Panel Revision Exercises

Exercise 1: Consider that we have a panel data set on n = 48 US states during T=7 periods, from 1982 up to and including 1988. The total number of observations is 336.

a) Is this a balanced panel? Explain.

b) For each state, in each time period, let y_{it} denote the number of annual traffic deaths per 10000 in the population. Let $\mathbf{x}_{it}$ denote the beer tax in 1988 US dollars. Temporarily ignore the data after 1982, so that we have a cross-section of 48 states. The estimated regression line is,

$$\widehat{y}_{i,1982} = 2.01 + 0.13\mathbf{x}_{i,1982}$$

If the Least Squares assumptions hold for this regression, how would you interpret the 0.13?

c) The estimated fixed effects regression line is

$$\widehat{y}_{it} = \widehat{\alpha}_i - \underset{(0.29)}{0.66}\mathbf{x}_{i,1982}$$

How would you interpret the -0.66?

d) Consider the results for the fixed effects regression. Do you think that the Least Squares assumptions hold, i.e., do you believe that the 0.13 in the first result comes from an unbiased estimator?

Exercise 2 (Arellano): Consider a first-order autoregressive model with individual and time effects of the form

$$y_{it} - \alpha_i - \delta_t = \rho(y_{i,t-1} - \alpha_i - \delta_{t-1}) + v_{it},\ \ i = 1, ..., N;\ \ t = 1, ..., T$$

with $E(v_{it}|y_{i0}, ..., y_{i,t-1},\ \delta_0, ..., \delta_t, \alpha_i) = 0$. Suppose that $T = 2$ so that for each individual we observe $y_{i0},\ y_{i1},\ y_{i2}$.

a) Obtain the within-groups estimate of ρ and discuss its properties.

b) Derive a consistent estimator of ρ for large N. How would your answer be modified if $T > 2$?

Exercise 3 (Arellano): Consider the following partial adjustment model with individual effects

$$y_{it} = \rho y_{i,t-1} + \beta_0 x_{it} + \beta_1 x_{i,t-1} + \alpha_i + v_{it},\ i = 1, ..., N;\ t = 1, ..., T.$$

Discuss the identification and estimation of the parameters of a model of this type when T is small and N is large, under the assumptions listed below. Set out carefully any additional assumptions that you make in each case.

a) x_{it} is a strictly exogenous variable uncorrelated with α_i, and v_{it} is a potentially serially correlated error.

b) The variable x_{it} is strictly exogenous but correlated with the individual effect α_i.

c) x_{it} is a predetermined variable correlated with α_i and v_{it} is a white noise error.

Exercise 4: Consider the following simple panel data model

$$y_{it} = x_{it}\beta + \alpha_i^* + v_{it},\ i = 1, ..., N;\ t = 1, ..., T$$

where β is one dimensional and where it is assumed that

$$\alpha_i^* = \overline{x}_i \lambda + \alpha_i$$

with $\alpha_i \sim NID(0, \sigma_\alpha^2)$ and $v_{it} \sim NID(0, \sigma_v^2)$ mutually independent and independent of all $x_{it}s$, where $\overline{x}_i = \sum_{t=1}^{T} x_{it}$. The parameter β can be estimated by the fixed effects (or within) estimator given by

$$\widehat{\beta}_{FE} = \frac{\sum_{n=1}^{N}\sum_{t=1}^{T}(y_{it} - \overline{y}_i)(x_{it} - \overline{x}_i)}{\sum_{n=1}^{N}\sum_{t=1}^{T}(x_{it} - \overline{x}_i)^2}.$$

As an alternative, the correlation between the error term $\alpha_i^* + v_{it}$ and x_{it} can be handled by instrumental variables.

a) Give an expression for the $\widehat{\beta}_{IV}$ using $(x_{it} - \overline{x}_i)$ as an instrument for x_{it}. Show that $\widehat{\beta}_{IV}$ and $\widehat{\beta}_{FE}$ are identical.

b) Another way to eliminate the individual effects α_i^* from the model is doing the following transformation:

$$y_{it} - \overline{y}_i = (x_{it} - \overline{x}_i)\beta + (v_{it} - \overline{v}_i),\ i = 1, ..., N;\ t = 1, ..., T$$

Which is the OLS estimator $(\widehat{\beta})$ based on this model? In which conditions is $\widehat{\beta}$ a consistent estimator of β?

c) Consider the between estimator $\widehat{\beta}_B$ for β. Give an expression for $\widehat{\beta}_B$ and show that it is unbiased for $\beta + \lambda$.

Exercise 5 (Dougherty): The NLSY2000 data set contains the following data for a sample of 2,427 males and 2,392 females for the years 1980{2000: years of work experience, EXP, years of schooling, S, and age, AGE. A researcher investigating the impact of schooling on willingness to work regresses EXP on S, including potential work experience, PWE, as a control. PWE was defined as:

$$PWE = AGE - S - 5$$

The following regressions were performed for males and females separately:

(1) an ordinary least squares (OLS) regression pooling the observations
(2) a within-groups fixed effects regression
(3) a random effects regression.

The results of these regressions are shown in the table below. Standard errors are given in parentheses.

	Males			Females		
	OLS	FE	RE	OLS	FE	RE
S	0.78	0.65	0.72	0.89	0.71	0.85
	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
PWE	0.83	0.94	0.94	0.74	0.88	0.87
	(0.003)	(0.001)	(0.001)	(0.004)	(0.002)	(0.002)
constant	-10.16	dropped	-10.56	-11.11	dropped	-12.39
	(0.09)		(0.14)	(0.12)		(0.19)
R^2	0.79	—	—	0.71	—	—
n	24,057	24,057	24,057	18,758	18,758	18,758
DHW $\chi^2(2)$			10.76			1.43

a) Explain why the researcher included PWE as a control.

b) Evaluate the results of the Durbin-Wu-Hausman tests.

c) For males and females separately, explain the differences in the coefficients of S in the OLS and FE regressions.

d) For males and females separately, explain the differences in the coefficients of PWE in the OLS and FE regressions.

Exercise 6 (Dougherty): A researcher has data on G, the average annual rate of growth of GDP 2001-2005, and S, the average years of schooling of the workforce in 2005, for 28 European Union countries. She believes that G depends on S and on E, the level of entrepreneurship in the country, and a disturbance term u:

$$G = 1 + 2S + 3E + u \tag{1}$$

u may be assumed to satisfy the usual regression model assumptions.

a) Unfortunately the researcher does not have data on E. Explain intuitively and mathematically the consequences of performing a simple regression of G on S. For this purpose S and E may be treated as nonstochastic variables.

The researcher does some more research and obtains data on G, the average annual rate of growth of GDP 1996-2000, and S, the average years of schooling of the workforce in 2000, for the same countries. She thinks that she can deal with the unobservable variable problem by regressing ΔG, the change in G, on ΔS, the change in S, where:

$$\begin{aligned} \Delta G &= G - G^* \\ \Delta S &= S - S^* \end{aligned}$$

assuming that E would be much the same for each country in the two periods.

She fits the equation:

$$\Delta G = \delta_1 + \delta_2 \Delta S + w \tag{2}$$

where w is a disturbance term that satisfies the usual regression model assumptions.

b) Compare the properties of the estimators of the coefficient of S in (1) and of the coefficient of S in (2).

c) Explain why in principle you would expect the estimate of δ_1 in (2) not to be significant. Suppose that nevertheless the researcher finds that the coefficient is significant. Give two possible explanations.

d) Random effects regressions have potential advantages over fixed effect regressions. Could the researcher have used a random effects regression in the present case?