Problem Set 4

Autocorrelation, Nonstationarity, and Causality

EC 421: Introduction to Econometrics

Due *before* midnight (11:59pm) on Sunday, 08 March 2020

DUE Your solutions to this problem set are due *before* midnight on Sunday, 08 March 2020. Your files must be uploaded to Canvas.

IMPORTANT You must submit **two files**:

1. your typed responses/answers to the question (in a Word file or something similar)
2. the R script you used to generate your answers. Each student must turn in her/his own answers.

OBJECTIVE This problem set has three purposes: (1) reinforce the econometrics topics we reviewed in class; (2) build your R toolset; (3) start building your intuition about causality and time series within econometrics.

Problem 1: Concepts

1a. Define the term **standard error**.

1b. If the *p*-value of a hypothesis test is 0.14, what do we conclude about the null hypothesis?

1c. If the *p*-value of a hypothesis test is 0.01, what do we conclude about the null hypothesis?

1d. Define **mean stationarity** in your own words.

1e. If the variance of a random variable increases over time, is that variable variance stationary? Explain.

1f. Define a **random walk** and show why/how it violates **variance stationarity**.

1f. How does the task of *understanding causality* differ from the task of *prediction*?

1g. Define the fundamental problem of causality in your own words.

1h. Define **covariance stationarity** in your own words.

Problem 2: Empirics

Load the energy-price dataset from the previous problem set (`004-data.csv`). The variables are explained on the last page.

2a. Write out a regression model that has

- outcome variable: the price of electricity
- explanatory variables: the price of natural gas and its first **two** lags (plus an intercept)

2b Estimate the model in **2a** and report your results. Interpret the coefficients and comment on their significance.

2c. If the disturbance in the model in **2a** is autocorrelated, will OLS be unbiased and/or consistent for the coefficients? Briefly explain your answer.

2d. What issues does OLS have if the disturbance in the model in **2a** is autocorrelated?

2e. Following the methods described in class and in the notes, use your residuals (from **2b**) to test the disturbance for first-order autocorrelation.

2f. Now add two lags of the price of electricity to your model in **2a**—so your explanatory variables should be:

- the price of natural gas
- the first two lags of the price of natural gas
- the first two lags of the price of electricity
- an intercept

Write out this model.

2g. Is OLS unbiased and/or consistent for the model you wrote out in **2f**? Briefly explain.

2h. Now estimate the model that you wrote out in **2f**. Interpret the coefficient on the lags of the price of electricity and comment on all of the variables' significance—how have they changed from the first model?

2i. Using the residuals from **2h**, test the disturbance for second-order autocorrelation. What do you conclude? What do your conclusions imply for our coefficient estimates?

Hint: Make sure you include the **full** set of explanatory variables in the test (the notes were unclear in class but have since been updated).

Problem 3: Causality

Following the Rubin causal model, imagine that we observe the following dataset.

Table: Imaginary dataset

i	Trt.	y_1	y_0
1	0	34	22
2	0	19	13
3	1	13	1
4	1	13	10

3a. Which numbers would be impossible to observe in real life?

3b. Define the term **average treatment effect**.

3c. Based upon this dataset, calculate the **true average treatment effect**.

3d. Based upon this dataset, **estimate** the average treatment effect by taking the difference between the mean of the treatment group and the mean of the control group.

3e. Under what conditions would the estimate in **3d** be unbiased? Explain.

3f. Is the treatment effect heterogeneous? Explain your answer.

Description of variables and names

Variable	Description
`t`	Time, measured by months in the dataset (`numeric`)
`date`	The observation's month and year (`character`)
`year`	The year (`numeric`)
`month`	The month (`numeric`)
`price_electricity`	The average residential electricity price in USD (`numeric`)
`price_coal`	The average price of coal, $ per short ton (`numeric`)
`price_gas`	The average price of natural gas, $ per cubic ft (`numeric`)