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)
EXPER²	−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

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

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) The results seem consistent between 1987 and 1988, also it seems that the standard errors are similar on 1987 and 1988, showing there might not be heterogeneity.

(b) Every variables now has time varying errors, and also a time invariant error u_i.

(c) The most difference: SOUTH from $-0.2128 \to -0.3261$. -0.1133.

(d) $F = 11.68 > 1.2 = F_{0.05}(715, 711)$

⇒ reject H_0, we should use fixed effects model

(e) within transformation: $\tilde{e}_{it} = e_{it} - \bar{e}_{it}$

⇒ all has larger se.

(f) $b_{FE,EXPER} - b_{RE,EXPER} = -0.0411$

$b_{FE,EXPER^2} - b_{RE,EXPER^2} = 0.0011$

$b_{FE,SOUTH} - b_{RE,SOUTH} = -0.0935$

$b_{FE,UNION} - b_{RE,UNION} = -0.0205$

⇒ The most different: SOUTH.

Hausman test:

$$EXPER = \frac{-0.0411}{(0.0330^2 - 0.0220^2)^{\frac{1}{2}}} = -1.670$$

$$EXPER^2 = \frac{0.0011}{(0.0011^2 - 0.0007^2)^{\frac{1}{2}}} = 1.296$$

$$SOUTH = \frac{-0.0935}{(0.1258^2 - 0.0317^2)^{\frac{1}{2}}} = -0.768$$

$$UNION = \frac{-0.0205}{(0.0312^2 - 0.0245^2)^{\frac{1}{2}}} = -1.061$$

⇒ RE is appropriate.