For Instructors

Solutions to End-of-Chapter Exercises

Chapter 2
Review of Probability

2.1. (a) Probability distribution function for Y

Outcome (number of heads)	$Y = 0$	$Y = 1$	$Y = 2$
Probability	0.25	0.50	0.25

(b) Cumulative probability distribution function for Y

Outcome (number of heads)	$Y < 0$	$0 \le Y < 1$	$1 \le Y < 2$	$Y \ge 2$
Probability	0	0.25	0.75	1.0

(c) $\mu_Y = E(Y) = (0\times 0.25) + (1\times 0.50) + (2\times 0.25) = 1.00$. $F \xrightarrow{d} Fq, \infty.$

Using Key Concept 2.3: $\text{var}(Y) = E(Y^2) - [E(Y)]^2$,

and

$$(u_i|X_i)$$

so that

$$\text{var}(Y) = E(Y^2) - [E(Y)]^2 = 1.50 - (1.00)^2 = 0.50.$$

2.2. 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)
$$\begin{aligned}\mu_Y &= E(Y) = 0\times \Pr(Y = 0) + 1\times \Pr(Y = 1)\\ &= 0\times 0.22 + 1\times 0.78 = 0.78,\\ \mu_X &= E(X) = 0\times \Pr(X = 0) + 1\times \Pr(X = 1)\\ &= 0\times 0.30 + 1\times 0.70 = 0.70.\end{aligned}$$

(b)
$$\begin{aligned}\sigma_X^2 &= E[(X - \mu_X)^2]\\ &= (0 - 0.70)^2 \times \Pr(X = 0) + (1 - 0.70)^2 \times \Pr(X = 1)\\ &= (-0.70)^2 \times 0.30 + 0.30^2 \times 0.70 = 0.21,\\ \sigma_Y^2 &= E[(Y - \mu_Y)^2]\\ &= (0 - 0.78)^2 \times \Pr(Y = 0) + (1 - 0.78)^2 \times \Pr(Y = 1)\\ &= (-0.78)^2 \times 0.22 + 0.22^2 \times 0.78 = 0.1716.\end{aligned}$$

(c) $\sigma_{XY} = \text{cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)]$

$$\begin{aligned}
&= (0 - 0.70)(0 - 0.78)\Pr(X = 0, Y = 0) \\
&\quad + (0 - 0.70)(1 - 0.78)\Pr(X = 0, Y = 1) \\
&\quad + (1 - 0.70)(0 - 0.78)\Pr(X = 1, Y = 0) \\
&\quad + (1 - 0.70)(1 - 0.78)\Pr(X = 1, Y = 1) \\
&= (-0.70)\times(-0.78)\times 0.15 + (-0.70)\times 0.22\times 0.15 \\
&\quad + 0.30\times(-0.78)\times 0.07 + 0.30\times 0.22\times 0.63 \\
&= 0.084,
\end{aligned}$$

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

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

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

2.4. (a) $E(X^3) = 0^3 \times (1 - p) + 1^3 \times p = p$

(b) $E(X^k) = 0^k \times (1 - p) + 1^k \times p = p$

(c) $E(X) = 0.3$, and var(X) = $E(X^2)-[E(X)]^2$ = 0.3 −0.09 = 0.21. Thus $\sigma = \sqrt{0.21} = 0.46$.

$\text{var}(X) = E(X^2) - [E(X)]^2 = 0.3 - 0.09 = 0.21$ $\sigma = \sqrt{0.21} = 0.46$. To compute the skewness, use the formula from exercise 2.21:

$$\begin{aligned}
E(X - \mu)^3 &= E(X^3) - 3[E(X^2)][E(X)] + 2[E(X)]^3 \\
&= 0.3 - 3\times 0.3^2 + 2\times 0.3^3 = 0.084
\end{aligned}$$

Alternatively, $E(X - \mu)^3 = [(1 - 0.3)^3 \times 0.3] + [(0 - 0.3)^3 \times 0.7] = 0.084$

Thus, skewness $= E(X - \mu)^3/\sigma^3 = 0.084/0.46^3 = 0.87.$

To compute the kurtosis, use the formula from exercise 2.21:

$$\begin{aligned}
E(X - \mu)^4 &= E(X^4) - 4[E(X)][E(X^3)] + 6[E(X)]^2[E(X^2)] - 3[E(X)]^4 \\
&= 0.3 - 4\times 0.3^2 + 6\times 0.3^3 - 3\times 0.3^4 = 0.0777
\end{aligned}$$

Alternatively, $E(X - \mu)^4 = [(1 - 0.3)^4 \times 0.3] + [(0 - 0.3)^4 \times 0.7] = 0.0777$

Thus, kurtosis is $E(X - \mu)^4/\sigma^4 = 0.0777/0.46^4 = 1.76$

2.5. Let X denote temperature in °F and Y denote temperature in °C. Recall that $Y = 0$ when $X = 32$ and $Y = 100$ when $X = 212$; this implies $Y = (100/180) \times (X - 32)$ or $Y = -17.78 + (5/9) \times X.$ Using Key Concept 2.3, $\mu_X = 70^{\mathrm{o}}$F implies that $\mu_Y = -17.78 + (5/9) \times 70 = 21.11°\mathrm{C},$ and $\sigma_X = 7^{\mathrm{o}}$F implies $\sigma_Y = (5/9) \times 7 = 3.89°\mathrm{C}.$

2.6. The table shows that $\Pr(X = 0, Y = 0) = 0.037,$ $\Pr(X = 0, Y = 1) = 0.622,$ $\Pr(X = 1, Y = 0) = 0.009,$ $\Pr(X = 1, Y = 1) = 0.332,$ $\Pr(X = 0) = 0.659,$ $\Pr(X = 1) = 0.341,$ $\Pr(Y = 0) = 0.046,$ $\Pr(Y = 1) = 0.954.$

(a) $$\begin{aligned} E(Y) = \mu_Y &= 0 \times \Pr(Y = 0) + 1 \times \Pr(Y = 1) \\ &= 0 \times 0.046 + 1 \times 0.954 = 0.954. \end{aligned}$$

(b) $$\begin{aligned} \text{Unemployment Rate} &= \frac{\#(\text{unemployed})}{\#(\text{labor force})} \\ &= \Pr(Y = 0) = 1 - \Pr(Y = 1) = 1 - E(Y) = 1 - 0.954 = 0.046. \end{aligned}$$

(c) Calculate the conditional probabilities first:

$$\Pr(Y = 0|X = 0) = \frac{\Pr(X = 0, Y = 0)}{\Pr(X = 0)} = \frac{0.037}{0.659} = 0.056,$$

$$\Pr(Y = 1|X = 0) = \frac{\Pr(X = 0, Y = 1)}{\Pr(X = 0)} = \frac{0.622}{0.659} = 0.944,$$

$$\Pr(Y = 0|X = 1) = \frac{\Pr(X = 1, Y = 0)}{\Pr(X = 1)} = \frac{0.009}{0.341} = 0.026,$$

$$\Pr(Y = 1|X = 1) = \frac{\Pr(X = 1, Y = 1)}{\Pr(X = 1)} = \frac{0.332}{0.341} = 0.974.$$

The conditional expectations are

$$\begin{aligned} E(Y|X = 1) &= 0 \times \Pr(Y = 0|X = 1) + 1 \times \Pr(Y = 1|X = 1) \\ &= 0 \times 0.026 + 1 \times 0.974 = 0.974, \\ E(Y|X = 0) &= 0 \times \Pr(Y = 0|X = 0) + 1 \times \Pr(Y = 1|X = 0) \\ &= 0 \times 0.056 + 1 \times 0.944 = 0.944. \end{aligned}$$

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

Unemployment rate for college graduates = $1 - E(Y|X = 1) = 1 - 0.974 = 0.026$

Unemployment rate for non-college graduates = $1 - E(Y|X = 0) = 1 - 0.944 = 0.056$

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

$$\Pr(X = 1|Y = 0) = \frac{\Pr(X = 1, Y = 0)}{\Pr(Y = 0)} = \frac{0.009}{0.046} = 0.196.$$

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

$$\Pr(X = 0|Y = 0) = 1 - \Pr(X = 1|Y = 0) = 1 - 0.196 = 0.804.$$

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

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

For example, from part (e) $\Pr(X = 0|Y = 0) = 0.804$, while from the table $\Pr(X = 0) = 0.659$.

2.7. Using obvious notation, $C = M + F$; thus $\mu_C = \mu_M + \mu_F$ and $\sigma_C^2 = \sigma_M^2 + \sigma_F^2 + 2\text{cov}(M, F)$. This implies

(a) $\mu_C = 40 + 45 = \$85,000$ per year.

(b) $\text{corr}(M, F) = \dfrac{\text{cov}(M, F)}{\sigma_M \sigma_F}$, so that $\text{cov}(M, F) = \sigma_M \sigma_F \text{corr}(M, F)$. Thus $\text{cov}(M, F) =$ $12 \times 18 \times 0.80 = 172.80$, where the units are squared thousands of dollars per year.

(c) $\sigma_C^2 = \sigma_M^2 + \sigma_F^2 + 2\text{cov}(M, F)$, so that $\sigma_C^2 = 12^2 + 18^2 + 2 \times 172.80 = 813.60$, and $\sigma_C = \sqrt{813.60} = 28.524$ thousand dollars per year.

(d) First you need to look up the current Euro/dollar exchange rate in the Wall Street Journal, the Federal Reserve web page, or other financial data outlet. Suppose that this exchange rate is e (say $e = 0.80$ Euros per dollar); each 1 dollar is therefore with e Euros. The mean is therefore $e \times \mu_C$ (in units of thousands of Euros per year), and the standard deviation is $e \times \sigma_C$ (in units of thousands of Euros per year). The correlation is unit-free, and is unchanged.

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

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

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

2.9.

		Value of *Y*					Probability Distribution of *X*
		14	**22**	**30**	**40**	**65**	
Value of *X*	1	0.02	0.05	0.10	0.03	0.01	0.21
	5	0.17	0.15	0.05	0.02	0.01	0.40
	8	0.02	0.03	0.15	0.10	0.09	0.39
Probability distribution of *Y*		0.21	0.23	0.30	0.15	0.11	1.00

(a) The probability distribution is given in the table above.

$$E(Y) = 14 \times 0.21 + 22 \times 0.23 + 30 \times 0.30 + 40 \times 0.15 + 65 \times 0.11 = 30.15$$

$$E(Y^2) = 14^2 \times 0.21 + 22^2 \times 0.23 + 30^2 \times 0.30 + 40^2 \times 0.15 + 65^2 \times 0.11 = 1127.23$$

$$\text{var}(Y) = E(Y^2) - [E(Y)]^2 = 218.21$$

$$\sigma_Y = 14.77$$

(b) The conditional probability of $Y|X = 8$ is given in the table below

Value of *Y*				
14	22	30	40	65
0.02/0.39	0.03/0.39	0.15/0.39	0.10/0.39	0.09/0.39

$$E(Y|X=8) = 14 \times (0.02/0.39) + 22 \times (0.03/0.39) + 30 \times (0.15/0.39)$$
$$+ 40 \times (0.10/0.39) + 65 \times (0.09/0.39) = 39.21$$
$$E(Y^2|X=8) = 14^2 \times (0.02/0.39) + 22^2 \times (0.03/0.39) + 30^2 \times (0.15/0.39)$$
$$+ 40^2 \times (0.10/0.39) + 65^2 \times (0.09/0.39) = 1778.7$$
$$\text{var}(Y) = 1778.7 - 39.21^2 = 241.65$$
$$\sigma_{Y|X=8} = 15.54$$

(c) $E(XY) = (1 \times 14 \times 0.02) + (1 \times 22 : 0.05) + \cdots + (8 \times 65 \times 0.09) = 171.7$
$$\text{cov}(X, Y) = E(XY) - E(X)E(Y) = 171.7 - 5.33 \times 30.15 = 11.0$$
$$\text{corr}(X, Y) = \text{cov}(X, Y)/(\sigma_X \sigma_Y) = 11.0 / (2.60 \times 14.77) = 0.286$$

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

(a) $\Pr(Y \le 3) = \Pr\left(\dfrac{Y-1}{2} \le \dfrac{3-1}{2}\right) = \Phi(1) = 0.8413.$

(b) $\Pr(Y > 0) = 1 - \Pr(Y \le 0) = 1 - \Pr\left(\dfrac{Y-3}{3} \le \dfrac{0-3}{3}\right)$
$$= 1 - \Phi(-1) = \Phi(1) = 0.8413.$$

(c) $\Pr(40 \le Y \le 52) = \Pr\left(\dfrac{40-50}{5} \le \dfrac{Y-50}{5} \le \dfrac{52-50}{5}\right)$
$$= \Phi(0.4) - \Phi(-2) = \Phi(0.4) - [1 - \Phi(2)]$$
$$= 0.6554 - 1 + 0.9772 = 0.6326.$$

(d) $\Pr(6 \le Y \le 8) = \Pr\left(\dfrac{6-5}{\sqrt{2}} \le \dfrac{Y-5}{\sqrt{2}} \le \dfrac{8-5}{\sqrt{2}}\right)$
$$= \Phi(2.1213) - \Phi(0.7071)$$
$$= 0.9831 - 0.7602 = 0.2229.$$

2.11. (a) 0.90
(b) 0.05
(c) 0.05
(d) When $Y \sim \chi^2_{10}$, then $Y/10 \sim F_{10,\infty}$.
(e) $Y = Z^2$, where $Z \sim N(0,1)$, thus $\Pr(Y \le 1) = \Pr(-1 \le Z \le 1) = 0.32.$

2.12. (a) 0.05
(b) 0.950
(c) 0.953
(d) The t_{df} distribution and $N(0, 1)$ are approximately the same when df is large.
(e) 0.10
(f) 0.01

2.13. (a) $E(Y^2) = \text{Var}(Y) + \mu_Y^2 = 1 + 0 = 1; E(W^2) = \text{Var}(W) + \mu_W^2 = 100 + 0 = 100.$
(b) Y and W are symmetric around 0, thus skewness is equal to 0; because their mean is zero, this means that the third moment is zero.
(c) The kurtosis of the normal is 3, so $3 = E(Y - \mu_Y)^4 / \sigma_Y^4$; solving yields $\text{E}(Y^4) = 3;$ a similar calculation yields the results for W.
(d) First, condition on $X = 0,$ so that $S = W$:

$$E(S|X=0)=0; E(S^2|X=0)=100, E(S^3|X=0)=0, E(S^4|X=0)=3\times 100^2.$$

Similarly,

$$E(S|X=1)=0; E(S^2|X=1)=1, E(S^3|X=1)=0, E(S^4|X=1)=3.$$

From the law of iterated expectations

$$E(S) = E(S|X=0)\times \Pr(\text{X}=0) + E(S|X=1)\times \Pr(X=1) = 0$$
$$E(S^2) = E(S^2|X=0)\times \Pr(\text{X}=0) + E(S^2|X=1)\times \Pr(X=1) = 100\times 0.01 + 1\times 0.99 = 1.99$$
$$E(S^3) = E(S^3|X=0)\times \Pr(\text{X}=0) + E(S^3|X=1)\times \Pr(X=1) = 0$$
$$E(S^4) = E(S^4|X=0)\times \Pr(\text{X}=0) + E(S^4|X=1)\times \Pr(X=1)$$
$$= 3\times 100^2 \times 0.01 + 3\times 1\times 0.99 = 302.97$$

(e) $\mu_S = E(S) = 0,$ thus $E(S - \mu_S)^3 = E(S^3) = 0$ from part (d). Thus skewness = 0. Similarly, $\sigma_S^2 = E(S - \mu_S)^2 = E(S^2) = 1.99,$ and $E(S - \mu_S)^4 = E(S^4) = 302.97.$ Thus, $\text{kurtosis} = 302.97 / (1.99^2) = 76.5$

2.14. 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 $\mu_Y = 100,$ $\sigma_Y^2 = 43.0,$

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

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

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

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

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

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

2.15. (a) $\Pr(9.6 \le \bar{Y} \le 10.4) = \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le \frac{\bar{Y}-10}{\sqrt{4/n}} \le \frac{10.4-10}{\sqrt{4/n}}\right)$

$$= \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right)$$

where $Z \sim N(0, 1)$. Thus,

(i) $n = 20;\ \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right) = \Pr(-0.89 \le Z \le 0.89) = 0.63$

(ii) $n = 100;\ \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right) = \Pr(-2.00 \le Z \le 2.00) = 0.954$

(iii) $n = 1000;\ \Pr\left(\frac{9.6-10}{\sqrt{4/n}} \le Z \le \frac{10.4-10}{\sqrt{4/n}}\right) = \Pr(-6.32 \le Z \le 6.32) = 1.000$

(b) $\Pr(10 - c \le \bar{Y} \le 10 + c) = \Pr\left(\frac{-c}{\sqrt{4/n}} \le \frac{\bar{Y}-10}{\sqrt{4/n}} \le \frac{c}{\sqrt{4/n}}\right)$

$$= \Pr\left(\frac{-c}{\sqrt{4/n}} \le Z \le \frac{c}{\sqrt{4/n}}\right).$$

As n get large $\frac{c}{\sqrt{4/n}}$ gets large, and the probability converges to 1.

(c) This follows from (b) and the definition of convergence in probability given in Key Concept 2.6.

2.16. There are several ways to do this. Here is one way. Generate n draws of Y, $Y_1, Y_2, \ldots Y_n$. Let $X_i = 1$ if $Y_i < 3.6$, otherwise set $X_i = 0$. Notice that X_i is a Bernoulli random variables with $\mu_X = \Pr(X = 1) = \Pr(Y < 3.6)$. Compute $\bar{X}$. Because $\bar{X}$ converges in probability to $\mu_X = \Pr(X = 1) = \Pr(Y < 3.6)$, $\bar{X}$ will be an accurate approximation if n is large.

2.17. $\mu_Y = 0.4$ and $\sigma_Y^2 = 0.4 \times 0.6 = 0.24$

(a) (i) $P(\bar{Y} \ge 0.43) = \Pr\left(\frac{\bar{Y}-0.4}{\sqrt{0.24/n}} \ge \frac{0.43-0.4}{\sqrt{0.24/n}}\right) = \Pr\left(\frac{\bar{Y}-0.4}{\sqrt{0.24/n}} \ge 0.6124\right) = 0.27$

(ii) $P(\overline{Y} \le 0.37) = \Pr\left(\frac{\overline{Y}-0.4}{\sqrt{0.24/n}} \le \frac{0.37-0.4}{\sqrt{0.24/n}}\right) = \Pr\left(\frac{\overline{Y}-0.4}{\sqrt{0.24/n}} \le -1.22\right) = 0.11$

(b) We know Pr(−1.96 ≤ Z ≤ 1.96) = 0.95, thus we want n to satisfy $0.41 = \frac{0.41-0.40}{\sqrt{24/n}} > -1.96$ and $\frac{0.39-0.40}{\sqrt{24/n}} < -1.96$. Solving these inequalities yields $n \ge 9220$.

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

(a) The mean of Y is

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

The variance of Y is

$$\begin{aligned}\sigma_Y^2 &= E\left[(Y-\mu_Y)^2\right]\\ &=(0-1000)^2\times\Pr(Y=0)+(20000-1000)^2\times\Pr(Y=20000)\\ &=(-1000)^2\times0.95+19000^2\times0.05=1.9\times10^7,\end{aligned}$$

so the standard deviation of Y is $\sigma_Y = (1.9\times10^7)^{\frac{1}{2}} = \$4359$.

(b) (i) $E(\overline{Y})=\mu_Y=\$1000$, $\sigma_{\overline{Y}}^2=\frac{\sigma_Y^2}{n}=\frac{1.9\times10^7}{100}=1.9\times10^5$.

(ii) Using the central limit theorem,

$$\begin{aligned}\Pr(\overline{Y}>2000) &= 1-\Pr(\overline{Y}\le 2000)\\ &=1-\Pr\left(\frac{\overline{Y}-1000}{\sqrt{1.9\times10^5}} \le \frac{2{,}000-1{,}000}{\sqrt{1.9\times10^5}}\right)\\ &\approx 1-\Phi(2.2942)=1-0.9891=0.0109.\end{aligned}$$

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

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

(c) When X and Y are independent,

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

so

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

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

2.20. (a) $\Pr(Y=y_i) = \sum_{j=1}^{l}\sum_{h=1}^{m}\Pr(Y=y_i|X=x_j, Z=z_h)\Pr(X=x_j, Z=z_h)$

(b)
$$
\begin{aligned}
E(Y) &= \sum_{i=1}^{k} y_i\Pr(Y=y_i)\Pr(Y=y_i) \\
&= \sum_{i=1}^{k} y_i\sum_{j=1}^{l}\sum_{h=1}^{m}\Pr(Y=y_i|X=x_j, Z=z_h)\Pr(X=x_j, Z=z_h) \\
&= \sum_{j=1}^{l}\sum_{h=1}^{m}\left[\sum_{i=1}^{k} y_i\Pr(Y=y_i|X=x_j, Z=z_h)\right]\Pr(X=x_j, Z=z_h) \\
&= \sum_{j=1}^{l}\sum_{h=1}^{m} E(Y|X=x_j, Z=z_h)\Pr(X=x_j, Z=z_h)
\end{aligned}
$$

where the first line in the definition of the mean, the second uses (a), the third is a rearrangement, and the final line uses the definition of the conditional expectation.

2.21. (a)
$$
\begin{aligned}
E(X-\mu)^3 &= E[(X-\mu)^2(X-\mu)] = E[X^3 - 2X^2\mu + X\mu^2 - X^2\mu + 2X\mu^2 - \mu^3] \\
&= E(X^3) - 3E(X^2)\mu + 3E(X)\mu^2 - \mu^3 = E(X^3) - 3E(X^2)E(X) \\
&\quad + 3E(X)[E(X)]^2 - [E(X)]^3 \\
&= E(X^3) - 3E(X^2)E(X) + 2E(X)^3
\end{aligned}
$$

(b)
$$
\begin{aligned}
E(X-\mu)^4 &= E[(X^3 - 3X^2\mu + 3X\mu^2 - \mu^3)(X-\mu)] \\
&= E[X^4 - 3X^3\mu + 3X^2\mu^2 - X\mu^3 - X^3\mu + 3X^2\mu^2 - 3X\mu^3 + \mu^4] \\
&= E(X^4) - 4E(X^3)E(X) + 6E(X^2)E(X)^2 - 4E(X)E(X)^3 + E(X)^4 \\
&= E(X^4) - 4[E(X)][E(X^3)] + 6[E(X)]^2[E(X^2)] - 3[E(X)]^4
\end{aligned}
$$

2.22. The mean and variance of R are given by

$$\mu = w\times 0.08 + (1-w)\times 0.05$$
$$\sigma^2 = w^2\times 0.07^2 + (1-w)^2\times 0.042 + 2\times w\times(1-w)\times[0.07\times 0.04\times 0.25]$$

where $0.07\times 0.04\times 0.25 = Cov(R_s,\ R_b)$ follows from the definition of the correlation between R_s and R_b.

(a) $\mu = 0.065;\ \sigma = 0.044$

(b) $\mu = 0.0725;\ \sigma = 0.056$

(c) $w = 1$ maximizes μ; $\sigma = 0.07$ for this value of w.

(d) The derivative of σ^2 with respect to w is

$$\begin{aligned}\frac{d\sigma^2}{dw} &= 2w\times 0.07^2 - 2(1-w)\times 0.04^2 + (2-4w)\times[0.07\times 0.04\times 0.25]\\ &= 0.0102w - 0.0018\end{aligned}$$

Solving for w yields $w = 18/102 = 0.18$. (Notice that the second derivative is positive, so that this is the global minimum.) With $w = 0.18, \sigma_R = 0.038$.

2.23. 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 $E(Y|X) = E(X^2 + Z|X) = E(X^2|X) + E(Z|X) = X^2 + 0 = X^2$.

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

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

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

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

2.24. (a) $E(Y_i^2) = \sigma^2 + \mu^2 = \sigma^2$ and the result follows directly.

(b) (Y_i/σ) is distributed i.i.d. $N(0,1)$, $W = \sum_{i=1}^n (Y_i/\sigma)^2$, and the result follows from the definition of a χ_n^2 random variable.

(c) $E(W) = E\sum_{i=1}^n \frac{Y_i^2}{\sigma^2} = \sum_{i=1}^n E\frac{Y_i^2}{\sigma^2} = n.$

(d) Write

$$V = \frac{Y_1}{\sqrt{\frac{\sum_{i=2}^n Y_i^2}{n-1}}} = \frac{Y_1/\sigma}{\sqrt{\frac{\sum_{i=2}^n (Y/\sigma)^2}{n-1}}}$$

which follows from dividing the numerator and denominator by σ. $Y_1/\sigma \sim N(0,1)$, $\sum_{i=2}^n (Y_i/\sigma)^2 \sim \chi^2_{n-1}$, and Y_1/σ and $\sum_{i=2}^n (Y_i/\sigma)^2$ are independent. The result then follows from the definition of the t distribution.

2.25. (a) $\sum_{i=1}^n ax_i = (ax_1 + ax_2 + ax_3 + \cdots + ax_n) = a(x_1 + x_2 + x_3 + \cdots + x_n) = a\sum_{i=1}^n x_i$

(b)
$$\begin{aligned}\sum_{i=1}^n (x_i + y_i) &= (x_1 + y_1 + x_2 + y_2 + \cdots x_n + y_n)\\ &= (x_1 + x_2 + \cdots x_n) + (y_1 + y_2 + \cdots y_n)\\ &= \sum_{i=1}^n x_i + \sum_{i=1}^n y_i\end{aligned}$$

(c) $\sum_{i=1}^n a = (a + a + a + \cdots + a) = na$

(d)
$$\begin{aligned}\sum_{i=1}^n (a + bx_i + cy_i)^2 &= \sum_{i=1}^n (a^2 + b^2 x_i^2 + c^2 y_i^2 + 2abx_i + 2acy_i + 2bcx_i y_i)\\ &= na^2 + b^2\sum_{i=1}^n x_i^2 + c^2\sum_{i=1}^n y_i^2 + 2ab\sum_{i=1}^n x_i + 2ac\sum_{i=1}^n y_i + 2bc\sum_{i=1}^n x_i y_i\end{aligned}$$

2.26. (a) corr(Y_i,Y_j) = $\text{cov}(Y_i, Y_j)/\sigma_{Y_i}\sigma_{Y_j} = \text{cov}(Y_i, Y_j)/\sigma_Y\sigma_Y = \text{cov}(Y_i, Y_j)/\sigma_Y^2 = \rho$, where the first equality uses the definition of correlation, the second uses the fact that Y_i and Y_j have the same variance (and standard deviation), the third equality uses the definition of standard deviation, and the fourth uses the correlation given in the problem. Solving for cov(Y_i, Y_j) from the last equality gives the desired result.

(b) $\overline{Y} = \frac{1}{2}Y_1 + \frac{1}{2}Y_2$, so that $E(\overline{Y}) = \frac{1}{2}E(Y)_1 + \frac{1}{2}E(Y_2) = \mu_Y$

$$\text{var}(\overline{Y}) = \frac{1}{4}\text{var}(Y_1) + \frac{1}{4}\text{var}(Y_2) + \frac{2}{4}\text{cov}(Y_1, Y_2) = \frac{\sigma_Y^2}{2} + \frac{\rho\sigma_Y^2}{2}$$

(c) $\overline{Y} = \frac{1}{n}\sum_{i=1}^n Y_i$, so that $E(\overline{Y}) = \frac{1}{n}\sum_{i=1}^n E(Y_i) = \frac{1}{n}\sum_{i=1}^n \mu_Y = \mu_Y$

$$\begin{aligned}
\text{var}(\overline{Y}) &= \text{var}\left(\frac{1}{n}\sum_{i=1}^{n} Y_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\text{var}(Y_i) + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\text{cov}(Y_i, Y_j) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma_Y^2 + \frac{2}{n^2}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\rho\sigma_Y^2 \\
&= \frac{\sigma_Y^2}{n} + \frac{n(n-1)}{n^2}\rho\sigma_Y^2 \\
&= \frac{\sigma_Y^2}{n} + \left(1 - \frac{1}{n}\right)\rho\sigma_Y^2
\end{aligned}$$

where the fourth line uses $\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} a = a(1+2+3+\cdots+n-1) = \frac{an(n-1)}{2}$ for any variable a.

(d) When n is large $\frac{\sigma_Y^2}{n} \approx 0$ and $\frac{1}{n} \approx 0$, and the result follows from (c).

2.27 (a) $E(W) = E[E(W|Z)] = E[E(X - \tilde{X})|Z] = E[E(X|Z) - E(X|Z)] = 0$.

(b) $E(WZ) = E[E(WZ|Z)] = E[ZE(W)|Z] = E[Z \times 0] = 0$

(c) Using the hint: $V = W - h(Z)$, so that $E(V^2) = E(W^2) + E[h(Z)^2] - 2 \times E[W \times h(Z)]$. Using an argument like that in (b), $E[W \times h(Z)] = 0$. Thus, $E(V^2) = E(W^2) + E[h(Z)^2]$, and the result follows by recognizing that $E[h(Z)^2] \geq 0$ because $h(z)^2 \leq 0$ for any value of z.

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 $N\left(\mu_Y, \sigma^2_{\bar{Y}}\right)$ with $\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n}$. Given a population $\mu_Y = 100$, $\sigma^2_Y = 43.0$, we have

(a) $n = 100$, $\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n} = \frac{43}{100} = 0.43$, and

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

(b) $n = 64$, $\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n} = \frac{43}{64} = 0.6719$, and

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

(c) $n = 165$, $\sigma^2_{\bar{Y}} = \frac{\sigma^2_Y}{n} = \frac{43}{165} = 0.2606$, and

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

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

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

and variance

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

(a) The fraction of successes is

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

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

The second equality uses the fact that $Y_1, \ldots, Y_n$ are i.i.d. draws and $\text{cov}(Y_i, Y_j) = 0,$ 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 $\Pr(Y = 0) = 1 - p.$ From the solution to Exercise 3.2, Y has mean p and variance $p(1-p).$

(a) $\hat{p} = \frac{215}{400} = 0.5375.$

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

(c) The computed t-statistic is

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

Because of the large sample size $(n = 400),$ we can use Equation (3.14) in the text to get the p-value for the test $H_0 : p = 0.5$ vs. $H_1 : p \neq 0.5:$

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

(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

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

(e) Part (c) is a two-sided test and the p-value is the area in the tails of the standard normal distribution outside ± (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 1.64 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

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

(b) 99% confidence interval for p is

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

(c) Mechanically, the interval in (b) is wider because of a larger critical value (2.57 versus 1.96). Substantively, a 99% confidence interval is wider than a 95% confidence because a 99% confidence interval must contain the true value of p in 99% of all possible samples, while a 95% confidence interval must contain the true value of p in only 95% of all possible samples.

(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. (a) (i) The size is given by $\Pr(|\hat{p} - 0.5| > .02)$, where the probability is computed assuming that $p = 0.5$.

$$\begin{aligned}\Pr(|\hat{p} - 0.5| > 0.02) &= 1 - \Pr(-0.02 \le \hat{p} - 0.5 \le .02)\\ &= 1 - \Pr\left(\frac{-0.02}{\sqrt{0.5 \times 0.5/1055}} \le \frac{\hat{p} - 0.5}{\sqrt{0.5 \times 0.5/1055}} \le \frac{0.02}{\sqrt{0.5 \times 0.5/1055}}\right)\\ &= 1 - \Pr\left(-1.30 \le \frac{\hat{p} - 0.5}{\sqrt{0.5 \times 0.5/1055}} \le 1.30\right)\\ &= 0.19\end{aligned}$$

where the final equality using the central limit theorem approximation.

(ii) The power is given by $\Pr(|\hat{p} - 0.5| > 0.02)$, where the probability is computed assuming that $p = 0.53$.

$$\begin{aligned}\Pr(|\hat{p} - 0.5| > 0.02) &= 1 - \Pr(-0.02 \le \hat{p} - 0.5 \le .02)\\ &= 1 - \Pr\left(\frac{-0.02}{\sqrt{0.53 \times 0.47/1055}} \le \frac{\hat{p} - 0.5}{\sqrt{0.53 \times 0.47/1055}} \le \frac{0.02}{\sqrt{0.53 \times 0.47/1055}}\right)\\ &= 1 - \Pr\left(\frac{-0.05}{\sqrt{0.53 \times 0.47/1055}} \le \frac{\hat{p} - 0.53}{\sqrt{0.53 \times 0.47/1055}} \le \frac{-0.01}{\sqrt{0.53 \times 0.47/1055}}\right)\\ &= 1 - \Pr\left(-3.25 \le \frac{\hat{p} - 0.53}{\sqrt{.53 \times 0.47/1055}} \le -0.65\right)\\ &= 0.74\end{aligned}$$

where the final equality using the central limit theorem approximation.

(b) (i) $t = \dfrac{0.54 - 0.50}{\sqrt{(0.54 \times 0.46)/1055}} = 2.61$, and $\Pr(|t| > 2.61) = 0.01$, so that the null is rejected at the 5% level.

(ii) $\Pr(t > 2.61) = .004$, so that the null is rejected at the 5% level.

(iii) $0.54 \pm 1.96\sqrt{(0.54 \times 0.46)/1055} = 0.54 \pm 0.03$, or 0.51 to 0.57.

(iv) $0.54 \pm 2.58\sqrt{(0.54 \times 0.46)/1055} = 0.54 \pm 0.04$, or 0.50 to 0.58.

(v) $0.54 \pm 0.67\sqrt{(0.54 \times 0.46)/1055} = 0.54 \pm 0.01$, or 0.53 to 0.55.

(c) (i) The probability is 0.95 is any single survey, there are 20 independent surveys, so the probability if $0.95^{20} = 0.36$

(ii) 95% of the 20 confidence intervals or 19.

(d) The relevant equation is $1.96 \times \mathrm{SE}(\hat{p}) < .01$ or $1.96 \times \sqrt{p(1-p)/n} < .01$. Thus n must be chosen so that $n > \frac{1.96^2 p(1-p)}{0.01^2}$, so that the answer depends on the value of p. Note that the largest value that $p(1 - p)$ can take on is 0.25 (that is, $p = 0.5$ makes $p(1 - p)$ as large as possible). Thus if $n > \frac{1.96^2 \times 0.25}{0.01^2} = 9604$, then the margin of error is less than 0.01 for all values of p.

3.6. (a) No. Because the p-value is less than 0.05 (= 5%), $\mu = 5$ is rejected at the 5% level and is therefore not contained in the 95% confidence interval.

(b) No. This would require calculation of the t-statistic for $\mu = 6$, which requires $\bar{Y}$ and $\mathrm{SE}(\bar{Y})$. Only the p-value for test that $\mu = 5$ is given in the problem.

3.7. The null hypothesis is that the survey is a random draw from a population with $p = 0.11$. The t-statistic is $t = \frac{\hat{p} - 0.11}{\mathrm{SE}(\hat{p})}$, where $\mathrm{SE}(\hat{p}) = \hat{p}(1-\hat{p})/n.$ (An alternative formula for SE($\hat{p}$) is $0.11 \times (1-0.11)/n$, which is valid under the null hypothesis that $p = 0.11$). The value of the t-statistic is 2.71, which has a p-value of that is less than 0.01. Thus the null hypothesis $p = 0.11$ (the survey is unbiased) can be rejected at the 1% level.

3.8 $1110 \pm 1.96\left(\frac{123}{\sqrt{1000}}\right)$ or 1110 ± 7.62.

3.9. 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

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

where $\Pr(\overline{Y} > 2100|\mu = 2000)$ 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:

$$\beta = \Pr(\overline{Y} \leq 2100|\mu = 2150) = \Pr\left(\frac{\overline{Y} - 2150}{20} \leq \frac{2100 - 2150}{20}|\mu = 2150\right)$$
$$= \Phi(-2.5) = 1 - \Phi(2.5) = 1 - 0.9938 = 0.0062.$$

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.

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

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%.

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

$$\mu_1 = \overline{Y}_1 \pm 1.96\text{SE}(\overline{Y}_1) = 58 \pm 1.96 \times 0.8 = (56.432, 59.568).$$

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

$$\mu_1 - \mu_2 = (\overline{Y}_1 - \overline{Y}_2) \pm 1.64\text{SE}(\overline{Y}_1 - \overline{Y}_2)$$
$$= (58 - 62) \pm 1.64 \times 1.1158 = (-5.8299, -2.1701).$$

(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 $\mathrm{SE}(\overline{Y}_1 - \overline{Y}_2) =$ 1.1158. The *t*-statistic for testing the null hypothesis is

$$t^{act} = \frac{\overline{Y}_1 - \overline{Y}_2}{\mathrm{SE}(\overline{Y}_1 - \overline{Y}_2)} = \frac{58 - 62}{1.1158} = -3.5849.$$

Use Equation (3.14) in the text to compute the *p*-value:

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

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.

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

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

$$= \frac{1}{n}\left(\frac{1}{2} \cdot \frac{n}{2} \cdot \mu_Y + \frac{3}{2} \cdot \frac{n}{2} \cdot \mu_Y\right) = \mu_Y$$

$$\mathrm{var}(\tilde{Y}) = \frac{1}{n^2}\left(\frac{1}{4}\mathrm{var}(Y_1) + \frac{9}{4}\mathrm{var}(Y_2) + \cdots \frac{1}{4}\mathrm{var}(Y_{n-1}) + \frac{9}{4}\mathrm{var}(Y_n)\right)$$

$$= \frac{1}{n^2}\left(\frac{1}{4} \cdot \frac{n}{2} \cdot \sigma_Y^2 + \frac{9}{4} \cdot \frac{n}{2} \cdot \sigma_Y^2\right) = 1.25\frac{\sigma_Y^2}{n}.$$

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

(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

$$t^{act} = \frac{\overline{Y}_1 - \overline{Y}_2}{\mathrm{SE}(\overline{Y}_1 - \overline{Y}_2)} = \frac{3100 - 2900}{44.721} = 4.4722.$$

Use Equation (3.14) in the text to get the *p*-value:

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

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, and a related calculation shows overwhelming evidence that mean earning for men are *greater* that 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.

3.13 (a) Sample size $n = 420,$ sample average $\overline{Y} = 646.2$ sample standard deviation $s_Y = 19.5.$ The standard error of $\overline{Y}$ is $\text{SE}(\overline{Y}) = \frac{s_Y}{\sqrt{n}} = \frac{19.5}{\sqrt{420}} = 0.9515.$ The 95% confidence interval for the mean test score in the population is

$$\mu = \overline{Y} \pm 1.96\text{SE}(\overline{Y}) = 646.2 \pm 1.96 \times 0.9515 = (644.34,\ 648.06).$$

(b) The data are: sample size for small classes $n_1 = 238,$ sample average $\overline{Y}_1 = 657.4,$ sample standard deviation $s_1 = 19.4;$ sample size for large classes $n_2 = 182,$ sample average $\overline{Y}_2 = 650.0,$ sample standard deviation $s_2 = 17.9.$ The standard error of $\overline{Y}_1 - \overline{Y}_2$ is $SE(\overline{Y}_1 - \overline{Y}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{\frac{19.4^2}{238} + \frac{17.9^2}{182}} = 1.8281.$ 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

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

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

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

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.

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

3.15. From the textbook equation (2.46), we know that $E(\bar{Y}) = \mu_Y$ and from (2.47) we know that $\text{var}(\bar{Y}) = \frac{\sigma_Y^2}{n}$. In this problem, because Y_a and Y_b are Bernoulli random variables, $\hat{p}_a = \bar{Y}_a$, $\hat{p}_b = \bar{Y}_b$, $\sigma_{Ya}^2 = p_a(1-p_a)$ and $\sigma_{Yb}^2 = p_b(1-p_b)$. The answers to (a) follow from this. For part (b), note that $\text{var}(\hat{p}_a - \hat{p}_b) = \text{var}(\hat{p}_a) + \text{var}(\hat{p}_b) - 2\text{cov}(\hat{p}_a, \hat{p}_b)$. But, they are independent (and thus have $\text{cov}(\hat{p}_a, \hat{p}_b) = 0$ because $\hat{p}_a$ and $\hat{p}_b$ are independent (they depend on data chosen from independent samples). Thus $\text{var}(\hat{p}_a - \hat{p}_b) = \text{var}(\hat{p}_a) + \text{var}(\hat{p}_b)$. For part (c), use equation 3.21 from the text (replacing $\bar{Y}$ with $\hat{p}$ and using the result in (b) to compute the SE). For (d), apply the formula in (c) to obtain

$$95\% \text{ CI is } (.859 - .374) \pm 1.96\sqrt{\frac{0.859(1-0.859)}{5801} + \frac{0.374(1-0.374)}{4249}} \text{ or } 0.485 \pm 0.017.$$

3.16. (a) The 95% confidence interval if $\bar{Y} \pm 1.96\,\text{SE}(\bar{Y})$ or $1013 \pm 1.96 \times \frac{108}{\sqrt{453}}$ or 1013 ± 9.95.

(b) The confidence interval in (a) does not include $\mu = 1000$, so the null hypothesis that $\mu = 1000$ (Florida students have the same average performance as students in the U.S.) can be rejected at the 5% level.)

(c) (i) The 95% confidence interval is $\bar{Y}_{prep} - \bar{Y}_{Non-prep} \pm 1.96\, SE\,(\bar{Y}_{prep} - \bar{Y}_{Non-prep})$ where $\text{SE}(\bar{Y}_{prep} - \bar{Y}_{Non-prep}) = \sqrt{\frac{s_{prep}^2}{n_{prep}} + \frac{s_{non-prep}^2}{n_{non-prep}}} = \sqrt{\frac{95^2}{503} + \frac{108^2}{453}} = 6.61$; the 95% confidence interval is $(1019 - 1013) \pm 12.96$ *or* 6 ± 12.96.

(ii) No. The 95% confidence interval includes $\mu_{prep} - \mu_{non-prep} = 0$.

(d) (i) Let X denote the change in the test score. The 95% confidence interval for μ_X is $\bar{X} \pm 1.96\, SE(\bar{X})$, where $\text{SE}(\bar{X}) = \frac{60}{\sqrt{453}} = 2.82$; thus, the confidence interval is 9 ± 5.52.

(ii) Yes. The 95% confidence interval does not include $\mu_X = 0$.

(iii) Randomly select n students who have taken the test only one time. Randomly select one half of these students and have them take the prep course. Administer the test again to all of the n students. Compare the gain in performance of the prep-course second-time test takers to the non-prep-course second-time test takers.

3.17. (a) The 95% confidence interval is $\bar{Y}_{m,2008} - \bar{Y}_{m,1992} \pm 1.96\,\text{SE}(\bar{Y}_{m,2008} - \bar{Y}_{m,1992})$ where $\text{SE}(\bar{Y}_{m,2008} - \bar{Y}_{m,1992}) = \sqrt{\frac{s_{m,2008}^2}{n_{m,2008}} + \frac{s_{m,1992}^2}{n_{m,1992}}} = \sqrt{\frac{11.78^2}{1838} + \frac{10.17^2}{1594}} = 0.37$; the 95% confidence interval is $(24.98 - 23.27) \pm 0.73$ or 1.71 ± 0.73.

(b) The 95% confidence interval is $\bar{Y}_{w,2008} - \bar{Y}_{w,1992} \pm 1.96\,\text{SE}(\bar{Y}_{w,2008} - \bar{Y}_{w,1992})$ where $\text{SE}(\bar{Y}_{w,2008} - \bar{Y}_{w,1992}) = \sqrt{\frac{s_{w,2008}^2}{n_{w,2008}} + \frac{s_{w,1992}^2}{n_{w,1992}}} = \sqrt{\frac{9.66^2}{1871} + \frac{7.78^2}{1368}} = 0.31$; the 95% confidence interval is $(20.87 - 20.05) \pm 0.60$ or 0.82 ± 0.60.

(c) The 95% confidence interval is $(\bar{Y}_{m,2004} - \bar{Y}_{m,1992}) - (\bar{Y}_{w,2004} - \bar{Y}_{w,1992}) \pm 1.96\,\mathrm{SE}[(\bar{Y}_{m,2008} - \bar{Y}_{m,1992}) - (\bar{Y}_{w,2008} - \bar{Y}_{w,1992})]$, where $\mathrm{SE}[(\bar{Y}_{m,2008} - \bar{Y}_{m,1992}) - (\bar{Y}_{w,2008} - \bar{Y}_{w,1992})] = \sqrt{\frac{s^2_{m,2008}}{n_{m,2008}} + \frac{s^2_{m,1992}}{n_{m,1993}} + \frac{s^2_{w,2008}}{n_{w,2008}} + \frac{s^2_{w,1992}}{n_{w,1992}}} = \sqrt{\frac{11.78^2}{1838} + \frac{10.17^2}{1594} + \frac{9.66^2}{1871} + \frac{7.78^2}{1368}} = 0.48$. The 95% confidence interval is (24.98-23.27) − (20.87−20.05) ± 1.96 × 0.48 or 0.89 ± 0.95.

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

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

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

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

3.19. (a) No. $E(Y_i^2) = \sigma_Y^2 + \mu_Y^2$ and $E(Y_iY_j) = \mu_Y^2$ for $i \neq j$. Thus

$$E(\bar{Y}^2) = E\left(\frac{1}{n}\sum_{i=1}^n Y_i\right)^2 = \frac{1}{n^2}\sum_{i=1}^n E(Y_i^2) + \frac{1}{n^2}\sum_{i=1}^n \sum_{j \neq i} E(Y_iY_j) = \mu_Y^2 + \frac{1}{n}\sigma_Y^2$$

(b) Yes. If $\bar{Y}$ gets arbitrarily close to μ_Y with probability approaching 1 as n gets large, then $\bar{Y}^2$ gets arbitrarily close to μ_Y^2 with probability approaching 1 as n gets large. (As it turns out, this is an example of the "continuous mapping theorem" discussed in Chapter 17.)

3.20. Using analysis like that in equation (3.29)

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

because $\bar{X} \xrightarrow{p} \mu_X$ and $\bar{Y} \xrightarrow{p} \mu_Y$ the final term converges in probability to zero.

Let $W_i = (X_i - \mu_x)(Y_i - \mu_Y)$. Note W_i is *iid* with mean σ_{XY} and second moment $E[(X_i - \mu_X)^2(Y_i - \mu_Y)^2]$. But $E[(X_i - \mu_X)^2(Y_i - \mu_Y)^2] \le \sqrt{E(X_i - \mu_X)^4}\sqrt{E(Y_i - \mu_Y)^4}$ from the Cauchy-Schwartz inequality. Because X and Y have finite fourth moments, the second moment of W_i is finite, so that it has finite variance. Thus $\frac{1}{n}\sum_{i=1}^{n} W_i \xrightarrow{p} E(W_i) = \sigma_{XY}$. Thus, $s_{XY} \xrightarrow{p} \sigma_{XY}$ (because the term $\frac{n}{n-1} \to 1$).

3.21. Set $n_m = n_w = n$, and use equation (3.19) write the squared SE of $\bar{Y}_m - \bar{Y}_w$ as

$$[SE(\bar{Y}_m - \bar{Y}_w)]^2 = \frac{\frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{mi} - \bar{Y}_m)^2}{n} + \frac{\frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{wi} - \bar{Y}_w)^2}{n}$$

$$= \frac{\sum_{i=1}^{n}(Y_{mi} - \bar{Y}_m)^2 + \sum_{i=1}^{n}(Y_{wi} - \bar{Y}_w)^2}{n(n-1)}.$$

Similarly, using equation (3.23)

$$[SE_{pooled}(\bar{Y}_m - \bar{Y}_w)]^2 = \frac{\frac{1}{2(n-1)}\left[\sum_{i=1}^{n}(Y_{mi} - \bar{Y}_m)^2 + \frac{1}{(n-1)}\sum_{i=1}^{n}(Y_{wi} - \bar{Y}_w)^2\right]}{2n}$$

$$= \frac{\sum_{i=1}^{n}(Y_{mi} - \bar{Y}_m)^2 + \sum_{i=1}^{n}(Y_{wi} - \bar{Y}_w)^2}{n(n-1)}.$$

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) Using the formula for $\hat{\beta}_0$ in Equation (4.8), we know the sample average of the test scores across the 100 classrooms is

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

(d) Use the formula for the standard error of the regression (SER) in Equation (4.19) 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 (4.16) 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. 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) We have the following relations: $1\,in = 2.54\,cm$ and $1\,lb = 0.4536\,kg.$ Suppose the regression equation in centimeter-kilogram units is

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

The coefficients are $\hat{\gamma}_0 = -99.41 \times 0.4536 = -45.092\,kg$; $\hat{\gamma}_1 = 3.94 \times \frac{0.4536}{2.54} = 0.7036\,kg$ per cm. The R^2 is unit free, so it remains at $R^2 = 0.81$. The standard error of the regression is $SER = 10.2 \times 0.4536 = 4.6267\,kg.$

4.3. (a) The coefficient 9.6 shows the marginal effect of *Age* on *AWE*; that is, *AWE* is expected to increase by \$9.6 for each additional year of age. 696.7 is the intercept of the regression line. It determines the overall level of the line.

(b) *SER* is in the same units as the dependent variable (Y, or *AWE* in this example). Thus *SER* is measured in dollars per week.

(c) R^2 is unit free.

(d) (i) $696.7 + 9.6 \times 25 = \$936.7$;

(ii) $696.7 + 9.6 \times 45 = \$1{,}128.7$

(e) No. The oldest worker in the sample is 65 years old. 99 years is far outside the range of the sample data.

(f) No. The distribution of earning is positively skewed and has kurtosis larger than the normal.

(g) $\hat{\beta}_0 = \overline{Y} - \hat{\beta}_1 \overline{X}$, so that $\overline{Y} = \hat{\beta}_0 + \hat{\beta}_1 \overline{X}$. Thus the sample mean of *AWE* is 696.7 + 9.6 × 41.6 = \$1,096.06.

4.4. (a) $(R - R_f) = \beta(R_m - R_f) + u$, so that $\text{var}(R - R_f) = \beta^2 \times \text{var}(R_m - R_f) + \text{var}(u) + 2\beta \times \text{cov}(u, R_m - R_f)$. But $\text{cov}(u, R_m - R_f) = 0$, thus $\text{var}(R - R_f) = \beta^2 \times \text{var}(R_m - R_f) + \text{var}(u)$. With $\beta > 1$, $\text{var}(R - R_f) > \text{var}(R_m - R_f)$, follows because $\text{var}(u) \geq 0$.

(b) Yes. Using the expression in (a) $\text{var}(R - R_f) - \text{var}(R_m - R_f) = (\beta^2 - 1) \times \text{var}(R_m - R_f) + \text{var}(u)$, which will be positive if $\text{var}(u) > (1 - \beta^2) \times \text{var}(R_m - R_f)$.

(c) $R_m - R_f = 7.3\% - 3.5\% = 3.8\%$. Thus, the predicted returns are

$$\hat{R} = R_f + \hat{\beta}(R_m - R_f) = 3.5\% + \hat{\beta} \times 3.8\%$$

Wal-Mart: $3.5\% + 0.3 \times 3.8\% = 4.6\%$
Kellogg: $3.5\% + 0.5 \times 3.8\% = 5.4\%$
Waste Management: $3.5\% + 0.6 \times 3.8\% = 5.8\%$
Verizon: $3.5\% + 0.6 \times 3.8\% = 5.8\%$
Microsoft: $3.5\% + 1.0 \times 3.8\% = 7.3\%$
Best Buy: $3.5\% + 1.3 \times 3.8\% = 8.4\%$
Bank of America: $3.5\% + 2.4 \times 3.8\% = 11.9\%$

4.5. (a) u_i represents factors other than time that influence the student's performance on the exam including amount of time studying, aptitude for the material, and so forth. Some students will have studied more than average, other less; some students will have higher than average aptitude for the subject, others lower, and so forth.

(b) Because of random assignment u_i is independent of X_i. Since u_i represents deviations from average $E(u_i) = 0$. Because u and X are independent $E(u_i|X_i) = E(u_i) = 0$.

(c) (2) is satisfied if this year's class is typical of other classes, that is, students in this year's class can be viewed as random draws from the population of students that enroll in the class. (3) is satisfied because $0 \leq Y_i \leq 100$ and X_i can take on only two values (90 and 120).

(d) (i) $49 + 0.24 \times 90 = 70.6$; $49 + 0.24 \times 120 = 77.8$; $49 + 0.24 \times 150 = 85.0$

(ii) $0.24 \times 10 = 2.4$.

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

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

4.7. The expectation of $\hat{\beta}_0$ is obtained by taking expectations of both sides of Equation (4.8):

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

where the third equality in the above equation has used the facts that $E(u_i) = 0$ and $E[(\hat{\beta}_1 - \beta_1)\bar{X}] = E[(E(\hat{\beta}_1 - \beta_1)|\bar{X})\bar{X}] =$ because $E[(\beta_1 - \hat{\beta}_1)|\bar{X}] = 0$ (see text equation (4.31).)

4.8. The only change is that the mean of $\hat{\beta}_0$ is now $\beta_0 + 2$. An easy way to see this is this is to write the regression model as

$$Y_i = (\beta_0 + 2) + \beta_1 X_i + (u_i - 2).$$

The new regression error is $(u_i - 2)$ and the new intercept is $(\beta_0 + 2)$. All of the assumptions of Key Concept 4.3 hold for this regression model.

4.9. (a) With $\hat{\beta}_1 = 0,\ \hat{\beta}_0 = \bar{Y},$ and $\hat{Y}_i = \hat{\beta}_0 = \bar{Y}.$ Thus ESS = 0 and $R^2 = 0$.

(b) If $R^2 = 0$, then ESS = 0, so that $\hat{Y}_i = \bar{Y}$ for all i. But $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i,$ so that $\hat{Y}_i = \bar{Y}$ for all i, which implies that $\hat{\beta}_1 = 0,$ or that X_i is constant for all i. If X_i is constant for all i, then $\sum_{i-1}^{n}(X_i - \bar{X})^2 = 0$ and $\hat{\beta}_1$ is undefined (see equation (4.7)).

4.10. (a) $E(u_i|X = 0) = 0$ and $E(u_i|X = 1) = 0$. (X_i, u_i) are i.i.d. so that (X_i, Y_i) are i.i.d. (because Y_i is a function of X_i and u_i). X_i is bounded and so has finite fourth moment; the fourth moment is non-zero because $\Pr(X_i = 0)$ and $\Pr(X_i = 1)$ are both non-zero so that X_i has finite, non-zero kurtosis. Following calculations like those exercise 2.13, u_i also has nonzero finite fourth moment.

(b) $\text{var}(X_i) = 0.2 \times (1 - 0.2) = 0.16$ and $\mu_X = 0.2.$ Also

$$\begin{aligned} \text{var}[(X_i - \mu_X)u_i] &= E[(X_i - \mu_X)u_i]^2 \\ &= E[(X_i - \mu_X)u_i|X_i = 0]^2 \times \Pr(X_i = 0) + E[(X_i - \mu_X)u_i|X_i = 1]^2 \times \Pr(X_i = 1) \end{aligned}$$

where the first equality follows because $E[(X_i - \mu_X)u_i] = 0$, and the second equality follows from the law of iterated expectations.

$$E[(X_i - \mu_X)u_i|X_i = 0]^2 = 0.2^2 \times 1, \text{ and } E[(X_i - \mu_X)u_i|X_i = 1]^2 = (1 - 0.2)^2 \times 4.$$

Putting these results together

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\frac{(0.2^2 \times 1 \times 0.8) + ((1 - 0.2)^2 \times 4 \times 0.2)}{0.16^2} = \frac{1}{n}21.25$$

4.11. (a) The least squares objective function is $\sum_{i=1}^{n}(Y_i - b_1X_i)^2$. Differentiating with respect to b_1 yields $\frac{\partial \sum_{i=1}^{n}(Y_i - b_1X_i)^2}{\partial b_1} = -2\sum_{i=1}^{n} X_i(Y_i - b_1X_i)$. Setting this zero, and solving for the least squares estimator yields $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_iY_i}{\sum_{i=1}^{n} X_i^2}$.

(b) Following the same steps in (a) yields $\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i(Y_i - 4)}{\sum_{i=1}^{n} X_i^2}$.

4.12. (a) Write

$$ESS = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{\beta}_0 + \hat{\beta}_1X_i - \bar{Y})^2 = \sum_{i=1}^{n}[\hat{\beta}_1(X_i - \bar{X})]^2$$
$$= \hat{\beta}_1^2\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{\left[\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\right]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

This implies

$$R^2 = \frac{ESS}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} = \frac{\left[\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})\right]^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$
$$= \left[\frac{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}\right]^2$$
$$= \left[\frac{s_{XY}}{s_Xs_Y}\right]^2 = r_{XY}^2$$

(b) This follows from part (a) because $r_{XY} = r_{YX}$.

(c) Because $r_{XY} = \frac{s_{XY}}{s_Xs_Y}$, $r_{XY}\frac{s_Y}{s_X} = \frac{s_{XY}}{s_X^2} = \frac{\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\frac{1}{(n-1)}\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \hat{\beta}_1$

4.13. The answer follows the derivations in Appendix 4.3 in "Large-Sample Normal Distribution of the OLS Estimator." In particular, the expression for ν_i is now $\nu_i = (X_i - \mu_X)\kappa u_i$, so that var($\nu_i$) = κ^3var$[(X_i - \mu_X)u_i]$, and the term κ^2 carry through the rest of the calculations.

4.14. Because $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$, $\bar{Y} = \hat{\beta}_0 + \beta_1\bar{X}$. The sample regression line is $y = \hat{\beta}_0 + \beta_1x$, so that the sample regression line passes through ($\bar{X}, \bar{Y}$).

Chapter 5
Regression with a Single Regressor: Hypothesis Tests and Confidence Intervals

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

(b) Calculate the t-statistic:

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

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

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

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.

(c) The t-statistic is

$$t^{act} = \frac{\hat{\beta}_1 - (-5.6)}{\text{SE}(\hat{\beta}_1)} = \frac{0.22}{2.21} = 0.10$$

The p-value for the test $H_0: \beta_1 = -5.6$ vs. $H_1: \beta_1 \neq -5.6$ is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-0.10) = 0.92$$

thatc —)

The p-value is larger than 0.10, so we cannot reject the null hypothesis at the 10%, 5% or 1% significance level. Because $\beta_1 = -5.6$ is not rejected at the 5% level, this value is contained in the 95% confidence interval.

(d) The 99% confidence interval for β_0 is $\{520.4 \pm 2.58 \times 20.4\}$, that is, $467.7 \le \beta_0 \le 573.0$.

5.2. (a) The estimated gender gap equals \$2.12/hour.

(b) The null and alternative hypotheses are $H_0: \beta_1 = 0$ vs. $H_1: \beta_1 \neq 0$. The t-statistic is

$$t^{act} = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)} = \frac{2.12}{0.36} = 5.89,$$

and the p-value for the test is

$$p\text{-value} = 2\Phi(-|t^{act}|) = 2\Phi(-5.89) = 2\times 0.0000 = 0.000 \text{ (to four decimal places)}$$

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 $\{2.12 \pm 1.96\times 0.36\}$, that is, $1.41 \le \beta_1 \le 2.83$.

(d) The sample average wage of women is $\hat{\beta}_0 = \$12.52/\text{hour}$. The sample average wage of men is $\hat{\beta}_0 + \hat{\beta}_1 = \$12.52 + \$2.12 = \$14.64/\text{hour}$.

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

$$Wage = \beta_0 + \beta_1 Male + u_i,$$

or

$$Wage = \gamma_0 + \gamma_1 Female + v_i.$$

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:

$$\gamma_0 = \beta_0 + \beta_1,$$
$$\gamma_0 + \gamma_1 = \beta_0.$$

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

$$\hat{\gamma}_0 = \hat{\beta}_0 + \hat{\beta}_1 = 14.64,$$
$$\hat{\gamma}_1 = \hat{\beta}_0 - \hat{\gamma}_0 = -\hat{\beta}_1 = -2.12.$$

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 $SER = \left(\frac{SSR}{n-1}\right)^{1/2}$ and $R^2 = 1 - \frac{SSR}{TSS}$ are unchanged.

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

$$\widehat{Wages} = 14.64 - 2.12 Female, \quad R^2 = 0.06, \quad SER = 4.2.$$

5.3. The 99% confidence interval is 1.5 × {3.94 ± 2.58 × 0.31) or 4.71 lbs ≤ WeightGain ≤ 7.11 lbs.

5.4. (a) – 5.38 + 1.76 × 16 – $22.78 per hour

(b) The wage is expected to increase by 1.76 × 2 = $3.52 per hour.

(c) The increase in wages for college education is $\beta_1 \times 4$. Thus, the counselor's assertion is that $\beta_1 = 10/4 = 2.50$. The t-statistic for this null hypothesis is $t^{act} = \frac{1.76 - 2.50}{0.08} - 9.25$, which has a p-value of 0.00. Thus, the counselor's assertion can be rejected at the 1% significance level. A 95% confidence for $\beta_1 \times 4$ is $4 \times (1.76 \pm 1.96 \times 0.08)$ or $\$6.41 \leq \text{Gain} \leq \$7.67$.

检验

5.5 (a) The estimated gain from being in a small class is 13.9 points. This is equal to approximately 1/5 of the standard deviation in test scores, a moderate increase. ① compared to mean of Y — 13.9/92.5 ≈
② compared to SD of Y — 20%

(b) The t-statistic is $t^{act} = \frac{13.9}{2.5} = 5.56$, which has a p-value of 0.00. Thus the null hypothesis is rejected at the 5% (and 1%) level.

(c) $13.9 \pm 2.58 \times 2.5 = 13.9 \pm 6.45$.

5.6. (a) The question asks whether the variability in test scores in large classes is the same as the variability in small classes. It is hard to say. On the one hand, teachers in small classes might able so spend more time bringing all of the students along, reducing the poor performance of particularly unprepared students. On the other hand, most of the variability in test scores might be beyond the control of the teacher.

(b) The formula in 5.3 is valid for heteroskesdasticity or homoskedasticity; thus inferences are valid in either case.

5.7. (a) The t-statistic is $\frac{3.2}{1.5} = 2.13$ with a p-value of 0.03; since the p-value is less than 0.05, the null hypothesis is rejected at the 5% level.

(b) $3.2 \pm 1.96 \times 1.5 = 3.2 \pm 2.94$

(c) Yes. If Y and X are independent, then $\beta_1 = 0$; but this null hypothesis was rejected at the 5% level in part (a).

(d) β_1 would be rejected at the 5% level in 5% of the samples; 95% of the confidence intervals would contain the value $\beta_1 = 0$.

5.8. (a) $43.2 \pm 2.05 \times 10.2$ or 43.2 ± 20.91, where 2.05 is the 5% two-sided critical value from the t_{28} distribution.

(b) The t-statistic is $t^{act} = \frac{61.5-55}{7.4} = 0.88$, which is less (in absolute value) than the critical value of 20.5. Thus, the null hypothesis is not rejected at the 5% level.

(c) The one sided 5% critical value is 1.70; t^{act} is less than this critical value, so that the null hypothesis is not rejected at the 5% level.

5.9. (a) $\bar{\beta} = \frac{\frac{1}{n}(Y_1 + Y_2 + \cdots + Y_n)}{\bar{X}}$ so that it is linear function of $Y_1, Y_2, \ldots, Y_n$.

(b) $E(Y_i|X_1, \ldots, X_n) = \beta X_i$, thus

$$E(\bar{\beta}|X_1,\ldots,X_n) = E\left[\left(\frac{\frac{1}{n}(Y_1+Y_2+\cdots+Y_n)}{\bar{X}}\right)|(X_1,\ldots,X_n)\right]$$

$$= \frac{\frac{1}{n}\beta(X_1+\cdots+X_n)}{\bar{X}} = \beta.$$

5.10. Let n_0 denote the number of observation with $X = 0$ and n_1 denote the number of observations with $X = 1$; note that $\sum_{i=1}^{n} X_i = n_1$; $\bar{X} = \frac{n_1}{n}$; $\frac{1}{n_1}\sum_{i=1}^{n} X_iY_i = \bar{Y}_1$; $\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2 = n_1 - \frac{n_1^2}{n} = n_1\left(1 - \frac{n_1}{n}\right) = \frac{n_1 n_0}{n}$; $n_1\bar{Y}_1 + n_0\bar{Y}_0 = \sum_{i=1}^{n} Y_i$, so that $\bar{Y} = \frac{n_1}{n}\bar{Y}_1 + \frac{n_0}{n}\bar{Y}_0$

From the least squares formula

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{\sum_{i=1}^{n}X_i(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} = \frac{\sum_{i=1}^{n}X_iY_i - \bar{Y}n_1}{n_1n_0/n}$$

$$= \frac{n}{n_0}(\bar{Y}_1 - \bar{Y}) = \frac{n}{n_0}\left(\bar{Y} - \frac{n_1}{n}\bar{Y}_1 - \frac{n_0}{n}\bar{Y}_0\right) = \bar{Y}_1 - \bar{Y}_0$$

and $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X} = \left(\frac{n_0}{n}\bar{Y}_0 + \frac{n_1}{n}\bar{Y}_1\right) - (\bar{Y}_1 - \bar{Y}_0)\frac{n_1}{n} = \frac{n_1+n_0}{n}\bar{Y}_0 = \bar{Y}_0$

5.11. Using the results from 5.10, $\hat{\beta}_0 = \bar{Y}_m$ and $\hat{\beta}_1 = \bar{Y}_w - \bar{Y}_m$. From Chapter 3, $\mathrm{SE}(\bar{Y}_m) = \frac{s_m}{\sqrt{n_m}}$ and $\mathrm{SE}(\bar{Y}_w - \bar{Y}_m) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}$. Plugging in the numbers $\hat{\beta}_0 = 523.1$ and $\mathrm{SE}(\hat{\beta}_0) = 6.22$; $\hat{\beta}_1 = -38.0$ and $\mathrm{SE}(\hat{\beta}_1) = 7.65$.

5.12. Equation (4.22) gives

$$\sigma_{\hat{\beta}_0}^2 = \frac{\mathrm{var}(H_iu_i)}{n\left[E\left(H_i^2\right)\right]^2}, \quad \text{where } H_i = 1 - \frac{\mu_X}{E\left(X_i^2\right)}X_i.$$

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

$$E(H_iu_i) = E\left(u_i - \frac{\mu_x}{E\left(X_i^2\right)}X_iu_i\right) = E(u_i) - \frac{\mu_x}{E\left(X_i^2\right)}E[X_iE(u_i|X_i)]$$

$$= 0 - \frac{\mu_x}{E\left(X_i^2\right)} \times 0 = 0,$$

and

$$E[(H_iu_i)^2] = E\left\{\left(u_i - \frac{\mu_X}{E\left(X_i^2\right)}X_iu_i\right)^2\right\}$$

$$= E\left\{u_i^2 - 2\frac{\mu_X}{E\left(X_i^2\right)}X_iu_i^2 + \left[\frac{\mu_X}{E\left(X_i^2\right)}\right]^2 X_i^2u_i^2\right\}$$

$$= E\left(u_i^2\right) - 2\frac{\mu_X}{E\left(X_i^2\right)}E\left[X_iE\left(u_i^2|X_i\right)\right] + \left[\frac{\mu_X}{E\left(X_i^2\right)}\right]^2 E\left[X_i^2E\left(u_i^2|X_i\right)\right]$$

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

Because $E(H_iu_i) = 0$, $\text{var}(H_iu_i) = E[(H_iu_i)^2]$, so

$$\text{var}(H_iu_i) = E[(H_iu_i)^2] = \left(1 - \frac{\mu_X^2}{E\left(X_i^2\right)}\right)\sigma_u^2.$$

Also

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

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

Thus

$$\sigma_{\hat\beta_0}^2 = \frac{\text{var}(H_iu_i)}{\left[nE\left(H_i^2\right)^2\right]} = \frac{\left(1 - \frac{\mu_X^2}{E\left(X_i^2\right)}\right)\sigma_u^2}{n\left(1 - \frac{\mu_X^2}{E\left(X_i^2\right)}\right)^2} = \frac{\sigma_u^2}{n\left(1 - \frac{\mu_X^2}{E\left(X_i^2\right)}\right)}$$

$$= \frac{E(X_i^2)\sigma_u^2}{n[E\left(X_i^2 - \mu_X^2\right)]} = \frac{E\left(X_i^2\right)\sigma_u^2}{n\sigma_X^2}.$$

5.13. (a) Yes, this follows from the assumptions in KC 4.3.

(b) Yes, this follows from the assumptions in KC 4.3 and conditional homoskedasticity

(c) They would be unchanged for the reasons specified in the answers to those questions.

(d) (a) is unchanged; (b) is no longer true as the errors are not conditionally homosckesdastic.

5.14. (a) From Exercise (4.11), $\hat\beta = \sum a_iY_i$ where $a_i = \frac{X_i}{\sum_{j=1}^n X_j^2}$. Since the weights depend only on X_i but not on Y_i, $\hat\beta$ is a linear function of *Y*.

(b) $E(\hat{\beta}|X_1,\ldots,X_n)=\beta+\frac{\sum_{i=1}^{n}X_iE(u_i|X_1,\ldots,X_n)}{\sum_{i=1}^{n}X_j^2}=\beta$ since $E(u_i|X_1,\ldots,X_n)=0$

(c) $\mathrm{var}(\hat{\beta}|X_1,\ldots,X_n)=\frac{\sum_{i=1}^{n}X_i^2\,\mathrm{var}(u_i|X_1,\ldots,X_n)}{\left[\sum_{i=1}^{n}X_j^2\right]^2}=\frac{\sigma^2}{\sum_{i=1}^{n}X_j^2}$

(d) This follows the proof in the appendix.

5.15. Because the samples are independent, $\hat{\beta}_{m,1}$ and $\hat{\beta}_{w,1}$ are independent. Thus $\mathrm{var}(\hat{\beta}_{m,1}-\hat{\beta}_{w,1})=\mathrm{var}(\hat{\beta}_{m,1})+\mathrm{var}(\hat{\beta}_{w,1})$. $\mathrm{Var}(\hat{\beta}_{m,1})$ is consistently estimated as $[\mathrm{SE}(\hat{\beta}_{m,1})]^2$ and $\mathrm{Var}(\hat{\beta}_{w,1})$ is consistently estimated as $[\mathrm{SE}(\hat{\beta}_{w,1})]^2$, so that $\mathrm{var}(\hat{\beta}_{m,1}-\hat{\beta}_{w,1})$ is consistently estimated by $[\mathrm{SE}(\hat{\beta}_{m,1})]^2+[\mathrm{SE}(\hat{\beta}_{w,1})]^2$, and the result follows by noting the SE is the square root of the estimated variance.

Chapter 6
Linear Regression with Multiple Regressors

6.1. By equation (6.15) in the text, we know

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

Thus, that values of $\bar{R}^2$ are 0.175, 0.189, and 0.193 for columns (1)–(3).

6.3. (a) On average, a worker earns \$0.29/hour more for each year he ages.

(b) Sally's earnings prediction is $4.40 + 5.48 \times 1 - 2.62 \times 1 + 0.29 \times 29 = 15.67$ dollars per hour. Betsy's earnings prediction is $4.40 + 5.48 \times 1 - 2.62 \times 1 + 0.29 \times 34 = 17.12$ dollars per hour. The difference is 1.45

6.4. (a) Workers in the Northeast earn \$0.69 more per hour than workers in the West, on average, controlling for other variables in the regression. Workers in the Northeast earn \$0.60 more per hour than workers in the West, on average, controlling for other variables in the regression. Workers in the South earn \$0.27 less than workers in the West.

(b) The regressor *West* is omitted to avoid perfect multicollinearity. If *West* is included, then the intercept can be written as a perfect linear function of the four regional regressors.

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

6.5. (a) \$23,400 (recall that *Price* is measured in \$1000s).

(b) In this case $\Delta BDR = 1$ and $\Delta Hsize = 100$. The resulting expected change in price is 23.4 + 0.156 × 100 = 39.0 thousand dollars or \$39,000.

(c) The loss is \$48,800.

(d) From the text $\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2)$, so $R^2 = 1 - \frac{n-k-1}{n-1}(1-\bar{R}^2)$, thus, $R^2 = 0.727$.

6.6. (a) There are other important determinants of a country's crime rate, including demographic characteristics of the population.

(b) Suppose that the crime rate is positively affected by the fraction of young males in the population, and that counties with high crime rates tend to hire more police. In this case, the size of the police force is likely to be positively correlated with the fraction of young males in the population leading to a positive value for the omitted variable bias so that $\hat{\beta}_1 > \beta_1$.

6.7. (a) The proposed research in assessing the presence of gender bias in setting wages is too limited. There might be some potentially important determinants of salaries: type of engineer, amount of work experience of the employee, and education level. The gender with the lower wages could reflect the type of engineer among the gender, the amount of work experience of the employee, or the education level of the employee. The research plan could be improved with the collection of additional data as indicated and an appropriate statistical technique for analyzing the data would be a multiple regression in which the dependent variable is wages and the independent variables would include a dummy variable for gender, dummy variables for type of engineer, work experience (time units), and education level (highest grade level completed). The potential importance of the suggested omitted variables makes a "difference in means" test inappropriate for assessing the presence of gender bias in setting wages.

(b) The description suggests that the research goes a long way towards controlling for potential omitted variable bias. Yet, there still may be problems. Omitted from the analysis are characteristics associated with behavior that led to incarceration (excessive drug or alcohol use, gang activity, and so forth), that might be correlated with future earnings. Ideally, data on these variables should be included in the analysis as additional control variables.

6.8. Omitted from the analysis are reasons *why* the survey respondents slept more or less than average. People with certain chronic illnesses might sleep more than 8 hours per night. People with other illnesses might sleep less than 5 hours. This study says nothing about the *causal* effect of sleep on mortality.

6.9. For omitted variable bias to occur, two conditions must be true: X_1 (the included regressor) is correlated with the omitted variable, and the omitted variable is a determinant of the dependent variable. Since X_1 and X_2 are uncorrelated, the estimator of β_1 does not suffer from omitted variable bias.

6.10. (a) $\sigma^2_{\hat{\beta}_1} = \frac{1}{n}\left[\frac{1}{1-\rho^2_{X_1,X_2}}\right]\frac{\sigma^2_u}{\sigma^2_{X_1}}$

Assume X_1 and X_2 are uncorrelated: $\rho^2_{X_1X_2} = 0$

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{400}\left[\frac{1}{1-0}\right]\frac{4}{6}$$
$$= \frac{1}{400}\cdot\frac{4}{6} = \frac{1}{600} = 0.00167$$

(b) With $\rho_{X_1,X_2} = 0.5$

$$\sigma^2_{\hat{\beta}_1} = \frac{1}{400}\left[\frac{1}{1-0.5^2}\right]\frac{4}{6}$$
$$= \frac{1}{400}\left[\frac{1}{0.75}\right]\frac{4}{6} = .0022$$

(c) The statement correctly says that the larger is the correlation between X_1 and X_2 the larger is the variance of $\hat{\beta}_1$, however the recommendation "it is best to leave X_2 out of the regression" is incorrect. If X_2 is a determinant of Y, then leaving X_2 out of the regression will lead to omitted variable bias in $\hat{\beta}_1$.

6.11. (a) $\sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2$

(b) $\dfrac{\partial \sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_1} = -2\sum X_{1i}(Y_i - b_1 X_{1i} - b_2 X_{2i})$

$\dfrac{\partial \sum (Y_i - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_2} = -2\sum X_{2i}(Y_i - b_1 X_{1i} - b_2 X_{2i})$

(c) From (b), $\hat{\beta}_1$ satisfies

$$\sum X_{1i}(Y_i - \hat{\beta}_1 X_{1i} - \hat{\beta}_1 X_{2i}) = 0$$

or

$$\hat{\beta}_1 = \frac{\sum X_{1i} Y_i - \hat{\beta}_2 \sum X_{1i} X_{2i}}{\sum X_{1i}^2}$$

and the result follows immediately.

(d) Following analysis as in (c)

$$\hat{\beta}_2 = \frac{\sum X_{2i} Y_i - \hat{\beta}_1 \sum X_{1i} X_{2i}}{\sum X_{2i}^2}$$

and substituting this into the expression for $\hat{\beta}_1$ in (c) yields

$$\hat{\beta}_1 = \frac{\sum X_{1i} Y \frac{\sum X_{2i} Y_i - \hat{\beta}_1 \sum X_{1i} X_{2i}}{\sum X_{2i}^2} \sum X_{1i} X_{2i}}{\sum X_{1i}^2}.$$

Solving for $\hat{\beta}_1$ yields:

$$\hat{\beta}_1 = \frac{\sum X_{2i}^2 \sum X_{1i} Y_i - \sum X_{1i} X_{2i} \sum X_{2i} Y_i}{\sum X_{1i}^2 \sum X_{2i}^2 - (\sum X_{1i} X_{2i})^2}$$

(e) The least squares objective function is $\sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2$ and the partial derivative with respect to b_0 is

$$\frac{\partial \sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i})^2}{\partial b_0} = -2\sum (Y_i - b_0 - b_1 X_{1i} - b_2 X_{2i}).$$

Setting this to zero and solving for $\hat{\beta}_0$ yields: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$.

(f) Substituting $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$ into the least squares objective function yields $\sum (Y_i - \hat{\beta}_0 - b_1 X_{1i} - b_2 X_{2i})^2 = \sum \left((Y_i - \bar{Y}) - b_1 (X_{1i} - \bar{X}_1) - b_2 (X_{2i} - \bar{X}_2) \right)^2$, which is identical to the least squares objective function in part (a), except that all variables have been replaced with deviations from sample means. The result then follows as in (c).

Notice that the estimator for β_1 is identical to the OLS estimator from the regression of Y onto X_1, omitting X_2. Said differently, when $\sum (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) = 0$, the estimated coefficient on X_1 in the OLS regression of Y onto both X_1 and X_2 is the same as estimated coefficient in the OLS regression of Y onto X_1.

Chapter 7
Hypothesis Tests and Confidence Intervals in Multiple Regression

7.1 and 7.2

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)
Ntheast (X_4)			0.69* (0.30)
Midwest (X_5)			0.60* (0.28)
South (X_6)			– 0.27 (0.26)
Intercept	12.69** (0.14)	4.40** (1.05)	3.75** (1.06)

(a) The t-statistic is 5.46/0.21 = 26.0, which exceeds 1.96 in absolute value. Thus, the coefficient is statistically significant at the 5% level. The 95% confidence interval is 5.46 ± 1.96 × 0.21.

(b) t-statistic is – 2.64/0.20 = –13.2, and 13.2 > 1.96, so the coefficient is statistically significant at the 5% level. The 95% confidence interval is –2.64 ± 1.96 × 0.20.

7.3. (a) Yes, age is an important determinant of earnings. Using a t-test, the t-statistic is $0.29/0.04 = 7.25$, with a p-value of 4.2×10^{-13}, implying that the coefficient on age is statistically significant at the 1% level. The 95% confidence interval is 0.29 ± 1.96 × 0.04.

(b) $\Delta Age \times [0.29 \pm 1.96 \times 0.04] = 5 \times [0.29 \pm 1.96 \times 0.04] = 1.45 \pm 1.96 \times 0.20$ = \$1.06 to \$1.84

7.4. (a) The F-statistic testing the coefficients on the regional regressors are zero is 6.10. The 1% critical value (from the $F_{3,\infty}$ distribution) is 3.78. Because 6.10 > 3.78, the regional effects are significant at the 1% level.

(b) The expected difference between Juanita and Molly is $(X_{6,\text{Juanita}} - X_{6,\text{Molly}}) \times \beta_6 = \beta_6$. Thus a 95% confidence interval is –0.27 ± 1.96 × 0.26.

(c) The expected difference between Juanita and Jennifer is $(X_{5,\text{Juanita}} - X_{5,\text{Jennifer}}) \times \beta_5 + (X_{6,\text{Juanita}} - X_{6,\text{Jennifer}}) \times \beta_6 = -\beta_5 + \beta_6$. A 95% confidence interval could be constructed using the general methods discussed in Section 7.3. In this case, an easy way to do this is to omit *Midwest* from the regression and replace it with $X_5 = West$. In this new regression the coefficient on *South* measures the difference in wages between the *South* and the *Midwest*, and a 95% confidence interval can be computed directly.

7.5. The *t*-statistic for the difference in the college coefficients is $t = \dfrac{\hat{\beta}_{\text{college},1998} - \hat{\beta}_{\text{college},1992}}{\text{SE}(\hat{\beta}_{\text{college},1998} - \hat{\beta}_{\text{college},1992})}$. Because $\hat{\beta}_{\text{college},1998}$ and $\hat{\beta}_{\text{college},1992}$ are computed from independent samples, they are independent, which means that $\text{cov}(\hat{\beta}_{\text{college},1998}, \hat{\beta}_{\text{college},1992}) = 0$ Thus, $\text{var}(\hat{\beta}_{\text{college},1998} - \hat{\beta}_{\text{college},1992}) =$ $\text{var}(\hat{\beta}_{\text{college},1998}) + \text{var}(\hat{\beta}_{\text{college},1998})$. This implies that $\text{SE}(\hat{\beta}_{\text{college},1998} - \hat{\beta}_{\text{college},1992}) = (0.21^2 + 0.20^2)^{\frac{1}{2}}$. Thus, $t^{act} = \dfrac{5.48 - 5.29}{\sqrt{0.21^2 + 0.20^2}} = 0.6552$. There is no significant change since the calculated *t*-statistic is less than 1.96, the 5% critical value.

7.6. In isolation, these results do imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. Thus, 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 in the U.S. labor market.) 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.

7.7. (a) The *t*-statistic is $0.485/2.61 = 0.186 < 1.96$. Therefore, the coefficient on BDR is not statistically significantly different from zero.

(b) The coefficient on *BDR* measures the *partial effect* of the number of bedrooms holding house size (*Hsize*) constant. Yet, the typical 5-bedroom house is much larger than the typical 2-bedroom house. Thus, the results in (a) says little about the conventional wisdom.

(c) The 99% confidence interval for effect of lot size on price is $2000 \times [0.002 \pm 2.58 \times 0.00048]$ or 1.52 to 6.48 (in thousands of dollars).

(d) Choosing the scale of the variables should be done to make the regression results easy to read and to interpret. If the lot size were measured in thousands of square feet, the estimate coefficient would be 2 instead of 0.002.

(e) The 10% critical value from the $F_{2,\infty}$ distribution is 2.30. Because $0.08 < 2.30$, the coefficients are not jointly significant at the 10% level.

7.8. (a) Using the expressions for R^2 and $\bar{R}^2$, algebra shows that

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1}(1-R^2), \text{ so } R^2 = 1 - \frac{n-k-1}{n-1}(1-\bar{R}^2).$$

Column 1: $R^2 = 1 - \dfrac{420-1-1}{420-1}(1-0.049) = 0.051$

Column 2: $\bar{R}^2 = 1 - \frac{420-2-1}{420-1}(1-0.424) = 0.427$

Column 3: $\bar{R}^2 = 1 - \frac{420-3-1}{420-1}(1-0.773) = 0.775$

Column 4: $\bar{R}^2 = 1 - \frac{420-3-1}{420-1}(1-0.626) = 0.629$

Column 5: $\bar{R}^2 = 1 - \frac{420-4-1}{420-1}(1-0.773) = 0.775$

(b) $H_0 : \beta_3 = \beta_4 = 0$

$H_1 : \beta_3 \neq, \beta_4 \neq 0$

Unrestricted regression (Column 5): $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4,\ R^2_{\text{unrestricted}} = 0.775$

Restricted regression (Column 2): $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2,\ R^2_{\text{restricted}} = 0.427$

$$F_{\text{HomoskedasticityOnly}} = \frac{(R^2_{\text{unrestricted}} - R^2_{\text{restricted}})/q}{(1-R^2_{\text{unrestricted}})/(n-k_{\text{unrestricted}}-1)},\ n = 420,\ k_{\text{unrestricted}} = 4,\ q = 2$$

$$= \frac{(0.775-0.427)/2}{(1-0.775)/(420-4-1)} = \frac{0.348/2}{(0.225)/415} = \frac{0.174}{0.00054} = 322.22$$

5% Critical value form $F_{2,00} = 4.61$; $F_{\text{HomoskedasticityOnly}} = F_{2,00}$ so H_o is rejected at the 5% level.

(c) $t_3 = -13.921$ and $t_4 = 0.814$, $q = 2$; $|t_3| > c$ (Where $c = 2.807$, the 1% Benferroni critical value from Table 7.3). Thus the null hypothesis is rejected at the 1% level.

(d) $-1.01 \pm 2.58 \times 0.27$

7.9. (a) Estimate

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

and test whether $\gamma = 0$.

(b) Estimate

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

and test whether $\gamma = 0$.

(c) Estimate

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

and test whether $\gamma = 0$.

7.10. Because $R^2 = 1 - \frac{SSR}{TSS}$, $R^2_{\text{unrestricted}} - R^2_{\text{restricted}} = \frac{SSR_{\text{restricted}} - SSR_{\text{unrestricted}}}{TSS}$ and $1 - R^2_{\text{unrestricted}} = \frac{SSR_{\text{unrestricted}}}{TSS}$. Thus

$$F = \frac{(R^2_{\text{unrestricted}} - R^2_{\text{restricted}})/q}{(1-R^2_{\text{unrestricted}})/(n-k_{\text{unrestricted}}-1)} = \frac{\frac{SSR_{\text{restricted}} - SSR_{\text{unrestricted}}}{TSS}/q}{\frac{SSR_{\text{unrestricted}}}{TSS}/(n-k_{\text{unrestricted}}-1)}$$

$$= \frac{(SSR_{\text{restricted}} - SSR_{\text{unrestricted}})/q}{SSR_{\text{unrestricted}}/(n-k_{\text{unrestricted}}-1)}$$

7.11. (a) Treatment (assignment to small classes) was not randomly assigned in the population (the continuing and newly-enrolled students) because of the difference in the proportion of treated continuing and newly-enrolled students. Thus, the treatment indicator X_1 is correlated with X_2. If newly-enrolled students perform systematically differently on standardized tests than continuing students (perhaps because of adjustment to a new school), then this becomes part of the error term u in (a). This leads to correlation between X_1 and u, so that $E(u|X_1) \neq 0$. Because $E(u|X_1) \neq 0$, the $\hat{\beta}_1$ is biased and inconsistent.

(b) Because treatment was randomly assigned conditional on enrollment status (continuing or newly-enrolled), $E(u \mid X_1, X_2)$ will not depend on X_1. This means that the assumption of conditional mean independence is satisfied, and $\hat{\beta}_1$ is unbiased and consistent. However, because X_2 was not randomly assigned (newly-enrolled students may, on average, have attributes other than being newly enrolled that affect test scores), $E(u \mid X_1, X_2)$ may depend of X_2, so that $\hat{\beta}_2$ may be biased and inconsistent.

Chapter 8
Nonlinear Regression Functions

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

(b) When $Sales_{2010}$ = 205, the percentage increase is $100 \times \frac{205-196}{196} = 4.5918\%$ and the approximation is 100 × [ln (205) – ln (196)] = 4.4895%. When $Sales_{2010}$ = 250, the percentage increase is $100 \times \frac{250-196}{196} = 27.551\%$ and the approximation is 100 × [ln (250) – ln (196)] = 24.335%. When $Sales_{2010}$ = 500, the percentage increase is $100 \times \frac{500-196}{196} = 155.1\%$ and the approximation is 100 × [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.

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

(b) Because the regressions in columns (1) and (2) have the same dependent variable, $\bar{R}^2$ can be used 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 so use ln(*Size*) to explain house prices.

(c) The house price is expected to increase by 7.1% (= 100% × 0.071 × 1). The 95% confidence interval for this effect is 100% × (0.071 ± 1.96 × 0.034) = [0.436% to 13.764%].

(d) The house price is expected to increase by 0.36% (100% × 0.0036 × 1 = 0.36%) with an additional bedroom while other factors are held constant. The effect is not statistically significant at a 5% significance level: $|t| = \frac{0.0036}{0.037} = 0.09730 < 1.96$. Note that this coefficient measures the effect of an additional bedroom holding the size of the house constant. Thus, it measures the effect of converting existing space (from, say a family room) into a 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% × 0.071 × 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% × (0.071 × 1 + 0.0022 × 1)) 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: $|t| = \frac{0.0022}{0.10} = 0.022 < 1.96$.

8.3 (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 $20 \le STR \le 25$; 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.

8.4. (a) With 2 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.1032 \times 16) - (0.451 \times 0) + (0.0134 \times 0 \times 16) + (0.0134 \times 2) - 0.000211 \times 2^2)$$
$$-(0.095 \times 0) - (0.092 \times 0) - (0.023 \times 1) + 1.503 = 3.159$$

With 3 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.1032 \times 16) - (0.451 \times 0) + (0.0134 \times 0 \times 16) + (0.0134 \times 3) - 0.000211 \times 3^2)$$
$$-(0.095 \times 0) - (0.092 \times 0) - (0.023 \times 1) + 1.503 = 3.172$$

Difference = 3.172 – 3.159 = 0.013 (or 1.3%)

(b) With 10 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.1032 \times 16) - (0.451 \times 0) + (0.0134 \times 0 \times 16) + (0.0134 \times 10) - 0.000211 \times 10^2)$$
$$-(0.095 \times 0) - (0.092 \times 0) - (0.023 \times 1) + 1.503 = 3.253$$

With 11 years of experience, the man's expected *AHE* is

$$\widehat{\ln(AHE)} = (0.1032 \times 16) - (0.451 \times 0) + (0.0134 \times 0 \times 16) + (0.0134 \times 11) - 0.000211 \times 11^2)$$
$$-(0.095 \times 0) - (0.092 \times 0) - (0.023 \times 1) + 1.503 = 3.263$$

Difference 3.263 – 0.010 (or 1.0%)

(c) The regression in nonlinear in experience (it includes *Potential experience*^{2}).

(d) Let β_1 denote the coefficient on *Potential Experience* and β_2 denote the coefficient on (*Potential Experience*)^2. Then, following the discussion in the paragraphs just about equation (8.7) in the text, the expected change in part (a) is given by $\beta_1 + 5\beta_2$ and the expected change in (b) is given by $\beta_1 + 21\beta_2$. The difference between these, say (b) – (a), is $16\beta_2$. Because the estimated value of β_2 is significant at the 5% level (the *t*-statistic for $\hat{\beta}_2$ is $-0.000211/0.000023 = -9.2$), the difference between the effects in (a) and (b) ($=16\beta_2$) is significant at the 5% level.

(e) No. This would affect the level of ln(*AHE*), but not the change associated with another year of experience.

(f) Include interaction terms *Female* × *Potential experience* and *Female* × (*Potential experience*)^2.

8.5. (a) (1) The demand for older journals is less elastic than for younger journals because the interaction term between the log of journal age and price per citation is positive. (2) There is a linear relationship between log price and log of quantity follows because the estimated coefficients on log price squared and log price cubed are both insignificant. (3) The demand is greater for journals with more characters follows from the positive and statistically significant coefficient estimate on the log of characters.

(b) (i) The effect of ln(*Price per citation*) is given by [– 0.899 + 0.141 × ln(*Age*)] × ln(*Price per citation*). Using *Age* = 80, the elasticity is [– 0.899 + 0.141 × ln(80)] = – 0.28.

(ii) As described in equation (8.8) and the footnote on page 261, the standard error can be found by dividing 0.28, the absolute value of the estimate, by the square root of the F-statistic testing $\beta_{ln(Price\ per\ citation)} + \ln(80) \times \beta_{ln(Age)\times ln(Price\ per\ citation)} = 0$.

(c) $\ln\left(\frac{Characters}{a}\right) = \ln(Characters) - \ln(a)$ for any constant a. Thus, estimated parameter on *Characters* will not change and the constant (intercept) will change.

8.6. (a) (i) There are several ways to do this. Here is one. Create an indicator variable, say $DV1$, that equals one if *%Eligible* is greater than 20% and less than 50%. Create another indicator, say $DV2$, that equals one if *%Eligible* is greater than 50%. Run the regression:

$$TestScore = \beta_0 + \beta_1 \%Eligible + \beta_2 DV1 \times \%Eligible + \beta_3 DV2 \times \%Eligible + \text{other regressors}$$

The coefficient β_1 shows the marginal effect of *%Eligible* on *TestScores* for values of *%Eligible* $< 20\%$, $\beta_1 + \beta_2$ shows the marginal effect for values of *%Eligible* between 20% and 50% and $\beta_1 + \beta_3$ shows the marginal effect for values of *%Eligible* greater than 50%.

(ii) The linear model implies that $\beta_2 = \beta_3 = 0$, which can be tested using an F-test.

(b) (i) There are several ways to do this, perhaps the easiest is to include an interaction term *STR* × ln(*Income*) to the regression in column (7).

(ii) Estimate the regression in part (b.i) and test the null hypothesis that the coefficient on the interaction term is equal to zero.

8.7. (a) (i) ln(*Earnings*) for females are, on average, 0.44 lower for men than for women.

(ii) The error term has a standard deviation of 2.65 (measured in log-points).

(iii) Yes. However the regression does not control for many factors (size of firm, industry, profitability, experience and so forth).

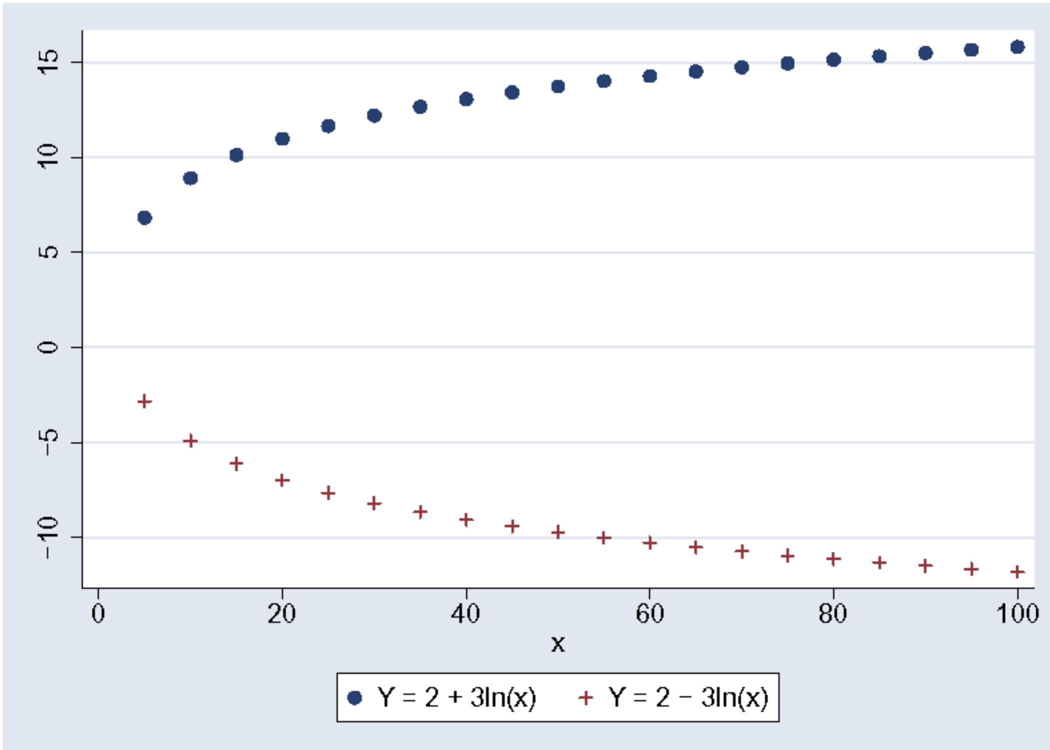
(iv) No. In isolation, these results do not imply gender discrimination. Gender discrimination means that two workers, identical in every way but gender, are paid different wages. Thus, it is also important to control for characteristics of the workers that may affect their productivity (education, years 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 in top corporate jobs.) 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.

(b) (i) If *MarketValue* increases by 1%, earnings increase by 0.37%

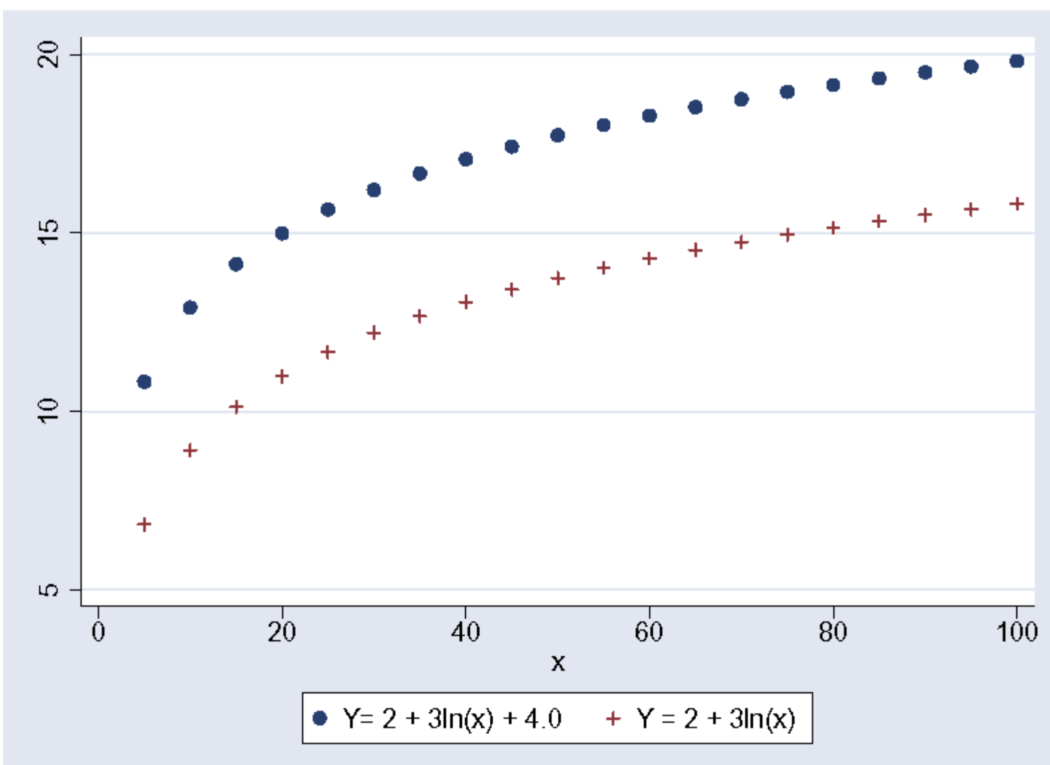
(ii) *Female* is correlated with the two new included variables and at least one of the variables is important for explaining ln(*Earnings*). Thus the regression in part (a) suffered from omitted variable bias.

(c) Forgetting about the effect or *Return*, whose effects seems small and statistically insignificant, the omitted variable bias formula (see equation (6.1)) suggests that *Female* is negatively correlated with ln(*MarketValue*).

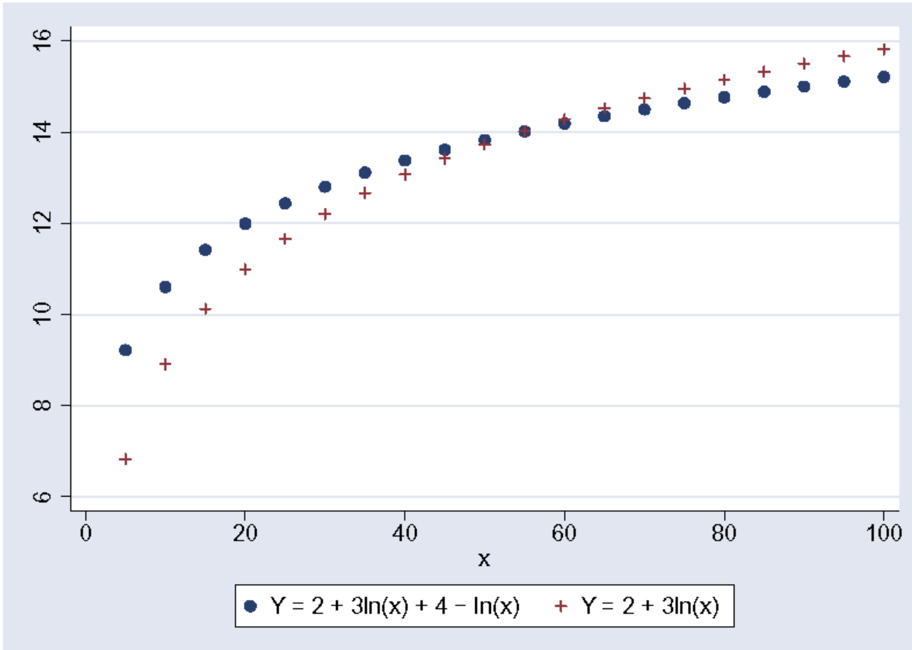
8.8 (a) and (b)


15
10
5
0
−5
−10
0
20
40
60
80
100
x
Y = 2 + 3ln(x)
Y = 2 − 3ln(x)

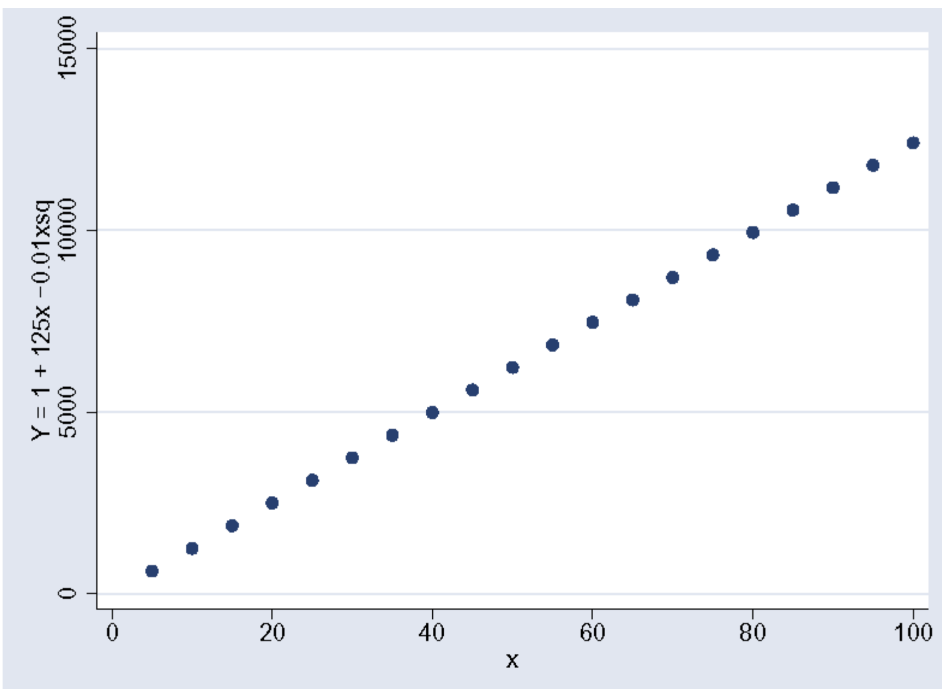
(c)


20
15
10
5
0
20
40
60
80
100
x
Y= 2 + 3ln(x) + 4.0
Y = 2 + 3ln(x)

(d)


16
14
12
10
8
6
0
20
40
60
80
100
x
Y = 2 + 3ln(x) + 4 − ln(x)
Y = 2 + 3ln(x)

(e)


15000
10000
5000
0
Y = 1 + 125x −0.01xsq
0
20
40
60
80
100
x

8.9. Note that

$$Y = \beta_0 + \beta_1 X + \beta_2 X^2$$
$$= \beta_0 + (\beta_1 + 21\beta_2)X + \beta_2(X^2 - 21X).$$

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 $\hat{\gamma} \pm 1.96 \times \mathrm{SE}(\hat{\gamma})$.

8.10. (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)
$$\begin{aligned}\Delta Y &= f(X_1 + \Delta X_1, X_2 + \Delta X_2) - f(X_1, X_2)\\ &= \beta_0 + \beta_1(X_1 + \Delta X_1) + \beta_2(X_2 + \Delta X_2) + \beta_3(X_1 + \Delta X_1)(X_2 + \Delta X_2)\\ &\quad - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2)\\ &= (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2.\end{aligned}$$

8.11. Linear model: $E(Y \mid X) = \beta_0 + \beta_1 X$, so that $\frac{dE(Y \mid X)}{dX} = \beta_1$ and the elasticity is $\frac{d \ln E(Y|X)}{d \ln Y} = \frac{dE(Y|X)}{dX} \times \frac{X}{E(Y|X)}$

$$\beta_1 \frac{X}{E(Y \mid X)} = \frac{\beta_1 X}{\beta_0 + \beta_1 X}$$

Log-Log Model: $E(Y \mid X) = E\left(e^{\beta_0 + \beta_1 \ln(X) + u} \mid X\right) = e^{\beta_0 + \beta_1 \ln(X)} E(e^u \mid X) = ce^{\beta_0 + \beta_1 \ln(X)}$, where $c =$ $E(e^u \mid X)$, which does not depend on X because u and X are assumed to be independent.

Thus $\frac{dE(Y \mid X)}{dX} = \frac{\beta_1}{X} ce^{\beta_0 + \beta_1 \ln(X)} = \beta_1 \frac{E(Y \mid X)}{X}$, and the elasticity is β_1.

8.12. (a) Because of random assignment within the group of returning students $E(X_{1i} \mid u_i) = 0$ in "γ-regression," so that $\hat{\gamma}_1$ is an unbiased estimator of γ_1.

(b) Because of random assignment within the group of returning students $E(X_{1i} \mid u_i) = 0$ in "δ- regression," so that $\hat{\delta}_1$ is an unbiased estimator of δ_1.

(c) Write $E(u_i \mid X_{1i}, X_{2i}) = E(u_i \mid X_{2i}) = \lambda_0 + \lambda_1 X_{2i}$, where linearity is assumed for the conditional expected value. Thus,

$$\begin{aligned}E(Y \mid X_1, X_2) &= \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + E(u_i \mid X_{1i}, X_{2i})\\ &= (\beta_0 + \lambda_0) + \beta_1 X_1 + (\beta_2 + \lambda_1) X_2 + \beta_3 X_1 X_2\end{aligned}.$$

Using this expression, $E(Y| X_1 = 1, X_2 = 0) - E(Y| X_1 = 0, X_2 = 0) = \beta_1$, which is equal to γ_1 from (a). Also, $E(Y| X_1 = 1, X_2 = 1) - E(Y| X_1 = 0, X_2 = 1) = \beta_1 + \beta_3$, which is equal to δ_1 from (b). Together, these results imply that $\beta_3 = \delta_1 - \gamma_1$.

(d) Defining $v_i = u_i - E(u_i \mid X_{1i}, X_{2i}) = u_i - E(u_i \mid X_{2i})$, then

$$Y_i = (\beta_0 + \lambda_0) + \beta_1 X_1 + (\beta_2 + \lambda_1) X_2 + \beta_3 X_1 X_2 + v_i,$$

where $E(v_i \mid X_{1i}, X_{2i}) = 0$. Thus, applying OLS to the equation will yield a biased estimate of the constant term [$E(\hat{\beta}_0) = \beta_0 + \lambda_0$], an unbiased estimate of β_1[$E(\hat{\beta}_1) = \beta_1$], a biased estimate of β_2 [$E(\hat{\beta}_2) = \beta_2 + \lambda_1$], and an unbiased estimate of β_3 [$E(\hat{\beta}_3) = \beta_3$].

Chapter 9
Assessing Studies Based on Multiple Regression

9.1. 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 1970s are likely to apply to Boston in the 1970s but not to Los Angeles in the 1970s. 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 1970s are unlikely to apply to New York in 2010. Attitudes towards smoking changed significantly from 1970 to 2010.

9.2. (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, $(X_i, \tilde{Y}_i),\ i = 1, \ldots, n$ are i.i.d. draws from their joint distribution.

(3) $v_i = u_i + w_i$ has a finite fourth moment because 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.

9.3. 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.

9.4.

			Estimated Effect of a 10% Increase in Average Income	
State	$\beta_{\ln(Income)}$	**Std. Dev. of Scores**	**In Points**	**In Std. Dev.**
Calif.	11.57 (1.81)	19.1	1.157 (0.18)	0.06 (0.001)
Mass.	16.53 (3.15)	15.1	1.65 (0.31)	0.11 (0.021)

The income effect in Massachusetts is roughly twice as large as the effect in California.

9.5. (a) $Q = \dfrac{\gamma_1\beta_0 - \gamma_0\beta_1}{\gamma_1 - \beta_1} + \dfrac{\gamma_1 u - \beta_1 v}{\gamma_1 - \beta_1}.$

and $P = \dfrac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1} + \dfrac{u - v}{\gamma_1 - \beta_1}.$

(b) $E(Q) = \dfrac{\gamma_1\beta_0 - \gamma_0\beta_1}{\gamma_1 - \beta_1}, \quad E(P) = \dfrac{\beta_0 - \gamma_0}{\gamma_1 - \beta_1}$

(c) $\operatorname{var}(Q) = \left(\dfrac{1}{\gamma_1 - \beta_1}\right)^2 (\gamma_1^2\sigma_u^2 + \beta_1^2\sigma_v^2), \ \operatorname{var}(P)$

$$= \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\sigma_u^2 + \sigma_v^2),$$

and

$$\operatorname{cov}(P, Q) = \left(\frac{1}{\gamma_1 - \beta_1}\right)^2 (\gamma_1\sigma_u^2 + \beta_1\sigma_V^2)$$

(d) (i) $\hat{\beta}_1 \xrightarrow{p} \dfrac{\operatorname{cov}(Q, P)}{\operatorname{var}(P)} = \dfrac{\gamma_1\sigma_u^2 + \beta_1\sigma_v^2}{\sigma_u^2 + \sigma_v^2}, \qquad \hat{\beta}_0 \xrightarrow{p} E(Q) - E(P)\dfrac{\operatorname{cov}(P, Q)}{\operatorname{var}(P)}$

(ii) $\hat{\beta}_1 - \beta_1 \xrightarrow{p} \dfrac{\sigma_u^2(\gamma_1 - \beta_1)}{\sigma_u^2 + \sigma_v^2} > 0,$ using the fact that $\gamma_1 > 0$ (supply curves slope up) and $\beta_1 < 0$ (demand curves slope down).

SSR_{100}

9.6. (a) The parameter estimates do not change. Nor does the R^2. The sum of squared residuals from the 100 observation regression is $SER_{200} = (100-2) \times 15.1^2 = 22,344.98,$ and the sum of squared residuals from the 200 observation regression is twice this value: $SSR_{200} = 2 \times 22,344.98.$ Thus, the *SER* from the 200 observation regression is $SER_{200} = \sqrt{\frac{1}{200-2} SSR_{200}} = 15.02.$ The standard errors for the regression coefficients are now computed using equation (5.4) where $\sum_{i=1}^{200}(X_i - \bar{X})^2 \hat{u}_i^2$ and $\sum_{i=1}^{200}(X_i - \bar{X})^2$ are twice their value from the 100 observation regression. Thus the standard errors for the 200 observation regression are the standard errors in the 100 observation regression multiplied by $\sqrt{\dfrac{100-2}{200-2}} = 0.704.$ In summary, the results for the 200 observation regression are

$$\underset{(10.63)}{\hat{Y} = 32.1} + \underset{(8.59)}{66.8X}, \quad SER = 15.02, \quad R^2 = 0.81$$

(b) The observations are not *i.i.d.*: half of the observations are identical to the other half, so that the observations are not *independent*.

9.7. (a) True. Correlation between regressors and error terms means that the OLS estimator is inconsistent.

(b) True.

9.8. Not directly. Test scores in California and Massachusetts are for different tests and have different means and variances. However, converting (9.5) into units for Massachusetts yields the implied regression to *TestScore*(*MA units*) = 740.9 − 1.80 × *STR*, which is similar to the regression using Massachusetts data shown in Column 1 of Table 9.2. After this adjustment the regression could be somewhat useful; however, the regression in Column 1 of Table 9.2 has a low R^2, suggesting that it will not provide an accurate forecast of test scores.

9.9. Both regressions suffer from omitted variable bias so that they will not provide reliable estimates of the causal effect of income on test scores. However, the nonlinear regression in (8.18) fits the data well, so that it could be used for forecasting.

9.10. There are several reasons for concern. Here are a few.

Internal consistency: omitted variable bias as explained in the last paragraph of the box.

Internal consistency: sample selection may be a problem as the regressions were estimated using a sample of full-time workers. (See exercise 9.3 for a related problem.)

External consistency: Returns to education may change over time because of the relative demands and supplies of skilled and unskilled workers in the economy. To the extent that this is important, the results shown in the box (based on 2008 data) may not accurately estimate today's returns to education.

9.11. Again, there are reasons for concern. Here are a few.

Internal consistency: To the extent that price is affected by demand, there may be simultaneous equation bias.

External consistency: The internet and introduction of "*E*-journals" may induce important changes in the market for academic journals so that the results for 2000 may not be relevant for today's market.

9.12. (a) See the answer to part (c) of exercise 2.27.

(b) Because $\tilde{X} = E(X|Z)$, then $E(X|\tilde{X}) = \tilde{X}$. Thus

$$E(w|\tilde{X}) = E[(X - \tilde{X} \mid \tilde{X})] = E(X|\tilde{X}) - E(\tilde{X} \mid \tilde{X}) = \tilde{X} - \tilde{X} = 0.$$

(c) $X_i = \tilde{X} + w_i$, so that $Y_i = \beta_0 + \beta_1(\tilde{X} + w_i) + u_i = \beta_0 + \beta_1 \tilde{X} + v_i$, where $v_i = u_i + \beta_1 w_i$. Because $E(u|Z) = 0$ and $\tilde{X}$ depends only on Z (that is, $\tilde{X} = E(X|Z)$), then $E(u|\tilde{X}) = 0$. Together with the result in (b), this implies that $E(v_i \mid \tilde{X}) = 0$. You can then verify the other assumptions in Key Concept (4.3), and the result follows from the consistency of the OLS estimator under these assumptions.

9.13. (a) $\hat{\beta}_1 = \dfrac{\sum_{i=1}^{300}(\tilde{X}_i - \bar{\tilde{X}})(Y_i - \bar{Y})}{\sum_{i=1}^{300}(\tilde{X}_i - \bar{\tilde{X}})^2}$. Because all of the X_i's are used (although some are used for the wrong values of Y_j), $\bar{\tilde{X}} = \bar{X}$, and $\sum_{i=1}^{n}(X_i - \bar{X})^2$. Also, $Y_i - \bar{Y} = \beta_1(X_i - \bar{X}) + u_i - \bar{u}$. Using these expressions:

$$\hat{\beta}_1 = \beta_1 \frac{\sum_{i=1}^{0.8n}(X_i - \bar{X})^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + \beta_1 \frac{\sum_{i=0.8n+1}^{n}(\tilde{X}_i - \bar{X})(X_i - \bar{X})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\sum_{i=1}^{n}(\tilde{X}_i - \bar{X})(u_i - \bar{u})}{\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$= \beta_1 \frac{\frac{1}{n}\sum_{i=1}^{0.8n}(X_i - \bar{X})^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} + \beta_1 \frac{\frac{1}{n}\sum_{i=0.8n+1}^{n}(\tilde{X}_i - \bar{X})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i - \bar{X})(u_i - \bar{u})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

where $n = 300$, and the last equality uses an ordering of the observations so that the first 240 observations ($= 0.8 \times n$) correspond to the correctly measured observations ($\tilde{X}_i = X_i$).

As is done elsewhere in the book, we interpret $n = 300$ as a large sample, so we use the approximation of n tending to infinity. The solution provided here thus shows that these expressions are approximately true for n large and hold in the limit that n tends to infinity. Each of the averages in the expression for $\hat{\beta}_1$ have the following probability limits:

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \xrightarrow{p} \sigma_X^2,$$

$$\frac{1}{n}\sum_{i=1}^{0.8n}(X_i - \bar{X})^2 \xrightarrow{p} 0.8\sigma_X^2,$$

$$\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i - \bar{X})(u_i - \bar{u}) \xrightarrow{p} 0, \text{ and}$$

$$\frac{1}{n}\sum_{i=0.8n+1}^{n}(\tilde{X}_i - \bar{X})(X_i - \bar{X}) \xrightarrow{p} 0,$$

where the last result follows because $\tilde{X}_i \neq X_i$ for the scrambled observations and X_j is independent of X_i for $i \neq j$. Taken together, these results imply $\hat{\beta}_1 \xrightarrow{p} 0.8\beta_1$.

(b) Because $\hat{\beta}_1 \xrightarrow{p} 0.8\beta_1$, $\hat{\beta}_1 / 0.8 \xrightarrow{p} \beta_1$, so a consistent estimator of β_1 is the OLS estimator divided by 0.8.

(c) Yes, the estimator based on the first 240 observations is better than the adjusted estimator from part (b). Equation (4.21) in Key Concept 4.4 (page 129) implies that the estimator based on the first 240 observations has a variance that is

$$\text{var}(\hat{\beta}_1(240obs)) = \frac{1}{240}\frac{\text{var}\left[(X_i-\mu_X)u_i\right]}{\left[\text{var}(X_i)\right]^2}.$$

From part (a), the OLS estimator based on all of the observations has two sources of sampling error. The first is $\frac{\sum_{i=1}^{300}(\tilde{X}_i-\bar{X})(u_i-\bar{u})}{\sum_{i=1}^{300}(X_i-\bar{X})^2}$ which is the usual source that comes from the omitted factors (u). The second is $\beta_1\frac{\sum_{i=241}^{300}(\tilde{X}_i-\bar{X})(X_i-\bar{X})}{\sum_{i=1}^{300}(X_i-\bar{X})^2}$, which is the source that comes from scrambling the data. These two terms are uncorrelated in large samples, and their respective large-sample variances are:

$$\text{var}\left(\frac{\sum_{i=1}^{300}(\tilde{X}_i-\bar{X})(u_i-\bar{u})}{\sum_{i=1}^{300}(X_i-\bar{X})^2}\right) = \frac{1}{300}\frac{\text{var}\left[(X_i-\mu_X)u_i\right]}{\left[\text{var}(X_i)\right]^2}$$

and

$$\text{var}\left(\beta_1\frac{\sum_{i=241}^{300}(\tilde{X}_i-\bar{X})(X_i-\bar{X})}{\sum_{i=1}^{300}(X_i-\bar{X})^2}\right) = \beta_1^2\frac{0.2}{300}.$$

Thus

$$\text{var}\left(\frac{\hat{\beta}_1(300obs)}{0.8}\right) = \frac{1}{0.64}\left[\frac{1}{300}\frac{\text{var}\left[(X_i-\mu_X)u_i\right]}{\left[\text{var}(X_i)\right]^2} + \beta_1^2\frac{0.2}{300}\right]$$

which is larger than the variance of the estimator that only uses the first 240 observations.

Chapter 10
Regression with Panel Data

10.1. (a) With a \$1 increase 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 × 810 = 364.5. The 95% confidence interval is (0.45 ± 1.96 × 0.22) × 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 ± 1.96 × 0.066 = [– 0.1014, 0.1574]. With a population of 8.1 million, the number of fatalities will increase by 0.028 × 810 = 22.68 with a 95% confidence interval [– 0.1014, 0.1574] × 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 ± 1.64 × 0.47 = [1.04, 2.58]. With a population of 8.1 million, the number of fatalities will increase by 1.81 × 810 = 1466.1 with a 90% confidence interval [1.04, 2.58] × 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) Define a binary variable *west* which equals 1 for the western states and 0 for the other states. Include the interaction term between the binary variable *west* and the unemployment rate, *west* × (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* × (unemployment rate) is γ. Then β captures the effect of the unemployment rate in the eastern states, and $\beta + \beta$ 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 β. Using the coefficient estimate $(\hat{\gamma})$ and the standard error $\mathrm{SE}(\hat{\gamma})$, you can calculate the *t*-statistic to test whether γ is statistically significant at a given significance level.

10.2. (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 perfect multicollinearity. 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?)

10.3. 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 threats 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.

10.4. (a) slope = β_1, intercept = β_0

(b) slope = β_1, intercept = β_0

(c) slope = β_1, intercept = $\beta_0 + \gamma_3$

(d) slope = β_1, intercept = $\beta_0 + \gamma_3$

10.5. Let $D2_i = 1$ if $i = 2$ and 0 otherwise; $D3_i = 1$ if $i = 3$ and 0 otherwise … $Dn_i = 1$ if $i = n$ and 0 otherwise. Let $B2_t = 1$ if $t = 2$ and 0 otherwise; $B3_t = 1$ if $t = 3$ and 0 otherwise … $BT_t = 1$ if $t = T$ and 0 otherwise. Let $\beta_0 = \alpha_1 + \lambda_1$; $\gamma_i = \alpha_i - \alpha_1$ and $\delta_t = \lambda_t - \beta_1$.

10.6. $\tilde{v}_{it} = \tilde{X}_{it}u_{it}$. First note that $E(\tilde{v}_{it}) = E(\tilde{X}_{it}u_{it}) = E[\tilde{X}_{it}E(u_{it}|X_{i1}, X_{i2},\ldots,X_{iT},\alpha_i)] = 0$ from assumption 1. Thus, $\text{cov}(\tilde{v}_{it}\tilde{v}_{is}) = E(\tilde{v}_{it}\tilde{v}_{is}) = E(\tilde{X}_{it}\tilde{X}_{is}u_{it}u_{is})$. The assumptions in Key Concept 10.3 do not imply that this term is zero. That is, Key Concept 10.3 allows the errors for individual *i* to be correlated across time periods.

10.7. (a) Average snow fall does not vary over time, and thus will be perfectly collinear with the state fixed effect.

(b) $Snow_{it}$ does vary with time, and so this method can be used along with state fixed effects.

10.8. There are several ways. Here is one: let $Y_{it} = \beta_0 + \beta_1 X_{1,it} + \beta_2 t + \gamma_2 D2_i + \cdots + \gamma_n Dn_i + \delta_2(D2_i \times t) + \cdots + \delta_n(Dn_i \times t) + u_{it}$, where $D2_i = 1$ if $i = 2$ and 0 otherwise and so forth. The coefficient $\lambda_i = \beta_2 + \delta_i$.

10.9. (a) $\hat{\alpha}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$ which has variance $\frac{\sigma_u^2}{T}$. Because *T* is not growing, the variance is not getting small. $\hat{\alpha}_i$ is not consistent.

(b) The average in (a) is computed over *T* observations. In this case *T* is small ($T = 4$), so the normal approximation from the CLT is not likely to be very good.

10.10. No, one of the regressors is Y_{it-1}. This depends on Y_{it-1}. This means that assumption (1) is violated.

10.11 Using the hint, equation (10.22) can be written as

$$\hat{\beta}_1^{DM} = \frac{\sum_{i=1}^{n}\left(\frac{1}{4}\left(X_{i2}-X_{i1}\right)\left(Y_{i2}-Y_{i1}\right)+\frac{1}{4}\left(X_{i2}-X_{i1}\right)\left(Y_{i2}-Y_{i1}\right)\right)}{\sum_{i=1}^{n}\left(\frac{1}{4}\left(X_{i2}-X_{i1}\right)^2+\frac{1}{4}\left(X_{i2}-X_{i1}\right)^2\right)}$$

$$= \frac{\sum_{i=1}^{n}\left(X_{i2}-X_{i1}\right)\left(Y_{i2}-Y_{i1}\right)}{\sum_{i=1}^{n}\left(X_{i2}-X_{i1}\right)^2} = \hat{\beta}_1^{BA}$$

Chapter 11
Regression with a Binary Dependent Variable

11.1. (a) The t-statistic for the coefficient on *Experience* is 0.031/0.009 = 3.44, which is significant at the 1% level.

(b) $z_{Matthew} = 0.712 + 0.031 \times 10 = 1.022; \Phi(1.022) = 0.847$

(c) $z_{Christopher} = 0.712 + 0.031 \times 0 = 0.712; \Phi(0.712) = 0.762$

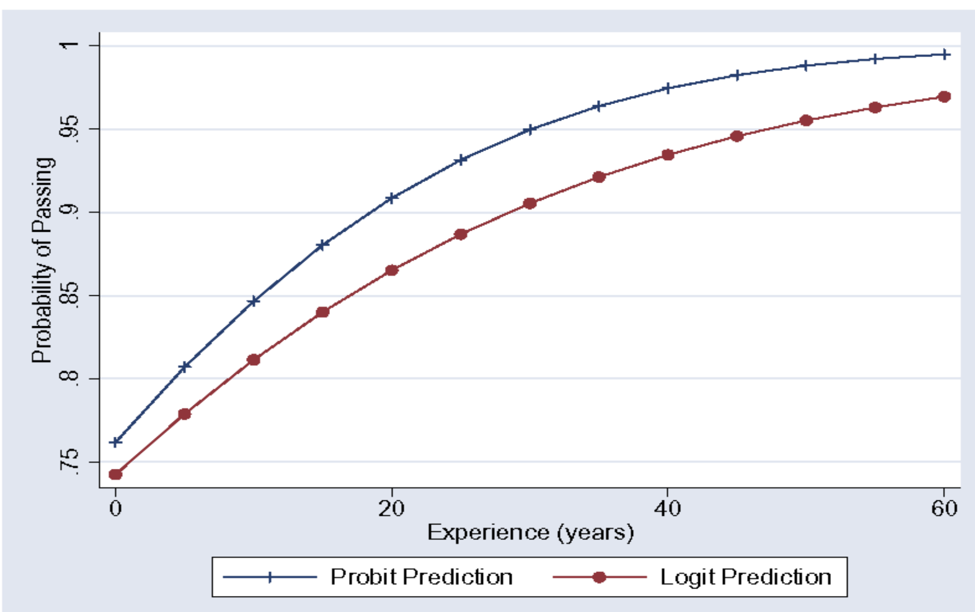
(d) $z_{Jed} = 0.712 + 0.031 \times 80 = 3.192; \Phi(3.192) = 0.999,$ this is unlikely to be accurate because the sample did not include anyone with more that 40 years of driving experience.

11.2. (a) The t-statistic for the coefficient on *Experience* is t = 0.040/0.016 = 2.5, which is significant at the 5% level.

$$\text{Prob}_{Matthew} = \frac{1}{1+e^{-(1.059+0.040\times 10)}} = \frac{1}{1+e^{-1.459}} = 0.811$$

$$\text{Prob}_{Christopher} = \frac{1}{1+e^{-(1.059+0.040\times 0)}} = \frac{1}{1+e^{-1.059}} = 0.742$$

(b)



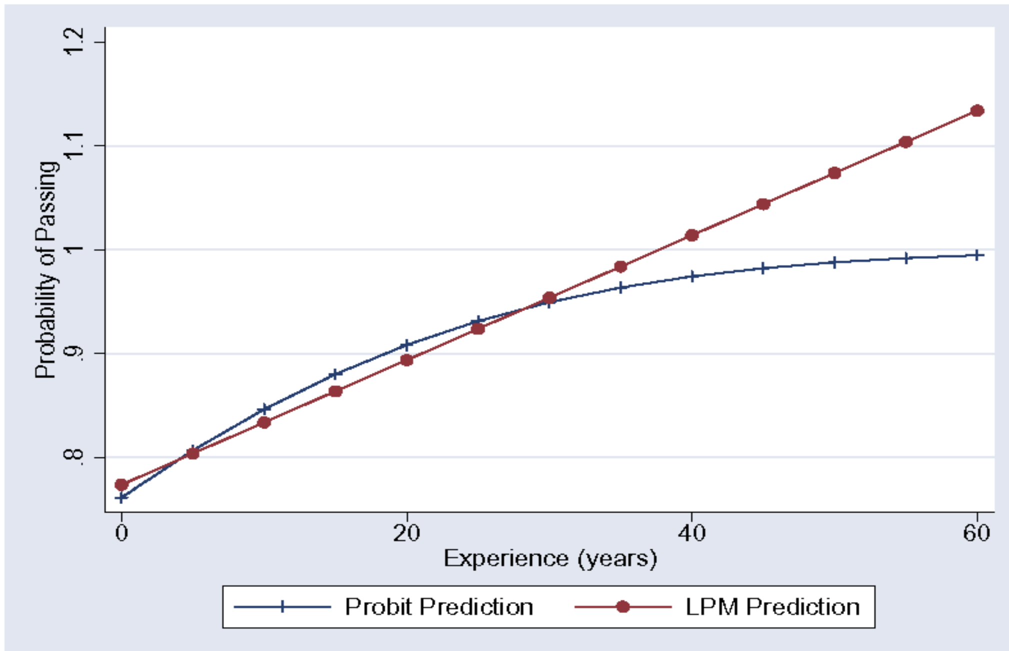
The shape of the regression functions are similar, but the logit regression lies below the probit regression for experience in the range of 0 = 60 years.

11.3. (a) The t-statistic for the coefficient on *Experience* is t = 0.006/0.002 = 3, which is significant a the 1% level.

$\text{Prob}_{Matther} = 0.774 + 0.006 \times 10 = 0.836$

$\text{Prob}_{Christopher} = 0.774 + 0.006 \times 0 = 0.774$

(b)



The probabilities are similar except when experience in large (> 40 years). In this case the LPM model produces nonsensical results (probabilities greater than 1.0).

11.4. (a)

Group	Probit	Logit	LPM
Men	$\Phi(1.282 - 0.333) = 0.829$	$\frac{1}{1+e^{-(2.197-0.622)}} = 0.829$	0.829
Women	$\Phi(1.282) = 0.900$	$\frac{1}{1+e^{-(2.197)}} = 0.900$	0.900

(b) Because there is only regressor and it is binary (*Male*), estimates for each model show the fraction on males and females passing the test. Thus, the results are identical for all models.

11.5. (a) $\Phi(0.806 + 0.041 \times 10 \times 0.174 \times 1 - 0.015 \times 1 \times 10) = 0.814$

(b) $\Phi(0.806 + 0.041 \times 2 - 0.174 \times 0 - 0.015 \times 0 \times 2) = 0.813$

(c) The *t*-stat on the interaction term is $-0.015/0.019 = -0.79$, which is not significant at the 10% level.

11.6. (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.

11.7. (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.

11.8. (a) Since Y_i is 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 ith 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., $i = 1,\ldots, n$, the joint probability distribution of $Y_1,\ldots, Y_n$ conditional on the X's is

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

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

11.9. (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 ± 1.96 × 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 these omitted variables are unlikely to bias the answer in (a). Other variables such as education, marital status, and occupation may also be related the probability of default, and these variables are omitted from the regression in column. Adding these variables (see columns (4)–(6)) have little effect on the estimated effect of *black* on the probability of mortgage denial.

11.10. (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 $Y_1,\ldots, Y_n$ is

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

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

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

Using calculus, the partial derivatives of L are

$$\frac{\partial L}{\partial p} = \frac{n_1}{p} - \frac{n-n_1-n_2}{1-p-q}, \text{ and}$$

$$\frac{\partial L}{\partial q} = \frac{n_2}{q} - \frac{n-n_1-n_2}{1-p-q}.$$

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

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

11.11. (a) This is a censored or truncated regression model (note the dependent variable might be zero).

(b) This is an ordered response model.

(c) This is the discrete choice (or multiple choice) model.

(d) This is a model with count data.

Chapter 12
Instrumental Variables Regression

12.1. (a) The change in the regressor, $\ln(P_{i,1995}^{\text{cigarettes}}) - \ln(P_{i,1985}^{\text{cigarettes}})$, from a \$0.50 per pack increase in the retail price is ln(8.00) – ln(7.50) = 0.0645. The expected percentage change in cigarette demand is $-0.94 \times 0.0645 \times 100\% = -6.07\%$. The 95% confidence interval is $(-0.94 \pm 1.96 \times 0.21) \times 0.0645 \times 100\% = [-8.72\%, -3.41\%]$.

(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).

12.2. (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 corr(X_i, u_i) = 0, and this implies 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 of conclusion 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

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

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

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

12.4. 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

$$s_{\hat{X}Y} = \sum_{i=1}^n (\hat{X}_i - \bar{\hat{X}})(Y_i - \bar{Y}) = \hat{\pi}_1 \sum_{i=1}^n (Z_i - \bar{Z})(Y_i - \bar{Y}) = \hat{\pi}_1 s_{ZY},$$

$$s_{\hat{X}}^2 = \sum_{i=1}^n (\hat{X}_i - \bar{\hat{X}})^2 = \hat{\pi}_1^2 \sum_{i=1}^n (Z_i - \bar{Z})^2 = \hat{\pi}_1^2 s_Z^2.$$

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

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

12.5. (a) Instrument relevance. Z_i does not enter the population regression for X_i

(b) Z is not a valid instrument. $\hat{X}$ will be perfectly collinear with W. (Alternatively, the first stage regression suffers from perfect multicollinearity.)

(c) W is perfectly collinear with the constant term.

(d) Z is not a valid instrument because it is correlated with the error term.

12.6. Use R^2 to compute the homoskedastic-only F statistic as

$F_{\text{HomoskedasitcOnly}} = \dfrac{R^2/k}{(1-R^2)/(T-k-1)} = \dfrac{0.05}{0.95/98} = 5.16$ with 100 observations in which case we conclude that the instrument may be week. With 500 observations the $F_{\text{HomoskedasitcOnly}} = 26.2$ so the instrument is not weak.

12.7. (a) Under the null hypothesis of instrument exogeneity, the J statistic is distributed as a χ_1^2 random variable, with a 1% critical value of 6.63. Thus the statistic is significant, and instrument exogeneity $E(u_i|Z_{1i}, Z_{2i}) = 0$ is rejected.

(b) The J test suggests that $E(u_i|Z_{1i}, Z_{2i}) \neq 0$, but doesn't provide evidence about whether the problem is with Z_1 or Z_2 or both.

12.8. (a) Solving for P yields $P = \dfrac{\gamma_0 - \beta_0}{\beta_1} + \dfrac{u_i^d - u_i^s}{\beta_1}$; thus $\text{cov}(P, u^s) = \dfrac{-\sigma_{u^s}^2}{\beta_1}$

(b) Because cov$(P,u) \neq 0$, the OLS estimator is inconsistent (see (6.1)).

(c) We need a instrumental variable, something that is correlated with P but uncorrelated with u^s. In this case Q can serve as the instrument, because demand is completely inelastic (so that Q is not affected by shifts in supply). γ_0 can be estimated by OLS (equivalently as the sample mean of Q_i).

12.9. (a) There are other factors that could affect both the choice to serve in the military and annual earnings. One example could be education, although this could be included in the regression as a control variable. Another variable is "ability" which is difficult to measure, and thus difficult to control for in the regression.

(b) The draft was determined by a national lottery so the choice of serving in the military was random. Because it was randomly selected, the lottery number is uncorrelated with individual characteristics that may affect earning and hence the instrument is exogenous. Because it affected the probability of serving in the military, the lottery number is relevant.

12.10. $$\hat{\beta}_{TSLS} = \frac{\text{cov}(Z_i, Y_i)}{\text{cov}(Z_i, X_i)} = \frac{\text{cov}(Z_i, \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i)}{\text{cov}(Z_i, X_i)} = \frac{\beta_1 \text{cov}(Z_i, X_i) + \beta_2 \text{cov}(Z_i, W_i)}{\text{cov}(Z_i, X_i)}$$

(a) If $\text{cov}(Z_i, W_i) = 0$ the IV estimator is consistent.

(b) If $\text{cov}(Z_i, W_i) \neq 0$ the IV estimator is not consistent.

Chapter 13
Experiments and Quasi-Experiments

13.1. 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 $15.59 \pm 1.96 \times 2.40 = [10.886, 20.294]$.

13.2. (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 in $15.89 + 0.66 \times (-5) = 12.59$. Because of random assignment, the estimators of the small class effect and the teacher experience effect are uncorrelated. Thus, the standard error for the difference in average test scores is $[2.16^2 + (-5)^2 \times 0.17^2]^{\frac{1}{2}} = 2.3212$. 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.

13.3. (a) The estimated average treatment effect is $\bar{X}_{\text{TreatmentGroup}} - \bar{X}_{\text{Control}} = 1241 - 1201 = 40$ points.

(b) There would be nonrandom assignment if men (or women) had different probabilities of being assigned to the treatment and control groups. Let p_{Men} denote the probability that a male is assigned to the treatment group. Random assignment means $p_{Men} = 0.5$. Testing this null hypothesis results in a t-statistic of $t_{Men} = \frac{\hat{p}_{Men} - 0.5}{\sqrt{\frac{1}{n_{Men}} \hat{p}_{Men}(1 - \hat{p}_{Men})}} = \frac{0.55 - 0.50}{\sqrt{\frac{1}{100} 0.55(1 - 0.55)}} = 1.00,$ so that the null of random assignment cannot be rejected at the 10% level. A similar result is found for women.

13.4. (a) (i) $X_{it} = 0, G_i = 1, D_t = 0$

(ii) $X_{it} = 1, G_i = 1, D_t = 1$

(iii) $X_{it} = 0, G_i = 0, D_t = 0$

(iv) $X_{it} = 0, G_i = 0, D_t = 1$

(b) (i) $\beta_0 + \beta_2$

(ii) $\beta_0 + \beta_1 + \beta_2 + \beta_3$

(iii) β_0

(iv) $\beta_0 + \beta_3$

(c) β_1

(d) "New Jersey after – New Jersey before" = $\beta_1 + \beta_3$, where β_3 denotes the time effect associated with changes in the economy between 1991 and 1993. "1993 New Jersey – 1993 Pennsylvania" = $\beta_1 + \beta_2$, where β_2 denotes the average difference in employment between NJ and PA.

13.5. (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.

13.6. 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 (5.27), it is straightforward to see that the variance for the differences estimator

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

(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 ith 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 ith 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

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

which is homoskedastic. Using Equation (5.27), it is straightforward to see that the variance for the differences-in-differences estimator

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

(c) When $\sigma_\alpha^2 > \sigma_u^2$, we'll have $\text{var}\left(\hat{\beta}_1^{\text{differences}}\right) > \text{var}\left(\hat{\beta}_1^{\text{diffs-in-diffs}}\right)$ 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.

13.7. From the population regression

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

we have

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

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 (13.5) in the case of a single W regressor.

13.8. 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 8.3

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

Thus

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

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

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

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

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

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

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

So

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

13.10. (a) This is achieved by adding and subtracting $\beta_0 + \beta_1X_i$ to the right hand side of the equation and rearranging terms.

(b) $E[u_i|X_i] = E[\beta_{0i} - \beta_0|X_i] + E[(\beta_{1i} - \beta_1)X_i|X_i]|X_i] + E[v_i|X_i] = 0$.

(c) (1) is shown in (b). (2) follows from the assumption that (v_i, X_i, β_{0i}, β_{1i}) are *i.i.d.* random variables.

(d) Yes, the assumptions in KC 4.3 are satisfied.

(e) If β_{1i} and X_i are positively correlated, $\text{cov}(\beta_{1i}, X_i) = E[(\beta_{1i} - \beta_1)(X_i - \beta_X)] = E[(\beta_{1i} - \beta_1)X_i\ ] > 0$, where the first equality follows because $\beta_1 = E(\beta_{1i})$ and the inequality follows because the covariance and correlation have the same sign. Note $0 < E[(\beta_{1i} - \beta_1)X_i] = E\{E[(\beta_{1i} - \beta_1)X_i|X_i]\}$ by the law of iterated expectations, so it must the case that $E[(\beta_{1i} - \beta_1)X_i|X_i] > 0$ for some values of X_i. Thus assumption (1) is violated. This induces positive correlation between the regressors and error term, leading the inconsistency in OLS. Thus, the methods in Chapter 4 are not appropriate.

13.11. Following the notation used in Chapter 13, let π_{1i} denote the coefficient on state sales tax in the "first stage" IV regression, and let $-\beta_{1i}$ denote cigarette demand elasticity. (In both cases, suppose that income has been controlled for in the analysis.) From (13.11)

$$\hat{\beta}^{TSLS} \xrightarrow{p} \frac{E(\beta_{1i}\ \pi_{1i})}{E(\pi_{1i})} = E(\beta_{1i}) + \frac{\text{cov}(\beta_{1i}, \pi_{1i})}{E(\pi_{1i})} = \text{Average Treatment Effect} + \frac{\text{cov}(\beta_{1i}, \pi_{1i})}{E(\pi_{1i})},$$

where the first equality uses the uses properties of covariances (equation (2.34)), and the second equality uses the definition of the average treatment effect. Evidently, the local average treatment effect will deviate from the average treatment effect when $\text{cov}(\beta_{1i}, \pi_{1i}) \neq 0$. As discussed in Section 13.6, this covariance is zero when β_{1i} or π_{1i} are constant. This seems likely. But, for the sake of argument, suppose that they are not constant; that is, suppose the demand elasticity differs from state to state (β_{1i} is not constant) as does the effect of sales taxes on cigarette prices (π_{1i} is not constant). Are β_{1i} and π_{1i} related? Microeconomics suggests that they might be. Recall from your microeconomics class that the lower is the demand elasticity, the larger fraction of a sales tax is passed along to consumers in terms of higher prices. This suggests that β_{1i} and π_{1i} are positively related, so that $\text{cov}(\beta_{1i}, \pi_{1i}) > 0$. Because $E(\pi_{1i}) > 0$, this suggests that the local average treatment effect is greater than the average treatment effect when β_{1i} varies from state to state.

Chapter 14
Introduction to Time Series Regression and Forecasting

14.1. (a) Since the probability distribution of Y_t is the same as the probability distribution of Y_{t-1} (this is the definition of stationarity), the means (and all other moments) are the same.

(b) $E(Y_t) = \beta_0 + \beta_1 E(Y_{t-1}) + E(u_t)$, but $E(u_t) = 0$ and $E(Y_t) = E(Y_{t-1})$. Thus $E(Y_t) = \beta_0 + \beta_1 E(Y_t)$, and solving for $E(Y_t)$ yields the result.

14.2. (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 $[\ln(IP_t) - \ln(IP_{t-1})] \times 100 = 100 \times \ln\left(\frac{IP_t}{IP_{t-1}}\right)$ when the change is small.

Converting this into an annual (12 month) change yields $1200 \times \ln\left(\frac{IP_t}{IP_{t-1}}\right)$.

(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 14.5, the QLR statistic 3.45 is larger than the 10% critical value (3.26), but less than the 5% critical value (3.66). 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 × 12 = 492 number of observations on the dependent variable. The BIC and AIC are calculated from the formulas $\text{BIC}(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{\ln T}{T}$ and $AIC(p) = \ln\left(\frac{SSR(p)}{T}\right) + (p+1)\frac{2}{T}$.

AR Order (*p*)	**1**	**2**	**3**	**4**	**5**	**6**
SSR (p)	29175	28538	28393	28391	28378	28317
$\ln\left[\frac{SSR(p)}{T}\right]$	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.

14.3. (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 14.4, 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 14.2. The use of first differences in Exercise 14.2 eliminates random walk trend in ln(IP).

14.4. (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 5% critical value, so we conclude at the 5% significance level that IP growth helps to predict future interest rates.

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

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

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

with the conditional variance $\sigma^2_{t|t-1} = E[(Y_t - Y_{t|t-1})^2]$. This equation is minimized when the second term equals zero, or when $f_{t-1} = Y_{t|t-1}$. (An alternative is to use the hint, and notice that the result follows immediately from exercise 2.27.)

(c) Applying Equation (2.27), we know the error u_t is uncorrelated with u_{t-1} if $E(u_t|u_{t-1}) = 0$. From Equation (14.14) for the AR(p) process, we have

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

a function of Y_{t-1} and its lagged values. The assumption $E(u_t|Y_{t-1}, Y_{t-2}, ...) = 0$ 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,

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

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.

14.6. 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 we found.

(b) When we did these simulations, the 5%, 50% and 95% quantiles of the R^2 were 0.00, 0.19, and 0.73. The 5%, 50% and 95% quantiles of the t-statistic 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 we did these simulations with $T = 50$, the 5%, 50% and 95% quantiles of the R^2 were 0.00, 0.16 and 0.68. The 5%, 50% and 95% quantiles of the t-statistic 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 we did these simulations with $T = 200$, the 5%, 50% and 95% quantiles of the R^2 were 0.00, 0.17, and 0.68. The 5%, 50% and 95% quantiles of the t-statistic were −16.8, −0.76 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 $\frac{t - statistic}{\sqrt{T}}$ has a well-behaved limiting distribution. This is consistent with the Monte Carlo presented in this problem.)

14.7. (a) From Exercise (14.1) $E(Y_t) = 2.5 + 0.7E(Y_{t-1}) + E(u_t)$, but $E(Y_t) = E(Y_{t-1})$ (stationarity) and $E(u_t) = 0$, so that $E(Y_t) = 2.5/(1 - 0.7)$. Also, because $Y_t = 2.5 + 0.7Y_{t-1} + u_t$, $\text{var}(Y_t) = 0.7^2\text{var}(Y_{t-1}) + \text{var}(u_t) + 2 \times 0.7 \times \text{cov}(Y_{t-1}, u_t)$. But $\text{cov}(Y_{t-1}, u_t) = 0$ and $\text{var}(Y_t) = \text{var}(Y_{t-1})$ (stationarity), so that $\text{var}(Y_t) = 9/(1 - 0.7^2) = 17.647$.

(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

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

(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 of Y_{T+1} given Y_T is

$$Y_{T+1/T} = 2.5 + 0.7\text{Y}_\text{T} = 2.5 + 0.7 \times 102.3 = 74.11.$$

14.8. Because $E(u_t|Feb_t, Mar_t, \ldots Dec_t) = 0$, $E(Y_t|\ Feb_t, Mar_t, \ldots Dec_t) = \beta_0 + \beta_1 Feb_t + \beta_2 Mar_t + \cdots + \beta_{11} Dec_t$. For observations in January, all of the regressors are equal to zero, thus $\mu_{\text{Jan}} = \beta_0$. For observations in February, $Feb_t = 1$ and the other regressors equal 0, so that $\mu_{\text{Feb}} = \beta_0 + \beta_1$. Similar calculations apply for the other months.

14.9. (a) $E(Y_t) = \beta_0 + E(e_t) + b_1E(e_{t-1}) + \cdots + b_qE(e_{t-q}) = \beta_0$ [because $E(e_t) = 0$ for all values of t].

(b)

$$\begin{aligned}\text{var}(Y_t) &= \text{var}(e_t) + b_1^2\ \text{var}(e_{t-1}) + \cdots + b_q^2\ \text{var}(e_{t-q}) \\ &\quad + 2b_1\,\text{cov}(e_t, e_{t-1}) + \cdots + 2b_{q-1}b_q\,\text{cov}(e_{t-q+1}, e_{t-q}) \\ &= \sigma_e^2(1 + b_1^2 + \cdots + b_q^2)\end{aligned}$$

where the final equality follows from $\text{var}(e_t) = \sigma_e^2$ for all t and $\text{cov}(e_t, e_i) = 0$ for $i \neq t$.

(c) $Y_t = \beta_0 + e_t + b_1e_{t-1} + b_2e_{t-2} + \cdots + b_qe_{t-q}$ and $Y_{t-j} = \beta_0 + e_{t-j} + b_1e_{t-1-j} + b_2e_{t-2-j} + \cdots + b_qe_{t-q-j}$ and $\text{cov}(Y_t, Y_{t-j}) = \sum_{k=0}^{q}\sum_{m=0}^{q} b_kb_m\,\text{cov}(e_{t-k}, e_{t-j-m})$, where $b_0 = 1$. Notice that $\text{cov}(e_{t-k}, e_{t-j-m}) = 0$ for all terms in the sum.

(d) $\text{var}(Y_t) = \sigma_e^2\left(1 + b_1^2\right)$, $\text{cov}(Y_t, Y_{t-j}) = \sigma_e^2 b_1$, and $\text{cov}(Y_t, Y_{t-j}) = 0$ for $j > 1$.

14.10. A few things to note: first, computing the QLR using 25% trimming will result in a statistic that is at least as large as just choosing one date (the usual F statistic) and a statistic that can be no larger than the QLR with 15% trimming (because with the test with 15% trimming chooses that maximum over a larger number of statistics). Thus the 25%-trimming critical values will be larger than the critical values for the F statistic and smaller than the critical values for the 15%-trimming QLR statistic.

(a) The F statistic is larger than the 5% CV for 15% trimming (3.66), so it must be larger than the critical value for 25% trimming (which must be less than 3.66), so the null is rejected.

(b) The F statistic is smaller than the 5% critical value from the $F_{5,\infty}$ distribution (2.21), so that it must be smaller than the critical value with 25% (which must be greater than 2.21), so the null is not rejected.

(c) This is the intermediate case. Critical values for the 25% trimming would have to be computed.

14.11. Write the model as $Y_t - Y_{t-1} = \beta_0 + \beta_1(Y_{t-1} - Y_{t-2}) + u_t$. Rearranging yields

$$Y_t = \beta_0 + (1 + \beta_1)Y_{t-1} - \beta_1 Y_{t-2} + u_t.$$

Chapter 15
Estimation of Dynamic Causal Effects

15.1. (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 $25 \times [\beta_i \pm 1.96SE(\beta_i)]$**
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 ith quarter from the 25% oil price jump is $25 \times [\beta_i \pm 1.96SE(\beta_i)]$ percentage points. The confidence interval is reported in the table in (a).

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

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

(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.

15.2. (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 ith quarter is $25\beta_i$.

Period ahead (i)	**Dynamic multiplier (β_i)**	**Predicted change in interest rates ($25\beta_i$)**	**95% confidence interval $25 \times [\beta_i \pm 1.96SE(\beta_i)]$**
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 ith quarter from the 25% oil price jump is $25 \times [\beta_i \pm 1.96\text{SE}(\beta_i)]$. 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.

15.3. The dynamic causal effects are for experiment A. The regression in exercise 15.1 does not control for interest rates, so that interest rates are assumed to evolve in their "normal pattern" given changes in oil prices.

15.4. 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.

15.5. 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 (15.4), we have

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

Comparing the above equation to Equation (15.7), we see $\delta_0 = \beta_0$, $\delta_1 = \beta_1$, $\delta_2 = \beta_1 + \beta_2$, $\delta_3 = \beta_1 + \beta_2 + \beta_3, \ldots,$ and $\delta_{r+1} = \beta_1 + \beta_2 + \cdots + \beta_r + \beta_{r+1}$.

15.6. (a) Write

$$\begin{aligned} \text{var}(u_t) &= \phi_1^2 \,\text{var}(u_{t-1}) + \text{var}(\tilde{u}_t) + 2\phi_1 \,\text{cov}(u_{t-1}, \tilde{u}_t) \\ &= \phi_1^2 \,\text{var}(u_t) + \sigma_{\tilde{u}}^2 \end{aligned}$$

where the second equality follows from stationarity (so that $\text{var}(u_t) = \text{var}(u_{t-1})$ and $\text{cov}(u_{t-1}, \tilde{u}_t) = 0$. The result follows by solving for $\text{var}(u_t)$. The result of $\text{var}(X_t)$ is similar.

(b) Write

$$\begin{aligned}\text{cov}(u_t, u_{t-1}) &= \phi_1 \text{ var}(u_{t-1}) + \text{cov}(u_{t-1}, \tilde{u}_t)\\ &= \phi_1 \text{ var}(u_t)\end{aligned}$$

showing the result for $j = 1$. For $j = 2$, write

$$\begin{aligned}\text{cov}(u_t, u_{t-2}) &= \phi_1 \text{ cov}(u_{t-1}, u_{t-2}) + \text{cov}(u_{t-2}, \tilde{u}_t)\\ &= \phi_1 \text{ cov}(u_t, u_{t-1})\\ &= \phi_1^2 \text{ var}(u_t)\end{aligned}$$

and similarly for other values of j. The result for X is similar.

(c) $\text{cor}(u_t, u_{t-j}) = \dfrac{\text{cov}(u_t, u_{t-j})}{\sqrt{\text{var}(u_t)\,\text{var}(u_{t-j})}} = \dfrac{\text{cov}(u_t, u_{t-j})}{\text{var}(u_t)}$ from stationarity. The result then follows from (b). The result for X is similar.

(d) $v_t = (X_t - \mu_X)u_t$

(i) $E(v_t^2) = E[(X_t - \mu_X)^2 u_t^2] = E[(X_t - \mu_X)^2]E[u_t^2] = \sigma_X^2 \sigma_u^2$, where the second equality follows because X_t and u_t are independent.

(ii) $\text{cov}(v_t v_{t-j}) = E[(X_t - \mu_X)(X_{t-j} - \mu_X)u_t u_{t-j}] = E[(X_t - \mu_X)(X_{t-j} - \mu_X)]E[u_t u_{t-j}] =$ $\gamma_1^j \text{ var}(X_t)\phi_1^j \text{ var}(u_t)$, so that $\text{cor}(v_t, v_{t-j}) = \gamma_1^j \phi_1^j$. Thus $f_\infty = 1 + 2\sum_{i=1}^{\infty}(\gamma_1\phi_1)^i = \dfrac{1+\gamma_1\phi_1}{1-\gamma_1\phi}$.

15.7. Write $u_t = \sum_{i=0}^{\infty} \phi_1^i \tilde{u}_{t-i}$

(a) Because $E(\tilde{u}_i | X_t) = 0$ for all i and t, $E(u_i|X_t) = 0$ for all i and t, so that X_t is strictly exogenous.

(b) Because $E(u_{t-j} | \tilde{u}_{t+1}) = 0$ for $j \geq 0$, X_t is exogenous. However $E(u_{t+1} | \tilde{u}_{t+1}) = \tilde{u}_{t+1}$ so that X_t is not strictly exogenous.

15.8. (a) Because X_t is exogenous, OLS is consistent.

(b) The GLS estimator regresses $Y_t - \phi_1 Y_{t-1}$ onto $X_t - \phi_1 X_{t-1}$. The error term in this regression is $\tilde{u}_t$. Because $X_t = \tilde{u}_{t+1}$, $X_t - \phi_1 X_{t-1} = \tilde{u}_{t+1} - \phi_1 \tilde{u}_t$, which is correlated with the regression error. Thus the GLS estimator will be inconsistent.

(c) $\hat{\beta}_1 \xrightarrow{p} \beta_1 + \dfrac{\text{cov}(X_t - \phi_1 X_{t-1}, \tilde{u}_t)}{\text{var}(X_t - \phi_1 X_{t-1})} = \beta_1 - \dfrac{\sigma_{\tilde{u}}^2 \phi_1}{\sigma_{\tilde{u}}^2(1+\phi_1^2)} = \beta_1 - \dfrac{\phi_1}{(1+\phi_1^2)}$

15.9. (a) This follows from the material around equation (3.2).

(b) Quasi-differencing the equation yields $Y_t - \phi_1 Y_{t-1} = (1 - \phi_1)\beta_0 + \tilde{u}_t$, and the GLS estimator of $(1 - \phi_1)\beta_0$ is the mean of $Y_t - \phi_1 Y_{t-1} = \frac{1}{T-1}\sum_{t=2}^{T}(Y_t - \phi_1 Y_{t-1})$. Dividing by $(1-\phi_1)$ yields the GLS estimator of β_0.

(c) This is a rearrangement of the result in (b).

(d) Write $\hat{\beta}_0 = \frac{1}{T}\sum_{t=1}^{T} Y_t = \frac{1}{T}(Y_T + Y_1) + \frac{T-1}{T}\frac{1}{T-1}\sum_{t=2}^{T-1} Y_t$, so that $\hat{\beta}_0 - \hat{\beta}_0^{GLS} = \frac{1}{T}(Y_T + Y_1) - \frac{1}{T}\frac{1}{T-1}\sum_{t=2}^{T-1} Y_t -$ $\frac{1}{1-\phi}\frac{1}{T-1}(Y_T - Y_1)$ and the variance is seen to be proportional to $\dfrac{1}{T^2}$.

15.10.

Multipliers		
Lag	**Multiplier**	**Cumulative Multiplier**
0	2	2
1	0	2
2	0	2
3	0	2
4	0	2
5	0	2

The long-run cumulative multiplier = 2.

Chapter 16
Additional Topics in Time Series Regression

16.1. 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 $Y_t, Y_{t+h|t} = E(Y_{t+h}|Y_t, Y_{t-1}, \ldots)$, is

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

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

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

16.2. (a) Because $R1_t$ follows a random walk ($R1_t = R1_{t-1} + u_t$), the i-period ahead forecast of $R1_t$ is $R1_{t+i|t} = R1_{t+i-1|t} = R1_{t+i-2|t} = \cdots\cdots = R1_t$. 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 $\Delta R1_t = 0.5\Delta + u_t$, $\Delta R1_t$ 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

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

The i-period ahead forecast of $R1_t$ is

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

Thus

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

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. $R1_{t+i|t} = 0.5^i R1_t$, 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 $I(0)$.

16.3. 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.

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

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

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

The last equation has used the fact that $E(u_t^2) = \mathrm{var}(u_t) + E(u_t)]^2 = \mathrm{var}(u_t),$ which follows because $E(u_t) = 0$. Because of the stationarity, $\mathrm{var}(u_{t-1}) = \mathrm{var}(u_t)$. Thus, $\mathrm{var}(u_t) = 1.0 + 0.5\mathrm{var}(u_t)$ which implies $\mathrm{var}(u_t) = 1.0/0.5 = 2.$

(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

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

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

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

16.4. Y_t follows an AR(p) model $Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + u_t$. $E(u_t | Y_{t-1}, Y_{t-2}, \ldots) = 0$ 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

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

16.5. 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:

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

16.6. (a) Rewrite the regression as

$Y_t = 3.0 + 2.3X_t + 1.7(X_{t+1} - X_t) + 0.2(X_t - X_{t-1}) + u_t$

Thus $\theta = 2.3$, $\delta_{-1} = 1.7$, $\delta_0 = 0.2$ and $\delta_1 = 0.0$.

(b) Cointegration requires X_t to be $I(1)$ and u_t to be $I(0)$.

(i) No

(ii) No

(iii) Yes

16.7. $\hat{\beta} = \dfrac{\sum_{t=1}^{T} Y_t X_t}{\sum_{t=1}^{T} X_t^2} = \dfrac{\sum_{t=1}^{T} Y_t \Delta Y_{t+1}}{\sum_{t=1}^{T} (\Delta Y_{t+1})^2} = \dfrac{\frac{1}{T}\sum_{t=1}^{T} Y_t \Delta Y_{t+1}}{\frac{1}{T}\sum_{t=1}^{T} (\Delta Y_{t+1})^2}$. Following the hint, the numerator is the same expression as (16.21) (shifted forward in time 1 period), so that $\frac{1}{T}\sum_{t=1}^{T} Y_t \Delta Y_{t+1} \xrightarrow{d} \frac{\sigma_u^2}{2}(\chi_1^2 - 1)$. The denominator is $\frac{1}{T}\sum_{t=1}^{T} (\Delta Y_{t+1})^2 = \frac{1}{T}\sum_{t=1}^{T} u_{t+1}^2 \xrightarrow{p} \sigma_u^2$ by the law of large numbers. The result follows directly.

16.8. (a) First note that $Y_{t-1/t-2} = \beta_{11} Y_{t-2} + \gamma_{11} X_{t-2}$ and $X_{t-1/-2} = \beta_{21} Y_{t-2} + \gamma_{21} X_{t-2}$. Also $Y_{t/t-2} = \beta_{11} Y_{t-1/t-2} + \gamma_{11} X_{t-1/t-2}$. Substituting yields

$$\begin{aligned} Y_{t/-2} &= \beta_{11}(\beta_{11} Y_{t-2} + \gamma_{11} X_{t-2}) + \gamma_{11}(\beta_{21} Y_{t-2} + \gamma_{21} X_{t-2}) \\ &= [\beta_{11}\beta_{11} + \gamma_{11}\beta_{21}] Y_{t-2} + [\beta_{11}\gamma_{11} + \gamma_{11}\gamma_{21}] X_{t-2} \end{aligned}$$

so that $\delta_1 = \beta_{11}\beta_{11} + \gamma_{11}\beta_{21}$ and $\delta_2 = \beta_{11}\gamma_{11} + \gamma_{11}\gamma_{21}$.

(b) There is no difference in iterated multistep or direct forecast *if* the values of δ_1 and δ_2 were known. (This is shown in (a).) But, these parameters must be estimated, and the implied VAR estimates of these parameters are more accurate (have lower standard errors) if the VAR model is correctly specified.

16.9. (a) From the law of iterated expectations

$$\begin{aligned} E(u_t^2) &= E\left(\sigma_t^2\right) \\ &= E\left(\alpha_0 + \alpha_1 u_{t-1}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_{t-1}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_t^2\right) \end{aligned}$$

where the last line uses stationarity of u. Solving for $E(u_t^2)$ gives the required result.

(b) As in (a)

$$\begin{aligned} E(u_t^2) &= E\left(\sigma_t^2\right) \\ &= E\left(\alpha_0 + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_p u_{t-p}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_{t-1}^2\right) + \alpha_2 E\left(u_{t-2}^2\right) + \cdots + \alpha_p E\left(u_{t-p}^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_t^2\right) + \alpha_2 E\left(u_t^2\right) + \cdots + \alpha_p E\left(u_t^2\right) \end{aligned}$$

so that $E(u_t^2) = \dfrac{\alpha_0}{1 - \sum_{t=1}^{p} \alpha_i}$.

(c) This follows from (b) and the restriction that $E(u_t^2) > 0$.

(d) As in (a)

$$\begin{aligned} E(u_t^2) &= E\left(\sigma_t^2\right) \\ &= \alpha_0 + \alpha_1 E\left(u_{t-1}^2\right) + \phi_1 E\left(\sigma_{t-1}^2\right) \\ &= \alpha_0 + (\alpha_1 + \phi_1) E\left(u_{t-1}^2\right) \\ &= \alpha_0 + (\alpha_1 + \phi_1) E\left(u_t^2\right) \\ &= \frac{\alpha_0}{1 - \alpha_1 - \phi_1} \end{aligned}$$

(e) This follows from (d) and the restriction that $E\left(u_t^2\right) > 0$.

16.10. Write $\Delta Y_t = \theta \Delta X_t + \Delta v_{1t}$ and $\Delta X_t = v_{2t}$; also $v_{1t-1} = Y_{t-1} - \theta X_{t-1}$. Thus $\Delta Y_t = -(Y_{t-1} - \theta X_{t-1}) + u_{1t}$ and $\Delta X_t = u_{2t}$, with $u_{1t} = v_{1t} + \theta v_{2t}$ and $u_{2t} = v_{2t}$.

Chapter 17
The Theory of Linear Regression with One Regressor

17.1. (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

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

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, $E(u_i|X_1,\ldots,X_n) = E(u_i|X_i)$ because the observations are i.i.d.

Under assumptions 1–3 of Key Concept 17.1, $\hat{\beta}_1^{RLS}$ is asymptotically normally distributed. The large sample normal approximation to the limiting distribution of $\hat{\beta}_1^{RLS}$ follows from considering

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

Consider first the numerator which is the sample average of $v_i = X_i u_i$. By assumption 1 of Key Concept 17.1, 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, 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

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

For the denominator, 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

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

(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 ($i = 1,\dots, n$) depends on $X_1,\dots, X_n$ but not on $Y_1,\dots, Y_n$.
Thus

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

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

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

The final equality used the fact that

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

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

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

$$\text{var}(\hat{\beta}_1^{RLS}|X1,\ldots,X_n) = \text{var}\left(\beta_1 + \frac{\sum_{i=1}^n X_i u_i}{\sum_{i=1}^n X_i^2}|X_1,\ldots,X_n\right)$$
$$= \frac{\sum_{i=1}^n X_i^2 \,\text{var}(u_i|X_1,\ldots,X_n)}{(\sum_{i=1}^n X_i^2)^2}$$
$$= \frac{\sum_{i=1}^n X_i^2 \sigma_u^2}{(\sum_{i=1}^n X_i^2)^2}$$
$$= \frac{\sigma_u^2}{\sum_{i=1}^n X_i^2}.$$

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

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

Since

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

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

(f) Under assumption 5 of Key Concept 17.1, 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:

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

Using the conditional mean and conditional variance of $\hat{\beta}_1^{RLS}$ derived in parts (c) and (d) respectively, 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

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

The conditional variance is

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

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

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

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

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

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 $n\Sigma_{i=1}^n X_i^2 \geq (\Sigma_{x=1}^n X_i)^2$, or $\text{var}(\tilde{\beta}_1|X_1,\ldots, X_n) \geq \text{var}(\hat{\beta}_1^{RLS}|X_1,\ldots, X_n)$.

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

17.2. The sample covariance is

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

where the final equality follows from the definition of $\overline{X}$ and $\overline{Y}$ which implies that $\Sigma_{i=1}^{n}(X_i - \mu_X) = n(\overline{X} - \mu_X)$ and $\Sigma_{i=1}^{n}(Y_i - \mu_Y) = n(\overline{Y} - \mu_Y)$, 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 $\overline{X} \xrightarrow{p} \mu_X$ and $\overline{Y} \xrightarrow{p} \mu_Y$ so $(\overline{X} - \mu_X)(\overline{Y} - \mu_Y) \xrightarrow{p} 0$ 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 $E[(X_i - \mu_X)(Y_i - \mu_Y)] = \sigma_{XY}$. To apply the law of large numbers on the first term, we need to have

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

which is satisfied since

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

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

17.3. (a) Using Equation (17.19), we have

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

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 $\bar{X} \xrightarrow{p} \mu_{X_2}$, or $\bar{X} - \mu_X \xrightarrow{p} 0$. By the consistency of sample variance, $\frac{1}{n}\Sigma_{i=1}^{n}(X_i - \bar{X})^2$ 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:

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

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

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

Using Slutsky's theorem,

$$\frac{\dfrac{\frac{1}{n}\sum_{i=1}^{n} v_t}{\sigma_v}}{\dfrac{\frac{1}{n}\sum_{i=1}^{n}(X_t-\bar{X})^2}{\sigma_X^2}} \xrightarrow{d} N(0,1),$$

or equivalently

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

Thus

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

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

17.4. (a) Write $(\hat{\beta}_1-\beta_1) = a_n S_n$ where $a_n = \frac{1}{\sqrt{n}}$ and $S_n = \sqrt{n}(\hat{B}_1-\beta_1)$. Now, $a_n \to 0$ 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 $\Pr(|\hat{\beta}_1-\beta_1|>\delta)\to 0$ 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

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

17.5. Because $E(W^4) = [E(W^2)]^2 + \mathrm{var}(W^2)$, $[E(W^2)]^2 \le E(W^4) < \infty$. Thus $E(W^2) < \infty$.

17.6. Using the law of iterated expectations, we have

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

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

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

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

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

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

$$f(u_i, u_j|X_i, X_j) = f(u_i|X_i)f(u_j|X_j)$$

into the definition of the conditional expectation, we have

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

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

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

and the result then follows from part (b).

17.8. (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

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

where

$$\tilde{Y}_i = \frac{Y_i}{\sqrt{\theta_0 + \theta_1 |X_i|}}, \quad \tilde{X}_{0i} = \frac{1}{\sqrt{\theta_0 + \theta_1 |X_i|}}$$

$$\tilde{X}_{1i} = \frac{X_i}{\sqrt{\theta_0 + \theta_1 |X_i|}}, \quad \text{and} \quad \tilde{u} = \frac{u_i}{\sqrt{\theta_0 + \theta_1 |X_i|}}.$$

(c) Using equations (17.2) and (17.19), we know the OLS estimator, $\hat{\beta}_1$, is

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

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

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

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

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

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

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

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

$$\sum_{i=1}^{n} 2(\tilde{Y}_i - b_0\tilde{X}_{0i} - b_1\tilde{X}_{1i})(-\tilde{X}_{0i}) = 0,$$
$$\sum_{i=1}^{n} 2(\tilde{Y}_i - b_0\tilde{X}_{0i} - b_1\tilde{X}_{1i})(-\tilde{X}_{1i}) = 0.$$

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

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

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

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

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

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

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

17.9. We need to prove

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

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

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

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

$$\hat{u}_i^2 = u_i^2 + (\hat{\beta}_0 - \beta_0)^2 + (\hat{\beta}_1 - \beta_1)^2 X_i^2 - 2u_i(\hat{\beta}_0 - \beta_0)$$
$$- 2u_i(\hat{\beta}_1 - \beta_1)X_i + 2(\hat{\beta}_0 - \beta_0)(\hat{\beta}_1 - \beta_1)X_i.$$

Substituting this into the expression for $\frac{1}{n}\Sigma_{i=1}^{n}[(X_i - \bar{X})^2 \hat{u}_i^2 - (X_i - \mu_X)^2 u_i^2]$ yields a series of terms each of which can be written as $a_n b_n$ where $a_n \xrightarrow{p} 0$ and $b_n = \frac{1}{n}\Sigma_{i=1}^{n} X_i^r u_i^s$ where r and s are integers. For example, $a_n = (\bar{X} - \mu_X)$, $a_n = (\hat{\beta}_1 - \beta_1)$ and so forth. The result then follows from Slutksy's theorem if $\frac{1}{n}\Sigma_{i=1}^{n} X_i^r u_i^s \xrightarrow{p} d$ 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.

17.10. Using (17.43) with $W = \hat{\theta} - \theta$ implies

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

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

17.11. Note: in early printing of the third edition there was a typographical error in the expression for $\mu_{Y|X}$. The correct expression is $\mu_{Y|X} = \mu_Y + (\sigma_{XY} / \sigma_X^2)(x - \mu_X)$.

(a) Using the hint and equation (17.38)

$$f_{Y|X=x}(y) = \frac{1}{\sqrt{\sigma_Y^2(1 - \rho_{XY}^2)}}$$
$$\times \exp\left(\frac{1}{-2(1 - \rho_{XY}^2)}\left(\left(\frac{x - \mu_X}{\sigma_X}\right)^2 - 2\rho_{XY}\left(\frac{x - \mu_X}{\sigma_X}\right)\left(\frac{y - \mu_Y}{\sigma_Y}\right) + \left(\frac{y - \mu_Y}{\sigma_Y}\right)^2\right) + \frac{1}{2}\left(\frac{x - \mu_X}{\sigma_X}\right)^2\right).$$

Simplifying yields the desired expression.

(b) The result follows by noting that $f_{Y|X=x}(y)$ is a normal density (see equation (17.36)) with $\mu = \mu_{T|X}$ and $\sigma^2 = \sigma^2_{Y|X}$.

(c) Let $b = \sigma_{XY}/\sigma^2_X$ and $a = \mu_Y - b\mu_X$.

17.12. (a)
$$E(e^u) = \int_{-\infty}^{\infty} \frac{1}{\sigma_u\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma_u^2} + u\right) du = \exp\left(\frac{\sigma_u^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sigma_u\sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma_u^2} + u - \frac{\sigma_u^2}{2}\right) du$$
$$= \exp\left(\frac{\sigma_u^2}{2}\right) \int_{-\infty}^{\infty} \frac{1}{\sigma_u\sqrt{2\pi}} \exp\left(-\frac{1}{2\sigma_u^2}\left(u - \sigma_u^2\right)^2\right) du = \exp\left(\frac{\sigma_u^2}{2}\right)$$

where the final equality follows because the integrand is the density of a normal random variable with mean and variance equal to σ_u^2. Because the integrand is a density, it integrates to 1.

(b) The result follows directly from (a).

17.13 (a) The answer is provided by equation (13.10) and the discussion following the equation. The result was also shown in Exercise 13.10, and the approach used in the exercise is discussed in part (b).

(b) Write the regression model as $Y_i = \beta_0 + \beta_1 X_i + v_i$, where $\beta_0 = E(\beta_{0i})$, $\beta_1 = E(\beta_{1i})$, and $v_i = u_i + (\beta_{0i} - \beta_0) + (\beta_{1i} - \beta_1)X_i$. Notice that

$$E(v_i|X_i) = E(u_i|X_i) + E(\beta_{0i} - \beta_0|X_i) + X_i E(\beta_{1i} - \beta_1|X_i) = 0$$

because β_{0i} and β_{1i} are independent of X_i. Because $E(v_i \mid X_i) = 0$, the OLS regression of Y_i on X_i will provide consistent estimates of $\beta_0 = E(\beta_{0i})$ and $\beta_1 = E(\beta_{1i})$. Recall that the weighted least squares estimator is the OLS estimator of Y_i/σ_i onto $1/\sigma_i$ and X_i/σ_i , where $\sigma_i = \sqrt{\theta_0 + \theta_1 X_i^2}$. Write this regression as

$$Y_i / \sigma_i = \beta_0 (1/\sigma_i) + \beta_1 (X_i / \sigma_i) + v_i / \sigma_i .$$

This regression has two regressors, $1/\sigma_i$ and X_i/σ_i. Because these regressors depend only on X_i, $E(v_i|X_i) = 0$ implies that $E(v_i/\sigma_i \mid (1/\sigma_i), X_i/\sigma_i) = 0$. Thus, weighted least squares provides a consistent estimator of $\beta_0 = E(\beta_{0i})$ and $\beta_1 = E(\beta_{1i})$.

Chapter 18
The Theory of Multiple Regression

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

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

with

$$\mathbf{Y} = \begin{pmatrix} \text{TestScore}_1 \\ \text{TestScore}_2 \\ \vdots \\ \text{TestScore}_n \end{pmatrix}, \qquad \mathbf{X} = \begin{pmatrix} 1 & \text{Income}_1 & \text{Income}_1^2 \\ 1 & \text{Income}_2 & \text{Income}_2^2 \\ \vdots & \vdots & \vdots \\ 1 & \text{Income}_n & \text{Income}_n^2 \end{pmatrix}$$

$$\mathbf{U} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_n \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

(b) The null hypothesis is H_0: $\boldsymbol{R\beta} = \boldsymbol{r}$ versus H_1: $\mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$, with

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

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

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

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.

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

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

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 the coefficient vector is

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

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

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

By the definition of sample variance

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

we know

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

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

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

and

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

By the definition of sample covariance

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

we know

$$\sum_{i=1}^{n} X_i Y_i = (n-1)s_{XY} + n\overline{X}\overline{Y}.$$

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

$$\sum_{i=1}^{n} X_{1i} Y_i = (n-1)s_{X_1Y} + n\overline{X}_1\overline{Y}$$
$$= (20-1)\times 0.22 + 20\times 7.24\times 6.39 = 929.45,$$
$$\sum_{i=1}^{n} X_{2i} Y_i = (n-1)s_{X_2Y} + n\overline{X}_2\overline{Y}$$
$$= (20-1)\times 0.32 + 20\times 4.00\times 6.39 = 517.28,$$

and

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

Therefore we have

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

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

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

The OLS estimator of the coefficient vector is

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

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 $\hat{\mathbf{U}} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{X}\hat{\beta}$, so

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

We have

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

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

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

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

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

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

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 $\tilde{\Sigma}_{\hat{\beta}}$ of the covariance matrix of $\hat{\beta}$, conditional on **X**, which is

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

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

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

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

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

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

18.3. (a)
$$\begin{aligned} \text{Var}(Q) &= E[(Q-\mu_Q)^2] \\ &= E[(Q-\mu_Q)(Q-\mu_Q)'] \\ &= E[(\mathbf{c}'\mathbf{W} - \mathbf{c}'\boldsymbol{\mu}_{\mathbf{W}})(\mathbf{c}'\mathbf{W} - \mathbf{c}'\boldsymbol{\mu}_{\mathbf{W}})'] \\ &= \mathbf{c}'E[(\mathrm{W} - \boldsymbol{\mu}_{\mathbf{W}})(\mathbf{W} - \boldsymbol{\mu}_{\mathbf{W}})']\mathbf{c} \\ &= \mathbf{c}'\,\text{var}(\mathbf{W})\,\mathbf{c} = \mathbf{c}'\Sigma_{\mathbf{w}}\mathbf{c} \end{aligned}$$

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 Σ_{W} is positive definite, we have $\mathbf{c}'\Sigma_{\mathbf{w}}\mathbf{c} > 0$ for every non-zero vector from the definition. Thus, var(Q) > 0. Both the vector **c** and the matrix Σ_{W} are finite, so var(Q) = $\mathbf{c}'\Sigma_{\mathbf{w}}\mathbf{c}$ is also finite. Thus, 0 < var(Q) < ∞.

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

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

with

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

(b) Because $X_i' = (1 \quad X_i)$, assumptions 1–3 in Key Concept 18.1 follow directly from assumptions 1–3 in Key Concept 4.3. Assumption 4 in Key Concept 18.1 is satisfied since observations X_i $(i = 1, 2, \dots n)$ are not constant and there is no perfect multicollinearity among the two vectors of the matrix **X**.

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

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

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

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

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

The estimator for the coefficient vector is

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X'X})^{-1}\mathbf{X'Y} \\ &= \frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2} \begin{pmatrix} \sum_{i=1}^n X_i^2/n & -\overline{X} \\ -\overline{X} & 1 \end{pmatrix} \begin{pmatrix} n\overline{Y} \\ \sum_{i=1}^n X_i Y_i \end{pmatrix} \\ &= \frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2} \begin{pmatrix} \overline{Y}\sum_{i=1}^n X_i^2 - \overline{X}\sum_{i=1}^n X_i Y_i \\ \sum_{i=1}^n X_i Y_i - n\overline{X}\overline{Y} \end{pmatrix}\end{aligned}$$

Therefore we have

$$\hat{\beta} = \frac{\sum_{i=1}^n X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^n (X_i - \overline{X})^2} = \frac{\sum_{i=1}^n (X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^n (X_i - \overline{X})^2},$$

and

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

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 $\hat{\beta}$, conditional on **X**, converges to

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

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

$$\mathbf{X}_i = \begin{pmatrix} 1 \\ X_i \end{pmatrix},$$

so we have

$$\mathbf{X}_i\mathbf{X}_i' = \begin{pmatrix} 1 \\ X_i \end{pmatrix} (1 \quad X_i) = \begin{pmatrix} 1 & X_i \\ X_i & X_i^2 \end{pmatrix},$$

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

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

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

and

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

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

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

The inverse of $\mathbf{Q_X}$ is

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

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

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

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

$$\begin{aligned}&\frac{1}{n[EX_i^2)-\mu_X^2]^2}\{[E(X_i^2)]^2\ \text{var}(u_i) - 2E(X_i^2)\mu_X\ \text{cov}(X_iu_i,\, u_i) + \mu_X^2\ \text{var}(X_iu_i)\}\\ &= \frac{1}{n[E(X_i^2)-\mu_X^2]^2}\text{var}[E(X_i^2)\, u_i - \mu_X X_iu_i]\\ &= \frac{[E(X_i^2)]^2}{n[E(X_i^2)-\mu_X^2]^2}\text{var}\left[u_i - \frac{\mu_X}{E(X_i^2)}X_iu_i\right]\\ &= \frac{1}{n\left[1-\frac{\mu_X^2}{E(X_i^2)}\right]^2}\text{var}\left[\left(1-\frac{\mu_X}{E(X_i^2)}X_i\right)u_i\right]\\ &= \frac{\text{var}(H_iu_i)}{n(E(H_i^2)]^2}, \text{(the same as the expression for } \sigma_{\beta_0}^2 \text{ given in Key Concept 4.4)}\end{aligned}$$

by defining

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

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

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

so

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

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

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

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

$\mathbf{M_X}$ is idempotent because

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

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

$$\mathbf{P_X M_X} \;=\; \mathbf{P_X}\,(\mathbf{I}_{\mathrm{n}} \;\; \mathbf{P_X}) = \mathbf{P_X} - \mathbf{P_X P_X} = \mathbf{P_X} - \mathbf{P_X} = 0_{n\times n}$$

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

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

which is Equation (18.27). The residual vector is

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

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

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

so the residual vector can be further written as

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

which is Equation (18.28).

18.6. The matrix form for Equation of (10.14) is

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

with

$$\tilde{\mathbf{Y}} = \begin{pmatrix} Y_{11} - \overline{Y}_1 \\ Y_{12} - Y_1 \\ \vdots \\ Y_{1T} - \overline{Y}_1 \\ Y_{21} - \overline{Y}_2 \\ Y_{22} - \overline{Y}_2 \\ \vdots \\ Y_{2T} - \overline{Y}_2 \\ \vdots \\ Y_{n1} - \overline{Y}_n \\ Y_{n2} - \overline{Y}_n \\ \vdots \\ Y_{nT} - \overline{Y}_n \end{pmatrix}, \quad \tilde{\mathbf{X}} = \begin{pmatrix} X_{11} - \overline{X}_1 \\ X_{12} - \overline{X}_1 \\ \vdots \\ X_{1T} - \overline{X}_1 \\ X_{21} - \overline{X}_2 \\ X_{22} - \overline{X}_2 \\ \vdots \\ X_{2T} - \overline{X}_2 \\ \vdots \\ X_{n1} - \overline{X}_n \\ X_{n2} - \overline{X}_n \\ \vdots \\ X_{nT} - \overline{X}_n \end{pmatrix}, \quad \tilde{\mathbf{U}} = \begin{pmatrix} u_{11} - \overline{u}_1 \\ u_{12} - \overline{u}_1 \\ \vdots \\ u_{1T} - \overline{u}_1 \\ u_{21} - \overline{u}_2 \\ u_{22} - \overline{u}_2 \\ \vdots \\ u_{2T} - \overline{u}_2 \\ \vdots \\ u_{n1} - \overline{u}_n \\ u_{n2} - \overline{u}_n \\ \vdots \\ u_{nT} - \overline{u}_n \end{pmatrix}$$

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

The OLS "de-meaning" fixed effects estimator is

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

Rewrite Equation (10.11) using n fixed effects as

$$Y_{it} = X_{it}\boldsymbol{\beta}_1 + D1_i\gamma_1 + D2_i\gamma_2 + \cdots + Dn_i\gamma_n + u_{it}.$$

In matrix form this is

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

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

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

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

Using the expression for $\hat{\boldsymbol{\beta}}$ given in the question, we have the estimator

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

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

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

and

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

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

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

18.7. (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

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

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); $\frac{1}{n}\sum_{i=1}^n X_i u_i \xrightarrow{p} E(Xu) = 0$ Thus

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

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

(c) Consider the population linear regression 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

$$\begin{pmatrix} \sqrt{n}(\hat{\beta}_1 - \beta_1) \\ \sqrt{n}(\hat{\beta}_2 - \theta) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_{i=1}^n X_i^2 & \frac{1}{n}\sum_{i=1}^n X_iW_i \\ \frac{1}{n}\sum_{i=1}^n W_iX_i & \frac{1}{n}\sum_{i=1}^n W_i^2 \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sqrt{n}}\sum_{i=1}^n X_ia_i \\ \frac{1}{\sqrt{n}}\sum_{i=1}^n W_ua_i \end{pmatrix}$$

$$\xrightarrow{d} \begin{pmatrix} E(X^2) & 0 \\ 0 & E(W^2) \end{pmatrix}^{-1} \begin{pmatrix} S_1 \\ S_2 \end{pmatrix}$$

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

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

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

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

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

18.8. (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

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

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 ($i = 1, 2, \ldots, n$):

$$E(u_i) = \sum_{j=1}^{i} 0.5^{i-j} E(\tilde{u}_j) = 0,$$

$$\sigma_i^2 = \mathrm{var}(u_i) = \sum_{j=1}^{i} (0.5^{i-j})^2 \,\mathrm{var}(\tilde{u}_j) = \sum_{j=1}^{i} (0.5^2)^{i-j} \times 1 = \frac{1-(0.5^2)^i}{1-0.5^2}.$$

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

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

Similarly we can get

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

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

$$E(\mathbf{UU}') = \begin{pmatrix} E(u_1^2) & E(u_1u_2) & \cdots & E(u_1u_n) \\ E(u_2u_1) & E(u_2^2) & \cdots & E(u_2u_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(u_nu_1) & E(u_nu_2) & \cdots & E(u_n^2) \end{pmatrix}.$$

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

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

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

(b) The original regression model is

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

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

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

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

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

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

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

using data for $i = 1,\ldots, n$. 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.

18.9. (a) $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{M_W X})^{-1}\mathbf{X}'\mathbf{M_W Y}$

$= (\mathbf{X}'\mathbf{M_W X})^{-1}\mathbf{X}'\mathbf{M_W}(\mathbf{X}\boldsymbol{\beta} + \mathbf{W}\boldsymbol{\gamma} + \mathbf{U})$

$= \boldsymbol{\beta} + (\mathbf{X}'\mathbf{M_W X})^{-1}\mathbf{X}'\mathbf{M_W U}.$

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

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

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

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

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 $\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_{ji}\mathbf{X}_{li}$. 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:

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

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

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

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

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

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

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

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

and

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

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

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

which is finite and invertible.

(c) The conditional expectation

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

$$= \begin{pmatrix} \mathbf{W}_1'\delta \\ \mathbf{W}_2'\delta \\ \vdots \\ \mathbf{W}_n'\delta \end{pmatrix} = \begin{pmatrix} \mathbf{W}_1' \\ \mathbf{W}_2' \\ \vdots \\ \mathbf{W}_n' \end{pmatrix}\delta = \mathbf{W}\delta.$$

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

(d) In the limit

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

because $\mathbf{M}_\mathbf{W}\mathbf{W} = \mathbf{0}$.

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

$$\hat{\beta} - \beta = (n^{-1}\mathbf{X}'\mathbf{M}_\mathbf{W}\mathbf{X})^{-1}\,(n^{-1}\mathbf{X}'\mathbf{M}_\mathbf{W}\mathbf{U}) \xrightarrow{p} \mathbf{0}.$$

This implies

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

18.10. (a) Using the hint: $\boldsymbol{Cq} = \lambda\boldsymbol{q}$, so that $\mathbf{0} = \boldsymbol{Cq} - \lambda\boldsymbol{q} = \boldsymbol{CCq} - \lambda\boldsymbol{q} = \lambda\boldsymbol{Cq} - \lambda\boldsymbol{q} = \lambda^2\boldsymbol{q} - \lambda\boldsymbol{q} = \lambda(1-\lambda)\boldsymbol{q}$, and the result follows by inspection.

(b) The trace of a matrix is equal to sum of its eigenvalues. The rank of a matrix is equal to the number of non-zero eigenvalues. Thus, the result follows from (a).

(c) Because $\boldsymbol{C}$ is symmetric with non-negative eigenvalues, $\boldsymbol{C}$ is positive semidefinite, and the result follows.

18.11. (a) Using the hint $\boldsymbol{C} = [\boldsymbol{Q}_1\,\boldsymbol{Q}_2]\begin{bmatrix} \mathbf{I}_r & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} \mathbf{Q}_1' \\ \mathbf{Q}_2' \end{bmatrix}$, where $\boldsymbol{Q}'\boldsymbol{Q} = \boldsymbol{I}$. The result follows with $\boldsymbol{A}=\boldsymbol{Q}_1$.

(b) $\boldsymbol{W} = \boldsymbol{A}'\boldsymbol{V} \sim \mathrm{N}(\boldsymbol{A}'\boldsymbol{0},\,\boldsymbol{A}'\boldsymbol{I}_n\boldsymbol{A})$ and the result follows immediately.

(c) $\boldsymbol{V}'\boldsymbol{CV} = \boldsymbol{V}'\boldsymbol{AA}'\boldsymbol{V} = (\boldsymbol{A}'\boldsymbol{V})'(\boldsymbol{A}'\boldsymbol{V}) = \boldsymbol{W}'\boldsymbol{W}$ and the result follows from (b).

18.12. (a) and (b) These mimic the steps using TSLS.

18.13. (a) This follows from the definition of the Lagrangian.

(b) The first order conditions are

$$\textbf{(*) } \mathbf{X}'(\boldsymbol{Y}-\boldsymbol{X}\tilde{\boldsymbol{\beta}}) + \boldsymbol{R}'\boldsymbol{\lambda} = \mathbf{0}$$

and

$$\textbf{(**) } \boldsymbol{R}\tilde{\boldsymbol{\beta}} - \boldsymbol{r} = \mathbf{0}$$

Solving (*) yields

$$(***)\ \tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} + (\mathbf{X}'\mathbf{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}.$$

Multiplying by $\boldsymbol{R}$ and using (**) yields $\boldsymbol{r} = \boldsymbol{R}\hat{\boldsymbol{\beta}} + \boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'\boldsymbol{\lambda}$, so that

$$\boldsymbol{\lambda} = -[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}).$$

Substituting this into (***) yields the result.

(c) Using the result in (b), $\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}} = (\mathbf{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})$, so that

$$(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}}) = (\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) + (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})$$
$$+ 2(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}).$$

But $(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{X} = \mathbf{0}$, so the last term vanishes, and the result follows.

(d) The result in (c) shows that $(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}']^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) = SSR_{Restricted} - SSR_{Unrestricted}$. Also $s_u^2 = SSR_{Unrestricted}/(n - k_{Unrestricted} - 1)$, and the result follows immediately.

18.14. (a) $\hat{\boldsymbol{\beta}}'(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} = \boldsymbol{Y}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \boldsymbol{Y}'\boldsymbol{X}_1\boldsymbol{H}\boldsymbol{X}_1'\boldsymbol{Y}$, where $\boldsymbol{H}$ is the upper $k_1 \times k_1$ block of $(\boldsymbol{X}'\boldsymbol{X})^{-1}$. Also $\boldsymbol{R}\hat{\boldsymbol{\beta}} = \boldsymbol{H}\boldsymbol{X}_1'\boldsymbol{Y}$ and $\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}' = \boldsymbol{H}$. Thus $(\boldsymbol{R}\hat{\boldsymbol{\beta}})'[\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}]^{-1}(\boldsymbol{R}\hat{\boldsymbol{\beta}}) = \boldsymbol{Y}'\boldsymbol{X}_1\boldsymbol{H}\boldsymbol{X}_1'\boldsymbol{Y}$.

(b) (i) Write the second stage regression as $Y = \hat{X}\boldsymbol{\beta} + U$, where $\hat{X}$ and the fitted values from the first stage regression. Note that $\hat{U}\hat{X} = \mathbf{0}$ where $\hat{U} = Y - \hat{X}\hat{\boldsymbol{\beta}}$ because OLS residual are orthogonal to the regressors. Now $\hat{U}^{TSLS} = Y - X\hat{\boldsymbol{\beta}} = \hat{U} - (X - \hat{X})\hat{\boldsymbol{\beta}} = \hat{U} - \hat{V}\hat{\boldsymbol{\beta}}$, where $\hat{V}$ is the residual from the first stage regression. But, since $\boldsymbol{W}$ is a regressor in the first stage regression, $\hat{V}'W = \mathbf{0}$ Thus $\hat{U}^{TSLS}{}'W = \hat{U}'W - \hat{\boldsymbol{\beta}}'\hat{V}'W = \mathbf{0}$.

(ii) $\hat{\boldsymbol{\beta}}'(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} = (\mathbf{R}\hat{\boldsymbol{\beta}})'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}}) = SSR_{Rest} - SSR_{Unrest}$ for the regression in KC 12.6, and the result follows directly.

18.15. (a) This follows from exercise (18.6).

(b) $\tilde{\mathbf{Y}}_i = \tilde{\mathbf{X}}_i\boldsymbol{\beta} + \tilde{\mathbf{u}}_i$, so that

$$\begin{aligned}\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} &= \left(\sum_{i=1}^{n} \tilde{\mathbf{X}}_{\mathbf{i}}{}'\tilde{\mathbf{X}}_{\mathbf{i}}\right)^{-1} \sum_{i=1}^{n} \tilde{\mathbf{X}}_{\mathbf{i}}{}'\tilde{\mathbf{u}}_{\mathbf{i}} \\ &= \left(\sum_{i=1}^{n} \tilde{\mathbf{X}}_{\mathbf{i}}{}'\tilde{\mathbf{X}}_{\mathbf{i}}\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_{\mathbf{i}}{}'\mathbf{M}'\mathbf{M}\mathbf{u}_{\mathbf{i}} \\ &= \left(\sum_{i=1}^{n} \tilde{\mathbf{X}}_{\mathbf{i}}{}'\tilde{\mathbf{X}}_{\mathbf{i}}\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_{\mathbf{i}}{}'\mathbf{M}'\mathbf{u}_{\mathbf{i}} \\ &= \left(\sum_{i=1}^{n} \tilde{\mathbf{X}}_{\mathbf{i}}{}'\tilde{\mathbf{X}}_{\mathbf{i}}\right)^{-1} \sum_{i=1}^{n} \tilde{\mathbf{X}}_{\mathbf{i}}{}'\mathbf{u}_{\mathbf{i}}\end{aligned}$$

(c) $\hat{Q}_{\tilde{X}} = \frac{1}{n}\sum_{i=1}^{n}(T^{-1}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)^2)$, where $(T^{-1}\sum_{t=1}^{T}(X_{it} - \bar{X}_i)^2)$ are i.i.d. with mean $\mathbf{Q}_{\tilde{\mathbf{X}}}$ and finite variance (because X_{it} has finite fourth moments). The result then follows from the law of large numbers.

(d) This follows the Central limit theorem.

(e) This follows from Slutsky's theorem.

(f) η_i^2 are i.i.d., and the result follows from the law of large numbers.

(g) Let $\hat{\eta}_i = T^{-1/2}\tilde{\mathbf{X}}_i{}'\hat{\tilde{\mathbf{u}}}_i = \eta_i - T^{-1/2}(\hat{\beta} - \beta)\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i$. Then

$$\hat{\eta}_i^2 = T^{-1/2}\tilde{\mathbf{X}}_i{}'\hat{\tilde{\mathbf{u}}}_i = \eta_i^2 + T^{-1}(\hat{\beta} - \beta)^2(\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i)^2 - 2T^{-1/2}(\hat{\beta} - \beta)\eta_i\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i$$

and $\frac{1}{n}\sum_{i=1}^{n}\hat{\eta}_i^2 - \frac{1}{n}\sum_{i=1}^{n}\eta_i^2 = T^{-1}(\hat{\beta} - \beta)^2 \frac{1}{n}\sum_{i=1}^{n}(\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i)^2 - 2T^{-1/2}(\hat{\beta} - \beta)\frac{1}{n}\sum_{i=1}^{n}\eta_i\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i$

Because $(\hat{\beta} - \beta)\xrightarrow{p}0$, the result follows from (a) $\frac{1}{n}\sum_{i=1}^{n}(\tilde{X}_i{}'\tilde{X}_i)^2 \xrightarrow{p} E[(\tilde{X}_i{}'\tilde{X}_i)^2]$ and (b) $\frac{1}{n}\sum_{i=1}^{n}\eta_i\tilde{X}_i{}'\tilde{X}_i \xrightarrow{p} E(\eta_i\tilde{X}_i{}'\tilde{X}_i)$. Both (a) and (b) follow from the law of large numbers; both (a) and (b) are averages of i.i.d. random variables. Completing the proof requires verifying that $(\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i)^2$ has two finite moments and $\eta_i\tilde{\mathbf{X}}_i{}'\tilde{\mathbf{X}}_i$ has two finite moments. These in turn follow from 8-moment assumptions for (X_{it}, u_{it}) and the Cauchy-Schwartz inequality. Alternatively, a "strong" law of large numbers can be used to show the result with finite fourth moments.

18.16 (a) Using analysis like that Appendices 4.2 and 4.3

$$\hat{\beta} = \frac{\sum X_i Y_i}{\sum X_i^2} = \beta + \frac{\sum X_i u_i}{\sum X_i^2} = \beta + \frac{\frac{1}{n}\sum X_i u_i}{\frac{1}{n}\sum X_i^2}$$

Consistency follows by analyzing the averages in the expression:

$$\frac{1}{n}\sum X_i u_i \xrightarrow{p} E(X_i u_i) = 0, \qquad \frac{1}{n}\sum X_i^2 \xrightarrow{p} E(X_i^2) > 0,$$

and Slutsky's theorem.

To show unbiasness, note that

$$E\left(\frac{\sum X_i u_i}{\sum X_i^2}\right) = E\left(E\left(\frac{\sum X_i u_i}{\sum X_i^2} | X_1, \ldots, X_n\right)\right)$$
$$= E\left(\frac{\sum X_i E(u_i | X_1, \ldots, X_n)}{\sum X_i^2}\right) = E\left(\frac{\sum X_i E(u_i | X_i)}{\sum X_i^2}\right) = 0$$

(b) (i) The result follows from the expression given in (a) and the definition of I_i.

(ii) – (iv)

As in (a), consistency follows by analyzing the sample averages in

$$\hat{\beta} = \frac{\sum I_i X_i Y_i}{\sum I_i X_i^2} = \beta + \frac{\sum I_i X_i u_i}{\sum I_i X_i^2} = \beta + \frac{\frac{1}{n}\sum I_i X_i u_i}{\frac{1}{n}\sum I_i X_i^2}.$$

Specifically, we have $\frac{1}{n}\sum I_i X_i u_i \xrightarrow{p} E(I_i X_i u_i)$ and $\frac{1}{n}\sum I_i X_i^2 \xrightarrow{p} E(I_i X_i^2)$.

In parts (ii)-(iv), $E(I_i X_i^2) > 0$.

In part (ii) $E(I_iX_iu_i) = E(E(I_iX_iu_i \mid X_i, u_i)) = pE(X_iu_i) = pE(E(X_iu_i \mid X_i)) = 0$

In part (iii) $E(I_iX_iu_i) = E(E(I_iX_iu_i \mid X_i, u_i)) = E(p(X_i)X_iu_i) = pE(E(p(X_i)X_iu_i \mid X_i)) = 0$.

In part (iv) $E(I_iX_iu_i) = E(E(I_iX_iu_i \mid X_i, u_i)) = E(p(u_i,X_i)X_iu_i) = pE(E(p(u_i,X_i)X_iu_i \mid X_i)) \neq 0$.

So that $\hat{\beta}$ is consistent in (ii) and (iii), but is not consistent in (iv).

The unbiasness analysis proceeds as in part (a), but it is necessary to condition on both I_i and X_i when carrying out the law of iterated expectations. A key result is that $E(u_i \mid X_i, I_i) = 0$ in (ii) and (iii), but not in (iv). This yields unbiasness in (ii) and (iii), but not in (iv). To see that $E(u_i \mid X_i, I_i) = 0$ in (ii) and (iii), notice that the conditional distribution of u given X, I can be written as

$$f(u \mid X, I) = \frac{f(u, I \mid X)}{f(I \mid X)} = \frac{f(I \mid X, u) f(u \mid X)}{f(I \mid X)} = \frac{f(I \mid X) f(u \mid X)}{f(I \mid X)} = f(u \mid X)$$

where the first equality is the definition of the conditional density, the second is Bayes rule, and the third equality follows because $\Pr(I = 1 \mid X, u)$ does not depend on u in (ii) and (iii). This implies that $E(u_i \mid X_i, I_i) = E(u_i \mid X_i) = 0$. Unbiasness then follows in (ii) and (iii) using an argument analogous that in part (a).

(c) In this example, $I_i = 1$ if $Y_i \geq 0$, or $X_i + u_i \geq 0$, or $u_i \geq -X_i$. Notice that $E(I_iX_iu_i|X_i) = X_iE(u_i|u_i \geq -X_i)$. A calculation based on the normal distribution shows that

$$E(u \mid u \geq -X) = \frac{\phi(-X)}{1 - \Phi(-X)}$$

where Φ is the standard normal CDF and ϕ is standard normal density.
Because $E(u|u \geq -X) \neq 0$, the OLS estimator is biased and inconsistent.

18.17 The results follow from the hints and matrix multiplication and addition.