CONTENTS

SOUTH-WESTERN CENGAGE Learning

PREFACE

This manual contains suggested course outlines, teaching notes, and detailed solutions to all of the problems and computer exercises in *Introductory Econometrics: A Modern Approach*, 4e. For several problems, I have added additional notes about interesting asides or suggestions for how to modify or extend the problem.

Some of the answers given here are subjective, and you may want to supplement or replace them with your own answers. I wrote all solutions as if I were preparing them for the students, so you may find some solutions a bit tedious (if not offensive). This way, if you prefer, you can distribute my answers to some of the even-numbered problems directly to the students. (The student study guide contains answers to all odd-numbered problems.) Many of the equations in the Word files were created using MathType, and the equations will not look quite right without MathType.

I solved the computer exercises using various versions of Stata, starting with version 4.0 and running through 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. 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.)

While I have endeavored to make the solutions mistake-free, some errors may have crept in. I would appreciate hearing from you if you find mistakes. I will keep a list of any substantive errors on the Web site for the book, www.international.cengage.com. I heard from many of you regarding the earlier editions of the text, and I incorporated many of your suggestions. I welcome any comments that will help me make improvements to future editions. I can be reached via e-mail at wooldri1@.msu.edu.

I hope you find this instructor's manual useful, and I look forward to hearing your reactions to the second edition.

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

SUGGESTED COURSE OUTLINES

For an introductory, one-semester course, I like to cover most of the material in Chapters 1 through 8 and Chapters 10 through 12, as well as parts of Chapter 9 (but mostly through examples). I do not typically cover all sections or subsections within each chapter. Under the chapter headings listed below, I provide some comments on the material I find most relevant for a first-semester course.

An alternative course ignores time series applications altogether, while delving into some of the more advanced methods that are particularly useful for policy analysis. This would consist of Chapters 1 through 8, much of Chapter 9, and the first four sections of Chapter 13. Chapter 9 discusses the practically important topics of proxy variables, measurement error, outlying observations, and stratified sampling. In addition, I have written a more careful description of the method of least absolute deviations, including a discussion of its strengths and weaknesses. Chapter 13 covers, in a straightforward fashion, methods for pooled cross sections (including the so-called "natural experiment" approach) and two-period panel data analysis. The basic cross-sectional treatment of instrumental variables in Chapter 15 is a natural topic for cross-sectional, policy-oriented courses. For an accelerated course, the nonlinear methods used for cross-sectional analysis in Chapter 17 can be covered.

I typically do not begin with a review of basic algebra, probability, and statistics. In my experience, this takes too long and the payoff is minimal. (Students tend to think that they are taking another statistics course, and start to drift.) Instead, when I need a tool (such as the summation or expectations operator), I briefly review the necessary definitions and key properties. Statistical inference is not more difficult to describe in terms of multiple regression than in tests of a population mean, and so I briefly review the principles of statistical inference during multiple regression analysis. Appendices A, B, and C are fairly extensive. When I cover asymptotic properties of OLS, I provide a brief discussion of the main definitions and limit theorems. If students need more than the brief review provided in class, I point them to the appendices.

For a master's level course, I include a couple of lectures on the matrix approach to linear regression. This could be integrated into Chapters 3 and 4 or covered after Chapter 4. Again, I do not summarize matrix algebra before proceeding. Instead, the material in Appendix D can be reviewed as it is needed in covering Appendix E.

A second semester course, at either the undergraduate or masters level, could begin with some of the material in Chapter 9, particularly with the issues of proxy variables and measurement error. The advanced chapters, starting with Chapter 13, are useful for students who have an interest in policy analysis. The pooled cross section and panel data chapters (Chapters 13 and 14) emphasize how these data sets can be used, in conjunction with econometric methods, for policy evaluation. Chapter 15, which introduces the method of instrumental variables, is also important for policy analysis. Most modern IV applications are used to address the problems of omitted variables (unobserved heterogeneity) or measurement error. I have intentionally separated out the conceptually more difficult topic of simultaneous equations models in Chapter 16.

Chapter 17, in particular the material on probit, logit, Tobit, and Poisson regression models, is a good introduction to nonlinear econometric methods. Specialized courses that emphasize applications in labor economics can use the material on sample selection corrections. Duration models are also briefly covered as an example of a censored regression model.

Chapter 18 is much different from the other advanced chapters, as it focuses on more advanced or recent developments in time series econometrics. Combined with some of the more advanced topics in Chapter 12, it can serve as the basis for a second semester course in time series topics, including forecasting.

Most second semester courses would include an assignment to write an original empirical paper, and Chapter 19 should be helpful in this regard.

CHAPTER 1

TEACHING NOTES

You have substantial latitude about what to emphasize in Chapter 1. I find it useful to talk about the economics of crime example (Example 1.1) and the wage example (Example 1.2) so that students see, at the outset, that econometrics is linked to economic reasoning, even if the economics is not complicated theory.

I like to familiarize students with the important data structures that empirical economists use, focusing primarily on cross-sectional and time series data sets, as these are what I cover in a first-semester course. It is probably a good idea to mention the growing importance of data sets that have both a cross-sectional and time dimension.

I spend almost an entire lecture talking about the problems inherent in drawing causal inferences in the social sciences. I do this mostly through the agricultural yield, return to education, and crime examples. These examples also contrast experimental and nonexperimental (observational) data. Students studying business and finance tend to find the term structure of interest rates example more relevant, although the issue there is testing the implication of a simple theory, as opposed to inferring causality. I have found that spending time talking about these examples, in place of a formal review of probability and statistics, is more successful (and more enjoyable for the students and me).

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.

1.3 (i) Here is one way to pose the question: If two firms, say *A* and *B*, are identical in all respects except that firm *A* supplies job training one hour per worker more than firm *B*, by how much would firm *A*'s output differ from firm *B*'s?

(ii) Firms are likely to choose job training depending on the characteristics of workers. Some observed characteristics are years of schooling, years in the workforce, and experience in a particular job. Firms might even discriminate based on age, gender, or race. Perhaps firms choose to offer training to more or less able workers, where "ability" might be difficult to quantify but where a manager has some idea about the relative abilities of different employees. Moreover, different kinds of workers might be attracted to firms that offer more job training on average, and this might not be evident to employers.

(iii) The amount of capital and technology available to workers would also affect output. So, two firms with exactly the same kinds of employees would generally have different outputs if they use different amounts of capital or technology. The quality of managers would also have an effect.

(iv) No, unless the amount of training is randomly assigned. The many factors listed in parts (ii) and (iii) can contribute to finding a positive correlation between *output* and *training* even if job training does not improve worker productivity.

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.

C1.2 (i) There are 1,388 observations in the sample. Tabulating the variable *cigs* shows that 212 women have *cigs* > 0.

(ii) The average of *cigs* is about 2.09, but this includes the 1,176 women who did not smoke. Reporting just the average masks the fact that almost 85 percent of the women did not smoke. It makes more sense to say that the "typical" woman does not smoke during pregnancy; indeed, the median number of cigarettes smoked is zero.

(iii) The average of *cigs* over the women with *cigs* > 0 is about 13.7. Of course this is much higher than the average over the entire sample because we are excluding 1,176 zeros.

(iv) The average of *fatheduc* is about 13.2. There are 196 observations with a missing value for *fatheduc*, and those observations are necessarily excluded in computing the average.

(v) The average and standard deviation of *faminc* are about 29.027 and 18.739, respectively, but *faminc* is measured in thousands of dollars. So, in dollars, the average and standard deviation are $29,027 and $18,739.

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.]

C1.4 (i) 185/445 $\approx$ .416 is the fraction of men receiving job training, or about 41.6%.

(ii) For men receiving job training, the average of *re78* is about 6.35, or $6,350. For men not receiving job training, the average of *re78* is about 4.55, or $4,550. The difference is $1,800, which is very large. On average, the men receiving the job training had earnings about 40% higher than those not receiving training.

(iii) About 24.3% of the men who received training were unemployed in 1978; the figure is 35.4% for men not receiving training. This, too, is a big difference.

(iv) The differences in earnings and unemployment rates suggest the training program had strong, positive effects. Our conclusions about economic significance would be stronger if we could also establish statistical significance (which is done in Computer Exercise C9.10 in Chapter 9).

CHAPTER 2

TEACHING NOTES

This is the chapter where I expect students to follow most, if not all, of the algebraic derivations. In class I like to derive at least the unbiasedness of the OLS slope coefficient, and usually I derive the variance. At a minimum, I talk about the factors affecting the variance. To simplify the notation, after I emphasize the assumptions in the population model, and assume random sampling, I just condition on the values of the explanatory variables in the sample. Technically, this is justified by random sampling because, for example, $\mathrm{E}(u_i|x_1,x_2,\ldots,x_n) = \mathrm{E}(u_i|x_i)$ by independent sampling. I find that students are able to focus on the key assumption SLR.4 and subsequently take my word about how conditioning on the independent variables in the sample is harmless. (If you prefer, the appendix to Chapter 3 does the conditioning argument carefully.) Because statistical inference is no more difficult in multiple regression than in simple regression, I postpone inference until Chapter 4. (This reduces redundancy and allows you to focus on the interpretive differences between simple and multiple regression.)

You might notice how, compared with most other texts, I use relatively few assumptions to derive the unbiasedness of the OLS slope estimator, followed by the formula for its variance. This is because I do not introduce redundant or unnecessary assumptions. For example, once SLR.4 is assumed, nothing further about the relationship between u and x is needed to obtain the unbiasedness of OLS under random sampling.

SOLUTIONS TO PROBLEMS

2.1 In the equation $y = \beta_0 + \beta_1 x + u$, add and subtract α_0 from the right hand side to get $y = (\alpha_0 + \beta_0) + \beta_1 x + (u - \alpha_0)$. Call the new error $e = u - \alpha_0$, so that E(e) = 0. The new intercept is $\alpha_0 + \beta_0$, but the slope is still β_1.

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 $\text{E}(u|educ) \neq 0$, 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 $\text{E}(u) = 0$, 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 $\text{E}(u|inc) = \text{E}(\sqrt{inc} \cdot e|inc) = \sqrt{inc} \cdot \text{E}(e|inc) = \sqrt{inc} \cdot 0$ because $\text{E}(e|inc) = \text{E}(e) = 0$.

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

(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.

2.6 (i) This derivation is essentially done in equation (2.52), once $(1/\mathrm{SST}_x)$ is brought inside the summation (which is valid because SST_x does not depend on *i*). Then, just define $w_i = d_i / \mathrm{SST}_x$.

(ii) Because $\mathrm{Cov}(\hat{\beta}_1, \bar{u}) = \mathrm{E}[(\hat{\beta}_1 - \beta_1)\bar{u}]$, we show that the latter is zero. But, from part (i), $\mathrm{E}[(\hat{\beta}_1 - \beta_1)\bar{u}] = \mathrm{E}\left[\left(\sum_{i=1}^n w_i u_i\right)\bar{u}\right] = \sum_{i=1}^n w_i \mathrm{E}(u_i \bar{u})$. Because the u_i are pairwise uncorrelated (they are independent), $\mathrm{E}(u_i \bar{u}) = \mathrm{E}(u_i^2 / n) = \sigma^2 / n$ (because $\mathrm{E}(u_i u_h) = 0,\ i \neq h$). Therefore, $\sum_{i=1}^n w_i \mathrm{E}(u_i \bar{u}) = \sum_{i=1}^n w_i (\sigma^2 / n) = (\sigma^2 / n) \sum_{i=1}^n w_i = 0.$

(iii) The formula for the OLS intercept is $\hat{\beta}_0 = \bar{y} - \hat{\beta}\bar{x}$ and, plugging in $\bar{y} = \beta_0 + \beta_1 \bar{x} + \bar{u}$ gives $\hat{\beta}_0 = (\beta_0 + \beta_1 \bar{x} + \bar{u}) - \hat{\beta}_1 \bar{x} = \beta_0 + \bar{u} - (\hat{\beta}_1 - \beta_1)\bar{x}.$

(iv) Because $\hat{\beta}_1$ and $\bar{u}$ are uncorrelated,

$$\mathrm{Var}(\hat{\beta}_0) = \mathrm{Var}(\bar{u}) + \mathrm{Var}(\hat{\beta}_1)\bar{x}^2 = \sigma^2 / n + (\sigma^2 / \mathrm{SST}_x)\bar{x}^2 = \sigma^2 / n + \sigma^2 \bar{x}^2 / \mathrm{SST}_x,$$

which is what we wanted to show.

(v) Using the hint and substitution gives $\mathrm{Var}(\hat{\beta}_0) = \sigma^2[(\mathrm{SST}_x / n) + \bar{x}^2] / \mathrm{SST}_x$

$$= \sigma^2 \left[\left(n^{-1}\sum_{i=1}^n x_i^2 - \bar{x}^2\right) + \bar{x}^2\right] / \mathrm{SST}_x = \sigma^2 \left(n^{-1}\sum_{i=1}^n x_i^2\right) / \mathrm{SST}_x.$$

2.7 (i) Yes. If living closer to an incinerator depresses housing prices, then being farther away increases housing prices.

(ii) If the city chose to locate the incinerator in an area away from more expensive neighborhoods, then log(*dist*) is positively correlated with housing quality. This would violate SLR.4, and OLS estimation is biased.

(iii) Size of the house, number of bathrooms, size of the lot, age of the home, and quality of the neighborhood (including school quality), are just a handful of factors. As mentioned in part (ii), these could certainly be correlated with *dist* [and log(*dist*)].

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

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

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

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

(ii) We use the same approach from part (i) along with the fact that $\overline{(c_1 + y)} = c_1 + \bar{y}$ and $\overline{(c_2 + x)} = c_2 + \bar{x}$. Therefore, $(c_1 + y_i) - \overline{(c_1 + y)} = (c_1 + y_i) - (c_1 + \bar{y}) = y_i - \bar{y}$ and $(c_2 + x_i) - \overline{(c_2 + x)} = x_i - \bar{x}$. 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 $\tilde{\beta}_0 = \overline{(c_1 + y)} - \tilde{\beta}_1\overline{(c_2 + x)} = (c_1 + \bar{y}) - \hat{\beta}_1(c_2 + \bar{x}) = (\bar{y} - \hat{\beta}_1\bar{x}) + c_1 - c_2\hat{\beta}_1 = \hat{\beta}_0 + c_1 - c_2\hat{\beta}_1$, which is what we wanted to show.

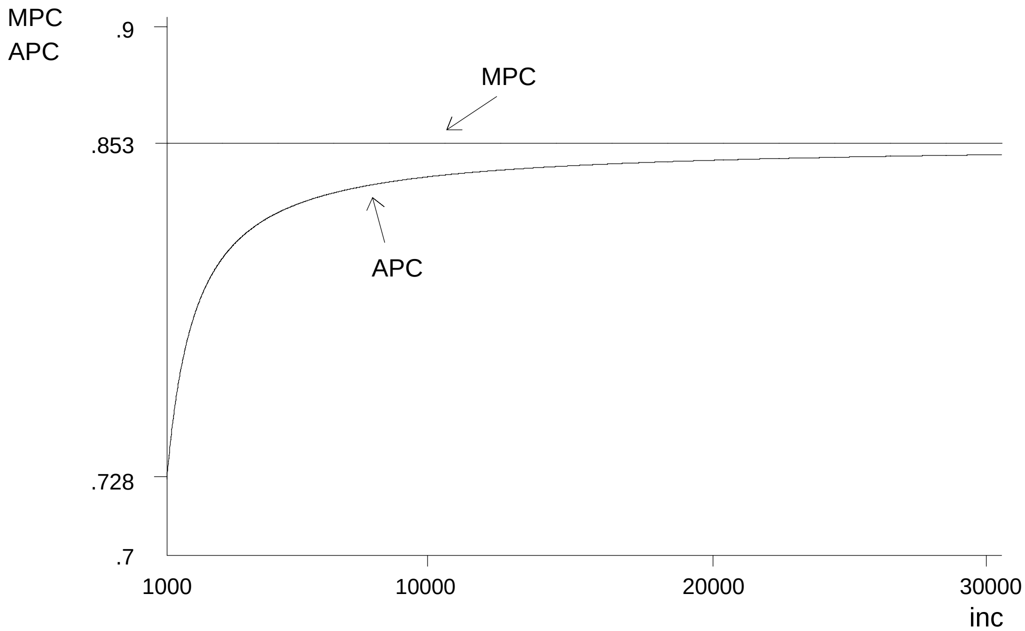
(iii) We can simply apply part (ii) because $\log(c_1 y_i) = \log(c_1) + \log(y_i)$. In other words, replace c_1 with $\log(c_1)$, y_i with $\log(y_i)$, and set $c_2 = 0$.

(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).



2.10 (i) From equation (2.66),

$$\tilde{\beta}_1 = \left(\sum_{i=1}^{n} x_i y_i\right) / \left(\sum_{i=1}^{n} x_i^2\right).$$

Plugging in $y_i = \beta_0 + \beta_1 x_i + u_i$ gives

$$\tilde{\beta}_1 = \left(\sum_{i=1}^{n} x_i(\beta_0 + \beta_1 x_i + u_i)\right) / \left(\sum_{i=1}^{n} x_i^2\right).$$

After standard algebra, the numerator can be written as

$$\beta_0 \sum_{i=1}^{n} x_i + \beta_1 \sum_{i=1}^{n} x_i^2 + \sum_{i=1}^{n} x_i u_i .$$

Putting this over the denominator shows we can write $\tilde{\beta}_1$ as

$$\tilde{\beta}_1 = \beta_0 \left(\sum_{i=1}^{n} x_i\right) / \left(\sum_{i=1}^{n} x_i^2\right) + \beta_1 + \left(\sum_{i=1}^{n} x_i u_i\right) / \left(\sum_{i=1}^{n} x_i^2\right).$$

Conditional on the x_i, we have

$$\mathrm{E}(\tilde{\beta}_1) = \beta_0\left(\sum_{i=1}^{n} x_i\right) / \left(\sum_{i=1}^{n} x_i^2\right) + \beta_1$$

because $\mathrm{E}(u_i) = 0$ for all i. Therefore, the bias in $\tilde{\beta}_1$ is given by the first term in this equation. This bias is obviously zero when $\beta_0 = 0$. It is also zero when $\sum_{i=1}^{n} x_i = 0$, which is the same as $\bar{x} = 0$. In the latter case, regression through the origin is identical to regression with an intercept.

(ii) From the last expression for $\tilde{\beta}_1$ in part (i) we have, conditional on the x_i,

$$\mathrm{Var}(\tilde{\beta}_1) = \left(\sum_{i=1}^{n} x_i^2\right)^{-2} \mathrm{Var}\left(\sum_{i=1}^{n} x_i u_i\right) = \left(\sum_{i=1}^{n} x_i^2\right)^{-2} \left(\sum_{i=1}^{n} x_i^2 \mathrm{Var}(u_i)\right)$$

$$= \left(\sum_{i=1}^{n} x_i^2\right)^{-2} \left(\sigma^2 \sum_{i=1}^{n} x_i^2\right) = \sigma^2 / \left(\sum_{i=1}^{n} x_i^2\right).$$

(iii) From (2.57), $\mathrm{Var}(\hat{\beta}_1) = \sigma^2/\left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$. From the hint, $\sum_{i=1}^{n} x_i^2 \geq \sum_{i=1}^{n}(x_i - \bar{x})^2$, and so $\mathrm{Var}(\tilde{\beta}_1) \leq \mathrm{Var}(\hat{\beta}_1)$. A more direct way to see this is to write $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n(\bar{x})^2$, which is less than $\sum_{i=1}^{n} x_i^2$ unless $\bar{x} = 0$.

(iv) For a given sample size, the bias in $\tilde{\beta}_1$ increases as $\bar{x}$ increases (holding the sum of the x_i^2 fixed). But as $\bar{x}$ increases, the variance of $\hat{\beta}_1$ increases relative to $\mathrm{Var}(\tilde{\beta}_1)$. The bias in $\tilde{\beta}_1$ is also small when β_0 is small. Therefore, whether we prefer $\tilde{\beta}_1$ or $\hat{\beta}_1$ on a mean squared error basis depends on the sizes of β_0, $\bar{x}$, and n (in addition to the size of $\sum_{i=1}^{n} x_i^2$).

2.11 (i) When *cigs* = 0, predicted birth weight is 119.77 ounces. When *cigs* = 20, $\widehat{bwght}$ = 109.49. This is about an 8.6% drop.

(ii) Not necessarily. There are many other factors that can affect birth weight, particularly overall health of the mother and quality of prenatal care. These could be correlated with cigarette smoking during birth. Also, something such as caffeine consumption can affect birth weight, and might also be correlated with cigarette smoking.

(iii) If we want a predicted *bwght* of 125, then *cigs* = (125 – 119.77)/(–.524) ≈ –10.18, or about –10 cigarettes! This is nonsense, of course, and it shows what happens when we are trying to predict something as complicated as birth weight with only a single explanatory variable. The largest predicted birth weight is necessarily 119.77. Yet almost 700 of the births in the sample had a birth weight higher than 119.77.

(iv) 1,176 out of 1,388 women did not smoke while pregnant, or about 84.7%. Because we are using only *cigs* to explain birth weight, we have only one predicted birth weight at *cigs* = 0. The predicted birth weight is necessarily roughly in the middle of the observed birth weights at *cigs* = 0, and so we will under predict high birth rates.

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 $\widehat{prate}$ = 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.

C2.2 (i) Average salary is about 865.864, which means $865,864 because *salary* is in thousands of dollars. Average *ceoten* is about 7.95.

(ii) There are five CEOs with *ceoten* = 0. The longest tenure is 37 years.

(iii) The estimated equation is

$$\widehat{\log(salary)} = 6.51 + .0097\, ceoten$$

$$n = 177,\ \ R^2 = .013.$$

We obtain the approximate percentage change in *salary* given Δ*ceoten* = 1 by multiplying the coefficient on *ceoten* by 100, 100(.0097) = .97%. Therefore, one more year as CEO is predicted to increase salary by almost 1%.

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.

C2.4 (i) Average salary is about \$957.95 and average IQ is about 101.28. The sample standard deviation of IQ is about 15.05, which is pretty close to the population value of 15.

(ii) This calls for a level-level model:

$$\widehat{wage} = 116.99 + 8.30\ IQ$$

$$n = 935,\ \ R^2 = .096.$$

An increase in *IQ* of 15 increases predicted monthly salary by 8.30(15) = \$124.50 (in 1980 dollars). *IQ* score does not even explain 10% of the variation in *wage*.

(iii) This calls for a log-level model:

$$\widehat{\log(wage)} = 5.89 + .0088\ IQ$$

$$n = 935,\ \ R^2 = .099.$$

If $\Delta IQ = 15$ then $\Delta\widehat{\log(wage)} = .0088(15) = .132$, which is the (approximate) proportionate change in predicted wage. The percentage increase is therefore approximately 13.2.

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

$$\widehat{\log(rd)} = -4.105 + 1.076\ \log(sales)$$

$$n = 32,\ \ R^2 = .910.$$

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%.

C2.6 (i) It seems plausible that another dollar of spending has a larger effect for low-spending schools than for high-spending schools. At low-spending schools, more money can go toward purchasing more books, computers, and for hiring better qualified teachers. At high levels of spending, we would expend little, if any, effect because the high-spending schools already have high-quality teachers, nice facilities, plenty of books, and so on.

(ii) If we take changes, as usual, we obtain

$$\Delta math10 = \beta_1 \Delta \log(expend) \approx (\beta_1/100)(\%\Delta expend),$$

just as in the second row of Table 2.3. So, if $\%\Delta expend = 10$, $\Delta math10 = \beta_1/10$.

(iii) The regression results are

$$\widehat{math10} = -69.34 \;+\; 11.16 \;\log(expend)$$
$$n = 408,\quad R^2 = .0297$$

(iv) If *expend* increases by 10 percent, $\widehat{math10}$ increases by about 1.1 percentage points. This is not a huge effect, but it is not trivial for low-spending schools, where a 10 percent increase in spending might be a fairly small dollar amount.

(v) In this data set, the largest value of *math10* is 66.7, which is not especially close to 100. In fact, the largest fitted values is only about 30.2.

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

TEACHING NOTES

For undergraduates, I do not work through most of the derivations in this chapter, at least not in detail. Rather, I focus on interpreting the assumptions, which mostly concern the population. Other than random sampling, the only assumption that involves more than population considerations is the assumption about no perfect collinearity, where the possibility of perfect collinearity in the sample (even if it does not occur in the population) should be touched on. The more important issue is perfect collinearity in the population, but this is fairly easy to dispense with via examples. These come from my experiences with the kinds of model specification issues that beginners have trouble with.

The comparison of simple and multiple regression estimates – based on the particular sample at hand, as opposed to their statistical properties – usually makes a strong impression. Sometimes I do not bother with the "partialling out" interpretation of multiple regression.

As far as statistical properties, notice how I treat the problem of including an irrelevant variable: no separate derivation is needed, as the result follows form Theorem 3.1.

I do like to derive the omitted variable bias in the simple case. This is not much more difficult than showing unbiasedness of OLS in the simple regression case under the first four Gauss-Markov assumptions. It is important to get the students thinking about this problem early on, and before too many additional (unnecessary) assumptions have been introduced.

I have intentionally kept the discussion of multicollinearity to a minimum. This partly indicates my bias, but it also reflects reality. It is, of course, very important for students to understand the potential consequences of having highly correlated independent variables. But this is often beyond our control, except that we can ask less of our multiple regression analysis. If two or more explanatory variables are highly correlated in the sample, we should not expect to precisely estimate their ceteris paribus effects in the population.

I find extensive treatments of multicollinearity, where one "tests" or somehow "solves" the multicollinearity problem, to be misleading, at best. Even the organization of some texts gives the impression that imperfect multicollinearity is somehow a violation of the Gauss-Markov assumptions: they include multicollinearity in a chapter or part of the book devoted to "violation of the basic assumptions," or something like that. I have noticed that master's students who have had some undergraduate econometrics are often confused on the multicollinearity issue. It is very important that students not confuse multicollinearity among the included explanatory variables in a regression model with the bias caused by omitting an important variable.

I do not prove the Gauss-Markov theorem. Instead, I emphasize its implications. Sometimes, and certainly for advanced beginners, I put a special case of Problem 3.12 on a midterm exam, where I make a particular choice for the function $g(x)$. Rather than have the students directly compare the variances, they should appeal to the Gauss-Markov theorem for the superiority of OLS over any other linear, unbiased estimator.

SOLUTIONS TO PROBLEMS

3.1 (i) Yes. Because of budget constraints, it makes sense that, the more siblings there are in a family, the less education any one child in the family has. To find the increase in the number of siblings that reduces predicted education by one year, we solve $1 = .094(\Delta sibs)$, so $\Delta sibs = 1/.094 \approx 10.6$.

(ii) Holding *sibs* and *feduc* fixed, one more year of mother's education implies .131 years more of predicted education. So if a mother has four more years of education, her son is predicted to have about a half a year (.524) more years of education.

(iii) Since the number of siblings is the same, but *meduc* and *feduc* are both different, the coefficients on *meduc* and *feduc* both need to be accounted for. The predicted difference in education between B and A is .131(4) + .210(4) = 1.364.

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, $\Delta\widehat{colgpa} = .00148\Delta sat$. Now, we want to find Δsat such that $\Delta\widehat{colgpa} = .5$, so $.5 = .00148(\Delta sat)$ or $\Delta sat = .5/(.00148) \approx 338$. 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.

3.3 (i) A larger rank for a law school means that the school has less prestige; this lowers starting salaries. For example, a rank of 100 means there are 99 schools thought to be better.

(ii) $\beta_1 > 0$, $\beta_2 > 0$. Both *LSAT* and *GPA* are measures of the quality of the entering class. No matter where better students attend law school, we expect them to earn more, on average. β_3, $\beta_4 > 0$. The number of volumes in the law library and the tuition cost are both measures of the school quality. (Cost is less obvious than library volumes, but should reflect quality of the faculty, physical plant, and so on.)

(iii) This is just the coefficient on *GPA*, multiplied by 100: 24.8%.

(iv) This is an elasticity: a one percent increase in library volumes implies a .095% increase in predicted median starting salary, other things equal.

(v) It is definitely better to attend a law school with a lower rank. If law school A has a ranking 20 less than law school B, the predicted difference in starting salary is 100(.0033)(20) = 6.6% higher for law school A.

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.)

3.5 Conditioning on the outcomes of the explanatory variables, we have $\text{E}(\hat{\theta}_1) = \text{E}(\hat{\beta}_1 + \hat{\beta}_2) = \text{E}(\hat{\beta}_1) + \text{E}(\hat{\beta}_2) = \beta_1 + \beta_2 = \theta_1$.

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.

3.7 We can use Table 3.2. By definition, $\beta_2 > 0$, and by assumption, Corr(x_1,x_2) < 0. Therefore, there is a negative bias in $\tilde{\beta}_1$: E($\tilde{\beta}_1$) < β_1. This means that, on average across different random samples, the simple regression estimator underestimates the effect of the training program. It is even possible that E($\tilde{\beta}_1$) is negative even though $\beta_1 > 0$.

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.

3.9 (i) Because x_1 is highly correlated with x_2 and x_3, and these latter variables have large partial effects on *y*, the simple and multiple regression coefficients on x_1 can differ by large amounts. We have not done this case explicitly, but given equation (3.46) and the discussion with a single omitted variable, the intuition is pretty straightforward.

(ii) Here we would expect $\tilde{\beta}_1$ and $\hat{\beta}_1$ to be similar (subject, of course, to what we mean by "almost uncorrelated"). The amount of correlation between x_2 and x_3 does not directly effect the multiple regression estimate on x_1 if x_1 is essentially uncorrelated with x_2 and x_3.

(iii) In this case we are (unnecessarily) introducing multicollinearity into the regression: x_2 and x_3 have small partial effects on *y* and yet x_2 and x_3 are highly correlated with x_1. Adding x_2 and x_3 like increases the standard error of the coefficient on x_1 substantially, so se($\hat{\beta}_1$) is likely to be much larger than se($\tilde{\beta}_1$).

(iv) In this case, adding x_2 and x_3 will decrease the residual variance without causing much collinearity (because x_1 is almost uncorrelated with x_2 and x_3), so we should see se($\hat{\beta}_1$) smaller than se($\tilde{\beta}_1$). The amount of correlation between x_2 and x_3 does not directly affect se($\hat{\beta}_1$).

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 $\mathrm{E}(u_i) = 0$, and so

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

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 $\mathrm{Corr}(x_1,x_2) < 0$, 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. [$\mathrm{E}(\tilde{\beta}_1)$ 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 $s_{zx} = \sum_{i=1}^{n}(z_i - \bar{z})x_i$; 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

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

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

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

$$= \frac{\beta_0 \sum_{i=1}^{n}(z_i - \overline{z}) + \beta_1 s_{zx} + \sum_{i=1}^{n}(z_i - \overline{z})u_i}{s_{zx}}$$

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

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),

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

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

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

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

3.13 (i) The shares, by definition, add to one. If we do not omit one of the shares then the equation would suffer from perfect multicollinearity. The parameters would not have a ceteris paribus interpretation, as it is impossible to change one share while holding *all* of the other shares fixed.

(ii) Because each share is a proportion (and can be at most one, when all other shares are zero), it makes little sense to increase $share_p$ by one unit. If $share_p$ increases by .01 – which is equivalent to a one percentage point increase in the share of property taxes in total revenue – holding $share_I$, $share_S$, and the *other factors* fixed, then *growth* increases by $\beta_1(.01)$. With the other shares fixed, the excluded share, $share_F$, must fall by .01 when $share_p$ increases by .01.

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.)

C3.2 (i) The estimated equation is

$$\widehat{price} = -19.32 + .128\, sqrft + 15.20\, bdrms$$

$$n = 88,\ \ R^2 = .632$$

(ii) Holding square footage constant, $\Delta\widehat{price}$ = 15.20 $\Delta bdrms$, and so $\widehat{price}$ increases by 15.20, which means $15,200.

(iii) Now $\Delta\widehat{price}$ = .128 $\Delta sqrft$ + 15.20 $\Delta bdrms$ = .128(140) + 15.20 = 33.12, or $33,120. Because the size of the house is increasing, this is a much larger effect than in (ii).

(iv) About 63.2%.

(v) The predicted price is –19.32 + .128(2,438) + 15.20(4) = 353.544, or $353,544.

(vi) From part (v), the estimated value of the home based only on square footage and number of bedrooms is $353,544. The actual selling price was $300,000, which suggests the buyer underpaid by some margin. But, of course, there are many other features of a house (some that we cannot even measure) that affect price, and we have not controlled for these.

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

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

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

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.

C3.4 (i) The minimum, maximum, and average values for these three variables are given in the table below:

Variable	Average	Minimum	Maximum
atndrte	81.71	6.25	100
priGPA	2.59	.86	3.93
ACT	22.51	13	32

(ii) The estimated equation is

$$\widehat{atndrte} = 75.70 + 17.26\ priGPA - 1.72\ ACT$$

$$n = 680,\ \ R^2 = .291.$$

The intercept means that, for a student whose prior GPA is zero and ACT score is zero, the predicted attendance rate is 75.7%. But this is clearly not an interesting segment of the population. (In fact, there are no students in the college population with *priGPA* = 0 and *ACT* = 0, or with values even close to zero.)

(iii) The coefficient on *priGPA* means that, if a student's prior GPA is one point higher (say, from 2.0 to 3.0), the attendance rate is about 17.3 percentage points higher. This holds *ACT* fixed. The negative coefficient on *ACT* is, perhaps initially a bit surprising. Five more points on the *ACT* is predicted to lower attendance by 8.6 percentage points at a given level of *priGPA*. As *priGPA* measures performance in college (and, at least partially, could reflect, past attendance rates), while *ACT* is a measure of potential in college, it appears that students that had more promise (which could mean more innate ability) think they can get by with missing lectures.

(iv) We have $\widehat{atndrte}$ = 75.70 + 17.267(3.65) – 1.72(20) $\approx$ 104.3. Of course, a student cannot have higher than a 100% attendance rate. Getting predictions like this is always possible when using regression methods for dependent variables with natural upper or lower bounds. In practice, we would predict a 100% attendance rate for this student. (In fact, this student had an actual attendance rate of 87.5%.)

(v) The difference in predicted attendance rates for A and B is 17.26(3.1 – 2.1) – (21 – 26) = 25.86.

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.

C3.6 (i) The slope coefficient from the regression *IQ* on *educ* is (rounded to five decimal places) $\tilde{\delta}_1 = 3.53383$.

(ii) The slope coefficient from log(*wage*) on *educ* is $\tilde{\beta}_1 = .05984$.

(iii) The slope coefficients from log(*wage*) on *educ* and *IQ* are $\hat{\beta}_1 = .03912$ and $\hat{\beta}_2 = .00586$, respectively.

(iv) We have $\hat{\beta}_1 + \tilde{\delta}_1\hat{\beta}_2 = .03912 + 3.53383(.00586) \approx .05983$, which is very close to .05984; the small difference is due to rounding error.

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.

C3.8 (i) The average of *prpblck* is .113 with standard deviation .182; the average of *income* is 47,053.78 with standard deviation 13,179.29. It is evident that *prpblck* is a proportion and that *income* is measured in dollars.

(ii) The results from the OLS regression are

$$\widehat{psoda} = .956 \ + \ .115 \ prpblck \ + \ .0000016 \ income$$

$$n \ = \ 401, \quad R^2 \ = \ .064.$$

If, say, *prpblck* increases by .10 (ten percentage points), the price of soda is estimated to increase by .0115 dollars, or about 1.2 cents. While this does not seem large, there are communities with no black population and others that are almost all black, in which case the difference in *psoda* is estimated to be almost 11.5 cents.

(iii) The simple regression estimate on *prpblck* is .065, so the simple regression estimate is actually lower. This is because *prpblck* and *income* are negatively correlated (-.43) and *income* has a positive coefficient in the multiple regression.

(iv) To get a constant elasticity, income should be in logarithmic form. I estimate the constant elasticity model:

$$\widehat{\log(psoda)} = -.794 + .122\ prpblck + .077\ \log(income)$$

$$n = 401, \quad R^2 = .068.$$

If *prpblck* increases by .20, log(*psoda*) is estimated to increase by .20(.122) = .0244, or about 2.44 percent.

(v) $\hat{\beta}_{prpblck}$ falls to about .073 when *prppov* is added to the regression.

(vi) The correlation is about $-.84$, which makes sense because poverty rates are determined by income (but not directly in terms of median income).

(vii) There is no argument that they are highly correlated, but we are using them simply as controls to determine if the is price discrimination against blacks. In order to isolate the pure discrimination effect, we need to control for as many measures of income as we can; including both variables makes sense.

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

TEACHING NOTES

At the start of this chapter is good time to remind students that a specific error distribution played no role in the results of Chapter 3. That is because only the first two moments were derived under the full set of Gauss-Markov assumptions. Nevertheless, normality is needed to obtain exact normal sampling distributions (conditional on the explanatory variables). I emphasize that the full set of CLM assumptions are used in this chapter, but that in Chapter 5 we relax the normality assumption and still perform approximately valid inference. One could argue that the classical linear model results could be skipped entirely, and that only large-sample analysis is needed. But, from a practical perspective, students still need to know where the *t* distribution comes from because virtually all regression packages report *t* statistics and obtain *p*-values off of the *t* distribution. I then find it very easy to cover Chapter 5 quickly, by just saying we can drop normality and still use *t* statistics and the associated *p*-values as being approximately valid. Besides, occasionally students will have to analyze smaller data sets, especially if they do their own small surveys for a term project.

It is crucial to emphasize that we test hypotheses about unknown population parameters. I tell my students that they will be punished if they write something like H_0: $\hat{\beta}_1 = 0$ on an exam or, even worse, H_0: .632 = 0.

One useful feature of Chapter 4 is its illustration of how to rewrite a population model so that it contains the parameter of interest in testing a single restriction. I find this is easier, both theoretically and practically, than computing variances that can, in some cases, depend on numerous covariance terms. The example of testing equality of the return to two- and four-year colleges illustrates the basic method, and shows that the respecified model can have a useful interpretation. Of course, some statistical packages now provide a standard error for linear combinations of estimates with a simple command, and that should be taught, too.

One can use an *F* test for single linear restrictions on multiple parameters, but this is less transparent than a *t* test and does not immediately produce the standard error needed for a confidence interval or for testing a one-sided alternative. The trick of rewriting the population model is useful in several instances, including obtaining confidence intervals for predictions in Chapter 6, as well as for obtaining confidence intervals for marginal effects in models with interactions (also in Chapter 6).

The major league baseball player salary example illustrates the difference between individual and joint significance when explanatory variables (*rbisyr* and *hrunsyr* in this case) are highly correlated. I tend to emphasize the *R*-squared form of the *F* statistic because, in practice, it is applicable a large percentage of the time, and it is much more readily computed. I do regret that this example is biased toward students in countries where baseball is played. Still, it is one of the better examples of multicollinearity that I have come across, and students of all backgrounds seem to get the point.

SOLUTIONS TO PROBLEMS

4.1 (i) H_0: $\beta_3 = 0$. H_1: $\beta_3 > 0$.

(ii) The proportionate effect on $\widehat{salary}$ is .00024(50) = .012. To obtain the percentage effect, we multiply this by 100: 1.2%. Therefore, a 50 point ceteris paribus increase in *ros* is predicted to increase salary by only 1.2%. Practically speaking, this is a very small effect for such a large change in *ros*.

(iii) The 10% critical value for a one-tailed test, using $df = \infty$, is obtained from Table G.2 as 1.282. The *t* statistic on *ros* is .00024/.00054 $\approx$ .44, which is well below the critical value. Therefore, we fail to reject H_0 at the 10% significance level.

(iv) Based on this sample, the estimated *ros* coefficient appears to be different from zero only because of sampling variation. On the other hand, including *ros* may not be causing any harm; it depends on how correlated it is with the other independent variables (although these are very significant even with *ros* in the equation).

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.

4.5 (i) With $df = n - 2 = 86$, we obtain the 5% critical value from Table G.2 with $df = 90$. Because each test is two-tailed, the critical value is 1.987. The *t* statistic for H_0: $\beta_0 = 0$ is about -.89, which is much less than 1.987 in absolute value. Therefore, we fail to reject $\beta_0 = 0$. The *t* statistic for H_0: $\beta_1 = 1$ is $(.976 - 1)/.049 \approx -.49$, which is even less significant. (Remember, we reject H_0 in favor of H_1 in this case only if $|t| > 1.987$.)

(ii) We use the SSR form of the *F* statistic. We are testing $q = 2$ restrictions and the *df* in the unrestricted model is 86. We are given $SSR_r = 209{,}448.99$ and $SSR_{ur} = 165{,}644.51$. Therefore,

$$F = \frac{(209{,}448.99 - 165{,}644.51)}{165{,}644.51} \cdot \left(\frac{86}{2}\right) \approx 11.37,$$

which is a strong rejection of H_0: from Table G.3c, the 1% critical value with 2 and 90 *df* is 4.85.

(iii) We use the R-squared form of the F statistic. We are testing $q = 3$ restrictions and there are 88 – 5 = 83 *df* in the unrestricted model. The F statistic is [(.829 – .820)/(1 – .829)](83/3) $\approx$ 1.46. The 10% critical value (again using 90 denominator *df* in Table G.3a) is 2.15, so we fail to reject H_0 at even the 10% level. In fact, the p-value is about .23.

(iv) If heteroskedasticity were present, Assumption MLR.5 would be violated, and the F statistic would not have an F distribution under the null hypothesis. Therefore, comparing the F statistic against the usual critical values, or obtaining the p-value from the F distribution, would not be especially meaningful.

4.6 (i) We need to compute the F statistic for the overall significance of the regression with n = 142 and k = 4: F = [.0395/(1 – .0395)](137/4) $\approx$ 1.41. The 5% critical value with 4 numerator *df* and using 120 for the numerator *df*, is 2.45, which is well above the value of F. Therefore, we fail to reject H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ at the 10% level. No explanatory variable is individually significant at the 5% level. The largest absolute t statistic is on *dkr*, $t_{dkr} \approx 1.60$, which is not significant at the 5% level against a two-sided alternative.

(ii) The F statistic (with the same *df*) is now [.0330/(1 – .0330)](137/4) $\approx$ 1.17, which is even lower than in part (i). None of the t statistics is significant at a reasonable level.

(iii) We probably should not use the logs, as the logarithm is not defined for firms that have zero for *dkr* or *eps*. Therefore, we would lose some firms in the regression.

(iv) It seems very weak. There are no significant t statistics at the 5% level (against a two-sided alternative), and the F statistics are insignificant in both cases. Plus, less than 4% of the variation in *return* is explained by the independent variables.

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* = ∞ 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* = ∞: *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.

4.9 (i) H_0: $\beta_3 = 0$. H_1: $\beta_3 \neq 0$.

(ii) Other things equal, a larger population increases the demand for rental housing, which should increase rents. The demand for overall housing is higher when average income is higher, pushing up the cost of housing, including rental rates.

(iii) The coefficient on log(*pop*) is an elasticity. A correct statement is that "a 10% increase in population increases *rent* by .066(10) = .66%."

(iv) With $df = 64 - 4 = 60$, the 1% critical value for a two-tailed test is 2.660. The t statistic is about 3.29, which is well above the critical value. So β_3 is statistically different from zero at the 1% level.

4.10 (i) We use Property VAR.3 from Appendix B: $\text{Var}(\hat{\beta}_1 - 3\hat{\beta}_2) = \text{Var}(\hat{\beta}_1) + 9\,\text{Var}(\hat{\beta}_2) - 6\,\text{Cov}(\hat{\beta}_1, \hat{\beta}_2)$.

(ii) $t = (\hat{\beta}_1 - 3\hat{\beta}_2 - 1)/\text{se}(\hat{\beta}_1 - 3\hat{\beta}_2)$, so we need the standard error of $\hat{\beta}_1 - 3\hat{\beta}_2$.

(iii) Because $\theta_1 = \beta_1 - 3\beta_2$, we can write $\beta_1 = \theta_1 + 3\beta_2$. Plugging this into the population model gives

$$y = \beta_0 + (\theta_1 + 3\beta_2)x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$

$$= \beta_0 + \theta_1 x_1 + \beta_2(3x_1 + x_2) + \beta_3 x_3 + u.$$

This last equation is what we would estimate by regressing y on x_1, $3x_1 + x_2$, and x_3. The coefficient and standard error on x_1 are what we want.

4.11 (i) Holding *profmarg* fixed, $\widehat{\Delta rdintens} = .321\,\Delta\log(sales) = (.321/100)[100\cdot\Delta\log(sales)] \approx .00321(\%\Delta sales)$. Therefore, if $\%\Delta sales = 10$, $\widehat{\Delta rdintens} \approx .032$, 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: $\beta_2 = -\beta_1$.

C4.2 (i) In the model

$$\log(salary) = \beta_0 + \beta_1 LSAT + \beta_2 GPA + \beta_3 \log(libvol) + \beta_4 \log(cost) + \beta_5 rank + u,$$

the hypothesis that *rank* has no effect on log(*salary*) is H_0: $\beta_5 = 0$. The estimated equation (now with standard errors) is

$$\begin{aligned}\widehat{\log(salary)} = \underset{(0.53)}{8.34} + \underset{(.0040)}{.0047}\, LSAT + \underset{(.090)}{.248}\, GPA + \underset{(.033)}{.095} \log(libvol)\\ + \underset{(.032)}{.038} \log(cost) - \underset{(.0003)}{.0033}\, rank\end{aligned}$$

$$n = 136, \quad R^2 = .842.$$

The *t* statistic on *rank* is –11, which is very significant. If *rank* decreases by 10 (which is a move up for a law school), median starting salary is predicted to increase by about 3.3%.

(ii) *LSAT* is not statistically significant (*t* statistic $\approx$ 1.18) but *GPA* is very significance (*t* statistic $\approx$ 2.76). The test for joint significance is moot given that *GPA* is so significant, but for completeness the *F* statistic is about 9.95 (with 2 and 130 *df*) and *p*-value $\approx$ .0001.

(iii) When we add *clsize* and *faculty* to the regression we lose five observations. The test of their joint significant (with 2 and 131 – 8 = 123 *df*) gives $F \approx$ .95 and *p*-value $\approx$ .39. So these two variables are not jointly significant unless we use a very large significance level.

(iv) If we want to just determine the effect of numerical ranking on starting law school salaries, we should control for other factors that affect salaries and rankings. The idea is that there is some randomness in rankings, or the rankings might depend partly on frivolous factors that do not affect quality of the students. LSAT scores and GPA are perhaps good controls for student quality. However, if there are differences in gender and racial composition across schools, and systematic gender and race differences in salaries, we could also control for these. However, it is unclear why these would be correlated with *rank*. Faculty quality, as perhaps measured by publication records, could be included. Such things do enter rankings of law schools.

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 $\mathrm{se}(\hat{\theta}_1) = .0268$. The 95% confidence interval reported by my software package is .0326 to .1390 (or about 3.3% to 13.9%).

C4.4 The R-squared from the regression *bwght* on *cigs*, *parity*, and *faminc*, using all 1,388 observations, is about .0348. This means that, if we mistakenly use this in place of .0364, which is the R-squared using the same 1,191 observations available in the unrestricted regression, we would obtain $F = [(.0387 - .0348)/(1 - .0387)](1{,}185/2) \approx 2.40$, which yields p-value $\approx .091$ in an F distribution with 2 and 1,1185 *df*. This is significant at the 10% level, but it is incorrect. The correct F statistic was computed as 1.42 in Example 4.9, with p-value $\approx .242$.

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

$$\begin{aligned}\widehat{\log(salary)} = \underset{(0.27)}{11.02} &+ \underset{(.0121)}{.0677}\, years + \underset{(.0016)}{.0158}\, gamesyr \\ &+ \underset{(.0011)}{.0014}\, bavg + \underset{(.0072)}{.0359}\, hrunsyr\end{aligned}$$
$$n = 353, \quad R^2 = .625.$$

Now *hrunsyr* is very statistically significant (t statistic ≈ 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.

C4.6 (i) In the model

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + u$$

the null hypothesis of interest is H_0: $\beta_2 = \beta_3$.

(ii) Let $\theta_2 = \beta_2 - \beta_3$. Then we can estimate the equation

$$\log(wage) = \beta_0 + \beta_1 educ + \theta_2 exper + \beta_3 (exper + tenure) + u$$

to obtain the 95% CI for θ_2. This turns out to be about .0020 ± 1.96(.0047), or about -.0072 to .0112. Because zero is in this CI, θ_2 is not statistically different from zero at the 5% level, and we fail to reject H_0: $\beta_2 = \beta_3$ at the 5% level.

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:

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

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

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.

C4.8 (i) There are 2,017 single people in the sample of 9,275.

(ii) The estimated equation is

$$\widehat{nettfa} = \underset{(4.08)}{-43.04} + \underset{(.060)}{.799}\, inc + \underset{(.092)}{.843}\, age$$

$$n = 2{,}017,\ R^2 = .119.$$

The coefficient on *inc* indicates that one more dollar in income (holding *age* fixed) is reflected in about 80 more cents in predicted *nettfa*; no surprise there. The coefficient on *age* means that, holding income fixed, if a person gets another year older, his/her *nettfa* is predicted to increase by about \$843. (Remember, *nettfa* is in thousands of dollars.) Again, this is not surprising.

(iii) The intercept is not very interesting as it gives the predicted *nettfa* for *inc* = 0 and *age* = 0. Clearly, there is no one with even close to these values in the relevant population.

(iv) The *t* statistic is $(.843 - 1)/.092 \approx -1.71$. Against the one-sided alternative H_1: $\beta_2 < 1$, the p-value is about .044. Therefore, we can reject H_0: $\beta_2 = 1$ at the 5% significance level (against the one-sided alternative).

(v) The slope coefficient on *inc* in the simple regression is about .821, which is not very different from the .799 obtained in part (ii). As it turns out, the correlation between *inc* and *age* in the sample of single people is only about .039, which helps explain why the simple and multiple regression estimates are not very different; refer back to page 84 of the text.

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.

C4.10 (i) Using the 1,848 observations, the simple regression estimate of β_{bs} is about $-.795$. The 95% confidence interval runs from -1.088 to $-.502$, which includes -1. Therefore, at the 5% level, we cannot reject that $\text{H}_0: \beta_{bs} = -1$ against the two-sided alternative.

(ii) When *lenrol* and *lstaff* are added to the regression, the coefficient on *bs* becomes about –.605; it is now statistically different from one, as the 95% CI is from about –.818 to –.392. The situation is very similar to that in Table 4.1, where the simple regression estimate is –.825 and the multiple regression estimate (with the logs of enrollment and staff included) is –.605. (It is a coincidence that the two multiple regression estimates are the same, as the data set in Table 4.1 is for an earlier year at the high school level.)

(iii) The standard error of the simple regression estimate is about .150, and that for the multiple regression estimate is about .109. When we add extra explanatory variables, two factors work in opposite directions on the standard errors. Multicollinearity – in this case, correlation between *bs* and the two variables *lenrol* and *lstaff* works to increase the multiple regression standard error. Working to reduce the standard error of $\hat{\beta}_{bs}$ is the smaller error variance when *lenrol* and *lstaff* are included in the regression; in effect, they are taken out of the simple regression error term. In this particular example, the multicollinearity is modest compared with the reduction in the error variance. In fact, the standard error of the regression goes from .231 for simple regression to .168 in the multiple regression. (Another way to summarize the drop in the error variance is to note that the R-squared goes from a very small .0151 for the simple regression to .4882 for multiple regression.) Of course, ahead of time we cannot know which effect will dominate, but we can certainly compare the standard errors after running both regressions.

(iv) The variable *lstaff* is the log of the number of staff per 1,000 students. As *lstaff* increases, there are more teachers per student. We can associate this with smaller class sizes, which are generally desirable from a teacher's perspective. It appears that, all else equal, teachers are willing to take less in salary to have smaller class sizes. The elasticity of *salary* with respect to *staff* is about –.714, which seems quite large: a ten percent increase in staff size (holding enrollment fixed) is associated with a 7.14 percent lower salary.

(v) When *lunch* is added to the regression, its coefficient is about –.00076, with $t = -4.69$. Therefore, other factors fixed (*bs*, *lenrol*, and *lstaff*), a hire poverty rate is associated with *lower* teacher salaries. In this data set, the average value of *lunch* is about 36.3 with standard deviation of 25.4. Therefore, a one standard deviation increase in *lunch* is associated with a change in *lsalary* of about $-.00076(25.4) \approx -.019$, or almost two percent lower. Certainly there is no evidence that teachers are compensated for teaching disadvantaged children.

(vi) Yes, the pattern obtained using ELEM94_95.RAW is very similar to that in Table 4.1, and the magnitudes are reasonably close, too. The largest estimate (in absolute value) is the simple regression estimate, and the absolute value declines as more explanatory variables are added. The final regressions in the two cases are not the same, because we do not control for *lunch* in Table 4.1, and graduation and dropout rates are not relevant for elementary school children.

CHAPTER 5

TEACHING NOTES

Chapter 5 is short, but it is conceptually more difficult than the earlier chapters, primarily because it requires some knowledge of asymptotic properties of estimators. In class, I give a brief, heuristic description of consistency and asymptotic normality before stating the consistency and asymptotic normality of OLS. (Conveniently, the same assumptions that work for finite sample analysis work for asymptotic analysis.) More advanced students can follow the proof of consistency of the slope coefficient in the bivariate regression case. Section E.4 contains a full matrix treatment of asymptotic analysis appropriate for a master's level course.

An explicit illustration of what happens to standard errors as the sample size grows emphasizes the importance of having a larger sample. I do not usually cover the *LM* statistic in a first-semester course, and I only briefly mention the asymptotic efficiency result. Without full use of matrix algebra combined with limit theorems for vectors and matrices, it is very difficult to prove asymptotic efficiency of OLS.

I think the conclusions of this chapter are important for students to know, even though they may not fully grasp the details. On exams I usually include true-false type questions, with explanation, to test the students' understanding of asymptotics. [For example: "In large samples we do not have to worry about omitted variable bias." (False). Or "Even if the error term is not normally distributed, in large samples we can still compute approximately valid confidence intervals under the Gauss-Markov assumptions." (True).]

SOUTH-WESTERN CENGAGE Learning

SOLUTIONS TO PROBLEMS

5.1 Write y = $\beta_0 + \beta_1 x + u$, and take the expected value: E(y) = $\beta_0 + \beta_1$E(x) + E(u), or μ_y = $\beta_0 + \beta_1\mu_x$, since E(u) = 0, where μ_y = E(y) and μ_x = E(x). We can rewrite this as $\beta_0 = \mu_y - \beta_1\mu_x$. Now, $\tilde{\beta}_0 = \bar{y} - \tilde{\beta}_1\bar{x}$. Taking the plim of this we have plim($\tilde{\beta}_0$) = plim($\bar{y} - \tilde{\beta}_1\bar{x}$) = plim($\bar{y}$) – plim($\tilde{\beta}_1$)·plim($\bar{x}$) = $\mu_y - \beta_1\mu_x$, where we use the fact that plim($\bar{y}$) = μ_y and plim($\bar{x}$) = μ_x by the law of large numbers, and plim($\tilde{\beta}_1$) = β_1. We have also used the parts of the Property PLIM.2 from Appendix C.

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.

5.3 A higher tolerance of risk means more willingness to invest in the stock market, so $\beta_2 > 0$. By assumption, *funds* and *risktol* are positively correlated. Now we use equation (5.5), where $\delta_1 > 0$: plim($\tilde{\beta}_1$) = $\beta_1 + \beta_2\delta_1 > \beta_1$, so $\tilde{\beta}_1$ has a positive inconsistency (asymptotic bias). This makes sense: if we omit *risktol* from the regression and it is positively correlated with *funds,* some of the estimated effect of *funds* is actually due to the effect of *risktol*.

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 μ_y = $\beta_0 + \beta_1\mu_x$ since E(u) = 0, where μ_y = E(y) and μ_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 plim($\hat{\beta}_0$) = plim($\bar{y} - \hat{\beta}_1\bar{x}_1$) = plim($\bar{y}$) – plim($\hat{\beta}_1$)·plim($\bar{x}_1$) = $\mu_y - \beta_1\mu_x$, where we use the fact that plim($\bar{y}$) = μ_y and plim($\bar{x}_1$) = μ_x by the law of large numbers, and plim($\hat{\beta}_1$) = β_1. We have also used the parts of Property PLIM.2 from Appendix C.

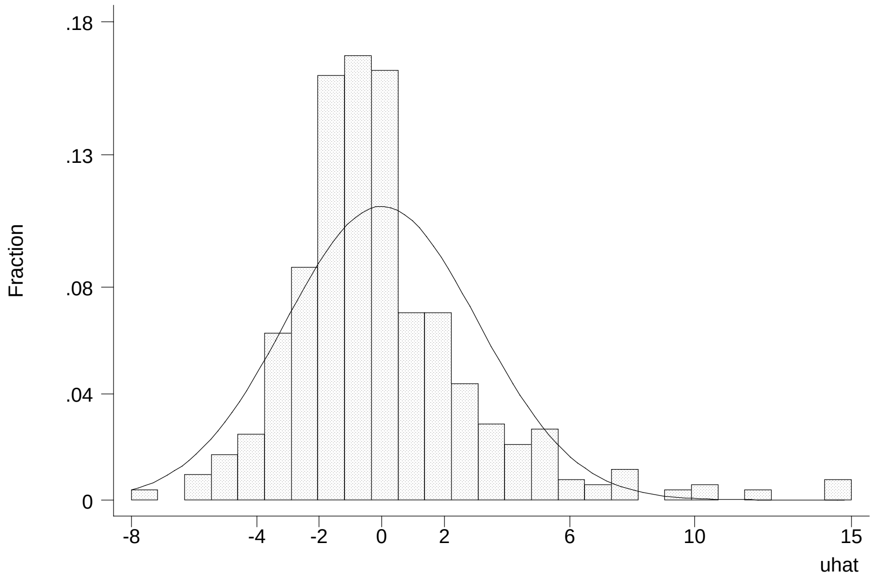
SOLUTIONS TO COMPUTER EXERCISES

C5.1 (i) The estimated equation is

$$\widehat{wage} = -2.87 + .599\, educ + .022\, exper + .169\, tenure$$
$$\quad\;\; (0.73)\;\; (.051) \qquad\quad (.012) \qquad\quad (.022)$$

$$n = 526, \quad R^2 = .306, \quad \hat{\sigma} = 3.085.$$

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.

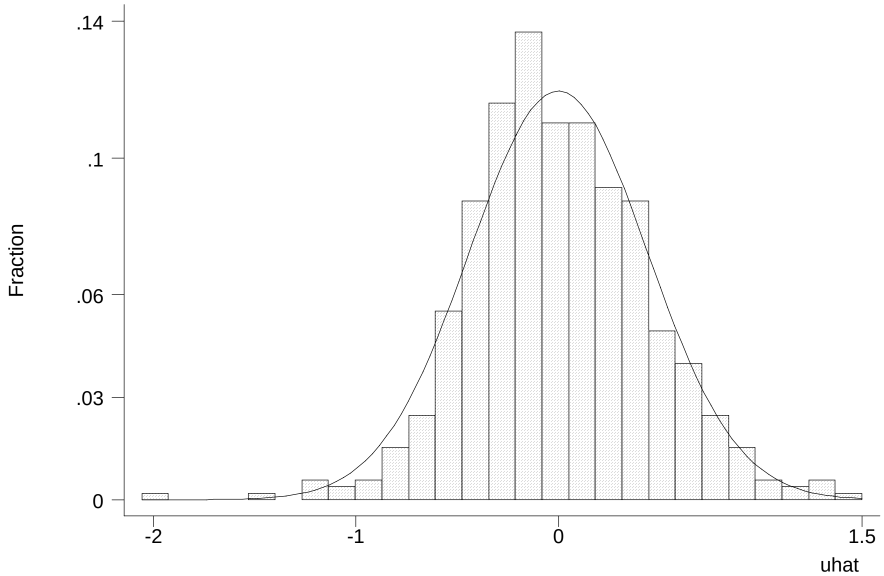


(ii) With log(*wage*) as the dependent variable the estimated equation is

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

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

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.

C5.2 (i) The regression with all 4,137 observations is

$$\widehat{colgpa} = \underset{(0.072)}{1.392} - \underset{(.00055)}{.01352}\, hsperc + \underset{(.00007)}{.00148}\, sat$$

$$n = 4{,}137, \quad R^2 = .273.$$

(ii) Using only the first 2,070 observations gives

$$\widehat{colgpa} = \underset{(0.098)}{1.436} - \underset{(.00072)}{.01275}\, hsperc + \underset{(.00009)}{.00147}\, sat$$

$$n = 2{,}070, \quad R^2 = .283.$$

(iii) The ratio of the standard error using 2,070 observations to that using 4,137 observations is about 1.31. From (5.10) we compute $\sqrt{(4{,}137/2{,}070)} \approx 1.41$, which is somewhat above the ratio of the actual standard errors.

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.

C5.4 (i) The measure of skewness for *inc* is about 1.86. When we use log(*inc*), the skewness measure is about .360. Therefore, there is much less skewness in log of income, which means *inc* is less likely to be normally distributed. (In fact, the skewness in income distributions is a well-documented fact across many countries and time periods.)

(ii) The skewness for *bwght* is about −.60. When we use log(*bwght*), the skewness measure is about −2.95. In this case, there is much more skewness after taking the natural log.

(iii) The example in part (ii) clearly shows that this statement cannot hold generally. It is possible to introduce skewness by taking the natural log. As an empirical matter, for many economic variables, particularly dollar values, taking the log often does help to reduce or eliminate skewness. But it does not *have* to.

(iv) For the purposes of regression analysis, we should be studying the *conditional* distributions; that is, the distributions of *y* and log(*y*) conditional on the explanatory variables, $x_1, \ldots, x_k$. If we think the mean is linear, as in Assumptions MLR.1 and MLR.3, then this is equivalent to studying the distribution of the population error, *u*. In fact, the skewness measure studied in this question often is applied to the residuals from and OLS regression.

CHAPTER 6

TEACHING NOTES

I cover most of Chapter 6, but not all of the material in great detail. I use the example in Table 6.1 to quickly run through the effects of data scaling on the important OLS statistics. (Students should already have a feel for the effects of data scaling on the coefficients, fitting values, and *R*-squared because it is covered in Chapter 2.) At most, I briefly mention beta coefficients; if students have a need for them, they can read this subsection.

The functional form material is important, and I spend some time on more complicated models involving logarithms, quadratics, and interactions. An important point for models with quadratics, and especially interactions, is that we need to evaluate the partial effect at interesting values of the explanatory variables. Often, zero is not an interesting value for an explanatory variable and is well outside the range in the sample. Using the methods from Chapter 4, it is easy to obtain confidence intervals for the effects at interesting x values.

As far as goodness-of-fit, I only introduce the adjusted *R*-squared, as I think using a slew of goodness-of-fit measures to choose a model can be confusing to novices (and does not reflect empirical practice). It is important to discuss how, if we fixate on a high *R*-squared, we may wind up with a model that has no interesting ceteris paribus interpretation.

I often have students and colleagues ask if there is a simple way to predict y when $\log(y)$ has been used as the dependent variable, and to obtain a goodness-of-fit measure for the $\log(y)$ model that can be compared with the usual *R*-squared obtained when y is the dependent variable. The methods described in Section 6.4 are easy to implement and, unlike other approaches, do not require normality.

The section on prediction and residual analysis contains several important topics, including constructing prediction intervals. It is useful to see how much wider the prediction intervals are than the confidence interval for the conditional mean. I usually discuss some of the residual-analysis examples, as they have real-world applicability.

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.

6.3 (i) Holding all other factors fixed we have

$$\Delta \log(wage) = \beta_1 \Delta educ + \beta_2 \Delta educ \cdot pareduc = (\beta_1 + \beta_2 pareduc)\Delta educ.$$

Dividing both sides by $\Delta educ$ gives the result. The sign of β_2 is not obvious, although $\beta_2 > 0$ if we think a child gets more out of another year of education the more highly educated are the child's parents.

(ii) We use the values *pareduc* = 32 and *pareduc* = 24 to interpret the coefficient on $educ \cdot pareduc$. The difference in the estimated return to education is .00078(32 – 24) = .0062, or about .62 percentage points.

(iii) When we add *pareduc* by itself, the coefficient on the interaction term is negative. The *t* statistic on $educ \cdot pareduc$ is about –1.33, which is not significant at the 10% level against a two-sided alternative. Note that the coefficient on *pareduc* is significant at the 5% level against a two-sided alternative. This provides a good example of how omitting a level effect (*pareduc* in this case) can lead to biased estimation of the interaction effect.

6.4 (i) The answer is not entirely obvious, but one must properly interpret the coefficient on *alcohol* in either case. If we include *attend*, then we are measuring the effect of alcohol consumption on college GPA, holding attendance fixed. Because attendance is likely to be an important mechanism through which drinking affects performance, we probably do not want to hold it fixed in the analysis. If we do include *attend*, then we interpret the estimate of $\beta_{alcohol}$ as measuring those effects on *colGPA* that are not due to attending class. (For example, we could be measuring the effects that drinking alcohol has on study time.) To get a total effect of alcohol consumption, we would leave *attend* out.

(ii) We would want to include *SAT* and *hsGPA* as controls, as these measure student abilities and motivation. Drinking behavior in college could be correlated with one's performance in high school and on standardized tests. Other factors, such as family background, would also be good controls.

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.

6.7 By definition of the OLS regression of $c_0 y_i$ on $c_1 x_{i1}$, …, $c_k x_{ik}$, $i = 2$, …, n, the $\tilde{\beta}_j$ solve

$$\sum_{i=1}^{n}[(c_0 y_i) - \tilde{\beta}_0 - \tilde{\beta}_1(c_1 x_{i1}) - \ldots - \tilde{\beta}_k(c_k x_{ik})] = 0$$

$$\sum_{i=1}^{n}(c_1 x_{i1})[(c_0 y_i) - \tilde{\beta}_0 - \tilde{\beta}_1(c_1 x_{i1}) - \ldots - \tilde{\beta}_k(c_k x_{ik})] = 0$$

$$\vdots$$

$$\sum_{i=1}^{n}(c_k x_{ik})[(c_0 y_i) - \tilde{\beta}_0 - \tilde{\beta}_1(c_1 x_{i1}) - \ldots - \tilde{\beta}_k(c_k x_{ik})] = 0.$$

[We obtain these from equations (3.13), where we plug in the scaled dependent and independent variables.] We now show that if $\tilde{\beta}_0 = c_0\hat{\beta}_0$ and $\tilde{\beta}_j = (c_0 / c_j)\tilde{\beta}_j$, $j = 1,\ldots,k$, then these $k + 1$ first order conditions are satisfied, which proves the result because we know that the OLS estimates are the unique solutions to the FOCs (once we rule out perfect collinearity in the independent variables). Plugging in these guesses for the $\tilde{\beta}_j$ gives the expressions

$$\sum_{i=1}^{n}[(c_0 y_i) - c_0\hat{\beta}_0 - (c_0 / c_1)\hat{\beta}_1(c_1 x_{i1}) - \ldots - (c_0 / c_k)\hat{\beta}_k(c_k x_{ik})]$$

$$\sum_{i=1}^{n}(c_j x_{ij})[(c_0 y_i) - c_0\hat{\beta}_0 - (c_0 / c_1)\hat{\beta}_1(c_1 x_{i1}) - \ldots - (c_0 / c_k)\hat{\beta}_k(c_k x_{ik})],$$

for $j = 1,2,\ldots,k$. Simple cancellation shows we can write these equations as

$$\sum_{i=1}^{n}[(c_0 y_i) - c_0\hat{\beta}_0 - c_0\hat{\beta}_1 x_{i1} - \ldots - c_0\hat{\beta}_k x_{ik}]$$

and

$$\sum_{i=1}^{n}(c_j x_{ij})[(c_0 y_i) - c_0\hat{\beta}_0 - c_0\hat{\beta}_1 x_{i1} - \ldots - c_0\hat{\beta}_k x_{ik}],\ j = 1,2,\ldots,k$$

or, factoring out constants,

$$c_0\left(\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik})\right)$$

and

$$c_0 c_j\left(\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik})\right),\ j = 1, 2, \ldots$$

But the terms multiplying c_0 and c_0c_j are identically zero by the first order conditions for the $\hat{\beta}_j$ since, by definition, they are obtained from the regression y_i on $x_{i1}, \ldots, x_{ik}$, $i = 1,2,..,n$. So we have shown that $\tilde{\beta}_0 = c_0\hat{\beta}_0$ and $\tilde{\beta}_j = (c_0/c_j)\hat{\beta}_j$, $j = 1, \ldots, k$ solve the requisite first order conditions.

6.8 The extended model has $df = 680 - 9 = 671$, and we are testing two restrictions. Therefore, $F = [(.232 - .229)/(1 - .232)](671/2) \approx 1.31$, which is well below the 10% critical value in the F distribution with 2 and ∞ df: $cv = 2.30$. Thus, $atndrte^2$ and $ACT \cdot atndrte$ are jointly insignificant. Because adding these terms complicates the model without statistical justification, we would not include them in the final model.

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 ($se \approx .058$). 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)

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

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

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.

C6.2 (i) The estimated equation is

$$\widehat{\log(wage)} = .128 + .0904\, educ + .0410\, exper - .000714\, exper^2$$
$$\quad\quad (.106) \quad (.0075) \quad\quad (.0052) \quad\quad (.000116)$$

$n = 526$, $R^2 = .300$, $\bar{R}^2 = .296$.

(ii) The t statistic on $exper^2$ is about –6.16, which has a p-value of essentially zero. So *exper* is significant at the 1% level(and much smaller significance levels).

(iii) To estimate the return to the fifth year of experience, we start at *exper* = 4 and increase *exper* by one, so $\Delta exper = 1$:

$$\%\Delta\widehat{wage} \approx 100[.0410 - 2(.000714)4] \approx 3.53\%.$$

Similarly, for the 20th year of experience,

$$\%\Delta\widehat{wage} \approx 100[.0410 - 2(.000714)19] \approx 1.39\%$$

(iv) The turnaround point is about $.041/[2(.000714)] \approx 28.7$ years of experience. In the sample, there are 121 people with at least 29 years of experience. This is a fairly sizeable fraction of the sample.

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.

C6.4 (i) The estimated equation is

$$\begin{aligned}\widehat{sat} &= \underset{(6.20)}{997.98} + \underset{(3.99)}{19.81}\, hsize - \underset{(0.55)}{2.13}\, hsize^2\\ n &= 4{,}137, \quad R^2 = .0076.\end{aligned}$$

The quadratic term is very statistically significant, with *t* statistic ≈ -3.87.

(ii) We want the value of *hsize*, say *hsize**, where $\hat{sat}$ reaches its maximum. This is the turning point in the parabola, which we calculate as *hsize** = 19.81/[2(2.13)] $\approx$ 4.65. Since *hsize* is in 100s, this means 465 students is the "optimal" class size. Of course, the very small R-squared shows that class size explains only a tiny amount of the variation in SAT score.

(iii) Only students who actually take the SAT exam appear in the sample, so it is not representative of all high school seniors. If the population of interest is all high school seniors, we need a random sample of such students who all took the same standardized exam.

(iv) With log(*sat*) as the dependent variable we get

$$\widehat{\log(sat)} = \underset{(0.006)}{6.896} + \underset{(.0040)}{.0196}\, hsize - \underset{(.00054)}{.00209}\, hsize^2$$

$$n = 4{,}137, \quad R^2 = .0078.$$

The optimal class size is now estimated as about 469, which is very close to what we obtained with the level-level model.

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.

C6.6 (i) For the model

$$voteA = \beta_0 + \beta_1 prtystrA + \beta_2 expendA + \beta_3 expendB + \beta_4 expendA \cdot expendB + u,$$

the ceteris paribus effect of *expendB* on *voteA* is obtained by taking changes and holding *prtystrA*, *expendA*, and *u* fixed:

$$\Delta voteA = \beta_3 \Delta expendB + \beta_4 expendA(\Delta expendB) = (\beta_3 + \beta_4 expendA) \Delta expendB,$$

or

$$\Delta voteA/\Delta expendB = \beta_3 + \beta_4 expendA.$$

We think $\beta_3 < 0$ if a ceteris paribus increase in spending by B lowers the share of the vote received by A. But the sign of β_4 is ambiguous: Is the effect of more spending by B smaller or larger for higher levels of spending by A?

(ii) The estimated equation is

$$\begin{aligned}\widehat{voteA} = \underset{(4.59)}{32.12} + \underset{(.088)}{.342}\, prtystrA + \underset{(.0050)}{.0383}\, expendA - \underset{(.0046)}{.0317}\, expendB \\ - \underset{(.0000072)}{.0000066}\, expendA \cdot expendB\end{aligned}$$

$$n = 173, \quad R^2 = .571, \quad \bar{R}^2 = .561.$$

The interaction term is not statistically significant, as its *t* statistic is less than one in absolute value.

(iii) The average value of *expendA* is about 310.61, or \$310,610. If we set *expendA* at 300, which is close to the average value, we have

$$\Delta\widehat{voteA} = [-.0317 - .0000066 \cdot (300)]\, \Delta expendB \approx -.0337(\Delta expendB).$$

So, when $\Delta expendB = 100$, $\Delta\widehat{voteA} \approx -3.37$, which is a fairly large effect. (Note that, given the insignificance of the interaction term, we would be justified in leaving it out and reestimating the model. This would make the calculation easier.)

(iv) Now we have

$$\Delta\widehat{voteA} = (\hat{\beta}_2 + \hat{\beta}_4 expendB)\Delta expendA \approx .0376(\Delta expendA) = 3.76$$

when $\Delta expendA = 100$. This does make sense, and it is a nontrivial effect.

(v) When we replace the interaction term with *shareA* we obtain

$$\widehat{voteA} = \underset{(2.57)}{18.20} + \underset{(.050)}{.157}\, prtystrA - \underset{(.0028)}{.0067}\, expendA + \underset{(.0026)}{.0043}\, expendB + \underset{(.025)}{.494}\, shareA$$

$n = 173$, $R^2 = .868$, $\bar{R}^2 = .865$.

Notice how much higher the goodness-of-fit measures are as compared with the equation estimated in part (ii), and how significant *shareA* is. To obtain the partial effect of *expendB* on $\widehat{voteA}$ we must compute the partial derivative. Generally, we have

$$\frac{\partial \widehat{voteA}}{\partial expendB} = \hat{\beta}_3 + \hat{\beta}_4\left(\frac{\partial shareA}{\partial expendB}\right),$$

where *shareA* = 100[*expendA*/(*expendA* + *expendB*)]. Now

$$\frac{\partial shareA}{\partial expendB} = -100\left(\frac{expendA}{(expendA + expendB)^2}\right).$$

Evaluated at *expendA* = 300 and *expendB* = 0, the partial derivative is $-100(300/300^2) = -1/3$, and therefore

$$\frac{\partial \widehat{voteA}}{\partial expendB} = \hat{\beta}_3 + \hat{\beta}_4(1/3) = .0043 - .494/3 \approx -.164.$$

So $\widehat{voteA}$ falls by .164 percentage points given the first thousand dollars of spending by candidate B, where A's spending is held fixed at 300 (or \$300,000). This is a fairly large effect, although it may not be the most typical scenario (because it is rare to have one candidate spend so much and another spend so little). The effect tapers off as *expendB* grows. For example, at *expendB* = 100, the effect of the thousand dollars of spending is only about $.0043 - .494(.188) \approx -.089$.

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 $\hat{\theta}_2$.

C6.8 (i) The estimated equation (where *price* is in dollars) is

$$\begin{aligned} \widehat{price} = &-21{,}770.3 + 2.068\, lotsize + 122.78\, sqrft + 13{,}852.5\, bdrms \\ &(29{,}475.0) \quad (0.642) \qquad (13.24) \qquad (9{,}010.1) \end{aligned}$$

$$n = 88,\quad R^2 = .672,\quad \bar{R}^2 = .661,\quad \hat{\sigma} = 59{,}833.$$

The predicted price at *lotsize* = 10,000, *sqrft* = 2,300, and *bdrms* = 4 is about \$336,714.

(ii) The regression is $price_i$ on ($lotsize_i$ – 10,000), ($sqrft_i$ – 2,300), and ($bdrms_i$ – 4). We want the intercept estimate and the associated 95% CI from this regression. The CI is approximately 336,706.7 ± 14,665, or about \$322,042 to \$351,372 when rounded to the nearest dollar.

(iii) We must use equation (6.36) to obtain the standard error of $\hat{e}^0$ and then use equation (6.37) (assuming that *price* is normally distributed). But from the regression in part (ii), se($\hat{y}^0$) ≈ 7,374.5 and $\hat{\sigma} \approx$ 59,833. Therefore, se($\hat{e}^0$) ≈ $[(7{,}374.5)^2 + (59{,}833)^2]^{1/2} \approx$ 60,285.8. Using 1.99 as the approximate 97.5th percentile in the t_{84} distribution gives the 95% CI for $price^0$, at the given values of the explanatory variables, as 336,706.7 ± 1.99(60,285.8) or, rounded to the nearest dollar, \$216,738 to \$456,675. This is a fairly wide prediction interval. But we have not used many factors to explain housing price. If we had more, we could, presumably, reduce the error standard deviation, and therefore $\hat{\sigma}$, to obtain a tighter prediction interval.

C6.9 (i) The estimated equation is

$$\widehat{points} = \underset{(6.99)}{35.22} + \underset{(.405)}{2.364}\, exper - \underset{(.0235)}{.0770}\, exper^2 - \underset{(.295)}{1.074}\, age - \underset{(.451)}{1.286}\, coll$$

$n = 269$, $R^2 = .141$, $\bar{R}^2 = .128$.

(ii) The turnaround point is 2.364/[2(.0770)] ≈ 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

$$\widehat{\log(wage)} = \underset{(.85)}{6.78} + \underset{(.007)}{.078}\, points + \underset{(.050)}{.218}\, exper - \underset{(.0028)}{.0071}\, exper^2 - \underset{(.035)}{.048}\, age - \underset{(.053)}{.040}\, coll$$

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

(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.

C6.10 (i) The estimated equation is

$$\widehat{\log(bwght)} = \underset{(.027)}{7.958} + \underset{(.0037)}{.0189}\, npvis - \underset{(.00012)}{.00043}\, npvis^2$$

$n = 1{,}764$, $R^2 = .0213$, $\bar{R}^2 = .0201$

The quadratic term is very significant as its t statistic is above 3.5 in absolute value.

(ii) The turning point calculation is by now familiar: $npvis^* = .0189/[2(.00043)] \approx 21.97$, or about 22. In the sample, 89 women had 22 or more prenatal visits.

(iii) While prenatal visits are a good thing for helping to prevent low birth weight, a woman's having many prenatal visits is a possible indicator of a pregnancy with difficulties. So it does make sense that the quadratic has a hump shape, provided we do not interpret the turnaround as implying that too many visits actually *causes* low birth weight.

(iv) With *mage* added in quadratic form, we get

$$\widehat{\log(bwght)} = \underset{(.137)}{7.584} + \underset{(.0037)}{.0180}\, npvis - \underset{(.00012)}{.00041}\, npvis^2 + \underset{(.0093)}{.0254}\, mage - \underset{(.00015)}{.00041}\, mage^2$$

$n = 1{,}764$, $R^2 = .0256$, $\bar{R}^2 = .0234$

The birth weight is maximized at $mage \approx 31$. 746 women are at least 31 years old; 605 are at least 32.

(v) These variables explain on the order of 2.6% of the variation in log(*bwght*) (or even less according to $\bar{R}^2$), which is not very much.

(vi) If we regress *bwght* on *npvis*, $npvis^2$, *mage*, and $mage^2$, then $R^2 = .0192$. But remember, we cannot compare this directly with the *R*-squared reported in part (iv). Instead, we compute an *R*-squared for the log(*bwght*) model that can be compared with .0192. From Section 6.4, we compute the squared correlation between *bwght* and $\exp(\widehat{lbwght})$, where $\widehat{lbwght}$ denotes the fitted values from the log(*bwght*) model. The correlation is .1362, so its square is about .0186. Therefore, for explaining *bwght*, the model with *bwght* as the dependent variable actually fits slightly better (but nothing to make a big deal about).

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.

C6.12 (i) The youngest age is 25, and there are 99 people of this age in the sample with *fsize* = 1.

(ii) One literal interpretation is that β_2 is the increase in *nettfa* when *age* increases by one year, holding fixed *inc* and age^2. Of course, it makes no sense to change *age* while keeping age^2 fixed. Alternatively, because $\partial nettfa / \partial age = \beta_2 + 2\beta_3 age$, β_2 is the approximate increase in *nettfa* when *age* increases from zero to one. But in this application, the partial effect starting at *age* = 0 is not interesting; the sample represents single people at least 25 years old.

(iii) The OLS estimates are

$$\widehat{nettfa} = \underset{(15.28)}{-1.20} + \underset{(.060)}{.825}\, inc - \underset{(0.767)}{1.322}\, age + \underset{(.0090)}{.0256}\, age^2$$

$$n = 2{,}017,\ R^2 = .1229,\ \bar{R}^2 = .1216$$

Initially, the negative coefficient on *age* may seem counterintuitive. The estimated relationship is a U-shape, but, to make sense of it, we need to find the turning point in the quadratic. From equation (6.13), the estimated turning point is 1.322/[2(.0256)] ≈ 25.8. Interestingly, this is very close to the youngest age in the sample. In other words, starting at roughly *age* = 25, the relationship between *nettfa* and *age* is positive – as we might expect. So, in this case, the negative coefficient on *age* makes sense when we compute the partial effect.

(iv) I follow the hint, form the new regressor $(age-25)^2$, and run the regression *nettfa* on *inc*, *age*, and $(age-25)^2$. This changes the intercept (which we are not concerned with, anyway) and the coefficient on *age*, which is simply $\beta_2 + 2\beta_3(25)$ – the partial effect evaluated at $age = 25$. The results are

$$\widehat{nettfa} = \underset{(9.97)}{-17.20} + \underset{(.060)}{.825\, inc} - \underset{(.767)}{.0437\, age} + \underset{(.0090)}{.0256\,(age - 25)^2}$$

$$n = 2{,}017,\ R^2 = .1229,\ \bar{R}^2 = .1216$$

Therefore, the estimated partial effect starting at *age* = 25 is only –.044 and very statistically insignificant ($t = -.13$). The two-sided *p*-value is about .89.

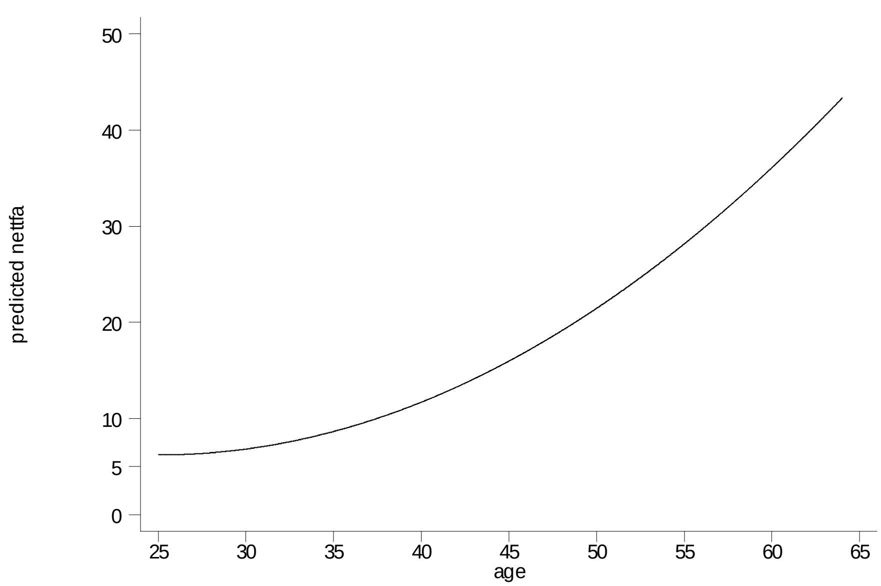
(v) If we drop *age* from the regression in part (iv) we get

$$\widehat{nettfa} = \underset{(2.18)}{-18.49} + \underset{(.060)}{.824\, inc} + \underset{(.0025)}{.0244\,(age - 25)^2}$$

$$n = 2{,}017,\ R^2 = .1229,\ \bar{R}^2 = .1220$$

The adjusted *R*-squared is slightly higher when we drop *age*. But the real reason for dropping *age* is that its *t* statistic is quite small, and the model without it has a straightforward interpretation.

(vi) The graph of the relationship estimated in (v), with *inc* = 30, is



The slope of the relationship between $\widehat{nettfa}$ and *age* is clearly increasing. That is, there is an increasing marginal effect. The model is constructed so that the slope is zero at *age* = 25; from there, the slope increases.

(vii) When inc^2 is added to the regression in part (v) its coefficient is only $-.00054$ with $t = -0.27$. Thus, the linear relationship between *nettfa* and *inc* is not rejected, and we would exclude the squared income term.

C6.13 (i) The estimated equation is

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

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

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 $p\text{-value} \approx .67$. 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

TEACHING NOTES

This is a fairly standard chapter on using qualitative information in regression analysis, although I try to emphasize examples with policy relevance (and only cross-sectional applications are included.).

In allowing for different slopes, it is important, as in Chapter 6, to appropriately interpret the parameters and to decide whether they are of direct interest. For example, in the wage equation where the return to education is allowed to depend on gender, the coefficient on the female dummy variable is the wage differential between women and men at zero years of education. It is not surprising that we cannot estimate this very well, nor should we want to. In this particular example we would drop the interaction term because it is insignificant, but the issue of interpreting the parameters can arise in models where the interaction term is significant.

In discussing the Chow test, I think it is important to discuss testing for differences in slope coefficients after allowing for an intercept difference. In many applications, a significant Chow statistic simply indicates intercept differences. (See the example in Section 7.4 on student-athlete GPAs in the text.) From a practical perspective, it is important to know whether the partial effects differ across groups or whether a constant differential is sufficient.

I admit that an unconventional feature of this chapter is its introduction of the linear probability model. I cover the LPM here for several reasons. First, the LPM is being used more and more because it is easier to interpret than probit or logit models. Plus, once the proper parameter scalings are done for probit and logit, the estimated effects are often similar to the LPM partial effects near the mean or median values of the explanatory variables. The theoretical drawbacks of the LPM are often of secondary importance in practice. Computer Exercise C7.9 is a good one to illustrate that, even with over 9,000 observations, the LPM can deliver fitted values strictly between zero and one for all observations.

If the LPM is not covered, many students will never know about using econometrics to explain qualitative outcomes. This would be especially unfortunate for students who might need to read an article where an LPM is used, or who might want to estimate an LPM for a term paper or senior thesis. Once they are introduced to purpose and interpretation of the LPM, along with its shortcomings, they can tackle nonlinear models on their own or in a subsequent course.

A useful modification of the LPM estimated in equation (7.29) is to drop *kidsge6* (because it is not significant) and then define two dummy variables, one for *kidslt6* equal to one and the other for *kidslt6* at least two. These can be included in place of *kidslt6* (with no young children being the base group). This allows a diminishing marginal effect in an LPM. I was a bit surprised when a diminishing effect did not materialize.

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 ≈ 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 ≈ –9.06, which is very statistically significant. The coefficient implies that one more hour of work (60 minutes) is associated with .163(60) ≈ 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.

7.2 (i) If $\Delta cigs$ = 10 then $\Delta\widehat{\log(bwght)}$ = –.0044(10) = –.044, which means about a 4.4% lower birth weight.

(ii) A white child is estimated to weigh about 5.5% more, other factors in the first equation fixed. Further, t_{white} ≈ 4.23, which is well above any commonly used critical value. Thus, the difference between white and nonwhite babies is also statistically significant.

(iii) If the mother has one more year of education, the child's birth weight is estimated to be .3% higher. This is not a huge effect, and the t statistic is only one, so it is not statistically significant.

(iv) The two regressions use different sets of observations. The second regression uses fewer observations because *motheduc* or *fatheduc* are missing for some observations. We would have to reestimate the first equation (and obtain the R-squared) using the same observations used to estimate the second equation.

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 $\widehat{sat}$ (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.

7.4 (i) The approximate difference is just the coefficient on *utility* times 100, or –28.3%. The *t* statistic is $-.283/.099 \approx -2.86$, which is very statistically significant.

(ii) $100\cdot[\exp(-.283)-1] \approx -24.7\%$, and so the estimate is somewhat smaller in magnitude.

(iii) The proportionate difference is .181 – .158 = .023, or about 2.3%. One equation that can be estimated to obtain the standard error of this difference is

$$\log(salary) = \beta_0 + \beta_1\log(sales) + \beta_2 roe + \delta_1 consprod + \delta_2 utility + \delta_3 trans + u,$$

where *trans* is a dummy variable for the transportation industry. Now, the base group is *finance*, and so the coefficient δ_1 directly measures the difference between the consumer products and finance industries, and we can use the *t* statistic on *consprod*.

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).

7.6 In Section 3.3 – in particular, in the discussion surrounding Table 3.2 – we discussed how to determine the direction of bias in the OLS estimators when an important variable (ability, in this case) has been omitted from the regression. As we discussed there, Table 3.2 only strictly holds with a single explanatory variable included in the regression, but we often ignore the presence of other independent variables and use this table as a rough guide. (Or, we can use the results of Problem 3.10 for a more precise analysis.) If less able workers are more likely to receive training, then *train* and *u* are negatively correlated. If we ignore the presence of *educ* and *exper*, or at least assume that *train* and *u* are negatively correlated after netting out *educ* and *exper*, then we can use Table 3.2: the OLS estimator of β_1 (with ability in the error term) has a downward bias. Because we think $\beta_1 \geq 0$, we are less likely to conclude that the training program was effective. Intuitively, this makes sense: if those chosen for training had not received training, they would have lowers wages, on average, than the control group.

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 $R^2 = 1 - \text{SSR/SST}$, the R-squared is the same in the two regressions.

7.8 (i) We want to have a constant semi-elasticity model, so a standard wage equation with marijuana usage included would be

$$\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u.$$

Then $100\cdot\beta_1$ is the approximate percentage change in *wage* when marijuana usage increases by one time per month.

(ii) We would add an interaction term in *female* and *usage*:

$$\log(wage) = \beta_0 + \beta_1 usage + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + \beta_6 female\cdot usage + u.$$

The null hypothesis that the effect of marijuana usage does not differ by gender is H_0: $\beta_6 = 0$.

(iii) We take the base group to be nonuser. Then we need dummy variables for the other three groups: *lghtuser*, *moduser*, and *hvyuser*. Assuming no interactive effect with gender, the model would be

$$\log(wage) = \beta_0 + \delta_1 lghtuser + \delta_2 moduser + \delta_3 hvyuser + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 female + u.$$

(iv) The null hypothesis is H_0: $\delta_1 = 0$, $\delta_2 = 0$, $\delta_3 = 0$, for a total of $q = 3$ restrictions. If n is the sample size, the *df* in the unrestricted model – the denominator *df* in the *F* distribution – is $n - 8$. So we would obtain the critical value from the $F_{q,n-8}$ distribution.

(v) The error term could contain factors, such as family background (including parental history of drug abuse) that could directly affect wages and also be correlated with marijuana usage. We are interested in the effects of a person's drug usage on his or her wage, so we would like to hold other confounding factors fixed. We could try to collect data on relevant background information.

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.

7.10 (i) Yes, simple regression does produce an unbiased estimator of the effect of the voucher program. Because participation was randomized, we can write

$$score = \beta_0 + \beta_1 voucher + u,$$

where *voucher* is independent of *u*, that is, all other factors affecting *score*. Therefore, the key assumption for unbiasedness of simple regression, Assumption SLR.3, is satisfied.

(ii) No, we do not need to control for background variables. In the equation from part (i), these are factors in the error term, *u*. But *voucher* was assigned to be independent of all factors, including the listed background variables.

(iii) We should include the background variables to reduce the sampling error of the estimated voucher effect. By pulling background variables out of the error term, we reduce the error variance – perhaps substantially. Further, we can be sure that multicollinearity is not a problem because the key variable of interest, *voucher*, is uncorrelated with all of the added explanatory variables. (This zero correlation will only be approximate in any random sample, but in large samples it should be very small.) The one case where we would not add these variables – or, at least, when there is no benefit from doing so – is when the background variables themselves have no affect on the test score. Given the list of background variables, this seems unlikely in the current application.

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.

C7.2 (i) The estimated equation is

$$\widehat{\log(wage)} = \underset{(0.11)}{5.40} + \underset{(.0063)}{.0654}\, educ + \underset{(.0032)}{.0140}\, exper + \underset{(.0025)}{.0117}\, tenure$$

$$+ \underset{(.039)}{.199}\, married - \underset{(.038)}{.188}\, black - \underset{(.026)}{.091}\, south + \underset{(.027)}{.184}\, urban$$

$n = 935,\ R^2 = .253.$

The coefficient on *black* implies that, at given levels of the other explanatory variables, black men earn about 18.8% less than nonblack men. The t statistic is about –4.95, and so it is very statistically significant.

(ii) The F statistic for joint significance of $exper^2$ and $tenure^2$, with 2 and 925 df, is about 1.49 with p-value $\approx$ .226. Because the p-value is above .20, these quadratics are jointly insignificant at the 20% level.

(iii) We add the interaction $black \cdot educ$ to the equation in part (i). The coefficient on the interaction is about $-.0226$ (se $\approx$ .0202). Therefore, the point estimate is that the return to another year of education is about 2.3 percentage points lower for black men than nonblack men. (The estimated return for nonblack men is about 6.7%.) This is nontrivial if it really reflects differences in the population. But the t statistic is only about 1.12 in absolute value, which is not enough to reject the null hypothesis that the return to education does not depend on race.

(iv) We choose the base group to be single, nonblack. Then we add dummy variables *marrnonblck*, *singblck*, and *marrblck* for the other three groups. The result is

$$\widehat{\log(wage)} = \underset{(0.11)}{5.40} + \underset{(.0063)}{.0655}\,educ + \underset{(.0032)}{.0141}\,exper + \underset{(.0025)}{.0117}\,tenure$$
$$- \underset{(.026)}{.092}\,south + \underset{(.027)}{.184}\,urban + \underset{(.043)}{.189}\,marrnonblck$$
$$- \underset{(.096)}{.241}\,singblck + \underset{(.0560)}{.0094}\,marrblck$$
$$n = 935, \quad R^2 = .253.$$

We obtain the ceteris paribus differential between married blacks and married nonblacks by taking the difference of their coefficients: .0094 − .189 = −.1796, or about −.18. That is, a married black man earns about 18% less than a comparable, married nonblack man.

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] ≈ 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).

C7.4 (i) The two signs that are pretty clear are $\beta_3 < 0$ (because *hsperc* is defined so that the smaller the number the better the student) and $\beta_4 > 0$. The effect of size of graduating class is not clear. It is also unclear whether males and females have systematically different GPAs. We may think that $\beta_6 < 0$, that is, athletes do worse than other students with comparable characteristics. But remember, we are controlling for ability to some degree with *hsperc* and *sat*.

(ii) The estimated equation is

$$\widehat{colgpa} = \underset{(0.079)}{1.241} - \underset{(.0164)}{.0569}\,hsize + \underset{(.00225)}{.00468}\,hsize^2 - \underset{(.0006)}{.0132}\,hsperc$$
$$+ \underset{(.00007)}{.00165}\,sat + \underset{(.018)}{.155}\,female + \underset{(.042)}{.169}\,athlete$$
$$n = 4{,}137, \quad R^2 = .293.$$

Holding other factors fixed, an athlete is predicted to have a GPA about .169 points *higher* than a nonathlete. The t statistic $.169/.042 \approx 4.02$, which is very significant.

(iii) With *sat* dropped from the model, the coefficient on *athlete* becomes about .0054 (se $\approx$ .0448), which is practically and statistically not different from zero. This happens because we do not control for SAT scores, and athletes score lower on average than nonathletes. Part (ii) shows that, once we account for SAT differences, athletes do better than nonathletes. Even if we do not control for SAT score, there is no difference.

(iv) To facilitate testing the hypothesis that there is no difference between women athletes and women nonathletes, we should choose one of these as the base group. We choose female nonathletes. The estimated equation is

$$\begin{aligned}\widehat{colgpa} = \underset{(0.076)}{1.396} &- \underset{(.0164)}{.0568}\, hsize + \underset{(.00225)}{.00467}\, hsize^2 - \underset{(.0006)}{.0132}\, hsperc \\ &+ \underset{(.00007)}{.00165}\, sat + \underset{(.084)}{.175}\, femath + \underset{(.049)}{.013}\, maleath - \underset{(.018)}{.155}\, malenonath\end{aligned}$$

$$n = 4{,}137, \quad R^2 = .293.$$

The coefficient on $femath = female \cdot athlete$ shows that *colgpa* is predicted to be about .175 points higher for a female athlete than a female nonathlete, other variables in the equation fixed. The hypothesis that there is no difference between female athletes and female nonathletes is testing by using the t statistic on *femath*. In this case, $t = 2.08$, which is statistically significant at the 5% level against a two-sided alternative.

(v) Whether we add the interaction $female \cdot sat$ to the equation in part (ii) or part (iv), the outcome is practically the same. For example, when $female \cdot sat$ is added to the equation in part (ii), its coefficient is about .000051 and its t statistic is about .40. There is very little evidence that the effect of *sat* differs by gender.

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.

C7.6 (i) The estimated equation for men is

$$\widehat{sleep} = \underset{(310.0)}{3{,}648.2} - \underset{(.024)}{.182}\ totwrk - \underset{(7.41)}{13.05}\ educ + \underset{(14.32)}{7.16}\ age - \underset{(.1684)}{.0448}\ age^2 + \underset{(59.02)}{60.38}\ yngkid$$

$n = 400$, $R^2 = .156$

and the estimated equation for women is

$$\widehat{sleep} = \underset{(384.9)}{4{,}238.7} - \underset{(.028)}{.140}\ totwrk - \underset{(9.59)}{10.21}\ educ - \underset{(18.53)}{30.36}\ age - \underset{(.223)}{.368}\ age^2 - \underset{(93.19)}{118.28}\ yngkid$$

$n = 306$, $R^2 = .098$.

There are certainly notable differences in the point estimates. For example, having a young child in the household leads to less sleep for women (about two hours a week) while men are estimated to sleep about an hour more. The quadratic in *age* is a hump-shape for men but a U-shape for women. The intercepts for men and women are also notably different.

(ii) The F statistic (with 6 and 694 *df*) is about 2.12 with p-value $\approx$ .05, and so we reject the null that the sleep equations are the same at the 5% level.

(iii) If we leave the coefficient on *male* unspecified under H_0, and test only the five interaction terms, $male \cdot totwrk$, $male \cdot educ$, $male \cdot age$, $male \cdot age^2$, and $male \cdot yngkid$, the F statistic (with 5 and 694 *df*) is about 1.26 and p-value $\approx$ .28.

(iv) The outcome of the test in part (iii) shows that, once an intercept difference is allowed, there is not strong evidence of slope differences between men and women. This is one of those cases where the practically important differences in estimates for women and men in part (i) do not translate into statistically significant differences. We need a larger sample size to confidently determine whether there are differences in slopes. For the purposes of studying the sleep-work tradeoff, the original model with *male* added as an explanatory variable seems sufficient.

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 $educ = 0$.

C7.8 (i) If the appropriate factors have been controlled for, $\beta_1 > 0$ signals discrimination against minorities: a white person has a greater chance of having a loan approved, other relevant factors fixed.

(ii) The simple regression results are

$$\begin{aligned}\widehat{approve} &= \underset{(.018)}{.708} + \underset{(.020)}{.201}\, white \\ n &= 1{,}989,\quad R^2 = .049.\end{aligned}$$

The coefficient on *white* means that, in the sample of 1,989 loan applications, an application submitted by a white application was 20.1% more likely to be approved than that of a nonwhite applicant. This is a practically large difference and the t statistic is about 10. (We have a large sample size, so standard errors are pretty small.)

(iii) When we add the other explanatory variables as controls, we obtain $\hat{\beta}_1 \approx .129$, se($\hat{\beta}_1$) $\approx$ .020. The coefficient has fallen by some margin because we are now controlling for factors that should affect loan approval rates, and some of these clearly differ by race. (On average, white people have financial characteristics – such as higher incomes and stronger credit histories – that make them better loan risks.) But the race effect is still strong and very significant (t statistic $\approx$ 6.45).

(iv) When we add the interaction $white \cdot obrat$ to the regression, its coefficient and t statistic are about .0081 and 3.53, respectively. Therefore, there is an interactive effect: a white applicant is penalized less than a nonwhite applicant for having other obligations as a larger percent of income.

(v) The trick should be familiar by now. Replace $white \cdot obrat$ with $white \cdot (obrat - 32)$; the coefficient on *white* is now the race differential when $obrat = 32$. We obtain about .113 and se $\approx$ .020. So the 95% confidence interval is about .113 ± 1.96(.020) or about .074 to .152. Clearly, this interval excludes zero, so at the average *obrat* there is evidence of discrimination (or, at least loan approval rates that differ by race for some other reason that is not captured by the control variables).

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

(ii) The estimated equation is

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

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

(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

$$\widehat{e401k} = \underset{(.081)}{-.502} + \underset{(.0006)}{.0123}\, inc - \underset{(.000005)}{.000061}\, inc^2 + \underset{(.0039)}{.0265}\, age - \underset{(.00005)}{.00031}\, age^2$$
$$- \underset{(.0121)}{.0038}\, male + \underset{(.0122)}{.0198}\, pira$$
$$n = 9{,}275, \quad R^2 = .095.$$

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.

C7.10 (i) The estimated equation is

$$\widehat{points} = \underset{(1.18)}{4.76} + \underset{(.33)}{1.28}\, exper - \underset{(.024)}{.072}\, exper^2 + \underset{(1.00)}{2.31}\, guard + \underset{(1.00)}{1.54}\, forward$$
$$n = 269, \quad R^2 = .091, \quad \bar{R}^2 = .077.$$

(ii) Including all three position dummy variables would be redundant, and result in the dummy variable trap. Each player falls into one of the three categories, and the overall intercept is the intercept for centers.

(iii) A guard is estimated to score about 2.3 points more per game, holding experience fixed. The t statistic is 2.31, so the difference is statistically different from zero at the 5% level, against a two-sided alternative.

(iv) When *marr* is added to the regression, its coefficient is about .584 (se = .740). Therefore, a married player is estimated to score just over half a point more per game (experience and position held fixed), but the estimate is not statistically different from zero (p-value = .43). So, based on points per game, we cannot conclude married players are more productive.

(v) Adding the terms $marr \cdot exper$ and $marr \cdot exper^2$ leads to complicated signs on the three terms involving *marr*. The F test for their joint significance, with 3 and 261 *df*, gives $F = 1.44$ and p-value = .23. Therefore, there is not very strong evidence that marital status has any partial effect on points scored.

(vi) If in the regression from part (iv) we use *assists* as the dependent variable, the coefficient on *marr* becomes .322 (se = .222). Therefore, holding experience and position fixed, a married man has almost one-third more assist per game. The p-value against a two-sided alternative is about .15, which is stronger, but not overwhelming, evidence that married men are more productive when it comes to assists.

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): 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 SSR_{ur} = 29,985,400. The Chow statistic is F = [(30,215,207.5 – 29,985,400)/ 29,985,400]*(9245/20) ≈ 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.)

C7.12 (i) For women, the fraction rated as having above average looks is about .33; for men, it is .29. The proportion of women rated as having below average looks is only .135; for men, it is even lower at about .117.

(ii) The difference is about .04, that is, the percent rated as having above average looks is about four percentage points higher for women than men. A simple way to test whether the difference is statistically significant is to run a simple regression of *abvavg* on *female* and do a t test (which is asymptotically valid). The t statistic is about 1.48 with two-sided *p*-value = .14. Therefore, there is not strong evidence against the null that the population fractions are the same, but there is some evidence.

(iii) The regression for men is

$$\widehat{\log(wage)} = \underset{(0.024)}{1.884} - \underset{(.060)}{.199}\, belavg - \underset{(.042)}{.044}\, abvavg$$

$$n = 824 \quad R^2 = .013$$

and the regression for women is

$$\widehat{\log(wage)} = \underset{(0.034)}{1.309} - \underset{(.076)}{.138}\, belavg + \underset{(.055)}{.034}\, abvavg$$

$$n = 436 \quad R^2 = .011.$$

Using the standard approximation, a man with below average looks earns almost 20% less than a man of average looks, and a woman with below average looks earns about 13.8% less than a woman with average looks. (The more accurate estimates are about 18% and 12.9%, respectively.) The null hypothesis H_0: $\beta_1 = 0$ against H_1: $\beta_1 < 0$ means that the null is that people with below average looks earn the same, on average, as people with average looks; the alternative is that people with below average looks earn less than people with average looks (in the population). The one-sided *p*-value for men is .0005 and for women it is .036. We reject H_0 more strongly for men because the estimate is larger in magnitude and the estimate has less sampling variation (as measured by the standard error).

(iv) Women with above average looks are estimated to earn about 3.4% more, on average, than women with average looks. But the one-sided *p*-value is .272, and this provides very little evidence against H_0: $\beta_2 = 0$.

(v) Given the number of added controls, with many of them very statistically significant, the coefficients on the looks variables do not change by much. For men, the coefficient on *belavg* becomes −.143 (t = −2.80) and the coefficient on *abvavg* becomes −.001 (t = −.03). For women, the changes in magnitude are similar: the coefficient on *belavg* becomes −.115 (t = −1.75) and the coefficient on *abvavg* becomes .058 (t = 1.18). In both cases, the estimates on *belavg* move closer to zero but are still reasonably large.

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%.)

C7.14 (i) The estimated LPM is

$$\widehat{respond} = \underset{(.009)}{.282} + \underset{(.015)}{.344}\, resplast + \underset{(.00009)}{.00015}\, avggift$$

n = 4,268, R^2 = .110

Holding the average gift fixed, the probability of a current response is estimated to be .344 higher if the person responded most recently.

(ii) Once we control for responding most recently, the effect of of *avggift* is very small. Even if *avggift* is 100 guilders more (the mean is about 18.2 with standard deviation 78.7), the probability of responding this period is only .015 higher. Plus, the t statistic has a two-sided p-value of about .09, so it is only marginally statistically significant.

(iii) The coefficient on *propresp* is about .747 (standard error = .034). If *propresp* increases by .1 (for example, from .4 to .5), the probability of responding is about .075 higher.

(iv) When *propresp* is added to the regression, the coefficient on *resplast* falls to about .095 (although it is still very statistically significant). This makes sense, because the relationship between responding currently and responding most recently should be weaker once the average response is controlled for. Certainly *resplast* and *propresp* are positively correlated.

(v) The coefficient on *mailsyear* is about .062 (t = 6.18). This is a reasonably large effect: each new mailing is estimated to increase the probability of responding by .062. Unfortunately, we do not know how the charitable organization determines the mailings sent. To the extent that it depends only on past gift giving, as controlled for by the average gift, the most recent response, and the response rate, the estimate could be a good (consistent) estimate of the causal effect. But if mailings are determined by other factors that are necessarily in the error term – such as income – then the estimate would be systematically biased. If, say, more mailings are sent to people with higher incomes, and higher income people are more likely to respond, then the regression that omits income produces an upward bias for the *mailsyear* coefficient.

CHAPTER 8

TEACHING NOTES

This is a good place to remind students that homoskedasticity played no role in showing that OLS is unbiased for the parameters in the regression equation. In addition, you probably should mention that there is nothing wrong with the *R*-squared or adjusted *R*-squared as goodness-of-fit measures. The key is that these are estimates of the population *R*-squared, $1 - [\text{Var}(u)/\text{Var}(y)]$, where the variances are the *unconditional* variances in the population. The usual *R*-squared, and the adjusted version, consistently estimate the population *R*-squared whether or not $\text{Var}(u|\mathbf{x}) = \text{Var}(y|\mathbf{x})$ depends on **x**. Of course, heteroskedasticity causes the usual standard errors, *t* statistics, and *F* statistics to be invalid, even in large samples, with or without normality.

By explicitly stating the homoskedasticity assumption as conditional on the explanatory variables that appear in the conditional mean, it is clear that only heteroskedasticity that depends on the explanatory variables in the model affects the validity of standard errors and test statistics. The version of the Breusch-Pagan test in the text, and the White test, are ideally suited for detecting forms of heteroskedasticity that invalidate inference obtained under homoskedasticity. If heteroskedasticity depends on an exogenous variable that does not also appear in the mean equation, this can be exploited in weighted least squares for efficiency, but only rarely is such a variable available. One case where such a variable is available is when an individual-level equation has been aggregated. I discuss this case in the text but I rarely have time to teach it.

As I mention in the text, other traditional tests for heteroskedasticity, such as the Park and Glejser tests, do not directly test what we want, or add too many assumptions under the null. The Goldfeld-Quandt test only works when there is a natural way to order the data based on one independent variable. This is rare in practice, especially for cross-sectional applications.

Some argue that weighted least squares estimation is a relic, and is no longer necessary given the availability of heteroskedasticity-robust standard errors and test statistics. While I am sympathetic to this argument, it presumes that we do not care much about efficiency. Even in large samples, the OLS estimates may not be precise enough to learn much about the population parameters. With substantial heteroskedasticity we might do better with weighted least squares, even if the weighting function is misspecified. As discussed in the text on pages 287-288, one can, and probably should, compute robust standard errors after weighted least squares. For asymptotic efficiency comparisons, these would be directly comparable to the heteroskedasiticity-robust standard errors for OLS.

Weighted least squares estimation of the LPM is a nice example of feasible GLS, at least when all fitted values are in the unit interval. Interestingly, in the LPM examples in the text and the LPM computer exercises, the heteroskedasticity-robust standard errors often differ by only small amounts from the usual standard errors. However, in a couple of cases the differences are notable, as in Computer Exercise C8.7.

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.

8.2 With $\text{Var}(u|inc,price,educ,female) = \sigma^2 inc^2$, $h(\mathbf{x}) = inc^2$, where $h(\mathbf{x})$ is the heteroskedasticity function defined in equation (8.21). Therefore, $\sqrt{h(\mathbf{x})} = inc$, and so the transformed equation is obtained by dividing the original equation by *inc*:

$$\frac{beer}{inc} = \beta_0(1/inc) + \beta_1 + \beta_2(price/inc) + \beta_3(educ/inc) + \beta_4(female/inc) + (u/inc).$$

Notice that β_1, which is the slope on *inc* in the original model, is now a constant in the transformed equation. This is simply a consequence of the form of the heteroskedasticity and the functional forms of the explanatory variables in the original equation.

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.

8.4 (i) These coefficients have the anticipated signs. If a student takes courses where grades are, on average, higher – as reflected by higher *crsgpa* – then his/her grades will be higher. The better the student has been in the past – as measured by *cumgpa* – the better the student does (on average) in the current semester. Finally, *tothrs* is a measure of experience, and its coefficient indicates an increasing return to experience.

The t statistic for *crsgpa* is very large, over five using the usual standard error (which is the largest of the two). Using the robust standard error for *cumgpa*, its t statistic is about 2.61, which is also significant at the 5% level. The t statistic for *tothrs* is only about 1.17 using either standard error, so it is not significant at the 5% level.

(ii) This is easiest to see without other explanatory variables in the model. If *crsgpa* were the only explanatory variable, H_0: $\beta_{crsgpa} = 1$ means that, without any information about the student, the best predictor of term GPA is the average GPA in the students' courses; this holds essentially by definition. (The intercept would be zero in this case.) With additional explanatory variables it is not necessarily true that $\beta_{crsgpa} = 1$ because *crsgpa* could be correlated with characteristics of the student. (For example, perhaps the courses students take are influenced by ability – as measured by test scores – and past college performance.) But it is still interesting to test this hypothesis.

The t statistic using the usual standard error is $t = (.900 - 1)/.175 \approx -.57$; using the heteroskedasticity-robust standard error gives $t \approx -.60$. In either case we fail to reject H_0: $\beta_{crsgpa} = 1$ at any reasonable significance level, certainly including 5%.

(iii) The in-season effect is given by the coefficient on *season*, which implies that, other things equal, an athlete's GPA is about .16 points lower when his/her sport is competing. The t statistic using the usual standard error is about –1.60, while that using the robust standard error is about –1.96. Against a two-sided alternative, the t statistic using the robust standard error is just significant at the 5% level (the standard normal critical value is 1.96), while using the usual standard error, the t statistic is not quite significant at the 10% level ($cv \approx 1.65$). So the standard error used makes a difference in this case. This example is somewhat unusual, as the robust standard error is more often the larger of the two.

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.)

8.6 (i) The proposed test is a hybrid of the BP and White tests. There are $k + 1$ regressors, each original explanatory variable and the squared fitted values. So, the number of restrictions tested is $k + 1$, and this is the numerator df. The denominator df is $n - (k + 2) = n - k - 2$.

(ii) For the BP test, this is easy: the hybrid test has an extra regressor, $\hat{y}^2$, and so the R-squared will be no less for the hybrid test than for the BP test. For the special case of the White test, the argument is a bit more subtle. In regression (8.20), the fitted values are a linear function of the regressors (where, of course, the coefficients in the linear function are the OLS estimates). So, we are putting a restriction on how the original explanatory variables appear in the regression. This means that the R-squared from (8.20) will be no greater than the R-squared from the hybrid regression.

(iii) No. The F statistic for joint significance of the regressors depends on $R^2_{\hat{u}^2}/(1 - R^2_{\hat{u}^2})$, and it is true that this ratio increases as $R^2_{\hat{u}^2}$ increases. But, the F statistic also depends on the df, and the *df* are different among all three tests: the BP test, the special case of the White test, and the hybrid test. So we do not know which test will deliver the smallest *p*-value.

(iv) As discussed in part (ii), the OLS fitted values are a linear combination of the original regressors. Because those regressors appear in the hybrid test, adding the OLS fitted values is redundant; perfect collinearity would result.

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

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

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 $male_i$, $i = 1,2, \ldots, 706$ (including, of course, an intercept). We can write the results as

$$\hat{u}^2 = \underset{(20{,}546.4)}{189{,}359.2} - \underset{(27{,}296.5)}{28{,}849.6}\, male + residual$$

$$n = 706, \quad R^2 = .0016.$$

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.

C8.2 (i) The estimated equation with both sets of standard errors (heteroskedasticity-robust standard errors in parentheses) is

$$\widehat{price} = \underset{\substack{(29.48)\\ [36.28]}}{-21.77} + \underset{\substack{(.00064)\\ [.00122]}}{.00207}\, lotsize + \underset{\substack{(.013)\\ [.017]}}{.123}\, sqrft + \underset{\substack{(9.01)\\ [8.28]}}{13.85}\, bdrms$$

$$n = 88, \quad R^2 = .672.$$

The robust standard error on *lotsize* is almost twice as large as the usual standard error, making *lotsize* much less significant (the *t* statistic falls from about 3.23 to 1.70). The *t* statistic on *sqrft* also falls, but it is still very significant. The variable *bdrms* actually becomes somewhat more significant, but it is still barely significant. The most important change is in the significance of *lotsize*.

(ii) For the log-log model,

$$\widehat{\log(price)} = \underset{(0.65)\ [0.76]}{5.61} + \underset{(.038)\ [.041]}{.168} \log(lotsize) + \underset{(.093)\ [.101]}{.700} \log(sqrft) + \underset{(.028)\ [.030]}{.037}\, bdrms$$

$$n = 88, \quad R^2 = .643.$$

Here, the heteroskedasticity-robust standard error is always slightly greater than the corresponding usual standard error, but the differences are relatively small. In particular, log(*lotsize*) and log(*sqrft*) still have very large *t* statistics, and the *t* statistic on *bdrms* is not significant at the 5% level against a one-sided alternative using either standard error.

(iii) As we discussed in Section 6.2, using the logarithmic transformation of the dependent variable often mitigates, if not entirely eliminates, heteroskedasticity. This is certainly the case here, as no important conclusions in the model for log(*price*) depend on the choice of standard error. (We have also transformed two of the independent variables to make the model of the constant elasticity variety in *lotsize* and *sqrft*.)

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.

C8.4 (i) The estimated equation is

$$\widehat{voteA} = \underset{(4.74)}{37.66} + \underset{(.071)}{.252}\, prtystrA + \underset{(1.407)}{3.793}\, democA + \underset{(0.392)}{5.779} \log(expendA)$$
$$- \underset{(0.397)}{6.238} \log(expendB) + \hat{u}$$

$$n = 173, \quad R^2 = .801, \quad \bar{R}^2 = .796.$$

You can convince yourself that regressing the $\hat{u}_i$ on all of the explanatory variables yields an *R*-squared of zero, although it might not be exactly zero in your computer output due to rounding error. Remember, OLS works by choosing the estimates, $\hat{\beta}_j$, such that the residuals are uncorrelated in the sample with each independent variable (and the residuals have a zero sample average, too).

(ii) The B-P test entails regressing the $\hat{u}_i^2$ on the independent variables in part (i). The *F* statistic for joint significant (with 4 and 168 *df*) is about 2.33 with *p*-value $\approx$ .058. Therefore, there is some evidence of heteroskedasticity, but not quite at the 5% level.

(iii) Now we regress $\hat{u}_i^2$ on $\widehat{voteA_i}$ and $(\widehat{voteA_i})^2$, where the $\widehat{voteA_i}$ are the OLS fitted values from part (i). The *F* test, with 2 and 170 *df*, is about 2.79 with *p*-value $\approx$ .065. This is slightly less evidence of heteroskedasticity than provided by the B-P test, but the conclusion is very similar.

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

$$\underset{}{\widehat{sprdcvr}} = \underset{(.045)}{.490} + \underset{(.050)}{.035}\, favhome + \underset{(.095)}{.118}\, neutral - \underset{(.050)}{.023}\, fav25 + \underset{(.092)}{.018}\, und25$$

$$n = 553, \quad R^2 = .0034.$$

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.

C8.6 (i) The estimates are given in equation (7.31). Rounded to four decimal places, the smallest fitted value is .0066 and the largest fitted value is .5577.

(ii) The estimated heteroskedasticity function for each observation *i* is $\hat{h}_i = \widehat{arr86_i}(1 - \widehat{arr86_i})$, which is strictly between zero and one because $0 < \widehat{arr86_i} < 1$ for all *i*.

The weights for WLS are $1/\hat{h}_i$. To show the WLS estimate of each parameter, we report the WLS results using the same equation format as for OLS:

$$\begin{aligned}\widehat{arr86} &= \underset{(.018)}{.448} - \underset{(.019)}{.168}\, pcnv + \underset{(.0051)}{.0054}\, avgsen - \underset{(.0033)}{.0018}\, tottime - \underset{(.003)}{.025}\, ptime86 \\ &\quad - \underset{(.005)}{.045}\, qemp86\end{aligned}$$

$n = 2{,}725, \quad R^2 = .0744.$

The coefficients on the significant explanatory variables are very similar to the OLS estimates. The WLS standard errors on the slope coefficients are generally lower than the nonrobust OLS standard errors. A proper comparison would be with the robust OLS standard errors.

(iii) After WLS estimation, the *F* statistic for joint significance of *avgsen* and *tottime*, with 2 and 2,719 *df*, is about .88 with *p*-value $\approx$ .41. They are not close to being jointly significant at the 5% level. If your econometrics package has a command for WLS and a test command for joint hypotheses, the *F* statistic and *p*-value are easy to obtain. Alternatively, you can obtain the restricted *R*-squared using the same weights as in part (ii) and dropping *avgsen* and *tottime* from the WLS estimation. (The unrestricted *R*-squared is .0744.)

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.

C8.8 (i) The equation estimated by OLS is

$$\widehat{colGPA} = \underset{(.33)}{1.36} + \underset{(.092)}{.412}\, hsGPA + \underset{(.010)}{.013}\, ACT - \underset{(.026)}{.071}\, skipped + \underset{(.057)}{.124}\, PC$$

$n = 141, \quad R^2 = .259, \quad \bar{R}^2 = .238$

(ii) The *F* statistic obtained for the White test is about 3.58. With 2 and 138 *df*, this gives *p*-value $\approx$ .031. So, at the 5% level, we conclude there is evidence of heteroskedasticity in the errors of the *colGPA* equation. (As an aside, note that the *t* statistics for each of the terms is very small, and we could have simply dropped the quadratic term without losing anything of value.)

(iii) In fact, the smallest fitted value from the regression in part (ii) is about .027, while the largest is about .165. Using these fitted values as the $\hat{h}_i$ in a weighted least squares regression gives the following:

$$\widehat{colGPA} = \underset{(.30)}{1.40} + \underset{(.083)}{.402}\, hsGPA + \underset{(.010)}{.013}\, ACT - \underset{(.022)}{.076}\, skipped + \underset{(.056)}{.126}\, PC$$

$n = 141,\ R^2 = .306,\ \bar{R}^2 = .286$

There is very little difference in the estimated coefficient on *PC*, and the OLS *t* statistic and WLS *t* statistic are also very close. Note that we have used the usual OLS standard error, even though it would be more appropriate to use the heteroskedasticity-robust form (since we have evidence of heteroskedasticity). The *R*-squared in the weighted least squares estimation is larger than that from the OLS regression in part (i), but, remember, these are not comparable.

(iv) With robust standard errors – that is, with standard errors that are robust to misspecifying the function $h(\mathbf{x})$ – the equation is

$$\widehat{colGPA} = \underset{(.31)}{1.40} + \underset{(.086)}{.402}\, hsGPA + \underset{(.010)}{.013}\, ACT - \underset{(.021)}{.076}\, skipped + \underset{(.059)}{.126}\, PC$$

$n = 141,\ R^2 = .306,\ \bar{R}^2 = .286$

The robust standard errors do not differ by much from those in part (iii); in most cases, they are slightly higher, but all explanatory variables that were statistically significant before are still statistically significant. But the confidence interval for β_{PC} is a bit wider.

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 $\breve{y}_i$, $\breve{y}_i^2$, $i = 1,\ldots,807$ 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

$$\widehat{cigs} = 5.64 + 1.30 \log(income) - 2.94 \log(cigpric) - .463\, educ$$

$$\begin{array}{llll} (37.31) & (.54) & (8.97) & (.149) \end{array}$$

$$+ \underset{(.115)}{.482}\, age - \underset{(.0012)}{.0056}\, age^2 - \underset{(.72)}{3.46}\, restaurn$$

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

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.

[Instructor's Note: You can also do this exercise with regression (8.34) used in place of (8.32). This gives a somewhat larger estimated income effect.]

C8.10 (i) In the following equation, estimated by OLS, the usual standard errors are in (·) and the heteroskedasticity-robust standard errors are in [·]:

$$\widehat{e401k} = \underset{\substack{(.081)\\ [.079]}}{-.506} + \underset{\substack{(.0006)\\ [.0006]}}{.0124}\, inc - \underset{\substack{(.000005)\\ [.000005]}}{.000062}\, inc^2 + \underset{\substack{(.0039)\\ [.0038]}}{.0265}\, age - \underset{\substack{(.00005)\\ [.00004]}}{.00031}\, age^2 - \underset{\substack{(.0121)\\ [.0121]}}{.0035}\, male$$

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

There are no important differences; if anything, the robust standard errors are smaller.

(ii) This is a general claim. Since Var(y|**x**) = $p(\mathbf{x})[1 - p(\mathbf{x})]$, we can write $\mathrm{E}(u^2 \mid \mathbf{x}) = p(\mathbf{x}) - [p(\mathbf{x})]^2$. Written in error form, $u^2 = p(\mathbf{x}) - [p(\mathbf{x})]^2 + v$. In other words, we can write this as a regression model $u^2 = \delta_0 + \delta_1 p(\mathbf{x}) + \delta_2 [p(\mathbf{x})]^2 + v$, with the restrictions $\delta_0 = 0$, $\delta_1 = 1$, and $\delta_2 = -1$. Remember that, for the LPM, the fitted values, $\hat{y}_i$, are estimates of $p(\mathbf{x}_i) = \beta_0 + \beta_1 x_{i1} + ... + \beta_k x_{ik}$. So, when we run the regression $\hat{u}_i^2$ on $\hat{y}_i, \hat{y}_i^2$ (including an intercept), the intercept estimates should be close to zero, the coefficient on $\hat{y}_i$ should be close to one, and the coefficient on $\hat{y}_i^2$ should be close to –1.

(iii) The White *F* statistic is about 310.32, which is very significant. The coefficient on $\widehat{e401k}$ is about 1.010, the coefficient on $(\widehat{e401k})^2$ is about –.970, and the intercept is about -.009. These estimates are quite close to what we expect to find from the theory in part (ii).

(iv) The smallest fitted value is about .030 and the largest is about .697. The WLS estimates of the LPM are

$$\widehat{e401k} = -.488 + .0126\, inc - .000062\, inc^2 + .0255\, age - .00030\, age^2 - .0055\, male$$
$$\quad\quad (.076) \quad (.0005) \quad (.000004) \quad (.0037) \quad (.00004) \quad (.0117)$$

$n = 9{,}275$, $R^2 = .108$.

There are no important differences with the OLS estimates. The largest relative change is in the coefficient on *male*, but this variable is very insignificant using either estimation method.

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.

C8.12 (i) The estimated equation is

$$\begin{array}{llllllll} \widehat{math4} = & 91.93 & - & .449\, lunch & - & 5.40\, lenroll & + & 3.52\, lexppp \\ & (19.96) & & (.015) & & (0.94) & & (2.10) \\ & [23.09] & & [.017] & & [1.13] & & [2.35] \end{array}$$

n = 1,692, R^2 = .373

The heteroskedasticity-robust standard errors are somewhat larger, in all cases, than the usual OLS standard errors. The robust t statistic on *lexppp* is about 1.50, which raises further doubt about whether performance is linked to spending.

(ii) The value of the F statistic is 132.7, which gives a p-value of zero to at least four decimal places. Therefore, there is strong evidence of heteroskedasticity.

(iii) The equation estimated by WLS is

$$\begin{array}{llllllll} \widehat{math4} = & 50.48 & - & .449\, lunch & - & 2.65\, lenroll & + & 6.47\, lexppp \\ & (16.51) & & (.015) & & (0.84) & & (1.69) \end{array}$$

n = 1,692, R^2 = .360

where the usual WLS standard errors are in (·). The OLS and WLS coefficients on *lunch* are the same to three decimal places, but the other coefficients differ in practically important ways. The most important is that the WLS coefficient on *lexppp* is much larger than the OLS coefficient. Now, a 10 percent increase in spending (so *lexppp* increases by .1) is associated with roughly a .65 percentage point increase in the math pass rate. The WLS t statistic is much larger, too: $t \approx$ 3.83.

(iv) Because our model of heteroskedasticity might be wrong, it is a good idea to compute the robust standard errors for WLS. On the key variable *lexppp*, the robust standard error is about 1.82, which is somewhat higher than the usual WLS standard error. The robust standard error on *lenroll* is also somewhat higher, 1.05. That on *lunch* is slightly lower: .014 to three decimal places. For *lexppp*, the robust t statistic is about 3.55, which is still very statistically significant.

(v) WLS is more precise: its robust standard error is 1.82, compare with the robust standard error for OLS of 2.35. Of course, the t for WLS is much larger partly because the coefficient estimate is much larger. The lower standard error has an effect, too.

CHAPTER 9

TEACHING NOTES

The coverage of RESET in this chapter recognizes that it is a test for neglected nonlinearities, and it should not be expected to be more than that. (Formally, it can be shown that if an omitted variable has a conditional mean that is linear in the included explanatory variables, RESET has no ability to detect the omitted variable. Interested readers may consult my chapter in *Companion to Theoretical Econometrics*, 2001, edited by Badi Baltagi.) I just teach students the F statistic version of the test.

The Davidson-MacKinnon test can be useful for detecting functional form misspecification, especially when one has in mind a specific alternative, nonnested model. It has the advantage of always being a one degree of freedom test.

I think the proxy variable material is important, but the main points can be made with Examples 9.3 and 9.4. The first shows that controlling for IQ can substantially change the estimated return to education, and the omitted ability bias is in the expected direction. Interestingly, education and ability do not appear to have an interactive effect. Example 9.4 is a nice example of how controlling for a previous value of the dependent variable – something that is often possible with survey and nonsurvey data – can greatly affect a policy conclusion. Computer Exercise 9.3 is also a good illustration of this approach.

The short section on random coefficient models is intended as a simple introduction to the idea; it is not needed for any other parts of the book.

I rarely get to teach the measurement error material in a first-semester course, although the attenuation bias result for classical errors-in-variables is worth mentioning. You might have analytically skilled students try Problem 9.7.

The result on exogenous sample selection is easy to discuss, with more details given in Chapter 17. The effects of outliers can be illustrated using the examples. I think the infant mortality example, Example 9.10, is useful for illustrating how a single influential observation can have a large effect on the OLS estimates. Studentized residuals are computed by many regression packages, and they can be informative if used properly.

With the growing importance of least absolute deviations, it makes sense to discuss the merits of LAD, at least in more advanced courses. Computer Exercise C9.9 is a good example to show how mean and median effects can be very different, even though there may not be “outliers” in the usual sense.

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.)

9.2 [Instructor's Note: Out of the 186 records in VOTE2.RAW, three have *voteA88* less than 50, which means the incumbent running in 1990 cannot be the candidate who received *voteA88* percent of the vote in 1988. You might want to reestimate the equation dropping these three observations.]

(i) The coefficient on *voteA88* implies that if candidate A had one more percentage point of the vote in 1988, she/he is predicted to have only .067 more percentage points in 1990. Or, 10 more percentage points in 1988 implies .67 points, or less than one point, in 1990. The *t* statistic is only about 1.26, and so the variable is insignificant at the 10% level against the positive one-sided alternative. (The critical value is 1.282.) While this small effect initially seems surprising, it is much less so when we remember that candidate A in 1990 is always the incumbent. Therefore, what we are finding is that, conditional on being the incumbent, the percent of the vote received in 1988 does not have a strong effect on the percent of the vote in 1990.

(ii) Naturally, the coefficients change, but not in important ways, especially once statistical significance is taken into account. For example, while the coefficient on log(*expendA*) goes from $-.929$ to $-.839$, the coefficient is not statistically or practically significant anyway (and its sign is not what we expect). The magnitudes of the coefficients in both equations are quite similar, and there are certainly no sign changes. This is not surprising given the insignificance of *voteA88*.

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 $\widehat{\Delta math10} \approx -(1.26/100)(\%\Delta enroll) = -.0126(\%\Delta enroll)$. 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.

9.4 (i) For the CEV assumptions to hold, we must be able to write $tvhours = tvhours^* + e_0$, where the measurement error e_0 has zero mean and is uncorrelated with $tvhours^*$ and each explanatory variable in the equation. (Note that for OLS to consistently estimate the parameters we do not need e_0 to be uncorrelated with $tvhours^*$.)

(ii) The CEV assumptions are unlikely to hold in this example. For children who do not watch TV at all, $tvhours^* = 0$, and it is very likely that reported TV hours is zero. So if $tvhours^* = 0$ then $e_0 = 0$ with high probability. If $tvhours^* > 0$, the measurement error can be positive or negative, but, since $tvhours \geq 0$, e_0 must satisfy $e_0 \geq -tvhours^*$. So e_0 and $tvhours^*$ are likely to be correlated. As mentioned in part (i), because it is the dependent variable that is measured with error, what is important is that e_0 is uncorrelated with the explanatory variables. But this is unlikely to be the case, because $tvhours^*$ depends directly on the explanatory variables. Or, we might argue directly that more highly educated parents tend to underreport how much television their children watch, which means e_0 and the education variables are negatively correlated.

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.)

9.6 Following the hint, we write

$$y_i = \alpha + \beta x_i + u_i$$

where $u_i = c_i + d_i x_i$ and $c_i = a_i - \alpha$ and $d_i = b_i - \beta$. Now, $\mathrm{E}(c_i) = 0$ and $\mathrm{E}(d_i x_i) = \mathrm{Cov}(b_i, x_i)$ because $d_i = b_i - \mathrm{E}(b_i)$. By assumption, b_i and x_i are uncorrelated, and so $\mathrm{E}(d_i x_i)$. Therefore, we have shown that u_i has a zero mean. Next, we need to show that u_i is uncorrelated with x_i. This follows if c_i and $d_i x_i$ is each uncorrelated with x_i. But we are to assume c_i (equivalently, a_i)

is uncorrelated with x_i. Further, $\text{Cov}(d_i x_i, x_i) = \text{E}(d_i x_i^2) = \text{Cov}(d_i, x_i^2) = \text{Cov}(b_i, x_i^2) = 0$, because we are to assume that b_i is uncorrelated with x_i^2. We have shown that u_i has zero mean and is uncorrelated with x_i, and so OLS is consistent for α and β. OLS is not necessarily unbiased in this case. Under (9.19), OLS is unbiased.

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 $\approx$ .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.

C9.2 [Instructor's Note: If $educ \cdot KWW$ is used along with *KWW*, the interaction term is significant. This is in contrast to when IQ is used as the proxy. You may want to pursue this as an additional part to the exercise.]

(i) We estimate the model from column (2) but with *KWW* in place of *IQ*. The coefficient on *educ* becomes about .058 (se $\approx$ .006), so this is similar to the estimate obtained with *IQ*, although slightly larger and more precisely estimated.

(ii) When *KWW* and *IQ* are both used as proxies, the coefficient on *educ* becomes about .049 (se $\approx$ .007). Compared with the estimate when only *KWW* is used as a proxy, the return to education has fallen by almost a full percentage point.

(iii) The *t* statistic on *IQ* is about 3.08 while that on *KWW* is about 2.07, so each is significant at the 5% level against a two-sided alternative. They are jointly very significant, with $F_{2,925} \approx$ 8.59 and *p*-value $\approx$ .0002.

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.

C9.4 (i) Adding DC to the regression in equation (9.37) gives

$$\begin{array}{llllllll} \widehat{infmort} = & 23.95 & - & .567\log(pcinc) & - & 2.74\log(physic) & + & .629\log(popul) & + & 16.03\,DC \\ & (12.42) & & (1.641) & & (1.19) & & (.191) & & (1.77) \end{array}$$

$$n = 51,\quad R^2 = .691,\quad \bar{R}^2 = .664.$$

The coefficient on DC means that even if there was a state that had the same per capita income, per capita physicians, and population as Washington D.C., we predict that D.C. has an infant mortality rate that is about 16 deaths per 1000 live births higher. This is a very large "D.C. effect."

(ii) In the regression from part (i), the intercept and all slope coefficients, along with their standard errors, are identical to those in equation (9.38), which simply excludes D.C. [Of course, equation (9.38) does not have DC in it, so we have nothing to compare with its coefficient and standard error.] Therefore, for the purposes of obtaining the effects and statistical significance of the other explanatory variables, including a dummy variable for a single observation is identical to just dropping that observation when doing the estimation.

The R-squareds and adjusted R-squareds from (9.38) and the regression in part (i) are not the same. They are much larger when DC is included as an explanatory variable because we are predicting the infant mortality rate *perfectly* for D.C. You might want to confirm that the residual for the observation corresponding to D.C. is identically zero.

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

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

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

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.

C9.6 (i) Only four of the 408 schools have *b/s* less than .01.

(ii) We estimate the model in column (3) of Table 4.3, omitting schools with *b/s* < .01:

$$\widehat{\log(salary)} = \underset{(0.26)}{10.71} - \underset{(.196)}{.421}\,(b/s) + \underset{(.007)}{.089}\log(enroll) - \underset{(.050)}{.219}\log(staff)$$
$$- \underset{(.00161)}{.00023}\, droprate + \underset{(.00066)}{.00090}\, gradrate$$

$$n = 404, \quad R^2 = .354.$$

Interestingly, the estimated tradeoff is reduced by a nontrivial amount (from .589 to .421). This is a pretty large difference considering only four of 408 observations, or less than 1%, were omitted.

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.

C9.8 (i) The mean of *stotal* is .047, its standard deviation is .854, the minimum value is –3.32, and the maximum value is 2.24.

(ii) In the regression *jc* on *stotal*, the slope coefficient is .011 (se = .011). Therefore, while the estimated relationship is positive, the t statistic is only one: the correlation between *jc* and *stotal* is weak at best. In the regression *univ* on *stotal*, the slope coefficient is 1.170 (se = .029), for a t statistic of 38.5. Therefore, *univ* and *stotal* are positively correlated (with correlation = .435).

(iii) When we add *stotal* to (4.17) and estimate the resulting equation by OLS, we get

$$\widehat{\log(wage)} = \underset{(.021)}{1.495} + \underset{(.0068)}{.0631}\, jc + \underset{(.0026)}{.0686}\, univ + \underset{(.00016)}{.00488}\, exper + \underset{(.0068)}{.0494}\, stotal$$

$n = 6{,}758,\ R^2 = .228$

For testing $\beta_{jc} = \beta_{univ}$, we can use the same trick as in Section 4.4 to get the standard error of the difference: replace *univ* with *totcoll* = *jc* + *univ*, and then the coefficient on *jc* is the difference in the estimated returns, along with its standard error. Let $\theta_1 = \beta_{jc} - \beta_{univ}$. Then $\hat{\theta}_1 = -.0055\ (\text{se} = .0069)$. Compared with what we found without *stotal*, the evidence is even weaker against H_1: $\beta_{jc} < \beta_{univ}$. The t statistic from equation (4.27) is about –1.48, while here we have obtained only –.80.

(iv) When $stotal^2$ is added to the equation, its coefficient is .0019 (t statistic = .40). Therefore, there is no reason to add the quadratic term.

(v) The F statistic for testing exclusion of the interaction terms $stotal \cdot jc$ and $stotal \cdot univ$ is about 1.96; with 2 and 6,756 *df*, this gives p-value = .141. So, even at the 10% level, the interaction terms are jointly insignificant. It is probably not worth complicating the basic model estimated in part (iii).

(vi) I would just use the model from part (iii), where *stotal* appears only in level form. The other embellishments were not statistically significant at small enough significance levels to warrant the additional complications.

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, 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 $R^2_{\hat{u}^2} = .0374$, 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 $\text{E}(u|x_1,x_2,\ldots,x_k) = 0$ 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.

C9.10 (i) About .416 of the mean receive training in JTRAIN2, whereas only .069 receive training in JTRAIN3. The men in JTRAIN2, who were low earners, were targeted to receive training in a special job training experiment. This is not a representative group from the entire population. The sample from JTRAIN3 is, for practical purposes, a random sample from the population of men working in 1978; we would expect a much smaller fraction to have participated in job training in the prior year.

(ii) The simple regression gives

$$\widehat{re78} = \underset{(0.41)}{4.55} + \underset{(0.63)}{1.79}\, train$$

$n = 445, R^2 = .018$

Because *re78* is measured in thousands, job training participation is estimated to increase real earnings in 1978 by $1,790 – a nontrivial amount.

(iii) Adding all of the control listed changes the coefficient on *train* to 1.68 (se = .63). This is not much of a change from part (ii), and we would not expect it to be. Because *train* was supposed to be assigned randomly, it should be roughly uncorrelated with all other explanatory variables. Therefore, the simple and multiple regression estimates are similar. (Interestingly, the standard errors are the same to two decimal places.)

(iv) The simple regression coefficient on *train* is −15.20 (se = 1.15). This implies a huge negative effect of job training, which is hard to believe. Because training was not randomly assigned for this group, we can assume self-selection into job training is at work. That is, it is the low earning group that tends to select itself (perhaps with the help of administrators) into job training. When we add the controls, the coefficient becomes .213 (se = .853). In other words, when we account for factors such as previous earnings and education, we obtain a small but insignificant positive effect of training. This is certainly more believable than the large negative effect obtained from simple regression.

(v) For JTRAIN2, the average is 1.74, the standard deviation is 3.90, the minimum is 0, and the maximum is 24.38. For JTRAIN3, the average is 18.04, the standard deviation is 13.29, the minimum is 0, and the maximum is 146.90. Clearly these samples are not representative of the same population. JTRAIN3, which represents a much broader population, has a much larger mean value and much larger standard deviation.

(vi) For JTRAIN2, which uses 427 observations, the estimate on *train* is similar to before, 1.58 (se = .63). For JTRAIN3, which uses 765 observations, the estimate is now much closer to the experimental estimate: 1.84 (se = .89).

(vii) The estimate for JTRAIN2, which uses 280 observations, is 1.84 (se = .89); it is a coincidence that this is the same, to two digits, as that obtained for JTRAIN3 in part (vi). For JTRAIN3, which uses 271 observations, the estimate is 3.80 (se = .88).

(viii) When we base our analysis on comparable samples – roughly representative of the same population, those with average real earnings less than $10,000 is 1974 and 1975 – we get positive, nontrivial training effects estimates using either sample. Using the full data set in JTRAIN3 can be misleading because it includes many men for whom training would never be beneficial. In effect, when we use the entire data set, we average in the zero effect for high earners with the positive effect for low-earning men. Of course, if we only have experimental data, it can be difficult to know how to find the part of the population where there is an effect. But for those who were unemployed in the two years prior to job training, the effect appears to be unambiguously positive.

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.

C9.12 (i) The coefficient on *bs* is about −.52, with usual standard error .110. Using this standard error, the *t* statistic is about –4.7. The heteroskedasticity-robust standard error is much larger, .283, and results in a marginally significant *t* statistic. Nevertheless, from these findings we might just conclude that there is substantial heteroskedasticity that affects the standard errors. The estimate of –.52 is not small (although it is not especially close to the hypothesized value −1).

(ii) Dropping the four observations with *bs* > .5 gives $\hat{\beta}_{bs}$ = –.186 with (robust) t = –1.27. The practical significance of the coefficient is much lower, and it is no longer statistically different from zero.

(iii) One can verify that "dummying out" the four observations does give the same estimates as in part (ii). The *t* statistic on the dummy *d*1508 is –6.16. None of the other dummies is significant at even the 15% level against a two-sided alternative.

(iv) The coefficient on *bs* with only observation 1,508 dropped is about –.20, which is a big difference from –.52. When only observation 68 is dropped, the coefficient is –.50, when only 1,127 is dropped, it is –.54, and when only 1,670 is dropped, the coefficient is about –.53. Thus, the estimate is only sensitive to the inclusion of observation 1,508.

(v) The findings in part (iv) show that, if an observation is extreme enough, it can exert much influence on the OLS estimate. One might hope that having over 1,800 observations would make OLS resilient, but that is not necessarily the case. The influential observation happens to be the observation with the lowest average salary and the highest *bs* ratio, and so including it clearly shifts the estimate toward –1.

(vi) The LAD estimate of β_{bs} using the full sample is about –.109 (t = –1.01). When observation 1,508 is dropped, the estimate is –.097 (t = –0.81). The change in the estimate is modest, and, besides, it is statistically insignificant in both cases.

CHAPTER 10

TEACHING NOTES

Because of its realism and its care in stating assumptions, this chapter puts a somewhat heavier burden on the instructor and student than traditional treatments of time series regression. Nevertheless, I think it is worth it. It is important that students learn that there are potential pitfalls inherent in using regression with time series data that are not present for cross-sectional applications. Trends, seasonality, and high persistence are ubiquitous in time series data. By this time, students should have a firm grasp of multiple regression mechanics and inference, and so you can focus on those features that make time series applications different from cross-sectional ones.

I think it is useful to discuss static and finite distributed lag models at the same time, as these at least have a shot at satisfying the Gauss-Markov assumptions. Many interesting examples have distributed lag dynamics. In discussing the time series versions of the CLM assumptions, I rely mostly on intuition. The notion of strict exogeneity is easy to discuss in terms of feedback. It is also pretty apparent that, in many applications, there are likely to be some explanatory variables that are not strictly exogenous. What the student should know is that, to conclude that OLS is unbiased – as opposed to consistent – we need to assume a very strong form of exogeneity of the regressors. Chapter 11 shows that only contemporaneous exogeneity is needed for consistency.

Although the text is careful in stating the assumptions, in class, after discussing strict exogeneity, I leave the conditioning on **X** implicit, especially when I discuss the no serial correlation assumption. As this is a new assumption I spend some time on it. (I also discuss why we did not need it for random sampling.)

Once the unbiasedness of OLS, the Gauss-Markov theorem, and the sampling distributions under the classical linear model assumptions have been covered – which can be done rather quickly – I focus on applications. Fortunately, the students already know about logarithms and dummy variables. I treat index numbers in this chapter because they arise in many time series examples.

A novel feature of the text is the discussion of how to compute goodness-of-fit measures with a trending or seasonal dependent variable. While detrending or deseasonalizing y is hardly perfect (and does not work with integrated processes), it is better than simply reporting the very high R-squareds that often come with time series regressions with trending variables.

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.

10.2 We follow the hint and write

$$gGDP_{t-1} = \alpha_0 + \delta_0 int_{t-1} + \delta_1 int_{t-2} + u_{t-1},$$

and plug this into the right-hand-side of the int_t equation:

$$\begin{aligned} int_t &= \gamma_0 + \gamma_1(\alpha_0 + \delta_0 int_{t-1} + \delta_1 int_{t-2} + u_{t-1} - 3) + v_t \\ &= (\gamma_0 + \gamma_1\alpha_0 - 3\gamma_1) + \gamma_1\delta_0 int_{t-1} + \gamma_1\delta_1 int_{t-2} + \gamma_1 u_{t-1} + v_t. \end{aligned}$$

Now by assumption, u_{t-1} has zero mean and is uncorrelated with all right-hand-side variables in the previous equation, except itself of course. So

$$\text{Cov}(int,u_{t-1}) = \text{E}(int_t \cdot u_{t-1}) = \gamma_1 \text{E}(u_{t-1}^2) > 0$$

because $\gamma_1 > 0$. If $\sigma_u^2 = \text{E}(u_t^2)$ for all t then $\text{Cov}(int,u_{t-1}) = \gamma_1\sigma_u^2$. This violates the strict exogeneity assumption, TS.2. While u_t is uncorrelated with int_t, int_{t-1}, and so on, u_t is correlated with int_{t+1}.

10.3 Write

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

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

10.4 We use the R-squared form of the F statistic (and ignore the information on $\bar{R}^2$). The 10% critical value with 3 and 124 degrees of freedom is about 2.13 (using 120 denominator *df* in Table G.3a). The F statistic is

$$F = [(.305 - .281)/(1 - .305)](124/3) \approx 1.43,$$

which is well below the 10% *cv*. Therefore, the event indicators are jointly insignificant at the 10% level. This is another example of how the (marginal) significance of one variable (*afdec6*) can be masked by testing it jointly with two very insignificant variables.

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.

10.6 (i) Given $\delta_j = \gamma_0 + \gamma_1 j + \gamma_2 j^2$ for $j = 0,1, \dots ,4$, we can write

$$\begin{aligned} y_t &= \alpha_0 + \gamma_0 z_t + (\gamma_0 + \gamma_1 + \gamma_2)z_{t-1} + (\gamma_0 + 2\gamma_1 + 4\gamma_2)z_{t-2} + (\gamma_0 + 3\gamma_1 + 9\gamma_2)z_{t-3} \\ &\quad + (\gamma_0 + 4\gamma_1 + 16\gamma_2)z_{t-4} + u_t \\ &= \alpha_0 + \gamma_0(z_t + z_{t-1} + z_{t-2} + z_{t-3} + z_{t-4}) + \gamma_1(z_{t-1} + 2z_{t-2} + 3z_{t-3} + 4z_{t-4}) \\ &\quad + \gamma_2(z_{t-1} + 4z_{t-2} + 9z_{t-3} + 16z_{t-4}) + u_t. \end{aligned}$$

(ii) This is suggested in part (i). For clarity, define three new variables: $z_{t0} = (z_t + z_{t-1} + z_{t-2} + z_{t-3} + z_{t-4})$, $z_{t1} = (z_{t-1} + 2z_{t-2} + 3z_{t-3} + 4z_{t-4})$, and $z_{t2} = (z_{t-1} + 4z_{t-2} + 9z_{t-3} + 16z_{t-4})$. Then, α_0, γ_0, γ_1, and γ_2 are obtained from the OLS regression of y_t on z_{t0}, z_{t1}, and z_{t2}, $t = 1, 2, \dots, n$. (Following our convention, we let $t = 1$ denote the first time period where we have a full set of regressors.) The $\hat{\delta}_j$ can be obtained from $\hat{\delta}_j = \hat{\gamma}_0 + \hat{\gamma}_1 j + \hat{\gamma}_2 j^2$.

(iii) The unrestricted model is the original equation, which has six parameters (α_0 and the five δ_j). The PDL model has four parameters. Therefore, there are two restrictions imposed in moving from the general model to the PDL model. (Note how we do not have to actually write out what the restrictions are.) The *df* in the unrestricted model is $n - 6$. Therefore, we would obtain the unrestricted R-squared, R^2_{ur} from the regression of y_t on z_t, z_{t-1}, $\dots$, z_{t-4} and the restricted R-squared from the regression in part (ii), R^2_r. The F statistic is

$$F = \frac{(R_{ur}^2 - R_r^2)}{(1 - R_{ur}^2)} \cdot \frac{(n-6)}{2}.$$

Under H_0 and the CLM assumptions, $F \sim F_{2,n\text{-}6}$.

10.7 (i) $pe_{t\text{-}1}$ and $pe_{t\text{-}2}$ 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 $pe_{t\text{-}2}$, $pe_{t\text{-}1}$, and pe_t by the same amount.

10.8 It is easiest to discuss this question in the context of correlations, rather than conditional means. The solution here does both.

(i) Strict exogeneity implies that the error at time t, u_t, is uncorrelated with the regressors in every time period: current, past, and future. Sequential exogeneity states that u_t is uncorrelated with current and past regressors, so it is implied by strict exogeneity. In terms of conditional means, strict exogeneity is $\mathrm{E}(u_t \mid ..., \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1}, ...) = 0$, and so u_t conditional on any subset of $(..., \mathbf{x}_{t-1}, \mathbf{x}_t, \mathbf{x}_{t+1}, ...)$, including $(\mathbf{x}_t, \mathbf{x}_{t-1}, ...)$, also has a zero conditional mean. But the latter condition is the definition of sequential exogeneity.

(ii) Sequential exogeneity implies that u_t is uncorrelated with $\mathbf{x}_t$, $\mathbf{x}_{t\text{-}1}$, …, which, of course, implies that u_t is uncorrelated with $\mathbf{x}_t$ (which is contemporaneous exogeneity stated in terms of zero correlation). In terms of conditional means, $\mathrm{E}(u_t \mid \mathbf{x}_t, \mathbf{x}_{t-1}, ...) = 0$ implies that u_t has zero mean conditional on any subset of variables in $(\mathbf{x}_t, \mathbf{x}_{t-1}, ...)$. In particular, $\mathrm{E}(u_t \mid \mathbf{x}_t) = 0$.

(iii) No, OLS is not generally unbiased under sequential exogeneity. To show unbiasedness, we need to condition on the entire matrix of explanatory variables, **X**, and use $\mathrm{E}(u_t \mid \mathbf{X}) = 0$ for all t. But this condition is strict exogeneity, and it is not implied by sequential exogeneity.

(iv) The model and assumption imply

$$\mathrm{E}(u_t \mid pccon_t, pccon_{t-1}, ...) = 0,$$

which means that $pccon_t$ is sequentially exogenous. (One can debate whether three lags is enough to capture the distributed lag dynamics, but the problem asks you to assume this.) But $pccon_t$ may very well fail to be strictly exogenous because of feedback effects. For example, a shock to the HIV rate this year – manifested as $u_t > 0$ – could lead to increased condom usage in the future. Such a scenario would result in positive correlation between u_t and $pccon_{t+h}$ for $h > 0$. OLS would still be consistent, but not unbiased.

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

$$\widehat{i3}_t = \underset{(0.43)}{1.30} + \underset{(.076)}{.608}\, inf_t + \underset{(.120)}{.363}\, def_t + \underset{(0.51)}{1.56}\, post79_t$$

$$n = 56, \quad R^2 = .664, \quad \bar{R}^2 = .644.$$

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.

C10.2 (i) Adding a linear time trend to (10.22) gives

$$\widehat{\log(chnimp)} = \underset{(20.78)}{-2.37} - \underset{(1.240)}{.686} \log(chempi) + \underset{(.876)}{.466} \log(gas) + \underset{(.472)}{.078} \log(rtwex)$$

$$+ \underset{(.251)}{.090}\, befile6 + \underset{(.257)}{.097}\, affile6 - \underset{(.282)}{.351}\, afdec6 + \underset{(.004)}{.013}\, t$$

$$n = 131, \quad R^2 = .362, \quad \bar{R}^2 = .325.$$

Only the trend is statistically significant. In fact, in addition to the time trend, which has a t statistic over three, only *afdec6* has a t statistic bigger than one in absolute value. Accounting for a linear trend has important effects on the estimates.

(ii) The F statistic for joint significance of all variables except the trend and intercept, of course) is about .54. The *df* in the F distribution are 6 and 123. The p-value is about .78, and so the explanatory variables other than the time trend are jointly very insignificant. We would have to conclude that once a positive linear trend is allowed for, nothing else helps to explain log(*chnimp*). This is a problem for the original event study analysis.

(iii) Nothing of importance changes. In fact, the p-value for the test of joint significance of all variables except the trend and monthly dummies is about .79. The 11 monthly dummies themselves are not jointly significant: p-value $\approx$ .59.

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

$$\widehat{\log(prepop_t)} = -6.66 - .212 \log(mincov_t) + .486 \log(usgnp_t) + .285 \log(prgnp_t)$$
$$\quad\quad (1.26) \quad (.040) \quad\quad\quad\quad (.222) \quad\quad\quad\quad (.080)$$
$$- .027\,t$$
$$(.005)$$

$$n = 38, \quad R^2 = .889, \quad \bar{R}^2 = .876.$$

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).

C10.4 If we run the regression of gfr_t on pe_t, ($pe_{t-1} - pe_t$), ($pe_{t-2} - pe_t$), $ww2_t$, and $pill_t$, the coefficient and standard error on pe_t are, rounded to four decimal places, .1007 and .0298, respectively. When rounded to three decimal places we obtain .101 and .030, as reported in the text.

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.

C10.6 (i) The regression of gfr_t on a quadratic in time gives

$$\widehat{gfr}_t = 107.06 + .072\,t - .0080\,t^2$$
$$\quad (6.05) \quad (.382) \quad (.0051)$$

$$n = 72, \quad R^2 = .314.$$

Although t and t^2 are individually insignificant, they are jointly very significant (p-value $\approx$ .0000).

(ii) Using $\ddot{gfr}_t$ as the dependent variable in (10.35) gives $R^2 \approx .602$, compared with about .727 if we do not initially detrend. Thus, the equation still explains a fair amount of variation in *gfr* even after we net out the trend in computing the total variation in *gfr*.

(iii) The coefficient and *t* statistic on t^3 are about –.00129 and .00019, respectively, which results in a very significant *t* statistic. It is difficult to know what to make of this. The cubic trend, like the quadratic, is not monotonic. So this almost becomes a curve-fitting exercise.

C10.7 (i) The estimated equation is

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

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

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

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

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

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.

C10.8 (i) The estimated equation is

$$\widehat{gfr}_t = \underset{(3.33)}{92.05} + \underset{(.126)}{.089}\, pe_t - \underset{(.1531)}{.0040}\, pe_{t-1} + \underset{(.1651)}{.0074}\, pe_{t-2} + \underset{(.154)}{.018}\, pe_{t-3} + \underset{(.105)}{.014}\, pe_{t-4}$$

$$- \underset{(11.54)}{21.34}\, ww2_t - \underset{(3.90)}{31.08}\, pill_t$$

$$n = 68,\quad R^2 = .537,\quad \bar{R}^2 = .483.$$

The p-value for the F statistic of joint significance of pe_{t-3} and pe_{t-4} is about .94, which is very weak evidence against H_0.

(ii) The LRP and its standard error can be obtained as the coefficient and standard error on pe_t in the regression

$$gfr_t \text{ on } pe_t,\ (pe_{t-1} - pe_t),\ (pe_{t-2} - pe_t),\ (pe_{t-3} - pe_t),\ (pe_{t-4} - pe_t),\ ww2_t,\ pill_t$$

We get $\widehat{LRP} \approx .129$ (se $\approx .030$), which is above the estimated LRP with only two lags (.101). The standard errors are the same rounded to three decimal places.

(iii) We estimate the PDL with the additional variables $ww2_2$ and $pill_t$. To estimate γ_0, γ_1, and γ_2, we define the variables

$$z0_t = pe_t + pe_{t-1} + pe_{t-2} + pe_{t-3} + pe_{t-4}$$

$$z1_t = pe_{t-1} + 2pe_{t-2} + 3pe_{t-3} + 4pe_{t-4}$$

$$z2_t = pe_{t-1} + 4pe_{t-2} + 9pe_{t-3} + 16pe_{t-4}.$$

Then, run the regression $gfrt_t$ on $z0_t$, $z1_t$, $z2_t$, $ww2_t$, $pill_t$. Using the data in FERTIL3.RAW gives (to three decimal places) $\hat{\gamma}_0 = .069$, $\hat{\gamma}_1 = -.057$, $\hat{\gamma}_2 = .012$. So $\hat{\delta}_0 = \hat{\gamma}_0 = .069$, $\hat{\delta}_1 = .069 - .057 + .012 = .024$, $\hat{\delta}_2 = .069 - 2(.057) + 4(.012) = .003$, $\hat{\delta}_3 = .069 - 3(.057) + 9(.012) = .006$, $\hat{\delta}_4 = .069 - 4(.057) + 16(.012) = .033$. Therefore, the LRP is .135. This is slightly above the .129 obtained from the unrestricted model, but not much.

Incidentally, the F statistic for testing the restrictions imposed by the PDL is about $[(.537 - .536)/(1 - .537)](60/2) \approx .065$, which is very insignificant. Therefore, the restrictions are not rejected by the data. Anyway, the only parameter we can estimate with any precision, the LRP, is not very different in the two models.

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

$$\widehat{rsp500}_t = \underset{(3.27)}{18.84} + \underset{(.129)}{.036}\, pcip_t - \underset{(0.54)}{1.36}\, i3_t$$

$$n = 557, \quad R^2 = .012.$$

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*.

C10.10 (i) The sample correlation between *inf* and *def* is only about .098, which is pretty small. Perhaps surprisingly, inflation and the deficit rate are practically uncorrelated over this period. Of course, this is a good thing for estimating the effects of each variable on *i*3, as it implies almost no multicollinearity.

(ii) The equation with the lags is

$$\widehat{i3}_t = \underset{(0.40)}{1.61} + \underset{(.125)}{.343}\, inf_t + \underset{(.134)}{.382}\, inf_{t-1} - \underset{(.221)}{.190}\, def_t + \underset{(.197)}{.569}\, def_{t-1}$$

$$n = 55, \quad R^2 = .685, \quad \bar{R}^2 = .660.$$

(iii) The estimated LRP of *i*3 with respect to *inf* is .343 + .382 = .725, which is somewhat larger than .606, which we obtain from the static model in (10.15). But the estimates are fairly close considering the size and significance of the coefficient on inf_{t-1}.

(iv) The F statistic for significance of inf_{t-1} and def_{t-1} is about 5.22, with p-value $\approx$ .009. So they are jointly significant at the 1% level. It seems that both lags belong in the model.

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

$$\widehat{\log(totacc)} = \underset{(.019)}{10.469} + \underset{(.00016)}{.00275}\, t - \underset{(.0244)}{.0427}\, feb + \underset{(.0244)}{.0798}\, mar + \underset{(.0245)}{.0185}\, apr$$

$$+ \underset{(.0245)}{.0321}\, may + \underset{(.0245)}{.0202}\, jun + \underset{(.0245)}{.0376}\, jul + \underset{(.0245)}{.0540}\, aug$$

$$+ \underset{(.0245)}{.0424}\, sep + \underset{(.0245)}{.0821}\, oct + \underset{(.0245)}{.0713}\, nov + \underset{(.0245)}{.0962}\, dec$$

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

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.

C10.12 (i) OLS estimation using all of the data gives

$$\widehat{inf} = \underset{(1.55)}{1.05} + \underset{(.266)}{.502}\, unem$$

$$n = 56,\ R^2 = .062,\ \bar{R}^2 = .045,$$

so there are 56 years of data.

(ii) The estimates are similar to those in equation (10.14). Adding the extra years does not help in finding a tradeoff between inflation and unemployment. In fact, the slope estimate becomes even larger (and is still positive) in the full sample.

(iii) Using only data from 1997 to 2003 gives

$$\widehat{inf} = \underset{(1.65)}{4.16} - \underset{(.334)}{.378}\, unem$$

$$n = 7,\ R^2 = .204,\ \bar{R}^2 = .044.$$

The equation now shows a tradeoff between inflation and unemployment: a one percentage point increase in *unem* is estimated to reduce *inf* by about .38 percentage points. Not surprisingly, with such a small sample size, the estimate is not statistically different from zero: the two-sided p-value is .31. So, while it is tempting to think that the inflation-unemployment tradeoff reemerges in the last part of the sample, the estimates are not precise enough to draw that conclusion.

(iv) The regressions in parts (i) and (iii) are an example of this setup, with $n_1 = 49$ and $n_2 = 7$. The weighted average of the slopes from the two different periods is $(49/56)\cdot(.468) + (7/56)\cdot(-.378) \approx .362$. But the slope estimate on the entire sample is .502. Generally, there is no simple relationship between the slope estimate on the entire sample and the slope estimates on two sub-samples.

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

TEACHING NOTES

Much of the material in this chapter is usually postponed, or not covered at all, in an introductory course. However, as Chapter 10 indicates, the set of time series applications that satisfy all of the classical linear model assumptions might be very small. In my experience, spurious time series regressions are the hallmark of many student projects that use time series data. Therefore, students need to be alerted to the dangers of using highly persistent processes in time series regression equations. (Spurious regression problem and the notion of cointegration are covered in detail in Chapter 18.)

It is fairly easy to heuristically describe the difference between a weakly dependent process and an integrated process. Using the MA(1) and the stable AR(1) examples is usually sufficient.

When the data are weakly dependent and the explanatory variables are contemporaneously exogenous, OLS is consistent. This result has many applications, including the stable AR(1) regression model. When we add the appropriate homoskedasticity and no serial correlation assumptions, the usual test statistics are asymptotically valid.

The random walk process is a good example of a unit root (highly persistent) process. In a one-semester course, the issue comes down to whether or not to first difference the data before specifying the linear model. While unit root tests are covered in Chapter 18, just computing the first-order autocorrelation is often sufficient, perhaps after detrending. The examples in Section 11.3 illustrate how different first-difference results can be from estimating equations in levels.

Section 11.4 is novel in an introductory text, and simply points out that, if a model is dynamically complete in a well-defined sense, it should not have serial correlation. Therefore, we need not worry about serial correlation when, say, we test the efficient market hypothesis. Section 11.5 further investigates the homoskedasticity assumption, and, in a time series context, emphasizes that what is contained in the explanatory variables determines what kind of hetero-skedasticity is ruled out by the usual OLS inference. These two sections could be skipped without loss of continuity.

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})] = $\gamma_h/(\sqrt{\gamma_0}\cdot\sqrt{\gamma_0}) = \gamma_h/\gamma_0$.

11.2 (i) E(x_t) = E(e_t) – (1/2)E(e_{t-1}) + (1/2)E(e_{t-2}) = 0 for t = 1,2, … Also, because the e_t are independent, they are uncorrelated and so Var(x_t) = Var(e_t) + (1/4)Var(e_{t-1}) + (1/4)Var(e_{t-2}) = 1 + (1/4) + (1/4) = 3/2 because Var (e_t) = 1 for all t.

(ii) Because x_t has zero mean, Cov(x_t,x_{t+1}) = E(x_tx_{t+1}) = E[(e_t – (1/2)e_{t-1} + (1/2)e_{t-2})(e_{t+1} – (1/2)e_t + (1/2)e_{t-1})] = E(e_te_{t+1}) – (1/2)E(e_t^2) + (1/2)E(e_te_{t-1}) – (1/2)E($e_{t-1}e_{t+1}$) + (1/4(E($e_{t-1}e_t$) – (1/4)E(e_{t-1}^2) + (1/2)E($e_{t-2}e_{t+1}$) – (1/4)E($e_{t-2}e_t$) +(1/4)E($e_{t-2}e_{t-1}$) = – (1/2)E(e_t^2) – (1/4)E(e_{t-1}^2) = –(1/2) – (1/4) = –3/4; the third to last equality follows because the e_t are pairwise uncorrelated and E(e_t^2) = 1 for all t. Using Problem 11.1 and the variance calculation from part (i), Corr(x_tx_{t+1}) = – (3/4)/(3/2) = –1/2.

Computing Cov(x_t,x_{t+2}) is even easier because only one of the nine terms has expectation different from zero: (1/2)E(e_t^2) = ½. Therefore, Corr(x_t,x_{t+2}) = (1/2)/(3/2) = 1/3.

(iii) Corr(x_t,x_{t+h}) = 0 for h >2 because, for h > 2, x_{t+h} depends at most on e_{t+j} for j > 0, while x_t depends on e_{t+j}, $j \leq 0$.

(iv) Yes, because terms more than two periods apart are actually uncorrelated, and so it is obvious that Corr(x_t,x_{t+h}) $\rightarrow$ 0 as $h \rightarrow \infty$.

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_ty_{t+h}) = E[(z + e_t)(z + e_{t+h})] = E(z^2) + E(ze_{t+h}) + E(e_tz) + E(e_te_{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) = $\sigma_z^2/(\sigma_z^2 + \sigma_e^2) > 0$.

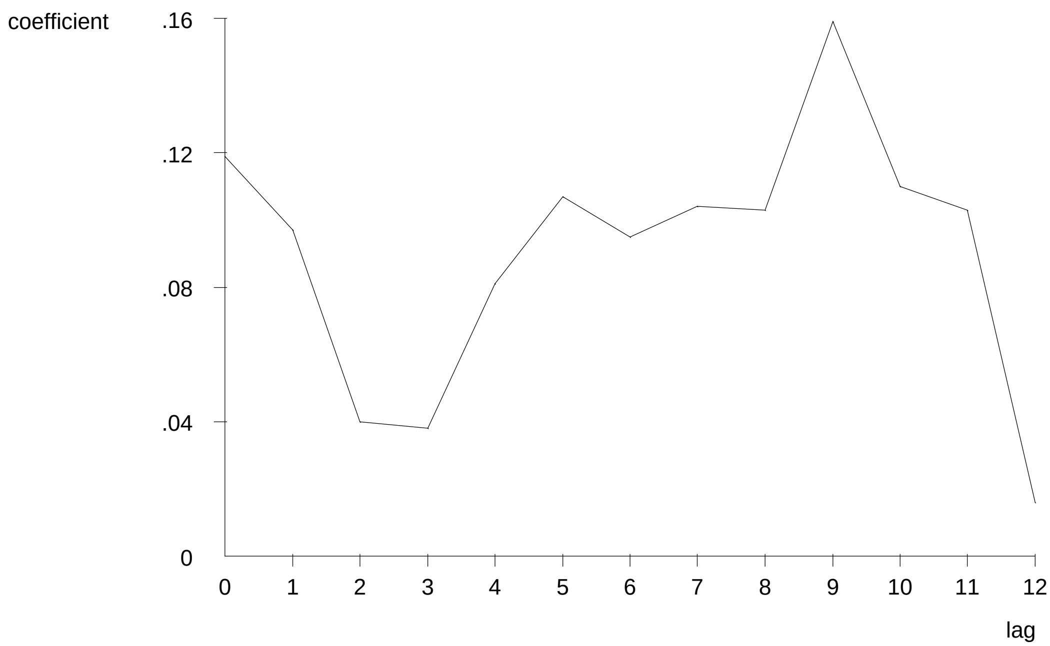
(iv) No. The correlation between y_t and y_{t+h} is the same positive value obtained in part (iii) now matter how large is h. In other words, no matter how far apart y_t and y_{t+h} are, their correlation is always the same. Of course, the persistent correlation across time is due to the presence of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

11.4 Assuming $y_0 = 0$ is a special case of assuming y_0 nonrandom, and so we can obtain the variances from (11.21): $\text{Var}(y_t) = \sigma_e^2 t$ and $\text{Var}(y_{t+h}) = \sigma_e^2(t+h)$, $h > 0$. Because $\text{E}(y_t) = 0$ for all t (since $\text{E}(y_0) = 0$), $\text{Cov}(y_t, y_{t+h}) = \text{E}(y_t y_{t+h})$ and, for $h > 0$,

$$\begin{aligned} \text{E}(y_t y_{t+h}) &= \text{E}[(e_t + e_{t-1} + \ldots\ e_1)(e_{t+h} + e_{t+h-1} + \ldots + e_1)] \\ &= \text{E}(e_t^2) + \text{E}(e_{t-1}^2) + \ldots + \text{E}(e_1^2) = \sigma_e^2 t, \end{aligned}$$

where we have used the fact that $\{e_t\}$ is a pairwise uncorrelated sequence. Therefore, $\text{Corr}(y_t, y_{t+h}) = \text{Cov}(y_t, y_{t+h}) / \sqrt{\text{Var}(y_t) \cdot \text{Var}(y_{t+h})} = t / \sqrt{t(t+h)} = \sqrt{t/(t+h}$.

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^2_{ur} be the R-squared from this regression. To obtain the restricted R-squared, R^2_r, we need to reestimate the model reported in the problem but with the same 267 observations used to estimate the unrestricted model. Then $F = [(R^2_{ur} - R^2_r)/(1 - R^2_{ur})](247/6)$. We would find the critical value from the $F_{6,247}$ distribution.

[Instructor's Note: As a computer exercise, you might have the students test whether all 13 lag coefficients in the population model are equal. The restricted regression is *gprice* on ($gwage + gwage_{-1} + gwage_{-2} + \ldots\ gwage_{-12}$), and the R-squared form of the F test, with 12 and 259 df, can be used.]

11.6 (i) The t statistic for H_0: $\beta_1 = 1$ is $t = (1.104 - 1)/.039 \approx 2.67$. Although we must rely on asymptotic results, we might as well use df = 120 in Table G.2. So the 1% critical value against a two-sided alternative is about 2.62, and so we reject H_0: $\beta_1 = 1$ against H_1: $\beta_1 \neq 1$ at the 1% level. It is hard to know whether the estimate is practically different from one without comparing investment strategies based on the theory ($\beta_1 = 1$) and the estimate ($\hat{\beta}_1 = 1.104$). But the estimate is 10% higher than the theoretical value.

(ii) The t statistic for the null in part (i) is now $(1.053 - 1)/.039 \approx 1.36$, so H_0: $\beta_1 = 1$ is no longer rejected against a two-sided alternative unless we are using more than a 10% significance level. But the lagged spread is very significant (contrary to what the expectations hypothesis predicts): $t = .480/.109 \approx 4.40$. Based on the estimated equation, when the lagged spread is positive, the predicted holding yield on six-month T-bills is above the yield on three-month T-bills (even if we impose $\beta_1 = 1$), and so we should invest in six-month T-bills.

(iii) This suggests unit root behavior for $\{hy3_t\}$, which generally invalidates the usual t-testing procedure.

(iv) We would include three quarterly dummy variables, say $Q2_t$, $Q3_t$, and $Q4_t$, and do an F test for joint significance of these variables. (The F distribution would have 3 and 117 *df*.)

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,

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

$$\equiv \beta_0 + \beta_1 y_{t-1} + \beta_2 x_t + u_t,$$

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 $E(e_t|x_t,y_{t-1},x_{t-1}, \ldots) = E(a_t|x_t,y_{t-1},x_{t-1}, \ldots) = 0$ it follows that $E(u_t|x_t,y_{t-1},x_{t-1}, \ldots) = 0$, which means that the model is dynamically complete [see equation (11.37)]. Therefore, the errors are serially uncorrelated. If the homoskedasticity assumption $\text{Var}(u_t|x_t,y_{t-1}) = \sigma^2$ 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 $\hat{\gamma}_1 = \hat{\beta}_2/\hat{\lambda} = .2/.3 \approx .67$.

11.8 (i) Sequential exogeneity does not rule out correlation between, say, u_{t-1} and x_{tj} for any regressors $j = 1, 2, \ldots, k$. The differencing generally induces correlation between the differenced errors and the differenced regressors. To see why, consider a single explanatory variable, x_t. Then $\Delta u_t = u_t - u_{t-1}$ and $\Delta x_t = x_t - x_{t-1}$. Under sequential exogeneity, u_t is uncorrelated with x_t and x_{t-1}, and u_{t-1} is uncorrelated with x_{t-1}. But u_{t-1} can be correlated with x_t, which means that Δu_t and Δx_t are generally correlated. In fact, under sequential exogeneity, it is always true that $\text{Cov}(\Delta x_t, \Delta u_t) = -\text{Cov}(x_t, u_{t-1})$.

(ii) Strict exogeneity of the regressors in the original equation is sufficient for OLS on the first-differenced equation to be consistent. Remember, strict exogeneity implies that the regressors in any time period are uncorrelated with the errors in any time period. Of course, we could make the weaker assumption: for any t, u_t is uncorrelated with $x_{t-1,j}$, x_{tj}, and $x_{t+1,j}$ for all $j = 1, 2, \ldots, k$. The strengthening beyond sequential exogeneity is the assumption that u_t is uncorrelated with all of next period's outcomes on all regressors. In practice, this is probably similar to just assuming strict exogeneity.

(iii) If we assume sequential exogeneity in a static model, the condition can be written as

$$\mathrm{E}(y_t \mid \mathbf{z}_t, \mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \ldots) = \mathrm{E}(y_t \mid \mathbf{z}_t)$$

which means that, once we control for the current (comtemporaneous) values of all explanatory variables, no lags matter. Although some relationships in economics are purely static, many admit distributed lag dynamics. Therefore, if one wants to capture lagged effects, it is a good idea to explore distributed lag models – whether or not we think there might be feedback from u_t to future values of the explanatory variables.

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

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

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

$$\begin{aligned}\Delta\widehat{\log(invpc_t)} &= \underset{(.048)}{.006} + \underset{(1.14)}{1.57}\,\Delta\log(price_t) + \underset{(.00190)}{.00004t} \\ n &= 41,\ \ R^2 = .048.\end{aligned}$$

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.

C11.2 (i) The estimated equation is

$$\widehat{ghrwage}_t = \underset{(.005)}{-.010} + \underset{(.167)}{.728}\, goutphr_t + \underset{(.166)}{.458}\, goutphr_{t-1}$$

$$n = 39,\ R^2 = .493,\ \bar{R}^2 = .465.$$

The t statistic on the lag is about 2.76, so the lag is very significant.

(ii) We follow the hint and write the LRP as $\theta = \beta_1 + \beta_2$, and then plug $\beta_1 = \theta - \beta_2$ into the original model:

$$ghrwage_t = \beta_0 + \theta\, goutphr_t + \beta_2(goutphr_{t-1} - goutphr_t) + u_t.$$

Therefore, we regress $ghrwage_t$ onto $goutphr_t$, and $(goutphr_{t-1} - goutphr_t)$ and obtain the standard error for $\hat{\theta}$. Doing this regression gives 1.186 [as we can compute directly from part (i)] and se($\hat{\theta}$) = .203. The t statistic for testing H_0: $\theta = 1$ is $(1.186 - 1)/.203 \approx .916$, which is not significant at the usual significance levels (not even 20% against a two-sided alternative).

(iii) When $goutphr_{t-2}$ is added to the regression from part (i), and we use the 38 observations now available for the regression, $\hat{\beta}_3 \approx .065$ with a t statistic of about .41. Therefore, $goutphr_{t-2}$ need not be in the model.

C11.3 (i) The estimated equation is

$$\widehat{return}_t = \underset{(.087)}{.226} + \underset{(.039)}{.049}\ return_{t-1} - \underset{(.0070)}{.0097}\ return_{t-1}^2$$

$$n = 689,\ R^2 = .0063.$$

(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 $return_t$.

C11.4 (i) The estimated equation in first differences is

$$\widehat{\Delta inf} = \underset{(.348)}{-.078} - \underset{(.314)}{.842}\,\Delta unem$$

$$n = 48, \quad R^2 = .135, \quad \bar{R}^2 = .116.$$

The coefficient on $\Delta unem$ has the sign that implies an inflation-unemployment tradeoff, and the coefficient is quite large in magnitude. The t statistic on $\Delta unem$ is about –2.68, which is very significant. In fact, the estimated coefficient is not statistically different from –1: $(-.842 + 1)/.314 \approx .5$, which would imply a one-for-one tradeoff.

(ii) Based on the R-squareds (or adjusted R-squareds), the model from part (i) explains Δinf better than (11.19): the model with $\Delta unem$ as the explanatory variable explains about three percentage points more of the variation in Δinf.

C11.5 (i) The estimated equation is

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

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

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

(iii) The estimated equation is

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

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

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.

[Instructor's Note: A variation on this exercise is to start with the model in levels and then difference *all* of the independent variables, including the dummy variables *ww2* and *pill*.]

C11.6 (i) The estimated accelerator model is

$$\widehat{\Delta inven_t} = \underset{(3.64)}{2.59} + \underset{(.023)}{.152}\, \Delta GDP_t$$

$$n = 36, \quad R^2 = .554.$$

Both *inven* and *GDP* are measured in billions of dollars, so a one billion dollar change in *GDP* changes inventory investment by \$152 million. $\hat{\beta}_1$ is very statistically significant, with $t \approx 6.61$.

(ii) When we add $r3_t$, we obtain

$$\widehat{\Delta inven_t} = \underset{(3.69)}{3.00} + \underset{(.025)}{.159}\, \Delta GDP_t - \underset{(1.101)}{.895}\, r3_t$$

$$n = 36, \quad R^2 = .562.$$

The sign of $\hat{\beta}_2$ is negative, as predicted by economic theory, and it seems practically large: a one percentage point increase in $r3_t$ reduces inventories by almost \$1 billion. However, $\hat{\beta}_2$ is not statistically different from zero. (Its t statistic is less than one in absolute value.)

If $\Delta r3_t$ is used instead, the coefficient becomes about $-.470$, se = 1.540. So this is even less significant than when $r3_t$ is in the equation. But, without more data, we cannot conclude that interest rates have a ceteris paribus effect on inventory investment.

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 $\mathrm{H}_0\colon \beta_1 = 0$ and the alternative is $\mathrm{H}_1\colon \beta_1 \neq 0$. Estimating the simple regression using the data in CONSUMP.RAW gives

$$\widehat{gc_t} = \underset{(.004)}{.011} + \underset{(.156)}{.446}\, gc_{t-1}$$

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

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.

C11.8 (i) The estimated AR(1) model is

$$\widehat{unem}_t = \underset{(0.52)}{1.49} + \underset{(.089)}{.742}\, unem_{t-1}$$

$$n = 55,\quad R^2 = .566,\quad \hat{\sigma} = .999.$$

In 2003 the unemployment rate was 6.0, so the predicted unemployment rate for 2004 is 1.49 + .742(6) $\approx$ 5.94. From the 2005 *Economic Report of the President* (Table B–42), the U.S. civilian unemployment rate was 5.5. Therefore, the equation overpredicts the 2004 unemployment rate by almost half a percentage point.

(ii) When we add inf_{t-1} to the equation we get

$$\widehat{unem}_t = \underset{(0.44)}{1.30} + \underset{(.078)}{.649}\, unem_{t-1} + \underset{(.039)}{.183}\, inf_{t-1}$$

$$n = 55,\quad R^2 = .696,\quad \hat{\sigma} = .843.$$

Lagged inflation is very statistically significant, with a t statistic of about 4.7.

(iii) To use the equation from part (ii) to predict unemployment in 2004, we also need the inflation rate for 2003. This is given in PHILLIPS.RAW as 2.3. Therefore, the prediction of *unem* in 2003 is 1.30 + .649(6) + .183(2.3) $\approx$ 5.61. While still too large, it is pretty close to the actual rate of 5.5 percent, and it is certainly better than the predication from part (i).

(iv) We use the model from part (iii) because inf_{t-1} is very significant. To use the 95% prediction interval from Section 6.4, we assume that $unem_t$ has a conditional normal distribution. As shown in equation (6.36), we need the standard error of the predicted value as well as the standard error of the regressions. The latter is given in part (ii), $\hat{\sigma}$ = .843. To obtain the standard error of the predicted value, se($\hat{y}_0$) in the notation of Chapter 6, we need to find the standard error of $\hat{\beta}_0 + (6.0)\hat{\beta}_1 + (2.3)\hat{\beta}_2$. We use the method described in Section 6.4: we run the regression $unem_t$ on ($unem_{t-1}$ – 6.0) and (inf_{t-1} – 2.3), and obtain the intercept and standard error from this regression. We know the intercept must be (approximately) 5.61 from part (iii). Its standard error is about .136. Therefore, from equation (6.36),

$$\text{se}(\hat{y}_0) = [(.136)^2 + (.843)^2]^{1/2} \approx .854.$$

Although the OLS estimators are only approximately normally distributed in the presence of a lagged dependent variable, we use the 97.5th percentile from the normal distribution, 1.96, in constructing the confidence interval. Therefore, the 95% prediction interval for 2004 unemployment is 5.61 ± 1.96(.854), or about 3.94 to 7.28. The actual unemployment rate for 2004, 5.5, is comfortably within this interval. (If we forget to include $\hat{\sigma}$ in obtaining the standard error of the future value, the CI would be about 5.34 to 5.59, which includes 5.5. But this would be an incorrect prediction interval as it ignores the unobservables that affect *unem* in 2004.)

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.

C11.10 (i) Using the data through 2003 gives

$$\Delta\widehat{inf}_t = \underset{(1.22)}{2.83} - \underset{(.209)}{.518}\, unem_t$$

$$n = 55,\ R^2 = .104$$

These estimates are similar to those obtained in equation (11.19), as we would hope. Both the intercept and slope have gotten a little smaller in magnitude.

(ii) The estimate of the natural rate is obtained as in Example 11.5. The new estimate is 2.83/.518 ≈ 5.46, which is slightly smaller than the 5.58 obtained using only the data through 1996.

(iii) The first order autocorrelation of *unem* is about .75. This is one of those tough cases: the correlation between $unem_t$ and $unem_{t-1}$ is large, but it is not especially close to one.

(iv) Just as when we use the data only through 1996, the model with $\Delta unem_t$ as the explanatory variable fits somewhat better (and yields a more pronounced tradeoff between inflation and unemployment):

$$\Delta\widehat{inf}_t = \underset{(.306)}{-.072} - \underset{(.290)}{.833}\ \Delta unem_t$$

$$n = 55,\ R^2 = .135$$

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.

C11.12 (i) The first order autocorrelation in *gmwage*232 is only about −.035, which is very small. There is little doubt that *gmwage*232 is I(0) (weakly dependent).

(ii) The estimated equation is

$$\widehat{gwage232_t} = \underset{(.0004)}{.0024} - \underset{(.034)}{.078}\, gwage232_{t\text{-}1} + \underset{(.0096)}{.1518}\, gmwage_t + \underset{(.082)}{.263}\, gcpi$$

$$n = 610, \quad R^2 = .299$$

Holding last period's wage growth (in sector 232) fixed, along with CPI growth, a 10 percentage point increase in the minimum wage is estimated to increase *gwage*232 by about 1.52 percentage points. The t statistic is over 15, so, statistically, the relationship is very strong.

(iii) When $gemp232_{t\text{-}1}$ is added to the regression, the coefficient (standard error) on *gmwage* are .1527 (.0095), which is very similar to the estimate and standard error from part (ii).

(iv) Without the lags of wage and employment growth, the estimate (standard error) are .1506 (.0097), which are very similar to the estimate with wage growth lagged or wage growth and employment growth lagged. Notice that $gwage232_{t\text{-}1}$ and $gemp_{t\text{-}1}$ are each very statistically significant. Therefore, the reason adding them does not much change the estimate of β_{gmwage} must be that *gmwage* is only slightly correlated with each of them. The regression $gmwage_t$ on $gwage232_{t\text{-}1}$ and $gemp_{t\text{-}1}$ verifies that the linear relationship is weak: the R-squared is only .0039. The collinearity between *gmwage* and the two lagged growth rates is almost nonexistent, so similar estimates with and without the lags is not surprising.

CHAPTER 12

TEACHING NOTES

Most of this chapter deals with serial correlation, but it also explicitly considers heteroskedasticity in time series regressions. The first section allows a review of what assumptions were needed to obtain both finite sample and asymptotic results. Just as with heteroskedasticity, serial correlation itself does not invalidate *R*-squared. In fact, if the data are stationary and weakly dependent, *R*-squared and adjusted *R*-squared consistently estimate the population *R*-squared (which is well-defined under stationarity).

Equation (12.4) is useful for explaining why the usual OLS standard errors are not generally valid with AR(1) serial correlation. It also provides a good starting point for discussing serial correlation-robust standard errors in Section 12.5. The subsection on serial correlation with lagged dependent variables is included to debunk the myth that OLS is always inconsistent with lagged dependent variables and serial correlation. I do not teach it to undergraduates, but I do to master's students.

Section 12.2 is somewhat untraditional in that it begins with an asymptotic *t* test for AR(1) serial correlation (under strict exogeneity of the regressors). It may seem heretical not to give the Durbin-Watson statistic its usual prominence, but I do believe the DW test is less useful than the *t* test. With nonstrictly exogenous regressors I cover only the regression form of Durbin's test, as the *h* statistic is asymptotically equivalent and not always computable.

Section 12.3, on GLS and FGLS estimation, is fairly standard, although I try to show how comparing OLS estimates and FGLS estimates is not so straightforward. Unfortunately, at the beginning level (and even beyond), it is difficult to choose a course of action when they are very different.

I do not usually cover Section 12.5 in a first-semester course, but, because some econometrics packages routinely compute fully robust standard errors, students can be pointed to Section 12.5 if they need to learn something about what the corrections do. I do cover Section 12.5 for a master's level course in applied econometrics (after the first-semester course).

I also do not cover Section 12.6 in class; again, this is more to serve as a reference for more advanced students, particularly those with interests in finance. One important point is that ARCH is heteroskedasticity and not serial correlation, something that is confusing in many texts. If a model contains no serial correlation, the usual heteroskedasticity-robust statistics are valid. I have a brief subsection on correcting for a known form of heteroskedasticity and AR(1) errors in models with strictly exogenous regressors.

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 $\sigma/\sqrt{SST_x}$.

12.2 This statement implies that we are still using OLS to estimate the β_j. But we are not using OLS; we are using feasible GLS (without or with the equation for the first time period). In other words, neither the Cochrane-Orcutt nor the Prais-Winsten estimators are the OLS estimators (and they usually differ from each other).

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.

12.4 This is false, and a source of confusion in several textbooks. (ARCH is often discussed as a way in which the errors can be serially correlated.) As we discussed in Example 12.9, the errors in the equation $return_t = \beta_0 + \beta_1 return_{t-1} + u_t$ are serially uncorrelated, but there is strong evidence of ARCH; see equation (12.51).

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.

12.6 With the strong heteroskedasticity in the errors it is not too surprising that the robust standard error for $\hat{\beta}_1$ differs from the OLS standard error by a substantial amount: the robust standard error is almost 82% larger. Naturally, this reduces the t statistic. The robust t statistic is .059/.069 $\approx$ .86, which is even less significant than before. Therefore, we conclude that, once heteroskedasticity is accounted for, there is very little evidence that $return_{t\text{-}1}$ is useful for predicting $return_t$.

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.

C12.2 (i) After estimating the FDL model by OLS, we obtain the residuals and run the regression $\hat{u}_t$ on $\hat{u}_{t-1}$, using 272 observations. We get $\hat{\rho} \approx .503$ and $t_{\hat{\rho}} \approx 9.60$, which is very strong evidence of positive AR(1) correlation.

(ii) When we estimate the model by iterated C-O, the LRP is estimated to be about 1.110.

(iii) We use the same trick as in Problem 11.5, except now we estimate the equation by iterated C-O. In particular, write

$$gprice_t = \alpha_0 + \theta_0 gwage_t + \delta_1(gwage_{t\text{-}1} - gwage_t) + \delta_2(gwage_{t\text{-}2} - gwage_t)$$
$$+ \ldots + \delta_{12}(gwage_{t\text{-}12} - gwage_t) + u_t,$$

Where θ_0 is the LRP and $\{u_t\}$ is assumed to follow an AR(1) process. Estimating this equation by C-O gives $\hat{\theta}_0 \approx 1.110$ and se($\hat{\theta}_0$) $\approx$ .191. The t statistic for testing H_0: $\theta_0 = 1$ is (1.110 – 1)/.191 $\approx$ .58, which is not close to being significant at the 5% level. So the LRP is not statistically different from one.

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).

C12.4 (i) After obtaining the residuals $\hat{u}_t$ from equation (11.16) and then estimating (12.48), we can compute the fitted values $\hat{h}_t$ = 4.66 – 1.104 $return_t$ for each t. This is easily done in a single command using most software packages. It turns out that 12 of 689 fitted values are negative. Among other things, this means we cannot directly apply weighted least squares using the heteroskedasticity function in (12.48).

(ii) When we add $return_{t-1}^2$ to the equation we get

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

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

The conditional variance is a quadratic in $return_{t-1}$, in this case a U-shape that bottoms out at .789/[2(.297)] ≈ 1.33. Now, there are no estimated variances less than zero.

(iii) Given our finding in part (ii) we can use WLS with the $\hat{h}_t$ obtained from the quadratic heteroskedasticity function. When we apply WLS to equation (12.47) we obtain $\hat{\beta}_0 \approx .155$ (se ≈ .078) and $\hat{\beta}_1 \approx .039$ (se ≈ .046). So the coefficient on $return_{t-1}$, once weighted least squares has been used, is even less significant (t statistic ≈ .85) than when we used OLS.

(iv) To obtain the WLS using an ARCH variance function we first estimate the equation in (12.51) and obtain the fitted values, $\hat{h}_t$. The WLS estimates are now $\hat{\beta}_0 \approx .159$ (se ≈ .076) and $\hat{\beta}_1 \approx .024$ (se ≈ .047). The coefficient and t statistic are even smaller. Therefore, once we account for heteroskedasticity via one of the WLS methods, there is virtually no evidence that E($return_t | return_{t-1}$) depends linearly on $return_{t-1}$.

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

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

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

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 [$\cdot$] 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*$\cdot$*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.

C12.6 (i) The regression $\hat{u}_t$ on $\hat{u}_{t-1}$ (with 35 observations) gives $\hat{\rho} \approx -.089$ and se($\hat{\rho}$) $\approx .178$; there is no evidence of AR(1) serial correlation in this equation, even though it is a static model in the growth rates.

(ii) We regress gc_t on gc_{t-1} and obtain the residuals $\hat{u}_t$. Then, we regress $\hat{u}_t^2$ on gc_{t-1} and gc_{t-1}^2 (using 35 observations), the F statistic (with 2 and 32 *df*) is about 1.08. The p-value is about .352, and so there is little evidence of heteroskedasticity in the AR(1) model for gc_t. This means that we need not modify our test of the PIH by correcting somehow for heteroskedasticity.

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:

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

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

(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.

C12.8 (i) This is the model that was estimated in part (vi) of Computer Exercise C10.11. After getting the OLS residuals, $\hat{u}_t$, we run the regression $\hat{u}_t$ on $\hat{u}_{t-1}, t = 2,...,108$. (Included an intercept, but that is unimportant.) The coefficient on $\hat{u}_{t-1}$ is $\hat{\rho} = .281$ (se = .094). Thus, there is evidence of some positive serial correlation in the errors ($t \approx 2.99$). I strong case can be made that all explanatory variables are strictly exogenous. Certainly there is no concern about the time trend, the seasonal dummy variables, or *wkends*, as these are determined by the calendar. It is seems safe to assume that unexplained changes in *prcfat* today do not cause future changes in the state-wide unemployment rate. Also, over this period, the policy changes were permanent once they occurred, so strict exogeneity seems reasonable for *spdlaw* and *beltlaw*. (Given legislative lags, it seems unlikely that the dates the policies went into effect had anything to do with recent, unexplained changes in *prcfat*.

(ii) Remember, we are still estimating the β_j by OLS, but we are computing different standard errors that have some robustness to serial correlation. Using Stata 7.0, I get

$\hat{\beta}_{spdlaw} = .0671$, $\text{se}(\hat{\beta}_{spdlaw}) = .0267$ and $\hat{\beta}_{beltlaw} = -.0295$, $\text{se}(\hat{\beta}_{beltlaw}) = .0331$. The *t* statistic for *spdlaw* has fallen to about 2.5, but it is still significant. Now, the *t* statistic on *beltlaw* is less than one in absolute value, so there is little evidence that *beltlaw* had an effect on *prcfat*.

(iii) For brevity, I do not report the time trend and monthly dummies. The final estimate of ρ is $\hat{\rho} = .289$:

$$\begin{aligned}\widehat{prcfat} = \underset{(.102)}{1.009} + \ldots + \underset{(.00500)}{.00062}\, wkends - \underset{(.0055)}{.0132}\, unem \\ + \underset{(.0268)}{.0641}\, spdlaw - \underset{(.0301)}{.0248}\, beltlaw\end{aligned}$$

$n = 108$, $R^2 = .641$

There are no drastic changes. Both policy variable coefficients get closer to zero, and the standard errors are bigger than the incorrect OLS standard errors [and, coincidentally, pretty close to the Newey-West standard errors for OLS from part (ii)]. So the basic conclusion is the same: the increase in the speed limit appeared to increase *prcfat*, but the seat belt law, while it is estimated to decrease *prcfat*, does not have a statistically significant effect.

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

$$\begin{aligned}\widehat{\log(avgprc)} = \underset{(.190)}{-.920} - \underset{(.0014)}{.0012}\, t - \underset{(.1141)}{.0182}\, mon - \underset{(.1121)}{.0085}\, tues + \underset{(.1117)}{.0500}\, wed + \underset{(.1110)}{.1225}\, thurs \\ + \underset{(.0218)}{.0909}\, wave2 + \underset{(.0208)}{.0474}\, wave3\end{aligned}$$

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

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 $\hat{\rho} = .618, \text{se}(\hat{\rho}) = .081, t_{\hat{\rho}} = 7.63.$ 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

$$\widehat{\log(avgprc)} = \underset{(.239)}{-.658} - \underset{(.0029)}{.0007}\, t + \underset{(.0652)}{.0099}\, mon + \underset{(.0744)}{.0025}\, tues + \underset{(.0746)}{.0624}\, wed + \underset{(.0621)}{.1174}\, thurs$$

$$+ \underset{(.0174)}{.0497}\, wave2 + \underset{(.0174)}{.0323}\, wave3$$

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

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.

C12.10 (i) OLS estimation using all of the data gives

$$\widehat{inf} = \underset{(1.55)}{1.05} + \underset{(.266)}{.502}\, unem$$

$n = 56$, $R^2 = .062$, $\bar{R}^2 = .045.$

(ii) I included an intercept and got $\hat{\rho}$ = .572 with t = 5.28, which is very strong evidence of positive serial correlation.

(iii) The iterative Prais-Winsten estimates are

$$\widehat{inf} = \underset{(2.05)}{8.00} - \underset{(.290)}{.714}\, unem$$

$$n = 56,\ R^2 = .135,\ \bar{R}^2 = .119.$$

The slope estimate, –.714, is almost identical to that using the data through 1996, –.716. (Adding more data has reduced the standard error.)

(iv) The iterative C-O estimates are

$$\widehat{inf} = \underset{(2.16)}{7.28} - \underset{(.294)}{.663}\, unem$$

$$n = 55,\ R^2 = .088,\ \bar{R}^2 = .070,$$

and the final estimate of ρ is .782. The final estimate of ρ for PW is .789, which is very close to the C-O estimate. The slope coefficients differ by more than we might expect: –.663 for CO and –.714 for PW. Using the first observation has a some effect, although the estimates give the same basic story.

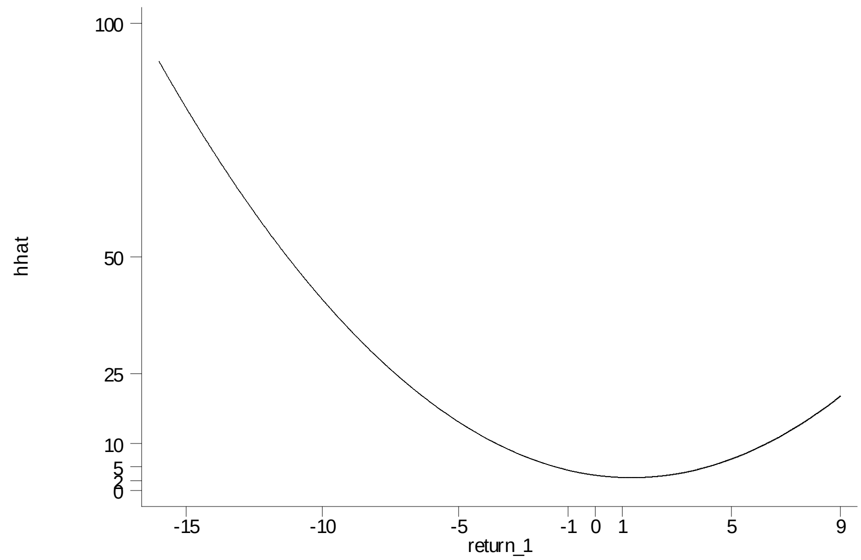
C12.11 (i) The average of $\hat{u}_i^2$ over the sample is 4.44, with the smallest value being .0000074 and the largest being 232.89.

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

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

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

(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).

C12.12 (i) The regression for AR(1) serial correlation gives $\hat{\rho} = -.110$ with $t = -.63$. The estimate of rho is small and statistically insignificant, so AR(1) serial correlation does not appear to be a problem (as is often the case for regressions using differenced data).

(ii) The OLS and PW estimates of β_1 are .152, rounded to three digits. Given that part (i) indicated little problem of serial correlation, we hope that they would be very similar.

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.

C12.14 (i) The test that maintains strict exogeneity gives $\hat{\rho}$ = –.097 (t = –2.41), whereas the regression that includes $gmwage_t$ and $gcpi_t$ gives $\hat{\rho}$ = –.098 (t = –2.42). Therefore, we find evidence of some negative serial correlation, and it does not matter which form of the test we use.

(ii) The estimated equation, with the usual OLS standard errors in (·) and the twelve-lag Newey-West standard errors in [·], is

$$\begin{array}{lcccccc} \widehat{gwage232} & = & .0022 & + & .151\, gmwage & + & .244\, gcpi \\ & & (.0004) & & (.010) & & (.082) \\ & & [.0003] & & [.045] & & [.064] \end{array}$$

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

The Newey-West standard error on *gmwage* is much larger than the usual OLS standard error; roughly, 4.5 times larger. Interestingly, the Newey-West standard error on *gcpi* is actually smaller than the usual OLS standard error.

(iii) The equation with heteroskedasticity-robust standard errors in {·} is

$$\begin{array}{lcccccc} \widehat{gwage232} & = & .0022 & + & .151\, gmwage & + & .244\, gcpi \\ & & (.0004) & & (.010) & & (.082) \\ & & [.0003] & & [.045] & & [.064] \\ & & \{.0004\} & & \{.045\} & & \{.091\} \end{array}$$

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

Certainly for the key variable, *gmwage*, heteroskedasticity is where all the action is. Adjusting the standard error for serial correlation in addition to heteroskedasticity – which is what Newey-West does – makes no difference. Probably because of the negative serial correlation, adjusting the standard error on *gcpi* actually reduces it. Heteroskedasticity does not have a major effect on the *gcpi* standard error.

(iv) The Breusch-Pagan test gives F = 233.8, which implies a p-value of essentially zero. There is very strong evidence of heteroskedasticity.

(v) Oddly, the *p*-value for the usual *F* test is about .058, while for the heteroskedasticity-robust test, it is zero to four decimal places (F = 4.53). In the static model, using the heteroskedasticity-robust *t* statistic led to a less significant minimum wage effect. But the heteroskedasticity-robust test for the lags finds very strong significance.

(vi) The Newey-West version of the *F* statistic is 7.79, which is even more statistically significant than just the heteroskedasticity-robust statistic. So, adjusting the *F* statistic for heteroskedasticity or heteroskedasticity and (twelfth-order) serial correlation leads to the conclusion that the lags are very statistically significant.

(vii) With the 12 lags, the estimated LRP is about .198, and without the lags the estimated LRP is just the coefficient on *gmwage*, .151. So it is about 31% larger with the lags included. Using the Newey-West standard error, the 95% confidence interval for the LRP is from .111 to .284, which easily contains the estimate from the static model.

CHAPTER 13

TEACHING NOTES

While this chapter falls under "Advanced Topics," most of this chapter requires no more sophistication than the previous chapters. (In fact, I would argue that, with the possible exception of Section 13.5, this material is easier than some of the time series chapters.)

Pooling two or more independent cross sections is a straightforward extension of cross-sectional methods. Nothing new needs to be done in stating assumptions, except possibly mentioning that random sampling in each time period is sufficient. The practically important issue is allowing for different intercepts, and possibly different slopes, across time.

The natural experiment material and extensions of the difference-in-differences estimator is widely applicable and, with the aid of the examples, easy to understand.

Two years of panel data are often available, in which case differencing across time is a simple way of removing g unobserved heterogeneity. If you have covered Chapter 9, you might compare this with a regression in levels using the second year of data, but where a lagged dependent variable is included. (The second approach only requires collecting information on the dependent variable in a previous year.) These often give similar answers. Two years of panel data, collected before and after a policy change, can be very powerful for policy analysis.

Having more than two periods of panel data causes slight complications in that the errors in the differenced equation may be serially correlated. (However, the traditional assumption that the errors in the original equation are serially uncorrelated is not always a good one. In other words, it is not always more appropriate to used fixed effects, as in Chapter 14, than first differencing.) With large N and relatively small T, a simple way to account for possible serial correlation after differencing is to compute standard errors that are robust to arbitrary serial correlation and heteroskedasticity. Econometrics packages that do cluster analysis (such as Stata) often allow this by specifying each cross-sectional unit as its own cluster.

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) ≈ –.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.

13.2 The first equation omits the 1981 year dummy variable, *y81*, and so does not allow any appreciation in nominal housing prices over the three year period in the absence of an incinerator. The interaction term in this case is simply picking up the fact that even homes that are near the incinerator site have appreciated in value over the three years. This equation suffers from omitted variable bias.

The second equation omits the dummy variable for being near the incinerator site, *nearinc*, which means it does not allow for systematic differences in homes near and far from the site before the site was built. If, as seems to be the case, the incinerator was located closer to less valuable homes, then omitting *nearinc* attributes lower housing prices too much to the incinerator effect. Again, we have an omitted variable problem. This is why equation (13.9) (or, even better, the equation that adds a full set of controls), is preferred.

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.

13.4 The sign of β_1 does not affect the direction of bias in the OLS estimator of $\hat{\beta}_1$, but only whether we underestimate or overestimate the effect of interest. If we write $\Delta crmrte_i = \delta_0 + \beta_1\Delta unem_i + \Delta u_i$, where Δu_i and $\Delta unem_i$ are negatively correlated, then there is a downward bias in the OLS estimator of β_1. Because $\beta_1 > 0$, we will tend to underestimate the effect of unemployment on crime.

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.

13.6 (i) Let *FL* be a binary variable equal to one if a person lives in Florida, and zero otherwise. Let *y90* be a year dummy variable for 1990. Then, from equation (13.10), we have the linear probability model

$$arrest = \beta_0 + \delta_0 y90 + \beta_1 FL + \delta_1 y90 \cdot FL + u.$$

The effect of the law is measured by δ_1, which is the change in the probability of drunk driving arrest due to the new law in Florida. Including *y90* allows for aggregate trends in drunk driving arrests that would affect both states; including *FL* allows for systematic differences between Florida and Georgia in either drunk driving behavior or law enforcement.

(ii) It could be that the populations of drivers in the two states change in different ways over time. For example, age, race, or gender distributions may have changed. The levels of education across the two states may have changed. As these factors might affect whether someone is arrested for drunk driving, it could be important to control for them. At a minimum, there is the possibility of obtaining a more precise estimator of δ_1 by reducing the error variance. Essentially, any explanatory variable that affects *arrest* can be used for this purpose. (See Section 6.3 for discussion.)

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*, …, *y84* using all 1,129 observations. The null hypothesis of homoskedasticity is H_0: $\gamma_1 = 0$, $\gamma_2 = 0$, … , $\gamma_6 = 0$. 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.

C13.2 (i) The coefficient on *y85* is roughly the proportionate change in *wage* for a male (*female* = 0) with zero years of education (*educ* = 0). This is not especially useful because the U.S. working population without any education is a small group; such people are in no way "typical."

(ii) What we want to estimate is $\theta_0 = \delta_0 + 12\delta_1$; this is the change in the intercept for a male with 12 years of education, where we also hold other factors fixed. If we write $\delta_0 = \theta_0 - 12\delta_1$, plug this into (13.1), and rearrange, we get

$$\begin{aligned}\log(wage) = {} & \beta_0 + \theta_0 y85 + \beta_1 educ + \delta_1 y85\cdot(educ - 12) + \beta_2 exper + \beta_3 exper^2 \\ & + \beta_4 union + \beta_5 female + \delta_5 y85\cdot female + u.\end{aligned}$$

Therefore, we simply replace $y85\cdot educ$ with $y85\cdot(educ - 12)$, and then the coefficient and standard error we want is on *y85*. These turn out to be $\hat{\theta}_0 = .339$ and se($\hat{\theta}_0$) = .034. Roughly, the nominal increase in wage is 33.9%, and the 95% confidence interval is 33.9 ± 1.96(3.4), or about 27.2% to 40.6%. (Because the proportionate change is large, we could use equation (7.10), which implies the point estimate 40.4%; but obtaining the standard error of this estimate is harder.)

(iii) Only the coefficient on *y85* differs from equation (13.2). The new coefficient is about –.383 (se $\approx$ .124). This shows that real wages have fallen over the seven year period, although less so for the more educated. For example, the proportionate change for a male with 12 years of education is –.383 + .0185(12) = –.161, or a fall of about 16.1%. For a male with 20 years of education there has been almost no change [–.383 + .0185(20) = –.013].

(iv) The R-squared when log(*rwage*) is the dependent variable is .356, as compared with .426 when log(*wage*) is the dependent variable. If the SSRs from the regressions are the same, but the R-squareds are not, then the total sum of squares must be different. This is the case, as the dependent variables in the two equations are different.

(v) In 1978, about 30.6% of workers in the sample belonged to a union. In 1985, only about 18% belonged to a union. Therefore, over the seven-year period, there was a notable fall in union membership.

(vi) When $y85 \cdot union$ is added to the equation, its coefficient and standard error are about –.00040 (se $\approx$ .06104). This is practically very small and the t statistic is almost zero. There has been no change in the union wage premium over time.

(vii) Parts (v) and (vi) are not at odds. They imply that while the economic return to union membership has not changed (assuming we think we have estimated a causal effect), the fraction of people reaping those benefits has fallen.

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.

C13.4 (i) In addition to *male* and *married*, we add the variables *head*, *neck*, *upextr*, *trunk*, *lowback*, *lowextr*, and *occdis* for injury type, and *manuf* and *construc* for industry. The coefficient on $afchnge \cdot highearn$ becomes .231 (se $\approx$ .070), and so the estimated effect and t statistic are now larger than when we omitted the control variables. The estimate .231 implies a substantial response of *durat* to the change in the cap for high-earnings workers.

(ii) The R-squared is about .041, which means we are explaining only a 4.1% of the variation in log(*durat*). This means that there are some very important factors that affect log(*durat*) that we are not controlling for. While this means that predicting log(*durat*) would be very difficult for a particular individual, it does not mean that there is anything biased about $\hat{\delta}_1$: it could still be an unbiased estimator of the causal effect of changing the earnings cap for workers' compensation.

(iii) The estimated equation using the Michigan data is

$$\begin{aligned}\widehat{\log(durat)} &= \underset{(0.057)}{1.413} + \underset{(.085)}{.097}\, afchnge + \underset{(.106)}{.169}\, highearn + \underset{(.154)}{.192}\, afchnge \cdot highearn \\ n &= 1{,}524,\quad R^2 = .012.\end{aligned}$$

The estimate of δ_1, .192, is remarkably close to the estimate obtained for Kentucky (.191). However, the standard error for the Michigan estimate is much higher (.154 compared with .069). The estimate for Michigan is not statistically significant at even the 10% level against $\delta_1 > 0$. Even though we have over 1,500 observations, we cannot get a very precise estimate. (For Kentucky, we have over 5,600 observations.)

C13.5 (i) Using pooled OLS we obtain

$$\begin{aligned}\widehat{\log(rent)} &= \underset{(.535)}{-.569} + \underset{(.035)}{.262}\, d90 + \underset{(.023)}{.041}\log(pop) + \underset{(.053)}{.571}\log(avginc) + \underset{(.0010)}{.0050}\, pctstu \\ n &= 128,\quad R^2 = .861.\end{aligned}$$

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

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

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

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 ≈ 4). Note that serial correlation is no longer an issue because we have no time component in the first-differenced equation.

C13.6 (i) You may use an econometrics software package that directly tests restrictions such as H_0: $\beta_1 = \beta_2$ after estimating the unrestricted model in (13.22). But, as we have seen many times, we can simply rewrite the equation to test this using any regression software. Write the differenced equation as

$$\Delta\log(crime) = \delta_0 + \beta_1\Delta clrprc_{-1} + \beta_2\Delta clrprc_{-2} + \Delta u.$$

Following the hint, we define $\theta_1 = \beta_1 - \beta_2$, and then write $\beta_1 = \theta_1 + \beta_2$. Plugging this into the differenced equation and rearranging gives

$$\Delta\log(crime) = \delta_0 + \theta_1\Delta clrprc_{-1} + \beta_2(\Delta clrprc_{-1} + \Delta clrprc_{-2}) + \Delta u.$$

Estimating this equation by OLS gives $\hat{\theta}_1 = .0091$, se($\hat{\theta}_1$) = .0085. The t statistic for H_0: $\beta_1 = \beta_2$ is .0091/.0085 $\approx$ 1.07, which is not statistically significant.

(ii) With $\beta_1 = \beta_2$ the equation becomes (without the i subscript)

$$\Delta\log(crime) = \delta_0 + \beta_1(\Delta clrprc_{-1} + \Delta clrprc_{-2}) + \Delta u$$
$$= \delta_0 + \delta_1[(\Delta clrprc_{-1} + \Delta clrprc_{-2})/2] + \Delta u,$$

where $\delta_1 = 2\beta_1$. But $(\Delta clrprc_{-1} + \Delta clrprc_{-2})/2 = \Delta avgclr$.

(iii) The estimated equation is

$$\widehat{\Delta\log(crime)} = \underset{(.063)}{.099} - \underset{(.0051)}{.0167}\,\Delta avgclr$$

$$n = 53,\quad R^2 = .175,\quad \bar{R}^2 = .159.$$

Since we did not reject the hypothesis in part (i), we would be justified in using the simpler model with *avgclr*. Based on adjusted *R*-squared, we have a slightly worse fit with the restriction imposed. But this is a minor consideration. Ideally, we could get more data to determine whether the fairly different unconstrained estimates of β_1 and β_2 in equation (13.22) reveal true differences in β_1 and β_2.

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

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

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

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 from 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

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

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

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.

C13.8 (i) The estimated equation using differences is

$$\widehat{\Delta vote} = -2.56 - 1.29\,\Delta\log(inexp) - .599\,\Delta\log(chexp) + .156\,\Delta incshr$$
$$(0.63) \quad (1.38) \qquad\qquad (.711) \qquad\qquad (.064)$$

$$n = 157,\quad R^2 = .244,\quad \bar{R}^2 = .229.$$

Only $\Delta incshr$ is statistically significant at the 5% level (t statistic $\approx$ 2.44, p-value $\approx$ .016). The other two independent variables have t statistics less than one in absolute value.

(ii) The F statistic (with 2 and 153 *df*) is about 1.51 with p-value $\approx$ .224. Therefore, $\Delta\log(inexp)$ and $\Delta\log(chexp)$ are jointly insignificant at even the 20% level.

(iii) The simple regression equation is

$$\widehat{\Delta vote} = -2.68 + .218\,\Delta incshr$$
$$(0.63) \qquad (.032)$$

$$n = 157,\quad R^2 = .229,\quad \bar{R}^2 = .224.$$

This equation implies that a 10 percentage point increase in the incumbent's share of total spending increases the percent of the incumbent's vote by about 2.2 percentage points.

(iv) Using the 33 elections with repeat challengers we obtain

$$\widehat{\Delta vote} = \underset{(1.00)}{-2.25} + \underset{(.085)}{.092}\,\Delta incshr$$

$$n = 33,\quad R^2 = .037,\quad \bar{R}^2 = .006.$$

The estimated effect is notably smaller and, not surprisingly, the standard error is much larger than in part (iii). While the direction of the effect is the same, it is not statistically significant (p-value $\approx$ .14 against a one-sided alternative).

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.

C13.10 (i) The estimated equation using the 1987 to 1988 and 1988 to 1989 changes, where we include a year dummy for 1989 in addition to an overall intercept, is

$$\widehat{\Delta hrsemp} = \underset{(1.942)}{-.740} + \underset{(2.65)}{5.42\, d89} + \underset{(2.97)}{32.60\, \Delta grant} + \underset{(5.55)}{2.00\, \Delta grant_{-1}} + \underset{(4.868)}{.744\, \Delta\log(employ)}$$

$$n = 251,\ \ R^2 = .476,\ \ \bar{R}^2 = .467.$$

There are 124 firms with both years of data and three firms with only one year of data used, for a total of 127 firms; 30 firms in the sample have missing information in both years and are not used at all. If we had information for all 157 firms, we would have 314 total observations in estimating the equation.

(ii) The coefficient on *grant* – more precisely, on $\Delta grant$ in the differenced equation – means that if a firm received a grant for the current year, it trained each worker an average of 32.6 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 grant 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 hours per employee by only .074. [Recall: $\widehat{\Delta hrsemp} \approx (.744/100)(\%\Delta employ)$.] This is very small, and the *t* statistic is also rather small.

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

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

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

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

$$\widehat{\Delta math4} = \underset{(.55)}{6.16} + \underset{(.77)}{5.70}\,y95 - \underset{(.79)}{6.80}\,y96 - \underset{(.74)}{8.99}\,y97 + \underset{(.74)}{8.45}\,y98$$

$$- \underset{(3.04)}{1.41}\,\Delta\log(rexpp) + \underset{(2.79)}{11.04}\,\Delta\log(rexpp_{-1}) + \underset{(1.18)}{2.14}\,\Delta\log(enroll)$$

$$+ \underset{(.061)}{.073}\,\Delta lunch$$

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

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.

C13.12. (i) The estimated equation using pooled OLS is

$$\widehat{mrdrte} = \underset{(4.43)}{-5.28} - \underset{(2.14)}{2.07}\,d93 + \underset{(.263)}{.128}\,exec + \underset{(0.78)}{2.53}\,unem$$

$$n = 102, \quad R^2 = .102, \quad \bar{R}^2 = .074.$$

Because the coefficient on *exec* is positive (but statistically insignificant), there is no evidence of a deterrent effect. In using pooled OLS, we are exploiting only the cross-sectional variation in the data. If states that have had high murder rates in the past have reacted by implementing capital punishment, we can see a positive relationship between murder rates and capital punishment even if there is a deterrent effect. (Yet again, we must distinguish between correlation and causality.)

(ii) If we difference away the unobserved state effects – which can include historical factors that lead to higher murder rates and aggressive use of capital punishment – the story is different. The FD estimates are

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

$$n = 51, \quad R^2 = .110, \quad \bar{R}^2 = .073.$$

Now we find a deterrent effect: one more execution in the prior three years is estimated to decrease the murder rate by about .10, or about one murder per million people (because *mrdrte* is measured as murders per 100,000 people). The t statistic on $\Delta exec$ is about −2.4, and so the effect is statistically significant. [The estimated deterrent effect turns out not to be robust to small changes in the data used. See Computer Exercise C14.7.] Note how the unemployment effect has become statistically insignificant.

(iii) The BP and White tests both test two restrictions in this case. The BP F statistic is .60 and the White F statistic is .58. Both have p-values above .50, so there is no evidence of heteroskedasticity in the FD equation.

(iv) The heteroskedasticity-robust t statistic on $\Delta exec$ is −6.11, which is a huge increase in magnitude. This is a bit puzzling for two reasons. First, the tests for heteroskedasticity find essentially no evidence for heteroskedasticity. Second, it is rare to find a heteroskedasticity-robust standard error that is so much smaller than the usual OLS standard error.

(v) I would tend to go with the usual OLS t statistic because it gives a more cautious conclusion and there is no evidence of heteroskedasticity that should affect the t statistics. The usual two-sided p-value is about .02. The heteroskedasticity-robust p-value is zero to many decimal places, and it is hard to believe we have that much confidence in finding an effect. This is a case where it is important to remember that the robust standard errors (and, therefore, the robust t statistics) are only justified in large samples. $n = 51$ may just not be a large enough sample size with this kind of data set to produce reliable heteroskedasticity-robust statistics.

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.)

C13.14 (i) The simple regression estimates, with usual OLS standard errors in (·) and heteroskedasticity-robust standard errors in [·], are

$$
\begin{array}{llll}
\widehat{re78} & = & 21.55 & -\ 15.20\ train \\
 & & (0.30) & \quad (1.15) \\
 & & [0.31] & \quad [0.66]
\end{array}
$$

$$n = 2{,}675, \quad R^2 = .061$$

The estimated training effect is actually negative, large in absolute value – some $15,200 less in real earnings in 1978 (just about one standard deviation, which is $15,633). Further, using either t statistic, the coefficient is very statistically significant.

(ii) With *cre* = *re*78 – *re*75, the simple regression results are

$$\widehat{cre} = \underset{\substack{(0.21)\\ [0.22]}}{2.49} + \underset{\substack{(0.81)\\ [0.64]}}{2.33}\, train$$

$$n = 2{,}675, \quad R^2 = .003$$

Now the estimated training effect is positive (about $2,330) and statistically significant. This is a huge difference compared with the original OLS estimate. It seems pretty clear that lower skilled workers were more likely to be in a job training program. The differencing, hopefully, eliminates most of the self-selection bias; it certainly seems to eliminate much of it.

(iii) The 95% CI using the usual OLS standard error is about .74 to 3.92. Using the heteroskedasticity-robust standard error, the 95% CI is about 1.08 to 3.58. The robust CI is narrower. To be cautious, we might want to use the wider confidence interval. (In addition, we should probably control for other factors in the FD regression.)

CHAPTER 14

TEACHING NOTES

My preference is to view the fixed and random effects methods of estimation as applying to the *same* underlying unobserved effects model. The name "unobserved effect" is neutral to the issue of whether the time-constant effects should be treated as fixed parameters or random variables. With large N and relatively small T, it almost always makes sense to treat them as random variables, since we can just view the unobserved a_i as being drawn from the population along with the observed variables. Especially for undergraduates and master's students, it seems sensible to not raise the philosophical issues underlying the professional debate. In my mind, the key issue in most applications is whether the unobserved effect is correlated with the observed explanatory variables. The fixed effects transformation eliminates the unobserved effect entirely whereas the random effects transformation accounts for the serial correlation in the composite error via GLS. (Alternatively, the random effects transformation only eliminates a fraction of the unobserved effect.)

As a practical matter, the fixed effects and random effects estimates are closer when T is large or when the variance of the unobserved effect is large relative to the variance of the idiosyncratic error. I think Example 14.4 is representative of what often happens in applications that apply pooled OLS, random effects, and fixed effects, at least on the estimates of the marriage and union wage premiums. The random effects estimates are below pooled OLS and the fixed effects estimates are below the random effects estimates.

Choosing between the fixed effects transformation and first differencing is harder, although useful evidence can be obtained by testing for serial correlation in the first-difference estimation. If the AR(1) coefficient is significant and negative (say, less than −.3, to pick a not quite arbitrary value), perhaps fixed effects is preferred.

Matched pairs samples have been profitably used in recent economic applications, and differencing or random effects methods can be applied. In an equation such as (14.12), there is probably no need to allow a different intercept for each sister provided that the labeling of sisters is random. The different intercepts might be needed if a certain feature of a sister that is not included in the observed controls is used to determine the ordering. A statistically significant intercept in the differenced equation would be evidence of this.

SOLUTIONS TO PROBLEMS

14.1 First, for each $t > 1$, $\text{Var}(\Delta u_{it}) = \text{Var}(u_{it} - u_{i,t-1}) = \text{Var}(u_{it}) + \text{Var}(u_{i,t-1}) = 2\sigma_u^2$, 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 $\text{E}(\Delta u_{it} \cdot \Delta u_{i,t+1}) = \text{E}[(u_{it} - u_{i,t-1})(u_{i,t+1} - u_{it})] = \text{E}(u_{it}u_{i,t+1}) - \text{E}(u_{it}^2) - \text{E}(u_{i,t-1}u_{i,t+1}) + \text{E}(u_{i,t-1}u_{it}) = -\text{E}(u_{it}^2) = -\sigma_u^2$ because of the no serial correlation assumption. Because the variance is constant across t, by Problem 11.1, $\text{Corr}(\Delta u_{it}, \Delta u_{i,t+1}) = \text{Cov}(\Delta u_{it}, \Delta u_{i,t+1})/\text{Var}(\Delta u_{it}) = -\sigma_u^2/(2\sigma_u^2) = -.5$.

14.2 (i) The between estimator is just the OLS estimator from the cross-sectional regression of $\bar{y}_i$ on $\bar{x}_i$ (including an intercept). Because we just have a single explanatory variable $\bar{x}_i$ and the error term is $a_i + \bar{u}_i$, we have, from Section 5.1,

$$\text{plim}(\tilde{\beta}_1) = \beta_1 + \text{Cov}(\bar{x}_i, a_i + \bar{u}_i)/\text{Var}(\bar{x}_i).$$

But $\text{E}(a_i + \bar{u}_i) = 0$ so $\text{Cov}(\bar{x}_i, a_i + \bar{u}_i) = \text{E}(\bar{x}_i(a_i + \bar{u}_i)] = \text{E}(\bar{x}_i a_i) + \text{E}(\bar{x}_i \bar{u}_i) = \text{E}(\bar{x}_i a_i)$ because $\text{E}(\bar{x}_i \bar{u}_i) = \text{Cov}(\bar{x}_i, \bar{u}_i) = 0$ by assumption. Now $\text{E}(\bar{x}_i a_i) = T^{-1}\sum_{t=1}^{T}\text{E}(x_{it}a_i) = \sigma_{xa}$. Therefore,

$$\text{plim}(\tilde{\beta}_1) = \beta_1 + \sigma_{xa}/\text{Var}(\bar{x}_i),$$

which is what we wanted to show.

(ii) If $\{x_{it}\}$ is serially uncorrelated with constant variance σ_x^2 then $\text{Var}(\bar{x}_i) = \sigma_x^2/T$, and so $\text{plim}\ \tilde{\beta}_1 = \beta_1 + \sigma_{xa}/(\sigma_x^2/T) = \beta_1 + T(\sigma_{xa}/\sigma_x^2)$.

(iii) As part (ii) shows, when the x_{it} are pairwise uncorrelated the magnitude of the inconsistency actually increases linearly with T. The sign depends on the covariance between x_{it} and a_i.

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

(ii) $\text{Var}(v_{it} - \lambda\bar{v}_i) = \text{Var}(v_{it}) + \lambda^2\text{Var}(\bar{v}_i) - 2\lambda\cdot\text{Cov}(v_{it}, \bar{v}_i) = \sigma_v^2 + \lambda^2\,\text{E}(\bar{v}_i^2) - 2\lambda\cdot\text{E}(v_{it}\bar{v}_i)$.

Now, $\sigma_v^2 = \text{E}(v_{it}^2) = \sigma_a^2 + \sigma_u^2$ and $\text{E}(v_{it}\bar{v}_i) = T^{-1}\sum_{s=1}^{T}E(v_{it}v_{is}) = T^{-1}[\sigma_a^2 + \sigma_a^2 + \ldots + (\sigma_a^2 + \sigma_u^2) + \ldots + \sigma_a^2] = \sigma_a^2 + \sigma_u^2/T$. Therefore, $\text{E}(\bar{v}_i^2) = T^{-1}\sum_{t=1}^{T}E(v_{it}\bar{v}_i) = \sigma_a^2 + \sigma_u^2/T$. 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

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

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):

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

14.4 (i) Men's athletics are still the most prominent, although women's sports, especially basketball but also gymnastics, softball, and volleyball, are very popular at some universities. Winning percentages for football and men's and women's basketball are good possibilities, as well as indicators for whether teams won conference championships, went to a visible bowl game (football), or did well in the NCAA basketball tournament (such as making the Sweet 16). We must be sure that we use measures of athletic success that are available prior to application deadlines. So, we would probably use football success from the previous school year; basketball success might have to be lagged one more year.

(ii) Tuition could be important: ceteris paribus, higher tuition should mean fewer applications. Measures of university quality that change over time, such as student/faculty ratios or faculty grant money, could be important.

(iii) An unobserved effects model is

$$\log(apps_{it}) = \delta_1 d90_t + \delta_2 d95_t + \beta_1 athsucc_{it} + \beta_2 \log(tuition_{it}) + \ldots + a_i + u_{it},\ t = 1,2,3.$$

The variable $athsucc_{it}$ is shorthand for a measure of athletic success; we might include several measures. If, for example, $athsucc_{it}$ is football winning percentage, then $100\beta_1$ is the percentage change in applications given a one percentage point increase in winning percentage. It is likely that a_i is correlated with athletic success, tuition, and so on, so fixed effects estimation is appropriate. Alternatively, we could first difference to remove a_i, as discussed in Chapter 13.

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.

14.6 (i) The fully robust standard errors are larger in each case, roughly double for the time-constant regressors *educ*, *black*, and *hispan*. On the time-varying explanatory variables *married* and *union*, the fully robust standard errors are roughly 60 percent larger. The differences are smaller for *exper* and $exper^2$ but hardly trivial. We expect this if we think the composite error term, v_{it}, contains an unobserved effect, a_i. This induces positive serial correlation and, as we saw in Section 12.1 for time series, the usual OLS standard errors tend to understate the actual sampling variation in the OLS estimates. The same holds true for pooled OLS with panel data.

(ii) On the time constant explanatory variables *educ*, *black*, and *hispan*, the RE standard errors and the robust standard errors for pooled OLS are roughly the same. (The coefficient estimates are very similar, too.) The main differences arise in the standard errors (and coefficients) on the time-varying explanatory variables. For example, the RE standard errors on the *married* and *union* coefficients are .017 and .018, respectively, compared with the robust standard errors for pooled OLS of .026 and .027. We expect this to be true because, under the under the RE assumptions, RE is more efficient than pooled OLS.

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.

[Instructor's Note: Some econometrics packages do report an intercept for fixed effects estimation; if so, it is usually the average of the estimated intercepts for the cross-sectional units, and it is not especially informative. If one obtains the FE estimates via the dummy variable regression, an intercept is reported for the base group, which is usually an arbitrarily chosen cross-sectional unit.]

C14.2 (i) We report the fixed effects estimates in equation form as

$$\widehat{\log(crmrte_{it})} = \underset{(.022)}{.013}\, d82_t - \underset{(.021)}{.079}\, d83_t - \underset{(.022)}{.118}\, d84_t - \underset{(.022)}{.112}\, d85_t$$
$$- \underset{(.021)}{.082}\, d86_t - \underset{(.021)}{.040}\, d87_t - \underset{(.032)}{.360} \log(prbarr_{it}) - \underset{(.021)}{.286} \log(prbconv_{it})$$
$$- \underset{(.032)}{.183} \log(prbpris_{it}) - \underset{(.0264)}{.0045} \log(avgsen_{it}) + \underset{(.026)}{.424} \log(polpc_{it})$$
$$N = 90, \quad T = 7.$$

There is no intercept because it gets swept away in the time demeaning. If your econometrics package reports a constant or intercept, it is choosing one of the cross-sectional units as the base group, and then the overall intercept is for the base unit in the base year. This overall intercept is not very informative because, without obtaining each $\hat{a}_i$, we cannot compare across units.

Remember that the coefficients on the year dummies are not directly comparable with those in the first-differenced equation because we did not difference the year dummies in (13.33). The fixed effects estimates are unbiased estimators of the parameters on the time dummies in the original model.

The first-difference and fixed effects slope estimates are broadly consistent. The variables that are significant with first differencing are significant in the FE estimation, and the signs are all the same. The magnitudes are also similar, although, with the exception of the insignificant variable log(*avgsen*), the FE estimates are all larger in absolute value. But we conclude that the estimates across the two methods paint a similar picture.

(ii) When the nine log wage variables are added and the equation is estimated by fixed effects, very little of importance changes on the criminal justice variables. The following table contains the new estimates and standard errors.

Independent Variable	Coefficient	Standard Error
log(*prbarr*)	–.356	.032
log(*prbconv*)	–.286	.021
log(*prbpris*)	–.175	.032
log(*avgsen*)	–.0029	.026
log(*polpc*)	.423	.026

The changes in these estimates are minor, even though the wage variables are jointly significant. The F statistic, with 9 and $N(T-1) - k = 90(6) - 20 = 520$ df, is $F \approx 2.47$ with p-value $\approx .0090$.

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

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

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

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

(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.

C14.4 (i) Write the equation for times t and $t - 1$ as

$$\log(uclms_{it}) = a_i + c_i t + \beta_1 ez_{it} + u_{it},$$

$$\log(uclms_{i,t\text{-}1}) = a_i + c_i(t-1) + \beta_1 ez_{i,t\text{-}1} + u_{i,t\text{-}1}$$

and subtract the second equation from the first. The a_i are eliminated and $c_i t - c_i(t-1) = c_i$. So, for each $t \geq 2$, we have

$$\Delta\log(uclms_{it}) = c_i + \beta_1 \Delta ez_{it} + u_{it}.$$

(ii) Because the differenced equation contains the fixed effect c_i, we estimate it by FE. We get $\hat{\beta}_1 = -.251$, se($\hat{\beta}_1$) = .121. The estimate is actually larger in magnitude than we obtain in Example 13.8 [where $\hat{\beta}_1 = -1.82$, se($\hat{\beta}_1$) = .078], but we have not yet included year dummies. In any case, the estimated effect of an EZ is still large and statistically significant.

(iii) Adding the year dummies reduces the estimated EZ effect, and makes it more comparable to what we obtained without $c_i t$ in the model. Using FE on the first-differenced equation gives $\hat{\beta}_1 = -.192$, se($\hat{\beta}_1$) = .085, which is fairly similar to the estimates without the city-specific trends.

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).

C14.6 First, the random effects estimate on $union_{it}$ becomes .174 (se $\approx$ .031), while the coefficient on the interaction term $union_{it} \cdot t$ is about –.0155 (se $\approx$ .0057). Therefore, the interaction between the union dummy and time trend is very statistically significant (t statistic $\approx$ –2.72), and is important economically. While at a given point in time there is a large union differential, the projected wage growth is less for unionized workers (on the order of 1.6% less per year).

The fixed effects estimate on $union_{it}$ becomes .148 (se $\approx$ .031), while the coefficient on the interaction $union_{it} \cdot t$ is about −.0157 (se $\approx$ .0057). Therefore, the story is very similar to that for the random effects estimates.

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

$$\begin{aligned} \widehat{mrdrte}_{it} &= \underset{(4.43)}{-5.28} - \underset{(2.14)}{2.07}\, d93_t + \underset{(.263)}{.128}\, exec_{it} + \underset{(0.78)}{2.53}\, unem_{it} \\ N &= 51,\ T = 2,\ R^2 = .102 \end{aligned}$$

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

$$\begin{aligned} \Delta\widehat{mrdrte}_i &= \underset{(.209)}{.413} - \underset{(.043)}{.104}\, \Delta exec_i - \underset{(.159)}{.067}\, \Delta unem_i \\ n &= 51,\ R^2 = .110 \end{aligned}$$

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 [·]:

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

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

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

$$\widehat{mrdrte_{it}} = \underset{(.75)}{1.56}\, d90_t + \underset{(.70)}{1.73}\, d93_t - \underset{(.177)}{.138}\, exec_{it} + \underset{(.296)}{.221}\, unem_{it}$$

$$N = 51,\ T = 3,\ R^2 = .073$$

The size of the deterrent effect is actually slightly larger than when 1987 is not used. However, the t statistic is only about $-.78$. Thus, while the magnitude of the effect is similar, the statistical significance is not. It is somewhat odd that adding another year of data causes the standard error on the *exec* coefficient to increase nontrivially.

C14.8 (i) The pooled OLS estimates are

$$\widehat{math4} = \underset{(10.30)}{-31.66} + \underset{(.74)}{6.38}\, y94 + \underset{(.79)}{18.65}\, y95 + \underset{(.77)}{18.03}\, y96 + \underset{(.78)}{15.34}\, y97 + \underset{(.78)}{30.40}\, y98$$

$$+ \underset{(2.428)}{.534} \log(rexpp) + \underset{(2.31)}{9.05} \log(rexpp_{-1}) + \underset{(.205)}{.593} \log(enrol) - \underset{(.014)}{.407}\, lunch$$

$N = 550,\ T = 6,\ R^2 = .505$

(ii) The *lunch* variable is the percent of students in the district eligible for free or reduced-price lunches, which is determined by poverty status. Therefore, *lunch* is effectively a poverty rate. We see that the district poverty rate has a large impact on the math pass rate: a one percentage point increase in *lunch* reduces the pass rate by about .41 percentage points.

(iii) I ran the pooled OLS regression $\hat{v}_{it}$ on $\hat{v}_{i,t-1}$ using the years 1994 through 1998 (since the residuals are first available for 1993). The coefficient on $\hat{v}_{i,t-1}$ is $\hat{\rho} = .504$ (se = .017), so there is very strong evidence of positive serial correlation. There are many reasons for positive serial correlation. In the context of panel data, it indicates the presences of a time-constant unobserved effect, a_i.

(iv) The fixed effects estimates are

$$\widehat{math4} = \underset{(.56)}{6.18}\, y94 + \underset{(.69)}{18.09}\, y95 + \underset{(.76)}{17.94}\, y96 + \underset{(.80)}{15.19}\, y97 + \underset{(.84)}{29.88}\, y98$$

$$- \underset{(2.458)}{.411} \log(rexpp) + \underset{(2.37)}{7.00} \log(rexpp_{-1}) + \underset{(1.100)}{.245} \log(enrol) + \underset{(.051)}{.062}\, lunch$$

$N = 550,\ T = 6,\ R^2 = .603$

The coefficient on the lagged spending variable has gotten somewhat smaller, but its *t* statistic is still almost three. Therefore, there is still evidence of a lagged spending effect after controlling for unobserved district effects.

(v) The change in the coefficient and significance on the *lunch* variable is most dramatic. Both *enrol* and *lunch* are slow to change over time, which means that their effects are largely captured by the unobserved effect, a_i. Plus, because of the time demeaning, their coefficients are hard to estimate. The spending coefficients can be estimated more precisely because of a policy change during this period, where spending shifted markedly in 1994 after the passage of Proposal A in Michigan, which changed the way schools were funded.

(vi) The estimated long-run spending effect is $\hat{\theta}_1 = 6.59$, se($\hat{\theta}_1$) = 2.64.

C14.9 (i) The OLS estimates are

$$
\begin{aligned}
\widehat{pctstck} = {} & \underset{(55.17)}{128.54} + \underset{(6.23)}{11.74}\, choice + \underset{(7.23)}{14.34}\, prftshr + \underset{(6.77)}{1.45}\, female - \underset{(.78)}{1.50}\, age \\
& + \underset{(1.20)}{.70}\, educ - \underset{(14.23)}{15.29}\, finc25 + \underset{(14.69)}{.19}\, finc35 - \underset{(14.55)}{3.86}\, finc50 \\
& - \underset{(16.02)}{13.75}\, finc75 - \underset{(15.72)}{2.69}\, finc100 - \underset{(17.80)}{25.05}\, finc101 - \underset{(.0128)}{.0026}\, wealth89 \\
& + \underset{(6.68)}{6.67}\, stckin89 - \underset{(6.38)}{7.50}\, irain89
\end{aligned}
$$

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

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

$$
\begin{aligned}
\Delta\widehat{pctstck} = {} & \underset{(10.94)}{15.93} + \underset{(15.00)}{2.28}\, \Delta choice - \underset{(16.92)}{9.27}\Delta prftshr + \underset{(21.49)}{21.55}\, \Delta female - \underset{(9.00)}{3.57}\, \Delta age \\
& \underset{(3.43)}{-1.22}\, \Delta educ
\end{aligned}
$$

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

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.

C14.10 (i) The pooled OLS estimate of β_1 is about .360. If $\Delta concen = .10$ then $\Delta\widehat{lfare} = .360(.10) = .036$, which means air fare is estimated to be about 3.6% higher.

(ii) The 95% CI obtained using the usual OLS standard error is .301 to .419. But the validity of this standard error requires the composite error to have no serial correlation, which effectively means a_i is not in the equation. The fully robust 95% CI, which allows any kind of serial correlation over the four years (and any kind of heteroskedasticity), is .245 to .475 – quite a bit wider than the usual CI. The wider CI is appropriate, as the neglected serial correlation introduces uncertainty into our parameter estimators.

(iii) The quadratic has a U-shape, and the turning point is about $.902/[2(.103)] \approx 4.38$. This is the value of log(*dist*) where the slope becomes positive. The value of *dist* is exp(4.38), or about 80. The shortest distance in the data set is 95 miles, so the turning point is outside the range of the data (a good thing in this case). What is being captured is an increasing elasticity of *fare* with respect to *dist* as fare increases.

(iv) The RE estimate of β_1 is about .209, which is quite a bit smaller than the pooled OLS estimate. Still, the estimate implies a positive relationship between fare and concentration. The estimate is very statistically significant, too, with $t = 7.88$.

(v) The FE estimate is .169, which is lower yet but not so different from the RE estimate. The value of $\hat{\lambda}$ in the RE estimation is about .900, and so we expect RE and FE to be fairly similar. [Remember, RE uses a quasi-demeaning that depends on the estimate of lamda; see equation (14.11).]

(vi) Factors about the cities near the two airports on a route could affect demand for air travel, such as population, education levels, types of employers, and so on. Of course, each of these can be time-varying, although, over a short stretch of time, they might be roughly constant. The quality of the freeway system and access to trains, along with geographical features (is the city near a river?) would roughly be time-constant. These could certainly be correlated with concentration.

(vii) Accounting for an unobserved effect and using fixed effects gives us a positive, statistically significant relationship. I would go with the FE estimate, .169, which allows for concentration to be correlated with all time-constant features that affect costs and demand.

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.)

C14.12 (i) The smallest number of schools is one; in fact, 271 of the 537 districts have only one elementary school in the sample. The largest number of schools in a district is 162. The average number of schools is about 22.3.

(ii) The OLS estimate of β_{bs} is about −.516 and the usual OLS standard error is .110.

(iii) The standard error on *bs* robust to within-district cluster correlation and heteroskedasticity is about .253, which drops the *t* statistic on *bs* to about −2.04, and so *bs* is only marginally statistically significant.

(iv) As we saw in Computer Exercise C9.12, the coefficient on *bs* is very sensitive to the inclusion of the observation with the highest *bs*. When the four observations with *bs* > .5 are dropped, $\hat{\beta}_{bs} = -.186$, and its robust standard error is .273. The resulting *t* statistic is −.68, which means there is very little evidence of any tradeoff.

(v) Using all of the observations, the FE estimate of β_{bs} is about −1.05 (se = .096), which is very much in line with the theoretical value of −1; certainly, the 95% CI easily contains −1. If we drop the four observations with *bs* > .5, the FE estimate of β_{bs} is about −.523 (se = .147). Now we can reject

(vi) If we just consider estimation with *bs* ≤ .5, pooled OLS gives a small (in absolute value), statistically insignificant estimate. When we allow for a district fixed effect – which eliminates all districts without only one school – the estimate is much larger and statistically different from zero. We reject $H_0: \beta_{bs} = -1$ against the two-sided alternative at the 5% level, but there is evidence of a tradeoff between salary and benefits. The district effect probably captures the reality that some districts pay higher salaries and higher benefits for reasons not captured by the small number of controls (enrollment, staff per student, poverty rate). Once we control for systematic differences across districts, we find a tradeoff. In effect, the fixed effects estimate is based on the variation in salary/benefits packages across schools within districts. One reason different compensation packages may exist within districts is because of different teacher age distributions within districts.

CHAPTER 15

TEACHING NOTES

When I wrote the first edition, I took the novel approach of introducing instrumental variables as a way of solving the omitted variable (or unobserved heterogeneity) problem. Traditionally, a student's first exposure to IV methods comes by way of simultaneous equations models. Occasionally, IV is first seen as a method to solve the measurement error problem. I have even seen texts where the first appearance of IV methods is to obtain a consistent estimator in an AR(1) model with AR(1) serial correlation.

The omitted variable problem is conceptually much easier than simultaneity, and stating the conditions needed for an IV to be valid in an omitted variable context is straightforward. Besides, most modern applications of IV have more of an unobserved heterogeneity motivation. A leading example is estimating the return to education when unobserved ability is in the error term. We are not thinking that education and wages are jointly determined; for the vast majority of people, education is completed before we begin collecting information on wages or salaries. Similarly, in studying the effects of attending a certain type of school on student performance, the choice of school is made and then we observe performance on a test. Again, we are primarily concerned with unobserved factors that affect performance and may be correlated with school choice; it is not an issue of simultaneity.

The asymptotics underlying the simple IV estimator are no more difficult than for the OLS estimator in the bivariate regression model. Certainly consistency can be derived in class. It is also easy to demonstrate how, even just in terms of inconsistency, IV can be worse than OLS if the IV is not completely exogenous.

At a minimum, it is important to always estimate the reduced form equation and test whether the IV is partially correlated with endogenous explanatory variable. The material on multicollinearity and 2SLS estimation is a direct extension of the OLS case. Using equation (15.43), it is easy to explain why multicollinearity is generally more of a problem with 2SLS estimation.

Another conceptually straightforward application of IV is to solve the measurement error problem, although, because it requires two measures, it can be hard to implement in practice.

Testing for endogeneity and testing any overidentification restrictions is something that should be covered in second semester courses. The tests are fairly easy to motivate and are very easy to implement.

While I provide a treatment for time series applications in Section 15.7, I admit to having trouble finding compelling time series applications. These are likely to be found at a less aggregated level, where exogenous IVs have a chance of existing. (See also Chapter 16 for examples.)

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.

15.2 (i) It seems reasonable to assume that *dist* and u are uncorrelated because classrooms are not usually assigned with convenience for particular students in mind.

(ii) The variable *dist* must be partially correlated with *atndrte*. More precisely, in the reduced form

$$atndrte = \pi_0 + \pi_1 priGPA + \pi_2 ACT + \pi_3 dist + v,$$

we must have $\pi_3 \neq 0$. Given a sample of data we can test H_0: $\pi_3 = 0$ against H_1: $\pi_3 \neq 0$ using a t test.

(iii) We now need instrumental variables for *atndrte* and the interaction term, *priGPA·atndrte*. (Even though *priGPA* is exogenous, *atndrte* is not, and so *priGPA·atndrte* is generally correlated with u.) Under the exogeneity assumption that $E(u|priGPA,ACT,dist) = 0$, any function of *priGPA*, *ACT*, and *dist* is uncorrelated with u. In particular, the interaction *priGPA·dist* is uncorrelated with u. If *dist* is partially correlated with *atndrte* then *priGPA·dist* is partially correlated with *priGPA·atndrte*. So, we can estimate the equation

$$stndfnl = \beta_0 + \beta_1 atndrte + \beta_2 priGPA + \beta_3 ACT + \beta_4 priGPA \cdot atndrte + u$$

by 2SLS using IVs *dist*, *priGPA*, *ACT*, and *priGPA·dist*. It turns out this is not generally optimal. It may be better to add $priGPA^2$ and *priGPA·ACT* to the instrument list. This would give us overidentifying restrictions to test. See Wooldridge (2002, Chapters 5 and 9) for further discussion.

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

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

15.4 (i) The state may set the level of its minimum wage at least partly based on past or expected current economic activity, and this could certainly be part of u_t. Then $gMIN_t$ and u_t are correlated, which causes OLS to be biased and inconsistent.

(ii) Because $gGDP_t$ controls for the overall performance of the U.S. economy, it seems reasonable that $gUSMIN_t$ is uncorrelated with the disturbances to employment growth for a particular state.

(iii) In some years, the U.S. minimum wage will increase in such a way so that it exceeds the state minimum wage, and then the state minimum wage will also increase. Even if the U.S. minimum wage is never binding, it may be that the state increases its minimum wage in response to an increase in the U.S. minimum. If the state minimum is always the U.S. minimum, then $gMIN_t$ is exogenous in this equation and we would just use OLS.

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 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).

15.6 (i) Plugging (15.26) into (15.22) and rearranging gives

$$\begin{aligned} y_1 &= \beta_0 + \beta_1(\pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2) + \beta_2 z_1 + u_1 \\ &= (\beta_0 + \beta_1\pi_0) + (\beta_1\pi_1 + \beta_2)z_1 + \beta_1\pi_2 z_2 + u_1 + \beta_1 v_2, \end{aligned}$$

and so $\alpha_0 = \beta_0 + \beta_1\pi_0$, $\alpha_1 = \beta_1\pi_1 + \beta_2$, and $\alpha_2 = \beta_1\pi_2$.

(ii) From the equation in part (i), $v_1 = u_1 + \beta_1 v_2$.

(iii) By assumption, u_1 has zero mean and is uncorrelated with z_1 and z_2, and v_2 has these properties by definition. So v_1 has zero mean and is uncorrelated with z_1 and z_2, which means that OLS consistently estimates the α_j. [OLS would only be unbiased if we add the stronger assumptions $E(u_1|z_1,z_2) = E(v_2|z_1,z_2) = 0$.]

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.

15.8 (i) Family income and background variables, such as parents' education.

(ii) The population model is

$$score = \beta_0 + \beta_1 girlhs + \beta_2 faminc + \beta_3 meduc + \beta_4 feduc + u_1,$$

where the variables are self-explanatory.

(iii) Parents who are supportive and motivated to have their daughters do well in school may also be more likely to enroll their daughters in a girls' high school. It seems likely that *girlhs* and u_1 are correlated.

(iv) Let *numghs* be the number of girls' high schools within a 20-mile radius of a girl's home. To be a valid IV for *girlhs*, *numghs* must satisfy two requirements: it must be uncorrelated with u_1 and it must be partially correlated with *girlhs*. The second requirement probably holds, and can be tested by estimating the reduced form

$$girlhs = \pi_0 + \pi_1 faminc + \pi_2 meduc + \pi_3 feduc + \pi_4 numghs + v_2$$

and testing *numghs* for statistical significance. The first requirement is more problematical. Girls' high schools tend to locate in areas where there is a demand, and this demand can reflect the seriousness with which people in the community view education. Some areas of a state have better students on average for reasons unrelated to family income and parents' education, and these reasons might be correlated with *numghs*. One possibility is to include community-level variables that can control for differences across communities.

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.

15.10 (i) Better and more serious students tend to go to college, and these same kinds of students may be attracted to private and, in particular, Catholic high schools. The resulting correlation between u and *CathHS* is another example of a self-selection problem: students self select toward Catholic high schools, rather than being randomly assigned to them.

(ii) A standardized score is a measure of student ability, so this can be used as a proxy variable in an OLS regression. Having this measure in an OLS regression should be an improvement over having no proxies for student ability.

(iii) The first requirement is that *CathRe1* must be uncorrelated with unobserved student motivation and ability (whatever is not captured by any proxies) and other factors in the error term. This holds if growing up Catholic (as opposed to attending a Catholic high school) does not make you a better student. It seems reasonable to assume that Catholics do not have more innate ability than non-Catholics. Whether being Catholic is unrelated to student motivation, or preparation for high school, is a thornier issue.

The second requirement is that being Catholic has an effect on attending a Catholic high school, controlling for the other exogenous factors that appear in the structural model. This can be tested by estimating the reduced form equation of the form $CathHS = \pi_0 + \pi_1 CathRel$ + (*other exogenous factors*) + (*reduced form error*).

(iv) Evans and Schwab (1995) find that being Catholic substantially increases the probability of attending a Catholic high school. Further, it seems that assuming *CathRe1* is exogenous in the structural equation is reasonable. See Evans and Schwab (1995) for an in-depth analysis.

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^*, $E(x_t e_t) = E[(x_t^* + e_t)e_t] = E(x_t^* e_t) + E(e_t^2) = \sigma_e^2$. Therefore, with v_t defined as above, $Cov(x_t,v_t) = Cov(x_t,u_t) - \beta_1 Cov(x_t,e_t) = -\beta_1 \sigma_e^2 < 0$ 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 $E(x_{t-1}^* u_t) = E(e_{t-1}u_t) = E(x_{t-1}^* e_t) = E(e_{t-1}e_t) = 0$, and so $E(x_{t-1}u_t) = E(x_{t-1}e_t) = 0$ because $x_t = x_t^* + e_t$. Therefore, $E(x_{t-1}v_t) = E(x_{t-1}u_t) - \beta_1 E(x_{t-1}e_t) = 0$.

(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): $\text{Cov}(x_{t-1},v_t) = \text{E}(x_{t-1}v_t) = 0$. 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.

C15.2 (i) The equation estimated by OLS is

$$\widehat{children} = \underset{(0.241)}{-4.138} - \underset{(.0059)}{.0906}\, educ + \underset{(.017)}{.332}\, age - \underset{(.00027)}{.00263}\, age^2$$

$$n = 4.361, \quad R^2 = .569.$$

Another year of education, holding *age* fixed, results in about .091 fewer children. In other words, for a group of 100 women, if each gets another of education, they collectively are predicted to have about nine fewer children.

(ii) The reduced form for *educ* is

$$educ = \pi_0 + \pi_1 age + \pi_2 age^2 + \pi_3 frsthalf + v,$$

and we need $\pi_3 \neq 0$. When we run the regression we obtain $\hat{\pi}_3$ = –.852 and se($\hat{\pi}_3$) = .113. Therefore, women born in the first half of the year are predicted to have almost one year less education, holding *age* fixed. The *t* statistic on *frsthalf* is over 7.5 in absolute value, and so the identification condition holds.

(iii) The structural equation estimated by IV is

$$\widehat{children} = \underset{(0.548)}{-3.388} - \underset{(.0532)}{.1715}\, educ + \underset{(.018)}{.324}\, age - \underset{(.00028)}{.00267}\, age^2$$

$$n = 4.361, \quad R^2 = .550.$$

The estimated effect of education on fertility is now much larger. Naturally, the standard error for the IV estimate is also bigger, about nine times bigger. This produces a fairly wide 95% CI for β_1.

(iv) When we add *electric*, *tv*, and *bicycle* to the equation and estimate it by OLS we obtain

$$\widehat{children} = \underset{(.0240)}{-4.390} - \underset{(.0064)}{.0767}\, educ + \underset{(.016)}{.340}\, age - \underset{(.00027)}{.00271}\, age^2 - \underset{(.076)}{.303}\, electric$$

$$- \underset{(.091)}{.253}\, tv + \underset{(.049)}{.318}\, bicycle$$

$$n = 4{,}356, \quad R^2 = .576.$$

The 2SLS (or IV) estimates are

$$\widehat{children} = \underset{(0.645)}{-3.591} - \underset{(.0655)}{.1640}\, educ + \underset{(.019)}{.328}\, age - \underset{(.00028)}{.00272}\, age^2 - \underset{(.166)}{.107}\, electric$$

$$- \underset{(.2092)}{.0026}\, tv + \underset{(.052)}{.332}\, bicycle$$

$$n = 4{,}356, \quad R^2 = .558.$$

Adding *electric*, *tv*, and *bicycle* to the model reduces the estimated effect of *educ* in both cases, but not by too much. In the equation estimated by OLS, the coefficient on *tv* implies that, other factors fixed, four families that own a television will have about one fewer child than four families without a TV. Television ownership can be a proxy for different things, including income and perhaps geographic location. A causal interpretation is that TV provides an alternative form of recreation.

Interestingly, the effect of TV ownership is practically and statistically insignificant in the equation estimated by IV (even though we are not using an IV for *tv*). The coefficient on *electric* is also greatly reduced in magnitude in the IV estimation. The substantial drops in the magnitudes of these coefficients suggest that a linear model might not be the best functional form, which would not be surprising since *children* is a count variable. (See Section 17.4.)

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.

C15.4 (i) The equation estimated by OLS, omitting the first observation, is

$$\widehat{i3}_t = \underset{(0.47)}{2.37} + \underset{(.091)}{.692}\, inf_t$$

$$n = 48, \quad R^2 = .555.$$

(ii) The IV estimates, where inf_{t-1} is an instrument for inf_t, are

$$\widehat{i3}_t = 1.50 + .907\, inf_t$$
$$\quad\ (0.65) \quad (.143)$$
$$n = 48, \quad R^2 = .501.$$

The estimate on inf_t is no longer statistically different from one. (If $\beta_1 = 1$, then one percentage point increase in inflation leads to a one percentage point increase in the three-month T-bill rate.)

(iii) In first differences, the equation estimated by OLS is

$$\Delta\widehat{i3}_t = .105 + .211\, \Delta inf_t$$
$$\quad\ (.186) \quad (.073)$$
$$n = 48, \quad R^2 = .154.$$

This is a much lower estimate than in part (i) or part (ii).

(iv) If we regress Δinf_t on $\Delta inf_{t\text{-}1}$ we obtain

$$\Delta\widehat{inf}_t = .088 + .0096\, \Delta inf_{t\text{-}1}$$
$$\quad\ (.325) \quad (.1266)$$
$$n = 47, \quad R^2 = .0001.$$

Therefore, Δinf_t and $\Delta inf_{t\text{-}1}$ are virtually uncorrelated, which means that $\Delta inf_{t\text{-}1}$ cannot be used as an IV for Δinf_t.

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) ≈ 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) ≈ .55, so the overidentifying restriction is not rejected.

C15.6 (i) Sixteen states executed at least one prisoner in 1991, 1992, or 1993. (That is, for 1993, *exec* is greater than zero for 16 observations.) Texas had by far the most executions with 34.

(ii) The results of the pooled OLS regression are

$$\widehat{mrdrte} = \underset{(4.43)}{-5.28} - \underset{(2.14)}{2.07}\, d93 + \underset{(.263)}{.128}\, exec + \underset{(0.78)}{2.53}\, unem$$

$$n = 102, \quad R^2 = .102, \quad \bar{R}^2 = .074.$$

The positive coefficient on *exec* is no evidence of a deterrent effect. Statistically, the coefficient is not different from zero. The coefficient on *unem* implies that higher unemployment rates are associated with higher murder rates.

(iii) When we difference (and use only the changes from 1990 to 1993), we obtain

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

$$n = 51, \quad R^2 = .110, \quad \bar{R}^2 = .073.$$

The coefficient on $\Delta exec$ is negative and statistically significant (p-value $\approx$ .02 against a two-sided alternative), suggesting a deterrent effect. One more execution reduces the murder rate by about .1, so 10 more executions reduce the murder rate by one (which means one murder per 100,000 people). The unemployment rate variable is no longer significant.

(iv) The regression $\Delta exec$ on $\Delta exec_{-1}$ yields

$$\Delta\widehat{exec} = \underset{(.370)}{.350} - \underset{(0.17)}{1.08}\, \Delta exec_{-1}$$

$$n = 51, \quad R^2 = .456, \quad \bar{R}^2 = .444,$$

which shows a strong negative correlation in the change in executions. This means that, apparently, states follow policies whereby if executions were high in the preceding three-year period, they are lower, one-for-one, in the next three-year period.

Technically, to test the identification condition, we should add $\Delta unem$ to the regression. But its coefficient is small and statistically very insignificant, and adding it does not change the outcome at all.

(v) When the differenced equation is estimated using $\Delta exec_{-1}$ as an IV for $\Delta exec$, we obtain

$$\Delta\widehat{mrdrte} = \underset{(.211)}{.411} - \underset{(.064)}{.100}\, \Delta exec - \underset{(.159)}{.067}\, \Delta unem$$

$$n = 51, \quad R^2 = .110, \quad \bar{R}^2 = .073.$$

This is very similar to when we estimate the differenced equation by OLS. Not surprisingly, the most important change is that the standard error on $\hat{\beta}_1$ is now larger and reduces the statistically significance of $\hat{\beta}_1$.

[Instructor's Note: As an illustration of how important a single observation can be, you might want the students to redo this exercise dropping Texas, which accounts for a large fraction of executions; see also Computer Exercise C14.7. The results are not nearly as significant. Does this mean Texas is an "outlier"? Not necessarily, especially given that we have differenced to remove the state effect. But we reduce the variation in the explanatory variable, $\Delta exec$, substantially by dropping Texas.]

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

$$\widehat{unem}_t = \underset{(0.58)}{1.57} + \underset{(.097)}{.732}\, unem_{t-1}$$

$$n = 48, \quad R^2 = .554.$$

Therefore, there is a strong, positive correlation between $unem_t$ and $unem_{t-1}$.

(iv) The expectations-augmented Phillips curve estimated by IV is

$$\Delta\widehat{inf}_t = \underset{(1.883)}{.694} - \underset{(.319)}{.138}\, unem_t$$

$$n = 48, \quad R^2 = .048.$$

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)].

C15.8 (i) The OLS results are

$$\widehat{pira} = -.198 + .054\, p401k + .0087\, inc - .000023\, inc^2 - .0016\, age + .00012\, age^2$$

$$\begin{array}{llllll} (.069) & (.010) & (.0005) & (.000004) & (.0033) & (.00004) \end{array}$$

$n = 9{,}275,\ R^2 = .180$

The coefficient on *p401k* implies that participation in a 401(k) plan is associate with a .054 higher probability of having an individual retirement account, holding income and age fixed.

(ii) While the regression in part (i) controls for income and age, it does not account for the fact that different people have different taste for savings, even within given income and age categories. People that tend to be savers will tend to have both a 401(k) plan as well as an IRA. (This means that the error term, u, is positively correlated with *p401k*.) What we would like to know is, for a given person, if that person participates in a 401(k) does it make it less likely or more likely that the person also has an IRA. This ceteris paribus question is difficult to answer by OLS without many more controls for the taste for saving.

(iii) First, we need $e401k$ to be partially correlated with $p401k$; not surprisingly, this is not an issue, as being eligible for a 401(k) plan is, by definition, necessary for participation. (The regression in part (iv) verifies that they are strongly positively correlated.) The more difficult issue is whether $e401k$ can be taken as exogenous in the structural model. In other words, is being *eligible* for a 401(k) correlated with unobserved taste for saving? If we think workers that like to save for retirement will match up with employers that provide vehicles for retirement saving, then u and $e401k$ would be positively correlated. Certainly we think that $e401k$ is less correlated with u than is $p401k$. But remember, this alone is not enough to ensure that the IV estimator has less asymptotic bias than the OLS estimator; see page 519.

(iv) The reduced form equation, estimated by OLS but with heteroskedasticity-robust standard errors, is

$$\begin{array}{l} \widehat{p401k} = \underset{(.046)}{.059} + \underset{(.008)}{.689}\, e401k + \underset{(.0003)}{.0011}\, inc - \underset{(.0000027)}{.0000018}\, inc^2 - \underset{(.0022)}{.0047}\, age + \underset{(.000026)}{.000052}\, age^2 \\ n = 9{,}275,\ R^2 = .596 \end{array}$$

The t statistic on $e401k$ is over 85, and its coefficient estimate implies that, holding income and age fixed, eligibility in a 401(k) plan increases the probability of participation in a 401(k) by .69. Clearly, $e401k$ passes one of the two requirements as an IV for $p401k$.

(v) When $e401k$ is used as an IV for $p401k$ we get the following, with heteroskedasticity-robust standard errors:

$$\begin{array}{l} \widehat{pira} = \underset{(.065)}{-.207} + \underset{(.013)}{.021}\, p401k + \underset{(.0005)}{.0090}\, inc - \underset{(.000004)}{.000024}\, inc^2 - \underset{(.0032)}{.0011}\, age + \underset{(.00004)}{.00011}\, age^2 \\ n = 9{,}275,\ R^2 = .180 \end{array}$$

The IV estimate of β_{p401k} is less than half as large as the OLS estimate, and the IV estimate has a t statistic roughly equal to 1.62. The reduction in $\hat{\beta}_{p401k}$ is what we expect given the unobserved taste for saving argument made in part (ii). But we still do not estimate a tradeoff between participating in a 401(k) plan and participating in an IRA. This conclusion has prompted some in the literature to claim that 401(k) saving is additional saving; it does not simply crowd out saving in other plans.

(vi) After obtaining the reduced form residuals from part (iv), say $\hat{v}_i$, we add these to the structural equation and run OLS. The coefficient on $\hat{v}_i$ is .075 with a heteroskedasticity-robust t = 3.92. Therefore, there is strong evidence that *p401k* is endogenous in the structural equation (assuming, of course, that the IV, *e*401*k*, is exogenous).

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.

C15.10 (i) The simple regression gives

$$\widehat{\log(wage)} = \underset{(.09)}{1.09} + \underset{(.007)}{.101}\, educ$$

$n = 1{,}230$, $R^2 = .162$

Given the above estimates, the 95% confidence interval for the return to education is roughly 8.7% to 11.5%.

(ii) The simple regression of *educ* on *ctuit* gives

$$\widehat{educ} = \underset{(.07)}{13.04} - \underset{(.084)}{.049}\, ctuit$$

$$n = 1{,}230, \;\; R^2 = .0003$$

While the correlation between *educ* and *ctuit* has the expected negative sign, the *t* statistic is only about –.59, and this is not nearly large enough to conclude that these variables are correlated. This means that, even if *ctuit* is exogenous in the simple wage equation, we cannot use it as an IV for *educ*.

(iii) The multiple regression equation, estimated by OLS, is

$$\widehat{\log(wage)} = \underset{(.241)}{-.507} + \underset{(.009)}{.137}\, educ + \underset{(.027)}{.112}\, exper - \underset{(.0012)}{.0030}\, exper^2 - \underset{(.086)}{.017}\, ne - \underset{(.071)}{.017}\, nc$$

$$+ \underset{(.081)}{.018}\, west + \underset{(.087)}{.156}\, ne18 + \underset{(.073)}{.011}\, nc18 - \underset{(.086)}{.030}\, west18 + \underset{(.042)}{.205}\, urban + \underset{(.049)}{.126}\, urban18$$

$n = 1{,}230$, $R^2 = .219$

The estimated return to a year of schooling is now higher, 13.7%.

(iv) In the multiple regression of *educ* on *ctuit* and the other explanatory variables in part (iii), the coefficient on *ctuit* is –.165, *t* statistic = –2.77. So an increase of $1,000 in tuition reduces years of education by about .165 (since the tuition variables are measured in thousands).

(v) Now we estimate the multiple regression model by IV, using *ctuit* as an IV for *educ*. The IV estimate of β_{educ} is .250 (se = .122). While the point estimate seems large, the 95% confidence interval is very wide: about 1.1% to 48.9%. Other than rejecting the value zero for β_{educ}, this confidence is too wide to be useful.

(vi) The very large standard error of the IV estimate in part (v) shows that the IV analysis is not very useful. This is as it should be, as *ctuit* is not especially convincing as an IV. While it is significant in the reduced form for *educ* with other controls, the fact that it was insignificant in part (ii) is troubling. If we changed the set of explanatory variables slightly, would *educ* and *ctuit* cease to be partially correlated?

CHAPTER 16

TEACHING NOTES

I spend some time in Section 16.1 trying to distinguish between good and inappropriate uses of SEMs. Naturally, this is partly determined by my taste, and many applications fall into a gray area. But students who are going to learn about SEMS should know that just because two (or more) variables are jointly determined does not mean that it is appropriate to specify and estimate an SEM. I have seen many bad applications of SEMs where no equation in the system can stand on its own with an interesting ceteris paribus interpretation. In most cases, the researcher either wanted to estimate a tradeoff between two variables, controlling for other factors – in which case OLS is appropriate – or should have been estimating what is (often pejoratively) called the "reduced form."

The identification of a two-equation SEM in Section 16.3 is fairly standard except that I emphasize that identification is a feature of the population. (The early work on SEMs also had this emphasis.) Given the treatment of 2SLS in Chapter 15, the rank condition is easy to state (and test).

Romer's (1993) inflation and openness example is a nice example of using aggregate cross-sectional data. Purists may not like the labor supply example, but it has become common to view labor supply as being a two-tier decision. While there are different ways to model the two tiers, specifying a standard labor supply function conditional on working is not outside the realm of reasonable models.

Section 16.5 begins by expressing doubts of the usefulness of SEMs for aggregate models such as those that are specified based on standard macroeconomic models. Such models raise all kinds of thorny issues; these are ignored in virtually all texts, where such models are still used to illustrate SEM applications.

SEMs with panel data, which are covered in Section 16.6, are not covered in any other introductory text. Presumably, if you are teaching this material, it is to more advanced students in a second semester, perhaps even in a more applied course. Once students have seen first differencing or the within transformation, along with IV methods, they will find specifying and estimating models of the sort contained in Example 16.8 straightforward. Levitt's example concerning prison populations is especially convincing because his instruments seem to be truly exogenous.

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.

16.2 Using simple economics, the first equation must be the demand function, as it depends on income, which is a common determinant of demand. The second equation contains a variable, *rainfall*, that affects crop production and therefore corn supply.

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.

16.4 We can easily see that the rank condition for identifying the second equation does not hold: there are no exogenous variables appearing in the first equation that are not also in the second equation. The first equation is identified provided $\gamma_3 \neq 0$ (and we would presume $\gamma_3 < 0$). This gives us an exogenous variable, log(*price*), that can be used as an IV for *alcohol* in estimating the first equation by 2SLS (which is just standard IV in this case).

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.

16.6 (i) It could be that the decision to unionize certain segments of workers is related to how a firm treats its employees. While the timing may not be contemporaneous, with the snapshot of a single cross section we might as well assume that it is.

(ii) One possibility is to collect information on whether workers' parents belonged to a union, and construct a variable that is the percentage of workers who had a parent in a union (say, *perpar*). This may be (partially) correlated with the percent of workers that belong to a union.

(iii) We would have to assume that *percpar* is exogenous in the pension equation. We can test whether *perunion* is partially correlated with *perpar* by estimating the reduced form for *perunion* and doing a t test on *perpar*.

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, $SEASPERC_{t-1}$ 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 $SEASPERC_{t-1}$ 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.)

16.8 We must first eliminate the unobserved effect, a_{i1}. If we difference, we have

$$\Delta 1HPRICE_{it} = \delta_t + \beta_1 \Delta lEXPEND_{\text{it}} + \beta_2 \Delta 1POLICE_{it} + \beta_3 \Delta 1MEDINC_{it}$$

$$+ \beta_4 \Delta PROPTAX_{it} + \Delta u_{it},$$

for $t = 2,3$. The δ_t here denotes different intercepts in the two years. The key assumption is that the change in the (log of) the state allocation, $\Delta 1STATEALL_{it}$, is exogenous in this equation. Naturally, $\Delta 1STATEALL_{it}$ is (partially) correlated with $\Delta 1EXPEND_{it}$ because local expenditures depend at least partly on the state subsidy. The policy change in 1994 means that there should be significant variation in $\Delta 1STATEALL_{it}$, at least for the 1994 to 1996 change. Therefore, we can estimate this equation by pooled 2SLS, using $\Delta 1STATEALL_{it}$ as an IV for $\Delta 1EXPEND_{it}$; of course, this assumes the other explanatory variables in the equation are exogenous. (We could certainly question the exogeneity of the policy and property tax variables.) Without a policy change, $\Delta 1STATEALL_{it}$ would probably not vary sufficiently across i or t.

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 $\gamma_5) = 0$. 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

$$\widehat{\log(income)} = \underset{(0.17)}{7.80} + \underset{(.0017)}{.0017}\, cigs + \underset{(.008)}{.060}\, educ + \underset{(.008)}{.058}\, age - \underset{(.00008)}{.00063}\, age^2$$

$$n = 807, \quad R^2 = .165.$$

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.

C16.2 (i) We estimate a constant elasticity version of the labor supply equation (naturally, only for *hours* > 0), again by 2SLS. We get

$$\widehat{\log(hours)} = \underset{(0.69)}{8.37} + \underset{(0.56)}{1.99} \log(wage) - \underset{(.071)}{.235}\, educ - \underset{(.011)}{.014}\, age$$
$$- \underset{(.219)}{.465}\, kidslt6 - \underset{(.008)}{.014}\, nwifeinc$$
$$n = 428,$$

which implies a labor supply elasticity of 1.99. This is even higher than the 1.26 we obtained from equation (16.24) at the mean value of hours (1303).

(ii) Now we estimate the equation by 2SLS but allow log(*wage*) and *educ* to both be endogenous. The full list of instrumental variables is *age*, *kidslt6*, *nwifeinc*, *exper*, $exper^2$, *motheduc*, and *fatheduc*. The result is

$$\widehat{\log(hours)} = \underset{(1.02)}{7.26} + \underset{(0.50)}{1.81}\log(wage) - \underset{(.087)}{.129}\,educ - \underset{(.011)}{.012}\,age$$
$$- \underset{(.211)}{.543}\,kidslt6 - \underset{(.009)}{.019}\,nwifeinc$$
$$n = 428.$$

The biggest effect is to reduce the size of the coefficient on *educ* as well as its statistical significance. The labor supply elasticity is only moderately smaller.

(iii) After obtaining the 2SLS residuals, $\hat{u}_1$, from the estimation in part (ii), we regress these on *age*, *kidslt6*, *nwifeinc*, *exper*, exper^2, *motheduc*, and *fatheduc*. The *n-R*-squared statistic is 408(.0010) = .428. We have two overidentifying restrictions, so the *p*-value is roughly P(χ^2_2 > .43) $\approx$ .81. There is no evidence against the exogeneity of the IVs.

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

$$\widehat{open} = \underset{(10.47)}{129.22} - \underset{(0.98)}{8.40}\log(land) + \underset{(.0000031)}{.0000043}\,land$$

$$n = 114,\ \ R^2 = .457.$$

While log(*land*) is very significant, *land* is not, so we might as well use only log(*land*) as the IV for open.

[Instructor's Note: You might ask students whether it is better to use log(*land*) as the single IV for *open* or to use both *land* and $land^2$. In fact, log(*land*) explains much more variation in *open*.]

(iii) When we add *oil* to the original model, and assume *oil* is exogenous, the IV estimates are

$$\widehat{inf} = \underset{(16.04)}{24.01} - \underset{(.144)}{.337}\,open + \underset{(2.12)}{.803}\log(pcinc) - \underset{(9.80)}{6.56}\,oil$$

$$n = 114,\ \ R^2 = .035.$$

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.

C16.4 (i) The usual form of the test assumes no serial correlation under H_0, and this appears to be the case. We also assume homoskedasticity. After estimating (16.35), we obtain the 2SLS residuals, $\hat{u}_t$. We then run the regression $\hat{u}_t$ on gc_{t-1}, gy_{t-1}, and $r3_{t-1}$. The *n-R*-squared statistic is 35(.0613) $\approx$ 2.15. With one *df* the (asymptotic) *p*-value is P(χ_1^2 > 2.15) $\approx$ .143, and so the instruments pass the overidentification test at the 10% level.

(ii) If we estimate (16.35) but with gc_{t-2}, gy_{t-2}, and $r3_{t-2}$ as the IVs, we obtain, with $n = 34$,

$$\widehat{gc_t} = \underset{(.0274)}{-.0054} + \underset{(1.272)}{1.204}\,gy_t - \underset{(.00196)}{.00043}\,r3_t.$$

The coefficient on gy_t has doubled in size compared with equation (16.35), but it is not statistically significant. The coefficient on $r3_t$ is still small and statistically insignificant.

(iii) If we regress gy_t on gc_{t-2}, gy_{t-2}, and $r3_{t-2}$ we obtain

$$\widehat{gy_t} = \underset{(.007)}{.021} - \underset{(.469)}{.070}\,gc_{t-2} + \underset{(.330)}{.094}\,gy_{t-2} + \underset{(.00166)}{.00074}\,r3_{t-2}$$

$$n = 34,\ \ R^2 = .0137.$$

The F statistic for joint significance of all explanatory variables yields p-value $\approx$ .94, and so there is no correlation between gy_t and the proposed IVs, gc_{t-2}, gy_{t-2}, and $r3_{t-2}$. Therefore, we never should have done the IV estimation in part (ii) in the first place.

[Instructor's Note: There may be serial correlation in this regression, in which case the F statistic is not valid. But the point remains that gy_t is not at all correlated with two lags of all variables.]

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.

C16.6 (i) If we estimate the inverse supply function by OLS we obtain (with the coefficients on the monthly dummies suppressed)

$$\begin{aligned}\widehat{gprc_t} &= \underset{(.0032)}{.0144} - \underset{(.0091)}{.0443}\, gcem_t + \underset{(.0153)}{.0628}\, gprcpet_t + \ldots \\ n &= 298,\ \ R^2 = .386.\end{aligned}$$

Several of the monthly dummy variables are very statistically significant, but their coefficients are not of direct interest here. The estimated supply curve slopes *down*, not up, and the coefficient on $gcem_t$ is very statistically significant (t statistic $\approx$ –4.87).

(ii) We need $gdefs_t$ to have a nonzero coefficient in the reduced form for $gcem_t$. More precisely, if we write

$$gcem_t = \pi_0 + \pi_1 gdefs_t + \pi_2 gprcpet_t + \pi_3 feb_t + \ldots + \pi_{13} dec_t + v_t,$$

then identification requires $\pi_1 \neq 0$. When we run this regression, $\hat{\pi}_1$ = –1.054 with a t statistic of about –0.294. Therefore, we cannot reject H_0: $\pi_1 = 0$ at any reasonable significance level, and we conclude that $gdefs_t$ is not a useful IV for $gcem_t$ (even if $grdefs_t$ is exogenous in the supply equation).

(iii) Now the reduced form for *gcem* is

$$gcem_t = \pi_0 + \pi_1 gres_t + \pi_2 gnon_t + \pi_3 gprcpet_t + \pi_4 feb_t + \ldots + \pi_{14} dec_t + v_t,$$

and we need at least one of π_1 and π_2 to be different from zero. In fact, $\hat{\pi}_1$ = .136, $t(\hat{\pi}_1)$ = .984 and $\hat{\pi}_2$ = 1.15, $t(\hat{\pi}_2)$ = 5.47. So $gnon_t$ is very significant in the reduced form for $gcem_t$, and we can proceed with IV estimation.

(iv) We use both $gres_t$ and $gnon_t$ as IVs for $gcem_t$ and apply 2SLS, even though the former is not significant in the RF. The estimated labor supply function (with seasonal dummy coefficients suppressed) is now

$$\begin{aligned} \widehat{gprc_t} &= \underset{(.0073)}{.0228} - \underset{(.0277)}{.0106}\, gcem_t + \underset{(.0157)}{.0605}\, gprcpet_t + \ldots \\ n &= 298,\ \ R^2 = .356. \end{aligned}$$

While the coefficient on $gcem_t$ is still negative, it is only about one-fourth the size of the OLS coefficient, and it is now very insignificant. At this point we would conclude that the static supply function is horizontal (with *gprc* on the vertical axis, as usual). Shea (1993) adds many lags of $gcem_t$ and estimates a finite distributed lag model by IV, using leads as well as lags of $gres_t$ and $gnon_t$ as IVs. He estimates a positive long run propensity.

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.

C16.8 (i) To estimate the demand equations, we need at least one exogenous variable that appears in the supply equation.

(ii) For $wave2_t$ and $wave3_t$ to be valid IVs for log($avgprc_t$), we need two assumptions. The first is that these can be properly excluded from the demand equation. This may not be entirely reasonable, and wave heights are determined partly by weather, and demand at a local fish market could depend on demand. The second assumption is that at least one of $wave2_t$ and $wave3_t$ appears in the supply equation. There is indirect evidence of this in part three, as the two variables are jointly significant in the reduced form for log($avgprc_t$).

(iii) The OLS estimates of the reduced form are

$$\widehat{\log(avgprc_t)} = \underset{(.14)}{-1.02} - \underset{(.114)}{.012}\, mon_t - \underset{(.1119)}{.0090}\, tues_t + \underset{(.112)}{.051}\, wed_t + \underset{(.111)}{.124}\, thurs_t$$

$$+ \underset{(.021)}{.094}\, wave2_t + \underset{(.020)}{.053}\, wave3_t$$

$n = 97$, $R^2 = .304$

The variables $wave2_t$ and $wave3_t$ are jointly very significant: $F = 19.1$, p-value = zero to four decimal places.

(iv) The 2SLS estimates of the demand function are

$$\widehat{\log(totqty_t)} = \underset{(.18)}{8.16} - \underset{(.327)}{.816}\log(avgprc_t) - \underset{(.229)}{.307}\, mon_t - \underset{(.226)}{.685}\, tues_t$$

$$- \underset{(.224)}{.521}\, wed_t + \underset{(.225)}{.095}\, thurs_t$$

$n = 97$, $R^2 = .193$

The 95% confidence interval for the demand elasticity is roughly –1.47 to –.17. The point estimate, –.82, seems reasonable: a 10 percent increase in price reduces quantity demanded by about 8.2%.

(v) The coefficient on $\hat{u}_{i,t-1}$ is about .294 (se = .103), so there is strong evidence of positive serial correlation, although the estimate of ρ is not huge. One could compute a Newey-West standard error for 2SLS in place of the usual standard error.

(vi) To estimate the supply elasticity, we would have to assume that the day-of-the-week dummies do not appear in the supply equation, but they do appear in the demand equation. Part (iii) provides evidence that there are day-of-the-week effects in the demand function. But we cannot know about the supply function.

(vii) Unfortunately, in the estimation of the reduced form for log($avgprc_t$) in part (iii), the variables *mon*, *tues*, *wed*, and *thurs* are jointly insignificant [$F(4,90) = .53$, p-value = .71.] This means that, while some of these dummies seem to show up in the demand equation, things cancel out in a way that they do not affect equilibrium price, once *wave2* and *wave3* are in the equation. So, without more information, we have no hope of estimating the supply equation.

[Instructor's Note: You could have the students try part (vii), anyway, to see what happens. Also, you could have them estimate the demand function by OLS, and compare the estimates with the 2SLS estimates in part (iv). You could also have them compute the test of the single overidentification condition.]

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

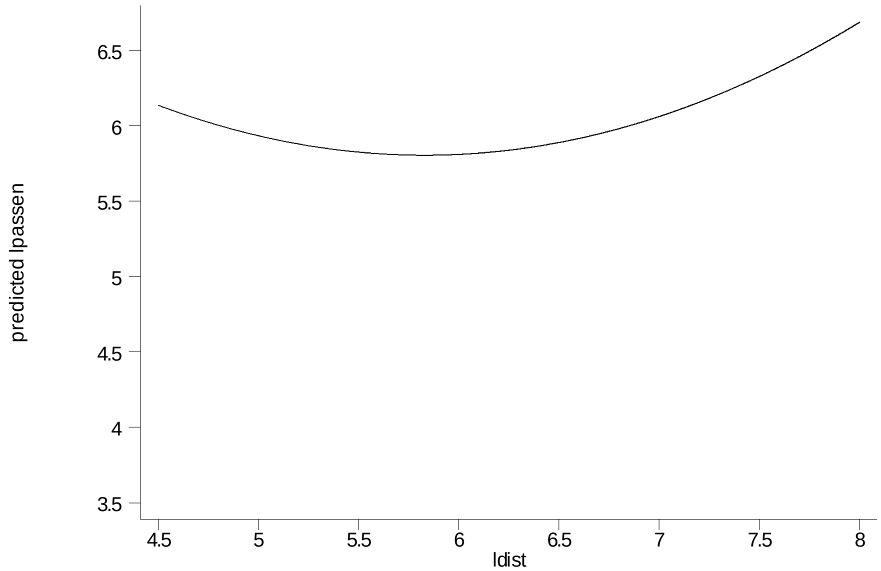
$$\widehat{\log(fare)} = \underset{(0.89)}{6.19} + \underset{(.063)}{.395}\, concen - \underset{(.272)}{.936}\, \log(dist) + \underset{(.021)}{.108}\, [\log(dist)]^2$$

$$n = 1{,}149, \quad R^2 = .408$$

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 ≈ 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.

C16.10 (i) The FE estimate of the elasticity is −1.155 with standard error .023, and so the estimate is economically large and very statistically significant.

(ii) The FE estimates of the reduced form are

$$\widehat{\log(fare)} = \underset{(.004)}{.023}\, y98 + \underset{(.004)}{.036}\, y99 + \underset{(.004)}{.098}\, y00 + \underset{(.029)}{.169}\, concen$$

$N = 1{,}149,\ T = 4,\ R^2 = .451,$

which shows a positive relationship between log(*fare*) and *concen*. In order to use *concen* as an IV for log(*fare*) in the passenger demand equation, we need $\pi_{21} \neq 0$. The null hypothesis H_0: $\pi_{21} = 0$ is strongly rejected with t = 5.74. Thus, we conclude that *concen* is a legitimate IV candidate for log(*fare*). [Of course, we are taking on faith that *concen* is uncorrelated with the idiosyncratic errors u_{it1} in the demand equation.]

(iii) Use fixed effects IV, the estimated elasticity is −.302, which is much smaller in magnitude than the usual FE estimate. Plus, the t statistic for the elasticity is now only −1.09, so the estimated elasticity is not statistically different from zero.

CHAPTER 17

TEACHING NOTES

I emphasize to the students that, first and foremost, the reason we use the probit and logit models is to obtain more reasonable functional forms for the response probability. Once we move to a nonlinear model with a fully specified conditional distribution, it makes sense to use the efficient estimation procedure, maximum likelihood. It is important to spend some time on interpreting probit and logit estimates. In particular, the students should know the rules-of-thumb for comparing probit, logit, and LPM estimates. Beginners sometimes mistakenly think that, because the probit and especially the logit estimates are much larger than the LPM estimates, the explanatory variables now have larger estimated effects on the response probabilities than in the LPM case. This may or may not be true, and can only be determined by computing partial effects for the probit and logit models. The two most common ways of compute the partial effects are the so-called "average partial effects" (APE), where the partial effects for each unit are averaged across all observations (or interesting subsets of observations), and the "partial effects at the average" (PAE). The PAEs are routinely computed by many econometrics packages, but they seem to be less useful than the APEs. The APEs have a more straightforward meaning in most cases and are more comparable to linear probability model estimates.

I view the Tobit model, when properly applied, as improving functional form for corner solution outcomes. (I believe this motivated Tobin's original work, too.) In most cases, it is wrong to view a Tobit application as a data-censoring problem (unless there is true data censoring in collecting the data or because of institutional constraints). For example, in using survey data to estimate the demand for a new product, say a safer pesticide to be used in farming, some farmers will demand zero at the going price, while some will demand positive pounds per acre. There is no data censoring here: some farmers simply find it optimal to use none of the new pesticide. The Tobit model provides more realistic functional forms for $E(y|\mathbf{x})$ and $E(y|y > 0,\mathbf{x})$ than a linear model for y. With the Tobit model, students may be tempted to compare the Tobit estimates with those from the linear model and conclude that the Tobit estimates imply larger effects for the independent variables. But, as with probit and logit, the Tobit estimates must be scaled down to be comparable with OLS estimates in a linear model. [See Equation (17.27); for examples, see Computer Exercise C17.3 and C17.12 The latter goes through calculation an average partial effect for a (roughly) continuous explanatory variable.]

Poisson regression with an exponential conditional mean is used primarily to improve over a linear functional form for $E(y|\mathbf{x})$ for count data. The parameters are easy to interpret as semi-elasticities or elasticities. If the Poisson distributional assumption is correct, we can use the Poisson distribution to compute probabilities, too. Unfortunately, overdispersion is often present in count regression models, and standard errors and likelihood ratio statistics should be adjusted to reflect overdispersion (and even underdispersion, which happens on occasion). Some reviewers of earlier editions complained about either the inclusion of this material or its location within the chapter. I think applications of count data models are on the rise: in microeconometric fields such as criminology, health economics, and industrial organization, many interesting response variables come in the form of counts. One suggestion was that

Poisson regression should not come between the Tobit model in Section 17.2 and Section 17.4, on censored and truncated regression. In fact, I put the Poisson regression model between these two topics on purpose: I hope it emphasizes that the material in Section 17.2 is purely about functional form, as is Poisson regression. Sections 17.4 and 17.5 deal with underlying linear models, but where there is a data-observability problem.

Censored regression, truncated regression, and incidental truncation are used for missing data problems. Censored and truncated data sets usually result from sample design, as in duration analysis. Incidental truncation often arises from self-selection into a certain state, such as employment or participating in a training program. It is important to emphasize to students that the underlying models are classical linear models; if not for the missing data or sample selection problem, OLS would be the efficient estimation procedure.

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).

17.2 We need to compute the estimated probability first at $hsGPA = 3.0$, $SAT = 1{,}200$, and $study = 10$ and subtract this from the estimated probability with $hsGPA = 3.0$, $SAT = 1{,}200$, and $study = 5$. To obtain the first probability, we start by computing the linear function inside $\Lambda(\cdot)$: $-1.77 + .24(3.0) + .00058(1{,}200) + .073(10) = .376$. Next, we plug this into the logit function: $\exp(.376)/[1 + \exp(.376)] \approx .593$. This is the estimated probability that a student-athlete with the given characteristics graduates in five years.

For the student-athlete who attended study hall five hours a week, we compute $-1.77 + .24(3.0) + .00058(1{,}200) + .073(5) = .011$. Evaluating the logit function at this value gives $\exp(.011)/[1 + \exp(.011)] \approx .503$. Therefore, the difference in estimated probabilities is $.593 - .503 = .090$, or just under .10. [Note how far off the calculation would be if we simply use the coefficient on *study* to conclude that the difference in probabilities is $.073(10 - 5) = .365$.]

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 $\partial E(y|y > 0,\mathbf{x})/\partial x_1 = \beta_1\{1 - \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}/\sigma + \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]\}$, $\partial E(y|y > 0,\mathbf{x})/\delta x_2 = \beta_2\{1 - \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)[\mathbf{x}\boldsymbol{\beta}/\sigma + \lambda(\mathbf{x}\boldsymbol{\beta}/\sigma)]\}$, $\partial x_1/\partial z_1 = 1$, and $\partial x_2/\partial z_1 = 2z_1$. Plugging these into the first formula and rearranging gives the answer.

17.4 Since $\log(\cdot)$ is an increasing function – that is, for positive w_1 and w_2, $w_1 > w_2$ if and only if $\log(w_1) > \log(w_2)$ – it follows that, for each i, $mvp_i > minwage_i$ if and only if $\log(mvp_i) > \log(minwage_i)$. Therefore, $\log(wage_i) = \max[\log(mvp_i), \log(minwage_i)]$.

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 $100(\beta_2 + 2\beta_3 RD)$.

17.6 (i) OLS will be unbiased, because we are choosing the sample on the basis of an exogenous explanatory variable. The population regression function for *sav* is the same as the regression function in the subpopulation with *age* > 25.

(ii) Assuming that marital status and number of children affect *sav* only through household size (*hhsize*), this is another example of exogenous sample selection. But, in the subpopulation of married people without children, *hhsize* = 2. Because there is no variation in *hhsize* in the subpopulation, we would not be able to estimate β_2; effectively, the intercept in the subpopulation becomes $\beta_0 + 2\beta_2$, and that is all we can estimate. But, assuming there is variation in *inc*, *educ*, and *age* among married people without children (and that we have a sufficiently varied sample from this subpopulation), we can still estimate β_1, β_3, and β_4.

(iii) This would be selecting the sample on the basis of the dependent variable, which causes OLS to be biased and inconsistent for estimating the β_j in the population model. We should instead use a truncated regression model.

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, where the dependent variable is a binary variable called something like *enroll*, which is one of a student enrolls at the university. The model can be estimated, by OLS or maximum likelihood, using the random sample from accepted students, and the estimators will be consistent and asymptotically normal. 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

$$\begin{aligned}\widehat{favwin} &= \underset{\substack{(.028)\\ [.032]}}{.577} + \underset{\substack{(.0023)\\ [.0019]}}{.0194}\, spread\\ n &= 553,\quad R^2 = .111.\end{aligned}$$

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^2_3 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.

C17.2 (i) The probit estimates from *approve* on *white* are given in the following table:

Dependent Variable: *approve*	
Independent Variable	Coefficient (Standard Error)
white	.784 (.087)
constant	.547 (.075)
Number of Observations	1,989
Log Likelihood Value	–700.88
Pseudo *R*-Squared	.053

As there is only one explanatory variable that takes on just two values, there are only two different predicted values: the estimated probabilities of loan approval for white and nonwhite applicants. Rounded to three decimal places these are .708 for nonwhites and .908 for whites. Without rounding errors, these are *identical* to the fitted values from the linear probability model. This must always be the case when the independent variables in a binary response model are mutually exclusive and exhaustive binary variables. Then, the predicted probabilities, whether we use the LPM, probit, or logit models, are simply the cell frequencies. (In other words, .708 is the proportion of loans approved for nonwhites and .908 is the proportion approved for whites.)

(ii) With the set of controls added, the probit estimate on *white* becomes about .520 (se $\approx$ .097). Therefore, there is still very strong evidence of discrimination against nonwhites. We can divide this by 2.5 to make it roughly comparable to the LPM estimate in part (iii) of Computer Exercise C7.8: .520/2.5 $\approx$ .208, compared with .129 in the LPM.

(iii) When we use logit instead of probit, the coefficient (standard error) on *white* becomes .938 (.173).

(iv) Recall that, to make probit and logit estimates roughly comparable, we can multiply the logit estimates by .625. The scaled logit coefficient becomes .625(.938) $\approx$ .586, which is reasonably close to the probit estimate. A better comparison would be to compare the predicted probabilities by setting the other controls at interesting values, such as their average values in the sample.

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.

[Instructor's Note: If we had just done a linear regression, we would add the coefficients on *white* and *male* to obtain the estimated difference. We get about 114.94 + 272.95 = 387.89, which is very close to the Tobit estimate. Provided that we focus on partial effects, Tobit and a linear model can give similar answers for explanatory variables near the mean values.]

(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.

C17.4 (i) The results for the Poisson regression model that includes $pcnv^2$, $ptime86^2$, and $inc86^2$ are given in the following table:

Dependent Variable: *narr86*	
Independent Variable	Coefficient (Standard Error)
pcnv	1.15 (0.28)
avgsen	−.026 (.021)
tottime	.012 (.016)
ptime86	.684 (.091)
qemp86	.023 (.033)
inc86	−.012 (.002)
black	.591 (.074)
hispan	.422 (.075)
born60	−.093 (.064)
$pcnv^2$	−1.80 (0.31)
$ptime86^2$	−.103 (.016)
$inc86^2$	.000021 (.000006)
constant	−.710 (.070)
Number of Observations	2,725
Log Likelihood Value	−2,168.87
$\hat{\sigma}$	1.179

(ii) $\hat{\sigma}^2 = (1.179)^2 \approx 1.39$, and so there is evidence of overdispersion. The maximum likelihood standard errors should be multiplied by $\hat{\sigma}$, which is about 1.179. Therefore, the MLE standard errors should be increased by about 18%.

(iii) From Table 17.3 we have the log-likelihood value for the restricted model, $\mathcal{L}_r$ = –2,248.76. The log-likelihood value for the unrestricted model is given in the above table as –2,168.87. Therefore, the usual likelihood ratio statistic is 159.78. The quasi-likelihood ratio statistic is 159.78/1.39 $\approx$ 114.95. In a χ_3^2 distribution this gives a *p*-value of essentially zero. Not surprisingly, the quadratic terms are jointly very significant.

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 $\approx$ .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.)

C17.6 The results of an OLS regression using only the uncensored durations are given in the following table:

Dependent Variable: log(*durat*)	
Independent Variable	Coefficient (Standard Error)
workprg	.092 (.083)
priors	–.048 (.014)
tserved	–.0068 (.0019)
felon	.119 (.103)
alcohol	–.218 (.097)
drugs	.018 (.089)
black	–.00085 (.08221)
married	.239 (.099)
educ	–.019 (.019)
age	.00053 (.00042)
constant	3.001 (0.244)
Number of Observations	552
R-Squared	.071

There are several important differences between the OLS estimates using the uncensored durations and the estimates from the censored regression in Table 17.4. For example, the binary indicator for drug usage, *drugs*, has become positive and insignificant, whereas it was negative (as we expect) and significant in Table 17.4. On the other hand, the work program dummy, *workprg*, becomes positive but is still insignificant. The remaining coefficients maintain the same sign, but they are all attenuated toward zero. The apparent attenuation bias of OLS for the coefficient on *black* is especially severe, where the estimate changes from –.543 in the (appropriate) censored regression estimation to –.00085 in the (inappropriate) OLS regression using only the uncensored durations.

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.

C17.8 (i) 185 out of 445 participated in the job training program. The longest time in the experiment was 24 months (obtained from the variable *mosinex*).

(ii) The F statistic for joint significance of the explanatory variables is $F(7,437) = 1.43$ with p-value = .19. Therefore, they are jointly insignificant at even the 15% level. Note that, even though we have estimated a linear probability model, the null hypothesis we are testing is that all slope coefficients are zero, and so there is no heteroskedasticity under H_0. This means that the usual F statistic is asymptotically valid.

(iii) After estimating the model $P(train = 1|\mathbf{x}) = \Phi(\beta_0 + \beta_1 unem74 + \beta_2 unem75 + \beta_3 age + \beta_4 educ + \beta_5 black + \beta_6 hisp + \beta_7 married)$ by probit maximum likelihood, the likelihood ratio test for joint significance is 10.18. In a χ^2_7 distribution this gives p-value = .18, which is very similar to that obtained for the LPM in part (ii).

(iv) Training eligibility was randomly assigned among the participants, so it is not surprising that *train* appears to be independent of other observed factors. (However, there can be a difference between eligibility and actual participation, as men can always refuse to participate if chosen.)

(v) The simple LPM results are

$$\widehat{unem78} = \underset{(.028)}{.354} - \underset{(.044)}{.111}\, train$$

$$n = 445,\ R^2 = .014$$

Participating in the job training program lowers the estimated probability of being unemployed in 1978 by .111, or 11.1 percentage points. This is a large effect: the probability of being unemployed without participation is .354, and the training program reduces it to .243. The differences is statistically significant at almost the 1% level against at two-sided alternative. (Note that this is another case where, because training was randomly assigned, we have confidence that OLS is consistently estimating a causal effect, even though the R-squared from the regression is very small. There is much about being unemployed that we are not explaining, but we can be pretty confident that this job training program was beneficial.)

(vi) The estimated probit model is

$$\widehat{\mathrm{P}(unem78 = 1 \mid train)} = \underset{(.080}{\Phi(-.375} \; - \; \underset{(.128)}{.321\, train)}$$

where standard errors are in parentheses. It does not make sense to compare the coefficient on *train* for the probit, –.321, with the LPM estimate. The probabilities have different functional forms. However, note that the probit and LPM t statistics are essentially the same (although the LPM standard errors should be made robust to heteroskedasticity).

(vii) There are only two fitted values in each case, and they are the same: .354 when *train* = 0 and .243 when *train* = 1. This has to be the case, because any method simply delivers the cell frequencies as the estimated probabilities. The LPM estimates are easier to interpret because they do not involve the transformation by $\Phi(\cdot)$, but it does not matter which is used provided the probability differences are calculated.

(viii) The fitted values are no longer identical because the model is not saturated, that is, the explanatory variables are not an exhaustive, mutually exclusive set of dummy variables. But, because the other explanatory variables are insignificant, the fitted values are highly correlated: the LPM and probit fitted values have a correlation of about .993.

(ix) To obtain the average partial effect of *train* using the probit model, we obtain fitted probabilities for each man for *train* = 1 and *train* = 0. Of course, on of these is a counterfactual, because the man was either in job training or not. Of course, we evaluate the other regressors at their actual outcomes. The APE is the average, over all observations, of the differences in the estimated probabilities. With the variables in part (ii) appearing in the probit, the estimated APE is about –.112. Interestingly, rounded to three decimal places, this is the same as the coefficient on *train* in the linear regression. In other words, the linear probability model and probit give virtually the same estimated APEs.

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.

[Instructor's Notes: This might be a good place to remind students about basic economics. You can ask them whether *reglbs* should be included as an additional explanatory variable in the demand equation for *ecolbs*, making the point that the resulting equation would no longer be a demand equation. In other words, *reglbs* and *ecolbs* are jointly determined, but it is not appropriate to write each as a function of the other. You could have the students compute heteroskedasticity-robust standard errors for the OLS estimates. Also, you could have them estimate a probit model for *ecolbs* = 0 versus *ecolbs* > 0, and have them compare the scaled Tobit slope estimates with the probit estimates.]

C17.10 (i) 497 people do not smoke at all. 101 people report smoking 20 cigarettes a day. Since one pack of cigarettes contains 20 cigarettes, it is not surprising that 20 is a focal point.

(ii) The Poisson distribution does not allow for the kinds of focal points that characterize *cigs*. If you look at the full frequency distribution, there are blips at half a pack, two packs, and so on. The probabilities in the Poisson distribution have a much smoother transition. Fortunately, the Poisson regression model has nice robustness properties.

(iii) The results of the Poisson regression are given in the following table, along with the OLS estimates of a linear model for later reference. The Poisson standard errors are the usual Poisson maximum likelihood standard errors, and the OLS standard errors are the usual (nonrobust) standard errors.

Dependent Variable: *cigs*		
Independent Variable	Poisson (Exponential Model)	OLS (Linear Model)
log(*cigpric*)	–.355 (.144)	–2.90 (5.70)
log(*income*)	.085 (.020)	.754 (.730)
white	–.0019 (.0372)	–.205 (1.458)
educ	–.060 (.004)	–.514 (.168)
age	.115 (.005)	.782 (.161)
age^2	–.00138 (.00006)	–.0091 (.0018)
constant	1.46 (.61)	5.77 (24.08)
Number of Observations	807	807
Log Likelihood Value	–8,184.03	———
$\hat{\sigma}$	4.54	13.46
R-squared	.043	.045

The estimated price elasticity is –.355 and the estimated income elasticity is .085.

(iv) If we use the maximum likelihood standard errors, the *t* statistic on log(*cigpric*) is about –2.47, which is significant at the 5% level against a two-sided alternative. The *t* statistic on log(*income*) is 4.25, which is very significant.

(v) $\hat{\sigma}^2 = 20.61$, and so $\hat{\sigma} = 4.54$. This is evidence of severe overdispersion, and means that all of the standard errors for Poisson regression should be multiplied by 4.54; the *t* statistics should be divided by 4.54.

(vi) The robust *t* statistic for log(*cigpric*) is about –.54, which makes it very insignificant. This is a good example of misleading the usual Poisson standard errors and test statistics can be. The robust *t* statistic for log(*income*) is about .94, which also makes the income elasticity statistically insignificant.

(vii) The education and age variables are still quite significant; the robust *t* statistic on *educ* over three in absolute value, and the robust *t* statistic on *age* is over five. The coefficient on *educ* implies that one more year of education reduces the expected number of cigarettes smoked by about 6.0%.

(viii) The minimum predicted value is .515 and the maximum is 18.84. The fact that we predict some smoking for anyone in the sample is a limitation with using the expected value for prediction. Further, we do not predict that anyone will smoke even one pack of cigarettes, even though more than 25% of the people in the sample report smoking a pack or more per day! This shows that smoking, especially heavy smoking, is difficult to predict based on the explanatory variables we have access to.

(ix) The squared correlation between $cigs_i$ and $\widehat{cigs}_i$ is the R-squared reported in the above table, .043.

(x) The linear model results are reported in the last column of the previous table. The R-squared is slightly higher for the linear model – but remember, the OLS estimates are chosen to maximize the R-squared, while the MLE estimates do not maximize the R-squared (as we have calculated it). In any case, both R-squareds are quite small.

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

$$\begin{aligned}\widehat{\log(wage)} = \underset{(.060)}{.649} + \underset{(.004)}{.099}\, educ + \underset{(.003)}{.020}\, exper - \underset{(.00008)}{.00035}\, exper^2 \\ - \underset{(.034)}{.030}\, black + \underset{(.036)}{.014}\, hispanic\end{aligned}$$

$n = 3{,}286,\ R^2 = .205$

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 $kidlt6$ 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.

C17.12 (i) Out of 4,248 respondents, 1,707 made a gift most recently. This is 40 percent of the sample.

(ii) All explanatory variables except *avggift* (which has a *t* statistic of 1.00) are statistically significant at the 5% level. The variable *resplast* has the smallest of the remaining *t* statistics of 2.20.

(iii) The average partial effect for *mailsyear* is

$$\hat{\beta}_{mailsyear} \cdot \left[n^{-1} \sum_{i=1}^{n} \phi(\hat{\beta}_0 + \mathbf{x}_i \hat{\boldsymbol{\beta}}) \right],$$

that is, the coefficient on *mailsyear* times the average of the standard normal density evaluated at the estimated index. This turns out to be about .046 for this application. In a linear probability model for *respond*, the coefficient on *mailsyear* is about .066, which is somewhat above the probit APE estimate.

(iv) In the Tobit estimation, all explanatory variables except *replast* are strongly statistically significant. The t statistic on *replast* is 1.24. The smallest t statistic on the other variables (in absolute value) is on *avggift*, 5.09.

(v) The APE in the Tobit case, again treating *mailsyear* as a continuous variable, is estimated as

$$\hat{\beta}_{mailsyear} \cdot \left\{ n^{-1} \sum_{i=1}^{n} \Phi[(\hat{\beta}_0 + \mathbf{x}_i \hat{\boldsymbol{\beta}}) / \hat{\sigma}] \right\}.$$

In this application, the APE is about 1.49. The linear regression coefficient is larger, about 1.99.

(vi) They are not entirely compatible with the Tobit model, although it is difficult to know whether the differences are important. Subject to sampling error, the coefficients from the probit, say $\hat{\gamma}_j$, should be roughly equal to $\hat{\beta}_j / \hat{\sigma}$, where the latter are from Tobit. Now, the signs of the probit and Tobit coefficients are all the same, but the scaled Tobit coefficient is not always close to the corresponding probit coefficient. For example, the scaled Tobit coefficient on *resplast* is about 1.67/27, or .062. The probit coefficient is .126, or about twice as large. However, the coefficient is not statistically significant in the Tobit case, and only marginally significant in the probit case. The agreement for *weekslast*, *mailsyear*, and *avggift* is much better. For example, for *mailsyear* – which is highly statistically significant in the probit and Tobit – the scaled Tobit coefficient is about .148 and the probit coefficient is .146. On *propresp*, the scaled Tobit coefficient is about 1.30, compared with 1.85 for the probit estimate. The discrepancies for *resplast* and *mailsyear* might be statistically significant, but they also might be explainable by sampling error. Investigating this issue is beyond the scope of the text.

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

TEACHING NOTES

Several of the topics in this chapter, including testing for unit roots and cointegration, are now staples of applied time series analysis. Instructors who like their course to be more time series oriented might cover this chapter after Chapter 12, if time permits. Or, the chapter can be used as a reference for ambitious students who wish to be versed in recent time series developments.

The discussion of infinite distributed lag models, and in particular geometric DL and rational DL models, gives one particular interpretation of dynamic regression models. But one must emphasize that only under fairly restrictive assumptions on the serial correlation in the error of the infinite DL model does the dynamic regression consistently estimate the parameters in the lag distribution. Computer Exercise C18.1 provides a good illustration of how the GDL model, and a simple RDL model, can be too restrictive.

Example 18.5 tests for cointegration between the general fertility rate and the value of the personal exemption. There is not much evidence of cointegration, which sheds further doubt on the regressions in levels that were used in Chapter 10. The error correction model for holding yields in Example 18.7 is likely to be of interest to students in finance. As a class project, or a term project for a student, it would be interesting to update the data to see if the error correction model is stable over time.

The forecasting section is heavily oriented towards regression methods and, in particular, autoregressive models. These can be estimated using any econometrics package, and forecasts and mean absolute errors or root mean squared errors are easy to obtain. The interest rate data sets (for example, INTQRT.RAW) can be updated to do much more recent out-of-sample forecasting exercises.

Several new time series data sets include OKUN.RAW (a small annual data set that can be used to illustrate Okun's Law), MINWAGE.RAW (U.S. monthly data, by sector, on wages, employment, and the federal minimum wage), and FEDFUND.RAW (quarterly data on the federal funds rate and the real GDP gap in the United States).

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.

18.2 (i) When we lag equation (18.68) once, multiply it by $(1 - \lambda)$, and subtract it from (18.68), we obtain

$$y_t - (1 - \lambda)y_{t-1} = \lambda\alpha_0 + \alpha_1[x_t^* - (1 - \lambda)x_{t-1}^*] + u_t - (1 - \lambda)u_{t-1}.$$

But we can rewrite (18.69) as

$$x_t^* - (1 - \lambda)x_{t-1}^* = \lambda x_{t-1};$$

when we plug this into the first equation we obtain the desired result.

(ii) If $\{u_t\}$ is serially uncorrelated, then $\{v_t = u_t - (1 - \lambda)u_{t-1}\}$ must be serially correlated. In fact, $\{v_t\}$ is an MA(1) process with $\alpha = -(1 - \lambda)$. Therefore, $\text{Cov}(v_t,v_{t-1}) = -(1 - \lambda)\sigma_u^2$, and the correlation between v_t and v_{t-h} is zero for $h > 1$.

(iii) Because $\{v_t\}$ follows an MA(1) process, it is correlated with the lagged dependent variable, y_{t-1}. Therefore, the OLS estimators of the β_j will be inconsistent (and biased, of course). Nevertheless, we can use x_{t-2} as an IV for y_{t-1} because x_{t-2} is uncorrelated with v_t (because u_t and u_{t-1} are both uncorrelated with x_{t-2}) and x_{t-2}) and x_{t-2} is partially correlated with y_{t-1}.

(iv) Because $\beta_1 = 1 - \lambda$, $\hat{\lambda} = 1 - \hat{\beta}_1$. Then, because $\beta_2 = \lambda\alpha_1$, $\hat{\alpha}_1 = \hat{\beta}_2 / \hat{\lambda} = \hat{\beta}_2 /(1 - \hat{\beta}_1)$. Because the plim passes through continuous functions, consistency of these estimators is immediate. None of the estimators is unbiased. For one, the OLS estimators of the β_j are not unbiased given that there is a lagged dependent variable. Of course, $\hat{\alpha}_1$ would not be unbiased even if $\hat{\beta}_1$ and $\hat{\beta}_2$ were.

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.

18.4 Following the hint, we show that $y_{t-2} - \beta x_{t-2}$ can be written as a linear function of $y_{t-1} - \beta x_{t-1}$, Δy_{t-1}, and Δx_{t-1}. That is,

$$y_{t-2} - \beta x_{t-2} = a_1(y_{t-1} - \beta x_{t-1}) + a_2\Delta y_{t-1} + a_3\Delta x_{t-1}$$

for constants a_1, a_2, and a_3. But

$$(y_{t-1} - \beta x_{t-1}) - \Delta y_{t-1} + \beta\Delta x_{t-1} = y_{t-1} - \beta x_{t-1} - (y_{t-1} - y_{t-2}) + \beta(x_{t-1} - x_{t-2}) = y_{t-2} - \beta x_{t-2},$$

and so $a_1 = 1$, $a_2 = -1$, and $a_3 = \beta$ work in the first equation.

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 $e_t = u_t + \beta v_t$.

18.6 (i) This is given by the estimated intercept, 1.54. Remember, this is the percentage growth at an annualized rate. It is statistically different from zero since $t = 1.54/.56 = 2.75$.

(ii) 1.54 + .031(10) = 1.85. As an aside, you could obtain the standard error of this estimate by running the regression.

$$pcip_t \text{ on } pcip_{t-1}, pcip_{t-2}, pcip_{t-3}, (pcsp_{t-1} - 10),$$

and obtaining the standard error on the intercept.

(iii) Growth in the S&P 500 index has a statistically significant effect on industrial production growth – in the Granger causality sense – because the t statistic on $pcsp_{t-1}$ is about 2.38. The economic effect is reasonably large.

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.

18.8 (i) Following the hint we have

$$\begin{aligned} y_t &= \alpha + \delta_1 z_{t-1} + u_t = \alpha + \delta_1 z_{t-1} + \rho u_{t-1} + e_t \\ &= \alpha + \delta_1 z_{t-1} + \rho(y_{t-1} - \alpha - \delta_1 z_{t-2}) + e_t \\ &= (1-\rho)\alpha + \rho y_{t-1} + \delta_1 z_{t-1} - \rho\delta_1 z_{t-2} + e_t. \end{aligned}$$

By assumption, $\mathrm{E}(e_t|I_{t-1}) = 0$, and since y_{t-1}, z_{t-1}, and z_{t-2} are all in I_{t-1}, we have

$$\mathrm{E}(y_t|I_{t-1}) = (1-\rho)\alpha + \rho y_{t-1} + \delta_1 z_{t-1} - \rho\delta_1 z_{t-2}.$$

We obtain the desired answer by adding one to the time index everywhere.

(ii) The forecasting equation for y_{n+1} is obtained by using part (i) with $t = n$, and then plugging in the estimates:

$$\hat{f}_n = (1-\hat{\rho})\hat{\alpha} + \hat{\rho}y_n + \hat{\delta}_1 z_n - \hat{\rho}\hat{\delta}_1 z_{n-1}$$

where I_{t-1} contains y and z dated at $t - 1$ and earlier.

(iii) From part (i), it follows that the model with one lag of z and AR(1) serial correlation in the errors can be obtained from

$$y_t = \alpha_0 + \rho y_{t-1} + \gamma_1 z_{t-1} + \gamma_2 z_{t-2} + e_t, \ \mathrm{E}(e_t|I_{t-1}) = 0$$

with $\alpha_0 = (1-\rho)\alpha$, $\gamma_1 = \delta_1$, and $\gamma_2 = -\rho\delta_1 = -\rho\gamma_1$. The key is that γ_2 is entirely determined (in a nonlinear way) by ρ and γ_1. So the model with a lag of z and AR(1) serial correlation is a special case of the more general model. (Note that the general model depends on four parameters, while the model from part (i) depends on only three.)

(iv) For forecasting, the AR(1) serial correlation model may be too restrictive. It may impose restrictions on the parameters that are not met. On the other hand, if the AR(1) serial correlation model holds, it captures the conditional mean $\mathrm{E}(y_t|I_{t-1})$ with one fewer parameter than the general model; in other words, the AR(1) serial correlation model is more parsimonious. [See Harvey (1990) for ways to test the restriction $\gamma_2 = -\rho\gamma_1$, which is called a *common factor restriction*.]

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}.

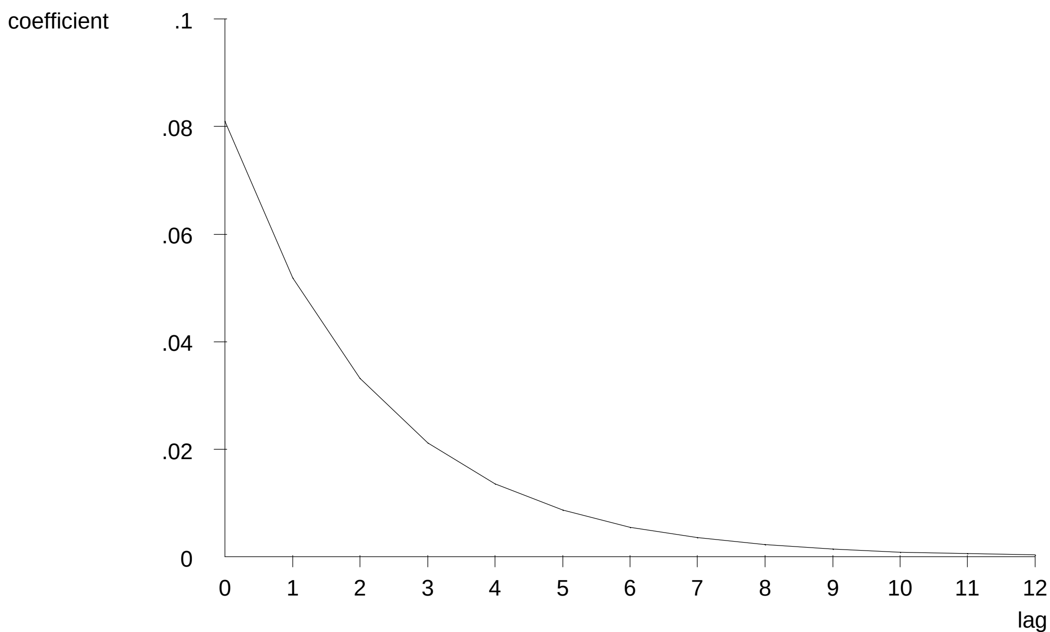
SOLUTIONS TO COMPUTER EXERCISES

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.



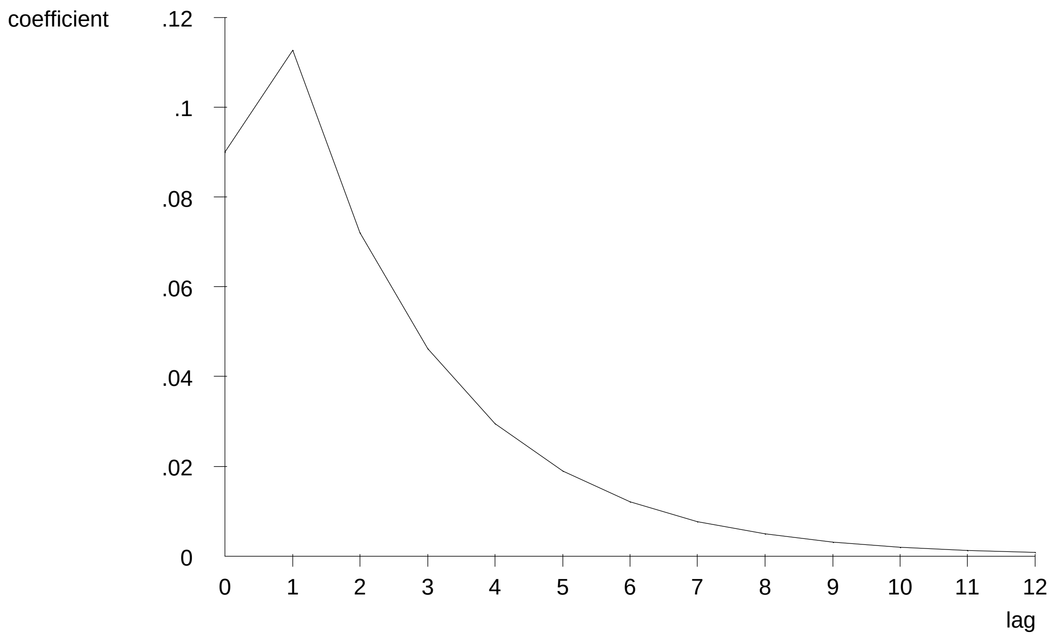
(ii) The IP for the FDL model estimated in Problem 11.5 was .119, which is substantially above the estimated IP for the GDL model. Further, the estimated LRP from GDL model is much lower than that for the FDL model, which we estimated as 1.172. Clearly we cannot think of the GDL model as a good approximation 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) ≈ .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.



C18.2 (i) We run the regression

$$\widehat{ginvpc_t} = \underset{(.170)}{-.786} - \underset{(.198)}{.956}\, \log(invpc_{t-1}) + \underset{(.0021)}{.0068}\, t + \underset{(.162)}{.532}\, ginvpc_{t-1} + \underset{(.165)}{.290}\, ginvpc_{t-2}$$

$$n = 39,\quad R^2 = .437,$$

where $ginvpc_t = \log(invpc_t) - \log(invpc_{t-1})$. The t statistic for the augmented Dickey-Fuller unit root test is $-.956/.198 \approx -4.82$, which is well below -3.96, the 1% critical value obtained from Table 18.3. Therefore, we strongly reject a unit root in $\log(invpc_t)$. (Incidentally, remember that the t statistics on the intercept and time trend in this estimated equation to not have approximate t distributions, although those on $ginvpc_{t-1}$ and $ginvpc_{t-2}$ do under the usual null hypothesis that the parameter is zero.)

(ii) When we apply the regression to $\log(price_t)$ we obtain

$$\widehat{gprice_t} = \underset{(.019)}{-.040} - \underset{(.092)}{.222} \log(price_{t-1}) + \underset{(.00049)}{.00097}\, t + \underset{(.155)}{.328}\, gprice_{t-1} + \underset{(.149)}{.130}\, gprice_{t-2}$$

$$n = 39, \quad R^2 = .200,$$

Now the Dickey-Fuller t statistic is about -2.41, which is above -3.12, the 10% critical value from Table 18.3. [The estimated root is $1 - .222 = .778$, which is much larger than for $\log(invpc_t)$.] We cannot reject the unit root null at a sufficiently small significance level.

(iii) Given the very strong evidence that $\log(invpc_t)$ does not contain a unit root, while $\log(price_t)$ may very well, it makes no sense to discuss cointegration between the two. If we take any nontrivial linear combination of an I(0) process (which may have a trend) and an I(1) process, the result will be an I(1) process (possibly with drift).

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 $\Delta i3$.

[Instructor's Note: The F test for joint significance of $\Delta i3_{t-1}$, $\Delta i3_{t-2}$, and $\Delta i3_{t-3}$ yields p-value = .228, and so $\Delta i3$ does not Granger cause $pcip$ conditional on past $pcsp$.]

C18.4 We first run the regression gfr_t on pe_t, t, and t^2, and obtain the residuals, $\hat{u}_t$. We then apply the augmented Dickey-Fuller test, with one lag of $\Delta\hat{u}_t$, by regressing $\Delta\hat{u}_t$ on $\hat{u}_{t-1}$ and $\Delta\hat{u}_{t-1}$. There are 70 observations available for this last regression, and it yields $-.165$ as the coefficient on $\hat{u}_{t-1}$ with t statistic = -2.76. This is well above -4.15, the 5% critical value [obtained from Davidson and MacKinnon (1993, Table 20.2)]. Therefore, we cannot reject the null hypothesis of no cointegration, so we conclude gfr_t and pe_t are not cointegrated even if we allow them to have different quadratic trends.

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.

[Instructor's Note: The standard errors on all slope coefficients can be used to construct t statistics with approximate t distributions, provided there is no serial correlation in $\{e_t\}$.]

(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).

C18.6 (i) The equations using data through 1997 are

$$\widehat{unem}_t = \underset{(0.572)}{1.549} + \underset{(.096)}{.734}\ unem_{t-1}$$

$$n = 49,\ \ R^2 = .554,\ \ \hat{\sigma} = 1.041$$

and

$$\widehat{unem}_t = \underset{(0.484)}{1.286} + \underset{(.083)}{.648}\ unem_{t-1} + \underset{(.041)}{.185}\ inf_{t-1}$$

$$n = 49,\ \ R^2 = .691,\ \ \hat{\sigma} = .876.$$

The parameter estimates do not change by much. This is not very surprising, as we have added only one year of data.

(ii) The forecast for $unem_{1998}$ from the first equation is 1.549 + .734(4.9) $\approx$ 5.15; from the second equation the forecast is 1.286 + .648(4.9) + .185(2.3) $\approx$ 4.89. The actual civilian unemployment rate for 1998 was 4.5. Once again the model that includes lagged inflation produces a better forecast.

(iii) There is no practical improvement in reestimating the parameters using data through 1997: 4.89 versus 4.90, which differs in a digit that is not even reported in the published unemployment series: our predicted unemployment rate would be 4.9% in both cases.

(iv) To obtain the two-step-ahead forecast we need the 1996 unemployment rate, which is 5.4. From equation (18.55), the forecast of $unem_{1998}$ made after we know $unem_{1996}$ is (1 + .732)(1.572) + ($.732^2$)(5.4) $\approx$ 5.62. The one-step ahead forecast is 1.572 + .732(4.9) $\approx$ 5.16, and so it is better to use the one-step-ahead forecast, as it is much closer to 4.5.

C18.7 (i) The estimated linear trend equation using the first 119 observations and excluding the last 12 months is

$$\widehat{chnimp}_t = \underset{(53.20)}{248.58} + \underset{(0.77)}{5.15}\ t$$

$$n = 119,\ \ R^2 = .277,\ \ \hat{\sigma} = 288.33.$$

The standard error of the regression is 288.33.

(ii) The estimated AR(1) model excluding the last 12 months is

$$\widehat{chnimp}_t = 329.18 + .416\, chnimp_{t-1}$$
$$(54.71) \qquad (.084)$$

$$n = 118, \quad R^2 = .174, \quad \hat{\sigma} = 308.17.$$

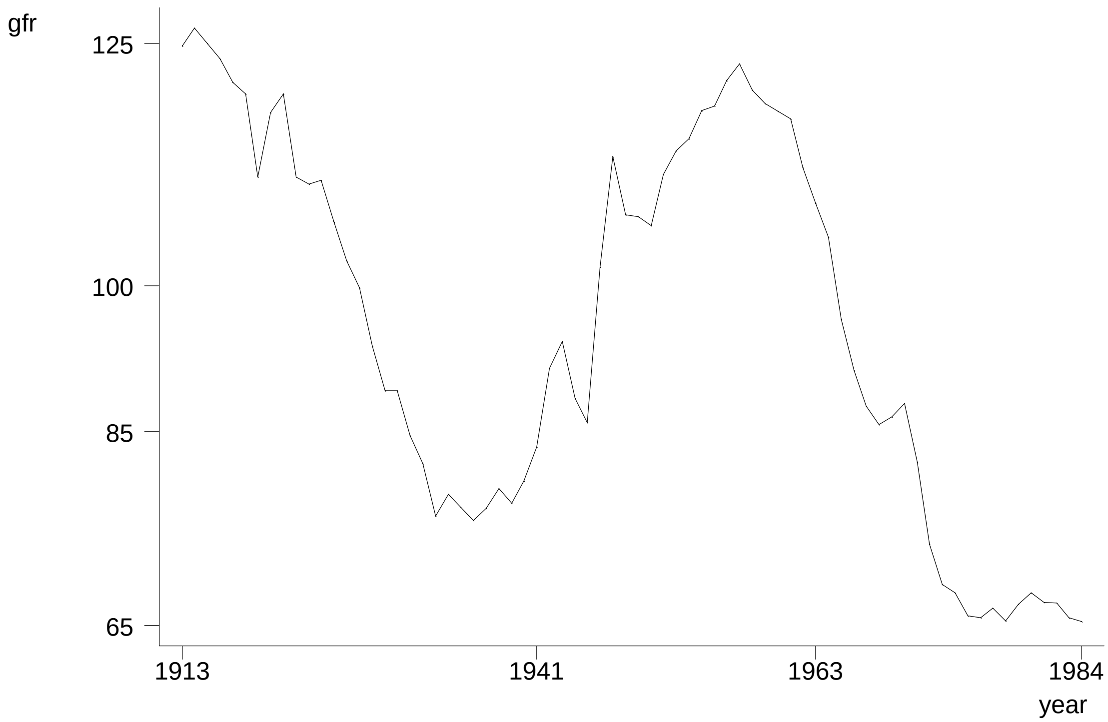
Because $\hat{\sigma}$ is lower for the linear trend model, it provides the better in-sample fit.

(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.

[Instructor's Note: In a model with a linear time trend and autoregressive term, both are statistically significant with $\hat{\sigma} = 285.03$ – a slightly better in-sample fit than the linear trend model. But, using the last 12 months for one-step-ahead forecasting, RMSE = 316.15 and MAE = 202.73. Therefore, one actually does a bit worse in using the more general model compared with the simple linear trend.]

(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*.

C18.8 (i) As can be seen from the following graph, *gfr* does not have a clear upward or downward trend. Starting from 1913, there is a sharp downward trend in fertility until the mid-1930s, when the fertility rate bottoms out. Fertility increased markedly until the end of the baby boom in the early 1960s, after which point it fell sharply and then leveled off.



(ii) The regression of gfr_t on a cubic in t, using the data up through 1979, gives

$$\widehat{gfr_t} = \underset{(5.09)}{148.71} - \underset{(0.64)}{6.90}\, t + \underset{(.022)}{.243}\, t^2 - \underset{(.0002)}{.0024}\, t^3$$

$$n = 67,\ \ R^2 = .739,\ \ \hat{\sigma} = 9.84.$$

If we use the usual t critical values, all terms are very statistically significant, and the R-squared indicates that this curve-fitting exercise tracks gfr_t pretty well, at least up through 1979.

(iii) The MAE is about 43.02.

(iv) The regression Δgfr_t on just an intercept, using data up through 1979, gives

$$\Delta\widehat{gfr_t} = \underset{(.543)}{-.871}$$

$$n = 66,\ \ \hat{\sigma} = 4.41.$$

(The R-squared is identically zero since there are no explanatory variables. But $\hat{\sigma}$, which estimates the standard deviation of the error, is comparable to that in part (ii), and we see that it is much smaller here.) The t statistic for the intercept is about –1.60, which is not significant at the 10% level against a two-sided alternative. Therefore, it is legitimate to treat gfr_t as having no drift, if it is indeed a random walk. (That is, if $gfr_t = \alpha_0 + gfr_{t-1} + e_t$, where $\{e_t\}$ is zero-mean, serially uncorrelated process, then we cannot reject H_0: $\alpha_0 = 0$.)

(v) The prediction of gfr_{n+1} is simply gfr_n, so the predication error is simply $\Delta gfr_{n+1} = gfr_{n+1} - gfr_n$. Obtaining the MAE for the five prediction errors for 1980 through 1984 gives MAE $\approx$ .840, which is much lower than the 43.02 obtained with the cubic trend model. The random walk is clearly preferred for forecasting.

(vi) The estimated AR(2) model for gfr_t is

$$\widehat{gfr_t} = \underset{(2.92)}{3.22} + \underset{(0.120)}{1.272}\, gfr_{t-1} - \underset{(.121)}{.311}\, gfr_{t-2}$$
$$n = 65,\ \ R^2 = .949,\ \ \hat{\sigma} = 4.25.$$

The second lag is significant. (Recall that its t statistic is valid even though gfr_t apparently contains a unit root: the coefficients on the two lags sum to .961.) The standard error of the regression is slightly below that of the random walk model.

(vii) The out-of-sample forecasting performance of the AR(2) model is worse than the random walk without drift: the MAE for 1980 through 1984 is about .991 for the AR(2) model.

[Instructor's Note: As a third possibility, you might have the students estimate an AR(1) model for Δgfr_t – that is, impose the unit root in the AR(2) model. The resulting MAE is about .879, so it is better to impose the unit root than to estimate the unrestricted AR(2). But it still does less well than the simple random walk without drift.]

C18.9 (i) Using the data up through 1989 gives

$$\hat{y}_t = \underset{(1,163.09)}{3,186.04} + \underset{(46.31)}{116.24}\, t + \underset{(.148)}{.630}\, y_{t-1}$$
$$n = 30,\ \ R^2 = .994,\ \ \hat{\sigma} = 223.95.$$

(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.

C18.10 (i) The AR(1) model for $\Delta r6$, estimated using all but the last 16 observations, is

$$\widehat{\Delta r6}_t = \underset{(.131)}{.047} - \underset{(.096)}{.179}\, \Delta r6_{t-1}$$

$$n = 106, \quad R^2 = .032, \quad \bar{R}^2 = .023.$$

The RMSE for forecasting one-step-ahead over the last 16 quarters is about .704.

(ii) The equation with spr_{t-1} included is

$$\widehat{\Delta r6}_t = \underset{(.195)}{.372} - \underset{(.095)}{.171}\, \Delta r6_{t-1} - \underset{(0.474)}{1.045}\, spr_{t-1}$$

$$n = 106, \quad R^2 = .076, \quad \bar{R}^2 = .058.$$

The RMSE is about .788, which is higher than the RMSE without the error correction term. Therefore, while the EC term improves the in-sample fit (and is statistically significant), it actually hampers out-of-sample forecasting.

(iii) To make the forecasting exercises comparable, we exclude the last 16 observations to estimate the cointegrating parameters. The CI coefficient is about 1.028. The estimated error correction model is

$$\widehat{\Delta r6}_t = \underset{(.195)}{.372} - \underset{(.095)}{.171}\, \Delta r6_{t-1} - \underset{(0.474)}{1.045}\, (r6_{t-1} - 1.028\, r3_{t-1})$$

$$n = 106, \quad R^2 = .058, \quad \bar{R}^2 = .040,$$

which shows that this fits worse than the EC model when the cointegrating parameter is assumed to be one. The RMSE for the last 16 quarters is .782, so this works slightly better. But both versions of the EC model are dominated by the AR(1) model for $\Delta r6_t$.

[Instructor's Note: Because $\Delta r6_{t-1}$ is only marginally significant in the AR(1) model, its coefficient is small, and the intercept is also very small and insignificant, you might have the students use zero to predict $\Delta r6$ for each of the last 16 quarters. In other words, the "model" for $r6$ is simply $\Delta r6_t = u_t$, where u_t is an unpredictable sequence with zero mean. The resulting

RMSE is about .657, which means this works best of all. The lesson is that econometric methods are not always called for, or even desirable.]

(iv) The conclusions would be identical because, as shown in Problem 18.9, the one-step-ahead errors for forecasting $r6_{n+1}$ are identical to those for forecasting $\Delta r6_{n+1}$.

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.

C18.12 (i) The F statistic for the second and third lags, with 2 and 550 degrees of freedom, gives $F = 3.76$ and p-value = .024.

(ii) When $pcsp_{t-1}$ is added to the AR(3) model in part (i), its coefficient is about .031 and its t statistic is about 2.40. Therefore, we conclude that *pcsp* does Granger cause *pcip*.

(iii) The heteroskedasticity-robust t statistic is 2.47, so the conclusion from part (ii) does not change.

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.

C18.14 (i) Using the ADF regression with one lag and a time trend, the coefficient on $lwage232_{t-1}$ is only −.0056 with $t = -1.39$. The estimated root is so close to one (.9944) that the test outcome is almost superfluous. Anyway, the 10% critical value is −3.12 (from Table 18.3), and the t statistic is well above that.

For the employment series, the coefficient on $lemp232_{t-1}$ in the ADF regression is actually slightly positive, .0008. For practical purposes, the root is estimated to be one.

With so little evidence against a unit root – including estimated roots that are virtually one – the only sensible approach to analyzing these series is to treat them as having unit roots.

(ii) Without a time trend, the regression $\Delta\hat{u}_t$ on $\hat{u}_{t-1}, \Delta\hat{u}_{t-1}, \Delta\hat{u}_{t-2}$ (where the $\hat{u}_t$ are the residuals from the cointegrating regression) gives a coefficient on $\hat{u}_{t-1}$ of .00008, and so the estimated root is one. When time is added to the initial regression, the coefficient on $\hat{u}_{t-1}$ becomes .0022; again, there is no evidence for cointegration. (This is one of those somewhat unusual cases where, even after obtaining the residuals from an OLS regression, the estimated root is actually greater than one.)

(iii) If we use the real wage, as defined in the problem, and include a time trend, the coefficient on $\hat{u}_{t-1}$ is −.044 with $t = -3.09$. Now, at least the estimated root is less than one (about .956), but not much less. The 10% critical value for the E-G test with a time trend is −3.50, and so we cannot reject the null that $lemp232_t$ and $lrwage232_t$ are not cointegrated. There is more evidence of cointegration when we use the real wage, but such a large root means that, if the two series are returning to an equilibrium (around a trend), it happens very slowly.

(iv) Generally, productivity is omitted from the cointegrating relationship. Economic theory tells us the demand for labor depends on worker productivity in addition to real wages.

Productivity is often modeled as following a random walk (with positive drift), and just including a linear time trend may not sufficiently capture changes in productivity. Factors such as other sources of income – that is, nonwage income – can affect the supply of labor, and other income (or its log) likely has a unit root.

CHAPTER 19

TEACHING NOTES

Students should read this chapter if you have assigned them a term paper. I used to allow students to choose their own topics, but this is difficult in a first-semester course, and places a heavy burden on instructors or teaching assistants, or both. A more manageable possibility is to assign a common topic and provide a data set with several weeks left in the term. When I have done this, I use a cross-sectional data set (because I teach time series at the end of the course), and I provide guidelines of the kinds of questions students should try to answer. For example, I might ask students to answer the following questions: Is there a marriage premium for NBA basketball players? If so, does it depend on race? Can the premium, if it exists, be explained by productivity differences? Or, using a data set on student performance, school inputs, and family background (which might require merging school-level data with census data), I might ask the students to study whether family structure (being in a single-parent home, having siblings, and so on) affects student performance.

I leave the specific econometric analysis up to the students, and they are to craft about a 10-page paper on their own. This gives them practice writing the results in a way that is easy-to-read, and forces them to interpret their findings. While leaving the topic to each student's discretion is more interesting, I find that many students flounder with an open-ended assignment until it is too late. Naturally, for a second-semester course, or a senior seminar, students would be expected to design their own topic, collect their own data, and then write a more substantial term paper.

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.2 (i) $P(X \leq 6) = P[(X - 5)/2 \leq (6 - 5)/2] = P(Z \leq .5) \approx .692$, where Z denotes a Normal (0,1) random variable. [We obtain $P(Z \leq .5)$ from Table G.1.]

(ii) $P(X > 4) = P[(X - 5)/2 > (4 - 5)/2] = P(Z > -.5) = P(Z \leq .5) \approx .692$.

(iii) $P(|X - 5| > 1) = P(X - 5 > 1) + P(X - 5 < -1) = P(X > 6) + P(X < 4) \approx (1 - .692) + (1 - .692) = .616$, where we have used answers from parts (i) and (ii).

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.4 We want $P(X \geq .6)$. Because X is continuous, this is the same as $P(X > .6) = 1 - P(X \leq .6) = F(.6) = 3(.6)^2 - 2(.6)^3 = .648$. One way to interpret this is that almost 65% of all counties have an elderly employment rate of .6 or higher.

B.5 (i) As stated in the hint, if X is the number of jurors convinced of Simpson's innocence, then $X \sim$ 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, θ = .2, and x = 1: P(X = 1) = 12· (.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.6 E(X) = $\int_0^3 xf(x)dx$ = $\int_0^3 x[(1/9)x^2]dx$ = (1/9) $\int_0^3 x^3 dx$. But $\int_0^3 x^3 dx = (1/4)x^4\Big|_0^3$ = 81/4. Therefore, E(X) = (1/9)(81/4) = 9/4, or 2.25 years.

B.7 In eight attempts the expected number of free throws is 8(.74) = 5.92, or about six free throws.

B.8 The weights for the two-, three-, and four-credit courses are 2/9, 3/9, and 4/9, respectively. Let Y_j be the grade in the j^{th} course, j = 1, 2, and 3, and let X be the overall grade point average. Then $X = (2/9)Y_1 + (3/9)Y_2 + (4/9)Y_3$ and the expected value is E(X) = (2/9)E(Y_1) + (3/9)E(Y_2) + (4/9)E(Y_3) = (2/9)(3.5) + (3/9)(3.0) + (4/9)(3.0) = (7 + 9 + 12)/9 ≈ 3.11.

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.

B.10 (i) E($GPA|SAT$ = 800) = .70 + .002(800) = 2.3. Similarly, E($GPA|SAT$ = 1,400) = .70 + .002(1400) = 3.5. The difference in expected GPAs is substantial, but the difference in SAT scores is also rather large.

(ii) Following the hint, we use the law of iterated expectations. Since E($GPA|SAT$) = .70 + .002 SAT, the (unconditional) expected value of GPA is .70 + .002 E(SAT) = .70 + .002(1100) = 2.9.

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.2 (i) $E(W_a) = a_1E(Y_1) + a_2E(Y_2) + \ldots + a_nE(Y_n) = (a_1 + a_2 + \ldots + a_n)\mu$. Therefore, we must have $a_1 + a_2 + \ldots + a_n = 1$ for unbiasedness.

(ii) $\text{Var}(W_a) = a_1^2\text{Var}(Y_1) + a_2^2\text{Var}(Y_2) + \ldots + a_n^2\text{Var}(Y_n) = (a_1^2 + a_2^2 + \ldots + a_n^2)\sigma^2$.

(iii) From the hint, when $a_1 + a_2 + \ldots + a_n = 1$ – the condition needed for unbiasedness of W_a – we have $1/n \le a_1^2 + a_2^2 + \ldots + a_n^2$. But then $\text{Var}(\bar{Y}) = \sigma^2/n \le \sigma^2(a_1^2 + a_2^2 + \ldots + a_n^2) = \text{Var}(W_a)$.

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.4 (i) Using the hint, $E(Z|X) = E(Y/X|X) = E(Y|X)/X = \theta X/X = \theta$. It follows by Property CE.4, the law of iterated expectations, that $E(Z) = E(\theta) = \theta$.

(ii) This follows from part (i) and the fact that the sample average is unbiased for the population average: write

$$W_1 = n^{-1}\sum_{i=1}^{n}(Y_i / X_i) = n^{-1}\sum_{i=1}^{n} Z_i,$$

where $Z_i = Y_i/X_i$. From part (i), $E(Z_i) = \theta$ for all i.

(iii) In general, the average of the ratios, Y_i/X_i, is not the ratio of averages, $W_2 = \bar{Y}/\bar{X}$. (This non-equivalence is discussed a bit on page 676.) Nevertheless, W_2 is also unbiased, as a simple application of the law of iterated expectations shows. First, $E(Y_i|X_1,\ldots,X_n) = E(Y_i|X_i)$ under random sampling because the observations are independent. Therefore, $E(Y_i|X_1,\ldots,X_n) = \theta X_i$ and so

$$E(\bar{Y} \mid X_1,\ldots,X_n) = n^{-1}\sum_{i=1}^{n} E(Y_i \mid X_1,\ldots,X_n) = n^{-1}\sum_{i=1}^{n}\theta X_i$$
$$= \theta n^{-1}\sum_{i=1}^{n} X_i = \theta \bar{X}.$$

Therefore, $E(W_2 \mid X_1,\ldots,X_n) = E(\bar{Y}/\bar{X} \mid X_1,\ldots,X_n) = \theta\bar{X}/\bar{X} = \theta$, which means that W_2 is actually unbiased conditional on $(X_1,\ldots,X_n)$, and therefore also unconditionally unbiased.

(iv) For the $n = 17$ observations given in the table – which are, incidentally, the first 17 observations in the file CORN.RAW – the point estimates are $w_1 = .418$ and $w_2 = 120.43/297.41 = .405$. These are pretty similar estimates. If we use w_1, we estimate $E(Y|X = x)$ for any $x > 0$ as $\widehat{E(Y \mid X = x)} = .418\,x$. For example, if $x = 300$ then the predicted yield is .418(300) = 125.4.

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.6 (i) H_0: $\mu = 0$.

(ii) H_1: $\mu < 0$.

(iii) The standard error of $\bar{y}$ is $s/\sqrt{n}$ = 466.4/30 ≈ 15.55. Therefore, the t statistic for testing H_0: $\mu = 0$ is $t = \bar{y}/\text{se}(\bar{y}) = -32.8/15.55 \approx -2.11$. We obtain the p-value as $P(Z \leq -2.11)$, where $Z \sim \text{Normal}(0,1)$. These probabilities are in Table G.1: p-value = .0174. Because the p-

value is below .05, we reject H_0 against the one-sided alternative at the 5% level. We do not reject at the 1% level because *p*-value = .0174 > .01.

(iv) The estimated reduction, about 33 ounces, does not seem large for an entire year's consumption. If the alcohol is beer, 33 ounces is less than three 12-ounce cans of beer. Even if this is hard liquor, the reduction seems small. (On the other hand, when aggregated across the entire population, alcohol distributors might not think the effect is so small.)

(v) The implicit assumption is that other factors that affect liquor consumption – such as income, or changes in price due to transportation costs, are constant over the two years.

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 ± 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.8 (i) For Mark Price, $\bar{y}$ = 188/429 $\approx$ .438.

(ii) $\mathrm{Var}(\bar{Y}) = \theta(1-\theta)/n$ [because the variance of each Y_i is $\theta(1-\theta)$ and so $\mathrm{sd}(\bar{Y})$ = $\sqrt{\theta(1-\theta)/n}$.

(iii) The asymptotic t statistic is $(\bar{Y} - .5)/\mathrm{se}(\bar{Y})$; when we plug in the estimate for Mark Price, $\mathrm{se}(\bar{y}) = \sqrt{\bar{y}(1-\bar{y})/n} = \sqrt{.438(1-.438)/429} \approx .024$. So the observed t statistic is $(.438 - .5)/.024 \approx -2.583$. This is well below the 5% critical value (based on the standard normal distribution), –1.645. In fact, the 1% critical value is –2.326, and so H_0 is rejected against H_1 at the 1% level.

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) $P(X \leq 115) = P[(X - 130)/6.75 \leq (115 - 130)/6.75] \approx P(Z \leq -2.22)$, where Z is a standard normal random variable. From Table G.1, $P(Z \leq -2.22) \approx .013$.

(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.

C.10 Since $\bar{y} = .394$, $\text{se}(\bar{y}) \approx .024$. We can use the standard normal approximation for the 95% CI: $.394 \pm 1.96(.024)$, or about .347 to .441. Therefore, based on Gwynn's average up to strike, there is not very strong evidence against $\theta = .400$, as this value is well within the 95% CI. (Of course, .350, say, is within this CI, too.)

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) $\mathbf{BA}$ does not exist because $\mathbf{B}$ is 3×3 and $\mathbf{A}$ is 2×3.

D.2 This result is easy to visualize. If $\mathbf{A}$ and $\mathbf{B}$ are $n \times n$ diagonal matrices, then $\mathbf{AB}$ is an $n \times n$ diagonal matrix with j^{th} diagonal element $a_j b_j$. Similarly, $\mathbf{BA}$ is an $n \times n$ diagonal matrix with j^{th} diagonal element $b_j a_j$, which, of course, is the same as $a_j b_j$.

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.4 (i) This follows from tr($\mathbf{BC}$) = tr($\mathbf{CB}$), when $\mathbf{B}$ is $n \times m$ and $\mathbf{C}$ is $m \times n$. Take $\mathbf{B} = \mathbf{A}'$ and $\mathbf{C} = \mathbf{A}$.

(ii) $\mathbf{A}'\mathbf{A} = \begin{pmatrix} 2 & 0 \\ 0 & 3 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 & -1 \\ 0 & 3 & 0 \end{pmatrix} = \begin{pmatrix} 4 & 0 & -2 \\ 0 & 9 & 0 \\ -2 & 0 & 1 \end{pmatrix}$; therefore, tr($\mathbf{A}'\mathbf{A}$) = 14.

Similarly, $\mathbf{AA}' = \begin{pmatrix} 2 & 0 & -1 \\ 0 & 3 & 0 \end{pmatrix}\begin{pmatrix} 2 & 0 \\ 0 & 3 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 5 & 0 \\ 0 & 9 \end{pmatrix}$, and so tr($\mathbf{AA}'$) = 14.

D.5 (i) The $n \times n$ matrix $\mathbf{C}$ is the inverse of $\mathbf{AB}$ 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.6 (i) Let $\mathbf{e}_j$ be the $n \times 1$ vector with j^{th} element equal to one and all other elements equal to zero. Then straightforward matrix multiplication shows that $\mathbf{e}'_j\mathbf{A}\mathbf{e}_j = a_{jj}$, where a_{jj} is the j^{th} diagonal element. But by definition of positive definiteness, $\mathbf{x}'\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$, including $\mathbf{x} = \mathbf{e}_j$. So $a_{jj} > 0$, $j = 1,2, \ldots, n$.

(ii) The matrix $\mathbf{A} = \begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}$ works because $\mathbf{x}'\mathbf{A}\mathbf{x} = -2 < 0$ for $\mathbf{x}' = (1\ 1)$.

D.7 We must show that, for any $n \times 1$ vector $\mathbf{x}$, $x \neq \mathbf{0}$, $\mathbf{x}'(\mathbf{P}'\mathbf{A}\mathbf{B})\,\mathbf{x} > 0$. But we can write this quadratic form as $(\mathbf{P}\,\mathbf{x})'\mathbf{A}(\mathbf{P}\,\mathbf{x}) = \mathbf{z}'\mathbf{A}\mathbf{z}$ where $\mathbf{z} \equiv \mathbf{P}\mathbf{x}$. Because $\mathbf{A}$ is positive definite by assumption, $\mathbf{z}'\mathbf{A}\mathbf{z} > 0$ 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 $\mathbf{P}\mathbf{x} = \mathbf{0}$ then, because $\mathbf{P}^{-1}$ exists, we have $\mathbf{P}^{-1}\mathbf{P}\mathbf{x} = \mathbf{0}$ or $\mathbf{x} = \mathbf{0}$, which completes the proof.

D.8 Let $\mathbf{z} = \mathbf{A}\mathbf{y} + \mathbf{b}$. Then, by the first property of expected values, $\mathrm{E}(\mathbf{z}) = \mathbf{A}\boldsymbol{\mu}_y + \mathbf{b}$, where $\boldsymbol{\mu}_y = \mathrm{E}(\mathbf{y})$. By Property (3) for variances, $\mathrm{Var}(\mathbf{z}) = \mathrm{E}[(\mathbf{z} - \boldsymbol{\mu}_z)(\mathbf{z} - \boldsymbol{\mu}_z)']$. But $\mathbf{z} - \boldsymbol{\mu}_z = \mathbf{A}\mathbf{y} + \mathbf{b} - (\mathbf{A}\boldsymbol{\mu}_y + \mathbf{b}) = \mathbf{A}(\mathbf{y} - \boldsymbol{\mu}_y)$. Therefore, $(\mathbf{z} - \boldsymbol{\mu}_z)' = (\mathbf{y} - \boldsymbol{\mu}_y)'\mathbf{A}'$, and so $(\mathbf{z} - \boldsymbol{\mu}_z)(\mathbf{z} - \boldsymbol{\mu}_z)' = \mathbf{A}(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)'\mathbf{A}'$. Now we can take the expectation and use the second property of expected value: $\mathrm{E}[\mathbf{A}(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)'\mathbf{A}'] = \mathbf{A}\mathrm{E}[(\mathbf{y} - \boldsymbol{\mu}_y)(\mathbf{y} - \boldsymbol{\mu}_y)']\mathbf{A}' = \mathbf{A}[\mathrm{Var}(\mathbf{y})]\mathbf{A}'$.

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.2 (i) Following the hint, we have SSR($\mathbf{b}$) = $(\mathbf{y} - \mathbf{X}\mathbf{b})'(\mathbf{y} - \mathbf{X}\mathbf{b}) = [\hat{\mathbf{u}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b})]'[\hat{\mathbf{u}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b})] = \hat{\mathbf{u}}'\hat{\mathbf{u}} + \hat{\mathbf{u}}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b}) + (\hat{\boldsymbol{\beta}} - \mathbf{b})'\mathbf{X}'\hat{\mathbf{u}} + (\hat{\boldsymbol{\beta}} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b})$. But by the first order conditions for OLS, $\mathbf{X}'\hat{\mathbf{u}} = \mathbf{0}$, and so $(\mathbf{X}'\hat{\mathbf{u}})' = \hat{\mathbf{u}}'\mathbf{X} = \mathbf{0}$. But then SSR($\mathbf{b}$) = $\hat{\mathbf{u}}'\hat{\mathbf{u}} + (\hat{\boldsymbol{\beta}} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b})$, which is what we wanted to show.

(ii) If $\mathbf{X}$ has a rank k then $\mathbf{X}'\mathbf{X}$ is positive definite, which implies that $(\hat{\boldsymbol{\beta}} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b}) > 0$ for all $\mathbf{b} \neq \hat{\boldsymbol{\beta}}$. The term $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ does not depend on $\mathbf{b}$, and so SSR($\mathbf{b}$) – SSR($\hat{\boldsymbol{\beta}}$) = $(\hat{\boldsymbol{\beta}} - \mathbf{b})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \mathbf{b}) > 0$ for $\mathbf{b} \neq \hat{\boldsymbol{\beta}}$.

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), $(\mathbf{Z'Z})^{-1} = \mathbf{A}^{-1}(\mathbf{X'X})^{-1}(\mathbf{A'})^{-1}$, and so $\tilde{\sigma}^2(\mathbf{Z'Z})^{-1} = \hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X'X})^{-1}(\mathbf{A}^{-1})'$, 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 $\hat{\beta}_j/a_j$, $j = 2, \dots, k$.

(v) From part (iii), the estimated variance matrix of $\tilde{\boldsymbol{\beta}}$ is $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X'X})^{-1}(\mathbf{A}^{-1})'$. 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 $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X'X})^{-1}\mathbf{A}^{-1}$, which is easily seen to be $= \hat{\sigma}^2 c_{jj}/a_j^{-2}$, where c_{jj} is the j^{th} diagonal element of $(\mathbf{X'X})^{-1}$. 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.4 (i) $\text{E}(\hat{\boldsymbol{\delta}}\,|\,\mathbf{X}) = \text{E}(\mathbf{G}\hat{\boldsymbol{\beta}}\,|\,\mathbf{X}) = \mathbf{G}\text{E}(\hat{\boldsymbol{\beta}}\,|\,\mathbf{X}) = \mathbf{G}\boldsymbol{\beta} = \boldsymbol{\delta}.$

(ii) $\text{Var}(\hat{\boldsymbol{\delta}}\,|\,\mathbf{X}) = \text{Var}(\mathbf{G}\hat{\boldsymbol{\beta}}\,|\,\mathbf{X}) = \mathbf{G}[\text{Var}(\hat{\boldsymbol{\beta}}\,|\,\mathbf{X})]\mathbf{G}' = \mathbf{G}[\sigma^2(\mathbf{X'X})^{-1}]\mathbf{G}' = \sigma^2\mathbf{G}[(\mathbf{X'X})^{-1}]\mathbf{G}'.$

(iii) The vector of regression coefficients from the regression $\mathbf{y}$ on $\mathbf{XG}^{-1}$ is

$$\begin{aligned}[(\mathbf{XG}^{-1})'\mathbf{XG}^{-1}]^{-1}(\mathbf{XG}^{-1})'\mathbf{y} &= [(\mathbf{G}^{-1})'\mathbf{X'XG}^{-1}]^{-1}(\mathbf{G}^{-1})'\mathbf{X'y}\\ &= \mathbf{G}(\mathbf{X'X})'[(\mathbf{G}^{-1})']^{-1}(\mathbf{G}')^{-1}\mathbf{X'y}\\ &= \mathbf{G}(\mathbf{X'X})'\mathbf{G}'(\mathbf{G}')^{-1}\mathbf{X'y} = \mathbf{G}(\mathbf{X'X})'\mathbf{X'y} = \hat{\boldsymbol{\delta}}.\end{aligned}$$

Further, as shown in Problem E.3, the residuals are the same as from the regression $\mathbf{y}$ on $\mathbf{X}$, and so the error variance estimate, $\hat{\sigma}^2$, is the same. Therefore, the estimated variance matrix is

$$\hat{\sigma}^2[(\mathbf{X}\mathbf{G}^{-1})'\mathbf{X}\mathbf{G}^{-1}]^{-1} = \hat{\sigma}^2\mathbf{G}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{G}',$$

which is the proper estimate of the expression in part (ii).

(iv) It is easily seen by matrix multiplication that choosing

$$\mathbf{G} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ 0 & \ldots & 0 & 1 & 0 \\ c_1 & c_2 & c_3 & \ldots & c_k \end{pmatrix}$$

does the trick: if $\boldsymbol{\delta} = \mathbf{G}\boldsymbol{\beta}$ then $\delta_j = \beta_j$, $j = 1,\ldots,k-1$, and $\delta_k = c_1\beta_1 + c_2\beta_2 + \ldots + c_k\beta_k$.

(v) Straightforward matrix multiplication shows that, for the suggested choice of $\mathbf{G}^{-1}$, $\mathbf{G}^{-1}\mathbf{G} = \mathbf{I}_n$. Also by multiplication, it is easy to see that, for each t,

$$\mathbf{x}_t\mathbf{G}^{-1} = [x_{t1} - (c_1/c_k)x_{tk}, x_{t2} - (c_2/c_k)x_{tk}, \ldots, x_{t,k-1} - (c_{k-1}/c_k)x_{tk}, x_{tk}/c_k].$$

E.5 (i) By plugging in for **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 **Z** is a function of **X** to pull **Z** outside of the conditional expectation:

$$\mathrm{E}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta} + \mathrm{E}[(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u} \mid \mathbf{X}] = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathrm{E}(\mathbf{u} \mid \mathbf{X}) = \boldsymbol{\beta}.$$

(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

$$\begin{aligned}\mathrm{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'[\mathrm{Var}(\mathbf{u} \mid \mathbf{X})]\mathbf{Z}[(\mathbf{Z}'\mathbf{X})^{-1}]' \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\sigma^2\mathbf{I}_n)\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1} = \sigma^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}.\end{aligned}$$

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 **y** and, as shown in part (i), it is unbiased (conditional on **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 $\mathrm{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X})$. Therefore, we prefer the OLS estimator.