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.

	(1) OLS 1987	(2) OLS 1988	(3) FE	(4) FE Robust	(5) RE	(3) - (5) Difference
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	0.0411
	(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)	(0.0011)	(0.0007)	
SOUTH	−0.2128	−0.2384	−0.3261	−0.3261	−0.2326	0.0935 ✓
	(0.0338)	(0.0344)	(0.1258)	(0.2495)	(0.0317)	
UNION	0.1445	0.1102	0.0822	0.0822	0.1027	0.0205
	(0.0382)	(0.0387)	(0.0312)	(0.0367)	(0.0245)	
N	716	716	1432	1432	1432	

(standard errors in parentheses)

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?

SOUTH 的係数差最多

H_0: There's no endogeneity ($\beta_{FE} = \beta_{RE}$)

H_1: There's endogeneity ($\beta_{FE} \neq \beta_{RE}$)

$$t_{EXPER} = \frac{0.0575 - 0.0986}{\sqrt{0.033^2 - 0.022^2}} = -1.67$$

$$t_{EXPER^2} = \frac{-0.0012 + 0.0023}{\sqrt{0.0011^2 - 0.0007^2}} = 1.3$$

$$t_{SOUTH} = \frac{-0.3261 + 0.2326}{\sqrt{0.1258^2 - 0.0317^2}} = -0.77$$

$$t_{UNION} = \frac{0.0822 - 0.027}{\sqrt{0.0312^2 - 0.0245^2}} = -1.06$$

⇒只有EXPER在10%显著水準下有显著差異

∴ random effect estimation is appropriate.

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.

```
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
```

95% CI : 0.0122 0.0409

c.

```
Lagrange Multiplier Test - (Breusch-Pagan)

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

∵ p-value < 0.05 ∴ We reject H_0, means there's no individual random effects.

d.

```
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
```

$$\widehat{LIQUORD}_{it} = 0.9163337 + 0.0207421\, income + \overset{\gamma}{0.0065792}\, INCOMED_{it}$$

∵ γ is not significantly different from 0,

∴ There's no evidence to show that the individaual random effect c_i are correlated with $INCOME_{it}$

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*.

✗ 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?

✗ 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.]

✗ 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.

```
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = newdata, 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 ***

        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
```

Most coefficients in the RE model closely match those from the pooled OLS & FE. This suggests that there's little correlation with unobserved school-level heterogeneity.

e.

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

small       :  t =  1.15,  p = 0.252
aide        :  t =  0.13,  p = 0.898
tchexper    :  t = -1.94,  p = 0.053
white_asian :  t =  1.22,  p = 0.223
freelunch   :  t = -0.10,  p = 0.924
```

∵ p-value < 0.05 & 13.809 > $\chi^2_{0.95,6}$ = 12.59

∴ We reject H_0, means that there's correlation between unobserved school effects & our regressors. Random effect is therefore inconsistent, and the fixed-effects estimator is perfect.

f.

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

Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch + small_m + aide_m + tchexper_m + boy_m + white_asian_m +
    freelunch_m, data = pdata_clean, model = "random")

Unbalanced Panel: n = 78, T = 34-136, N = 5681

Effects:
                 var std.dev share
idiosyncratic 756.11   27.50 0.817
individual    169.40   13.02 0.183
theta:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.6593  0.7327  0.7615  0.7630  0.7892  0.8217

Residuals:
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-98.886 -17.051  -3.166   0.039  12.846 193.321

Coefficients:
               Estimate Std. Error  z-value  Pr(>|z|)
(Intercept)  459.462989  20.529888  22.3802 < 2.2e-16 ***
small          6.637460   0.922068   7.1985 6.090e-13 ***
aide           1.157620   0.889542   1.3014    0.1931
tchexper       0.289286   0.071754   4.0316 5.539e-05 ***
boy           -5.386109   0.735063  -7.3274 2.346e-13 ***
white_asian    8.081423   1.550155   5.2133 1.855e-07 ***
freelunch    -14.699025   0.892109 -16.4767 < 2.2e-16 ***
small_m      -18.410060  22.273923  -0.8265    0.4085
aide_m        16.811358  20.793685   0.8085    0.4188
tchexper_m     1.006007   0.625690   1.6078    0.1079
boy_m        -53.353521  25.221654  -2.1154    0.0344 *
white_asian_m -6.648191   6.320012  -1.0519    0.2928
freelunch_m   -3.318853   8.779553  -0.3780    0.7054
```

Among all school-level average variables, only student gender (boy_m) is significant at 5% level (p=0.0344), means a significant correlation with school-specific effects (individual effect).

This suggest that the variable boy violates the exogeneity assumption, and the estimation results may be biased.