Problem Set 3 Solutions

Consistency and Time Series

EC 421: Introduction to Econometrics

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

Problem 1: Consistency and unbiasedness

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

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

Answer Formally, $\text{Bias}_\beta\left(\hat{\beta}\right) = E\left[\hat{\beta}\right] - \beta$, so an **unbiased** estimator is one with zero bias, *i.e.*, $E\left[\hat{\beta}\right] = \beta$.

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

Answer An unbiased estimator will, on average, return the parameter it is estimating.

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

Answer Consistency tells us about an estimator's behavior as the sample size increases to infinity, whereas unbiasedness tells us about the estimator's mean for a fixed (finite) sample size.

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

Answer We generally want our estimator to return the "right" answer—the parameter it is estimating. If we the estimator is biased or inconsistent, then we cannot trust its results as being close to the true parameter.

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

Answer Without autocorrelation: OLS is unbiased for static models and models with lagged explanatory variables. OLS is biased for models with lagged explanatory variables because they violate exogeneity.

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

Answer Without autocorrelation: OLS is consistent for all three models—consistency only requires contemporaneous exogeneity.

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

Answer With autocorrelation, OLS is consistent for static models and models with lagged explanatory variables. OLS is inconsistent for models with lagged explanatory variables in presence of autocorrelation because they violate contemporaneous exogeneity.

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.

Answer First we load packages and our data. Then we estimate our model.

```
# Setup
library(pacman)
p_load(tidyverse, broom, lubridate, magrittr, here)
# Load energy data
energy_df = here("003-data.csv") %>% read_csv()
# Estimate static model
est_2a = lm(price_electricity ~ price_coal + price_gas, data = energy_df)
# Tidy results
est_2a %>% tidy()
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)  12.8       0.348      36.6  5.36e-72
## 2 price_coal    0.0685    0.0309      2.21 2.85e- 2
## 3 price_gas    -0.249     0.0352     -7.07 7.47e-11
```

In this part, we regress the price of electricity on the price of coal and the price of gas.

We estimate the coefficient on `price_coal` is approximately 0.068, which is statistically significant (different from zero) at the five-percent level. The interpretation of this coefficient: holding everything else constant, a one-dollar increase in coal prices is typically associated with an increase in residential electricity prices by approximately 0.068 dollars.

We estimate the coefficient on `price_gas` is approximately -0.249, which is also statistically significant (different from zero) at the five-percent level. The interpretation of this coefficient: holding everything else constant, a one-dollar increase in natural gas prices is typically associated with a **reduction** in residential electricity prices by approximately 0.249 dollars.

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.

Answer If **2a.** is the true model—meaning the true model is a static model—then OLS will be unbiased for the effect of electricity prices because OLS is unbiased for static models.

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.

Answer If **2a.** is the true model—meaning the true model is a static model—then OLS will be consistent for the effect of electricity prices because OLS is consistent for static models.

2d. Explain the limitations of a static model.

Answer The drawback of static models is that they do not allow for effects across time periods—the entire effect of a variable must happen in the same period (it cannot carry over to other periods).

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

Answer We now add the lag of each explanatory variable

$$
\begin{aligned}
(\text{Price of electricity})_t = & \beta_0 + \beta_1(\text{Price of coal})_t + \beta_2(\text{Price of nat. gas})_t + \\
& \beta_3(\text{Price of coal})_{t-1} + \beta_4(\text{Price of nat. gas})_{t-1} + u_t
\end{aligned}
$$

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

Answer OLS is unbiased for models with lagged explanatory variables, so it will be unbiased for the model in **2e**.

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

Answer Estimating the model:

```
# Estimate model with lagged explanatory variables
est_2g = lm(
  price_electricity ~ price_coal + price_gas + lag(price_coal) + lag(price_gas),
  data = energy_df
)
# Tidy results
est_2g %>% tidy()
```

```
## # A tibble: 5 x 5
##   term            estimate std.error statistic  p.value
##   <chr>              <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)     12.9        0.362     35.5     2.37e-69
## 2 price_coal       0.0385     0.0713     0.540   5.90e- 1
## 3 price_gas       -0.273      0.128     -2.13    3.47e- 2
## 4 lag(price_coal)  0.0348     0.0714     0.488   6.27e- 1
## 5 lag(price_gas)  -0.00606    0.119     -0.0510  9.59e- 1
```

In this part, we regress the price of electricity on the price of coal and the price of gas and their lags.

We estimate the coefficient on the (first) lag of `price_coal` is approximately 0.035, which is not statistically significant (different from zero) at the five-percent level. The interpretation of this coefficient: holding everything else constant, a one-dollar increase in *last month's* price of coal is typically associated with an increase in residential electricity prices by approximately 0.035 dollars.

We estimate the coefficient on the (first) lag of `price_gas` is approximately -0.006, which is also not statistically significant (different from zero) at the five-percent level. The interpretation of this coefficient: holding everything else constant, a one-dollar increase in the last month's price of natural gas is typically associated with a reduction in residential electricity prices by approximately 0.006 dollars.

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

Answer We now add the lag of each explanatory variable

$$(\text{Price of electricity})_t = \beta_0 + \beta_1(\text{Price of coal})_t + \beta_2(\text{Price of nat. gas})_t + \beta_3(\text{Price of electricity})_{t-1} + u_t$$

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

Answer In "ADL" notation, this model is ADL(1,0).

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

Answer Even without autocorrelation, OLS tends to be biased for the coefficients in a model with lagged explanatory variables because these lagged explanatory variables violate exogeneity.

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

Answer Estimating the ADL(1,0) model:

```r
# Estimate model with lagged explanatory variables
est_2k = lm(
  price_electricity ~ price_coal + price_gas + lag(price_electricity),
  data = energy_df
)
# Tidy results
est_2k %>% tidy()
```

```
## # A tibble: 4 x 5
##   term                   estimate std.error statistic  p.value
##   <chr>                     <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)              1.20      0.521       2.30 2.32e- 2
## 2 price_coal               0.0176    0.0137      1.29 2.01e- 1
## 3 price_gas               -0.0300    0.0198     -1.52 1.31e- 1
## 4 lag(price_electricity)   0.899     0.0383     23.5  3.38e-49
```

In this part, we regress the price of electricity on the price of coal and the price of gas and the lag of the price of electricity.

We estimate the coefficient on the (first) lag of `price_electricity` is approximately 0.899, which is (highly) statistically significant (different from zero at the five-percent level). The interpretation of this coefficient: holding everything else constant, a one-dollar increase in *the previous month's* price of electricity is typically associated with an increase in electricity prices by approximately 0.899 dollars.

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.

Answer In the first model (the static model), each of the explanatory variables (prices of coal and natural gas) is statistically significant. When we add their lags, the lags are not statistically significant, and only the price of natural gas remains statistically significant. Finally, when we include the price of electricity, the prices of natural gas and coal cease to be signficant. So, it seems pretty important to include a lag of the dependent variable in this case. That said, we know OLS is biased for the coefficient on the lagged dependent, so we should be careful.

Alternatively, we could consider the R^2 from each regression: 0.27, 0.291, 0.862, respectively. Thus, the model with the lagged price of electricity is explaining *substantially* more variation in the price of electricity.

More formally: if we use an LM test (which uses R^2), we find the same result: use a lag of the price of electricity.

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?

Answer The sum of the coefficients on a variable and its lags tell us a variable's *total* effect on the outcome—across all time periods. This *total* effect differs from the one-period effect.

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`)