

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)	(2)	(3)	(4)	(5)
	OLS 1987	OLS 1988	FE	FE Robust	RE
C	0.9348 (0.2010)	0.8993 (0.2407)	1.5468 (0.2522)	1.5468 (0.2688)	1.1497 (0.1597)
$EXPER$	0.1270 (0.0295)	0.1265 (0.0323)	0.0575 (0.0330)	0.0575 (0.0328)	0.0986 (0.0220)
$EXPER^2$	-0.0033 (0.0011)	-0.0031 (0.0011)	-0.0012 (0.0011)	-0.0012 (0.0011)	-0.0023 (0.0007)
$SOUTH$	-0.2128 (0.0338)	-0.2384 (0.0344)	-0.3261 (0.1258)	-0.3261 (0.2495)	-0.2326 (0.0317)
$UNION$	0.1445 (0.0382)	0.1102 (0.0387)	0.0822 (0.0312)	0.0822 (0.0367)	0.1027 (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

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

- c. Explain any differences in assumptions between this model and the models in part (a). 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. OLS 在 1987 和 1988 的估計差異不大
個體間無異質性

b. 加入時間對各變數的影響

c. 截距差異最大從約 0.9 上升至 1.546

d. $F = 11.68 > 1.2 = F_{0.05}(715, 711)$
reject H_0 , 支持使用固定效果

e. within transformation $\tilde{e}_{it} = e_{it} - \bar{e}_i$
column (4) 標準誤變大

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.

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

a. $[-0.0284, 0.0819]$

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

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
```

$p\text{-value} < 0.05$, reject H_0

表示學校間存在差異, 學校固定效果在模型中不可忽略

a.

```
Call:
lm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = star)

Residuals:
    Min       1Q   Median       3Q      Max
-107.220  -20.214   -3.935   14.339   185.956

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

Residual standard error: 30.19 on 5759 degrees of freedom
(因為不存在 20 個觀察量被刪除了)
Multiple R-squared:  0.09685,    Adjusted R-squared:  0.09591
F-statistic: 102.9 on 6 and 5759 DF,  p-value: < 2.2e-16
```

small 係數為 5.82 且顯著, 表示小班學生分數顯著較高

aide 係數 0.82, 不顯著, 表示無明顯證據顯示教師助理能提升成績

tchexper 係數 0.49, 且顯著, 表示教師經驗越多, 學生成績越好

boy 係數 -6.15, 男生成績明顯低於女生, 顯著性高

white_asian +3.91, 白人/亞裔學生顯著優於其它族群

b.

```
Call:
plm(formula = readscore ~ small + aide + tchexper + boy + white_asian +
    freelunch, data = star_panel, model = "within")

Unbalanced Panel: n = 79, T = 34-137, N = 5766

Residuals:
    Min.    1st Qu.    Median     3rd Qu.     Max.
-102.6381 -16.7834  -2.8473  12.7591  198.4169

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

Total Sum of Squares:    4628000
Residual Sum of Squares: 4268900
R-Squared:              0.077592
Adj. R-Squared:         0.063954
F-statistic: 79.6471 on 6 and 5681 DF, p-value: < 2.22e-16
```

small 仍顯著為正

tchexper, boy, white_asian 依然影響成績

領免費午餐的學生成績較差

white_asian 係數在固定效果模型中上升, 顯示相同學校族群差異更大

tchexper 效果變小, 代表效果來自學校差異, 而非學校內部差異