Problem Set 3

Time Series, Autocorrelation, and Consistency

EC 421: Introduction to Econometrics

Due *before* midnight (11:59pm) on Friday, 29 April 2020

DUE Upload your answer on Canvas *before* midnight on Friday, 29 May 2020.

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.

If you are using RMarkdown, you can turn in one file, but it must be an HTML or PDF that includes your responses and R code.

OBJECTIVE This problem set has three purposes: (1) reinforce the topics of time series and statistical inference; (2) build your R toolset; (3) start building your intuition about causality within econometrics/regression.

INTEGRITY If you are suspected of cheating, then you will receive a zero. We may report you to the dean. Everything you turn in must be in your own words.

Conceptual Questions

1. Remember that we've discussed three types of time-series models: (**1**) static models, (**2**) dynamic models with lagged explanatory variables, (**3**) dynamic models with lagged outcome variables.

> **1a.** If the disturbance u_t **is not autocorrelated**, for which of the 3 types of models is OLS **unbiased**? If any of the models are biased, explain why.
>
> **1b.** If the disturbance u_t **is not autocorrelated**, for which of the 3 types of models is OLS **consistent**? If any of the models are inconsistent, explain why.
>
> **1c.** If the disturbance u_t **is autocorrelated**, for which of the 3 types of models is OLS **unbiased**? If any of the models are biased, explain why.
>
> **1d.** If the disturbance u_t **is autocorrelated**, for which of the 3 types of models is OLS **consistent**? If any of the models are inconsistent, explain why.

2. In our time-series lecture, we discussed how static time-series models are a pretty restrictive and simplistic way to model time-series data.

> **2a.** Explain why static time-series models are generally restrictive and simplistic.
>
> **2b.** Give an example of a reasonable **static** time-series model. By *reasonable* we mean that it would be reasonable to model the relationship as a static relationship. Explain why it is reasonable to model the relationship as static rather than dynamic—and make sure you tell us what t would represent (*e.g.*, days, months, years).
>
> *Note:* The model should look something like $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + u_t$
>
> **2c.** Give an example of a reasonable **dynamic** time-series model. By *reasonable* we mean that it would be reasonable to model the relationship as a dynamic relationship. Explain why this relationship should be modeled as a dynamic relationship. Make sure you tell us what t would represent (*e.g.*, days, months, years).
>
> *Note:* The model should look something like $\text{Births}_t = \beta_0 + \beta_1 \text{Income}_t + \text{Income}_{t-1} + u_t$

3. Time-series models frequently include the lag of a variable, *e.g.*, x_{t-1}. Explain why we usually do not use lags in cross-sectional models, *e.g.*, x_{i-1}.

Some Real Data

Now we're going to work with some real data. The data come from the Environmental Protection Agency (EPA). Specifically, the data describe electricity generation in the United States at a monthly level—the amount of electricity generated, associated emissions, the number of retirements, *etc*.

For more information on the dataset, see the table on the last page of this problem set.

Why? Electricity generation is obviously important for day-to-day life: it runs our heating and air conditioning, it allows us to have computers/phones/internet/refrigerators/*etc.*, and it supports many businesses and critical parts of our health systems and economy.

Emissions are important, because burning fossil fuels (*e.g.*, coal and natural gas) produces toxic gases that are released above the plant. These gases (emissions) have been traced to a bunch of negative outcomes—for people, animals, plants, and the general environment (*e.g.*, acid rain). Economics is about thinking on the margin: Where do the marginal benefits from something equal the marginal costs? We know we need electricity, so we do not want to make it too expensive for electricity generators to operate, but if we do not regulate electricity generation, then the power plants may poison our air and water. Thus, one job of economists (specially environmental and energy economists) is figuring out how regulations affect health, environment, and energy costs.

4. Load packages and your dataset `003-data.csv`.

5. Which dates does the dataset cover (what are the start and end dates)? How many months?

6. How many plants retired during this sample?

7. Create (and include) **three figures**: (**1**) the time series of total monthly generation (`generation_gwh`), (**2**) the time series of NO_X (Nitrogen Oxide) emissions (`emissions_nox`), and (**3**) the time series for the number of electricity generators who retired in the given month (`n_retirements`).

Hint: A time-series graph has time on the x axis and a variable on the y axis. Your x axis can have either time `t` (time relative to the beginning of the sample) or date (`month`).

8. For each of the three time-series graphs in **7**, explain whether the variable appears to be positively autocorrelated, negatively autocorrelated, or *not* autocorrelated. Make sure you explain your reasoning.

9. Estimate a **static** time-series model where monthly *NO_X emissions* (`emissions_nox`) are the outcome variable and our two explanatory variables are the *number of retirements* in the month (`n_retirements`) and the amount of electricity generation in the month (`generation_gwh`).

Report your coefficient estimates and their statistical significance.

10. Now estimate a **dynamic** model in which you include the first lag for each of your explanatory variables (number of retirements and amount of electricity generation). *Note:* You still want the non-lagged version of the variables too—*i.e.*, include x_t *and* x_{t-1}. Interpret the coefficient on the lagged number of retirements.

11. Why might it make sense to include lags of the variable *number of retirements*? In other words: Why might we want a dynamic model with lagged explanatory variables in this setting?

12. If the disturbance is autocorrelated, what problems does it cause for OLS regression estimates in **10**?

13. Use the residuals from the regression in **10** to test for first-order autocorrelation in your disturbance. Report the results from the hypothesis test.

Hint: Don't forget about the missing values due to lags (see lecture notes).

14. Now estimate a dynamic model (still with NO_X emissions as the outcome variable) with **0, 1, 2, and 3** lags of the **number of retirements** and also the current month's electricity generation (no lags). Interpret the coefficient on the third lag of the number of retirements.

15. Based upon your estimates in **14**, what is the *total* effect of a retirement on NO_X emissions?

16. Now estimate an ADL(1,1) model with NO_X emissions as the outcome and with *number of retirements* and *electricity generation* as the explanatory variables. Report/interpret the coefficient on the lag of NO_X emissions.

Hint: Your regression should have an intercept plus five more terms.

17. Does it make sense to regress current NO_X emissions on the previous month's emissions? Explain your answer.

18. If the disturbance is autocorrelated, then OLS is not consistent for the coefficients in **16**. Explain how you could test for an autocorrelated disturbance using the model from **16**.

Note: You do not actually need to run this test.

19. Try to find the "best" model for explaining the relationship between monthly NO_X emissions (your outcome variable) and retirements. Include lags, other variables, interactions, logs—whatever you want. Report your final model and explain why you chose it.

20. Return to your figures in **7**: Do any of the three figures suggest a violation of mean stationarity? Explain.

Description of Variables

Variable	Description
`t`	Time, relative to the first month of the sample (1, 2, ...)
`month`	Month of the sample (e.g., 2015-12-01)
`generation_gwh`	Total monthly electricity generation (Gigawatt hours, GWh)
`emissions_so2`	Total monthly emissions of SO_2 (in tons)
`emissions_nox`	Total monthly emissions of NO_X (in tons)
`n_plants`	Number of unique electricity-generating units (EGUs) operating in the month
`n_retirements`	Number of retired electricity generating units in the month
`cumulative_retirements`	Cumulative number of retirements (through the given month)
`i_cair`	Binary indicator for months during the Clean Air Interstate Rule (CAIR)
`i_csapr`	Binary indicator for months during the Cross-State Air Pollution Rule (CSAPR)