

Problem Set 3

Consistency and Time Series

EC 421: Introduction to Econometrics

Due *before* midnight (11:59pm) on Thursday, 27 February 2020

DUE Your solutions to this problem set are due *before* midnight on Thursday, 27 February 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.

If your answers are not **in your own words**, then you will receive a zero for the assignment.

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: Consistency and unbiasedness

Let's first review the many ways we evaluate an estimator's performance.

1a. What is the formal (mathematical) definition of **unbiasedness**? (Your answer should include expected values.)

1b. Give a more intuitive definition of **unbiasedness** (no expected values).

1c. How does **consistency** differ from **unbiasedness**?

1d. Why do we care if an estimator is unbiased/consistent?

1e. Assume we **do not have autocorrelated disturbances**. Assume contemporaneous exogeneity is satisfied. For which of the following time-series models is OLS **unbiased**? Briefly explain your answer.

- static models
- models with lagged explanatory variables
- models with lagged outcome variables

1f. Assume we **do not have autocorrelated disturbances**. Assume contemporaneous exogeneity is satisfied. For which of the following time-series models is OLS **consistent**? Briefly explain your answer.

- static models
- models with lagged explanatory variables
- models with lagged outcome variables

1g. Now assume we **have autocorrelated disturbances**.

For which of the following time-series models is OLS **consistent**? Briefly explain your answer.

- static models
- models with lagged explanatory variables
- models with lagged outcome variables

Problem 2: Time Series

Imagine that we are interested in estimating the effect of monthly natural gas and coal prices on monthly residential electricity prices in the United States. The dataset `003-data.csv` contains these prices—the monthly average national (US) price for natural gas (the "citygate price"), the monthly average national price of coal (dollars per short ton in the spot market), the monthly national average price of electricity, and dates.

The table on the last page also describes the variables in this dataset.

Start: Read the `003-data.csv` file into R. Name the dataset `energy_df`.

2a. Estimate a static time-series model where the price of electricity is our outcome variable, and the prices of natural gas and coal are explanatory variables. Report/interpret your results.

2b. If the model in **2a.** is the true model, is OLS unbiased for the effect of natural gas prices on electricity prices? **Briefly explain** your answer.

2c. If the model in **2a.** is the true model, is OLS consistent for the effect of natural gas prices on electricity prices? **Briefly explain** your answer.

2d. Explain the limitations of a static model.

2e. Write out a model that includes both explanatory variables as their "current-period" effects (subscript t) and their "previous-period" effects (lags).

2f. Will OLS be unbiased for the model you wrote out in **2e**? Briefly explain your answer.

2g. Estimate the model you wrote out in **2e**. Interpret the coefficients on the lagged explanatory variables.

2h. Return to your static model and write out a model that **only** includes a lag of the dependent variable.

2i. In "ADL" notation, how would we write out the model from **2h.**, i.e., ADL(p,q)?

2j. Is OLS unbiased for the model in **2h**? Explain.

2k. Estimate the model from **2h**. Interpret the coefficient on the lag of the price of electricity.

2l. Based upon the three models you've estimated, which do you think is the "best" model for the price of electricity? Explain your reasoning.

2m. If we have a model such as

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{t-1} + \beta_3 x_{t-2} + \beta_4 x_{t-3} + u_t$$

what does $\beta_1 + \beta_2 + \beta_3 + \beta_4$ tell us?

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 (numeric)
price_coal	The average price of coal, \$ per short ton (numeric)
price_gas	The average price of natural gas, \$ per cubic ft (numeric)