CHAPTER 14

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, $\text{Corr}(\Delta u_{it}, \Delta u_{i,t+1}) = \text{Cov}(\Delta u_{it}, \Delta u_{i,t+1})/\text{Var}(\Delta u_{it}) = -\sigma_u^2/(2\sigma_u^2) = -.5$.

14.3 (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

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

This is what we wanted to show.

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

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

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

(ii) An unobserved effects model is

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

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

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

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

SOLUTIONS TO COMPUTER EXERCISES

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

(ii) See Computer Exercise 13.5(ii).

(iii) See Computer Exercise 13.5(iii).

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

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

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

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

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

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

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

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

(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.

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

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

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

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

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

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

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

(iii) The first-differenced equation is

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

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

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

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

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

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

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

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

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

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

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

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

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

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

C14.9 (i) The OLS estimates are

$$\widehat{pctstck} = \underset{(55.17)}{128.54} + \underset{(6.23)}{11.74}\, choice + \underset{(7.23)}{14.34}\, prftshr + \underset{(6.77)}{1.45}\, female - \underset{(.78)}{1.50}\, age$$

$$+ \underset{(1.20)}{.70}\, educ - \underset{(14.23)}{15.29}\, finc25 + \underset{(14.69)}{.19}\, finc35 - \underset{(14.55)}{3.86}\, finc50$$

$$- 13.75\, finc75 - 2.69\, finc100 - 25.05\, finc101 - .0026\, wealth89$$

$$\qquad (16.02) \qquad (15.72) \qquad (17.80) \qquad (.0128)$$

$$+ \quad \underset{(6.68)}{6.67}\ stckin89 - \underset{(6.38)}{7.50}\ irain89$$

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

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

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

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

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

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

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

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

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

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

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

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

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

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