

Estimation of Production Functions – Fixed Effects in Panel Data: Lecture VIII

Charles B. Moss¹

¹University of Florida

September 13, 2018

- 1 Analysis of Covariance
 - Pooling Assumptions
 - Individual Regressions
 - Same Slopes
 - Same Slopes & Same Intercepts
- 2 Tests for Pooling
 - Both Slope and Intercept
 - Different Intercepts and Same Slopes
- 3 Other Approaches – Same Thing
 - Dummy Variable Formulation
 - Sweeping the Data

Analysis of Covariance

- Looking at a representative regression model

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

- It is well known that ordinary least squares (OLS) regressions of y on x and z are best linear unbiased estimators (BLUE) of α , β , and γ .
- However, the results are corrupted if we do not observe z . Specifically if the covariance of x and z is non-zero and we do not observe z the OLS estimates are biased.
- However, if repeated observations of a group of individuals are available (i.e., panel or longitudinal data) they may help us eliminate the effect of z .

Individual Effects

- For example, if $z_{it} = z_i$ (or the unobserved variable is the same for each individual across time), the effect of the unobserved variables can be removed by first-differencing the dependent and independent variables

$$y_{it} - y_{i,t-1} = \beta' (x_{it} - x_{i,t-1}) + \gamma' (z_{it} - z_{i,t-1}) + (u_{it} - u_{i,t-1}) \quad (2)$$

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

$$y_{it} - y_{i,t-1} = \beta' (x_{it} - x_{i,t-1}) + (u_{it} - u_{i,t-1}) \Leftrightarrow \begin{cases} i = 1, \dots, N \\ t = 2, \dots, T \end{cases} \quad (3)$$

Individual Effect - Mean Corrected

- Similarly, if $z_{it} = z_t$ (or the unobserved variables are the same for every individual at any point in time) we can derive a consistent estimator by subtracting the mean of each of the dependent and independent variables for each individual

$$y_{it} - \bar{y}_i = \beta' (x_{it} - \bar{x}_i) + \gamma' (z_{it} - \bar{z}_i) + (u_{it} - \bar{u}_i) \quad (4)$$

Since $z_{it} = \bar{z}_i$

$$y_{it} - \bar{y}_i = \beta' (x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i) \quad (5)$$
$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}, \bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}, \bar{u}_i = \frac{1}{T} \sum_{t=1}^T u_{it}$$

OLS estimators then provide unbiased and consistent estimates of β .

Pooling Assumptions

- In order to derive a little different form of the panel estimator, we start from the pooled estimates

$$y_{it} = \alpha^* + \beta' x_{it} + \nu_{it} \quad (6)$$

we envision two sets of restrictions or hypotheses:

- **Case I:** Heterogenous intercepts ($\alpha_i^* \neq \alpha^*$) and a homogenous slope ($\beta_i = \beta$).
- **Case II:** Heterogenous slopes and intercepts ($\alpha_i^* \neq \alpha_i, \beta_i \neq \beta$)

$$y_{it} = \alpha_i^* + \beta_i' x_{it} + \nu_{it} \quad (7)$$

- **Case III:** Homogenous slopes and heterogenous slopes (not typically addressed)

$$y_{it} = \alpha^* + \beta_i' + \nu_{it} \quad (8)$$

Empirical Procedure

- This leads to an empirical system with three hypotheses from the general model:
 - 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 (again infrequently tested).
 - H_3 : Both slopes and intercepts are the same.

Analysis of Variance – Individual Regressions

$$\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}, \quad \bar{x}_i = \frac{1}{T} \sum_{t=1}^T x_{it}$$

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

$$W_{XX,i} = \sum_{t=1}^T (x_{it} - \bar{x}_i)', \quad W_{XY,i} = \sum_{t=1}^T (x_{it} - \bar{x}_i) (y_{it} - \bar{y}_i) \quad (9)$$

$$W_{YY,i} = \sum_{t=1}^T (y_{it} - \bar{y}_i)^2, \quad RSS_i = W_{YY,i} - W_{XY,i}' W_{XX,i}^{-1} W_{XY,i}$$

$$S_1 = \sum_{i=1}^N RSS_i$$

Cotton Estimates

	Covariance Matrix			$X'Y$	β	α
Variable	Nitrogen	Phosphorous	Potash			
	Arkansas					
Nitrogen	0.0314	0.0016	0.0283	0.0356	1.0731	2.9101
Phosphorous	0.0016	0.0091	0.0036	-0.0011	-0.4322	
Potash	0.0283	0.0036	0.0355	0.0359	0.4770	
	Louisiana					
Nitrogen	0.0330	0.0009	0.0195	0.0163	0.3025	5.9802
Phosphorous	0.0009	0.0135	0.0033	-0.0066	-0.5964	
Potash	0.0195	0.0033	0.0345	0.0161	0.3519	
	Mississippi					
Nitrogen	0.0091	- 0.0011	0.0173	0.0088	0.3156	6.0824
Phosphorous	-0.0011	0.0074	-0.0035	-0.0058	-0.5877	
Potash	0.0173	-0.0035	0.0566	-0.0249	0.3073	

Cotton Estimates, Continued

	Covariance Matrix					
Variable	Nitrogen	Phosphorous	Potash	$X'Y$	β	α
	Texas					
Nitrogen	0.0343	0.0032	0.0191	0.0254	0.9489	4.6369
Phosphorous	0.0032	0.0089	0.0042	-0.0021	-0.4376	
Potash	0.0191	0.0042	0.0323	0.0068	-0.2966	
	Pooled					
Nitrogen	0.1078	0.0046	0.0841	0.0861	0.6670	AR – 4.6187
Phosphorous	0.0046	0.0388	0.0075	-0.0155	-0.5165	LA – 4.7065
Potash	0.0841	0.0075	0.1589	0.0839	0.1979	MS – 4.6429
						TX – 4.6490

Analysis of Variance – Same Slopes

$$\begin{aligned}\hat{\beta}_W &= W_{XX}^{-1} W_{XY}, \quad \bar{x}_i^* = \bar{y}_i - \hat{\beta}_W' \bar{x}_i, \quad i = 1, \dots, N \\ W_{XX} &= \sum_{i=1}^N W_{XX,i}, \quad W_{XY} = \sum_{i=1}^N W_{XY,i} \\ W_{YY} &= \sum_{i=1}^N W_{YY,i}, \quad S_2 = W_{YY} - W_{XY}' W_{XX}^{-1} W_{XY}\end{aligned}\tag{10}$$

Analysis of Variance – Same Slopes & Same Intercepts

$$\begin{aligned}\hat{\beta} &= T_{XX}^{-1} T_{XY}, \hat{\alpha}^* = \bar{y} - \hat{\beta} \bar{x} \\ T_{XX} &= \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})(x_{it} - \bar{x})', T_{XY} = \sum_{i=1}^N \sum_{t=1}^T (x_{it} - \bar{x})(y_{it} - \bar{y}) \\ T_{YY} &= \sum_{i=1}^N \sum_{t=1}^T (y_{it} - \bar{y})^2 \\ \bar{y} &= \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T y_{it}, \bar{x} = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T t = 1^T x_{it} \\ S_e &= T_{YY} - T_{XY}' T_{XX}^{-1} T_{XY}\end{aligned}\tag{11}$$

Test for Pooling – Both Slope and Intercept

$$H_3 : \begin{aligned} \alpha_1^* &= \alpha_2^* = \cdots \alpha_N^* \\ \beta_1 &= \beta_2 = \cdots \beta_N \end{aligned}$$

$$F_3 \frac{\frac{S_3 - S_1}{[(N-1)(K+1)]}}{\frac{S_1}{[NT - N(K+1)]}} \sim F([N-1][K+1], NT - N[K+1]) \quad (12)$$

Test for Pooling – Different Intercepts and Same Slopes

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

$$H_1; \beta_1 = \beta_2 = \cdots \beta_N$$
$$F_1 = \frac{\frac{S_2 - S_1}{[(N-1)K]}}{\frac{S_1}{[NT - N(K+1)]}} \sim F([N-1]K, NT - N[K+1]) \quad (13)$$

Dummy Variable Formulation

- The analysis of covariance can be accomplished using dummy variables. Specifically, we could reformulate the regression model as

$$\begin{aligned}
 Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} &+ \alpha_1^* \begin{bmatrix} e \\ 0 \\ \vdots \\ 0 \end{bmatrix} + \alpha_2^* \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix} + \\
 &\cdots \alpha_N^* \begin{bmatrix} 0 \\ 0 \\ \vdots \\ e \end{bmatrix} + \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{bmatrix}
 \end{aligned} \tag{14}$$

Dummy Variable Formulation, Continued

- Where

$$y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}, x_i = \begin{bmatrix} x_{11,i} & x_{21,i} & \cdots & x_{K1,i} \\ x_{12,i} & x_{22,i} & \cdots & x_{K2,i} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T,i} & x_{2T,i} & \cdots & x_{KT,i} \end{bmatrix} \quad (15)$$

$$e = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, u_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{3N} \end{bmatrix}, E[u_i] = 0, E[u_i u_i'] = \sigma^2 I_T$$

Dummy Variable Formulation, Continued

- Given this formulation, we know the OLS Estimation of

$$y_{it} = \alpha_i^* + \beta x_{it} + u_{it} \quad (16)$$

- The OLS estimation of α and β are obtained by minimizing

$$S = \sum_{i=1}^N u_i' u_i = \sum_{i=1}^N (y_i - e\alpha^* - x_i\beta)' (y_i - e\alpha^* - x_i\beta) \quad (17)$$

Sweeping the Data

- The sweep matrix which would remove the individual effects is written as

$$Q_T = I_T - \frac{1}{T}ee' \quad (18)$$

where e is a $T \times 1$ vector of ones. If $T = 4$

$$Q_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad (19)$$

Sweeping the Data, Continued

$$Q_4 = \begin{bmatrix} 1 - \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & 1 - \frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & 1 - \frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & -\frac{1}{4} & 1 - \frac{1}{4} \end{bmatrix} \quad (20)$$

Sweeping the Data, Continued

$$\begin{aligned}
 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}' \left[I_4 - \frac{1}{4} ee' \right] &= \begin{bmatrix} \left(1 - \frac{1}{4}\right) x_1 - \frac{1}{4} x_2 - \frac{1}{4} x_3 - \frac{1}{4} x_4 \\ -\frac{1}{4} x_1 + \left(1 - \frac{1}{4}\right) x_2 - \frac{1}{4} x_3 - \frac{1}{4} x_4 \\ -\frac{1}{4} x_1 - \frac{1}{4} x_2 - \left(1 - \frac{1}{4}\right) x_3 - \frac{1}{4} x_4 \\ -\frac{1}{4} x_2 - \frac{1}{4} x_3 - \frac{1}{4} x_3 + \left(1 - \frac{1}{4}\right) x_4 \end{bmatrix} \\
 &= \begin{bmatrix} x_1 - \frac{1}{4} (x_1 + x_2 + x_3 + x_4) \\ x_2 - \frac{1}{4} (x_1 + x_2 + x_3 + x_4) \\ x_3 - \frac{1}{4} (x_1 + x_2 + x_3 + x_4) \\ x_4 - \frac{1}{4} (x_1 + x_2 + x_3 + x_4) \end{bmatrix}
 \end{aligned}
 \tag{21}$$

Sweeping the Data, Continued

- The least square estimator without the intercept could then be expressed (for a single individual) as

$$\hat{\beta} = (x'Q_Tx)^{-1} (x'Q_Ty) \quad (22)$$

- Extending the formulation for panel estimation

$$\tilde{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \tilde{x} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ x_{21} & x_{22} & \cdots & x_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix} \quad (23)$$

$$\hat{\beta} = [\tilde{x}' (I_N \otimes Q_T) \tilde{x}]^{-1} [\tilde{x}' (I_N \otimes Q_T) \tilde{y}]$$