CH15 Q6

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?

(a) 兩年的OLS結果相近，且假設所有個体之參數相同，每年視為獨立的迴歸.

(b) 這個模型假設殘差為 $(u_i + e_{it})$，假設每個人有一個個体效果 u_i。而橫斷 OLS 假設 $u_i = 0$，且每年互不相關.

(c)

EXPER 與 $EXPER^2$ 變化最明顯，因為 FE 僅以「同一人」跨期變化辨識

(d) $df_1 = N-1 = 715$, $df_2 = NT-N-K = 716 \times 2 - 716 - 4 = 712$

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

$11.68 > 1.36$ ，在 1% level 下 reject no individul differences 的假設，表示個体 FE 顯著

(e)

與原本 FE 的 SE 相比，大部分變數的 SE 都變大，Robust FE 只要求不同個体之間獨立，同一個体在不同時点可以相關

(f)

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

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

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

$$t_{UNION} = \frac{0.0822 - 0.1027}{\sqrt{0.0312^2 - 0.0245^2}} = -1.06$$

EXPER |t|值最大，但仍小於 1.96 ，結果不顯著，代表 FE 和 RE 無明顯差異。

CH15 Q17

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.

(a)

```
Call:
lm(formula = LIQUORD ~ 0 + INCOMED, data = df_diff)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6852 -0.9196 -0.0323  0.9027  3.3620 

Coefficients:
        Estimate Std. Error t value Pr(>|t|)
INCOMED  0.02975    0.02922   1.018    0.312

Residual standard error: 1.417 on 79 degrees of freedom
Multiple R-squared:  0.01295,	Adjusted R-squared:  0.0004544 
F-statistic: 1.036 on 1 and 79 DF,  p-value: 0.3118
```

```
> confint(mod, level = 0.95)
              2.5 %     97.5 %
INCOMED -0.02841457 0.08790818
```

CH15 Q20

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?

(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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

small 係數為正，分配到小班平均多 5.8分

aide 係數不顯著

tchexper 為正，平均增加 0.49分

boy 為負，男生平均少 6.16分

white_asian 為正，代表白人或亞裔平均多3.9分

(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 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

控制學校的 Fixed effect 之後

small 效果仍顯著，tchexper 係數下降且顯著性下降

white_asian 係數增加

(c)

```
        F test for individual effects

data:  readscore ~ small + aide + tchexper + boy + white_asian + freelunch
F = 16.698, df1 = 78, df2 = 5681, p-value < 2.2e-16
alternative hypothesis: significant effects
```

學校固定效果顯著。

當主要自變數在學校之間的分佈相近，或

效果來自於校內的隨機分配（內生性低），

則即使加上學校固定效果影響也很小。