CHAPTER 2

SOLUTIONS TO PROBLEMS

2.1 (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 $E(u|educ) \neq 0$, and so SLR.4 fails.

2.3 (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

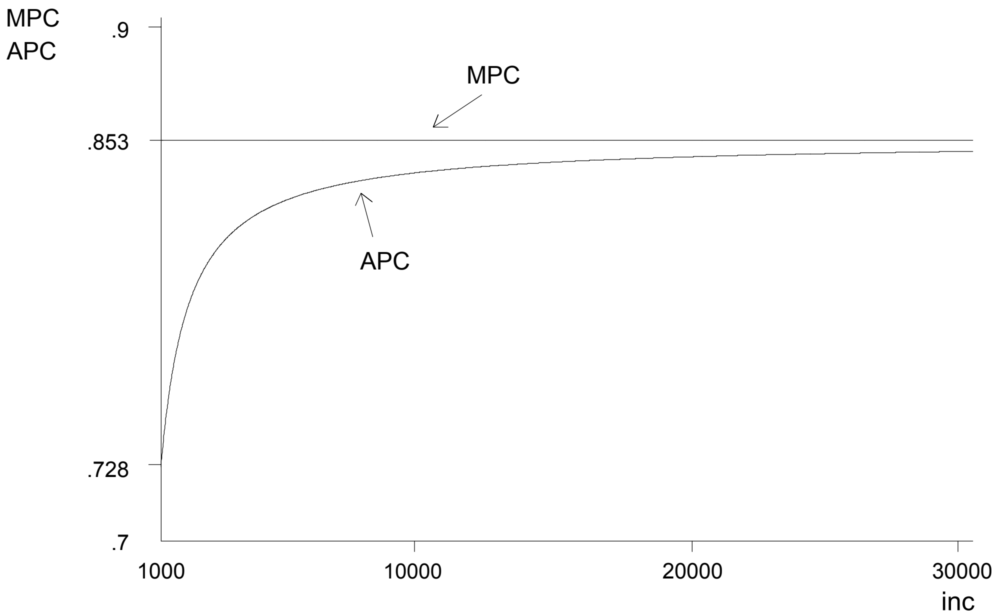
$$R^2 = 1 - \text{SSR/SST} \approx 1 - (.4347/1.0288) \approx .577.$$

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

2.5 (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).



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

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

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

SOLUTIONS TO COMPUTER EXERCISES

2.10 (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.

2.12 (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.

2.14 (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%.

CHAPTER 3

SOLUTIONS TO PROBLEMS

3.1 (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) 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 (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 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) $\beta_1 < 0$ 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 Corr(x_1,x_2) < 0, the simple regression estimator $\tilde{\beta}_1$ has a downward bias. But because $\beta_1 < 0$, this means that the simple regression, on average, overstates the importance of pollution. [E($\tilde{\beta}_1$) is more negative than β_1.]

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

3.10 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 $\hat{r}_{i1}$.

3.12 (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:

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

where we use the fact that $\sum_{i=1}^{n}(z_i - \bar{z}) = 0$ 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,

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

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),

$$Var(\tilde{\beta}_1) = Var\frac{\sum_{i=1}^{n}(z_i - \bar{z})u_i}{s_{zx}^2} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2 Var(u_i)}{s_{zx}^2}$$

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

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

(iii) We know that Var($\hat{\beta}_1$) = $\sigma^2/[\sum_{i=1}^{n}(x_i - \bar{x})^2]$. 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($\tilde{\beta}_1$) ≥ Var($\hat{\beta}_1$), which is what we wanted to show.

SOLUTIONS TO COMPUTER EXERCISES

3.13 (i) Probably $\beta_2 > 0$, 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.)

3.15 (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.

3.17 The regression of *educ* on *exper* and *tenure* yields

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

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

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.

CHAPTER 4

SOLUTIONS TO PROBLEMS

4.1 (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) Holding *profmarg* fixed, $\Delta \widehat{rdintens}$ = .321 Δlog(*sales*) = (.321/100)[$100 \cdot \Delta \log(sales)$] ≈ .00321(%Δ*sales*). Therefore, if %Δ*sales* = 10, $\Delta \widehat{rdintens}$ ≈ .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: β_1 = 0 versus H_1: β_1 > 0, where β_1 is the population slope on log(*sales*). The *t* statistic is .321/.216 ≈ 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.

4.5 (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.

4.7 (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 β_2 log(*employ*) from the right-hand-side and collect terms, we have

$$\log(scrap) = \beta_0 + \beta_1 hrsemp + [\beta_2 \log(\text{sales}) - \beta_2 \log(employ)]$$

$$+ [\beta_2 \log(employ) + \beta_3 \log(employ)] + u$$

$$= \beta_0 + \beta_1 hrsemp + \beta_2 \log(sales/employ)$$

$$+ (\beta_2 + \beta_3)\log(employ) + u,$$

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.

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

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

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

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

4.11 (i) 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.

SOLUTIONS TO COMPUTER EXERCISES

4.12 (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

$$\begin{aligned} \widehat{voteA} = &\ \underset{(3.93)}{45.08} + \underset{(0.382)}{6.083} \log(expendA) - \underset{(0.379)}{6.615} \log(expendB) + \underset{(.062)}{.152}\, prtystrA \\ &n = 173,\ R^2 = .793. \end{aligned}$$

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: $\beta_2 = -\beta_1$.

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

4.16 (i) If we drop *rbisyr* the estimated equation becomes

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

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

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

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

$$\widehat{\log(salary)} = 10.41 + .0700\, years + .0079\, gamesyr$$
$$(2.00)\quad (.0120) \qquad (.0027)$$

$$+ .00053\, bavg + .0232\, hrunsyr$$
$$(.00110) \qquad (.0086)$$

$$+ .0174\, runsyr + .0010\, fldperc - .0064\, sbasesyr$$
$$(.0051) \qquad (.0020) \qquad (.0052)$$

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

Of the three additional independent variables, only *runsyr* is statistically significant (t statistic = .0174/.0051 ≈ 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 ≈ .56. Therefore, these variables are jointly very insignificant.

4.18 (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:

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

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

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.

CHAPTER 5

SOLUTIONS TO PROBLEMS

5.1 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($\hat{\beta}_1$)$\cdot$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.

5.3 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.

SOLUTIONS TO COMPUTER EXERCISES

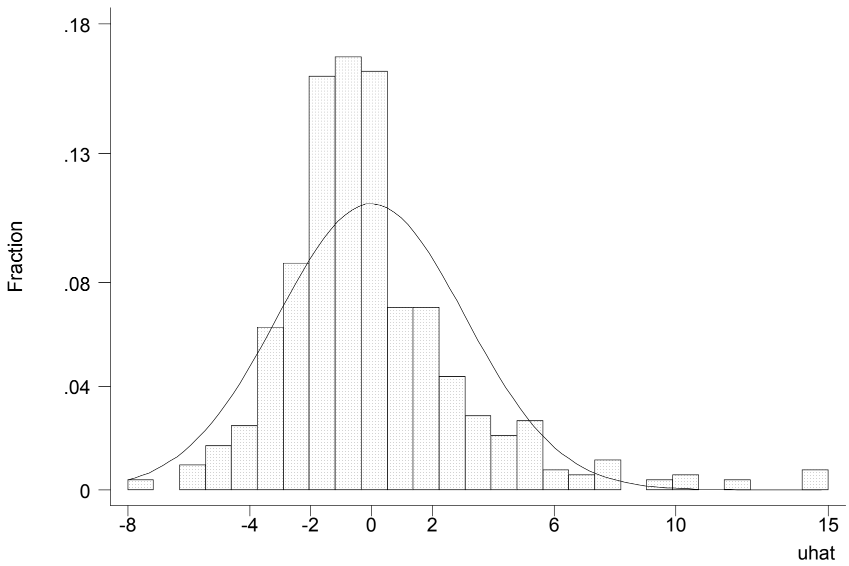
5.5 (i) The estimated equation is

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

$\widehat{wage}$ = –2.87	+.599 *educ*	+.022 *exper*	+.169 *tenure*
(0.73)	(.051)	(.012)	(.022)

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

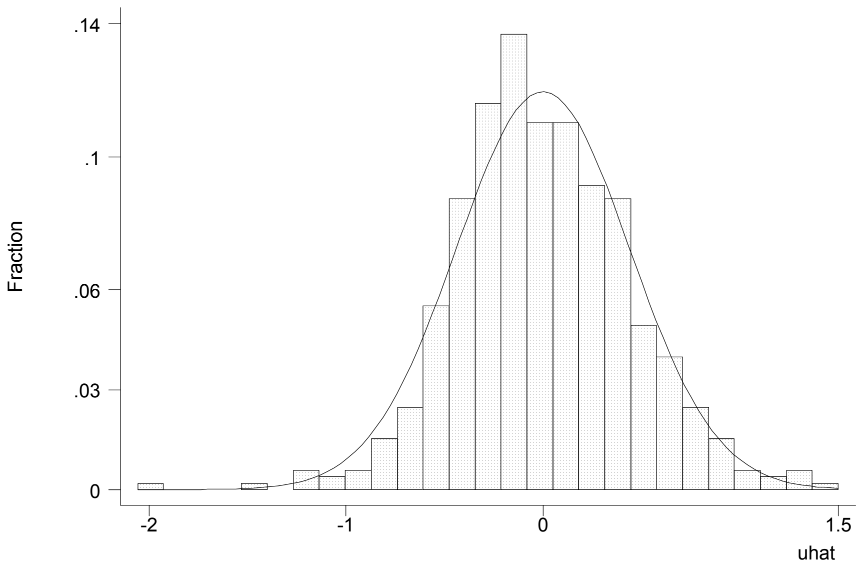


(ii) With log(*wage*) as the dependent variable the estimated equation is

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

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

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.

5.7 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.

Κεφάλαιο 6

6.1 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.)

6.3 (i) The turnaround point is given by $\hat{\beta}_1/(2|\hat{\beta}_2|)$, or .0003/(.000000014) ≈ 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* ≈ –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.

6.5 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.

6.7 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.

SOLUTIONS TO COMPUTER EXERCISES

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

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

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

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

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

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

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

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

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

We can find the value of log(*inst*) where the effect on log(*price*) actually becomes negative: 2.073/[2(.1193)] ≈ 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.

6.10 (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

$$\widehat{\log(wage)} = \underset{(0.24)}{5.95} + \underset{(.0174)}{.0440}\, educ - \underset{(.0200)}{.0215}\, exper + \underset{(.00153)}{.00320}\, educ \cdot exper$$

$n = 935$, $R^2 = .135$, $\bar{R}^2 = .132$.

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.

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

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

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

(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 $\widehat{price} = \hat{\alpha}_0 \exp(\widehat{lprice})$, 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.

6.14 (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

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

(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 $\hat{\theta}_2$.

6.16 (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.

Κεφάλαιο 7

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.

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

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

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

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

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

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

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

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

$$inlf = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsage6 + u,$$

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

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

or

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

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

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

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

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

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

SOLUTIONS TO COMPUTER EXERCISES

7.9 (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 ≈ .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.

7.11 (i) H_0: $\beta_{13} = 0$. Using the data in MLB1.RAW gives $\hat{\beta}_{13} \approx .254$, se($\hat{\beta}_{13}$) ≈ .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).

7.13 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.

7.15 (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.

7.17 (i) About .392, or 39.2%.

(ii) The estimated equation is

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

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

(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

$$\begin{aligned}\widehat{e401k} = &-.502 + .0123\, inc - .000061\, inc^2 + .0265\, age - .00031\, age^2 \\ &(.081) \quad (.0006) \quad (.000005) \quad (.0039) \quad (.00005) \\ &- .0038\, male + .0198\, pira \\ &(.0121) \quad (.0122)\end{aligned}$$

$n = 9{,}275, \quad R^2 = .095.$

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.

7.19 (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

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

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

$n = 9{,}275$, $R^2 = .204$, 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 $F = [(30{,}215{,}207.5 - 29{,}985{,}400)/29{,}985{,}400]*(9245/20) \approx 3.54$. With 20 and 9,245 *df*, the p-value is essentially zero. In this case, there is strong evidence that the slopes change across family size. Allowing for intercept changes alone is not sufficient. (If you look at the individual regressions, you will see that the signs on the income variables actually change across family size.)

APPENDIX A

SOLUTIONS TO PROBLEMS

A.1 (i) \$566.

(ii) The two middle numbers are 480 and 530; when these are averaged, we obtain 505, or \$505.

(iii) 5.66 and 5.05, respectively.

(iv) The average increases to \$586 while the median is unchanged (\$505).

A.3 If *price* = 15 and *income* = 200, *quantity* = 120 – 9.8(15) + .03(200) = –21, which is nonsense. This shows that linear demand functions generally cannot describe demand over a wide range of prices and income.

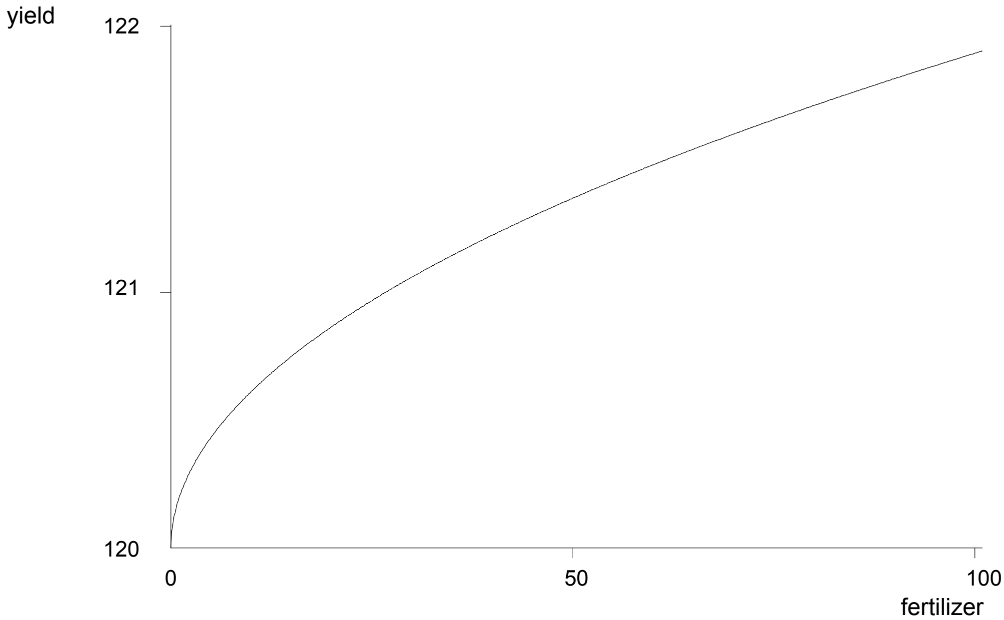
A.5 The majority shareholder is referring to the percentage point increase in the stock return, while the CEO is referring to the change relative to the initial return of 15%. To be precise, the shareholder should specifically refer to a 3 percentage *point* increase.

A.7 (i) When *exper* = 0, log(*salary*) = 10.6; therefore, *salary* = exp(10.6) $\approx$ \$40,134.84. When *exper* = 5, *salary* = exp[10.6 + .027(5)] $\approx$ \$45,935.80.

(ii) The approximate proportionate increase is .027(5) = .135, so the approximate percentage change is 13.5%.

(iii) 100[(45,935.80 – 40,134.84)/40,134.84) $\approx$ 14.5%, so the exact percentage increase is about one percentage point higher.

A.9 (i) The relationship between *yield* and *fertilizer* is graphed below.



(ii) Compared with a linear function, the function

$$yield = .120 + .19\sqrt{fertilizer}$$

has a diminishing effect, and the slope approaches zero as *fertilizer* gets large. The initial pound of fertilizer has the largest effect, and each additional pound has an effect smaller than the previous pound.

APPENDIX B

SOLUTIONS TO PROBLEMS

B.1 Before the student takes the SAT exam, we do not know – nor can we predict with certainty – what the score will be. The actual score depends on numerous factors, many of which we cannot even list, let alone know ahead of time. (The student's innate ability, how the student feels on exam day, and which particular questions were asked, are just a few.) The eventual SAT score clearly satisfies the requirements of a random variable.

B.3 (i) Let Y_{it} be the binary variable equal to one if fund i outperforms the market in year t. By assumption, $P(Y_{it} = 1) = .5$ (a 50-50 chance of outperforming the market for each fund in each year). Now, for any fund, we are also assuming that performance relative to the market is independent across years. But then the probability that fund i outperforms the market in all 10 years, $P(Y_{i1} = 1, Y_{i2} = 1, \ldots, Y_{i,10} = 1)$, is just the product of the probabilities: $P(Y_{i1} = 1) \cdot P(Y_{i2} = 1) \ldots P(Y_{i,10} = 1) = (.5)^{10} = 1/1024$ (which is slightly less than .001). In fact, if we define a binary random variable Y_i such that $Y_i = 1$ if and only if fund i outperformed the market in all 10 years, then $P(Y_i = 1) = 1/1024$.

(ii) Let X denote the number of funds out of 4,170 that outperform the market in all 10 years. Then $X = Y_1 + Y_2 + \ldots + Y_{4,170}$. If we assume that performance relative to the market is independent across funds, then X has the Binomial (n,θ) distribution with $n = 4,170$ and $\theta = 1/1024$. We want to compute $P(X \geq 1) = 1 - P(X = 0) = 1 - P(Y_1 = 0, Y_2 = 0, \ldots, Y_{4,170} = 0) = 1 - P(Y_1 = 0) \cdot P(Y_2 = 0) \cdots P(Y_{4,170} = 0) = 1 - (1023/1024)^{4170} \approx .983$. This means, if performance relative to the market is random and independent across funds, it is almost certain that at least one fund will outperform the market in all 10 years.

(iii) Using the Stata command Binomial(4170,5,1/1024), the answer is about .385. So there is a nontrivial chance that at least five funds will outperform the market in all 10 years.

B.5 (i) As stated in the hint, if X is the number of jurors convinced of Simpson's innocence, then $X \sim$ Binomial(12,.20). We want $P(X \geq 1) = 1 - P(X = 0) = 1 - (.8)^{12} \approx .931$.

(ii) Above, we computed $P(X = 0)$ as about .069. We need $P(X = 1)$, which we obtain from (B.14) with $n = 12$, $\theta = .2$, and $x = 1$: $P(X = 1) = 12 \cdot (.2)(.8)^{11} \approx .206$. Therefore, $P(X \geq 2) \approx 1 - (.069 + .206) = .725$, so there is almost a three in four chance that the jury had at least two members convinced of Simpson's innocence prior to the trial.

B.7 In eight attempts the expected number of free throws is 8(.74) = 5.92, or about six free throws.

B.9 If Y is salary in dollars then $Y = 1000 \cdot X$, and so the expected value of Y is 1,000 times the expected value of X, and the standard deviation of Y is 1,000 times the standard deviation of X. Therefore, the expected value and standard deviation of salary, measured in dollars, are \$52,300 and \$14,600, respectively.

APPENDIX C

SOLUTIONS TO PROBLEMS

C.1 (i) This is just a special case of what we covered in the text, with $n = 4$: $E(\bar{Y}) = \mu$ and $Var(\bar{Y}) = \sigma^2/4$.

(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,

$$\begin{aligned} Var(W) &= Var(Y_1)/64 + Var(Y_2)/64 + Var(Y_3)/16 + Var(Y_4)/4 \\ &= \sigma^2[(1/64) + (1/64) + (4/64) + (16/64)] = \sigma^2(22/64) = \sigma^2(11/32). \end{aligned}$$

(iii) Because $11/32 > 8/32 = 1/4$, $Var(W) > Var(\bar{Y})$ for any $\sigma^2 > 0$, so $\bar{Y}$ is preferred to W because each is unbiased.

C.3 (i) $E(W_1) = [(n - 1)/n]E(\bar{Y}) = [(n - 1)/n]\mu$, and so $Bias(W_1) = [(n - 1)/n]\mu - \mu = -\mu/n$. Similarly, $E(W_2) = E(\bar{Y})/2 = \mu/2$, and so $Bias(W_2) = \mu/2 - \mu = -\mu/2$. 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) $plim(W_1) = plim[(n - 1)/n]\cdot plim(\bar{Y}) = 1\cdot\mu = \mu$. $plim(W_2) = plim(\bar{Y})/2 = \mu/2$. Because $plim(W_1) = \mu$ and $plim(W_2) = \mu/2$, W_1 is consistent whereas W_2 is inconsistent.

(iii) $Var(W_1) = [(n - 1)/n]^2 Var(\bar{Y}) = [(n - 1)^2/n^3]\sigma^2$ and $Var(W_2) = Var(\bar{Y})/4 = \sigma^2/(4n)$.

(iv) Because $\bar{Y}$ is unbiased, its mean squared error is simply its variance. On the other hand, $MSE(W_1) = Var(W_1) + [Bias(W_1)]^2 = [(n - 1)^2/n^3]\sigma^2 + \mu^2/n^2$. When $\mu = 0$, $MSE(W_1) = Var(W_1) = [(n - 1)^2/n^3]\sigma^2 < \sigma^2/n = Var(\bar{Y})$ because $(n - 1)/n < 1$. Therefore, $MSE(W_1)$ is smaller than $Var(\bar{Y})$ for μ close to zero. For large n, the difference between the two estimators is trivial.

C.5 (i) While the expected value of the numerator of G is $E(\bar{Y}) = \theta$, and the expected value of the denominator is $E(1 - \bar{Y}) = 1 - \theta$, the expected value of the ratio is not the ratio of the expected value.

(ii) By Property PLIM.2(iii), the plim of the ratio is the ratio of the plims (provided the plim of the denominator is not zero): $plim(G) = plim[\bar{Y}/(1 - \bar{Y})] = plim(\bar{Y})/[1 - plim(\bar{Y})] = \theta/(1 - \theta) = \gamma$.

C.7 (i) The average increase in wage is $\bar{d} = .24$, or 24 cents. The sample standard deviation is about .451, and so, with $n = 15$, the standard error of $\bar{d}$ is $.451\sqrt{15} \approx .1164$. From Table G.2, the 97.5th percentile in the t_{14} distribution is 2.145. So the 95% CI is .24 ± 2.145(.1164), or about –.010 to .490.

(ii) If μ = E(D_i) then H_0: μ = 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 ≈ 2.062. The 5% critical value for a one-tailed test with df = 14 is 1.761, while the 1% critical value is 2.624. Therefore, H_0 is rejected in favor of H_1 at the 5% level but not the 1% level.

(iv) The p-value obtained from Stata is .029; this is half of the p-value for the two-sided alternative. (Econometrics packages, including Stata, report the p-value for the two-sided alternative.)

C.9 (i) X is distributed as Binomial(200,.65), and so E(X) = 200(.65) = 130.

(ii) Var(X) = 200(.65)(1 − .65) = 45.5, so sd(X) ≈ 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.

APPENDIX D

SOLUTIONS TO PROBLEMS

D.1 (i) $\mathbf{AB} = \begin{pmatrix} 2 & -1 & 7 \\ -4 & 5 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 6 \\ 1 & 8 & 0 \\ 3 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 20 & -6 & 12 \\ 5 & 36 & -24 \end{pmatrix}$

(ii) **BA** does not exist because **B** is 3 × 3 and **A** is 2 × 3.

D.3 Using the basic rules for transpose, $(\mathbf{X}'\mathbf{X}') = (\mathbf{X}')(\mathbf{X}')' = \mathbf{X}'\mathbf{X}$, which is what we wanted to show.

D.5 (i) The $n \times n$ matrix **C** is the inverse of **AB** if and only if $\mathbf{C}(\mathbf{AB}) = \mathbf{I}_n$ and $(\mathbf{AB})\mathbf{C} = \mathbf{I}_n$. We verify both of these equalities for $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. First, $(\mathbf{B}^{-1}\mathbf{A}^{-1})(\mathbf{AB}) = \mathbf{B}^{-1}(\mathbf{A}^{-1}\mathbf{A})\mathbf{B} = \mathbf{B}^{-1}\mathbf{I}_n\mathbf{B} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}_n$. Similarly, $(\mathbf{AB})(\mathbf{B}^{-1}\mathbf{A}^{-1}) = \mathbf{A}(\mathbf{BB}^{-1})\mathbf{A}^{-1} = \mathbf{AI}_n\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}_n$.

(ii) $(\mathbf{ABC})^{-1} = (\mathbf{BC})^{-1}\mathbf{A}^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1}$.

D.7 We must show that, for any $n \times 1$ vector **x**, $\boldsymbol{x} \neq \mathbf{0}$, $\mathbf{x}'(\mathbf{P}'\mathbf{AB})\,\mathbf{x} > 0$. But we can write this quadratic form as $(\mathbf{P\,x})'\mathbf{A}(\mathbf{P\,x}) = \mathbf{z}'\mathbf{Az}$ where $\mathbf{z} \equiv \mathbf{Px}$. Because **A** is positive definite by assumption, $\mathbf{z}'\mathbf{Az} > 0$ for $\mathbf{z} \neq \mathbf{0}$. So, all we have to show is that $\mathbf{x} \neq \mathbf{0}$ implies that $\mathbf{z} \neq \mathbf{0}$. We do this by showing the contrapositive, that is, if $\mathbf{z} = \mathbf{0}$ then $\mathbf{x} = \mathbf{0}$. If $\mathbf{Px} = \mathbf{0}$ then, because $\mathbf{P}^{-1}$ exists, we have $\mathbf{P}^{-1}\mathbf{Px} = \mathbf{0}$ or $\mathbf{x} = \mathbf{0}$, which completes the proof.

APPENDIX E

SOLUTIONS TO PROBLEMS

E.1 This follows directly from partitioned matrix multiplication in Appendix D. Write

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_n \end{pmatrix}, \mathbf{X}' = (\mathbf{x}_1' \; \mathbf{x}_2' \ldots \mathbf{x}_n'), \text{ and } \mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}$$

Therefore, $\mathbf{X}'\mathbf{X} = \sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t$ and $\mathbf{X}'\mathbf{y} = \sum_{t=1}^{n} \mathbf{x}_t'\mathbf{y}_t$. An equivalent expression for $\hat{\boldsymbol{\beta}}$ is

$$\hat{\boldsymbol{\beta}} = \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t\right)^{-1} \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t' y_t\right)$$

which, when we plug in $y_t = \mathbf{x}_t\boldsymbol{\beta} + u_t$ for each t and do some algebra, can be written as

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t'\mathbf{x}_t\right)^{-1} \left(n^{-1}\sum_{t=1}^{n} \mathbf{x}_t' u_t\right).$$

As shown in Section E.4, this expression is the basis for the asymptotic analysis of OLS using matrices.

E.3 (i) We use the placeholder feature of the OLS formulas. By definition, $\tilde{\boldsymbol{\beta}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y} = [(\mathbf{XA})' (\mathbf{XA})]^{-1}(\mathbf{XA})'\mathbf{y} = [\mathbf{A}'(\mathbf{X}'\mathbf{X})\mathbf{A}]^{-1}\mathbf{A}'\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}')^{-1}\mathbf{A}'\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$.

(ii) By definition of the fitted values, $\hat{y}_t = \mathbf{x}_t\hat{\boldsymbol{\beta}}$ and $\tilde{y}_t = \mathbf{z}_t\tilde{\boldsymbol{\beta}}$. Plugging $\mathbf{z}_t$ and $\tilde{\boldsymbol{\beta}}$ into the second equation gives $\tilde{y}_t = (\mathbf{x}_t\mathbf{A})(\mathbf{A}^{-1}\hat{\boldsymbol{\beta}}) = \mathbf{x}_t\hat{\boldsymbol{\beta}} = \hat{y}_t$.

(iii) The estimated variance matrix from the regression of $\mathbf{y}$ and $\mathbf{Z}$ is $\tilde{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1}$ where $\tilde{\sigma}^2$ is the error variance estimate from this regression. From part (ii), the fitted values from the two regressions are the same, which means the residuals must be the same for all t. (The dependent variable is the same in both regressions.) Therefore, $\tilde{\sigma}^2 = \hat{\sigma}^2$. Further, as we showed in part (i), $(\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}')^{-1}$, and so $\tilde{\sigma}^2(\mathbf{Z}'\mathbf{Z})^{-1} = \hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}^{-1})'$, which is what we wanted to show.

(iv) The $\tilde{\beta}_j$ are obtained from a regression of $\mathbf{y}$ on $\mathbf{XA}$, where $\mathbf{A}$ is the $k \times k$ diagonal matrix with 1, $a_2, \ldots, a_k$ 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, $a_2^{-1}, \ldots, a_k^{-1}$ down its diagonal. Straightforward multiplication shows that the first element of $\mathbf{A}^{-1}\hat{\boldsymbol{\beta}}$ is $\hat{\beta}_1$ and the j^{th} element is $\hat{\beta}_j/a_j$, $j = 2, \ldots, k$.

(v) From part (iii), the estimated variance matrix of $\tilde{\boldsymbol{\beta}}$ is $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{A}^{-1})'$. But $\mathbf{A}^{-1}$ is a symmetric, diagonal matrix, as described above. The estimated variance of $\tilde{\beta}_j$ is the j^{th} diagonal element of $\hat{\sigma}^2\mathbf{A}^{-1}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{A}^{-1}$, which is easily seen to be = $\hat{\sigma}^2 c_{jj}/a_j^{-2}$, where c_{jj} is the j^{th} diagonal element of $(\mathbf{X}'\mathbf{X})^{-1}$. The square root of this, $\hat{\sigma}\sqrt{c_{jj}}/|a_j|$, is se($\tilde{\beta}_j$), which is simply se($\tilde{\beta}_j$)/$|a_j|$.

(vi) The t statistic for $\tilde{\beta}_j$ is, as usual,

$$\tilde{\beta}_j/\text{se}(\tilde{\beta}_j) = (\hat{\beta}_j/a_j)/[\text{se}(\hat{\beta}_j)/|a_j|],$$

and so the absolute value is $(|\hat{\beta}_j|/|a_j|)/[\text{se}(\hat{\beta}_j)/|a_j|] = |\hat{\beta}_j|/\text{se}(\hat{\beta}_j)$, which is just the absolute value of the t statistic for $\hat{\beta}_j$. If $a_j > 0$, the t statistics themselves are identical; if $a_j < 0$, the t statistics are simply opposite in sign.

E.5 (i) By plugging in for **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:

$$\text{E}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) = \boldsymbol{\beta} + \text{E}[(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u} \mid \mathbf{X}] = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\text{E}(\mathbf{u} \mid \mathbf{X}) = \boldsymbol{\beta}.$$

(ii) We start from the same representation in part (i): $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{u}$ and so

$$\begin{aligned}\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X}) &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'[\text{Var}(\mathbf{u} \mid \mathbf{X})]\mathbf{Z}[(\mathbf{Z}'\mathbf{X})^{-1}]' \\ &= (\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'(\sigma^2\mathbf{I}_n)\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1} = \sigma^2(\mathbf{Z}'\mathbf{X})^{-1}\mathbf{Z}'\mathbf{Z}(\mathbf{X}'\mathbf{Z})^{-1}.\end{aligned}$$

A common mistake is to forget to transpose the matrix $\mathbf{Z}'\mathbf{X}$ in the last term.

(iii) The estimator $\tilde{\boldsymbol{\beta}}$ is linear in **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 $\text{Var}(\tilde{\boldsymbol{\beta}} \mid \mathbf{X})$. Therefore, we prefer the OLS estimator.