ECONOMETRICS II
Practice Problem Set 1
Spring 2022-2023

The textbook is Jeffrey M. Wooldridge Introductory Econometrics: A Modern Approach, 6th ed., 2016, Cengage

1. Based on the simple regression model with only one explanatory variable and a constant, the following quantities are calculated (SSR is the Sum of Squared Residuals):

$$\boldsymbol{X}^\top \boldsymbol{X} = \begin{bmatrix} 7 & 2 \\ 2 & 2 \end{bmatrix}, \quad (\boldsymbol{X}^\top \boldsymbol{X})^{-1} = \begin{bmatrix} 0.2 & -0.2 \\ -0.2 & 0.7 \end{bmatrix}, \quad \boldsymbol{X}^\top \boldsymbol{y} = \begin{bmatrix} 22 \\ 12 \end{bmatrix}, \quad SSR = 30$$

$\boldsymbol{\theta} = [\theta_0, \theta_1]^\top$ is the vector of unknown population parameters and $\hat{\boldsymbol{\theta}} = [\hat{\theta}_0, \hat{\theta}_1]^\top$ is the vector of OLS estimators. Answer the following questions:

   (a) Find the sample mean of the dependent variable, $\bar{y}$ =...
   (b) Find the vector of OLS estimators, $\hat{\boldsymbol{\theta}}$ and write the estimation result of this regression in an equation form.
   (c) Find the variance-covariance matrix of OLS estimators and compute the standard error vector, $se(\hat{\boldsymbol{\theta}})$.
   (d) Find the vector of OLS estimators, $\hat{\boldsymbol{\theta}}$ and write the estimation result of this regression in an equation form.

2. Consider the following simple regression model with only one explanatory variable and a constant:

$$y = \theta_0 + \theta_1 x + u$$

We want to estimate this model using the following data given in vector form:

$$y = \begin{bmatrix} 4 & 8 & 8 & 6 \end{bmatrix}^\top \quad x = \begin{bmatrix} 1 & 2 & 3 & 4 \end{bmatrix}^\top$$

Let 2×1 vector of unknown population parameters be $\boldsymbol{\theta} = [\theta_0, \theta_1]^\top$, and OLS estimators $\hat{\boldsymbol{\theta}} = [\hat{\theta}_0, \hat{\theta}_1]^\top$. Using these:

   (a) Compute OLS estimates, $\hat{\boldsymbol{\theta}}$, and write the sample regression function in equation form.

   (b) Compute the residual vector and the standard errors of OLS estimates.

3. Let $\{\epsilon_t : \ t = 0, 1, 2, \ldots\}$ be a white noise process with mean 0 and variance 1. The following stochastic process is defined:

$$x_t = 2 + \epsilon_t - 0.8\epsilon_{t-1}, \quad t = 1, 2, \ldots$$

Answer the following questions.

   (a) $E(x_t)$ =?

   (b) $Var(x_t)$ =?

(c) $Cov(x_t, x_{t-1}) =?$ and $Cov(x_t, x_{t-2}) =?$

(d) Is this process (covariance) stationary? Is x_t weakly dependent?

4. Consider the following stochastic process

$$y_t = y_{t-1} + \epsilon_t, \quad y_0 = 0$$

where ϵ_t is a white noise process with mean 0 and variance $\sigma_\epsilon^2 = 0.5$ and y_0 is the initial condition.

(a) Show that the unconditional mean of the process is 0.
(b) Show that the unconditional variance of the process is a function of time and hence it cannot be stationary.
(c) This process is called ....
(d) What kind of transformation is required to make the process stationary? Briefly explain.
(e) What's your best forecast for $t+1$? Find $\mathrm{E}[y_{t+1}|y_t] =$...

5. Now, consider the same stochastic process in the previous question but add a constant (drift) term:

$$y_t = 4 + y_{t-1} + \epsilon_t, \quad y_0 = 0$$

where ϵ_t is a white noise process with mean 0 and variance $\sigma_\epsilon^2 = 0.5$ and y_0 is the initial condition.

(a) Show that the unconditional mean of the process is now a function of time.
(b) Show that the unconditional variance of the process is a function of time.
(c) This process is called ....
(d) What kind of transformation is required to make the process stationary? Briefly explain.

6. Consider the following stochastic process

$$y_t = 2 + 0.1t + \epsilon_t, \quad t = 1, 2, \ldots$$

where ϵ_t is a white noise process with mean 0 and variance $\sigma_\epsilon^2 = 2$.

(a) Show that the unconditional mean of the process is a function of time. Also find $\mathrm{E}[y_{t+1}] =?$
(b) Find the unconditional variance of the process.
(c) Is this process stationary? Briefly explain.
(d) This process is called ....
(e) What kind of transformation is required to make the process stationary? Briefly explain.

7. Exercise from the textbook, ch.10 Exercise 1 on p. 339.

Decide if you agree or disagree with each of the following statements and give a brief explanation of your decision:

i. Like cross-sectional observations, we can assume that most time series observations are independently distributed.

ii. The OLS estimator in a time series regression is unbiased under the first three Gauss-Markov assumptions.

iii. A trending variable cannot be used as the dependent variable in multiple regression analysis.

iv. Seasonality is not an issue when using annual time series observations.

8. Exercise from the textbook, ch.10 Exercise 2 on p. 339 .

Let $gGDP_t$ denote the annual percentage change in gross domestic product, and let int_t denote a short-term interest rate. Suppose that $gGDP_t$ is related to interest rates by

$$gGDP_t = \alpha_0 + \delta_0 int_t + \delta_1 int_{t-1} + u_t, \tag{1}$$

where u_t is uncorrelated with int_t, int_{t-1}, and all other past values of interest rates. Suppose that the Federal Reserve / Central Bank follows the policy rule:

$$int_t = \gamma_0 + \gamma_1(gGDP_{t-1} - 3) + \nu_t,$$

where $\gamma_1 > 0$. (When last year?s GDP growth is above 3%, the FED increases interest rates to prevent an ?overheated? economy.) If v_t is uncorrelated with all past values of int_t and u_t, argue that int_t must be correlated with u_{t-1}. (Hint: Lag the first equation for one time period and substitute for $gGDP_{t-1}$ in the second equation.) Which Gauss-Markov assumption does this violate?

9. Exercise from the textbook, ch.10 Exercise 8 on p. 340.

In the linear model given in equation (10.8), the explanatory variables $x_t = (x_{t1}, \dots, x_{tk})$ are said to be *sequentially exogenous* (sometimes called *weakly exogenous*) if

$$\mathrm{E}(u_t|x_t, x_{t-1}, \dots, x_1) = 0, \quad t = 1, 2, \dots,$$

so that the errors are unpredictable given current and all *past* values of the explanatory variables.

i. Explain why sequential exogeneity is implied by strict exogeneity.

ii. Explain why contemporaneous exogeneity is implied by sequential exogeneity.

iii. Are the OLS estimators generally unbiased under the sequential exogeneity assumption? Explain.

iv. Consider a model to explain the annual rate of HIV infections (HIVrate) as a distributed lag of per capita condom usage (pccon) for a state, region, or province:

$$\mathrm{E}(HIVrate_t|pccon_t, pccon_{t-1}, \dots,) = \alpha_0 + \delta_0 pccon_t + \delta_1 pccon_{t-1} + \delta_2 pccon_{t-2} + \delta_3 pccon_{t-3}.$$

Explain why this model satisfies the sequential exogeneity assumption. Does it seem likely that strict exogeneity holds too?

10. Exercise from the textbook, ch.11 Exercise 2 on p. 365.

Let $\{e_t : t = -1, 0, 1, ...\}$ be a sequence of independent, identically distributed random variables with mean zero and variance one. Define a stochastic process by

$$x_t = e_t - \frac{1}{2}e_{t-1} + \frac{1}{2}e_{t-2}, \quad t = 1, 2, ...$$

i. Find $\mathrm{E}(x_t)$ and $\mathrm{Var}(x_t)$. Do either of these depend on t?

ii. Show that $\mathrm{Corr}(x_t, x_{t+1}) = -1/2$ and $\mathrm{Corr}(x_t, x_{t+2}) = 1/3$.

iii. What is $\mathrm{Corr}(x_t, x_{t+h})$for $h > 2$?

iv. Is $\{x_t\}$ an asymptotically uncorrelated process?

11. Exercise from the textbook, ch.11 Exercise 3 on p. 366 Suppose that a time series process y_t is generated by $y_t = z + e_t$, for all $t = 1, 2, \ldots$, where e_t is an i.i.d. sequence with mean zero and variance σ_e^2. The random variable z does not change over time; it has mean zero and variance σ_z^2. Assume that each e_t is uncorrelated with z.

    i. Find the expected value and variance of y_t. Do your answers depend on t?

    ii. Find $Cov(y_t, y_{t+h})$ for any t and h. Is y_t covariance stationary?

    iii. Use parts (i) and (ii) to show that $Corr(y_t, y_{t+h}) = \sigma_z^2/(\sigma_z^2 + \sigma_e^2)$ for all t and h.

    iv. Does y_t satisfy the intuitive requirement for being asymptotically uncorrelated? Explain.

12. Exercise from the textbook, ch.11 Exercise 6 on p. 366.

    Let $hy6_t$ denote the three-month holding yield (in percent) from buying a six-month T-bill at time $t-1$ and selling it at time t (three months hence) as a three-month T-bill. Let $hy3_{t-1}$ be the three-month holding yield from buying a three-month T-bill at time $t-1$. At time $t-1$, $hy3_{t-1}$ is known, whereas $hy6_t$ is unknown because $p3_t$ (the price of three-month T-bills) is unknown at time $t-1$. The *expectations hypothesis* (EH) says that these two different three-month investments should be the same, on average. Mathematically, we can write this as a conditional expectation:

$$\mathrm{E}(hy6_t | I_{t-1}) = hy3_{t-1},$$

    where I_{t-1} denotes all observable information up through time $t-1$. This suggests estimating the model

$$hy6_t = \beta_0 + \beta_1 hy3_{t-1} + u_t,$$

    and testing $\mathrm{H}_0 : \beta_1 = 1$. (We can also test $\mathrm{H}_0 : \beta_0 = 0$, but we often allow for a *term premium* for buying assets with different maturities, so that $\beta_0 \neq 0$.)

    i. Estimating the previous equation by OLS using the data in INTQRT (spaced every three months) gives

$$\widehat{hy6}_t = \underset{(.070)}{-.058} + \underset{(.039)}{1.104} hy3_{t-1}$$
$$n = 123, \quad R^2 = .866.$$

    Do you reject $\mathrm{H}_0 : \beta_1 = 1$ against $\mathrm{H}_0 : \beta_1 \neq 1$ at the 1% significance level? Does the estimate seem practically different from one?

ii. Another implication of the EH is that no other variables dated as $t-1$ or earlier should help explain $hy6_t$, once $hy3t-1$ has been controlled for. Including one lag of the *spread* between six-month and three-month T-bill rates gives

$$\widehat{hy6}_t = \underset{(.067)}{-.123} + \underset{(.039)}{1.053} hy3_{t-1} + \underset{(.109)}{.480} (r6_{t-1} - r3_{t-1})$$
$$n = 123, \quad R^2 = .865.$$

Now, is the coefficient on $hy3_{t-1}$ statistically different from one? Is the lagged spread term significant? According to this equation, if, at time $t-1$, $r6$ is above $r3$, should you invest in six-month or three-month T-bills?

iii. The sample correlation between $hy3_t$ and $hy3_{t-1}$ is .914. Why might this raise some concerns with the previous analysis?

iv. How would you test for seasonality in the equation estimated in part (ii)?

13. Computer exercise from the textbook, ch.10 Exercise C5 on p. 341.
Use the data in EZANDERS for this exercise. The data are on monthly unemployment claims in Anderson Township in Indiana, from January 1980 through November 1988. In 1984, an enterprise zone (EZ) was located in Anderson (as well as other cities in Indiana). [See Papke (1994) for details.]

i. Regress log($uclms$) on a linear time trend and 11 monthly dummy variables. What was the overall trend in unemployment claims over this period? (Interpret the coefficient on the time trend.) Is there evidence of seasonality in unemployment claims?

ii. Add ez, a dummy variable equal to one in the months Anderson had an EZ, to the regression in part (i). Does having the enterprise zone seem to decrease unemployment claims? By how much? [You should use formula (7.10) from Chapter 7.]

iii. What assumptions do you need to make to attribute the effect in part (ii) to the creation of an EZ?