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?

① $EXPER^2$ 係數差了 $\frac{-0.0023}{-0.0012} = 1.92$ 倍

② Hausman test-statistic: $\frac{\hat{\beta}_{FE} - \hat{\beta}_{RE}}{\sqrt{Var(\hat{\beta}_{FE}) - Var(\hat{\beta}_{RE})}}$

$t_{EXPER} = \frac{0.0575 - 0.0984}{\sqrt{0.033^2 - 0.022^2}} = -1.67$ $\quad t_{EXPER2} = \frac{-0.0012 - (-0.0023)}{\sqrt{0.0011^2 - 0.0007^2}} = 1.296$

$t_{SOUTH} = \frac{-0.3261 - (-0.2326)}{\sqrt{0.1258^2 - 0.0317^2}} = -0.77$ $\quad t_{EXPER2} = \frac{0.0822 - 0.1029}{\sqrt{0.0312^2 - 0.0245^2}} = -1.06$

Only EXPER is significant at 10% level

⇒ suggesting weak endogeneity, so random effect estimation is 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.

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

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = pdata, effect = "individual", model = "random")

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 ***
white_asian   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

        Lagrange Multiplier Test - (Breusch-Pagan)

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

⇒ Random effect model gives similar results

SMALL, TCHEXPER, BOY, WHITE_ASIAN, and FREELUNCH stay significant.

AIDE still not significant.

LM test confirms random effects ($p < 2.2e-16$)

If variables correlate with school effects, fixed effects are better.

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*?

```
     small        aide    tchexper white_asian   freelunch         boy
1.14600764  0.12843803 -1.93771666  1.21807432 -0.09555102  6.61727520
```

Most variables show no significant difference between FE and RE.

Only BOY differs significantly ($t = 6.62$), suggesting FE is better for gender.

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

```
Pooling Model

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

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-106.5793  -19.8022   -3.8505   14.8261  191.1282

Coefficients:
                     Estimate Std. Error  t-value  Pr(>|t|)
(Intercept)        464.079450   5.470575  84.8319 < 2.2e-16 ***
small_within         6.561772   1.003668   6.5378 6.788e-11 ***
aide_within          1.092017   0.967941   1.1282 0.2592893
tchexper_within      0.294604   0.078067   3.7737 0.0001625 ***
boy_within          -5.408590   0.799695  -6.7633 1.484e-11 ***
white_asian_within   8.196367   1.685491   4.8629 1.188e-06 ***
freelunch_within   -14.642487   0.969397 -15.1047 < 2.2e-16 ***
small_mean         -15.989792   5.736201  -2.7875 0.0053289 **
aide_mean           11.980924   5.272676   2.2723 0.0231076 *
tchexper_mean        1.119963   0.163393   6.8544 7.917e-12 ***
boy_mean           -56.646562   7.117671  -7.9586 2.084e-15 ***
white_asian_mean     1.173353   1.640677   0.7152 0.4745371
freelunch_mean     -18.271502   2.340331  -7.8072 6.911e-15 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:    5781700
Residual Sum of Squares: 5105300
R-Squared:      0.11699
Adj. R-Squared: 0.11513
F-statistic: 62.8119 on 12 and 5689 DF, p-value: < 2.22e-16
```

⇒ The Mundlak test indicates that several school-level means, such as small_mean, tchexper_mean, and freelunch_mean, are statistically significant.

This provides strong evidence that these variables are correlated with unobserved school effects, implying that the random effects model is inconsistent.

Therefore, the fixed effects model is prefered 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)

Call:
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:
     Min.   1st Qu.    Median   3rd Qu.      Max.
-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 .
income      0.0265755  0.0070126  3.7897 0.0001508 ***
---
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
income       0.01283111 0.04031983
```

$\widehat{LIQUOR}_{it} = 0.969 + 0.0266\, INCOME_{it} + u_i + e_{it}$

$\Rightarrow$

95% CI = [0.0128, 0.0401]

$\Rightarrow$ Compared to (a): Unlike the first-difference model, the effect is significant and the CI excludes zero

*

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

By R code $\Rightarrow$ LM-statistic $= 4.5475 > 1.645 = Z_{0.05}$

$\Rightarrow$ reject H_0 $\Rightarrow$ random effects are present.

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)

Call:
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

Coefficients:
             Estimate Std. Error z-value Pr(>|z|)
(Intercept) 0.9163337  0.5524439  1.6587  0.09718 .
income      0.0207421  0.0209083  0.9921  0.32117
INCOMEM     0.0065792  0.0222048  0.2963  0.76700
---
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
```

$\Rightarrow$ The coefficient on INCOME is not statistically significant ($p = 0.767$), suggesting no evidence of correlation between income and the individual random effect. Therefore, it is appropriate to use the random effects estimator in (b).