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?

f. Column (5) contains the random effects estimates. Which coefficients, apart from the intercepts, show the most difference from the fixed effects estimates? Use the Hausman test statistic (15.36) to test whether there are significant differences between the random effects estimates and the fixed effects estimates in column (3) (Why that one?). Based on the test results, is random effects estimation in this model appropriate?

f. ① The coefficient of SOUTH show the most difference

② $\dfrac{-0.3261 + 0.2326}{\sqrt{0.1258^2 - 0.0317^2}} = -0.768 \Rightarrow$ p-valu=a44248

$\Rightarrow$ prefer fixed effects model.

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)?

b. The interval in part (a) is wider

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

```
> confint(mod_RE)
                  2.5 %     97.5 %
(Intercept) -0.05211904 1.99018381
income       0.01283111 0.04031983
```

c.

```
> plmtest(mod_pool, type="bp")

        Lagrange Multiplier Test - (Breusch-Pagan)

data:  liquor ~ income
chisq = 20.68, df = 1, p-value = 5.429e-06
alternative hypothesis: significant effects
```

d.

The coefficiet is significant $\Rightarrow$ The correlation is not zero, prefer to use fixed effect estimator

```
> coeftest(modD, vcov=vcovHC(modD, type="HC1"))

t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.9163337  0.5657538  1.6197   0.1080
income      0.0207421  0.0194273  1.0677   0.2879
incomeM     0.0065792  0.0206317  0.3189   0.7504
```

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.

d. It is Not significant, the variables might not be correlated with the school effects.

```
> plmtest(modD, type="bp")

        Lagrange Multiplier Test - (Breusch-Pagan)

data:  readscore ~ small + aide + tchexper + boy + whiteAsian + freelunch
chisq = 6677.4, df = 1, p-value < 2.2e-16
alternative hypothesis: significant effects
```

```
> summary(modA)

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
```

```
> summary(modB)
Oneway (individual) effect Within Model

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + whiteAsian +
    freelunch, data = theData, model = "within", index = c("schid",
    "id"))

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-102.6381  -16.7834   -2.8473   12.7591  198.4169

Coefficients:
             Estimate Std. Error  t-value  Pr(>|t|)
small        6.490231   0.912962   7.1090 1.313e-12 ***
aide         0.996087   0.881693   1.1297    0.2586
tchexper     0.285567   0.070845   4.0309 5.629e-05 ***
boy         -5.455941   0.727589  -7.4987 7.440e-14 ***
whiteAsian   8.028019   1.535656   5.2277 1.777e-07 ***
freelunch  -14.593572   0.880006 -16.5835 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    4628000
Residual Sum of Squares: 4268900
R-Squared:      0.077592
Adj. R-Squared: 0.063954
F-statistic: 79.6471 on 6 and 5681 DF, p-value: < 2.22e-16
```

```
> summary(modD)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + whiteAsian +
    freelunch, data = theData, model = "random", index = c("schid",
    "id"))

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Effects:
                 var std.dev share
idiosyncratic 751.43   27.41 0.829
individual    155.31   12.46 0.171
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6470  0.7225  0.7523  0.7541  0.7831  0.8153

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-97.483 -17.236  -3.282   0.037  12.803 192.346

Coefficients:
              Estimate Std. Error  z-value  Pr(>|z|)
(Intercept) 436.126774   2.064782 211.2217 < 2.2e-16 ***
small         6.458722   0.912548   7.0777 1.466e-12 ***
aide          0.992146   0.881159   1.1260    0.2602
tchexper      0.302679   0.070292   4.3060 1.662e-05 ***
boy          -5.512081   0.727639  -7.5753 3.583e-14 ***
whiteAsian    7.350477   1.431376   5.1353 2.818e-07 ***
freelunch   -14.584332   0.874676 -16.6740 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    6158000
Residual Sum of Squares: 4332100
R-Squared:      0.29655
Adj. R-Squared: 0.29582
Chisq: 493.205 on 6 DF, p-value: < 2.22e-16
```

```
> phtest(modB, modD)

        Hausman Test

data:  readscore ~ small + aide + tchexper + boy + whiteAsian + freelunch
chisq = 13.809, df = 6, p-value = 0.03184
alternative hypothesis: one model is inconsistent
```

e. There is no significant difference.

This implies the fixed effects and random effects model have similar estimations.

Test for BOY:

	estimation	standard error
fixed:	-6.455	0.727589
random:	-5.512	0.72763

$$\Rightarrow \frac{-5.512+5.455}{\sqrt{(0.727589)^2+(0.72763)^2}} = -0.0553938$$

Thus, No significant difference, the unobserved school effects are not correlated with gender.

f.

```
> summary(modF)
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)

Call:
plm(formula = theData$readscore ~ theData$small + theData$aide +
    theData$tchexper + theData$boy + theData$whiteAsian + theData$freelunch +
    mean_small + mean_aide + mean_tchexper + mean_boy + mean_whiteAsian +
    mean_freelunch, data = theData, model = "random", index = c("schid",
    "Id"))

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Effects:
                  var std.dev share
idiosyncratic 895.919  29.932 0.987
individual     11.617   3.408 0.013
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.1669  0.2445  0.2856  0.2941  0.3364  0.3998

Residuals:
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
-107.552  -20.047   -3.918    0.032   14.221  183.967

Coefficients:
                      Estimate Std. Error  z-value  Pr(>|z|)
(Intercept)         446.753370   7.688308  58.1082 < 2.2e-16 ***
theData$small         5.673677   0.985620   5.7565 8.590e-09 ***
theData$aide          0.823488   0.948317   0.8684   0.38519
theData$tchexper      0.488510   0.069213   7.0581 1.688e-12 ***
theData$boy          -6.302558   0.794033  -7.9374 2.065e-15 ***
theData$whiteAsian    4.260684   0.987020   4.3167 1.584e-05 ***
theData$freelunch   -14.750280   0.891166 -16.5517 < 2.2e-16 ***
mean_small           -8.254292   8.074338  -1.0223   0.30665
mean_aide             2.187658   7.574176   0.2888   0.77271
mean_tchexper        -0.467995   0.227411  -2.0579   0.03960 *
mean_boy            -14.330275   9.637735  -1.4869   0.13704
mean_whiteAsian       5.529475   2.266007   2.4402   0.01468 *
mean_freelunch        1.085569   3.223253   0.3368   0.73627
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    9074000
Residual Sum of Squares: 5161000
R-Squared:      0.43132
Adj. R-Squared: 0.43013
Chisq: 627.07 on 12 DF, p-value: < 2.22e-16
```