$v_{i2} = u_i + e_{i2}$. Is this correlation relatively high, or relatively low?

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.

影响較小 → 誤差結構較乾淨

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).

a. 雖然各係數在1987, 1988年差別不大，但因為這是Panel Data，每個個体之間有固定的差異，而原假設是個体的差異被視為誤差，所以須加入固定效果控制個体的變異（能力教育背景……）。

b. a的OLS模型忽略了個体之間的差異，b加入u_i項控制住被遺漏的個體差異（cross-section, time varient）。

c. 可以觀察到EXPER原本被高估，可能源自於未控制住本來能力就好的人（比較早找到工作）EXPER^{2}絕對值減少，代表曲線變平緩SOUTH變數變負的更多，說明原本沒有控制到家庭背景等因素UNION則變得更小。

d.

分子自由度 $J = N - 1 = 716 - 1 = 715$ → exper, exper², south, union

分母自由度 $NT - N - K_S = 716 \times 2 - 716 - 4 = 712$

$F_{0.99}(715, 712) \approx 1.17$

$\because F = 11.08 > 1.17 \therefore$ reject $H_0: \beta_1 = \beta_2 = \cdots = \beta_K$

代表各個体之間的截距項不同

→ 應該採用固定效果控制個体不隨時間變動的異質性

e.

一般 FE SE 假設 e_i 各獨立，且滿足 homoskedasticity

Cluster-robust FE SE 允許 e_i 各相關且允許異質性（SE一般較大）

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.

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)

	2.5 %	97.5 %
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?

(a)

```
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 ***
```

小班制、教師經驗、白人/亞裔有助於提升學生閱讀成績

(b)

```
Coefficients:
             Estimate Std. Error  t-value  Pr(>|t|)
small        6.490231 大 0.912962   7.1090 1.313e-12 ***
aide         0.996087 不 0.881693   1.1297    0.2586
tchexper     0.285567 小 0.070845   4.0309 5.629e-05 ***
boy         -5.455941 小 0.727589  -7.4987 7.440e-14 ***
white_asian  8.028019 大 1.535656   5.2277 1.777e-07 ***
freelunch  -14.593572 小 0.880006 -16.5835 < 2.2e-16 ***
---
```

(c)

$F = 16.698$, p-value $< 2.2 \times 10^{-16}$

∴ reject H_0, fixed effect應該納入模型