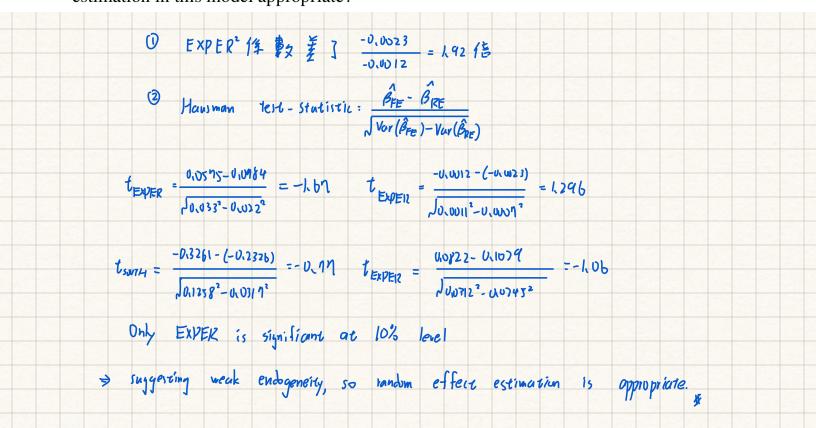
15.6 Using the NLS panel data on N = 716 young women, we consider only years 1987 and 1988. We are interested in the relationship between ln(WAGE) and experience, its square, and indicator variables for living in the south and union membership. Some estimation results are in Table 15.10.

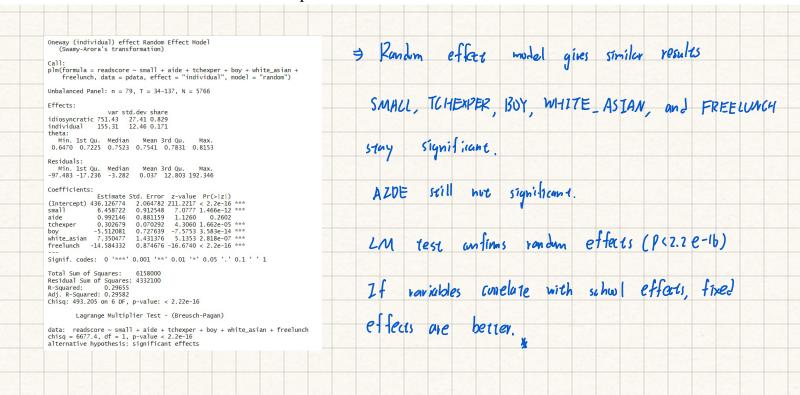
TABLE 15.10 Estimation Results for Exercise 15.6

	(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

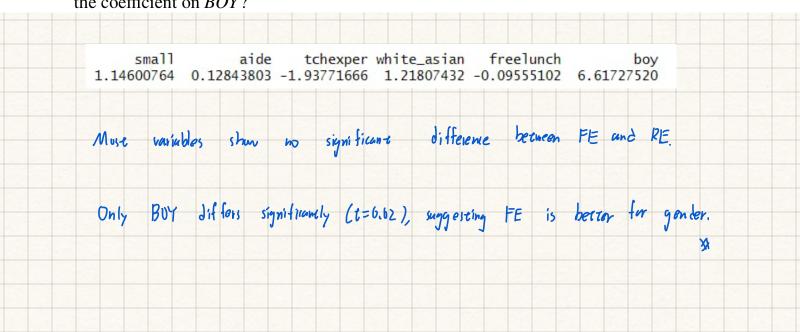
f. Column (5) contains the random effects estimates. Which coefficients, apart from the intercepts, show the most difference from the fixed effects estimates? Use the Hausman test statistic (15.36) to test whether there are significant differences between the random effects estimates and the fixed effects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



- 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*.
 - **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.

```
Pooling Model
Call:
Call:

plm(formula = readscore ~ small_within + aide_within + tchexper_within +

boy_within + white_asian_within + freelunch_within + small_mean +

aide_mean + tchexper_mean + boy_mean + white_asian_mean +

freelunch_mean, data = pdata, model = "pooling")
Unbalanced Panel: n = 78, T = 34-137, N = 5702
Min. 1st Qu.
-106.5793 -19.8022
                        Median 3rd Qu. Max.
-3.8505 14.8261 191.1282
Coefficients:
                        Estimate Std. Error t-value Pr(>|t|)
                                     5.470575 84.8319 < 2.2e-16 ***
(Intercept)
                      464.079450
small_within
                        6.561772
                                     1.003668
                                                6.5378 6.788e-11 ***
                                                1.1282 0.2592893
3.7737 0.0001625 ***
-6.7633 1.484e-11 ***
aide_within
                        1.092017
                                     0.967941
tchexper_within
                       0.294604
                                     0.078067
tcnexper_witnin 0.294604
boy_within -5.408590
white_asian_within 8.196367
freelunch_within -14.642487
small_mean -15.989792
                                     0.799695
                                     1.685491
                                                4.8629 1.188e-06 ***
                                     0.969397 -15.1047 < 2.2e-16 ***
5.736201 -2.7875 0.0053289 **
aide_mean
                      11.980924
                                     5.272676
                                                 2.2723 0.0231076 *
                                     0.163393 6.8544 7.917e-12 ***
7.117671 -7.9586 2.084e-15 ***
1.640677 0.7152 0.4745371
tchexper_mean
                        1.119963
boy_mean
white_asian_mean
                     -56.646562
1.173353
                     -18.271502
freelunch_mean
                                    2.340331 -7.8072 6.911e-15 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
                            5781700
Total Sum of Squares:
Residual Sum of Squares: 5105300
R-Squared: 0.11699
R-Squared: 0.11699
Adj. R-Squared: 0.11513
F-statistic: 62.8119 on 12 and 5689 DF, p-value: < 2.22e-16
                      Mundlak test indicates that several school-level means, such as
            small_men, tehexper_mean, and free lunch_meun, one
                                                                                                             statiscally
         This provides scrong evidence that these variables are concluded with
                                                                                                 random effects mudel
       unobserved subout effects, implying there
                                                                                       the
         inconsistent.
         Therefore, the fixed effects model is prefared for consistent estimation.
```

- **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.
 - **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)?

```
Oneway (individual) effect Random Effect Model
   (Swamy-Arora's transformation)
plm(formula = liquor ~ income, data = pdata, model = "random")
Balanced Panel: n = 40, T = 3, N = 120
Effects:
               var std.dev share
idiosyncratic 0.9640 0.9819 0.571
individual 0.7251 0.8515 0.429
theta: 0.4459
Residuals:
          1st Qu. Median 3rd Qu.
-2.263634 -0.697383 0.078697 0.552680 2.225798
Coefficients:
            Estimate Std. Error z-value Pr(>|z|)
(Intercept) 0.9690324 0.5210052 1.8599 0.0628957
          0.0265755 0.0070126 3.7897 0.0001508 ***
income
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
Total Sum of Squares:
                      126.61
Residual Sum of Squares: 112.88
R-Squared:
               0.1085
Adj. R-Squared: 0.10095
Chisq: 14.3618 on 1 DF, p-value: 0.00015083
                2.5 % 97.5 %
(Intercept) -0.05211904 1.99018381
            0.01283111 0.04031983
            LIQUOR; = 0.969 + 0,0266 INCOMERC + Ui + Ciz
      3
           95% (2 = [0,0128,0,0401]
      =) (umpried to (a); Unlike the first-difference model, the effect is
                              significant and the (I excludes zero
```

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)?

```
Oneway (individual) effect Random Effect Model
  (Swamy-Arora's transformation)
plm(formula = liquor ~ income + INCOMEM, data = pdata, model = "random")
Balanced Panel: n = 40, T = 3, N = 120
Effects:
var std.dev share
idiosyncratic 0.9640 0.9819 0.571
individual 0.7251 0.8515 0.429
theta: 0.4459
Residuals:
Min. 1st Qu. Median 3rd Qu. Max.
-2.300955 -0.703840 0.054992 0.560255 2.257325
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1
Total Sum of Squares:
                     126.61
Residual Sum of Squares: 112.79
R-Squared:
             0.10917
Adj. R-Squared: 0.093945
Chisq: 14.3386 on 2 DF, p-value: 0.00076987
                                              is not statistically significant (p=0, non)
              wefficient un ZNCOME
       suggesting no evidence of correlation between income and the individual
       rundom effect. Therefore, it is appropriate to use the random effects
      estimutur in (b).
```