15.6 Using the NLS panel data on $N = 716$ young women, we consider only years 1987 and 1988. We are interested in the relationship between $\ln(WAGE)$ and experience, its square, and indicator variables for living in the south and union membership. Some estimation results are in Table 15.10.

TABLE 15.10 Estimation Results for Exercise 15.6

	(1) OLS 1987	(2) OLS 1988	(3) FE	(4) FE Robust	(5) RE
C	0.9348	0.8993	1.5468	1.5468	1.1497
	(0.2010)	(0.2407)	(0.2522)	(0.2688)	(0.1597)
$EXPER$	0.1270	0.1265	0.0575	0.0575	0.0986
	(0.0295)	(0.0323)	(0.0330)	(0.0328)	(0.0220)
$EXPER^2$	−0.0033	−0.0031	−0.0012	−0.0012	−0.0023
	(0.0011)	(0.0011)	(0.0011)	(0.0011)	(0.0007)
$SOUTH$	−0.2128	−0.2384	−0.3261	−0.3261	−0.2326
	(0.0338)	(0.0344)	(0.1258)	(0.2495)	(0.0317)
$UNION$	0.1445	0.1102	0.0822	0.0822	0.1027
	(0.0382)	(0.0387)	(0.0312)	(0.0367)	(0.0245)
N	716	716	1432	1432	1432

(standard errors in parentheses)

a. The OLS estimates of the $\ln(WAGE)$ model for each of the years 1987 and 1988 are reported in columns (1) and (2). How do the results compare? For these individual year estimations, what are you assuming about the regression parameter values across individuals (heterogeneity)?

b. The $\ln(WAGE)$ equation specified as a panel data regression model is

$$\ln(WAGE_{it}) = \beta_1 + \beta_2 EXPER_{it} + \beta_3 EXPER_{it}^2 + \beta_4 SOUTH_{it} + \beta_5 UNION_{it} + (u_i + e_{it}) \qquad \text{(XR15.6)}$$

Explain any differences in assumptions between this model and the models in part (a).

c. Column (3) contains the estimated fixed effects model specified in part (b). Compare these estimates with the OLS estimates. Which coefficients, apart from the intercepts, show the most difference?

d. The F-statistic for the null hypothesis that there are no individual differences, equation (15.20), is 11.68. What are the degrees of freedom of the F-distribution if the null hypothesis (15.19) is true? What is the 1% level of significance critical value for the test? What do you conclude about the null hypothesis.

e. Column (4) contains the fixed effects estimates with cluster-robust standard errors. In the context of this sample, explain the different assumptions you are making when you estimate with and without cluster-robust standard errors. Compare the standard errors with those in column (3). Which ones are substantially different? Are the robust ones larger or smaller?

a. The estimations of 1987 and 1988 are similar according to the table.

That is, the relationship was fairly stable across 1987 ~ 1988

Assumed that no unobserved individual heterogeneity,

i.e. each year's cross-section is treated independently.

b. The panal model includes an individual-specific effect

u_i, which captures time-invariant unobserved heterogeneity.

c.

Since the coefficient of EXPER drops from 0.127 → 0.0575

and the coefficient of $EXPER^2$ drops from −0.0032 to −0.0012,

they show the most difference

d. the degree of freedom of Numerator:

716 − 1 = 715

the degree of freedom of Denominator:

716(2−1) − 5 = 716 − 5 = 711

1% critical value = 1.86

Since 11.68 > 1.86, we reject Null hypothesis

e. When errors are correlated over time,

regular standard errors may underestimate the true variability.

It consider the error terms in 1987 and 1988 are correlated.

The cluster-robust sd are generally larger.

15.20 This exercise uses data from the STAR experiment introduced to illustrate fixed and random effects for grouped data. In the STAR experiment, children were randomly assigned within schools into three types of classes: small classes with 13–17 students, regular-sized classes with 22–25 students, and regular-sized classes with a full-time teacher aide to assist the teacher. Student scores on achievement tests were recorded as well as some information about the students, teachers, and schools. Data for the kindergarten classes are contained in the data file *star*.

a. Estimate a regression equation (with no fixed or random effects) where *READSCORE* is related to *SMALL*, *AIDE*, *TCHEXPER*, *BOY*, *WHITE_ASIAN*, and *FREELUNCH*. Discuss the results. Do students perform better in reading when they are in small classes? Does a teacher's aide improve scores? Do the students of more experienced teachers score higher on reading tests? Does the student's sex or race make a difference?

b. Reestimate the model in part (a) with school fixed effects. Compare the results with those in part (a). Have any of your conclusions changed? [*Hint*: specify *SCHID* as the cross-section identifier and *ID* as the "time" identifier.]

c. Test for the significance of the school fixed effects. Under what conditions would we expect the inclusion of significant fixed effects to have little influence on the coefficient estimates of the remaining variables?

d. Reestimate the model in part (a) with school random effects. Compare the results with those from parts (a) and (b). Are there any variables in the equation that might be correlated with the school effects? Use the LM test for the presence of random effects.

e. Using the t-test statistic in equation (15.36) and a 5% significance level, test whether there are any significant differences between the fixed effects and random effects estimates of the coefficients on *SMALL*, *AIDE*, *TCHEXPER*, *WHITE_ASIAN*, and *FREELUNCH*. What are the implications of the test outcomes? What happens if we apply the test to the fixed and random effects estimates of the coefficient on *BOY*?

f. Create school-averages of the variables and carry out the Mundlak test for correlation between them and the unobserved heterogeneity.

a.

The coefficient of SMALL is 5.82 > 0 and t-value = 5.886

⇒ The students perform better in reading when they are in small class

The coefficient of AIDE is 0.817 and t-value = 0.858

⇒ Teacher's aide does not efficiently improve the score

The coefficient of TCHEXPER is 0.492 ⇒ subtly affect the score.

The t-value of BOY and WHITEASIAN is −7.7 and 4.09

⇒ Sex and Race make the score different.

```
Call:
lm(formula = readscore ~ small + aide + tchexper + boy + whiteAsian +
    freelunch, data = theData)

Residuals:
     Min       1Q   Median       3Q      Max
-107.220  -20.214   -3.935   14.339  185.956

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 437.76425    1.34622 325.180  < 2e-16 ***
small         5.82282    0.98933   5.886 4.19e-09 ***
aide          0.81784    0.95299   0.858    0.391
tchexper      0.49247    0.06956   7.080 1.61e-12 ***
boy          -6.15642    0.79613  -7.733 1.23e-14 ***
whiteAsian    3.90581    0.95361   4.096 4.26e-05 ***
freelunch   -14.77134    0.89025 -16.592  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 30.19 on 5759 degrees of freedom
  (因為不存在，20 個觀察量被刪除了)
Multiple R-squared:  0.09685,	Adjusted R-squared:  0.09591
F-statistic: 102.9 on 6 and 5759 DF,  p-value: < 2.2e-16
```

```
Call:
lm(formula = readscore ~ small + aide + tchexper + boy + whiteAsian +
    freelunch + factor(schid), data = theData)

Residuals:
     Min       1Q   Median       3Q      Max
-102.638  -16.783   -2.847   12.759  198.417

Coefficients:
                     Estimate Std. Error t value Pr(>|t|)
(Intercept)         411.16263    4.12826  99.597  < 2e-16 ***
small                 6.49023    0.91296   7.109 1.31e-12 ***
aide                  0.99609    0.88169   1.130 0.258632
tchexper              0.28557    0.07084   4.031 5.63e-05 ***
boy                  -5.45594    0.72759  -7.499 7.44e-14 ***
whiteAsian            8.02802    1.53566   5.228 1.78e-07 ***
freelunch           -14.59357    0.88001 -16.583  < 2e-16 ***
factor(schid)123056  12.23625    5.44228   2.248 0.024591 *
factor(schid)128068  15.01494    5.37116   2.795 0.005200 **
factor(schid)128076   4.67988    5.30228   0.883 0.377481
factor(schid)128079  13.63444    5.25478   2.595 0.009492 **
factor(schid)130085   8.45528    5.05953   1.671 0.094745 .
factor(schid)159171  47.40846    4.73054  10.022  < 2e-16 ***
factor(schid)161176   6.02154    4.96504   1.213 0.225262
factor(schid)161183  33.83940    4.67460   7.239 5.12e-13 ***
factor(schid)162184  15.86799    5.26127   3.016 0.002573 **
factor(schid)164198  26.35726    5.36818   4.910 9.37e-07 ***
factor(schid)165199  41.46659    5.44990   7.609 3.22e-14 ***
factor(schid)166203  14.19143    5.11416   2.775 0.005539 **
factor(schid)168211  17.60864    4.78756   3.678 0.000237 ***
factor(schid)168214  31.58991    5.35414   5.900 3.84e-09 ***
factor(schid)169219  34.68703    5.55106   6.249 4.44e-10 ***
factor(schid)169229  34.05621    4.50933   7.552 4.95e-14 ***
factor(schid)169231   5.32569    5.36560   0.993 0.320966
factor(schid)169280  32.12661    5.39084   5.959 2.68e-09 ***
factor(schid)170295  33.34073    5.15981   6.462 1.12e-10 ***
factor(schid)173312  64.88834    5.30497  12.232  < 2e-16 ***
factor(schid)176329  37.52219    5.12963   7.315 2.94e-13 ***
factor(schid)180344  24.42050    4.79958   5.088 3.73e-07 ***
factor(schid)189378  10.85152    5.08696   2.133 0.032951 *
factor(schid)189382  28.43563    5.19413   5.475 4.57e-08 ***
factor(schid)189396   8.47877    5.22957   1.621 0.105006
factor(schid)191411  34.38757    5.65481   6.081 1.27e-09 ***
factor(schid)193422  34.27524    5.28596   6.484 9.67e-11 ***
factor(schid)193423  21.05474    5.02051   4.194 2.79e-05 ***
factor(schid)201449  36.46515    4.63921   7.860 4.56e-15 ***
factor(schid)203452  24.59026    4.83114   5.090 3.70e-07 ***
factor(schid)203457  43.11623    5.76070   7.485 8.28e-14 ***
factor(schid)205488  22.57067    5.40286   4.178 2.99e-05 ***
factor(schid)205489  30.00209    5.44649   5.509 3.78e-08 ***
factor(schid)205490   6.77322    5.26929   1.285 0.198700
factor(schid)205491  14.90854    5.12638   2.908 0.003649 **
factor(schid)205492  39.89450    5.00274   7.975 1.84e-15 ***
factor(schid)208501  17.89130    5.19956   3.441 0.000584 ***
factor(schid)208503   2.49135    5.27313   0.472 0.636615
factor(schid)209510  16.88248    4.82678   3.498 0.000473 ***
factor(schid)212522  23.30361    5.15247   4.523 6.23e-06 ***
factor(schid)215533  32.66486    4.61343   7.080 1.61e-12 ***
factor(schid)216536  14.75764    4.76204   3.099 0.001951 **
factor(schid)218562  30.58211    5.32494   5.743 9.77e-09 ***
factor(schid)221571  -3.09472    4.73503  -0.654 0.513409
factor(schid)221574   7.63646    5.23074   1.460 0.144367
factor(schid)225585   8.81323    4.98063   1.770 0.076864 .
factor(schid)228606  24.34759    4.97148   4.897 9.98e-07 ***
factor(schid)230612  36.46971    5.29092   6.893 6.06e-12 ***
factor(schid)231616  24.69067    5.32207   4.639 3.57e-06 ***
factor(schid)234628  24.26273    4.77028   5.086 3.77e-07 ***
factor(schid)244697  18.91941    4.81305   3.931 8.56e-05 ***
factor(schid)244708  15.00339    4.76130   3.151 0.001635 **
factor(schid)244723  14.01645    4.74131   2.956 0.003127 **
factor(schid)244727  27.91203    5.00774   5.574 2.61e-08 ***
factor(schid)244728  19.76377    5.53800   3.569 0.000362 ***
factor(schid)244736  39.81653    5.51900   7.214 6.12e-13 ***
factor(schid)244745  32.47486    4.96056   6.547 6.40e-11 ***
factor(schid)244746  36.41700    5.42541   6.712 2.10e-11 ***
factor(schid)244755  33.30654    4.58974   7.257 4.50e-13 ***
factor(schid)244764  29.42589    6.13291   4.798 1.64e-06 ***
factor(schid)244774  31.91029    4.81869   6.622 3.86e-11 ***
factor(schid)244776  27.96923    4.59533   6.086 1.23e-09 ***
factor(schid)244780  61.13317    5.31024  11.512  < 2e-16 ***
factor(schid)244796  23.15802    5.37706   4.307 1.68e-05 ***
factor(schid)244799  17.32434    5.23386   3.310 0.000939 ***
factor(schid)244801  13.37781    5.03098   2.659 0.007857 **
factor(schid)244806  52.06905    4.57187  11.389  < 2e-16 ***
factor(schid)244818  16.41957    4.93277   3.329 0.000878 ***
factor(schid)244831  17.44809    5.30436   3.289 0.001010 **
factor(schid)244839  33.91715    5.04460   6.723 1.95e-11 ***
factor(schid)252885  28.59116    5.11500   5.590 2.38e-08 ***
factor(schid)253888  18.52718    5.85047   3.167 0.001549 **
factor(schid)257899  12.05127    4.73657   2.544 0.010976 *
factor(schid)257905  40.39019    4.59866   8.783  < 2e-16 ***
factor(schid)259915  14.34775    5.30438   2.705 0.006853 **
factor(schid)261927  21.14747    4.92342   4.295 1.77e-05 ***
factor(schid)262937  45.47139    5.14882   8.831  < 2e-16 ***
factor(schid)264945  29.99884    4.80356   6.245 4.54e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 27.41 on 5681 degrees of freedom
  (因為不存在，20 個觀察量被刪除了)
Multiple R-squared:  0.2653,	Adjusted R-squared:  0.2544
F-statistic: 24.42 on 84 and 5681 DF,  p-value: < 2.2e-16
```

15.17 The data file *liquor* contains observations on annual expenditure on liquor (*LIQUOR*) and annual income (*INCOME*) (both in thousands of dollars) for 40 randomly selected households for three consecutive years.

a. Create the first-differenced observations on *LIQUOR* and *INCOME*. Call these new variables *LIQUORD* and *INCOMED*. Using OLS regress *LIQUORD* on *INCOMED* without a constant term. Construct a 95% interval estimate of the coefficient.

b. Estimate the model $LIQUOR_{it} = \beta_1 + \beta_2 INCOME_{it} + u_i + e_{it}$ using random effects. Construct a 95% interval estimate of the coefficient on *INCOME*. How does it compare to the interval in part (a)?

c. Test for the presence of random effects using the LM statistic in equation (15.35). Use the 5% level of significance.

d. For each individual, compute the time averages for the variable *INCOME*. Call this variable *INCOMEM*. Estimate the model $LIQUOR_{it} = \beta_1 + \beta_2 INCOME_{it} + \gamma INCOMEM_i + c_i + e_{it}$ using the random effects estimator. Test the significance of the coefficient γ at the 5% level. Based on this test, what can we conclude about the correlation between the random effect u_i and *INCOME*? Is it OK to use the random effects estimator for the model in (b)?

a.

```
> confint(modA)
                  2.5 %    97.5 %
(Intercept) -1.85305193 0.2739794
incomeD     -0.04196685 0.3447086
```

b.

The coefficient of TCHEXPER become much lower than previous one,

and the coefficient of WHITEASIAN become much greater.

c. Since the F-statistic = 24.42 and p-value $< 2.2\times10^{-16}$ ⟹ statistically significant

If the explanatory variables are uncorrelated with school-level effects,

then including fixed effects will not chang the coefficients much.