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)?

The OLS results for 1987 and 1988 are very similar in signs and magnitudes, especially for EXPER, EXPER², SOUTH and UNION. This suggests stable relationships across years.

However, by running seperate OLS each year, we assume that all individuals share the same parameter values within each year and ignore unobserved individual heterogeneity. This can lead to biased estimates if such heterogeneity is correlated with the regressors.

b. The ln(*WAGE*) equation specified as a panel data regression model is

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

Explain any differences in assumptions between this model and the models in part (a).

The panel model allows for individual-specific effects u_i, capturing unobserved heterogeneity.

OLS in part (a) ignores these effects, assuming all individuals are identical apart from observed variables.

*

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?

Compared to the OLS estimates, the fixed effects estimates in column (3) show the biggest differences in the coefficients for SOUTH and EXPER.

① SOUTH drops from around −0.121 (OLS) to −0.33 (FE), indicating the OLS may have underestimated the region wage gap

② EXPER falls from ~ 0.127 (OLS) to 0.057 (FE), suggesting upward bias in OLS due to unobserved individual traits correlated with experience.

*

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.

Numerator df = $N - 1 = 716 - 1 = 715$

Denominator df $= NT - N - k_s = 1432 - 716 - 5 = 711$

1% critical value for $F(715, 711) \approx 1.98$

Since $F = 11.98 > 1.98$, we reject the null

⇒ There are significant individual effects ⇒ fixed effects model is preferred.

*

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?

Cluster-robust SEs (column 4) allow for serial correlation within individuals, unlike regular SEs (column 3), which assume i.i.d. errors.

Robust SEs are larger, especially for SOUTH (0.1258 → 0.2495), meaning inference may change.

15.7 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. We form first differences of the variables, such as $\Delta\ln(WAGE) = \ln(WAGE_{i,1988}) - \ln(WAGE_{i,1987})$, and specify the regression

$$\Delta\ln(WAGE) = \beta_2\Delta EXPER + \beta_3\Delta EXPER^2 + \beta_4\Delta SOUTH + \beta_5\Delta UNION + \Delta e \quad \text{(XR15.7)}$$

Table 15.11 reports OLS estimates of equation (XR15.7) as Model (1), with conventional standard errors in parentheses.

TABLE 15.11 **Estimates for Exercise 15.7**

Model	C	$\Delta EXPER$	$\Delta EXPER^2$	$\Delta SOUTH$	$\Delta UNION$	$SOUTH_{i,1988}$	$UNION_{i,1988}$
(1)		0.0575	−0.0012	−0.3261	0.0822		
		(0.0330)	(0.0011)	(0.1258)	(0.0312)		
(2)	−0.0774	0.1187	−0.0014	−0.3453	0.0814		
	(0.0524)	(0.0530)	(0.0011)	(0.1264)	(0.0312)		
(3)		0.0668	−0.0012	−0.3157	0.0887	−0.0220	−0.0131
		(0.0338)	(0.0011)	(0.1261)	(0.0333)	(0.0185)	(0.0231)

a. The ability of first differencing to eliminate unobservable time-invariant heterogeneity is illustrated in equation (15.8). Explain why the strict form of exogeneity, FE2, is required for the difference estimator to be consistent. You may wish to reread the start of Section 15.1.2 to help clarify the assumption.

FE2 requires that the idiosyncratic error e_{it} is uncorrelated with all explanatory variables across time:

$$E(e_{it} \mid x_{i1}, x_{i2}, \dots, x_{iT}, w_{i1}, u_i) = 0$$

In (15.8),

$$(y_{i2} - y_{i1}) = \beta_2 (x_{i2} - x_{i1}) + (e_{i2} - e_{i1})$$

the new error term is $\Delta e_i = e_{i2} - e_{i1}$

To ensure Δx_i is uncorrelated with Δe_i, we must have $x_{i2} \perp e_{i1}$, in addition to $x_{i2} \perp e_{i2}$.

This is exactly what FE2 ensures. Without it, Δx_i and Δe_i would be correlated, and the OLS estimator in the first-difference model would be inconsistent. ※