to transform your data.

7. Briefly discuss what information you've learned from the summary statistics you computed in #5 and the preliminary graph you created in #6.

8. Choose one of the explanatory variables that represents the focus of your particular study. Using your economic model, explain why this variable should be included in your analysis. What does this variable measure in the context of your model? Based on your model, should the relationship between your explanatory variable be positive or negative? Use your economic model to justify your answers.

9. For the explanatory variable you chose in Question 8, repeat the exercises from Questions 5-7. You may want to include the summary statistics in the same table you created for Question 5.

10. Compute the correlation between your dependent variable and the explanatory variable from question 8. Comment on the correlation you've computed – is it consistent with your economic model?

11. Create a scatter plot with a regression line of your dependent variable and the explanatory variable from question 7. The explanatory variable should appear on the horizontal axis. Is the slope of the line (positive/negative/zero) consistent with the correlation you computed in question 8? Explain why or why not.

12. Provide an explanation of what the preliminary statistics from 5-9 reveal about your research question. Identify at least one other variables you plan to control for in your regression analysis. Describe link between these explanatory variables and your dependent variable using your economic model.

Research Paper
In addition to submitting this assignment, you should integrate the preliminary data analysis. It is your decision as to which observations from this preliminary analysis are most important for answering your research question. Think carefully about which graphs and statistics help to answer your research question for the reader. Your graphs and tables should always stand alone. The reader should be able to look at tables and graphs without having to read through your paper to interpret them.

Oral Presentation
You will present your preliminary data analysis, along with the other parts of your paper (Introduction, Literature Review, and Economic Model). As with your paper (see above), you must decide which observations from this preliminary analysis to include in the presentation. Think carefully about which graphs and statistics help to answer your research question for the audience. Also, as in your paper, your tables and graphs should be labeled in a way that allows the audience to read them easily.