Solutions Manual

to Accompany

Introduction to
Econometrics

Stock • Watson

Jiangfeng Zhang

University of California at Berkeley

Addison Wesley

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

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

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

Solutions for Selected Exercises

prepared by Jiangfeng Zhang

July 2002

Chapter 2
Review of Probability

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

(a)

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

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

(b)

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

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

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

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

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

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

(a)

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

(c) We calculate the conditional probabilities first:

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

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

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

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

The conditional expectations are

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

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

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

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

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

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

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

(c)

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

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

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

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

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

Chapter 3
Review of Statistics

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

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

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

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

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

(c) The computed t-statistic is

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

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

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

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

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

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

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

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

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

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

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

The t-statistic is

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

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

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

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.

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

(b)

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

Chapter 4
Linear Regression with One Regressor

4.1. (a) The predicted average test score is

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

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

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

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

(d) Calculate the t-statistic first

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

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

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

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

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

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

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

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

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

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

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

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

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

Chapter 5
Linear Regression with Multiple Regressors

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

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

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

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

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

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

Chapter 6
Nonlinear Regression Functions

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

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

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

6.4. Note that

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

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

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

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

Chapter 7
Assessing Studies Based on Multiple Regression

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

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

Chapter 8
Regression with Panel Data

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

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

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

Chapter 9
Regression with a Binary Dependent Variable

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

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

Chapter 10
Instrumental Variables Regression

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

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

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

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

Chapter 11
Experiments and Quasi-Experiments

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

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

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

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

11.3. (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.

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

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

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

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

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

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

Chapter 12
Introduction to Time Series Regression and Forecasting

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

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

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

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

(b) The 1st autocovariance is

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

The 2nd autocovariance is

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

(c) The 1st autocorrelation is

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

The 2nd autocorrelation is

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

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

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

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

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

Chapter 13
Estimation of Dynamic Causal Effects

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

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

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

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

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

percentage points.

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

Chapter 14
Additional Topics in Time Series Regression

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

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

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

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

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

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

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

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

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

Chapter 15
The Theory of Linear Regression with One Regressor

15.2. The sample covariance is

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

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

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

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

which is satisfied since

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

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

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

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

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

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

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

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

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

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

Chapter 16
The Theory of Multiple Regression

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

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

with

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

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

(b) The null hypothesis is

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

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

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

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

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

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

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

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

16.6. The matrix form for Equation (8.14) is

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

with

$$\begin{aligned}
\tilde{\mathbf{Y}} &= \begin{pmatrix} Y_{11}-\bar{Y}_1 \\ Y_{12}-\bar{Y}_1 \\ \vdots \\ Y_{1T}-\bar{Y}_1 \\ Y_{21}-\bar{Y}_2 \\ Y_{22}-\bar{Y}_2 \\ \vdots \\ Y_{2T}-\bar{Y}_2 \\ \vdots \\ Y_{n1}-\bar{Y}_n \\ Y_{n2}-\bar{Y}_n \\ \vdots \\ Y_{nT}-\bar{Y}_n \end{pmatrix}, \qquad \tilde{\mathbf{X}} = \begin{pmatrix} X_{11}-\bar{X}_1 \\ X_{12}-\bar{X}_1 \\ \vdots \\ X_{1T}-\bar{X}_1 \\ X_{21}-\bar{X}_2 \\ X_{22}-\bar{X}_2 \\ \vdots \\ X_{2T}-\bar{X}_2 \\ \vdots \\ X_{n1}-\bar{X}_n \\ X_{n2}-\bar{X}_n \\ \vdots \\ X_{nT}-\bar{X}_n \end{pmatrix}, \qquad \tilde{\mathbf{U}} = \begin{pmatrix} u_{11}-\bar{u}_1 \\ u_{12}-\bar{u}_1 \\ \vdots \\ u_{1T}-\bar{u}_1 \\ u_{21}-\bar{u}_2 \\ u_{22}-\bar{u}_2 \\ \vdots \\ u_{2T}-\bar{u}_2 \\ \vdots \\ u_{n1}-\bar{u}_n \\ u_{n2}-\bar{u}_n \\ \vdots \\ u_{nT}-\bar{u}_n \end{pmatrix}, \\
\tilde{\boldsymbol{\beta}} &= \beta_1
\end{aligned}$$

The OLS "de-meaning" fixed effects estimator is

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

Rewrite Equation (8.11) using n fixed effects as

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

In matrix form this is

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

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

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

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

Using Equation (16.45), we have the estimator

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

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

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

and

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

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

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

16.8. (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 $u_{i-j} = 0.5u_{i-j-1} + \tilde{u}_{i-j}$ $(j = 1, 2, ...i-2)$ and $u_1 = \tilde{u}_1$ into the expression $u_i = 0.5u_{i-1} + \tilde{u}_i$ yields

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

Though we get the expression $u_i = \sum_{j=1}^{i} 0.5^{i-j}\tilde{u}_j$ for $i > 1$, it is apparent that it also holds for $i = 1$. Thus we can get mean and variance of random variables u_i $(i = 1, 2, \ldots, n)$:

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

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

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

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

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

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

Similarly we can get

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

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

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

It is straightforward to get

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

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

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

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