Problem Set 3

EC 421: Introduction to Econometrics

Due *before* midnight on Friday, 05 March 2021

DUE Upload your answer on Canvas *before* midnight on Friday, 05 March 2021.

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 (not just the RMD file).

If we ask you to create a figure or run a regression, then the figure or the regression results should be in the document that you submit (not just the code—we want the actual figure or regression output with coefficients, standard errors, *etc.*).

OBJECTIVE In this problem set, we want to (1) reinforce key topics to time-series econometrics; (2) continue to build your R toolset; (3) keep building your intuition about causality and inference within econometrics/regression.

INTEGRITY If you are suspected of cheating, then you will receive a zero. We may report you to the dean.

Theory/review

Q01. First, let's review some of the key concepts and results of time-series econometrics. For all of the sub-questions, you can assume that all of our assumptions are satisfied **unless we explicitly say otherwise**.

A. How does autocorrelation affect **static models**? Consider unbiasedness and consistency.
B. How does autocorrelation affect **dynamic models with lagged explanatory variables**? Consider unbiasedness and consistency.
C. How does autocorrelation affect **dynamic models with lagged outcome variables**? Consider unbiasedness and consistency.
D. With no autocorrelation, can **dynamic models with lagged outcome variables** be unbiased and/or consistent?
E. With no autocorrelation, can **dynamic models with lagged explanatory variables** be unbiased and/or consistent?
F. Why do we care about **nonstationarity**?

Crime and policing

In this problem set, we're going to compare the historical time series data for crime, population, and policing (specifically the size of the police force) in Illinois between 1985 and 2019.

Q02. Load your packages. You'll probably going to need/want `tidyverse` and `here` (among others). Now load the data (`003-data.csv`). The last page of this problem set describes the variables.

Important: Please note that most of the variables (excluding `year` and `unemployment`) are in tens of thousands. For example, in 1985 the population of Illinois (the state from which we have data) was `1139.981` **times 10,000** (which is 11,399,806).

Q03. Create a time-series graph (time, which is `year` here, should be in the *x* axis) for each of the following four variables: population (`pop`, in 10,000s), number of employed police officers (`n_officers`, in 10,000s), and both types of crime (violent: `n_crime_violent`; property: `n_crime_property`; both are in 10,000s).

Don't forget to label your figures.

Q04. Based upon your figures in **03**, which, if any, of your variables appear to be positively autocorrelated? Which, if any, appear to be negatively autocorrelated?

Q05. Based upon your time-series figure for property crimes in **03**: Does the variable appear to be stationary? If so: Explain how it satisfies each of the three requirements for stationarity. If not: Explain how it violates stationarity.

Q06. Let's start with a static model. Regress the **log** of property crimes (*i.e.*, `log(n_crime_property)`) on an intercept and on the **log** of the number of police officers (*i.e.*, `log(n_officers)`).

Include your regression results, interpret the coefficient on police officers, and comment on the significance.

Q07. We're currently using a log-log model. Explain why this might make sense in this setting, compared to a log-linear model.

Q08. Now add **log** population (*i.e.*, `log(pop)`) into your regression model—you now are regressing the **log** of the number of property crimes on the number of police officers and population.

Include your regression results. Interpret the coefficient on population, and comment on its significance.

Q09. When you added the log of population (moving from the regression in **06** to the regression in **08**), the coefficient on the log number of police officers should have changed signs. Give an explanation for why this happened.

Hint: Think about the lecture on signing the bias from omitted variables.

Q10. These previous models have all been static models. Explain why you think a static model in this setting could be appropriate or inappropriate.

Q11. Time for a dynamic model. Add the lags of your two explanatory variables (the log of police officers and the log of population). Note: You should have four regressors now (plus the intercept)—the contemporaneous variables *and* their lags. Everything is still logged.

Include your regression results. Interpret the coefficient on the lag of police officers, and comment on its significance.

Q12. Based upon your regression in **11**, what is the *total effect* of a 1-percent increase in the number of police officers (on property crime)?

Q13. Should we be worries about reverse causality here? Explain your answer.

Q14. Use the residuals from your regression model in **11** to test for first-order autocorrelation in the disturbance. Include the steps of your test and clearly state what you conclude from your test.

Hint: You test for autocorrelation in dynamic models with lagged *explanatory variables* just like you test for autocorrelation in static models.

Another hint: Don't forget to add an `NA` when adding the residuals to your dataset (see slide 49/64 from the autocorrelation notes if you don't remember).

Q15. If we have autocorrelation in the regression in **11**, what "problems" does it cause? What are our options for "living with" (or "fixing") the issues caused by autocorrelation?

Q16. Let's make sure our inference in **11** is robust to autocorrelation.

Step 1: Load the `lmtest` and `sandwich` packages, *i.e.*, `p_load(lmtest, sandwich)` (you must have loaded the `pacman` packge to use `p_load()`).
Step 2: Combine the `coeftest()` and `NeweyWest()` functions with your regression output from **11** to get **autocorrelation-robust standard errors** (also called *Newey West* standard errors).

For example, if your regression output in **11** is called `reg11` (the output of `lm()`), then you should run

```
# Load the packages
library(pacman)
p_load(lmtest, sandwich)
# Autocorrelation-robust standard errors
coeftest(reg11, NeweyWest(reg11))
```

Output your results. Does anything change?

Q17. One way the disturbance can have autocorrelation is when we've omitted a variable that is autocorrelated. Let's now specify our model as an ADL(1,1) (include the lag of the outcome and the lag for both explanatory variables). Keep everything in logs.

Adjust your standard errors to be robust to autocorrelation as we did above in **16**.

Include your regression results. Interpret the coefficient on the lag of logged property crime and comment on its significance.

Note: This regression should have an intercept two explanatory variables, the lags of the two explanatory variables, *and* the lag of the outcome variable.

Q18. Now that we've included the lag of our outcome variable (in **17**), use the residuals for evidence of autocorrelation. Remember that testing models with a lagged outcome variable differs from testing models without a lagged outcome variable.

Report your results/conclusions from the test.

Q19. Based upon everything you've seen in this assignment, what is your conclusion? Does a larger police force increase or decrease property crime? Or does it do nothing? Or do you need more data/information? Explain your answer.

Description of variables and names

Variable	Description
`year`	The year
`pop`	The population of Illinois (in 10,000s)
`unemployment`	The unemployment rate (between 0 and 1)
`n_officers`	The number of police officers employed in Illinois (in 10,000s)
`n_crime_violent`	The number of property crimes in Illinois (in 10,000s)
`n_crime_property`	The number of violent crimes in Illinois (in 10,000s)