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)

V **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)?

V **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).

V **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?

V **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.

V **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?

V **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?

(f)

The test can be carried out coefficient by coefficient using a *t*-test, or jointly, using a chi-square test. Let us consider the *t*-test first. Denote the fixed effects estimate of β_k as $b_{FE,k}$, and let the random effects estimate be $b_{RE,k}$. Then the *t*-statistic for testing that there is no difference between the estimators, and thus that there is no correlation between u_i and any of the explanatory variables, is

$$t = \frac{b_{FE,k} - b_{RE,k}}{\left[\widehat{\operatorname{var}}(b_{FE,k}) - \widehat{\operatorname{var}}(b_{RE,k})\right]^{1/2}} = \frac{b_{FE,k} - b_{RE,k}}{\left[\operatorname{se}(b_{FE,k})^2 - \operatorname{se}(b_{RE,k})^2\right]^{1/2}} \quad (15.36)$$

	difference ($\beta_{FE}-\beta_{RE}$)	$\sqrt{SE_{FE}^2 - SE_{RE}^2}$	t	
EXPER	$0.0575 - 0.0986 = -0.0411$	0.0246	−1.67	10%.
EXPER^2	$-0.0012 - (-0.0023) = 0.0011$	0.00085	1.30	
SOUTH	$-0.3261 - (-0.2326) = -0.0935$	0.1216	−0.77	
UNION	$0.0822 - 0.1027 = -0.0205$	0.0193	−1.06	

若保守分析，使用10% sig. level 測試，則EXPER項顯著，拒絕H_0: 所有explainable variables 和 u_i 不相關，所以建議採用FE；若是under 10% level，則可使用RE.(效率較高)

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)?

樣本資訊量增加

(b) 95% C.I. $= [0.013, 0.040]$ → 因為RE考量到個体差異,所以C.I. 变窄,提供更精準的證據 $\beta_2 \neq 0$

(a) $[-0.028, 0.088]$

(c) $\chi^2 = 20.68 > \chi^2_{1,0.05} = 3.84 \Rightarrow$ reject $H_0: u_i = 0$ (不存在隨机效果)

∴ in 5% level of sig., RE exists, we should use random effects models rather than pure pooled-OLS.

(d)

```
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
```

incomem係數不顯著,故 can't reject $H_0: \gamma = 0$

→ There's no evidence that the random effect c_i is correlated with incomem → 符合random effect假設

→ (b)的model沒問題

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.

(d)

```
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
```

```
        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
```

→ 拒絕 $H_0: \sigma_u^2 = 0$，即 $\sigma_u^2 > 0$，故納入 RE

(e)

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

boy NAN.

皆 < 1.96，故無法拒絕 H_0 → no endogeniety，故可使用 RE model.

唯 boy ∵ $SE_{FE} < SE_{RE}$，違反 RE 假設 x_{it} 和 u_{it} 不相關，所以要用 FE

(f)

```
  Res.Df Df      F  Pr(>F)
1   5695
2   5689  6 2.2541 0.03557 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
>
```

Mundlak $\begin{cases} H_0: \gamma_i = 0 \\ H_1: \gamma_i \neq 0 \end{cases}$ reject H_0

$\rightarrow$ random effect is inappropriate