STUDENT SOLUTIONS MANUAL

Jeffrey M. Wooldridge

Introductory Econometrics: A Modern Approach, 4e

CONTENTS

PREFACE

This manual contains solutions to the odd-numbered problems and computer exercises in *Introductory Econometrics: A Modern Approach*, 4e. Hopefully, you will find that the solutions are detailed enough to act as a study supplement to the text. Rather than just presenting the final answer, I usually provide detailed steps, emphasizing where the chapter material is used in solving the problems.

Some of the answers given here are subjective, and you or your instructor may have perfectly acceptable alternative answers or opinions.

I obtained the solutions to the computer exercises using Stata, starting with version 4.0 and ending with version 9.0. Nevertheless, almost all of the estimation methods covered in the text have been standardized, and different econometrics or statistical packages should give the same answers to the reported degree of accuracy. There can be differences when applying more advanced techniques, as conventions sometimes differ on how to choose or estimate auxiliary parameters. (Examples include heteroskedasticity-robust standard errors, estimates of a random effects model, and corrections for sample selection bias.) Any differences in estimates or test statistics should be practically unimportant, provided you are using a reasonably large sample size.

While I have endeavored to make the solutions free of mistakes, some errors may have crept in. I would appreciate hearing from students who find mistakes. I will keep a list of any notable errors on the Web site for the book, www.international.cengage.com. I would also like to hear from students who have suggestions for improving either the solutions or the problems themselves. I can be reached via e-mail at wooldri1@.msu.edu.

I hope that you find this solutions manual helpful when used in conjunction with the text. I look forward to hearing from you.

Jeffrey M. Wooldridge
Department of Economics
Michigan State University
110 Marshall-Adams Hall
East Lansing, MI 48824-1038

CHAPTER 1

SOLUTIONS TO PROBLEMS

1.1 It does not make sense to pose the question in terms of causality. Economists would assume that students choose a mix of studying and working (and other activities, such as attending class, leisure, and sleeping) based on rational behavior, such as maximizing utility subject to the constraint that there are only 168 hours in a week. We can then use statistical methods to measure the association between studying and working, including regression analysis that we cover starting in Chapter 2. But we would not be claiming that one variable "causes" the other. They are both choice variables of the student.

1.2 (i) Ideally, we could randomly assign students to classes of different sizes. That is, each student is assigned a different class size without regard to any student characteristics such as ability and family background. For reasons we will see in Chapter 2, we would like substantial variation in class sizes (subject, of course, to ethical considerations and resource constraints).

(ii) A negative correlation means that larger class size is associated with lower performance. We might find a negative correlation because larger class size actually hurts performance. However, with observational data, there are other reasons we might find a negative relationship. For example, children from more affluent families might be more likely to attend schools with smaller class sizes, and affluent children generally score better on standardized tests. Another possibility is that, within a school, a principal might assign the better students to smaller classes. Or, some parents might insist their children are in the smaller classes, and these same parents tend to be more involved in their children's education.

(iii) Given the potential for confounding factors – some of which are listed in (ii) – finding a negative correlation would not be strong evidence that smaller class sizes actually lead to better performance. Some way of controlling for the confounding factors is needed, and this is the subject of multiple regression analysis.

SOLUTIONS TO COMPUTER EXERCISES

C1.1 (i) The average of *educ* is about 12.6 years. There are two people reporting zero years of education, and 19 people reporting 18 years of education.

(ii) The average of *wage* is about $5.90, which seems low in the year 2008.

(iii) Using Table B-60 in the 2004 *Economic Report of the President*, the CPI was 56.9 in 1976 and 184.0 in 2003.

(iv) To convert 1976 dollars into 2003 dollars, we use the ratio of the CPIs, which is $184/56.9 \approx 3.23$. Therefore, the average hourly wage in 2003 dollars is roughly $3.23(\$5.90) \approx \$19.06$, which is a reasonable figure.

(v) The sample contains 252 women (the number of observations with *female* = 1) and 274 men.

C1.3 (i) The largest is 100, the smallest is 0.

(ii) 38 out of 1,823, or about 2.1 percent of the sample.

(iii) 17

(iv) The average of *math4* is about 71.9 and the average of *read4* is about 60.1. So, at least in 2001, the reading test was harder to pass.

(v) The sample correlation between *math4* and *read4* is about .843, which is a very high degree of (linear) association. Not surprisingly, schools that have high pass rates on one test have a strong tendency to have high pass rates on the other test.

(vi) The average of *exppp* is about $5,194.87. The standard deviation is $1,091.89, which shows rather wide variation in spending per pupil. [The minimum is $1,206.88 and the maximum is $11,957.64.]

CHAPTER 2

SOLUTIONS TO PROBLEMS

2.2 (i) Let $y_i = GPA_i$, $x_i = ACT_i$, and $n = 8$. Then $\bar{x} = 25.875$, $\bar{y} = 3.2125$, $\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) =$ 5.8125, and $\sum_{i=1}^{n}(x_i - \bar{x})^2 = 56.875$. From equation (2.9), we obtain the slope as $\hat{\beta}_1 =$ 5.8125/56.875 $\approx$ .1022, rounded to four places after the decimal. From (2.17), $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} \approx 3.2125 - (.1022)25.875 \approx .5681$. So we can write

$$\widehat{GPA} = .5681 + .1022\, ACT$$
$$n = 8.$$

The intercept does not have a useful interpretation because *ACT* is not close to zero for the population of interest. If *ACT* is 5 points higher, $\widehat{GPA}$ increases by .1022(5) = .511.

(ii) The fitted values and residuals — rounded to four decimal places — are given along with the observation number *i* and *GPA* in the following table:

i	GPA	$\widehat{GPA}$	$\hat{u}$
1	2.8	2.7143	.0857
2	3.4	3.0209	.3791
3	3.0	3.2253	–.2253
4	3.5	3.3275	.1725
5	3.6	3.5319	.0681
6	3.0	3.1231	–.1231
7	2.7	3.1231	–.4231
8	3.7	3.6341	.0659

You can verify that the residuals, as reported in the table, sum to –.0002, which is pretty close to zero given the inherent rounding error.

(iii) When *ACT* = 20, $\widehat{GPA} = .5681 + .1022(20) \approx 2.61$.

(iv) The sum of squared residuals, $\sum_{i=1}^{n}\hat{u}_i^2$, is about .4347 (rounded to four decimal places), and the total sum of squares, $\sum_{i=1}^{n}(y_i - \bar{y})^2$, is about 1.0288. So the R-squared from the regression is

$$R^2 = 1 - \text{SSR/SST} \approx 1 - (.4347/1.0288) \approx .577.$$

Therefore, about 57.7% of the variation in *GPA* is explained by *ACT* in this small sample of students.

2.3 (i) Income, age, and family background (such as number of siblings) are just a few possibilities. It seems that each of these could be correlated with years of education. (Income and education are probably positively correlated; age and education may be negatively correlated because women in more recent cohorts have, on average, more education; and number of siblings and education are probably negatively correlated.)

(ii) Not if the factors we listed in part (i) are correlated with *educ*. Because we would like to hold these factors fixed, they are part of the error term. But if u is correlated with *educ* then $\mathrm{E}(u|educ) \neq 0$, and so SLR.4 fails.

2.4 (i) We would want to randomly assign the number of hours in the preparation course so that *hours* is independent of other factors that affect performance on the SAT. Then, we would collect information on SAT score for each student in the experiment, yielding a data set $\{(sat_i, hours_i) : i = 1, \ldots, n\}$, where n is the number of students we can afford to have in the study. From equation (2.7), we should try to get as much variation in $hours_i$ as is feasible.

(ii) Here are three factors: innate ability, family income, and general health on the day of the exam. If we think students with higher native intelligence think they do not need to prepare for the SAT, then ability and *hours* will be negatively correlated. Family income would probably be positively correlated with *hours*, because higher income families can more easily afford preparation courses. Ruling out chronic health problems, health on the day of the exam should be roughly uncorrelated with hours spent in a preparation course.

(iii) If preparation courses are effective, β_1 should be positive: other factors equal, an increase in *hours* should increase *sat*.

(iv) The intercept, β_0, has a useful interpretation in this example: because $\mathrm{E}(u) = 0$, β_0 is the average SAT score for students in the population with *hours* = 0.

2.5 (i) When we condition on *inc* in computing an expectation, $\sqrt{inc}$ becomes a constant. So $E(u|inc) = E(\sqrt{inc} \cdot e|inc) = \sqrt{inc} \cdot E(e|inc) = \sqrt{inc} \cdot 0$ because $E(e|inc) = E(e) = 0$.

(ii) Again, when we condition on *inc* in computing a variance, $\sqrt{inc}$ becomes a constant. So $\mathrm{Var}(u|inc) = \mathrm{Var}(\sqrt{inc} \cdot e|inc) = (\sqrt{inc})^2 \mathrm{Var}(e|inc) = \sigma_e^2\, inc$ because $\mathrm{Var}(e|inc) = \sigma_e^2$.

(iii) Families with low incomes do not have much discretion about spending; typically, a low-income family must spend on food, clothing, housing, and other necessities. Higher income people have more discretion, and some might choose more consumption while others more saving. This discretion suggests wider variability in saving among higher income families.

2.8 (i) We follow the hint, noting that $\overline{c_1 y} = c_1\overline{y}$ (the sample average of $c_1 y_i$ is c_1 times the sample average of y_i) and $\overline{c_2 x} = c_2\overline{x}$. When we regress $c_1 y_i$ on $c_2 x_i$ (including an intercept) we use equation (2.19) to obtain the slope:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(c_2x_i - c_2x)(c_1\overline{y}_i - c_1\overline{y})}{\sum_{i=1}^{n}(c_2x_i - c_2\overline{x})^2} = \frac{\sum_{i=1}^{n} c_1c_2(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} c_2^2(x_i - \overline{x})^2}$$

$$= \frac{c_1}{c_2} \cdot \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2} = \frac{c_1}{c_2}\hat{\beta}_1.$$

From (2.17), we obtain the intercept as $\tilde{\beta}_0 = (c_1\overline{y}) - \tilde{\beta}_1(c_2\overline{x}) = (c_1\overline{y}) - [(c_1/c_2)\hat{\beta}_1](c_2\overline{x}) = c_1(\overline{y} - \hat{\beta}_1\overline{x}) = c_1\hat{\beta}_0)$ because the intercept from regressing y_i on x_i is $(\overline{y} - \hat{\beta}_1\overline{x})$.

(ii) We use the same approach from part (i) along with the fact that $(\overline{c_1 + y}) = c_1 + \overline{y}$ and $(\overline{c_2 + x}) = c_2 + \overline{x}$. Therefore, $(c_1 + y_i) - (\overline{c_1 + y}) = (c_1 + y_i) - (c_1 + \overline{y}) = y_i - \overline{y}$ and $(c_2 + x_i) - (\overline{c_2 + x}) = x_i - \overline{x}$. So c_1 and c_2 entirely drop out of the slope formula for the regression of $(c_1 + y_i)$ on $(c_2 + x_i)$, and $\tilde{\beta}_1 = \hat{\beta}_1$. The intercept is $\tilde{\beta}_0 = (\overline{c_1 + y}) - \tilde{\beta}_1(\overline{c_2 + x}) = (c_1 + \overline{y}) - \hat{\beta}_1(c_2 + \overline{x}) = (\overline{y} - \hat{\beta}_1\overline{x}) + c_1 - c_2\hat{\beta}_1 = \hat{\beta}_0 + c_1 - c_2\hat{\beta}_1$, which is what we wanted to show.

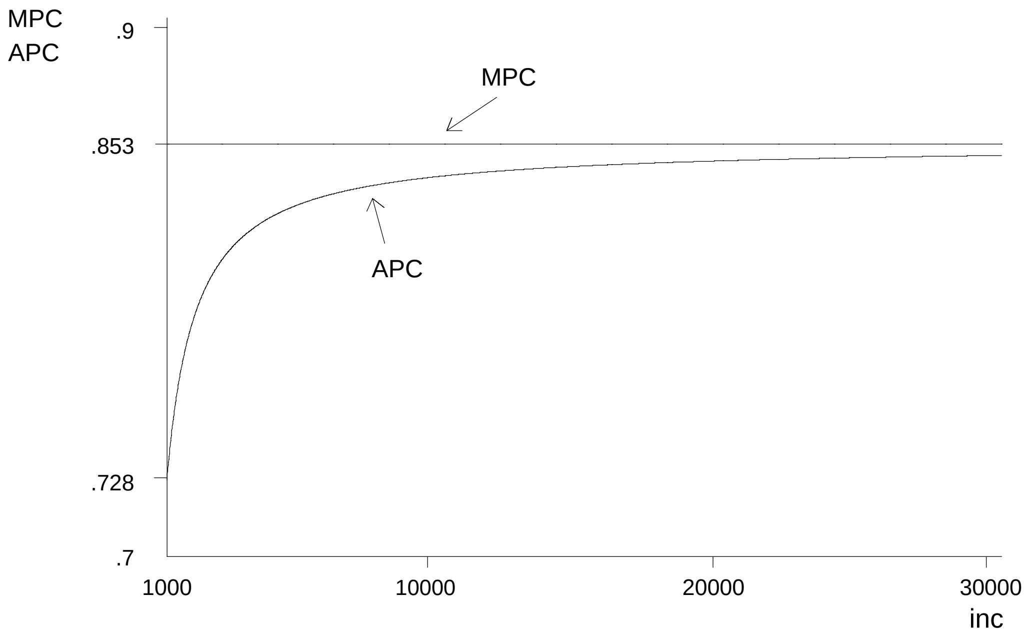
(iii) We can simply apply part (ii) because $\log(c_1 y_i) = \log(c_1) + \log(y_i)$. In other words, replace c_1 with $\log(c_1)$, y_i with $\log(y_i)$, and set $c_2 = 0$.

(iv) Again, we can apply part (ii) with $c_1 = 0$ and replacing c_2 with $\log(c_2)$ and x_i with $\log(x_i)$. If $\hat{\beta}_0$ and $\hat{\beta}_1$ are the original intercept and slope, then $\tilde{\beta}_1 = \hat{\beta}_1$ and $\tilde{\beta}_0 = \hat{\beta}_0 - \log(c_2)\hat{\beta}_1$.

2.9 (i) The intercept implies that when *inc* = 0, *cons* is predicted to be negative $124.84. This, of course, cannot be true, and reflects that fact that this consumption function might be a poor predictor of consumption at very low-income levels. On the other hand, on an annual basis, $124.84 is not so far from zero.

(ii) Just plug 30,000 into the equation: $\widehat{cons} = -124.84 + .853(30{,}000) = 25{,}465.16$ dollars.

(iii) The MPC and the APC are shown in the following graph. Even though the intercept is negative, the smallest APC in the sample is positive. The graph starts at an annual income level of $1,000 (in 1970 dollars).



SOLUTIONS TO COMPUTER EXERCISES

C2.1 (i) The average *prate* is about 87.36 and the average *mrate* is about .732.

(ii) The estimated equation is

$$\widehat{prate} = 83.05 + 5.86\ mrate$$

$$n = 1{,}534,\ \ R^2 = .075.$$

(iii) The intercept implies that, even if *mrate* = 0, the predicted participation rate is 83.05 percent. The coefficient on *mrate* implies that a one-dollar increase in the match rate – a fairly large increase – is estimated to increase *prate* by 5.86 percentage points. This assumes, of course, that this change *prate* is possible (if, say, *prate* is already at 98, this interpretation makes no sense).

(iv) If we plug *mrate* = 3.5 into the equation we get $\hat{prate}$ = 83.05 + 5.86(3.5) = 103.59. This is impossible, as we can have at most a 100 percent participation rate. This illustrates that, especially when dependent variables are bounded, a simple regression model can give strange predictions for extreme values of the independent variable. (In the sample of 1,534 firms, only 34 have *mrate* ≥ 3.5.)

(v) *mrate* explains about 7.5% of the variation in *prate*. This is not much, and suggests that many other factors influence 401(k) plan participation rates.

C2.3 (i) The estimated equation is

$$\widehat{sleep} = 3{,}586.4 - .151\ totwrk$$

$$n = 706,\ \ R^2 = .103.$$

The intercept implies that the estimated amount of sleep per week for someone who does not work is 3,586.4 minutes, or about 59.77 hours. This comes to about 8.5 hours per night.

(ii) If someone works two more hours per week then $\Delta totwrk$ = 120 (because *totwrk* is measured in minutes), and so $\Delta\widehat{sleep}$ = –.151(120) = –18.12 minutes. This is only a few minutes a night. If someone were to work one more hour on each of five working days, $\Delta\widehat{sleep}$ = –.151(300) = –45.3 minutes, or about five minutes a night.

C2.5 (i) The constant elasticity model is a log-log model:

$$\log(rd) = \beta_0 + \beta_1 \log(sales) + u,$$

where β_1 is the elasticity of *rd* with respect to *sales*.

(ii) The estimated equation is

$$\widehat{\log(rd)} = -4.105 + 1.076 \log(sales)$$
$$n = 32, \quad R^2 = .910.$$

The estimated elasticity of *rd* with respect to *sales* is 1.076, which is just above one. A one percent increase in *sales* is estimated to increase *rd* by about 1.08%.

C2.7 (i) The average gift is about 7.44 Dutch guilders. Out of 4,268 respondents, 2,561 did not give a gift, or about 60 percent.

(ii) The average mailings per year is about 2.05. The minimum value is .25 (which presumably means that someone has been on the mailing list for at least four years) and the maximum value is 3.5.

(iii) The estimated equation is

$$\widehat{gift} = 2.01 + 2.65\ mailsyear$$
$$n = 4{,}268, \quad R^2 = .0138$$

(iv) The slope coefficient from part (iii) means that each mailing per year is associated with – perhaps even "causes" – an estimated 2.65 additional guilders, on average. Therefore, if each mailing costs one guilder, the expected profit from each mailing is estimated to be 1.65 guilders. This is only the average, however. Some mailings generate no contributions, or a contribution less than the mailing cost; other mailings generated much more than the mailing cost.

(v) Because the smallest *mailsyear* in the sample is .25, the smallest predicted value of *gifts* is 2.01 + 2.65(.25) ≈ 2.67. Even if we look at the overall population, where some people have received no mailings, the smallest predicted value is about two. So, with this estimated equation, we never predict zero charitable gifts.

CHAPTER 3

SOLUTIONS TO PROBLEMS

3.2 (i) *hsperc* is defined so that the smaller it is, the lower the student's standing in high school. Everything else equal, the worse the student's standing in high school, the lower is his/her expected college GPA.

(ii) Just plug these values into the equation:

$$\widehat{colgpa} = 1.392 - .0135(20) + .00148(1050) = 2.676.$$

(iii) The difference between A and B is simply 140 times the coefficient on *sat*, because *hsperc* is the same for both students. So A is predicted to have a score .00148(140) $\approx$ .207 higher.

(iv) With *hsperc* fixed, $\Delta\widehat{colgpa}$ = .00148Δsat. Now, we want to find Δsat such that $\Delta\widehat{colgpa}$ = .5, so .5 = .00148(Δsat) or Δsat = .5/(.00148) $\approx$ 338. Perhaps not surprisingly, a large ceteris paribus difference in SAT score – almost two and one-half standard deviations – is needed to obtain a predicted difference in college GPA or a half a point.

3.4 (i) If adults trade off sleep for work, more work implies less sleep (other things equal), so $\beta_1 < 0$.

(ii) The signs of β_2 and β_3 are not obvious, at least to me. One could argue that more educated people like to get more out of life, and so, other things equal, they sleep less ($\beta_2 < 0$). The relationship between sleeping and age is more complicated than this model suggests, and economists are not in the best position to judge such things.

(iii) Since *totwrk* is in minutes, we must convert five hours into minutes: $\Delta totwrk$ = 5(60) = 300. Then *sleep* is predicted to fall by .148(300) = 44.4 minutes. For a week, 45 minutes less sleep is not an overwhelming change.

(iv) More education implies less predicted time sleeping, but the effect is quite small. If we assume the difference between college and high school is four years, the college graduate sleeps about 45 minutes less per week, other things equal.

(v) Not surprisingly, the three explanatory variables explain only about 11.3% of the variation in *sleep*. One important factor in the error term is general health. Another is marital status, and whether the person has children. Health (however we measure that), marital status, and number and ages of children would generally be correlated with *totwrk*. (For example, less healthy people would tend to work less.)

3.6 (i) No. By definition, *study* + *sleep* + *work* + *leisure* = 168. Therefore, if we change *study*, we must change at least one of the other categories so that the sum is still 168.

(ii) From part (i), we can write, say, *study* as a perfect linear function of the other independent variables: *study* = 168 − *sleep* − *work* − *leisure*. This holds for every observation, so MLR.3 violated.

(iii) Simply drop one of the independent variables, say *leisure*:

$$GPA = \beta_0 + \beta_1 study + \beta_2 sleep + \beta_3 work + u.$$

Now, for example, β_1 is interpreted as the change in *GPA* when *study* increases by one hour, where *sleep*, *work*, and *u* are all held fixed. If we are holding *sleep* and *work* fixed but increasing *study* by one hour, then we must be reducing *leisure* by one hour. The other slope parameters have a similar interpretation.

3.8 Only (ii), omitting an important variable, can cause bias, and this is true only when the omitted variable is correlated with the included explanatory variables. The homoskedasticity assumption, MLR.5, played no role in showing that the OLS estimators are unbiased. (Homoskedasticity was used to obtain the usual variance formulas for the $\hat{\beta}_j$.) Further, the degree of collinearity between the explanatory variables in the sample, even if it is reflected in a correlation as high as .95, does not affect the Gauss-Markov assumptions. Only if there is a *perfect* linear relationship among two or more explanatory variables is MLR.3 violated.

3.10 From equation (3.22) we have

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} y_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2},$$

where the $\hat{r}_{i1}$ are defined in the problem. As usual, we must plug in the true model for y$_i$:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n} \hat{r}_{i1} (\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + u_i}{\sum_{i=1}^{n} \hat{r}_{i1}^2}.$$

The numerator of this expression simplifies because $\sum_{i=1}^{n}\hat{r}_{i1} = 0$, $\sum_{i=1}^{n}\hat{r}_{i1}x_{i2} = 0$, and $\sum_{i=1}^{n}\hat{r}_{i1}x_{i1} = \sum_{i=1}^{n}\hat{r}_{i1}^2$. These all follow from the fact that the $\hat{r}_{i1}$ are the residuals from the regression of x_{i1} on x_{i2}: the $\hat{r}_{i1}$ have zero sample average and are uncorrelated in sample with x_{i2}. So the numerator of $\tilde{\beta}_1$ can be expressed as

$$\beta_1\sum_{i=1}^{n}\hat{r}_{i1}^2 + \beta_3\sum_{i=1}^{n}\hat{r}_{i1}x_{i3} + \sum_{i=1}^{n}\hat{r}_{i1}u_i.$$

Putting these back over the denominator gives

$$\tilde{\beta}_1 = \beta_1 + \beta_3\frac{\sum_{i=1}^{n}\hat{r}_{i1}x_{i3}}{\sum_{i=1}^{n}\hat{r}_{i1}^2} + \frac{\sum_{i=1}^{n}\hat{r}_{1}u_i}{\sum_{i=1}^{n}\hat{r}_{i1}^2}.$$

Conditional on all sample values on x_1, x_2, and x_3, only the last term is random due to its dependence on u_i. But $E(u_i) = 0$, and so

$$E(\tilde{\beta}_1) = \beta_1 + \beta_3\frac{\sum_{i=1}^{n}\hat{r}_{i1}x_{i3}}{\sum_{i=1}^{n}\hat{r}_{i1}^2},$$

which is what we wanted to show. Notice that the term multiplying β_3 is the regression coefficient from the simple regression of x_{i3} on $\hat{r}_{i1}$.

3.11 (i) $\beta_1 < 0$ because more pollution can be expected to lower housing values; note that β_1 is the elasticity of *price* with respect to *nox*. β_2 is probably positive because *rooms* roughly measures the size of a house. (However, it does not allow us to distinguish homes where each room is large from homes where each room is small.)

(ii) If we assume that *rooms* increases with quality of the home, then log(*nox*) and *rooms* are negatively correlated when poorer neighborhoods have more pollution, something that is often true. We can use Table 3.2 to determine the direction of the bias. If $\beta_2 > 0$ and $\text{Corr}(x_1,x_2) < 0$, the simple regression estimator $\tilde{\beta}_1$ has a downward bias. But because $\beta_1 < 0$, this means that the simple regression, on average, overstates the importance of pollution. [$E(\tilde{\beta}_1)$ is more negative than β_1.]

(iii) This is what we expect from the typical sample based on our analysis in part (ii). The simple regression estimate, –1.043, is more negative (larger in magnitude) than the multiple regression estimate, –.718. As those estimates are only for one sample, we can never know which is closer to β_1. But if this is a "typical" sample, β_1 is closer to –.718.

3.12 (i) For notational simplicity, define $s_{zx} = \sum_{i=1}^{n}(z_i - \overline{z})x_i$; this is not quite the sample covariance between z and x because we do not divide by n – 1, but we are only using it to simplify notation. Then we can write $\tilde{\beta}_1$ as

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{n}(z_i - \overline{z})y_i}{s_{zx}}.$$

This is clearly a linear function of the y_i: take the weights to be $w_i = (z_i - \overline{z})/s_{zx}$. To show unbiasedness, as usual we plug $y_i = \beta_0 + \beta_1 x_i + u_i$ into this equation, and simplify:

$$\begin{aligned}\tilde{\beta}_1 &= \frac{\sum_{i=1}^{n}(z_i - \overline{z})(\beta_0 + \beta_1 x_i + u_i)}{s_{zx}} \\ &= \frac{\beta_0\sum_{i=1}^{n}(z_i - \overline{z}) + \beta_1 s_{zx} + \sum_{i=1}^{n}(z_i - \overline{z})u_i}{s_{zx}} \\ &= \beta_1 + \frac{\sum_{i=1}^{n}(z_i - \overline{z})u_i}{s_{zx}}\end{aligned}$$

where we use the fact that $\sum_{i=1}^{n}(z_i - \overline{z}) = 0$ always. Now s_{zx} is a function of the z_i and x_i and the expected value of each u_i is zero conditional on all z_i and x_i in the sample. Therefore, conditional on these values,

$$\mathrm{E}(\tilde{\beta}_1) = \beta_1 + \frac{\sum_{i=1}^{n}(z_i - \overline{z})\mathrm{E}(u_i)}{s_{zx}} = \beta_1$$

because $\mathrm{E}(u_i) = 0$ for all i.

(ii) From the fourth equation in part (i) we have (again conditional on the z_i and x_i in the sample),

$$\operatorname{Var}(\tilde{\beta}_1) = \frac{\operatorname{Var}\left[\sum_{i=1}^{n}(z_i - \overline{z})u_i\right]}{s_{zx}^2} = \frac{\sum_{i=1}^{n}(z_i - \overline{z})^2 \operatorname{Var}(u_i)}{s_{zx}^2}$$

$$= \sigma^2 \frac{\sum_{i=1}^{n}(z_i - \overline{z})^2}{s_{zx}^2}$$

because of the homoskedasticity assumption [Var(u_i) = σ^2 for all i]. Given the definition of s_{zx}, this is what we wanted to show.

(iii) We know that Var($\hat{\beta}_1$) = σ^2/[$\sum_{i=1}^{n}(x_i - \overline{x})^2$]. Now we can rearrange the inequality in the hint, drop $\overline{x}$ from the sample covariance, and cancel n^{-1} everywhere, to get [$\sum_{i=1}^{n}(z_i - \overline{z})^2$]/$s_{zx}^2$ ≥ 1/[$\sum_{i=1}^{n}(x_i - \overline{x})^2$]. When we multiply through by σ^2 we get Var($\tilde{\beta}_1$) ≥ Var($\hat{\beta}_1$), which is what we wanted to show.

SOLUTIONS TO COMPUTER EXERCISES

C3.1 (i) Probably β_2 > 0, as more income typically means better nutrition for the mother and better prenatal care.

(ii) On the one hand, an increase in income generally increases the consumption of a good, and *cigs* and *faminc* could be positively correlated. On the other, family incomes are also higher for families with more education, and more education and cigarette smoking tend to be negatively correlated. The sample correlation between *cigs* and *faminc* is about −.173, indicating a negative correlation.

(iii) The regressions without and with *faminc* are

$$\widehat{bwght} = 119.77 - .514\, cigs$$

$$n = 1{,}388, \;\; R^2 = .023$$

and

$$\widehat{bwght} = 116.97 - .463\, cigs + .093\;\, faminc$$

$$n = 1{,}388, \;\; R^2 = .030.$$

The effect of cigarette smoking is slightly smaller when *faminc* is added to the regression, but the difference is not great. This is due to the fact that *cigs* and *faminc* are not very correlated, and the coefficient on *faminc* is practically small. (The variable *faminc* is measured in thousands, so \$10,000 more in 1988 income increases predicted birth weight by only .93 ounces.)

C3.3 (i) The constant elasticity equation is

$$\widehat{\log(salary)} = 4.62 + .162\log(sales) + .107\log(mktval)$$

$$n = 177,\ \ R^2 = .299.$$

(ii) We cannot include profits in logarithmic form because profits are negative for nine of the companies in the sample. When we add it in levels form we get

$$\widehat{\log(salary)} = 4.69 + .161\log(sales) + .098\log(mktval) + .000036\,profits$$

$$n = 177, R^2 = .299.$$

The coefficient on *profits* is very small. Here, *profits* are measured in millions, so if profits increase by \$1 billion, which means $\Delta profits$ = 1,000 – a huge change – predicted salary increases by about only 3.6%. However, remember that we are holding sales and market value fixed.

Together, these variables (and we could drop *profits* without losing anything) explain almost 30% of the sample variation in log(*salary*). This is certainly not "most" of the variation.

(iii) Adding *ceoten* to the equation gives

$$\widehat{\log(salary)} = 4.56 + .162\log(sales) + .102\log(mktval) + .000029\,profits + .012ceoten$$

$$n = 177,\ \ R^2 = .318.$$

This means that one more year as *CEO* increases predicted salary by about 1.2%.

(iv) The sample correlation between log(*mktval*) and *profits* is about .78, which is fairly high. As we know, this causes no bias in the OLS estimators, although it can cause their variances to be large. Given the fairly substantial correlation between market value and firm profits, it is not too surprising that the latter adds nothing to explaining CEO salaries. Also, *profits* is a short term measure of how the firm is doing while *mktval* is based on past, current, and expected future profitability.

C3.5 The regression of *educ* on *exper* and *tenure* yields

$$educ = 13.57 - .074\ exper + .048\ tenure + \hat{r}_1.$$

$$n = 526,\quad R^2 = .101.$$

Now, when we regress log(*wage*) on $\hat{r}_1$ we obtain

$$\widehat{\log(wage)} = 1.62 + .092\ \hat{r}_1$$

$$n = 526,\quad R^2 = .207.$$

As expected, the coefficient on $\hat{r}_1$ in the second regression is identical to the coefficient on *educ* in equation (3.19). Notice that the *R*-squared from the above regression is below that in (3.19). In effect, the regression of log(*wage*) on $\hat{r}_1$ explains log(*wage*) using only the part of *educ* that is uncorrelated with *exper* and *tenure*; separate effects of *exper* and *tenure* are not included.

C3.7 (i) The results of the regression are

$$\widehat{math10} = -20.36 + 6.23\ \log(expend) - .305\ lnchprg$$

$$n = 408,\quad R^2 = .180.$$

The signs of the estimated slopes imply that more spending increases the pass rate (holding *lnchprg* fixed) and a higher poverty rate (proxied well by *lnchprg*) decreases the pass rate (holding spending fixed). These are what we expect.

(ii) As usual, the estimated intercept is the predicted value of the dependent variable when all regressors are set to zero. Setting *lnchprg* = 0 makes sense, as there are schools with low poverty rates. Setting log(*expend*) = 0 does not make sense, because it is the same as setting *expend* = 1, and spending is measured in dollars per student. Presumably this is well outside any sensible range. Not surprisingly, the prediction of a -20 pass rate is nonsensical.

(iii) The simple regression results are

$$\widehat{math10} = -69.34 + 11.16\ \log(expend)$$

$$n = 408,\quad R^2 = .030$$

and the estimated spending effect is larger than it was in part (i) – almost double.

(iv) The sample correlation between *lexpend* and *lnchprg* is about $-.19$, which means that, on average, high schools with poorer students spent less per student. This makes sense, especially in 1993 in Michigan, where school funding was essentially determined by local property tax collections.

(v) We can use equation (3.23). Because Corr(x_1,x_2) < 0, which means $\tilde{\delta}_1 < 0$, and $\hat{\beta}_2 < 0$, the simple regression estimate, $\tilde{\beta}_1$, is larger than the multiple regression estimate, $\hat{\beta}_1$. Intuitively, failing to account for the poverty rate leads to an overestimate of the effect of spending.

C3.9 (i) The estimated equation is

$$\widehat{gift} = -4.55 \ + \ 2.17 \ mailsyear + .0059 \ giftlast + \ 15.36 \ propresp$$
$$n = 4{,}268, \quad R^2 = .0834$$

The *R*-squared is now about .083, compared with about .014 for the simple regression case. Therefore, the variables *giftlast* and *propresp* help to explain significantly more variation in *gifts* in the sample (although still just over eight percent).

(ii) Holding *giftlast* and *propresp* fixed, one more mailing per year is estimated to increase *gifts* by 2.17 guilders. The simple regression estimate is 2.65, so the multiple regression estimate is somewhat smaller. Remember, the simple regression estimate holds no other factors fixed.

(iii) Because *propresp* is a proportion, it makes little sense to increase it by one. Such an increase can happen only if *propresp* goes from zero to one. Instead, consider a .10 increase in *propresp*, which means a 10 percentage point increase. Then, *gift* is estimated to be 15.36(.1) ≈ 1.54 guilders higher.

(iv) The estimated equation is

$$\widehat{gift} = -7.33 \ + \ 1.20 \ mailsyear - .261 \ giftlast + \ 16.20 \ propresp + .527 \ avggift$$
$$n = 4{,}268, \quad R^2 = .2005$$

After controlling for the average past gift level, the effect of mailings becomes even smaller: 1.20 guilders, or less than half the effect estimated by simple regression.

(v) After controlling for the average of past gifts – which we can view as measuring the “typical” generosity of the person and is positively related to the current gift level – we find that the current gift amount is negatively related to the most recent gift. A negative relationship makes some sense, as people might follow a large donation with a smaller one.

CHAPTER 4

SOLUTIONS TO PROBLEMS

4.2 (i) and (iii) generally cause the *t* statistics not to have a *t* distribution under H_0. Homoskedasticity is one of the CLM assumptions. An important omitted variable violates Assumption MLR.3. The CLM assumptions contain no mention of the sample correlations among independent variables, except to rule out the case where the correlation is one.

4.3 (i) While the standard error on *hrsemp* has not changed, the magnitude of the coefficient has increased by half. The *t* statistic on *hrsemp* has gone from about –1.47 to –2.21, so now the coefficient is statistically less than zero at the 5% level. (From Table G.2 the 5% critical value with 40 *df* is –1.684. The 1% critical value is –2.423, so the *p*-value is between .01 and .05.)

(ii) If we add and subtract $\beta_2\log(employ)$ from the right-hand-side and collect terms, we have

$$\begin{aligned}\log(scrap) &= \beta_0 + \beta_1 hrsemp + [\beta_2\log(\text{sales}) - \beta_2\log(employ)] \\ &\quad + [\beta_2\log(employ) + \beta_3\log(employ)] + u \\ &= \beta_0 + \beta_1 hrsemp + \beta_2\log(sales/employ) \\ &\quad + (\beta_2 + \beta_3)\log(employ) + u,\end{aligned}$$

where the second equality follows from the fact that log(*sales*/*employ*) = log(*sales*) – log(*employ*). Defining $\theta_3 \equiv \beta_2 + \beta_3$ gives the result.

(iii) No. We are interested in the coefficient on log(*employ*), which has a *t* statistic of .2, which is very small. Therefore, we conclude that the size of the firm, as measured by employees, does not matter, once we control for training *and* sales per employee (in a logarithmic functional form).

(iv) The null hypothesis in the model from part (ii) is H_0: $\beta_2 = -1$. The *t* statistic is [–.951 – (–1)]/.37 = (1 – .951)/.37 $\approx$ .132; this is very small, and we fail to reject whether we specify a one- or two-sided alternative.

4.4 (i) In columns (2) and (3), the coefficient on *profmarg* is actually negative, although its *t* statistic is only about –1. It appears that, once firm sales and market value have been controlled for, profit margin has no effect on CEO salary.

(ii) We use column (3), which controls for the most factors affecting salary. The *t* statistic on log(*mktval*) is about 2.05, which is just significant at the 5% level against a two-sided alternative.

(We can use the standard normal critical value, 1.96.) So log(*mktval*) is statistically significant. Because the coefficient is an elasticity, a ceteris paribus 10% increase in market value is predicted to increase *salary* by 1%. This is not a huge effect, but it is not negligible, either.

(iii) These variables are individually significant at low significance levels, with $t_{ceoten} \approx 3.11$ and $t_{comten} \approx -2.79$. Other factors fixed, another year as CEO with the company increases salary by about 1.71%. On the other hand, another year with the company, but not as CEO, lowers salary by about .92%. This second finding at first seems surprising, but could be related to the "superstar" effect: firms that hire CEOs from outside the company often go after a small pool of highly regarded candidates, and salaries of these people are bid up. More non-CEO years with a company makes it less likely the person was hired as an outside superstar.

4.7 (i) .412 ± 1.96(.094), or about .228 to .596.

(ii) No, because the value .4 is well inside the 95% CI.

(iii) Yes, because 1 is well outside the 95% CI.

4.8 (i) With df = 706 – 4 = 702, we use the standard normal critical value ($df = \infty$ in Table G.2), which is 1.96 for a two-tailed test at the 5% level. Now $t_{educ} = -11.13/5.88 \approx -1.89$, so $|t_{educ}|$ = 1.89 < 1.96, and we fail to reject H_0: $\beta_{educ} = 0$ at the 5% level. Also, $t_{age} \approx 1.52$, so *age* is also statistically insignificant at the 5% level.

(ii) We need to compute the R-squared form of the F statistic for joint significance. But F = $[(.113 - .103)/(1 - .113)](702/2) \approx 3.96$. The 5% critical value in the $F_{2,702}$ distribution can be obtained from Table G.3b with denominator $df = \infty$: cv = 3.00. Therefore, *educ* and *age* are jointly significant at the 5% level (3.96 > 3.00). In fact, the p-value is about .019, and so *educ* and *age* are jointly significant at the 2% level.

(iii) Not really. These variables are jointly significant, but including them only changes the coefficient on *totwrk* from –.151 to –.148.

(iv) The standard t and F statistics that we used assume homoskedasticity, in addition to the other CLM assumptions. If there is heteroskedasticity in the equation, the tests are no longer valid.

4.11 (i) Holding *profmarg* fixed, $\Delta\widehat{rdintens} = .321\,\Delta\log(sales) =$ $(.321/100)[100\cdot\Delta\log(sales)] \approx .00321(\%\Delta sales)$. Therefore, if $\%\Delta sales = 10$, $\Delta\widehat{rdintens} \approx$ .032, or only about 3/100 of a percentage point. For such a large percentage increase in sales, this seems like a practically small effect.

(ii) H_0: $\beta_1 = 0$ versus H_1: $\beta_1 > 0$, where β_1 is the population slope on log(*sales*). The t statistic is $.321/.216 \approx 1.486$. The 5% critical value for a one-tailed test, with df = 32 – 3 = 29, is obtained from Table G.2 as 1.699; so we cannot reject H_0 at the 5% level. But the 10% critical value is 1.311; since the *t* statistic is above this value, we reject H_0 in favor of H_1 at the 10% level.

(iii) Not really. Its *t* statistic is only 1.087, which is well below even the 10% critical value for a one-tailed test.

SOLUTIONS TO COMPUTER EXERCISES

C4.1 (i) Holding other factors fixed,

$$\begin{aligned}\Delta voteA &= \beta_1 \Delta \log(expendA) = (\beta_1/100)[100 \cdot \Delta \log(expendA)] \\ &\approx (\beta_1/100)(\%\Delta expendA),\end{aligned}$$

where we use the fact that $100 \cdot \Delta \log(expendA) \approx \%\Delta expendA$. So $\beta_1/100$ is the (ceteris paribus) percentage point change in *voteA* when *expendA* increases by one percent.

(ii) The null hypothesis is H_0: $\beta_2 = -\beta_1$, which means a z% increase in expenditure by A and a z% increase in expenditure by B leaves *voteA* unchanged. We can equivalently write H_0: $\beta_1 + \beta_2 = 0$.

(iii) The estimated equation (with standard errors in parentheses below estimates) is

$$\begin{aligned}\widehat{voteA} = \underset{(3.93)}{45.08} + \underset{(0.382)}{6.083} \log(expendA) - \underset{(0.379)}{6.615} \log(expendB) + \underset{(.062)}{.152}\, prtystrA\end{aligned}$$

$$n = 173, \quad R^2 = .793.$$

The coefficient on log(*expendA*) is very significant (t statistic $\approx$ 15.92), as is the coefficient on log(*expendB*) (t statistic $\approx$ –17.45). The estimates imply that a 10% ceteris paribus increase in spending by candidate A increases the predicted share of the vote going to A by about .61 percentage points. [Recall that, holding other factors fixed, $\Delta\widehat{voteA} \approx (6.083/100)\%\Delta expendA$).] Similarly, a 10% ceteris paribus increase in spending by B reduces $\widehat{voteA}$ by about .66 percentage points. These effects certainly cannot be ignored.

While the coefficients on log(*expendA*) and log(*expendB*) are of similar magnitudes (and opposite in sign, as we expect), we do not have the standard error of $\hat{\beta}_1 + \hat{\beta}_2$, which is what we would need to test the hypothesis from part (ii).

(iv) Write $\theta_1 = \beta_1 + \beta_2$, or $\beta_1 = \theta_1 - \beta_2$. Plugging this into the original equation, and rearranging, gives

$$\widehat{voteA} = \beta_0 + \theta_1 \log(expendA) + \beta_2[\log(expendB) - \log(expendA)] + \beta_3\, prtystrA + u,$$

When we estimate this equation we obtain $\hat{\theta}_1 \approx -.532$ and se($\hat{\theta}_1$) $\approx$ .533. The t statistic for the hypothesis in part (ii) is –.532/.533 $\approx$ –1. Therefore, we fail to reject H_0: $\beta_2 = -\beta_1$.

C4.3 (i) The estimated model is

$$\widehat{\log(price)} = \underset{(0.10)}{11.67} + \underset{(.000043)}{.000379}\, sqrft + \underset{(.0296)}{.0289}\, bdrms$$

$n = 88,\ R^2 = .588.$

Therefore, $\hat{\theta}_1 = 150(.000379) + .0289 = .0858$, which means that an additional 150 square foot bedroom increases the predicted price by about 8.6%.

(ii) $\beta_2 = \theta_1 - 150\beta_1$, and so

$$\begin{aligned} \log(price) &= \beta_0 + \beta_1 sqrft + (\theta_1 - 150\beta_1) bdrms + u \\ &= \beta_0 + \beta_1 (sqrft - 150\ bdrms) + \theta_1 bdrms + u. \end{aligned}$$

(iii) From part (ii), we run the regression

log(*price*) on (*sqrft* – 150 *bdrms*), *bdrms*,

and obtain the standard error on *bdrms*. We already know that $\hat{\theta}_1 = .0858$; now we also get se($\hat{\theta}_1$) = .0268. The 95% confidence interval reported by my software package is .0326 to .1390 (or about 3.3% to 13.9%).

C4.5 (i) If we drop *rbisyr* the estimated equation becomes

$$\begin{aligned} \widehat{\log(salary)} = {} & \underset{(0.27)}{11.02} + \underset{(.0121)}{.0677}\, years + \underset{(.0016)}{.0158}\, gamesyr \\ & + \underset{(.0011)}{.0014}\, bavg + \underset{(.0072)}{.0359}\, hrunsyr \end{aligned}$$

$n = 353,\ R^2 = .625.$

Now *hrunsyr* is very statistically significant (t statistic $\approx$ 4.99), and its coefficient has increased by about two and one-half times.

(ii) The equation with *runsyr*, *fldperc*, and *sbasesyr* added is

$$\widehat{\log(salary)} = \underset{(2.00)}{10.41} + \underset{(.0120)}{.0700}\, years + \underset{(.0027)}{.0079}\, gamesyr$$
$$+ \underset{(.00110)}{.00053}\, bavg + \underset{(.0086)}{.0232}\, hrunsyr$$
$$+ \underset{(.0051)}{.0174}\, runsyr + \underset{(.0020)}{.0010}\, fldperc - \underset{(.0052)}{.0064}\, sbasesyr$$

$n = 353, \quad R^2 = .639.$

Of the three additional independent variables, only *runsyr* is statistically significant (t statistic = $.0174/.0051 \approx 3.41$). The estimate implies that one more run per year, other factors fixed, increases predicted salary by about 1.74%, a substantial increase. The stolen bases variable even has the "wrong" sign with a t statistic of about –1.23, while *fldperc* has a t statistic of only .5. Most major league baseball players are pretty good fielders; in fact, the smallest *fldperc* is 800 (which means .800). With relatively little variation in *fldperc*, it is perhaps not surprising that its effect is hard to estimate.

(iii) From their t statistics, *bavg*, *fldperc*, and *sbasesyr* are individually insignificant. The F statistic for their joint significance (with 3 and 345 *df*) is about .69 with p-value $\approx$ .56. Therefore, these variables are jointly very insignificant.

C4.7 (i) The minimum value is 0, the maximum is 99, and the average is about 56.16.

(ii) When *phsrank* is added to (4.26), we get the following:

$$\widehat{\log(wage)} = \underset{(0.024)}{1.459} - \underset{(.0070)}{.0093}\, jc + \underset{(.0026)}{.0755}\, totcoll + \underset{(.0002)}{.0049}\, exper + \underset{(.00024)}{.00030}\, phsrank$$

$n = 6{,}763, \quad R^2 = .223$

So *phsrank* has a t statistic equal to only 1.25; it is not statistically significant. If we increase *phsrank* by 10, log(*wage*) is predicted to increase by (.0003)10 = .003. This implies a .3% increase in *wage*, which seems a modest increase given a 10 percentage point increase in *phsrank*. (However, the sample standard deviation of *phsrank* is about 24.)

(iii) Adding *phsrank* makes the t statistic on *jc* even smaller in absolute value, about 1.33, but the coefficient magnitude is similar to (4.26). Therefore, the base point remains unchanged: the return to a junior college is estimated to be somewhat smaller, but the difference is not significant and standard significant levels.

(iv) The variable *id* is just a worker identification number, which should be randomly assigned (at least roughly). Therefore, *id* should not be correlated with any variable in the regression equation. It should be insignificant when added to (4.17) or (4.26). In fact, its t statistic is about .54.

C4.9 (i) The results from the OLS regression, with standard errors in parentheses, are

$$\widehat{\log(psoda)} = \underset{(0.29)}{-1.46} + \underset{(.031)}{.073}\, prpblck + \underset{(.027)}{.137}\log(income) + \underset{(.133)}{.380}\, prppov$$

$n = 401,\ R^2 = .087$

The p-value for testing H$_0$: $\beta_1 = 0$ against the two-sided alternative is about .018, so that we reject H$_0$ at the 5% level but not at the 1% level.

(ii) The correlation is about −.84, indicating a strong degree of multicollinearity. Yet each coefficient is very statistically significant: the t statistic for $\hat{\beta}_{\log(income)}$ is about 5.1 and that for $\hat{\beta}_{prppov}$ is about 2.86 (two-sided p-value = .004).

(iii) The OLS regression results when log(*hseval*) is added are

$$\widehat{\log(psoda)} = \underset{(.29)}{-.84} + \underset{(.029)}{.098}\, prpblck - \underset{(.038)}{.053}\log(income)$$
$$+ \underset{(.134)}{.052}\, prppov + \underset{(.018)}{.121}\log(hseval)$$

$n = 401,\ R^2 = .184$

The coefficient on log(*hseval*) is an elasticity: a one percent increase in housing value, holding the other variables fixed, increases the predicted price by about .12 percent. The two-sided *p*-value is zero to three decimal places.

(iv) Adding log(*hseval*) makes log(*income*) and *prppov* individually insignificant (at even the 15% significance level against a two-sided alternative for log(*income*), and *prppov* is does not have a *t* statistic even close to one in absolute value). Nevertheless, they are jointly significant at the 5% level because the outcome of the $F_{2,396}$ statistic is about 3.52 with *p*-value = .030. All of the control variables – log(*income*), *prppov*, and log(*hseval*) – are highly correlated, so it is not surprising that some are individually insignificant.

(v) Because the regression in (iii) contains the most controls, log(*hseval*) is individually significant, and log(*income*) and *prppov* are jointly significant, (iii) seems the most reliable. It holds fixed three measure of income and affluence. Therefore, a reasonable estimate is that if the proportion of blacks increases by .10, *psoda* is estimated to increase by 1%, other factors held fixed.

CHAPTER 5

SOLUTIONS TO PROBLEMS

5.2 The variable *cigs* has nothing close to a normal distribution in the population. Most people do not smoke, so *cigs* = 0 for over half of the population. A normally distributed random variable takes on no particular value with positive probability. Further, the distribution of *cigs* is skewed, whereas a normal random variable must be symmetric about its mean.

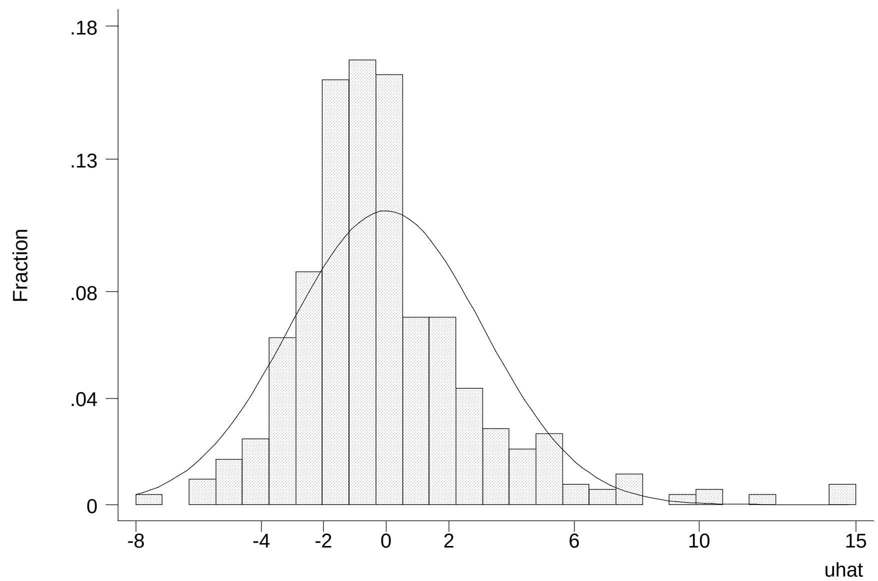
5.4 Write $y = \beta_0 + \beta_1 x_1 + u$, and take the expected value: $E(y) = \beta_0 + \beta_1 E(x_1) + E(u)$, or $\mu_y = \beta_0 + \beta_1 \mu_x$ since $E(u) = 0$, where $\mu_y = E(y)$ and $\mu_x = E(x_1)$. We can rewrite this as $\beta_0 = \mu_y - \beta_1 \mu_x$. Now, $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1$. Taking the plim of this we have $\text{plim}(\hat{\beta}_0) = \text{plim}(\bar{y} - \hat{\beta}_1 \bar{x}_1) = \text{plim}(\bar{y}) - \text{plim}(\hat{\beta}_1) \cdot \text{plim}(\bar{x}_1) = \mu_y - \beta_1 \mu_x$, where we use the fact that $\text{plim}(\bar{y}) = \mu_y$ and $\text{plim}(\bar{x}_1) = \mu_x$ by the law of large numbers, and $\text{plim}(\hat{\beta}_1) = \beta_1$. We have also used the parts of Property PLIM.2 from Appendix C.

SOLUTIONS TO COMPUTER EXERCISES

C5.1 (i) The estimated equation is

$$\widehat{wage} = -2.87 + .599\, educ + .022\, exper + .169\, tenure$$
$$(0.73)\quad (.051)\qquad\quad (.012)\qquad\quad (.022)$$
$$n = 526,\quad R^2 = .306,\quad \hat{\sigma} = 3.085.$$

Below is a histogram of the 526 residual, $\hat{u}_i$, $i = 1, 2, ..., 526$. The histogram uses 27 bins, which is suggested by the formula in the Stata manual for 526 observations. For comparison, the normal distribution that provides the best fit to the histogram is also plotted.

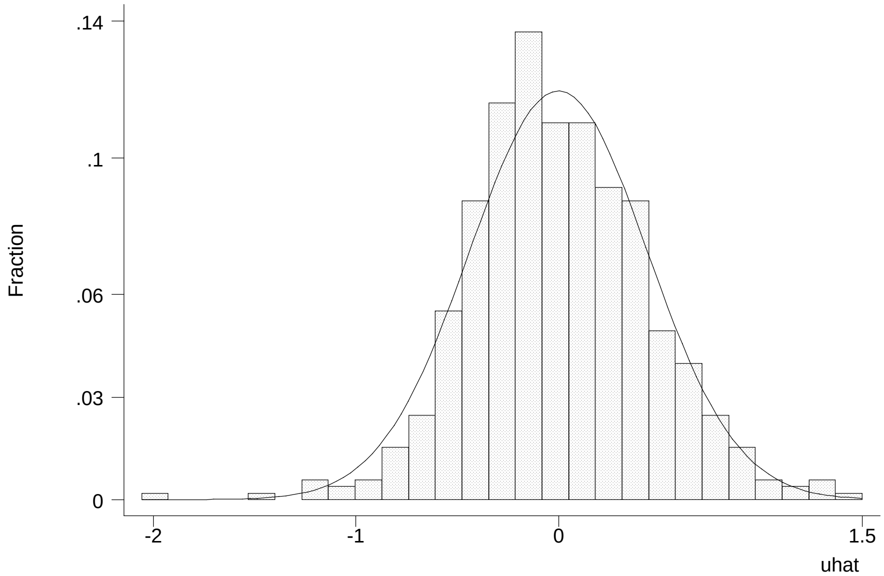


(ii) With log(*wage*) as the dependent variable the estimated equation is

$$\begin{array}{lcccccccc}
\widehat{\log(wage)} & = & .284 & + & .092\, educ & + & .0041\, exper & + & .022\, tenure \\
 & & (.104) & & (.007) & & (.0017) & & (.003)
\end{array}$$

$$n = 526, \quad R^2 = .316, \quad \hat{\sigma} = .441.$$

The histogram for the residuals from this equation, with the best-fitting normal distribution overlaid, is given below:



(iii) The residuals from the log(*wage*) regression appear to be more normally distributed. Certainly the histogram in part (ii) fits under its comparable normal density better than in part (i), and the histogram for the *wage* residuals is notably skewed to the left. In the *wage* regression there are some very large residuals (roughly equal to 15) that lie almost five estimated standard deviations ($\hat{\sigma}$ = 3.085) from the mean of the residuals, which is identically zero, of course. Residuals far from zero does not appear to be nearly as much of a problem in the log(*wage*) regression.

C5.3 We first run the regression *colgpa* on *cigs*, *parity*, and *faminc* using only the 1,191 observations with nonmissing observations on *motheduc* and *fatheduc*. After obtaining these residuals, $\tilde{u}_i$, these are regressed on $cigs_i$, $parity_i$, $faminc_i$, $motheduc_i$, and $fatheduc_i$, where, of course, we can only use the 1,197 observations with nonmissing values for both *motheduc* and *fatheduc*. The *R*-squared from this regression, R_u^2, is about .0024. With 1,191 observations, the chi-square statistic is (1,191)(.0024) $\approx$ 2.86. The *p*-value from the χ_2^2 distribution is about .239, which is very close to .242, the *p*-value for the comparable *F* test.

CHAPTER 6

SOLUTIONS TO PROBLEMS

6.1 This would make little sense. Performances on math and science exams are measures of outputs of the educational process, and we would like to know how various educational inputs and school characteristics affect math and science scores. For example, if the staff-to-pupil ratio has an effect on both exam scores, why would we want to hold performance on the science test fixed while studying the effects of *staff* on the math pass rate? This would be an example of controlling for too many factors in a regression equation. The variable *scill* could be a dependent variable in an identical regression equation.

6.2 (i) Because $\exp(-1.96\hat{\sigma}) < 1$ and $\exp(\hat{\sigma}^2/2) > 1$, the point prediction is always above the lower bound. The only issue is whether the point prediction is below the upper bound. This is the case when $\exp(\hat{\sigma}^2/2) \leq \exp(1.96\hat{\sigma})$ or, taking logs, $\hat{\sigma}^2/2 \leq 1.96\hat{\sigma}$, or $\hat{\sigma} \leq 2(1.96) = 3.92$. Therefore, the point prediction is in the approximate 95% prediction interval for $\hat{\sigma} \leq 3.92$. Because $\hat{\sigma}$ is the estimated standard deviation in the regression with log(*y*) as the dependent variable, 3.92 is a very large value for the estimated standard deviation of the error, which is on the order of 400 percent. Most of the time, the estimated SER is well below that.

(ii) In the CEO salary regression, $\hat{\sigma} = .505$, which is well below 3.92.

6.5 (i) The turnaround point is given by $\hat{\beta}_1/(2|\hat{\beta}_2|)$, or .0003/(.000000014) $\approx$ 21,428.57; remember, this is sales in millions of dollars.

(ii) Probably. Its *t* statistic is about –1.89, which is significant against the one-sided alternative H_0: $\beta_1 < 0$ at the 5% level (*cv* $\approx$ –1.70 with *df* = 29). In fact, the *p*-value is about .036.

(iii) Because *sales* gets divided by 1,000 to obtain *salesbil*, the corresponding coefficient gets multiplied by 1,000: (1,000)(.00030) = .30. The standard error gets multiplied by the same factor. As stated in the hint, $salesbil^2 = sales/1{,}000{,}000$, and so the coefficient on the quadratic gets multiplied by one million: (1,000,000)(.0000000070) = .0070; its standard error also gets multiplied by one million. Nothing happens to the intercept (because *rdintens* has not been rescaled) or to the R^2:

$$\widehat{rdintens} = \underset{(0.429)}{2.613} + \underset{(.14)}{.30}\, salesbil - \underset{(.0037)}{.0070}\, salesbil^2$$

$$n = 32, \quad R^2 = .1484.$$

(iv) The equation in part (iii) is easier to read because it contains fewer zeros to the right of the decimal. Of course the interpretation of the two equations is identical once the different scales are accounted for.

6.6 The second equation is clearly preferred, as its adjusted *R*-squared is notably larger than that in the other two equations. The second equation contains the same number of estimated parameters as the first, and the one fewer than the third. The second equation is also easier to interpret than the third.

6.9 The generality is not necessary. The *t* statistic on roe^2 is only about −.30, which shows that roe^2 is very statistically insignificant. Plus, having the squared term has only a minor effect on the slope even for large values of *roe*. (The approximate slope is .0215 − .00016 *roe*, and even when *roe* = 25 – about one standard deviation above the average *roe* in the sample – the slope is .211, as compared with .215 at *roe* = 0.)

SOLUTIONS TO COMPUTER EXERCISES

C6.1 (i) The causal (or ceteris paribus) effect of *dist* on *price* means that $\beta_1 \geq 0$: all other relevant factors equal, it is better to have a home farther away from the incinerator. The estimated equation is

$$\widehat{\log(price)} = \underset{(0.65)}{8.05} + \underset{(.066)}{.365}\log(dist)$$

$$n = 142,\ R^2 = .180,\ \bar{R}^2 = .174,$$

which means a 1% increase in distance from the incinerator is associated with a predicted price that is about .37% higher.

(ii) When the variables log(*inst*), log(*area*), log(*land*), *rooms*, *baths*, and *age* are added to the regression, the coefficient on log(*dist*) becomes about .055 (*se* ≈ .058). The effect is much smaller now, and statistically insignificant. This is because we have explicitly controlled for several other factors that determine the quality of a home (such as its size and number of baths) and its location (distance to the interstate). This is consistent with the hypothesis that the incinerator was located near less desirable homes to begin with.

(iii) When $[\log(inst)]^2$ is added to the regression in part (ii), we obtain (with the results only partially reported)

$$\widehat{\log(price)} = \underset{(2.65)}{-3.32} + \underset{(.062)}{.185}\log(dist) + \underset{(0.501)}{2.073}\log(inst) - \underset{(.0282)}{.1193}[\log(inst)]^2 + \ldots$$

$$n = 142,\ R^2 = .778,\ \bar{R}^2 = .764.$$

The coefficient on log(*dist*) is now very statistically significant, with a *t* statistic of about three. The coefficients on log(*inst*) and $[\log(inst)]^2$ are both very statistically significant, each with *t* statistics above four in absolute value. Just adding $[\log(inst)]^2$ has had a very big effect on the coefficient important for policy purposes. This means that distance from the incinerator and distance from the interstate are correlated in some nonlinear way that also affects housing price.

We can find the value of log(*inst*) where the effect on log(*price*) actually becomes negative: $2.073/[2(.1193)] \approx 8.69$. When we exponentiate this we obtain about 5,943 feet from the interstate. Therefore, it is best to have your home away from the interstate for distances less than just over a mile. After that, moving farther away from the interstate lowers predicted house price.

(iv) The coefficient on $[\log(dist)]^2$, when it is added to the model estimated in part (iii), is about -.0365, but its *t* statistic is only about -.33. Therefore, it is not necessary to add this complication.

C6.3 (i) Holding *exper* (and the elements in *u*) fixed, we have

$$\Delta \log(wage) = \beta_1 \Delta educ + \beta_3 (\Delta educ) exper = (\beta_1 + \beta_3 exper)\Delta educ,$$

or

$$\frac{\Delta \log(wage)}{\Delta educ} = (\beta_1 + \beta_3 exper).$$

This is the approximate proportionate change in *wage* given one more year of education.

(ii) H_0: $\beta_3 = 0$. If we think that education and experience interact positively – so that people with more experience are more productive when given another year of education – then $\beta_3 > 0$ is the appropriate alternative.

(iii) The estimated equation is

$$\begin{aligned} \widehat{\log(wage)} &= \underset{(0.24)}{5.95} + \underset{(.0174)}{.0440}\, educ - \underset{(.0200)}{.0215}\, exper + \underset{(.00153)}{.00320}\, educ \cdot exper \\ n &= 935, \quad R^2 = .135, \quad \bar{R}^2 = .132. \end{aligned}$$

The *t* statistic on the interaction term is about 2.13,which gives a *p*-value below .02 against H_1: $\beta_3 > 0$. Therefore, we reject H_0: $\beta_3 = 0$ against H_1: $\beta_3 > 0$ at the 2% level.

(iv) We rewrite the equation as

$$\log(wage) = \beta_0 + \theta_1 educ + \beta_2 exper + \beta_3 educ(exper - 10) + u,$$

and run the regression log(*wage*) on *educ*, *exper*, and *educ*(*exper* – 10). We want the coefficient on *educ*. We obtain $\hat{\theta}_1 \approx$ .0761 and $se(\hat{\theta}_1) \approx$ .0066. The 95% CI for θ_1 is about .063 to .089.

C6.5 (i) The results of estimating the log-log model (but with *bdrms* in levels) are

$$\widehat{\log(price)} = \underset{(0.65)}{5.61} + \underset{(.038)}{.168}\log(lotsize) + \underset{(.093)}{.700}\log(sqrft) + \underset{(.028)}{.037}\,bdrms$$

$$n = 88, \quad R^2 = .634, \quad \bar{R}^2 = .630.$$

(ii) With *lotsize* = 20,000, *sqrft* = 2,500, and *bdrms* = 4, we have

$$\widehat{lprice} = 5.61 + .168\cdot\log(20{,}000) + .700\cdot\log(2{,}500) + .037(4) \approx 12.90$$

where we use *lprice* to denote log(*price*). To predict *price*, we use the equation $\hat{price}$ = $\hat{\alpha}_0\exp(\widehat{lprice})$, where $\hat{\alpha}_0$ is the slope on $\hat{m}_i \equiv \exp(\widehat{lprice})$ from the regression $price_i$ on $\hat{m}_i$, i = 1,2, …, 88 (without an intercept). When we do this regression we get $\hat{\alpha}_0 \approx$ 1.023. Therefore, for the values of the independent variables given above, $\widehat{price} \approx (1.023)\exp(12.90) \approx$ \$409,519 (rounded to the nearest dollar). If we forget to multiply by $\hat{\alpha}_0$ the predicted price would be about \$400,312.

(iii) When we run the regression with all variables in levels, the *R*-squared is about .672. When we compute the correlation between $price_i$ and the $\hat{m}_i$ from part (ii), we obtain about .859. The square of this, or roughly .738, is the comparable goodness-of-fit measure for the model with log(*price*) as the dependent variable. Therefore, for predicting *price*, the log model is notably better.

C6.7 (i) If we hold all variables except *priGPA* fixed and use the usual approximation $\Delta(priGPA^2) \approx 2(priGPA)\cdot\Delta priGPA$, then we have

$$\begin{aligned}\Delta stndfnl &= \beta_2\Delta priGPA + \beta_4\Delta(priGPA^2) + \beta_6(\Delta priGPA)atndrte \\ &\approx (\beta_2 + 2\beta_4 priGPA + \beta_6 atndrte)\Delta priGPA;\end{aligned}$$

dividing by $\Delta priGPA$ gives the result. In equation (6.19) we have $\hat{\beta}_2 = -1.63$, $\hat{\beta}_4$ = .296, and $\hat{\beta}_6$ = .0056. When *priGPA* = 2.59 and *atndrte* = .82 we have

$$\frac{\widehat{\Delta stndfnl}}{\Delta priGPA} = -1.63 + 2(.296)(2.59) + .0056(.82) \approx -.092.$$

(ii) First, note that $(priGPA - 2.59)^2 = priGPA^2 - 2(2.59)priGPA + (2.59)^2$ and $priGPA(atndrte - .82) = priGPA \cdot atndrte - (.82)priGPA$. So we can write equation 6.18) as

$$\begin{aligned} stndfnl &= \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 (priGPA - 2.59)^2 \\ &\quad + \beta_4 [2(2.59) priGPA] - \beta_4 (2.59)^2 + \beta_5 ACT^2 \\ &\quad + \beta_6 priGPA(atndrte - .82) + \beta_6 (.82) priGPA + u \\ &= [\beta_0 - \beta_4 (2.59)^2] + \beta_1 atndrte \\ &\quad + [\beta_2 + 2\beta_4 (2.59) + \beta_6 (.82)] priGPA + \beta_3 ACT \\ &\quad + \beta_4 (priGPA - 2.59)^2 + \beta_5 ACT^2 + \beta_6 priGPA(atndrte - .82) + u \\ &\equiv \theta_0 + \beta_1 atndrte + \theta_2 priGPA + \beta_3 ACT + \beta_4 (priGPA - 2.59)^2 \\ &\quad + \beta_5 ACT^2 + \beta_6 priGPA(atndrte - .82) + u. \end{aligned}$$

When we run the regression associated with this last model, we obtain $\hat{\theta}_2 \approx$ -.091 [which differs from part (i) by rounding error] and $se(\hat{\theta}_2) \approx .363$. This implies a very small t statistic for $\hat{\theta}_2$.

C6.9 (i) The estimated equation is

$$\begin{aligned} \widehat{points} = \; & \underset{(6.99)}{35.22} + \underset{(.405)}{2.364}\, exper - \underset{(.0235)}{.0770}\, exper^2 - \underset{(.295)}{1.074}\, age - \underset{(.451)}{1.286}\, coll \\ & n = 269, \quad R^2 = .141, \quad \bar{R}^2 = .128. \end{aligned}$$

(ii) The turnaround point is $2.364/[2(.0770)] \approx 15.35$. So, the increase from 15 to 16 years of experience would actually reduce salary. This is a very high level of experience, and we can essentially ignore this prediction: only two players in the sample of 269 have more than 15 years of experience.

(iii) Many of the most promising players leave college early, or, in some cases, forego college altogether, to play in the NBA. These top players command the highest salaries. It is not more college that hurts salary, but less college is indicative of super-star potential.

(iv) When age^2 is added to the regression from part (i), its coefficient is .0536 (se = .0492). Its t statistic is barely above one, so we are justified in dropping it. The coefficient on *age* in the same regression is –3.984 (se = 2.689). Together, these estimates imply a negative, increasing, return to *age*. The turning point is roughly at 74 years old. In any case, the linear function of *age* seems sufficient.

(v) The OLS results are

$$\widehat{\log(wage)} = 6.78 + .078\, points + .218\, exper - .0071\, exper^2 - .048\, age - .040\, coll$$

$$(.85) \quad (.007) \qquad (.050) \qquad (.0028) \qquad (.035) \qquad (.053)$$

$n = 269$, $R^2 = .488$, $\bar{R}^2 = .478$

(vi) The joint F statistic produced by Stata is about 1.19. With 2 and 263 *df*, this gives a *p*-value of roughly .31. Therefore, once scoring and years played are controlled for, there is no evidence for wage differentials depending on age or years played in college.

C6.11 (i) The results of the OLS regression are

$$\widehat{ecolbs} = \underset{(0.38)}{1.97} - \underset{(0.59)}{2.93}\, ecoprc + \underset{(0.71)}{3.03}\, regprc$$

$n = 660$, $R^2 = .036$, $\bar{R}^2 = .034$

As predicted by economic theory, the own price effect is negative and the cross price effect is positive. In particular, an increase in *ecoprc* of .10, or 10 cents per pound, reduces the estimated demand for eco-labeled apples by about .29 lbs. A ceteris paribus increase of 10 cents per lb. for regular applies increases the estimated demand for eco-labeled apples by about .30 lbs. These effects, which are essentially the same magnitude but of opposite sign, are fairly large.

(ii) Each price variable is individually statistically significant with t statistics greater than four (in absolute value) in both cases. The *p*-values are zero to at least three decimal places.

(iii) The fitted values range from a low of about .86 to a high of about 2.09. This is much less variation than *ecoblbs* itself, which ranges from 0 to 42 (although 42 is a bit of an outlier). There are 248 out of 660 observations with $ecolbs = 0$ and these observations are clearly not explained well by the model.

(iv) The R-squared is only about 3.6% (and it does not really matter whether we use the usual or adjusted R-squared). This is a very small explained variation in *ecolbs*. So the two price variables do not do a good job of explaining why $ecolbs_i$ varies across families.

(v) When *faminc*, *hhsize*, *educ*, and *age* are added to the regression, the R-squared only increases to about .040 (and the adjusted R-squared falls from .034 to .031). The *p*-value for the joint F test (with 4 and 653 *df*) is about .63, which provides no evidence that these additional variables belong in the regression. Evidently, in addition to the two price variables, the factors that explain variation in *ecolbs* (which is, remember, a counterfactual quantity), are not captured by the demographic and economic variables collected in the survey. Almost 97 percent of the variation is due to unobserved "taste" factors.

C6.13 (i) The estimated equation is

$$\widehat{math4} = \underset{(19.96)}{91.93} + \underset{(2.10)}{3.52}\, lexppp - \underset{(0.94)}{5.40}\, lenroll - \underset{(.015)}{.449}\, lunch$$

$$n = 1{,}692,\ R^2 = .3729,\ \bar{R}^2 = .3718$$

The *lenroll* and *lunch* variables are individually significant at the 5% level, regardless of whether we use a one-sided or two-sided test; in fact, their *p*-values are very small. But *lexppp*, with t = 1.68, is not significant against a two-sided alternative. Its one-sided *p*-value is about .047, so it is statistically significant at the 5% level against the positive one-sided alternative.

(ii) The range of fitted values is from about 42.41 to 92.67, which is much narrower than the rage of actual math pass rates in the sample, which is from zero to 100.

(iii) The largest residual is about 51.42, and it belongs to building code 1141. This residual is the difference between the actual pass rate and our best prediction of the pass rate, given the values of spending, enrollment, and the free lunch variable. If we think that per pupil spending, enrollment, and the poverty rate are sufficient controls, the residual can be interpreted as a "value added" for the school. That is, for school 1141, its pass rate is over 51 points higher than we would expect, based on its spending, size, and student poverty.

(iv) The joint F statistic, with 3 and 1,685 *df*, is about .52, which gives *p*-value ≈ .67. Therefore, the quadratics are jointly very insignificant, and we would drop them from the model.

(v) The beta coefficients for *lexppp*, *lenroll*, and *lunch* are roughly .035, −.115, and −.613, respectively. Therefore, in standard deviation units, *lunch* has by far the largest effect. The spending variable has the smallest effect.

CHAPTER 7

SOLUTIONS TO PROBLEMS

7.1 (i) The coefficient on *male* is 87.75, so a man is estimated to sleep almost one and one-half hours more per week than a comparable woman. Further, $t_{male} = 87.75/34.33 \approx 2.56$, which is close to the 1% critical value against a two-sided alternative (about 2.58). Thus, the evidence for a gender differential is fairly strong.

(ii) The t statistic on *totwrk* is $-.163/.018 \approx -9.06$, which is very statistically significant. The coefficient implies that one more hour of work (60 minutes) is associated with $.163(60) \approx 9.8$ minutes less sleep.

(iii) To obtain R_r^2, the R-squared from the restricted regression, we need to estimate the model without *age* and age^2. When *age* and age^2 are both in the model, *age* has no effect only if the parameters on both terms are zero.

7.3 (i) The t statistic on $hsize^2$ is over four in absolute value, so there is very strong evidence that it belongs in the equation. We obtain this by finding the turnaround point; this is the value of *hsize* that maximizes $\hat{sat}$ (other things fixed): $19.3/(2 \cdot 2.19) \approx 4.41$. Because *hsize* is measured in hundreds, the optimal size of graduating class is about 441.

(ii) This is given by the coefficient on *female* (since *black* = 0): nonblack females have SAT scores about 45 points lower than nonblack males. The t statistic is about –10.51, so the difference is very statistically significant. (The very large sample size certainly contributes to the statistical significance.)

(iii) Because *female* = 0, the coefficient on *black* implies that a black male has an estimated SAT score almost 170 points less than a comparable nonblack male. The t statistic is over 13 in absolute value, so we easily reject the hypothesis that there is no ceteris paribus difference.

(iv) We plug in *black* = 1, *female* = 1 for black females and *black* = 0 and *female* = 1 for nonblack females. The difference is therefore –169.81 + 62.31 = –107.50. Because the estimate depends on two coefficients, we cannot construct a t statistic from the information given. The easiest approach is to define dummy variables for three of the four race/gender categories and choose nonblack females as the base group. We can then obtain the t statistic we want as the coefficient on the black female dummy variable.

7.5 (i) Following the hint, $\widehat{colGPA} = \hat{\beta}_0 + \hat{\delta}_0(1 - noPC) + \hat{\beta}_1 hsGPA + \hat{\beta}_2 ACT = (\hat{\beta}_0 + \hat{\delta}_0) - \hat{\delta}_0 noPC + \hat{\beta}_1 hsGPA + \hat{\beta}_2 ACT$. For the specific estimates in equation (7.6), $\hat{\beta}_0 = 1.26$ and $\hat{\delta}_0 = .157$, so the new intercept is 1.26 + .157 = 1.417. The coefficient on *noPC* is –.157.

(ii) Nothing happens to the *R*-squared. Using *noPC* in place of *PC* is simply a different way of including the same information on *PC* ownership.

(iii) It makes no sense to include both dummy variables in the regression: we cannot hold *noPC* fixed while changing *PC*. We have only two groups based on *PC* ownership so, in addition to the overall intercept, we need only to include one dummy variable. If we try to include both along with an intercept we have perfect multicollinearity (the dummy variable trap).

7.7 (i) Write the population model underlying (7.29) as

$$\begin{aligned} inlf = {} & \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age \\ & + \beta_6 kidslt6 + \beta_7 kidsage6 + u, \end{aligned}$$

plug in $inlf = 1 - outlf$, and rearrange:

$$\begin{aligned} 1 - outlf = {} & \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age \\ & + \beta_6 kidslt6 + \beta_7 kidsage6 + u, \end{aligned}$$

or

$$\begin{aligned} outlf = {} & (1 - \beta_0) - \beta_1 nwifeinc - \beta_2 educ - \beta_3 exper - \beta_4 exper^2 - \beta_5 age \\ & - \beta_6 kidslt6 - \beta_7 kidsage6 - u, \end{aligned}$$

The new error term, $-u$, has the same properties as u. From this we see that if we regress *outlf* on all of the independent variables in (7.29), the new intercept is 1 − .586 = .414 and each slope coefficient takes on the opposite sign from when *inlf* is the dependent variable. For example, the new coefficient on *educ* is −.038 while the new coefficient on *kidslt6* is .262.

(ii) The standard errors will not change. In the case of the slopes, changing the signs of the estimators does not change their variances, and therefore the standard errors are unchanged (but the *t* statistics change sign). Also, $\text{Var}(1 - \hat{\beta}_0) = \text{Var}(\hat{\beta}_0)$, so the standard error of the intercept is the same as before.

(iii) We know that changing the units of measurement of independent variables, or entering qualitative information using different sets of dummy variables, does not change the *R*-squared. But here we are changing the *dependent* variable. Nevertheless, the *R*-squareds from the regressions are still the same. To see this, part (i) suggests that the squared residuals will be identical in the two regressions. For each i the error in the equation for $outlf_i$ is just the negative of the error in the other equation for $inlf_i$, and the same is true of the residuals. Therefore, the SSRs are the same. Further, in this case, the total sum of squares are the same. For *outlf* we have

$$\text{SST} = \sum_{i=1}^{n}(outlf_i - \overline{outlf})^2 = \sum_{i=1}^{n}[(1 - inlf_i) - (1 - \overline{inlf})]^2 = \sum_{i=1}^{n}(-inlf_i + \overline{inlf})^2 = \sum_{i=1}^{n}(inlf_i - \overline{inlf})^2 ,$$

which is the SST for *inlf*. Because R^2 = 1 – SSR/SST, the *R*-squared is the same in the two regressions.

7.9 (i) Plugging in *u* = 0 and *d* = 1 gives $f_1(z) = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)z$.

(ii) Setting $f_0(z^*) = f_1(z^*)$ gives $\beta_0 + \beta_1 z^* = (\beta_0 + \delta_0) + (\beta_1 + \delta_1)z^*$ or $0 = \delta_0 + \delta_1 z^*$. Therefore, provided $\delta_1 \neq 0$, we have $z^* = -\delta_0 / \delta_1$. Clearly, z^* is positive if and only if δ_0 / δ_1 is negative, which means δ_0 and δ_1 must have opposite signs.

(iii) Using part (ii) we have $totcoll^* = .357 / .030 = 11.9$ years.

(iv) The estimated years of college where women catch up to men is much too high to be practically relevant. While the estimated coefficient on $female \cdot totcoll$ shows that the gap is reduced at higher levels of college, it is never closed – not even close. In fact, at four years of college, the difference in predicted log wage is still $-.357 + .030(4) = -.237$, or about 21.1% less for women.

SOLUTIONS TO COMPUTER EXERCISES

C7.1 (i) The estimated equation is

$$\begin{aligned}\widehat{colGPA} = \underset{(0.34)}{1.26} + \underset{(.059)}{.152}\,PC + \underset{(.094)}{.450}\,hsGPA + \underset{(.0107)}{.0077}\,ACT - \underset{(.0603)}{.0038}\,mothcoll \\ + \underset{(.0613)}{.0418}\,fathcoll\end{aligned}$$

$$n = 141, \quad R^2 = .222.$$

The estimated effect of *PC* is hardly changed from equation (7.6), and it is still very significant, with $t_{pc} \approx 2.58$.

(ii) The *F* test for joint significance of *mothcoll* and *fathcoll*, with 2 and 135 *df*, is about .24 with *p*-value $\approx$ .78; these variables are jointly very insignificant. It is not surprising the estimates on the other coefficients do not change much when *mothcoll* and *fathcoll* are added to the regression.

(iii) When $hsGPA^2$ is added to the regression, its coefficient is about .337 and its *t* statistic is about 1.56. (The coefficient on *hsGPA* is about –1.803.) This is a borderline case. The quadratic in *hsGPA* has a U-shape, and it only turns up at about $hsGPA^*$ = 2.68, which is hard to interpret. The coefficient of main interest, on *PC*, falls to about .140 but is still significant. Adding $hsGPA^2$ is a simple robustness check of the main finding.

C7.3 (i) H_0: $\beta_{13} = 0$. Using the data in MLB1.RAW gives $\hat{\beta}_{13} \approx .254$, se($\hat{\beta}_{13}$) $\approx .131$. The t statistic is about 1.94, which gives a p-value against a two-sided alternative of just over .05. Therefore, we would reject H_0 at just about the 5% significance level. Controlling for the performance and experience variables, the estimated salary differential between catchers and outfielders is huge, on the order of 100·[exp(.254) – 1] $\approx$ 28.9% [using equation (7.10)].

(ii) This is a joint null, H_0: $\beta_9 = 0$, $\beta_{10} = 0$, …, $\beta_{13} = 0$. The F statistic, with 5 and 339 df, is about 1.78, and its p-value is about .117. Thus, we cannot reject H_0 at the 10% level.

(iii) Parts (i) and (ii) are roughly consistent. The evidence against the joint null in part (ii) is weaker because we are testing, along with the marginally significant *catcher*, several other insignificant variables (especially *thrdbase* and *shrtstop*, which has absolute t statistics well below one).

C7.5 The estimated equation is

$$\widehat{\log(salary)} = \underset{(0.29)}{4.30} + \underset{(.034)}{.288}\log(sales) + \underset{(.0040)}{.0167}\,roe - \underset{(.109)}{.226}\,rosneg$$

$$n = 209, \quad R^2 = .297, \quad \bar{R}^2 = .286.$$

The coefficient on *rosneg* implies that if the CEO's firm had a negative return on its stock over the 1988 to 1990 period, the CEO salary was predicted to be about 22.6% lower, for given levels of *sales* and *roe*. The t statistic is about –2.07, which is significant at the 5% level against a two-sided alternative.

C7.7 (i) When $educ = 12.5$, the approximate proportionate difference in estimated *wage* between women and men is $-.227 - .0056(12.5) = -.297$. When $educ = 0$, the difference is $-.227$. So the differential at 12.5 years of education is about 7 percentage points greater.

(ii) We can write the model underlying (7.18) as

$$\begin{aligned}
\log(wage) &= \beta_0 + \delta_0\, female + \beta_1 educ + \delta_1\, female \cdot educ + other\ factors \\
&= \beta_0 + (\delta_0 + 12.5\delta_1)\, female + \beta_1 educ + \delta_1\, female \cdot (educ - 12.5) \\
&\quad + other\ factors \\
&\equiv \beta_0 + \theta_0\, female + \beta_1 educ + \delta_1\, female \cdot (educ - 12.5) + other\ factors,
\end{aligned}$$

where $\theta_0 \equiv \delta_0 + 12.5\delta_1$ is the gender differential at 12.5 years of education. When we run this regression we obtain about –.294 as the coefficient on *female* (which differs from –.297 due to rounding error). Its standard error is about .036.

(iii) The t statistic on *female* from part (ii) is about –8.17, which is very significant. This is because we are estimating the gender differential at a reasonable number of years of education, 12.5, which is close to the average. In equation (7.18), the coefficient on *female* is the gender differential when *educ* = 0. There are no people of either gender with close to zero years of education, and so we cannot hope – nor do we want to – to estimate the gender differential at *educ* = 0.

C7.9 (i) About .392, or 39.2%.

(ii) The estimated equation is

$$\begin{array}{llllllll} \widehat{e401k} = & -.506 & + .0124\, inc & - .000062\, inc^2 & + .0265\, age & - .00031\, age^2 & - .0035\, male \\ & (.081) & (.0006) & (.000005) & (.0039) & (.00005) & (.0121) \end{array}$$

$n = 9{,}275$, $R^2 = .094$.

(iii) 401(k) eligibility clearly depends on income and age in part (ii). Each of the four terms involving *inc* and *age* have very significant t statistics. On the other hand, once income and age are controlled for, there seems to be no difference in eligibility by gender. The coefficient on *male* is very small – at given income and age, males are estimated to have a .0035 lower probability of being 401(k) eligible – and it has a very small t statistic.

(iv) Somewhat surprisingly, out of 9,275 fitted values, none is outside the interval [0,1]. The smallest fitted value is about .030 and the largest is about .697. This means one theoretical problem with the LPM – the possibility of generating silly probability estimates – does not materialize in this application.

(v) Using the given rule, 2,460 families are predicted to be eligible for a 401(k) plan.

(vi) Of the 5,638 families actually ineligible for a 401(k) plan, about 81.7 are correctly predicted not to be eligible. Of the 3,637 families actually eligible, only 39.3 percent are correctly predicted to be eligible.

(vii) The overall percent correctly predicted is a weighted average of the two percentages obtained in part (vi). As we saw there, the model does a good job of predicting when a family is ineligible. Unfortunately, it does less well – predicting correctly less than 40% of the time – in predicting that a family is eligible for a 401(k).

(viii) The estimated equation is

$$\widehat{e401k} = -.502 + .0123\, inc - .000061\, inc^2 + .0265\, age - .00031\, age^2$$

(.081) (.0006) (.000005) (.0039) (.00005)

$$- .0038\, male + .0198\, pira$$
$$(.0121) \quad (.0122)$$

n = 9,275, R^2 = .095.

The coefficient on *pira* means that, other things equal, IRA ownership is associated with about a .02 higher probability of being eligible for a 401(k) plan. However, the t statistic is only about 1.62, which gives a two-sided p-value = .105. So *pira* is not significant at the 10% level against a two-sided alternative.

C7.11 (i) The average is 19.072, the standard deviation is 63.964, the smallest value is –502.302, and the largest value is 1,536.798. Remember, these are in thousands of dollars.

(ii) This can be easily done by regressing *nettfa* on *e401k* and doing a t test on $\hat{\beta}_{e401k}$; the estimate is the average difference in *nettfa* for those eligible for a 401(k) and those not eligible. Using the 9,275 observations gives $\hat{\beta}_{e401k} = 18.858$ and $t_{e401k} = 14.01$. Therefore, we strongly reject the null hypothesis that there is no difference in the averages. The coefficient implies that, on average, a family eligible for a 401(k) plan has $18,858 more on net total financial assets.

(iii) The equation estimated by OLS is

$$\widehat{nettfa} = \underset{(9.96)}{23.09} + \underset{(1.277)}{9.705}\, e401k - \underset{(.075)}{.278}\, inc + \underset{(.0006)}{.0103}\, inc^2 - \underset{(.483)}{1.972}\, age + \underset{(.0055)}{.0348}\, age^2$$

n = 9,275, R^2 = .202

Now, holding income and age fixed, a 401(k)-eligible family is estimated to have $9,705 more in wealth than a non-eligible family. This is just more than half of what is obtained by simply comparing averages.

(iv) Only the interaction $e401k\cdot(age - 41)$ is significant. Its coefficient is .654 (t = 4.98). It shows that the effect of 401(k) eligibility on financial wealth increases with age. Another way to think about it is that *age* has a stronger positive effect on *nettfa* for those with 401(k) eligibility. The coefficient on $e401k\cdot(age - 41)^2$ is –.0038 (t statistic = –.33), so we could drop this term.

(v) The effect of e401k in part (iii) is the same for all ages, 9.705. For the regression in part (iv), the coefficient on *e401k* from part (iv) is about 9.960, which is the effect at the average age, *age* = 41. Including the interactions increases the estimated effect of *e401k*, but only by $255. If we evaluate the effect in part (iv) at a wide range of ages, we would see more dramatic differences.

(vi) I chose *fsize1* as the base group. The estimated equation is

$$\widehat{nettfa} = \underset{(10.12)}{16.34} + \underset{(1.278)}{9.455}\, e401k - \underset{(.075)}{.240}\, inc + \underset{(.0006)}{.0100}\, inc^2 - \underset{(.483)}{1.495}\, age + \underset{(.0055)}{.0290}\, age^2$$

$$- \underset{(1.818)}{.859}\, fsize2 - \underset{(1.877)}{4.665}\, fsize3 - \underset{(1.868)}{6.314}\, fsize4 - \underset{(2.101)}{7.361}\, fsize5$$

$n = 9{,}275,\ R^2 = .204,\ \text{SSR} = 30{,}215{,}207.5$

The F statistic for joint significance of the four family size dummies is about 5.44. With 4 and 9,265 *df*, this gives p-value = .0002. So the family size dummies are jointly significant.

(vii) The SSR for the restricted model is from part (vi): $\text{SSR}_r = 30{,}215{,}207.5$. The SSR for the unrestricted model is obtained by adding the SSRs for the five separate family size regressions. I get $\text{SSR}_{ur} = 29{,}985{,}400$. The Chow statistic is $F = [(30{,}215{,}207.5 - 29{,}985{,}400)/29{,}985{,}400]*(9245/20) \approx 3.54$. With 20 and 9,245 *df*, the p-value is essentially zero. In this case, there is strong evidence that the slopes change across family size. Allowing for intercept changes alone is not sufficient. (If you look at the individual regressions, you will see that the signs on the income variables actually change across family size.)

C7.13 (i) 412/660 ≈ .624.

(ii) The OLS estimates of the LPM are

$$\widehat{ecobuy} = \underset{(.165)}{.424} - \underset{(.109)}{.803}\, ecoprc + \underset{(.132)}{.719}\, regprc + \underset{(.00053)}{.00055}\, faminc + \underset{(.013)}{.024}\, hhsize$$

$$+ \underset{(.008)}{.025}\, educ - \underset{(.00125)}{.00050}\, age$$

$n = 660,\ R^2 = .110$

If *ecoprc* increases by, say, 10 cents (.10), then the probability of buying eco-labeled apples falls by about .080. If *regprc* increases by 10 cents, the probability of buying eco-labeled apples increases by about .072. (Of course, we are assuming that the probabilities are not close to the boundaries of zero and one, respectively.)

(iii) The F test, with 4 and 653 *df*, is 4.43, with p-value = .0015. Thus, based on the usual F test, the four non-price variables are jointly very significant. Of the four variables, *educ* appears to have the most important effect. For example, a difference of four years of education implies an increase of .025(4) = .10 in the estimated probability of buying eco-labeled apples. This suggests that more highly educated people are more open to buying produce that is environmentally friendly, which is perhaps expected. Household size (*hhsize*) also has an effect. Comparing a couple with two children to one that has no children – other factors equal – the couple with two children has a .048 higher probability of buying eco-labeled apples.

(iv) The model with log(*faminc*) fits the data slightly better: the R-squared increases to about .112. (We would not expect a large increase in R-squared from a simple change in the functional form.) The coefficient on log(*faminc*) is about .045 (t = 1.55). If log(*faminc*) increases by .10, which means roughly a 10% increase in *faminc*, then P(*ecobuy* = 1) is estimated to increase by about .0045, a pretty small effect.

(v) The fitted probabilities range from about .185 to 1.051, so none are negative. There are two fitted probabilities above 1, which is not a source of concern with 660 observations.

(vi) Using the standard prediction rule – predict one when $\widehat{ecobuy}_i \geq .5$ and zero otherwise – gives the fraction correctly predicted for *ecobuy* = 0 as 102/248 ≈ .411, so about 41.1%. For *ecobuy* = 1, the fraction correctly predicted is 340/412 ≈ .825, or 82.5%. With the usual prediction rule, the model does a much better job predicting the decision to buy eco-labeled apples. (The overall percent correctly predicted is about 67%.)

CHAPTER 8

SOLUTIONS TO PROBLEMS

8.1 Parts (ii) and (iii). The homoskedasticity assumption played no role in Chapter 5 in showing that OLS is consistent. But we know that heteroskedasticity causes statistical inference based on the usual *t* and *F* statistics to be invalid, even in large samples. As heteroskedasticity is a violation of the Gauss-Markov assumptions, OLS is no longer BLUE.

8.3 False. The unbiasedness of WLS and OLS hinges crucially on Assumption MLR.4, and, as we know from Chapter 4, this assumption is often violated when an important variable is omitted. When MLR.4 does not hold, both WLS and OLS are biased. Without specific information on how the omitted variable is correlated with the included explanatory variables, it is not possible to determine which estimator has a small bias. It is possible that WLS would have more bias than OLS or less bias. Because we cannot know, we should not claim to use WLS in order to solve "biases" associated with OLS.

8.5 (i) No. For each coefficient, the usual standard errors and the heteroskedasticity-robust ones are practically very similar.

(ii) The effect is –.029(4) = –.116, so the probability of smoking falls by about .116.

(iii) As usual, we compute the turning point in the quadratic: .020/[2(.00026)] $\approx$ 38.46, so about 38 and one-half years.

(iv) Holding other factors in the equation fixed, a person in a state with restaurant smoking restrictions has a .101 lower chance of smoking. This is similar to the effect of having four more years of education.

(v) We just plug the values of the independent variables into the OLS regression line:

$$\widehat{smokes} = .656 - .069\cdot\log(67.44) + .012\cdot\log(6{,}500) - .029(16) + .020(77) - .00026(77^2) \approx .0052.$$

Thus, the estimated probability of smoking for this person is close to zero. (In fact, this person is not a smoker, so the equation predicts well for this particular observation.)

8.7 (i) This follows from the simple fact that, for uncorrelated random variables, the variance of the sum is the sum of the variances: $\mathrm{Var}(f_i + v_{i,e}) = \mathrm{Var}(f_i) + \mathrm{Var}(v_{i,e}) = \sigma_f^2 + \sigma_v^2$.

(ii) We compute the covariance between any two of the composite errors as

$$\begin{aligned}\mathrm{Cov}(u_{i,e}, u_{i,g}) &= \mathrm{Cov}(f_i + v_{i,e},\, f_i + v_{i,g}) = \mathrm{Cov}(f_i, f_i) + \mathrm{Cov}(f_i, v_{i,g}) + \mathrm{Cov}(v_{i,e}, f_i) + \mathrm{Cov}(v_{i,e}, v_{i,g}) \\ &= \mathrm{Var}(f_i) + 0 + 0 + 0 = \sigma_f^2,\end{aligned}$$

where we use the fact that the covariance of a random variable with itself is its variance and the assumptions that $f_i, v_{i,e},$ and $v_{i,g}$ are pairwise uncorrelated.

(iii) This is most easily solved by writing

$$m_i^{-1}\sum_{e=1}^{m_i} u_{i,e} = m_i^{-1}\sum_{e=1}^{m_i}(f_i + u_{i,e}) = f_i + m_i^{-1}\sum_{e=1}^{m_i} v_{i,e}.$$

Now, by assumption, f_i is uncorrelated with each term in the last sum; therefore, f_i is uncorrelated with $m_i^{-1}\sum_{e=1}^{m_i} v_{i,e}$. It follows that

$$\begin{aligned}\mathrm{Var}\left(f_i + m_i^{-1}\sum_{e=1}^{m_i} v_{i,e}\right) &= \mathrm{Var}(f_i) + \mathrm{Var}\left(m_i^{-1}\sum_{e=1}^{m_i} v_{i,e}\right)\\ &= \sigma_f^2 + \sigma_v^2 / m_i,\end{aligned}$$

where we use the fact that the variance of an average of m_i uncorrelated random variables with common variance (σ_v^2 in this case) is simply the common variance divided by m_i – the usual formula for a sample average from a random sample.

(iv) The standard weighting ignores the variance of the firm effect, σ_f^2. Thus, the (incorrect) weight function used is $1/h_i = m_i$. A valid weighting function is obtained by writing the variance from (iii) as $\mathrm{Var}(\bar{u}_i) = \sigma_f^2[1 + (\sigma_v^2/\sigma_f^2)/m_i] = \sigma_f^2 h_i$. But obtaining the proper weights requires us to know (or be able to estimate) the ratio σ_v^2/σ_f^2. Estimation is possible, but we do not discuss that here. In any event, the usual weight is incorrect. When the m_i are large or the ratio σ_v^2/σ_f^2 is small – so that the firm effect is more important than the individual-specific effect – the correct weights are close to being constant. Thus, attaching large weights to large firms may be quite inappropriate.

SOLUTIONS TO COMPUTER EXERCISES

C8.1 (i) Given the equation

$$sleep = \beta_0 + \beta_1 totwrk + \beta_2 educ + \beta_3 age + \beta_4 age^2 + \beta_5 yngkid + \beta_6 male + u,$$

the assumption that the variance of u given all explanatory variables depends only on gender is

$$Var(u \mid totwrk, educ, age, yngkid, male) = Var(u \mid male) = \delta_0 + \delta_1 male$$

Then the variance for women is simply δ_0 and that for men is $\delta_0 + \delta_1$; the difference in variances is δ_1.

(ii) After estimating the above equation by OLS, we regress $\hat{u}_i^2$ on $male_i$, i = 1,2, …,706 (including, of course, an intercept). We can write the results as

$$\hat{u}^2 = 189{,}359.2 - 28{,}849.6\, male + residual$$
$$\qquad (20{,}546.4) \quad (27{,}296.5)$$
$$n = 706, \quad R^2 = .0016.$$

Because the coefficient on *male* is negative, the estimated variance is higher for women.

(iii) No. The t statistic on *male* is only about –1.06, which is not significant at even the 20% level against a two-sided alternative.

C8.3 After estimating equation (8.18), we obtain the squared OLS residuals $\hat{u}^2$. The full-blown White test is based on the *R*-squared from the auxiliary regression (with an intercept),

$$\hat{u}^2 \text{ on } llotsize,\ lsqrft,\ bdrms,\ llotsize^2,\ lsqrft^2,\ bdrms^2,$$
$$llotsize \cdot lsqrft,\ llotsize \cdot bdrms, \text{ and } lsqrft \cdot bdrms,$$

where "*l*" in front of *lotsize* and *sqrft* denotes the natural log. [See equation (8.19).] With 88 observations the *n*-*R*-squared version of the White statistic is 88(.109) $\approx$ 9.59, and this is the outcome of an (approximately) χ_9^2 random variable. The *p*-value is about .385, which provides little evidence against the homoskedasticity assumption.

C8.5 (i) By regressing *sprdcvr* on an intercept only we obtain $\hat{\mu} \approx .515$ se $\approx .021$). The asymptotic t statistic for H_0: $\mu = .5$ is $(.515 - .5)/.021 \approx .71$, which is not significant at the 10% level, or even the 20% level.

(ii) 35 games were played on a neutral court.

(iii) The estimated LPM is

$$\widehat{sprdcvr} = \underset{(.045)}{.490} + \underset{(.050)}{.035}\, favhome + \underset{(.095)}{.118}\, neutral - \underset{(.050)}{.023}\, fav25 + \underset{(.092)}{.018}\, und25$$
$$n = 553, \quad R^2 = .0034.$$

The variable *neutral* has by far the largest effect – if the game is played on a neutral court, the probability that the spread is covered is estimated to be about .12 higher – and, except for the intercept, its t statistic is the only t statistic greater than one in absolute value (about 1.24).

(iv) Under H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$, the response probability does not depend on any explanatory variables, which means neither the mean nor the variance depends on the explanatory variables. [See equation (8.38).]

(v) The F statistic for joint significance, with 4 and 548 *df*, is about .47 with *p*-value $\approx$ .76. There is essentially no evidence against H_0.

(vi) Based on these variables, it is not possible to predict whether the spread will be covered. The explanatory power is very low, and the explanatory variables are jointly very insignificant. The coefficient on *neutral* may indicate something is going on with games played on a neutral court, but we would not want to bet money on it unless it could be confirmed with a separate, larger sample.

C8.7 (i) The heteroskedasticity-robust standard error for $\hat{\beta}_{white} \approx .129$ is about .026, which is notably higher than the nonrobust standard error (about .020). The heteroskedasticity-robust 95% confidence interval is about .078 to .179, while the nonrobust CI is, of course, narrower, about .090 to .168. The robust CI still excludes the value zero by some margin.

(ii) There are no fitted values less than zero, but there are 231 greater than one. Unless we do something to those fitted values, we cannot directly apply WLS, as $\hat{h}_i$ will be negative in 231 cases.

C8.9 (i) I now get R^2 = .0527, but the other estimates seem okay.

(ii) One way to ensure that the unweighted residuals are being provided is to compare them with the OLS residuals. They will not be the same, of course, but they should not be wildly different.

(iii) The *R*-squared from the regression $\breve{u}_i^2$ on $\breve{y}_i,\ \breve{y}_i^2,\ i = 1,...,807$ is about .027. We use this as $R^2_{\hat{u}^2}$ in equation (8.15) but with *k* = 2. This gives *F* = 11.15, and so the *p*-value is essentially zero.

(iv) The substantial heteroskedasticity found in part (iii) shows that the feasible GLS procedure described on page 279 does not, in fact, eliminate the heteroskedasticity. Therefore, the usual standard errors, *t* statistics, and *F* statistics reported with weighted least squares are not valid, even asymptotically.

(v) Weighted least squares estimation with robust standard errors gives

$$\begin{aligned}\widehat{cigs} = \underset{(37.31)}{5.64} &+ \underset{(.54)}{1.30}\log(income) - \underset{(8.97)}{2.94}\log(cigpric) - \underset{(.149)}{.463}\,educ \\ &+ \underset{(.115)}{.482}\,age - \underset{(.0012)}{.0056}\,age^2 - \underset{(.72)}{3.46}\,restaurn\end{aligned}$$

$n = 807,\ R^2 = .1134$

The substantial differences in standard errors compared with equation (8.36) further indicate that our proposed correction for heteroskedasticity did not fully solve the heteroskedasticity problem. With the exception of *restaurn*, all standard errors got notably bigger; for example, the standard error for log(*cigpric*) doubled. All variables that were statistically significant with the nonrobust standard errors remain significant, but the confidence intervals are much wider in several cases.

C8.11 (i) The usual OLS standard errors are in (·), the heteroskedasticity-robust standard errors are in [·]:

$$\widehat{nettfa} = \underset{\substack{(2.82)\\ [3.23]}}{-17.20} + \underset{\substack{(.080)\\ [.098]}}{.628}\, inc + \underset{\substack{(.0026)\\ [.0044]}}{.0251}\,(age - 25)^2 + \underset{\substack{(2.04)\\ [2.06]}}{2.54}\, male$$

$$- \underset{\substack{(4.40)\\ [6.25]}}{3.83}\, e401k + \underset{\substack{(.124)\\ [.220]}}{.343}\, e401k\cdot inc$$

$n = 2{,}017$, $R^2 = .131$

Although the usual OLS t statistic on the interaction term is about 2.8, the heteroskedasticity-robust t statistic is just under 1.6. Therefore, using OLS, we must conclude the interaction term is only marginally significant. But the coefficient is nontrivial: it implies a much more sensitive relationship between financial wealth and income for those eligible for a 401(k) plan.

(ii) The WLS estimates, with usual WLS standard errors in (·) and the robust ones in [·], are

$$\widehat{nettfa} = \underset{\substack{(2.27)\\ [2.53]}}{-14.09} + \underset{\substack{(.084)\\ [.091]}}{.619}\, inc + \underset{\substack{(.0019)\\ [.0026]}}{.0175}\,(age - 25)^2 + \underset{\substack{(1.56)\\ [1.31]}}{1.78}\, male$$

$$- \underset{\substack{(3.66)\\ [3.51]}}{2.17}\, e401k + \underset{\substack{(.130)\\ [.160]}}{.295}\, e401k\cdot inc$$

$n = 2{,}017$, $R^2 = .114$

The robust t statistic is about 1.84, and so the interaction term is marginally significant (two-sided p-value is about .066).

(iii) The coefficient on $e401k$ literally gives the estimated difference in financial wealth at $inc = 0$, which obviously is not interesting. It is not surprising that it is not statistically different from zero; we obviously cannot hope to estimate the difference at $inc = 0$, nor do we care to.

(iv) When we replace $e401k\cdot inc$ with $e401k\cdot(inc - 30)$, the coefficient on $e401k$ becomes 6.68 (robust t = 3.20). Now, this coefficient is the estimated difference in *nettfa* between those with and without 401(k) eligibility at roughly the average income, \$30,000. Naturally, we can estimate this much more precisely, and its magnitude (\$6,680) makes sense.

CHAPTER 9

SOLUTIONS TO PROBLEMS

9.1 There is functional form misspecification if $\beta_6 \neq 0$ or $\beta_7 \neq 0$, where these are the population parameters on *ceoten*^{2} and *comten*^{2}, respectively. Therefore, we test the joint significance of these variables using the *R*-squared form of the *F* test: $F = [(.375 - .353)/(1 - .375)][(177 - 8)/2] \approx 2.97$. With 2 and ∞ *df*, the 10% critical value is 2.30 awhile the 5% critical value is 3.00. Thus, the *p*-value is slightly above .05, which is reasonable evidence of functional form misspecification. (Of course, whether this has a practical impact on the estimated partial effects for various levels of the explanatory variables is a different matter.)

9.3 (i) Eligibility for the federally funded school lunch program is very tightly linked to being economically disadvantaged. Therefore, the percentage of students eligible for the lunch program is very similar to the percentage of students living in poverty.

(ii) We can use our usual reasoning on omitting important variables from a regression equation. The variables log(*expend*) and *lnchprg* are negatively correlated: school districts with poorer children spend, on average, less on schools. Further, $\beta_3 < 0$. From Table 3.2, omitting *lnchprg* (the proxy for *poverty*) from the regression produces an upward biased estimator of β_1 [ignoring the presence of log(*enroll*) in the model]. So when we control for the poverty rate, the effect of spending falls.

(iii) Once we control for *lnchprg*, the coefficient on log(*enroll*) becomes negative and has a *t* of about –2.17, which is significant at the 5% level against a two-sided alternative. The coefficient implies that $\Delta\widehat{math10} \approx -(1.26/100)(\%\Delta enroll) = -.0126(\%\Delta enroll)$. Therefore, a 10% increase in enrollment leads to a drop in *math10* of .126 percentage points.

(iv) Both *math10* and *lnchprg* are percentages. Therefore, a ten percentage point increase in *lnchprg* leads to about a 3.23 percentage point fall in *math10*, a sizeable effect.

(v) In column (1) we are explaining very little of the variation in pass rates on the MEAP math test: less than 3%. In column (2), we are explaining almost 19% (which still leaves much variation unexplained). Clearly most of the variation in *math10* is explained by variation in *lnchprg*. This is a common finding in studies of school performance: family income (or related factors, such as living in poverty) are much more important in explaining student performance than are spending per student or other school characteristics.

9.5 The sample selection in this case is arguably endogenous. Because prospective students may look at campus crime as one factor in deciding where to attend college, colleges with high crime rates have an incentive not to report crime statistics. If this is the case, then the chance of appearing in the sample is negatively related to *u* in the crime equation. (For a given school size, higher *u* means more crime, and therefore a smaller probability that the school reports its crime figures.)

9.7 (i) Following the hint, we compute $\text{Cov}(w, y)$ and $\text{Var}(w)$, where $y = \beta_0 + \beta_1 x^* + u$ and $w = (z_1 + ... + z_m)/m$. First, because $z_h = x^* + e_h$, it follows that $w = x^* + \bar{e}$, where $\bar{e}$ is the average of the *m* measures (in the population). Now, by assumption, x^* is uncorrelated with each e_h, and the e_h are pairwise uncorrelated. Therefore,

$$\text{Var}(w) = \text{Var}(x^*) + \text{Var}(\bar{e}) = \sigma_{x^*}^2 + \sigma_e^2 / m,$$

where we use $\text{Var}(\bar{e}) = \sigma_e^2 / m$. Next,

$$\text{Cov}(w, y) = \text{Cov}(x^* + \bar{e}, \beta_0 + \beta_1 x^* + u) = \beta_1 \text{Cov}(x^*, x^*) = \beta_1 \text{Var}(x^*),$$

where we use the assumption that e_h is uncorrelated with *u* for all *h* and x^* is uncorrelated with *u*. Combining the two pieces gives

$$\frac{\text{Cov}(w, y)}{\text{Var}(w)} = \beta_1 \left\{ \frac{\sigma_{x^*}^2}{[\sigma_{x^*}^2 + (\sigma_e^2 / m)]} \right\},$$

which is what we wanted to show.

(ii) Because $\sigma_e^2 / m < \sigma_e^2$ for all $m > 1$, $\sigma_{x^*}^2 + (\sigma_e^2 / m) < \sigma_{x^*}^2 + \sigma_e^2$ for all $m > 1$. Therefore

$$1 > \frac{\sigma_{x^*}^2}{[\sigma_{x^*}^2 + (\sigma_e^2 / m)]} > \frac{\sigma_{x^*}^2}{\sigma_{x^*}^2 + \sigma_e^2},$$

which means the term multiplying β_1 is closer to one when *m* is larger. We have shown that the bias in $\bar{\beta}_1$ is smaller as *m* increases. As *m* grows, the bias disappears completely. Intuitively, this makes sense. The average of several mismeasured variables has less measurement error than a single mismeasured variable. As we average more and more such variables, the attenuation bias can become very small.

SOLUTIONS TO COMPUTER EXERCISES

C9.1 (i) To obtain the RESET *F* statistic, we estimate the model in Computer Exercise 7.5 and obtain the fitted values, say $\widehat{lsalary}_i$. To use the version of RESET in (9.3), we add $(\widehat{lsalary}_i)^2$ and $(\widehat{lsalary}_i)^3$ and obtain the *F* test for joint significance of these variables. With 2 and 203 *df*, the *F* statistic is about 1.33 and *p*-value ≈ .27, which means that there is not much concern about functional form misspecification.

(ii) Interestingly, the heteroskedasticity-robust F-type statistic is about 2.24 with p-value $\approx$ .11, so there is stronger evidence of some functional form misspecification with the robust test. But it is probably not strong enough to worry about.

C9.3 (i) If the grants were awarded to firms based on firm or worker characteristics, *grant* could easily be correlated with such factors that affect productivity. In the simple regression model, these are contained in u.

(ii) The simple regression estimates using the 1988 data are

$$\widehat{\log(scrap)} = \underset{(.241)}{.409} + \underset{(.406)}{.057}\, grant$$

$$n = 54, \quad R^2 = .0004.$$

The coefficient on *grant* is actually positive, but not statistically different from zero.

(iii) When we add $\log(scrap_{87})$ to the equation, we obtain

$$\widehat{\log(scrap_{88})} = \underset{(.089)}{.021} - \underset{(.147)}{.254}\, grant_{88} + \underset{(.044)}{.831}\log(scrap_{87})$$

$$n = 54, \quad R^2 = .873,$$

where the year subscripts are for clarity. The t statistic for H_0: $\beta_{grant} = 0$ is $-.254/.147 \approx -1.73$. We use the 5% critical value for 40 df in Table G.2: -1.68. Because $t = -1.73 < -1.68$, we reject H_0 in favor of H_1: $\beta_{grant} < 0$ at the 5% level.

(iv) The t statistic is $(.831 - 1)/.044 \approx -3.84$, which is a strong rejection of H_0.

(v) With the heteroskedasticity-robust standard error, the t statistic for $grant_{88}$ is $-.254/.142 \approx -1.79$, so the coefficient is even more significantly less than zero when we use the heteroskedasticity-robust standard error. The t statistic for H_0: $\beta_{\log(scrap_{87})} = 1$ is $(.831 - 1)/.071 \approx -2.38$, which is notably smaller than before, but it is still pretty significant.

C9.5 With *sales* defined to be in billions of dollars, we obtain the following estimated equation using all companies in the sample:

$$\widehat{rdintens} = \underset{(0.63)}{2.06} + \underset{(.139)}{.317}\, sales - \underset{(.0037)}{.0074}\, sales^2 + \underset{(.044)}{.053}\, profmarg$$

$$n = 32, \quad R^2 = .191, \quad \bar{R}^2 = .104.$$

When we drop the largest company (with sales of roughly $39.7 billion), we obtain

$$\widehat{rdintens} = \underset{(0.72)}{1.98} + \underset{(.239)}{.361}\, sales - \underset{(.0131)}{.0103}\, sales^2 + \underset{(.046)}{.055}\, profmarg$$

$$n = 31, \quad R^2 = .191, \quad \bar{R}^2 = .101.$$

When the largest company is left in the sample, the quadratic term is statistically significant, even though the coefficient on the quadratic is less in absolute value than when we drop the largest firm. What is happening is that by leaving in the large sales figure, we greatly increase the variation in both *sales* and $sales^2$; as we know, this reduces the variances of the OLS estimators (see Section 3.4). The *t* statistic on $sales^2$ in the first regression is about –2, which makes it almost significant at the 5% level against a two-sided alternative. If we look at Figure 9.1, it is not surprising that a quadratic is significant when the large firm is included in the regression: *rdintens* is relatively small for this firm even though its sales are very large compared with the other firms. Without the largest firm, a linear relationship between *rdintens* and *sales* seems to suffice.

C9.7 (i) 205 observations out of the 1,989 records in the sample have *obrate* > 40. (Data are missing for some variables, so not all of the 1,989 observations are used in the regressions.)

(ii) When observations with *obrat* > 40 are excluded from the regression in part (iii) of Problem 7.16, we are left with 1,768 observations. The coefficient on *white* is about .129 (se ≈ .020). To three decimal places, these are the same estimates we got when using the entire sample (see Computer Exercise C7.8). Perhaps this is not very surprising since we only lost 203 out of 1,971 observations. However, regression results can be very sensitive when we drop over 10% of the observations, as we have here.

(iii) The estimates from part (ii) show that $\hat{\beta}_{white}$ does not seem very sensitive to the sample used, although we have tried only one way of reducing the sample.

C9.9 (i) The equation estimated by OLS is

$$\widehat{nettfa} = \underset{(9.992)}{21.198} - \underset{(.075)}{.270}\, inc + \underset{(.0006)}{.0102}\, inc^2 - \underset{(.483)}{1.940}\, age + \underset{(.0055)}{.0346}\, age^2$$
$$+ \underset{(1.486)}{3.369}\, male + \underset{(1.277)}{9.713}\, e401k$$

$$n = 9{,}275, \quad R^2 = .202$$

The coefficient on *e401k* means that, holding other things in the equation fixed, the average level of net financial assets is about $9,713 higher for a family eligible for a 401(k) than for a family not eligible.

(ii) The OLS regression of $\hat{u}_i^2$ on inc_i, inc_i^2, age_i, age_i^2, $male_i$, and $e401k_i$ gives $R_{\hat{u}^2}^2$ = .0374, which translates into F = 59.97. The associated p-value, with 6 and 9,268 df, is essentially zero. Consequently, there is strong evidence of heteroskedasticity, which means that u and the explanatory variables cannot be independent [even though $E(u|x_1,x_2,\ldots,x_k) = 0$ is possible].

(iii) The equation estimated by LAD is

$$\widehat{nettfa} = \underset{(1.382)}{12.491} - \underset{(.010)}{.262}\, inc + \underset{(.00008)}{.00709}\, inc^2 - \underset{(.067)}{.723}\, age + \underset{(.0008)}{.0111}\, age^2$$
$$+ \underset{(.205)}{1.018}\, male + \underset{(.177)}{3.737}\, e401k$$

n = 9,275, Psuedo R^2 = .109

Now, the coefficient on *e401k* means that, at given income, age, and gender, the median difference in net financial assets between families with and without 401(k) eligibility is about \$3,737.

(iv) The findings from parts (i) and (iii) are not in conflict. We are finding that 401(k) eligibility has a larger effect on mean wealth than on median wealth. Finding different mean and median effects for a variable such as *nettfa*, which has a highly skewed distribution, is not surprising. Apparently, 401(k) eligibility has some large effects at the upper end of the wealth distribution, and these are reflected in the mean. The median is much less sensitive to effects at the upper end of the distribution.

C9.11 (i) The regression gives $\hat{\beta}_{exec}$ = .085 with t = .30. The positive coefficient means that there is no deterrent effect, and the coefficient is not statistically different from zero.

(ii) Texas had 34 executions over the period, which is more than three times the next highest state (Virginia with 11). When a dummy variable is added for Texas, its t statistic is –.32, which is not unusually large. (The coefficient is large in magnitude, –8.31, but the studentized residual is not large.) We would not characterize Texas as an outlier.

(iii) When the lagged murder rate is added, $\hat{\beta}_{exec}$ becomes –.071 with t = –2.34. The coefficient changes sign and becomes nontrivial: each execution is estimated to reduce the murder rate by .071 (murders per 100,000 people).

(iv) When a Texas dummy is added to the regression from part (iii), its t is only –.37 (and the coefficient is only –1.02). So, it is not an outlier here, either. Dropping TX from the regression reduces the magnitude of the coefficient to –.045 with t = –0.60. Texas accounts for much of the sample variation in *exec*, and dropping it gives a very imprecise estimate of the deterrent effect.

CHAPTER 10

SOLUTIONS TO PROBLEMS

10.1 (i) Disagree. Most time series processes are correlated over time, and many of them strongly correlated. This means they cannot be independent across observations, which simply represent different time periods. Even series that do appear to be roughly uncorrelated – such as stock returns – do not appear to be independently distributed, as you will see in Chapter 12 under dynamic forms of heteroskedasticity.

(ii) Agree. This follows immediately from Theorem 10.1. In particular, we do not need the homoskedasticity and no serial correlation assumptions.

(iii) Disagree. Trending variables are used all the time as dependent variables in a regression model. We do need to be careful in interpreting the results because we may simply find a spurious association between y_t and trending explanatory variables. Including a trend in the regression is a good idea with trending dependent or independent variables. As discussed in Section 10.5, the usual R-squared can be misleading when the dependent variable is trending.

(iv) Agree. With annual data, each time period represents a year and is not associated with any season.

10.3 Write

$$y^* = \alpha_0 + (\delta_0 + \delta_1 + \delta_2)z^* = \alpha_0 + LRP \cdot z^*,$$

and take the change: $\Delta y^* = LRP \cdot \Delta z^*$.

10.5 The functional form was not specified, but a reasonable one is

$$\log(hsestrts_t) = \alpha_0 + \alpha_1 t + \delta_1 Q2_t + \delta_2 Q3_t + \delta_3 Q3_t + \beta_1 int_t + \beta_2 \log(pcinc_t) + u_t,$$

Where $Q2_t$, $Q3_t$, and $Q4_t$ are quarterly dummy variables (the omitted quarter is the first) and the other variables are self-explanatory. This inclusion of the linear time trend allows the dependent variable and $\log(pcinc_t)$ to trend over time (int_t probably does not contain a trend), and the quarterly dummies allow all variables to display seasonality. The parameter β_2 is an elasticity and $100 \cdot \beta_1$ is a semi-elasticity.

10.7 (i) pe_{t-1} and pe_{t-2} must be increasing by the same amount as pe_t.

(ii) The long-run effect, by definition, should be the change in gfr when pe increases permanently. But a permanent increase means the level of pe increases and stays at the new level, and this is achieved by increasing pe_{t-2}, pe_{t-1}, and pe_t by the same amount.

SOLUTIONS TO COMPUTER EXERCISES

C10.1 Let *post79* be a dummy variable equal to one for years after 1979, and zero otherwise. Adding *post79* to equation 10.15) gives

$$\begin{aligned}\widehat{i3}_t &= \underset{(0.43)}{1.30} + \underset{(.076)}{.608}\, inf_t + \underset{(.120)}{.363}\, def_t + \underset{(0.51)}{1.56}\, post79_t \\ n &= 56,\ \ R^2 = .664,\ \ \bar{R}^2 = .644.\end{aligned}$$

The coefficient on *post79* is statistically significant (t statistic $\approx$ 3.06) and economically large: accounting for inflation and deficits, *i3* was about 1.56 points higher on average in years after 1979. The coefficient on *def* falls once *post79* is included in the regression.

C10.3 Adding log(*prgnp*) to equation (10.38) gives

$$\begin{aligned}\widehat{\log(prepop_t)} = &\underset{(1.26)}{-6.66} - \underset{(.040)}{.212}\log(mincov_t) + \underset{(.222)}{.486}\log(usgnp_t) + \underset{(.080)}{.285}\log(prgnp_t) \\ &- \underset{(.005)}{.027}\, t \\ n = 38,\ \ &R^2 = .889,\ \ \bar{R}^2 = .876.\end{aligned}$$

The coefficient on log($prgnp_t$) is very statistically significant (t statistic $\approx$ 3.56). Because the dependent and independent variable are in logs, the estimated elasticity of *prepop* with respect to *prgnp* is .285. Including log(*prgnp*) actually increases the size of the minimum wage effect: the estimated elasticity of *prepop* with respect to *mincov* is now –.212, as compared with –.169 in equation (10.38).

C10.5 (i) The coefficient on the time trend in the regression of log(*uclms*) on a linear time trend and 11 monthly dummy variables is about –.0139 (se $\approx$ .0012), which implies that monthly unemployment claims fell by about 1.4% per month on average. The trend is very significant. There is also very strong seasonality in unemployment claims, with 6 of the 11 monthly dummy variables having absolute t statistics above 2. The F statistic for joint significance of the 11 monthly dummies yields p-value $\approx$ .0009.

(ii) When *ez* is added to the regression, its coefficient is about –.508 (se $\approx$ .146). Because this estimate is so large in magnitude, we use equation (7.10): unemployment claims are estimated to fall 100[1 – exp(–.508)] $\approx$ 39.8% after enterprise zone designation.

(iii) We must assume that around the time of *EZ* designation there were not other external factors that caused a shift down in the trend of log(*uclms*). We have controlled for a time trend and seasonality, but this may not be enough.

C10.7 (i) The estimated equation is

$$\widehat{gc}_t = .0081 + .571\, gy_t$$
$$(.0019) \quad (.067)$$
$$n = 36, \quad R^2 = .679.$$

This equation implies that if income growth increases by one percentage point, consumption growth increases by .571 percentage points. The coefficient on gy_t is very statistically significant (t statistic $\approx$ 8.5).

(ii) Adding gy_{t-1} to the equation gives

$$\widehat{gc}_t = .0064 + .552\, gy_t + .096\, gy_{t-1}$$
$$(.0023) \quad (.070) \quad (.069)$$
$$n = 35, \quad R^2 = .695.$$

The t statistic on gy_{t-1} is only about 1.39, so it is not significant at the usual significance levels. (It is significant at the 20% level against a two-sided alternative.) In addition, the coefficient is not especially large. At best there is weak evidence of adjustment lags in consumption.

(iii) If we add $r3_t$ to the model estimated in part (i) we obtain

$$\widehat{gc}_t = .0082 + .578\, gy_t + .00021\, r3_t$$
$$(.0020) \quad (.072) \quad (.00063)$$
$$n = 36, \quad R^2 = .680.$$

The t statistic on $r3_t$ is very small. The estimated coefficient is also practically small: a one-point increase in $r3_t$ reduces consumption growth by about .021 percentage points.

C10.9 (i) The sign of β_2 is fairly clear-cut: as interest rates rise, stock returns fall, so $\beta_2 < 0$. Higher interest rates imply that T-bill and bond investments are more attractive, and also signal a future slowdown in economic activity. The sign of β_1 is less clear. While economic growth can be a good thing for the stock market, it can also signal inflation, which tends to depress stock prices.

(ii) The estimated equation is

$$\widehat{rsp500}_t = 18.84 + .036\, pcip_t - 1.36\, i3_t$$
$$(3.27) \quad (.129) \quad (0.54)$$
$$n = 557, \quad R^2 = .012.$$

A one percentage point increase in industrial production growth is predicted to increase the stock market return by .036 percentage points (a very small effect). On the other hand, a one percentage point increase in interest rates decreases the stock market return by an estimated 1.36 percentage points.

(iii) Only *i3* is statistically significant with *t* statistic ≈ -2.52.

(iv) The regression in part (i) has nothing directly to say about predicting stock returns because the explanatory variables are dated contemporaneously with *rsp500*. In other words, we do not know $i3_t$ before we know $rsp500_t$. What the regression in part (i) says is that a change in *i3* is associated with a contemporaneous change in *rsp500*.

C10.11 (i) The variable *beltlaw* becomes one at $t = 61$, which corresponds to January, 1986. The variable *spdlaw* goes from zero to one at $t = 77$, which corresponds to May, 1987.

(ii) The OLS regression gives

$$\begin{aligned}\widehat{\log(totacc)} = \underset{(.019)}{10.469} &+ \underset{(.00016)}{.00275}\, t - \underset{(.0244)}{.0427}\, feb + \underset{(.0244)}{.0798}\, mar + \underset{(.0245)}{.0185}\, apr \\ &+ \underset{(.0245)}{.0321}\, may + \underset{(.0245)}{.0202}\, jun + \underset{(.0245)}{.0376}\, jul + \underset{(.0245)}{.0540}\, aug \\ &+ \underset{(.0245)}{.0424}\, sep + \underset{(.0245)}{.0821}\, oct + \underset{(.0245)}{.0713}\, nov + \underset{(.0245)}{.0962}\, dec\end{aligned}$$

$n = 108$, $R^2 = .797$

When multiplied by 100, the coefficient on *t* gives roughly the average monthly percentage growth in *totacc*, ignoring seasonal factors. In other words, once seasonality is eliminated, *totacc* grew by about .275% per month over this period, or, 12(.275) = 3.3% at an annual rate.

There is pretty clear evidence of seasonality. Only February has a lower number of total accidents than the base month, January. The peak is in December: roughly, there are 9.6% accidents more in December over January in the average year. The *F* statistic for joint significance of the monthly dummies is $F = 5.15$. With 11 and 95 *df*, this give a *p*-value essentially equal to zero.

(iii) I will report only the coefficients on the new variables:

$$\widehat{\log(totacc)} = \underset{(.063)}{10.640} + \ldots + \underset{(.00378)}{.00333}\, wkends - \underset{(.0034)}{.0212}\, unem$$

$$- \underset{(.0126)}{.0538}\, spdlaw + \underset{(.0142)}{.0954}\, beltlaw$$

$n = 108$, $R^2 = .910$

The negative coefficient on *unem* makes sense if we view *unem* as a measure of economic activity. As economic activity increases – *unem* decreases – we expect more driving, and therefore more accidents. The estimate that a one percentage point increase in the unemployment rate reduces total accidents by about 2.1%. A better economy does have costs in terms of traffic accidents.

(iv) At least initially, the coefficients on *spdlaw* and *beltlaw* are not what we might expect. The coefficient on *spdlaw* implies that accidents dropped by about 5.4% *after* the highway speed limit was increased from 55 to 65 miles per hour. There are at least a couple of possible explanations. One is that people because safer drivers after the increased speed limiting, recognizing that the must be more cautious. It could also be that some other change – other than the increased speed limit or the relatively new seat belt law – caused lower total number of accidents, and we have not properly accounted for this change.

The coefficient on *beltlaw* also seems counterintuitive at first. But, perhaps people became less cautious once they were forced to wear seatbelts.

(v) The average of *prcfat* is about .886, which means, on average, slightly less than one percent of all accidents result in a fatality. The highest value of *prcfat* is 1.217, which means there was one month where 1.2% of all accidents resulting in a fatality.

(vi) As in part (iii), I do not report the coefficients on the time trend and seasonal dummy variables:

$$\widehat{prcfat} = \underset{(.103)}{1.030} + \ldots + \underset{(.00616)}{.00063}\, wkends - \underset{(.0055)}{.0154}\, unem$$

$$+ \underset{(.0206)}{.0671}\, spdlaw - \underset{(.0232)}{.0295}\, beltlaw$$

$n = 108$, $R^2 = .717$

Higher speed limits are estimated to increase the percent of fatal accidents, by .067 percentage points. This is a statistically significant effect. The new seat belt law is estimated to decrease the percent of fatal accidents by about .03, but the two-sided p-value is about .21.

Interestingly, increased economic activity also increases the percent of fatal accidents. This may be because more commercial trucks are on the roads, and these probably increase the chance that an accident results in a fatality.

C10.13 (i) The estimated equation is

$$\widehat{gwage232} = \underset{(.0004)}{.0022} + \underset{(.001)}{.151}\, gmwage + \underset{(.082)}{.244}\, gcpi$$

$n = 611$, $R^2 = .293$

The coefficient on *gmwage* implies that a one percentage point growth in the minimum wage is estimated to increase the growth in *wage*232 by about .151 percentage points.

(ii) When 12 lags of *gmwage* are added, the sum of all coefficients is about .198, which is somewhat higher than the .151 obtained from the static regression. Plus, the *F* statistic for lags 1 through 12 given *p*-value = .058, which shows they are jointly, marginally statistically significant. (Lags 8 through 12 have fairly large coefficients, and some individual *t* statistics are significant at the 5% level.)

(iii) The estimated equation is

$$\widehat{gemp232} = \underset{(.0010)}{-.0004} - \underset{(.0228)}{.0019}\, gmwage - \underset{(.1938)}{.0055}\, gcpi$$

$n = 611$, $R^2 = .000$

The coefficient on *gmwage* is puny with a very small *t* statistic. In fact, the *R*-squared is practically zero, which means neither *gmwage* nor *gcpi* has any effect on employment growth in sector 232.

(iv) Adding lags of *gmwage* does not change the basic story. The *F* test of joint significance of *gmwage* and lags 1 through 12 of *gmwage* gives *p*-value = .439. The coefficients change sign and none is individually statistically significant at the 5% level. Therefore, there is little evidence that minimum wage growth affects employment growth in sector 232, either in the short run or the long run.

CHAPTER 11

SOLUTIONS TO PROBLEMS

11.1 Because of covariance stationarity, γ_0 = Var(x_t) does not depend on t, so sd(x_{t+h}) = $\sqrt{\gamma_0}$ for any $h \geq 0$. By definition, Corr(x_t,x_{t+h}) = Cov(x_t,x_{t+h})/[sd(x_t)$\cdot$sd(x_{t+h})] = $\gamma_h/(\sqrt{\gamma_0}\cdot\sqrt{\gamma_0}) = \gamma_h/\gamma_0$.

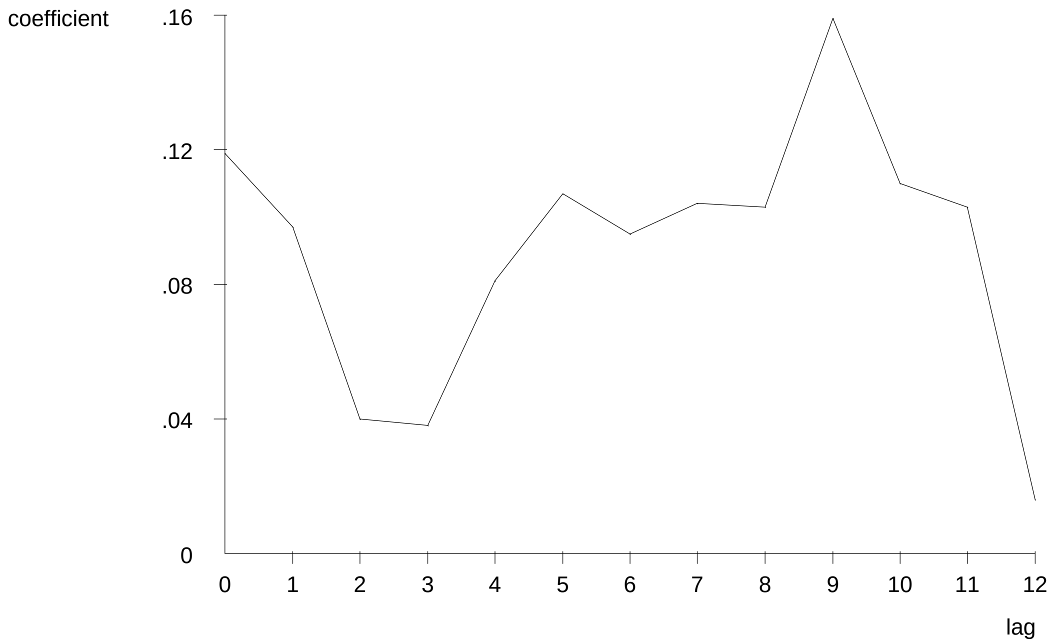
11.3 (i) E(y_t) = E($z + e_t$) = E(z) + E(e_t) = 0. Var(y_t) = Var($z + e_t$) = Var(z) + Var(e_t) + 2Cov(z,e_t) = $\sigma_z^2 + \sigma_e^2 + 2\cdot 0 = \sigma_z^2 + \sigma_e^2$. Neither of these depends on t.

(ii) We assume $h > 0$; when $h = 0$ we obtain Var(y_t). Then Cov(y_t,y_{t+h}) = E($y_t y_{t+h}$) = E[(z + e_t)(z + e_{t+h})] = E(z^2) + E(ze_{t+h}) + E($e_t z$) + E($e_t e_{t+h}$) = E(z^2) = σ_z^2 because $\{e_t\}$ is an uncorrelated sequence (it is an independent sequence and z is uncorrelated with e_t for all t. From part (i) we know that E(y_t) and Var(y_t) do not depend on t and we have shown that Cov(y_t,y_{t+h}) depends on neither t nor h. Therefore, $\{y_t\}$ is covariance stationary.

(iii) From Problem 11.1 and parts (i) and (ii), Corr(y_t,y_{t+h}) = Cov(y_t,y_{t+h})/Var(y_t) = $\sigma_z^2/(\sigma_z^2 + \sigma_e^2) > 0$.

(iv) No. The correlation between y_t and y_{t+h} is the same positive value obtained in part (iii) now matter how large is h. In other words, no matter how far apart y_t and y_{t+h} are, their correlation is always the same. Of course, the persistent correlation across time is due to the presence of the time-constant variable, z.

11.5 (i) The following graph gives the estimated lag distribution:



By some margin, the largest effect is at the ninth lag, which says that a temporary increase in wage inflation has its largest effect on price inflation nine months later. The smallest effect is at the twelfth lag, which hopefully indicates (but does not guarantee) that we have accounted for enough lags of *gwage* in the FLD model.

(ii) Lags two, three, and twelve have *t* statistics less than two. The other lags are statistically significant at the 5% level against a two-sided alternative. (Assuming either that the CLM assumptions hold for exact tests or Assumptions TS.1′ through TS.5′ hold for asymptotic tests.)

(iii) The estimated LRP is just the sum of the lag coefficients from zero through twelve: 1.172. While this is greater than one, it is not much greater, and the difference from unity could be due to sampling error.

(iv) The model underlying and the estimated equation can be written with intercept α_0 and lag coefficients $\delta_0, \delta_1, \ldots, \delta_{12}$. Denote the LRP by $\theta_0 = \delta_0 + \delta_1 + \ldots + \delta_{12}$. Now, we can write $\delta_0 = \theta_0 - \delta_1 - \delta_2 - \ldots - \delta_{12}$. If we plug this into the FDL model we obtain (with $y_t = gprice_t$ and $z_t = gwage_t$)

$$\begin{aligned} y_t &= \alpha_0 + (\theta_0 - \delta_1 - \delta_2 - \ldots - \delta_{12})z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \ldots + \delta_{12} z_{t-12} + u_t \\ &= \alpha_0 + \theta_0 z_t + \delta_1(z_{t-1} - z_t) + \delta_2(z_{t-2} - z_t) + \ldots + \delta_{12}(z_{t-12} - z_t) + u_t. \end{aligned}$$

Therefore, we regress y_t on z_t, $(z_{t-1} - z_t)$, $(z_{t-2} - z_t)$, …, $(z_{t-12} - z_t)$ and obtain the coefficient and standard error on z_t as the estimated LRP and its standard error.

(v) We would add lags 13 through 18 of $gwage_t$ to the equation, which leaves 273 – 6 = 267 observations. Now, we are estimating 20 parameters, so the *df* in the unrestricted model is df_{ur} = 267. Let R_{ur}^2 be the *R*-squared from this regression. To obtain the restricted *R*-squared, R_r^2, we need to reestimate the model reported in the problem but with the same 267 observations used to estimate the unrestricted model. Then $F = [(R_{ur}^2 - R_r^2)/(1 - R_{ur}^2)](247/6)$. We would find the critical value from the $F_{6,247}$ distribution.

11.7 (i) We plug the first equation into the second to get

$$y_t - y_{t-1} = \lambda(\gamma_0 + \gamma_1 x_t + e_t - y_{t-1}) + a_t,$$

and, rearranging,

$$\begin{aligned} y_t &= \lambda\gamma_0 + (1 - \lambda)y_{t-1} + \lambda\gamma_1 x_t + a_t + \lambda e_t, \\ &\equiv \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t, \end{aligned}$$

where $\beta_0 \equiv \lambda\gamma_0$, $\beta_1 \equiv (1 - \lambda)$, $\beta_2 \equiv \lambda\gamma_1$, and $u_t \equiv a_t + \lambda e_t$.

(ii) An OLS regression of y_t on y_{t-1} and x_t produces consistent, asymptotically normal estimators of the β_j. Under $\mathrm{E}(e_t|x_t,y_{t-1},x_{t-1}, \dots) = \mathrm{E}(a_t|x_t,y_{t-1},x_{t-1}, \dots) = 0$ it follows that $\mathrm{E}(u_t|x_t,y_{t-1},x_{t-1}, \dots) = 0$, which means that the model is dynamically complete [see equation (11.37)]. Therefore, the errors are serially uncorrelated. If the homoskedasticity assumption $\mathrm{Var}(u_t|x_t,y_{t-1}) = \sigma^2$ holds, then the usual standard errors, *t* statistics and *F* statistics are asymptotically valid.

(iii) Because $\beta_1 = (1 - \lambda)$, if $\hat{\beta}_1 = .7$ then $\hat{\lambda} = .3$. Further, $\hat{\beta}_2 = \hat{\lambda}\hat{\gamma}_1$, or $\hat{\gamma}_1 = \hat{\beta}_2/\hat{\lambda} = .2/.3 \approx .67$.

SOLUTIONS TO COMPUTER EXERCISES

C11.1 (i) The first order autocorrelation for log(*invpc*) is about .639. If we first detrend log(*invpc*) by regressing on a linear time trend, $\hat{\rho}_1 \approx .485$. Especially after detrending there is little evidence of a unit root in log(*invpc*). For log(*price*), the first order autocorrelation is about .949, which is very high. After detrending, the first order autocorrelation drops to .822, but this is still pretty large. We cannot confidently rule out a unit root in log(*price*).

(ii) The estimated equation is

$$\widehat{\log(invpc_t)} = -.853 + 3.88\,\Delta\log(price_t) + .0080\,t$$
$$(.040) \quad (0.96) \qquad\qquad (.0016)$$
$$n = 41, \quad R^2 = .501.$$

The coefficient on $\Delta\log(price_t)$ implies that a one percentage point increase in the growth in price leads to a 3.88 percent increase in housing investment above its trend. [If $\Delta\log(price_t) = .01$ then $\Delta\widehat{\log(invpc_t)} = .0388$; we multiply both by 100 to convert the proportionate changes to percentage changes.]

(iii) If we first linearly detrend $\log(invpc_t)$ before regressing it on $\Delta\log(price_t)$ and the time trend, then $R^2 = .303$, which is substantially lower than that when we do not detrend. Thus, $\Delta\log(price_t)$ explains only about 30% of the variation in $\log(invpc_t)$ about its trend.

(iv) The estimated equation is

$$\Delta\widehat{\log(invpc_t)} = .006 + 1.57\,\Delta\log(price_t) + .00004t$$
$$(.048) \quad (1.14) \qquad\qquad (.00190)$$
$$n = 41, \quad R^2 = .048.$$

The coefficient on $\Delta\log(price_t)$ has fallen substantially and is no longer significant at the 5% level against a positive one-sided alternative. The R-squared is much smaller; $\Delta\log(price_t)$ explains very little variation in $\Delta\log(invpc_t)$. Because differencing eliminates linear time trends, it is not surprising that the estimate on the trend is very small and very statistically insignificant.

C11.3 (i) The estimated equation is

$$\widehat{return}_t = .226 + .049\,return_{t-1} - .0097\,return_{t-1}^2$$
$$(.087) \qquad (.039) \qquad\qquad (.0070)$$
$$n = 689, \quad R^2 = .0063.$$

(ii) The null hypothesis is H_0: $\beta_1 = \beta_2 = 0$. Only if both parameters are zero does $E(return_t|return_{t-1})$ not depend on $return_{t-1}$. The F statistic is about 2.16 with p-value $\approx$ .116. Therefore, we cannot reject H_0 at the 10% level.

(iii) When we put $return_{t-1} \cdot return_{t-2}$ in place of $return_{t-1}^2$ the null can still be stated as in part (ii): no past values of $return$, or any functions of them, should help us predict $return_t$. The R-squared is about .0052 and $F \approx 1.80$ with p-value $\approx$ .166. Here, we do not reject H_0 at even the 15% level.

(iv) Predicting $return_t$ based on past returns does not appear promising. Even though the F statistic from part (ii) is almost significant at the 10% level, we have many observations. We cannot even explain 1% of the variation in $return_t$.

C11.5 (i) The estimated equation is

$$\Delta\widehat{gfr} = -1.27 - .035\,\Delta pe - .013\,\Delta pe_{-1} - .111\,\Delta pe_{-2} + .0079\,t$$
$$(1.05)\quad (.027)\quad (.028)\quad (.027)\quad (.0242)$$

$$n = 69,\quad R^2 = .234,\quad \bar{R}^2 = .186.$$

The time trend coefficient is very insignificant, so it is not needed in the equation.

(iii) The estimated equation is

$$\Delta\widehat{gfr} = -.650 - .075\,\Delta pe - .051\,\Delta pe_{-1} + .088\,\Delta pe_{-2} + 4.84\,ww2 - 1.68\,pill$$
$$(.582)\quad (.032)\quad (.033)\quad (.028)\quad (2.83)\quad (1.00)$$

$$n = 69,\quad R^2 = .296,\quad \bar{R}^2 = .240.$$

The F statistic for joint significance is $F = 2.82$ with p-value $\approx .067$. So $ww2$ and $pill$ are not jointly significant at the 5% level, but they are at the 10% level.

(iii) By regressing Δgfr on Δpe, $(\Delta pe_{-1} - \Delta pe)$. $(\Delta pe_{-2} - \Delta pe)$, $ww2$, and $pill$, we obtain the LRP and its standard error as the coefficient on Δpe: $-.075$, se = .032. So the estimated LRP is now negative and significant, which is very different from the equation in levels, (10.19) (the estimated LRP was .101 with a t statistic of about 3.37). This is a good example of how differencing variables before including them in a regression can lead to very different conclusions than a regression in levels.

C11.7 (i) If $\mathrm{E}(gc_t|I_{t-1}) = \mathrm{E}(gc_t)$ – that is, $\mathrm{E}(gc_t|I_{t-1})$ = does not depend on gc_{t-1}, then $\beta_1 = 0$ in $gc_t = \beta_0 + \beta_1 gc_{t-1} + u_t$. So the null hypothesis is H_0: $\beta_1 = 0$ and the alternative is H_1: $\beta_1 \neq 0$. Estimating the simple regression using the data in CONSUMP.RAW gives

$$\widehat{gc}_t = .011 + .446\,gc_{t-1}$$
$$(.004)\quad (.156)$$

$$n = 35,\quad R^2 = .199.$$

The t statistic for $\hat{\beta}_1$ is about 2.86, and so we strongly reject the PIH. The coefficient on gc_{t-1} is also practically large, showing significant autocorrelation in consumption growth.

(ii) When gy_{t-1} and $i3_{t-1}$ are added to the regression, the R-squared becomes about .288. The F statistic for joint significance of gy_{t-1} and $i3_{t-1}$, obtained using the Stata "test" command, is

1.95, with p-value $\approx$ .16. Therefore, gy_{t-1} and $i3_{t-1}$ are not jointly significant at even the 15% level.

C11.9 (i) The first order autocorrelation for *prcfat* is .709, which is high but not necessarily a cause for concern. For *unem*, $\hat{\rho}_1 = .950$, which is cause for concern in using *unem* as an explanatory variable in a regression.

(ii) If we use the first differences of *prcfat* and *unem*, but leave all other variables in their original form, we get the following:

$$\begin{aligned}\Delta\widehat{prcfat} = \; & \underset{(.105)}{-.127} + \ldots + \underset{(.0072)}{.0068}\, wkends + \underset{(.0161)}{.0125}\, \Delta unem \\ & - \underset{(.0238)}{.0072}\, spdlaw + \underset{(.0265)}{.0008}\, bltlaw\end{aligned}$$

$$n = 107,\; R^2 = .344,$$

where I have again suppressed the coefficients on the time trend and seasonal dummies. This regression basically shows that the change in *prcfat* cannot be explained by the change in *unem* or any of the policy variables. It does have some seasonality, which is why the R-squared is .344.

(iii) This is an example about how estimation in first differences loses the interesting implications of the model estimated in levels. Of course, this is not to say the levels regression is valid. But, as it turns out, we can reject a unit root in *prcfat*, and so we can at least justify using it in level form; see Computer Exercise 18.13. Generally, the issue of whether to take first differences is very difficult, even for professional time series econometricians.

C11.11 (i) The estimated equation is

$$\widehat{pcrgdp}_t = \underset{(0.163)}{3.344} - \underset{(0.182)}{1.891}\; \Delta unem_t$$

$$n = 46,\; R^2 = .710$$

Naturally, we do not get the exact estimates specified by the theory. Okun's Law is expected to hold, at best, on average. The estimates are not particularly far from their hypothesized values of 3 (intercept) and −2 (slope).

(ii) The t statistic for testing $\mathrm{H}_0 : \beta_1 = -2$ is about .60, which gives a two-sided p-value of about .55. This is very little evidence against H_0; the null is not rejected at any reasonable significance level.

(iii) The *t* statistic for $\mathrm{H}_0: \beta_0 = 3$ is about 2.11, and the two-sided *p*-value is about .04. Therefore, the null is rejected at the 5% level, although it is not much stronger than that.

(iv) The joint test underlying Okun's Law gives $F = 2.41$. With (2,44) *df*, we get, roughly, *p*-value = .10. Therefore, Okun's Law passes at the 5% level, but only just at the 10% level.

CHAPTER 12

SOLUTIONS TO PROBLEMS

12.1 We can reason this from equation (12.4) because the usual OLS standard error is an estimate of $\sigma/\sqrt{SST_x}$. When the dependent and independent variables are in level (or log) form, the AR(1) parameter, ρ, tends to be positive in time series regression models. Further, the independent variables tend to be positive correlated, so $(x_t - \bar{x})(x_{t+j} - \bar{x})$ – which is what generally appears in (12.4) when the $\{x_t\}$ do not have zero sample average – tends to be positive for most t and j. With multiple explanatory variables the formulas are more complicated but have similar features.

If $\rho < 0$, or if the $\{x_t\}$ is negatively autocorrelated, the second term in the last line of (12.4) could be negative, in which case the true standard deviation of $\hat{\beta}_1$ is actually less than $\sigma/\sqrt{SST_x}$.

12.3 (i) Because U.S. presidential elections occur only every four years, it seems reasonable to think the unobserved shocks – that is, elements in u_t – in one election have pretty much dissipated four years later. This would imply that $\{u_t\}$ is roughly serially uncorrelated.

(ii) The t statistic for H_0: $\rho = 0$ is $-.068/.240 \approx -.28$, which is very small. Further, the estimate $\hat{\rho} = -.068$ is small in a practical sense, too. There is no reason to worry about serial correlation in this example.

(iii) Because the test based on $t_{\hat{\rho}}$ is only justified asymptotically, we would generally be concerned about using the usual critical values with $n = 20$ in the original regression. But any kind of adjustment, either to obtain valid standard errors for OLS as in Section 12.5 or a feasible GLS procedure as in Section 12.3, relies on large sample sizes, too. (Remember, FGLS is not even unbiased, whereas OLS is under TS.1 through TS.3.) Most importantly, the estimate of ρ is *practically* small, too. With $\hat{\rho}$ so close to zero, FGLS or adjusting the standard errors would yield similar results to OLS with the usual standard errors.

12.5 (i) There is substantial serial correlation in the errors of the equation, and the OLS standard errors almost certainly underestimate the true standard deviation in $\hat{\beta}_{EZ}$. This makes the usual confidence interval for β_{EZ} and t statistics invalid.

(ii) We can use the method in Section 12.5 to obtain an approximately valid standard error. [See equation (12.43).] While we might use $g = 2$ in equation (12.42), with monthly data we might want to try a somewhat longer lag, maybe even up to $g = 12$.

SOLUTIONS TO COMPUTER EXERCISES

C12.1 Regressing $\hat{u}_t$ on $\hat{u}_{t-1}$, using the 69 available observations, gives $\hat{\rho} \approx .292$ and se($\hat{\rho}$) $\approx$ .118. The t statistic is about 2.47, and so there is significant evidence of positive AR(1) serial correlation in the errors (even though the variables have been differenced). This means we should view the standard errors reported in equation (11.27) with some suspicion.

C12.3 (i) The test for AR(1) serial correlation gives (with 35 observations) $\hat{\rho} \approx -.110$, se($\hat{\rho}$) $\approx$ .175. The t statistic is well below one in absolute value, so there is no evidence of serial correlation in the accelerator model. If we view the test of serial correlation as a test of dynamic misspecification, it reveals no dynamic misspecification in the accelerator model.

(ii) It is worth emphasizing that, if there is little evidence of AR(1) serial correlation, there is no need to use feasible GLS (Cochrane-Orcutt or Prais-Winsten).

C12.5 (i) Using the data only through 1992 gives

$$\begin{aligned}\widehat{demwins} = \underset{(.107)}{.441} - \underset{(.354)}{.473}\, partyWH + \underset{(.205)}{.479}\, incum + \underset{(.036)}{.059}\, partyWH\cdot gnews \\ - \underset{(.028)}{.024}\, partyWH\cdot inf\end{aligned}$$

$$n = 20, \quad R^2 = .437, \quad \bar{R}^2 = .287.$$

The largest t statistic is on *incum*, which is estimated to have a large effect on the probability of winning. But we must be careful here. *incum* is equal to 1 if a Democratic incumbent is running and –1 if a Republican incumbent is running. Similarly, *partyWH* is equal to 1 if a Democrat is currently in the White House and –1 if a Republican is currently in the White House. So, for an incumbent Democrat running, we must add the coefficients on *partyWH* and *incum* together, and this nets out to about zero.

The economic variables are less statistically significant than in equation (10.23). The *gnews* interaction has a t statistic of about 1.64, which is significant at the 10% level against a one-sided alternative. (Since the dependent variable is binary, this is a case where we must appeal to asymptotics. Unfortunately, we have only 20 observations.) The inflation variable has the expected sign but is not statistically significant.

(ii) There are two fitted values less than zero, and two fitted values greater than one.

(iii) Out of the 10 elections with *demwins* = 1, 8 of these are correctly predicted. Out of the 10 elections with *demwins* = 0, 7 are correctly predicted. So 15 out of 20 elections through 1992 are correctly predicted. (But, remember, we used data from these years to obtain the estimated equation.)

(iv) The explanatory variables are *partyWH* = 1, *incum* = 1, *gnews* = 3, and *inf* = 3.019. Therefore, for 1996,

$$\widehat{demwins} = .441 - .473 + .479 + .059(3) - .024(3.019) \approx .552.$$

Because this is above .5, we would have predicted that Clinton would win the 1996 election, as he did.

(v) The regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ produces $\hat{\rho} \approx$ -.164 with heteroskedasticity-robust standard error of about .195. (Because the LPM contains heteroskedasticity, testing for AR(1) serial correlation in an LPM generally requires a heteroskedasticity-robust test.) Therefore, there is little evidence of serial correlation in the errors. (And, if anything, it is negative.)

(vi) The heteroskedasticity-robust standard errors are given in [·] below the usual standard errors:

$$\begin{aligned}\widehat{demwins} = \underset{\substack{(.107)\\ [.086]}}{.441} - \underset{\substack{(.354)\\ [.301]}}{.473}\, partyWH + \underset{\substack{(.205)\\ [.185]}}{.479}\, incum + \underset{\substack{(.036)\\ [.030]}}{.059}\, partyWH \cdot gnews \\ - \underset{\substack{(.028)\\ [.019]}}{.024}\, partyWH \cdot inf\end{aligned}$$

$$n = 20, \quad R^2 = .437, \quad \bar{R}^2 = .287.$$

In fact, all heteroskedasticity-robust standard errors are less than the usual OLS standard errors, making each variable more significant. For example, the *t* statistic on *partyWH·gnews* becomes about 1.97, which is notably above 1.64. But we must remember that the standard errors in the LPM have only asymptotic justification. With only 20 observations it is not clear we should prefer the heteroskedasticity-robust standard errors to the usual ones.

C12.7 (i) The iterated Prais-Winsten estimates are given below. The estimate of ρ is, to three decimal places, .293, which is the same as the estimate used in the final iteration of Cochrane-Orcutt:

$$\begin{aligned}\widehat{\log(chnimp)} = \underset{(22.78)}{-37.08} + \underset{(.63)}{2.94}\log(chempi) + \underset{(.98)}{1.05}\log(gas) + \underset{(.51)}{1.13}\log(rtwex) \\ - \underset{(.319)}{.016}\, befile6 - \underset{(.322)}{.033}\, affile6 - \underset{(.342)}{.577}\, afdec6\end{aligned}$$

$$n = 131, \quad R^2 = .202$$

(ii) Not surprisingly, the C-O and P-W estimates are quite similar. To three decimal places, they use the same value of $\hat{\rho}$ (to four decimal places it is .2934 for C-O and .2932 for P-W). The only practical difference is that P-W uses the equation for t = 1. With n = 131, we hope this makes little difference.

C12.9 (i) Here are the OLS regression results:

$$\widehat{\log(avgprc)} = \underset{(.115)}{-.073} - \underset{(.0014)}{.0040}\, t - \underset{(.1294)}{.0101}\, mon - \underset{(.1273)}{.0088}\, tues + \underset{(.1257)}{.0376}\, wed + \underset{(.1257)}{.0906}\, thurs$$

$n = 97,\ R^2 = .086$

The test for joint significance of the day-of-the-week dummies is F = .23, which gives p-value = .92. So there is no evidence that the average price of fish varies systematically within a week.

(ii) The equation is

$$\widehat{\log(avgprc)} = \underset{(.190)}{-.920} - \underset{(.0014)}{.0012}\, t - \underset{(.1141)}{.0182}\, mon - \underset{(.1121)}{.0085}\, tues + \underset{(.1117)}{.0500}\, wed + \underset{(.1110)}{.1225}\, thurs$$

$$+ \underset{(.0218)}{.0909}\, wave2 + \underset{(.0208)}{.0474}\, wave3$$

$n = 97,\ R^2 = .310$

Each of the wave variables is statistically significant, with *wave2* being the most important. Rough seas (as measured by high waves) would reduce the supply of fish (shift the supply curve back), and this would result in a price increase. One might argue that bad weather reduces the demand for fish at a market, too, but that would reduce price. If there are demand effects captured by the wave variables, they are being swamped by the supply effects.

(iii) The time trend coefficient becomes much smaller and statistically insignificant. We can use the omitted variable bias table from Chapter 3, Table 3.2 to determine what is probably going on. Without *wave2* and *wave3*, the coefficient on t seems to have a downward bias. Since we know the coefficients on *wave2* and *wave3* are positive, this means the wave variables are negatively correlated with t. In other words, the seas were rougher, on average, at the beginning of the sample period. (You can confirm this by regressing *wave2* on t and *wave3* on t.)

(iv) The time trend and daily dummies are clearly strictly exogenous, as they are just functions of time and the calendar. Further, the height of the waves is not influenced by past unexpected changes in log(*avgprc*).

(v) We simply regress the OLS residuals on one lag, getting $\hat{\rho} = .618, \mathrm{se}(\hat{\rho}) = .081, t_{\hat{\rho}} = 7.63.$ Therefore, there is strong evidence of positive serial correlation.

(vi) The Newey-West standard errors are $\text{se}(\hat{\beta}_{wave2}) = .0234$ and $\text{se}(\hat{\beta}_{wave3}) = .0195$. Given the significant amount of AR(1) serial correlation in part (v), it is somewhat surprising that these standard errors are not much larger compared with the usual, incorrect standard errors. In fact, the Newey-West standard error for $\hat{\beta}_{wave3}$ is actually smaller than the OLS standard error.

(vii) The Prais-Winsten estimates are

$$\begin{aligned}\widehat{\log(avgprc)} = \underset{(.239)}{-.658} &- \underset{(.0029)}{.0007}\, t + \underset{(.0652)}{.0099}\, mon + \underset{(.0744)}{.0025}\, tues + \underset{(.0746)}{.0624}\, wed + \underset{(.0621)}{.1174}\, thurs \\ &+ \underset{(.0174)}{.0497}\, wave2 + \underset{(.0174)}{.0323}\, wave3\end{aligned}$$

$n = 97,\ R^2 = .135$

The coefficient on *wave2* drops by a nontrivial amount, but it still has a *t* statistic of almost 3. The coefficient on *wave3* drops by a relatively smaller amount, but its *t* statistic (1.86) is borderline significant. The final estimate of ρ is about .687.

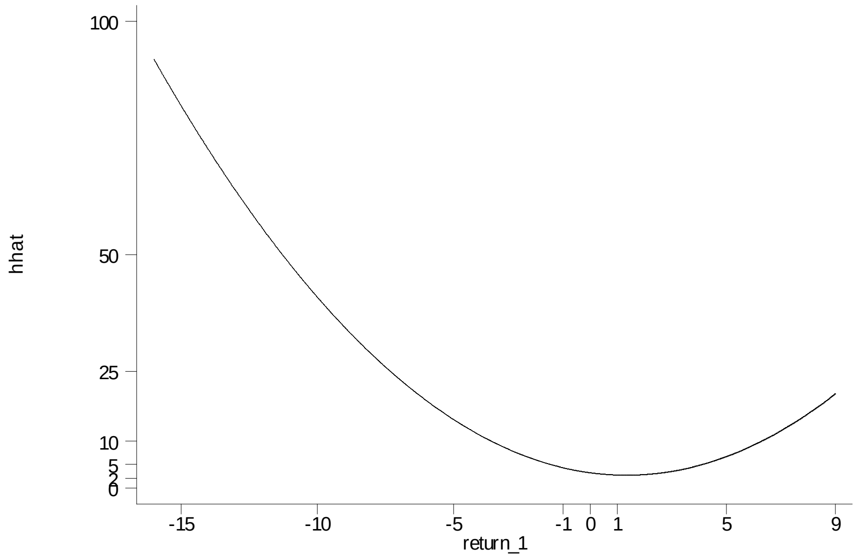
C12.11 (i) The average of $\hat{u}_i^2$ over the sample is 4.44, with the smallest value being .0000074 and the largest being 232.89.

(ii) This is the same as C12.4, part (ii):

$$\hat{u}_i^2 = \underset{(0.44)}{3.26} - \underset{(.196)}{.789}\, return_{t\text{-}1} + \underset{(.036)}{.297}\, return_{t-1}^2 + residual_t$$

$$n = 689,\ R^2 = .130.$$

(iii) The graph of the estimated variance function is



The variance is smallest when *return*$_{-1}$ is about 1.33, and the variance is then about 2.74.

(iv) No. The graph in part (iii) makes this clear, as does finding that the smallest variance estimate is 2.74.

(v) The *R*-squared for the ARCH(1) model is .114, compared with .130 for the quadratic in *return*$_{-1}$. We should really compare adjusted *R*-squareds, because the ARCH(1) model contains only two total parameters. For the ARCH(1) model, $\bar{R}^2$ is about .112; for the model in part (ii), $\bar{R}^2$ = .128. Therefore, after adjusting for the different *df*, the quadratic in *return*$_{-1}$ fits better than the ARCH(1) model.

(vi) The coefficient on $\hat{u}_{t-2}^2$ is only .042, and its *t* statistic is barely above one (t = 1.09). Therefore, an ARCH(2) model does not seem warranted. The adjusted *R*-squared is about .113, so the ARCH(2) fits worse than the model estimated in part (ii).

C12.13 (i) The regression $\hat{u}_t$ on $\hat{u}_{t-1}$, $\Delta unem_t$ gives a coefficient on $\hat{u}_{t-1}$ of .073 with t = .42. Therefore, there is very little evidence of first-order serial correlation.

(ii) The simple regression $\hat{u}_t^2$ on $\Delta unem_t$ gives a slope coefficient of about .452 with t = 2.07, and so, at the 5% significance level, we find that there is heteroskedasticity. The variance of the error appears to be larger when the change in unemployment is larger.

(iii) The heteroskedasticity-robust standard error is about .223, compared with the usual OLS standard error of .182. So, the robust standard error is more than 20% larger than the usual OLS one. Of course, a larger standard error leads to a wider confidence interval for β_1.

CHAPTER 13

SOLUTIONS TO PROBLEMS

13.1 Without changes in the averages of *any* explanatory variables, the average fertility rate fell by .545 between 1972 and 1984; this is simply the coefficient on *y84*. To account for the increase in average education levels, we obtain an additional effect: $-.128(13.3 - 12.2) \approx -.141$. So the drop in average fertility if the average education level increased by 1.1 is .545 + .141 = .686, or roughly two-thirds of a child per woman.

13.3 We do not have repeated observations on the *same* cross-sectional units in each time period, and so it makes no sense to look for pairs to difference. For example, in Example 13.1, it is very unlikely that the same woman appears in more than one year, as new random samples are obtained in each year. In Example 13.3, some houses may appear in the sample for both 1978 and 1981, but the overlap is usually too small to do a true panel data analysis.

13.5 No, we cannot include age as an explanatory variable in the original model. Each person in the panel data set is exactly two years older on January 31, 1992 than on January 31, 1990. This means that $\Delta age_i = 2$ for all i. But the equation we would estimate is of the form

$$\Delta saving_i = \delta_0 + \beta_1 \Delta age_i + \ldots,$$

where δ_0 is the coefficient the year dummy for 1992 in the original model. As we know, when we have an intercept in the model we cannot include an explanatory variable that is constant across i; this violates Assumption MLR.3. Intuitively, since age changes by the same amount for everyone, we cannot distinguish the effect of age from the aggregate time effect.

13.7 (i) It is not surprising that the coefficient on the interaction term changes little when *afchnge* is dropped from the equation because the coefficient on *afchnge* in (3.12) is only .0077 (and its t statistic is very small). The increase from .191 to .198 is easily explained by sampling error.

(ii) If *highearn* is dropped from the equation [so that $\beta_1 = 0$ in (3.10)], then we are assuming that, prior to the change in policy, there is no difference in average duration between high earners and low earners. But the very large (.256), highly statistically significant estimate on *highearn* in (3.12) shows this presumption to be false. Prior to the policy change, the high earning group spent about 29.2% [$\exp(.256) - 1 \approx .292$] longer on unemployment compensation than the low earning group. By dropping *highearn* from the regression, we attribute to the policy change the difference between the two groups that would be observed without any intervention.

SOLUTIONS TO COMPUTER EXERCISES

C13.1 (i) The F statistic (with 4 and 1,111 *df*) is about 1.16 and p-value $\approx$ .328, which shows that the living environment variables are jointly insignificant.

(ii) The F statistic (with 3 and 1,111 *df*) is about 3.01 and p-value $\approx$ .029, and so the region dummy variables are jointly significant at the 5% level.

(iii) After obtaining the OLS residuals, $\hat{u}$, from estimating the model in Table 13.1, we run the regression $\hat{u}^2$ on $y74, y76, \ldots, y84$ using all 1,129 observations. The null hypothesis of homoskedasticity is H_0: $\gamma_1 = 0, \gamma_2 = 0, \ldots, \gamma_6 = 0$. So we just use the usual F statistic for joint significance of the year dummies. The R-squared is about .0153 and $F \approx 2.90$; with 6 and 1,122 *df*, the p-value is about .0082. So there is evidence of heteroskedasticity that is a function of time at the 1% significance level. This suggests that, at a minimum, we should compute heteroskedasticity-robust standard errors, t statistics, and F statistics. We could also use weighted least squares (although the form of heteroskedasticity used here may not be sufficient; it does not depend on *educ*, *age*, and so on).

(iv) Adding $y74 \cdot educ, \ldots, y84 \cdot educ$ allows the relationship between fertility and education to be different in each year; remember, the coefficient on the interaction gets added to the coefficient on *educ* to get the slope for the appropriate year. When these interaction terms are added to the equation, $R^2 \approx .137$. The F statistic for joint significance (with 6 and 1,105 *df*) is about 1.48 with p-value $\approx$ .18. Thus, the interactions are not jointly significant at even the 10% level. This is a bit misleading, however. An abbreviated equation (which just shows the coefficients on the terms involving *educ*) is

$$\begin{aligned}\widehat{kids} = \underset{(3.13)}{-8.48} - \underset{(.054)}{.023}\, educ + \ldots - \underset{(.073)}{.056}\, y74 \cdot educ - \underset{(.071)}{.092}\, y76 \cdot educ \\ - \underset{(.075)}{.152}\, y78 \cdot educ - \underset{(.070)}{.098}\, y80 \cdot educ - \underset{(.068)}{.139}\, y82 \cdot educ - \underset{(.070)}{.176}\, y84 \cdot educ.\end{aligned}$$

Three of the interaction terms, $y78 \cdot educ$, $y82 \cdot educ$, and $y84 \cdot educ$ are statistically significant at the 5% level against a two-sided alternative, with the p-value on the latter being about .012. The coefficients are large in magnitude as well. The coefficient on *educ* – which is for the base year, 1972 – is small and insignificant, suggesting little if any relationship between fertility and education in the early seventies. The estimates above are consistent with fertility becoming more linked to education as the years pass. The F statistic is insignificant because we are testing some insignificant coefficients along with some significant ones.

C13.3 (i) Other things equal, homes farther from the incinerator should be worth more, so $\delta_1 > 0$. If $\beta_1 > 0$, then the incinerator was located farther away from more expensive homes.

(ii) The estimated equation is

$$\widehat{\log(price)} = \underset{(0.51)}{8.06} - \underset{(.805)}{.011}\, y81 + \underset{(.052)}{.317} \log(dist) + \underset{(.082)}{.048}\, y81 \cdot \log(dist)$$

$$n = 321, \quad R^2 = .396, \quad \bar{R}^2 = .390.$$

While $\hat{\delta}_1 = .048$ is the expected sign, it is not statistically significant (t statistic $\approx .59$).

(iii) When we add the list of housing characteristics to the regression, the coefficient on $y81 \cdot \log(dist)$ becomes .062 (se = .050). So the estimated effect is larger – the elasticity of *price* with respect to *dist* is .062 after the incinerator site was chosen – but its t statistic is only 1.24. The p-value for the one-sided alternative H_1: $\delta_1 > 0$ is about .108, which is close to being significant at the 10% level.

C13.5 (i) Using pooled OLS we obtain

$$\widehat{\log(rent)} = \underset{(.535)}{-.569} + \underset{(.035)}{.262}\, d90 + \underset{(.023)}{.041} \log(pop) + \underset{(.053)}{.571} \log(avginc) + \underset{(.0010)}{.0050}\, pctstu$$

$$n = 128, \quad R^2 = .861.$$

The positive and very significant coefficient on *d90* simply means that, other things in the equation fixed, nominal rents grew by over 26% over the 10 year period. The coefficient on *pctstu* means that a one percentage point increase in *pctstu* increases *rent* by half a percent (.5%). The t statistic of five shows that, at least based on the usual analysis, *pctstu* is very statistically significant.

(ii) The standard errors from part (i) are not valid, unless we thing a_i does not really appear in the equation. If a_i is in the error term, the errors across the two time periods for each city are positively correlated, and this invalidates the usual OLS standard errors and t statistics.

(iii) The equation estimated in differences is

$$\Delta\widehat{\log(rent)} = \underset{(.037)}{.386} + \underset{(.088)}{.072}\, \Delta\log(pop) + \underset{(.066)}{.310} \log(avginc) + \underset{(.0041)}{.0112}\, \Delta pctstu$$

$$n = 64, \quad R^2 = .322.$$

Interestingly, the effect of *pctstu* is over twice as large as we estimated in the pooled OLS equation. Now, a one percentage point increase in *pctstu* is estimated to increase rental rates by about 1.1%. Not surprisingly, we obtain a much less precise estimate when we difference (although the OLS standard errors from part (i) are likely to be much too small because of the positive serial correlation in the errors within each city). While we have differenced away a_i, there may be other unobservables that change over time and are correlated with $\Delta pctstu$.

(iv) The heteroskedasticity-robust standard error on $\Delta pctstu$ is about .0028, which is actually much smaller than the usual OLS standard error. This only makes *pctstu* even more significant (robust *t* statistic $\approx$ 4). Note that serial correlation is no longer an issue because we have no time component in the first-differenced equation.

C13.7 (i) Pooling across semesters and using OLS gives

$$\begin{aligned}\widehat{trmgpa} = &-1.75 - .058\, spring + .00170\, sat - .0087\, hsperc \\ &\;\;(0.35)\;\;(.048) \qquad\quad (.00015) \qquad (.0010) \\ &+ .350\, female - .254\, black - .023\, white - .035\, frstsem \\ &\;\;(.052) \qquad\quad (.123) \qquad (.117) \qquad (.076) \\ &- .00034\, tothrs + 1.048\, crsgpa - .027\, season \\ &\;\;(.00073) \qquad\; (0.104) \qquad (.049)\end{aligned}$$

$$n = 732, \quad R^2 = .478, \quad \bar{R}^2 = .470.$$

The coefficient on *season* implies that, other things fixed, an athlete's term GPA is about .027 points lower when his/her sport is in season. On a four point scale, this a modest effect (although it accumulates over four years of athletic eligibility). However, the estimate is not statistically significant (*t* statistic $\approx -.55$).

(ii) The quick answer is that if omitted ability is correlated with *season* then, as we know form Chapters 3 and 5, OLS is biased and inconsistent. The fact that we are pooling across two semesters does not change that basic point.

If we think harder, the direction of the bias is not clear, and this is where pooling across semesters plays a role. First, suppose we used only the fall term, when football is in season. Then the error term and season would be negatively correlated, which produces a downward bias in the OLS estimator of β_{season}. Because β_{season} is hypothesized to be negative, an OLS regression using only the fall data produces a downward biased estimator. [When just the fall data are used, $\hat{\beta}_{season} = -.116$ (se = .084), which is in the direction of more bias.] However, if we use just the spring semester, the bias is in the opposite direction because ability and season would be positive correlated (more academically able athletes are in season in the spring). In fact, using just the spring semester gives $\hat{\beta}_{season} = .00089$ (se = .06480), which is practically and statistically equal to zero. When we pool the two semesters we cannot, with a much more detailed analysis, determine which bias will dominate.

(iii) The variables *sat*, *hsperc*, *female*, *black*, and *white* all drop out because they do not vary by semester. The intercept in the first-differenced equation is the intercept for the spring. We have

$$\widehat{\Delta trmgpa} = -.237 + .019\,\Delta frstsem + .012\,\Delta tothrs + 1.136\,\Delta crsgpa - .065\,season$$
$$\quad (.206) \quad (.069) \qquad (.014) \qquad (0.119) \qquad (.043)$$

$$n = 366, \quad R^2 = .208, \quad \bar{R}^2 = .199.$$

Interestingly, the in-season effect is larger now: term GPA is estimated to be about .065 points lower in a semester that the sport is in-season. The t statistic is about –1.51, which gives a one-sided p-value of about .065.

(iv) One possibility is a measure of course load. If some fraction of student-athletes take a lighter load during the season (for those sports that have a true season), then term GPAs may tend to be higher, other things equal. This would bias the results away from finding an effect of *season* on term GPA.

C13.9 (i) When we add the changes of the nine log wage variables to equation (13.33) we obtain

$$\begin{aligned}
\widehat{\Delta\log(crmrte)} = {} & \underset{(.021)}{.020} - \underset{(.027)}{.111}\, d83 - \underset{(.025)}{.037}\, d84 - \underset{(.0241)}{.0006}\, d85 + \underset{(.025)}{.031}\, d86 + \underset{(.025)}{.039}\, d87 \\
& - \underset{(.030)}{.323}\, \Delta\log(prbarr) - \underset{(.018)}{.240}\, \Delta\log(prbconv) - \underset{(.026)}{.169}\, \Delta\log(prbpris) \\
& - \underset{(.022)}{.016}\, \Delta\log(avgsen) + \underset{(.027)}{.398}\, \Delta\log(polpc) - \underset{(.030)}{.044}\, \Delta\log(wcon) \\
& + \underset{(0.14)}{.025}\, \Delta\log(wtuc) - \underset{(.031)}{.029}\, \Delta\log(wtrd) + \underset{(.0212)}{.0091}\, \Delta\log(wfir) \\
& + \underset{(.014)}{.022}\, \Delta\log(wser) - \underset{(.102)}{.140}\, \Delta\log(wmfg) - \underset{(.172)}{.017}\, \Delta\log(wfed) \\
& - \underset{(.096)}{.052}\, \Delta\log(wsta) - \underset{(.102)}{.031}\, \Delta\log(wloc)
\end{aligned}$$

$$n = 540, \quad R^2 = .445, \quad \bar{R}^2 = .424.$$

The coefficients on the criminal justice variables change very modestly, and the statistical significance of each variable is also essentially unaffected.

(ii) Since some signs are positive and others are negative, they cannot all really have the expected sign. For example, why is the coefficient on the wage for transportation, utilities, and communications (*wtuc*) positive and marginally significant (t statistic $\approx$ 1.79)? Higher manufacturing wages lead to lower crime, as we might expect, but, while the estimated coefficient is by far the largest in magnitude, it is not statistically different from zero (t statistic $\approx$ –1.37). The F test for joint significance of the wage variables, with 9 and 529 *df*, yields $F \approx 1.25$ and p-value $\approx .26$.

C13.11 (i) Take changes as usual, holding the other variables fixed: $\Delta math4_{it} = \beta_1 \Delta\log(rexpp_{it}) = (\beta_1/100)\cdot[\,100\cdot\Delta\log(rexpp_{it})] \approx (\beta_1/100)\cdot(\,\%\Delta rexpp_{it})$. So, if $\%\Delta rexpp_{it} = 10$, then $\Delta math4_{it} = (\beta_1/100)\cdot(10) = \beta_1/10$.

(ii) The equation, estimated by pooled OLS in first differences (except for the year dummies), is

$$\begin{aligned}\widehat{\Delta math4} = \underset{(.52)}{5.95} + \underset{(.73)}{.52}\, y94 + \underset{(.78)}{6.81}\, y95 - \underset{(.73)}{5.23}\, y96 - \underset{(.72)}{8.49}\, y97 + \underset{(.72)}{8.97}\, y98 \\ - \underset{(2.76)}{3.45}\, \Delta\log(rexpp) + \underset{(1.029)}{.635}\, \Delta\log(enroll) + \underset{(.055)}{.025}\, \Delta lunch\end{aligned}$$

$n = 3{,}300,\ \ R^2 = .208.$

Taken literally, the spending coefficient implies that a 10% increase in real spending per pupil decreases the *math4* pass rate by about $3.45/10 \approx .35$ percentage points.

(iii) When we add the lagged spending change, and drop another year, we get

$$\begin{aligned}\widehat{\Delta math4} = \underset{(.55)}{6.16} + \underset{(.77)}{5.70}\, y95 - \underset{(.79)}{6.80}\, y96 - \underset{(.74)}{8.99}\, y97 + \underset{(.74)}{8.45}\, y98 \\ - \underset{(3.04)}{1.41}\, \Delta\log(rexpp) + \underset{(2.79)}{11.04}\, \Delta\log(rexpp_{-1}) + \underset{(1.18)}{2.14}\, \Delta\log(enroll) \\ + \underset{(.061)}{.073}\, \Delta lunch\end{aligned}$$

$n = 2{,}750,\ \ R^2 = .238.$

The contemporaneous spending variable, while still having a negative coefficient, is not at all statistically significant. The coefficient on the lagged spending variable is very statistically significant, and implies that a 10% increase in spending last year increases the *math4* pass rate by about 1.1 percentage points. Given the timing of the tests, a lagged effect is not surprising. In Michigan, the fourth grade math test is given in January, and so if preparation for the test begins a full year in advance, spending when the students are in third grade would at least partly matter.

(iv) The heteroskedasticity-robust standard error for $\hat{\beta}_{\Delta\log(rexpp)}$ is about 4.28, which reduces the significance of $\Delta\log(rexpp)$ even further. The heteroskedasticity-robust standard error of

$\hat{\beta}_{\Delta\log(rexpp_{-1})}$ is about 4.38, which substantially lowers the t statistic. Still, $\Delta\log(rexpp_{-1})$ is statistically significant at just over the 1% significance level against a two-sided alternative.

(v) The fully robust standard error for $\hat{\beta}_{\Delta\log(rexpp)}$ is about 4.94, which even further reduces the t statistic for $\Delta\log(rexpp)$. The fully robust standard error for $\hat{\beta}_{\Delta\log(rexpp_{-1})}$ is about 5.13, which gives $\Delta\log(rexpp_{-1})$ a t statistic of about 2.15. The two-sided p-value is about .032.

(vi) We can use four years of data for this test. Doing a pooled OLS regression of $\hat{r}_{it}$ on $\hat{r}_{i,t-1}$, using years 1995, 1996, 1997, and 1998 gives $\hat{\rho} = -.423$ (se = .019), which is strong negative serial correlation.

(vii) The fully robust "F" test for $\Delta\log(enroll)$ and $\Delta lunch$, reported by Stata 7.0, is .93. With 2 and 549 *df*, this translates into p-value = .40. So we would be justified in dropping these variables, but they are not doing any harm.

C13.13 (i) We can estimate all parameters except β_0 and β_1: the intercept for the base year cannot be estimated, and neither can coefficients on the time-constant variable $educ_i$.

(ii) We want to test $H_0: \gamma_1 = \gamma_2, ..., \gamma_7 = 0$, so there are seven restrictions to be tested. Using FD (which eliminates $educ_i$) and obtaining the F statistic gives F = .31 (p-value = .952). Therefore, there is no evidence that the return to education varied over this time period. (Also, each coefficient is individuall statistically insignificant at the 25% level.)

(iii) The fully robust F statistic is about 1.00, with p-value = .432. So the conclusion really does not change: the γ_j are jointly insignificant.

(iv) The estimated union differential in 1980 is simply the coefficient on $\Delta union_{it}$, or about .106 (10.6%). For 1987, we add the coefficients on $\Delta union_t$ and $\Delta d87_t \cdot union_{it}$, or $-.041$ (–4.1%). The difference, –14.7%, is statistically significant (t = –2.15, whether we use the usual pooled OLS standard error or the fully robust one).

(v) The usual F statistic is 1.03 (p-value = .405) and the statistic robust to heteroskedasticity and serial correlation is 1.15 (p-value = .331). Therefore, when we test all interaction terms as a group (seven of them), we fail to reject the null that the union differential was constant over this period. Most of the interactions are individually insignificant; in fact, only those for 1986 and 1987 are close. We can get joint insignificance by lumping several statistically insignificant variables in with one or two statistically significant ones. But it is hard to ignore the practically large change from 1980 to 1987. (There might be a problem in this example with the strict exogeneity assumption: perhaps union membership next year depends on unexpected wage changes this year.)

CHAPTER 14

SOLUTIONS TO PROBLEMS

14.1 First, for each $t > 1$, Var(Δu_{it}) = Var($u_{it} - u_{i,t\text{-}1}$) = Var($u_{it}$) + Var($u_{i,t\text{-}1}$) = $2\sigma_u^2$, where we use the assumptions of no serial correlation in $\{u_t\}$ and constant variance. Next, we find the covariance between Δu_{it} and $\Delta u_{i,t+1}$. Because these each have a zero mean, the covariance is E($\Delta u_{it} \cdot \Delta u_{i,t+1}$) = E[($u_{it} - u_{i,t\text{-}1}$)($u_{i,t+1} - u_{it}$)] = E($u_{it}u_{i,t+1}$) – E($u_{it}^2$) – E($u_{i,t\text{-}1}u_{i,t+1}$) + E($u_{i,t\text{-}1}u_{it}$) = –E($u_{it}^2$) = $-\sigma_u^2$ because of the no serial correlation assumption. Because the variance is constant across t, by Problem 11.1, Corr(Δu_{it}, $\Delta u_{i,t+1}$) = Cov(Δu_{it}, $\Delta u_{i,t+1}$)/Var(Δu_{it}) = $-\sigma_u^2/(2\sigma_u^2)$ = –.5.

14.3 (i) E(e_{it}) = E($v_{it} - \lambda\bar{v}_i$) = E($v_{it}$) – λE($\bar{v}_i$) = 0 because E(v_{it}) = 0 for all t.

(ii) Var($v_{it} - \lambda\bar{v}_i$) = Var($v_{it}$) + λ^2Var($\bar{v}_i$) – $2\lambda\cdot$Cov(v_{it}, $\bar{v}_i$) = $\sigma_v^2 + \lambda^2$ E($\bar{v}_i^2$) – $2\lambda\cdot$E($v_{it}\bar{v}_i$). Now, $\sigma_v^2 = \text{E}(v_{it}^2) = \sigma_a^2 + \sigma_u^2$ and E($v_{it}\bar{v}_i$) = $T^{-1}\sum_{s=1}^{T} E(v_{it}v_{is})$ = $T^{-1}[\sigma_a^2 + \sigma_a^2 + \ldots + (\sigma_a^2 + \sigma_u^2) + \ldots + \sigma_a^2] = \sigma_a^2 + \sigma_u^2/T$. Therefore, E($\bar{v}_i^2$) = $T^{-1}\sum_{t=1}^{T} E(v_{it}\bar{v}_i)$ = $\sigma_a^2 + \sigma_u^2/T$. Now, we can collect terms:

$$\text{Var}(v_{it} - \lambda\bar{v}_i) = (\sigma_a^2 + \sigma_u^2) + \lambda^2(\sigma_a^2 + \sigma_u^2/T) - 2\lambda(\sigma_a^2 + \sigma_u^2/T).$$

Now, it is convenient to write $\lambda = 1 - \sqrt{\eta}/\sqrt{\gamma}$, where $\eta \equiv \sigma_u^2/T$ and $\gamma \equiv \sigma_a^2 + \sigma_u^2/T$. Then

$$
\begin{aligned}
\text{Var}(v_{it} - \lambda\bar{v}_i) &= (\sigma_a^2 + \sigma_u^2) - 2\lambda(\sigma_a^2 + \sigma_u^2/T) + \lambda^2(\sigma_a^2 + \sigma_u^2/T) \\
&= (\sigma_a^2 + \sigma_u^2) - 2(1 - \sqrt{\eta}/\sqrt{\gamma})\gamma + (1 - \sqrt{\eta}/\sqrt{\gamma})^2\gamma \\
&= (\sigma_a^2 + \sigma_u^2) - 2\gamma + 2\sqrt{\eta}\cdot\sqrt{\gamma} + (1 - 2\sqrt{\eta}/\sqrt{\gamma} + \eta/\gamma)\gamma \\
&= (\sigma_a^2 + \sigma_u^2) - 2\gamma + 2\sqrt{\eta}\cdot\sqrt{\gamma} + (1 - 2\sqrt{\eta}/\sqrt{\gamma} + \eta/\gamma)\gamma \\
&= (\sigma_a^2 + \sigma_u^2) - 2\gamma + 2\sqrt{\eta}\cdot\sqrt{\gamma} + \gamma - 2\sqrt{\eta}\cdot\sqrt{\gamma} + \eta \\
&= (\sigma_a^2 + \sigma_u^2) + \eta - \gamma = \sigma_u^2.
\end{aligned}
$$

This is what we wanted to show.

(iii) We must show that $E(e_{it}e_{is}) = 0$ for $t \neq s$. Now $E(e_{it}e_{is}) = E[(v_{it} - \lambda\bar{v}_i)(v_{is} - \lambda\bar{v}_i)] = E(v_{it}v_{is}) - \lambda E(\bar{v}_i v_{is}) - \lambda E(v_{it}\bar{v}_i) + \lambda^2 E(\bar{v}_i^2) = \sigma_a^2 - 2\lambda(\sigma_a^2 + \sigma_u^2/T) + \lambda^2 E(\bar{v}_i^2) = \sigma_a^2 - 2\lambda(\sigma_a^2 + \sigma_u^2/T) + \lambda^2(\sigma_a^2 + \sigma_u^2/T)$. The rest of the proof is very similar to part (ii):

$$\begin{aligned}
E(e_{it}e_{is}) &= \sigma_a^2 - 2\lambda(\sigma_a^2 + \sigma_u^2/T) + \lambda^2(\sigma_a^2 + \sigma_u^2/T) \\
&= \sigma_a^2 - 2(1 - \sqrt{\eta}/\sqrt{\gamma})\gamma + (1 - \sqrt{\eta}/\sqrt{\gamma})^2\gamma \\
&= \sigma_a^2 - 2\gamma + 2\sqrt{\eta}\cdot\sqrt{\gamma} + (1 - 2\sqrt{\eta}/\sqrt{\gamma} + \eta/\gamma)\gamma \\
&= \sigma_a^2 - 2\gamma + 2\sqrt{\eta}\cdot\sqrt{\gamma} + (1 - 2\sqrt{\eta}/\sqrt{\gamma} + \eta/\gamma)\gamma \\
&= \sigma_a^2 - 2\gamma + 2\sqrt{\eta}\cdot\sqrt{\gamma} + \gamma - 2\sqrt{\eta}\cdot\sqrt{\gamma} + \eta \\
&= \sigma_a^2 + \eta - \gamma = 0.
\end{aligned}$$

14.5 (i) For each student we have several measures of performance, typically three or four, the number of classes taken by a student that have final exams. When we specify an equation for each standardized final exam score, the errors in the different equations for the same student are certain to be correlated: students who have more (unobserved) ability tend to do better on all tests.

(ii) An unobserved effects model is

$$score_{sc} = \theta_c + \beta_1 atndrte_{sc} + \beta_2 major_{sc} + \beta_3 SAT_s + \beta_4 cumGPA_s + a_s + u_{sc},$$

where a_s is the unobserved student effect. Because SAT score and cumulative GPA depend only on the student, and not on the particular class he/she is taking, these do not have a c subscript. The attendance rates do generally vary across class, as does the indicator for whether a class is in the student's major. The term θ_c denotes different intercepts for different classes. Unlike with a panel data set, where time is the natural ordering of the data within each cross-sectional unit, and the aggregate time effects apply to all units, intercepts for the different classes may not be needed. If all students took the same set of classes then this is similar to a panel data set, and we would want to put in different class intercepts. But with students taking different courses, the class we label as "1" for student A need have nothing to do with class "1" for student B. Thus, the different class intercepts based on arbitrarily ordering the classes for each student probably are not needed. We can replace θ_c with β_0, an intercept constant across classes.

(iii) Maintaining the assumption that the idiosyncratic error, u_{sc}, is uncorrelated with all explanatory variables, we need the unobserved student heterogeneity, a_s, to be uncorrelated with $atndrte_{sc}$. The inclusion of SAT score and cumulative GPA should help in this regard, as a_s, is the part of ability that is not captured by SAT_s and $cumGPA_s$. In other words, controlling for SAT_s and $cumGPA_s$ could be enough to obtain the ceteris paribus effect of class attendance.

(iv) If SAT_s and $cumGPA_s$ are not sufficient controls for student ability and motivation, a_s is correlated with $atndrte_{sc}$, and this would cause pooled OLS to be biased and inconsistent. We could use fixed effects instead. Within each student we compute the demeaned data, where, for each student, the means are computed across classes. The variables SAT_s and $cumGPA_s$ drop out of the analysis.

SOLUTIONS TO COMPUTER EXERCISES

C14.1 (i) This is done in Computer Exercise 13.5(i).

(ii) See Computer Exercise 13.5(ii).

(iii) See Computer Exercise 13.5(iii).

(iv) This is the only new part. The fixed effects estimates, reported in equation form, are

$$\widehat{\log(rent_{it})} = \underset{(.037)}{.386}\, y90_t + \underset{(.088)}{.072} \log(pop_{it}) + \underset{(.066)}{.310} \log(avginc_{it}) + \underset{(.0041)}{.0112}\, pctstu_{it},$$

$$N = 64, \quad T = 2.$$

(There are $N = 64$ cities and $T = 2$ years.) We do not report an intercept because it gets removed by the time demeaning. The coefficient on $y90_t$ is identical to the intercept from the first difference estimation, and the slope coefficients and standard errors are identical to first differencing. We do not report an R-squared because none is comparable to the R-squared obtained from first differencing.

C14.3 (i) 135 firms are used in the FE estimation. Because there are three years, we would have a total of 405 observations if each firm had data on all variables for all three years. Instead, due to missing data, we can use only 390 observations in the FE estimation. The fixed effects estimates are

$$\widehat{hrsemp}_{it} = \underset{(1.98)}{-1.10}\, d88_t + \underset{(2.48)}{4.09}\, d89_t + \underset{(2.86)}{34.23}\, grant_{it}$$

$$+ \underset{(4.127)}{.504}\, grant_{i,t-1} - \underset{(4.288)}{.176} \log(employ_{it})$$

$$n = 390, \quad N = 135, \quad T = 3.$$

(ii) The coefficient on *grant* means that if a firm received a grant for the current year, it trained each worker an average of 34.2 hours more than it would have otherwise. This is a practically large effect, and the t statistic is very large.

(iii) Since a grant last year was used to pay for training last year, it is perhaps not surprising that the grants does not carry over into more training this year. It would if inertia played a role in training workers.

(iv) The coefficient on the employees variable is very small: a 10% increase in *employ* increases predicted hours per employee by only about .018. [Recall: $\Delta \widehat{hrsemp} \approx (.176/100)$ $(\%\Delta employ)$.] This is very small, and the t statistic is practically zero.

C14.5 (i) Different occupations are unionized at different rates, and wages also differ by occupation. Therefore, if we omit binary indicators for occupation, the union wage differential may simply be picking up wage differences across occupations. Because some people change occupation over the period, we should include these in our analysis.

(ii) Because the nine occupational categories (*occ1* through *occ9*) are exhaustive, we must choose one as the base group. Of course the group we choose does not affect the estimated union wage differential. The fixed effect estimate on *union*, to four decimal places, is .0804 with standard error = .0194. There is practically no difference between this estimate and standard error and the estimate and standard error without the occupational controls ($\hat{\beta}_{union}$ = .0800, se = .0193).

C14.7 (i) If there is a deterrent effect then $\beta_1 < 0$. The sign of β_2 is not entirely obvious, although one possibility is that a better economy means less crime in general, including violent crime (such as drug dealing) that would lead to fewer murders. This would imply $\beta_2 > 0$.

(ii) The pooled OLS estimates using 1990 and 1993 are

$$\begin{aligned}\widehat{mrdrte}_{it} = \underset{(4.43)}{-5.28} - \underset{(2.14)}{2.07}\, d93_t + \underset{(.263)}{.128}\, exec_{it} + \underset{(0.78)}{2.53}\, unem_{it}\end{aligned}$$

$N = 51,\ T = 2,\ R^2 = .102$

There is no evidence of a deterrent effect, as the coefficient on *exec* is actually positive (though not statistically significant).

(iii) The first-differenced equation is

$$\Delta\widehat{mrdrte}_i = \underset{(.209)}{.413} - \underset{(.043)}{.104}\, \Delta exec_i - \underset{(.159)}{.067}\, \Delta unem_i$$

$n = 51,\ R^2 = .110$

Now, there is a statistically significant deterrent effect: 10 more executions is estimated to reduce the murder rate by 1.04, or one murder per 100,000 people. Is this a large effect? Executions are relatively rare in most states, but murder rates are relatively low on average, too.

In 1993, the average murder rate was about 8.7; a reduction of one would be nontrivial. For the (unknown) people whose lives might be saved via a deterrent effect, it would seem important.

(iv) The heteroskedasticity-robust standard error for $\Delta exec_i$ is .017. Somewhat surprisingly, this is well below the nonrobust standard error. If we use the robust standard error, the statistical evidence for the deterrent effect is quite strong ($t \approx -6.1$). See also Computer Exercise 13.12.

(v) Texas had by far the largest value of *exec*, 34. The next highest state was Virginia, with 11. These are three-year totals.

(vi) Without Texas in the estimation, we get the following, with heteroskedasticity-robust standard errors in [·]:

$$\begin{aligned}\Delta \widehat{mrdrte}_i = \underset{\substack{(.211)\\ [.200]}}{.413} - \underset{\substack{(.105)\\ [.079]}}{.067}\,\Delta exec_i - \underset{\substack{(.160)\\ [.146]}}{.070}\,\Delta unem_i\end{aligned}$$

$$n = 50,\ R^2 = .013$$

Now the estimated deterrent effect is smaller. Perhaps more importantly, the standard error on $\Delta exec_i$ has increased by a substantial amount. This happens because when we drop Texas, we lose much of the variation in the key explanatory variable, $\Delta exec_i$.

(vii) When we apply fixed effects using all three years of data and all states we get

$$\widehat{mrdrte}_{it} = \underset{(.75)}{1.73}\, d90_t + \underset{(.71)}{1.70}\, d93_t - \underset{(.160)}{.054}\, exec_{it} + \underset{(.285)}{.395}\, unem_{it}$$

$$N = 51,\ T = 3,\ R^2 = .068$$

The size of the deterrent effect is only about half as big as when 1987 is not used. Plus, the t statistic, about $-.34$, is very small. The earlier finding of a deterrent effect is not robust to the time period used. Oddly, adding another year of data causes the standard error on the *exec* coefficient to markedly increase.

C14.9 (i) The OLS estimates are

$$\begin{aligned}\widehat{pctstck} = {} & \underset{(55.17)}{128.54} + \underset{(6.23)}{11.74}\, choice + \underset{(7.23)}{14.34}\, prftshr + \underset{(6.77)}{1.45}\, female - \underset{(.78)}{1.50}\, age \\ & + \underset{(1.20)}{.70}\, educ - \underset{(14.23)}{15.29}\, finc25 + \underset{(14.69)}{.19}\, finc35 - \underset{(14.55)}{3.86}\, finc50 \\ & - 13.75\, finc75 - 2.69\, finc100 - 25.05\, finc101 - .0026\, wealth89\end{aligned}$$

$$\begin{array}{cccc} (16.02) & (15.72) & (17.80) & (.0128) \end{array}$$

$$+ \quad 6.67\ stckin89 \ - \ 7.50\ irain89$$
$$\quad (6.68) \qquad\qquad (6.38)$$

$n = 194,\ R^2 = .108$

Investment choice is associated with about 11.7 percentage points more in stocks. The t statistic is 1.88, and so it is marginal significant.

(ii) These variables are not very important. The F test for joint significant is 1.03. With 9 and 179 *df*, this gives p-value = .42. Plus, when these variables are dropped from the regression, the coefficient on *choice* only falls to 11.15.

(iii) There are 171 different families in the sample.

(iv) I will only report the cluster-robust standard error for *choice*: 6.20. Therefore, it is essentially the same as the usual OLS standard error. This is not very surprising because at least 171 of the 194 observations can be assumed independent of one another. The explanatory variables may adequately capture the within-family correlation.

(v) There are only 23 families with spouses in the data set. Differencing within these families gives

$$\Delta\widehat{pctstck} = \underset{(10.94)}{15.93} + \underset{(15.00)}{2.28}\ \Delta choice - \underset{(16.92)}{9.27}\Delta prftshr + \underset{(21.49)}{21.55}\ \Delta female - \underset{(9.00)}{3.57}\ \Delta age$$

$$\underset{(3.43)}{-1.22}\ \Delta educ$$

$n = 23,\ R^2 = .206,\ \ \bar{R}^2 = -.028$

All of the income and wealth variables, and the stock and IRA indicators, drop out, as these are defined at the family level (and therefore are the same for the husband and wife).

(vi) None of the explanatory variables is significant in part (v), and this is not too surprising. We have only 23 observations, and we are removing much of the variation in the explanatory variables (except the gender variable) by using within-family differences.

C14.11 (i) The robust standard errors on *educ*, *married*, and *union* are all quite a bit larger than the usual OLS standard errors. In the case of *educ*, the robust standard error is about .0111, compared with the usual OLS standard error .0052; this is more than a doubling. For *married*, the robust standard error is about .0260, which again is much higher than the usual standard error, .0157. A similar change is evident for *union* (from .0172 to .0274).

(ii) For *married*, the usual FE standard error is .0183, and the fully robust one is .0210. For *union*, these are .0193 and .0227, respectively. In both cases, the robust standard error is somewhat higher.

(iii) The relative increase in standard errors when we go from the usual standard error to the robust version is much higher for pooled OLS than for FE. For FE, the increases are on the order of 15%, or slightly higher. For pooled OLS, the increases for *married* and *union* are on the order of at least 60%. Typically, the adjustment for FE has a smaller relative effect because FE removes the main source of positive serial correlation: the unobserved effect, a_i. Remember, pooled OLS leaves a_i in the error term. The usual standard errors for both pooled OLS and FE are invalid with serial correlation in the idiosyncratic errors, u_{it}, but this correlation is usually of a smaller degree. (And, in some applications, it is not unreasonable to think the u_{it} have no serial correlation. However, if we are being careful, we allow this possibility in computing our standard errors and test statistics.)

CHAPTER 15

SOLUTIONS TO PROBLEMS

15.1 (i) It has been fairly well established that socioeconomic status affects student performance. The error term u contains, among other things, family income, which has a positive effect on *GPA* and is also very likely to be correlated with PC ownership.

(ii) Families with higher incomes can afford to buy computers for their children. Therefore, family income certainly satisfies the second requirement for an instrumental variable: it is correlated with the endogenous explanatory variable [see (15.5) with $x = PC$ and $z = faminc$]. But as we suggested in part (i), *faminc* has a positive affect on *GPA*, so the first requirement for a good IV, (15.4), fails for *faminc*. If we had *faminc* we would include it as an explanatory variable in the equation; if it is the only important omitted variable correlated with *PC*, we could then estimate the expanded equation by OLS.

(iii) This is a natural experiment that affects whether or not some students own computers. Some students who buy computers when given the grant would not have without the grant. (Students who did not receive the grants might still own computers.) Define a dummy variable, *grant*, equal to one if the student received a grant, and zero otherwise. Then, if *grant* was randomly assigned, it is uncorrelated with u. In particular, it is uncorrelated with family income and other socioeconomic factors in u. Further, *grant* should be correlated with *PC*: the probability of owning a PC should be significantly higher for student receiving grants. Incidentally, if the university gave grant priority to low-income students, *grant* would be negatively correlated with u, and IV would be inconsistent.

15.3 It is easiest to use (15.10) but where we drop $\bar{z}$. Remember, this is allowed because $\sum_{i=1}^{n}(z_i - \bar{z})\,(x_i - \bar{x}) = \sum_{i=1}^{n} z_i(x_i - \bar{x})$ and similarly when we replace x with y. So the numerator in the formula for $\hat{\beta}_1$ is

$$\sum_{i=1}^{n} z_i(y_i - \bar{y}) = \sum_{i=1}^{n} z_i y_i - \left(\sum_{i=1}^{n} z_i\right)\bar{y} \;=\; n_1\bar{y}_1 - n_1\bar{y},$$

where $n_1 = \sum_{i=1}^{n} z_i$ is the number of observations with $z_i = 1$, and we have used the fact that $\left(\sum_{i=1}^{n} z_i y_i\right)/n_1 = \bar{y}_1$, the average of the y_i over the i with $z_i = 1$. So far, we have shown that the numerator in $\hat{\beta}_1$ is $n_1(\bar{y}_1 - \bar{y})$. Next, write $\bar{y}$ as a weighted average of the averages over the two subgroups:

$$\bar{y} \;=\; (n_0/n)\,\bar{y}_0 + (n_1/n)\,\bar{y}_1,$$

where $n_0 = n - n_1$. Therefore,

$$\bar{y}_1 - \bar{y} = [(n - n_1)/n]\ \bar{y}_1 - (n_0/n)\ \bar{y}_0 = (n_0/n)\ (\bar{y}_1 - \bar{y}_0).$$

Therefore, the numerator of $\hat{\beta}_1$ can be written as

$$(n_0 n_1/n)(\bar{y}_1 - \bar{y}_0).$$

By simply replacing y with x, the denominator in $\hat{\beta}_1$ can be expressed as $(n_0 n_1/n)(\bar{x}_1 - \bar{x}_0)$. When we take the ratio of these, the terms involving n_0, n_1, and n, cancel, leaving

$$\hat{\beta}_1 = (\bar{y}_1 - \bar{y}_0)/(\bar{x}_1 - \bar{x}_0).$$

15.5 (i) From equation (15.19) with $\sigma_u = \sigma_x$, plim $\hat{\beta}_1 = \beta_1 + (.1/.2) = \beta_1 + .5$, where $\hat{\beta}_1$ is the IV estimator. So the asymptotic bias is .5.

(ii) From equation (15.20) with $\sigma_u = \sigma_x$, plim $\tilde{\beta}_1 = \beta_1 + \text{Corr}(x,u)$, where $\tilde{\beta}_1$ is the OLS estimator. So we would have to have $\text{Corr}(x,u) > .5$ before the asymptotic bias in OLS exceeds that of IV. This is a simple illustration of how a seemingly small correlation (.1 in this case) between the IV (z) and error (u) can still result in IV being more biased than OLS if the correlation between z and x is weak (.2).

15.7 (i) Even at a given income level, some students are more motivated and more able than others, and their families are more supportive (say, in terms of providing transportation) and enthusiastic about education. Therefore, there is likely to be a self-selection problem: students that would do better anyway are also more likely to attend a choice school.

(ii) Assuming we have the functional form for *faminc* correct, the answer is yes. Since u_1 does not contain income, random assignment of grants within income class means that grant designation is not correlated with unobservables such as student ability, motivation, and family support.

(iii) The reduced form is

$$choice = \pi_0 + \pi_1 faminc + \pi_2 grant + v_2,$$

and we need $\pi_2 \neq 0$. In other words, after accounting for income, the grant amount must have some affect on *choice*. This seems reasonable, provided the grant amounts differ within each income class.

(iv) The reduced form for score is just a linear function of the exogenous variables (see Problem 15.6):

$$score = \alpha_0 + \alpha_1 faminc + \alpha_2 grant + v_1.$$

This equation allows us to directly estimate the effect of increasing the grant amount on the test score, holding family income fixed. From a policy perspective this is itself of some interest.

15.9 Just use OLS on an expanded equation, where *SAT* and *cumGPA* are added as proxy variables for student ability and motivation; see Chapter 9.

15.11 (i) We plug $x_t^* = x_t - e_t$ into $y_t = \beta_0 + \beta_1 x_t^* + u_t$:

$$\begin{aligned} y_t &= \beta_0 + \beta_1(x_t - e_t) + u_t = \beta_0 + \beta_1 x_t + u_t - \beta_1 e_t \\ &\equiv \beta_0 + \beta_1 x_t + v_t, \end{aligned}$$

where $v_t \equiv u_t - \beta_1 e_t$. By assumption, u_t is uncorrelated with x_t^* and e_t; therefore, u_t is uncorrelated with x_t. Since e_t is uncorrelated with x_t^*, $\mathrm{E}(x_t e_t) = \mathrm{E}[(x_t^* + e_t)e_t] = \mathrm{E}(x_t^* e_t) + \mathrm{E}(e_t^2) = \sigma_e^2$. Therefore, with v_t defined as above, $\mathrm{Cov}(x_t,v_t) = \mathrm{Cov}(x_t,u_t) - \beta_1 \mathrm{Cov}(x_t,e_t) = -\beta_1 \sigma_e^2 < 0$ when $\beta_1 > 0$. Because the explanatory variable and the error have negative covariance, the OLS estimator of β_1 has a downward bias [see equation (5.4)].

(ii) By assumption $\mathrm{E}(x_{t-1}^* u_t) = \mathrm{E}(e_{t-1}u_t) = \mathrm{E}(x_{t-1}^* e_t) = \mathrm{E}(e_{t-1}e_t) = 0$, and so $\mathrm{E}(x_{t-1}u_t) = \mathrm{E}(x_{t-1}e_t) = 0$ because $x_t = x_t^* + e_t$. Therefore, $\mathrm{E}(x_{t-1}v_t) = \mathrm{E}(x_{t-1}u_t) - \beta_1 \mathrm{E}(x_{t-1}e_t) = 0$.

(iii) Most economic time series, unless they represent the first difference of a series or the percentage change, are positively correlated over time. If the initial equation is in levels or logs, x_t and x_{t-1} are likely to be positively correlated. If the model is for first differences or percentage changes, there still may be positive or negative correlation between x_t and x_{t-1}.

(iv) Under the assumptions made, x_{t-1} is exogenous in

$$y_t = \beta_0 + \beta_1 x_t + v_t,$$

as we showed in part (ii): $\mathrm{Cov}(x_{t-1},v_t) = \mathrm{E}(x_{t-1}v_t) = 0$. Second, x_{t-1} will often be correlated with x_t, and we can check this easily enough by running a regression of x_t of x_{t-1}. This suggests estimating the equation by instrumental variables, where x_{t-1} is the IV for x_t. The IV estimator will be consistent for β_1 (and β_0), and asymptotically normally distributed.

SOLUTIONS TO COMPUTER EXERCISES

C15.1 (i) The regression of log(*wage*) on *sibs* gives

$$\widehat{\log(wage)} = \underset{(0.022)}{6.861} - \underset{(.0059)}{.0279}\, sibs$$

$$n = 935, \quad R^2 = .023.$$

This is a reduced form simple regression equation. It shows that, controlling for no other factors, one more sibling in the family is associated with monthly salary that is about 2.8% lower. The *t* statistic on *sibs* is about –4.73. Of course *sibs* can be correlated with many things that should have a bearing on wage including, as we already saw, years of education.

(ii) It could be that older children are given priority for higher education, and families may hit budget constraints and may not be able to afford as much education for children born later. The simple regression of *educ* on *brthord* gives

$$\widehat{educ} = \underset{(0.13)}{14.15} - \underset{(.046)}{.283}\, brthord$$

$$n = 852, \quad R^2 = .042.$$

(Note that *brthord* is missing for 83 observations.) The equation predicts that every one-unit increase in *brthord* reduces predicted education by about .28 years. In particular, the difference in predicted education for a first-born and fourth-born child is about .85 years.

(iii) When *brthord* is used as an IV for *educ* in the simple wage equation we get

$$\widehat{\log(wage)} = \underset{(0.43)}{5.03} + \underset{(.032)}{.131}\, educ$$

$$n = 852.$$

(The *R*-squared is negative.) This is much higher than the OLS estimate (.060) and even above the estimate when *sibs* is used as an IV for *educ* (.122). Because of missing data on *brthord*, we are using fewer observations than in the previous analyses.

(iv) In the reduced form equation

$$educ = \pi_0 + \pi_1 sibs + \pi_2 brthord + v,$$

we need $\pi_2 \neq 0$ in order for the β_j to be identified. We take the null to be H_0: $\pi_2 = 0$, and look to reject H_0 at a small significance level. The regression of *educ* on *sibs* and *brthord* (using 852 observations) yields $\hat{\pi}_2 = -.153$ and se($\hat{\pi}_2$) = .057. The *t* statistic is about –2.68, which rejects H_0 fairly strongly. Therefore, the identification assumptions appears to hold.

(v) The equation estimated by IV is

$$\widehat{\log(wage)} = \underset{(1.06)}{4.94} + \underset{(.075)}{.137}\, educ + \underset{(.0174)}{.0021}\, sibs$$

$$n = 852.$$

The standard error on $\hat{\beta}_{educ}$ is much larger than we obtained in part (iii). The 95% CI for β_{educ} is roughly –.010 to .284, which is very wide and includes the value zero. The standard error of $\hat{\beta}_{sibs}$ is very large relative to the coefficient estimate, rendering *sibs* very insignificant.

(vi) Letting $\widehat{educ}_i$ be the first-stage fitted values, the correlation between $\widehat{educ}_i$ and $sibs_i$ is about –.930, which is a very strong negative correlation. This means that, for the purposes of using IV, multicollinearity is a serious problem here, and is not allowing us to estimate β_{educ} with much precision.

C15.3 (i) IQ scores are known to vary by geographic region, and so does the availability of four year colleges. It could be that, for a variety of reasons, people with higher abilities grow up in areas with four year colleges nearby.

(ii) The simple regression of *IQ* on *nearc4* gives

$$\widehat{IQ} = \underset{(0.63)}{100.61} + \underset{(0.74)}{2.60}\, nearc4$$

$$n = 2{,}061,\quad R^2 = .0059,$$

which shows that predicted *IQ* score is about 2.6 points higher for a man who grew up near a four-year college. The difference is statistically significant (*t* statistic $\approx$ 3.51).

(iii) When we add *smsa66*, *reg662*, …, *reg669* to the regression in part (ii), we obtain

$$\widehat{IQ} = \underset{(1.62)}{104.77} + \underset{(.814)}{.348}\, nearc4 + \underset{(0.81)}{1.09}\, smsa66 + \ldots$$

$$n = 2{,}061,\quad R^2 = .0626,$$

where, for brevity, the coefficients on the regional dummies are not reported. Now, the relationship between *IQ* and *nearc4* is much weaker and statistically insignificant. In other words, once we control for region and environment while growing up, there is no apparent link between IQ score and living near a four-year college.

(iv) The findings from parts (ii) and (iii) show that it is important to include *smsa66*, *reg662*, …, *reg669* in the wage equation to control for differences in access to colleges that might also be correlated with ability.

C15.5 (i) When we add $\hat{v}_2$ to the original equation and estimate it by OLS, the coefficient on $\hat{v}_2$ is about –.057 with a *t* statistic of about –1.08. Therefore, while the difference in the estimates of the return to education is practically large, it is not statistically significant.

(ii) We now add *nearc2* as an IV along with *nearc4*. (Although, in the reduced form for *educ*, *nearc2* is not significant.) The 2SLS estimate of β_1 is now .157, se($\hat{\beta}_1$) = .053. So the estimate is even larger.

(iii) Let $\hat{u}_i$ be the 2SLS residuals. We regress these on all exogenous variables, including *nearc2* and *nearc4*. The *n*-*R*-squared statistic is (3,010)(.0004) $\approx$ 1.20. There is one over-identifying restriction, so we compute the *p*-value from the χ_1^2 distribution: *p*-value = P(χ_1^2 > 1.20) $\approx$ .55, so the overidentifying restriction is not rejected.

C15.7 (i) As usual, if $unem_t$ is correlated with e_t, OLS will be biased and inconsistent for estimating β_1.

(ii) If $\mathrm{E}(e_t|inf_{t-1},unem_{t-1},\ \ldots) = 0$ then $unem_{t-1}$ is uncorrelated with e_t, which means $unem_{t-1}$ satisfies the first requirement for an IV in

$$\Delta inf_t = \beta_0 + \beta_1 unem_t + e_t.$$

(iii) The second requirement for $unem_{t-1}$ to be a valid IV for $unem_t$ is that $unem_{t-1}$ must be sufficiently correlated. The regression $unem_t$ on $unem_{t-1}$ yields

$$\begin{aligned} \widehat{unem}_t &= \underset{(0.58)}{1.57} + \underset{(.097)}{.732}\, unem_{t-1} \\ n &= 48,\quad R^2 = .554. \end{aligned}$$

Therefore, there is a strong, positive correlation between $unem_t$ and $unem_{t-1}$.

(iv) The expectations-augmented Phillips curve estimated by IV is

$$\begin{aligned} \Delta\widehat{inf}_t &= \underset{(1.883)}{.694} - \underset{(.319)}{.138}\, unem_t \\ n &= 48,\quad R^2 = .048. \end{aligned}$$

The IV estimate of β_1 is much lower in magnitude than the OLS estimate (−.543), and $\hat{\beta}_1$ is not statistically different from zero. The OLS estimate had a *t* statistic of about –2.36 [see equation (11.19)].

C15.9 (i) The IV (2SLS) estimates are

$$\widehat{\log(wage)} = \underset{(.54)}{5.22} + \underset{(.0337)}{.0936}\, educ + \underset{(.0084)}{.0209}\, exper + \underset{(.0027)}{.0115}\, tenure - \underset{(.050)}{.183}\, black$$

$n = 935,\ R^2 = .169$

(ii) The coefficient on $\widehat{educ_i}$ in the second stage regression is, naturally, .0936. But the reported standard error is .0353, which is slightly too large.

(iii) When instead we (incorrectly) use $\widetilde{educ_i}$ in the second stage regression, its coefficient is .0700 and the corresponding standard error is .0264. Both are too low. The reduction in the estimated return to education from about 9.4% to 7.0% is not trivial. This illustrates that it is best to avoid doing 2SLS manually.

CHAPTER 16

SOLUTIONS TO PROBLEMS

16.1 (i) If $\alpha_1 = 0$ then $y_1 = \beta_1 z_1 + u_1$, and so the right-hand-side depends only on the exogenous variable z_1 and the error term u_1. This then is the reduced form for y_1. If $\alpha_1 = 0$, the reduced form for y_1 is $y_1 = \beta_2 z_2 + u_2$. (Note that having both α_1 and α_2 equal zero is not interesting as it implies the bizarre condition $u_2 - u_1 = \beta_1 z_1 - \beta_2 z_2$.)

If $\alpha_1 \neq 0$ and $\alpha_2 = 0$, we can plug $y_1 = \beta_2 z_2 + u_2$ into the first equation and solve for y_2:

$$\beta_2 z_2 + u_2 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$

or

$$\alpha_1 y_2 = \beta_1 z_1 - \beta_2 z_2 + u_1 - u_2.$$

Dividing by α_1 (because $\alpha_1 \neq 0$) gives

$$\begin{aligned} y_2 &= (\beta_1/\alpha_1)z_1 - (\beta_2/\alpha_1)z_2 + (u_1 - u_2)/\alpha_1 \\ &\equiv \pi_{21} z_1 + \pi_{22} z_2 + v_2, \end{aligned}$$

where $\pi_{21} = \beta_1/\alpha_1$, $\pi_{22} = -\beta_2/\alpha_1$, and $v_2 = (u_1 - u_2)/\alpha_1$. Note that the reduced form for y_2 generally depends on z_1 and z_2 (as well as on u_1 and u_2).

(ii) If we multiply the second structural equation by (α_1/α_2) and subtract it from the first structural equation, we obtain

$$\begin{aligned} y_1 - (\alpha_1/\alpha_2)y_1 &= \alpha_1 y_2 - \alpha_1 y_2 + \beta_1 z_1 - (\alpha_1/\alpha_2)\beta_2 z_2 + u_1 - (\alpha_1/\alpha_2)u_2 \\ &= \beta_1 z_1 - (\alpha_1/\alpha_2)\beta_2 z_2 + u_1 - (\alpha_1/\alpha_2)u_2 \end{aligned}$$

or

$$[1 - (\alpha_1/\alpha_2)]y_1 = \beta_1 z_1 - (\alpha_1/\alpha_2)\beta_2 z_2 + u_1 - (\alpha_1/\alpha_2)u_2.$$

Because $\alpha_1 \neq \alpha_2$, $1 - (\alpha_1/\alpha_2) \neq 0$, and so we can divide the equation by $1 - (\alpha_1/\alpha_2)$ to obtain the reduced form for y_1: $y_1 = \pi_{11} z_1 + \pi_{12} z_2 + v_1$, where $\pi_{11} = \beta_1/[1 - (\alpha_1/\alpha_2)]$, $\pi_{12} = -(\alpha_1/\alpha_2)\beta_2/[1 - (\alpha_1/\alpha_2)]$, and $v_1 = [u_1 - (\alpha_1/\alpha_2)u_2]/[1 - (\alpha_1/\alpha_2)]$.

A reduced form does exist for y_2, as can be seen by subtracting the second equation from the first:

$$0 = (\alpha_1 - \alpha_2)y_2 + \beta_1 z_1 - \beta_2 z_2 + u_1 - u_2;$$

because $\alpha_1 \neq \alpha_2$, we can rearrange and divide by $\alpha_1 - \alpha_2$ to obtain the reduced form.

(iii) In supply and demand examples, $\alpha_1 \neq \alpha_2$ is very reasonable. If the first equation is the supply function, we generally expect $\alpha_1 > 0$, and if the second equation is the demand function, $\alpha_2 < 0$. The reduced forms can exist even in cases where the supply function is not upward sloping and the demand function is not downward sloping, but we might question the usefulness of such models.

16.3 No. In this example, we are interested in estimating the tradeoff between sleeping and working, controlling for some other factors. OLS is perfectly suited for this, provided we have been able to control for all other relevant factors. While it is true individuals are assumed to optimally allocate their time subject to constraints, this does not result in a system of simultaneous equations. If we wrote down such a system, there is no sense in which each equation could stand on its own; neither would have an interesting ceteris paribus interpretation. Besides, we could not estimate either equation because economic reasoning gives us no way of excluding exogenous variables from either equation. See Example 16.2 for a similar discussion.

16.5 (i) Other things equal, a higher rate of condom usage should reduce the rate of sexually transmitted diseases (STDs). So $\beta_1 < 0$.

(ii) If students having sex behave rationally, and condom usage does prevent STDs, then condom usage should increase as the rate of infection increases.

(iii) If we plug the structural equation for *infrate* into $conuse = \gamma_0 + \gamma_1 infrate + \ldots$, we see that *conuse* depends on $\gamma_1 u_1$. Because $\gamma_1 > 0$, *conuse* is positively related to u_1. In fact, if the structural error (u_2) in the *conuse* equation is uncorrelated with u_1, $\text{Cov}(conuse,u_1) = \gamma_1\text{Var}(u_1) > 0$. If we ignore the other explanatory variables in the *infrate* equation, we can use equation (5.4) to obtain the direction of bias: $\text{plim}(\hat{\beta}_1) - \beta_1 > 0$ because $\text{Cov}(conuse,u_1) > 0$, where $\hat{\beta}_1$ denotes the OLS estimator. Since we think $\beta_1 < 0$, OLS is biased towards zero. In other words, if we use OLS on the *infrate* equation, we are likely to underestimate the importance of condom use in reducing STDs. (Remember, the more negative is β_1, the more effective is condom usage.)

(iv) We would have to assume that *condis* does not appear, in addition to *conuse*, in the *infrate* equation. This seems reasonable, as it is usage that should directly affect STDs, and not just having a distribution program. But we must also assume *condis* is exogenous in the *infrate*: it cannot be correlated with unobserved factors (in u_1) that also affect *infrate*.

We must also assume that *condis* has some partial effect on *conuse*, something that can be tested by estimating the reduced form for *conuse*. It seems likely that this requirement for an IV – see equations (15.30) and (15.31) – is satisfied.

16.7 (i) Attendance at women's basketball may grow in ways that are unrelated to factors that we can observe and control for. The taste for women's basketball may increase over time, and this would be captured by the time trend.

(ii) No. The university sets the price, and it may change price based on expectations of next year's attendance; if the university uses factors that we cannot observe, these are necessarily in the error term u_t. So even though the supply is fixed, it does not mean that price is uncorrelated with the unobservables affecting demand.

(iii) If people only care about how this year's team is doing, $SEASPERC_{t\text{-}1}$ can be excluded from the equation once $WINPERC_t$ has been controlled for. Of course, this is not a very good assumption for all games, as attendance early in the season is likely to be related to how the team did last year. We would also need to check that $1PRICE_t$ is partially correlated with $SEASPERC_{t\text{-}1}$ by estimating the reduced form for $1PRICE_t$.

(iv) It does make sense to include a measure of men's basketball ticket prices, as attending a women's basketball game is a substitute for attending a men's game. The coefficient on $1MPRICE_t$ would be expected to be positive: an increase in the price of men's tickets should increase the demand for women's tickets. The winning percentage of the men's team is another good candidate for an explanatory variable in the women's demand equation.

(v) It might be better to use first differences of the logs, which are then growth rates. We would then drop the observation for the first game in each season.

(vi) If a game is sold out, we cannot observe true demand for that game. We only know that desired attendance is some number above capacity. If we just plug in capacity, we are understating the actual demand for tickets. (Chapter 17 discusses censored regression methods that can be used in such cases.)

SOLUTIONS TO COMPUTER EXERCISES

C16.1 (i) Assuming the structural equation represents a causal relationship, $100 \cdot \beta_1$ is the approximate percentage change in income if a person smokes one more cigarette per day.

(ii) Since consumption and price are, ceteris paribus, negatively related, we expect $\gamma_5 \leq 0$ (allowing for γ_5) = 0. Similarly, everything else equal, restaurant smoking restrictions should reduce cigarette smoking, so $\gamma_5 \leq 0$.

(iii) We need γ_5 or γ_6 to be different from zero. That is, we need at least one exogenous variable in the *cigs* equation that is not also in the log(*income*) equation.

(iv) OLS estimation of the log(*income*) equation gives

$$\begin{array}{rcccccc} \widehat{\log(income)} = & 7.80 & + .0017\, cigs & + .060\, educ & + .058\, age & - .00063\, age^2 \\ & (0.17) & (.0017) & (.008) & (.008) & (.00008) \end{array}$$

$n = 807, \quad R^2 = .165.$

The coefficient on *cigs* implies that cigarette smoking causes income to increase, although the coefficient is not statistically different from zero. Remember, OLS ignores potential simultaneity between income and cigarette smoking.

(v) The estimated reduced form for *cigs* is

$$\widehat{cigs} = \underset{(23.70)}{1.58} - \underset{(.162)}{.450}\, educ + \underset{(.154)}{.823}\, age - \underset{(.0017)}{.0096}\, age^2 - \underset{(5.766)}{.351} \log(cigpric)$$
$$- \underset{(1.11)}{2.74}\, restaurn$$
$$n = 807, \quad R^2 = .051.$$

While log(*cigpric*) is very insignificant, *restaurn* had the expected negative sign and a *t* statistic of about –2.47. (People living in states with restaurant smoking restrictions smoke almost three fewer cigarettes, on average, given education and age.) We could drop log(*cigpric*) from the analysis but we leave it in. (Incidentally, the *F* test for joint significance of log(*cigpric*) and *restaurn* yields *p*-value $\approx$ .044.)

(vi) Estimating the log(*income*) equation by 2SLS gives

$$\widehat{\log(income)} = \underset{(0.23)}{7.78} - \underset{(.026)}{.042}\, cigs + \underset{(.016)}{.040}\, educ + \underset{(.023)}{.094}\, age - \underset{(.00027)}{.00105}\, age^2$$
$$n = 807.$$

Now the coefficient on *cigs* is negative and almost significant at the 10% level against a two-sided alternative. The estimated effect is very large: each additional cigarette someone smokes lowers predicted income by about 4.2%. Of course, the 95% CI for β_{cigs} is very wide.

(vii) Assuming that state level cigarette prices and restaurant smoking restrictions are exogenous in the income equation is problematical. Incomes are known to vary by region, as do restaurant smoking restrictions. It could be that in states where income is lower (after controlling for education and age), restaurant smoking restrictions are less likely to be in place.

C16.3 (i) The OLS estimates are

$$\widehat{inf} = \underset{(4.10)}{25.23} - \underset{(.093)}{.215}\, open$$
$$n = 114, \quad R^2 = .045.$$

The IV estimates are

$$\widehat{inf} = \underset{(5.66)}{29.61} - \underset{(.140)}{.333}\, open$$

$$n = 114, \quad R^2 = .032.$$

The OLS coefficient is the same, to three decimal places, when log(*pcinc*) is included in the model. The IV estimate with log(*pcinc*) in the equation is −.337, which is very close to −.333. Therefore, dropping log(*pcinc*) makes little difference.

(ii) Subject to the requirement that an IV be exogenous, we want an IV that is as highly correlated as possible with the endogenous explanatory variable. If we regress *open* on *land* we obtain R^2 = .095. The simple regression of *open* on log(*land*) gives R^2 = .448. Therefore, log(*land*) is much more highly correlated with *open*. Further, if we regress open on log(*land*) and *land* we get

$$\widehat{open} = \underset{(10.47)}{129.22} - \underset{(0.98)}{8.40}\log(land) + \underset{(.0000031)}{.0000043}\, land$$

$$n = 114, \quad R^2 = .457.$$

While log(*land*) is very significant, *land* is not, so we might as well use only log(*land*) as the IV for open.

(iii) When we add *oil* to the original model, and assume *oil* is exogenous, the IV estimates are

$$\widehat{inf} = \underset{(16.04)}{24.01} - \underset{(.144)}{.337}\, open + \underset{(2.12)}{.803}\log(pcinc) - \underset{(9.80)}{6.56}\, oil$$

$$n = 114, \quad R^2 = .035.$$

Being an oil producer is estimated to reduce average annual inflation by over 6.5 percentage points, but the effect is not statistically significant. This is not too surprising, as there are only seven oil producers in the sample.

C16.5 This is an open-ended question without a single answer. Even if we settle on extending the data through a particular year, we might want to change the disposable income and nondurable consumption numbers in earlier years, as these are often recalculated. For example, the value for real disposable personal income in 1995, as reported in Table B-29 of the 1997 *Economic Report of the President* (*ERP*), is $4,945.8 billions. In the 1999 *ERP*, this value has been changed to $4,906.0 billions (see Table B-31). All series can be updated using the latest edition of the *ERP*. The key is to use real values and make them per capita by dividing by population. Make sure that you use nondurable consumption.

C16.7 (i) If county administrators can predict when crime rates will increase, they may hire more police to counteract crime. This would explain the estimated positive relationship between $\Delta\log(crmrte)$ and $\Delta\log(polpc)$ in equation (13.33).

(ii) This may be reasonable, although tax collections depend in part on income and sales taxes, and revenues from these depend on the state of the economy, which can also influence crime rates.

(iii) The reduced form for $\Delta\log(polpc_{it})$, for each i and t, is

$$\begin{aligned}\Delta\log(polpc_{it}) \;=\;& \pi_0 + \pi_1 d83_t + \pi_2 d84_t + \pi_3 d85_t + \pi_4 d86_t + \pi_5 d87_t \\ &+ \pi_6\Delta\log(prbarr_{it}) + \pi_7\Delta\log(prbconv_{it}) + \pi_8\Delta\log(prbpris_{it}) \\ &+ \pi_9\Delta\log(avgsen_{it}) + \pi_{10}\Delta\log(taxpc_{it}) + v_{it}.\end{aligned}$$

We need $\pi_{10} \neq 0$ for $\Delta\log(taxpc_{it})$ to be a reasonable IV candidate for $\Delta\log(polpc_{it})$. When we estimate this equation by pooled OLS (N = 90, T = 6 for n = 540), we obtain $\hat{\pi}_{10}$ = .0052 with a t statistic of only .080. Therefore, $\Delta\log(taxpc_{it})$ is not a good IV for $\Delta\log(polpc_{it})$.

(iv) If the grants were awarded randomly, then the grant amounts, say $grant_{it}$ for the dollar amount for county i and year t, will be uncorrelated with Δu_{it}, the changes in unobservables that affect county crime rates. By definition, $grant_{it}$ should be correlated with $\Delta\log(polpc_{it})$ across i and t. This means we have an exogenous variable that can be omitted from the crime equation and that is (partially) correlated with the endogenous explanatory variable. We could reestimate (13.33) by IV.

C16.9 (i) The demand function should be downward sloping, so $\alpha_1 < 0$: as price increases, quantity demanded for air travel decreases.

(ii) The estimated price elasticity is –.391 (t statistic = –5.82).

(iii) We must assume that passenger demand depends only on air fare, so that, once price is controlled for, passengers are indifferent about the fraction of travel accounted for by the largest carrier.

(iv) The reduced form equation for $\log(fare)$ is

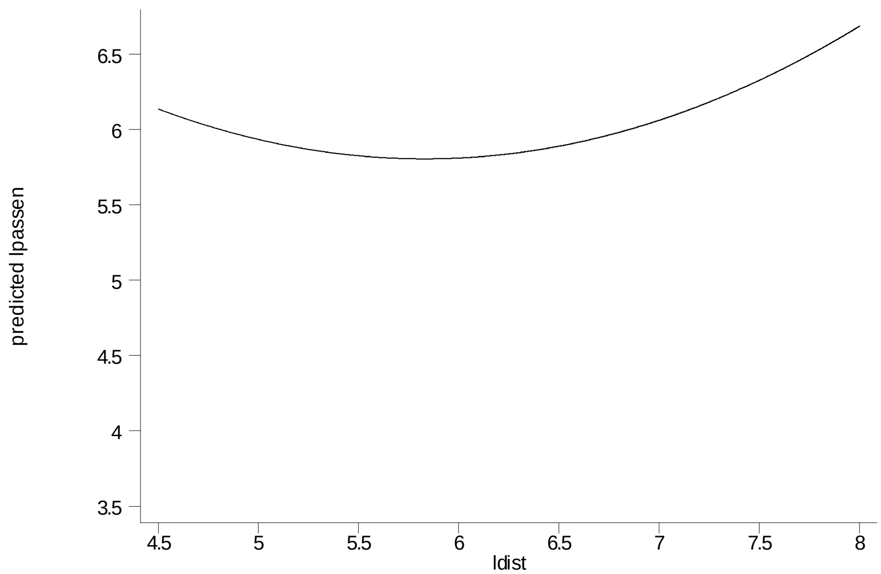
$$\begin{aligned}\widehat{\log(fare)} = \underset{(0.89)}{6.19} + \underset{(.063)}{.395}\, concen - \underset{(.272)}{.936}\log(dist) + \underset{(.021)}{.108}\,[\log(dist)]^2\end{aligned}$$

n = 1,149, R^2 = .408

The coefficient on *concen* shows a pretty strong link between concentration and fare. If *concen* increases by .10 (10 percentage points), *fare* is estimated to increase by almost 4%. The t statistic is about 6.3.

(v) Using *concen* as an IV for log(*fare*) [and where the distance variables act as their own IVs], the estimated price elasticity is -1.17, which shows much greater price sensitivity than did the OLS estimate. The IV estimate suggests that a one percent increase in fare leads to a slightly more than one percent increase drop in passenger demand. Of course, the standard error of the IV estimate is much larger (about .389 compared with the OLS standard error of .067), but the IV estimate is statistically significant (t is about -3.0).

(vi) The coefficient on *ldist* = log(*dist*) is about -2.176 and that on *ldistsq* = $[\log(dist)]^2$ is about .187. Therefore, the relationship between log(*passen*) and log(*dist*) has a U-shape, as given in the following graph:



The minimum is at about *ldist* = $2.176/.187 \approx 5.82$, which, in terms of distance, is about 337 miles. About 11.3% of the routes are less than 337 miles long. If the estimated quadratic is believable, the lowest demand occurs for short, but not very short, routes (holding price fixed). It is possible, of course, that we should ignore the quadratic to the left of the turning point, but it does contain a nontrivial fraction of the observations.

CHAPTER 17

SOLUTIONS TO PROBLEMS

17.1 (i) Let m_0 denote the number (not the percent) correctly predicted when $y_i = 0$ (so the prediction is also zero) and let m_1 be the number correctly predicted when $y_i = 1$. Then the proportion correctly predicted is $(m_0 + m_1)/n$, where n is the sample size. By simple algebra, we can write this as $(n_0/n)(m_0/n_0) + (n_1/n)(m_1/n_1) = (1 - \bar{y})(m_0/n_0) + \bar{y}(m_1/n_1)$, where we have used the fact that $\bar{y} = n_1/n$ (the proportion of the sample with $y_i = 1$) and $1 - \bar{y} = n_0/n$ (the proportion of the sample with $y_i = 0$). But m_0/n_0 is the proportion correctly predicted when $y_i = 0$, and m_1/n_1 is the proportion correctly predicted when $y_i = 1$. Therefore, we have

$$(m_0 + m_1)/n = (1 - \bar{y})(m_0/n_0) + \bar{y}(m_1/n_1).$$

If we multiply through by 100 we obtain

$$\hat{p} = (1 - \bar{y})\hat{q}_0 + \bar{y}\cdot\hat{q}_1,$$

where we use the fact that, by definition, $\hat{p} = 100[(m_0 + m_1)/n]$, $\hat{q}_0 = 100(m_0/n_0)$, and $\hat{q}_1 = 100(m_1/n_1)$.

(ii) We just use the formula from part (i): $\hat{p} = .30(80) + .70(40) = 52$. Therefore, overall we correctly predict only 52% of the outcomes. This is because, while 80% of the time we correctly predict $y = 0$, $y_i = 0$ accounts for only 30 percent of the outcomes. More weight (.70) is given to the predictions when $y_i = 1$, and we do much less well predicting that outcome (getting it right only 40% of the time).

17.3 (i) We use the chain rule and equation (17.23). In particular, let $x_1 \equiv \log(z_1)$. Then, by the chain rule,

$$\frac{\partial E(y|y>0,\mathbf{x})}{\partial z_1} = \frac{\partial E(y|y>0,\mathbf{x})}{\partial x_1}\cdot\frac{\partial x_1}{\partial z_1} = \frac{\partial E(y|y>0,\mathbf{x})}{\partial x_1}\cdot\frac{1}{z_1},$$

where we use the fact that the derivative of $\log(z_1)$ is $1/z_1$. When we plug in (17.23) for $\partial E(y|y > 0,\mathbf{x})/\partial x_1$, we obtain the answer.

(ii) As in part (i), we use the chain rule, which is now more complicated:

$$\frac{\partial E(y|y>0,\mathbf{x})}{\partial z_1} = \frac{\partial E(y|y>0,\mathbf{x})}{\partial x_1}\cdot\frac{\partial x_1}{\partial z_1} + \frac{\partial E(y|y>0,\mathbf{x})}{\partial x_2}\cdot\frac{\partial x_2}{\partial z_1},$$

where $x_1 = z_1$ and $x_2 = z_1^2$. But $\partial E(y|y > 0,\mathbf{x})/\partial x_1 = \beta_1\{1 - \lambda(\mathbf{x\beta}/\sigma)[\mathbf{x\beta}/\sigma + \lambda(\mathbf{x\beta}/\sigma)]\}$, $\partial E(y|y > 0,\mathbf{x})/\delta x_2 = \beta_2\{1 - \lambda(\mathbf{x\beta}/\sigma)[\mathbf{x\beta}/\sigma + \lambda(\mathbf{x\beta}/\sigma)]\}$, $\partial x_1/\partial z_1 = 1$, and $\partial x_2/\partial z_1 = 2z_1$. Plugging these into the first formula and rearranging gives the answer.

17.5 (i) *patents* is a count variable, and so the Poisson regression model is appropriate.

(ii) Because β_1 is the coefficient on log(*sales*), β_1 is the elasticity of *patents* with respect to *sales*. (More precisely, β_1 is the elasticity of E(*patents*|*sales*,*RD*) with respect to *sales*.)

(iii) We use the chain rule to obtain the partial derivative of $\exp[\beta_0 + \beta_1\log(sales) + \beta_2 RD + \beta_3 RD^2]$ with respect to *RD*:

$$\frac{\partial E(patents\,|\,sales, RD)}{\partial RD} = (\beta_2 + 2\beta_3 RD)\exp[\beta_0 + \beta_1\log(sales) + \beta_2 RD + \beta_3 RD^2].$$

A simpler way to interpret this model is to take the log and then differentiate with respect to *RD*: this gives $\beta_2 + 2\beta_3 RD$, which shows that the semi-elasticity of *patents* with respect to *RD* is $100(\beta_2 + 2\beta_3 RD)$.

17.7 For the immediate purpose of determining the variables that explain whether accepted applicants choose to enroll, there is not a sample selection problem. The population of interest is applicants accepted by the particular university, and you have a random sample from this population. Therefore, it is perfectly appropriate to specify a model for this group, probably a linear probability model, a probit model, or a logit model, and estimate the model using the data at hand. OLS or maximum likelihood estimation will produce consistent, asymptotically normal estimators. This is a good example of where many data analysts' knee-jerk reaction might be to conclude that there is a sample selection problem, which is why it is important to be very precise about the purpose of the analysis, which requires one to clearly state the population of interest.

If the university is hoping the applicant pool changes in the near future, then there is a potential sample selection problem: the current students that apply may be systematically different from students that may apply in the future. As the nature of the pool of applicants is unlikely to change dramatically over one year, the sample selection problem can be mitigated, if not entirely eliminated, by updating the analysis after each first-year class has enrolled.

SOLUTIONS TO COMPUTER EXERCISES

C17.1 (i) If *spread* is zero, there is no favorite, and the probability that the team we (arbitrarily) label the favorite should have a 50% chance of winning.

(ii) The linear probability model estimated by OLS gives

$$\widehat{favwin} = \underset{(.028)\;[.032]}{.577} + \underset{(.0023)\;[.0019]}{.0194}\, spread$$

$$n = 553,\quad R^2 = .111.$$

where the usual standard errors are in (·) and the heteroskedasticity-robust standard errors are in [·]. Using the usual standard error, the t statistic for H_0: $\beta_0 = .5$ is $(.577 - .5)/.028 = 2.75$, which leads to rejecting H_0 against a two-sided alternative at the 1% level (critical value ≈ 2.58). Using the robust standard error reduces the significance but nevertheless leads to strong rejection of H_0 at the 2% level against a two-sided alternative: $t = (.577 - .5)/.032 \approx 2.41$ (critical value ≈ 2.33).

(iii) As we expect, *spread* is very statistically significant using either standard error, with a t statistic greater than eight. If *spread* = 10 the estimated probability that the favored team wins is $.577 + .0194(10) = .771$.

(iv) The probit results are given in the following table:

Dependent Variable: *favwin*	
Independent Variable	Coefficient (Standard Error)
spread	.0925 (.0122)
constant	–.0106 (.1037)
Number of Observations	553
Log Likelihood Value	–263.56
Pseudo *R*-Squared	.129

In the probit model

$$P(favwin = 1|spread) = \Phi(\beta_0 + \beta_1 spread),$$

where $\Phi(\cdot)$ denotes the standard normal cdf, if $\beta_0 = 0$ then

$$P(favwin = 1|spread) = \Phi(\beta_1 spread)$$

and, in particular, $P(favwin = 1|spread = 0) = \Phi(0) = .5$. This is the analog of testing whether the intercept is .5 in the LPM. From the table, the t statistic for testing H_0: $\beta_0 = 0$ is only about -.102, so we do not reject H_0.

(v) When *spread* = 10 the predicted response probability from the estimated probit model is $\Phi[-.0106 + .0925(10)] = \Phi(.9144) \approx .820$. This is somewhat above the estimate for the LPM.

(vi) When *favhome*, *fav25*, and *und25* are added to the probit model, the value of the log-likelihood becomes –262.64. Therefore, the likelihood ratio statistic is 2[–262.64 – (–263.56)] = 2(263.56 – 262.64) = 1.84. The *p*-value from the χ_3^2 distribution is about .61, so *favhome*, *fav25*, and *und25* are jointly very insignificant. Once *spread* is controlled for, these other factors have no additional power for predicting the outcome.

C17.3 (i) Out of 616 workers, 172, or about 18%, have zero pension benefits. For the 444 workers reporting positive pension benefits, the range is from \$7.28 to \$2,880.27. Therefore, we have a nontrivial fraction of the sample with $pension_t = 0$, and the range of positive pension benefits is fairly wide. The Tobit model is well-suited to this kind of dependent variable.

(ii) The Tobit results are given in the following table:

Dependent Variable: *pension*		
Independent Variable	(1)	(2)
exper	5.20 (6.01)	4.39 (5.83)
age	–4.64 (5.71)	–1.65 (5.56)
tenure	36.02 (4.56)	28.78 (4.50)
educ	93.21 (10.89)	106.83 (10.77)
depends	(35.28 (21.92)	41.47 (21.21)
married	(53.69 (71.73)	19.75 (69.50)
white	144.09 (102.08)	159.30 (98.97)
male	308.15 (69.89)	257.25 (68.02)
union	———	439.05 (62.49)
constant	–1,252.43 (219.07)	–1,571.51 (218.54)
Number of Observations	616	616
Log Likelihood Value	–3,672.96	–3648.55
$\hat{\sigma}$	677.74	652.90

In column (1), which does not control for *union*, being white or male (or, of course, both) increases predicted pension benefits, although only *male* is statistically significant ($t \approx 4.41$).

(iii) We use equation (17.22) with *exper* = *tenure* = 10, *age* = 35, *educ* = 16, *depends* = 0, *married* = 0, *white* = 1, and *male* = 1 to estimate the expected benefit for a white male with the given characteristics. Using our shorthand, we have

$$\mathbf{x}\hat{\beta} = -1{,}252.5 + 5.20(10) - 4.64(35) + 36.02(10) + 93.21(16) + 144.09 + 308.15 = 940.90.$$

Therefore, with $\hat{\sigma}$ = 677.74 we estimate E(*pension*|**x**) as

$$\Phi(940.9/677.74)\cdot(940.9) + (677.74)\cdot\phi(940.9/677.74) \approx 966.40.$$

For a nonwhite female with the same characteristics,

$$\mathbf{x}\hat{\beta} = -1{,}252.5 + 5.20(10) - 4.64(35) + 36.02(10) + 93.21(16) = 488.66.$$

Therefore, her predicted pension benefit is

$$\Phi(488.66/677.74)\cdot(488.66) + (677.74)\cdot\phi(488.66/677.74) \approx 582.10.$$

The difference between the white male and nonwhite female is 966.40 – 582.10 = $384.30.

(iv) Column (2) in the previous table gives the results with *union* added. The coefficient is large, but to see exactly how large, we should use equation (17.22) to estimate E(*pension*|**x**) with *union* = 1 and *union* = 0, setting the other explanatory variables at interesting values. The t statistic on *union* is over seven.

(v) When *peratio* is used as the dependent variable in the Tobit model, *white* and *male* are individually and jointly insignificant. The p-value for the test of joint significance is about .74. Therefore, neither whites nor males seem to have different tastes for pension benefits as a fraction of earnings. White males have higher pension benefits because they have, on average, higher earnings.

C17.5 (i) The Poisson regression results are given in the following table:

Dependent Variable: *kids*		
Independent Variable	Coefficient	Standard Error
educ	–.048	.007
age	.204	.055
age^2	–.0022	.0006
black	.360	.061
east	.088	.053
northcen	.142	.048
west	.080	.066
farm	–.015	.058
othrural	–.057	.069
town	.031	.049
smcity	.074	.062
y74	.093	.063
y76	–.029	.068
y78	–.016	.069
y80	–.020	.069
y82	–.193	.067
y84	–.214	.069
constant	–3.060	1.211
n = 1,129		
$\mathcal{L}$ = –2,070.23		
$\hat{\sigma}$ = .944		

The coefficient on *y82* means that, other factors in the model fixed, a woman's fertility was about 19.3% lower in 1982 than in 1972.

(ii) Because the coefficient on *black* is so large, we obtain the estimated proportionate difference as exp(.36) – 1 ≈ .433, so a black woman has 43.3% more children than a comparable nonblack woman. (Notice also that *black* is very statistically significant.)

(iii) From the above table, $\hat{\sigma}$ = .944, which shows that there is actually underdispersion in the estimated model.

(iv) The sample correlation between $kids_i$ and $\widehat{kids}_i$ is about .348, which means the *R*-squared (or, at least one version of it), is about $(.348)^2 \approx .121$. Interestingly, this is actually smaller than the R-squared for the linear model estimated by OLS. (However, remember that OLS obtains the highest possible R-squared for a linear model, while Poisson regression does not obtain the highest possible R-squared for an exponential regression model.)

C17.7 (i) When log(*wage*) is regressed on *educ*, *exper*, $exper^2$, *nwifeinc*, *age*, *kidslt6*, and *kidsge6*, the coefficient and standard error on *educ* are .0999 (se = .0151).

(ii) The Heckit coefficient on *educ* is .1187 (se = .0341), where the standard error is just the usual OLS standard error. The estimated return to education is somewhat larger than without the Heckit corrections, but the Heckit standard error is over twice as large.

(iii) Regressing $\hat{\lambda}$ on *educ, exper, exper*^{2}*, nwifeinc, age, kidslt6,* and *kidsge6* (using only the selected sample of 428) produces $R^2 \approx .962$, which means that there is substantial multicollinearity among the regressors in the second stage regression. This is what leads to the large standard errors. Without an exclusion restriction in the log(*wage*) equation, $\hat{\lambda}$ is almost a linear function of the other explanatory variables in the sample.

C17.9 (i) 248.

(ii) The distribution is not continuous: there are clear focal points, and rounding. For example, many more people report one pound than either two-thirds of a pound or 1 1/3 pounds. This violates the latent variable formulation underlying the Tobit model, where the latent error has a normal distribution. Nevertheless, we should view Tobit in this context as a way to possibly improve functional form. It may work better than the linear model for estimating the expected demand function.

(ii) The following table contains the Tobit estimates and, for later comparison, OLS estimates of a linear model:

Dependent Variable: *ecolbs*		
Independent Variable	Tobit	OLS (Linear Model)
ecoprc	–5.82 (.89)	–2.90 (.59)
regprc	5.66 (1.06)	3.03 (.71)
faminc	.0066 (.0040)	.0028 (.0027)
hhsize	.130 (.095)	.054 (.064)
constant	1.00 (.67)	1.63 (.45)
Number of Observations	660	660
Log Likelihood Value	–1,266.44	———
$\hat{\sigma}$	3.44	2.48
R-squared	.0369	.0393

Only the price variables, *ecoprc* and *regprc*, are statistically significant at the 1% level.

(iv) The signs of the price coefficients accord with basic demand theory: the own-price effect is negative, the cross price effect for the substitute good (regular apples) is positive.

(v) The null hypothesis can be stated as H_0: $\beta_1 + \beta_2 = 0$. Define $\theta_1 = \beta_1 + \beta_2$. Then $\hat{\theta}_1 =$ –.16. To obtain the *t* statistic, I write $\beta_2 = \theta_1 - \beta_1$, plug in, and rearrange. This results in doing Tobit of *ecolbs* on (*ecoprc* – *regprc*), *regprc*, *faminc*, and *hhsize*. The coefficient on *regprc* is $\hat{\theta}_1$ and, of course we get its standard error: about .59. Therefore, the *t* statistic is about –.27 and *p*-value = .78. We do not reject the null.

(vi) The smallest fitted value is .798, while the largest is 3.327.

(vii) The squared correlation between $ecolbs_i$ and $\widehat{ecolbs_i}$ is about .0369. This is one possible *R*-squared measure.

(viii) The linear model estimates are given in the table for part (ii). The OLS estimates are smaller than the Tobit estimates because the OLS estimates are estimated partial effects on E(*ecolbs*|**x**), whereas the Tobit coefficients must be scaled by the term in equation (17.27). The scaling factor is always between zero and one, and often substantially less than one. The Tobit model does not fit better, at least in terms of estimating E(*ecolbs*|**x**): the linear model R-squared is a bit larger (.0393 versus .0369).

(ix) This is not a correct statement. We have another case where we have confidence in the ceteris paribus price effects (because the price variables are exogenously set), yet we cannot explain much of the variation in *ecolbs*. The fact that demand for a fictitious product is hard to explain is not very surprising.

C17.11 (i) The fraction of women in the work force is 3,286/5,634 ≈ .583.

(ii) The OLS results using the selected sample are

$$\widehat{\log(wage)} = \underset{(.060)}{.649} + \underset{(.004)}{.099}\, educ + \underset{(.003)}{.020}\, exper - \underset{(.00008)}{.00035}\, exper^2$$

$$- \underset{(.034)}{.030}\, black + \underset{(.036)}{.014}\, hispanic$$

$n = 3{,}286,\;\; R^2 = .205$

While the point estimates imply blacks earn, on average, about 3% less and Hispanics about 1.3% more than the base group (non-black, non-Hispanic), neither coefficient is statistically significant – or even very close to statistical significance at the usual levels. The joint F test gives a p-value of about .63. So, there is little evidence for differences by race and ethnicity once education and experience have been controlled for.

(iii) The coefficient on *nwifeinc* is –.0091 with t = –13.47 and the coefficient on *kidlt*6 is –.500 with t = –11.05. We expect both coefficients to be negative. If a woman's spouse earns more, she is less likely to work. Having a young child in the family also reduces the probability that the woman works. Each variable is very statistically significant. (Not surprisingly, the joint test also yields a p-value of essentially zero.)

(iv) We need at least one variable to affect labor force participation that does not have a direct effect on the wage offer. So, we must assume that, controlling for education, experience, and the race/ethnicity variables, other income and the presence of a young children do not affect wage. These propositions could be false if, say, employers discriminate against women who have young children or whose husbands work. Further, if having a young child reduces productivity – through, say, having to take time off for sick children and appointments – then it would be inappropriate to exclude *kidlt6* from the wage equation.

(v) The t statistic on the inverse Mills ratio is 1.77 and the p-value against the two-sided alternative is .077. With 3,286 observations, this is not a very small p-value. The test on $\hat{\lambda}$ does not provide strong evidence against the null hypothesis of no selection bias.

(vi) Just as important, the slope coefficients do not change much when the inverse Mills ratio is added. For example, the coefficient on *educ* increases from .099 to .103 – a change within the 95% confidence interval for the original OLS estimate. [The 95% CI is (.092,.106.)]. The changes on the experience coefficients are also pretty small; the Heckman estimates are well within the 95% confidence intervals of the OLS estimates. Superficially, the *black* and *hispanic* coefficients change by larger amounts, but these estimates are statistically insignificant. Based on the wide confidence intervals, we expect rather wide changes in the estimates to even minor changes in the specification.

The most substantial change is in the intercept estimate – from .649 to .539 – but it is hard to know what to make of this. Remember, in this example, the intercept is the estimated value of log(*wage*) for a non-black, non-Hispanic woman with zero years of education and experience. No one in the full sample even comes close to this description. Because the slope coefficients do change somewhat, we cannot say that the Heckman estimates imply a lower average wage offer than the uncorrected estimates. Even if this were true, the estimated marginal effects of the explanatory variables are hardly affected.

C17.13 (i) Using the entire sample, the estimated coefficient on *educ* is .1037 with standard error = .0097.

(ii) 166 observations are lost when we restrict attention to the sample with *educ* < 16. This is about 13.5% of the original sample. The coefficient on *educ* becomes .1182 with standard error = .0126. This is a slight increase in the estimated return to education, and it is estimated less precisely (because we have reduced the sample variation in *educ*).

(iii) If we restrict attention to those with *wage* < 20, we lose 164 observations [about the same number in part (ii)]. But now the coefficient on *educ* is much smaller, .0579, with standard error = .0093.

(iv) If we use the sample in part (iii) but account for the known truncation point, log(20), the coefficient on *educ* is .1060 (standard error = .0168). This is very close to the estimate on the original sample. We obtain a less precise estimate because we have dropped 13.3% of the original sample.

CHAPTER 18

SOLUTIONS TO PROBLEMS

18.1 With z_{t1} and z_{t2} now in the model, we should use one lag each as instrumental variables, $z_{t-1,1}$ and $z_{t-1,2}$. This gives one overidentifying restriction that can be tested.

18.3 For $\delta \neq \beta$, $y_t - \delta z_t = y_t - \beta z_t + (\beta - \delta)z_t$, which is an I(0) sequence ($y_t - \beta z_t$) plus an I(1) sequence. Since an I(1) sequence has a growing variance, it dominates the I(0) part, and the resulting sum is an I(1) sequence.

18.5 Following the hint, we have

$$y_t - y_{t-1} = \beta x_t - \beta x_{t-1} + \beta x_{t-1} - y_{t-1} + u_t$$

or

$$\Delta y_t = \beta \Delta x_t - (y_{t-1} - \beta x_{t-1}) + u_t.$$

Next, we plug in $\Delta x_t = \gamma \Delta x_{t-1} + v_t$ to get

$$\begin{aligned} \Delta y_t &= \beta(\gamma \Delta x_{t-1} + v_t) - (y_{t-1} - \beta x_{t-1}) + u_t \\ &= \beta\gamma \Delta x_{t-1} - (y_{t-1} - \beta x_{t-1}) + u_t + \beta v_t \\ &\equiv \gamma_1 \Delta x_{t-1} + \delta(y_{t-1} - \beta x_{t-1}) + e_t, \end{aligned}$$

where $\gamma_1 = \beta\gamma$, $\delta = -1$, and $e_t = u_t + \beta v_t$.

18.7 If $unem_t$ follows a stable AR(1) process, then this is the null model used to test for Granger causality: under the null that gM_t does not Granger cause $unem_t$, we can write

$$unem_t = \beta_0 + \beta_1 unem_{t-1} + u_t$$

$$\mathrm{E}(u_t | unem_{t-1}, gM_{t-1}, unem_{t-2}, gM_{t-2}, \ldots) = 0$$

and $|\beta_1| < 1$. Now, it is up to us to choose how many lags of gM to add to this equation. The simplest approach is to add gM_{t-1} and to do a t test. But we could add a second or third lag (and probably not beyond this with annual data), and compute an F test for joint significance of all lags of gM_t.

18.9 Let $\hat{e}_{n+1}$ be the forecast error for forecasting y_{n+1}, and let $\hat{a}_{n+1}$ be the forecast error for forecasting Δy_{n+1}. By definition, $\hat{e}_{n+1} = y_{n+1} - \hat{f}_n = y_{n+1} - (\hat{g}_n + y_n) = (y_{n+1} - y_n) - \hat{g}_n = \Delta y_{n+1} - \hat{g}_n = \hat{a}_{n+1}$, where the last equality follows by definition of the forecasting error for Δy_{n+1}.

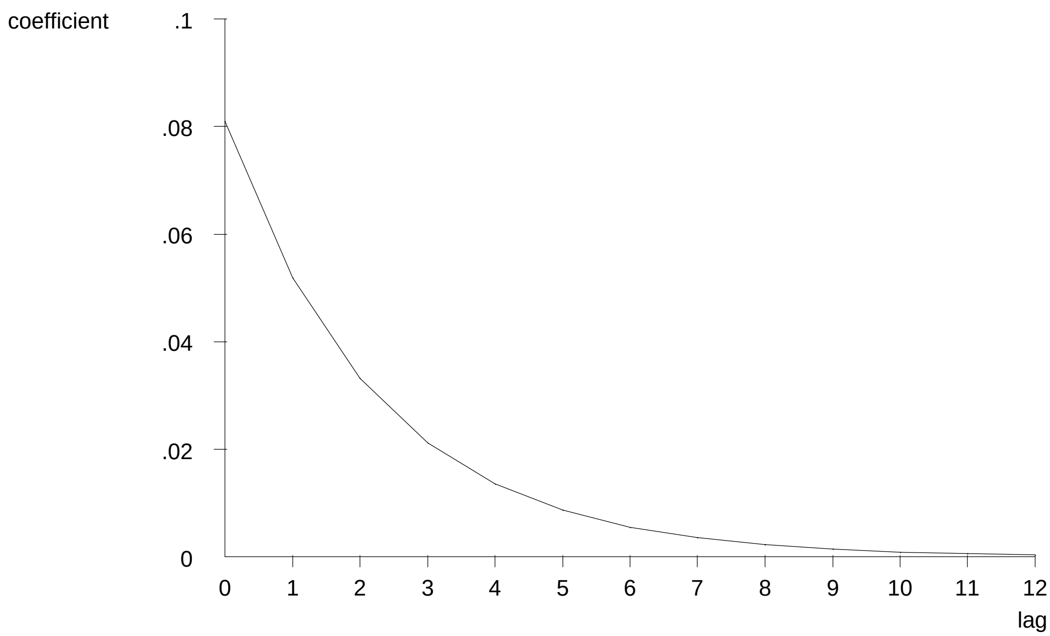
SOLUTIONS TO COMPUTER EXERCISES

C18.1 (i) The estimated GDL model is

$$\widehat{gprice} = \underset{(.0003)}{.0013} + \underset{(.031)}{.081}\, gwage + \underset{(.045)}{.640}\, gprice_{-1}$$

$$n = 284, \quad R^2 = .454.$$

The estimated impact propensity is .081 while the estimated LRP is .081/(1 – .640) = .225. The estimated lag distribution is graphed below.



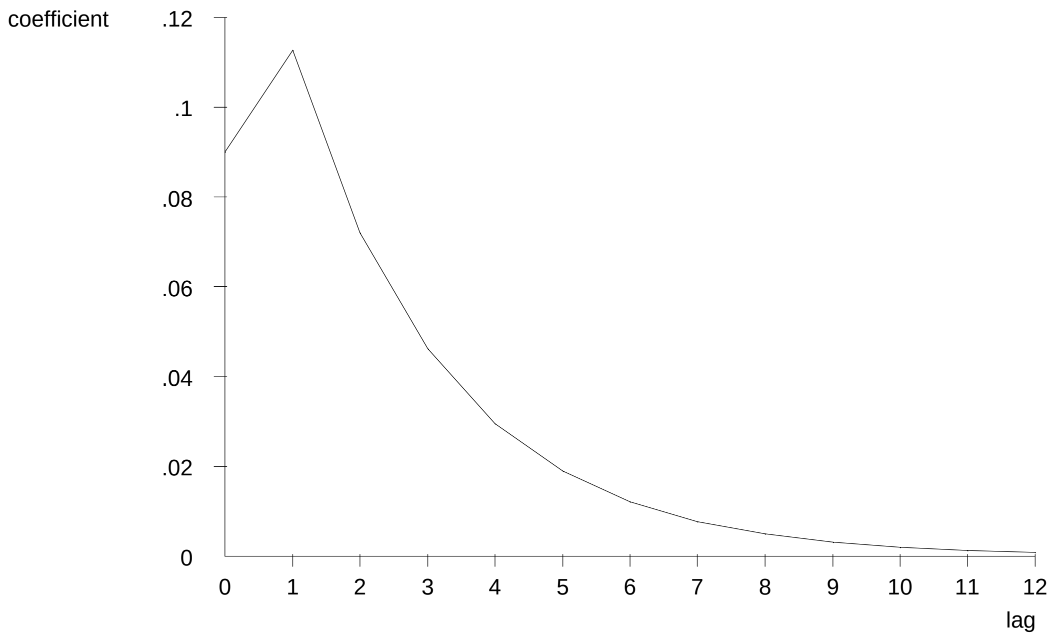
(ii) The IP for the FDL model estimated in Problem 11.5 was .119, which is substantially above the estimated IP for the GDL model. Further, the estimated LRP from GDL model is much lower than that for the FDL model, which we estimated as 1.172. Clearly we cannot think of the GDL model as a good approximation to the FDL model. One reason these are so different can be seen by comparing the estimated lag distributions (see below for the GDL model). With the FDL, the largest lag coefficient is at the ninth lag, which is impossible with the GDL model (where the largest impact is always at lag zero). It could also be that $\{u_t\}$ in equation (18.8) does not follow an AR(1) process with parameter ρ, which would cause the dynamic regression to produce inconsistent estimators of the lag coefficients.

(iii) When we estimate the RDL from equation (18.16) we obtain

$$\widehat{gprice} = \underset{(.0003)}{.0011} + \underset{(.031)}{.090}\, gwage + \underset{(.046)}{.619}\, gprice_{-1} + \underset{(.032)}{.055}\, gwage_{-1}$$

$$n = 284, \quad R^2 = .460.$$

The coefficient on $gwage_{-1}$ is not especially significant but we include it in obtaining the estimated LRP. The estimated IP is .09 while the LRP is $(.090 + .055)/(1 - .619) \approx .381$. These are both slightly higher than what we obtained for the GDL, but the LRP is still well below what we obtained for the FDL in Problem 11.5. While this RDL model is more flexible than the GDL model, it imposes a maximum lag coefficient (in absolute value) at lag zero or one. For the estimates given above, the maximum effect is at the first lag. (See the estimated lag distribution below.) This is not consistent with the FDL estimates in Problem 11.5.



C18.3 (i) The estimated AR(3) model for $pcip_t$ is

$$\widehat{pcip}_t = \underset{(0.55)}{1.80} + \underset{(.043)}{.349}\, pcip_{t-1} + \underset{(.045)}{.071}\, pcip_{t-2} + \underset{(.043)}{.067}\, pcip_{t-2}$$

$$n = 554, \quad R^2 = .166, \quad \hat{\sigma} = 12.15.$$

When $pcip_{t-4}$ is added, its coefficient is .0043 with a t statistic of about .10.

(ii) In the model

$$pcip_t = \delta_0 + \alpha_1 pcip_{t-1} + \alpha_2 pcip_{t-2} + \alpha_3 pcip_{t-3} + \gamma_1 pcsp_{t-1} + \gamma_2 pcsp_{t-2} + \gamma_3 pcsp_{t-3} + u_t,$$

The null hypothesis is that *pcsp* does not Granger cause *pcip*. This is stated as H_0: $\gamma_1 = \gamma_2 = \gamma_3 = 0$. The F statistic for joint significance of the three lags of $pcsp_t$, with 3 and 547 *df*, is $F = 5.37$ and p-value = .0012. Therefore, we strongly reject H_0 and conclude that *pcsp* does Granger cause *pcip*.

(iii) When we add $\Delta i3_{t-1}$, $\Delta i3_{t-2}$, and $\Delta i3_{t-3}$ to the regression from part (ii), and now test the joint significance of $pcsp_{t-1}$, $pcsp_{t-2}$, and $pcsp_{t-3}$, the F statistic is 5.08. With 3 and 544 *df* in the F distribution, this gives p-value = .0018, and so *pcsp* Granger causes *pcip* even conditional on past $\Delta i3$.

C18.5 (i) The estimated equation is

$$\widehat{hy6}_t = \underset{(.028)}{.078} + \underset{(0.016)}{1.027}\, hy3_{t-1} - \underset{(0.038)}{1.021}\, \Delta hy3_t - \underset{(.037)}{.085}\, \Delta hy3_{t-1} - \underset{(.037)}{.104}\, \Delta hy3_{t-2}$$

$$n = 121, \quad R^2 = .982, \quad \hat{\sigma} = .123.$$

The t statistic for H_0: $\beta = 1$ is $(1.027 - 1)/.016 \approx 1.69$. We do not reject H_0: $\beta = 1$ at the 5% level against a two-sided alternative, although we would reject at the 10% level.

(ii) The estimated error correction model is

$$\widehat{hy6}_t = \underset{(.049)}{.070} + \underset{(.278)}{1.259}\, \Delta hy3_{t-1} - \underset{(.256)}{.816}\, (hy6_{t-1} - hy3_{t-2})$$
$$+ \underset{(.272)}{.283}\, \Delta hy3_{t-2} + \underset{(.256)}{.127}\, (hy6_{t-2} - hy3_{t-3})$$

$$n = 121, \quad R^2 = .795.$$

Neither of the added terms is individually significant. The F test for their joint significance gives $F = 1.35$, p-value = .264. Therefore, we would omit these terms and stick with the error correction model estimated in (18.39).

C18.7 (i) The estimated linear trend equation using the first 119 observations and excluding the last 12 months is

$$\widehat{chnimp}_t = \underset{(53.20)}{248.58} + \underset{(0.77)}{5.15}\, t$$

$$n = 119, \quad R^2 = .277, \quad \hat{\sigma} = 288.33.$$

The standard error of the regression is 288.33.

(ii) The estimated AR(1) model excluding the last 12 months is

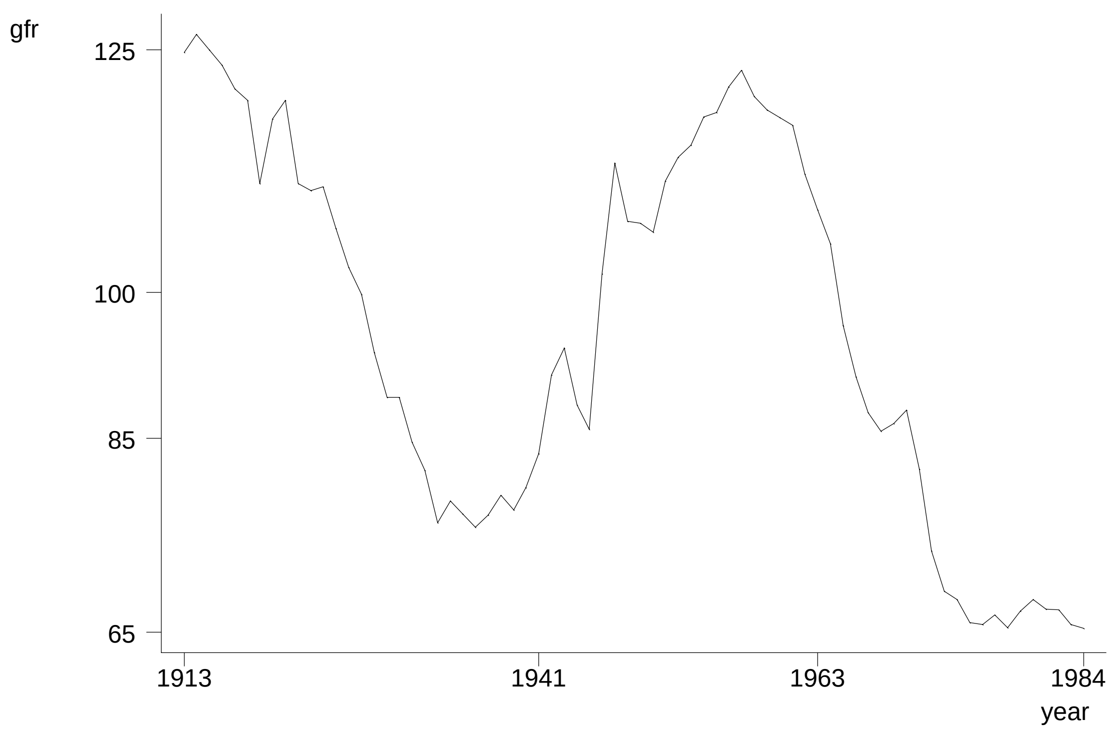
$$\widehat{chnimp_t} = \underset{(54.71)}{329.18} + \underset{(.084)}{.416}\, chnimp_{t\text{-}1}$$

$$n = 118,\quad R^2 = .174,\quad \hat{\sigma} = 308.17.$$

Because $\hat{\sigma}$ is lower for the linear trend model, it provides the better in-sample fit.

(iii) Using the last 12 observations for one-step-ahead out-of-sample forecasting gives an RMSE and MAE for the linear trend equation of about 315.5 and 201.9, respectively. For the AR(1) model, the RMSE and MAE are about 388.6 and 246.1, respectively. In this case, the linear trend is the better forecasting model.

(iv) Using again the first 119 observations, the F statistic for joint significance of feb_t, mar_t, …, dec_t when added to the linear trend model is about 1.15 with p-value $\approx$ .328. (The df are 11 and 107.) So there is no evidence that seasonality needs to be accounted for in forecasting *chnimp*.



C18.9 (i) Using the data up through 1989 gives

$$\hat{y}_t = \underset{(1{,}163.09)}{3{,}186.04} + \underset{(46.31)}{116.24}\, t + \underset{(.148)}{.630}\, y_{t-1}$$

$$n = 30, \quad R^2 = .994, \quad \hat{\sigma} = 223.95.$$

(Notice how high the R-squared is. However, it is meaningless as a goodness-of-fit measure because $\{y_t\}$ has a trend, and possibly a unit root.)

(ii) The forecast for 1990 (t = 32) is 3,186.04 + 116.24(32) + .630(17,804.09) $\approx$ 18,122.30, because y is \$17,804.09 in 1989. The actual value for real per capita disposable income was \$17,944.64, and so the forecast error is –\$177.66.

(iii) The MAE for the 1990s, using the model estimated in part (i), is about 371.76.

(iv) Without y_{t-1} in the equation, we obtain

$$\hat{y}_t = \underset{(103.38)}{8{,}143.11} + \underset{(5.64)}{311.26}\, t$$

$$n = 31, \quad R^2 = .991, \quad \hat{\sigma} = 280.87.$$

The MAE for the forecasts in the 1990s is about 718.26. This is much higher than for the model with y_{t-1}, so we should use the AR(1) model with a linear time trend.

C18.11 (i) For *lsp*500, the ADF statistic without a trend is t = –.79; with a trend, the t statistic is –2.20. This are both well above their respective 10% critical values. In addition, the estimated roots are quite close to one. For *lip*, the ADF statistic without a trend is –1.37 without a trend and –2.52 with a trend. Again, these are not close to rejecting even at the 10% levels, and the estimated roots are very close to one.

(ii) The simple regression of *lsp*500 on *lip* gives

$$\widehat{lsp500} = \underset{(.095)}{-2.402} + \underset{(.024)}{1.694}\, lip$$

$$n = 558, R^2 = .903$$

The t statistic for *lip* is over 70, and the R-squared is over .90. These are hallmarks of spurious regressions.

(iii) Using the residuals $\hat{u}_t$ obtained in part (ii), the ADF statistic (with two lagged changes) is –1.57, and the estimated root is over .99. There is no evidence of cointegration. (The 10% critical value is –3.04.)

(iv) After adding a linear time trend to the regression from part (ii), the ADF statistic applied to the residuals is −1.88, and the estimated root is again about .99. Even with a time trend there is no evidence of cointegration.

(v) It appears that *lsp*500 and *lip* do not move together in the sense of cointegration, even if we allow them to have unrestricted linear time trends. The analysis does not point to a long-run equilibrium relationship.

C18.13 (i) The DF statistic is about −3.31, which is to the left of the 2.5% critical value (−3.12), and so, using this test, we can reject a unit root at the 2.5% level. (The estimated root is about .81.)

(ii) When two lagged changes are added to the regression in part (i), the *t* statistic becomes −1.50, and the root is larger (about .915). Now, there is little evidence against a unit root.

(iii) If we add a time trend to the regression in part (ii), the ADF statistic becomes −3.67, and the estimated root is about .57. The 2.5% critical value is −3.66, and so we are back to fairly convincingly rejecting a unit root.

(iv) The best characterization seems to be an I(0) process about a linear trend. In fact, a stable AR(3) about a linear trend is suggested by the regression in part (iii).

(v) For $prcfat_t$, the ADF statistic without a trend is −4.74 (estimated root = .62) and with a time trend the statistic is −5.29 (estimated root = .54). Here, the evidence is strongly in favor of an I(0) process whether or not we include a trend.

APPENDIX B

SOLUTIONS TO PROBLEMS

B.1 Before the student takes the SAT exam, we do not know – nor can we predict with certainty – what the score will be. The actual score depends on numerous factors, many of which we cannot even list, let alone know ahead of time. (The student's innate ability, how the student feels on exam day, and which particular questions were asked, are just a few.) The eventual SAT score clearly satisfies the requirements of a random variable.

B.3 (i) Let Y_{it} be the binary variable equal to one if fund i outperforms the market in year t. By assumption, $P(Y_{it} = 1) = .5$ (a 50-50 chance of outperforming the market for each fund in each year). Now, for any fund, we are also assuming that performance relative to the market is independent across years. But then the probability that fund i outperforms the market in all 10 years, $P(Y_{i1} = 1, Y_{i2} = 1, \ldots, Y_{i,10} = 1)$, is just the product of the probabilities: $P(Y_{i1} = 1) \cdot P(Y_{i2} = 1) \ldots P(Y_{i,10} = 1) = (.5)^{10} = 1/1024$ (which is slightly less than .001). In fact, if we define a binary random variable Y_i such that $Y_i = 1$ if and only if fund i outperformed the market in all 10 years, then $P(Y_i = 1) = 1/1024$.

(ii) Let X denote the number of funds out of 4,170 that outperform the market in all 10 years. Then $X = Y_1 + Y_2 + \ldots + Y_{4,170}$. If we assume that performance relative to the market is independent across funds, then X has the Binomial (n,θ) distribution with $n = 4,170$ and $\theta = 1/1024$. We want to compute $P(X \geq 1) = 1 - P(X = 0) = 1 - P(Y_1 = 0, Y_2 = 0, \ldots, Y_{4,170} = 0) = 1 - P(Y_1 = 0) \cdot P(Y_2 = 0) \cdots P(Y_{4,170} = 0) = 1 - (1023/1024)^{4170} \approx .983$. This means, if performance relative to the market is random and independent across funds, it is almost certain that at least one fund will outperform the market in all 10 years.

(iii) Using the Stata command Binomial(4170,5,1/1024), the answer is about .385. So there is a nontrivial chance that at least five funds will outperform the market in all 10 years.

B.5 (i) As stated in the hint, if X is the number of jurors convinced of Simpson's innocence, then $X \sim \text{Binomial}(12,.20)$. We want $P(X \geq 1) = 1 - P(X = 0) = 1 - (.8)^{12} \approx .931$.

(ii) Above, we computed $P(X = 0)$ as about .069. We need $P(X = 1)$, which we obtain from (B.14) with $n = 12$, $\theta = .2$, and $x = 1$: $P(X = 1) = 12 \cdot (.2)(.8)^{11} \approx .206$. Therefore, $P(X \geq 2) \approx 1 - (.069 + .206) = .725$, so there is almost a three in four chance that the jury had at least two members convinced of Simpson's innocence prior to the trial.

B.7 In eight attempts the expected number of free throws is 8(.74) = 5.92, or about six free throws.

B.9 If Y is salary in dollars then $Y = 1000 \cdot X$, and so the expected value of Y is 1,000 times the expected value of X, and the standard deviation of Y is 1,000 times the standard deviation of X. Therefore, the expected value and standard deviation of salary, measured in dollars, are \$52,300 and \$14,600, respectively.

APPENDIX C

SOLUTIONS TO PROBLEMS

C.1 (i) This is just a special case of what we covered in the text, with $n = 4$: $E(\bar{Y}) = \mu$ and $\text{Var}(\bar{Y}) = \sigma^2/4$.

(ii) $E(W) = E(Y_1)/8 + E(Y_2)/8 + E(Y_3)/4 + E(Y_4)/2 = \mu[(1/8) + (1/8) + (1/4) + (1/2)] = \mu(1 + 1 + 2 + 4)/8 = \mu$, which shows that W is unbiased. Because the Y_i are independent,

$$\begin{aligned}\text{Var}(W) &= \text{Var}(Y_1)/64 + \text{Var}(Y_2)/64 + \text{Var}(Y_3)/16 + \text{Var}(Y_4)/4 \\ &= \sigma^2[(1/64) + (1/64) + (4/64) + (16/64)] = \sigma^2(22/64) = \sigma^2(11/32).\end{aligned}$$

(iii) Because $11/32 > 8/32 = 1/4$, $\text{Var}(W) > \text{Var}(\bar{Y})$ for any $\sigma^2 > 0$, so $\bar{Y}$ is preferred to W because each is unbiased.

C.3 (i) $E(W_1) = [(n-1)/n]E(\bar{Y}) = [(n-1)/n]\mu$, and so $\text{Bias}(W_1) = [(n-1)/n]\mu - \mu = -\mu/n$. Similarly, $E(W_2) = E(\bar{Y})/2 = \mu/2$, and so $\text{Bias}(W_2) = \mu/2 - \mu = -\mu/2$. The bias in W_1 tends to zero as $n \to \infty$, while the bias in W_2 is $-\mu/2$ for all n. This is an important difference.

(ii) $\text{plim}(W_1) = \text{plim}[(n-1)/n]\cdot\text{plim}(\bar{Y}) = 1\cdot\mu = \mu$. $\text{plim}(W_2) = \text{plim}(\bar{Y})/2 = \mu/2$. Because $\text{plim}(W_1) = \mu$ and $\text{plim}(W_2) = \mu/2$, W_1 is consistent whereas W_2 is inconsistent.

(iii) $\text{Var}(W_1) = [(n-1)/n]^2\text{Var}(\bar{Y}) = [(n-1)^2/n^3]\sigma^2$ and $\text{Var}(W_2) = \text{Var}(\bar{Y})/4 = \sigma^2/(4n)$.

(iv) Because $\bar{Y}$ is unbiased, its mean squared error is simply its variance. On the other hand, $\text{MSE}(W_1) = \text{Var}(W_1) + [\text{Bias}(W_1)]^2 = [(n-1)^2/n^3]\sigma^2 + \mu^2/n^2$. When $\mu = 0$, $\text{MSE}(W_1) = \text{Var}(W_1) = [(n-1)^2/n^3]\sigma^2 < \sigma^2/n = \text{Var}(\bar{Y})$ because $(n-1)/n < 1$. Therefore, $\text{MSE}(W_1)$ is smaller than $\text{Var}(\bar{Y})$ for μ close to zero. For large n, the difference between the two estimators is trivial.

C.5 (i) While the expected value of the numerator of G is $E(\bar{Y}) = \theta$, and the expected value of the denominator is $E(1 - \bar{Y}) = 1 - \theta$, the expected value of the ratio is not the ratio of the expected value.

(ii) By Property PLIM.2(iii), the plim of the ratio is the ratio of the plims (provided the plim of the denominator is not zero): $\text{plim}(G) = \text{plim}[\bar{Y}/(1 - \bar{Y})] = \text{plim}(\bar{Y})/[1 - \text{plim}(\bar{Y})] = \theta/(1 - \theta) = \gamma$.

C.7 (i) The average increase in wage is $\bar{d} = .24$, or 24 cents. The sample standard deviation is about .451, and so, with $n = 15$, the standard error of $\bar{d}$ is $.451\sqrt{15} \approx .1164$. From Table G.2, the 97.5th percentile in the t_{14} distribution is 2.145. So the 95% CI is $.24 \pm 2.145(.1164)$, or about –.010 to .490.

(ii) If $\mu = \mathrm{E}(D_i)$ then H_0: $\mu = 0$. The alternative is that management's claim is true: H_1: $\mu > 0$.

(iii) We have the mean and standard error from part (i): $t = .24/.1164 \approx 2.062$. The 5% critical value for a one-tailed test with $df = 14$ is 1.761, while the 1% critical value is 2.624. Therefore, H_0 is rejected in favor of H_1 at the 5% level but not the 1% level.

(iv) The *p*-value obtained from Stata is .029; this is half of the *p*-value for the two-sided alternative. (Econometrics packages, including Stata, report the *p*-value for the two-sided alternative.)

C.9 (i) X is distributed as Binomial(200,.65), and so $\mathrm{E}(X) = 200(.65) = 130$.

(ii) $\mathrm{Var}(X) = 200(.65)(1 - .65) = 45.5$, so $\mathrm{sd}(X) \approx 6.75$.

(iii) $\mathrm{P}(X \leq 115) = \mathrm{P}[(X - 130)/6.75 \leq (115 - 130)/6.75] \approx \mathrm{P}(Z \leq -2.22)$, where Z is a standard normal random variable. From Table G.1, $\mathrm{P}(Z \leq -2.22) \approx .013$.

(iv) The evidence is pretty strong against the dictator's claim. If 65% of the voting population actually voted yes in the plebiscite, there is only about a 1.3% chance of obtaining 115 or fewer voters out of 200 who voted yes.

APPENDIX D

SOLUTIONS TO PROBLEMS

D.1 (i) $\mathbf{AB} = \begin{pmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 6 \\ 1 & 8 & 0 \\ 3 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 20 & -6 & 12 \\ 5 & 36 & -24 \end{pmatrix}$

(ii) **BA** does not exist because **B** is 3 × 3 and **A** is 2 × 3.

D.3 Using the basic rules for transpose, $(\mathbf{X}'\mathbf{X}') = (\mathbf{X}')(\mathbf{X}')' = \mathbf{X}'\mathbf{X}$, which is what we wanted to show.

D.5 (i) The $n \times n$ matrix **C** is the inverse of **AB** if and only if $\mathbf{C}(\mathbf{AB}) = \mathbf{I}_n$ and $(\mathbf{AB})\mathbf{C} = \mathbf{I}_n$. We verify both of these equalities for $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. First, $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{I}_n\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_n$. Similarly, $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AI}_n\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}_n$.

(ii) $(\mathbf{ABC})^{-1} = (\mathbf{BC})^{-1}\mathbf{A}^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$.

D.7 We must show that, for any $n \times 1$ vector **x**, $x \neq \mathbf{0}$, $\mathbf{x}'(\mathbf{P}'\mathbf{AB})\,\mathbf{x} > 0$. But we can write this quadratic form as $(\mathbf{P\,x})'\mathbf{A}(\mathbf{P\,x}) = \mathbf{z}'\mathbf{Az}$ where $\mathbf{z} \equiv \mathbf{Px}$. Because **A** is positive definite by assumption, $\mathbf{z}'\mathbf{Az} > 0$ for $\mathbf{z} \neq \mathbf{0}$. So, all we have to show is that $\mathbf{x} \neq \mathbf{0}$ implies that $\mathbf{z} \neq \mathbf{0}$. We do this by showing the contrapositive, that is, if $\mathbf{z} = \mathbf{0}$ then $\mathbf{x} = \mathbf{0}$. If $\mathbf{Px} = \mathbf{0}$ then, because $\mathbf{P}^{-1}$ exists, we have $\mathbf{P}^{-1}\mathbf{Px} = \mathbf{0}$ or $\mathbf{x} = \mathbf{0}$, which completes the proof.

APPENDIX E

SOLUTIONS TO PROBLEMS

E.1 This follows directly from partitioned matrix multiplication in Appendix D. Write

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}, \mathbf{X}' = (\mathbf{x}_1' \ \mathbf{x}_2' \ \ldots \ \mathbf{x}_n'), \text{ and } \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}$$

Therefore, $\mathbf{X}'\mathbf{X} = \sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t$ and $\mathbf{X}'\mathbf{y} = \sum_{t=1}^{n} \mathbf{x}_t'\mathbf{y}_t$. An equivalent expression for $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} = \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t \right)^{-1} \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t' y_t \right)$$

which, when we plug in $y_t = \mathbf{x}_t\boldsymbol{\beta} + u_t$ for each t and do some algebra, can be written as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t \right)^{-1} \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t' u_t \right).$$

As shown in Section E.4, this expression is the basis for the asymptotic analysis of OLS using matrices.

E.3 (i) We use the placeholder feature of the OLS formulas. By definition, $\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = [(\mathbf{XA})' (\mathbf{XA})]^{-1}(\mathbf{XA})'\mathbf{y} = [\mathbf{A}'(\mathbf{X}'\mathbf{X})\mathbf{A}]^{-1}\mathbf{A}'\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}')^{-1}\mathbf{A}'\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$.

(ii) By definition of the fitted values, $\hat{y}_t = \mathbf{x}_t\hat{\boldsymbol{\beta}}$ and $\tilde{y}_t = \mathbf{z}_t\tilde{\boldsymbol{\beta}}$. Plugging $\mathbf{z}_t$ and $\tilde{\boldsymbol{\beta}}$ into the second equation gives $\tilde{y}_t = (\mathbf{x}_t\mathbf{A})(\mathbf{A}^{-1}\hat{\boldsymbol{\beta}}) = \mathbf{x}_t\hat{\boldsymbol{\beta}} = \hat{y}_t$.

(iii) The estimated variance matrix from the regression of $\mathbf{y}$ and $\mathbf{Z}$ is $\tilde{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1}$ where $\tilde{\sigma}^2$ is the error variance estimate from this regression. From part (ii), the fitted values from the two regressions are the same, which means the residuals must be the same for all t. (The dependent variable is the same in both regressions.) Therefore, $\tilde{\sigma}^2 = \hat{\sigma}^2$. Further, as we showed in part (i), $(\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}')^{-1}$, and so $\tilde{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1} = \hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}^{-1})'$, which is what we wanted to show.

(iv) The $\tilde{\beta}_j$ are obtained from a regression of $\mathbf{y}$ on $\mathbf{XA}$, where $\mathbf{A}$ is the $k \times k$ diagonal matrix with 1, a_2, …, a_k down the diagonal. From part (i), $\tilde{\boldsymbol{\beta}} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$. But $\mathbf{A}^{-1}$ is easily seen to be the $k \times k$ diagonal matrix with 1, a_2^{-1}, …, a_k^{-1} down its diagonal. Straightforward multiplication shows that the first element of $\mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$ is $\hat{\beta}_1$ and the j^{th} element is $\hat{\beta}_j/a_j$, $j = 2, \ldots, k$.

(v) From part (iii), the estimated variance matrix of $\tilde{\boldsymbol{\beta}}$ is $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}^{-1})'$. But $\mathbf{A}^{-1}$ is a symmetric, diagonal matrix, as described above. The estimated variance of $\tilde{\beta}_j$ is the j^{th} diagonal element of $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}^{-1}$, which is easily seen to be = $\hat{\sigma}^2 c_{jj}/a_j^{-2}$, where c_{jj} is the j^{th} diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. The square root of this, $\hat{\sigma}\sqrt{c_{jj}}/|a_j|$, is se($\tilde{\beta}_j$), which is simply se($\tilde{\beta}_j$)/$|a_j|$.

(vi) The t statistic for $\tilde{\beta}_j$ is, as usual,

$$\tilde{\beta}_j/\text{se}(\tilde{\beta}_j) = (\hat{\beta}_j/a_j)/[\text{se}(\hat{\beta}_j)/|a_j|],$$

and so the absolute value is $(|\hat{\beta}_j|/|a_j|)/[\text{se}(\hat{\beta}_j)/|a_j|] = |\hat{\beta}_j|/\text{se}(\hat{\beta}_j)$, which is just the absolute value of the t statistic for $\hat{\beta}_j$. If $a_j > 0$, the t statistics themselves are identical; if $a_j < 0$, the t statistics are simply opposite in sign.

E.5 (i) By plugging in for $\mathbf{y}$, we can write

$$\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{y} = (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\mathbf{X}\boldsymbol{\beta} + \mathbf{u}) = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}.$$

Now we use the fact that $\mathbf{Z}$ is a function of $\mathbf{X}$ to pull $\mathbf{Z}$ outside of the conditional expectation:

$$\text{E}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta} + \text{E}[(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u} \mid \mathbf{X}] = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\text{E}(\mathbf{u} \mid \mathbf{X}) = \boldsymbol{\beta}.$$

(ii) We start from the same representation in part (i): $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}$ and so

$$\begin{aligned}\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'[\text{Var}(\mathbf{u} \mid \mathbf{X})]\mathbf{Z}[(\mathbf{Z}'\mathbf{X})^{-1}]' \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\sigma^2\mathbf{I}_n)\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1} = \sigma^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}.\end{aligned}$$

A common mistake is to forget to transpose the matrix $\mathbf{Z}'\mathbf{X}$ in the last term.

(iii) The estimator $\tilde{\boldsymbol{\beta}}$ is linear in $\mathbf{y}$ and, as shown in part (i), it is unbiased (conditional on $\mathbf{X}$). Because the Gauss-Markov assumptions hold, the OLS estimator, $\hat{\boldsymbol{\beta}}$, is best linear unbiased. In particular, its variance-covariance matrix is "smaller" (in the matrix sense) than $\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X})$. Therefore, we prefer the OLS estimator.