Solutions Manual

to Accompany

Introduction to
Econometrics

Stock • Watson

Jiangfeng Zhang

University of California at Berkeley

Addison Wesley

Boston San Francisco New York
London Toronto Sydney Tokyo Singapore Madrid
Mexico City Munich Paris Cape Town Hong Kong Montreal

Solutions Manual to accompany Stock/Watson, *Introduction to Econometrics*

ISBN: 0-201-79612-0
1 2 3 4 5 6 7 8 9 10-DPC-0605040302

Exercise Solutions for Stock and Watson's Introduction to Econometrics

by Jiangfeng Zhang

July 2002

Chapter 1

Economic Questions and Data

There are no exercises in Chapter 1.

Chapter 2
Review of Probability

2.1. We know from Table 2.2 that $\Pr(Y=0) = 0.22$, $\Pr(Y=1) = 0.78$, $\Pr(X=0) = 0.30$, $\Pr(X=1) = 0.70$. So

(a)

$$\begin{aligned}\mu_Y &= E(Y) = 0 \times \Pr(Y=0) + 1 \times \Pr(Y=1) \\ &= 0 \times 0.22 + 1 \times 0.78 = 0.78,\end{aligned}$$

$$\begin{aligned}\mu_X &= E(X) = 0 \times \Pr(X=0) + 1 \times \Pr(X=1) \\ &= 0 \times 0.30 + 1 \times 0.70 = 0.70.\end{aligned}$$

(b)

$$\begin{aligned}\sigma_X^2 &= E\left[(X-\mu_X)^2\right] \\ &= (0-0.70)^2 \times \Pr(X=0) + (1-0.70)^2 \times \Pr(X=1) \\ &= (-0.70)^2 \times 0.30 + 0.30^2 \times 0.70 = 0.21,\end{aligned}$$

$$\begin{aligned}\sigma_Y^2 &= E\left[(Y-\mu_Y)^2\right] \\ &= (0-0.78)^2 \times \Pr(Y=0) + (1-0.78)^2 \times \Pr(Y=1) \\ &= (-0.78)^2 \times 0.22 + 0.22^2 \times 0.78 = 0.1716.\end{aligned}$$

(c) Table 2.2 shows $\Pr(X=0, Y=0) = 0.15$, $\Pr(X=0, Y=1) = 0.15$, $\Pr(X=1, Y=0) = 0.07$, $\Pr(X=1, Y=1) = 0.63$. So

$$\begin{aligned}\sigma_{XY} &= \text{cov}(X,Y) = E[(X-\mu_X)(Y-\mu_Y)] \\ &= (0-0.70)(0-0.78)\Pr(X=0, Y=0) \\ &\quad + (0-0.70)(1-0.78)\Pr(X=0, Y=1) \\ &\quad + (1-0.70)(0-0.78)\Pr(X=1, Y=0) \\ &\quad + (1-0.70)(1-0.78)\Pr(X=1, Y=1) \\ &= (-0.70) \times (-0.78) \times 0.15 + (-0.70) \times 0.22 \times 0.15 \\ &\quad + 0.30 \times (-0.78) \times 0.07 + 0.30 \times 0.22 \times 0.63 \\ &= 0.084,\end{aligned}$$

$$\text{corr}(X,Y) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{0.084}{\sqrt{0.21 \times 0.1716}} = 0.4425.$$

2.2. For the two new random variables $W = 3 + 6X$ and $V = 20 - 7Y$, we have:

(a)

$$E(V) = E(20 - 7Y) = 20 - 7E(Y) = 20 - 7 \times 0.78 = 14.54,$$

$$E(W) = E(3 + 6X) = 3 + 6E(X) = 3 + 6 \times 0.70 = 7.2.$$

(b)

$$\sigma_W^2 = \text{var}(3 + 6X) = 6^2 \cdot \sigma_X^2 = 36 \times 0.21 = 7.56,$$

$$\sigma_V^2 = \text{var}(20 - 7Y) = (-7)^2 \cdot \sigma_Y^2 = 49 \times 0.1716 = 8.4084.$$

(c)

$$\sigma_{WV} = \text{cov}(3 + 6X, 20 - 7Y) = 6(-7)\,\text{cov}(X, Y) = -42 \times 0.084 = -3.528$$

$$\text{corr}(W, V) = \frac{\sigma_{WV}}{\sigma_W \sigma_V} = \frac{-3.528}{\sqrt{7.56 \times 8.4084}} = -0.4425.$$

2.3. The table shows that $\Pr(X = 0, Y = 0) = 0.045$, $\Pr(X = 0, Y = 1) = 0.709$, $\Pr(X = 1, Y = 0) = 0.005$, $\Pr(X = 1, Y = 1) = 0.241$, $\Pr(X = 0) = 0.754$, $\Pr(X = 1) = 0.246$, $\Pr(Y = 0) = 0.050$, $\Pr(Y = 1) = 0.950$.

(a)

$$\begin{aligned} E(Y) &= \mu_Y = 0 \times \Pr(Y = 0) + 1 \times \Pr(Y = 1) \\ &= 0 \times 0.050 + 1 \times 0.950 = 0.950. \end{aligned}$$

(b)

$$\begin{aligned} \text{Unemployment Rate} &= \frac{\#\,(\text{unemployed})}{\#\,(\text{labor force})} \\ &= \Pr(Y = 0) = 0.050 = 1 - 0.950 = 1 - E(Y). \end{aligned}$$

(c) We calculate the conditional probabilities first:

$$\Pr(Y=0|X=0)=\frac{\Pr(X=0,\,Y=0)}{\Pr(X=0)}=\frac{0.045}{0.754}=0.0597,$$

$$\Pr(Y=1|X=0)=\frac{\Pr(X=0,\,Y=1)}{\Pr(X=0)}=\frac{0.709}{0.754}=0.9403,$$

$$\Pr(Y=0|X=1)=\frac{\Pr(X=1,\,Y=0)}{\Pr(X=1)}=\frac{0.005}{0.246}=0.0203,$$

$$\Pr(Y=1|X=1)=\frac{\Pr(X=1,\,Y=1)}{\Pr(X=1)}=\frac{0.241}{0.246}=0.9797.$$

The conditional expectations are

$$\begin{aligned} E(Y|X=1) &= 0\times\Pr(Y=0|X=1)+1\times\Pr(Y=1|X=1) \\ &= 0\times 0.0203+1\times 0.9797=0.9797, \end{aligned}$$

$$\begin{aligned} E(Y|X=0) &= 0\times\Pr(Y=0|X=0)+1\times\Pr(Y=1|X=0) \\ &= 0\times 0.0597+1\times 0.9403=0.9403. \end{aligned}$$

(d) Use the solution to part (b),

$$\begin{aligned} &\text{Unemployment rate for college grads} \\ =\ &1-E(Y|X=1)=1-0.9797=0.0203. \end{aligned}$$

$$\begin{aligned} &\text{Unemployment rate for non-college grads} \\ =\ &1-E(Y|X=0)=1-0.9403=0.0597. \end{aligned}$$

(e) The probability that a randomly selected worker who is reported being unemployed is a college graduate is

$$\Pr(X=1|Y=0)=\frac{\Pr(X=1,\,Y=0)}{\Pr(Y=0)}=\frac{0.005}{0.050}=0.1.$$

The probability that this worker is a non-college graduate is

$$\Pr(X=0|Y=0)=1-\Pr(X=1|Y=0)=1-0.1=0.9.$$

(f) Educational achievement and employment status are not independent because they do not satisfy that, for all values of x and y,

$$\Pr(Y=y|X=x)=\Pr(Y=y).$$

For example,

$$\Pr(Y = 0|X = 0) = 0.0597 \neq \Pr(Y = 0) = 0.050.$$

2.4. $\mu_Y = E(Y) = 1$, σ_Y^2 =var$(Y) = 4$. With $Z = \frac{1}{2}(Y - 1)$,

$$\mu_Z = E\left(\frac{1}{2}(Y-1)\right) = \frac{1}{2}(\mu_Y - 1) = \frac{1}{2}(1-1) = 0,$$

$$\sigma_Z^2 = \text{var}\left(\frac{1}{2}(Y-1)\right) = \frac{1}{4}\sigma_Y^2 = \frac{1}{4} \times 4 = 1.$$

2.5. Using the fact that if $Y \sim N(\mu_Y, \sigma_Y^2)$ then $\frac{Y-\mu_Y}{\sigma_Y} \sim N(0, 1)$ and Appendix Table 1, we have

(a)

$$\Pr(Y \leq 3) = \Pr\left(\frac{Y-1}{2} \leq \frac{3-1}{2}\right) = \Phi(1) = 0.8413.$$

(b)

$$\begin{aligned}\Pr(Y > 0) &= 1 - \Pr(Y \leq 0)\\ &= 1 - \Pr\left(\frac{Y-3}{3} \leq \frac{0-3}{3}\right) = 1 - \Phi(-1) = \Phi(1) = 0.8413.\end{aligned}$$

(c)

$$\begin{aligned}\Pr(40 \leq Y \leq 52) &= \Pr\left(\frac{40-50}{5} \leq \frac{Y-50}{5} \leq \frac{52-50}{5}\right)\\ &= \Phi(0.4) - \Phi(-2) = \Phi(0.4) - [1 - \Phi(2)]\\ &= 0.6554 - 1 + 0.9772 = 0.6326.\end{aligned}$$

(d)

$$\begin{aligned}\Pr(6 \leq Y \leq 8) &= \Pr\left(\frac{6-5}{\sqrt{2}} \leq \frac{Y-5}{\sqrt{2}} \leq \frac{8-5}{\sqrt{2}}\right)\\ &= \Phi(2.1213) - \Phi(0.7071)\\ &= 0.9831 - 0.7602 = 0.2229.\end{aligned}$$

2.6. (a) When Y is distributed χ_1^2, $\Pr(Y \leq 6.63) = 0.99$.

(b) When Y is distributed χ_4^2, $\Pr(Y \leq 7.78) = 0.90$.

(c) When Y is distributed $F_{10,\infty}$, $\Pr(Y > 2.32) = 1 - \Pr(Y \leq 2.32) = 1 - 0.99 = 0.01$.

2.7. The central limit theorem suggests that when the sample size (n) is large, the distribution of the sample average ($\bar{Y}$) is approximately $N\left(\mu_Y, \sigma_{\bar{Y}}^2\right)$ with $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$. Given a population $\mu_Y = 100$, $\sigma_Y^2 = 43.0$, we have

(a) $n = 100$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{100} = 0.43$, and

$$\Pr\left(\bar{Y} \leq 101\right) = \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.43}} \leq \frac{101 - 100}{\sqrt{0.43}}\right) \approx \Phi\left(1.525\right) = 0.9364.$$

(b) $n = 165$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{165} = 0.2606$, and

$$\begin{aligned}\Pr\left(\bar{Y} > 98\right) &= 1 - \Pr\left(\bar{Y} \leq 98\right) = 1 - \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.2606}} \leq \frac{98 - 100}{\sqrt{0.2606}}\right) \\ &\approx 1 - \Phi\left(-3.9178\right) = \Phi\left(3.9178\right) = 1.000 \quad \text{(rounded to four decimal places).}\end{aligned}$$

(c) $n = 64$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{64} = \frac{43}{64} = 0.6719$, and

$$\begin{aligned}\Pr\left(101 \leq \bar{Y} \leq 103\right) &= \Pr\left(\frac{101 - 100}{\sqrt{0.6719}} \leq \frac{\bar{Y} - 100}{\sqrt{0.6719}} \leq \frac{103 - 100}{\sqrt{0.6719}}\right) \\ &\approx \Phi\left(3.6599\right) - \Phi\left(1.2200\right) = 0.9999 - 0.8888 = 0.1111.\end{aligned}$$

2.8. $\Pr(Y = \$0) = 0.95$, $\Pr(Y = \$20000) = 0.05$.

(a) The mean of Y is

$$\mu_Y = 0 \times \Pr\left(Y = \$0\right) + 20,000 \times \Pr\left(Y = \$20000\right) = \$1000.$$

The variance of Y is

$$\begin{aligned}\sigma_Y^2 &= E\left[\left(Y - \mu_Y\right)^2\right] \\ &= \left(0 - 1000\right)^2 \times \Pr\left(Y = 0\right) + \left(20000 - 1000\right)^2 \times \Pr\left(Y = 20000\right) \\ &= (-1000)^2 \times 0.95 + 19000^2 \times 0.05 = 1.9 \times 10^7,\end{aligned}$$

so the standard deviation of Y is $\sigma_Y = \left(1.9 \times 10^7\right)^{\frac{1}{2}} = \$4359$.

(b) (i) $E\left(\bar{Y}\right) = \mu_Y = \$1000$, $\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n} = \frac{1.9\times10^7}{100} = 1.9 \times 10^5$.
(ii) Using the central limit theorem,

$$\begin{aligned}
\Pr\left(\bar{Y} > 2000\right) &= 1 - \Pr\left(\bar{Y} \le 2000\right) \\
&= 1 - \Pr\left(\frac{\bar{Y} - 1000}{\sqrt{1.9 \times 10^5}} \le \frac{2,000 - 1,000}{\sqrt{1.9 \times 10^5}}\right) \\
&\approx 1 - \Phi\left(2.2942\right) = 1 - 0.9891 = 0.0109.
\end{aligned}$$

2.9. (a)

$$\begin{aligned}
\Pr\left(Y = y_j\right) &= \sum_{i=1}^{l} \Pr\left(X = x_i,\ Y = y_j\right) \\
&= \sum_{i=1}^{l} \Pr\left(Y = y_j | X = x_i\right) \Pr\left(X = x_i\right)
\end{aligned}$$

(b)

$$\begin{aligned}
E\left(Y\right) &= \sum_{j=1}^{k} y_j \Pr\left(Y = y_j\right) = \sum_{j=1}^{k} y_j \sum_{i=1}^{l} \Pr\left(Y = y_j | X = x_i\right) \Pr\left(X = x_i\right) \\
&= \sum_{i=1}^{l} \left(\sum_{j=1}^{k} y_j \Pr\left(Y = y_j | X = x_i\right)\right) \Pr\left(X = x_i\right) \\
&= \sum_{i=1}^{l} E\left(Y | X = x_i\right) \Pr\left(X = x_i\right).
\end{aligned}$$

(c) When X and Y are independent,

$$\Pr\left(X = x_i, Y = y_j\right) = \Pr\left(X = x_i\right) \Pr\left(Y = y_j\right),$$

so

$$
\begin{aligned}
\sigma_{XY} &= E\left[(X-\mu_X)(Y-\mu_Y)\right] \\
&= \sum_{i=1}^{l}\sum_{j=1}^{k}(x_i-\mu_X)(y_j-\mu_Y)\Pr(X=x_i,Y=y_j) \\
&= \sum_{i=1}^{l}\sum_{j=1}^{k}(x_i-\mu_X)(y_j-\mu_Y)\Pr(X=x_i)\,Pr(Y=y_j) \\
&= \left(\sum_{i=1}^{l}(x_i-\mu_X)\Pr(X=x_i)\right)\left(\sum_{j=1}^{k}(y_j-\mu_Y)\Pr(Y=y_j)\right) \\
&= E(X-\mu_X)E(Y-\mu_Y) = 0\times 0 = 0,
\end{aligned}
$$

$$
\operatorname{corr}(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{0}{\sigma_X\sigma_Y} = 0.
$$

2.10. X and Z are two independently distributed standard normal random variables, so $\mu_X = \mu_Z = 0$, $\sigma_X^2 = \sigma_Z^2 = 1$, $\sigma_{XZ} = 0$.

(a) Because of the independence between X and Z, $\Pr(Z = z|X = x)$=$\Pr(Z = z)$, and $E(Z|X) = E(Z) = 0$. Thus $E(Y|X) = E\left(X^2 + Z|X\right) = E\left(X^2|X\right) + E(Z|X) = X^2 + 0 = X^2$.

(b) $E\left(X^2\right) = \sigma_X^2 + \mu_X^2 = 1$, and $\mu_Y = E\left(X^2 + Z\right) = E\left(X^2\right) + \mu_Z = 1 + 0 = 1$.

(c) $E(XY) = E\left(X^3 + ZX\right) = E\left(X^3\right) + E(ZX)$. Using the fact that the odd moments of a standard normal random variable are all zero, we have $E\left(X^3\right) = 0$. Using the independence between X and Z, we have $E(ZX) = \mu_Z\mu_X = 0$. Thus $E(XY) = E\left(X^3\right) + E(ZX) = 0$.

(d)

$$
\begin{aligned}
\operatorname{cov}(XY) &= E\left[(X-\mu_X)(Y-\mu_Y)\right] = E\left[(X-0)(Y-1)\right] \\
&= E(XY - X) = E(XY) - E(X) \\
&= 0 - 0 = 0.
\end{aligned}
$$

$\operatorname{corr}(X,Y) = \frac{\sigma_{XY}}{\sigma_X\sigma_Y} = \frac{0}{\sigma_X\sigma_Y} = 0.$

Chapter 3
Review of Statistics

3.1. The central limit theorem suggests that when the sample size (n) is large, the distribution of the sample average ($\bar{Y}$) is approximately $N\left(\mu_Y, \sigma_{\bar{Y}}^2\right)$ with $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$. Given a population $\mu_Y = 100$, $\sigma_Y^2 = 43.0$, we have

(a) $n = 100$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{100} = 0.43$, and

$$\Pr\left(\bar{Y} < 101\right) = \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.43}} < \frac{101 - 100}{\sqrt{0.43}}\right) \approx \Phi\left(1.525\right) = 0.9364.$$

(b) $n = 64$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{64} = \frac{43}{64} = 0.6719$, and

$$\begin{aligned} \Pr\left(101 < \bar{Y} < 103\right) &= \Pr\left(\frac{101 - 100}{\sqrt{0.6719}} < \frac{\bar{Y} - 100}{\sqrt{0.6719}} < \frac{103 - 100}{\sqrt{0.6719}}\right) \\ &\approx \Phi\left(3.6599\right) - \Phi\left(1.2200\right) = 0.9999 - 0.8888 = 0.1111. \end{aligned}$$

(c) $n = 165$, $\sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n} = \frac{43}{165} = 0.2606$, and

$$\begin{aligned} \Pr\left(\bar{Y} > 98\right) &= 1 - \Pr\left(\bar{Y} \leq 98\right) = 1 - \Pr\left(\frac{\bar{Y} - 100}{\sqrt{0.2606}} \leq \frac{98 - 100}{\sqrt{0.2606}}\right) \\ &\approx 1 - \Phi\left(-3.9178\right) = \Phi\left(3.9178\right) = 1.0000 \;\; \text{(rounded to four decimal places).} \end{aligned}$$

3.2. Each random draw Y_i from the Bernoulli distribution takes a value of either zero or one with probability $\Pr(Y_i = 1) = p$ and $\Pr(Y_i = 0) = 1-p$. The random variable Y_i has mean

$$E\left(Y_i\right) = 0 \times \Pr\left(Y = 0\right) + 1 \times \Pr\left(Y = 1\right) = p,$$

and variance

$$\begin{aligned} \operatorname{var}\left(Y_i\right) &= E\left[\left(Y_i - \mu_Y\right)^2\right] \\ &= \left(0 - p\right)^2 \times \Pr\left(Y_i = 0\right) + \left(1 - p\right)^2 \times \Pr\left(Y_i = 1\right) \\ &= p^2\left(1 - p\right) + \left(1 - p\right)^2 p = p\left(1 - p\right). \end{aligned}$$

(a) The fraction of successes is

$$\hat{p} = \frac{\#\left(\text{success}\right)}{n} = \frac{\#\left(Y_i = 1\right)}{n} = \frac{\sum_{i=1}^{n} Y_i}{n} = \bar{Y}.$$

(b)

$$E(\hat{p}) = E\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(Y_i) = \frac{1}{n}\sum_{i=1}^{n} p = p.$$

(c)

$$\operatorname{var}(\hat{p}) = \operatorname{var}\left(\frac{\sum_{i=1}^{n} Y_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{var}(Y_i) = \frac{1}{n^2}\sum_{i=1}^{n} p(1-p) = \frac{p(1-p)}{n}.$$

The second equality uses the fact that $Y_1, ..., Y_n$ are i.i.d. draws and $\operatorname{cov}(Y_i, Y_j) = 0,$ for $i \neq j$.

3.3. Denote each voter's preference by Y. $Y = 1$ if the voter prefers the incumbent and $Y = 0$ if the voter prefers the challenger. Y is a Bernoulli random variable with probability $\Pr(Y = 1) = p$ and $\Pr(Y = 0) = 1 - p$. As we have seen from the solution to Exercise 3.2, Y has mean p and variance $p(1-p)$.

(a) From the solution to Exercise 3.2, we know an unbiased estimator of p is $\hat{p} = \frac{215}{400} = 0.5375$.

(b) $\operatorname{var}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n} = \frac{0.5375\times(1-0.5375)}{400} = 6.2148 \times 10^{-4}$. The standard error is $\operatorname{SE}(\hat{p}) = (\operatorname{var}(\hat{p}))^{\frac{1}{2}} = 0.0249$.

(c) The computed t-statistic is

$$t^{act} = \frac{\hat{p} - \mu_{p,0}}{\operatorname{SE}(\hat{p})} = \frac{0.5375 - 0.5}{0.0249} = 1.506.$$

Because of the large sample size ($n = 400$), we can use Equation (3.13) in the text to get the p-value for the test $H_0 : p = 0.5$ vs. $H_1 : p \neq 0.5$:

$$p\text{-value} = 2\Phi\left(-\left|t^{act}\right|\right) = 2\Phi(-1.506) = 2 \times 0.066 = 0.132.$$

(d) Using Equation (3.17) in the text, the p-value for the test $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$ is

$$p\text{-value} = 1 - \Phi\left(t^{act}\right) = 1 - \Phi(1.506) = 1 - 0.934 = 0.066.$$

(e) Part (c) is a two-sided test and the p-value is the area in the tails of the standard normal distribution outside $\pm$(calculated t-statistic). Part (d) is a one-sided test and the p-value is the area under the standard normal distribution to the right of the calculated t-statistic.

(f) For the test $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$, we cannot reject the null hypothesis at the 5% significance level. The p-value 0.066 is larger than 0.05. Equivalently the calculated t-statistic 1.506 is less than the critical value 1.645 for a one-sided test with a 5% significance level. The test suggests that the survey did not contain statistically significant evidence that the incumbent was ahead of the challenger at the time of the survey.

3.4. Using Key Concept 3.7 in the text

(a) 95% confidence interval for p is

$$\hat{p} \pm 1.96\text{SE}(\hat{p}) = 0.5375 \pm 1.96 \times 0.0249 = (0.4887,\ 0.5863).$$

(b) 99% confidence interval for p is

$$\hat{p} \pm 2.57\text{SE}(\hat{p}) = 0.5375 \pm 2.57 \times 0.0249 = (0.4735,\ 0.6015).$$

(c) The interval in (b) is wider because of a larger critical value due to a lower significance level.

(d) Since 0.50 lies inside the 95% confidence interval for p, we cannot reject the null hypothesis at a 5% significance level.

3.5. Denote the life of a light bulb from the new process by Y. The mean of Y is μ and the standard deviation of Y is $\sigma_Y = 200$ hours. $\bar{Y}$ is the sample mean with a sample size $n = 100$. The standard deviation of the sampling distribution of $\bar{Y}$ is $\sigma_{\bar{Y}} = \frac{\sigma_Y}{\sqrt{n}} = \frac{200}{\sqrt{100}} = 20$ hours. The hypothesis test is $H_0 : \mu = 2000$ vs. $H_1 : \mu > 2000$. The manager will accept the alternative hypothesis if $\bar{Y} > 2100$ hours.

(a) The size of a test is the probability of erroneously rejecting a null hypothesis when it is valid. The size of the manager's test is

$$\begin{aligned}
\text{size} &= \Pr\left(\bar{Y} > 2100|\mu = 2000\right) = 1 - \Pr\left(\bar{Y} \leq 2100|\mu = 2000\right) \\
&= 1 - \Pr\left(\frac{\bar{Y} - 2000}{20} \leq \frac{2100 - 2000}{20}|\mu = 2000\right) \\
&= 1 - \Phi(5) = 1 - 0.999999713 = 2.87 \times 10^{-7}.
\end{aligned}$$

$\Pr\left(\bar{Y} > 2100|\mu = 2000\right)$ means the probability that the sample mean is greater than 2100 hours when the new process has a mean of 2000 hours.

(b) The power of a test is the probability of correctly rejecting a null hypothesis when it is invalid. We calculate first the probability of the manager erroneously accepting the null hypothesis when it is invalid:

$$\begin{aligned}\beta &= \Pr\left(\bar{Y} \le 2100|\mu = 2150\right) = \Pr\left(\frac{\bar{Y} - 2150}{20} \le \frac{2100 - 2150}{20}|\mu = 2150\right) \\ &= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062.\end{aligned}$$

The power of the manager's testing is $1 - \beta = 1 - 0.0062 = 0.9938$.

(c) For a test with 5%, the rejection region for the null hypothesis contains those values of the t-statistic exceeding 1.645.

$$t^{act} = \frac{\bar{Y}^{act} - 2000}{20} > 1.645 \Rightarrow \bar{Y}^{act} > 2000 + 1.645 \times 20 = 2032.9.$$

The manager should believe the inventor's claim if the sample mean life of the new product is greater than 2032.9 hours if she wants the size of the test to be 5%.

3.6. (a) New Jersey sample size $n_1 = 100$, sample average $\bar{Y}_1 = 58$, sample standard deviation $s_1 = 8$. The standard error of $\bar{Y}_1$ is $\mathrm{SE}(\bar{Y}_1) = \frac{s_1}{\sqrt{n_1}} = \frac{8}{\sqrt{100}} = 0.8$. The 95% confidence interval for the mean score of all New Jersey third graders is

$$\mu_1 = \bar{Y}_1 \pm 1.96\mathrm{SE}\left(\bar{Y}_1\right) = 58 \pm 1.96 \times 0.8 = (56.432,\ 59.568).$$

(b) Iowa sample size $n_2 = 200$, sample average $\bar{Y}_2 = 62$, sample standard deviation $s_2 = 11$. The standard error of $\bar{Y}_1 - \bar{Y}_2$ is $\mathrm{SE}(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{64}{100} + \frac{121}{200}} = 1.1158$. The 90% confidence interval for the difference in mean score between the two states is

$$\begin{aligned}\mu_1 - \mu_2 &= \left(\bar{Y}_1 - \bar{Y}_2\right) \pm 1.64\mathrm{SE}\left(\bar{Y}_1 - \bar{Y}_2\right) \\ &= (58 - 62) \pm 1.64 \times 1.1158 = (-5.8299,\ -2.1701).\end{aligned}$$

(c) The hypothesis tests for the difference in mean scores is

$$H_0: \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1: \mu_1 - \mu_2 \neq 0.$$

From part (b) the standard error of the difference in the two sample means is $\text{SE}(\bar{Y}_1 - \bar{Y}_2) = 1.1158$. The t-statistic for testing the null hypothesis is

$$t^{act} = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{SE}\left(\bar{Y}_1 - \bar{Y}_2\right)} = \frac{58 - 62}{1.1158} = -3.5849.$$

Use Equation (3.13) in the text to compute the p-value:

$$p\text{-value} = 2\Phi\left(-\left|t^{act}\right|\right) = 2\Phi\left(-3.5849\right) = 2 \times 0.00017 = 0.00034.$$

Because of the extremely low p-value, we can reject the null hypothesis with a very high degree of confidence. That is, the population means for Iowa and New Jersey students are different.

3.7. Assume that n is an even number. Then $\tilde{Y}$ is constructed by applying a weight of $\frac{1}{2}$ to the $\frac{n}{2}$ "odd" observations and a weight of $\frac{3}{2}$ to the remaining $\frac{n}{2}$ observations.

$$\begin{aligned} E\left(\tilde{Y}\right) &= \frac{1}{n}\left(\frac{1}{2}E\left(Y_1\right) + \frac{3}{2}E\left(Y_2\right) + \cdots \frac{1}{2}E\left(Y_{n-1}\right) + \frac{3}{2}E\left(Y_n\right)\right) \\ &= \frac{1}{n}\left(\frac{1}{2} \cdot \frac{n}{2} \cdot \mu_Y + \frac{3}{2} \cdot \frac{n}{2} \cdot \mu_Y\right) = \mu_Y \end{aligned}$$

$$\begin{aligned} \text{var}\left(\tilde{Y}\right) &= \frac{1}{n^2}\left(\frac{1}{4}\text{var}\left(Y_1\right) + \frac{9}{4}\text{var}\left(Y_2\right) + \cdots \frac{1}{4}\text{var}\left(Y_{n-1}\right) + \frac{9}{4}\text{var}\left(Y_n\right)\right) \\ &= \frac{1}{n^2}\left(\frac{1}{4} \cdot \frac{n}{2} \cdot \sigma_Y^2 + \frac{9}{4} \cdot \frac{n}{2} \cdot \sigma_Y^2\right) = 1.25\frac{\sigma_Y^2}{n}. \end{aligned}$$

3.8. Sample size for men $n_1 = 100$, sample average $\bar{Y}_1 = 3100$, sample standard deviation $s_1 = 200$. Sample size for women $n_2 = 64$, sample average $\bar{Y}_2 = 2900$, sample standard deviation $s_2 = 320$. The standard error of $\bar{Y}_1 - \bar{Y}_2$ is $\text{SE}(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{200^2}{100} + \frac{320^2}{64}} = 44.721$.

(a) The hypothesis test for the difference in mean monthly salaries is

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 \neq 0.$$

The t-statistic for testing the null hypothesis is

$$t^{act} = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{SE}\left(\bar{Y}_1 - \bar{Y}_2\right)} = \frac{3100 - 2900}{44.721} = 4.4722.$$

Use Equation (3.13) in the text to get the p-value:

$$p\text{-value} = 2\Phi\left(-\left|t^{act}\right|\right) = 2\Phi\left(-4.4722\right) = 2 \times \left(3.8744 \times 10^{-6}\right) = 7.7488 \times 10^{-6}.$$

The extremely low level of p-value implies that the difference in the monthly salaries for men and women is statistically significant. We can reject the null hypothesis with a high degree of confidence.

(b) From part (a), there is overwhelming statistical evidence that mean earnings for men differ from mean earnings for women. To examine whether there is gender discrimination in the compensation policies, we take the following one-sided alternative test

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 > 0.$$

With the t-statistic $t^{act} = 4.4722$, the p-value for the one-sided test is:

$$p\text{-value} = 1 - \Phi\left(t^{act}\right) = 1 - \Phi\left(4.4722\right) = 1 - 0.999996126 = 3.874 \times 10^{-6}.$$

With the extremely small p-value, the null hypothesis can be rejected with a high degree of confidence. There is overwhelming statistical evidence that mean earnings for men are greater than mean earnings for women.

However, by itself, this does not imply gender discrimination by the firm. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. The data description suggests that some care has been taken to make sure that workers with similar jobs are being compared. But, it is also important to control for characteristics of the workers that may affect their productivity (education, years of experience, etc.). If these characteristics are systematically different between men and women, then they may be responsible for the difference in mean wages. (If this is true, it raises an interesting and important question of why women tend to have less education or less experience than men, but that is a question about something other than gender discrimination by this firm.) Since these characteristics are not controlled for in the statistical analysis, it is premature to reach a conclusion about gender discrimination.

3.9. (a) Sample size $n = 420$, sample average $\bar{Y} = 654.2$, sample standard deviation $s_Y = 19.5$. The standard error of $\bar{Y}$ is $\text{SE}(\bar{Y}) = \frac{s_Y}{\sqrt{n}} = \frac{19.5}{\sqrt{420}} = 0.9515$. The 95% confidence interval for the mean test score in the population is

$$\mu = \bar{Y} \pm 1.96\text{SE}\left(\bar{Y}\right) = 654.2 \pm 1.96 \times 0.9515 = (652.34,\, 656.06)\,.$$

(b) The data are: sample size for small classes $n_1 = 238$, sample average $\bar{Y}_1 = 657.4$, sample standard deviation $s_1 = 19.4$; sample size for large classes $n_2 =$

182, sample average $\bar{Y}_2 = 650.0$, sample standard deviation $s_2 = 17.9$. The standard error of $\bar{Y}_1 - \bar{Y}_2$ is $\text{SE}(\bar{Y}_1 - \bar{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}} =$ 1.8281. The hypothesis tests for higher average scores in smaller classes is

$$H_0 : \mu_1 - \mu_2 = 0 \quad \text{vs.} \quad H_1 : \mu_1 - \mu_2 > 0.$$

The t-statistic is

$$t^{act} = \frac{\bar{Y}_1 - \bar{Y}_2}{\text{SE}\left(\bar{Y}_1 - \bar{Y}_2\right)} = \frac{657.4 - 650.0}{1.8281} = 4.0479.$$

The p-value for the one-sided test is:

$$p\text{-value} = 1 - \Phi\left(t^{act}\right) = 1 - \Phi\left(4.0479\right) = 1 - 0.999974147 = 2.5853 \times 10^{-5}.$$

With the small p-value, the null hypothesis can be rejected with a high degree of confidence. There is statistically significant evidence that the districts with smaller classes have higher average test scores.

3.10. We have the following relations: $1\text{in} = 0.0254\,\text{m}$ (or $1\text{m} = 39.37\,\text{in}$), $1\text{lb} = 0.4536\,\text{kg}$ (or $1\text{kg} = 2.2046\,\text{lb}$). The summary statistics in the metric system are $\bar{X} = 70.5 \times 0.0254 = 1.79\,\text{m}$; $\bar{Y} = 158 \times 0.4536 = 71.669\,\text{kg}$; $s_X = 1.8 \times 0.0254 = 0.0457\,\text{m}$; $s_Y = 14.2 \times 0.4536 = 6.4411\,\text{kg}$; $s_{XY} = 21.73 \times 0.0254 \times 0.4536 = 0.2504\,\text{m} \times \text{kg}$, and $r_{XY} = 0.85$.

3.11. $Y_1, \ldots, Y_n$ are i.i.d. with mean μ_Y and variance σ_Y^2. The covariance $\text{cov}(Y_j, Y_i) = 0$, $j \neq i$. The sampling distribution of the sample average $\bar{Y}$ has mean μ_Y and variance $\text{var}\left(\bar{Y}\right) = \sigma_{\bar{Y}}^2 = \frac{\sigma_Y^2}{n}$.

(a)

$$\begin{aligned}
E\left[\left(Y_i - \bar{Y}\right)^2\right] &= E\left\{\left[\left(Y_i - \mu_Y\right) - \left(\bar{Y} - \mu_Y\right)\right]^2\right\} \\
&= E\left[\left(Y_i - \mu_Y\right)^2 - 2\left(Y_i - \mu_Y\right)\left(\bar{Y} - \mu_Y\right) + \left(\bar{Y} - \mu_Y\right)^2\right] \\
&= E\left[\left(Y_i - \mu_Y\right)^2\right] - 2E\left[\left(Y_i - \mu_Y\right)\left(\bar{Y} - \mu_Y\right)\right] + E\left[\left(\bar{Y} - \mu_Y\right)^2\right] \\
&= \text{var}\left(Y_i\right) - 2\text{cov}\left(Y_i, \bar{Y}\right) + \text{var}\left(\bar{Y}\right).
\end{aligned}$$

(b)

$$
\begin{aligned}
\operatorname{cov}\left(\bar{Y}, Y_i\right) &= E\left[\left(\bar{Y}-\mu_Y\right)\left(Y_i-\mu_Y\right)\right] \\
&= E\left[\left(\frac{\sum_{j=1}^{n} Y_j}{n}-\mu_Y\right)\left(Y_i-\mu_Y\right)\right] \\
&= E\left[\left(\frac{\sum_{j=1}^{n}\left(Y_j-\mu_Y\right)}{n}\right)\left(Y_i-\mu_Y\right)\right] \\
&= \frac{1}{n} E\left[\left(Y_i-\mu_Y\right)^2\right]+\frac{1}{n} \sum_{j \neq i} E\left[\left(Y_j-\mu_Y\right)\left(Y_i-\mu_Y\right)\right] \\
&= \frac{1}{n} \sigma_Y^2+\frac{1}{n} \sum_{j \neq i} \operatorname{cov}\left(Y_j, Y_i\right) \\
&= \frac{\sigma_Y^2}{n}.
\end{aligned}
$$

(c)

$$
\begin{aligned}
E\left(s_Y^2\right) &= E\left(\frac{1}{n-1} \sum_{i=1}^{n}\left(Y_i-\bar{Y}\right)^2\right) \\
&= \frac{1}{n-1} \sum_{i=1}^{n} E\left[\left(Y_i-\bar{Y}\right)^2\right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n}\left[\operatorname{var}\left(Y_i\right)-2 \operatorname{cov}\left(Y_i, \bar{Y}\right)+\operatorname{var}\left(\bar{Y}\right)\right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n}\left[\sigma_Y^2-2 \times \frac{\sigma_Y^2}{n}+\frac{\sigma_Y^2}{n}\right] \\
&= \frac{1}{n-1} \sum_{i=1}^{n}\left(\frac{n-1}{n} \sigma_Y^2\right) \\
&= \sigma_Y^2.
\end{aligned}
$$

Chapter 4
Linear Regression with One Regressor

4.1. (a) The predicted average test score is

$$\widehat{TestScore} = 520.4 - 5.82 \times 22 = 392.36.$$

(b) The predicted change in the classroom average test score is

$$\Delta\widehat{TestScore} = (-5.82 \times 19) - (-5.82 \times 23) = 23.28.$$

(c) The 95% confidence interval for β_1 is $\{-5.82 \pm 1.96 \times 2.21\}$, that is, $-10.152 \leq \beta_1 \leq -1.4884$.

(d) Calculate the t-statistic first

$$t^{act} = \frac{\hat{\beta}_1 - 0}{\mathrm{SE}\left(\hat{\beta}_1\right)} = \frac{-5.82}{2.21} = -2.6335.$$

The p-value for the test $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$ is

$$p - \text{value} = 2\Phi\left(-\left|t^{act}\right|\right) = 2\Phi\left(-2.6335\right) = 2 \times 0.0042 = 0.0084.$$

The p-value is less than 0.01, so we can reject the null hypothesis at the 5% significance level, and also at the 1% significance level.

(e) Using the formula for $\hat{\beta}_0$ in Equation (4.9), we know the sample average of the test scores across the 100 classrooms is

$$\widehat{TestScore} = \hat{\beta}_0 + \hat{\beta}_1 \times CS = 520.4 - 5.82 \times 21.4 = 395.85.$$

(f) Use the formula for the standard error of the regression (SER) in Equation (4.40) to get the sum of squared residuals:

$$SSR = (n - 2)\,SER^2 = (100 - 2) \times 11.5^2 = 12961.$$

Use the formula for R^2 in Equation (4.39) to get the total sum of squares:

$$TSS = \frac{SSR}{1 - R^2} = \frac{12961}{1 - 0.08^2} = 13044.$$

The sample variance is $s_Y^2 = \frac{\text{TSS}}{n-1} = \frac{13044}{99} = 131.8$. Thus, standard deviation is $s_Y = \sqrt{s_Y^2} = 11.5$.

4.2. (a) The estimated gender gap equals $\hat{\beta}_1 = \$2.79/\text{hour}$.

(b) The hypothesis testing for the gender gap is $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$. With a t-statistic

$$t^{act} = \frac{\hat{\beta}_1 - 0}{SE\left(\hat{\beta}_1\right)} = \frac{2.79}{0.84} = 3.3214,$$

the p-value for the test is

$$p\text{-value} = 2\Phi\left(-\left|t^{act}\right|\right) = 2\Phi\left(-3.3214\right) = 2 \times 0.0004 = 0.0008.$$

The p-value is less than 0.01, so we can reject the null hypothesis that there is no gender gap at a 1% significance level.

(c) The 95% confidence interval for the gender gap β_1 is $\{2.79 \pm 1.96 \times 0.84\}$, that is, $1.1436 \leq \beta_1 \leq 4.4364$.

(d) The sample average wage of women is $\hat{\beta}_0 = \$12.68/\text{hour}$. The sample average wage of men is $\hat{\beta}_0 + \hat{\beta}_1 = \$15.47/\text{hour}$.

(e) The binary variable regression model relating wages to gender can be written as either

$$\text{Wage} = \beta_0 + \beta_1 \text{Male} + u_i,$$

or

$$\text{Wage} = \gamma_0 + \gamma_1 \text{Female} + v_i.$$

In the first regression equation, $Male$ equals 1 for men and 0 for women; β_0 is the population mean of wages for women and $\beta_0 + \beta_1$ is the population mean of wages for men. In the second regression equation, $Female$ equals 1 for women and 0 for men; γ_0 is the population mean of wages for men and $\gamma_0 + \gamma_1$ is the population mean of wages for women. We have the following relationship for the coefficients in the two regression equations:

$$\begin{aligned} \gamma_0 &= \beta_0 + \beta_1, \\ \gamma_0 + \gamma_1 &= \beta_0. \end{aligned}$$

Given the coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$, we have

$$\begin{aligned}\hat{\gamma}_0 &= \hat{\beta}_0 + \hat{\beta}_1 = 15.47, \\ \hat{\gamma}_1 &= \hat{\beta}_0 - \hat{\gamma}_0 = -\hat{\beta}_1 = -2.79.\end{aligned}$$

Due to the relationship among coefficient estimates, for each individual observation, the OLS residual is the same under the two regression equations: $\hat{u}_i = \hat{v}_i$. Thus the sum of squared residuals, $SSR = \sum_{i=1}^{n} \hat{u}_i^2$, is the same under the two regressions. This implies that both $SER = \left(\frac{SSR}{n-1}\right)^{\frac{1}{2}}$ and $R^2 = 1 - \frac{SSR}{TSS}$ are unchanged.

In summary, in regressing $Wages$ on $Female$, we will get

$$\widehat{Wages} = 15.47 - 2.79 Female, \qquad R^2 = 0.06, \quad \text{SER} = 3.10.$$

4.3. Using $E(u_i|X_i) = 0$, we have

$$E(Y_i|X_i) = E(\beta_0 + \beta_1 X_i + u_i|X_i) = \beta_0 + \beta_1 E(X_i|X_i) + E(u_i|X_i) = \beta_0 + \beta_1 X_i.$$

4.4. The expectation of $\hat{\beta}_0$ is obtained by taking expectations of both sides of Equation (4.9):

$$\begin{aligned}E\left(\hat{\beta}_0\right) &= E\left(\bar{Y} - \widehat{\beta}_1 \bar{X}\right) = E\left[\left(\beta_0 + \beta_1 \bar{X} + \frac{1}{n}\sum_{i=1}^{n} u_i\right) - \hat{\beta}_1 \bar{X}\right] \\ &= \beta_0 + E\left(\beta_1 - \hat{\beta}_1\right)\bar{X} + \frac{1}{n}\sum_{i=1}^{n} E(u_i|X_i) = \beta_0,\end{aligned}$$

where the third equality in the above equation has used the facts that $\hat{\beta}_1$ is unbiased so $E\left(\beta_1 - \hat{\beta}_1\right) = 0$ and $E(u_i|X_i) = 0$.

4.5. The sample size $n = 200$. The estimated regression equation is

$$\widehat{Weight} = \underset{(2.15)}{-99.41} + \underset{(0.31)}{3.94}\, Height, \qquad R^2 = 0.81, \quad \text{SER} = 10.2.$$

(a) Substituting $Height = 70$, 65, and 74 inches into the equation, the predicted weights are 176.39, 156.69, and 192.15 pounds.

(b) $\Delta\widehat{Weight} = 3.94 \times \Delta Height = 3.94 \times 1.5 = 5.91$.

(c) The 99% confidence interval for the weight increase is

$$\begin{aligned} 5.91 \pm 2.58 \times \text{SE}(\beta_1) \times \Delta\text{Height} &= 5.91 \pm 2.58 \times 0.31 \times 1.5 \\ &= [4.7103,\ 7.1097]. \end{aligned}$$

(d) We have the following relations: 1in = 2.54 cm and 1lb = 0.4536 kg. Suppose the regression equation in the centimeter-kilogram space is

$$\widehat{\text{Weight}} = \hat{\gamma}_0 + \hat{\gamma}_1 \text{Height}.$$

The coefficients are $\hat{\gamma}_0 = -99.41 \times 0.4536 = -45.092$ kg with a standard error $\text{SE}(\hat{\gamma}_0) = 2.15 \times 0.4536 = 0.9752$ kg; $\hat{\gamma}_1 = 3.94 \times \frac{0.4536}{2.54} = 0.7036$ kg with a standard error $\text{SE}(\hat{\gamma}_1) = 0.31 \times \frac{0.4536}{2.54} = 0.0554$ kg / cm. R^2 is respective of the units of variables, so it remains at $R^2 = 0.81$. The standard error of the regression is $SER = 10.2 \times 0.4536 = 4.6267$.

4.6. Equation (4.15) gives

$$\sigma^2_{\hat{\beta}_0} = \frac{\text{var}(H_i u_i)}{n\left[E\left(H_i^2\right)\right]^2}, \qquad \text{where } H_i = 1 - \frac{\mu_x}{E\left(X_i^2\right)} X_i.$$

Using the facts that $E(u_i|X_i) = 0$ and $\text{var}(u_i|X_i) = \sigma_u^2$ (homoskedasticity), we have

$$\begin{aligned} E(H_i u_i) &= E\left(u_i - \frac{\mu_x}{E\left(X_i^2\right)} X_i u_i\right) = E(u_i) - \frac{\mu_x}{E\left(X_i^2\right)} E\left[X_i E(u_i|X_i)\right] \\ &= 0 - \frac{\mu_x}{E\left(X_i^2\right)} \times 0 = 0, \end{aligned}$$

and

$$\begin{aligned} E\left[(H_i u_i)^2\right] &= E\left\{\left(u_i - \frac{\mu_x}{E\left(X_i^2\right)} X_i u_i\right)^2\right\} \\ &= E\left\{u_i^2 - 2\frac{\mu_x}{E\left(X_i^2\right)} X_i u_i^2 + \left[\frac{\mu_x}{E\left(X_i^2\right)}\right]^2 X_i^2 u_i^2\right\} \\ &= E\left(u_i^2\right) - 2\frac{\mu_x}{E\left(X_i^2\right)} E\left[X_i E\left(u_i^2|X_i\right)\right] + \left[\frac{\mu_x}{E\left(X_i^2\right)}\right]^2 E\left[X_i^2 E\left(u_i^2|X_i\right)\right] \\ &= \sigma_u^2 - 2\frac{\mu_x}{E\left(X_i^2\right)} \mu_x \sigma_u^2 + \left[\frac{\mu_x}{E\left(X_i^2\right)}\right]^2 E\left(X_i^2\right) \sigma_u^2 \\ &= \left(1 - \frac{\mu_x^2}{E\left(X_i^2\right)}\right) \sigma_u^2. \end{aligned}$$

Because $E\left(H_i u_i\right) = 0$, $\operatorname{var}(H_i u_i) = E\left[(H_i u_i)^2\right]$,so

$$\operatorname{var}\left(H_i u_i\right) = E\left[(H_i u_i)^2\right] = \left(1 - \frac{\mu_x^2}{E\left(X_i^2\right)}\right)\sigma_u^2.$$

We can also get

$$\begin{aligned} E\left(H_i^2\right) &= E\left\{\left(1 - \frac{\mu_x}{E\left(X_i^2\right)}X_i\right)^2\right\} = E\left\{1 - 2\frac{\mu_x}{E\left(X_i^2\right)}X_i + \left[\frac{\mu_x}{E\left(X_i^2\right)}\right]^2 X_i^2\right\} \\ &= 1 - 2\frac{\mu_x^2}{E\left(X_i^2\right)} + \left[\frac{\mu_x}{E\left(X_i^2\right)}\right]^2 E\left(X_i^2\right) = 1 - \frac{\mu_x^2}{E\left(X_i^2\right)}. \end{aligned}$$

Thus

$$\begin{aligned} \sigma_{\hat{\beta}_0}^2 &= \frac{\operatorname{var}\left(H_i u_i\right)}{n\left[E\left(H_i^2\right)\right]^2} = \frac{\left(1 - \frac{\mu_x^2}{E\left(X_i^2\right)}\right)\sigma_u^2}{n\left(1 - \frac{\mu_x^2}{E\left(X_i^2\right)}\right)^2} = \frac{\sigma_u^2}{n\left(1 - \frac{\mu_x^2}{E\left(X_i^2\right)}\right)} \\ &= \frac{E\left(X_i^2\right)\sigma_u^2}{n\left[E\left(X_i^2\right) - \mu_x^2\right]} = \frac{E\left(X_i^2\right)\sigma_u^2}{n\sigma_X^2}. \end{aligned}$$

Chapter 5
Linear Regression with Multiple Regressors

5.1. The hypothesis testing for the significance of a coefficient (β) is $H_0 : \beta = 0$ vs. $H_1 : \beta \neq 0$. The coefficient estimate is significantly different from 0 if the computed t-statistic $t^{act} = \frac{\hat{\beta}}{\mathrm{SE}(\hat{\beta})}$ is larger than the critical value, which is 1.96 at a 5% significance level and 2.58 at a 1% significance level. See the table for the added "*" (5%) or "**" (1%) to indicate statistical significance of the coefficient.

Dependent Variable: Average Hourly Earning (AHE)			
Regressor	(1)	(2)	(3)
College (X_1)	5.46** (0.21)	5.48** (0.21)	5.44** (0.21)
Female (X_2)	−2.64** (0.20)	−2.62** (0.20)	−2.62** (0.20)
Age (X_3)		0.29** (0.04)	0.29** (0.04)
Northeast (X_4)			0.69* (0.30)
Midwest (X_5)			0.60* (0.28)
South (X_6)			−0.27 (0.26)
Intercept (X_7)	12.69** (0.14)	4.40** (1.05)	3.75** (1.06)
Summary Statistics and Joint Tests			
F-statistic for regional effects = 0			6.10
SER	6.27	6.22	6.21
R^2	0.176	0.190	0.194
$\bar{R}^2$	0.175	0.189	0.193
n	4000	4000	4000

5.2. By equations (5.29) and (5.30) in the text, we know

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}\left(1 - R^2\right).$$

See the table above for the computed $\bar{R}^2$ for each of the regressions.

5.3. (a) Workers with college degrees earn \$5.46/hour more, on average, than workers with only high school degrees. The significance test from Exercise (5.1) suggests that the earnings difference is statistically significant at the 5% level.

(b) Men earn \$2.64/hour more, on average, than women. The significance test from Exercise (5.1) suggests that the earnings difference is statistically significant at the 5% level.

5.4. (a) Age is an important determinant of earnings since the coefficient associated with age is found to be statistically significant at the 1% level. On average, a worker earns \$0.29/hour more for each year he ages.

(b) Sally's earnings prediction is $4.40+5.48\times1-2.62\times1+0.29\times29 = 15.67$ dollars per hour. Betsy's earnings prediction is $4.40+5.48\times1-2.62\times1+0.29\times34 = 17.12$ dollars per hour. The difference is 1.45, and the 95% confidence interval is $1.45 \pm 1.96 \times 0.04 \times 5 = [1.058,\ 1.842]$.

5.5. (a) The F-statistic for the null hypothesis that the coefficients on the regional effects are jointly equal to zero is 6.10. This is larger than the 1% critical value of 3.78, so that the regional effects are jointly significant. Inspection of the results for each region shows that, at the 5% level, earnings in the Northeast and Midwest are significantly different from earnings in the West; there is no significant difference between the South and the West.

(b) The regressor $West$ is omitted to avoid the perfect multicollinearity problem. If $West$ is included, then the intercept can be written as a perfect linear function of the four regional regressors. Because of perfect multicollinearity, the OLS estimator cannot be computed.

(c.i) The 95% confidence interval for the difference in expected earnings between Juanita and Molly is $-0.27 \pm 1.96 \times 0.26 = [-0.7796,\ 0.2396]$.

(c.ii) The expected difference in earnings between Juanita and Jennifer is $-0.27- 0.6 = -0.87$.

(c.iii) To construct a 95% confidence interval for the difference in expected earnings between Juanita and Jennifer, we could include $West$ and exclude $Midwest$ from the regression. The estimated coefficient associated with $South$ would then give the expected difference in earnings between Juanita and Jennifer. The estimated coefficient and its standard error could be used to compute the confidence interval as in part (c.i).

5.6. The t-statistic for the difference in the college coefficients is

$$t = (\widehat{\beta}_{college,1998} - \widehat{\beta}_{college,1992})/SE(\widehat{\beta}_{college,1998} - \widehat{\beta}_{college,1992})$$

Because $\widehat{\beta}_{college,1998}$ and $\widehat{\beta}_{college,1992}$ are computed from independent samples, they are independent, which means that $\text{cov}(\widehat{\beta}_{college,1998}, \widehat{\beta}_{college,1992}) = 0$ Thus, $\text{var}(\widehat{\beta}_{college,1998} - \widehat{\beta}_{college,1992}) = \text{var}(\widehat{\beta}_{college,1998}) + \text{var}(\widehat{\beta}_{college,1998})$. This implies that $SE(\widehat{\beta}_{college,1998} - \widehat{\beta}_{college,1992}) = \left(0.21^2 + 0.20^2\right)^{\frac{1}{2}}$. Thus,

$$t^{act} = \frac{5.48 - 5.29}{(0.21^2 + 0.20^2)^{\frac{1}{2}}} = 0.6552.$$

There is no significant change since the calculated t-statistic is less than 1.96, the 5% critical value.

5.7. In isolation, these results do imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. It is also important to control for characteristics of the workers that may affect their productivity (education, years of experience, etc.) If these characteristics are systematically different between men and women, then they may be responsible for the difference in mean wages. (If this were true, it would raise an interesting and important question of why women tend to have less education or less experience than men, but that is a question about something other than gender discrimination.) These are potentially important omitted variables in the regression that will lead to bias in the OLS coefficient estimator for $Female$. Since these characteristics were not controlled for in the statistical analysis, it is premature to reach a conclusion about gender discrimination.

5.8. (a) Estimate

$$Y_i = \beta_0 + \gamma X_{1i} + \beta_2 \left(X_{1i} + X_{2i} \right) + u_i$$

and test whether $\gamma = 0$.

(b) Estimate

$$Y_i = \beta_0 + \gamma X_{1i} + \beta_2 \left(X_{2i} - aX_{1i} \right) + u_i$$

and test whether $\gamma = 0$.

(c) Estimate

$$Y_i - X_{1i} = \beta_0 + \gamma X_{1i} + \beta_2 \left(X_{2i} - X_{1i} \right) + u_i$$

and test whether $\gamma = 0$.

5.9. Because $R^2 = 1 - \frac{SSR}{TSS}$, $R^2_{unrestricted} - R^2_{restricted} = \frac{SSR_{restricted} - SSR_{unrestricted}}{TSS}$ and $1 - R^2_{unrestricted} = \frac{SSR_{unrestricted}}{TSS}$. Thus from Equation (5.39) we have

$$\begin{aligned} F &= \frac{\left(R^2_{unrestricted} - R^2_{restricted} \right) / q}{\left(1 - R^2_{unrestricted} \right) / \left(n - k_{unrestricted} - 1 \right)} \\ &= \frac{\frac{SSR_{restricted} - SSR_{unrestricted}}{TSS} / q}{\frac{SSR_{unrestricted}}{TSS} / \left(n - k_{unrestricted} - 1 \right)} \\ &= \frac{\left(SSR_{restricted} - SSR_{unrestricted} \right) / q}{SSR_{unrestricted} / \left(n - k_{unrestricted} - 1 \right)} \end{aligned}$$

which is Equation (5.38).

Chapter 6
Nonlinear Regression Functions

6.1. (a) The percentage increase in sales is $100 \times \frac{198-196}{196} = 1.0204\%$. The approximation is $100 \times [\ln(198) - \ln(196)] = 1.0152\%$.

(b) When $Sales_{2002} = 205$, the percentage increase is $100 \times \frac{205-196}{196} = 4.5918\%$ and the approximation is $100 \times [\ln(205) - \ln(196)] = 4.4895\%$.

When $Sales_{2002} = 250$, the percentage increase is $100 \times \frac{250-196}{196} = 27.551\%$ and the approximation is $100 \times [\ln(250) - \ln(196)] = 24.335\%$.

When $Sales_{2002} = 500$, the percentage increase is $100 \times \frac{500-196}{196} = 155.1\%$ and the approximation is $100 \times [\ln(500) - \ln(196)] = 93.649\%$.

(c) The approximation works well when the change is small. The quality of the approximation deteriorates as the percentage change increases.

6.2. (a) According to the regression results in column (1), the house price is expected to increase by 21% ($100\% \times 0.00042 \times 500 = 21\%$) with an additional 500 square feet and other factors held constant. The 95% confidence interval for the percentage change is $100\% \times 500 \times (0.00042 \pm 1.96 \times 0.000038) = [17.276\%, 24.724\%]$.

(b) Because the regressions in columns (1) and (2) have the same dependent variable, we can use $\bar{R}^2$ to compare the fit of these two regressions. The log-log regression in column (2) has the higher $\bar{R}^2$, so it is better to use $\ln(Size)$ to explain house prices.

(c) The house price is expected to increase by 7.1% ($100\% \times 0.071 \times 1 = 7.1\%$) if the house has a swimming pool with other factors held constant. The 95% confidence interval for this effect is $100\% \times (0.071 \pm 1.96 \times 0.034) = [0.436\%, 13.764\%]$.

(d) The house price is expected to increase by 0.36% ($100\% \times 0.0036 \times 1 = 0.36\%$) with an additional bedroom while other factors are held constant. The effect is not statistically significant at a 5% significance level: $|t| = \frac{0.0036}{0.037} = 0.09730 < 1.96$. A reason is that we have held the size of the house constant while considering the effect of adding an additional bedroom.

(e) The quadratic term $\ln(Size)^2$ is not important. The coefficient estimate is not statistically significant at a 5% significance level: $|t| = \frac{0.0078}{0.14} = 0.05571 < 1.96$.

(f) The house price is expected to increase by 7.1% ($100\% \times 0.071 \times 1 = 7.1\%$) when a swimming pool is added to a house without a view and other factors are held constant. The house price is expected to increase by 7.32% ($100\% \times (0.071 \times 1 + 0.0022 \times 1) = 7.32\%$) when a swimming pool is added to a house with a view and other factors are held constant. The difference in the expected percentage change in price is 0.22%. The difference is not statistically significant at a 5% significance level: $|t| = \frac{0.0022}{0.10} = 0.022 < 1.96$.

6.3. (a) The regression functions for hypothetical values of the regression coefficients that are consistent with the educator's statement are: $\beta_1 > 0$ and $\beta_2 < 0$. When $TestScore$ is plotted against STR the regression will show three horizontal segments. The first segment will be for values of $STR < 20$; the next segment for $20 \leq STR \leq 25$; the final segment for $STR > 25$. The first segment will be higher than the second, and the second segment will be higher than the third.

(b) It happens because of perfect multicollinearity. With all three class size binary variables included in the regression, it is impossible to compute the OLS estimates because the intercept is a perfect linear function of the three class size regressors.

6.4. Note that

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \beta_2 X^2 \\ &= \beta_0 + (\beta_1 + 21\beta_2) X + \beta_2 \left(X^2 - 21X\right). \end{aligned}$$

We can define a new independent variable $Z = X^2 - 21X$, and estimate

$$Y = \beta_0 + \gamma X + \beta_2 Z + u_i.$$

The confidence interval is $\hat{\gamma} \pm 1.96 \times \mathrm{SE}(\hat{\gamma})$.

6.5. (a) $\Delta Y = f(X_1 + \Delta X_1, X_2) - f(X_1, X_2) = \beta_1 \Delta X_1 + \beta_3 \Delta X_1 \times X_2$, so $\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$.

(b) $\Delta Y = f(X_1, X_2 + \Delta X_2) - f(X_1, X_2) = \beta_2 \Delta X_2 + \beta_3 X_1 \times \Delta X_2$, so $\frac{\Delta Y}{\Delta X_2} = \beta_2 + \beta_3 X_1$.

(c)

$$\begin{aligned} \Delta Y &= f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2) \\ &= \beta_0 + \beta_1 (X_1 + \Delta X_1) + \beta_2 (X_2 + \Delta X_2) + \beta_3 (X_1 + \Delta X_1)(X_2 + \Delta X_2) \\ &\quad - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2) \\ &= (\beta_1 + \beta_3 X_2) \Delta X_1 + (\beta_2 + \beta_3 X_1) \Delta X_2 + \beta_3 \Delta X_1 \Delta X_2. \end{aligned}$$

Chapter 7
Assessing Studies Based on Multiple Regression

7.1. As explained in the text, potential threats to external validity arise from differences between the population and setting studied and the population and setting of interest. The statistical results based on New York in the 1970's are likely to apply to Boston in the 1970's but not to Los Angeles in the 1970's. In 1970, New York and Boston had large and widely used public transportation systems. Attitudes about smoking were roughly the same in New York and Boston in the 1970s. In contrast, Los Angeles had a considerably smaller public transportation system in 1970. Most residents of Los Angeles relied on their cars to commute to work, school, and so forth.

The results from New York in the 1970's are unlikely to apply to New York in 2002. Attitudes towards smoking changed significantly from 1970 to 2002.

7.2. (a) When Y_i is measured with error, we have $\tilde{Y}_i = Y_i + w_i$, or $Y_i = \tilde{Y}_i - w_i$. Substituting the 2nd equation into the regression model $Y_i = \beta_0 + \beta_1 X_i + u_i$ gives $\tilde{Y}_i - w_i = \beta_0 + \beta_1 X_i + u_i$, or $\tilde{Y}_i = \beta_0 + \beta_1 X_i + u_i + w_i$. Thus $v_i = u_i + w_i$.

(b) (1) The error term v_i has conditional mean zero given X_i:

$$E(v_i|X_i) = E(u_i + w_i|X_i) = E(u_i|X_i) + E(w_i|X_i) = 0 + 0 = 0.$$

(2) $\tilde{Y}_i = Y_i + w_i$ is i.i.d since both Y_i and w_i are i.i.d. and mutually independent; X_i and $\tilde{Y}_j$ $(i \neq j)$ are independent since X_i is independent of both Y_j and w_j. Thus, $\left(X_i, \tilde{Y}_i\right)$, $i = 1, ..., n$ are i.i.d. draws from their joint distribution.

(3) $v_i = u_i + w_i$ has a finite fourth moment given that both u_i and w_i have finite fourth moments and are mutually independent. So (X_i, v_i) have nonzero finite fourth moments.

(c) The OLS estimators are consistent because the least squares assumptions hold.

(d) Because of the validity of the least squares assumptions, we can construct the confidence intervals in the usual way.

(e) The answer here is the economists' "On the one hand, and on the other hand." On the one hand, the statement is true: i.i.d. measurement error in X means that the OLS estimators are inconsistent and inferences based on OLS are invalid. OLS estimators are consistent and OLS inference is valid when Y has i.i.d. measurement error. On the other hand, even if the measurement error in Y is i.i.d. and independent of Y_i and X_i, it increases the variance of the regression error ($\sigma_v^2 = \sigma_u^2 + \sigma_w^2$), and this will increase the variance of the OLS estimators. Also, measurement error that is not i.i.d. may change these results, although this would need to be studied on a case-by-case basis.

7.3. The key is that the selected sample contains only employed women. Consider two women, Beth and Julie. Beth has no children, Julie has one child. Beth and Julie are otherwise identical. Both can earn $25,000 per year in the labor market. Each must compare the $25,000 benefit to the costs of working. For Beth, the cost of working is forgone leisure. For Julie, it is forgone leisure and the costs (pecuniary and other) of child care. If Beth is just on the margin between working in the labor market or not, then Julie, who has a higher opportunity cost, will decide not to work in the labor market. Instead, Julie will work in "home production," caring for children, and so forth. Thus, on average, women with children who decide to work are women who earn higher wages in the labor market.

Chapter 8
Regression with Panel Data

8.1. (a) With \$1 increases in the beer tax, the expected number of lives that would be saved is 0.45 per $10,000$ people. Since New Jersey has a population of 8.1 million, the expected number of lives saved is $0.45 \times 810 = 364.5$. The 95% confidence interval is $(0.45 \pm 1.96 \times 0.22) \times 810 = [15.228,\ 713.77]$.

(b) When New Jersey lowers its drinking age from 21 to 18, the expected fatality rate increases by 0.028 deaths per 10,000. The 95% confidence interval for the change in death rate is $0.028 \pm 1.96 \times 0.066 = [-0.1014,\ 0.1574]$. With a population of 8.1 million, the number of fatalities will increase by $0.028 \times 810 = 22.68$ with a 95% confidence interval $[-0.1014,\ 0.1574] \times 810 = [-82.134, 127.49]$.

(c) When real income per capita in New Jersey increases by 1%, the expected fatality rate increases by 1.81 deaths per 10,000. The 90% confidence interval for the change in death rate is $1.81 \pm 1.64 \times 0.47 = [1.04,\ 2.58]$. With a population of 8.1 million, the number of fatalities will increase by $1.81 \times 810 = 1466.1$ with a 90% confidence interval $[1.04,\ 2.58] \times 810 = [840, 2092]$.

(d) The low p-value (or high F-statistic) associated with the F-test on the assumption that time effects are zero suggests that the time effects should be included in the regression.

(e) The difference in the significance levels arises primarily because the estimated coefficient is higher in (5) than in (4). However, (5) leaves out two variables (unemployment rate and real income per capita) that are statistically significant. Thus, the estimated coefficient on $Beer\ Tax$ in (5) may suffer from omitted variable bias. The results from (4) seem more reliable. In general, statistical significance should be used to measure reliability only if the regression is well-specified (no important omitted variable bias, correct functional form, no simultaneous causality or selection bias, and so forth.)

(f) Define a binary variable $west$ which equals 1 for the western states and 0 for the other states. Include the cross term between the binary variable $west$ and the unemployment rate, $west \times$(unemployment rate), in the regression equation corresponding to column (4). Suppose the coefficient associated with unemployment rate is β, and the coefficient associated with $west \times$ (unemployment rate) is γ. Then β captures the effect of the unemployment rate in the eastern states, and $\beta + \gamma$ captures the effect of the unemployment rate in the western states. The difference in the effect of the unemployment rate in the western and eastern states is γ. Using the coefficient estimate ($\hat{\gamma}$) and the standard error SE($\hat{\gamma}$), we can calculate the t-statistic to test whether γ is statistically significant at a given significance level.

8.2. (a) For each observation, there is one and only one binary regressor equal to one. That is, $D1_i + D2_i + D3_i = 1 = X_{0,it}$.

(b) For each observation, there is one and only one binary regressor that equals 1. That is, $D1_i + D2_i + \cdots + Dn_i = 1 = X_{0,it}$.

(c) The inclusion of all the binary regressors and the "constant" regressor causes the perfect multicollinearity problem. The constant regressor is a perfect linear function of the n binary regressors. OLS estimators cannot be computed in this case. Your computer program should print out a message to this effect. (Different programs print different messages for this problem. Why not try this, and see what your program says?)

8.3. The five potential threats to the internal validity of a regression study are: omitted variables, misspecification of the functional form, imprecise measurement of the independent variables, sample selection, and simultaneous causality. You should think about these one-by-one. Are there important omitted variables that affect traffic fatalities and that may be correlated with the other variables included in the regression? The most obvious candidates are the safety of roads, weather, and so forth. These variables are essentially constant over the sample period, so their effect is captured by the state fixed effects. You may think of something that we missed. Since most of the variables are binary variables, the largest functional form choice involves the $Beer\ Tax$ variable. A linear specification is used in the text, which seems generally consistent with the data in Figure 8.2. To check the reliability of the linear specification, it would be useful to consider a log specification or a quadratic. Measurement error does not appear to a problem, as variables like traffic fatalities and taxes are accurately measured. Similarly, sample selection is a not a problem because data were used from all of the states. Simultaneous causality could be a potential problem. That is, states with high fatality rates might decide to increase taxes to reduce consumption. Expert knowledge is required to determine if this is a problem.

Chapter 9
Regression with a Binary Dependent Variable

9.1. Using the probit model in Equation (9.8):
(a) For a black applicant having a P/I ratio of 0.35, the probability that the application will be denied is $\Phi(-2.26 + 2.74 \times 0.35 + 0.71) = \Phi(-0.59) = 27.76\%$.

(b) With the P/I ratio reduced to 0.30, the probability of being denied is $\Phi(-2.26 + 2.74 \times 0.30 + 0.71) = \Phi(-0.73) = 23.27\%$. The difference in denial probabilities compared to (a) is 4.4 percentage points lower.

(c) For a white applicant having a P/I ratio of 0.35, the probability that the application will be denied is $\Phi(-2.26 + 2.74 \times 0.35) = 9.7\%$. If the P/I ratio is reduced to 0.30, the probability of being denied is $\Phi(-2.26 + 2.74 \times 0.30) = 7.5\%$. The difference in denial probabilities is 2.2 percentage points lower.

(d) From the results in parts (a)-(c), we can see that the marginal effect of the P/I ratio on the probability of mortgage denial depends on race. In the probit regression functional form, the marginal effect depends on the level of probability which in turn depends on the race of the applicant. The coefficient on *black* is statistically significant at the 1% level.

9.2. Using the logit model in Equation (9.10):
(a) For a black applicant having a P/I ratio of 0.35, the probability that the application will be denied is $F(-4.13 + 5.37 \times 0.35 + 1.27) = \frac{1}{1+e^{0.9805}} = 27.28\%$.

(b) With the P/I ratio reduced to 0.30, the probability of being denied is $F(-4.13 + 5.37 \times 0.30 + 1.27) = \frac{1}{1+e^{1.249}} = 22.29\%$. The difference in denial probabilities compared to (a) is 4.99 percentage points lower.

(c) For a white applicant having a P/I ratio of 0.35, the probability that the application will be denied is $F(-4.13 + 5.37 \times 0.35) = \frac{1}{1+e^{2.2505}} = 9.53\%$. If the P/I ratio is reduced to 0.30, the probability of being denied is $F(-4.13 + 5.37 \times 0.30) = \frac{1}{1+e^{2.519}} = 7.45\%$. The difference in denial probabilities is 2.08 percentage points lower.

(d) From the results in parts (a)-(c), we can see that the marginal effect of the P/I ratio on the probability of mortgage denial depends on race. In the logit regression functional form, the marginal effect depends on the level of probability which in turn depends on the race of the applicant. The coefficient on *black* is statistically significant at the 1% level. The logit and probit results are similar.

9.3. (a) Since Y_i is a binary variable, we know $E(Y_i|X_i) = 1 \times \Pr(Y_i = 1|X_i) + 0 \times \Pr(Y_i = 0|X_i) = \Pr(Y_i = 1|X_i) = \beta_0 + \beta_1 X_i$. Thus

$$\begin{aligned} E(u_i|X_i) &= E[Y_i - (\beta_0 + \beta_1 X_i)|X_i] \\ &= E(Y_i|X_i) - (\beta_0 + \beta_1 X_i) = 0 \end{aligned}$$

.

(b) Using Equation (2.7), we have

$$\begin{aligned} \text{var}(Y_i|X_i) &= \Pr(Y_i = 1|X_i)[1 - \Pr(Y_i = 1|X_i)] \\ &= (\beta_0 + \beta_1 X_i)[1 - (\beta_0 + \beta_1 X_i)]. \end{aligned}$$

Thus

$$\begin{aligned} \text{var}(u_i|X_i) &= \text{var}[Y_i - (\beta_0 + \beta_1 X_i)_i |X_i] \\ &= \text{var}(Y_i|X_i) = (\beta_0 + \beta_1 X_i)[1 - (\beta_0 + \beta_1 X_i)]. \end{aligned}$$

(c) var$(u_i|X_i)$ depends on the value of X_i, so u_i is heteroskedastic.

(d) The probability that $Y_i = 1$ conditional on X_i is $p_i = \beta_0 + \beta_1 X_i$. The conditional probability distribution for the i^{th} observation is $\Pr(Y_i = y_i|X_i) = p_i^{y_i}(1 - p_i)^{1-y_i}$. Assuming that (X_i, Y_i) are i.i.d., $i = 1, ..., n$, the joint probability distribution of $Y_1, ..., Y_n$ conditional on the $X's$ is

$$\begin{aligned} \Pr(Y_1 = y_1, ..., Y_n = y_n|X_1, ..., X_n) &= \prod_{i=1}^{n} \Pr(Y_i = y_i|X_i) \\ &= \prod_{i=1}^{n} p_i^{y_i}(1 - p_i)^{1-y_i} \\ &= \prod_{i=1}^{n} (\beta_0 + \beta_1 X_i)^{y_i}[1 - (\beta_0 + \beta_1 X_i)]^{1-y_i}. \end{aligned}$$

The likelihood function is the above joint probability distribution treated as a function of the unknown coefficients (β_0 and β_1).

9.4. (a) The coefficient on *black* is 0.084, indicating an estimated denial probability that is 8.4 percentage points higher for the black applicant.

(b) The 95% confidence interval is $0.084 \pm 1.96 \times 0.023 = [3.89\%, 12.91\%]$.

(c) The answer in (a) will be biased if there are omitted variables which are race-related and have impacts on mortgage denial. Such variables would have to be related with race and also be related with the probability of default on the mortgage (which in turn would lead to denial of the mortgage application). Standard measures of default probability (past credit history and employment variables) are included in the regressions shown in Table 9.2, so omitted variables are unlikely to bias the answer in (a).

9.5. (a) Let $n_1 = \#(Y = 1)$, the number of observations on the random variable Y which equals 1; and $n_2 = \#(Y = 2)$. Then $\#(Y = 3) = n - n_1 - n_2$. The joint probability distribution of $Y_1, ..., Y_n$ is

$$\Pr(Y_1 = y_1, ..., Y_n = y_n) = \prod_{i=1}^{n} \Pr(Y_i = y_i) = p^{n_1} q^{n_2} (1 - p - q)^{n - n_1 - n_2}.$$

The likelihood function is the above joint probability distribution treated as a function of the unknown coefficients (p and q).

(b) The MLEs of p and q maximize the likelihood function. Let's use the log-likelihood function

$$\begin{aligned} L &= \ln[\Pr(Y_1 = y_1, ..., Y_n = y_n)] \\ &= n_1 \ln p + n_2 \ln q + (n - n_1 - n_2) \ln(1 - p - q). \end{aligned}$$

Using calculus, the partial derivatives of L are

$$\begin{aligned} \frac{\partial L}{\partial p} &= \frac{n_1}{p} - \frac{n - n_1 - n_2}{1 - p - q}, \text{ and} \\ \frac{\partial L}{\partial q} &= \frac{n_2}{q} - \frac{n - n_1 - n_2}{1 - p - q}. \end{aligned}$$

Setting these two equations equal to zero and solving the resulting equations yield the MLE of p and q:

$$\hat{p} = \frac{n_1}{n}, \qquad \hat{q} = \frac{n_2}{n}.$$

Chapter 10
Instrumental Variables Regression

10.1. (a) The change in the regressor, $\ln\left(P_{i,1995}^{cigarettes}\right) - \ln\left(P_{i,1985}^{cigarettes}\right)$, from a \$0.10 per pack increase in the retail price is $\ln 2.10 - \ln 2.00 = 0.0488$. The expected percentage change in cigarette demand is $-0.94 \times 0.0488 \times 100\% = -4.5872\%$. The 95% confidence interval is $(-0.94 \pm 1.96 \times 0.21) \times 0.0488 \times 100\% = [-6.60\%, -2.58\%]$.

(b) With a 2% reduction in income, the expected percentage change in cigarette demand is $0.53 \times (-0.02) \times 100\% = -1.06\%$.

(c) The regression in column (1) will not provide a reliable answer to the question in (b) when recessions last less than 1 year. The regression in column (1) studies the long-run price and income elasticity. Cigarettes are addictive. The response of demand to an income decrease will be smaller in the short run than in the long run.

(d) The instrumental variable would be too weak (irrelevant) if the F-statistic in column (1) was 3.6 instead of 33.6, and we cannot rely on the standard methods for statistical inference. Thus the regression would not provide a reliable answer to the question posed in (a).

10.2. (a) When there is only one X, we only need to check that the instrument enters the first stage population regression. Since the instrument is $Z = X$, the regression of X onto Z will have a coefficient of 1.0 on Z, so that the instrument enters the first stage population regression. Key Concept 4.3 implies $\text{corr}(X_i, u_i) = 0$, and this implies $\text{corr}(Z_i, u_i) = 0$. Thus, the instrument is exogenous.

(b) Condition 1 is satisfied because there are no W's. Key Concept 4.3 implies that condition 2 is satisfied because (X_i, Z_i, Y_i) are i.i.d. draws from their joint distribution. Condition 3 is also satisfied by applying assumption 3 in Key Concept 4.3. Condition 4 is satisfied because there are no W's. Condition 5 is satisfied because of conclusions in part (a).

(c) The TSLS estimator is $\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}}$ using Equation (10.4) in the text. Since $Z_i = X_i$, we have

$$\hat{\beta}_1^{TSLS} = \frac{s_{ZY}}{s_{ZX}} = \frac{s_{XY}}{s_X^2} = \hat{\beta}_1^{OLS}.$$

10.3. (a) The estimator $\hat{\sigma}_a^2 = \frac{1}{n-2}\sum_{i=1}^n \left(Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}\hat{X}_i\right)^2$ is not consistent. Write this as $\widehat{\sigma}_a^2 = \frac{1}{n-2}\sum_{i=1}^n \left(\widehat{u}_i - \hat{\beta}_1^{TSLS}(\hat{X}_i - X_i)\right)^2$, where $\widehat{u}_i = Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}X_i$. Replacing $\hat{\beta}_1^{TSLS}$ with β_1, as suggested in the question, write this as $\hat{\sigma}_a^2 \approx \frac{1}{n}\sum_{i=1}^n \left(u_i - \beta_1(\hat{X}_i - X_i)\right)^2 = \frac{1}{n}\sum_{i=1}^n u_i^2 + \frac{1}{n}\sum_{i=1}^n [\beta_1^2(\hat{X}_i - X_i)^2 + 2u_i\beta_1(\hat{X}_i - X_i)]$. The first term on the right hand side of the equation converges to σ_u^2, but the second term converges to something that is non-zero. Thus $\hat{\sigma}_a^2$ is not consistent.

(b) The estimator $\widehat{\sigma}_b^2 = \frac{1}{n-2}\sum_{i=1}^n \left(Y_i - \hat{\beta}_0^{TSLS} - \hat{\beta}_1^{TSLS}X_i\right)^2$ is consistent. Using the same notation as in (a), we can write $\hat{\sigma}_b^2 \approx \frac{1}{n}\sum_{i=1}^n u_i^2$, and this estimator converges in probability to σ_u^2.

10.4. Using $\hat{X}_i = \hat{\pi}_0 + \hat{\pi}_1 Z_i$, we have $\bar{\hat{X}} = \hat{\pi}_0 + \hat{\pi}_1\bar{Z}$ and

$$\begin{aligned} s_{\hat{X}Y} &= \sum_{i=1}^n \left(\widehat{X}_i - \bar{\hat{X}}\right)\left(Y_i - \bar{Y}\right) = \hat{\pi}_1 \sum_{i=1}^n \left(Z_i - \bar{Z}\right)\left(Y_i - \bar{Y}\right) = \hat{\pi}_1 s_{ZY}, \\ s_{\hat{X}}^2 &= \sum_{i=1}^n \left(\widehat{X}_i - \bar{\hat{X}}\right)^2 = \hat{\pi}_1^2 \sum_{i=1}^n \left(Z_i - \bar{Z}\right)^2 = \hat{\pi}_1^2 s_Z^2. \end{aligned}$$

Using the formula for the OLS estimator in Key Concept 4.2, we have

$$\hat{\pi}_1 = \frac{s_{ZX}}{s_Z^2}.$$

Thus the TSLS estimator

$$\hat{\beta}_1^{TSLS} = \frac{s_{\hat{X}Y}}{s_{\hat{X}}^2} = \frac{\hat{\pi}_1 s_{ZY}}{\hat{\pi}_1^2 s_Z^2} = \frac{s_{ZY}}{\hat{\pi}_1 s_Z^2} = \frac{s_{ZY}}{\frac{s_{ZX}}{s_Z^2} \times s_Z^2} = \frac{s_{ZY}}{s_{ZX}}.$$

Chapter 11
Experiments and Quasi-Experiments

11.1. For students in kindergarten, the estimated small class treatment effect relative to being in a regular class is an increase of 13.90 points on the test with a standard error 2.45. The 95% confidence interval is $13.90 \pm 1.96 \times 2.45 = [9.098, 18.702]$.

For students in grade 1, the estimated small class treatment effect relative to being in a regular class is an increase of 29.78 points on the test with a standard error 2.83. The 95% confidence interval is $29.78 \pm 1.96 \times 2.83 = [24.233, 35.327]$.

For students in grade 2, the estimated small class treatment effect relative to being in a regular class is an increase of 19.39 points on the test with a standard error 2.71. The 95% confidence interval is $19.39 \pm 1.96 \times 2.71 = [14.078, 24.702]$.

For students in grade 3, the estimated small class treatment effect relative to being in a regular class is an increase of 15.59 points on the test with a standard error 2.40. The 95% confidence interval is $15.59 \pm 1.96 \times 2.40 = [10.886, 20.294]$.

11.2. (a) On average, a student in class A (the "small class") is expected to score higher than a student in class B (the "regular class") by 15.89 points with a standard error 2.16. The 95% confidence interval for the predicted difference in average test scores is $15.89 \pm 1.96 \times 2.16 = [11.656, 20.124]$.

(b) On average, a student in class A taught by a teacher with 5 years of experience is expected to score lower than a student in class B taught by a teacher with 10 years of experience by $0.66 \times 5 = 3.3$ points. The standard error for the score difference is $0.17 \times 5 = 0.85$. The 95% confidence interval for the predicted lower score for students in classroom A is $3.3 \pm 1.96 \times 0.85 = [1.634, 4.966]$.

(c) The expected difference in average test scores is $15.89 + 0.66 \times (-5) = 12.59$. Because of random assignment, the estimators of the small class effect and the teacher experience effect are uncorreleated. Thus, the standard error for the difference in average test scores is $\left[2.16^2 + (-5)^2 \times 0.17^2\right]^{\frac{1}{2}} = 2.3212$. The 95% confidence interval for the predicted difference in average test scores in classrooms A and B is $12.59 \pm 1.96 \times 2.3212 = [8.0404, 17.140]$.

(d) The intercept is not included in the regression to avoid the perfect multicollinearity problem that exists among the intercept and school indicator variables.

11.3. (a) This is an example of attrition, which poses a threat to internal validity. After the male athletes leave the experiment, the remaining subjects are representative of a population that excludes male athletes. If the average causal effect for this population is the same as the average causal effect for the population that includes the male athletes, then the attrition does not affect the internal validity of the experiment. On the other hand, if the average causal effect for male athletes differs from the rest of population, internal validity has been compromised.

(b) This is an example of partial compliance which is a threat to internal validity. The local area network is a failure to follow treatment protocol, and this leads to bias in the OLS estimator of the average causal effect.

(c) This poses no threat to internal validity. As stated, the study is focused on the effect of dorm room Internet connections. The treatment is making the connections available in the room; the treatment is not the use of the Internet. Thus, the art majors received the treatment (although they chose not to use the Internet).

(d) As in part (b) this is an example of partial compliance. Failure to follow treatment protocol leads to bias in the OLS estimator.

11.4. The treatment effect is modeled using the fixed effects specification

$$Y_{it} = \alpha_i + \beta_1 X_{it} + u_{it}.$$

(a) α_i is an individual-specific intercept. The random effect in the regression has variance

$$\begin{aligned} \text{var}(\alpha_i + u_{it}) &= \text{var}(\alpha_i) + \text{var}(u_{it}) + 2\text{cov}(\alpha_i, u_{it}) \\ &= \sigma_\alpha^2 + \sigma_u^2 \end{aligned}$$

which is homoskedastic. The differences estimator is constructed using data from time period $t = 2$. Using Equation (4.60), it is straightforward to see that the variance for the differences estimator

$$n\text{var}\left(\hat{\beta}_1^{differences}\right) \longrightarrow \frac{\text{var}(\alpha_i + u_{i2})}{\text{var}(X_{i2})} = \frac{\sigma_\alpha^2 + \sigma_u^2}{\text{var}(X_{i2})}.$$

(b) The regression equation using the differences-in-differences estimator is

$$\Delta Y_i = \beta_1 \Delta X_i + v_i$$

with $\Delta Y_i = Y_{i2} - Y_{i1}$, $\Delta X_i = X_{i2} - X_{i1}$, and $v_i = u_{i2} - u_{i1}$. If the i^{th} individual is in the treatment group at time $t = 2$, then $\Delta X_i = X_{i2} - X_{i1} = 1 - 0 = 1 = X_{i2}$. If the i^{th} individual is in the control group at time $t = 2$, then $\Delta X_i = X_{i2} - X_{i1} = 0 - 0 = 0 = X_{i2}$. Thus ΔX_i is a binary treatment variable and $\Delta X_i = X_{i2}$, which in turn implies $\text{var}(\Delta X_i) = \text{var}(X_{i2})$. The variance for the new error term is

$$\begin{aligned} \sigma_v^2 &= \text{var}(u_{i2} - u_{i1}) = \text{var}(u_{i2}) + \text{var}(u_{i1}) - 2\text{cov}(u_{i2}, u_{i1}) \\ &= 2\sigma_u^2 \end{aligned}$$

which is homoskedastic. Using Equation (4.60), it is straightforward to see that the variance for the differences-in-differences estimator

$$n\text{var}\left(\hat{\beta}_1^{diffs-in-diffs}\right) \longrightarrow \frac{\sigma_v^2}{\text{var}\left(\Delta X_i\right)} = \frac{2\sigma_u^2}{\text{var}\left(X_{i2}\right)}.$$

(c) When $\sigma_\alpha^2 > \sigma_u^2$, we'll have $\text{var}\left(\hat{\beta}_1^{differences}\right) > \text{var}\left(\hat{\beta}_1^{diffs-in-diffs}\right)$ and the differences-in-differences estimator is more efficient then the differences estimator. Thus, if there is considerable large variance in the individual-specific fixed effects, it is better to use the differences-in-differences estimator.

11.5. From the population regression Equation (11.13)

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2\left(D_t \times W_i\right) + \beta_0 D_t + v_{it},$$

we have

$$Y_{i2} - Y_{i1} = \beta_1\left(X_{i2} - X_{i1}\right) + \beta_2\left[\left(D_2 - D_1\right) \times W_i\right] + \beta_0\left(D_2 - D_1\right) + \left(v_{i2} - v_{i1}\right).$$

By defining $\Delta Y_i = Y_{i2} - Y_{i1}$, $\Delta X_i = X_{i2} - X_{i1}$ (a binary treatment variable) and $u_i = v_{i2} - v_{i1}$, and using $D_1 = 0$ and $D_2 = 1$, we can rewrite this equation as

$$\Delta Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i,$$

which is Equation (11.5) in the case of a single W regressor.

11.6. The regression model is

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_2 G_i + \beta_3 B_t + u_{it},$$

Using the results in Section 6.3

$$\begin{aligned}
\overline{Y}^{control,before} &= \widehat{\beta}_0 \\
\overline{Y}^{control,after} &= \widehat{\beta}_0 + \widehat{\beta}_3 \\
\overline{Y}^{treatment,before} &= \widehat{\beta}_0 + \widehat{\beta}_2 \\
\overline{Y}^{treatment,after} &= \widehat{\beta}_0 + \widehat{\beta}_1 + \widehat{\beta}_2 + \widehat{\beta}_3
\end{aligned}$$

Thus

$$\begin{aligned}
\widehat{\beta}^{diffs-in-diffs} &= (\overline{Y}^{treatment,after} - \overline{Y}^{treatment,before}) \\
&\quad -(\overline{Y}^{control,after} - \overline{Y}^{control,before}) \\
&= (\widehat{\beta}_1 + \widehat{\beta}_3) - (\widehat{\beta}_3) = \widehat{\beta}_1
\end{aligned}$$

11.7. The covariance between $\beta_{1i}X_i$ and X_i is

$$\begin{aligned}\operatorname{cov}(\beta_{1i}X_i,\, X_i) &= E\{[\beta_{1i}X_i - E(\beta_{1i}X_i)][X_i - E(X_i)]\} \\ &= E\left\{\beta_{1i}X_i^2 - E(\beta_{1i}X_i)X_i - \beta_{1i}X_iE(X_i) + E(\beta_{1i}X_i)E(X_i)\right\} \\ &= E\left(\beta_{1i}X_i^2\right) - E(\beta_{1i}X_i)E(X_i)\end{aligned}$$

Because X_i is randomly assigned, X_i is distributed independently of β_{1i}. The independence means

$$E(\beta_{1i}X_i) = E(\beta_{1i})E(X_i) \qquad \text{and} \qquad E\left(\beta_{1i}X_i^2\right) = E(\beta_{1i})E\left(X_i^2\right).$$

Thus $\operatorname{cov}(\beta_{1i}X_i,\, X_i)$ can be further simplified:

$$\begin{aligned}\operatorname{cov}(\beta_{1i}X_i,\, X_i) &= E(\beta_{1i})\left[E\left(X_i^2\right) - E^2(X_i)\right] \\ &= E(\beta_{1i})\sigma_X^2.\end{aligned}$$

So

$$\frac{\operatorname{cov}(\beta_{1i}X_i,\, X_i)}{\sigma_X^2} = \frac{E(\beta_{1i})\sigma_X^2}{\sigma_X^2} = E(\beta_{1i}).$$

Chapter 12
Introduction to Time Series Regression and Forecasting

12.1. (a) Continuing to substitute $Y_{t-j} = 2.5 + 0.7Y_{t-j-1} + u_{t-j}$, $j = 1, 2, ...\infty$, into the expression $Y_t = 2.5 + 0.7Y_{t-1} + u_t$ yields

$$\begin{aligned}
Y_t &= 2.5 + 0.7\left(2.5 + 0.7Y_{t-2} + u_{t-1}\right) + u_t \\
&= (1 + 0.7)\, 2.5 + 0.7^2\left(2.5 + 0.7Y_{t-3} + u_{t-2}\right) + u_t + 0.7u_{t-1} \\
&= \ldots\ldots \\
&= \left(1 + 0.7 + 0.7^2 + \cdots\right) 2.5 + \left(u_t + 0.7u_{t-1} + 0.7^2 u_{t-1} + \cdots\right) \\
&= 2.5\sum_{i=0}^{\infty} 0.7^i + \sum_{i=0}^{\infty} 0.7^i u_{t-i} \\
&= 2.5 \times \frac{1}{1-0.7} + \sum_{i=0}^{\infty} 0.7^i u_{t-i} \\
&= \frac{25}{3} + \sum_{i=0}^{\infty} 0.7^i u_{t-i}.
\end{aligned}$$

Because u_t is i.i.d. with $E(u_t) = 0$ and $\mathrm{var}(u_t) = 9$, the mean and variance of Y_t are

$$\begin{aligned}
\mu_Y &= E(Y_t) = E\left(\frac{25}{3} + \sum_{i=0}^{\infty} 0.7^i u_{t-i}\right) \\
&= \frac{25}{3} + \sum_{i=0}^{\infty} 0.7^i E(u_{t-i}) \\
&= \frac{25}{3} = 8.333. \\
\sigma_Y^2 &= \mathrm{var}(Y_t) = \mathrm{var}\left(\frac{25}{3} + \sum_{i=0}^{\infty} 0.7^i u_{t-i}\right) \\
&= \sum_{i=0}^{\infty} 0.7^{2i}\mathrm{var}(u_{t-i}) \\
&= \sum_{i=0}^{\infty} 0.7^{2i} \times 9 \\
&= \frac{9}{1-0.7^2} = 17.647.
\end{aligned}$$

(b) The 1st autocovariance is

$$\begin{aligned}
\text{cov}(Y_t, Y_{t-1}) &= \text{cov}(2.5 + 0.7Y_{t-1} + u_t,\, Y_{t-1}) \\
&= 0.7\text{var}(Y_{t-1}) + \text{cov}(u_t, Y_{t-1}) \\
&= 0.7\sigma_Y^2 \\
&= 0.7 \times 17.647 = 12.353.
\end{aligned}$$

The 2nd autocovariance is

$$\begin{aligned}
\text{cov}(Y_t, Y_{t-2}) &= \text{cov}\left[(1 + 0.7)\, 2.5 + 0.7^2 Y_{t-2} + u_t + 0.7u_{t-1},\, Y_{t-2}\right] \\
&= 0.7^2 \text{var}(Y_{t-2}) + \text{cov}(u_t + 0.7u_{t-1}, Y_{t-2}) \\
&= 0.7^2 \sigma_Y^2 \\
&= 0.7^2 \times 17.647 = 8.6471.
\end{aligned}$$

(c) The 1st autocorrelation is

$$\text{corr}(Y_t, Y_{t-1}) = \frac{\text{cov}(Y_t, Y_{t-1})}{\sqrt{\text{var}(Y_t)\,\text{var}(Y_{t-1})}} = \frac{0.7\sigma_Y^2}{\sigma_Y^2} = 0.7.$$

The 2nd autocorrelation is

$$\text{corr}(Y_t, Y_{t-2}) = \frac{\text{cov}(Y_t, Y_{t-2})}{\sqrt{\text{var}(Y_t)\,\text{var}(Y_{t-2})}} = \frac{0.7^2\sigma_Y^2}{\sigma_Y^2} = 0.49.$$

(d) The conditional expectation Y_{T+1} given Y_T is

$$Y_{T+1|T} = 2.5 + 0.7Y_T = 2.5 + 0.7 \times 102.3 = 74.11.$$

12.2. (a) The statement is correct. The monthly percentage change in IP is $\frac{IP_t - IP_{t-1}}{IP_{t-1}} \times 100$ which can be approximated by $[\ln(IP_t) - \ln(IP_{t-1})] \times 100 = 100 \times \ln(\frac{IP_t}{IP_{t-1}})$ when the change is small. Converting this into an annual (12 month) change yields $1200 \times \ln(\frac{IP_t}{IP_{t-1}})$.

(b) The values of Y from the table are

Date	2000:7	2000:8	2000:9	2000:10	2000:11	2000:12
IP	147.595	148.650	148.973	148.660	148.206	146.300
Y		8.55	2.60	−2.52	−3.67	−7.36

The forecasted value of Y_t in January 2001 is

$$\begin{aligned}\hat{Y}_{t|t-1} &= 1.377 + [0.318 \times (-7.36)] + [0.123 \times (-3.67)] \\ &\quad +[0.068 \times (-2.52)] + [0.001 \times (2.60)] \\ &= -1.58.\end{aligned}$$

(c) The t-statistic on Y_{t-12} is $t = \frac{-0.054}{0.053} = -1.0189$ with an absolute value less than 1.96, so the coefficient is not statistically significant at the 5% level.

(d) For the QLR test, there are 5 coefficients (including the constant) that are being allowed to break. Compared to the critical values for $q = 5$ in Table 12.5, the QLR statistic 3.45 is larger than the critical value 3.26 at the 10% level, but less than the critical value 3.66 at the 5% level. Thus the hypothesis that these coefficients are stable is rejected at the 10% significance level, but not at the 5% significance level.

(e) There are $41 \times 12 = 492$ number of observations on the dependent variable. The BIC and AIC are calculated from the formulas $\text{BIC}(p) = \ln\left[\frac{\text{SSR}(p)}{T}\right] + (p+1)\frac{\ln T}{T}$ and $\text{AIC}(p) = \ln\left[\frac{\text{SSR}(p)}{T}\right] + (p+1)\frac{2}{T}$.

AR Order (p)	1	2	3	4	5	6
SSR(p)	29175	28538	28393	28391	28378	28317
$\ln\left[\frac{\text{SSR}(p)}{T}\right]$	4.0826	4.0605	4.0554	4.0553	4.0549	4.0527
$(p+1)\frac{\ln T}{T}$	0.0252	0.0378	0.0504	0.0630	0.0756	0.0882
$(p+1)\frac{2}{T}$	0.0081	0.0122	0.0163	0.0203	0.0244	0.0285
BIC	4.1078	4.0983	4.1058	4.1183	4.1305	4.1409
AIC	4.0907	4.0727	4.0717	4.0757	4.0793	4.0812

The BIC is smallest when $p = 2$. Thus the BIC estimate of the lag length is 2. The AIC is smallest when $p = 3$. Thus the AIC estimate of the lag length is 3.

12.3. (a) To test for a stochastic trend (unit root) in $\ln(IP)$, the ADF statistic is the t-statistic testing the hypothesis that the coefficient on $\ln(IP_{t-1})$ is zero versus the alternative hypothesis that the coefficient on $\ln(IP_{t-1})$ is less than zero. The calculated t-statistic is $t = \frac{-0.018}{0.007} = -2.5714$. From Table 12.4, the 10% critical value with a time trend is -3.12. Because $-2.5714 > -3.12$, the test does not reject the null hypothesis that $\ln(IP)$ has a unit autoregressive root at the 10% significance level. That is, the test does not reject the null hypothesis that $\ln(IP)$ contains a stochastic trend, against the alternative that it is stationary.

(b) The ADF test supports the specification used in Exercise 12.2. The use of first differences in Exercise 12.2 eliminates random walk trend in $\ln(IP)$.

12.4. (a) The critical value for the F-test is 2.372 at a 5% significance level. Since the Granger-causality F-statistic 2.35 is less than the critical value, we cannot reject the null hypothesis that interest rates have no predictive content for IP growth at the 5% level. The Granger causality statistic is significant at the 10% level.

(b) The Granger-causality F-statistic of 2.87 is larger than the critical value, so we conclude at the 5% significance level that IP growth helps to predict future interest rates.

12.5. (a)

$$\begin{aligned} E\left[(W-c)^2\right] &= E\left\{[(W-\mu_W)+(\mu_W-c)]^2\right\} \\ &= E\left[(W-\mu_W)^2\right]+2E\left(W-\mu_W\right)(\mu_W-c)+(\mu_W-c)^2 \\ &= \sigma_W^2+(\mu_W-c)^2 . \end{aligned}$$

(b) Using the result in part (a), the conditional mean squared error

$$E\left[(Y_t-f_{t-1})^2 \,|Y_{t-1},Y_{t-2},\ldots\right]=\sigma_{t|t-1}^2+\left(Y_{t|t-1}-f_{t-1}\right)^2$$

with the conditional variance $\sigma_{t|t-1}^2 = E\left[\left(Y_t-Y_{t|t-1}\right)^2\right]$. This equation is minimized when the second term equals zero, or when $f_{t-1}=Y_{t|t-1}$.

(c) Applying Equation (2.25), we know the error u_t is uncorrelated with u_{t-1} if $E\left(u_t|u_{t-1}\right)=0$. From Equation (12.14) for the AR(p) process, we have

$$u_{t-1}=Y_{t-1}-\beta_0-\beta_1 Y_{t-2}-\beta_2 Y_{t-3}-\cdots-\beta_p Y_{t-p-1}=f\left(Y_{t-1},Y_{t-2},\cdots,Y_{t-p-1}\right),$$

a function of Y_{t-1} and its lagged values. The assumption $E\left(u_t|Y_{t-1},Y_{t-2},\cdots\right)=0$ means that conditional on Y_{t-1} and its lagged values, or any functions of Y_{t-1} and its lagged values, u_t has mean zero. That is,

$$E\left(u_t|u_{t-1}\right)=E\left[u_t|f\left(Y_{t-1},Y_{t-2},\cdots,Y_{t-p-2}\right)\right]=0.$$

Thus u_t and u_{t-1} are uncorrelated. A similar argument shows that u_t and u_{t-j} are uncorrelated for all $j\geq 1$. Thus u_t is serially uncorrelated.

12.6. This exercise requires a Monte Carlo simulation on spurious regression. The answer to (a) will depend on the particular “draw” from your simulation, but your answers should be similar to the ones that I found.

(b) When I did these simulations, the 5%, 50% and 95% quantiles of the R^2 were .00, .19, and .73. The 5%, 50% and 95% quantiles of the t-statisic were

-12.9, -0.02 and 13.01. Your simulations should yield similar values. In 76% of the draws the absolute value of the t-statistic exceeded 1.96.

(c) When I did these simulations with $T = 50$, the 5%, 50% and 95% quantiles of the R^2 were .00, .16, and .68. The 5%, 50% and 95% quantiles of the t-statisic were -8.3, -0.20 and 7.8. Your simulations should yield similar values. In 66% of the draws the absolute value of the t-statistic exceeded 1.96.

When I did these simulations with $T = 200$, the 5%, 50% and 95% quantiles of the R^2 were .00, .17, and .68. The 5%, 50% and 95% quantiles of the t-statisic were -16.8, $-.76$ and 17.24. Your simulations should yield similar values. In 83% of the draws the absolute value of the t-statistic exceeded 1.96.

The quantiles of the R^2 do not seem to change as the sample size changes. However the distribution of the t-statistic becomes more dispersed. In the limit as T grows large, the fraction of the t-statistics that exceed 1.96 in absolute values seems to approach 1.0. (You might find it interesting that $\frac{t-statistc}{\sqrt{T}}$ has a well-behaved limiting distribution. This is consistent with the Monte Carlo presented in this problem.)

Chapter 13
Estimation of Dynamic Causal Effects

13.1. (a) See the table below. β_i is the dynamic multiplier. With the 25% oil price jump, the predicted effect on output growth for the ith quarter is $25\beta_i$ percentage points.

Period ahead (i)	Dynamic multiplier (β_i)	Predicted effect on output growth ($25\beta_i$)	95% confidence interval $25 \times [\beta_i \pm 1.96\text{SE}(\beta_i)]$
0	-0.055	-1.375	$[-4.021, 1.271]$
1	-0.026	-0.65	$[-3.443, 2.143]$
2	-0.031	-0.775	$[-3.127, 1.577]$
3	-0.109	-2.725	$[-4.783, -0.667]$
4	-0.128	-3.2	$[-5.797, -0.603]$
5	0.008	0.2	$[-1.025, 1.425]$
6	0.025	0.625	$[-1.727, 2.977]$
7	-0.019	-0.475	$[-2.386, 1.436]$
8	0.067	1.675	$[-0.015, 0.149]$

(b) The 95% confidence interval for the predicted effect on output growth for the i'th quarter from the 25% oil price jump is $25 \times [\beta_i \pm 1.96\text{SE}(\beta_i)]$ percentage points. The confidence interval is reported in the table in (a).

(c) The predicted cumulative change in GDP growth over eight quarters is

$$25 \times (-0.055 - 0.026 - 0.031 - 0.109 - 0.128 + 0.008 + 0.025 - 0.019) = -8.375\%$$

percentage points.

(d) The 1% critical value for the F-test is 2.407. Since the HAC F-statistic 3.49 is larger than the critical value, we reject the null hypothesis that all the coefficients are zero at the 1% level.

13.2. (a) See the table below. β_i is the dynamic multiplier. With the 25% oil price jump, the predicted change in interest rates for the i'th quarter is $25\beta_i$.

Period ahead (i)	Dynamic multiplier (β_i)	Predicted change in interest rates ($25\beta_i$)	95% confidence interval $25 \times [\beta_i \pm 1.96\mathrm{SE}(\beta_i)]$
0	0.062	1.55	$[-0.655, 3.755]$
1	0.048	1.2	$[-0.466, 2.866]$
2	−0.014	−0.35	$[-1.722, 1.022]$
3	−0.086	−2.15	$[-10.431, 6.131]$
4	−0.000	0	$[-2.842, 2.842]$
5	0.023	0.575	$[-2.61, 3.76]$
6	−0.010	−0.25	$[-2.553, 2.053]$
7	−0.100	−2.5	$[-4.362, -0.638]$
8	−0.014	−0.35	$[-1.575, 0.875]$

(b) The 95% confidence interval for the predicted change in interest rates for the i'th quarter from the 25% oil price jump is $25 \times [\beta_i \pm 1.96\mathrm{SE}(\beta_i)]$. The confidence interval is reported in the table in (a).

(c) The effect of this change in oil prices on the level of interest rates in period $t + 8$ is the price change implied by the cumulative multiplier:

$$25 \times (0.062 + 0.048 - 0.014 - 0.086 - 0.000 + 0.023 - 0.010 - 0.100 - 0.014) = -2.275.$$

(d) The 1% critical value for the F-test is 2.407. Since the HAC F-statistic 4.25 is larger than the critical value, we reject the null hypothesis that all the coefficients are zero at the 1% level.

13.3. The dynamic causal effects are for experiment A. The regression in exercise 13.1 does not control for interest rates, so that interest rates are assumed to evolve in their "normal pattern" given changes in oil prices.

13.4. When oil prices are strictly exogenous, there are two methods to improve upon the estimates. The first method is to use OLS to estimate the coefficients in an ADL model, and to calculate the dynamic multipliers from the estimated ADL coefficients. The second method is to use generalized least squares (GLS) to estimate the coefficients of the distributed lag model.

13.5. Substituting

$$\begin{aligned} X_t &= \Delta X_t + X_{t-1} = \Delta X_t + \Delta X_{t-1} + X_{t-2} \\ &= \cdots\cdots \\ &= \Delta X_t + \Delta X_{t-1} + \cdots + \Delta X_{t-p+1} + X_{t-p} \end{aligned}$$

into Equation (13.4), we have

$$
\begin{aligned}
Y_t &= \beta_0 + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \cdots + \beta_{r+1} X_{t-r} + u_t \\
&= \beta_0 + \beta_1 \left(\Delta X_t + \Delta X_{t-1} + \cdots + \Delta X_{t-r+1} + X_{t-r}\right) \\
&\quad + \beta_2 \left(\Delta X_{t-1} + \cdots + \Delta X_{t-r+1} + X_{t-r}\right) \\
&\quad + \cdots + \beta_r \left(\Delta X_{t-r+1} + X_{t-r}\right) + \beta_{r+1} X_{t-r} + u_t \\
&= \beta_0 + \beta_1 \Delta X_t + \left(\beta_1 + \beta_2\right) \Delta X_{t-1} + \left(\beta_1 + \beta_2 + \beta_3\right) \Delta X_{t-2} \\
&\quad + \cdots + \left(\beta_1 + \beta_2 + \cdots + \beta_r\right) \Delta X_{t-r+1} \\
&\quad + \left(\beta_1 + \beta_2 + \cdots + \beta_r + \beta_{r+1}\right) X_{t-r} + u_t.
\end{aligned}
$$

Comparing the above equation to Equation (13.8), we see $\delta_0 = \beta_0$, $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3$, ..., $\delta_{r+1} = \beta_1 + \beta_2 + \cdots + \beta_r + \beta_{r+1}$.

Chapter 14
Additional Topics in Time Series Regression

14.1. Y_t follows a stationary AR(1) model, $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$. The mean of Y_t is $\mu_Y = E(Y_t) = \frac{\beta_0}{1-\beta_1}$, and $E(u_t|Y_t) = 0$.

(a) The h-period ahead forecast of Y_t, $Y_{t+h|t} = E(Y_{t+h}|Y_t, Y_{t-1}, ...)$, is

$$\begin{aligned}
Y_{t+h|t} = E(Y_{t+h}|Y_t, Y_{t-1}, ...) &= E(\beta_0 + \beta_1 Y_{t+h-1} + u_t|Y_t, Y_{t-1}, ...) \\
&= \beta_0 + \beta_1 Y_{t+h-1|t} = \beta_0 + \beta_1 \left(\beta_0 + \beta_1 Y_{t+h-2|t}\right) \\
&= (1+\beta_1)\beta_0 + \beta_1^2 Y_{t+h-2|t} \\
&= (1+\beta_1)\beta_0 + \beta_1^2 \left(\beta_0 + \beta_1 Y_{t+h-3|t}\right) \\
&= \left(1+\beta_1+\beta_1^2\right)\beta_0 + \beta_1^3 Y_{t+h-3|t} \\
&= \ldots\ldots \\
&= \left(1+\beta_1+\cdots+\beta_1^{h-1}\right)\beta_0 + \beta_1^h Y_t \\
&= \frac{1-\beta_1^h}{1-\beta_1}\beta_0 + \beta_1^h Y_t \\
&= \mu_Y + \beta_1^h (Y_t - \mu_Y).
\end{aligned}$$

(b) Substituting the result from part (a) into X_t gives

$$\begin{aligned}
X_t &= \sum_{i=0}^{\infty} \delta^i Y_{t+i|t} = \sum_{i=0}^{\infty} \delta^i \left[\mu_Y + \beta_1^i (Y_t - \mu_Y)\right] \\
&= \mu_Y \sum_{i=0}^{\infty} \delta^i + (Y_t - \mu_Y) \sum_{i=0}^{\infty} (\beta_1 \delta)^i \\
&= \frac{\mu_Y}{1-\delta} + \frac{Y_t - \mu_Y}{1-\beta_1\delta}.
\end{aligned}$$

14.2. (a) Because $R1_t$ follows a random walk ($R1_t = R1_{t-1} + u_t$), the i-period ahead forecast of $R1_t$ is

$$R1_{t+i|t} = R1_{t+i-1|t} = R1_{t+i-2|t} = \ldots\ldots = R1_t.$$

Thus

$$Rk_t = \frac{1}{k}\sum_{i=1}^{k} R1_{t+i|t} + e_t = \frac{1}{k}\sum_{i=1}^{k} R1_t + e_t = R1_t + e_t.$$

(b) $R1_t$ follows a random walk and is $I(1)$. Rk_t is also $I(1)$. Given that both Rk_t and $R1_t$ are integrated of order one, and $Rk_t - R1_t = e_t$ is integrated of order zero, we can conclude that Rk_t and $R1_t$ are cointegrated. The cointegrating coefficient is 1.

(c) When $\Delta R1_t = 0.5\Delta R1_{t-1} + u_t$, $\Delta R1_t$ is stationary but $R1_t$ is not stationary. $R1_t = 1.5R1_{t-1} - 0.5R1_{t-2} + u_t$, an AR(2) process with a unit autoregressive root. That is, $R1_t$ is $I(1)$. The i-period ahead forecast of $\Delta R1_t$ is

$$\Delta R1_{t+i|t} = 0.5\Delta R1_{t+i-1|t} = 0.5^2\Delta R1_{t+i-2|t} = \ldots\ldots = 0.5^i\Delta R1_t.$$

The i-period ahead forecast of $R1_t$ is

$$\begin{aligned} R1_{t+i|t} &= R1_{t+i-1|t} + \Delta R1_{t+i|t} \\ &= R1_{t+i-2|t} + \Delta R1_{t+i-1|t} + \Delta R1_{t+i|t} \\ &= \ldots\ldots \\ &= R1_t + \Delta R1_{t+1|t} + \cdots + \Delta R1_{t+i|t} \\ &= R1_t + \left(0.5 + \cdots + 0.5^i\right)\Delta R1_t \\ &= R1_t + \frac{0.5(1-0.5^i)}{1-0.5}\Delta R1_t. \end{aligned}$$

Thus

$$\begin{aligned} Rk_t &= \frac{1}{k}\sum_{i=1}^{k} R1_{t+i|t} + e_t = \frac{1}{k}\sum_{i=1}^{k}\left[R1_t + (1-0.5^i)\Delta R1_t\right] + e_t \\ &= R1_t + \phi\Delta R1_t + e_t. \end{aligned}$$

where $\phi = \frac{1}{k}\sum_{i=1}^{k}(1-0.5^i)$. Thus $Rk_t - R1_t = \phi\Delta R1_t + e_t$. Thus Rk_t and $R1_t$ are cointegrated. The cointegrating coefficient is 1.

(d) When $R1_t = 0.5R1_{t-1} + u_t$, $R1_t$ is stationary and does not have a stochastic trend. $R1_{t+i|t} = 0.5^iR1_t$, so that $Rk_t = \theta R1_t + e_t$, where $\theta = \frac{1}{k}\sum_{i=1}^{k}0.5^i$. Since $R1_t$ and e_t are $I(0)$, then Rk_t is $I(0)$.

14.3. u_t follows the ARCH process with mean $E(u_t) = 0$ and variance $\sigma_t^2 = 1.0 + 0.5u_{t-1}^2$.

(a) For the specified ARCH process, u_t has the conditional mean $E(u_t|u_{t-1}) = 0$ and the conditional variance

$$\text{var}(u_t|u_{t-1}) = \sigma_t^2 = 1.0 + 0.5u_{t-1}^2.$$

The unconditional mean of u_t is $E(u_t) = 0$, and the unconditional variance of u_t is

$$\begin{aligned} \text{var}(u_t) &= \text{var}\left[E(u_t|u_{t-1})\right] + E\left[\text{var}(u_t|u_{t-1})\right] \\ &= 0 + 1.0 + 0.5E\left(u_{t-1}^2\right) \\ &= 1.0 + 0.5\text{var}(u_{t-1}). \end{aligned}$$

The last equation has used the fact that $E\left(u_t^2\right) = \text{var}(u_t) + [E(u_t)]^2 = \text{var}(u_t)$ since $E(u_t) = 0$. Because of the stationarity, we have $\text{var}(u_{t-1}) = \text{var}(u_t)$. Thus, $\text{var}(u_t) = 1.0 + 0.5\text{var}(u_t)$ which implies $\text{var}(u_t) = \frac{1.0}{0.5} = 2$.

(b) When $u_{t-1} = 0.2$, $\sigma_t^2 = 1.0 + 0.5 \times 0.2^2 = 1.02$. The standard deviation of u_t is $\sigma_t = 1.01$. Thus

$$\begin{aligned}
\Pr(-3 \le u_t \le 3) &= \Pr\left(\frac{-3}{1.01} \le \frac{u_t}{\sigma_t} \le \frac{3}{1.01}\right) \\
&= \Phi(2.9703) - \Phi(-2.9703) = 0.9985 - 0.0015 = 0.9970.
\end{aligned}$$

When $u_{t-1} = 2.0$, $\sigma_t^2 = 1.0 + 0.5 \times 2.0^2 = 3.0$. The standard deviation of u_t is $\sigma_t = 1.732$. Thus

$$\begin{aligned}
\Pr(-3 \le u_t \le 3) &= \Pr\left(\frac{-3}{1.732} \le \frac{u_t}{\sigma_t} \le \frac{3}{1.732}\right) \\
&= \Phi(1.732) - \Phi(-1.732) = 0.9584 - 0.0416 = 0.9168.
\end{aligned}$$

14.4. Y_t follows an AR(p) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + \ldots + \beta_p Y_{t-p} + u_t$. $E(u_t | Y_{t-1}, Y_{t-2}, \ldots) = 0$ implies $E(u_{t+h} | Y_t, Y_{t-1}, \ldots) = 0$ for $h \ge 1$. The h-period ahead forecast of Y_t is

$$\begin{aligned}
Y_{t+h|t} &= E(Y_{t+h} | Y_t, Y_{t-1}, \ldots) \\
&= E\left(\beta_0 + \beta_1 Y_{t+h-1} + \ldots + \beta_p Y_{t+h-p} + u_{t+h} | Y_t, Y_{t-1}, \ldots\right) \\
&= \beta_0 + \beta_1 E(Y_{t+h-1} | Y_t, Y_{t-1}, \ldots) + \ldots \\
&\quad + \beta_p E(Y_{t+h-p} | Y_t, Y_{t-1}, \ldots) + E(u_{t+h} | Y_t, Y_{t-1}, \ldots) \\
&= \beta_0 + \beta_1 Y_{t+h-1|t} + \ldots + \beta_p Y_{t+h-p|t}.
\end{aligned}$$

14.5. Because $Y_t = Y_t - Y_{t-1} + Y_{t-1} = Y_{t-1} + \Delta Y_t$,

$$\sum_{t=1}^{T} Y_t^2 = \sum_{t=1}^{T} (Y_{t-1} + \Delta Y_t)^2 = \sum_{t=1}^{T} Y_{t-1}^2 + \sum_{t=1}^{T} (\Delta Y_t)^2 + 2\sum_{t=1}^{T} Y_{t-1}\Delta Y_t.$$

So

$$\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t = \frac{1}{T} \times \frac{1}{2}\left[\sum_{t=1}^{T} Y_t^2 - \sum_{t=1}^{T} Y_{t-1}^2 - \sum_{t=1}^{T} (\Delta Y_t)^2\right].$$

Note that $\sum_{t=1}^T Y_t^2 - \sum_{t=1}^T Y_{t-1}^2 = \left(\sum_{t=1}^{T-1} Y_t^2 + Y_T^2\right) - \left(Y_0^2 + \sum_{t=1}^{T-1} Y_t^2\right) = Y_T^2 - Y_0^2 = Y_T^2$ because $Y_0 = 0$. Thus:

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} Y_{t-1}\Delta Y_t &= \frac{1}{T} \times \frac{1}{2}\left[Y_T^2 - \sum_{t=1}^{T} (\Delta Y_t)^2\right] \\
&= \frac{1}{2}\left[\left(\frac{Y_T}{\sqrt{T}}\right)^2 - \frac{1}{T}\sum_{t=1}^{T} (\Delta Y_t)^2\right].
\end{aligned}
$$

Chapter 15
The Theory of Linear Regression with One Regressor

15.1. (a) Suppose there are n observations. Let b_1 be an arbitrary estimator of β_1. Given the estimator b_1, the sum of squared errors for the given regression model is

$$\sum_{i=1}^{n} (Y_i - b_1 X_i)^2 .$$

$\hat{\beta}_1^{RLS}$, the restricted least squares estimator of β_1, minimizes the sum of squared errors. That is, $\hat{\beta}_1^{RLS}$ satisfies the first order condition for the minimization which requires the differential of the sum of squared errors with respect to b_1 equals zero:

$$\sum_{i=1}^{n} 2 (Y_i - b_1 X_i)(-X_i) = 0.$$

Solving for b_1 from the first order condition leads to the restricted least squares estimator

$$\hat{\beta}_1^{RLS} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2}.$$

(b) We show first that $\hat{\beta}_1^{RLS}$ is unbiased. We can represent the restricted least squares estimator $\hat{\beta}_1^{RLS}$ in terms of the regressors and errors:

$$\hat{\beta}_1^{RLS} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{\sum_{i=1}^{n} X_i (\beta_1 X_i + u_i)}{\sum_{i=1}^{n} X_i^2} = \beta_1 + \frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}.$$

Thus

$$E\left(\hat{\beta}_1^{RLS}\right) = \beta_1 + E\left(\frac{\sum_{i=1}^{n} X_i u_i}{\sum_{i=1}^{n} X_i^2}\right) = \beta_1 + E\left[\frac{\sum_{i=1}^{n} X_i E(u_i|X_1, \ldots, X_n)}{\sum_{i=1}^{n} X_i^2}\right] = \beta_1,$$

where the second equality follows by using the law of iterated expectations, and the third equality follows from

$$\frac{\sum_{i=1}^{n} X_i E(u_i|X_1, \ldots, X_n)}{\sum_{i=1}^{n} X_i^2} = 0$$

because the observations are i.i.d. and $E(u_i|X_i) = 0$. (Note, $E(u_i|X_1, ..., X_n) = E(u_i|X_i)$ because the observations are i.i.d.)

Under assumptions 1-3 of Key Concept 15.1, $\hat{\beta}_1^{RLS}$ is asymptotically normally distributed. The large sample normal approximation to the limiting distribution of $\hat{\beta}_1^{RLS}$ follows from considering

$$\hat{\beta}_1^{RLS} - \beta_1 = \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2} = \frac{\frac{1}{n}\sum_{i=1}^n X_i u_i}{\frac{1}{n}\sum_{i=1}^n X_i^2}.$$

Let's consider first the numerator which is the sample average of $v_i = X_i u_i$. By assumption 1 of Key Concept 15.1, v_i has mean zero: $E(X_i u_i) = E[X_i E(u_i|X_i)] = 0$. By assumption 2, v_i is i.i.d. By assumption 3, var(v_i) is finite. Let $\bar{v} = \frac{1}{n}\sum_{i=1}^n X_i u_i$, then $\sigma_{\bar{v}}^2 = \sigma_v^2/n$. Using the central limit theorem, we have the sample average

$$\bar{v}/\sigma_{\bar{v}} = \frac{1}{\sigma_v \sqrt{n}} \sum_{i=1}^n v_i \xrightarrow{d} N(0, 1)$$

or

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i u_i \xrightarrow{d} N\left(0, \sigma_v^2\right).$$

For the denominator, we have X_i^2 is i.i.d. with finite second variance (because X has a finite fourth moment), so that by the law of large numbers

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{p} E(X^2).$$

Combining the results on the numerator and the denominator and applying Slutsky's theorem lead to

$$\sqrt{n}(\hat{\beta}_1^{RLS} - \beta_u) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^n X_i u_i}{\frac{1}{n}\sum_{i=1}^n X_i^2} \xrightarrow{d} N\left(0, \frac{\text{var}(X_i u_i)}{E(X^2)}\right).$$

(c) $\hat{\beta}_1^{RLS}$ is a linear estimator:

$$\hat{\beta}_1^{RLS} = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} = \sum_{i=1}^n a_i Y_i, \qquad \text{where } a_i = \frac{X_i}{\sum_{i=1}^n X_i^2}.$$

The weight a_i ($i = 1, \ldots, n$) depends on $X_1, \ldots, X_n$ but not on $Y_1, \ldots, Y_n$.

We have shown

$$\hat{\beta}_1^{RLS} = \beta_1 + \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2}.$$

$\hat{\beta}_1^{RLS}$ is conditionally unbiased because

$$\begin{aligned} E\left(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n\right) &= E\left(\beta_1+\frac{\sum_{i=1}^n X_iu_i}{\sum_{i=1}^n X_i^2}|X_1,\ldots,X_n\right) \\ &= \beta_1+E\left(\frac{\sum_{i=1}^n X_iu_i}{\sum_{i=1}^n X_i^2}|X_1,\ldots,X_n\right) \\ &= \beta_1. \end{aligned}$$

The final equality used the fact that

$$E\left(\frac{\sum_{i=1}^n X_iu_i}{\sum_{i=1}^n X_i^2}|X_1,\ldots,X_n\right)=\frac{\sum_{i=1}^n X_iE\left(u_i|X_1,\ldots,X_n\right)}{\sum_{i=1}^n X_i^2}=0$$

because the observations are i.i.d. and $E\left(u_i|X_i\right)=0$.

(d) The conditional variance of $\hat{\beta}_1^{RLS}$, given $X_1,\ldots,X_n$, is

$$\begin{aligned} \text{var}\left(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n\right) &= \text{var}\left(\beta_1+\frac{\sum_{i=1}^n X_iu_i}{\sum_{i=1}^n X_i^2}|X_1,\ldots,X_n\right) \\ &= \frac{\sum_{i=1}^n X_i^2\text{var}\left(u_i|X_1,\ldots,X_n\right)}{\left(\sum_{i=1}^n X_i^2\right)^2} \\ &= \frac{\sum_{i=1}^n X_i^2\sigma_u^2}{\left(\sum_{i=1}^n X_i^2\right)^2} \\ &= \frac{\sigma_u^2}{\sum_{i=1}^n X_i^2}. \end{aligned}$$

(e) The conditional variance of the OLS estimator $\hat{\beta}_1$ is

$$\text{var}\left(\hat{\beta}_1|X_1,\ldots,X_n\right)=\frac{\sigma_u^2}{\sum_{i=1}^n\left(X_i-\bar{X}\right)^2}.$$

Since

$$\sum_{i=1}^n\left(X_i-\bar{X}\right)^2=\sum_{i=1}^n X_i^2-2\bar{X}\sum_{i=1}^n X_i+n\bar{X}^2=\sum_{i=1}^n X_i^2-n\bar{X}^2<\sum_{i=1}^n X_i^2,$$

the OLS estimator has a larger conditional variance: $\text{var}\left(\hat{\beta}_1|X_1,\ldots,X_n\right)>\text{var}\left(\hat{\beta}_1^{RLS}|X_1,\ldots,X_n\right)$. The restricted least squares estimator $\hat{\beta}_1^{RLS}$ is more efficient.

(f) Under assumption 5 of Key Concept 15.1, we know that, conditional on $X_1,\ldots,X_n$, $\hat{\beta}_1^{RLS}$ is normally distributed since it is a weighted average of normally distributed variables u_i:

$$\hat{\beta}_1^{RLS}=\beta_1+\frac{\sum_{i=1}^n X_iu_i}{\sum_{i=1}^n X_i^2}.$$

Using the conditional mean and conditional variance of $\hat{\beta}_1^{RLS}$ derived in parts (c) and (d) respectively, we have the sampling distribution of $\hat{\beta}_1^{RLS}$, conditional on $X_1, \ldots, X_n$, is

$$\hat{\beta}_1^{RLS} \sim N\left(\beta_1, \frac{\sigma_u^2}{\sum_{i=1}^n X_i^2}\right).$$

(g) The estimator

$$\widetilde{\beta}_1 = \frac{\sum_{i=1}^n Y_i}{\sum_{i=1}^n X_i} = \frac{\sum_{i=1}^n (\beta_1 X_i + u_i)}{\sum_{i=1}^n X_i} = \beta_1 + \frac{\sum_{i=1}^n u_i}{\sum_{i=1}^n X_i}.$$

The conditional variance is

$$\begin{aligned}
\operatorname{var}\left(\widetilde{\beta}_1 | X_1, \ldots, X_n\right) &= \operatorname{var}\left(\beta_1 + \frac{\sum_{i=1}^n u_i}{\sum_{i=1}^n X_i} | X_1, \ldots, X_n\right) \\
&= \frac{\sum_{i=1}^n \operatorname{var}(u_i | X_1, \ldots, X_n)}{\left(\sum_{i=1}^n X_i\right)^2} \\
&= \frac{n\sigma_u^2}{\left(\sum_{i=1}^n X_i\right)^2}.
\end{aligned}$$

The difference in the conditional variance of $\widetilde{\beta}_1$ and $\hat{\beta}_1^{RLS}$ is

$$\operatorname{var}\left(\widetilde{\beta}_1 | X_1, \ldots, X_n\right) - \operatorname{var}\left(\hat{\beta}_1^{RLS} | X_1, \ldots, X_n\right) = \frac{n\sigma_u^2}{\left(\sum_{i=1}^n X_i\right)^2} - \frac{\sigma_u^2}{\sum_{i=1}^n X_i^2}.$$

In order to prove $\operatorname{var}\left(\widetilde{\beta}_1 | X_1, \ldots, X_n\right) \geq \operatorname{var}\left(\hat{\beta}_1^{RLS} | X_1, \ldots, X_n\right)$, we need to show

$$\frac{n}{\left(\sum_{i=1}^n X_i\right)^2} \geq \frac{1}{\sum_{i=1}^n X_i^2}$$

or equivalently

$$n \sum_{i=1}^n X_i^2 \geq \left(\sum_{i=1}^n X_i\right)^2.$$

This inequality comes directly by applying the Cauchy-Schwartz inequality

$$\left[\sum_{i=1}^n (a_i \cdot b_i)\right]^2 \leq \sum_{i=1}^n a_i^2 \cdot \sum_{i=1}^n b_i^2$$

which implies

$$\left(\sum_{i=1}^{n} X_i\right)^2 = \left(\sum_{i=1}^{n} 1 \cdot X_i\right)^2 \leq \sum_{i=1}^{n} 1^2 \cdot \sum_{i=1}^{n} X_i^2 = n \sum_{i=1}^{n} X_i^2.$$

That is $n \sum_{i=1}^{n} X_i^2 \geq \left(\sum_{i=1}^{n} X_i\right)^2$, or $\mathrm{var}\left(\widetilde{\beta}_1 | X_1, \ldots, X_n\right) \geq \mathrm{var}\left(\hat{\beta}_1^{RLS} | X_1, \ldots, X_n\right)$.

Note: because $\widetilde{\beta}_1$ is linear and conditionally unbiased, the result $\mathrm{var}\left(\widetilde{\beta}_1 | X_1, \ldots, X_n\right) \geq \mathrm{var}\left(\hat{\beta}_1^{RLS} | X_1, \ldots, X_n\right)$ follows directly from the Gauss-Markov theorem.

15.2. The sample covariance is

$$\begin{aligned}
s_{XY} &= \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left\{\left[(X_i - \mu_X) - \left(\bar{X} - \mu_X\right)\right]\left[(Y_i - \mu_Y) - \left(\bar{Y} - \mu_Y\right)\right]\right\} \\
&= \frac{1}{n-1} \left\{\sum_{i=1}^{n} (X_i - \mu_X)(Y_i - \mu_Y) - \sum_{i=1}^{n} \left(\bar{X} - \mu_X\right)(Y_i - \mu_Y)\right. \\
&\qquad \left. - \sum_{i=1}^{n} (X_i - \mu_X)\left(\bar{Y} - \mu_Y\right) + \sum_{i=1}^{n} \left(\bar{X} - \mu_X\right)\left(\bar{Y} - \mu_Y\right)\right\} \\
&= \frac{n}{n-1} \left[\frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_X)(Y_i - \mu_Y)\right] - \frac{n}{n-1} \left(\bar{X} - \mu_X\right)\left(\bar{Y} - \mu_Y\right)
\end{aligned}$$

where the final equality follows from the definition of $\bar{X}$ and $\bar{Y}$ which implies that $\sum_{i=1}^{n} (X_i - \mu_X) = n\left(\bar{X} - \mu_X\right)$ and $\sum_{i=1}^{n} (Y_i - \mu_Y) = n\left(\bar{Y} - \mu_Y\right)$, and by collecting terms.

We apply the law of large numbers on s_{XY} to check its convergence in probability. It is easy to see the second term converges in probability to zero because $\bar{X} \xrightarrow{p} \mu_X$ and $\bar{Y} \xrightarrow{p} \mu_Y$ so $\left(\bar{X} - \mu_X\right)\left(\bar{Y} - \mu_Y\right) \xrightarrow{p} 0$ by Slutsky's theorem. Let's look at the first term. Since (X_i, Y_i) are i.i.d., the random sequence $(X_i - \mu_X)(Y_i - \mu_Y)$ are i.i.d. By the definition of covariance, we have $E\left[(X_i - \mu_X)(Y_i - \mu_Y)\right] = \sigma_{XY}$. To apply the law of large numbers on the first term, we need to have

$$\mathrm{var}\left[(X_i - \mu_X)(Y_i - \mu_Y)\right] < \infty$$

which is satisfied since

$$\begin{aligned}
\mathrm{var}\left[(X_i - \mu_X)(Y_i - \mu_Y)\right] &< E\left[(X_i - \mu_X)^2 (Y_i - \mu_Y)^2\right] \\
&\leq \sqrt{E\left[(X_i - \mu_X)^4\right] E\left[(Y_i - \mu_Y)^4\right]} < \infty.
\end{aligned}$$

The second inequality follows by applying the Cauchy-Schwartz inequality, and the third inequality follows because of the finite fourth moments for (X_i, Y_i). Applying the law of large numbers, we have

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_X)(Y_i-\mu_Y) \xrightarrow{p} E[(X_i-\mu_X)(Y_i-\mu_Y)] = \sigma_{XY}.$$

Also, $\frac{n}{n-1} \longrightarrow 1$, so the first term for s_{XY} converges in probability to σ_{XY}. Combining results on the two terms for s_{XY}, we have $s_{XY} \xrightarrow{p} \sigma_{XY}$.

15.3. (a) Using Equation (15.19), we have

$$\begin{aligned}
\sqrt{n}\left(\hat{\beta}_1-\beta_1\right) &= \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)u_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \\
&= \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\left[(X_i-\mu_X)-\left(\bar{X}-\mu_X\right)\right]u_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \\
&= \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}(X_i-\mu_X)u_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} - \frac{\left(\bar{X}-\mu_X\right)\sqrt{\frac{1}{n}}\sum_{i=1}^{n}u_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \\
&= \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}v_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} - \frac{\left(\bar{X}-\mu_X\right)\sqrt{\frac{1}{n}}\sum_{i=1}^{n}u_i}{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}
\end{aligned}$$

by defining $v_i = (X_i-\mu_X)u_i$.

(b) The random variables $u_1, \ldots, u_n$ are i.i.d. with mean $\mu_u = 0$ and variance $0 < \sigma_u^2 < \infty$. By the central limit theorem,

$$\frac{\sqrt{n}(\bar{u}-\mu_u)}{\sigma_u} = \frac{\sqrt{\frac{1}{n}}\sum_{i=1}^{n}u_i}{\sigma_u} \xrightarrow{d} N(0,\,1).$$

The law of large numbers implies $\bar{X} \xrightarrow{p} \mu_X$, or $\bar{X}-\mu_X \xrightarrow{p} 0$. By the consistency of sample variance, $\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ converges in probability to population variance, var(X_i), which is finite and non-zero. The result then follows from Slutsky's theorem.

(c) The random variable $v_i = (X_i-\mu_X)u_i$ has finite variance:

$$\begin{aligned}
\text{var}(v_i) &= \text{var}[(X_i-\mu_X)u_i] \\
&\leq E\left[(X_i-\mu_X)^2 u_i^2\right] \\
&\leq \sqrt{E\left[(X_i-\mu_X)^4\right]E\left[(u_i)^4\right]} < \infty.
\end{aligned}$$

The inequality follows by applying the Cauchy-Schwartz inequality, and the second inequality follows because of the finite fourth moments for (X_i, u_i). The finite variance along with the fact that v_i has mean zero (by assumption 1 of Key Concept 15.1) and v_i is i.i.d. (by assumption 2) implies that the sample average $\bar{v}$ satisfies the requirements of the central limit theorem. Thus,

$$\frac{\bar{v}}{\sigma_{\bar{v}}} = \frac{\sqrt{\frac{1}{n}} \sum_{i=1}^{n} v_i}{\sigma_v}$$

satisfies the central limit theorem.

(d) Applying the central limit theorem, we have

$$\frac{\sqrt{\frac{1}{n}} \sum_{i=1}^{n} v_i}{\sigma_v} \xrightarrow{d} N(0, 1).$$

Because the sample variance is a consistent estimator of the population variance, we have

$$\frac{\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{\operatorname{var}(X_i)} \xrightarrow{p} 1.$$

Using Slutsky's theorem,

$$\frac{\frac{\sqrt{\frac{1}{n}} \sum_{i=1}^{n} v_i}{\sigma_v}}{\frac{\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2}{\sigma_X^2}} \xrightarrow{d} N(0, 1),$$

or equivalently

$$\frac{\sqrt{\frac{1}{n}} \sum_{i=1}^{n} v_i}{\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \xrightarrow{d} N\left(0, \frac{\operatorname{var}(v_i)}{[\operatorname{var}(X_i)]^2}\right).$$

Thus

$$\begin{aligned} \sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) &= \frac{\sqrt{\frac{1}{n}} \sum_{i=1}^{n} v_i}{\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} - \frac{\left(\bar{X} - \mu_X\right) \sqrt{\frac{1}{n}} \sum_{i=1}^{n} u_i}{\frac{1}{n} \sum_{i=1}^{n} \left(X_i - \bar{X}\right)^2} \\ &\xrightarrow{d} N\left(0, \frac{\operatorname{var}(v_i)}{[\operatorname{var}(X_i)]^2}\right) \end{aligned}$$

since the the second term for $\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right)$ converges in probability to zero as shown in part (b).

15.4. (a) Write $\left(\hat{\beta}_1 - \beta_1\right) = a_n S_n$ where $a_n = \frac{1}{\sqrt{n}}$ and $S_n = \sqrt{n}\left(\hat{\beta}_1 - \beta_1\right)$. Now, $a_n \longrightarrow 0$ and $S_n \xrightarrow{d} S$ where S is distributed $N(0, a^2)$. By Slutsky's theorem $a_n S_n \xrightarrow{d} 0 \times S$. Thus $\Pr(|\hat{\beta}_1 - \beta_1| > \delta) \longrightarrow 0$ for any $\delta > 0$, so that $\hat{\beta}_1 - \beta_1 \xrightarrow{p} 0$ and $\hat{\beta}_1$ is consistent.

(b) We have (i) $\frac{s_u^2}{\sigma_u^2} \xrightarrow{p} 1$ and (ii) $g(x) = \sqrt{x}$ is a continuous function; thus from the continuous mapping theorem

$$\sqrt{\frac{s_u^2}{\sigma_u^2}} = \frac{s_u}{\sigma_u} \xrightarrow{p} 1.$$

15.5. Because $E\left(W^4\right) = \left[E\left(W^2\right)\right]^2 + \text{var}\left(W^2\right)$, $\left[E\left(W^2\right)\right]^2 \leq E\left(W^4\right) < \infty$. Thus $E\left(W^2\right) < \infty$.

15.6. Using the law of iterated expectations, we have

$$E\left(\hat{\beta}_1\right) = E\left[E\left(\hat{\beta}_1 | X_1, \ldots, X_n\right)\right] = E\left(\beta_1\right) = \beta_1.$$

15.7. (a) The joint probability distribution function of u_i, u_j, X_i, X_j is $f\left(u_i, u_j, X_i, X_j\right)$. The conditional probability distribution function of u_i and X_i given u_j and X_j is $f\left(u_i, X_i | u_j, X_j\right)$. Since u_i, X_i, $i = 1, \ldots, n$ are i.i.d., $f\left(u_i, X_i | u_j, X_j\right) = f\left(u_i, X_i\right)$. By definition of the conditional probability distribution function, we have

$$\begin{aligned} f\left(u_i, u_j, X_i, X_j\right) &= f\left(u_i, X_i | u_j, X_j\right) f\left(u_j, X_j\right) \\ &= f\left(u_i, X_i\right) f\left(u_j, X_j\right). \end{aligned}$$

(b) The conditional probability distribution function of u_i and u_j given X_i and X_j equals

$$f\left(u_i, u_j | X_i, X_j\right) = \frac{f\left(u_i, u_j, X_i, X_j\right)}{f\left(X_i, X_j\right)} = \frac{f\left(u_i, X_i\right) f\left(u_j, X_j\right)}{f\left(X_i\right) f\left(X_j\right)} = f\left(u_i | X_i\right) f\left(u_j | X_j\right).$$

The first and third equalities used the definition of the conditional probability distribution function. The second equality used the conclusion the from part (a) and the independence between X_i and X_j. Substituting

$$f\left(u_i, u_j | X_i, X_j\right) = f\left(u_i | X_i\right) f\left(u_j | X_j\right)$$

into the definition of the conditional expectation, we have

$$\begin{aligned}
E\left(u_i u_j | X_i, X_j\right) &= \int\int u_i u_j f\left(u_i, u_j | X_i, X_j\right) du_i du_j \\
&= \int\int u_i u_j f\left(u_i | X_i\right) f\left(u_j | X_j\right) du_i du_j \\
&= \int u_i f\left(u_i | X_i\right) du_i \int u_j f\left(u_j | X_j\right) du_j \\
&= E\left(u_i | X_i\right) E\left(u_j | X_j\right).
\end{aligned}$$

(c) Let $Q = (X_1, X_2, ..., X_{i-1}, X_{i+1}, ..., X_n)$, so that $f(u_i|X_1, ..., X_n) = f(u_i|X_i, Q)$. Write

$$\begin{aligned}
f(u_i|X_i, Q) &= \frac{f(u_i, X_i, Q)}{f(X_i, Q)} \\
&= \frac{f(u_i, X_i) f(Q)}{f(X_i) f(Q)} \\
&= \frac{f(u_i, X_i)}{f(X_i)} \\
&= f(u_i|X_i)
\end{aligned}$$

where the first equality uses the definition of the conditional density, the second uses the fact that (u_i, X_i) and Q are independent, and the final equality uses the definition of the conditional density. The result then follows directly.

(d) An argument like that used in (c) implies

$$f\left(u_i u_j | X_1, \ldots X_n\right) = f\left(u_i u_j | X_i, X_j\right)$$

and the result then follows from part (b).

15.8. (a) Because the errors are heteroskedastic, the Gauss-Markov theorem does not apply. The OLS estimator of β_1 is not BLUE.

(b) We obtain the BLUE estimator of β_1 from OLS in the following

$$\widetilde{Y}_i = \beta_0 \widetilde{X}_{0i} + \beta_1 \widetilde{X}_{1i} + \widetilde{u}_i$$

where

$$\begin{aligned}
\widetilde{Y}_i &= \frac{Y_i}{\sqrt{\theta_0 + \theta_1 |X_i|}}, \widetilde{X}_{0i} = \frac{1}{\sqrt{\theta_0 + \theta_1 |X_i|}} \\
\widetilde{X}_{1i} &= \frac{X_i}{\sqrt{\theta_0 + \theta_1 |X_i|}}, \text{ and } \widetilde{u} = \frac{u_i}{\sqrt{\theta_0 + \theta_1 |X_i|}}.
\end{aligned}$$

(c) Using equations (15.2) and (15.19), we know the OLS estimator, $\hat{\beta}_1$, is

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} = \beta_1 + \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}.$$

As a weighted average of normally distributed variables u_i, $\hat{\beta}_1$ is normally distributed with mean $E\left(\hat{\beta}_1\right) = \beta_1$. The conditional variance of $\hat{\beta}_1$, given $X_1, \ldots, X_n$, is

$$\begin{aligned}
\operatorname{var}\left(\hat{\beta}_1 | X_1, \ldots, X_n\right) &= \operatorname{var}\left(\beta_1 + \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)u_i}{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2} | X_1, \ldots, X_n\right) \\
&= \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \operatorname{var}\left(u_i | X_1, \ldots, X_n\right)}{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2} \\
&= \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \operatorname{var}\left(u_i | X_i\right)}{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2} \\
&= \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \left(\theta_0 + \theta_1 \left|X_i\right|\right)}{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2}.
\end{aligned}$$

Thus the exact sampling distribution of the OLS estimator, $\hat{\beta}_1$, conditional on $X_1, \ldots, X_n$, is

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 \left(\theta_0 + \theta_1 \left|X_i\right|\right)}{\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right]^2}\right).$$

(d) The weighted least squares (WLS) estimators, $\hat{\beta}_0^{WLS}$ and $\hat{\beta}_1^{WLS}$, are solutions to

$$\min_{b_0, b_1} \sum_{i=1}^{n}\left(\widetilde{Y}_i - b_0 \widetilde{X}_{0i} - b_1 \widetilde{X}_{1i}\right)^2,$$

the minimization of the sum of squared errors of the weighted regression. The first order conditions of the minimization with respect to b_0 and b_1 are

$$\begin{aligned}
\sum_{i=1}^{n} 2\left(\widetilde{Y}_i - b_0 \widetilde{X}_{0i} - b_1 \widetilde{X}_{1i}\right)\left(-\widetilde{X}_{0i}\right) &= 0, \\
\sum_{i=1}^{n} 2\left(\widetilde{Y}_i - b_0 \widetilde{X}_{0i} - b_1 \widetilde{X}_{1i}\right)\left(-\widetilde{X}_{1i}\right) &= 0.
\end{aligned}$$

Solving for b_1 gives the WLS estimator

$$\hat{\beta}_1^{WLS} = \frac{-Q_{01}S_0 + Q_{00}S_1}{Q_{00}Q_{11} - Q_{01}^2}$$

where $Q_{00} = \sum_{i=1}^n \widetilde{X}_{0i}\widetilde{X}_{0i}$, $Q_{01} = \sum_{i=1}^n \widetilde{X}_{0i}\widetilde{X}_{1i}$, $Q_{11} = \sum_{i=1}^n \widetilde{X}_{1i}\widetilde{X}_{1i}$, $S_0 = \sum_{i=1}^n \widetilde{X}_{0i}\widetilde{Y}_i$, and $S_1 = \sum_{i=1}^n \widetilde{X}_{1i}\widetilde{Y}_i$. Substituting $\widetilde{Y}_i = \beta_0\widetilde{X}_{0i} + \beta_1\widetilde{X}_{1i} + \widetilde{u}_i$ yields

$$\hat{\beta}_1^{WLS} = \beta_1 + \frac{-Q_{01}Z_0 + Q_{00}Z_1}{Q_{00}Q_{11} - Q_{01}^2}$$

where $Z_0 = \sum_{i=1}^n \widetilde{X}_{0i}\widetilde{u}_i$, and $Z_1 = \sum_{i=1}^n \widetilde{X}_{1i}\widetilde{u}_i$ or

$$\hat{\beta}_1^{WLS} - \beta_1 = \frac{\sum_{i=1}^n (Q_{00}\widetilde{X}_{1i} - Q_{01}\widetilde{X}_{0i})\widetilde{u}_i}{Q_{00}Q_{11} - Q_{01}^2}.$$

From this we see that the distribution of $\hat{\beta}_1^{WLS}|X_1, \ldots X_n$ is $N(\beta_1, \sigma^2_{\hat{\beta}_1^{WLS}})$, where

$$\begin{aligned}
\sigma^2_{\hat{\beta}_1^{WLS}} &= \frac{\sigma^2_{\widetilde{u}} \sum_{i=1}^n (Q_{00}\widetilde{X}_{1i} - Q_{01}\widetilde{X}_{0i})^2}{(Q_{00}Q_{11} - Q_{01}^2)^2} \\
&= \frac{Q_{00}^2 Q_{11} + Q_{01}^2 Q_{00} - 2Q_{00}Q_{01}^2}{(Q_{00}Q_{11} - Q_{01}^2)^2} \\
&= \frac{Q_{00}}{Q_{00}Q_{11} - Q_{01}^2}
\end{aligned}$$

where the first equality uses the fact that the observations are independent, the second uses $\sigma^2_{\widetilde{u}} = 1$, the definition of Q_{00}, Q_{11}, and Q_{01}, and the third is an algebraic simplification.

15.9. We need to prove

$$\frac{1}{n}\sum_{i=1}^n \left[\left(X_i - \bar{X}\right)^2 \hat{u}_i^2 - \left(X_i - \mu_X\right)^2 u_i^2\right] \xrightarrow{p} 0.$$

Using the identity $\bar{X} = \mu_X + (\bar{X} - \mu_X)$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \left[\left(X_i - \bar{X}\right)^2 \hat{u}_i^2 - \left(X_i - \mu_X\right)^2 u_i^2\right] &= (\bar{X} - \mu_X)^2 \frac{1}{n}\sum_{i=1}^n \widehat{u}_i^2 \\
&\quad -2(\bar{X} - \mu_X)\frac{1}{n}\sum_{i=1}^n (X_i - \mu_X)\widehat{u}_i^2 \\
&\quad +\frac{1}{n}\sum_{i=1}^n (X_i - \mu_X)^2(\widehat{u}_i^2 - u_i^2).
\end{aligned}$$

The definition of $\widehat{u}_i$ implies

$$\begin{aligned}\widehat{u}_i^2 &= u_i^2 + (\widehat{\beta}_0 - \beta_0)^2 + (\widehat{\beta}_1 - \beta_1)^2 X_i^2 - 2u_i(\widehat{\beta}_0 - \beta_0) \\ &\quad -2u_i(\widehat{\beta}_1 - \beta_1)X_i + 2(\widehat{\beta}_0 - \beta_0)(\widehat{\beta}_1 - \beta_1)X_i.\end{aligned}$$

Substituting this into the expression for $\frac{1}{n}\sum_{i=1}^{n}\left[\left(X_i - \bar{X}\right)^2 \hat{u}_i^2 - \left(X_i - \mu_X\right)^2 u_i^2\right]$ yields a series of terms each of which can be written as $a_n b_n$ where $a_n \xrightarrow{p} 0$ and $b_n = \frac{1}{n}\sum_{i=1}^{n} X_i^r u_i^s$ where r and s are integers. For example, $a_n = (\bar{X} - \mu_X)$, $a_n = (\widehat{\beta}_1 - \beta_1)$ and so forth. The result then follows from Slutksy's theorem if $\frac{1}{n}\sum_{i=1}^{n} X_i^r u_i^s \xrightarrow{p} d$ where d is a finite constant. Let $w_i = X_i^r u_i^s$ and note that w_i is i.i.d. The law of large numbers can then be used for the desired result if $E(w_i^2) < \infty$. There are two cases that need to be addressed. In the first, both r and s are non-zero. In this case write

$$E(w_i^2) = E(X_i^{2r} u_i^{2s}) < \sqrt{[E(X_i^{4r})][E(u_i^{4s})]}$$

and this term is finite if r and s are less than 2. Inspection of the terms shows that this is true. In the second case, either $r = 0$ or $s = 0$. In this case the result follows directly if the non-zero exponent (r or s) is less than 4. Inspection of the terms shows that this is true.

15.10 Using (15.48) with $W = \widehat{\theta} - \theta$ implies

$$\Pr(|\widehat{\theta} - \theta| \geq \delta) \leq \frac{E[(\widehat{\theta} - \theta)^2]}{\delta^2}$$

Since $E[(\widehat{\theta} - \theta)^2] \longrightarrow 0$, $\Pr(|\widehat{\theta} - \theta| > \delta) \longrightarrow 0$, so that $\widehat{\theta} - \theta \xrightarrow{p} 0$.

Chapter 16
The Theory of Multiple Regression

16.1. (a) The regression in the matrix form is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$$

with

$$\mathbf{Y} = \begin{pmatrix} TestScore_1 \\ TestScore_2 \\ \vdots \\ TestScore_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & Income_1 & Income_1^2 \\ 1 & Income_2 & Income_2^2 \\ \vdots & \vdots & \vdots \\ 1 & Income_n & Income_n^2 \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

(b) The null hypothesis is

$$\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$$

versus $\mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$ with

$$\mathbf{R} = (0 \quad 0 \quad 1) \qquad \text{and} \qquad \mathbf{r} = 0.$$

The heteroskedasticity-robust F-statistic testing the null hypothesis is

$$F = \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)' \left[\mathbf{R}\hat{\Sigma}_{\hat{\boldsymbol{\beta}}}\mathbf{R}'\right]^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}\right)/q$$

with $q = 1$. Under the null hypothesis,

$$F \xrightarrow{d} F_{q,\infty}.$$

We reject the null hypothesis if the calculated F-statistic is larger than the critical value of the $F_{q,\infty}$ distribution at a given significance level.

16.2. (a) The sample size $n = 20$. We write the regression in the matrix form:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$$

with

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & X_{1,1} & X_{2,1} \\ 1 & X_{1,2} & X_{2,2} \\ \vdots & \vdots & \vdots \\ 1 & X_{1,n} & X_{2,n} \end{pmatrix},$$

$$\mathbf{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

The OLS estimator of the coefficient vector is

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{Y}.$$

with

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{2i} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}^2 & \sum_{i=1}^n X_{1i}X_{2i} \\ \sum_{i=1}^n X_{1i} & \sum_{i=1}^n X_{1i}X_{2i} & \sum_{i=1}^n X_{2i}^2 \end{pmatrix},$$

and

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{1i}Y_i \\ \sum_{i=1}^n X_{2i}Y_i \end{pmatrix}.$$

Note

$$\begin{aligned} \sum_{i=1}^n X_{1i} &= n\bar{X}_1 = 20 \times 7.24 = 144.8, \\ \sum_{i=1}^n X_{2i} &= n\bar{X}_2 = 20 \times 4.00 = 80.0, \\ \sum_{i=1}^n Y_i &= n\bar{Y} = 20 \times 6.39 = 127.8. \end{aligned}$$

By the definition of sample variance

$$s_Y^2 = \frac{1}{n-1} \sum_{i=1}^n \left(Y_i - \bar{Y}\right)^2 = \frac{1}{n-1} \sum_{i=1}^n Y_i^2 - \frac{n}{n-1} \bar{Y}^2,$$

we know

$$\sum_{i=1}^n Y_i^2 = (n-1)\, s_Y^2 + n\bar{Y}^2.$$

Thus using the sample means and sample variances, we can get

$$\begin{aligned}\sum_{i=1}^{n} X_{1i}^2 &= (n-1)\, s_{X_1}^2 + n\bar{X}_1^2 \\ &= (20-1)\times 0.80 + 20\times 7.24^2 = 1063.6,\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{n} X_{2,i}^2 &= (n-1)\, s_{X_2}^2 + n\bar{X}_2^2 \\ &= (20-1)\times 2.40 + 20\times 4.00^2 = 365.6.\end{aligned}$$

By the definition of sample covariance

$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right) = \frac{1}{n-1}\sum_{i=1}^{n} X_iY_i - \frac{n}{n-1}\bar{X}\bar{Y},$$

we know

$$\sum_{i=1}^{n} X_iY_i = (n-1)\, s_{XY} + n\bar{X}\bar{Y}.$$

Thus using the sample means and sample covariances, we can get

$$\begin{aligned}\sum_{i=1}^{n} X_{1i}Y_i &= (n-1)\, s_{X_1Y} + n\bar{X}_1\bar{Y} \\ &= (20-1)\times 0.22 + 20\times 7.24\times 6.39 = 929.45,\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{n} X_{2i}Y_i &= (n-1)\, s_{X_2Y} + n\bar{X}_2\bar{Y} \\ &= (20-1)\times 0.32 + 20\times 4.00\times 6.39 = 517.28,\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^{n} X_{1i}X_{2i} &= (n-1)\, s_{X_1X_2} + n\bar{X}_1\bar{X}_2 \\ &= (20-1)\times 0.28 + 20\times 7.24\times 4.00 = 584.52.\end{aligned}$$

Therefore we have

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} 20 & 144.8 & 80.0 \\ 144.8 & 1063.6 & 584.52 \\ 80.0 & 584.52 & 365.6 \end{pmatrix}, \qquad \mathbf{X}'\mathbf{Y} = \begin{pmatrix} 127.8 \\ 929.45 \\ 517.28 \end{pmatrix}.$$

The inverse of matrix $\mathbf{X}'\mathbf{X}$ is

$$\left(\mathbf{X}'\mathbf{X}\right)^{-1} = \begin{pmatrix} 3.5373 & -0.4631 & -0.0337 \\ -0.4631 & 0.0684 & -0.0080 \\ -0.0337 & -0.0080 & 0.0229 \end{pmatrix}.$$

The OLS estimator of the coefficient vector is

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \\ &= \begin{pmatrix} 3.5373 & -0.4631 & -0.0337 \\ -0.4631 & 0.0684 & -0.0080 \\ -0.0337 & -0.0080 & 0.0229 \end{pmatrix} \begin{pmatrix} 127.8 \\ 929.45 \\ 517.28 \end{pmatrix} = \begin{pmatrix} 4.2063 \\ 0.2520 \\ 0.1033 \end{pmatrix}. \end{aligned}$$

That is, $\hat{\beta}_0 = 4.2063$, $\hat{\beta}_1 = 0.2520$, and $\hat{\beta}_2 = 0.1033$.

With the number of slope coefficients $k = 2$, the squared standard error of the regression $s_{\hat{u}}^2$ is

$$s_{\hat{u}}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n}\hat{u}_i = \frac{1}{n-k-1}\hat{\mathbf{U}}'\hat{\mathbf{U}}.$$

The OLS residuals $\hat{\mathbf{U}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}$, so

$$\hat{\mathbf{U}}'\hat{\mathbf{U}} = \left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)'\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) = \mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}.$$

We have

$$\begin{aligned} \mathbf{Y}'\mathbf{Y} &= \sum_{i=1}^{n} Y_i^2 = (n-1)\,s_Y^2 + n\bar{Y}^2 \\ &= (20-1)\times 0.26 + 20 \times 6.39^2 = 821.58, \end{aligned}$$

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} = \begin{pmatrix} 4.2063 \\ 0.2520 \\ 0.1033 \end{pmatrix}' \begin{pmatrix} 127.8 \\ 929.45 \\ 517.28 \end{pmatrix} = 825.22,$$

and

$$\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \begin{pmatrix} 4.2063 \\ 0.2520 \\ 0.1033 \end{pmatrix}' \begin{pmatrix} 20 & 144.8 & 80.0 \\ 144.8 & 1063.6 & 584.52 \\ 80.0 & 584.52 & 365.6 \end{pmatrix} \begin{pmatrix} 4.2063 \\ 0.2520 \\ 0.1033 \end{pmatrix} = 832.23.$$

Therefore the sum of squared residuals

$$\begin{aligned} SSR &= \sum_{i=1}^{n}\hat{u}_i^2 = \hat{\mathbf{U}}'\hat{\mathbf{U}} = \mathbf{Y}'\mathbf{Y} - 2\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{Y} + \hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} \\ &= 821.58 - 2\times 825.22 + 832.23 = 3.37. \end{aligned}$$

The squared standard error of the regression $s_{\hat{u}}^2$ is

$$s_{\hat{u}}^2 = \frac{1}{n-k-1}\hat{\mathbf{U}}'\hat{\mathbf{U}} = \frac{1}{20-2-1} \times 3.37 = 0.1982.$$

With the total sum of squares

$$TSS = \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right)^2 = (n-1)\, s_Y^2 = (20-1) \times 0.26 = 4.94,$$

the R^2 of the regression is

$$R^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{3.37}{4.94} = 0.3178.$$

(b) When all six assumptions in Key Concept 16.1 hold, we can use the homoskedasticity-only estimator $\tilde{\Sigma}_{\hat{\boldsymbol{\beta}}}$ of the covariance matrix of $\hat{\boldsymbol{\beta}}$, conditional on $\mathbf{X}$, which is

$$\begin{aligned}
\tilde{\Sigma}_{\hat{\boldsymbol{\beta}}} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} s_{\hat{u}}^2 &= \begin{pmatrix} 3.5373 & -0.4631 & -0.0337 \\ -0.4631 & 0.0684 & -0.0080 \\ -0.0337 & -0.0080 & 0.0229 \end{pmatrix} \times 0.1982 \\
&= \begin{pmatrix} 0.7011 & -0.09179 & -0.0067 \\ -0.09179 & 0.0136 & -0.0016 \\ -0.0067 & -0.0016 & 0.0045 \end{pmatrix}.
\end{aligned}$$

The homoskedasticity-only standard error of $\hat{\beta}_1$ is

$$\widetilde{SE}\left(\hat{\beta}_1\right) = 0.0136^{\frac{1}{2}} = 0.1166.$$

The t-statistic testing the hypothesis $\beta_1 = 0$ has a $t_{n-k-1} = t_{17}$ distribution under the null hypothesis. The value of the t-statistic is

$$\tilde{t} = \frac{\hat{\beta}_1}{\widetilde{SE}\left(\hat{\beta}_1\right)} = \frac{0.2520}{0.1166} = 2.1612,$$

and the 5% two-sided critical value (from Appendix 2) is 2.11. Thus we can reject the null hypothesis $\beta_1 = 0$ at the 5% significance level.

16.3. (a)

$$\begin{aligned}
\text{var}(Q) &= E[(Q-\mu_Q)^2] \\
&= E[(Q-\mu_Q)(Q-\mu_Q)'] \\
&= E[\left(\mathbf{c}'\mathbf{W} - \mathbf{c}'\boldsymbol{\mu}_{\mathbf{W}}\right)\left(\mathbf{c}'\mathbf{W} - \mathbf{c}'\boldsymbol{\mu}_{\mathbf{W}}\right)'] \\
&= \mathbf{c}'E[\left(\mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}}\right)\left(\mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}}\right)']\mathbf{c} \\
&= \mathbf{c}'\text{var}(\mathbf{W})\,\mathbf{c} = \mathbf{c}'\Sigma_{\mathbf{W}}\mathbf{c}
\end{aligned}$$

where the second equality uses the fact that Q is a scalar and the third equality uses the fact that $\mu_Q = \mathbf{c}'\boldsymbol{\mu}_{\mathbf{W}}$.

(b) Because the covariance matrix $\Sigma_{\mathbf{W}}$ is positive definite, we have $\mathbf{c}'\Sigma_{\mathbf{W}}\mathbf{c} > 0$ for every nonzero vector from the definition. Thus, $\text{var}(Q) > 0$. Both the vector $\mathbf{c}$ and the matrix $\Sigma_{\mathbf{W}}$ are finite, so $\text{var}(Q) = \mathbf{c}'\Sigma_{\mathbf{W}}\mathbf{c}$ is also finite. Thus, $0 < \text{var}(Q) < \infty$.

16.4. (a) The regression in the matrix form is

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}$$

with

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{pmatrix}, \quad \mathbf{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}.$$

(b) Because $\mathbf{X}_i' = (1 \quad X_i)$, assumptions 1-3 in Key Concept 16.1 follow directly from assumptions 1-3 in Key Concept 4.3. Assumption 4 in Key Concept 16.1 is satisfied since observations X_i $(i = 1, 2, \cdots n)$ are not constant and there is no perfect multicollinearity among the two vectors of the matrix $\mathbf{X}$.

(c) Matrix multiplication of $\mathbf{X}'\mathbf{X}$ and $\mathbf{X}'\mathbf{Y}$ yields

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}.$$

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} = \begin{pmatrix} n\bar{Y} \\ \sum_{i=1}^n X_i Y_i \end{pmatrix}.$$

The inverse of $\mathbf{X}'\mathbf{X}$ is

$$\begin{aligned} (\mathbf{X}'\mathbf{X})^{-1} &= \begin{pmatrix} n & \sum_{i=1}^n X_i \\ \sum_{i=1}^n X_i & \sum_{i=1}^n X_i^2 \end{pmatrix}^{-1} \\ &= \frac{1}{n\sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i\right)^2} \begin{pmatrix} \sum_{i=1}^n X_i^2 & -\sum_{i=1}^n X_i \\ -\sum_{i=1}^n X_i & n \end{pmatrix} \\ &= \frac{1}{\sum_{i=1}^n \left(X_i - \bar{X}\right)^2} \begin{pmatrix} \sum_{i=1}^n X_i^2/n & -\bar{X} \\ -\bar{X} & 1 \end{pmatrix}. \end{aligned}$$

The estimator for the coefficient vector is

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} \\
&= \frac{1}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\left(\begin{array}{cc} \sum_{i=1}^{n}X_i^2/n & -\bar{X} \\ -\bar{X} & 1 \end{array}\right)\left(\begin{array}{c} n\bar{Y} \\ \sum_{i=1}^{n}X_iY_i \end{array}\right) \\
&= \frac{1}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\left(\begin{array}{c} \bar{Y}\sum_{i=1}^{n}X_i^2-\bar{X}\sum_{i=1}^{n}X_iY_i \\ \sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y} \end{array}\right).
\end{aligned}
$$

Therefore we have

$$
\hat{\beta}_1 = \frac{\sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} = \frac{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)\left(Y_i-\bar{Y}\right)}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2},
$$

and

$$
\begin{aligned}
\hat{\beta}_0 &= \frac{\bar{Y}\sum_{i=1}^{n}X_i^2-\bar{X}\sum_{i=1}^{n}X_iY_i}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \\
&= \frac{\bar{Y}\sum_{i=1}^{n}\left(X_i-\bar{X}+\bar{X}\right)^2-\bar{X}\sum_{i=1}^{n}X_iY_i}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \\
&= \frac{\bar{Y}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2+n\bar{X}^2\bar{Y}-\bar{X}\sum_{i=1}^{n}X_iY_i}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2} \\
&= \bar{Y}-\left[\frac{\sum_{i=1}^{n}X_iY_i-n\bar{X}\bar{Y}}{\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2}\right]\bar{X} \\
&= \bar{Y}-\hat{\beta}_1\bar{X}.
\end{aligned}
$$

We get the same expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$ as given in Key Concept 4.2.

(d) The large-sample covariance matrix of $\hat{\boldsymbol{\beta}}$, conditional on $\mathbf{X}$, converges to

$$
\Sigma_{\hat{\boldsymbol{\beta}}} = \frac{1}{n}\mathbf{Q}_{\mathbf{X}}^{-1}\Sigma_V\mathbf{Q}_{\mathbf{X}}^{-1}
$$

with $\mathbf{Q}_{\mathbf{X}} = E\left(\mathbf{X}_i\mathbf{X}_i'\right)$ and $\Sigma_{\mathbf{V}} = E\left(\mathbf{V}_i\mathbf{V}_i'\right) = E\left(\mathbf{X}_iu_iu_i'\mathbf{X}_i\right)$. The column vector $\mathbf{X}_i$ for the ith observation is

$$
\mathbf{X}_i = \left(\begin{array}{c} 1 \\ X_i \end{array}\right),
$$

so we have

$$
\mathbf{X}_i\mathbf{X}_i' = \left(\begin{array}{c} 1 \\ X_i \end{array}\right)\left(\begin{array}{cc} 1 & X_i \end{array}\right) = \left(\begin{array}{cc} 1 & X_i \\ X_i & X_i^2 \end{array}\right),
$$

$$\mathbf{V}_i = \mathbf{X}_i u_i = \begin{pmatrix} u_i \\ X_i u_i \end{pmatrix},$$

and

$$\mathbf{V}_i \mathbf{V}_i' = \begin{pmatrix} u_i \\ X_i u_i \end{pmatrix} \begin{pmatrix} u_i & X_i u_i \end{pmatrix} = \begin{pmatrix} u_i^2 & X_i u_i^2 \\ X_i u_i^2 & X_i^2 u_i^2 \end{pmatrix}.$$

Taking expectations, we get

$$\mathbf{Q}_{\mathbf{X}} = E\left(\mathbf{X}_i \mathbf{X}_i'\right) = \begin{pmatrix} 1 & \mu_X \\ \mu_X & E\left(X_i^2\right) \end{pmatrix},$$

and

$$\begin{aligned} \Sigma_{\mathbf{V}} &= E\left(\mathbf{V}_i \mathbf{V}_i'\right) \\ &= \begin{pmatrix} E\left(u_i^2\right) & E\left(X_i u_i^2\right) \\ E\left(X_i u_i^2\right) & E\left(X_i^2 u_i^2\right) \end{pmatrix} \\ &= \begin{pmatrix} \operatorname{var}\left(u_i\right) & \operatorname{cov}\left(X_i u_i,\, u_i\right) \\ \operatorname{cov}\left(X_i u_i,\, u_i\right) & \operatorname{var}\left(X_i u_i\right) \end{pmatrix}. \end{aligned}$$

In the above equation, the third equality has used the fact that $E\left(u_i|X_i\right) = 0$ so

$$\begin{aligned} E\left(u_i\right) &= E\left[E\left(u_i|X_i\right)\right] = 0, \\ E\left(X_i u_i\right) &= E\left[X_i E\left(u_i|X_i\right)\right] = 0, \\ E\left(u_i^2\right) &= \operatorname{var}\left(u_i\right) + \left[E\left(u_i\right)\right]^2 = \operatorname{var}\left(u_i\right) + \left[E\left(u_i\right)\right]^2 = \operatorname{var}\left(u_i\right), \\ E\left(X_i^2 u_i^2\right) &= \operatorname{var}\left(X_i u_i\right) + \left[E\left(X_i u_i\right)\right]^2 = \operatorname{var}\left(X_i u_i\right), \\ E\left(X_i u_i^2\right) &= \operatorname{cov}\left(X_i u_i,\, u_i\right) + E\left(X_i u_i\right) E\left(u_i\right) = \operatorname{cov}\left(X_i u_i,\, u_i\right). \end{aligned}$$

The inverse of $\mathbf{Q}_{\mathbf{X}}$ is

$$\mathbf{Q}_{\mathbf{X}}^{-1} = \begin{pmatrix} 1 & \mu_X \\ \mu_X & E\left(X_i^2\right) \end{pmatrix}^{-1} = \frac{1}{E\left(X_i^2\right) - \mu_X^2} \begin{pmatrix} E\left(X_i^2\right) & -\mu_X \\ -\mu_X & 1 \end{pmatrix}.$$

We now can calculate the large-sample covariance matrix of $\hat{\boldsymbol{\beta}}$, conditional on $\mathbf{X}$, from

$$\begin{aligned} \Sigma_{\hat{\boldsymbol{\beta}}} &= \frac{1}{n} \mathbf{Q}_{\mathbf{X}}^{-1} \Sigma_V \mathbf{Q}_{\mathbf{X}}^{-1} \\ &= \frac{1}{n\left[E\left(X_i^2\right) - \mu_X^2\right]^2} \\ &\quad \times \begin{pmatrix} E\left(X_i^2\right) & -\mu_X \\ -\mu_X & 1 \end{pmatrix} \begin{pmatrix} \operatorname{var}\left(u_i\right) & \operatorname{cov}\left(X_i u_i,\, u_i\right) \\ \operatorname{cov}\left(X_i u_i,\, u_i\right) & \operatorname{var}\left(X_i u_i\right) \end{pmatrix} \begin{pmatrix} E\left(X_i^2\right) & -\mu_X \\ -\mu_X & 1 \end{pmatrix}. \end{aligned}$$

The $(1,1)$ element of $\Sigma_{\hat{\boldsymbol{\beta}}}$ is

$$\begin{aligned}
& \frac{1}{n\left[E\left(X_i^2\right)-\mu_X^2\right]^2}\left\{\left[E\left(X_i^2\right)\right]^2 \operatorname{var}\left(u_i\right)-2E\left(X_i^2\right)\mu_X \operatorname{cov}\left(X_i u_i,\, u_i\right)+\mu_X^2 \operatorname{var}\left(X_i u_i\right)\right\} \\
= & \frac{1}{n\left[E\left(X_i^2\right)-\mu_X^2\right]^2} \operatorname{var}\left[E\left(X_i^2\right) u_i-\mu_X X_i u_i\right] \\
= & \frac{\left[E\left(X_i^2\right)\right]^2}{n\left[E\left(X_i^2\right)-\mu_X^2\right]^2} \operatorname{var}\left[u_i-\frac{\mu_X}{E\left(X_i^2\right)} X_i u_i\right] \\
= & \frac{1}{n\left[1-\frac{\mu_X^2}{E\left(X_i^2\right)}\right]^2} \operatorname{var}\left[\left(1-\frac{\mu_X}{E\left(X_i^2\right)} X_i\right) u_i\right] \\
= & \frac{\operatorname{var}\left(H_i u_i\right)}{n\left[E\left(H_i^2\right)\right]^2}, \text{ (the same as the expression for } \sigma_{\hat{\beta}_0}^2 \text{ given in Key Concept 4.4)}
\end{aligned}$$

by defining

$$H_i=1-\frac{\mu_X}{E\left(X_i^2\right)} X_i .$$

The denominator in the last equality for the $(1,1)$ element of $\Sigma_{\hat{\boldsymbol{\beta}}}$ has used the facts that

$$H_i^2=\left(1-\frac{\mu_X}{E\left(X_i^2\right)} X_i\right)^2=1+\frac{\mu_X^2}{E^2\left(X_i^2\right)} X_i^2-\frac{2 \mu_X}{E\left(X_i^2\right)} X_i,$$

so

$$E\left(H_i^2\right)=1+\frac{\mu_X^2}{\left[E\left(X_i^2\right)\right]^2} E\left(X_i^2\right)-\frac{2 \mu_X}{E\left(X_i^2\right)} \mu_X=1-\frac{\mu_X^2}{E\left(X_i^2\right)} .$$

16.5. $\mathbf{P_X}=\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'$, $\mathbf{M}_X=\mathbf{I}_n-\mathbf{P_X}$.

(a) $\mathbf{P_X}$ is idempotent because

$$\mathbf{P_X P_X}=\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'=\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'=\mathbf{P_X}.$$

$\mathbf{M_X}$ is idempotent because

$$\begin{aligned}
\mathbf{M_X M_X} & =\left(\mathbf{I}_n-\mathbf{P_X}\right)\left(\mathbf{I}_n-\mathbf{P_X}\right)=\mathbf{I}_n-\mathbf{P_X}-\mathbf{P_X}+\mathbf{P_X P_X} \\
& =\mathbf{I}_n-2\mathbf{P_X}+\mathbf{P_X}=\mathbf{I}_n-\mathbf{P_X}=\mathbf{M_X}.
\end{aligned}$$

$\mathbf{P_X M_X}=\mathbf{0}_{nxn}$ because

$$\mathbf{P_X M_X}=\mathbf{P_X}\left(\mathbf{I}_n-\mathbf{P_X}\right)=\mathbf{P_X}-\mathbf{P_X P_X}=\mathbf{P_X}-\mathbf{P_X}=\mathbf{0}_{n\times n}.$$

(b) Because $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, we have

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{P_X}\mathbf{Y}$$

which is Equation (16.27). The residual vector is

$$\hat{\mathbf{U}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{P_X}\mathbf{Y} = (\mathbf{I}_n - \mathbf{P_X})\mathbf{Y} = \mathbf{M_X}\mathbf{Y}.$$

We know that $\mathbf{M_X}\mathbf{X}$ is orthogonal to the columns of $\mathbf{X}$:

$$\mathbf{M_X}\mathbf{X} = (\mathbf{I}_n - \mathbf{P_X})\mathbf{X} = \mathbf{X} - \mathbf{P_X}\mathbf{X} = \mathbf{X} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} = \mathbf{X} - \mathbf{X} = \mathbf{0},$$

so the residual vector can be further written as

$$\hat{\mathbf{U}} = \mathbf{M_X}\mathbf{Y} = \mathbf{M_X}(\mathbf{X}\boldsymbol{\beta} + \mathbf{U}) = \mathbf{M_X}\mathbf{X}\boldsymbol{\beta} + \mathbf{M_X}\mathbf{U} = \mathbf{M_X}\mathbf{U}$$

which is Equation (16.28).

16.6. The matrix form for Equation (8.14) is

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{U}}$$

with

$$\tilde{\mathbf{Y}} = \begin{pmatrix} Y_{11} - \bar{Y}_1 \\ Y_{12} - \bar{Y}_1 \\ \vdots \\ Y_{1T} - \bar{Y}_1 \\ Y_{21} - \bar{Y}_2 \\ Y_{22} - \bar{Y}_2 \\ \vdots \\ Y_{2T} - \bar{Y}_2 \\ \vdots \\ Y_{n1} - \bar{Y}_n \\ Y_{n2} - \bar{Y}_n \\ \vdots \\ Y_{nT} - \bar{Y}_n \end{pmatrix}, \quad \tilde{\mathbf{X}} = \begin{pmatrix} X_{11} - \bar{X}_1 \\ X_{12} - \bar{X}_1 \\ \vdots \\ X_{1T} - \bar{X}_1 \\ X_{21} - \bar{X}_2 \\ X_{22} - \bar{X}_2 \\ \vdots \\ X_{2T} - \bar{X}_2 \\ \vdots \\ X_{n1} - \bar{X}_n \\ X_{n2} - \bar{X}_n \\ \vdots \\ X_{nT} - \bar{X}_n \end{pmatrix}, \quad \tilde{\mathbf{U}} = \begin{pmatrix} u_{11} - \bar{u}_1 \\ u_{12} - \bar{u}_1 \\ \vdots \\ u_{1T} - \bar{u}_1 \\ u_{21} - \bar{u}_2 \\ u_{22} - \bar{u}_2 \\ \vdots \\ u_{2T} - \bar{u}_2 \\ \vdots \\ u_{n1} - \bar{u}_n \\ u_{n2} - \bar{u}_n \\ \vdots \\ u_{nT} - \bar{u}_n \end{pmatrix},$$

$$\tilde{\boldsymbol{\beta}} = \beta_1$$

The OLS "de-meaning" fixed effects estimator is

$$\hat{\beta}_1^{DM} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}}.$$

Rewrite Equation (8.11) using n fixed effects as

$$Y_{it} = X_{it}\beta_1 + D1_i\gamma_1 + D2_i\gamma_2 + \cdots + Dn_i\gamma_n + u_{it}.$$

In matrix form this is

$$\mathbf{Y}_{nT\times 1} = \mathbf{X}_{nT\times 1}\boldsymbol{\beta}_{1\times 1} + \mathbf{W}_{nT\times n}\boldsymbol{\gamma}_{n\times 1} + \mathbf{U}_{nT\times 1}$$

with the subscripts denoting the size of the matrices. The matrices for variables and coefficients are

$$\mathbf{Y} = \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{1T} \\ Y_{21} \\ Y_{22} \\ \vdots \\ Y_{2T} \\ \vdots \\ Y_{n1} \\ Y_{n2} \\ \vdots \\ Y_{nT} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} X_{11} \\ X_{12} \\ \vdots \\ X_{1T} \\ X_{21} \\ X_{22} \\ \vdots \\ X_{2T} \\ \vdots \\ X_{n1} \\ X_{n2} \\ \vdots \\ X_{nT} \end{pmatrix}, \mathbf{W} = \begin{pmatrix} D1_1 & D2_1 & \cdots & Dn_1 \\ D1_1 & D2_1 & \cdots & Dn_1 \\ \vdots & \vdots & \cdots & \vdots \\ D1_1 & D2_1 & \cdots & Dn_1 \\ D1_2 & D2_2 & \cdots & Dn_2 \\ D1_2 & D2_2 & \cdots & Dn_2 \\ \vdots & \vdots & \cdots & \vdots \\ D1_2 & D2_2 & \cdots & Dn_2 \\ \vdots & \vdots & \vdots & \vdots \\ D1_n & D2_n & \cdots & Dn_n \\ D1_n & D2_n & \cdots & Dn_n \\ \vdots & \vdots & \cdots & \vdots \\ D1_n & D2_n & \cdots & Dn_n \end{pmatrix}, \mathbf{U} = \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T} \\ u_{21} \\ u_{22} \\ \vdots \\ u_{2T} \\ \vdots \\ u_{n1} \\ u_{n2} \\ \vdots \\ u_{nT} \end{pmatrix},$$

$$\boldsymbol{\beta} = \beta_1, \qquad \boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}.$$

Using Equation (16.45), we have the estimator

$$\begin{aligned} \widehat{\beta}_1^{BV} &= \hat{\boldsymbol{\beta}} = \left(\mathbf{X}'\mathbf{M_W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{Y} \\ &= \left((\mathbf{M_W}\mathbf{X})'(\mathbf{M_W}\mathbf{X})\right)^{-1}(\mathbf{M_W}\mathbf{X})'(\mathbf{M_W}\mathbf{Y}). \end{aligned}$$

where the second equality uses the fact that $\mathbf{M_W}$ is idempotent. Using the definition of $\mathbf{W}$,

$$\mathbf{P_W X} = \begin{pmatrix} \overline{X}_1 & 0 & \cdots & 0 \\ \overline{X}_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ \overline{X}_1 & 0 & \cdots & 0 \\ 0 & \overline{X}_2 & \cdots & 0 \\ 0 & \overline{X}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & \overline{X}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \overline{X}_n \\ 0 & 0 & \cdots & \overline{X}_n \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \overline{X}_n \end{pmatrix}$$

and

$$\mathbf{M_W X} = \begin{pmatrix} X_{11} - \overline{X}_1 & 0 & \cdots & 0 \\ X_{12} - \overline{X}_1 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ X_{1T} - \overline{X}_1 & 0 & \cdots & 0 \\ 0 & X_{21} - \overline{X}_2 & \cdots & 0 \\ 0 & X_{22} - \overline{X}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & X_{2T} - \overline{X}_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & X_{n1} - \overline{X}_n \\ 0 & 0 & \cdots & X_{n2} - \overline{X}_n \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & X_{nT} - \overline{X}_n \end{pmatrix}$$

so that $\mathbf{M_W X} = \tilde{\mathbf{X}}$. A similar calculation shows $\mathbf{M_W Y} = \tilde{\mathbf{Y}}$. Thus

$$\widehat{\beta}_1^{BV} = \left(\tilde{\mathbf{X}}'\tilde{\mathbf{X}}\right)^{-1}\tilde{\mathbf{X}}'\tilde{\mathbf{Y}} = \hat{\beta}_1^{DM}.$$

16.7. (a) We write the regression model, $Y_i = \beta_1 X_i + \beta_2 W_i + u_i$, in the matrix form as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{U}$$

with

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix}, \qquad \mathbf{W} = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix}, \qquad \mathbf{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix},$$

$$\boldsymbol{\beta} = \beta_1, \qquad \boldsymbol{\gamma} = \beta_2.$$

The OLS estimator is

$$\begin{aligned}
\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\mathbf{Y} \\ \mathbf{W}'\mathbf{Y} \end{pmatrix} \\
&= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}'\mathbf{U} \\ \mathbf{W}'\mathbf{U} \end{pmatrix} \\
&= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \frac{1}{n}\mathbf{X}'\mathbf{X} & \frac{1}{n}\mathbf{X}'\mathbf{W} \\ \frac{1}{n}\mathbf{W}'\mathbf{X} & \frac{1}{n}\mathbf{W}'\mathbf{W} \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{n}\mathbf{X}'\mathbf{U} \\ \frac{1}{n}\mathbf{W}'\mathbf{U} \end{pmatrix} \\
&= \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i^2 & \frac{1}{n}\sum_{i=1}^n X_i W_i \\ \frac{1}{n}\sum_{i=1}^n W_i X_i & \frac{1}{n}\sum_{i=1}^n W_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i u_i \\ \frac{1}{n}\sum_{i=1}^n W_i u_i \end{pmatrix}
\end{aligned}$$

By the law of large numbers $\frac{1}{n}\sum_{i=1}^n X_i^2 \xrightarrow{p} E(X^2)$; $\frac{1}{n}\sum_{i=1}^n W_i^2 \xrightarrow{p} E(W^2)$; $\frac{1}{n}\sum_{i=1}^n X_i W_i \xrightarrow{p} E(XW) = 0$ (because X and W are independent with means of zero); $\frac{1}{n}\sum_{i=1}^n X_i u_i \xrightarrow{p} E(Xu) = 0$ (because X and u are independent with means of zero); $\frac{1}{n}\sum_{i=1}^n W_i u_i \xrightarrow{p} E(Wu)$. Thus

$$\begin{aligned}
&\begin{pmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{pmatrix} \xrightarrow{p} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} E(X^2) & 0 \\ 0 & E(W^2) \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ E(Wu) \end{pmatrix} \\
= &\begin{pmatrix} \beta_1 \\ \beta_2 + \frac{E(Wu)}{E(W^2)} \end{pmatrix}.
\end{aligned}$$

(b) From the answer to (a) $\widehat{\beta}_2 \xrightarrow{p} \beta_2 + \frac{E(Wu)}{E(W^2)} \neq \beta_2$ if $E(Wu)$ is nonzero.

(c) Consider the population linear regression of u_i onto W_i:

$$u_i = \lambda W_i + a_i$$

where $\lambda = E(Wu)/E(W^2)$. In this population regression, by construction, $E(aW) = 0$. Using this equation for u_i rewrite the equation to be estimated as

$$\begin{aligned}
Y_i &= X_i\beta_1 + W_i\beta_2 + u_i \\
&= X_i\beta_1 + W_i(\beta_2 + \lambda) + a_i \\
&= X_i\beta_1 + W_i\theta + a_i
\end{aligned}$$

where $\theta = \beta_2 + \lambda$. A calculation like that used in part (a) can be used to show that

$$\begin{pmatrix} \sqrt{n}(\widehat{\beta}_1 - \beta_1) \\ \sqrt{n}(\widehat{\beta}_2 - \theta) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i^2 & \frac{1}{n}\sum_{i=1}^n X_i W_i \\ \frac{1}{n}\sum_{i=1}^n W_i X_i & \frac{1}{n}\sum_{i=1}^n W_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i a_i \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n W_i a_i \end{pmatrix}$$
$$\xrightarrow{d} \begin{pmatrix} E(X^2) & 0 \\ 0 & E(W^2) \end{pmatrix}^{-1} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$$

where S_1 is distributed $N(0, \sigma_a^2 E(X_2))$. Thus by Slutsky's theorem

$$\sqrt{n}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} N(0, \frac{\sigma_a^2}{E(X^2)})$$

Now consider the regression that omits W, which can be written as:

$$Y_i = X_i \beta_1 + d_i$$

where $d_i = W_i \theta + a_i$. Calculations like those used above imply that

$$\sqrt{n}(\widehat{\beta}_1^r - \beta_1) \xrightarrow{d} N\left(0, \frac{\sigma_d^2}{E(X^2)}\right).$$

Since $\sigma_d^2 = \sigma_a^2 + \theta^2 E(W^2)$, the asymptotic variance of $\widehat{\beta}_1^r$ is never smaller than the asymptotic variance of $\widehat{\beta}_1$.

16.8. (a) The regression errors satisfy $u_1 = \tilde{u}_1$ and $u_i = 0.5u_{i-1} + \tilde{u}_i$ for $i = 2, 3, \ldots, n$ with the random variables $\tilde{u}_i$ $(i = 1, 2, \ldots, n)$ being i.i.d. with mean 0 and variance 1. For $i > 1$, continuing substituting $u_{i-j} = 0.5u_{i-j-1} + \tilde{u}_{i-j}$ $(j = 1, 2, \ldots i-2)$ and $u_1 = \tilde{u}_1$ into the expression $u_i = 0.5u_{i-1} + \tilde{u}_i$ yields

$$\begin{aligned} u_i &= 0.5u_{i-1} + \tilde{u}_i \\ &= 0.5\,(0.5u_{i-2} + \tilde{u}_{i-1}) + \tilde{u}_i \\ &= 0.5^2\,(0.5u_{i-3} + \tilde{u}_{i-2}) + 0.5\tilde{u}_{i-1} + \tilde{u}_i \\ &= 0.5^3\,(0.5u_{i-4} + \tilde{u}_{i-3}) + 0.5^2\tilde{u}_{i-2} + 0.5\tilde{u}_{i-1} + \tilde{u}_i \\ &= \ldots\ldots \\ &= 0.5^{i-1}\tilde{u}_1 + 0.5^{i-2}\tilde{u}_2 + 0.5^{i-3}\tilde{u}_3 + \cdots + 0.5^2\tilde{u}_{i-2} + 0.5\tilde{u}_{i-1} + \tilde{u}_i \\ &= \sum_{j=1}^{i} 0.5^{i-j}\tilde{u}_j. \end{aligned}$$

Though we get the expression $u_i = \sum_{j=1}^{i} 0.5^{i-j}\tilde{u}_j$ for $i > 1$, it is apparent that it also holds for $i = 1$. Thus we can get mean and variance of random variables

u_i $(i = 1, 2, \dots, n)$:

$$\begin{aligned} E(u_i) &= \sum_{j=1}^{i} 0.5^{i-j} E(\tilde{u}_j) = 0, \\ \sigma_i^2 = \text{var}(u_i) &= \sum_{j=1}^{i} \left(0.5^{i-j}\right)^2 \text{var}(\tilde{u}_j) = \sum_{j=1}^{i} \left(0.5^2\right)^{i-j} \times 1 = \frac{1 - \left(0.5^2\right)^i}{1 - 0.5^2}. \end{aligned}$$

In calculating the variance, the second equality has used the fact that $\tilde{u}_i$ is i.i.d.

Since $u_i = \sum_{j=1}^{i} 0.5^{i-j} \tilde{u}_j$, we know for $k > 0$,

$$\begin{aligned} u_{i+k} &= \sum_{j=1}^{i+k} 0.5^{i+k-j} \tilde{u}_j = 0.5^k \sum_{j=1}^{i} 0.5^{i-j} \tilde{u}_j + \sum_{j=i+1}^{i+k} 0.5^{i+k-j} \tilde{u}_j \\ &= 0.5^k u_i + \sum_{j=i+1}^{i+k} 0.5^{i+k-j} \tilde{u}_j. \end{aligned}$$

Because $\tilde{u}_i$ is i.i.d., the covariance between random variables u_i and u_{i+k} is

$$\begin{aligned} \text{cov}(u_i, u_{i+k}) &= \text{cov}\left(u_i, 0.5^k u_i + \sum_{j=i+1}^{i+k} 0.5^{i+k-j} \tilde{u}_j\right) \\ &= 0.5^k \sigma_i^2. \end{aligned}$$

Similarly we can get

$$\text{cov}(u_i, u_{i-k}) = 0.5^k \sigma_{i-k}^2.$$

The column vector $\mathbf{U}$ for the regression error is

$$\mathbf{U} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

It is straightforward to get

$$E\left(\mathbf{U}\mathbf{U}'\right) = \begin{pmatrix} E\left(u_1^2\right) & E(u_1 u_2) & \cdots & E(u_1 u_n) \\ E(u_2 u_1) & E\left(u_2^2\right) & \cdots & E(u_2 u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_n u_1) & E(u_n u_2) & \cdots & E\left(u_n^2\right) \end{pmatrix}.$$

Because $E(u_i) = 0$, we have $E\left(u_i^2\right) =$var(u_i) and $E(u_i u_j) =$cov(u_i, u_j). Substituting in the results on variances and covariances, we have

$$\Omega = E\left(\mathbf{U}\mathbf{U}'\right) = \begin{pmatrix} \sigma_1^2 & 0.5\sigma_1^2 & 0.5^2\sigma_1^2 & 0.5^3\sigma_1^2 & \cdots & 0.5^{n-1}\sigma_1^2 \\ 0.5\sigma_1^2 & \sigma_2^2 & 0.5\sigma_2^2 & 0.5^2\sigma_2^2 & \cdots & 0.5^{n-2}\sigma_2^2 \\ 0.5^2\sigma_1^2 & 0.5\sigma_2^2 & \sigma_3^2 & 0.5\sigma_3^2 & \cdots & 0.5^{n-3}\sigma_3^2 \\ 0.5^3\sigma_1^2 & 0.5^2\sigma_2^2 & 0.5\sigma_3^2 & \sigma_4^2 & \cdots & 0.5^{n-4}\sigma_4^2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0.5^{n-1}\sigma_1^2 & 0.5^{n-2}\sigma_2^2 & 0.5^{n-3}\sigma_3^2 & 0.5^{n-4}\sigma_4^2 & \cdots & \sigma_n^2 \end{pmatrix}$$

with $\sigma_i^2 = \frac{1-\left(0.5^2\right)^i}{1-0.5^2}$.

(b) The original regression model is

$$Y_i = \beta_0 + \beta_1 X_i + u_i.$$

Lagging each side of the regression equation and substracting 0.5 times this lag from each side gives

$$Y_i - 0.5Y_{i-1} = 0.5\beta_0 + \beta_1\left(X_i - 0.5X_{i-1}\right) + u_i - 0.5u_{i-1}$$

for $i = 2, \ldots, n$ with $u_i - 0.5u_{i-1} = \tilde{u}_i$. Also

$$Y_1 = \beta_0 + \beta_1 X_1 + u_1$$

with $u_1 = \widetilde{u}_1$. Thus we can define a pair of new variables

$$\left(\tilde{Y}_i,\, \tilde{X}_{1i}, \tilde{X}_{2i}\right) = \left(Y_i - 0.5Y_{i-1},\, 0.5,\; X_i - 0.5X_{i-1}\right),$$

for $i = 2, ..., n$ and $\left(\tilde{Y}_1,\, \tilde{X}_{11}, \tilde{X}_{21}\right) = \left(Y_1, 1,\, X_{1i}\right)$, and estimate the regression equation

$$\tilde{Y}_i = \beta_0\tilde{X}_{1i} + \beta_1\tilde{X}_{2i} + \tilde{u}_i$$

using data for $i = 1, ..., n$. The regression error $\tilde{u}_i$ is i.i.d. and distributed independently of $\tilde{X}_i$, thus the new regression model can be estimated directly by the OLS.

16.9. (a) Using Equation (16.45) we know

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= \left(\mathbf{X}'\mathbf{M_W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{Y} \\ &= \left(\mathbf{X}'\mathbf{M_W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M_W}\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{U}\right) \\ &= \boldsymbol{\beta} + \left(\mathbf{X}'\mathbf{M_W}\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{U}. \end{aligned}$$

The last equality has used the orthogonality $\mathbf{M_W W} = \mathbf{0}$. Thus

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(\mathbf{X}'\mathbf{M_W X}\right)^{-1} \mathbf{X}'\mathbf{M_W U} = \left(n^{-1}\mathbf{X}'\mathbf{M_W X}\right)^{-1} \left(n^{-1}\mathbf{X}'\mathbf{M_W U}\right).$$

(b) Using $\mathbf{M_W} = \mathbf{I}_n - \mathbf{P_W}$ and $\mathbf{P_W} = \mathbf{W}\left(\mathbf{W}'\mathbf{W}\right)^{-1}\mathbf{W}'$ we can get

$$\begin{aligned} n^{-1}\mathbf{X}'\mathbf{M_W X} &= n^{-1}\mathbf{X}'\left(\mathbf{I}_n - \mathbf{P_W}\right)\mathbf{X} \\ &= n^{-1}\mathbf{X}'\mathbf{X} - n^{-1}\mathbf{X}'\mathbf{P_W X} \\ &= n^{-1}\mathbf{X}'\mathbf{X} - \left(n^{-1}\mathbf{X}'\mathbf{W}\right)\left(n^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(n^{-1}\mathbf{W}'\mathbf{X}\right). \end{aligned}$$

First consider $n^{-1}\mathbf{X}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i'$. The (j,l) element of this matrix is $\frac{1}{n}\sum_{i=1}^{n}X_{ji}X_{li}$. By Assumption (ii), $\mathbf{X}_i$ is i.i.d., so $X_{ji}X_{li}$ is i.i.d. By Assumption (iii) each element of $\mathbf{X}_i$ has four moments, so by the Cauchy-Schwarz inequality $X_{ji}X_{li}$ has two moments:

$$E\left(X_{ji}^2 X_{li}^2\right) \le \sqrt{E\left(X_{ji}^4\right) \cdot E\left(X_{li}^4\right)} < \infty.$$

Because $X_{ji}X_{li}$ is i.i.d. with two moments, $\frac{1}{n}\sum_{i=1}^{n}X_{ji}X_{li}$ obeys the law of large numbers, so

$$\frac{1}{n}\sum_{i=1}^{n}X_{ji}X_{li} \xrightarrow{p} E\left(X_{ji}X_{li}\right).$$

This is true for all the elements of $n^{-1}\mathbf{X}'\mathbf{X}$, so

$$n^{-1}\mathbf{X}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{X}_i' \xrightarrow{p} E\left(\mathbf{X}_i\mathbf{X}_i'\right) = \Sigma_{\mathbf{XX}}.$$

Applying the same reasoning and using Assumption (ii) that $(\mathbf{X}_i, \mathbf{W}_i, Y_i)$ are i.i.d. and Assumption (iii) that $(\mathbf{X}_i, \mathbf{W}_i, u_i)$ have four moments, we have

$$n^{-1}\mathbf{W}'\mathbf{W} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_i\mathbf{W}_i' \xrightarrow{p} E\left(\mathbf{W}_i\mathbf{W}_i'\right) = \Sigma_{\mathbf{WW}},$$

$$n^{-1}\mathbf{X}'\mathbf{W} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i\mathbf{W}_i' \xrightarrow{p} E\left(\mathbf{X}_i\mathbf{W}_i'\right) = \Sigma_{\mathbf{XW}},$$

and

$$n^{-1}\mathbf{W}'\mathbf{X} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{W}_i\mathbf{X}_i' \xrightarrow{p} E\left(\mathbf{W}_i\mathbf{X}_i'\right) = \Sigma_{\mathbf{WX}}.$$

From Assumption (iii) we know $\Sigma_{\mathbf{XX}}$, $\Sigma_{\mathbf{WW}}$, $\Sigma_{\mathbf{XW}}$, and $\Sigma_{\mathbf{WX}}$ are all finite nonzero. Slutsky's theorem implies

$$\begin{aligned} n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{X} &= n^{-1}\mathbf{X}'\mathbf{X} - \left(n^{-1}\mathbf{X}'\mathbf{W}\right)\left(n^{-1}\mathbf{W}'\mathbf{W}\right)^{-1}\left(n^{-1}\mathbf{W}'\mathbf{X}\right) \\ &\xrightarrow{p} \Sigma_{\mathbf{XX}} - \Sigma_{\mathbf{XW}}\Sigma_{\mathbf{WW}}^{-1}\Sigma_{\mathbf{WX}} \end{aligned}$$

which is finite and invertible.

(c) The conditional expectation

$$\begin{aligned} E(\mathbf{U}|\mathbf{X},\mathbf{W}) &= \begin{pmatrix} E(u_1|\mathbf{X},\mathbf{W}) \\ E(u_2|\mathbf{X},\mathbf{W}) \\ \vdots \\ E(u_n|\mathbf{X},\mathbf{W}) \end{pmatrix} = \begin{pmatrix} E(u_1|\mathbf{X}_1,\mathbf{W}_1) \\ E(u_2|\mathbf{X}_2,\mathbf{W}_2) \\ \vdots \\ E(u_n|\mathbf{X}_n,\mathbf{W}_n) \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{W}_1'\boldsymbol{\delta} \\ \mathbf{W}_2'\boldsymbol{\delta} \\ \vdots \\ \mathbf{W}_n'\boldsymbol{\delta} \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1' \\ \mathbf{W}_2' \\ \vdots \\ \mathbf{W}_n' \end{pmatrix}\boldsymbol{\delta} = \mathbf{W}\boldsymbol{\delta}. \end{aligned}$$

The second equality used Assumption (ii) that $(\mathbf{X}_i, \mathbf{W}_i, Y_i)$ are i.i.d., and the third equality applied the conditional mean independence assumption (i).

(d) In the limit

$$n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{U} \xrightarrow{p} E(\mathbf{X}'\mathbf{M_W}\mathbf{U}|\mathbf{X},\mathbf{W}) = \mathbf{X}'\mathbf{M_W}E(\mathbf{U}|\mathbf{X},\mathbf{W}) = \mathbf{X}'\mathbf{M_W}\mathbf{W}\boldsymbol{\delta} = \mathbf{0}_{k_1\times 1}$$

because $\mathbf{M_W}\mathbf{W} = \mathbf{0}$.

(e) $n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{X}$ converges in probability to a finite invertible matrix, and $n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{U}$ converges in probability to a zero vector. Applying Slutsky's theorem,

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \left(n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{X}\right)^{-1}\left(n^{-1}\mathbf{X}'\mathbf{M_W}\mathbf{U}\right) \xrightarrow{p} \mathbf{0}.$$

This implies

$$\hat{\boldsymbol{\beta}} \xrightarrow{p} \boldsymbol{\beta}.$$