Panel Data Analysis
Part II – Additional Results

James J. Heckman
University of Chicago

Econ 312, Spring 2021

- **Mundlak Problem**: Example of a use of panel data to ferret out and identify a cross sectional relationship.
- Mundlak posed the problem that X_{it} is correlated with f_i (*e.g.*.
- f_i is managerial ability; X_{it} is inputs)

$E(f_i(X_{it})) \neq 0$, and we have a specification error bias:

$$E\left[\hat{\beta}\right] = \beta + E\left[(X'X)^{-1}X'\varepsilon\right] \neq 0 \qquad \text{since}$$

$E(X'f) \neq 0$.

THE UNIVERSITY OF CHICAGO

- OLS is inconsistent and biased.
- One way to eliminate the problem: Use the within estimator:

$$\left[\mathrm{I} - \frac{\iota\iota'}{T}\right] Y_i = \left[\mathrm{I} - \frac{\iota\iota'}{T}\right] X_i \beta + \left[\mathrm{I} - \frac{\iota\iota'}{T}\right] \varepsilon_i$$

$$Y_{it} - Y_{i\cdot} = [X_i - \iota X_{i\cdot}]'\beta + U_{i\cdot}$$

- On the transformed data, we get an estimator that is unbiased and consistent.
- Estimator of fixed effect not consistent (we acquire an incidental parameters problem but we can eliminate it as $N \to \infty$, T fixed, we get a consistent estimator.)
- Genesis of the first component of differences in differences.

- Notice, however, if there exists a variable that stays constant over the spell for *all* persons, we cannot estimate the associated β.

$$\hat{f}_i = f_i + X_{i\cdot}^f \beta^f$$

where $X_{i\cdot}^f$ and β^f are variables and associated coefficients that stay fixed over spells (we can regress estimated fixed effects on the X provided that they stay constant over the spell are not corrected with the f_i).

- In a cross section context, we have that without some other information, the model is not identified unless we can invoke *IV* estimation.
- F.E. estimator is a conditional version of R.E. estimator.
- R.E. estimator: $f_i + U_{it}$ both random values, we condition on values of f_i.

How To Test For the Presence of Bias?

- H_0 : No Bias in the OLS estimator
- H_A :OLS and between estimator are biased, Within estimator is unbiased.

$$\hat{\beta}_W \text{ vs. } \hat{\beta}_B \; \hat{\beta}_W = \beta + (W_{XX})^{-1} \sum_{i=1}^{\mathrm{I}} X_i'[\mathrm{I} - \frac{1}{T} \, \iota\iota']\varepsilon_i \; \hat{\beta}_{\mathrm{B}} = \beta + (\mathrm{B}_{XX})^{-1} \left[\sum_{i=1}^{T} X_i' \frac{1}{T} \iota\iota' \; \varepsilon_i \right].$$

- Under H_0

$$\text{COV}\left(\hat{\beta}_W, \hat{\beta}_B\right) = 0$$

- Independently distributed under a normality assumption.
- $\therefore$ we can test (just pool the standard errors).

Strict Exogeneity Test

- Basic idea

$$E(f_i \mid X_i) \neq 0. \quad \text{where } X_i \equiv (X_{i1}, \dots, X_{iT})$$

- Failure of this is failure of strict exogeneity in the time series literature.
- Regression Function (Scalar Case)

$$E^*(f_i \mid X_i) = \varphi_0 + X_{i1}\varphi_1 + X_{i2}\varphi_2 + X_{i3}\varphi_3 + \dots$$

where E^* denotes linear projection.

- Then, in Mundlak's problem, we get that $t = 1, \dots, T$

$$Y_{it} = \beta_0 + X_{it}\beta_1 + [\varphi_0 + X_{i1}\varphi_1 + X_{i2}\varphi_2 + \dots] + U_{it}.$$

- Then, we can test to see whether or not future and past values of X_{it} enter the equation (if so, we get a violation of strict exogeneity in this set up).
- Notice we can estimate φ_2 (from first equation), φ_1 (from second equation) and so forth
- $\therefore$ can estimate β_1 but, we cannot separate out the intercepts in this equation.
- Nor can we identify variables that don't vary over time.
- This is just a control function in the sense of Heckman and Robb (1985, 1986).

Chamberlain's Strict Exogeneity Test

$$Y_{it} = X_{it}\beta + \varepsilon_{it}, \qquad i = 1, \ldots, \mathrm{I}, \qquad t = 1, \ldots, T$$

X_{it} is strictly exogenous if $E(\varepsilon_{it} \mid X_i) = 0$

- $\therefore$ model can be fitted by OLS.
- We can test, in time series

$$Y_{it} = X_{it}\beta + \underbrace{X_{it+j}\gamma}_{\text{an extraneous variable}} + \varepsilon_{it},$$
$$i = 1, \ldots, I, t = 1, \ldots, T$$

- We have strict exogeneity in the process if $\gamma = 0$ (assumption: X_{it} is correlated over time: $X_{i,t+j}$) a future value of a variable is in the equation (that doesn't belong) $\therefore$ we can do an *exact* test.

- Consider special error structure: (one factor setup)

$$\varepsilon_{it} = f_i + U_{it}, \qquad U_{it} \text{ i.i.d.}$$

- $E^*(f_i \,|X_{1i}, X_{2i}, X_{3i}, \ldots, X_{Ti}) = \sum_{j=1}^{T} \varphi_j X_{ji}$
- Then if we relax the strict exogeneity assumption, we have that

$$E^*(Y_{it} \,|X_{it},\ [X_{i1}, \ldots, X_{iT}]) = X_{it}\beta + \sum_{j=1}^{T} \varphi_j X_{ji} \ E(Y_{it} \mid X_i) = X_i\pi$$

- Array the X_{it} into a supervector

$$\Pi = \text{DIAG}\{\beta, \beta, \ldots, \beta\} + \iota_T \varphi$$

$\therefore$ in all T regressions, we have that φ_j stays fixed $\therefore$ we can test this assumption.

- When applying this test in particular economic situations, we must interpret the results with caution.
- For e.g., in the application of this test to the situation in the permanent income hypothesis, the significance of the coefficients of future values can not be ruled out under the model.

Example: Chamberlain test with $T = 3$ periods

- Simple regression setting with $\varepsilon_{it} = f_i + U_{it,} U_{it}$ i.i.d., $U_{it} \not\perp\!\!\!\perp f_i$ we have:

$$f_i = \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3 + V$$

- Then

$$\begin{aligned} Y_1 &= \beta_1 X_1 + \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3 + V + U_1 \\ Y_2 &= \beta_2 X_2 + \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3 + V + U_2 \\ Y_3 &= \beta_3 X_3 + \sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3 + V + U_3 \end{aligned}$$

- For a factor structure,
- $\varepsilon_{it} = \lambda_t f_i + U_{it} \qquad U_{it}$ i.i.d. $\perp\!\!\!\perp f_i$.

- Then: $Y_1 = \beta_1 X_1 + \lambda_1(\sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3) + \lambda_1 V + U_1$
  $Y_2 = \beta_2 X_2 + \lambda_2(\sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3) + \lambda_2 V + U_2$
  $Y_3 = \beta_3 X_3 + \lambda_3(\sigma_1 X_1 + \sigma_2 X_2 + \sigma_3 X_3) + \lambda_3 V + U_3$

Can Identify

$$\begin{bmatrix} \lambda_1\sigma_2 & \lambda_1\sigma_3 & (\beta_1+\lambda_1\sigma_1) \\ \lambda_2\sigma_1 & \lambda_2\sigma_3 & \beta_2+\lambda_2\sigma_2 \\ (\beta_3+\lambda_3\sigma_3) & \lambda_3\sigma_1 & \lambda_3\sigma_2\cdots\cdots\cdots \end{bmatrix}$$

- Normalize: set $\lambda_1 \equiv 1$ then we can identify, λ_2, λ_3, σ_1, σ_2 and σ_3.

Maximum likelihood panel data estimators

- Consider these models from a more general viewpoint, we can form different maximum likelihood estimators of the parameters of interest.
- Assume $\varepsilon_{it} = f_i + U_{it}$. Write
- $\underset{\sim}{Z}_i = (Y_{i1}, \ldots, Y_{iT}, X_{i1}, \ldots, X_{iT}) \qquad i = 1, \ldots, \mathrm{I}$

- $\underset{\sim}{Z}_i$ is an i.i.d. random vector with distribution depending on

$$\underset{\sim}{\theta} = (\beta, f_1, \ldots, f_i, \ldots, f_{\mathrm{I}}) = (\beta, f)$$

(treat f_i as a parameter)

$$\mathcal{L} = \prod_{i=1}^{N} f(\underset{\sim}{Z}_i \Big| \underset{\sim}{\theta})\ f(Z_i \,|\beta, f_1, \ldots, f_I). \text{ Max } \mathcal{L}\ w.r.t.\theta \Longrightarrow \hat{\theta}_{ML}.$$

$$\left(\hat{f}_i\right)_{ML} \overset{P}{\nrightarrow} f_i \text{ as } \begin{array}{c} I \to \infty \\ T \text{ fixed} \end{array} .$$

- In general, $\hat{\beta}_{ML} \overset{P}{\nrightarrow} \beta$ as I $\to \infty$ because of this.
- Not like in linear models (in general, roots of these equations interconnected and we have problems).
- A joint system of equations

$$\frac{\partial \ell n \mathcal{L}}{\partial \beta} = 0 \qquad \frac{\partial \ell n \mathcal{L}}{\partial f_i} = 0, \qquad i = 1, \dots, I.$$

- This set of likelihood equations can be solved using three distinct concepts:
  1. Marginal Likelihood;
  2. Conditional Likelihood; and
  3. Integrated Likelihood.

Marginal Likelihood (or Ancillary Likelihood):

- Find (if possible) $g(Y, X)$ independent of the f
- *i.e.* find some statistic $S_i = S(Y_i, X_i)$ such that $f(S_i|\beta)$

$$\mathcal{L}_{\text{Marginal}} = \prod_{i=1}^{I} f(S_i|\beta) \ \underset{\beta}{Max} \mathcal{L}_M \to \hat{\beta}_M$$

- Then we can form the ML estimators for β (the parameters of interest) without worrying about the $f_i s$.

- We say that s_i is ancillary for f given β with respect to original model.
- (This is really b-ancillarity).
- An example of this is the *within* estimator.

$$Y_{it} = X_{it}\beta + f_i + U_{it}$$

$$U_{it} \text{ i.i.d. } \mathcal{N}(0, \sigma_U^2)$$

- $T = 2$

$$S_i = Y_{i2} - Y_{i1}$$

- S_i is called an ancillary statistic distribution is independent is f_i:

$$S_i \,|X \overset{d}{\sim} \mathcal{N}(\beta(X_{i2} - X_{i1}) \,\big|2\sigma^2) \,.$$

- Thus an example of the Marginal likelihood estimator is the first difference estimator, which is almost identical to the "within" estimator.
- Here, the within estimator would also be a Marginal likelihood estimator.

- Because

$$U_i'[I - \frac{\iota\iota'}{T}]\frac{\iota\iota'}{T}U_i = 0.$$

- We can always break up the distribution of Y_i into two pieces

$$Y_i = \left[\mathrm{I} - \frac{\iota\iota'}{T}\right]Y_i + \frac{\iota\iota'}{T}Y_i$$

$$g\left(Y_i \mid X_i, \beta, f_i\right) = \underbrace{g\left(FY_i \mid X_i, \beta\right)}_{\substack{\text{This portion ind of } f_i \\ \text{Marginal Likelihood}}} \underbrace{g\left(\iota Y_{i\cdot} \mid X_i, \beta\right)}_{\substack{\text{This is a sufficient} \\ \text{statistic for } f_i}}$$

Maximum Likelihood: Second Principle

- Find s, a sufficient statistic for f_i such that

$$f(Y_i \mid \text{sufficient statistic for } f_i) \text{ is ind. of } f_i.$$

- Find

$$s_i = s(Y_i, X_i) \text{ so that } f(Z_i \,|\beta, f_i, s_i) = f(Z_i \,|\beta, s_i)$$

- Can throw away s_i, e.g., $S_i = Y_{i1} + Y_{i2}$

$$Y_{i1} + Y_{i2} \overset{d}{\sim} N(f_i + \beta\,(X_{i1} + X_{i2})\,, 2\sigma_U^2 + 4\sigma_f^2).$$

Transform observation

$$\begin{pmatrix} Y_{i1} \\ Y_{i2} \end{pmatrix} \rightarrow \begin{pmatrix} Y_{i2} - Y_{i1} \\ Y_{i2} + Y_{i1} \end{pmatrix}$$

$$Cov(Y_{i2} - Y_{i1}, Y_{i2} + Y_{i1} \,|X) = 0$$

$$\begin{aligned} f(Y_{i1}, Y_{i2}) &= f(Y_{i2} - Y_{i1}, Y_{i1} + Y_{i2}) \\ &= f(Y_{i2} - Y_{i1} \,|X)\, f(Y_{i1} + Y_{i2} \,|X) \end{aligned}$$

but

$$\begin{aligned} & f\left(Y_{i2} - Y_{i1}, Y_{i2} + Y_{i1} \mid X, S_i\right) \\ = \; & f\left(Y_{i2} - Y_{i1} \mid X_i\right) \end{aligned}$$

$\therefore$ conditional likelihood function is the same as in previous case.

Integrated L.F. or Random Effects Estimator

- Pick a density for f_i (other methods do not require this)

$$\text{pdf of } f_i \equiv g\left(f_i \mid X\right).$$

- For each person

$$g\left(Y_i \mid X_i, \beta\right) = \int g\left(Y_i \mid X_i, \beta, f\right) g\left(f \mid X\right) df$$

$$\mathcal{L}_I = \prod_{i=1}^{I} g\left(Y_i \mid X_i, \beta\right)$$

- Suppose it is normal, $f_i \sim \mathcal{N}\left(0, \sigma_f^2\right)$.

- When we integrate out f_i in the above using normality, we get

$$Y_i \sim \mathcal{N}\left(X_i\beta, (\sigma_f^2 + \sigma_U^2)\begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \cdots & \cdots & 1 \end{pmatrix}\right).$$

- Problem becomes one of estimating
- (β, σ_U^2 and distribution function of f_i).

Two possible methods:

1. Assume $g(f \mid X, \eta)$ is a known finite parameter distribution (function of η) and estimate $(\beta, \sigma_U^2, \eta)$ (maybe f too).
2. Nonparametric estimation (e.g., Heckman-Singer). Then estimate $\beta, \sigma_U^2, dg(f)$.

Mundlak Point:

- The within estimator is the GLS estimator in all cases if

$$f_i = \alpha \bar{X}_i + W_i.$$

- The more general point is that if we permit fixed effects to be functions of exogenous variables, the between and within estimators will in general differ.
- Lee (as cited in Judge, et al.) shows how special the Mundlak point is.

- Suppose $f_i = \alpha Z_i + W_i$

$$W_i \sim \mathcal{N}(0, \sigma_f^2).$$

If $Z_i = X_{i\cdot}$

$$\beta_{\text{Marginal}} = \beta_{\text{Con}} = \beta_{\text{Int.}} = \beta_{\text{Within}} = \hat{\beta}_{\text{MLE}}$$

in a regression setting.

- Mundlak's point is this: Suppose that

$$f_i = \varphi X_{i\cdot} + V_i \text{ (then } V_i \text{ is ind of } U_{it}\text{)}$$

$$Y_i = X_i \beta + \varphi X_{i\cdot} + \iota V_i + U_i.$$

THE UNIVERSITY OF CHICAGO

- Now what is the random effect estimator?

$$Y_i = (X_i - iX_{i\cdot})\beta + iX_{i\cdot}(\varphi + \beta) + \iota V_i + U_i$$

intuitively: you get info only on β from within.

- Apply GLS

$$A^* = \left[I - c\frac{\iota\iota'}{T} \right] = \tilde{F}$$

where $c = \left[1 - \sqrt{\frac{1-\rho}{1-\rho+\rho T}} \right]$ (refer to Section 3.2 of Part I).

- Thus, we get the GLS transformation as:

$$\tilde{F} Y_i = \tilde{F}(X_i - iX_{i\cdot})\beta + \tilde{F} X_{i\cdot}(\varphi + \beta) + \tilde{F}(\iota V_i + U_i)$$

- In general,

$$\left[I - c\frac{\iota\iota'}{T} \right] X_{i\cdot} = X_{i\cdot}(1 - c).$$