

Fixed Effect Models for Panel Data: Lecture XXIII

Charles B. Moss ¹

¹University of Florida

March 14, 2012

- 1 Readings
- 2 Panel Data
 - Individual Effects
 - Time Effects
 - Modified Ordinary Least Squares
- 3 Pooling Assumptions
 - Empirical Procedure
- 4 Covariance Estimator
 - Different Slopes and Intercepts
 - Different Intercepts - Same Slopes
 - Corn Estimates
 - Same Slopes – Same Intercepts
- 5 Testing for Pooling
- 6 Dummy Variable Formulation
 - Dummy Variables for Panel Data

Readings

- Arellano, M. 2003. *Panel Data Econometrics*. Oxford University Press. (Chapters 1 and 2 pp. 1-29).
- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata Revised Edition*. Stata Press. (Chapter 8 [Section 8.1- 8.2] pp. 235- 240).
- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition. (Chapter 11 Section 11.1 11.4 pp. 314 335)
- * Hsiao, C. 1990. *Analysis of Panel Data*. Cambridge University Press. (Chapters 1, 2, and 3 [Section 3.1-3.2] pp. 1-31).

Panel Data

- There are two primary types of data used in econometrics:
 - Data on actions across individuals – Cross Sectional Data

$$y_i = \alpha + \beta'x_i + \gamma'z_i + u_i \quad \exists: i = 1, \dots, N \quad (1)$$

where y_i is the dependent variable, x_i are exogenous (and observed) variables, z_i are exogenous (possibly unobserved) variables, and u_i is the unobserved error term.

- Repeated observations of the actions on an individual over time – Time Series Data

$$y_t = \alpha + \beta'x_t + \gamma'z_t + u_t \quad \exists: t = 1, \dots, T \quad (2)$$

where y_t is the dependent variable, x_t are exogenous (and observed) variables, z_t are exogenous (possibly unobserved) variables, and u_t is the unobserved error term.

Panel Data - Continued

- Panel or longitudinal data then involves repeated observation on a cross-section of individuals over time

$$y_{it} = \alpha^* + \beta' x_{it} + \gamma' z_{it} + u_{it} \quad \ni \quad \begin{matrix} i = 1, \dots, N \\ t = 1, \dots, T \end{matrix} \quad (3)$$

- We know that Ordinary Least Squares estimates of α^* , β , and γ are Best Linear Unbiased (with the standard caveats about homoscedasticity and exogeneity).
- Of course problems are potentially introduced if we cannot observe z – If x and z are correlated then OLS estimates of β are biased.
- However, the fact that we have repeated observations on individuals may allow us to remove the effect of z .

Individual Effects

- First, if we assume that $z_{it} = z_i$ (or that the value of z is the same for each individual over time – maybe z is a measure of ethnicity), the effect can be removed by first-differencing the data

$$y_{it} - y_{i,t-1} = \beta' (x_{it} - x_{i,t-1}) + \gamma' (z_{it} - z_{i,t-1}) + (u_{it} - u_{i,t-1}) \quad (4)$$

Since $z_{it} = z_{i,t-1} = z_i$

$$y_{it} - y_{i,t-1} = \beta' (x_{it} - x_{i,t-1}) + (u_{it} - u_{i,t-1}) \quad \ni: \begin{matrix} i = 1, \dots, N \\ t = 2, \dots, T \end{matrix} \quad (5)$$

Time Effects

- Similarly, if $z_{it} = z_t$ (or the unobserved variables are the same for every individual at each point in time) we can remove the effect of the unobserved variables by subtracting the average effect for each time period

$$y_{it} - \bar{y}_t = \beta' (x_{it} - \bar{x}_t) + \gamma' (z_{it} - \bar{z}_t) + (u_{it} - \bar{u}_t) \quad (6)$$

Like the individual effect $z_{it} = \bar{z}_t$

$$y_{it} - \bar{y}_t = \beta' (x_{it} - \bar{x}_t) + (u_{it} - \bar{u}_t)$$

$$\bar{y}_t = \frac{1}{N} \sum_{i=1}^N y_{it}, \quad \bar{x}_t = \frac{1}{N} \sum_{i=1}^N x_{it}, \quad \bar{u}_t = \frac{1}{N} \sum_{i=1}^N u_{it} \quad (7)$$

Modified Ordinary Least Squares

- Using either transformation (or modification) Ordinary Least Squares provides unbiased and consistent estimates of β .
- Unfortunately, if we have a cross-sectional dataset (i.e., $T = 1$) or a single time-series (i.e., $N = 1$) these transformations cannot be used.
 - We cannot remove the unobserved effect.

Pooling Assumptions

- Consider first the complete pool – the model is homogeneous in intercept and slopes

$$y_{it}^* = \alpha^* + \beta' x_{it} + u_{it} \quad (8)$$

- Case I: Heterogeneous intercepts ($\alpha_i \neq \alpha$) and homogeneous in slopes ($\beta_i = \beta$)

$$y_{it} = \alpha_i^* + \beta' x_{it} + u_{it} \quad (9)$$

- Case II: Heterogeneous in slopes and intercepts ($\alpha_i \neq \alpha$, $\beta_i \neq \beta$)

$$y_{it} = \alpha_i^* + \beta_i' x_{it} + u_{it} \quad (10)$$

Empirical Procedure

- From the general model we pose three hypotheses:
 - H_1 : Regression slope coefficients are identical and the intercepts are not.
 - H_2 : Regression intercepts are the same and the slope coefficients are not.
 - H_3 : Both slopes and intercepts are the same.

Covariance Estimator - Different Slopes and Intercepts

- Computing the variance matrices for each individual

$$\begin{aligned}\bar{y}_i &= \frac{1}{T} \sum_{t=1}^T y_{it} & \bar{x}_i &= \frac{1}{T} \sum_{t=1}^T x_{it} \\ W_{XX,i} &= \sum_{t=1}^T (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' & W_{XY,i} &= \sum_{t=1}^T (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i) \\ W_{YY,i} &= \sum_{t=1}^T (y_{it} - \bar{y}_i)^2\end{aligned}\tag{11}$$

- Ordinary Least Squares Estimators

$$\hat{\beta}_i = W_{XX,i}^{-1} W_{XY,i} \quad \alpha_i = \bar{y}_i - \hat{\beta}_i \bar{x}_i \quad i = 1, \dots, N \tag{12}$$

Different Slopes and Intercepts

- Sum Squared Residuals

$$\begin{aligned} W_{YY,i} &= \sum_{t=1}^T (y_{it} - \bar{y}_i)^2 \\ RSS_i &= W_{YY,i} - W'_{XY,i} W_{XX,i}^{-1} W_{XY,i} \\ S_1 &= \sum_{i=1}^N RSS_i \end{aligned} \tag{13}$$

Different Intercepts - Same Slopes

- Aggregating the variance matrices (taking the "average" covariance matrices)

$$\begin{aligned} W_{XX} &= \sum_{i=1}^N W_{XX,i} & W_{XY} &= \sum_{i=1}^N W_{XY,i} \\ W_{YY} &= \sum_{i=1}^N W_{YY,i} \end{aligned} \quad (14)$$

- The aggregated estimator is then

$$\hat{\beta}_W = W_{XX}^{-1} W_{XY} \quad \bar{\alpha}_i^* = \bar{y}_i - \hat{\beta}'_W \bar{x}_i \quad i = 1, \dots, N \quad (15)$$

- Sum Squared Residuals

$$S_2 = W_{YY} - W'_{XY} W_{XX}^{-1} W_{XY} \quad (16)$$

Corn Estimates

Variable	Covariance Matrix			$X'Y$	β	α
	Nitrogen	Phosphorous	Potash			
Illinois						
Nitrogen	1.2823	0.7194	1.5488	0.7415	0.7985	3.7917
Phosphorous	0.7160	0.6410	1.0156	0.2204	-0.9813	
Potash	1.5427	1.0174	2.0326	0.7894	0.2734	
Indiana						
Nitrogen	1.0346	0.2489	0.7220	0.6577	0.4386	3.6162
Phosphorous	0.2348	0.3717	0.2320	-0.0913	-0.8905	
Potash	0.7268	0.2448	0.6072	0.4587	0.5894	
Pooled						
Nitrogen	2.3168	0.9683	2.2708	1.3992	0.5924	3.9789
Phosphorous	0.9508	1.0128	1.2475	0.1291	-0.9335	3.8851
Potash	2.2695	1.2622	2.6398	1.2481	0.4008	

Same Slopes – Same Intercepts

- Taking the averages over the complete sample

$$\bar{y} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it} \quad \bar{x} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T x_{it} \quad (17)$$

- Next, redefining the covariance matrices

$$\begin{aligned} T_{XX} &= \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})(x_{it} - \bar{x})' \\ T_{XY} &= \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})(y_{it} - \bar{y})' \\ T_{YY} &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y})^2 \end{aligned} \quad (18)$$

Same Slopes – Same Intercepts - Continued

- Estimator

$$\hat{\beta} = T_{XX}^{-1} T_{XY} \quad \hat{\alpha}^* = \bar{y} - \hat{\beta} \bar{x} \quad (19)$$

- Sum Squared Residuals

$$S_3 = T_{YY} - T'_{XY} T_{XX}^{-1} T_{XY} \quad (20)$$

Testing for Pooling

- Testing first for pooling both the slope and intercept terms:

$$H_3 : \alpha_1^* = \alpha_2^* = \cdots \alpha_N^* \text{ and } \beta_1 = \beta_2 = \cdots \beta_N$$

$$F_3 = \frac{\frac{S_3 - S_1}{(N-1)(K+1)}}{\frac{S_1}{NT - N(K+1)}} \sim F([N-1][K+1], NT - N[K+1]) \quad (21)$$

- If this hypothesis is rejected, we then test for homogeneity of the slopes but heterogeneity of the intercepts:

$$H_1 : \beta_1 = \beta_2 = \cdots \beta_N$$

$$F_1 = \frac{\frac{S_2 - S_1}{(N-1)K}}{\frac{S_1}{NT - N(K+1)}} \sim F([N-1]K, NT - N[K+1]) \quad (22)$$

Dummy Variable Formulation

- Dummy variables are typically formulated as $D_i = 1$ if $i \in G_i$ and $D_i = 0$ if $i \notin G_i$.
 - For example, suppose that I want to test whether participation in a development program increases income:

$$y_{it} = \alpha_0 + \beta^* D_i + \tilde{\beta} x_{it} + \epsilon_{it} \quad (23)$$

In this formulation β^* is the effect of the development program on household income.

- We can extend the formulation to test whether the effect of x_{it} depends on the program

$$y_{it} = \alpha_0 + \beta^* D_i + (1 + \gamma D_i) \beta x_{it} + \epsilon_{it} \quad (24)$$

In this case the effect of $\gamma\beta$ is the effect of x_{it} in the presence of the program.

Dummy Variables for Panel Data

- First, consider the standard Panel Data formulation

$$\begin{aligned}
 y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix} \quad x_i = \begin{bmatrix} x_{i,11} & x_{i,12} & \cdots & x_{i,1K} \\ x_{i,21} & x_{i,22} & \cdots & x_{i,2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{i,T1} & x_{i,T2} & \cdots & x_{i,TK} \end{bmatrix} \\
 e \in M_{T \times 1} = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad u_i \in M_{T \times 1} = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix} \quad (25) \\
 E[u_i] = 0, \quad E[u_i u_i'] = \sigma^2 I_T, \quad E[u_i u_j'] = 0 \forall i \neq j
 \end{aligned}$$

Dummy Variables for Panel Data - Continued

- The fixed effect model can then be formulated as a dummy variable problem

$$\begin{aligned}
 Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} &= \begin{bmatrix} e \\ 0 \\ \vdots \\ 0 \end{bmatrix} \alpha_1^* + \begin{bmatrix} 0 \\ e \\ \vdots \\ 0 \end{bmatrix} \alpha_2^* + \cdots + \begin{bmatrix} 0 \\ 0 \\ \vdots \\ e \end{bmatrix} \alpha_N^* \\
 &+ \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}
 \end{aligned} \tag{26}$$

Readings – Next Time

- Arellano, M. 2003. *Panel Data Econometrics*. Oxford University Press. (Chapters 3 and 4 pp. 31-54).
- * Cameron, A. C. and P. K. Trivedi. 2010. *Microeconometrics Using Stata* Revised Edition. Stata Press. (Chapter 8 [Section 8.3 - 8.7] pp. 240- 253).
- * Greene, W.H. 2012. *Econometric Analysis* Seventh Edition. (Chapter 11 Section 11.5 pp. 359 364)
- * Hsiao, C. 1990. *Analysis of Panel Data*. Cambridge University Press. (Chapters 3 [Section 3.4-3.9] pp. 32-68)