Exercises (with Solutions) · Chapter 5

1. **Problem**
   To answer the following questions, you can refer to Assumptions 1-4 of Chapter 3 or Assumptions 1-4* of Chapter 5 (both is possible).
   (a) Which assumptions are needed for the unbiasedness of the Ordinary Least Squares (OLS) estimator in the linear regression model?
   (b) Which additional assumptions are needed to make OLS the best linear unbiased estimator?
   (c) Correct or false: The phrase "linear" in part (b) refers to the fact that we are estimating a linear model.

   Solution

   (a) Assumption 1 must hold, i.e. that the relationship between Y_i and X_i is given by

   $$Y_i = \sum_{k=1}^{K} \beta_k X_{ik} + \varepsilon_i \,.$$

   with $X_{i1} = 1$ for all i and (Y_i, X_i) being i.i.d. across $i = 1, \dots, n$.
   Unbiasedness of the Ordinary Least Squares (OLS) estimator $\hat{\beta}$ follows then from the strict exogeneity assumption $E(\varepsilon\,|X) = 0$, where strict exogeneity follows from our Assumption 2 (i.e. $E(\varepsilon_i\,|X_i) = 0$ for all i) together with Assumption 1 (i.e. (Y_i, X_i) is i.i.d. across $i = 1, \dots, n$).

   (b) The Gauss-Markov theorem states that, under certain assumptions, OLS is the best linear unbiased estimator. "Best" refers to the fact that OLS has the smallest variance conditional on X of all linear unbiased estimators, given the assumptions are true. The assumptions needed for the Gauss-Markov theorem are:
   1. Assumption 1 (correct data generating process)
   2. $E(\varepsilon\,|X) = 0$ (strict exogeneity, already required for unbiasedness. Follows from Assumptions 1 and 2)
   3. Spherical errors, i.e. $\mathrm{Var}(\varepsilon\,|X) = \sigma^2 I_n$ (homoscedastic, uncorrelated error terms.)

   So, the additional assumption are spherical errors.

   (c) This statement is false. The phrase 'linear' here refers to the fact that the OLS estimator is a linear function of Y since we can write $\hat{\beta} = CY$ where C is the $(K \times n)$ matrix $C = (X'X)^{-1}X'$. (By contrast, the model $E(Y|X) = X\beta$ is called linear because the mean of $Y|X$ is a linear function of the unknown parameters β.)

2. **Problem**

   (a) Explain, why the elasticity

   $$El_x f(x) = \frac{x}{f(x)} \frac{\partial f(x)}{\partial x}$$

   of a deterministic and differentiable function f, with $y = f(x)$, can be interpreted as the approximate percentage change in y per 1% change in x.

   Moreover, show that the elasticity with respect to x can be written as

   $$El_x f(x) = \frac{\partial \log(f(x))}{\partial \log(x)}.$$

(b) Consider the following log-log regression model (our regularity Assumptions 1-4 are assumed to be true):

$$\log(Y_i) = \gamma_1 + \gamma_2 \log(X_i) + \varepsilon_i, \quad i = 1, \dots, n. \tag{1}$$

The parameter γ_2 in this regression model is often interpreted as the elasticity of $f(x) = E(Y|X = x)$ with respect to $x > 0$, i.e. as the approximate percentage change in $E(Y|X = x)$ per 1% change in x. When is this interpretation true?

Solution

(a) In the deterministic case, $y = f(x)$, the elasticity of the differentiable function $f(x)$ with respect to x is defined as

$$\begin{aligned} El_x f(x) &= \frac{x}{f(x)} \frac{\partial f(x)}{\partial x} = \frac{\frac{\partial f(x)}{f(x)}}{\frac{\partial x}{x}} && \text{(for } f(x) \neq 0) \\ &\approx \frac{\frac{f(x+\Delta)-f(x)}{f(x)} \cdot 100\%}{\frac{x+\Delta-x}{x} \cdot 100\%} && \text{(for small } \Delta > 0) \\ &= \frac{\%\text{ change in } f(x)}{\%\text{ change in } x} \end{aligned}$$

Thus, the elasticity $El_x f(x)$ can be interpreted as the approximate percentage change of $f(x)$ for 1% change of x.

The following alternative expression for the elasticity $El_x f(x)$ is often useful (remember $\partial \log(x)/\partial(x) = 1/x$):

$$\begin{aligned} El_x f(x) &= \quad x \quad \frac{1}{f(x)} \quad \frac{\partial f(x)}{\partial x} && \text{(for } f(x) \neq 0) \\ &= \underbrace{\frac{\partial x}{\partial \log(x)}}_{=x} \underbrace{\frac{\partial \log(f(x))}{\partial f(x)}}_{=1/f(x)} \frac{\partial f(x)}{\partial x} && \text{(only if } f(x) > 0 \text{ and } x > 0\text{; otherwise the } \log\text{'s aren't defined.)} \\ &= \frac{\partial \log(f(x))}{\partial \log(x)} \end{aligned}$$

Note. In a stochastic setting (e.g., $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$) one considers the elasticity $El_x f(x)$ with respect to x for the *conditional mean function* $f(x) = E(Y_i|X_i = x)$.

(b) We are asked to show when γ_2 equals the elasticity of $f(x) = E(Y|X = x)$. First, observe that

$$\begin{aligned} \log(Y) &= \gamma_1 + \gamma_2 \log(X) + \varepsilon \\ \Leftrightarrow \; Y &= \exp(\gamma_1 + \gamma_2 \log(X) + \varepsilon) \\ &= \exp(\gamma_1 + \gamma_2 \log(X)) \exp(\varepsilon), \end{aligned}$$

where we skipped the index i since the same result applies anyways to each $i = 1, \dots, n$.
We can now use this equivalent expression for model (1), i.e. $Y = \exp(\gamma_1 + \gamma_2 \log(X)) \exp(\varepsilon)$, and the above alternative elasticity expression, i.e. $El_x f(x) = \partial \log(f(x))/\partial \log(x)$, to

do the following derivations:

$$\begin{aligned}
El_x f(x) &= El_x E[Y|X=x] \\
&= \frac{\partial \log\big\{E[Y|X=x]\big\}}{\partial \log(x)} \\
&= \frac{\partial \log\big\{E\left[\exp(\gamma_1+\gamma_2\log(X))\exp(\varepsilon)|X=x\right]\big\}}{\partial \log(x)} \quad \text{(using } Y=\exp(\gamma_1+\gamma_2\log(X))\exp(\varepsilon)\text{)} \\
&= \frac{\partial \log\big\{\exp(\gamma_1+\gamma_2\log(X))\cdot E\left[\exp(\varepsilon)|X=x\right]\big\}}{\partial \log(x)} \\
&= \frac{\partial\Big(\log\{\exp(\gamma_1+\gamma_2\log(x))\}+\log\{E[\exp(\varepsilon)|X=x]\}\Big)}{\partial \log(x)} \\
&= \frac{\partial\Big(\gamma_1+\gamma_2\log(x)+\log\{E[\exp(\varepsilon)|X=x]\}\Big)}{\partial \log(x)} \\
&= \gamma_2 + \frac{\partial\Big(\log\{E[\exp(\varepsilon)|X=x]\}\Big)}{\partial \log(x)}
\end{aligned}$$

So, γ_2 equals the elasticity of the conditional mean function $f(x)=E(Y|X=x)$ if

$$\frac{\partial(\log\{E[\exp(\varepsilon)|X=x]\})}{\partial\log(x)}=0.$$

This is the case, for instance, if ε and X are *independent* since then $E[\exp(\varepsilon)|X=x]=c$ for some constant $c>0$ which does not depend on x, such that

$$\begin{aligned}
\frac{\partial(\log\{E[\exp(\varepsilon)|X=x]\})}{\partial\log(x)} &= \frac{\partial\Big(\log\{c\}\Big)}{\partial\log(x)} \\
&= \frac{\partial\Big(\log\{c\}\Big)/\partial x}{\partial\log(x)/\partial x} = \frac{0}{\frac{1}{x}} = 0\cdot x = 0
\end{aligned}$$

Note. Strict exogeneity $E[\varepsilon\,|X=x]=0$ does not imply that $E[\exp(\varepsilon)|X=x]=c$ for some constant $c>0$.

3. **Problem**

Consider the following multiple linear regression model

$$Y_i=\beta_1+\beta_2X_{i2}+\beta_3X_{i3}+\varepsilon_i,\quad i=1,\dots,n.$$

Assumptions 1-4* of Chapter 5 are assumed to hold. Consider a smallish sample size of $n=25$ and let

$$\begin{aligned}
\beta=(\beta_1,\beta_2,\beta_3)' &= (2,3,4)' \\
X_{i2} &\sim U[1,4] \\
V_i &\sim N(0,1) \\
X_{i3} &= 2X_{i2}+V_i \\
\varepsilon_i\,|X &\sim N(0,2/3)
\end{aligned}$$

Hint: Observe that X_2 and X_3 are constructed to correlate with each other.

This is how you can draw $n = 25$ realizations of the regressors and the error term. Observe the strong correlation between X_2 and X_3.

```
> set.seed(1234)
> n   <- 25
> X_2 <- runif(n, 1, 4)
> V   <- rnorm(n)
> X_3 <- 2 * X_2 + V
> eps <- rnorm(n, sd=sqrt(2/3))
> ## Sample correlation between X_2 and X_3:
> cor(X_2, X_3)

[1] 0.8547267
```

Take the following X matrix as the observed X to condition on.

```
> X   <- cbind(rep(1, n), X_2, X_3)
```

(a) Write a Monte Carlo (MC) simulation with 500 MC replications to produce 500 realizations of the OLS estimator $\hat{\beta}_2$ for β_2 conditional on X. Approximate the conditional bias of $\hat{\beta}_2$ given X, i.e. $\mathrm{Bias}(\hat{\beta}_2|X)$, using the 500 realizations of $\hat{\beta}_2|X$. What do you observe?

(b) Repeat the MC simulation in (a), but, when computing the OLS estimates, omit X_{i3} from the estimation formula for all $i = 1, \dots, n$. Approximate again conditional bias of $\hat{\beta}_2$ given X, i.e. $\mathrm{Bias}(\hat{\beta}_2|X)$, using the 500 realizations of $\hat{\beta}_2|X$. What do you observe?

(c) Does omitting X_{i3}, $i = 1, \dots, n$, violate the exogeneity assumption (Assumption 2)? Explain your answer using mathematical derivations.

Solution

(a) R-Coding:

```
> myDataGenerator <- function(n, beta, condition_on_X){
+   ##
+   X    <- condition_on_X
+   ##
+   eps  <- rnorm(n, mean = 0, sd=sqrt(2/3))
+   Y    <- X %*% beta + eps
+   ##
+   data <- data.frame("Y"=Y, "X_2"=X[,2], "X_3"=X[,3])
+   ##
+   return(data)
+ }
> ## Defining the data generating process:
> n          <- 25
> beta_true  <- c(2,3,4)
> rep        <- 500 # MC replications
> beta_hat_2 <- rep(NA, times=rep)
> ## Generate data and compute estimates
> ## 500 times and safe all
```

```
> ## 500 estimates:
> for(r in 1:rep){
+   MC_data       <- myDataGenerator(n=n, beta = beta_true, condition_on_X=X)
+   lm_obj        <- lm(Y ~ X_2 + X_3, data = MC_data)
+   beta_hat_2[r] <- coef(lm_obj)[2]
+ }
> ## MC approximation of the bias of \hat\beta_2|X
> mean(beta_hat_2) - beta_true[2]

[1] -0.01566869
```

The Monte Carlo simulation indicates that the bias of $\hat{\beta}_2|X$ is zero.

(b) R-Coding:

```
> ## Now X_3 is omitted from the regression equation:
> for(r in 1:rep){
+   ## Data generation using both X2 and X3:
+   MC_data <- myDataGenerator(n= n, beta = beta_true, condition_on_X=X)
+   ## estimation with X3 omitted
+   lm_obj        <- lm(Y ~ X_2, data = MC_data)
+   beta_hat_2[r] <- coef(lm_obj)[2]
+ }
> ## The sample mean of the 500 estimates for \beta_2|X
> mean(beta_hat_2)  - beta_true[2]

[1] 7.578456
```

The Monte Carlo simulation indicates a **large** bias of $\hat{\beta}_2|X$.

(c) The simulation results indicate a bias problem in (b). This bias is called **Omitted Variable Bias**.
The regression model in (b) is

$$Y_i = \beta_1 + \beta_2 X_{i2} + u_i$$

with

$$u_i = \varepsilon_i + \beta_3 X_{i3}.$$

Let $\beta = (\beta_1, \beta_2)'$, $Y = (Y_1, \dots, Y_n)'$, $X_2 = (X_{12}, \dots, X_{n2})'$, $u = (u_1, \dots, u_n)'$, and

$$X = \begin{pmatrix} 1 & X_{12} \\ \vdots & \vdots \\ 1 & X_{n2} \end{pmatrix}$$

Due to omitting X_{i3}, which is correlated with X_{i2}, the exogeneity assumption is violated:

$$\begin{aligned} E(u_i|X) &= E(\varepsilon_i + \beta_3 X_{i3}|X) \\ &= E(\varepsilon_i + \beta_3(2X_{i2} + V_i)|X) \\ &= E(\varepsilon_i + \beta_3 2X_{i2} + \beta_3 V_i|X) \\ &= \underbrace{E(\varepsilon_i\,|X)}_{=0} + \underbrace{E(2\beta_3 X_{i2}|X)}_{=2\beta_3 X_{i2}} + \underbrace{E(\beta_3 V_i|X)}_{=\beta_3 E(V_i)=0} = 2\beta_3 X_{i2} \end{aligned}$$

Thus, staking all observations, $i = 1, \dots, n$, we have that

$$E(u|X) = \underbrace{2\beta_3 X_2}_{(n\times 1)}$$

This allows us to derive the conditional mean of the estimator $\hat\beta = (\hat\beta_1, \hat\beta_2)'$, given X, as following

$$\begin{aligned} E(\hat\beta|X) &= E((X'X)^{-1}X'Y|X) \\ &= \beta + E((X'X)^{-1}X'u|X) \\ &= \beta + (X'X)^{-1}X'E(u|X) \\ &= \beta + (X'X)^{-1}X'2\beta_3 X_2 \\ \Rightarrow \quad \mathrm{Bias}(\hat\beta|X) &= (X'X)^{-1}X'2\beta_3 X_2 \end{aligned}$$

where $(X'X)^{-1}X'2\beta_3X_2$ is generally not zero. Thus, $\hat\beta$ is a biased estimator.

Note: Since this estimation bias is caused by an omitted variable, this type of a bias is called **omitted variables bias**.

4. **Problem**

Correct or false? Justify your answers.

a) A regression of the OLS residuals on the regressors included in the model yields, by construction, an R^2 of zero.

b) If an estimate $\hat\beta_k$ is significantly different from zero at the 10% level, it is also significantly different from zero at the 5% level.

Solution

a) Correct. This follows from the *orthogonality* between X_k and $\hat\varepsilon$ for all $k = 1, \dots, K$, i.e. $X'\hat\varepsilon = 0$, where 0 is here a $K \times 1$ column-vector (see Chapter 3.3). In a regression of $\hat\varepsilon$ on X, the OLS estimator would be $\hat\gamma = (X'X)^{-1}X'\hat\varepsilon$ which is a vector of zeros ($\hat\gamma = 0$) since $X'\hat\varepsilon = 0$. Let denote the fitted residuals by $\hat e = X\hat\gamma$ and note that $\hat e = 0$ since $\hat\gamma = 0$. Consequently, we also have that $\bar{\hat e} = n^{-1}\sum_{i=1}^n \hat e_i = 0$. This now allows us to show that $R^2 = 0$:

$$R^2 = \frac{\sum_{i=1}^n \left(\hat e_i - \bar{\hat e}\right)^2}{\sum_{i=1}^n \left(\hat\varepsilon_i - \bar{\hat\varepsilon}\right)^2} = \frac{0}{\sum_{i=1}^n \hat\varepsilon_i^2} = 0,$$

where $\bar{\hat\varepsilon} = n^{-1}\sum_{i=1}^n \hat\varepsilon_i = 0$ since the intercept $(1, \dots, 1)'$ and $\hat\varepsilon$ are orthogonal.

b) False. Actually, it works the other way around. If an estimate is significantly different from zero at the 5% level, it is also significantly different from zero at the 10% level. This is most easily explained on the basis of the p-value. If e.g. the p-value belonging to the observed value of the t test statistic, t_{obs}, is smaller than 0.05 (5%), the null hypothesis $H_0\colon \beta_k = 0$ can be rejected and we say that the estimate $\hat\beta_k$ is significantly different from zero at the 5% significance level. Clearly, if p is smaller than 0.05 it is also smaller than 0.10 (10%) but not vice versa.

5. **Problem**

Consider the following multiple linear regression model:

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i \quad i = 1, \dots, n,$$

where n is a small sample size (e.g. $n = 15$). Assume the model assumptions of Chapter 5 are fulfilled.

(a) How can one test the hypothesis that $\beta_3 = 1$ against a two-sided alternative. State H_0, H_A, the test-statistic, its distribution, and explain how the test decision is conduced.

(b) Consider the null hypothesis $\beta_2 + \beta_3 = 0$ against a two-sided alternative. Use mathematical derivations to show that in this case an F-test simplifies to a t-test.

(c) How can one test the hypothesis that $\beta_2 = \beta_3 = 0$. State H_0, H_A, the test-statistic, its distribution, and explain how the test decision is conduced.

Solution

(a) The null hypothesis $H_0 : \beta_3 = 1$ can be tested versus the two-sided alternative $H_A : \beta_3 \neq 1$ by means of a t-test. The test statistic is

$$t = \frac{\hat{\beta}_3 - 1}{\widehat{\mathrm{SE}}\left(\hat{\beta}_3 | X\right)},$$

where

$$\widehat{\mathrm{SE}}\left(\hat{\beta}_3 | X\right) = \sqrt{s_{UB}^2 \left[(X'X)^{-1}\right]_{33}}.$$

Under the null hypothesis and under Assumptions 1-4*, the test statistic t has a t distribution with $(n - K) = (n - 3)$ degrees of freedom. At the $\alpha = 5\%$ significance level, we reject the null if $|t_{\text{obs}}| > c_{1-\alpha/2}$ (two-tailed test), where t_{obs} is the observed value of the test statistic, and where $c_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the t distribution with $(n - 3)$ degrees of freedom.

(b) The hypothesis $H_0 : \beta_2 + \beta_3 = 0$ can be tested by means of a t-test since it corresponds effectively to one single hypothesis (i.e., $q = 1$). The null hypothesis is equivalent to $H_0 : R\beta - r = 0$ with $R = (0, 1, 1)$ and $r = 0$. That is, the F-test simplifies to a t-test as following:

$$F = (R\hat{\beta}_n - r)'[R(s_{UB}^2(X'X)^{-1})R']^{-1}(R\hat{\beta}_n - r)/q$$

$$F = \frac{(R\hat{\beta}_n - r)'(R\hat{\beta}_n - r)}{R(s_{UB}^2(X'X)^{-1})R'} \quad \text{(Since } R(s_{UB}^2(X'X)^{-1})R' \text{ is a scalar as } q = 1.)$$

$$F = \frac{\left(\hat{\beta}_2 + \hat{\beta}_3\right)^2}{s_{UB}^2\left[(X'X)^{-1}\right]_{22} + s_{UB}^2\left[(X'X)^{-1}\right]_{33} + 2s_{UB}^2\left[(X'X)^{-1}\right]_{23}}$$

$$F = \frac{\left(\hat{\beta}_2 + \hat{\beta}_3\right)^2}{\widehat{\mathrm{Var}}\left(\hat{\beta}_2 | X\right) + \widehat{\mathrm{Var}}\left(\hat{\beta}_3 | X\right) + 2 \cdot \widehat{\mathrm{Cov}}\left(\hat{\beta}_2, \hat{\beta}_3 | X\right)}$$

$$F = \frac{\left(\hat{\beta}_2 + \hat{\beta}_3\right)^2}{\widehat{\mathrm{Var}}\left(\hat{\beta}_2 + \hat{\beta}_3 | X\right)} \overset{H_0}{\sim} F_{(1,n-3)}$$

$$\Leftrightarrow \quad t = \frac{\hat{\beta}_2 + \hat{\beta}_3}{\widehat{\mathrm{SE}}\left(\hat{\beta}_2 + \hat{\beta}_3 | X\right)} \overset{H_0}{\sim} t_{(n-3)}$$

(Remember that for two random variables Y and Z we have $\mathrm{Var}(Y + Z) = \mathrm{Var}(Y) + \mathrm{Var}(Z) + 2\,\mathrm{Cov}(Y, Z)$)

(c) The joint null hypothesis that $\beta_2 = \beta_3 = 0$ can be tested by means of an F-test. The null hypothesis is $H_0 : R\beta - r = 0$ with the following $(q \times K) = (2 \times 3)$ matrix R and $q = 2$ vector r :

$$R = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

The alternative hypothesis is $H_A : R\beta - r \neq 0$. The test statistic is given by

$$F = (R\hat{\beta}_n - r)'[R(s^2_{UB}(X'X)^{-1})R']^{-1}(R\hat{\beta}_n - r)/q$$

We compare the observed test statistic, F_{obs}, with the critical value $c_{1-\alpha}$ from an F distribution with $q = 2$ (the number of restrictions) and $n - 3$ degrees of freedom, where $c_{1-\alpha}$ is the $(1 - \alpha)$ quantile of the F distribution with $(q, n - K) = (2, n - 3)$ degrees of freedom. If $F_{\text{obs}} > c_{1-\alpha}$, we reject the null hypothesis.