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

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

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?

6

(a) 兩年的係數值相近，表示 ln(WAGE) 與解釋變數間關係較穩定

並假設回歸無異質性

(b)

panel data regression 的殘差是 $(u_i + e_{it})$，

相比 OLS 多考慮了個體的 u_i

(c) EXPER & $EXPER^2$ 項　　　　C15Q6(d, e)

✓(d) $df_1 = N - 1 = 715$ → 2年觀察值

$df_2 = 1432 - 716 - 4 = 712$

$F_{0.99}(715, 712) > 1.36$

$F = 11.68$, $F_{0.99}(715, 712) \approx 1.19$

$11.68 > 1.19$. reject H_0.

因此使用 FE model 較合適，因資料的個體異質性

✓

(e) cluster standard error 較大.

因為其控制了個體內的異質性

與自我相關

(f) $t_{Exper} = \dfrac{0.0575 - 0.0986}{\sqrt{0.033^2 - 0.022^2}} = -1.67$

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

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

$t_{union} = \dfrac{0.0822 - 0.1027}{\sqrt{0.0312^2 - 0.0245^2}} = -1.06$

⇒ EXPER t 值最大，但其結果在統計上

不顯著，FE、RE 無明顯差異.

17. (a)

estimated regression with differented data is

$$\widehat{LIQUAD_{it}} = 0.02975\ INCOMED_{it}$$
$$(se) \qquad (0.02922)$$

The 95% interval estimate of the coefficient of Income is [-0.02841, 0.0879]

The interval covers zero. we have no evidence against the hypothesis that income doesn't affect liquor expenditure

```
95% 信賴區間：
> cat("          2.5 %          97.5 %\n")
          2.5 %          97.5 %
> cat(formatted_conf_int, sep = "\n")
incomed    -0.02841457 0.08790818
> 
```

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)

```
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 ***
white_asian   3.90581    0.95361   4.096 4.26e-05 ***
freelunch   -14.77134    0.89025 -16.592  < 2e-16 ***
```

small 正顯著，小班平均分較高 (5.82分)

aide. 不顯著

tchexper 正顯著，experienced teacher 分數較高

boy 為負顯著，男生分數較低

white_asian 為正且顯著、白人 or 亞洲人平均分較多

(b)

```
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 ***
white_asian   8.028019   1.535656   5.2277 1.777e-07 ***
freelunch   -14.593572   0.880006 -16.5835 < 2.2e-16 ***
```

控制FE後，small 仍顯著，小班分數較高，

tchexper 係數則下降

white_asian 係數提升很多且有顯著性

(c) $F=16.698$，學校的 FE 在統計上顯著